TOPOLOGICALLY-ENHANCED DEEP LEARNING FOR SPATIAL IMAGE PROCESSING AND GENERATION

A THESIS
SUBMITTED TO THE FACULTY OF THE GRADUATE SCHOOL
OF THE UNIVERSITY OF MINNESOTA
BY

Mohsen Ahmadkhani

IN PARTIAL FULFILLMENT OF THE REQUIREMENTS
FOR THE DEGREE OF
DOCTOR OF PHILOSOPHY

Dr. Eric Shook

February, 2025

Acknowledgements

Completing this dissertation represents not only years of research and dedication but also the culmination of support, guidance, and encouragement from numerous individuals who have been instrumental in my academic journey. First and foremost, I would like to express my profound gratitude to my advisor, Dr. Eric Shook, whose unwavering support, insightful guidance, and endless patience have been invaluable throughout my doctoral studies. His mentorship has not only shaped my research but has also profoundly influenced my growth as a researcher. I am deeply grateful to my dissertation committee members, Dr. Ying Song, Dr. Di Zhu, and Dr. David Haynes, for their constructive feedback, expertise, and time invested in reviewing my work. Their diverse perspectives and thoughtful suggestions have significantly enhanced the quality of this research and broadened my academic horizons. I gratefully acknowledge the Minnesota Supercomputing Institute (MSI) and South Dakota State University Research Cyberinfrastructure (RCi) for providing the high-performance computing resources essential for this dissertation research. I am particularly thankful to Dr. Daniel Griffin, Fan Wang, Pegah Salehi, Sajad Amouei, and Xiaoling Hu for their valuable insights and enlightening discussions that helped shape various aspects of this research. My deepest gratitude goes to my family for their unconditional love and endless encouragement. To my parents, who instilled in me the value of education and perseverance, and to my wife, who has been my steadfast supporter throughout this journey - your patience, understanding, and unwavering support have been my anchor in this challenging but rewarding pursuit. Finally, I extend my appreciation to all those who have contributed to this work in ways both large and small. While not all names can be mentioned here, your support has been crucial to the completion of this dissertation and will always be remembered with gratitude.

Dedication

*To my beloved wife, **Farnoosh Dadashi**, for walking beside me on this journey with endless love and understanding,*

and

*To my father, **Ali Ahmadkhani**, who taught me the value of perseverance and the courage to pursue my dreams.*

Abstract

This dissertation explores the integration of topological awareness into deep learning frameworks for spatial image segmentation and generation. Addressing the critical need for topological consistency, the research develops novel methods to enhance the accuracy and reliability of segmentation outputs in geospatial and dendrological contexts. Key contributions include the introduction of TopoSinGAN, a topology-aware generative adversarial network, and TopoSegNet, a scalable segmentation model that incorporates topology-preserving loss functions. These models are evaluated across diverse datasets, including ultra-high-resolution tree-ring images and agricultural field boundaries. Building on the individual contributions of TopoSinGAN and TopoSegNet, the dissertation implements a comprehensive workflow that combines these models. TopoSinGAN is used to generate topologically consistent synthetic datasets, which are then employed to improve the training and accuracy of TopoSegNet. This integrated approach demonstrates significant advancements in segmentation performance, emphasizing the synergy between synthetic data generation and topological enhancement. The dissertation's findings establish a robust foundation for applying topological principles in deep learning, with implications for geospatial analysis, dendrochronology, and other domains requiring precise spatial delineation. By leveraging topology-aware techniques, this work advances the state-of-the-art in deep learning for image analysis, ensuring both geometric and structural fidelity in challenging applications.

Contents

List of Tables

List of Figures

Chapter 1

Introduction

1.1 Problem Statement and Motivation

The increasing sophistication of digital imaging technologies, from satellite-based remote sensing to ultra-high-resolution microscopic imaging systems, has revolutionized our ability to observe and analyze complex spatial phenomena across diverse scales [8, 9]. These advances have driven rapid progress in computer vision tasks, particularly in image generation - the task of creating new synthetic images from text descriptions or other inputs [10], and image segmentation - the process of partitioning an image into multiple segments by assigning a label to each pixel [11, 12]. However, despite impressive achievements in both domainswith image generation models creating visually appealing synthetic images, and segmentation models achieving high pixel-level accuracy (the percentage of correctly classified pixels compared to ground truth), current deep learning methods still struggle with fundamental errors. In image generation, models often produce outputs with physically impossible structures or inconsistent spatial relationships[13]. Similarly, in image segmentation, while models can achieve high accuracy in classifying individual pixels into their respective categories, they frequently fail to maintain the continuity and connectivity of important features such as linear boundaries or loops[14, 7]. These inaccuracies in both domains are often referred to as a lack of "topological accuracy."

The concept of topological accuracy varies depending on the field of study. In image generation, where new images are created rather than processing existing ones, topological accuracy is primarily assessed through connectivity, specifically, whether features that should form continuous structures (e.g., field boundaries) maintain proper connections in the generated output, rather than appearing fragmented or disconnected [15, 16]. This slightly differs from semantic segmentation in image processing, where topological accuracy typically means maintaining the topological structure between original images and their corresponding segmentation predictions [17, 18]. Despite these different definitions, both approaches can leverage sophisticated mathematical tools for quantifying topological errors, including topological data analysis (TDA) (detailed in Section 4.3.4) and persistent homology (detailed in Section 2.1.4). These methods provide rigorous frameworks for measuring and comparing topological features (e.g. loops and connected components) across different scales, enabling precise evaluation of topological accuracy in both generated and segmented images [15, 16, 17, 18].

Topological accuracy differs from pixel-level accuracy in image segmentation, which measures how correctly individual pixels are classified without considering their relationships to neighboring pixels [19, 20]. In the case of image generation, topological accuracy is also different than the common image generation evaluation techniques such as FID (Section 3.3.2) which compare visual patterns such as textures, shapes, and colors between generated and real images [21]. While FID captures overall visual similarity and image quality by measuring how close the generated images' characteristics are to those of real images, it does not specifically assess whether topological properties are preserved in the generated outputs.

Therefore, topological accuracy persists as a critical challenge in both image generation and segmentation fields [22, 23, 7, 19, 24]. This limitation is particularly evident in practical applications across various domains. For instance, in synthetic image generation, even state-of-the-art models can produce structurally inconsistent results, as demonstrated by the AI-generated image produced by DALL-E 2, OpenAI [1] in response to the prompt in ChatGPT, OpenAI, Version GPT-4 as "generate an image of a bicycle wheel with no dangling spoke" in November 2024 (Figure 1.1). In Figure 1.1, the spokes highlighted in red appear disconnected which is a physical impossibility that violates the basic topological requirements of the object. Similarly, in delineating linear boundaries of objects of interest from remote sensing images (Figure 1.2), current segmentation models often produce fragmented or disconnected boundaries, signifying topological errors in the field delineation task [12].

Figure 1.1: An example of topological errors in an image generated by DALL-E 2 [1], based on a textual description of "generate an image of a bicycle wheel with no dangling spoke" in ChatGPT, OpenAI, Version GPT-4 in November 2024.

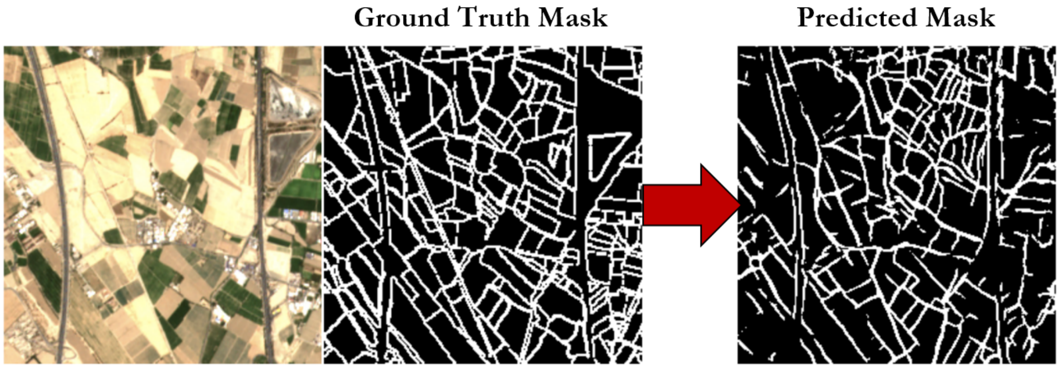


Figure 1.2: An example of topological errors in image segmentation.

Topological accuracy in image segmentation is essential across various domains such as remote sensing [25], urban planning [26], medical imaging [11], and neuroscience [27].

In remote sensing, topological accuracy is critical for the segmentation of natural and man-made environments. For example, accurately delineating water bodies and land cover is essential for reliable environmental assessments and resource management. Fragmented or imprecise boundaries can lead to substantial errors in applications such as flood modeling or ecosystem monitoring [28, 25, 12]. Studies have extensively demonstrated how topological errors in environmental and geospatial analyses can significantly impact the accuracy of resource management and urban planning decisions [28, 25, 12, 10, 29, 30, 31, 32]. Similarly, the segmentation of road networks in aerial imagery is heavily dependent on topological accuracy, as errors can disrupt the connectivity and coherence of mapped networks, adversely affecting urban planning and infrastructure development [33, 34, 25, 26, 35, 29].

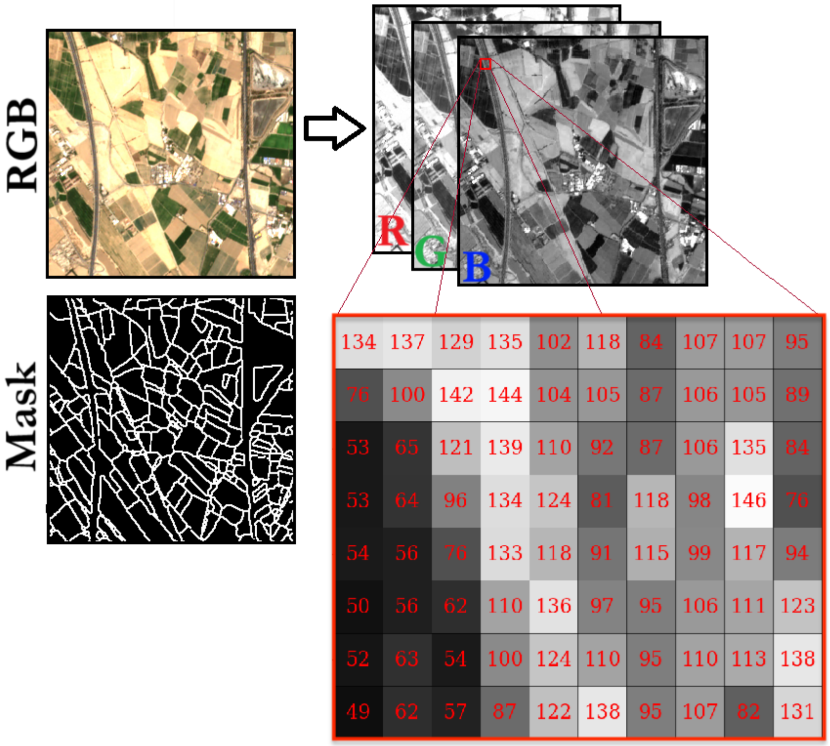
In urban planning, the role of topological accuracy extends to understanding structural layouts and inter-object relationships. Mapping transportation networks, delineating building footprints, and modeling urban environments rely on maintaining continuity and connectivity. Errors in topological representation can result in misinterpretations that impact zoning, construction, and city planning initiatives. Studies emphasize the need for topology-preserving segmentation in these contexts to prevent misrepresentation of critical urban infrastructure [19, 36].

In medical imaging, the stakes for topological accuracy are perhaps the highest. Accurate segmentation of anatomical structures such as blood vessels, organs, and neural pathways is paramount, as these structures' topological integrity directly influences patient diagnoses and treatment outcomes. For instance, topology-preserving approaches in MRI segmentation help ensure the continuity of brain structures, a necessity for neurosurgical planning and neurological assessments [37, 24, 38, 39, 20, 40, 41, 42, 43, 44]. Similarly, accurate delineation of skin lesions or lung abnormalities (i.e., precise identification and marking of boundaries between different tissue types) enables differentiation between benign and malignant conditions, which is critical for effective clinical decision-making [36]. In neuroscience, the segmentation of fine-scale neuronal connections from electron microscopy images further highlights the importance of topological accuracy. Neural connectivity maps rely on preserving the structure and continuity of segmented features, as inaccuracies can obscure critical pathways essential for understanding brain function and disease pathology. Studies in this domain underscore the necessity of topology-preserving segmentation to maintain the accuracy and interpretability of neural network models [27, 42].

Existing models commonly incorporate topological awareness through persistence homology (PH) - a mathematical method that tracks how topological features evolve as a shape or dataset is analyzed at different resolutions [45]. For example, when analyzing an image at different thresholds, PH measures how long features such as holes or connected regions persist - features that quickly disappear might be noise, while those that persist across many thresholds are likely to be meaningful structural characteristics [45]. PH is explored further in Sections 2.1.4 and 4.3.4. PH-based methods face two significant limitations. First, these methods primarily focus on identifying and analyzing loops (i.e., closed regions or polygons) as the key topological feature [45]. While loops are important topological elements, this focus on computing and tracking them makes these methods computationally intensive and difficult to apply to large-scale datasets [7, 37]. Second, their emphasis on loops means they often overlook other important topological features, particularly terminal nodes or endpoints of lines that do not form complete loops (hanging lines) in binary masks. These hanging lines are crucial indicators of topological errors in many real-world applications, such as detecting disconnected road segments [46] or broken field boundaries [30]. This dissertation addresses these limitations by introducing novel approaches to integrate scalable topological constraints into deep learning frameworks for both image generation and segmentation tasks, with particular attention to capturing a broader range of topological features beyond just loops.

1.2 Digital Image Representation and Spatial Topology

Digital images are represented as discrete two-dimensional arrays of pixels, where each pixel contains numerical values representing color or intensity information [47]. For grayscale images, each pixel typically holds a single value ranging from 0 (black) to 255 (white), while color images often use three channels (Red, Green, Blue) to represent the full-color spectrum. In image segmentation tasks, binary masks are commonly used to represent the presence or absence of specific features, where pixels are assigned values of 1 to indicate the presence of a feature and 0 for the background [48]. For instance, in agricultural field boundary detection, a binary mask would have white pixels (value 1) marking the field boundaries and black pixels (value 0) for non-boundary regions. This mask effectively transforms the complex visual information of the field edges in the original image into a simplified binary representation that highlights the topological structure of the boundaries. This discrete representation transforms continuous real-world scenes into a finite grid of numerical values that computers can process and analyze [49]. Figure 1.3 illustrates the representation of a color image of agricultural fields along with the mask channel in a two-dimensional space.



134	137	129	135	102	118	84	107	107	95
76	100	142	144	104	105	87	106	105	89
53	65	121	139	110	92	87	106	135	84
53	64	96	134	124	81	118	98	146	76
54	56	76	133	118	91	115	99	117	94
50	56	62	110	136	97	95	106	111	123
52	63	54	100	124	110	95	110	113	138
49	62	57	87	122	138	95	107	82	131

Figure 1.3: Digital image representation in two-dimensional space.

In the context of spatial imaging, this representation takes on additional significance. By spatial images, we refer to any image where the location of objects and their spatial relationships can be described using a reference coordinate system, regardless of scale or subject matter [48, 47]. This encompasses everything from microscopic images of cellular structures to medical imaging and conventional Earth-based remote sensing data. In remote sensing applications, the spatial resolution determines the ground sampling distance (GSD), which represents the real-world size of each pixel [49]. For instance, a 10-meter resolution satellite image means that each pixel represents a 10x10 m^2 area on the ground. Similar principles apply to any spatial imaging context, though the scale and reference systems vary based on the subject being observed [48]. This relationship between pixel space and physical space is maintained through appropriate spatial reference systems, whether georeferencing for Earth observation imagery or local coordinate systems for microscopic imaging, which enables the analysis of coherent features such as agricultural fields, urban areas, or xylem cells in microscopic images of tree rings [48].

Topology in the two-dimensional digital space differs from classical mathematical topology in a key aspect. While mathematical topology deals with continuous spaces and transformations, digital topology operates on discrete pixel grids where connectivity and neighborhood relationships become fundamental [50]. In digital images, topological properties are typically defined through pixel adjacency and connectivity rules [51, 50]. The most common connectivity patterns are 4-connectivity (considering only orthogonal neighbors) and 8-connectivity (including diagonal neighbors), which determine how features are connected and how boundaries are defined [51]. The topological characteristics of features in spatial images include [51]:

1. **Connectivity:** How pixels belonging to the same feature are connected, forming continuous regions or networks.
2. **Holes and Enclosures:** The presence of complete boundaries that separate internal regions from their surroundings.
3. **Adjacency Relationships:** How different features share boundaries or interact spatially.

These topological characteristics are particularly important in remote sensing applications where the accurate representation of spatial relationships is crucial [25]. For instance, in agricultural field mapping, maintaining the connectivity of field boundaries and preserving their relationships is essential for accurate area calculations and crop monitoring [28, 25]. Similarly, in urban feature extraction, the proper connectivity of road networks and building footprints is vital for transportation planning and urban development analysis [26].

The preservation of topological properties in digital image processing presents unique challenges due to the discrete nature of pixel-based representation [51]. While continuous topological transformations in mathematical topology preserve properties such as connectivity and the number of holes, discrete operations on pixel grids must carefully maintain these properties through specific algorithms and constraints. This becomes particularly challenging in tasks such as image segmentation and feature extraction, where the goal is to identify and delineate continuous features from discrete pixel data while preserving their topological characteristics [51].

1.3 Topology and Image Analysis

Topology is a branch of mathematics that studies properties of geometric objects that remain unchanged under continuous deformations such as stretching, bending, and twisting, but not tearing or gluing [52]. This concept is elegantly illustrated in Figure 1.4, which shows the topological equivalence between a coffee mug and a doughnut (torus) [3]. The sequence demonstrates the gradual deformation (function f) of a coffee mug into a doughnut shape, and its reverse transformation (function g) back to a mug. From a topological perspective, both objects are equivalent because they each have exactly one hole - the handle of the mug corresponds to the hole in the doughnut. This property, known as the genus of the surface, remains invariant throughout the transformation [3]. This famous "mug-to-doughnut" transformation demonstrates one of topology's fundamental concepts: it is not the exact shape (geometry) that matters, but rather the fundamental structural properties that persist under continuous deformation [52, 50, 3].

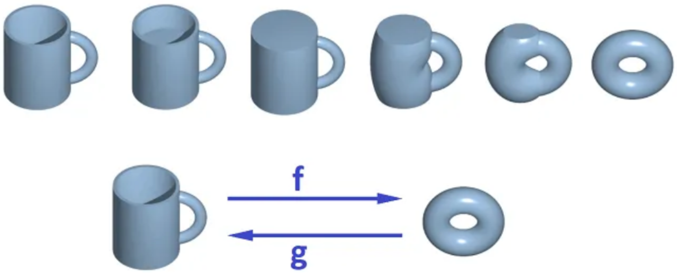


Figure 1.4: Persistence of topological features under continuous deformation showing topological equivalence between a coffee mug and a torus [2, 3].

In the context of image analysis, topology provides a framework for understanding and preserving the structural relationships between different parts of an image, such as connectivity, holes, and boundaries [53, 54]. These topological features are essential for maintaining the integrity of segmented structures and generated images, particularly in applications where accurate representation of spatial relationships is crucial [45].

Topological accuracy in image generation and segmentation

Topological accuracy varies in definition depending on the field, such as graph theory or computer science. In models for image semantic segmentation, it typically denotes the maintenance of topological relationships between the original image and its segmented output [17, 18, 55]. Figure 1.2 illustrates this concept in the context of agricultural field boundary segmentation, where a satellite image is processed to extract field boundaries. The comparison between the ground truth mask and predicted mask demonstrates how segmentation models must preserve the topological structure of field boundaries, maintaining proper connectivity and avoiding breaks or discontinuities in the boundary lines.

On the other hand, in the realm of image generation where there is no actual image to compare with each synthetic output, topological accuracy primarily concerns connectivity. This means it is crucial to ensure that elements expected to be continuous, such as roads or borders of agricultural fields, remain uninterrupted in the synthesized data [15, 16, 56]. In the ChatGPT example shown in Figure 1.1, while the overall visual appearance of the wheel is realistic, several spokes (highlighted in red circles) appear disconnected or dangling.

These topological errors in the spoke connections violate the structural integrity expected in a real bicycle wheel, where all spokes should form continuous connections between the hub and rim, demonstrating how even visually appealing generated images can contain fundamental structural inconsistencies. The importance of topology in image generation and segmentation tasks can be understood from several perspectives:

1. **Structural Preservation:** Topological features help maintain the structural integrity of generated or segmented images, ensuring that critical relationships between different parts of the image are preserved [57, 4].
2. **Scale Invariance:** Topological properties remain consistent across different scales, making them particularly valuable for analyzing complex structures in medical imaging, remote sensing, and other domains where features may appear at multiple scales [24, 38].
3. **Robustness to Noise:** Topological methods can distinguish between significant structural features and noise, helping to produce more reliable segmentation results and higher-quality synthetic images [43, 58].
4. **Quality Assessment:** Topological metrics provide meaningful measures for evaluating the quality of generated images and segmentation results, particularly in cases where traditional pixel-based metrics may not capture important structural properties [4, 59].

The integration of topological constraints into deep learning models represents a significant advancement in computer vision, enabling more accurate and structurally consistent results in both image generation and segmentation tasks. This is particularly relevant in applications such as medical imaging, where preserving the connectivity of anatomical structures is crucial, and in remote sensing, where maintaining the integrity of geographic features is essential for accurate analysis [40, 41, 42].

1.4 Background

Image semantic segmentation, also referred to as image segmentation in this proposal, involves separating an image into non-overlapping regions that completely comprise the image, enabling boundary delineation and identification of areas of interest critical for downstream processing [60]. Prior to the advent of deep learning, traditional machine learning methodologies demonstrated commendable accuracy in image segmentation tasks [61, 62, 63, 64]. These methods, however, necessitated extensive feature engineering, which involves the process of selecting, modifying, and creating relevant features from raw data to enhance the performance of machine learning algorithms, making it complex and task-specific [65]. In contrast, deep convolutional neural networks, such as Fully Convolutional Networks (FCN), have the capability to autonomously learn hierarchical features directly from the data, thereby eliminating the need for manual feature engineering [11]. This shift has been propelled by advancements in computational power and the increased availability of large training datasets [11]. Consequently, these deep learning approaches have gained widespread popularity and have been recognized for delivering marked improvements in accuracy across a variety of segmentation tasks, surpassing the performance of previous conventional methods [66, 67].

Advances in computer vision and the pivotal role of image segmentation have paved the way for breakthroughs across various scientific domains in areas as diverse as robotics [68, 69], remote sensing [12, 10, 70], and biomedical imaging [71, 72]. However, accurately partitioning images into distinct segments or classes remains a considerable challenge [73]. Difficulties are especially evident for significantly class-imbalanced data where minority classes, such as boundary pixels in remote sensing, are disproportionately penalized for misalignments [73].

Current segmentation models are often trained using loss functions that compare ground truth and prediction images via metrics such as Dice score and Intersection over Union (IOU). These may inadequately capture the topological properties of image features. A slight shift for a linear feature gets fully counted as false positives, significantly reducing accuracy. Incorporating topological relationships into loss functions could alleviate limitations for imbalanced data. This represents an area requiring further investigation [73].

Topology is a field in mathematics focused on preserving geometric relationships between objects under continuous deformations, rather than specific metric distances or dimensions [53]. Independent of precise metrics, topology shares conceptual links with spatial descriptions in language and geography that rarely reference exact magnitudes or directions [54]. In GIS, topology refers to the spatial relationships and rules that govern how points, lines, and polygons share coincident geometry within geospatial data [53]. While geometry deals with continuous coordinate-based representations, topological approaches in GIS facilitate efficient modeling and analysis of spatial entities based on their connections and adjacencies rather than precise size or shape [74].

Incorporating topological constraints and relationships into image segmentation frameworks could enable more robust learning for tasks such as delineating linear features by preserving the notion of adjacency instead of strictly penalizing minor localization errors. This provides scope for significant improvements in accuracy, especially for minority imbalanced classes [75, 7, 73, 4]. However, the challenges of intricate structures, whether it be the delicate weave of human cell membranes, xylem cells, or the intricate sprawl of urban roadways and interwoven structure of agricultural fields in the Middle East, expose the limitations of contemporary methods [76, 77]. This dissertation delves into the intricate balance between geometric precision (how accurately the predicted shapes and their size and area align with the true object [78]) and topological accuracy of image segmentation, aiming to enhance the overall efficacy of deep learning models. Building upon this foundation, the following research questions guide the investigations in this proposal:

1.5 Research Questions

1. How can topological constraints be integrated into deep learning models to enhance spatial image generation and segmentation?
2. To what extent can a topologically enhanced deep learning model impact spatial and topological attributes when generating synthetic images of irregularly shaped agricultural fields in Iran?
3. To what extent can a topologically enhanced deep learning model impact spatial and topological attributes when delineating vessels in ultra-high-resolution tree-ring dendrology images?

1.6 Significance and Contributions

This research addresses critical gaps in image generation and segmentation tasks by integrating topological constraints to improve the spatial coherence and structural integrity of outputs. While deep learning models have significantly advanced segmentation and generation accuracy, they often fail to preserve topological properties that are essential for practical applications [15, 16, 17, 18]. This limitation impacts a wide range of fields, including environmental monitoring, urban planning, medical imaging, and dendrochronology, where fragmented boundaries and misaligned features can undermine analysis and decision-making [28, 25, 12]. By incorporating topology-preserving methods, the research aims to address topological accuracy in tasks with complex structures in image generation and semantic segmentation tasks. The expected contributions of this work include the development of novel methodologies that embed topological constraints into deep learning models, striking a balance between geometric and topological accuracy. These methods will be evaluated through a comprehensive framework that measures both segmentation accuracy and the preservation of topological properties, offering a holistic view of model performance. The tools and techniques developed will be applicable across various domains, facilitating advancements in tasks such as remote sensing, medical imaging, and climate studies through dendrochronology. By demonstrating the broad applicability and transformative potential of topology-enhanced segmentation, this research will not only advance the state of the art but also provide practical solutions to longstanding challenges across interdisciplinary fields.

Research question 1, which addresses the incorporation of topological constraints, comprises two main components in image generation and segmentation domains (Figure 1.5, Section 1.5). The first component of research question 1, concerning topological constraints in image generation models, is addressed in Chapter 3 through the development of TopoSinGAN, a novel topology-aware generative architecture. Research question 2, investigating agricultural field synthesis through topology-aware modeling, is also resolved in Chapter 3 through experimental validation of TopoSinGAN.

The second component of research question 1, focusing on the implementation of topological constraints in image segmentation models, is addressed in Chapter 4 with the development of TopoSegNet, a scalable topology-aware segmentation framework. Research question 3, examining the segmentation of ultra-high-resolution tree-ring dendrology images, is also addressed in Chapter 4 through the implementation of TopoSegNet.

Research questions 2 and 3 are further investigated in Chapter 5 through the integration of TopoSinGAN and TopoSegNet methodologies. This synthesis enabled the generation of synthetic training data to augment datasets for both agricultural field and dendrology experiments. The augmented datasets facilitated additional validation of findings through comprehensive experimentation, thereby strengthening the conclusions for research questions 2 and 3.

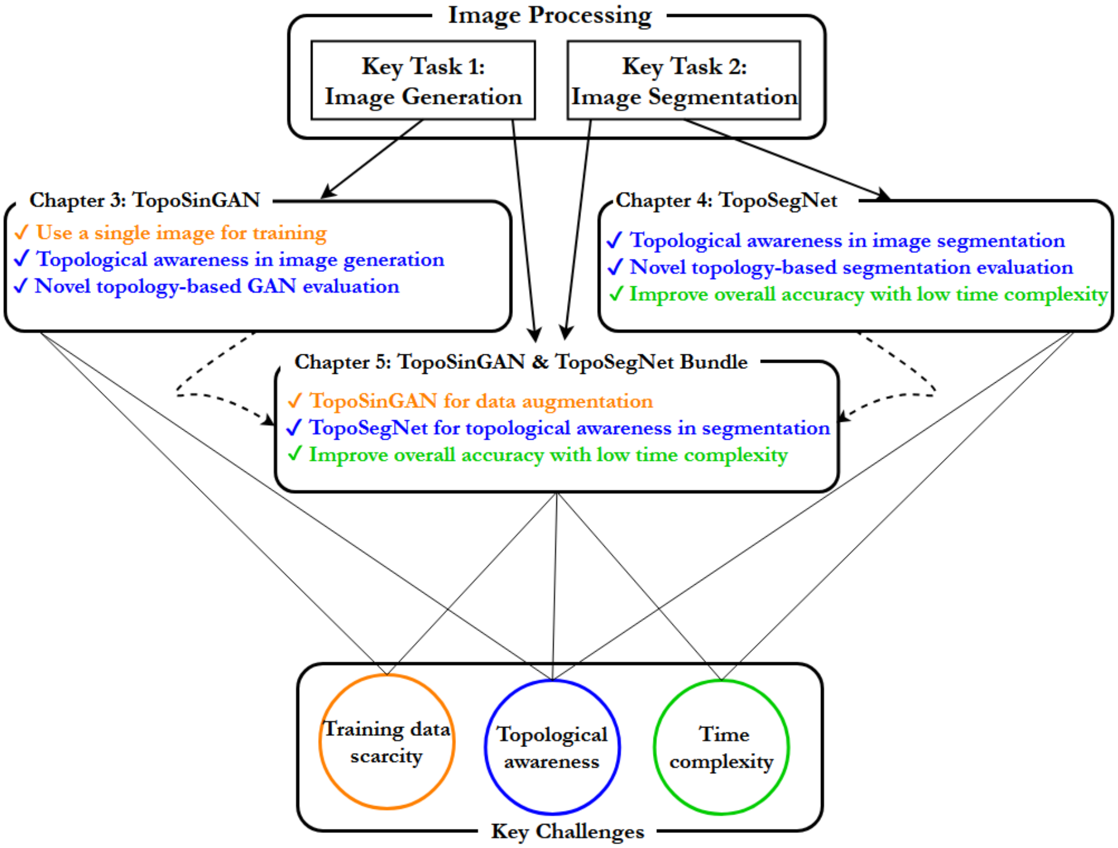


Figure 1.5: An overview of the dissertation structure.

Chapter 2

Literature Review

2.1 Background & Literature Review

2.1.1 Integration of Image Semantic Segmentation and Generation

Image semantic segmentation and synthetic image generation stand as distinct yet interrelated fields of study that underpin the practical applications of computer vision. They encompass the techniques and methodologies necessary for the processing, analysis, and synthesis of visual data (i.e., images) [79]. While distinct in objectives, these fields have notable interplay. Segmentation models, being data-driven, rely extensively on dataset availability and variability [80]. Synthetic data can thus boost segmentation accuracy by providing abundant additional fake training images [80]. Further, synthetic images, being perfectly labeled, allow for more control over data variability and domain generalization with reduced labeling effort unlike real-world datasets [81]. However, synthetic images often lack topological consistencies, as evidenced by discontinuities along generated region boundaries, which can limit their utility [4]. Enhancing topology preservation within synthetic data generation itself can alleviate this limitation [7, 26, 4].

Additionally, segmentation frameworks face difficulties fully separating fine-grained spatial boundaries and patterns, especially in remotely sensed imagery covering vast regions [75, 7, 26]. Incorporating topological constraints directly within segmentation models, ensuring contiguity along detected edges, can improve boundary delineation without requiring greater training data [75, 7].

2.1.2 Image Semantic Segmentation

The field of semantic segmentation has greatly advanced with CNNs surpassing older methods such as random forests and unsupervised algorithms in classifying image pixels. The FCN [67] and the U-Net [6] have been particularly transformative. FCNs allowed end-to-end training with significant performance boosts, while U-Net's dual-path structure became crucial in biomedical imaging, merging context learning with spatial detail preservation. Subsequent models such as ResNet [82], SegNet [66], DeepLab [83] further evolved the landscape of semantic segmentation, setting the stage for sophisticated applications such as the precise delineation of agricultural fields from satellite imagery.

Deep learning has significantly advanced the delineation of agricultural fields from satellite imagery. A fully convolutional network was utilized for the upscaling of field boundaries from Sentinel 2 imagery, enhancing the resolution from 10 meters to 5 meters in the Dutch landscapes [84]. A U-Net architecture was successfully implemented for the separate demarcation of field areas and boundaries [85]. The cutting-edge field boundary detection largely incorporates variations of U-Net integrated with advanced features evidenced by the ResUNet-a used in [30], application of ResU-Net in [86], the R2U-Net utilized by [87], and the DSCUnet developed by [88]. Further advancements in the field have seen the integration of dilated or atrous convolutions, as explored by [89] and [90] to gather multi-scale contextual information without compromising resolution. Attention mechanisms, which model spatial dependencies and feature map channel interdependencies, have also become critical. Models such as EncNet [91] and DANet [92] showcase the effectiveness of attention in scene understanding. Concurrent developments include employing graph neural networks for relational reasoning [93], multiple loss functions for better supervision [94], and the fusion of multi-scale features [95]. Despite these significant advancements, challenges remain in the field of semantic segmentation. These include discontinuities in the boundaries generated by these networks [96]. Additionally, Transformer-based models including the Vision Transformer (ViT) have revolutionized the approach to image processing. ViT, employing Multi-head Self-Attention (MSA) blocks, marks a departure from traditional CNNs but struggles in capturing high-frequency image components [97, 98]. This limitation arises from ViT's method of processing images, which is less effective than CNNs in extracting local features. Integrated models combining CNN and ViT features, such as TransUNet [99] and HiFormer [100], have been proposed to address this. However, these models face their own limitations, including computational inefficiency and inadequacies in multi-scale information processing, often leading to less effective network learning in applications such as remotely sensed image segmentation [97, 101, 98].

2.1.3 Generative Adversarial Networks

Generative Adversarial Networks (GANs), introduced by [102], have significantly impacted image synthesis, employing two convolutional networks, the generator and the discriminator, in an adversarial process to produce fake images nearly indistinguishable from the real ones. This capability has been leveraged for tasks such as image inpainting [103], image-to-image translation [104], and generating synthetic training images to mitigate the shortage of extensive datasets, particularly in image semantic segmentation [105].

Subsequent studies have incorporated GANs to enhance training for segmentation models. For instance, synthetic medical images were generated using GANs for transfer learning-based segmentation [106]. In another study, synthetic sky scenes for statistical ground-truth mapping were produced using GANs [107]. In a study by [80] Segmentation detail was significantly improved with a GAN framework, surpassing previous methods. GANs with transfer learning and data augmentation have been employed for segmenting microscopy images of iRPE[1] stem cells, illustrating the diverse applications of GANs in this domain [108]. Recently, a single-image-based synthetic data generator, SinGAN-Seg, has been proposed to effectively produce medical image datasets, demonstrating its utility in improving segmentation models where data is limited or sensitive [109]. Challenges in synthesizing remote sensing images, which demand high visual fidelity and diversity due to semantic similarity and data imbalance, have been addressed through a novel method, CSEBGAN, which enhances the synthesis of remote sensing images by decoupling semantic classes and employing a tripartite adversarial training scheme, proving effective in data augmentation for image segmentation tasks [110].

---

[1] Human induced pluripotent Retinal Pigment Epithelial

2.1.4 Topological Integrity

Topological discrepancies in synthetically augmented training sets can compromise segmentation model performance. This is particularly critical in fields requiring detailed and accurate structural mappings, such as cartographic representation, intricate medical imaging, and precision in robotic pathfinding. The high-stakes nature of these areas, from the fine-tuned requirements of robotic movements to the faithful representation of anatomical structures in medical imaging, underscores the need for flawless topological consistency [4].

Persistent Homology

Persistent Homology (PH) techniques are esteemed for their precision in preserving topological features of data, capturing essential attributes such as shape and connectivity. These characteristics are crucial for the analysis of complex data structures. The methodology systematically identifies consistent topological elements throughout various spatial resolutions [45]. Studies, including the work by [7], have leveraged PH to refine features such as loopiness in generated imagery. Yet, a notable drawback of PH is its computational intensity. The approach necessitates multifaceted analysis across numerous scales and dimensions, which can lead to extensive processing times. In the worst-case scenarios, the computational complexity of PH algorithms may escalate to a cubic proportion relative to the dataset's size, symbolized as $O(n^3)$, where n represents the quantity of data points or simplices being processed. Consequently, as the size of the dataset grows, the time required for computation increases dramatically, potentially making PH impractical for very large datasets [75, 4].

Other Topology Enhancement Endeavors

A layered approach to feature extraction was adopted, leveraging the strengths of a ResNet50 architecture [26]. This system is uniquely tailored with a directional connectivity label transformation that aims to fortify the integrity of topological representation. While this approach marks a significant advancement, it also highlights a persistent gap in achieving the dual goals of accurate segmentation and robust topological consistency.

Separately, an innovative approach combining deep learning with heuristic search strategies has been taken, exemplified by the incorporation of the A^* algorithm to refine topological anomalies [35]. This hybrid technique suggests a new direction for enhancing topological accuracy through the synergy of AI and classical algorithmic approaches. Further, Guo and Zhou (2023) have underscored the importance of data augmentation in addressing topological challenges, such as the continuity of road networks obscured by occlusions [56]. Their findings advocate for the adoption of data augmentation as a strategy to strengthen the topological accuracy of segmentation models, particularly in environments constrained by incomplete or obscured datasets.

2.2 Case Studies

This dissertation focuses on two primary case studies that address research questions 2 and 3: agricultural fields in Tehran Province, Iran, and gigapixel dendrological tree rings (GMP) [111]. The agricultural fields dataset, addressing research question 2, presents challenges in segmenting irregularly shaped field boundaries from satellite imagery in Iran. The GMP dataset, addressing research question 3, provides ultra-high-resolution tree-ring imagery to evaluate the effectiveness of our topology-aware approach in delineating complex vessel structures in dendrology. These diverse case studies demonstrate the broad applicability of our topologically-enhanced methods (further details in Section 4.6.1). In Chapter 4, the evaluation is extended using CREMI-B [112] and DRIVE [113], well-established benchmark datasets, to further assess segmentation performance on varied and complex tasks, supporting the findings from both research questions through additional validation.

2.3 Agricultural Fields

This case study utilizes a shapefile of agricultural field boundaries derived from precise land surveys conducted with Total Station equipment in Tehran Province, Iran, situated in the Middle East's diverse landscape. The study area spans a geographic extent from $50°25'10''$ W to $52°27'5''$ E and $35°43'38''$ N to $35°0'12''$ S, encompassing approximately 19,500 km^2. For the RGB imagery, 10-meter resolution remote sensing data from Sentinel-2 was employed. The segmentation task in this region poses significant challenges due to the irregular geometries and distinct topological layouts of the agricultural fields. These fields feature heterogeneous patterns and thicker, uniformly distributed line features compared to other datasets, adding to the complexity of satellite image segmentation. The dataset consists of 447 images, each measuring 250 × 250 pixels with four channels, where the fourth channel is a rasterized mask layer representing the boundaries of the agricultural plots. This case study provides a robust testbed for evaluating the topological accuracy of segmentation models in real-world agricultural applications.

2.4 Tree Ring Gigapixel Macro Photography (GMP)

The second case study involves a gigapixel macro photography (GMP) dataset, a subset of the dataset collected by [111] which includes 1,000 high-resolution images measuring 175 × 240 pixels each. Captured at an ultra-high resolution of 19,812 dpi, the dataset was manually cropped and annotated to include four channels: RGB and an annotated mask layer that highlights xylem vessels in hardwood samples. With approximately 750 vessels per 1 mm^2, the complexity and density of these features make manual annotation arduous, underscoring the need for automated deep learning solutions to advance dendrological analysis. The line features in this dataset exhibit significant variation in thickness, presenting additional challenges for segmentation models. By addressing these variations and improving segmentation precision, this case study highlights the potential for deep learning techniques to streamline the analysis of tree-ring data for applications in climate science and forestry.

2.5 Evaluation Metrics

Figure 2.1 and Table 2.1 present an overview of the existing evaluation metrics for image semantic segmentation and generation models utilized in this dissertation. Given the limitations of these metrics in addressing specific use cases and objectives, this dissertation introduces novel evaluation methods that are not included here but are detailed in subsequent chapters.

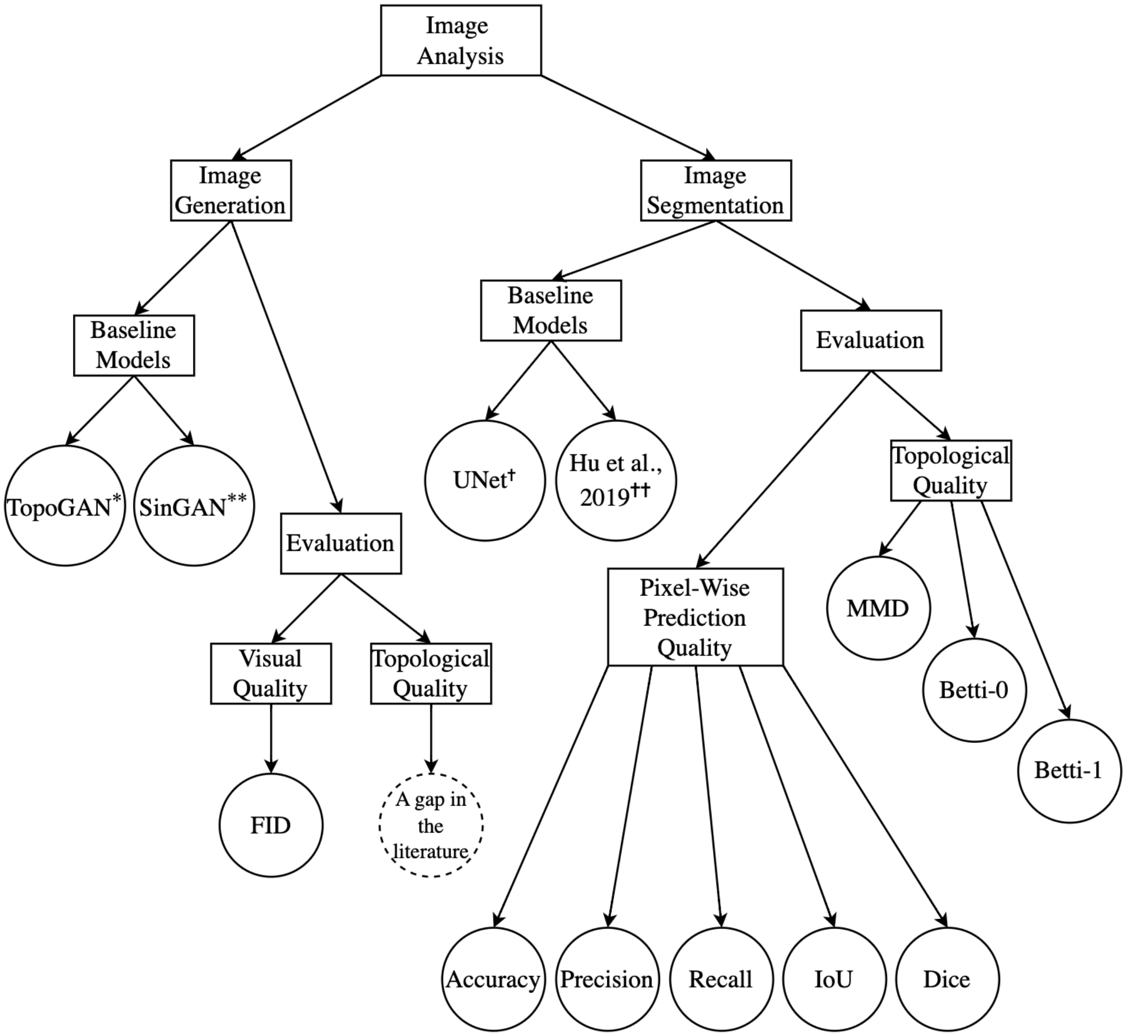


Figure 2.1: An overview of the existing evaluation metrics and baseline models (*: [4], **: [5], †: [6], ††: [7]) used in this research and detailed in Table 2.1. Notably, quality metrics for image generation differ fundamentally from those used in image segmentation tasks, as generated images cannot be evaluated through pixel-wise comparison with ground truth images, which are absent in generative frameworks.

Table 2.1: The existing evaluation metrics used in this dissertation.

Metric	**Definition**	**Use Case**	**Limitations**
Accuracy	Ratio of correctly classified pixels to total pixels.	General performance assessment in balanced datasets.	Misleading for imbalanced datasets; doesn't account for spatial relationships.
Precision	Ratio of true positives to all positive predictions.	Applications where false positives are costly (e.g., medical diagnosis).	Doesn't consider false negatives; sensitive to class imbalance.
Recall	Ratio of true positives to all actual positives.	Applications where missing positive cases is costly (e.g., tumor detection).	Doesn't consider false positives; can be high even with poor precision.
IoU	Area of intersection divided by area of union between predicted and ground truth segmentation.	Object detection and segmentation evaluation; standard in computer vision.	May not capture fine structural details; sensitive to small deviations in boundary.
Dice	$2 \times \text{IoU}/(1+\text{IoU})$.	Medical image segmentation; widely used in healthcare applications.	Similar limitations to IoU; doesn't account for topological correctness.
MMD	Average minimum distance between points in two distributions.	Comparing shape distributions; evaluating generated vs. real data.	Computationally expensive; may miss global structural differences.

Continued on next page

Table 2.1 continued

Metric	Definition	Use Case	Limitations
Betti-0 (β_0)	Number of connected components in segmentation.	Evaluating topological correctness; ensuring connectivity preservation.	Only captures component count; doesn't assess shape or size accuracy; sensitive to noise.
Betti-1 (β_1)	Number of loops in segmentation.	Assessing preservation of circular structures and holes.	Doesn't capture higher-dimensional topological features; sensitive to noise.
FID	Measure of similarity between feature distributions of real and generated images (used for GANs).	Evaluating quality of generated synthetic images (style transfer assessment).	Requires pre-trained network; may not capture domain-specific features.

Chapter 3

TopoSinGAN: Learning a Topology-Aware Generative Model from a Single Image

3.1 Abstract

Generative adversarial networks (GANs) have significantly advanced synthetic image generation, yet ensuring topological coherence remains a challenge. This paper introduces TopoSinGAN, a topology-aware extension of the SinGAN framework, designed to enhance the topological accuracy of generated images. TopoSinGAN incorporates a novel, differentiable topology loss function that minimizes terminal node counts along predicted segmentation boundaries, thereby addressing topological anomalies not captured by traditional losses. We evaluate TopoSinGAN using agricultural and dendrological case studies, demonstrating its capability to maintain boundary continuity and reduce undesired loop openness. A novel evaluation metric, Node Topology Clustering (NTC), is proposed to assess topological attributes independently of geometric variations. TopoSinGAN significantly improves topological accuracy, reducing NTC index values from 15.15 to 3.94 for agriculture and 14.55 to 2.44 for dendrology, compared to the baseline SinGAN. Modified FID evaluations also show improved realism, with lower FID scores: 0.1914 for agricultural fields compared to 0.2485 for SinGAN, and 0.0013 versus 0.0014 for

dendrology. The topology loss enables end-to-end training with direct topological feedback. This new framework advances the generation of topologically accurate synthetic images, with applications in fields requiring precise structural representations, such as geographic information systems (GIS) and medical imaging.

3.2 Introduction

Since its inception by [102], Generative Adversarial Networks (GANs) have revolutionized image generation. At the core of GANs are two neural networks, a generator and a discriminator, which are trained concurrently using a game theory strategy. This innovative structure enables GANs to produce synthetic images that are nearly indistinguishable from real ones. GANs have been applied in various domains, including unconditional and conditional image synthesis [105], image inpainting [103], and image-to-image translation [104]. They have been particularly valuable in generating training images to address the scarcity of comprehensive datasets in image processing [105, 106, 107, 80, 108, 109]. However, dealing with diverse datasets that include multiple object classes, such as ImageNet [114], remains challenging. This often requires training models tailored to specific tasks, such as super resolution [115], inpainting [103], and retargeting [116]. Training GANs typically demands large datasets, which can be resource-intensive. As a result, the focus can shift to training generative models on a limited number of images, or even a single image to capture unique characteristics and address data scarcity [5].

The innovative SinGAN [5] provides a unique approach by training a GAN on a single image, focusing on unconditional image generation and harmonization. SinGAN uses a multitiered and multiresolution method, beginning with training at low resolutions such as 26×35 pixels. As training progresses through various scales, the generator's complexity increases, and the image resolution is enhanced. Each new scale builds on the training from previous scales, with only the newly introduced layers being trained. This multiscale architecture is illustrated in Figure 3.1. Building on this concept, the SinGAN-Seg model [109] was developed to enhance SinGAN by adding a mask channel to the conventional RGB channels. This modification has proven effective, particularly in creating annotated images for medical segmentation tasks.

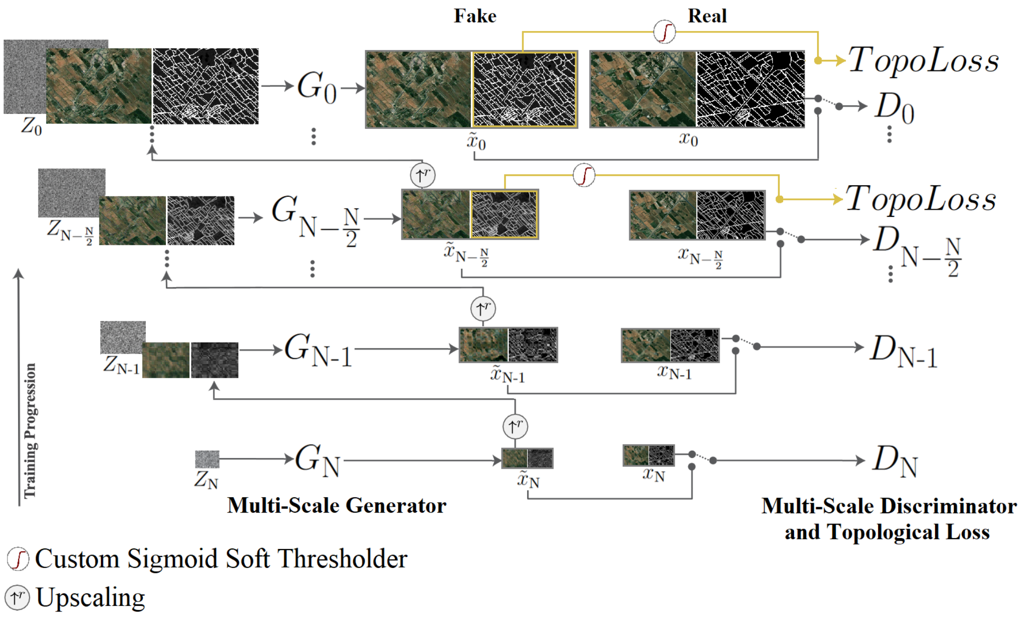


Figure 3.1: ToposInGAN's multiresolution architecture.

Despite their success in generating visually convincing images, these models struggle with ensuring topological integrity in their predictions, which is crucial for applications such as agricultural fields, detailed medical delineations, and robotics. The precise nature of these tasks, whether they involve accurate robotic movements or detailed vascular representations in medical imaging, requires high topological accuracy [76, 77].

To address this issue, we introduce TopoSinGAN. This model builds on the SinGAN framework and incorporates a novel differentiable topology loss function. The core of this loss function is its ability to identify and count terminals within the generated boundary mask, where a terminal is defined as a boundary pixel with only one neighboring boundary pixel. By reducing the number of these terminal nodes, TopoSinGAN aims to improve topological accuracy. We further validate that this approach is differentiable, making it suitable for optimization through backpropagation.

3.2.1 SinGAN

The original SinGAN architecture, as referenced in [5], is notable for its ability to generate synthetic data using a GAN model trained on a single image. The training process in SinGAN employs an image pyramid, which consists of differently scaled versions of the same input image, progressing from low to high resolution. In this hierarchical structure, SinGAN creates a GAN pyramid, where each level corresponds to an image from the image pyramid and undergoes training. SinGAN can input a three-channel RGB image and generate a three-channel RGB output. The use of multiple scales enables the creation of various synthetic versions of the input image across different resolutions. The number of scales is determined by the dimensions of the input image. This approach allows SinGAN to produce multiple synthetic variations from a single real image, illustrated in Figure 3.1 as a 1:N generation, where N represents the number of layers in the GAN pyramid.

Additionally, the intrinsic design and training mechanism of the SinGAN model provide it with versatility, which has been utilized and modified in studies such as SinGAN-Seg [109]. While SinGAN-Seg introduces modifications such as expanding to four channels (RGB plus a segmentation mask), the core architecture and the image pyramid-based training of the original SinGAN remain fundamental to its operation. SinGAN relies on two central loss functions. The adversarial loss L_{adv} measures the differences between patch distributions in the original image and the generated images. It uses a Markovian discriminator, D_N, that classifies patches as real or fake, similar to patch-based GANs. This adversarial method employs the WGAN-GP loss, known for its training stability, and focuses on the entire image, facilitating the learning of boundary conditions [117]. Meanwhile, the reconstruction loss L_{rec} ensures the model can recreate the original image using specific noise maps, with variations based on the scale to emphasize the pixel-wise similarity between the original and generated images. The reconstructed image further aids in determining the noises standard deviation across scales. While L_{adv} and L_{rec} are effective at capturing visual details, they may neglect topological nuances, which can present challenges in applications that require precise topological accuracy.

3.2.2 The Need for Topological Accuracy

Topological accuracy can have different meanings across fields such as graph theory and computer science. In image semantic segmentation models, it often refers to preserving the topological relationships between the input image and the predicted mask [17, 18, 55]. However, in the context of GANs, where there is no real image corresponding to each generated image for direct comparison, topological accuracy is generally defined by connectivity. This involves ensuring that objects expected to be connected, such as roads or agricultural field boundaries, do not display disconnections in the generated datasets [15, 16, 56]. Similarly, in this study, we define topological accuracy as minimizing open loops in the generated images. The case studies presented here, agricultural fields and dendrology data, involve scenarios where the generated masks should consist entirely of closed loops, as both agricultural fields and xylem cells in dendrology images are expected to appear as fully enclosed regions. Thus, our research focuses on reducing these discontinuities and enhancing the topological integrity of the generated masks.

Ensuring topological accuracy is crucial in GAN applications across various domains. While GANs are highly effective at generating visually compelling outputs, they may overlook the intrinsic structure and connectivity of linear features that define the topology. This distinction from geometric representations is critical. In contexts such as road network mapping or medical imaging, maintaining topological accuracy in GAN outputs is essential to prevent the misrepresentation of crucial structures. For instance, an inaccurately depicted road junction in an autonomous driving map or an incorrect representation of a neuronal pathway in a medical image can have serious consequences [59]. Therefore, maintaining topological accuracy in GAN outputs not only preserves the interconnected nature of the data but also enhances the reliability and practical utility of the models in real-world applications.

Attempts to improve the structural coherence of generated images began soon after the introduction of GANs. The work by Arjovsky and colleagues [118], introducing Wasserstein GAN (WGAN), was an early example of this direction. WGAN aimed to improve training stability and address issues such as mode collapse by minimizing the Earth Mover (Wasserstein) distance, offering a more stable optimization process compared to traditional GANs. In the domain of GANs that emphasize geometry, several models have been developed to capture and utilize geometric nuances from images. For example, geometricGAN [119] integrates the expansive margin principle from SVMs to guide both discriminator and generator training. Similarly, Localized GAN (LGAN) [120] utilizes local coordinates to focus on the specific geometry of the data manifold. GAGAN [121] is notable for its application to facial image generation, drawing on established facial geometries. Meanwhile, GcGAN [122] enforces geometry-consistency regulation to preserve the inherent semantics of images, limiting geometric modifications to simple rotations and flips.

Recent advancements further highlight the integration of topological features in GAN models to enhance their application across different fields. The TR-GAN framework, for instance, improves retinal artery/vein (A/V) classification by employing a topology ranking discriminator and triplet loss, which significantly enhances vessel connectivity, outperforming conventional methods on the AV-DRIVE dataset [123]. TW-GAN further advances A/V classification by incorporating both topology and vessel width awareness, markedly improving performance on datasets such as AV-DRIVE, which is crucial for diagnosing conditions such as hypertension and diabetes [124]. Similarly, Liu and colleagues [125] employed a hierarchical architecture to extract and preserve both local and global graph features, leading to more accurate graph reconstruction.

Recently, the application of persistence homology (PH) in learning topology-aware studies has gained attention [4, 7, 75]. TopoGAN has recently been proposed by [4]. TopoGAN is the pioneering GAN model that explicitly learns the topology of real images, such as connectedness and loops, by introducing a novel topological loss function based on PH. Their approach showed improvements in preserving the topological integrity of generated images [4]. In another study, the authors developed PHGAN that enhances image generation by incorporating topological features from PH into a GAN. These topological features capture global structural information, which is combined with convolutional features in the discriminator. The topological features extracted by a PH module transform into vectors that are fed into the network [126].

Although PH is a significant tool in topological data analysis (TDA) for addressing inaccuracies in GANs, its integration presents challenges. PH-based methods focus primarily on features such as connected components and loops, often overlooking other crucial aspects such as terminal nodes or disconnections in the generated masks. An example is a bicycle wheel with a missing spoke compared to one with a dangling spoke. In terms of TDA and PH, these configurations are treated as topologically identical because they present the same connected components and loops. However, from our perspective, the presence of a dangling spoke (i.e., a terminal node) is a critical difference that indicates an issue embedded in the structure. This limitation makes the application of PH in GANs less appropriate for certain cases where disconnections are considered important. In addition, the computational demands of PH can be substantial, especially with large datasets, potentially reaching $O(n^3)$ in complexity [4, 7, 75].

3.3 Materials and Methods

3.3.1 Topology Loss Algorithm

A differentiable topology loss was developed to evaluate the number of terminal nodes present in the generated boundary mask. This loss was integrated into the SinGAN architecture as an auxiliary loss term, L_{topo} for the generator. Equation (3.1) shows all three terms of the loss function. In this equation, λ_1, λ_2, and λ_3 represent the coefficients of the loss terms, which can be adjusted manually:

$$\text{Loss} = \lambda_1 L_{\text{adv}} + \lambda_2 L_{\text{rec}} + \lambda_3 L_{\text{topo}} \tag{3.1}$$

We propose L_{topo} that primarily focuses on minimizing the presence of terminal nodes in the generated mask. The algorithm consists of three components as follows.

I **Soft Thresholding with Sigmoid Function ($\sigma(x)$)**

To detect terminals, the input mask (M) undergoes a custom sigmoid function $\sigma(x)$, resulting in a soft binarization of the tensor. $\sigma(x)$ acts as a soft thresholding mechanism ensuring continuity across the entire domain. The mathematical representation is as follows:

$$\sigma(x) = \frac{1}{1 + e^{\alpha(\beta - x)}} \tag{3.2}$$

$$\lim_{x\to\infty,\alpha>0} \sigma(x) = \begin{cases} 1, & x > \beta \\ 0.5, & x = \beta \\ 0, & x < \beta \end{cases} \tag{3.3}$$

$$\lim_{x\to\infty,\alpha<0} \sigma(x) = \begin{cases} 0, & x > \beta \\ 0.5, & x = \beta \\ 1, & x < \beta \end{cases} \tag{3.4}$$

where β is the threshold point and α is the thresholding slope. The sigmoid function is smooth and differentiable across its entire domain, ensuring that its gradient can be computed. Specifically, the derivative of $\sigma(x)$ with respect to x is given by:

$$\frac{d\sigma(x)}{dx} = \alpha \cdot \sigma(x) \cdot (1 - \sigma(x)) \tag{3.5}$$

II Detection of Terminal Nodes via Convolution

After applying the sigmoid function, the soft-binarized mask $M' = \sigma(M)$ is convolved with a set of eight fixed kernels, $K_1, K_2, \ldots, K_8$, which are depicted in Figure 3.2. These kernels are specifically designed to target the eight main cardinal and intercardinal directions in a 2D space (north, northeast, east, southeast, south, southwest, west, northwest). Empirically, the use of these kernels has proven effective in identifying discontinuities, thereby enhancing the model's ability to evaluate topological performance. This evaluation is integral to the core of the developed loss function, enabling the application of corrective penalties during the training phase for improved topological accuracy.

$$C_i(M') = M' \otimes K_i \tag{3.6}$$

where $\otimes$ denotes the convolution operator. The convolution operation is differentiable, and thus, gradients can be propagated through this layer.

III Loss Computation

The topology loss is computed as:

$$L_{\text{topo}} = \frac{1}{8} \sum_{i=1}^{8} \sum_{(x,y)} C_i(M')_{x,y} \tag{3.7}$$

where $C_i(M')_{x,y}$ represents the value of the convolution output at position (x, y). Since convolution and summation are linear operations, L_{topo} remains differentiable with respect to M'. For each convolution operation, the gradient of the loss L_{topo} with respect to M' is computed using the chain rule:

$$\frac{\partial L_{\text{topo}}}{\partial M'} = \sum_{i=1}^{8} K_i \otimes \frac{\partial L_{\text{topo}}}{\partial C_i(M')} \tag{3.8}$$

The final gradient of the topology loss with respect to the original mask M is:

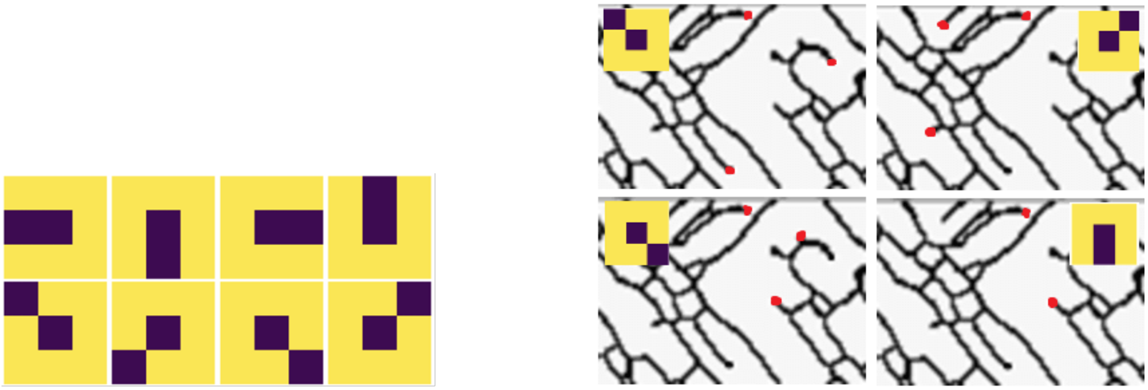
$$\frac{\partial L_{\text{topo}}}{\partial M} = \frac{\partial L_{\text{topo}}}{\partial M'} \cdot \frac{\partial M'}{\partial M} \tag{3.9}$$

Since both components are differentiable, the overall gradient is well-defined, and the optimization function can be represented as:

$$\theta^* = \arg\min_{\theta} L_{\text{topo}}(\theta) \tag{3.10}$$

where θ represents the parameters of the model and θ^* is the optimal set of parameters that minimize the loss.

Including L_{topo} in the model's training process is conditional on the scale belonging to the second half of the GAN pyramid, which in our case studies corresponds to level four and beyond, as we had a total of eight levels according to the dimensions of the input image. This implementation strategy is based on empirical results indicating that introducing the topology loss at earlier stages can disrupt the learning of foundational features. By applying L_{topo} starting from the second half of the pyramid, the model has already developed basic structures, allowing L_{topo} to effectively enhance topological accuracy without impeding initial convergence. The architecture of the developed loss function is depicted in Figure 3.3.



(**a**) Terminal node detector kernels.

(**b**) Detected terminal nodes overlaid with binary mask.

Figure 3.2: (**a**) The eight 3 × 3 convolution kernels employed in the topology enhancer loss. Dark pixels represent 1 s and 0 s otherwise. (**b**) The convolution kernels' performance for topology loss. The output of the kernel convolution (highlighted points) is overlaid on the input tensor for better visualization.

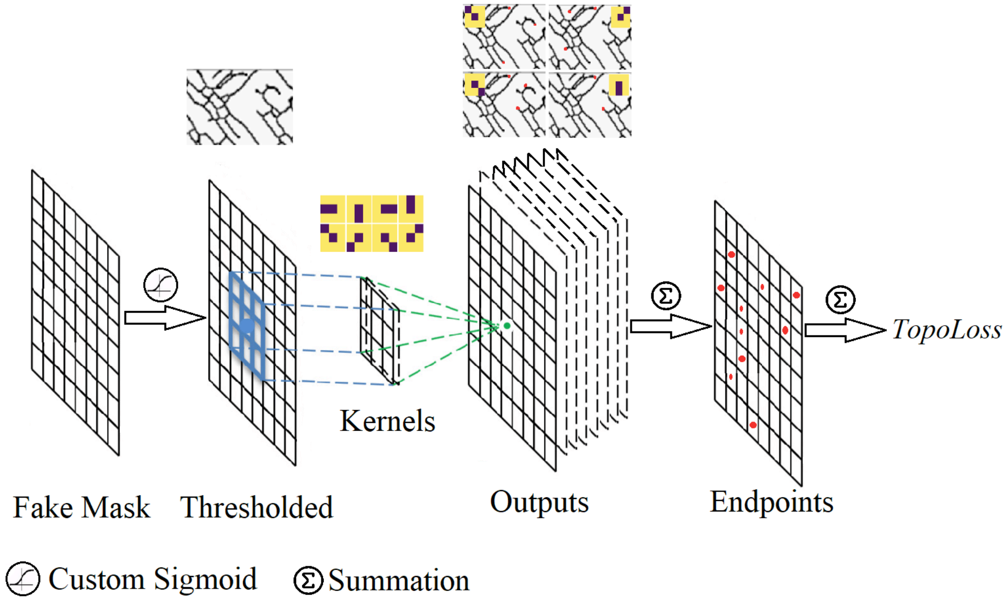


Figure 3.3: The architecture of the developed loss function.

3.3.2 Evaluation

Node Topology Clustering (NTC)

GAN evaluation metrics focusing on the topological structure of the synthetic images are rare. Conventionally, pretrained CNN-based models [127] are used to evaluate the quality of images generated by GANs. Inception score (IS) [128] and Frechet Inception Distance (FID) [21] are common methods of this type. IS calculates the Kullback–Leibler divergence between the conditional and marginal class distributions across the generated data. Likewise, FID uses an Inception network pre-trained on ImageNet to map images into a feature space [13]. These models are biased toward the Inception models feature space, while the topological properties are not guaranteed to be preserved in the feature space [4]. Therefore, recent evaluation metrics have been developed to focus more on the topological structure of the generated images to provide a more reliable quantitative description of the model performance [59, 4]. These metrics are mostly based on topological data analysis (TDA), which generally uses persistent homology (PH) to measure the similarity of the distribution of persistence diagrams for real and generated images.

However, while PH-based metrics focusing on topology can provide reliable evaluations, they have limitations. These metrics are typically based on the birth and death of holes and simplices at different scales, forming the final barcodes. Consequently, the formation of these barcodes is influenced not only by topological relationships but also by geometric properties such as shape and size [129, 130]. In the context of GANs, the inherent stochasticity produces diverse images, which may include both geometric and topological variations. Moreover, TDA- and PH-based evaluation methods, such as Betti number error, are not well-suited for assessing models like ours, which consider the presence of a terminal node as a topological "error". TDA primarily focuses on features such as connected components and holes without explicitly accounting for terminal nodes, which we consider critical in evaluating topological accuracy. Therefore, relying solely on TDA-based metrics may not fully capture the quality and variability of the generated images. In this paper, we propose Node Topology Clustering (NTC), a novel topology-aware metric to evaluate the performance of GAN models focusing on topology and structural integrity.

The NTC calculation begins with converting the input binary mask into an undirected graph. In this graph, nodes represent critical points such as terminals and junctions, and edges represent the connections between these points using straight lines. Nodes with a degree of one, referred to as terminals, are of particular interest because they indicate points of disconnectivity within the graph. Using these terminal nodes, we define a new measure called Terminal Distance (TD), which captures the distance from each node to its nearest terminal (Figure 3.4).

Definition 1 (Terminal Distance (TD))**.** *The Terminal Distance (TD) quantifies the geodesic distance [131] in a graph, derived from a binary image mask, between each node and its nearest terminal node. Let G be an undirected graph constructed from a binary mask M, where nodes represent critical points such as terminals and junctions. The Terminal Distance $TD(v)$ for a node $vs. \in G$ is computed as:*

$$TD(v) = \min_{t \in T} d(v, t) \tag{3.11}$$

where T is the set of all terminal nodes and $d(v, t)$ is the geodesic distance between nodes v and t.

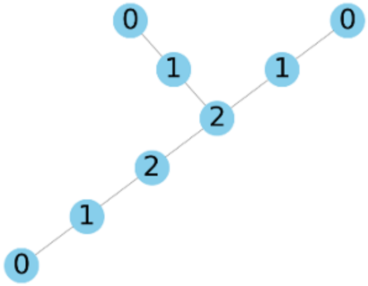


Figure 3.4: Illustration of defined terminal distance (TD) in a simple graph.

Using the TD values and the graph's adjacency matrix, a local Moran's I clustering analysis, also known as Local Indicators of Spatial Association (LISA) [132], is conducted to identify patterns and clusters based on topological relationships among nodes. This analysis categorizes nodes into High-High (HH), Low-High (LH), Low-Low (LL), and High-Low (HL) clusters, reflecting nodes with similar or contrasting TD values and their neighbors with a significance level of 95%. Due to the nature of the TD metric, the input graphs must have at least one terminal node. To ensure this, we randomly connected a single terminal node to each graph. In Appendix A.1, we demonstrate that the addition and position of this terminal node do not significantly affect the functionality of the proposed evaluation metric (Table A.1). Since the predicted masks may have different node counts, we normalize cluster sizes into percentages. Finally, the NTC index is measured by calculating the multidimensional Euclidean distance between the four LISA cluster values of the predicted mask and the ground truth (GT) mask. This Euclidean distance reflects differences in the LISA attributes in multidimensional space, not spatial features.

To evaluate the significance of our proposed NTC topological features, we created graphs in seven groups of extreme cases based on the topological distribution of terminal nodes in a two-dimensional grid of 15 by 15 nodes. These cases include random, dispersed, clustered, clustered-star, clustered-web, isolated edges, and randomly split-edge distributions, detailed in Appendix A.2, Figure A.1. We generated 1000 instances of each group. We used Support Vector Machines (SVM) [133] with 10-fold cross-validation to ensure consistency with previous studies that developed feature extraction methods for graph structural clustering [134, 135]. We employed a suite of five network metrics including network density, average clustering coefficient, transitivity, modularity, and assortativity, as proposed by [134]. Network density measures the ratio of actual connections to possible connections, indicating how interconnected the network is [136]. The average clustering coefficient indicates the degree to which nodes cluster together, offering insights into tightly connected groups [137]. Transitivity assesses the probability that adjacent nodes are interconnected, indicating global clustering [138]. Modularity measures the strength of division into communities, and assortativity evaluates the tendency of nodes to connect with others of similar degrees, highlighting hierarchical structures [139].

Additionally, we calculated interval probabilities, known as Degree Distribution Quantile Concentration (DDQC) features, to provide a detailed view of the degree distribution within the network [134]. DDQC, by segmenting node degrees into specific intervals based on statistical properties such as the mean and standard deviation, generates a histogram quantifying the proportion of nodes in each interval. These interval probabilities normalize the histogram values and offer a probabilistic perspective of how node degrees are distributed across the network. Later in this paper, we use these graph features to measure the similarity of graphs generated by GAN models.

Modified Fréchet Inception Distance (FID)

In this work, we employed a modified version of the Fréchet Inception Distance (FID) [21] to evaluate the quality of the images generated by TopoSinGAN compared to those from the baseline SinGAN. This modification accounts for having only a single real input image alongside multiple generated images. Feature representations were extracted using a pre-trained InceptionV3 model for both the real and generated images, and the mean and covariance of these activations were computed accordingly.

Next, we calculated the mean and covariance of the activations for both the real image and the generated images. Since there was only one real image, its covariance matrix was set as the identity matrix of appropriate dimensions to ensure a valid comparison. The modified FID score was then computed by measuring the distance between the means and covariances of the real and generated activations, incorporating the Fréchet distance formulation. To handle numerical issues arising from the matrix square root calculation in the covariance product, adjustments were made to work with real components only if complex values appeared. This modified approach enabled us to evaluate the quality and diversity of generated images relative to the single reference image in a statistically meaningful manner.

3.3.3 Experimental Setup

For this research, we considered two experiments to showcase the practicality of the developed model. These two experiments include an agricultural fields' image, and the second experiment uses a dendrology image. We also use the CREMI [111] dataset to compare our model with TopoGAN [4] and WGAN.

Agricultural Fields

For this experiment, we used a 175 × 240 pixel 3-meter resolution XYZ tile downloaded using QGIS 3.34.0 software from agricultural fields of Tehran province, Iran. This image contains red, green, and blue (RGB) channels. We manually annotated the image delineating agricultural fields in the image and appended it as the fourth channel to this image. Notably, the agricultural plots within this region present considerable variation in shape and size, making them intricate subjects for accurate prediction. This study area is especially challenging given the intrinsic differences in field geometries, topological arrangements, irrigation systems, and the diverse range of crops.

Dendrology

In our research, we employed a single gigapixel macro photography (GMP) image, measuring 175 × 240, as developed by [111]. This image, a part of a large dataset featuring an ultra-high resolution of 19,812 dpi, was manually cropped and annotated. It encompasses a four-channel composition, including RGB and an additional mask channel that delineates cellular structures of xylem vessels in hardwood tree sections. In the field of dendrology, such images are invaluable, offering insights ranging from tree age to environmental adaptations, as discussed by [140]. A typical 1 mm^2 area of a tree ring contains approximately 750 vessels, presenting significant challenges in manual delineation. This complexity underscores the necessity of leveraging deep learning methodologies for efficient and accurate data synthesis in dendrology studies.

3.3.4 System Setup

System Configurations

In our study, we utilized the high-performance computing resources of the Minnesota Supercomputing Institute (MSI). The computational backbone of our research was supported by the NVIDIA A100 GPU (NVIDIA, Santa Clara, CA, USA) on the Agate cluster, specifically within the a100-4 partition. For reproducibility purposes, detailed documentation about the system configuration can be found at their website [141].

Training Procedure

Our training framework employs a multiscale approach. It starts at a base scale and progressively refines through larger scales. This progression is parameter-controlled, which defines the number of scales and when to stop refining. The Discriminator is trained with both real and synthetically generated images, incorporating traditional adversarial training steps and a gradient penalty for stability. The Generator is then trained to produce images that increasingly challenge the Discriminator's detection capabilities. Additional loss functions, such as topological and reconstruction losses, are applied conditionally based on the scale and training parameters. Coefficients λ_1, λ_2, and λ_3 in the loss function (Equation (3.1)) were set to 10, 5, and 0.145, respectively, in our experiments. These values were determined through extensive hyperparameter tuning, carefully balancing adversarial, topological, and reconstruction components to preserve topological features and maintain overall image quality, as reflected in improved NTC scores discussed in Section 3.4.2. In Section 3.4.5, we provide a more detailed discussion on the hyperparameter tuning process for λ_3.

3.4 Results

3.4.1 NTC-Based Graph Classification

The graph classification results, as depicted in Figure 3.5, illustrate the performance of different feature sets in categorizing graph structures using a Support Vector Machine (SVM) classifier. The "Features" set, which includes basic network attributes, achieved an accuracy of 71.35%. The addition of Degree Distribution Quantile Concentration (DDQC) features, labeled as "Features + DDQC", significantly improved the classification accuracy to 94.73%. The addition of Node Topology Clustering (NTC) features shown as "Features + DDQC + NTC" achieved the highest accuracy of 98.39%. These results suggest that incorporating NTC graph features enhances the model's ability to distinguish between different graph categories with disparate topological structures.

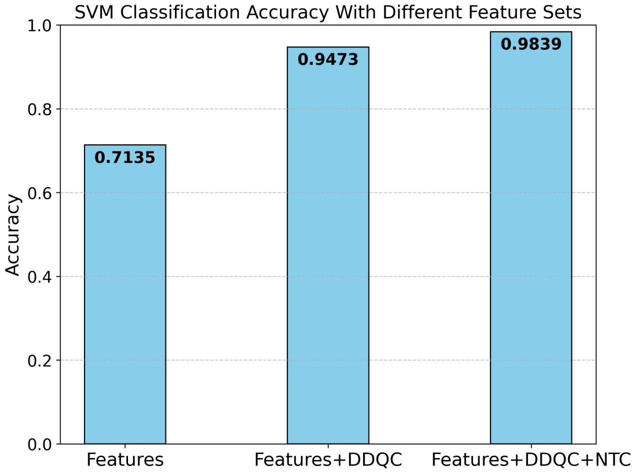


Figure 3.5: The result of SVM graph classification using different feature sets.

3.4.2 TopoSinGAN Performance Evaluation Using NTC

The qualitative evaluation of synthetic images generated by SinGAN and TopoSinGAN reveals notable differences, as illustrated in Figure 3.6. While SinGAN outputs visually resemble the input images, they exhibit topological inaccuracies upon closer inspection. These inaccuracies include discontinuities along field boundaries and isolated line segments within the generated masks. Such anomalies pose a risk to applications that depend on precise topological structures.

In contrast, TopoSinGAN demonstrates improved topological coherence. It maintains the continuity of agricultural field boundaries without introducing unintended gaps. Additionally, it minimized the number of abnormally isolated line segments within fields. Quantitative assessment using the NTC metric reveals a substantial improvement in accuracy with TopoSinGAN. Specifically, for agricultural and dendrology cases, the NTC indexes between the GT and the 1000 generated images were 15.15 and 14.55, respectively, for SinGAN. These distances are significantly reduced to 3.94 and 2.44 with TopoSinGAN. Lower NTC values indicate higher topological accuracy, thereby demonstrating the effectiveness of TopoSinGAN in enhancing topological coherence. The detailed NTC evaluation results are summarized in Table 3.1.

Table 3.1: Measured NTC indexes for 1000 synthetic images.

	SinGAN		**TopoSinGAN**	
	Mean	***STD***	***Mean***	***STD***
Agricultural	15.15	3.41	**3.94**	**1.81**
Dendrology	14.55	3.07	**2.44**	**1.35**

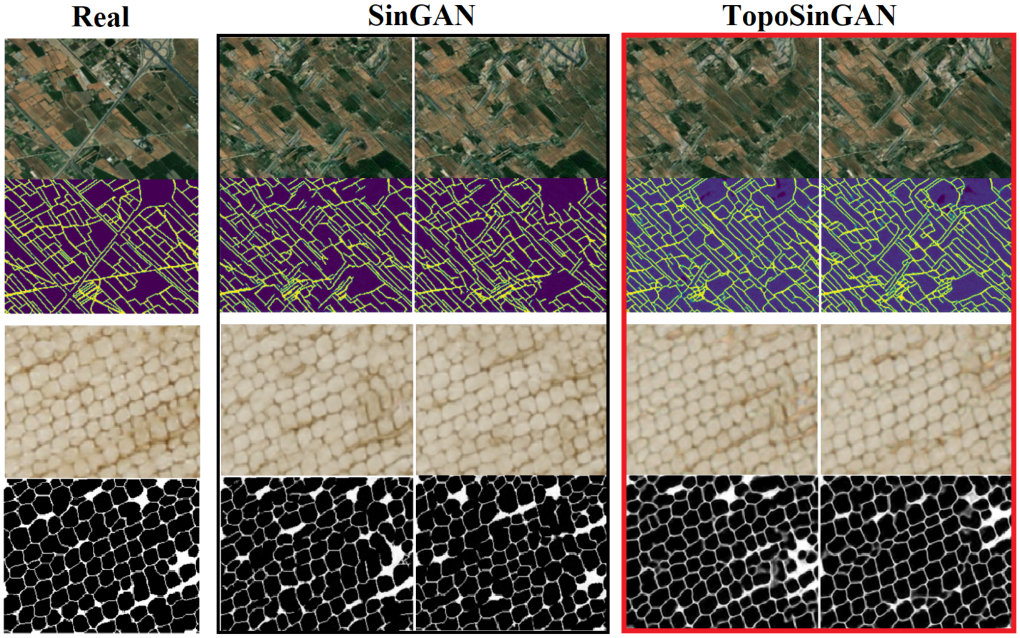


Figure 3.6: The outputs of the SinGAN and TopoSinGAN models trained on a single real image for agricultural fields and dendrology experiments. The red frame highlights the TopoSinGAN results, emphasizing the improved topological accuracy in comparison with SinGAN.

To further evaluate the performance of our model and facilitate comparison with existing approaches, we conducted experiments using the CREMI dataset, assessing the results through the NTC metric compared with the TopoGAN and WGAN models. The comparative results are summarized in Table 3.2. Our model demonstrated improved NTC scores, recording an NTC of 8.15, compared to TopoGAN, which achieved an NTC of 11.05.

The WGAN and plain SinGAN models recorded similar NTC scores of 15.76 and 15.46, respectively. Additionally, Figure 3.7 presents sample generated masks produced by each model, visually comparing their performance.

Table 3.2: NTC evaluation of TopoSinGAN in comparison with other models on the CREMI dataset.

	SinGAN		**WGAN**		**TopoGAN**		**TopoSinGAN**	
	Mean	***STD***	***Mean***	***STD***	***Mean***	***STD***	***Mean***	***STD***
CREMI	15.46	2.86	15.76	2.77	11.05	2.80	**8.15**	**2.01**

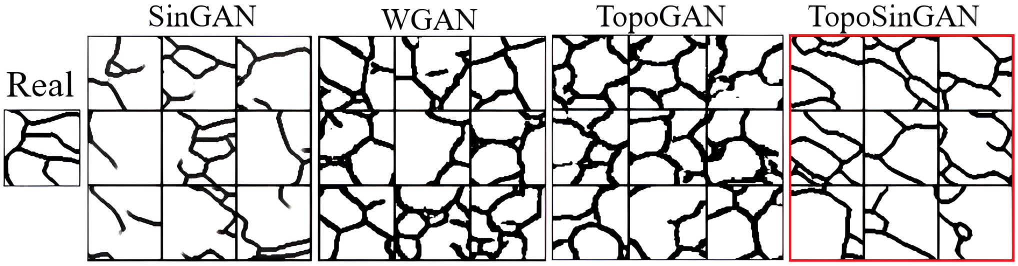


Figure 3.7: Sample outputs generated by TopoSinGAN and other GAN models. The single input used for the SinGAN and TopoSinGAN training is displayed on the left, while WGAN and TopoGAN were trained using the CREMI dataset. The red frame highlights the outputs of the TopoSinGAN model, showcasing its topological consistency relative to other models.

3.4.3 Modified FID Evaluation

The experiments were performed on both case studies, dendrology, and agricultural fields, with the TopoSinGAN and the baseline SinGAN models trained on each input image for 10 runs, generating 1000 synthetic images per experiment. The measured average FID values were recorded in Table 3.3.

Table 3.3: Average FID values over 10 runs for 1000 synthetic images per model for each case study.

	SinGAN		**TopoSinGAN**	
	Mean	***STD***	***Mean***	***STD***
Agricultural	0.2485	0.0086	0.1914	0.0083
Dendrology	0.0014	0.0013	0.0014	0.0018

As shown in Table 3.3, the average FID for the agricultural fields image, TopoSinGAN achieved an average FID value of 0.1914, compared to 0.2485 for the baseline SinGAN. The agricultural fields' image is characterized by significant variations in color, texture, and geometric patterns, leading to inherently higher FID values. The lower FID score for TopoSinGAN suggests that the topologically improved model produced more realistic images compared to SinGAN. Specifically, TopoSinGAN's ability to preserve topological structures while maintaining visual diversity led to a reduction of 0.0571 in FID, demonstrating its effectiveness in generating images with enhanced topological accuracy while retaining natural variation.

The average FID for the dendrology experiment with TopoSinGAN was 0.0013, while the baseline SinGAN produced a similar value of 0.0014. Given the inherent similarity in the structure and geometry of xylem cells in the dendrology image, small FID values are expected. The similarity between the two models indicates that TopoSinGAN maintains the structural fidelity of the generated xylem images.

3.4.4 Comparative Efficiency Analysis

To assess the efficiency of our method relative to the original SinGAN model, we conducted experiments where the model processed input images of varying dimensions: 175 × 240, 350 × 400, and 500 × 500. The SinGAN models architecture, dictated by a scale factor parameter set by default to 0.75, generated 8, 11, and 12 scales for the respective image dimensions. At each scale of the pyramid, the training process ran for 1000 iterations. In all three experimental setups, topology loss was deliberately excluded during the first four scales to allow the model to form basic topological features without constraints. It was then introduced at finer scales (5–12) to refine the details while preserving structural integrity. All other configurations and parameters remained consistent with those detailed in Section 3.2.1. The experiment involved running the model on each image size ten times. Table 3.4 outlines the average recorded execution times for each configuration. On average, our TopoSinGAN model demonstrated an approximate 9.41%, 7.29%, and 5.98% increase in training time compared to the baseline SinGAN model for 175 × 240, 350 × 400, and 500 × 500 dimensions, respectively.

Table 3.4: The results of comparative performance analysis.

Input Dimensions	**Pyramid Scales**	**SinGAN (minutes)**	**TopoSinGAN (minutes)**
$175 \times 240 \times 4$	8	20.93	22.90
$350 \times 400 \times 4$	11	65.13	69.88
$500 \times 500 \times 4$	12	108.11	114.58

3.4.5 Hyperparameter Tuning of λ_3 for CREMI Experiment

According to our experiments, the value of λ_3, the coefficient of L_{topo} in Equation (3.1), depends on factors such as input image dimensions and the density of line features in the real mask. Therefore, hyperparameter tuning is crucial for each case study. Figure 3.8 shows the tuning process of the λ_3 parameter for the CREMI experiment. In this experiment, we gradually increased λ_3 by 0.2, starting from 0 (baseline SinGAN), while keeping the other two hyperparameters constant, and recorded the average FID and NTC values for each trial.

In Figure 3.8, the blue line represents FID scores, which are initially low for SinGAN. As λ_3 increases to 0.2, the FID worsens to 0.1801, while the NTC slightly improves to 14.9. The best balance of FID and NTC occurs at $\lambda_3 = 0.4$, with FID at 0.1127 and NTC at 8.15. When λ_3 is further increased to 0.6, NTC decreases to 6.91, but the FID jumps to 0.15, indicating a negative impact on the realism of the generated contexts. At $\lambda_3 = 0.8$, this trend continues, with NTC dropping to 6.02 and FID rising sharply to 0.21. It is important to note that higher values of λ_3 lead to model divergence.

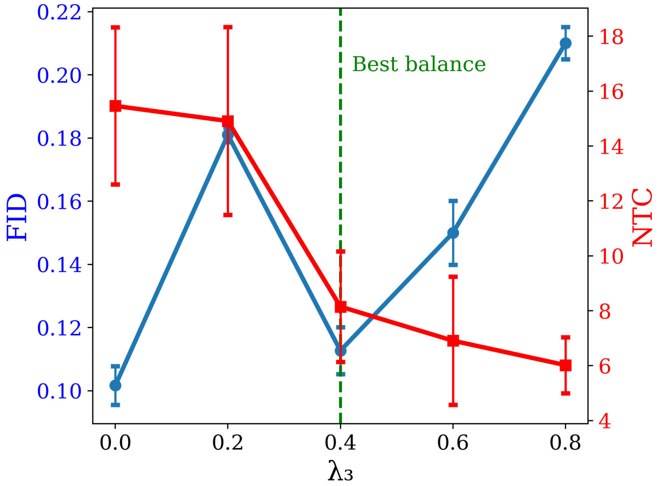


Figure 3.8: Tuning of the λ_3 parameter for the CREMI experiment, showing its impact on FID (blue line) and NTC (red line). The best balance between FID and NTC is achieved at $\lambda_3 = 0.4$ for this experiment.

3.5 Discussion

The research presented demonstrates the effectiveness of TopoSinGAN in enhancing topological accuracy through its integrated topology loss function. Unlike conventional losses, such as adversarial or reconstruction losses [5, 142], the topology loss optimizes for lower terminal counts using its specialized convolutional kernels. This approach allows TopoSinGAN to minimize topological anomalies that could otherwise go unnoticed, thereby addressing a critical need in applications requiring high topological integrity. While SinGAN retains visual similarities, structural discrepancies can arise without dedicated topological supervision, which TopoSinGAN successfully mitigates by penalizing these errors.

A key feature of the topology loss function is its full differentiability, which facilitates end-to-end training via backpropagation. This seamless integration within the training cycle enables direct optimizations based on topological feedback, surpassing methods that rely solely on post-processing. The end-to-end nature of this approach ensures that topological feedback is continually refined throughout the training process, leading to more robust and accurate models. Furthermore, our study introduces a novel GAN evaluation metric focused purely on the topological attributes of the images, excluding geometric variations. This metric, termed the NTC score, provides a more realistic judgment of GAN models by quantifying topological integrity. Consistent improvements in NTC scores indicate the quantitative effectiveness of TopoSinGAN, as lower NTC values correspond to reduced loop openness and fewer topological anomalies. Importantly, these lower NTC scores denote higher topological accuracy, not merely structural similarity. Improved NTC scores reported in the comparative evaluation of TopoSinGAN with plain SinGAN, WGAN, and TopoGAN further highlight TopoSinGAN's effectiveness in generating topologically enhanced images. According to Table 3.3, SinGAN and WGAN recorded similar NTC scores as they both use similar adversarial losses [5].

In Appendix A.3, we use the common graph metrics to demonstrate that our proposed topology loss has no significant negative impact on the learning flow of the generator compared to the original SinGAN (Table 3.1). The high similarity of common graph metrics for SinGAN and TopoSinGAN outputs suggests two observations. First, both models effectively replicate the general structure of real input images. This signifies that our topology loss has not significantly affected the realism of the generated image contexts which aligns with the result of modified FID evaluation. Second, it also reveals a limitation of common graph metrics to capture the topological inaccuracies within mask graphs. The NTC metric addresses this gap by focusing on a graph's topological structure, offering a more nuanced evaluation of topological accuracy. Unlike traditional metrics that compare graphs or graphlets for similarity [143, 143, 144, 145], the NTC metric emphasizes topological integrity, making it a specialized tool for assessing the precision of topological features in a graph.

Additionally, the topological loss function in TopoSinGAN is uniquely characterized by its standalone application, assessing the models topological loss based solely on the generated mask channel. This approach penalizes the model for topological inaccuracies without comparing the generated masks to real inputs, focusing on minimizing the occurrence of terminal nodes. Despite the modest additional computational cost, its benefits in preserving the integrity of topological features make it a valuable tool for various applications demanding precise structural fidelity.

Nevertheless, TopoSinGAN has certain limitations that should be considered. Its treatment of hanging lines as anomalies makes it particularly suitable for applications requiring closed loops, such as agricultural field delineation or membrane cell segmentation. However, for applications where hanging lines are expected, such as in road network analysis, TopoSinGAN may not be the optimal choice. This specificity underscores the importance of context in applying any approach including TopoSinGAN, highlighting its potential limitations in certain scenarios. This limitation is due to the design of the current optimization function (Equation (3.10)), which is designed to minimize the number of terminal nodes in the generated mask. To extend the applicability of our method to scenarios where hanging lines are expected, a possible adaptation could modify the optimization function to minimize the difference between the number of terminal nodes in the real input mask and those in the generated mask. For example, an adaptation of the optimization function (Equation (3.10)) could be expressed as:

$$\theta^* = \arg\min_{\theta} \left(L_{\text{topo}}^{\text{Real}}(\theta) - L_{\text{topo}}^{\text{Generated}}(\theta) \right) \tag{3.12}$$

This approach could allow the model to preserve a desired number of terminal nodes rather than eliminating them entirely. However, implementing and validating this adaptation is beyond the scope of the current research, but could be addressed in future work. Another limitation of TopoSinGAN relates to the thickness of the boundary mask in the input real image. Our loss function uses fixed 3 × 3 terminal detector kernels, which are sensitive to boundary thickness, potentially leading to suboptimal performance in certain cases. Two potential adaptations could be to convert the mask layer to a centerline (skeletonized) binary mask or introduce larger versions of the kernels, such as 4 × 4 or 5 × 5. Developing a version of TopoSinGAN that is robust to mask thickness will also be a focus of future work.

3.6 Conclusions

In this study, we introduced TopoSinGAN, an extension of SinGAN that incorporates a differentiable topology loss function to enhance topological accuracy in synthetic image generation. By minimizing terminal nodes, TopoSinGAN effectively addresses topological anomalies that conventional GAN losses often overlook, without significantly increasing computational complexity. Another contribution of this work is the proposal of the Node Topology Clustering (NTC) metric, which focuses on evaluating topological integrity rather than just geometric similarity. Our experiments, involving agricultural and dendrological images, demonstrated that TopoSinGAN achieves significantly better topological fidelity than SinGAN, as evidenced by lower NTC indexes. This indicates fewer instances of loop openness and isolated line segments, which are crucial for applications requiring precise topological features. While TopoSinGAN excels in preserving topological integrity, it may not be ideal for contexts where open structures are expected. Nevertheless, it represents a crucial advancement in generating topologically accurate synthetic data, with potential applications in GIS, medical imaging, and other fields where topological accuracy is essential. Future work will focus on refining the topology loss function and extending its application to a broader range of tasks, such as image semantic segmentation. Additionally, we plan to adapt the proposed model to overcome the limitations outlined in Section 3.5. Moreover, the development of more sophisticated evaluation metrics for topological accuracy will remain a priority, aiming to enhance the robustness and applicability of generative models across diverse fields. The code and data for this study are available in the projects GitHub repository [146].

Chapter 4

TopoSegNet: Scalable Topology Preservation in Image Segmentation via Critical Points

4.1 Abstract

Image segmentation is crucial in computer vision, with applications in various fields. Despite advancements in deep learning techniques, maintaining topological consistency in segmented outputs remains a significant challenge. Traditional topology-aware methods, such as those using persistent homology (PH), preserve topological features such as loops and connected components but are often computationally expensive. We introduce a new approach that shifts focus from topological features like loops and connected components to points. We term points that play a crucial role in topology, topologically critical points (TCPs) such as junctions and terminal nodes that form these topological features. Focusing attention on points reduces computational overhead while preserving the key topological structures of the segmented outputs. We propose TopoSegNet, a novel topology-aware loss function for image segmentation that emphasizes the preservation of TCPs to maintain topological integrity without the computational complexity of traditional approaches. Evaluated on datasets such as DRIVE, CREMI-B, GMP for Tree Rings, and satellite imagery of agricultural fields, TopoSegNet consistently outperforms existing methods on

both pixel-wise accuracy and topology-focused metrics demonstrating TopoSegNet as a scalable and efficient approach for topology-aware image segmentation.

4.2 Introduction

In recent years, image semantic segmentation has become a fundamental task in computer vision with applications spanning medical imaging, autonomous driving, robotics, and geographic information systems (GIS). The objective of image segmentation is to partition an image into meaningful segments by assigning a label to each pixel, thus enabling the extraction of object boundaries, structures, or regions of interest. While deep learning techniques, particularly convolutional neural networks (CNNs), have revolutionized segmentation tasks by achieving remarkable pixel-wise accuracy, challenges remain, particularly in maintaining structural and topological consistency in the segmented output [41, 58, 43].

The preservation of topological features is particularly important when dealing with fine-scale, curvilinear structures such as xylem cells in gigapixel images of tree cross-sections (GMP Images) [111], irregular agricultural fields in the Middle East, road networks in satellite imagery, blood vessels in medical imaging, and neurons in microscopy. Small topological errors such as missing connections, broken branches, or merged structures can have a significant impact on the functional interpretation of segmentation results. While these errors may contribute only marginally to pixel-wise metrics such as accuracy or precision, they can lead to serious issues in practical applications [7]. For example, in medical imaging, topological mistakes in vessel segmentation can result in incorrect diagnoses or treatment plans, and in GIS, topological errors in road networks can disrupt navigation and transportation planning systems.

To address these challenges, this research introduces a novel topology-aware loss function specifically designed for image semantic segmentation. The proposed loss function offers two key advantages: first, it is computationally efficient, especially in comparison to methods that rely on persistent homology; second, it significantly enhances segmentation performance, particularly in terms of Dice and Intersection over Union (IoU) metrics. These metrics are widely recognized for evaluating image segmentation tasks, especially when dealing with imbalanced class distributions. The improvement becomes even more pronounced when using topology-based evaluation metrics such as Betti numbers and maximum mean discrepancy (MMD). By integrating topological constraints into the segmentation process, our approach not only enhances the structural integrity of segmentation outputs but also maintains high accuracy across various benchmarks.

4.3 Background and Related Work

4.3.1 Image segmentation challenges

Image segmentation is a fundamental task across various domains that require detailed image analysis, such as biomedical imaging and satellite imagery. While traditional segmentation methods have been successful in delineating coarse structures, they often struggle with capturing fine, intricate topological features such as thin connections, small parts, or elongated structures [14, 147]. These limitations are particularly pronounced in applications such as blood vessel segmentation or neuron mapping, where structural errors, such as broken connections or missing branches, can significantly affect downstream tasks such as disease diagnosis or neuroanatomical analysis [147, 38].

Modern deep learning models, such as U-Net [6], Mask R-CNN [148], and segment anything model (SAM) [76] have significantly improved pixel-wise segmentation accuracy. However, despite their success, these models often fail to preserve the topological integrity of fine-scale structures. This failure arises because traditional segmentation techniques typically optimize pixel-wise loss functions, such as cross-entropy or Dice loss, which do not account for topological correctness [7, 18]. These limitations have led to the exploration of new methods, such as uncertainty quantification (UQ), Discrete Morse theory (DMT), and topology-aware approaches, aimed at addressing the gaps in preserving structural consistency.

4.3.2 Uncertainty quantification

Uncertainty Quantification (UQ) refers to the process of estimating the confidence or uncertainty associated with predictions made by machine learning models. In segmentation tasks, UQ helps to identify areas where the model is less certain, which is critical for making reliable decisions in high-stakes applications such as medical imaging [149, 150]. In recent years, significant advancements have been made in the field of uncertainty quantification (UQ) for deep neural networks, particularly in semantic segmentation tasks [149, 151, 152]. Since semantic segmentation involves per-pixel classification, most UQ techniques focus on estimating uncertainty at the pixel level. For instance, a Bayesian framework using Monte Carlo (MC) dropout [153] combined with a learned loss attenuation is proposed in [154] to capture both model and data uncertainty. Other recent approaches use generative models to produce multiple segmentation hypotheses, with the per-pixel variance across these hypotheses treated as uncertainty. Examples include ensemble methods with multiple networks [155], single networks with multiple heads [156], Prob.-UNet [157], and PHiSeg [158]. Prob.-UNet integrates a conditional variational autoencoder [159] with UNet, generating multiple hypotheses via latent variable sampling, while PHiSeg introduces latent variables at every level of the UNet to increase the diversity of generated samples.

4.3.3 Discrete morse theory

Discrete Morse theory (DMT) has also emerged as a valuable tool for improving the topological fidelity of image segmentation [14, 160, 34, 161, 162, 163]. DMT identifies critical topological structures such as ridges and valleys in the likelihood maps of predicted segmentation and incorporates these structures into a loss function that penalizes errors at key topological locations. This approach helps mitigate segmentation errors in regions where the topology is most fragile, such as near-thin connections or blurred boundaries [14, 147]. A notable method integrating DMT, proposed by [14], leverages this theory to improve both pixel-wise and topological accuracy by focusing on critical structures such as neuron membranes and vessel connections. Furthermore, probabilistic DMT has been used to model uncertainty in segmentation, allowing for more accurate human-in-the-loop proofreading by highlighting error-prone regions based on their topological uncertainty [147].

4.3.4 Topological data analysis

Topological Data Analysis (TDA) through persistence homology (PH) offers a promising approach to address the challenges of preserving topological structures through deep learning [7, 24]. Topology, in the context of TDA, refers to the characteristics of an object that remain unchanged under continuous deformations, such as the number of connected components and loops, commonly quantified using Betti numbers [45]. Ensuring that these topological features are preserved in segmentation outputs helps ensure that fine-scale structural integrity is maintained [14]. PH provides a powerful framework for capturing topological features across multiple scales by tracking the birth and death of topological structures such as connected components, holes, and voids throughout the segmentation process. By analyzing the evolution of these structures across different intensity or confidence thresholds, PH enables a detailed, multi-scale analysis of the segmentation's topology [147, 164].

As an emerging analytical framework, TDA is particularly useful for studying data shapes by identifying and quantifying their topological features, including connected components, loops, and voids. This approach holds significant value in data-intensive domains where the structural characteristics of data are critical such as in image segmentation, where maintaining connectivity and structural integrity is essential [7, 165]. By constructing mathematical structures that represent the underlying topology of data, TDA allows for insights into connectivity, clustering, and continuity at various scales, ultimately aiding in more robust segmentation and structural analysis.

Simplices and Topological Structures in TDA

In TDA, simplices serve as the basic building blocks. These geometrical constructs describe the relationships between data points and help form complex structures called simplicial complexes:

0-Simplex (Point): A single point in space, representing an individual data element.

1-Simplex (Line Segment): Created by connecting two points, forming a line or edge.

2-Simplex (Triangle): Generated by linking three points in a triangular shape.

3-Simplex (Tetrahedron): Constructed by joining four points, forming a three-dimensional solid.

Each k-dimensional simplex represents a multi-point connection that builds up the data structure. A collection of simplices forms a simplicial complex, which organizes these elements into a topological space [166]. In TDA, simplicial complexes are analyzed through the concept of persistent homology (PH), which tracks the birth and death of features as the data structure evolves across multiple scales, ultimately capturing connected components, cycles, and voids within data structures.

Persistent Homology

Persistent homology (PH) provides a framework for understanding how features within data, such as connected components or loops, persist across scales, an essential factor for analyzing topological properties in segmentation tasks [167]. PH employs filtration processes, progressively adding simplices to form higher-dimensional complexes. Through this, features such as 0-dimensional connected components, 1-dimensional loops, and 2-dimensional voids are quantified in a way that allows segmentation tasks to prioritize robust, scale-invariant structures [168]. The use of persistence diagrams and Betti numbers allows for the quantification of these features, providing a means to measure topological stability across the data.

Role of 0-Simplices in Image Processing

0-simplices (i.e., points or pixels in 2D-images) are the simplest and most fundamental structures in TDA, representing isolated components within data. In segmentation applications, 0-simplices serve as the essential units for connected component analysis. By examining points in this manner, segmentation tasks can: Identify individual data regions before more complex structures are introduced. Establish connectivity between data points, ensuring that distinct regions are accurately preserved. Create a foundation for higher-order relationships, such as edges and loops, essential for preserving continuity and avoiding fragmentation in segmentation models [165]. The significance of 0-simplices is underscored in PH as they form the baseline connected components (Betti-0 features), which are crucial for establishing structural consistency within segmentation tasks, especially in image analysis where connectivity errors can severely affect the accuracy of object boundaries [165, 168]. Literature employing PH for image processing tasks

[7, 75, 169, 170, 171] typically emphasizes 1-dimensional features (i.e., loops) as critical topological features.

4.3.5 Topology-aware loss functions

Most studies in these fields incorporate topological constraints through custom loss functions, with recent efforts integrating topological information directly into segmentation model training [7, 18, 75, 40, 172, 173, 55]. Topology-aware loss functions enforce per-pixel constraints to enhance the preservation of topological integrity. One of the earliest approaches, proposed by [18], introduced a topology-aware loss function designed to preserve elongated and thin structures, such as blood vessels, by using feature maps from pre-trained CNNs. Although this method improved the preservation of these structures, it did not explicitly enforce topological invariants such as Betti numbers, leading to suboptimal results in more complex topological scenarios [38].

Hu and colleagues [7] introduced a persistent homology-based loss function that directly integrates topological information into the training of deep neural networks for image segmentation. Their method achieved impressive results in ensuring topological correctness, particularly in biomedical datasets with fine structures such as neuron membranes [7]. This approach serves as one of the baselines against which we evaluate our proposed method. In another study, [75] developed a method to integrate prior topological knowledge into neural network training for image and volume segmentation. Using PH, they specified topological features such as Betti numbers to guide the segmentation process without requiring ground-truth labels. Demonstrated across various datasets, their approach improved performance in challenging tasks, showing that topological priors can enhance segmentation accuracy in both semi-supervised and post-processing contexts [75]. Another notable approach is the topology-aware focal loss (TAFL), which combines focal loss with a topological constraint based on the Wasserstein distance between the persistence diagrams of predicted and ground truth masks.This approach has shown promise in resolving common topological errors such as disconnected regions and broken boundaries, particularly in 3D medical image segmentation [24]. Another example is TopoSeg, which applies PH to guide the learning process for nuclear instance segmentation, ensuring topological correctness by focusing on the persistence of structural features in the segmentation probability maps [38].

Homotopy theory has also been employed to develop loss functions that explicitly target the preservation of topological consistency between predicted and ground truth segmentations [55]. The homotopy warping loss, for instance, focuses on identifying and correcting topologically critical pixels by transforming the predicted mask towards the ground truth without altering its topology. This method has shown significant promise in applications such as satellite and neuron image segmentation, where maintaining the topology of fine structures is essential [55].

4.4 Identified Gaps

Despite the promise of topology-aware segmentation methods, several challenges remain. The computational complexity of tools such as PH makes these methods difficult to apply in large-scale settings. Efficient computation of PHs and critical points remains a bottleneck for the widespread adoption of these methods in practical applications [14, 147]. These methods often necessitate the use of smaller patches (e.g., 65×65) as they are computationally expensive with respect to the input patch dimensions, with complexity up to $O(n^3)$ [4]. For instance, Hu et al.'s method leverages persistent homology, where the matching of persistence diagrams requires $O(n \log n)$ and can reach $O\left(n^{3/2}\right)$ in some cases, depending on the complexity of the structures being analyzed [7]. The computation of persistence diagrams, which captures topological features across multiple scales, can involve cubic complexity $O(n^3)$ in certain steps, particularly when handling high-dimensional data [4]. While this method excels in capturing detailed topological information, it can be computationally intensive, specifically for large-scale datasets. Recent advances, such as the use of the Sinkhorn-Knopp algorithm to approximate Wasserstein distances in topological loss functions, have made strides in alleviating some of these computational challenges [24, 55, 174], but further optimization is needed to scale these techniques effectively.

Moreover, PH-based methods emphasize 1-dimensional features (i.e., loops) in the feature space and therefore primarily capture topological features such as loops and connected components but often overlook terminal nodes (i.e., dangling lines), which can be critical for certain applications. A bicycle wheel provides a useful analogy where one with a missing spoke is considered identical to one with a dangling spoke in PH-based methods, as both configurations share the same number of connected components and loops. However, from a structural perspective, the dangling spoke represents a crucial difference. This limitation of PH makes it less suitable for applications where disconnections, such as terminal nodes, are important. For instance, in datasets such as DRIVE [113] thin vessels often appear as dangling lines, not part of any loop, and thus receive less emphasis in PH-based methods.

In this paper, we focus on 0-simplices, the foundational elements of topological structures, and implement a workflow that preserves key connectivity and segmentation details without the computational burden of persistent homology (PH) computation, making it suitable for large-scale datasets [166]. To address the limitations of PH-based approaches, we introduce a novel, fast topology-aware loss function that provides a computationally efficient alternative to PH while enhancing segmentation performance. By focusing on convolution operations and iterative erosion, TopoSegNet achieves a more manageable time complexity of $O(kn^2)$, where k is the number of erosion iterations. This linear scaling ensures that our model remains computationally efficient, even for large datasets, without compromising performance. Our approach enforces topological consistency during training, ensuring that the segmented structures closely match the topology of the ground truth. Notably, our method is significantly faster than traditional PH methods, making it ideal for large datasets. Additionally, we propose a new evaluation metric, the Centroid Displacement Measure (CDM), which assesses topological and geometric similarity between predicted and ground truth masks, further validating the effectiveness of our approach.

4.5 Methods

4.5.1 Design

Building on the success of TopoSinGAN [175] in embedding topological awareness within GAN-based generative models, which underscores the importance of structural consistency in computer vision, we extend this topological focus to image segmentation with TopoSegNet. Unlike generative models, where topological awareness ensures connectivity and loop closeness in generated images, segmentation tasks benefit from the targeted preservation of structural features essential for accurately representing segmented objects. Our approach introduces a novel method for estimating the topological structure of binary masks, based on the observation that "not all pixels in a binary mask are equally important" for preserving topology. We define these essential pixels as *topologically critical points (TCPs)* including terminal nodes, junction points, and thinner regions of line features, all of which are prone to errors that can disrupt the mask's topology. We introduce a loss function designed to preserve these TCPs in predicted segmentation masks by aligning them with those of the ground truth.

Definition 2 (Topologically Critical Point)**.** *Let B be a binary mask representing a segmented structure, and let $f : \mathbb{R}^2 \to \{0, 1\}$ be the function mapping the image domain to the binary mask, where 1 represents the foreground and 0 represents the background. Define a set of critical pixels $C \subset B$ such that $p \in C$ if and only if the removal of p results in a topological change in the structure of B, such as the birth or death of connected components and loops. These points include terminal nodes, junction points, and thin regions of line features. The TCP set C can be mathematically formalized by identifying pixels p where:*

$$C = \{p \in B \mid \textit{removal of } p \textit{ changes the Betti numbers of } B\}$$

Here, Betti numbers refer to the number of connected components (Betti 0) and the number of loops or holes (Betti 1).

In other words, we shift our focus from traditional topological features, such as loops and connected components, to the TCPs, where the presence or absence of these points leads to the creation or elimination of such topological features. Figure 4.1 illustrates how the removal of each of these TCPs under the prediction function f can lead to a change in the original topological structure in three sample cases: the left panel shows a junction TCP, the middle panel demonstrates thin TCPs and the right panel presents a terminal TCP, all highlighted in red.

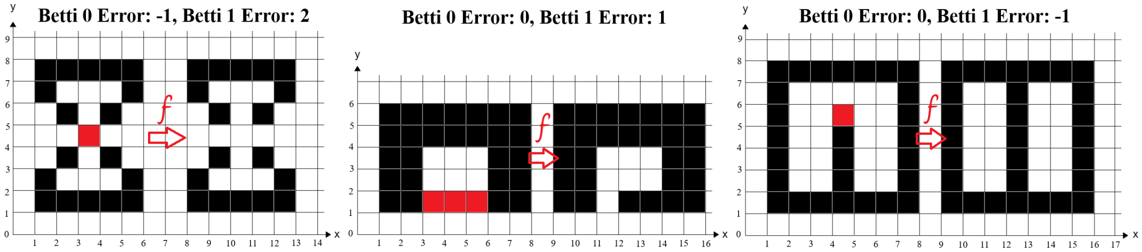


Figure 4.1: Illustration of topologically critical points (TCPs) and their impact on the topological structure through the prediction function f. The figure demonstrates three sample cases of TCPs, all highlighted in red: (Left) a junction TCP, (Middle) thin TCPs, and (Right) a terminal TCP. These TCPs represent key structural points where their presence or absence can result in significant changes to the topological features.

4.5.2 TCP extraction and erosion process

Our method employs eight distinct 3 × 3 kernels (Figure 4.2) to detect terminal nodes, thin regions, and junction points. These kernels are specifically designed to target the eight main cardinal and inter-cardinal directions in a 2D space (north, northeast, east, southeast, south, southwest, west, and northwest). Empirically, the use of these kernels has proven effective in identifying discontinuities, thereby enhancing the models ability to evaluate topological performance. This evaluation is integral to the core of the developed loss function, enabling the application of corrective penalties during the training phase for improved topological accuracy.

As illustrated in Figure 4.3, after convolving the binary mask T with each kernel, the resulting output T' is shifted and scaled by applying parameters $\gamma = 10$ and $\delta = 1$ to enhance feature detection. The shifted output T'' is then clamped to an upper bound $U = 10$, producing a clamped map C. This clamping is achieved using PyTorchs "clamp" function, which ensures that values exceeding U are restricted to this bound. Importantly, the "clamp" function is differentiable almost everywhere, allowing gradients to propagate during backpropagation, except at the boundaries where gradients are set to zero for values at the clamp limit [176]. Consequently, this clamping step integrates smoothly with gradient-based optimization techniques. Subsequently, C undergoes a sigmoid transformation to produce C', which is element-wise multiplied with the original mask T, yielding a TCP map specific to each kernel. Finally, these individual TCP maps are aggregated into a single comprehensive TCP map, which will then be used for computing topology loss for each prediction.

The TCP detection process consists of two phases: forward and inverted. In the forward phase, the binary mask directly goes through the TCP detection process for detecting endpoints and thin features. In the inverted phase, the binary mask is inverted, and the same TCP detection process is applied again, effectively detecting junction points that may not have been identified in the forward phase. By inverting the input mask (i.e., subtracting each pixels value from the masks maximum value), the junctions in the original mask form sharp edges, making them detectable by the TCP detection kernels. Each TCP map is then compared to both the ground truth (GT) and predicted maps and the loss is calculated by aggregating the computed loss per image. Additionally, after the convolution of the mask with the 3×3 erosion filter, junction points that may not be visible in one binary mask can be identified in the binary mask of the inverse structure.

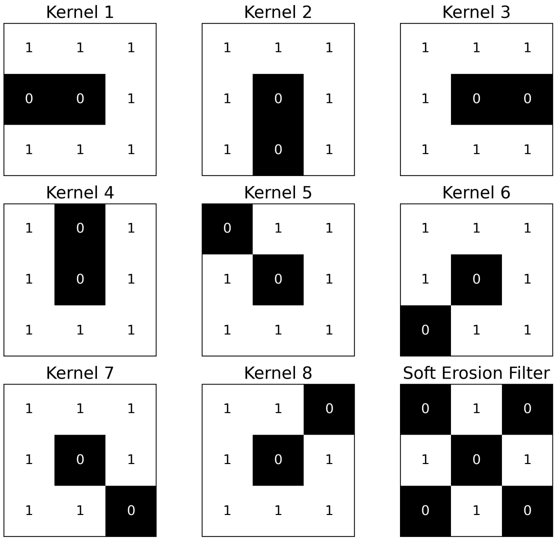
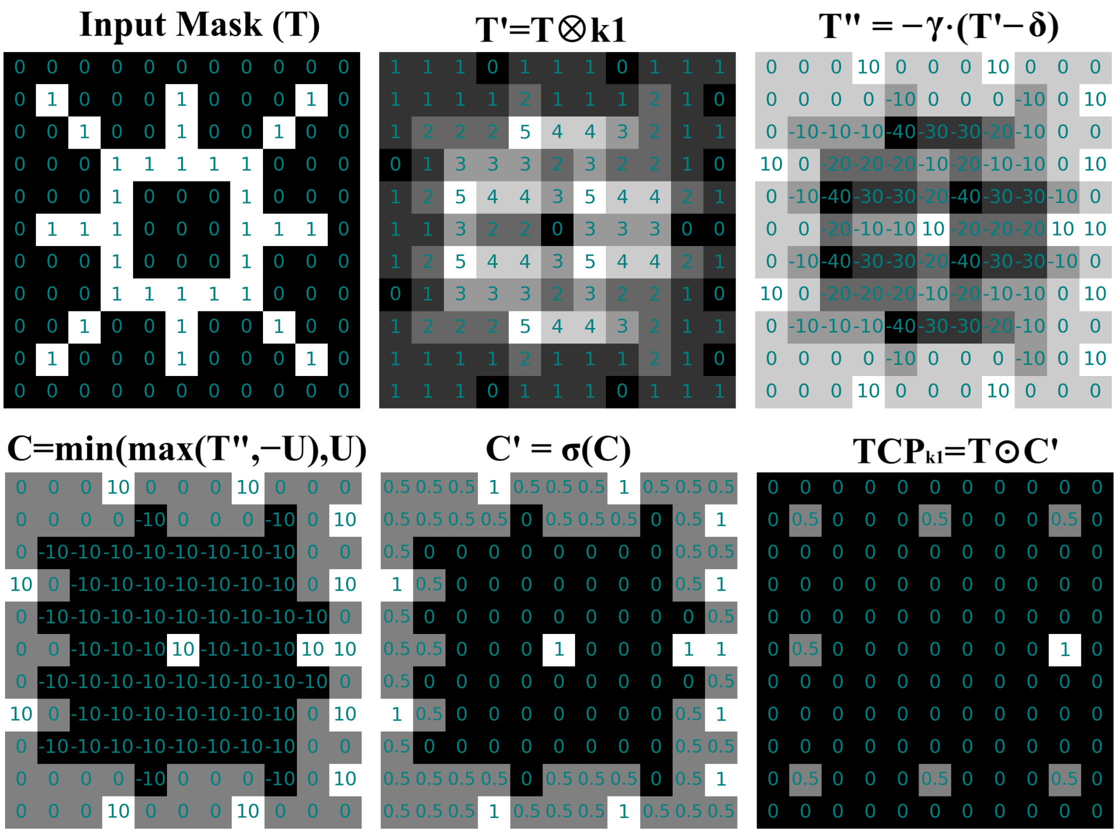


Figure 4.2: The visualization of the 3x3 TCP mapper kernels and the soft erosion filter.



Input Mask (T)

0	0	0	0	0	0	0	0	0	0	0
0	1	0	0	0	1	0	0	0	1	0
0	0	1	0	0	1	0	0	1	0	0
0	0	0	1	1	1	1	1	0	0	0
0	0	0	1	0	0	0	1	0	0	0
0	1	1	1	0	0	0	1	1	1	0
0	0	0	1	0	0	0	1	0	0	0
0	0	0	1	1	1	1	1	0	0	0
0	0	1	0	0	1	0	0	1	0	0
0	1	0	0	0	1	0	0	0	1	0
0	0	0	0	0	0	0	0	0	0	0

$T'=T\otimes k1$

1	1	1	0	1	1	1	0	1	1	1
1	1	1	1	2	1	1	1	2	1	0
1	2	2	2	5	4	4	3	2	1	1
0	1	3	3	3	2	3	2	2	1	0
1	2	5	4	4	3	5	4	4	2	1
1	1	3	2	2	0	3	3	3	0	0
1	2	5	4	4	3	5	4	4	2	1
0	1	3	3	3	2	3	2	2	1	0
1	2	2	2	5	4	4	3	2	1	1
1	1	1	1	2	1	1	1	2	1	0
1	1	1	0	1	1	1	0	1	1	1

$T'' = -\gamma\cdot(T'-\delta)$

0	0	0	10	0	0	0	10	0	0	0
0	0	0	0	-10	0	0	0	-10	0	10
0	-10	-10	-10	-40	-30	-30	-20	-10	0	0
10	0	-20	-20	-20	-10	-20	-10	-10	0	10
0	-10	-40	-30	-30	-20	-40	-30	-30	-10	0
0	0	-20	-10	-10	10	-20	-20	-20	10	10
0	-10	-40	-30	-30	-20	-40	-30	-30	-10	0
10	0	-20	-20	-20	-10	-20	-10	-10	0	10
0	-10	-10	-10	-40	-30	-30	-20	-10	0	0
0	0	0	0	-10	0	0	0	-10	0	10
0	0	0	10	0	0	0	10	0	0	0

$C=\min(\max(T'',-U),U)$

0	0	0	10	0	0	0	10	0	0	0
0	0	0	0	-10	0	0	0	-10	0	10
0	-10	-10	-10	-10	-10	-10	-10	-10	0	0
10	0	-10	-10	-10	-10	-10	-10	-10	0	10
0	-10	-10	-10	-10	-10	-10	-10	-10	-10	0
0	0	-10	-10	-10	10	-10	-10	-10	10	10
0	-10	-10	-10	-10	-10	-10	-10	-10	-10	0
10	0	-10	-10	-10	-10	-10	-10	-10	0	10
0	-10	-10	-10	-10	-10	-10	-10	-10	0	0
0	0	0	0	-10	0	0	0	-10	0	10
0	0	0	10	0	0	0	10	0	0	0

$C' = \sigma(C)$

0.5	0.5	0.5	1	0.5	0.5	0.5	1	0.5	0.5	0.5
0.5	0.5	0.5	0.5	0	0.5	0.5	0.5	0	0.5	1
0.5	0	0	0	0	0	0	0	0	0.5	0.5
1	0.5	0	0	0	0	0	0	0	0.5	1
0.5	0	0	0	0	0	0	0	0	0	0.5
0.5	0.5	0	0	0	1	0	0	0	1	1
0.5	0	0	0	0	0	0	0	0	0	0.5
1	0.5	0	0	0	0	0	0	0	0.5	1
0.5	0	0	0	0	0	0	0	0	0.5	0.5
0.5	0.5	0.5	0.5	0	0.5	0.5	0.5	0	0.5	1
0.5	0.5	0.5	1	0.5	0.5	0.5	1	0.5	0.5	0.5

$TCP_{k1}=T\odot C'$

0	0	0	0	0	0	0	0	0	0	0
0	0.5	0	0	0	0.5	0	0	0	0.5	0
0	0	0	0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0	0	0	0
0	0.5	0	0	0	0	0	0	0	1	0
0	0	0	0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0	0	0	0
0	0.5	0	0	0	0.5	0	0	0	0.5	0
0	0	0	0	0	0	0	0	0	0	0

Figure 4.3: Illustration of the TCP mapping process for input mask T using kernel 1 in Figure 4.2 (k_1). T' is the result of the convolution of T and k_1. T'' is obtained by shifting and scaling T' with parameters $\gamma = 10$ and $\delta = 1$. C is the clamped version of T'' with an upper bound $U = 10$. C' is the sigmoid-transformed C, and TCP_{k_1} is the result of element-wise multiplication of C' and the original mask T.

4.5.3 Loss function

We applied the erosion operation across three levels, with TCPs extracted at each level to enable the model to learn the topological structure at multiple granularities illustrated in Figure 4.5. The choice of three levels of erosion was based on our experimental results, and the number of iterations was adjusted according to the intrinsic characteristics of the label mask, such as the thickness of linear features. To assess the impact on model performance, we conducted experiments using both three and six levels of erosion across various datasets. Following the soft erosion process, the outputs need to be binarized before proceeding to the TCP detection stage. To ensure smooth integration into the computational graph while maintaining differentiability, we employed a soft binarization technique using a custom sigmoid function (Equation 4.1). This function, with a steep slope ($\alpha = 10$) and a threshold of $\beta = 0.5$, allows for precise separation of key features, ensuring the binarization process remains differentiable.

$$\sigma(x) = \frac{1}{1 + e^{\alpha(\beta - x)}} \tag{4.1}$$

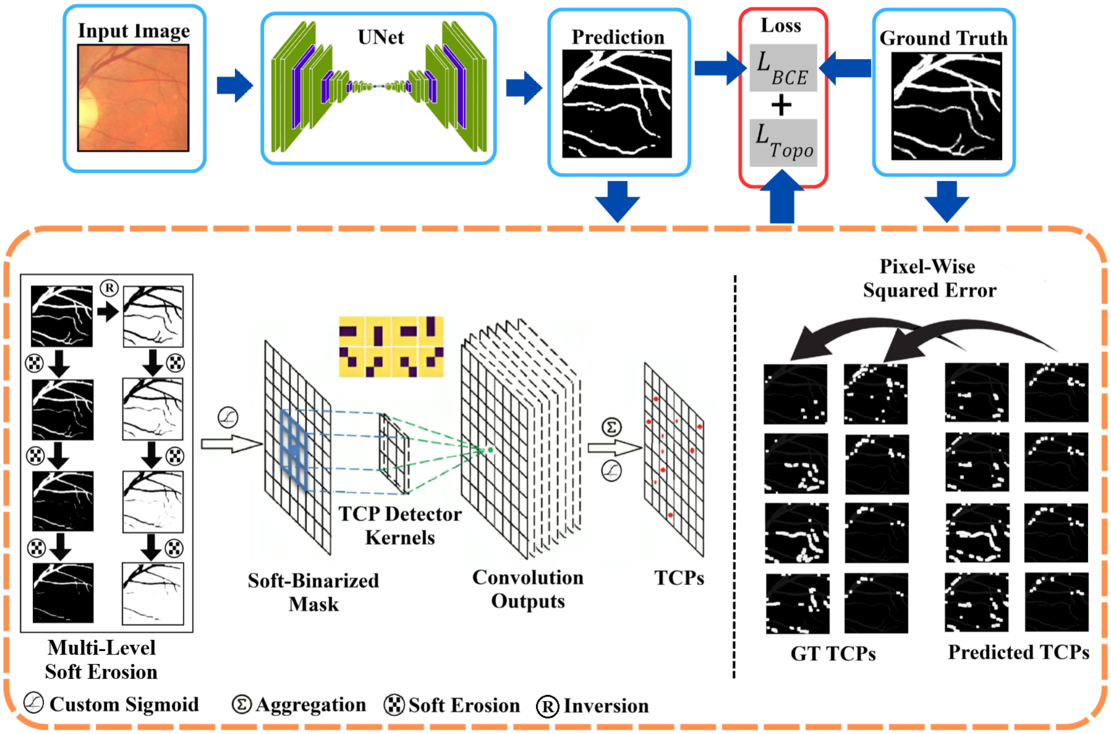


Figure 4.4: The overview of the proposed method.

The whole topological loss is differentiable and remains connected to the network's computational graph, allowing for end-to-end training of deep neural networks. Equations

4.1- 4.10 present the mathematical formulas for calculating this loss, where operations such as erosion and TCP detection are applied to both the ground truth and predicted masks. The workflow of the proposed method is depicted in Figure 4.4. In these equations GT_{E_0} and $Pred_{E_0}$ represent the original ground truth and predicted tensors, and GT_{E_i} and $Pred_{E_i}$ represent the i-th level of erosion of ground truth and predicted mask, respectively. Also, EF represents the erosion filter, and K_i are the TCP detector kernels shown in Figure 4.2, and $Inv\_GT_{E_i}$ and $Inv\_Pred_{E_i}$ are the i-th level of erosion of inverted ground truth and predicted mask, respectively.

$$Pred_{E_i} = \sigma(EF \otimes Pred_{E_{i-1}}), \quad GT_{E_i} = \sigma(EF \otimes GT_{E_{i-1}}) \tag{4.2}$$

$$Inv\_Pred_{E_0} = Max(Pred_{E_0}) - Pred_{E_0}, \quad Inv\_GT_{E_0} = Max(GT_{E_0}) - GT_{E_0} \tag{4.3}$$

$$Inv\_Pred_{E_i} = \sigma(EF \otimes Inv\_Pred_{E_{i-1}}), \quad Inv\_GT_{E_i} = \sigma(EF \otimes Inv\_GT_{E_{i-1}}) \tag{4.4}$$

$$\begin{aligned} TCP_{Pred} = &\sum_{i=0}^{8}\sum_{j=0}^{3}\left(C'_{i,j} \odot Pred_{E_j}\right), \\ &\text{and } C'_{i,j} = \sigma\left(clamp\left(-\gamma \cdot (K_i \otimes Pred_{E_j} - \delta), -U, U\right)\right) \end{aligned} \tag{4.5}$$

$$TCP_{GT} = \sum_{i=0}^{8}\sum_{j=0}^{3}\left(C'_{i,j} \odot GT_{E_j}\right), \quad \text{and } C'_{i,j} = \sigma\left(clamp\left(-\gamma \cdot (K_i \otimes GT_{E_j} - \delta), -U, U\right)\right) \tag{4.6}$$

$$\begin{aligned} Inv\_TCP_{Pred} = &\sum_{i=0}^{8}\sum_{j=0}^{3}\left(C'_{i,j} \odot Inv\_Pred_{E_j}\right), \\ &\text{and } C'_{i,j} = \sigma\left(clamp\left(-\gamma \cdot (K_i \otimes Inv\_Pred_{E_j} - \delta), -U, U\right)\right) \end{aligned} \tag{4.7}$$

$$Inv\_TCP_{GT} = \sum_{i=0}^{8}\sum_{j=0}^{3}\left(C'_{i,j} \odot Inv\_GT_{E_j}\right),$$
$$\text{and } C'_{i,j} = \sigma\left(clamp\left(-\gamma \cdot (K_i \otimes Inv\_GT_{E_j} - \delta), -U, U\right)\right) \tag{4.8}$$

$$TopoLoss = \sum (TCP_{Pred} - TCP_{GT})^2 + \sum (Inv\_TCP_{Pred} - Inv\_TCP_{GT})^2 \tag{4.9}$$

In Equations 4.5- 4.8, U is the limiting bound for the clamping operation that is set to 10. Also, γ and δ are scale and shift arguments that are set to 10 and 1, respectively. Finally, we add the calculated TopoLoss to the binary cross entropy loss with tunable hyperparameters λ_1 and λ_2 as shown below.

$$Loss = \lambda_1 \times BCELoss + \lambda_2 \times TopoLoss \tag{4.10}$$

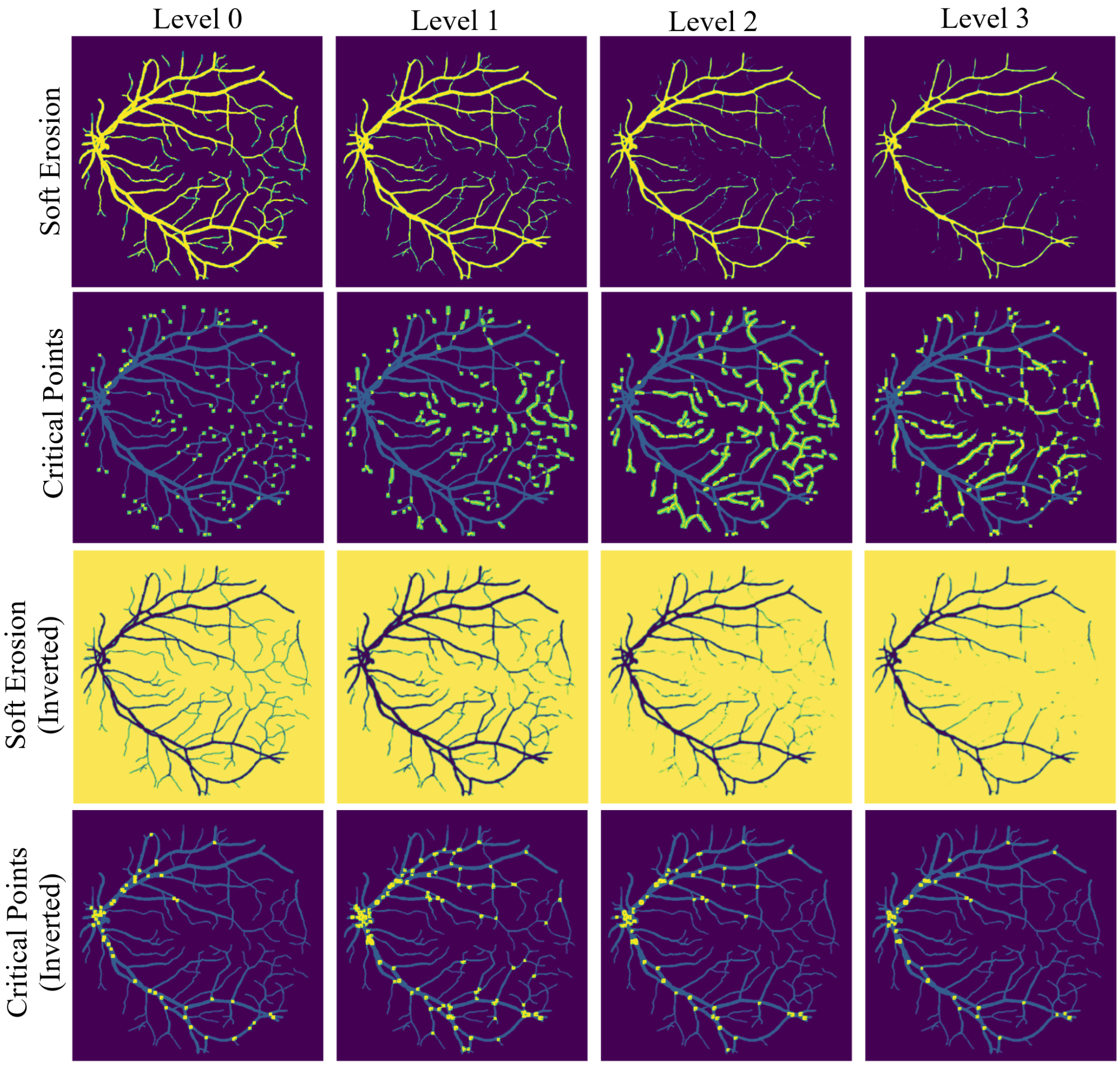


Figure 4.5: Visualization of tensors in the TCP mapping process. The first row shows the original tensor (Level 0) followed by three iterations of erosion. The second column presents the TCP maps generated from the convolution of the eight TCP detector kernels. The third row illustrates the TCP maps obtained from convolutions applied to the reversed masks. In this figure, the TCP maps are overlaid with the original mask for enhanced visualization, where bright pixels represent the detected TCPs.

4.6 Experiments

4.6.1 Datasets

We evaluate our method on four datasets: DRIVE [113], CREMI-B [112], GMP Tree rings [111], and agricultural fields in Tehran, Iran.

The DRIVE (Digital Retinal Images for Vessel Extraction) dataset is a benchmark dataset commonly used for evaluating the performance of retinal vessel segmentation methods [113]. The primary features of interest in the dataset are blood vessels, rather than loops or network-like structures. While there are some loops present, they are not the dominant characteristic of this dataset. Unlike PH-based methods that focus on loops, this provides an opportunity to assess how well our model performs on elongated vessel structures without significant emphasis on loops. In addition, this dataset includes line features with varying thicknesses which makes it a challenging dataset for accurately delineating the thinner parts through the learning process.

The CREMI-B dataset (Circuit Reconstruction from Electron Microscopy Images), part of the CREMI challenge for neuron segmentation in electron microscopy volumes, consists of 125 images captured from the adult *Drosophila* brain at a resolution of $4 \times 4 \times 40$ nm, with a physical size of $1250 \times 1250 \times 125$ pixels ($5 \times 5 \times 5$ μm^3). CREMI-B is particularly challenging due to its obscure cell boundaries and significant variations in cell size and shape, making it ideal for evaluating models on complex neuron structures. We selected CREMI-B for this study to push our model's ability to handle diverse geometric shapes and varying foreground feature areas. Also, this dataset contains line features with mostly constant thicknesses making it an appropriate case study for evaluating our model in which the thickness of the boundaries is considered an influential factor on the performance of our model.

We used a gigapixel macro photography (GMP) dataset consisting of 1,000 images, each measuring 175×240, from a larger collection by [111]. This ultra-high-resolution dataset (19,812 dpi) was manually cropped and annotated, with each image containing four channels: RGB and an annotated mask highlighting the xylem vessels in hardwood. These high-resolution dendrological images are crucial for studying tree age and environmental adaptations [177]. With around 750 vessels per 1 mm^2 of tree ring, manual annotation is difficult, highlighting the need for deep learning to streamline dendrology analysis. This dataset also contains line features with significant variations in terms of their thickness making it a good case study for evaluating our model.

Lastly, we utilized a shapefile of agricultural field boundaries, created through precise land surveys using Total Station equipment in Tehran province, located in the diverse landscape of Irans Middle East region. The study area spans from $50°25'10''$ W to $52°27'5''$ E and $35°43'38''$ N to $35°0'12''$ S, covering approximately 19,500 km^2. For the RGB channels, we employed 10-meter resolution remote sensing imagery from Sentinel-2. Segmentation analysis in this region is particularly challenging due to the distinct and irregular agricultural layouts. This heterogeneity in the geometrical and topological structure, coupled with the presence of uniformly thicker line features compared to the other datasets, adds complexity to the satellite image segmentation task. The dataset comprises 447 images, each sized 250×250 pixels with four channels, where the fourth channel represents a rasterized mask layer. Figure 4.6 provides a snapshot of the study area. For all case studies, we randomly selected 70% of the data for training, 15% for validation, and 15% for testing.

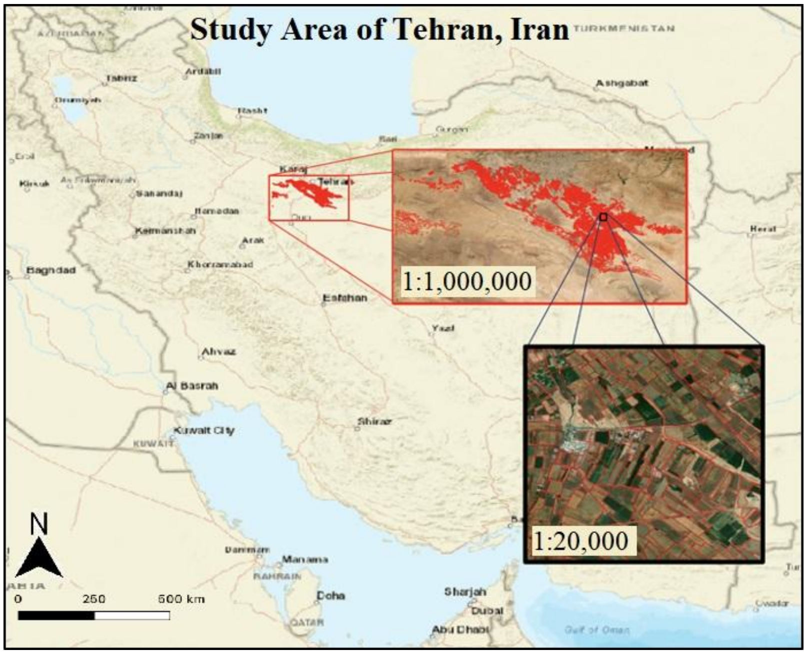


Figure 4.6: The study area of Tehran, Iran, and the field mask data in multiple scales.

4.6.2 Baselines and comparative models

We use UNet as the baseline model and implement our custom loss function within this framework. We deliberately chose UNet as our base architecture for several reasons. First, its encoder-decoder structure with skip connections provides a clear, interpretable framework that allows us to systematically demonstrate the impact of our topological innovations. Second, unlike more complex architectures, UNet's straightforward design enables us to isolate and evaluate the specific contributions of our topology-aware components. This methodological clarity is essential, as our primary goal is not to achieve state-of-the-art performance through architectural sophistication, but rather to demonstrate how topological awareness can enhance segmentation performance regardless of the underlying architecture. Finally, UNet's widespread use and well-understood behavior also makes it an ideal platform for showcasing our methodological innovations in topology preservation.

Additionally, we implement the topology loss function developed by Hu et al. (2019)

in the same UNet model [14]. For comparison, we evaluate three versions: the plain UNet, the UNet with Hu et al.'s topology loss, and our UNet model with the newly designed loss function with three levels of erosions. We also provide an additional experiment of the same model with 8 erosions to test the impact of the number of erosions in the developed algorithm. For reproducibility, all code is made available in our GitHub repository at https://github.com/mohsenumn/TopoSegNet.git.

4.6.3 Evaluation

Pixel-wise metrics

We employ standard evaluation metrics such as accuracy, precision, dice, recall, and Intersection over Union (IoU). While these metrics are widely used in image segmentation tasks, they rely on pixel-wise comparisons between the ground truth (GT) and predictions, which can be misleading in cases of significant class imbalance, such as the segmentation of retinal veins or thin linear structures [178]. The challenge arises because deep learning models may produce varying thicknesses and positions for these features, depending on the loss function, and often fail to align perfectly.

Topology-focused metrics

In datasets with class imbalance, such as in medical image segmentation or the delineation of agricultural fields from satellite imagery, the number of true positives is typically much smaller than true negatives. This can lead to inflated performance scores for metrics that heavily rely on true negatives, even if the model struggles with accurately segmenting the minority class [178]. To address this, there is a need for evaluation metrics that are less sensitive to class imbalance and more focused on structural accuracy. In our study, we incorporate Maximum Mean Discrepancy (MMD) [4] and Betti numbers to emphasize the topological properties of the segmentation masks. Betti 0 measures the difference in the number of connected components between the ground truth and predicted masks, which a negative Betti 0 indicates more connected components in the predicted mask. Betti 1 measures the difference in the number of loops between the ground truth and predictions, with a negative value indicating more loops in the predictions, suggesting a potential over-segmentation. Additionally, we introduce a new metric, the Centroid Displacement Measure (CDM), which evaluates both topological and geometric properties to provide a more nuanced assessment of how accurately the predicted structures align with the ground truth.

Centroid displacement measure (CDM) We introduce a new evaluation metric called the Centroid Displacement Measure (CDM), which evaluates segmented objects by considering both their geometric and topological features. CDM transforms a binary mask into a representation where loops are modeled as convex polygons and then calculates the centroid of each polygon. For each centroid in the ground truth mask, the distance to the nearest centroid in the predicted mask is measured. This process is repeated for all centroids, and the average of these distances serves as the evaluation metric. A lower CDM indicates higher segmentation accuracy. Figure 4.7 illustrates the application of CDM on two predictions of a GT GMP instance. CDM is particularly robust to variations in boundary thickness, as the centroid of a loop remains consistent when the thickness changes uniformly throughout the loop. During the polygonization stage, if there is an undesired break or opening in a loop within the predicted feature (indicating an error), the feature is merged with the nearest neighboring loop to form a convex shape. This merging process shifts the centroid of that feature away from its true position, resulting in a higher CDM, thus effectively penalizing inaccuracies. However, CDM has a limitation in its sensitivity to over-segmentation. When a prediction is heavily over-segmented and contains an excessive number of loops, there will likely be at least one centroid close to the primary centroid in the original ground truth. This proximity can result in a lower CDM, as the average displacement is reduced, giving a potentially misleading impression of segmentation accuracy. To address this sensitivity, CDM should be used in conjunction with Betti 1, which evaluates the number of loops in the prediction. This combined approach provides a more reliable assessment of segmentation quality.

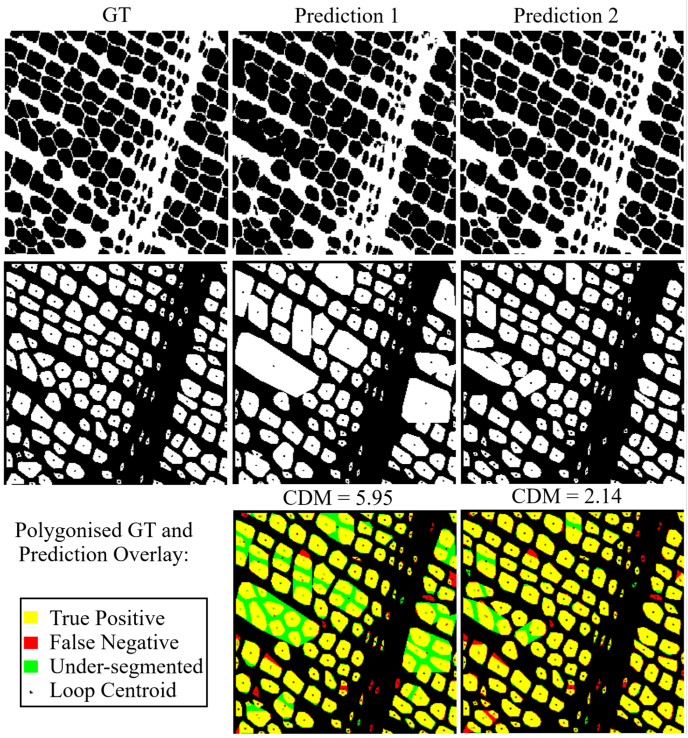


Figure 4.7: The illustration of the developed centroid displacement measure (CDM).

Maximum mean discrepancy Maximum Mean Discrepancy (MMD) evaluation metric is employed to measure the topological differences between GT and predicted masks in terms of their structural features [4]. MMD computes the distance between two distributions, in this case, the persistence diagrams of GT and predicted masks, by embedding them in a reproducing kernel Hilbert space. The goal is to assess how closely the predicted masks match the topological structure of the GT masks. Prior research indicates that sample-based MMD closely approximates its continuous counterpart [179]. For persistence diagrams, various kernels have been used in the literature [180]; here, we used the Gaussian kernel, predicated on the 1-Wasserstein distance as follows.

$$k_{w_1}(dgm_i, dgm_j) = \exp\left(-\frac{w_1(dgm_i, dgm_j)}{\sigma^2}\right) \tag{4.11}$$

In this equation, dgm_i is the persistence diagram for mask i and w_1 is the 1-Wasserstein distance function.

4.7 Results

The results of the experiments across four datasets of GMP, CREMI, DRIVE, and Fields are summarized in Table 4.1, comparing three models: a plain UNet, the model by [7], and our proposed model. The metrics evaluated include standard segmentation metrics such as accuracy, precision, recall, Dice, IoU, as well as more topology-focused metrics such as MMD, Betti numbers (Betti 0 and Betti 1), and the newly proposed CDM. The training times for each model are also reported. The sample outputs for the models are provided in Figure 4.8.

On the GMP dataset, our model achieved the highest scores across most metrics. Specifically, it improved accuracy to 97.03%, surpassing Hu et al.'s model (95.74%) and the plain UNet (94.07%). Precision, recall, Dice, and IoU also saw substantial improvements, with the Dice coefficient reaching 93.24%, significantly higher than both Hu et al.'s (89.3%) and the plain UNet (86.94%). For topology-based metrics, our model outperformed in MMD (0.34 ± 0.38 vs. 0.42 ± 0.26 for Hu et al.) and slightly improved CDM, reducing it to 1.09 ± 0.93 from Hu et al.'s 1.32 ± 1.01. Regarding Betti numbers, our model showed smaller Betti 0 values compared to the plain UNet (-11.68 ± 10.59 vs. -10.01 ± 33.05), while Hus model showed a much higher value (30.24 ± 13.49). For Betti 1, our model improved over the plain UNet (-1.15 ± 4.15 vs. 4.61 ± 17.66), though Hus model achieved the best score (0.32 ± 6.67). In terms of training time, our model completed training in 3 hours and 23 minutes, much faster than Hu et al.'s model (21 hours and 51 minutes), though slightly longer than the plain UNet (2 hours and 15 minutes).

For the CREMI dataset, our model also demonstrated competitive performance, achieving an accuracy of 98.51%, slightly better than Hu et al. (98.32%) and the plain UNet (98.48%). It performed better in recall (76.35%), Dice (78.46%), and IoU (64.73%), improving upon both the baseline and Hu et al.'s results. Our model significantly reduced the MMD score to 0.37 ± 0.26, indicating tighter alignment with ground truth topological features compared to Hu et al. (0.44 ± 0.19) and the plain UNet (0.43 ± 0.2). Similarly, CDM dropped to 22.38 ± 11.95, indicating better structural consistency. In terms of Betti numbers, our model performed slightly better in Betti 1 (44.22 ± 23.18), showing better handling of loops, while maintaining similar Betti 0 results. Importantly, our model required only 38 minutes to train, a significant reduction compared to Hu et al.'s training time (1 hour and 25 minutes).

On the DRIVE dataset, our model achieved an accuracy of 95%, a slight improvement over the plain UNet (94.69%) and Hu et al. (93.81%). Both precision and recall saw balanced improvements, with recall reaching 62.61%, compared to 58.1% for Hu et al. and 58.6% for the plain UNet. The Dice score of 75.24% and IoU of 60.39% were also higher than those of the competing models. Regarding topology-based metrics, our model had a lower CDM (21.41 ± 7.46) and better MMD performance (0.47 ± 0.09) than both Hu et al. and the plain UNet. The Betti 1 score was 14.87 ± 6.16, better representing topological loops compared to the plain UNet (17 ± 6.63) and comparable to Hu et al. (13.93 ± 9). Training time was also reduced to 25 minutes, faster than Hu et al. (51 minutes) and competitive with the plain UNet (15 minutes).

On the Fields dataset, which posed more challenging segmentation conditions, our model performed similarly to Hus model but showed significant improvements over the plain UNet. Our model achieved similar values to Hus in accuracy (82.17%), precision (40.86%), recall (46.69%), Dice (42.72%), and IoU (28.69%), with Hus model recording 82.19%, 41.12%, 46.9%, 42.74%, and 28.61%, respectively. In contrast, the plain UNet performed worse, with lower recall (37.27%), Dice (39.23%), and IoU (25.73%), though it had slightly higher accuracy (83.12%) and precision (44.66%). For topological metrics, although Hus model showed the best CDM (18.75 ± 21.91) and MMD (5.49 ± 9.05), our model recorded similar but slightly lower values (20.46 ± 25.75 for CDM and 5.95 ± 13.14 for MMD), still significantly outperforming the plain UNet (CDM 26.86 ± 28.28 and MMD 6.48 ± 9.74). Regarding Betti numbers, Hus model had the smallest difference in the number of components and loops (Betti 0: -2.1 ± 20.7, Betti 1: -3.13 ± 15.56), compared to our model (Betti 0: -25.67 ± 24.17, Betti 1: -6.82 ± 17.01) and the plain UNet (Betti 0: -64.15 ± 35.58, Betti 1: 8.03 ± 17.75). Our models training time of 35 minutes was a significant improvement over Hu et al.s 2 hours and 38 minutes.

We also evaluated the impact of increasing the number of erosions by doubling it to six levels across all datasets, with the results summarized in Table 4.2. Doubling the erosions caused model divergence in the CREMI dataset, but improved most of the standard pixel-wise metrics for the other datasets. In the GMP experiment, metrics such as accuracy, precision, recall, Dice, and IoU showed slight improvements, reaching 97%, 93.1%, 93.6%, 93.4%, and 87.71%, respectively. However, topological metrics declined, with CDM and MMD dropping to 1.34 and 0.37, respectively. The DRIVE dataset exhibited a similar trend, with pixel-wise metrics improving to 95.4%, 94.5%, 66%, 77.7%, and 63.59%, while CDM and MMD decreased to 22.64 and 0.61. For the Fields dataset, although accuracy and precision dropped to 80.5% and 38.3%, the IoU remained relatively stable at 28.63%, and both topological metrics of CDM and MMD improved, reaching 14.16 and 5.1, respectively. Interestingly, the average absolute number of connected components across all datasets also decreased (GMP: 0.62, DRIVE: 17.44, Fields: 74.54), suggesting a trend toward better topological consistency at the cost of some pixel-wise performance.

Table 4.1: Performance comparison of TopoSegNet vs. Hu et al. (2019), and Plain U-Net Models Across Datasets.

	Acc.	Prec.	Recall	Dice	IoU	CDM	MMD×100	Betti 0	Betti 1
GMP									
Simple	94.07	84.86	91.33	86.94	77.46	2.6±4.01	0.56±0.55	-2.57±7.26	3.54±14.21
Hu et al., 2019	95.74	85.51	**93.58**	89.3	80.7	1.32±1.01	0.42±0.26	1.59±1.38	2.89±7.01
Ours	**97.03**	**93.11**	93.41	**93.24**	**87.41**	**1.09±0.93**	**0.34±0.38**	**-0.11±1.26**	**-0.27±6.24**
CREMI									
Simple	98.48	81.76	73.6	77.39	63.25	40.23±21.85	0.43±0.2	-120.28±61.59	73.89±19.36
Hu et al., 2019	98.32	77.32	74.7	75.86	61.3	26.26±15.47	0.44±0.19	**-25.0±13.73**	**38.67±23.1**
Ours	**98.51**	80.72	**76.35**	**78.46**	**64.73**	**22.38±11.95**	**0.37±0.26**	-86.22±51.69	39.83±18.42
DRIVE									
Simple	94.69	**96.28**	58.6	72.81	57.3	25.98±11.01	0.51±0.14	-102.93±15.89	17±6.63
Hu et al., 2019	93.81	87.94	58.1	69.83	53.73	26.49±7.84	0.9±0.32	-22.5±15.4	**13.93±9**
Ours	**95**	94.38	**62.61**	**75.24**	**60.39**	**21.41±7.46**	**0.47±0.09**	**-59.8±8.57**	14.87±6.16
Fields									
Simple	83.12	44.66	37.27	39.23	25.73	26.86±28.28	6.48±9.74	-80.45±94.99	107.04±149.54
Hu et al., 2019	82.19	41.12	**46.9**	42.74	28.61	**18.75±21.91**	**5.49±9.05**	9.16±86.02	56.55±128.02
Ours	82.17	40.86	46.69	**42.72**	28.69	20.46±25.75	5.95±13.14	-21.51±85.56	**52.48±126.02**

Table 4.2: The performance of the TopoSegNet with doubling the number of erosion levels to 6.

	Acc.	Prec.	Recall	Dice	IoU	CDM	MMD× 100	Betti 0	Betti 1
GMP	97.0↑	93.1↑	93.6↑	93.4↑	87.71↑	1.34±1.21↓	0.37±0.87↓	-0.62±1.57↓	1.77±7.43↓
DRIVE	95.4↑	94.5↑	66↑	77.7↑	63.59↑	22.64±5.61↓	0.61±0.2↓	-17.44±4.7↓	17.62±11.76↓
Fields	80.5↓	38.3↓	51.4↑	42.8↑	28.63↓	14.16±12.5↓	5.1±8.18↑	-74.54±90.47↓	40.01±122.92↑
CREMI	–	–	–	–	–	Divergence	–	–	–

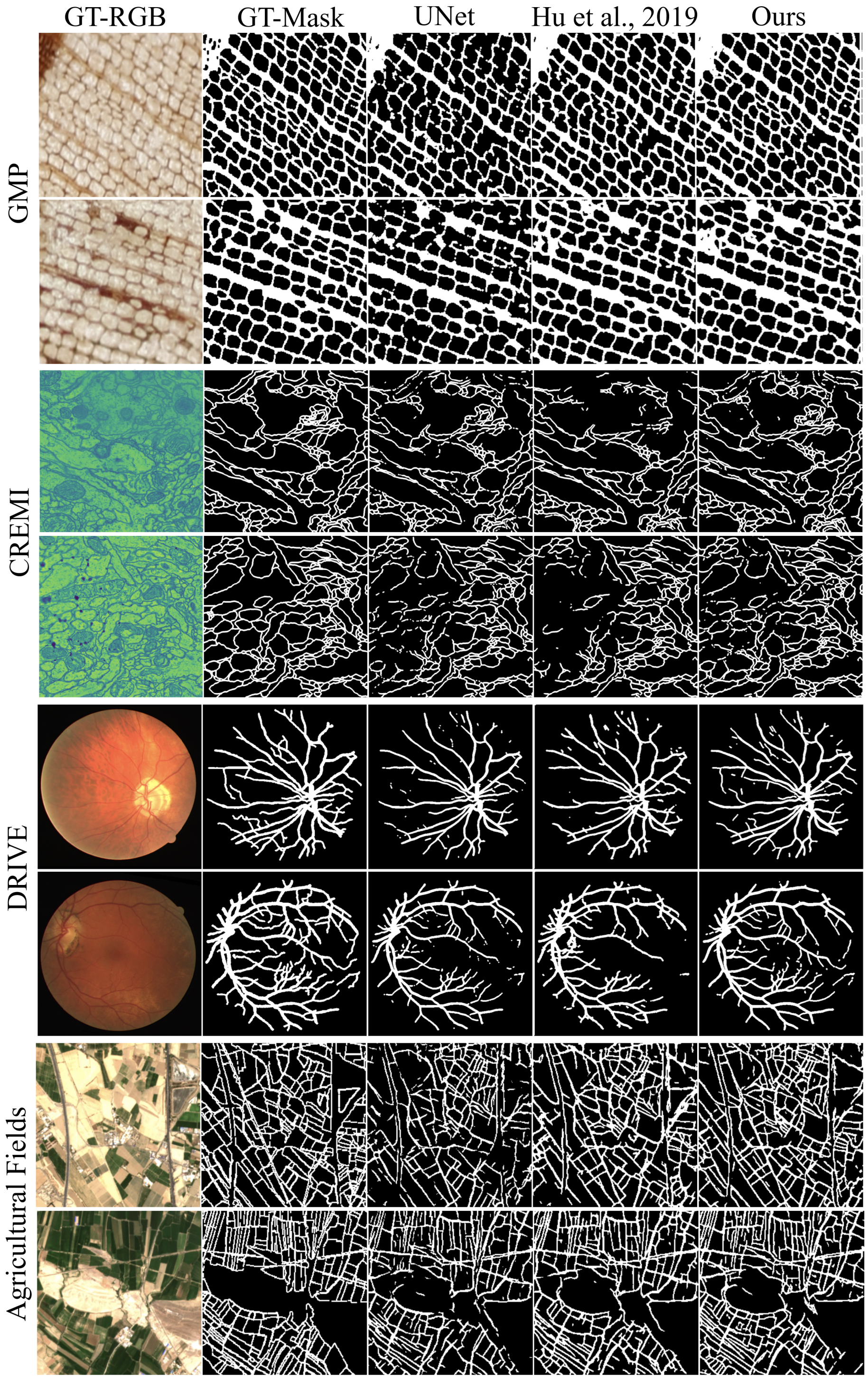


Figure 4.8: Performance comparison of our model against the baseline UNet and Hu et al. (2019) models on the test datasets.

4.8 Discussion

The objective of this research was to improve image segmentation by developing and evaluating a novel topology-aware loss function that focused on topological properties in binary masks while maintaining computational efficiency. Through extensive experiments across diverse datasets, including DRIVE, CREMI-B, GMP, and agricultural fields in Tehran, our proposed method consistently outperformed traditional models. Specifically, it showed improvements in both pixel-wise accuracy and topological consistency. The foundation of our approach lies in the integration of erosions and convolutions combined with the detection of topologically critical points (TCPs). We introduced a three-level soft-erosion strategy followed by TCP detection. The rationale behind using multiple erosion levels is to highlight the most "important" pixels in the binary mask. By focusing on these pixels at various erosion levels, the model gains insights into the persistence of the mask's topological structure under gradual transformations. This multi-level approach enables the model to identify critical points that play a crucial role in defining the topological features of the binary mask. As thinner parts of the mask erode, the model learns about the shifts in topological structures, allowing for a more nuanced understanding of the segmentation task.

Current deep learning-based segmentation methods with a topological focus typically rely on persistent homology (PH)–based losses. While these methods are effective at capturing topological features, they are often computationally expensive and difficult to scale [14, 147]. Consequently, our objective was twofold: first, to enhance segmentation accuracy, and second, to reduce computational complexity. While PH-based methods effectively capture the topological structure of binary masks, they have two key limitations that our method addresses. First, these methods are computationally intensive. The balance between computational efficiency and topological precision makes TopoSegNet highly suitable for real-time applications and large-scale image segmentation tasks, offering a practical and scalable alternative while maintaining competitive performance across various benchmarks. Second, While PH-based methods excel at capturing topological features, they tend to overlook critical elements such as terminal nodes. This limitation can be problematic in applications where disconnections or endpoints are important. Our approach effectively captures these subtleties, as evidenced by our experiments on the GMP and DRIVE datasets.

Uneven thickness in line features can present significant challenges in segmentation methods, as variations in thickness can lead to inconsistencies in how features are detected and preserved. The GMP and DRIVE datasets, for example, exhibit such challenges due to the uneven thickness of line features within their binary masks. In these cases, thin features are more likely to be missed in predictions. The DRIVE dataset, in particular, has a high number of terminal nodes (e.g., retinal vein endpoints) compared to loops, making PH-based methods less effective. As demonstrated in the results, our method uses three levels of erosion that successfully capture these thin features. In contrast, Hu's PH-based model performed better on the agricultural fields and CREMI datasets but struggled with GMP and DRIVE, where thin features are more prevalent. Referring to the average Betti 0 numbers reported in Table 4.1, Hus model often yielded positive values, indicating fewer connected components in the predictions compared to the ground truth masks. Our model, however, consistently produced negative values. Coupled with improved pixel-wise metrics, this suggests our approach was better at detecting thinner features, even when they appeared disconnected. Furthermore, our model improved the absolute value of Betti 0 compared to the plain U-Net, demonstrating improved connectivity of linear features and topological structures.

To assess the impact of erosion iterations, we doubled the number of erosion levels to six across all datasets. The results, presented in Table 4.2, reveal some important trends. While our model diverged on the CREMI dataset, performance on the other datasets generally improved in terms of pixel-wise metrics. However, topological accuracy decreased for the GMP and DRIVE datasets, while the Fields dataset showed improvements. These results are consistent with the characteristics of the datasets. GMP and DRIVE contain thin edges that quickly disappear after the earlier erosion levels, leading to poorer performance. In contrast, the agricultural fields dataset contains thicker, more uniform boundaries, where additional erosions helped in detecting TCPs more effectively. An added complexity arises when computing the loss over multiple iterations of erosion. With six erosion levels, our method calculates the loss at each level, both forward and reverse, resulting in 14 losses per image (6×2 for the eroded masks, plus two for the original mask). As the erosion levels increase, the weight assigned to the TCPs diminishes, reducing emphasis on thinner structures and shifting the focus to thicker ones that persist through the later stages of erosion. The improvement in pixel-wise metrics during the six-erosion experiments can be attributed to this focus on the thicker, more persistent features. Figure 4.9 illustrates a sample prediction for a DRIVE image, comparing results from three and six erosion levels. This pattern is further reflected in the absolute Betti 0 values, which consistently decreased across all experiments, indicating that thinner features were often missed as the number of erosions increased. The CREMI datasets divergence further supports this observation, as its uniformly thin boundaries were lost during the earlier erosion stages, leading to training instability.

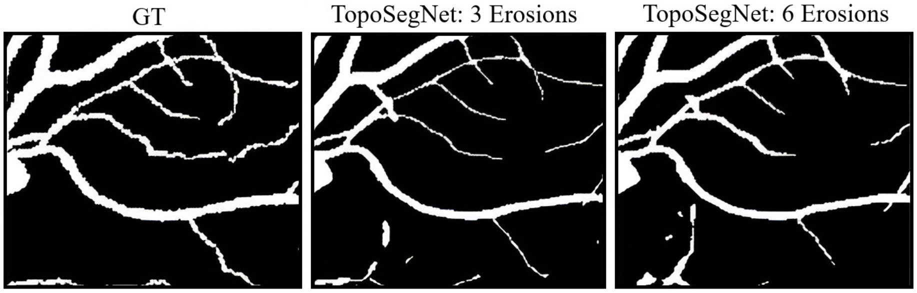


Figure 4.9: A sample test prediction with 3 and 6 erosion iterations.

Our results highlight an important trade-off between commonly used metrics such as Dice and IoU, and other metrics including precision and accuracy, especially in cases with class imbalance or fine, intricate structures. Dice and IoU prioritize overlap between predicted and ground truth regions, often favoring recall at the expense of precision. This tendency can lead to more false positives, especially in scenarios where the predicted regions extend beyond the true class boundaries. Consequently, while Dice and IoU may improve, precision and accuracy can suffer. This phenomenon was particularly evident in our experiments with topology-aware losses. By emphasizing connectivity and structural correctness, the model sometimes overestimated the thickness of linear features, leading to a higher false positive rate and, in turn, lower precision and accuracy scores. This trade-off was especially noticeable in Hus model and our model when we increased the erosion levels. These findings also underscore the limitations of traditional pixel-wise metrics and highlight the importance of incorporating topological evaluation metrics, such as CDM and MMD, for a more comprehensive assessment.

While our model has demonstrated notable improvements, several limitations and challenges remain. One key challenge is its sensitivity to the thickness of linear features and the number of erosions used. As seen in the six-erosion experiments, using too many erosions can lead to diminishing returns, with fine details being lost. This suggests that there is an optimal number of erosions that balances pixel-wise accuracy and topological performance. For datasets with varying feature thicknesses, careful selection of erosion levels is critical. Another limitation is the model's applicability to datasets where the foreground consists of the areas inside the loops rather than the boundaries. A potential remedy could be converting label masks to boundary representations using edge detection techniques, but this approach is outside the scope of this study. Additionally, detecting junction points remains a challenge, especially for rounded structures. Our method relies on the inversion of masks for TCP detection, which may lead to missed junctions when the mask structures are less sharp. Addressing these limitations will be the focus of future research.

4.9 Conclusion

In this study, we introduced TopoSegNet, a novel topology-aware loss function designed to enhance the accuracy and computational efficiency of image segmentation. We showed that not all pixels in a binary mask are equally important for preserving topology. Our method addresses the limitations of traditional pixel-wise loss functions, which often fail to capture topological consistency, especially in applications involving thin, elongated, or complex structures such as xylem cells, irregular agricultural fields, and blood vessels. To circumvent the complexities associated with persistent homology (PH)-based methods, which focus on topological features such as loops and connected components, we shift our attention to topologically critical points (TCPs), such as junctions and terminal nodes, that are crucial in forming these structures. By incorporating TCPs into the segmentation process using a multi-level erosion strategy, TopoSegNet not only improves segmentation accuracy but also preserves the structural integrity of the output. Through extensive experiments across diverse datasets, including medical imaging, satellite imagery, and gigapixel macro photography of tree rings, TopoSegNet demonstrated significant improvements in both standard segmentation metrics, such as Dice and IoU, and topology-focused metrics, including MMD and the newly developed CDM. Our approach supports topological correctness while remaining computationally efficient, making it suitable for large-scale applications. Despite these promising results, TopoSegNet faces challenges, particularly in its sensitivity to the thickness of line structures and the need for fine-tuning the number of erosion levels for optimal performance. Future work will focus on refining the loss function and addressing these limitations to further improve its robustness and versatility.

Chapter 5

Training Set Augmentation with TopoSinGAN: Impact of Synthetic Data on Image Segmentation

5.1 Abstract

Data scarcity remains a significant challenge in training image segmentation models, particularly in fields such as medical imaging and remote sensing, where annotated datasets are costly and difficult to acquire. Generative Adversarial Networks (GANs) have emerged as a powerful tool for data augmentation, but their reliance on large labeled datasets and potential for introducing artifacts limit their effectiveness. This study explores the application of TopoSinGAN, a topology-aware generative model capable of synthesizing structurally consistent images from a single source in data augmentation tasks. By augmenting datasets for agricultural field segmentation and gigapixel dendrological imagery, we demonstrate the impact of TopoSinGAN-generated data on improving the performance of segmentation models, including UNet and the topology-aware TopoSegNet. Our experiments reveal substantial gains in both pixel-wise and topology-aware metrics, with Intersection over Union (IoU) improvements of up to 2%, emphasizing the models ability to enhance segmentation accuracy while preserving structural integrity. This approach offers a scalable, efficient solution for augmenting datasets in domains requiring

topological precision.

5.2 Introduction

Image segmentation is a crucial task in various fields, such as medical imaging, remote sensing, and environmental monitoring, where the accurate delineation of object boundaries directly impacts downstream analyses and decision-making [178]. However, training segmentation models require large, high-quality annotated datasets, a need that poses significant challenges due to the labor-intensive and costly nature of data labeling. In domains such as medical imaging, where patient privacy and expertise requirements further constrain data availability [181], and in remote sensing, where diverse, high-resolution images complicate annotation processes [182], innovative solutions are necessary to overcome these limitations. Generative Adversarial Networks (GANs) have emerged as a powerful tool for data augmentation in image segmentation tasks, addressing data scarcity by synthesizing realistic images that expand the diversity and volume of training datasets [183]. GANs, consisting of a generator and discriminator in competitive learning, have been effectively applied to medical imaging for generating synthetic MRI, CT, and ultrasound scans [184, 185], thereby enhancing segmentation model performance on limited datasets. Similarly, in remote sensing, GANs have been used to generate high-resolution synthetic satellite and aerial imagery [186, 187], improving model robustness in diverse environments. By simulating variations in real-world scenes, GAN-augmented datasets enable segmentation models to generalize better across a broader range of conditions [188]. Despite their potential, conventional GAN models introduce challenges that limit their effectiveness in certain applications [189]. GANs typically require a substantial amount of labeled training data to generate high-quality synthetic images, creating a paradox where data augmentation methods still depend on large annotated datasets. Moreover, GAN-generated images may contain errors, including artifacts and topological inaccuracies that compromise the quality of the synthetic data [183]. In segmentation applications requiring topological precision, such as identifying vascular structures in medical images [190] or extracting road networks from satellite data [46], even minor topological errors in synthetic images can propagate into model training, degrading segmentation performance in critical areas. To address these challenges, this study explores the application of the TopoSinGAN model [175], which synthesizes topologically accurate images from a single source image, thereby alleviating the need for large labeled datasets while ensuring topological integrity in the generated data. We further apply TopoSinGAN-augmented datasets to train UNet and TopoSegNet, a topologically aware segmentation model, to assess their performance in applications where topological accuracy is essential. We chose UNet as the baseline architecture to maintain methodological clarity. While more sophisticated architectures such as SAM [76] might achieve higher absolute performance, our focus is on demonstrating the effectiveness of topological constraints and data augmentation strategies. UNet's transparent architecture allows us to clearly isolate and measure these effects without the confounding factors introduced by more complex models. This choice aligns with our research objective of advancing methodological understanding rather than competing for state-of-the-art performance metrics. By leveraging TopoSinGAN, we aim to bridge the gaps in current GAN-based data augmentation approaches, providing a robust, scalable solution for high-precision segmentation in data-constrained domains.

5.3 Background and Literature Review

The application of Generative Adversarial Networks (GANs) for expanding training datasets in image segmentation has emerged as a transformative approach across various fields [183, 189], notably medical imaging and remote sensing. Both domains face significant challenges regarding limited annotated data, diverse imaging characteristics, and high-dimensional information, which complicate the training of robust segmentation models [181, 182]. GANs, composed of a generator and discriminator in a competitive learning setup, have proven effective in generating synthetic images that closely resemble real ones [184, 185]. These synthetic images serve as data augmentations, enriching the training datasets and enhancing model performance in diverse applications, from medical diagnostics [191, 192] to environmental monitoring [186, 188].

5.3.1 Medical Imaging Applications of GANs

In medical imaging, GANs play a critical role in addressing the scarcity of annotated datasets [183]. The demand for labeled data in medical fields is particularly challenging due to the expertise required for accurate annotation and privacy concerns related to patient data [181, 184]. GANs offer a viable solution by creating synthetic images that closely resemble real medical scans [185]. These generated images expand the training datasets and support model generalization [191], making models more adaptable to the variability in medical imaging data, such as differences in imaging modalities, patient demographics, and imaging equipment [190, 192].

One advantage of using GANs in medical image segmentation is their capacity to generate high-fidelity synthetic images that help overcome the limited availability of annotated medical data. [191] leveraged a conditional GAN (cGAN) to generate additional samples for breast tumor segmentation in ultrasound images, which significantly improved the models accuracy by incorporating EfficientUNet into the framework to enhance spatial consistency. Similarly, [181] utilized a 2.5D patch-based GAN model to segment multiple organs in thoracic CT images, illustrating the ability of GANs to handle intricate anatomical structures and improve segmentation accuracy. GANs not only address the issue of dataset scarcity but also enable model generalization across various imaging modalities, as demonstrated by [185], who applied CycleGANs to enhance CT segmentation tasks across different datasets and imaging conditions, effectively mitigating the adverse effects of dataset shifts.

Beyond data augmentation, GANs also support ethical data sharing in medical research [184] showed that GANs could generate synthetic radiographs for segmentation tasks, enabling data sharing without compromising patient privacy. This ethical benefit is vital in medical imaging, where concerns around data privacy often limit the availability of real patient data. By generating synthetic datasets that can be shared and utilized without privacy concerns, GANs contribute to advancing medical research while maintaining ethical standards.

The flexibility of GANs extends across various imaging modalities, enabling segmentation applications in specialized areas. For example, [190] employed GAN-based algorithms to segment corneal subbasal nerves using confocal microscopy, showcasing the adaptability of GANs to different imaging techniques. The ability to train segmentation models across different types of medical images, from CT and MRI to confocal microscopy, demonstrates the robustness of GANs in handling diverse medical applications. Moreover, [193] highlighted the ability of GANs to overcome challenges related to small datasets in biomedical imaging. Their survey emphasized that GANs not only increase dataset size but also enable the development of more sophisticated segmentation models that can achieve high performance with limited annotated data.

GANs have also contributed to advancing segmentation model architectures by integrating with other deep learning frameworks. [184] employed GAN-based synthetic augmentation for liver lesion classification, where GAN-generated samples supported the learning process of convolutional neural networks (CNNs), demonstrating a synergistic approach that leverages the strengths of both GANs and CNNs for improved performance. [192] introduced a GAN framework that simultaneously performed segmentation and image synthesis, streamlining the training process and enhancing segmentation performance by exploiting the complementary roles of generators and segments.

5.3.2 Remote Sensing Applications of GANs

The application of GANs in remote sensing is similarly promising, given the challenges associated with limited labeled data and the complex, high-resolution nature of remote sensing images [182, 188]. The need for robust segmentation models in this field is driven by the variety of tasks that benefit from precise image segmentation, including environmental monitoring, urban planning, and disaster response [46, 194]. GANs address these challenges by creating synthetic images that simulate real-world scenarios [186, 187], allowing segmentation models to generalize effectively across diverse remote sensing environments [195, 196].

One of the main advantages of GANs in remote sensing is their capacity to produce high-resolution synthetic images that enhance the diversity of training datasets. [186] proposed the Relative Generation Adversarial Network (REL-SAGAN) to augment the NWPU remote sensing image dataset, achieving a 3% improvement in classification accuracy, a notable increase over traditional data augmentation techniques. Such findings highlight the utility of GANs in enriching datasets for remote sensing tasks that require segmentation models to handle the complexity of natural and urban scenes.

GANs have also been effectively integrated with other deep-learning models to improve segmentation accuracy in remote sensing. [182] developed a Wasserstein GAN-based method for high-resolution remote sensing image segmentation, demonstrating GANs' potential to refine CNNs' feature extraction capabilities in complex environments with diverse object appearances. This integration underscores the pivotal role of GANs in enhancing segmentation performance, especially in remote sensing scenarios where traditional augmentation methods may fall short. Specific segmentation tasks in remote sensing also benefit from GANs. [46] employed a GAN framework to address road network extraction, combining a generator to identify road structures with a discriminator to differentiate between generated and actual labels, thereby improving segmentation accuracy. This adaptability of GANs to task-specific challenges illustrates their versatility in handling the diverse requirements of remote sensing segmentation applications.

GANs also support labeled data generation, which is crucial for training supervised models in remote sensing. Ma and colleagues [195] introduced SiftingGAN to generate labeled samples for remote sensing image classification, demonstrating that GANs can effectively mitigate the issue of limited labeled data, a common bottleneck in model training. In their work, they proposed a GAN and a Fully Convolutional Network (FCN) framework that efficiently segments remote sensing images by generating additional labeled data, aiding in data-scarce scenarios where labeling is costly and time-intensive[195] introduced SiftingGAN, a model that generates and filters labeled samples to address the challenge of limited labeled data in remote sensing, thereby improving classification model performance[194] observed that neural network-based segmentation methods outperform traditional unsupervised methods, especially when trained on large datasets. Using GANs to augment these datasets exposes segmentation models to a broader range of scenarios, enhancing generalization and accuracy.

Furthermore, the integration of attention mechanisms within GAN frameworks, such as the Skip Attention GAN by [187], has shown improvements in capturing structural features in remote sensing images, resulting in higher-quality synthetic images that enhance segmentation tasks. In addition to supervised learning, GANs have facilitated semi-supervised learning in remote sensing segmentation. [189] employed a semi-supervised

GAN-based method for segmentation, using GAN-generated data to augment limited labeled datasets and improve classification accuracy. This approach is particularly relevant in remote sensing, where obtaining labeled data is often labor-intensive and costly. The advancements in GAN architectures for remote sensing have also contributed to their effectiveness. [188] developed the Deeply-supervised GAN (D-sGAN), designed to augment remote sensing datasets by generating images rich in detail, supporting semantic interpretation tasks critical to remote sensing analysis.

Despite the advantages of GANs in data augmentation for image segmentation, several gaps remain in this field [183, 189]. Notably, GAN models themselves typically require large, high-quality training datasets, which paradoxically reintroduces the challenge of limited annotated data [184]. This dependency limits GAN applicability in data-scarce environments. Additionally, GAN-generated images may contain errors, including inaccuracies that could degrade model performance [185]. In applications demanding topological accuracy, such as remote sensing [46, 182] and medical imaging [190, 191], the presence of topological errors in synthetic images risks propagating these inaccuracies into downstream tasks, potentially affecting the reliability of segmentation outputs [183].

In this study, we address these gaps by applying the TopoSinGAN model, which requires only a single image to generate an arbitrary number of topologically accurate synthetic images, ensuring topological correctness in the generated data. To further improve segmentation performance, we integrate these images into training pipelines for plain UNet and TopoSegNet, a model designed with topological awareness. This approach mitigates the limitations of data scarcity and error propagation, advancing segmentation accuracy in applications where topological precision is critical.

5.4 Methods

5.4.1 Synthetic Data Generation Using TopoSinGAN

TopoSinGAN [175] builds on the single-image generative capabilities of the SinGAN framework but incorporates a topology-aware loss function to ensure the topological coherence of generated images. The model is trained on a single base image using a hierarchical multi-scale approach, progressively generating synthetic variations across increasing resolutions. This architecture allows for the creation of multiple synthetic images from a single base image, while the topology loss ensures structural accuracy by minimizing terminal node counts in generated boundaries. This is critical for applications where maintaining connectivity and structural fidelity is essential, such as agricultural field delineations and cellular structures in dendrology. TopoSinGAN has been validated in prior studies, demonstrating significant improvements in topological accuracy through its Node Topology Clustering (NTC) index and reduced Fréchet Inception Distance (FID) scores compared to traditional SinGAN. The model was adapted for this study to generate diverse synthetic datasets tailored to the segmentation tasks.

5.4.2 Segmentation Models

Two segmentation models were employed to evaluate the impact of synthetic data:

1. **UNet [6]:**
   This model, widely used for semantic segmentation, follows an encoder-decoder architecture with skip connections to preserve spatial information across layers. UNet optimizes pixel-wise metrics such as accuracy and IoU but lacks mechanisms for maintaining structural and topological consistency.

2. **TopoSegNet (explained in Chapter 4):**
   Designed specifically for topology-aware segmentation, TopoSegNet incorporates a novel loss function based on topologically critical points (TCPs). These points include junctions, terminal nodes, and thin features, which are prone to errors that can disrupt structural integrity. By explicitly preserving these features during training, TopoSegNet achieves improved performance in both pixel-wise and topology-focused metrics.

5.4.3 Study Design

This study evaluates the impact of synthetic data generated by TopoSinGAN on the semantic segmentation performance of UNet and TopoSegNet across two datasets: agricultural fields and gigapixel dendrological (GMP) images. The experimental design involved systematically varying the proportion of synthetic data and the number of base images used to generate these synthetic datasets. The dataset proportions were set at 10%, 25%, 50%, and 75% synthetic data, with the total training dataset size fixed at 245 images for GMP and 313 images for the agricultural dataset, excluding the validation and test datasets. For both validation and testing, all real images not used for generating synthetic data were employed, ensuring that only real data was used for model evaluation. Models were trained using combinations of synthetic and real data, and their performance was evaluated against baseline models trained exclusively on real data.

The size of the real dataset (R) was fixed, while the total dataset size (D) and the number of synthetic images (F) varied based on the synthetic proportion (P). The total dataset size (D) was determined as D = round(R / 1 - P), ensuring consistency with the fixed size of the real dataset while rounding D to the nearest integer. The synthetic dataset size (F) was then computed as F = D - R. The number of synthetic images generated per base image (S) was calculated as S = round(F / B), where B represents the number of base images. Table 5.1 illustrates the study design for GMP case study.

Table 5.1: The study design for GMP case study.

Base Image Count (B)	**Fake Dataset (F)**		**Real Dataset (R)**		**Total Size (D)**
	%	**Count**	**%**	**Count**	
0	0%	0	100%	245	245
1	10%	27	90%	245	272
1	25%	82	75%	245	327
1	50%	245	50%	245	490
1	75%	735	25%	245	980
2	10%	$13 \times 2 = 26$	90%	245	271
2	25%	$41 \times 2 = 82$	75%	245	327
...	...	...	...	...	...
9	50%	$27 \times 9 = 243$	50%	245	488
9	75%	$82 \times 9 = 738$	25%	245	983
10	10%	$3 \times 10 = 30$	90%	245	375

Continued on next page

Table 5.1 continued

Base Image Count (B)	Fake Dataset (F)		Real Dataset (R)		Total Size (D)
	%	Count	%	Count	
10	25%	$8 \times 10 = 80$	75%	245	325
10	50%	$24 \times 10 = 240$	50%	245	485
10	75%	$73 \times 10 = 730$	25%	245	975

5.4.4 Datasets and Experimental Procedure

Two datasets were used for the experiments (further details in Section 4.6.1):

1. **Agricultural Fields Dataset:**
   This dataset consisted of 10-meter resolution satellite imagery depicting agricultural fields in Tehran Province, Iran. The images featured irregular field geometries, making them challenging for segmentation. A manually annotated mask delineating field boundaries was appended to each image as the ground truth.
2. **Gigapixel Dendrological Dataset (GMP):**
   This dataset featured ultra-high-resolution gigapixel images of tree cross-sections, with annotations highlighting cellular structures within xylem vessels. These complex structures required precise segmentation to maintain fine-scale connectivity, posing a unique challenge for the models [111].

For each configuration, synthetic images were generated using TopoSinGAN and combined with real images as described in Table 5.1 to form the training dataset. Models were trained using these datasets with consistent hyperparameters to ensure fair comparisons. The test datasets, containing only real images, remained constant across all configurations.

5.4.5 Evaluation Metrics

Performance was evaluated using a combination of pixel-wise and topology-aware metrics to ensure a comprehensive assessment of the segmentation models. Pixel-wise metrics focused on the overlap between predicted and ground truth segmentations, while topology-aware metrics captured the structural integrity and topological consistency of the predictions.

Pixel-Wise Metrics

Pixel-wise evaluation metrics for image segmentation tasks included precision, recall, and Intersection over Union (IoU), which collectively provide a robust assessment of model performance [178]. Precision measures the proportion of true positive pixels among all pixels classified as positive, reflecting the model's ability to minimize false positives. Recall evaluates the proportion of true positive pixels among all actual positive pixels, indicating the model's capacity to capture relevant features [183]. IoU quantifies the overlap between predicted and ground truth masks, defined as the ratio of their intersection to their union, offering a balanced perspective on segmentation accuracy [178, 196]. However, in datasets with class imbalances such as medical image segmentation or delineating agricultural fields from satellite imagery, true positives are often greatly outnumbered by true negatives. This imbalance can lead to inflated performance metrics that disproportionately reflect true negatives, even when the model struggles to segment the minority class effectively. Due to this limitation, accuracy was excluded from our evaluation metrics, as it could provide misleading insights in such scenarios.

Topology-Aware Metrics

To capture the structural fidelity of the segmentation outputs, topology-aware metrics were employed, evaluating the alignment of predicted masks with the ground truth in terms of connected components and loops [26, 35]. Betti numbers were used as key indicators of topological accuracy, with Betti0 measuring errors in connected components by reflecting discrepancies in the number of distinct regions between the predicted and ground truth masks, and Betti1 assessing errors in loops by indicating deviations in the number of enclosed structures [98]. These metrics are critical for ensuring the topological consistency of segmented outputs, particularly for datasets with fine-scale structures such as dendrological vessels or agricultural boundaries.

Centroid Displacement Measure (CDM)

The Centroid Displacement Measure (CDM) is a topology-aware metric specifically designed to quantify the geometric and topological alignment of predicted masks with ground truth masks by evaluating the spatial displacement of key structural features, particularly the centroids of connected components. For each connected component in both the predicted and ground truth masks, centroids are identified as the geometric center of the region. The Euclidean distance between each centroid in the ground truth mask and the nearest centroid in the predicted mask is then calculated, and the average of these distances is used as the CDM value. Lower CDM values indicate enhanced topological accuracy, reflecting minimal misalignment between predicted and ground truth components. Unlike Betti numbers, which focus on errors in connected components and loops, CDM provides a complementary geometric perspective by capturing the spatial arrangement of structures, ensuring that predictions preserve not only topological integrity but also spatial coherence. This metric complements Betti numbers by providing a geometric perspective on topological consistency. For further details please see Section 4.6.3.

5.4.6 Statistical Analysis

Effect Size of Synthetic Data Proportions

In machine learning research, effect sizes, such as Cohen's d and eta-squared (η^2), are critical metrics for quantifying the strength of relationships or differences between groups. To evaluate the impact of synthetic data on model performance, we calculated η^2 effect sizes for each metric across varying synthetic data proportions (10%, 25%, 50%, and 75%). η^2 was derived using one-way ANOVA, representing the proportion of variance in performance metrics attributable to changes in synthetic data proportions. This analysis was performed separately for each combination of model (UNet, TopoSegNet) and dataset (Fields, GMP) and included six key metrics of precision, recall, IoU, Betti0, Betti1 and CDM. Effect size benchmarks vary across fields [197, 198, 199, 200]. In machine learning applications, η^2 values less than 0.2, between 0.2 and 0.5, and greater than 0.5 are typically interpreted as small, medium, and large effects, respectively [201, 198, 199]. This categorization provides researchers with a framework to assess the practical significance of their findings, complementing traditional measures of statistical significance.

Performance Distribution Analysis

To visualize and analyze the impact of synthetic data proportions on model performance across different datasets, we employed split violin plots combining box plot information with kernel density estimation. For each metric, we compared the performance distributions between UNet and TopoSegNet across four synthetic data proportions (10%, 25%, 50%, 75%). The violin plots were vertically segmented to simultaneously display performance on Fields and GMP datasets. This visualization approach allows the examination of central tendency, spread, and probability density of performance metrics across both datasets, with the width of each segment indicating the probability density of observations at that performance level.

Model Comparison Analysis

To evaluate the performance differences between UNet and TopoSegNet, we conducted comprehensive statistical analyses across both Fields and GMP datasets. We performed paired t-tests to compare the models across multiple metrics including accuracy, precision, recall, IoU, CDM, and Betti number errors. For each metric, Cohen's d effect sizes were calculated to quantify the magnitude of the differences between the models. In line with standard guidelines, d values of 0.2, 0.5, and 0.8 were used to represent small, medium, and large effects, respectively [193, 202]. This provided a standardized measure to assess the practical significance of the observed differences between the models. Statistical significance was conservatively set at $p < 0.01$ for all statistical tests in this study. For metrics where lower values indicate better performance (CDM, Betti0, and Betti1), we calculated improvements as the absolute difference between the models' mean values. For all other metrics where higher values indicate better performance (accuracy, precision, recall, and IoU), we directly compared the mean values. This statistical framework allowed us to not only identify significant differences between the models but also quantify the practical significance of these differences through effect size measurements.

Statistical Analysis of Base Image Impact

To determine the optimal number of base images for synthetic data generation, we conducted a comprehensive statistical analysis grouping base images into ranges (1-2, 3-5, 6-7, and 8-10 images) to ensure sufficient sample sizes for statistical testing. This analysis was performed independently of the synthetic data proportions to provide a generalized recommendation for the number of base images to use. For each metric we employed the Kruskal-Wallis H-test, a non-parametric alternative to one-way ANOVA, to evaluate whether there were statistically significant differences between base image ranges per model-dataset combination [203, 204, 205]. This test was chosen over parametric alternatives due to its robustness with small sample sizes and no assumption of normal distribution. For each range of base images, we calculated descriptive statistics including mean performance, standard deviation, and sample count. The optimal range was determined based on the metric type, with lower values considered better for error-based metrics (CDM, Betti0, and Betti1) and higher values better for pixel-wise metrics (precision, recall, and IoU).

5.5 Results

This section presents the results of the experiments conducted on two datasets, Fields and GMP, to evaluate the performance of UNet and TopoSegNet under varying configurations of synthetic data proportions and the number of base images used to generate these synthetic datasets. The performance of these models is compared not only across these configurations but also against baseline models trained with 100% real data, represented as blue horizontal dashed lines in Figures 5.1 and 5.2. This analysis spans pixel-wise metrics such as precision, recall, and IoU, as well as topology-aware metrics such as Betti numbers and CDM. The results highlight trends within each group (based on the number of base images) and across proportions (10% synthetic: A, 25%: B, 50%: C, and 75%: D), with a focus on identifying how synthetic data influences segmentation performance relative to the real-data baseline. The results presented in Figures 5.1 and 5.2 demonstrate that data augmentation significantly improved the IoU, a critical pixel-wise metric for imbalanced datasets when compared to the baseline experiments with real datasets. Specifically, IoU increased from 25.51% to 27.19% (B-7) for UNet-Fields, from 26.22% to 28.00% (A-5) for TopoSegNet-Fields, from 78.54% to 80.06% (B-2) for UNet-GMP, and from 79.91% to 81.25% for TopoSegNet-GMP.

5.5.1 Fields Dataset

The evaluation of UNet and TopoSegNet on the Fields dataset provides valuable insights into their performance across varying proportions of synthetic data. In Group A (10% synthetic data), UNet demonstrated higher fluctuations in precision and recall compared to TopoSinGAN, particularly across base image groups. For Groups A and B, similar precision values of 38.56% and 38.58% were recorded. However, as the proportion of synthetic data increased, average precision dropped to 38.24% in Group C and 37.41% in Group D. Compared to the baseline UNet performance with real images (precision: 39.40%), most experiments with synthetic datasets yielded lower values. Exceptions were observed in Group A, where experiments such as A-4 and A-5 exceeded the baseline, achieving precision values above 39.9%. In contrast, TopoSinGAN exhibited a more stable trend in precision, with incremental improvements as synthetic data proportions and the number of base images increased. The average precision for Groups A and B exceeded the baseline (39.73% and 40.02% vs. 39.39%). Notably, TopoSinGAN achieved its highest precision of 40.98% in Group A when using four base images, over 1% higher than the group average of 39.73%. Group B (25% synthetic data) achieved the highest average precision (40.02%), with all experiments involving more than four base images yielding precision values above 40%. Although Groups C and D began with lower precision (36.76% and 34.86%), sharp improvements were observed, with Group D reaching 41.28% precision when trained with six base images.

For recall, UNet displayed wider fluctuations, ranging from 48.09% (A-3) to 38.64% (C-5), reflecting variability across experiments. While the global trend for recall across proportions was declining, within-group trends showed improvement as the number of base images increased. All group averages for UNet were above the baseline recall with real images (40.63%), recording values of 42.85%, 43.40%, 42.56%, and 41.86% for Groups A, B, C, and D, respectively. Conversely, TopoSegNet exhibited a steady within-group trend and a gradual decline across proportion groups. TopoSegNets recall fluctuations widened with increasing synthetic data proportions. The highest average recall for TopoSegNet (46.23%) was observed in Group A, while the lowest (39.32%) occurred in Group D, where 75% of the data was synthetic. Comparing group averages to the baseline recall for TopoSegNet (43.98%), Groups A, B, and C achieved higher recall values of 46.23%, 45.20%, and 45.05%, respectively.

The IoU metric, which accounts for class imbalance and provides a realistic pixel-wise evaluation, revealed a declining trend across synthetic data proportions for both UNet and TopoSegNet. Within groups, both models demonstrated significant improvements in IoU as the number of base images increased, except in Group A (10% synthetic data). For UNet, the highest IoU was recorded in B-7 (25% synthetic data with seven base images) at 27.19%, significantly exceeding the baseline IoU of 25.52%. The average IoU for Groups A and B (25.64% and 25.90%, respectively) was slightly higher than the baseline. In contrast, TopoSegNet achieved its best IoU in Group A, reaching a maximum of 28.00% in A-5 (five base images) and a group average of 27.41%. Group B followed closely with an average IoU of 27.19%, while Group C recorded 26.58%. All these group averages exceeded the baseline IoU of 26.20%, highlighting significant improvements in the first three groups, peaking in Group A. TopoSegNets lowest IoU (24.15%) occurred in D-1 (75% synthetic data with one base image). However, a sharp increase of 1.57% was observed when the number of base images increased to two.

Betti0, a topology-aware metric evaluating connected components, showed mixed trends. For UNet, only a few experiments in Groups A and B improved upon the baseline Betti0 of 80. The lowest Betti0 (best performance) for UNet was observed in B-6 at 57.27. However, Betti0 values sharply worsened as synthetic data proportions increased, peaking at 197.72 in D-3. For TopoSegNet, the Betti0 fluctuations were significantly smaller, and substantial improvements were observed in Groups A and B. The lowest Betti0 for TopoSegNet was recorded in A-6 at 22.15, substantially better than the baseline Betti0 of 65. Like UNet, TopoSegNets Betti0 degraded with higher synthetic proportions, particularly in Group D.

Betti1, which evaluates loops, showed an opposite trend for UNet compared to Betti0. Group averages revealed improvements across proportions, with Group D recording the best Betti1 (61.13), followed by Group C (63.12) and Group B (62.50). The best Betti1 value (54.55) was achieved in C-6. For TopoSegNet, however, the highest average Betti1 was recorded in Group A (54.02), followed by Group C (54.80), which matched the baseline Betti1 of 54.80. Fluctuations in Betti1 were significantly smaller for TopoSegNet compared to UNet. The best Betti1 value for TopoSegNet was achieved in A-6 (50.01), highlighting its ability to preserve loop structures.

The CDM metric, which assesses both topological and geometric consistency, showed consistent improvement for UNet across all groups, with averages of 19.79, 19.12, 20.09, and 18.05 for Groups A, B, C, and D, respectively. These values were better than the baseline CDM of 21.7, with the lowest CDM observed in D-4 at 12.24. Conversely, TopoSegNet showed a reversed trend, with worsening CDM as the proportion of synthetic data increased. Group A recorded the lowest average CDM (14.85), followed by Group B (15.16), while Groups C and D averaged 15.59 and 18.68, respectively. Groups A and B demonstrated improvements compared to the baseline CDM of 15.61, reflecting TopoSegNets advantage at lower synthetic proportions.

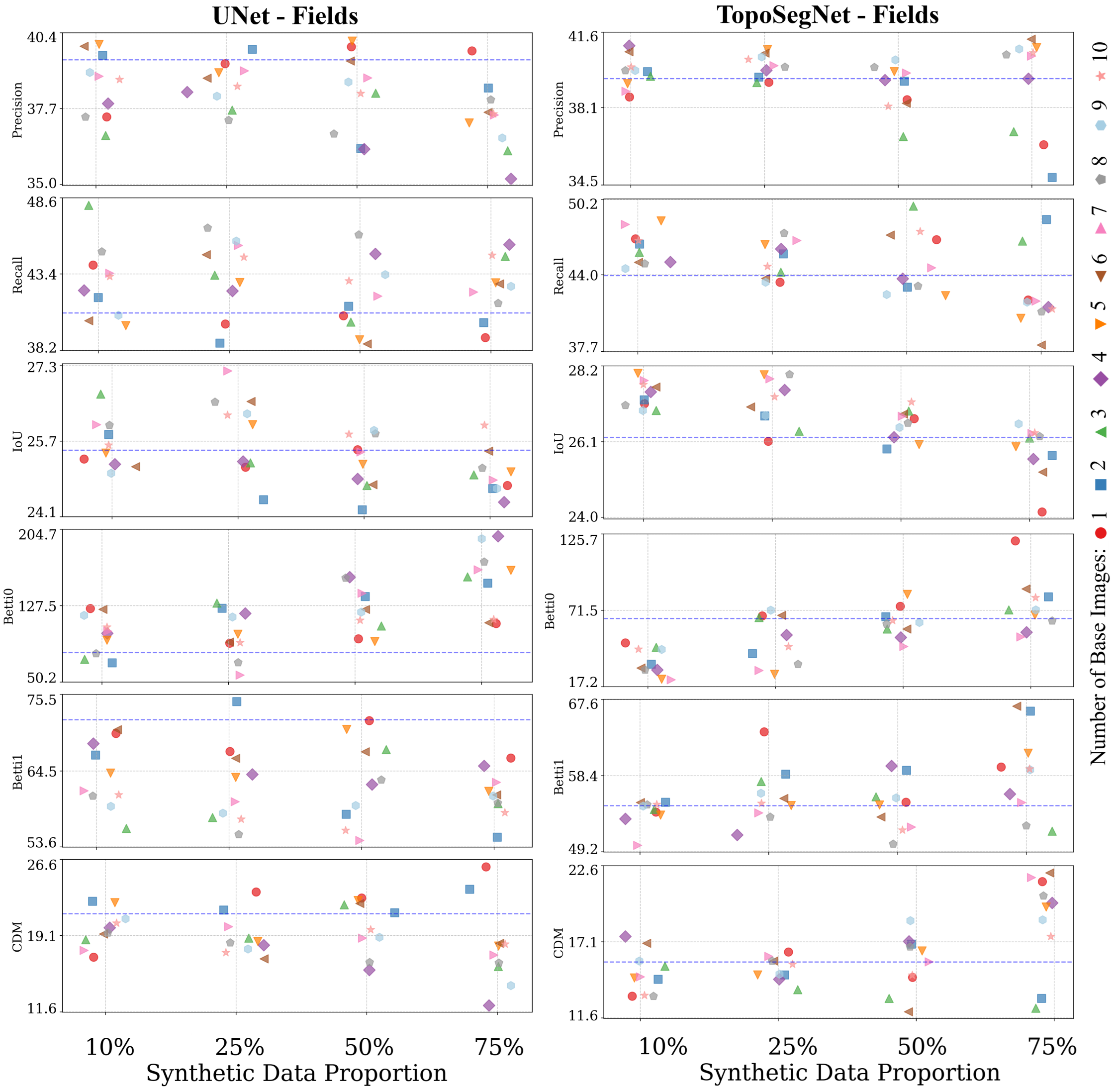


Figure 5.1: Performance comparison between UNet (left) and TopoSegNet (right) on the Fields dataset across different synthetic data proportions. Each plot shows a different evaluation metric against base image configurations (x-axis). Blue dashed lines represent baseline performance with no synthetic data. Results are grouped by synthetic data proportions: A (10%), B (25%), C (50%), and D (75%) shown in distinct colors and line styles.

5.5.2 GMP Dataset

For the GMP dataset, both UNet and TopoSegNet demonstrated consistent improvements across all metrics compared to the baseline experiment trained on 100% real data. The baseline precision for UNet was recorded at 86.73%, which improved gradually across the synthetic data proportion groups (A to D), culminating in an average precision of 89.07% for Group D. Notably, nearly all experiments exceeded the baseline precision, highlighting the effectiveness of synthetic data augmentation. TopoSegNet exhibited a similar upward trend, achieving a maximum precision of 89.79% in experiments D-1, D-3, and D-10. Interestingly, the group average precision for Groups A, B, and C remained relatively stable, approximately 88.63%, while a notable increase was observed for Group D, which recorded a group average of 89.45%. This suggests that a higher proportion of synthetic data positively influenced precision for TopoSegNet, particularly in Group D.

Recall, on the other hand, exhibited more variability, particularly for UNet. Group averages for recall showed a gradual decline across proportion groups, from 88.79% in Group A to 88.32% in Group D, indicating a downward trend as the proportion of synthetic data increased. However, within-group trends demonstrated an upward trajectory, suggesting that increasing the number of base images used for data augmentation mitigated the negative impact of higher synthetic proportions. In comparison, TopoSegNet demonstrated more stable recall values, with group averages for Groups A, B, and C slightly surpassing the baseline recall of 89.64%, achieving 89.78%, 89.79%, and 89.88%, respectively. The best recall for TopoSegNet was recorded at 91.31% in experiment C-4, showcasing the model's ability to generalize effectively with moderate synthetic data augmentation.

The IoU metric revealed consistent improvements for both models across all experiments compared to their respective baselines. For UNet, the baseline IoU was 78.58%, which improved to 79.35% in Group A and decreased slightly to 79.09% in Group B. IoU then recovered to 79.18% in Group C and rose sharply to 79.72% in Group D. The highest IoU for UNet was observed in experiment D-2, which achieved 80.06%. TopoSegNet followed a similar pattern but exhibited smaller standard deviations, reflecting greater consistency. Its baseline IoU was 79.95%, with group averages of 80.53% for Groups A and B, increasing to 80.61% for Group C and peaking at 81.04% for Group D. The highest IoU for TopoSegNet was recorded in experiment D-3 at 81.25%, emphasizing its capacity to benefit from higher proportions of synthetic data.

The Betti0 error, which measures errors in connected components, revealed an upward trend for UNet as the proportion of synthetic data increased. The best group average for UNet was observed in Group A at 20.20, with the minimum Betti0 recorded in experiment A-10 at 17.56, significantly improving upon the baseline Betti0 of 25.01. For TopoSegNet, Group D demonstrated the best performance, with a group average of 10.12 and a minimum value below 9, reflecting substantial topological improvements. The other groups for TopoSegNet exhibited similar average Betti0 values, indicating stable performance across lower proportions of synthetic data.

For Betti1, UNet exhibited an arc-shaped trend, with the best performance observed in Group D (75% synthetic data). This group achieved a significantly improved Betti1 of 40.89, compared to the baseline of 44.92. TopoSegNet showed smaller fluctuations and significantly improved Betti1 values overall. Group D recorded the best performance for TopoSegNet, with a group average of 14.71 and a minimum Betti1 of 13.29 in experiment C-2. This represents a marked improvement compared to the baseline Betti1 of 20.00, further highlighting TopoSegNets ability to preserve topological structures even at high synthetic proportions.

The CDM metric, which evaluates both topological and geometric consistency, displayed an upward trend for UNet across the proportion groups. The baseline CDM for UNet was 2.96, which improved slightly to 2.67 in experiment A-5. For TopoSegNet, however, there was no discernible trend, as the range of CDM values remained narrow. The best CDM for TopoSegNet was recorded at 1.91 in experiment A-10, reflecting an improvement over the baseline CDM of 2.09. The smaller variability in TopoSegNets CDM values suggests that it consistently maintains geometric and topological coherence regardless of the proportion of synthetic data.

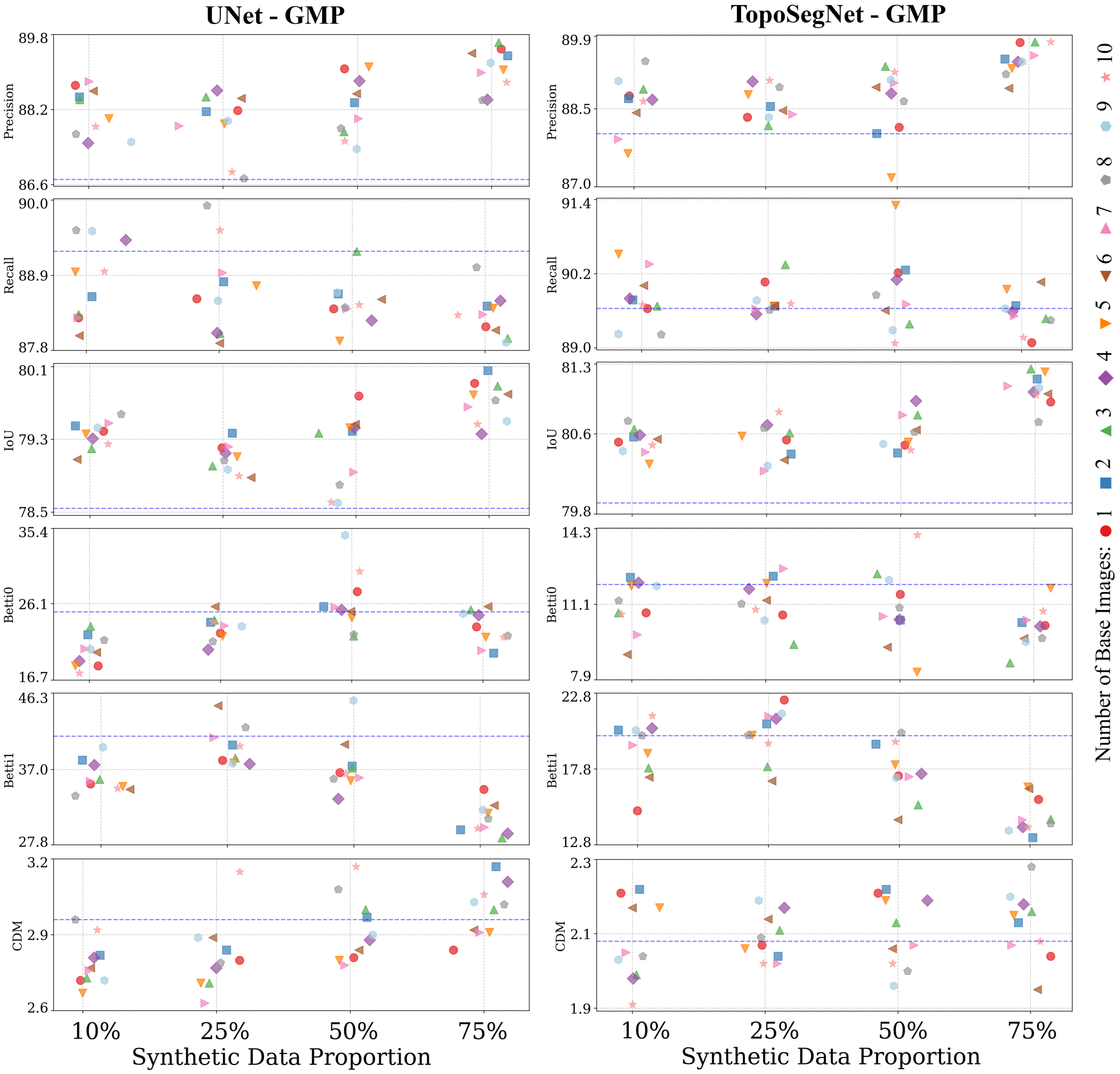


Figure 5.2: Performance comparison between UNet (left) and TopoSegNet (right) on the GMP dataset across different synthetic data proportions. Each plot shows a different evaluation metric against base image configurations (x-axis). Blue dashed lines represent baseline performance with no synthetic data. Results are grouped by synthetic data proportions: A (10%), B (25%), C (50%), and D (75%) shown in distinct colors and line styles.

5.5.3 Model Comparison Analysis

TopoSegNet demonstrated improved performance across all metrics in both datasets, with all differences being statistically significant ($p < 0.01$). In the Fields dataset, as reported in Table 5.2, TopoSegNet showed substantial improvements in topological metrics, reducing Betti0 errors by 61.68 units (d = 2.22) and Betti1 errors by 7.02 units (d = 1.55). Pixel-wise metrics showed moderate improvements, with increases in precision (1.27%, d = -0.99), recall (2.06%, d = -0.83), and IoU (1.29%, d = -1.73). The performance gap was more pronounced in the GMP dataset, where TopoSegNet achieved notably larger effect sizes (Table 5.3). Topological improvements were particularly strong, with substantial reductions in Betti0 (12.34 units, d = 5.14) and Betti1 (18.18 units, d = 5.37) errors. The model also demonstrated significant improvements in CDM (reduction of 0.78, d = 6.70). Pixel-wise metrics showed consistent enhancements, with improved precision (0.50%, d = -0.76), recall (1.15%, d = -2.45), and a substantial improvement in IoU (1.34%, d = -4.07). As shown in Tables 5.2 and 5.3, the sign of Cohens d reflects the direction of improvement. Negative d values for standard metrics (e.g., precision, recall, IoU) indicate that TopoSegNet outperformed UNet, as higher values are better for these metrics. Conversely, positive d values for topological metrics (e.g., Betti0, Betti1, CDM), where lower values are better, similarly reflect improvements for TopoSegNet. The magnitude of Cohen's d values highlights the extent of performance differences, which were most pronounced in the GMP dataset, particularly for topological metrics. This suggests that TopoSegNet's architectural design offers significant advantages, especially for complex dendrological structures.

Table 5.2: Comparative analysis of UNet and TopoSegNet performance across the Fields dataset.

Metric	**Dataset**	**UNet Mean**	**TopoSegNet Mean**	**Cohen's d**	**Better Model**	**Improvement**	**p_value**
Precision	Fields	38.23	39.50	-0.99	TopoSegNet	1.27	0.0001
Recall	Fields	42.62	44.68	-0.83	TopoSegNet	2.06	0.0008
IoU	Fields	25.45	26.74	-1.73	TopoSegNet	1.29	0.0000
CDM	Fields	19.31	16.07	1.22	TopoSegNet	3.24	0.0000

Continued on next page

Table 5.2 continued

Metric	**Dataset**	**UNet Mean**	**TopoSegNet Mean**	**Cohen's d**	**Better Model**	**Improvement**	**p_value**
Betti0	Fields	117.69	56.01	2.22	TopoSegNet	61.68	0.0000
Betti1	Fields	62.89	55.87	1.55	TopoSegNet	7.02	0.0000

Table 5.3: Comparative analysis of UNet and TopoSegNet performance across the GMP dataset.

Metric	**Dataset**	**UNet Mean**	**TopoSegNet Mean**	**Cohen's d**	**Better Model**	**Improvement**	**p_value**
precision	GMP	88.30	88.81	-0.76	TopoSegNet	0.50	0.0005
recall	GMP	88.60	89.75	-2.45	TopoSegNet	1.15	0.0000
iou	GMP	79.32	80.65	-4.07	TopoSegNet	1.34	0.0000
CDM	GMP	2.88	2.10	6.70	TopoSegNet	0.78	0.0000
Betti0	GMP	23.22	10.87	5.14	TopoSegNet	12.34	0.0000
Betti1	GMP	36.14	17.96	5.37	TopoSegNet	18.18	0.0000

5.5.4 Performance Distribution Analysis

The performance distribution analysis revealed distinct trends in segmentation metrics for UNet and TopoSegNet across synthetic data proportions (10%, 25%, 50%, and 75%). Violin plots presented in Figure 5.3 and Figure 5.4 for Fields and GMP datasets, respectively, highlighted how metrics such as precision, recall, IoU, and topological metrics (Betti numbers and CDM) were distributed, offering insights into the central tendency, variability, and overall reliability of the models in the Fields and GMP datasets. In the Fields dataset, TopoSegNet showed narrower distributions and higher medians for precision and recall at lower synthetic data proportions (10% and 25%), reflecting greater consistency and reliability. However, as synthetic data increased to 50% and 75%, TopoSegNets performance became less consistent, with wider recall distributions. By 75%, UNet surpassed TopoSegNets median recall, reversing the earlier trend of TopoSegNet outperforming UNet.

For IoU, TopoSegNet consistently demonstrated narrower distributions and higher medians across all synthetic proportions, particularly at 10%, 25%, and 50%, where it displayed reliable performance in capturing overlaps between predictions and ground truth. At 75%, while TopoSegNet maintained a higher median IoU than UNet, its distributions became broader, reflecting increased variability. Topological metrics such as Betti0 and Betti1 further underscored the differences between the models. TopoSegNet consistently maintained lower error rates and tighter distributions, while UNet exhibited substantial variability and pronounced degradation in these metrics, especially at higher synthetic proportions (25% to 75%). The CDM distributions reinforced TopoSegNet's advantages, showing lower median values and compact distributions compared to UNet's broader and more variable performance. Overall, the analysis highlights TopoSegNet's enhanced and more reliable performance, particularly in preserving structural and topological integrity, though its consistency diminishes slightly at the highest synthetic data proportions.

In the GMP dataset (Figure 5.4), TopoSegNet consistently outperformed UNet across multiple metrics, demonstrating improved precision, recall, IoU, and topological performance. For precision, TopoSegNet exhibited more compact distributions and slightly higher median values across all synthetic data proportions. As synthetic data increased to 75%, both models maintained stable precision, but TopoSegNet's distributions remained particularly concentrated around 89-90%, reflecting robust performance even with high synthetic data usage. In terms of recall, both models maintained neutral trends, but TopoSegNet consistently showed more compact distributions and higher medians compared to UNet across most proportions. IoU measurements highlighted one of the most notable differences, with TopoSegNet achieving consistently higher values (around 80-81%) and maintaining compact distributions across all synthetic data proportions. Both models showed progressively improved IoU medians after the 25% synthetic data proportion, though TopoSegNet displayed greater reliability. The topological metrics further underscored TopoSegNet's enhanced structural preservation capabilities in the

GMP dataset. All three topological metrics demonstrated stable trends, with Betti1 values slightly decreasing as synthetic data proportions increased, reflecting enhanced structural precision in TopoSegNet's predictions.

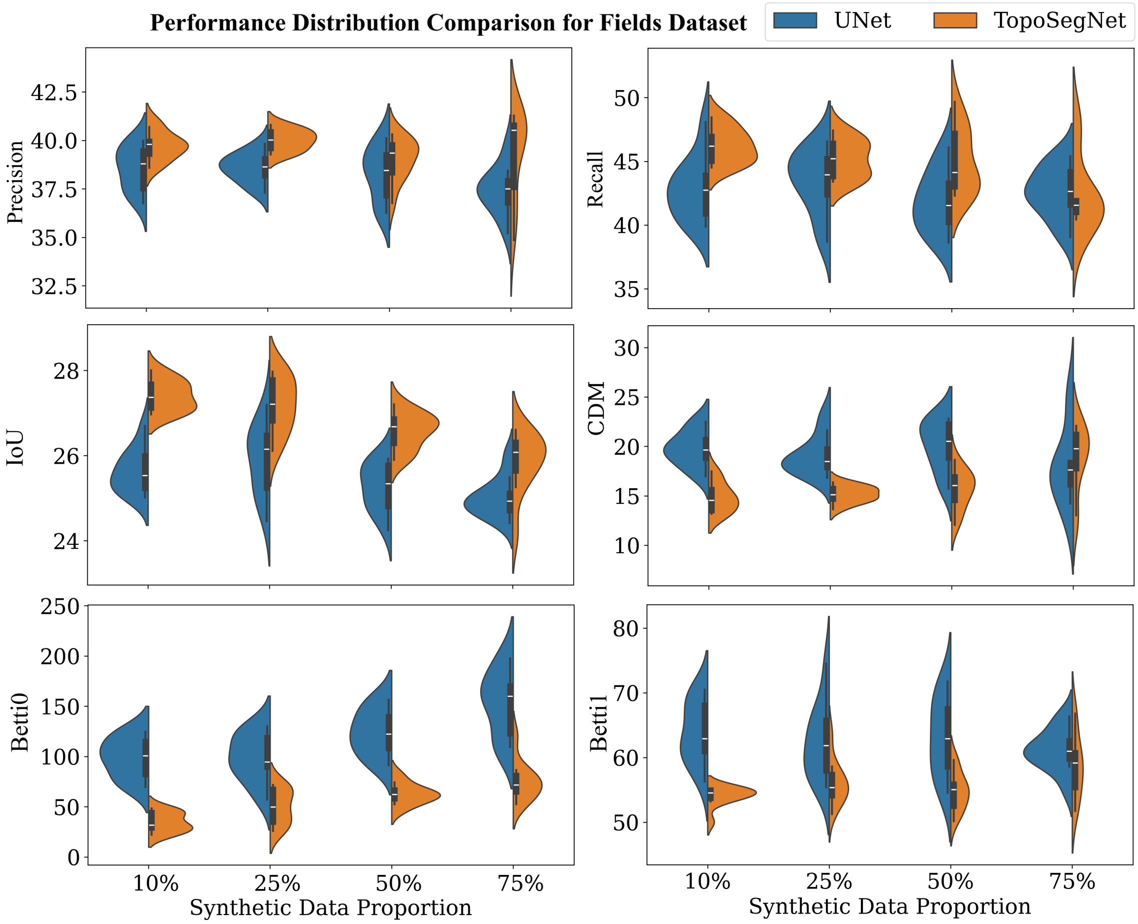


Figure 5.3: Performance distribution comparison between UNet and TopoSegNet across synthetic data proportions (10-75%) for Fields dataset. Violin plots with embedded box plots show the probability density of each metric.

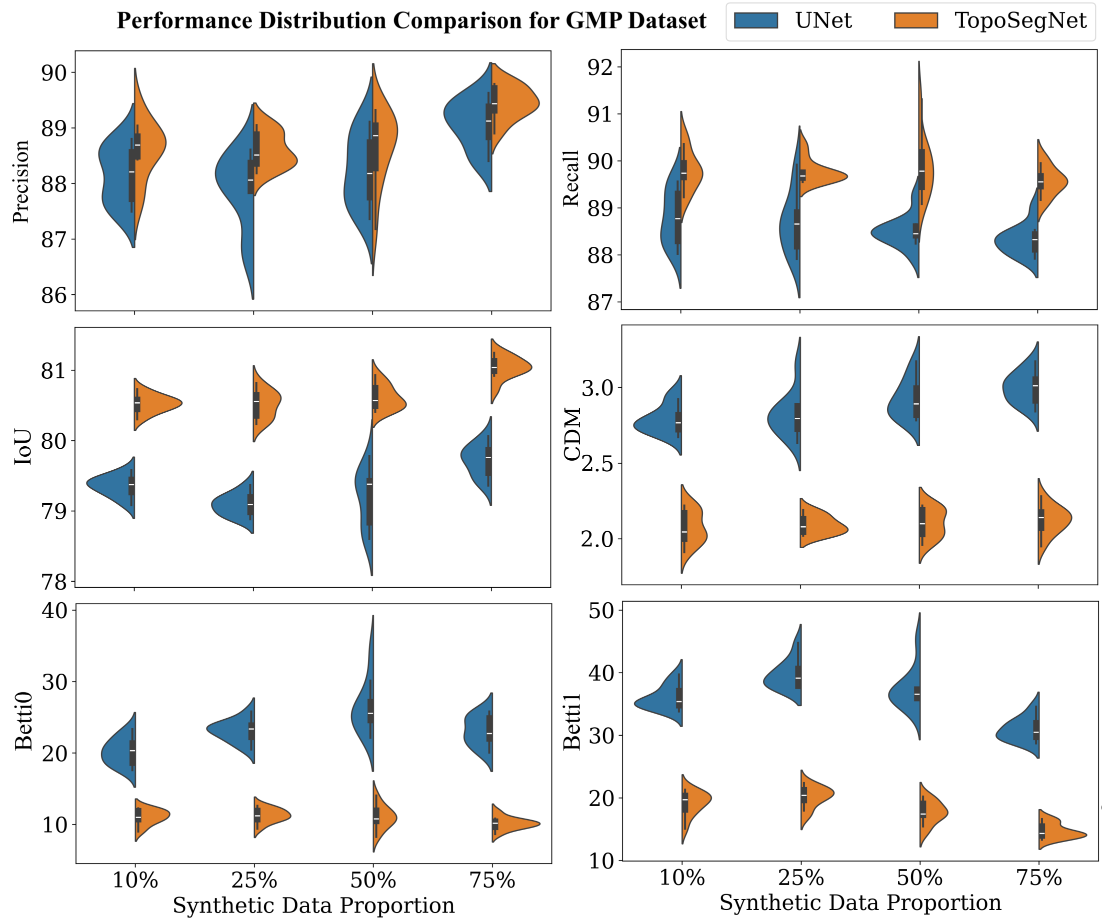


Figure 5.4: Performance distribution comparison between UNet (blue) and TopoSegNet (orange) across synthetic data proportions (10-75%) for GMP dataset. Violin plots with embedded box plots show the probability density of each metric.

5.5.5 Effect Size Analysis

The effect size analysis, as illustrated in Figure 5.5, highlights the influence of synthetic data proportions on the performance metrics of UNet and TopoSegNet across the Fields and GMP datasets. For precision, TopoSegNet demonstrates a medium effect in the GMP dataset ($\eta^2 = 0.40$) and a small effect in the Fields dataset ($\eta^2 = 0.09$). UNet exhibits a similar trend, with a medium effect in GMP ($\eta^2 = 0.40$) and a small effect in Fields (η^2

$= 0.16$). In recall, TopoSegNet shows a medium effect in the Fields dataset ($\eta^2 = 0.32$) but a small effect in GMP ($\eta^2 = 0.08$). Conversely, UNet achieves a small effect in GMP ($\eta^2 = 0.13$) and exhibits minimal sensitivity in the Fields dataset ($\eta^2 = 0.06$). For IoU, TopoSegNet outperforms UNet with consistently higher effect sizes in both datasets. In GMP, TopoSegNet achieves a large effect ($\eta^2 = 0.66$) compared to UNet's medium effect ($\eta^2 = 0.49$). Similarly, in the Fields dataset, TopoSegNet reaches a large effect ($\eta^2 = 0.57$), while UNet shows a medium effect ($\eta^2 = 0.26$). Topological metrics reveal distinct trends. For Betti0, TopoSegNet maintains greater stability in the Fields dataset with a large effect ($\eta^2 = 0.53$), compared to UNet's medium effect ($\eta^2 = 0.46$). In GMP, TopoSegNet's η^2 drops to a small effect ($\eta^2 = 0.13$), whereas UNet demonstrates a medium effect ($\eta^2 = 0.47$). For Betti1, TopoSegNet achieves a large effect in GMP ($\eta^2 = 0.63$), reflecting its strong ability to preserve loops with synthetic data. UNet shows a slightly larger η^2 in GMP ($\eta^2 = 0.67$), but this suggests greater dependence on synthetic data proportions, potentially reflecting performance variability. In the Fields dataset, TopoSegNet achieves a small effect ($\eta^2 = 0.22$) compared to UNet's minimal effect ($\eta^2 = 0.04$). The CDM metric emphasizes TopoSegNet's enhanced geometric consistency, showing a medium effect in the Fields dataset ($\eta^2 = 0.35$). In GMP, the impact is minimal for TopoSegNet ($\eta^2 = 0.04$). UNet demonstrates a medium effect in GMP ($\eta^2 = 0.36$) but only a small effect in Fields ($\eta^2 = 0.08$).

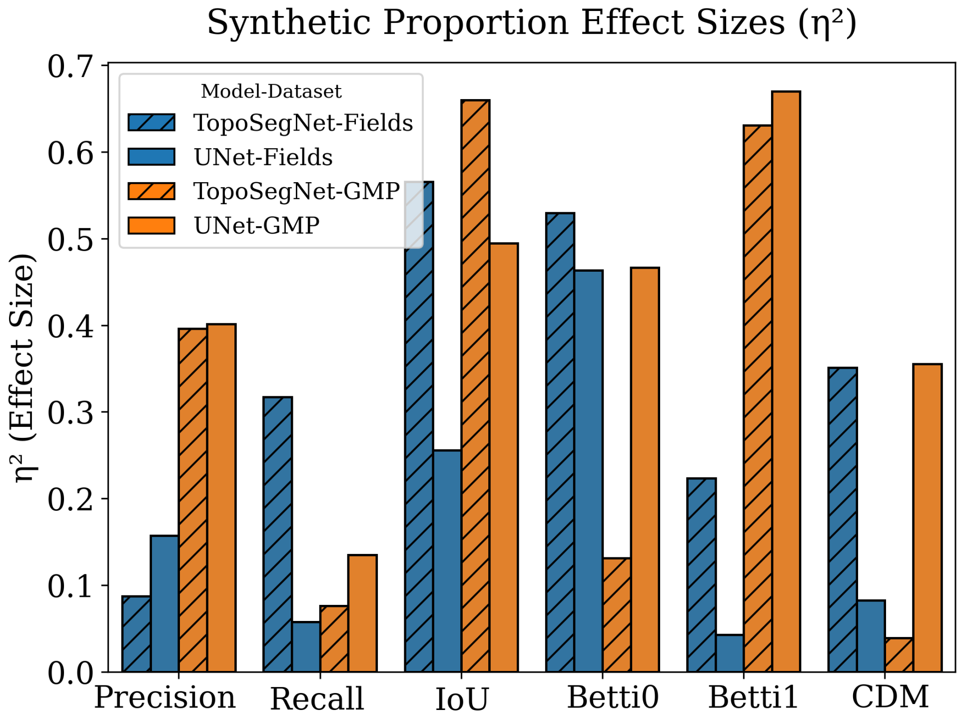


Figure 5.5: Effect sizes (η^2) plot, illustrating the impact of synthetic data proportions on segmentation metrics for UNet and TopoSegNet across the Fields and GMP datasets. Higher η^2 values indicate greater sensitivity of a metric to changes in synthetic data proportions.

5.5.6 Statistical Analysis of Base Image Impact

The results of the Kruskal-Wallis H-test applied across different base image ranges with a significance level of 95% for determining the optimal number of base images for synthetic data generation are summarized in Table 5.4. This table highlights the statistically significant results, including the H-statistic, p-value, and the optimal base image range for each combination. For the UNet-Fields model, metrics such as Recall, CDM, and Betti1 consistently indicated the 810 base image range as optimal, with p-values below 0.05. Similarly, for the TopoSegNet-Fields model, Precision showed a statistically significant preference for the 810 range. These results suggest that for these datasets, larger numbers of base images yield better performance for both pixel-wise and error-based metrics.

Table 5.4: Statistically significant (p_value < 0.01) optimal base image ranges for each model-dataset and metric combination based on the Kruskal-Wallis H-test.

Model-Dataset	**Metric**	**H-statistic**	**p-value**	**Optimal Range**
UNet-Fields	Recall	9.8726	0.0197	8-10
UNet-Fields	CDM	10.3451	0.0158	8-10
UNet-Fields	Betti1	9.8983	0.0195	8-10
TopoSegNet-Fields	Precision	10.7118	0.0134	8-10

5.6 Discussion

The results shown in Figures 5.1 and 5.2 suggest that using TopoSinGAN for generating synthetic images for expanding the training datasets for agricultural fields and dendrological (GMP) case studies has been effective in improving the performance of both UNet and TopoSegNet models significantly when evaluated by most pixel-wise and topology metrics. Statistical analysis confirms TopoSegNet's enhanced performance over UNet across all metrics with particularly strong improvements in topological metrics (Table 5.2 and 5.3). The magnitude of these improvements, quantified by Cohen's d effect sizes, varies between datasets and metrics. In the Fields dataset, TopoSegNet achieved substantial improvements in topological metrics, with large effect sizes for Betti0 and Betti1 errors, while showing moderate improvements in pixel-wise metrics. The GMP dataset exhibited even more pronounced differences, with large effect sizes for topological metrics and substantial improvements in pixel-wise metrics. These improvements are not merely statistically significant but practically meaningful, particularly in the context of topological preservation. The reduction in Betti number errors (61.68 for Betti0 in Fields dataset) represents a substantial improvement in structural accuracy that could significantly impact applications requiring precise feature connectivity, such as agricultural field boundary detection or cellular structure analysis.

The contrasting performance patterns between GMP and Fields datasets reveal important insights about how dataset characteristics influence the effectiveness of synthetic data augmentation. According to Figure 5.4, the GMP dataset, featuring more regular and consistent structures, showed more stable improvements across synthetic data proportions, with both models maintaining relatively consistent performance even at higher synthetic ratios. This stability can be attributed to the more uniform nature of dendrological structures, making them more amenable to synthetic reproduction.

Conversely, the Fields dataset, characterized by irregular agricultural boundaries and variable structures, exhibited more volatile performance patterns, particularly at higher synthetic proportions (Figure 5.3). This volatility suggests that complex, irregular structures and geometric properties pose greater challenges for synthetic data generation and subsequent segmentation. TopoSegNet's enhanced performance in maintaining structural consistency, especially in the Fields dataset, underscores the importance of topology-aware approaches when dealing with irregular features.

The influence of the number of base images used for generating synthetic data on segmentation performance was initially hypothesized to improve with an increasing number of base images. To statistically evaluate this hypothesis, the Kruskal-Wallis H-test was conducted across four bins of base image counts (12, 35, 67, and 810 images). The results, summarized in Table 5.4, support this hypothesis for the Fields dataset at a 95% significance level. Specifically, the test revealed a statistically significant optimal range of 810 base images for this dataset. However, no significant optimal ranges were detected for other model-dataset combinations in the experiments.

Visual inspection of the experimental results in Figure 5.1 further underscores the benefit of a higher number of base images for the Fields dataset. Each additional base image appears to introduce novel configurations of features and contexts, which likely enhance model generalization and segmentation performance. This aligns with the observed statistical significance for this dataset. Conversely, for the GMP dataset, Figure 5.2 shows less evidence of a clear improvement in segmentation performance as the number of base images increases. This can be attributed to the relatively uniform and monotonic context and structural configurations of xylem cells in the GMP images. In this case, the addition of new base images may not significantly increase the diversity of features presented to the model, thereby resulting in a less pronounced impact on performance.

The effect size analysis (η^2) presented in Figure 5.5 reveals distinct patterns in how synthetic data proportions influence various performance metrics. TopoSegNet demonstrated greater sensitivity to data augmentation proportions in both datasets, indicating its capacity to leverage synthetic data more effectively but also its susceptibility to overfitting when exposed to excessive synthetic data. This increased sensitivity can be attributed to the introduction of more topological structures into the model as the proportion of synthetic data grows. Metrics such as IoU and topological metrics (Betti0 and Betti1) were notably more sensitive to synthetic data proportions, while recall exhibited the least responsiveness.

For UNet, which lacks topological awareness, the Fields dataset displayed small effect sizes across most metrics, except for Betti0, where a large effect size was observed. This indicates that synthetic data proportions significantly impact the number of connected components (or noise) in the predictions. Conversely, when UNet was applied to the GMP dataset, significantly higher effect sizes were observed for all metrics, suggesting that the impact of data augmentation is greater in datasets with more homogeneous topological structures. For applications like the Fields dataset, where the features exhibit high topological variability, synthetic data augmentation has a more limited effect on model performance. In contrast, for more homogeneous datasets like GMP, synthetic data augmentation yields larger improvements in segmentation performance. Additionally, the medium effect size for precision and small effect size for recall in the GMP dataset highlights that homogeneous feature arrangements result in synthetic data proportions affecting false positive rates more than false negatives.

For TopoSegNet, the effect size analysis demonstrated differing results between the two datasets. In the Fields dataset, large effect sizes were observed for IoU and Betti0, while medium effect sizes were found for recall, Betti1, and CDM. This indicates that when a segmentation model incorporates topological awareness, synthetic data augmentation has a substantially higher impact on most metrics, except for precision. The lower sensitivity of precision suggests that false positive rates are less affected by synthetic data proportions compared to UNet. The higher sensitivity of topological metrics, particularly

Betti0, underscores that increasing the proportion of synthetic data introduces more diverse topological configurations into the model, thereby enhancing its topological accuracy. Additionally, the medium effect sizes for recall and IoU suggest that synthetic data augmentation can significantly affect the model's performance by increasing variations in false negative rates, ultimately impacting IoU.

For GMP, TopoSegNet exhibited a different pattern. The sensitivity of precision and recall remained similar to that of UNet, while IoU displayed higher sensitivity. Topological metrics revealed that Betti0s sensitivity to synthetic data proportions was significantly reduced, indicating improved stability in predicting the number of connected components. However, a large effect size for Betti1 and a small effect size for CDM suggest that synthetic data proportions primarily influence over-segmentation rates in the GMP dataset. This is because CDM, by design, is less sensitive to over-segmentation. Additionally, the higher effect sizes for IoU and precision suggest that variations in IoU may be driven by false positives, supporting the over-segmentation scenario.

Our exploratory analysis of data augmentation using TopoSinGAN has several limitations and caveats that should be considered. One primary limitation lies in the number of base images used for training the TopoSinGAN model. In this study, we restricted the number of base images to a maximum of 10 to ensure the feasibility of completing all planned experiments within the constraints of the study. While this limitation allowed for a manageable experimental design, it may have restricted the diversity and robustness of the synthetic data generated. Another important caveat relates to the selection of these base images. It is crucial to carefully curate base images to ensure they represent diverse configurations, styles, and contexts. This diversity is essential for reducing specificity and overfitting, as it broadens the models exposure to different feature variations. A lack of such careful selection may diminish the effectiveness of data augmentation.

Additionally, the study employed binning of synthetic data proportions into four groups including 10%, 25%, 50%, and 75% synthetic data, due to practical constraints on computational resources and time. While this binning strategy provided a structured framework for analysis, finer-grained bins could potentially yield additional insights into the relationship between synthetic data proportions and model performance. Another limitation stems from the choice of models used to evaluate the impact of data augmentation. The analysis focused on two models, UNet and TopoSegNet, with the latter utilizing a UNet backbone. While these models serve as strong baselines, exploring other architectures with different backbones, such as Transformer-based models or those incorporating ResNet, could provide a more comprehensive understanding of how data augmentation influences various image segmentation frameworks. These limitations suggest several avenues for future work.

5.7 Conclusion

This study highlights the potential of TopoSinGAN as an effective augmentation technique for improving image segmentation performance, particularly in data-scarce environments where topological accuracy is critical. Unlike conventional data augmentation methods, TopoSinGAN generates topologically consistent synthetic images from as few as a single base image, ensuring diversity and structural integrity in the augmented dataset. The experiments conducted on the Fields and GMP datasets demonstrate that incorporating TopoSinGAN-augmented data significantly enhances segmentation performance across various metrics, including both pixel-wise measures such as IoU and topological metrics such as Betti numbers and CDM. The augmentation approach proposed in this study not only enriches datasets but also reduces overfitting by exposing models to novel configurations and contexts, particularly evident in datasets with complex and irregular structures like agricultural fields. For more homogeneous datasets, such as GMP, the synthetic data further stabilizes model performance, showcasing its adaptability across diverse segmentation challenges. By addressing common limitations in synthetic data generation such as topological inaccuracies and the dependency on large annotated datasets TopoSinGAN establishes itself as a scalable and robust solution for training segmentation models in domains requiring high precision. These findings underscore the significant role of topology-aware data augmentation in advancing segmentation capabilities, paving the way for broader applications in fields such as medical imaging, remote sensing, and geographic information science (GIS). Future work should explore the scalability of this technique across other datasets and segmentation models to fully unlock its potential in diverse domains.

Chapter 6

Discussion and Conclusion

6.1 Discussion

This dissertation explored how topological constraints can enhance deep learning models for spatial image processing and generation, focusing on tasks where preserving topological integrity is critical. It addressed the research questions by developing novel approaches that bridge theoretical topology and its application in deep learning for image segmentation and generation. Through these efforts, the study contributed to the advancement of geospatial dendrological image analysis. The research questions were as follows:

1. How can topological constraints be integrated into deep learning models to enhance spatial image generation and segmentation?
2. To what extent can a topologically enhanced deep learning model impact spatial and topological attributes when generating synthetic images of irregularly shaped agricultural fields in Iran?
3. To what extent can a topologically enhanced deep learning model impact spatial and topological attributes when delineating vessels in ultra-high-resolution tree-ring dendrology images?

These questions were addressed through the development and evaluation of two novel models:

- **TopoSinGAN:** A topology-aware generative adversarial network designed to generate synthetic images with preserved topological structures from a single image.
- **TopoSegNet:** A topology-enhanced segmentation framework aimed at maintaining structural fidelity in segmentation tasks at minimal computational complexity.

The dissertation systematically addressed the first research question by introducing differentiable topology-aware loss functions. These functions were integrated into generative and segmentation models, enabling the preservation of topological features such as junctions, terminal nodes, continuity, and closed loops. Specifically, minimizing terminal nodes, identified as key indicators of topological errors, allowed the models to overcome structural inconsistencies often missed by traditional pixel-based loss metrics such as binary cross entropy.

To further enhance segmentation tasks, TopoSegNet was developed with a focus on topologically critical points (TCPs), including junctions and terminal nodes. By incorporating a multi-level erosion strategy, TopoSegNet ensured the structural integrity of complex and thin structures while maintaining computational efficiency. Unlike conventional topology-aware methods relying on persistent homology, which focus on 2-simplices such as loops and connected components, TopoSegNet shifts its focus to 0-simplices, such as junctions and terminal nodes. This shift significantly reduces computational complexity while maintaining topological correctness, making TopoSegNet faster and more suitable for large-scale applications.

The second research question was examined through the segmentation and generation of satellite imagery from Tehran Province. Agricultural fields in this region feature irregular geometries, requiring models capable of handling diverse topological patterns. TopoSinGAN successfully generated synthetic images with consistent field boundaries, reducing open loops and isolated line segments. Similarly, TopoSegNet maintained boundary continuity and outperformed baseline models on both topology-focused and pixel-wise metrics. These advancements have practical implications for remote sensing applications, enabling improved delineation of agricultural fields and other geospatial features.

The third research question was addressed through the application of TopoSinGAN and TopoSegNet to gigapixel dendrology images. These images present a complex task due to the intricate structure of tree-ring xylem vessels. The models demonstrated significant improvements in delineating these vessels, maintaining topological continuity, and reducing errors such as disconnections and boundary distortions.

By answering these research questions, this dissertation has advanced the field of topology-aware deep learning through the development and validation of novel methods for both image generation and segmentation tasks. The research progressed through three main phases, each building upon the previous work to create a comprehensive framework for topology-aware image processing. The systematic integration of topological constraints, coupled with the development of new evaluation metrics and computational strategies, has demonstrated significant improvements in both accuracy and efficiency compared to existing approaches.

In Chapter 3, the introduction of TopoSinGAN represented a significant advancement in topology-aware image generation. The model's novel differentiable topology loss function effectively minimized terminal nodes in generated images, addressing a critical gap in existing GAN architectures. The development of the Node Topology Clustering (NTC) metric provided a new way to evaluate topological accuracy in generated images, independent of geometric variations. TopoSinGAN demonstrated substantial improvements in both agricultural and dendrological applications, reducing NTC index values from 15.15 to 3.94 for agricultural fields and 14.55 to 2.44 for dendrology, compared to baseline SinGAN. These improvements were achieved while maintaining computational efficiency, with only a 9.41% increase in training time compared to the baseline model. However, TopoSinGAN faced several limitations. Its treatment of hanging lines as anomalies made it particularly suited for applications requiring closed loops, such as agricultural field delineation or membrane cell segmentation, but less optimal for scenarios where hanging lines are expected, such as road network analysis which dead-end road segments can be normal. The model also showed sensitivity to the thickness of boundary masks in input images, as it relied on fixed 3×3 terminal detector kernels. This sensitivity could lead to suboptimal performance when processing features with varying thicknesses.

Chapter 4 extended the topology-aware approach to image segmentation through TopoSegNet. The model's focus on topologically critical points (TCPs) represented a paradigm shift from traditional persistent homology-based methods. By emphasizing

0-simplices (points) rather than higher-dimensional features, TopoSegNet achieved improved computational efficiency while maintaining high segmentation accuracy. The model demonstrated consistent improvements across multiple datasets, with particular success in preserving fine structural details in both agricultural field boundaries and dendrological vessels. The introduction of the Centroid Displacement Measure (CDM) provided a new metric for evaluating both geometric and topological accuracy in segmentation tasks. However, TopoSegNet exhibited several key limitations in its implementation and application. To overcome the sensitivity of the terminal node detection process to the thickness of linear features in input masks, TopoSegNet implemented a three-level erosion operation within the TCP detection workflow. Although this adaptation made TopoSegNet less sensitive to line thickness, the selection of the number of erosion levels remained a manual process. This highlighted the model's sensitivity to erosion levels, as demonstrated in the six-erosion experiments where fine details were often lost with too many erosion iterations (Figure 4.9). This sensitivity necessitated careful selection of erosion levels based on dataset characteristics, potentially limiting the model's generalizability. Another notable limitation was the model's reduced applicability to datasets where the foreground consists of areas inside loops rather than boundaries themselves. While this could potentially be addressed through edge detection techniques, it represented a constraint on the model's versatility. Additionally, the detection of junction points remained challenging, particularly for rounded structures, as the model relied on mask inversion for TCP detection, which could lead to missed junctions when mask structures lacked sharp features.

Chapter 5 brought these developments together by investigating the impact of TopoSinGAN-generated synthetic data on segmentation performance. The study revealed that augmenting training datasets with topologically-accurate synthetic images improved segmentation performance for both UNet and TopoSegNet. The research demonstrated that optimal results were achieved with moderate proportions of synthetic data (10-25%), while higher proportions could lead to diminishing returns. The effect size analysis revealed different patterns of improvement between the agricultural and dendrological datasets, suggesting that the effectiveness of synthetic data augmentation varies with dataset characteristics. However, several limitations in the experimental design warrant consideration. The random selection of base images, while providing a baseline for evaluation, may not have captured the full range of topological variations present in the datasets. A more systematic approach to base image selection could potentially enhance the representation of predominant topological and contextual variations. Additionally, the inherent complexity of features within each base image potentially limited the generalizability of the synthetic data generated. The study was also constrained by practical limitations in the number of base images used, being restricted to 10 images due to computational and time constraints. A more comprehensive analysis with a larger number of base images could provide deeper insights into the relationship between base image quantity and augmentation effectiveness. Furthermore, the binning of synthetic data proportions into four groups (10%, 25%, 50%, and 75%) while sufficient for demonstrating general trends, may have missed more nuanced relationships between synthetic data proportion and model performance that could be revealed through finer-grained analysis.

6.2 Conclusion

This dissertation advances the integration of topological awareness into deep learning through novel approaches in image generation and segmentation. Through the development of TopoSinGAN and TopoSegNet, the research demonstrates that incorporating topological constraints can significantly enhance the accuracy and reliability of deep learning models for spatial image processing. The developed frameworks introduce differentiable topology loss functions that effectively preserve structural integrity while maintaining computational efficiency, addressing a critical gap in existing approaches. The comprehensive evaluation on complex spatial datasets, including agricultural fields and dendrological imagery, validates these approaches' effectiveness in real-world applications. The research makes several key contributions to the field of topology-aware deep learning. First, it establishes new methodologies for incorporating topological constraints into both generative and segmentation models, demonstrating that structural accuracy can be maintained without sacrificing computational efficiency. The introduction of two novel metrics - Node Topology Clustering (NTC) for quantifying topological integrity in generated images, and Centroid Displacement Measure (CDM) for assessing both geometric and topological accuracy in segmentation tasks - provides essential tools for evaluating structural consistency in deep learning outputs. The success of TopoSinGAN in generating topologically accurate synthetic images from a single source image represents a significant advancement in data augmentation capabilities, particularly valuable in domains where annotated data is scarce. The model's ability to maintain structural consistency while generating diverse variations demonstrates the potential for expanding limited datasets without compromising topological integrity. Similarly, TopoSegNet's focus on topologically critical points introduces a more efficient approach to preserving structural features in segmentation tasks, offering a practical alternative to computationally intensive persistent homology-based methods. The practical applications demonstrated through agricultural field delineation and dendrology analysis showcase the versatility and effectiveness of these approaches across different domains. In agricultural applications, the improved boundary detection and structural preservation capabilities contribute to more accurate field mapping and monitoring. In dendrology, the enhanced ability to delineate complex vessel structures advances the potential for automated analysis of tree-ring imagery, supporting research in climate science and forest ecology. While the developed methods show limitations in certain scenarios, such as handling open structures or varying feature thicknesses, they represent significant progress toward more robust and topology-aware deep learning approaches. The research establishes a foundation for future developments in topology-aware deep learning, particularly in applications requiring precise structural preservation. The demonstrated improvements in both computational efficiency and topological accuracy suggest promising directions for further research and practical applications across various fields requiring accurate structural analysis and representation. This work contributes to bridging the gap between theoretical topology and practical deep learning applications, offering scalable solutions for maintaining structural integrity in both generated and segmented images. The successful implementation and validation of these approaches across different applications establish a framework for future developments in topology-aware deep learning, potentially impacting fields ranging from remote sensing and medical imaging to environmental monitoring and beyond.

6.3 Future Directions

The research presented in this dissertation opens several promising avenues for future exploration:

Extension of Topology-Aware Generative Models The TopoSinGAN model could be adapted to handle scenarios where open structures are expected, such as road networks or branching patterns in medical imagery. This would involve modifying the topology loss function to maintain a desired number of terminal nodes rather than minimizing them entirely. Additionally, future work could focus on making the model more robust to variations in boundary thickness, potentially through the development of adaptive kernels or multi-scale topology detection approaches. Moreover, the novel topology loss function introduced in Chapter 3 has been integrated into the multi-scale architecture of the original SinGAN model, designed to learn from a single image. Future work could extend this approach by incorporating the loss function into general adversarial frameworks trained on a dataset of images.

Enhanced Segmentation Framework Future developments of TopoSegNet could explore the integration of adaptive erosion levels based on the intrinsic characteristics of input images. This could improve the model's performance across datasets with varying feature thicknesses and structural complexities. Another promising direction is the development of methods to better handle datasets where the foreground consists of areas inside loops rather than boundaries, potentially through automated boundary extraction techniques.

Data Augmentation Strategies Further research could investigate optimal strategies for synthetic data generation by focusing on several key areas. One critical direction is determining the ideal ratios of synthetic to real data for different application domains, as these ratios likely vary based on the complexity and characteristics of each field. Additionally, developing methods to automatically select the most effective base images for synthesis would streamline the data generation process and improve output quality. This could be complemented by creating hybrid approaches that combine traditional augmentation techniques with topology-aware generation, potentially leading to more robust and diverse training datasets.

Evaluation Metrics The Node Topology Clustering (NTC) and Centroid Displacement Measure (CDM) metrics could be enhanced to address current limitations and expand their applicability. A key improvement of CDM metric would be developing better mechanisms for handling over-segmentation scenarios, where current metric may not fully capture the complexity of structural errors. Additionally, extending these metrics to incorporate three-dimensional structural analysis for volumetric data would broaden their utility in medical imaging and materials science. Further development could also focus on accounting for scale-dependent topological features, making these metrics more versatile across different spatial scales and resolutions.

Application Domains The methods developed in this dissertation show promise for extension into several new application areas, each presenting unique challenges and opportunities. In urban planning, these techniques could enhance the analysis of transportation networks and building footprints, while in neuroscience, they could improve the mapping of neural pathways and connectivity patterns. In medical imaging, these approaches could advance the segmentation of anatomical structures such as blood vessels and organ boundaries, where topological accuracy is crucial for diagnosis and treatment planning. Additionally, these methods could be particularly valuable for analyzing medical scans across different modalities (CT, MRI, ultrasound), where preserving the structural integrity of anatomical features is essential for clinical decision-making. Each of these domains presents specific requirements that could drive further methodological advances in topology-aware deep learning.

These future directions aim to address current limitations while expanding the applicability and effectiveness of topology-aware deep learning methods across diverse fields and applications. The continued development of these approaches will contribute to more robust and accurate image analysis tools for both research and practical applications.

References

[1] Aditya Ramesh, Mikhail Pavlov, Gabriel Goh, Scott Gray, Chelsea Voss, Alec Radford, Mark Chen, and Ilya Sutskever. Zero-shot text-to-image generation. In *International conference on machine learning*, pages 8821–8831. Pmlr, 2021.

[2] Alexander Del Toro Barba. Topological data analysis with persistent homology. `http://bit.ly/3BEYRK6`, 2024. Accessed: 2024-12-18.

[3] Navketan Batra and Goutam Sheet. Physics with coffee and doughnuts: Understanding the physics behind topological insulators through su-schrieffer-heeger model. *Resonance*, 25(6):765–786, 2020.

[4] Fan Wang, Huidong Liu, Dimitris Samaras, and Chao Chen. Topogan: A topology-aware generative adversarial network. In *Computer Vision–ECCV 2020: 16th European Conference, Glasgow, UK, August 23–28, 2020, Proceedings, Part III 16*, pages 118–136. Springer, 2020.

[5] Tamar Rott Shaham, Tali Dekel, and Tomer Michaeli. Singan: Learning a generative model from a single natural image. In *Proceedings of the IEEE/CVF international conference on computer vision*, pages 4570–4580, 2019.

[6] Olaf Ronneberger, Philipp Fischer, and Thomas Brox. U-net: Convolutional networks for biomedical image segmentation. In *Medical image computing and computer-assisted intervention–MICCAI 2015: 18th international conference, Munich, Germany, October 5-9, 2015, proceedings, part III 18*, pages 234–241. Springer, 2015.

[7] Xiaoling Hu, Fuxin Li, Dimitris Samaras, and Chao Chen. Topology-preserving deep image segmentation. *Advances in neural information processing systems*, 32, 2019.

[8] JP Agrawal, BJ Erickson, and CE Kahn Jr. Imaging informatics: 25 years of progress. *Yearbook of medical informatics*, 25(S 01):S23–S31, 2016.

[9] Bing Zhang, Yuanfeng Wu, Boya Zhao, Jocelyn Chanussot, Danfeng Hong, Jing Yao, and Lianru Gao. Progress and challenges in intelligent remote sensing satellite systems. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, 15:1814–1822, 2022.

[10] Ioannis Kotaridis and Maria Lazaridou. Remote sensing image segmentation advances: A meta-analysis. *ISPRS Journal of Photogrammetry and Remote Sensing*, 173:309–322, 2021.

[11] Chen Chen, Chen Qin, Huaqi Qiu, Giacomo Tarroni, Jinming Duan, Wenjia Bai, and Daniel Rueckert. Deep learning for cardiac image segmentation: a review. *Frontiers in cardiovascular medicine*, 7:25, 2020.

[12] Mohammad D Hossain and Dongmei Chen. Segmentation for object-based image analysis (obia): A review of algorithms and challenges from remote sensing perspective. *ISPRS Journal of Photogrammetry and Remote Sensing*, 150:115–134, 2019.

[13] Danijela Horak, Simiao Yu, and Gholamreza Salimi-Khorshidi. Topology distance: A topology-based approach for evaluating generative adversarial networks. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 35, pages 7721–7728, 2021.

[14] Xiaoling Hu, Yusu Wang, Li Fuxin, Dimitris Samaras, and Chao Chen. Topology-aware segmentation using discrete morse theory. *arXiv preprint arXiv:2103.09992*, 2021.

[15] Dragos Costea, Alina Marcu, Emil Slusanschi, and Marius Leordeanu. Creating roadmaps in aerial images with generative adversarial networks and smoothing-based

optimization. In *Proceedings of the IEEE International Conference on Computer Vision Workshops*, pages 2100–2109, 2017.

[16] Edward Park et al. *Refining inferred road maps using GANs.* PhD thesis, Massachusetts Institute of Technology, 2019.

[17] Chuni Liu, Boyuan Ma, Xiaojuan Ban, Yujie Xie, Hao Wang, Weihua Xue, Jingchao Ma, and Ke Xu. Enhancing boundary segmentation for topological accuracy with skeleton-based methods. *arXiv preprint arXiv:2404.18539*, 2024.

[18] Agata Mosinska, Pablo Marquez-Neila, Mateusz Koziński, and Pascal Fua. Beyond the pixel-wise loss for topology-aware delineation. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 3136–3145, 2018.

[19] Viren Jain, Benjamin Bollmann, Mark Richardson, Daniel R Berger, Moritz N Helmstaedter, Kevin L Briggman, Winfried Denk, Jared B Bowden, John M Mendenhall, Wickliffe C Abraham, et al. Boundary learning by optimization with topological constraints. In *2010 IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, pages 2488–2495. IEEE, 2010.

[20] Xinyang Liu, Pierre-Louis Bazin, Aaron Carass, and Jerry Prince. Topology preserving brain tissue segmentation using graph cuts. In *2012 IEEE Workshop on Mathematical Methods in Biomedical Image Analysis*, pages 185–190. IEEE, 2012.

[21] Martin Heusel, Hubert Ramsauer, Thomas Unterthiner, Bernhard Nessler, and Sepp Hochreiter. Gans trained by a two time-scale update rule converge to a local nash equilibrium. *Advances in neural information processing systems*, 30, 2017.

[22] Hao Wang, Guosheng Lin, Steven CH Hoi, and Chunyan Miao. Structure-aware generation network for recipe generation from images. In *Computer Vision–ECCV 2020: 16th European Conference, Glasgow, UK, August 23–28, 2020, Proceedings, Part XXVII 16*, pages 359–374. Springer, 2020.

[23] Baoyu Zhu, Qunbo Lv, Yuanbo Yang, Kai Zhang, Xuefu Sui, Yinhui Tang, and Zheng Tan. Gradient structure information-guided attention generative adversarial networks for remote sensing image generation. *Remote Sensing*, 15(11):2827, 2023.

[24] Andac Demir, Elie Massaad, and Bulent Kiziltan. Topology-aware focal loss for 3d image segmentation. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 580–589, 2023.

[25] Kai Geng, Xian Sun, Zhiyuan Yan, Wenhui Diao, and Xin Gao. Topological space knowledge distillation for compact road extraction in optical remote sensing images. *Remote Sensing*, 12(19):3175, 2020.

[26] Xingang Li, Yuebin Wang, Liqiang Zhang, Suhong Liu, Jie Mei, and Yang Li. Topology-enhanced urban road extraction via a geographic feature-enhanced network. *IEEE Transactions on Geoscience and Remote Sensing*, 58(12):8819–8830, 2020.

[27] Juan Nunez-Iglesias, Ryan Kennedy, Toufiq Parag, Jianbo Shi, and Dmitri B Chklovskii. Machine learning of hierarchical clustering to segment 2d and 3d images. *PloS one*, 8(8):e71715, 2013.

[28] Wenying Du, Nengcheng Chen, and Dandan Liu. Topology adaptive water boundary extraction based on a modified balloon snake: using gf-1 satellite images as an example. *Remote Sensing*, 9(2):140, 2017.

[29] Loris Nanni, Sheryl Brahnam, and Andrea Loreggia. An enhanced loss function for semantic road segmentation in remote sensing images. *IEEE Access*, 2024.

[30] François Waldner and Foivos I Diakogiannis. Deep learning on edge: Extracting field boundaries from satellite images with a convolutional neural network. *Remote sensing of environment*, 245:111741, 2020.

[31] Ende Wang, Yanmei Jiang, Yong Li, Jingchao Yang, Mengcheng Ren, and Qingchun Zhang. Mfcsnet: Multi-scale deep features fusion and cost-sensitive loss function based segmentation network for remote sensing images. *Applied Sciences*, 9(19):4043, 2019.

[32] Chuanhao Wei, Dezhao Kong, Xuelian Sun, and Yu Zhou. Rsfnet: a method for remote sensing image semantic segmentation based on fully convolutional neural

networks. In *Third International Conference on Image Processing and Intelligent Control (IPIC 2023)*, volume 12782, pages 350–355. SPIE, 2023.

[33] Favyen Bastani, Songtao He, Sofiane Abbar, Mohammad Alizadeh, Hari Balakrishnan, Sanjay Chawla, Sam Madden, and David DeWitt. Roadtracer: Automatic extraction of road networks from aerial images. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 4720–4728, 2018.

[34] Tamal K Dey, Jiayuan Wang, and Yusu Wang. Road network reconstruction from satellite images with machine learning supported by topological methods. In *Proceedings of the 27th ACM SIGSPATIAL International Conference on Advances in Geographic Information Systems*, pages 520–523, 2019.

[35] Gellért Máttyus, Wenjie Luo, and Raquel Urtasun. Deeproadmapper: Extracting road topology from aerial images. In *Proceedings of the IEEE international conference on computer vision*, pages 3438–3446, 2017.

[36] Robin Vandaele, Guillaume Adrien Nervo, and Olivier Gevaert. Topological image modification for object detection and topological image processing of skin lesions. *Scientific Reports*, 10(1):21061, 2020.

[37] Nick Byrne, James R Clough, Giovanni Montana, and Andrew P King. A persistent homology-based topological loss function for multi-class cnn segmentation of cardiac mri. In *Statistical Atlases and Computational Models of the Heart. M&Ms and EMIDEC Challenges: 11th International Workshop, STACOM 2020, Held in Conjunction with MICCAI 2020, Lima, Peru, October 4, 2020, Revised Selected Papers 11*, pages 3–13. Springer, 2021.

[38] Hongliang He, Jun Wang, Pengxu Wei, Fan Xu, Xiangyang Ji, Chang Liu, and Jie Chen. Toposeg: Topology-aware nuclear instance segmentation. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pages 21307–21316, 2023.

[39] Yufan He, Dong Yang, Holger Roth, Can Zhao, and Daguang Xu. Dints: Differentiable neural network topology search for 3d medical image segmentation. In

Proceedings of the IEEE/CVF conference on computer vision and pattern recognition, pages 5841–5850, 2021.

[40] Suprosanna Shit, Johannes C Paetzold, Anjany Sekuboyina, Ivan Ezhov, Alexander Unger, Andrey Zhylka, Josien PW Pluim, Ulrich Bauer, and Bjoern H Menze. cldice-a novel topology-preserving loss function for tubular structure segmentation. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pages 16560–16569, 2021.

[41] Guangpu Wang, Peng Zhou, Hui Gao, Zewei Qin, Shuo Wang, Jinglai Sun, and Hui Yu. Coronary vessel segmentation in coronary angiography with a multi-scale u-shaped transformer incorporating boundary aggregation and topology preservation. *Physics in Medicine & Biology*, 69(2):025012, 2024.

[42] Huei-Fang Yang and Yoonsuck Choe. Ground truth estimation by maximizing topological agreements in electron microscopy data. In *International Symposium on Visual Computing*, pages 371–380. Springer, 2011.

[43] Jiaqi Yang, Xiaoling Hu, Chao Chen, and Chialing Tsai. A topology-attention convlstm network and its application to em images. *arXiv preprint arXiv:2202.03430*, 2022.

[44] Daoping Zhang and Lok Ming Lui. Topology-preserving 3d image segmentation based on hyperelastic regularization. *Journal of Scientific Computing*, 87(3):74, 2021.

[45] Afra Zomorodian and Gunnar Carlsson. Computing persistent homology. In *Proceedings of the twentieth annual symposium on Computational geometry*, pages 347–356, 2004.

[46] Abolfazl Abdollahi, Biswajeet Pradhan, Gaurav Sharma, Khairul Nizam Abdul Maulud, and Abdullah Alamri. Improving road semantic segmentation using generative adversarial network. *IEEE Access*, 9:64381–64392, 2021.

[47] Kenneth R Castleman. *Digital image processing.* Prentice Hall Press, 1996.

[48] Gustavo Camps-Valls, Devis Tuia, Luis Gómez-Chova, Sandra Jiménez, and Jesús Malo. *Remote sensing image processing.* Springer, 2011.

[49] Wei Yi, Yuhao Wang, Yong Zeng, Yaqin Wang, and Jianfei Xu. Comprehensive evaluation of the gf-4 satellite image quality from 2015 to 2020. *ISPRS International Journal of Geo-Information*, 10(6):406, 2021.

[50] Vanessa Robins, Peter John Wood, and Adrian P Sheppard. Theory and algorithms for constructing discrete morse complexes from grayscale digital images. *IEEE Transactions on pattern analysis and machine intelligence*, 33(8):1646–1658, 2011.

[51] P Soille. Morphological image analysis: Principles and applications, 1999.

[52] William F Basener. *Topology and its Applications*, volume 70. John Wiley & Sons, 2006.

[53] Anna-Katharina Lautenschütz, Clare Davies, Martin Raubal, Angela Schwering, and Eric Pederson. The influence of scale, context and spatial preposition in linguistic topology. In *Spatial Cognition V Reasoning, Action, Interaction: International Conference Spatial Cognition 2006, Bremen, Germany, September 24-28, 2006, Revised Selected Papers 5*, pages 439–452. Springer, 2007.

[54] Lenord Talmy. *Toward a cognitive semantics: Concept structuring systems.* MIT press, 2000.

[55] Xiaoling Hu. Structure-aware image segmentation with homotopy warping. *Advances in Neural Information Processing Systems*, 35:24046–24059, 2022.

[56] Xuejun Guo and Ruisen Zhou. Data augmentation method for extracting partially occluded roads from high spatial resolution remote sensing images. *IEEE Access*, 2023.

[57] James R Clough, Nicholas Byrne, Ilkay Oksuz, Veronika A Zimmer, Julia A Schnabel, and Andrew P King. A topological loss function for deep-learning based image segmentation using persistent homology. *IEEE transactions on pattern analysis and machine intelligence*, 44(12):8766–8778, 2020.

[58] Madeleine K Wyburd, Nicola K Dinsdale, Ana IL Namburete, and Mark Jenkinson. Teds-net: enforcing diffeomorphisms in spatial transformers to guarantee topology

preservation in segmentations. In *International Conference on Medical Image Computing and Computer-Assisted Intervention*, pages 250–260. Springer, 2021.

[59] Hrishi Patel, Colleen Farrelly, Quincy A Hathaway, Jennifer Z Rozenblit, Deepa Deepa, Yashbir Singh, Ashok Chaudhary, Yassine Himeur, Wathiq Mansoor, and Shadi Atalls. Topology-aware gan (topogan): Transforming medical imaging advances. In *2023 Tenth International Conference on Social Networks Analysis, Management and Security (SNAMS)*, pages 1–3. IEEE, 2023.

[60] R Haralick. Computer and robot vision. *Addison-Wesley Publishing Company google schola*, 2:2061–2068, 1992.

[61] Gustavo Camps-Valls. Machine learning in remote sensing data processing. In *2009 IEEE international workshop on machine learning for signal processing*, pages 1–6. IEEE, 2009.

[62] Mohamed A Hamada, Yeleussiz Kanat, and Adejor Egahi Abiche. Multi-spectral image segmentation based on the k-means clustering. *Int. J. Innov. Technol. Explor. Eng*, 9(2):1016–1019, 2019.

[63] Sveinn R Joelsson, Jon A Benediktsson, and Johannes R Sveinsson. Random forest classification of remote sensing data. In *Signal and Image Processing for Remote Sensing*, pages 344–361. CRC Press, 2006.

[64] Caroline Petitjean, Maria A Zuluaga, Wenjia Bai, Jean-Nicolas Dacher, Damien Grosgeorge, Jérôme Caudron, Su Ruan, Ismail Ben Ayed, M Jorge Cardoso, Hsiang-Chou Chen, et al. Right ventricle segmentation from cardiac mri: a collation study. *Medical image analysis*, 19(1):187–202, 2015.

[65] Guozhu Dong and Huan Liu. *Feature engineering for machine learning and data analytics*. CRC press, 2018.

[66] Vijay Badrinarayanan, Alex Kendall, and Roberto Cipolla. Segnet: A deep convolutional encoder-decoder architecture for image segmentation. *IEEE transactions on pattern analysis and machine intelligence*, 39(12):2481–2495, 2017.

[67] Jonathan Long, Evan Shelhamer, and Trevor Darrell. Fully convolutional networks for semantic segmentation. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 3431–3440, 2015.

[68] Stefan Ainetter and Friedrich Fraundorfer. End-to-end trainable deep neural network for robotic grasp detection and semantic segmentation from rgb. In *2021 IEEE International Conference on Robotics and Automation (ICRA)*, pages 13452–13458. IEEE, 2021.

[69] Bowen Pan, Jiankai Sun, Ho Yin Tiga Leung, Alex Andonian, and Bolei Zhou. Cross-view semantic segmentation for sensing surroundings. *IEEE Robotics and Automation Letters*, 5(3):4867–4873, 2020.

[70] Aijin Li, Licheng Jiao, Hao Zhu, Lingling Li, and Fang Liu. Multitask semantic boundary awareness network for remote sensing image segmentation. *IEEE Transactions on Geoscience and Remote Sensing*, 60:1–14, 2021.

[71] Intisar Rizwan I Haque and Jeremiah Neubert. Deep learning approaches to biomedical image segmentation. *Informatics in Medicine Unlocked*, 18:100297, 2020.

[72] Weihao Weng and Xin Zhu. Inet: convolutional networks for biomedical image segmentation. *Ieee Access*, 9:16591–16603, 2021.

[73] Hoel Kervadec, Jihene Bouchtiba, Christian Desrosiers, Eric Granger, Jose Dolz, and Ismail Ben Ayed. Boundary loss for highly unbalanced segmentation. In *International conference on medical imaging with deep learning*, pages 285–296. PMLR, 2019.

[74] Levi Westerveld and Anne K Knowles. Loosening the grid: topology as the basis for a more inclusive gis. *International Journal of Geographical Information Science*, 35(10):2108–2127, 2021.

[75] James R Clough, Nicholas Byrne, Ilkay Oksuz, Veronika A Zimmer, Julia A Schnabel, and Andrew P King. A topological loss function for deep-learning based image segmentation using persistent homology. *IEEE transactions on pattern analysis and machine intelligence*, 44(12):8766–8778, 2020.

[76] Alexander Kirillov, Eric Mintun, Nikhila Ravi, Hanzi Mao, Chloe Rolland, Laura Gustafson, Tete Xiao, Spencer Whitehead, Alexander C Berg, Wan-Yen Lo, et al. Segment anything. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pages 4015–4026, 2023.

[77] Yujian Mo, Yan Wu, Xinneng Yang, Feilin Liu, and Yujun Liao. Review the state-of-the-art technologies of semantic segmentation based on deep learning. *Neurocomputing*, 493:626–646, 2022.

[78] Xiaojun Shan, Ping Tang, and Changmiao Hu. An automatic geometric precision correction system based on hierarchical registration for hj-1 a/b ccd images. *International Journal of Remote Sensing*, 35(20):7154–7178, 2014.

[79] Paolo Andreini, Simone Bonechi, Monica Bianchini, Alessandro Mecocci, and Franco Scarselli. Image generation by gan and style transfer for agar plate image segmentation. *Computer methods and programs in biomedicine*, 184:105268, 2020.

[80] ZengShun Zhaoa, Yulong Wang, Ke Liu, Haoran Yang, Qian Sun, and Heng Qiao. Semantic segmentation by improved generative adversarial networks. *arXiv preprint arXiv:2104.09917*, 2021.

[81] Kaiyang Zhou, Yongxin Yang, Timothy Hospedales, and Tao Xiang. Learning to generate novel domains for domain generalization. In *Computer Vision–ECCV 2020: 16th European Conference, Glasgow, UK, August 23–28, 2020, Proceedings, Part XVI 16*, pages 561–578. Springer, 2020.

[82] Kaiming He, Xiangyu Zhang, Shaoqing Ren, and Jian Sun. Identity mappings in deep residual networks. In *Computer Vision–ECCV 2016: 14th European Conference, Amsterdam, The Netherlands, October 11–14, 2016, Proceedings, Part IV 14*, pages 630–645. Springer, 2016.

[83] Liang-Chieh Chen, George Papandreou, Iasonas Kokkinos, Kevin Murphy, and Alan L Yuille. Deeplab: Semantic image segmentation with deep convolutional nets, atrous convolution, and fully connected crfs. *IEEE transactions on pattern analysis and machine intelligence*, 40(4):834–848, 2017.

[84] Khairiya Mudrik Masoud, Claudio Persello, and Valentyn A Tolpekin. Delineation of agricultural field boundaries from sentinel-2 images using a novel super-resolution contour detector based on fully convolutional networks. *Remote sensing*, 12(1):59, 2019.

[85] Han Lin Aung, Burak Uzkent, Marshall Burke, David Lobell, and Stefano Ermon. Farm parcel delineation using spatio-temporal convolutional networks. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition workshops*, pages 76–77, 2020.

[86] Alireza Taravat, Matthias P Wagner, Rogerio Bonifacio, and David Petit. Advanced fully convolutional networks for agricultural field boundary detection. *Remote sensing*, 13(4):722, 2021.

[87] Huanxue Zhang, Mingxu Liu, Yuji Wang, Jiali Shang, Xiangliang Liu, Bin Li, Aiqi Song, and Qiangzi Li. Automated delineation of agricultural field boundaries from sentinel-2 images using recurrent residual u-net. *International Journal of Applied Earth Observation and Geoinformation*, 105:102557, 2021.

[88] Lu Xu, Dongping Ming, Tongyao Du, Yangyang Chen, Dehui Dong, and Chenghu Zhou. Delineation of cultivated land parcels based on deep convolutional networks and geographical thematic scene division of remotely sensed images. *Computers and Electronics in Agriculture*, 192:106611, 2022.

[89] Liang-Chieh Chen, Yukun Zhu, George Papandreou, Florian Schroff, and Hartwig Adam. Encoder-decoder with atrous separable convolution for semantic image segmentation. In *Proceedings of the European conference on computer vision (ECCV)*, pages 801–818, 2018.

[90] F Yu. Multi-scale context aggregation by dilated convolutions. *arXiv preprint arXiv:1511.07122*, 2015.

[91] Henghui Ding, Xudong Jiang, Bing Shuai, Ai Qun Liu, and Gang Wang. Semantic segmentation with context encoding and multi-path decoding. *IEEE Transactions on Image Processing*, 29:3520–3533, 2020.

[92] Jun Fu, Jing Liu, Jie Jiang, Yong Li, Yongjun Bao, and Hanqing Lu. Scene segmentation with dual relation-aware attention network. *IEEE Transactions on Neural Networks and Learning Systems*, 32(6):2547–2560, 2020.

[93] Yunsheng Li, Lu Yuan, and Nuno Vasconcelos. Bidirectional learning for domain adaptation of semantic segmentation. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pages 6936–6945, 2019.

[94] Towaki Takikawa, David Acuna, Varun Jampani, and Sanja Fidler. Gated-scnn: Gated shape cnns for semantic segmentation. In *Proceedings of the IEEE/CVF international conference on computer vision*, pages 5229–5238, 2019.

[95] Tao Song, Xudong Zhang, Mao Ding, Alfonso Rodriguez-Paton, Shudong Wang, and Gan Wang. Deepfusion: A deep learning based multi-scale feature fusion method for predicting drug-target interactions. *Methods*, 204:269–277, 2022.

[96] Maxwell Jong, Kaiyu Guan, Sibo Wang, Yizhi Huang, and Bin Peng. Improving field boundary delineation in resunets via adversarial deep learning. *International Journal of Applied Earth Observation and Geoinformation*, 112:102877, 2022.

[97] Reza Azad, Amirhossein Kazerouni, Babak Azad, Ehsan Khodapanah Aghdam, Yury Velichko, Ulas Bagci, and Dorit Merhof. Laplacian-former: Overcoming the limitations of vision transformers in local texture detection. In *International Conference on Medical Image Computing and Computer-Assisted Intervention*, pages 736–746. Springer, 2023.

[98] Peihao Wang, Wenqing Zheng, Tianlong Chen, and Zhangyang Wang. Anti-oversmoothing in deep vision transformers via the fourier domain analysis: From theory to practice. *arXiv preprint arXiv:2203.05962*, 2022.

[99] Jieneng Chen, Yongyi Lu, Qihang Yu, Xiangde Luo, Ehsan Adeli, Yan Wang, Le Lu, Alan L Yuille, and Yuyin Zhou. Transunet: Transformers make strong encoders for medical image segmentation. *arXiv preprint arXiv:2102.04306*, 2021.

[100] Moein Heidari, Amirhossein Kazerouni, Milad Soltany, Reza Azad, Ehsan Khodapanah Aghdam, Julien Cohen-Adad, and Dorit Merhof. Hiformer:

Hierarchical multi-scale representations using transformers for medical image segmentation. In *Proceedings of the IEEE/CVF winter conference on applications of computer vision*, pages 6202–6212, 2023.

[101] Jiawang Bai, Li Yuan, Shu-Tao Xia, Shuicheng Yan, Zhifeng Li, and Wei Liu. Improving vision transformers by revisiting high-frequency components. In *European Conference on Computer Vision*, pages 1–18. Springer, 2022.

[102] Ian Goodfellow, Jean Pouget-Abadie, Mehdi Mirza, Bing Xu, David Warde-Farley, Sherjil Ozair, Aaron Courville, and Yoshua Bengio. Generative adversarial nets. *Advances in neural information processing systems*, 27, 2014.

[103] Hongyu Liu, Ziyu Wan, Wei Huang, Yibing Song, Xintong Han, and Jing Liao. Pd-gan: Probabilistic diverse gan for image inpainting. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pages 9371–9381, 2021.

[104] Yuxi Ren, Jie Wu, Peng Zhang, Manlin Zhang, Xuefeng Xiao, Qian He, Rui Wang, Min Zheng, and Xin Pan. Ugc: Unified gan compression for efficient image-to-image translation. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pages 17281–17291, 2023.

[105] Mingle Xu, Sook Yoon, Alvaro Fuentes, and Dong Sun Park. A comprehensive survey of image augmentation techniques for deep learning. *Pattern Recognition*, 137:109347, 2023.

[106] Dwarikanath Mahapatra and Zongyuan Ge. Training data independent image registration with gans using transfer learning and segmentation information. In *2019 IEEE 16th International Symposium on Biomedical Imaging (ISBI 2019)*, pages 709–713. IEEE, 2019.

[107] Mayank Jain, Conor Meegan, and Soumyabrata Dev. Using gans to augment data for cloud image segmentation task. In *2021 IEEE International Geoscience and Remote Sensing Symposium IGARSS*, pages 3452–3455. IEEE, 2021.

[108] Michael Majurski, Petru Manescu, Sarala Padi, Nicholas Schaub, Nathan Hotaling, Carl Simon Jr, and Peter Bajcsy. Cell image segmentation using generative adversarial networks, transfer learning, and augmentations. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition workshops*, pages 0–0, 2019.

[109] Vajira Thambawita, Pegah Salehi, Sajad Amouei Sheshkal, Steven A Hicks, Hugo L Hammer, Sravanthi Parasa, Thomas de Lange, Pål Halvorsen, and Michael A Riegler. Singan-seg: Synthetic training data generation for medical image segmentation. *PloS one*, 17(5):e0267976, 2022.

[110] Chendan Wang, Bowen Chen, Zhengxia Zou, and Zhenwei Shi. Remote sensing image synthesis via semantic embedding generative adversarial networks. *IEEE Transactions on Geoscience and Remote Sensing*, 61:1–11, 2023.

[111] Daniel Griffin, Samantha T Porter, Matthew L Trumper, Kate E Carlson, Daniel J Crawford, Daniel Schwalen, and Colin H McFadden. Gigapixel macro photography of tree rings. *Tree-Ring Research*, 77(2):86–94, 2021.

[112] Ignacio Arganda-Carreras, Srinivas C Turaga, Daniel R Berger, Dan Cireşan, Alessandro Giusti, Luca M Gambardella, Jürgen Schmidhuber, Dmitry Laptev, Sarvesh Dwivedi, Joachim M Buhmann, et al. Crowdsourcing the creation of image segmentation algorithms for connectomics. *Frontiers in neuroanatomy*, 9:152591, 2015.

[113] Joes Staal, Michael D Abràmoff, Meindert Niemeijer, Max A Viergever, and Bram Van Ginneken. Ridge-based vessel segmentation in color images of the retina. *IEEE transactions on medical imaging*, 23(4):501–509, 2004.

[114] Jia Deng, Wei Dong, Richard Socher, Li-Jia Li, Kai Li, and Li Fei-Fei. Imagenet: A large-scale hierarchical image database. In *2009 IEEE conference on computer vision and pattern recognition*, pages 248–255. Ieee, 2009.

[115] Chenyu You, Guang Li, Yi Zhang, Xiaoliu Zhang, Hongming Shan, Mengzhou Li, Shenghong Ju, Zhen Zhao, Zhuiyang Zhang, Wenxiang Cong, et al. Ct

super-resolution gan constrained by the identical, residual, and cycle learning ensemble (gan-circle). *IEEE transactions on medical imaging*, 39(1):188–203, 2019.

[116] Jilyan Bianca Dy, John Jethro Virtusio, Daniel Stanley Tan, Yong-Xiang Lin, Joel Ilao, Yung-Yao Chen, and Kai-Lung Hua. Mcgan: mask controlled generative adversarial network for image retargeting. *Neural Computing and Applications*, 35(14):10497–10509, 2023.

[117] Ishaan Gulrajani, Faruk Ahmed, Martin Arjovsky, Vincent Dumoulin, and Aaron C Courville. Improved training of wasserstein gans. *Advances in neural information processing systems*, 30, 2017.

[118] Martin Arjovsky, Soumith Chintala, and Léon Bottou. Wasserstein generative adversarial networks. In *International conference on machine learning*, pages 214–223. PMLR, 2017.

[119] Jae Hyun Lim and Jong Chul Ye. Geometric gan. *arXiv preprint arXiv:1705.02894*, 2017.

[120] Guo-Jun Qi, Liheng Zhang, Hao Hu, Marzieh Edraki, Jingdong Wang, and Xian-Sheng Hua. Global versus localized generative adversarial nets. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 1517–1525, 2018.

[121] Jean Kossaifi, Linh Tran, Yannis Panagakis, and Maja Pantic. Gagan: Geometry-aware generative adversarial networks. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 878–887, 2018.

[122] Huan Fu, Mingming Gong, Chaohui Wang, Kayhan Batmanghelich, Kun Zhang, and Dacheng Tao. Geometry-consistent generative adversarial networks for one-sided unsupervised domain mapping. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pages 2427–2436, 2019.

[123] Wenting Chen, Shuang Yu, Junde Wu, Kai Ma, Cheng Bian, Chunyan Chu, Linlin Shen, and Yefeng Zheng. Tr-gan: Topology ranking gan with triplet loss for retinal artery/vein classification. In *Medical Image Computing and Computer Assisted*

Intervention–MICCAI 2020: 23rd International Conference, Lima, Peru, October 4–8, 2020, Proceedings, Part V 23, pages 616–625. Springer, 2020.

[124] Wenting Chen, Shuang Yu, Kai Ma, Wei Ji, Cheng Bian, Chunyan Chu, Linlin Shen, and Yefeng Zheng. Tw-gan: Topology and width aware gan for retinal artery/vein classification. *Medical Image Analysis*, 77:102340, 2022.

[125] Weiyi Liu, Pin-Yu Chen, Fucai Yu, Toyotaro Suzumura, and Guangmin Hu. Learning graph topological features via gan. *IEEE Access*, 7:21834–21843, 2019.

[126] Jinri Bao, Zicong Wang, Junli Wang, and Chungang Yan. Persistent homology based generative adversarial network. In *VISIGRAPP (4: VISAPP)*, pages 196–203, 2023.

[127] Christian Szegedy, Vincent Vanhoucke, Sergey Ioffe, Jon Shlens, and Zbigniew Wojna. Rethinking the inception architecture for computer vision. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 2818–2826, 2016.

[128] Tim Salimans, Ian Goodfellow, Wojciech Zaremba, Vicki Cheung, Alec Radford, and Xi Chen. Improved techniques for training gans. *Advances in neural information processing systems*, 29, 2016.

[129] Andrea Cerri, Barbara Di Fabio, Grzegorz Jabłoński, and Filippo Medri. Comparing shapes through multi-scale approximations of the matching distance. *Computer Vision and Image Understanding*, 121:43–56, 2014.

[130] Don Sheehy, Oliver Kisielius, and Nicholas J Cavanna. Computing the shift-invariant bottleneck distance for persistence diagrams. In *CCCG*, pages 78–84, 2018.

[131] Jérémie Bouttier, Philippe Di Francesco, and Emmanuel Guitter. Geodesic distance in planar graphs. *Nuclear physics B*, 663(3):535–567, 2003.

[132] Luc Anselin. Local indicators of spatial associationlisa. *Geographical analysis*, 27(2):93–115, 1995.

[133] Marti A. Hearst, Susan T Dumais, Edgar Osuna, John Platt, and Bernhard Scholkopf. Support vector machines. *IEEE Intelligent Systems and their applications*, 13(4):18–28, 1998.

[134] Sadegh Aliakbary, Jafar Habibi, and Ali Movaghar. Feature extraction from degree distribution for comparison and analysis of complex networks. *The Computer Journal*, 58(9):2079–2091, 2015.

[135] Niousha Attar and Sadegh Aliakbary. Classification of complex networks based on similarity of topological network features. *Chaos: An interdisciplinary journal of nonlinear science*, 27(9), 2017.

[136] ERDŐS PAUL and Alfréd Rényi. On the central limit theorem for samples from a finite population. *Selected papers of Alfréd Rényi*, page 353, 1959.

[137] Duncan J Watts and Steven H Strogatz. Collective dynamics of small-worldnetworks. *nature*, 393(6684):440–442, 1998.

[138] L da F Costa, Francisco A Rodrigues, Gonzalo Travieso, and Paulino Ribeiro Villas Boas. Characterization of complex networks: A survey of measurements. *Advances in physics*, 56(1):167–242, 2007.

[139] Mark EJ Newman. Modularity and community structure in networks. *Proceedings of the national academy of sciences*, 103(23):8577–8582, 2006.

[140] Uwe G Hacke, Rachel Spicer, Stefan G Schreiber, and Lenka Plavcová. An ecophysiological and developmental perspective on variation in vessel diameter. *Plant, cell & environment*, 40(6):831–845, 2017.

[141] Minnesota Supercomputing Institute. Research computing, research and innovation office. minnesota supercomputing institute (msi)-agate cluster. `https://msi.umn.edu/about-msi-services/high-performance-computing/agate`, 2024. Accessed: 2024-10-23.

[142] Yanchun Li, Nanfeng Xiao, and Wanli Ouyang. Improved generative adversarial networks with reconstruction loss. *Neurocomputing*, 323:363–372, 2019.

[143] Ömer Nebil Yaveroğlu, Tijana Milenković, and Nataša Pržulj. Proper evaluation of alignment-free network comparison methods. *Bioinformatics*, 31(16):2697–2704, 2015.

[144] Nataša Pržulj. Biological network comparison using graphlet degree distribution. *Bioinformatics*, 23(2):e177–e183, 2007.

[145] Oleksii Kuchaiev, Aleksandar Stevanović, Wayne Hayes, and Nataša Pržulj. Graphcrunch 2: software tool for network modeling, alignment and clustering. *BMC bioinformatics*, 12:1–13, 2011.

[146] Mohsen Ahmadkhani and Eric Shook. Toposingan github repository. `https://github.com/mohsenumn/TopoSinGAN`, 2024. Accessed: 2024-10-23.

[147] Saumya Gupta, Yikai Zhang, Xiaoling Hu, Prateek Prasanna, and Chao Chen. Topology-aware uncertainty for image segmentation. *Advances in Neural Information Processing Systems*, 36, 2024.

[148] Kaiming He, Georgia Gkioxari, Piotr Dollár, and Ross Girshick. Mask r-cnn. In *Proceedings of the IEEE international conference on computer vision*, pages 2961–2969, 2017.

[149] Moloud Abdar, Farhad Pourpanah, Sadiq Hussain, Dana Rezazadegan, Li Liu, Mohammad Ghavamzadeh, Paul Fieguth, Xiaochun Cao, Abbas Khosravi, U Rajendra Acharya, et al. A review of uncertainty quantification in deep learning: Techniques, applications and challenges. *Information fusion*, 76:243–297, 2021.

[150] Yarin Gal and Zoubin Ghahramani. Dropout as a bayesian approximation: Representing model uncertainty in deep learning. In *international conference on machine learning*, pages 1050–1059. PMLR, 2016.

[151] Jakob Gawlikowski, Cedrique Rovile Njieutcheu Tassi, Mohsin Ali, Jongseok Lee, Matthias Humt, Jianxiang Feng, Anna Kruspe, Rudolph Triebel, Peter Jung, Ribana Roscher, et al. A survey of uncertainty in deep neural networks. *Artificial Intelligence Review*, 56(Suppl 1):1513–1589, 2023.

[152] Chen Li, Xiaoling Hu, and Chao Chen. Confidence estimation using unlabeled data. *arXiv preprint arXiv:2307.10440*, 2023.

[153] Yarin Gal and Zoubin Ghahramani. Bayesian convolutional neural networks with bernoulli approximate variational inference. *arXiv preprint arXiv:1506.02158*, 2015.

[154] Alex Kendall and Yarin Gal. What uncertainties do we need in bayesian deep learning for computer vision? *Advances in neural information processing systems*, 30, 2017.

[155] Balaji Lakshminarayanan, Alexander Pritzel, and Charles Blundell. Simple and scalable predictive uncertainty estimation using deep ensembles. *Advances in neural information processing systems*, 30, 2017.

[156] Christian Rupprecht, Iro Laina, Robert DiPietro, Maximilian Baust, Federico Tombari, Nassir Navab, and Gregory D Hager. Learning in an uncertain world: Representing ambiguity through multiple hypotheses. In *Proceedings of the IEEE international conference on computer vision*, pages 3591–3600, 2017.

[157] Simon Kohl, Bernardino Romera-Paredes, Clemens Meyer, Jeffrey De Fauw, Joseph R Ledsam, Klaus Maier-Hein, SM Eslami, Danilo Jimenez Rezende, and Olaf Ronneberger. A probabilistic u-net for segmentation of ambiguous images. *Advances in neural information processing systems*, 31, 2018.

[158] Christian F Baumgartner, Kerem C Tezcan, Krishna Chaitanya, Andreas M Hötker, Urs J Muehlematter, Khoschy Schawkat, Anton S Becker, Olivio Donati, and Ender Konukoglu. Phiseg: Capturing uncertainty in medical image segmentation. In *Medical Image Computing and Computer Assisted Intervention–MICCAI 2019: 22nd International Conference, Shenzhen, China, October 13–17, 2019, Proceedings, Part II 22*, pages 119–127. Springer, 2019.

[159] Kihyuk Sohn, Honglak Lee, and Xinchen Yan. Learning structured output representation using deep conditional generative models. *Advances in neural information processing systems*, 28, 2015.

[160] Olaf Delgado-Friedrichs, Vanessa Robins, and Adrian Sheppard. Skeletonization and

partitioning of digital images using discrete morse theory. *IEEE transactions on pattern analysis and machine intelligence*, 37(3):654–666, 2014.

[161] Vanessa Robins, Peter John Wood, and Adrian P Sheppard. Theory and algorithms for constructing discrete morse complexes from grayscale digital images. *IEEE Transactions on pattern analysis and machine intelligence*, 33(8):1646–1658, 2011.

[162] Suyi Wang, Yusu Wang, and Yanjie Li. Efficient map reconstruction and augmentation via topological methods. In *Proceedings of the 23rd SIGSPATIAL international conference on advances in geographic information systems*, pages 1–10, 2015.

[163] Samik Banerjee, Lucas Magee, Dingkang Wang, Xu Li, Bing-Xing Huo, Jaikishan Jayakumar, Katherine Matho, Meng-Kuan Lin, Keerthi Ram, Mohanasankar Sivaprakasam, et al. Semantic segmentation of microscopic neuroanatomical data by combining topological priors with encoder–decoder deep networks. *Nature machine intelligence*, 2(10):585–594, 2020.

[164] Chi Seng Pun, Si Xian Lee, and Kelin Xia. Persistent-homology-based machine learning: a survey and a comparative study. *Artificial Intelligence Review*, 55(7):5169–5213, 2022.

[165] Elizabeth Munch. A users guide to topological data analysis. *Journal of Learning Analytics*, 4(2):47–61, 2017.

[166] Alan Said and Vicenç Torra. Data science: An introduction. *Data Science in Practice*, pages 1–6, 2019.

[167] Frédéric Chazal and Bertrand Michel. An introduction to topological data analysis: fundamental and practical aspects for data scientists. *Frontiers in artificial intelligence*, 4:667963, 2021.

[168] Li M Chen, Zhixun Su, Bo Jiang, and Li M Chen. Topological data analysis. *Mathematical Problems in Data Science: Theoretical and Practical Methods*, pages 101–124, 2015.

[169] Iryna Yurchuk. Digital image segmentation based on the persistent homologies. In *ICTES*, pages 226–232, 2019.

[170] Aras Asaad. *Persistent homology tools for image analysis.* The University of Buckingham (United Kingdom), 2020.

[171] Nick Byrne, James R Clough, Israel Valverde, Giovanni Montana, and Andrew P King. A persistent homology-based topological loss for cnn-based multiclass segmentation of cmr. *IEEE transactions on medical imaging*, 42(1):3–14, 2022.

[172] Jiaqi Yang, Xiaoling Hu, Chao Chen, and Chialing Tsai. A topological-attention convlstm network and its application to em images. In *Medical Image Computing and Computer Assisted Intervention–MICCAI 2021: 24th International Conference, Strasbourg, France, September 27–October 1, 2021, Proceedings, Part I 24*, pages 217–228. Springer, 2021.

[173] Saumya Gupta, Xiaoling Hu, James Kaan, Michael Jin, Mutshipay Mpoy, Katherine Chung, Gagandeep Singh, Mary Saltz, Tahsin Kurc, Joel Saltz, et al. Learning topological interactions for multi-class medical image segmentation. In *European Conference on Computer Vision*, pages 701–718. Springer, 2022.

[174] Xiaoling Hu, Dimitris Samaras, and Chao Chen. Learning probabilistic topological representations using discrete morse theory. *arXiv preprint arXiv:2206.01742*, 2022.

[175] Mohsen Ahmadkhani and Eric Shook. Toposingan: Learning a topology-aware generative model from a single image. *Applied Sciences (2076-3417)*, 14(21), 2024.

[176] Athindran Ramesh Kumar and Peter J Ramadge. Diffloop: Tuning pid controllers by differentiating through the feedback loop. In *2021 55th Annual Conference on Information Sciences and Systems (CISS)*, pages 1–6. IEEE, 2021.

[177] Uwe G Hacke, Rachel Spicer, Stefan G Schreiber, and Lenka Plavcová. An ecophysiological and developmental perspective on variation in vessel diameter. *Plant, cell & environment*, 40(6):831–845, 2017.

[178] Dominik Müller, Iñaki Soto-Rey, and Frank Kramer. Towards a guideline for evaluation metrics in medical image segmentation. *BMC Research Notes*, 15(1):210, 2022.

[179] Ilya Tolstikhin, Olivier Bousquet, Sylvain Gelly, and Bernhard Schoelkopf. Wasserstein auto-encoders. *arXiv preprint arXiv:1711.01558*, 2017.

[180] Mathieu Carriere, Marco Cuturi, and Steve Oudot. Sliced wasserstein kernel for persistence diagrams. In *International conference on machine learning*, pages 664–673. PMLR, 2017.

[181] Xue Dong, Yang Lei, Tonghe Wang, Matthew Thomas, Leonardo Tang, Walter J Curran, Tian Liu, and Xiaofeng Yang. Automatic multiorgan segmentation in thorax ct images using u-net-gan. *Medical physics*, 46(5):2157–2168, 2019.

[182] HR Hosseinpour, F Samadzadegan, F Dadrass Javan, and S Motayyeb. Improving semantic segmentation of high-resolution remote sensing images using wasserstein generative adversarial network. *The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, 48:45–51, 2023.

[183] Ahmed Iqbal, Muhammad Sharif, Mussarat Yasmin, Mudassar Raza, and Shabib Aftab. Generative adversarial networks and its applications in the biomedical image segmentation: a comprehensive survey. *International Journal of Multimedia Information Retrieval*, 11(3):333–368, 2022.

[184] Maayan Frid-Adar, Idit Diamant, Eyal Klang, Michal Amitai, Jacob Goldberger, and Hayit Greenspan. Gan-based synthetic medical image augmentation for increased cnn performance in liver lesion classification. *Neurocomputing*, 321:321–331, 2018.

[185] Veit Sandfort, Ke Yan, Perry J Pickhardt, and Ronald M Summers. Data augmentation using generative adversarial networks (cyclegan) to improve generalizability in ct segmentation tasks. *Scientific reports*, 9(1):16884, 2019.

[186] Yungang Cao, Baikai Sui, and Wei Zhang. Rel-sagan: relative generation adversarial network integrated with attention mechanism for scene data augmentation of remote

sensing. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, 15:3107–3119, 2022.

[187] Kai Deng, Kun Zhang, Ping Yao, Siyuan Cheng, and Peng He. Skip attention gan for remote sensing image synthesis. In *ICASSP 2021-2021 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, pages 2305–2309. IEEE, 2021.

[188] Ning Lv, Hongxiang Ma, Chen Chen, Qingqi Pei, Yang Zhou, Fenglin Xiao, and Ji Li. Remote sensing data augmentation through adversarial training. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, 14:9318–9333, 2021.

[189] Nasim Souly, Concetto Spampinato, and Mubarak Shah. Semi supervised semantic segmentation using generative adversarial network. In *Proceedings of the IEEE international conference on computer vision*, pages 5688–5696, 2017.

[190] Erdost Yıldız, Abdullah Taha Arslan, Ayşe Yıldız Taş, Ali Faik Acer, Sertac Demir, Afsun Şahin, and Duygun Erol Barkana. Generative adversarial network based automatic segmentation of corneal subbasal nerves on in vivo confocal microscopy images. *Translational Vision Science & Technology*, 10(6):33–33, 2021.

[191] Guizeng You, Yuanyuan Qin, Congrui Zhao, Yufei Zhao, Kongqiang Zhu, Xinwu Yang, and Yan lin Li. A cgan-based tumor segmentation method for breast ultrasound images. *Physics in Medicine & Biology*, 68(13):135002, 2023.

[192] Zizhao Zhang, Lin Yang, and Yefeng Zheng. Translating and segmenting multimodal medical volumes with cycle-and shape-consistency generative adversarial network. In *Proceedings of the IEEE conference on computer vision and pattern Recognition*, pages 9242–9251, 2018.

[193] Sil Aarts, Marjan Van Den Akker, and Bjorn Winkens. The importance of effect sizes. *The European Journal of General Practice*, 20(1):61–64, 2014.

[194] Jian Zhou, Yu Su, Qinglan Ding, Yuhe Qiu, and Qing Wang. Research on segmentation algorithm of uav remote sensing image based on deep learning.

In *Fourth International Conference on Geoscience and Remote Sensing Mapping (GRSM 2022)*, volume 12551, pages 704–713. SPIE, 2023.

[195] Dongao Ma, Ping Tang, and Lijun Zhao. Siftinggan: Generating and sifting labeled samples to improve the remote sensing image scene classification baseline in vitro. *IEEE Geoscience and Remote Sensing Letters*, 16(7):1046–1050, 2019.

[196] Liang Tian, Xiaorou Zhong, and Ming Chen. Semantic segmentation of remote sensing image based on gan and fcn network model. *Scientific Programming*, 2021(1):9491376, 2021.

[197] Iraís Castillo, Santos Solano, and Ana R Sepúlveda. A controlled study of an integrated prevention program for improving disordered eating and body image among mexican university students: A 3-month follow-up. *European Eating Disorders Review*, 27(5):541–556, 2019.

[198] Daniël Lakens. Calculating and reporting effect sizes to facilitate cumulative science: a practical primer for t-tests and anovas. *Frontiers in psychology*, 4:863, 2013.

[199] Amy J Morgan, Anna Ross, and Nicola J Reavley. Systematic review and meta-analysis of mental health first aid training: Effects on knowledge, stigma, and helping behaviour. *PloS one*, 13(5):e0197102, 2018.

[200] Simon M Rice, David Kealy, John S Ogrodniczuk, Zac E Seidler, Linda Denehy, and John L Oliffe. The cost of bottling it up: Emotion suppression as a mediator in the relationship between anger and depression among men with prostate cancer. *Cancer management and research*, pages 1039–1046, 2020.

[201] Carolyn J Hill, Howard S Bloom, Alison Rebeck Black, and Mark W Lipsey. Empirical benchmarks for interpreting effect sizes in research. *Child development perspectives*, 2(3):172–177, 2008.

[202] Joshua M Cherubini and Maureen J MacDonald. Statistical inferences using effect sizes in human endothelial function research. *Artery Research*, 27:176–185, 2021.

[203] Patrick E McKight and Julius Najab. Kruskal-wallis test. *The corsini encyclopedia of psychology*, pages 1–1, 2010.

[204] Eva Ostertagova, Oskar Ostertag, and Jozef Kováč. Methodology and application of the kruskal-wallis test. *Applied mechanics and materials*, 611:115–120, 2014.

[205] Philip Sedgwick. Non-parametric statistical tests for two independent groups: numerical data. *Bmj*, 348, 2014.

Appendix A

TopoSinGAN Supplementary Materials

A.1 Demonstration of Independence of NTC Metric on the Placement of the Randomly Added Terminal Node

To verify that the addition of a terminal node does not significantly impact the proposed topological evaluation metric, we randomly added a terminal node to both case study graphs 1000 times and calculated the percentage of LISA clusters. The results from both experiments, presented in Table A.1, indicate that the percentage of clusters remained consistent, with a similar mean and a small standard deviation (STD).

Table A.1: LISA cluster analysis results for 1000 iterations with randomly added terminal nodes.

Cluster Type	**Agricultural Fields**		**Dendrology**	
	Mean	***STD***	***Mean***	***STD***
HH	22.14	2.98	20.04	1.71
LL	18.91	1.84	18.76	1.78
LH	0.017	0.22	0	0
HL	0	0	0	0

A.2 Extreme Cases Considered for Graph Classification

Figure A.1 illustrates an instance of the seven extreme cases used to simulate various topological structures of 2D graphs.

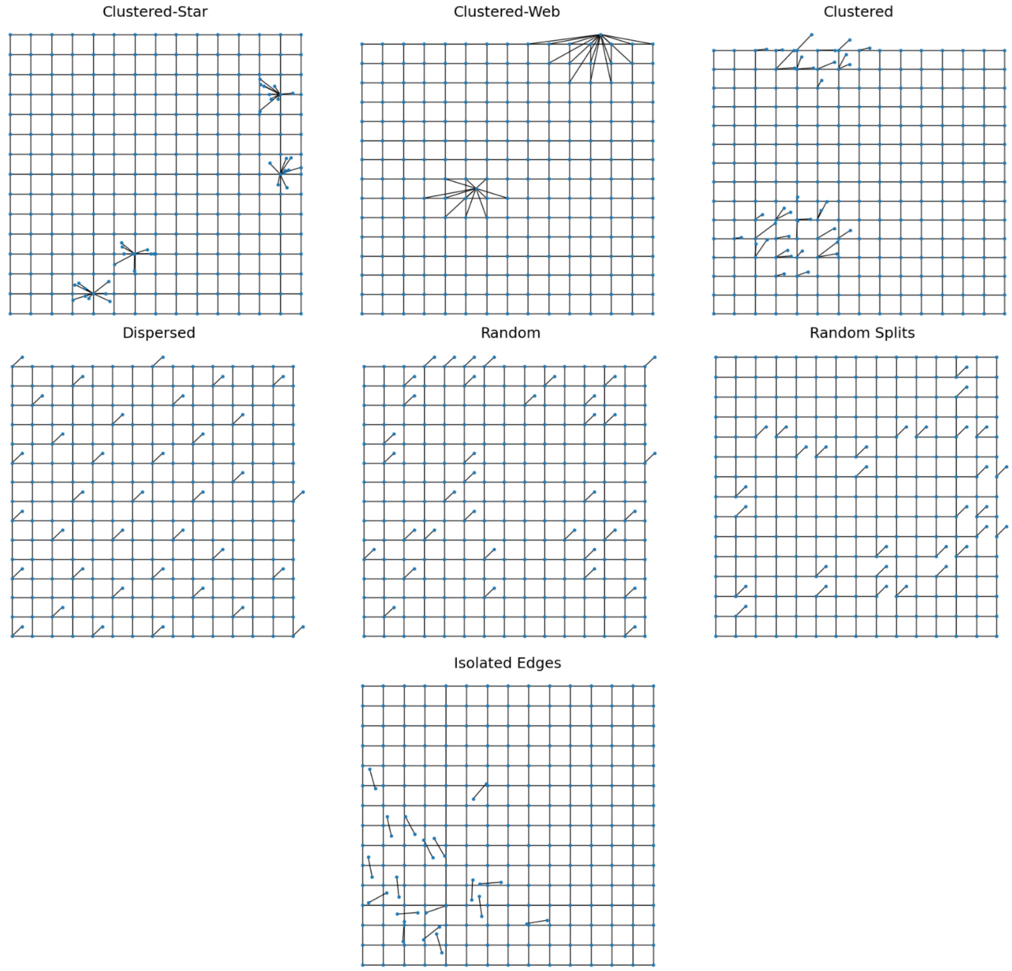


Figure A.1: The seven extreme topological distributions of terminal nodes in a 15 × 15 2D grid, including random, dispersed, clustered, clustered-star, clustered-web, isolated edges, and randomly split-edge distributions.

A.3 Demonstrating the Non-Disruptive Effect of Topology Loss on SinGAN Learning

To ensure that our proposed TopoLoss does not negatively impact the learning process, we first measure the common graph metrics specified in Section 3.3.2 (excluding NTC features). We exclude the NTC features from the calculations here to evaluate the similarity of the general attributes in the generated masks. Next, we use the cosine similarity metric to measure the similarity between the GTs and synthetic images generated by both GAN models. The results shown in Table A.2 indicate that SinGAN and TopoSinGAN exhibit similar performance when NTC features are excluded from the evaluation, indicating comparable efficacy in replicating common graph metrics.

Table A.2: Cosine similarity between GT and 1000 synthetic images for SinGAN and TopoSinGAN.

	SinGAN		**TopoSinGAN**	
	Mean	***STD***	***Mean***	***STD***
Agricultural	0.9873	0.0023	0.9895	0.0001
Dendrology	0.9892	0.0011	0.9891	0.0004