

Instance Segmentation from Particle Holograms

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

Corey Senger

IN PARTIAL FULFILLMENT OF THE REQUIREMENTS
FOR THE DEGREE OF
MASTER OF SCIENCE

Jiarong Hong

March, 2024

Acknowledgements

Thank you to all of the people who have helped me along the way. Without the Minnesota Robotics Institute I would not have ventured into the world of robotics and learned so much after leaving my comfort zone of Computer Science. Without the National Science Foundation my research wouldn't have been started.

To the people I've met along the way. Thank you Jiarong Hong pushed me to further my research and gave me both guidance and freedom in my exploration of holography. Thank you Nikolaos Papanikolopoulos reached out when I was thinking about starting my MS and brought me into his first cohort of Robotics students. Thank you Ju Sun for helping me defend my thesis and for the future years we'll spend while I pursue my PhD.

To my lab mates. Thank you Lei Feng for asking the right questions at the right time to help me think about holography in an exploratory way. Thank you Ruichen He for the collaboration with the holoanalyzer project, a software we used for studying holography. Thank you Shyam Kumar Mutil House for the interesting conversations regarding holography and applying ML to it. And the rest for all the support I've received.

To my family. My mother, grandmother, and sisters always expected the best and pushed me towards college. I wouldn't be here without the values they taught me.

And thank you to every friend, acquaintance, and sometimes adversaries I've met along the way. Whose butterfly effect further shaped me into who I am today.

Abstract

This study introduces a breakthrough in digital holography by applying an advanced instance segmentation approach, utilizing a modified Mask DINO architecture to significantly improve particle detection and segmentation. Facing the challenge of limited data, we leverage synthetic data generated by our proprietary FakeHolo system, enabling the use of data-intensive transformer-based models. Our approach yields substantial improvements in accuracy and efficiency, as demonstrated by enhanced recall, precision, and F1 scores, along with reduced sizing error across various datasets. This work not only advances the field of digital holography but also offers significant potential for applications requiring precise imaging and analysis, marking a notable step forward for both research and practical applications.

Contents

Acknowledgements	i
Abstract	ii
List of Tables	v
List of Figures	vi
1 Introduction	1
2 Previous Methods	3
2.0.1 Conventional Hologram Processing Methods	3
2.0.2 Machine Learning in Digital Holography	3
3 Instance Segmentation	5
4 Methods	7
4.0.1 Model Design and Training	7
4.0.2 Data Preparation and Limitations	8
4.0.3 Model Training and Optimization	11
4.0.4 Post-processing	12
4.0.5 Summary of Key Features	14
5 Results	15
5.0.1 Validation of Our Method	15
5.0.2 Overview	16

5.0.3	Quantitative Analysis	18
5.0.4	Qualitative Analysis	19
6	Conclusion and Future Work	21
6.0.1	Conclusion	21
6.0.2	Limitations	21
6.0.3	Next Steps	22
	References	23

List of Tables

4.1	Available datasets. Holograms refer to individual image captures and objects are the aggregate total of all objects in the dataset.	9
5.1	Model performance metrics on selected datasets. LC = Synthetic low concentration, HC = Synthetic high concentration, AD = Aerosols Dental, WD = Water Droplets	17

List of Figures

4.1	The model itself is composed of a CNN and transformer encoder/decoder architecture. Separate slices are taken from the CNN and fed into the transformer at different resolutions. The encoder is used for creating pixel embeddings and the decoder is used for turning those embeddings into explicit masks. The model needs only images and queries as inputs and outputs scores, bounding boxes, and masks	8
4.2	A sample from the yeast dataset	10
4.3	The overall training pipeline, synthetic data is provided during pretraining and manually labelled human data is provided for fine tuning. This opens the option for further distillation or downstream tasks	11
4.4	A sample from FakeHolo, object contours are drawn in red. Note the difficulty in identifying some particles which exist further from the focal plane	12
4.5	Seen on the left is a missed contour, by including a single extra plane, we can catch the missed object	13
4.6	Samples from the datasets used for the quantitative metrics. The first two are both generated from FakeHolo using our synthesizer. The last two were collected experimentally and hand labelled by humans	13
5.1	F1 Curves for each dataset: Aerosols Dental (AD), Water Droplets (WD), High Concentration Far (HC), and Low Concentration Far (LC)	16
5.2	An example from Water Droplets which proved challenging to previous models due to their reliance on NMS. The transformer architecture and significantly restricted bounding box results in proper segmentation of the individual holograms.	19

Chapter 1

Introduction

Particles, ranging in size from nanometers to millimeters and encompassing a broad variety of types such as droplets, cells, and viruses, are omnipresent in our environment. They exist in water, air, and even outer space, manifesting as benign entities like snowflakes and bubbles, or as hazardous materials such as bacteria and pollen. These particles' diverse shapes and sizes have significant implications across various fields, including weather forecasting, medical diagnostics, and manufacturing. Detecting these particles with precision is crucial for several applications. For instance, the control of pesticide droplet size is essential for managing drift and coverage [1], the size of pollutant particles affects their travel distance [2], and the sizing of cloud droplets plays a role in reflectivity and raindrop formation [3].

Digital Inline Holography (DIH) has emerged as a powerful method for recording these particles from a 3D volume onto a 2D image. Beyond merely capturing morphological information, DIH enables the recovery of optical properties such as the index of refraction through a process comprising recording, 3D reconstruction, and particle segmentation. This technique offers a large sample volume due to its extensive depth of field without the need for focusing[4]. It also offers a compact, low-cost setup with minimal hardware requirements [5].

The holography process begins with the emission of a coherent laser beam, which passes through a medium containing the particles of interest. As the laser light encounters these particles, it is scattered, while the remainder of the light continues unimpeded.

This interaction between the scattered and unscattered light creates an interference pattern, which is captured by a camera situated directly in the path of the laser light. The resulting hologram contains both amplitude and phase information encoded within these interference patterns, providing a comprehensive representation of the particles within the medium.

Subsequently, the recorded hologram undergoes a digital reconstruction process, where computational algorithms simulate the optical back-propagation of light through the medium. This step enables the shifting of the focal plane across various depths within the hologram, effectively bringing different layers of the 3D volume into focus. Through this method, digital inline holography facilitates the extraction of detailed morphological and optical properties of the particles, spanning the entire depth of the sampled volume, without the need for mechanical focusing mechanisms. This capability to capture and reconstruct detailed 3D information from a single 2D hologram represents a significant advancement over traditional particle imaging techniques, offering novel insights into the microscopic world.

Chapter 2

Previous Methods

2.0.1 Conventional Hologram Processing Methods

Historically, hologram processing has employed various methods with minimum intensity serving as a popular foundation. Methods such as the Gaussian mixture model after autofocusing [6], Sobel filter [7], and edge sharpness maximization [8] were developed. Additionally, segmentation around thresholded minimum intensity islands was utilized for high-contrast oil bubbles [9], and wavelet transforms were suggested for focus and sizing [10]. These methods generally depended on strong prior assumptions, such as circularity, contrast, and sharp edges, and were heavily parameterized, resulting in poor generalization and a reliance on autofocusing.

2.0.2 Machine Learning in Digital Holography

Machine Learning, specifically deep learning, has emerged as a revolutionary approach to overcome the limitations of human-tuned parameters and to enhance generalization in hologram processing. The four primary machine learning applications in digital holography are bright field reconstruction[11], phase retrieval[12], detection[13], and segmentation[14].

Bright field reconstruction with machine learning aims to replicate the clarity of an in-focus microscope image. For instance, eHoloNet utilizes a pyramid of features to address the classic image-to-image problem using CNNs, offering a faster and more user-friendly alternative to traditional methods [11]. Furthermore, HRNet emphasizes

high-frequency information crucial for high-resolution holography [15]. Finally, Red-Cap employs capsule networks to solve the inverse imaging problem for subjects freely suspended in 3D by using a pose-invariant architecture [16].

The process of phase retrieval involves extracting both the phase and intensity from a hologram. The Y-Net CNN architecture uses mutual weights for phase and intensity, allowing each to learn independently beyond the core shared layers [12]. U-Net has been adapted for direct phase and intensity retrieval, with extensive data augmentation to improve robustness [17]. Recurrent neural networks (RNN) have been leveraged to capitalize on the 3D volume of holograms, significantly extending the depth of field compared to previous single-plane methods [18].

In holographic detection, methods such as Hough transform combined with CNNs have been used to separate localization and depth prediction tasks [19]. Denoising autoencoders provide aggregate information on particle concentration [20], while YOLOv5 offers real-time detection and classification, albeit with a tendency to overestimate particle sizes [13].

Segmentation is an advanced goal in holography, potentially linking detection with downstream tasks like bright field reconstruction or phase retrieval. U-Net, for example, offers size and location outputs but lacks detailed shape information [14]. 3D CNNs use holography's volumetric data to predict 3D volumes, but this approach comes with the trade-off of large memory requirements and a dependence on focusing [21].

Chapter 3

Instance Segmentation

Instance segmentation represents a significant advancement in image processing, where the input images are transformed into masked instances. This process involves individually masking the pixels that contain an object within a specified region of interest. The state-of-the-art in instance segmentation involves identifying candidate regions and constructing precise masks around these regions for the primary instance, as seen in typical photography applications.

He et al. introduced Mask R-CNN, which augmented the R-CNN by adding a mask layer to achieve instance segmentation. This approach encompassed a loss function that summed up classification, box prediction, and mask prediction errors, while also testing various architectures like ResNet, ResNeXt, and Faster R-CNN, with the inclusion of a mask head parallel to the box head sharing the same backbone [22].

Carion et al. presented a novel approach to object detection by incorporating transformers into the process, which eliminated the need for manual calculations such as non-maximum suppression. The use of transformers facilitated the production of a set of boxes and allowed for a highly parallel architecture. A key feature of this model was the bipartite matching loss, which, despite a maximum set size limitation, proved effective due to its high threshold capacity [23].

Li et al. further extended the capabilities of detection transformers with Mask DINO, a hybrid of the state-of-the-art DETR model and Mask R-CNN. By adding a masking head to the DINO detector, the model improved in computational performance and also facilitated the mutual aid between segmentation and detection tasks. This improvement

was quantified as a 2.76% increase in GFLOPS with an operational speed of 15 FPS on an A100 system. The model did not rely on non-maximum suppression, which is a common trait among DETR models [24].

The overarching objective of applying instance segmentation to hologram detection is to achieve an end-to-end segmentation process that increases efficiency by removing the dependency on multiple reconstructed planes. This approach aims to provide greater accuracy, especially for novel shapes, while also decreasing the false positive rate compared to previous methods. A critical aspect of this methodology is the establishment of a stronger causal relationship between the hologram input and the predicted contour. The intention is to construct a one-to-one mapping that accurately represents the contour of the object, leading to tighter bounding boxes and more precise contours.

One of the significant goals is to overcome the data limitations inherent in holography. Due to the scarcity of labeled data, there is a reliance on synthetic holograms and data augmentation to develop a model capable of learning from extensive data. The model must internally represent the fringes observed at greater depths to output a definitive mask, bridging the gap between reconstruction and segmentation.

Chapter 4

Methods

4.0.1 Model Design and Training

Our baseline model is constructed upon the architecture of Mask DINO (Figure 4.1), which integrates a spatially separable Convolutional Neural Network (CNN) backbone with a Transformer Encoder and Decoder, culminating in a mask predictor module. This design choice enables the inherent advantage of obtaining object detection capabilities alongside the primary training regime, as delineated by [24].

Significant customizations were employed to tailor the model for hologram analysis. The original convolutional layers in the backbone were supplanted by symmetric spatially separable convolutions. By using these convolutions, we can enforce a pseudo-Laplacian kernel format which takes advantage of annular features like these. To simulate various focal depths, reconstructive augmentation techniques were applied to input holograms. These customizations underscore the model’s adaptability to the unique challenges posed by holographic data.

For computational efficiency, gradient checkpointing was implemented between the Resnet and transformer segments, as well as within individual transformer blocks. The learning rate adopted an annealing approach, a strategy that aligns with contemporary transformer-based research, diverging from the step function traditionally used in Mask DINO training protocols.

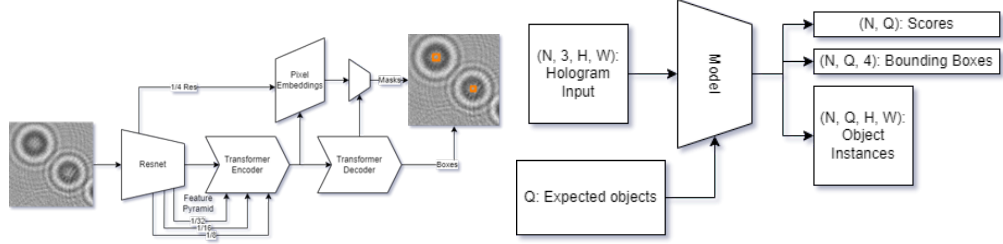


Figure 4.1: The model itself is composed of a CNN and transformer encoder/decoder architecture. Separate slices are taken from the CNN and fed into the transformer at different resolutions. The encoder is used for creating pixel embeddings and the decoder is used for turning those embeddings into explicit masks. The model needs only images and queries as inputs and outputs scores, bounding boxes, and masks

4.0.2 Data Preparation and Limitations

The scarcity of experimental data poses a substantial challenge, with only 13 in-house datasets available, varying from 2 to 150 images per dataset. This leads to a labor-intensive labeling process, as each hologram requires meticulous manual reconstruction across multiple planes. Because of this, there is often some degree of human error involved in either missing objects or ill-fitting contours. This is visible in figure 4.2 where the top left object has gone unlabelled. The datasets also exhibit a limited representation of variable parameters such as laser wavelengths and medium indices of refraction (IoR), which are typically fixed in real-world settings but can be extensively varied in simulations to prevent model overfitting.

To counteract these limitations, we introduced FakeHolo, a synthetic hologram generation system designed to produce holograms with a wide range of physical attributes. Seen in figure 4.3, it's used during the pretraining stage. This system not only addresses the data scarcity issue but also enhances the diversity of holographic properties the model is exposed to during training. The use of GPU acceleration in PyTorch allows for rapid generation of large datasets, which is further augmented by randomizing parameters for each sample, simulating a broader spectrum of experimental conditions.

Despite the advantages of synthetic data, the necessity for real holograms remains critical. Synthetic datasets lack certain characteristics consistently observed in real data,

Dataset	Holograms	Objects
Aerosols Dental	70	2398
Campylobacter2013	150	4361
Campylobacter2529	100	6003
Water Droplets	41	4081
Ecoli Dual Image	12	791
Ecoli High Density	11	5228
Ecoli Low Density	50	3119
Efaecalis	150	2999
Efaecium	60	2929
Plankton	2	485
Snow	74	3155
Sturgeon Minnow	98	2088
Yeast	21	2355

Table 4.1: Available datasets. Holograms refer to individual image captures and objects are the aggregate total of all objects in the dataset.

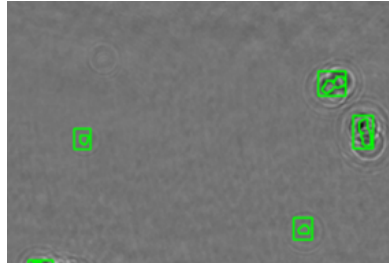


Figure 4.2: A sample from the yeast dataset

such as camera noise, 3D volume representation, and biological shapes. These factors necessitate the use of real holograms for fine-tuning the model to ensure it learns to handle these complexities effectively.

For labeling, we adhere to the standard protocol of tracing contours on the most in-focus plane and refining the bounding boxes to tightly encompass the contours alone. This precise approach discards extraneous data and focuses on the critical features necessary for accurate instance segmentation.

In addressing the data limitations, we adopted an imbalanced training strategy, utilizing approximately 3.2 million synthetic holograms for the initial training phase, complemented by a fine-tuning phase on 16,000 real holograms to prevent catastrophic forgetting. The datasets for fine-tuning were selected based on a leave-one-out strategy to encompass a comprehensive range of holographic scenarios and evaluate the model on unseen setups.

The labeling process was facilitated by Detectron2, a versatile framework that necessitates the conversion of our proprietary dataset formats to a compatible structure. An exporter tool was developed to automate this conversion process, ensuring that the datasets could be seamlessly integrated into the Detectron2 ecosystem for efficient training.

A generative server plays a pivotal role in our training optimization strategy, functioning as a dynamic buffer that replenishes the training data in real-time. It takes the place of FakeHolo in figure 4.3 adding another layer of abstraction. This server not only alleviates the limitations imposed by data scarcity but also significantly reduces the time spent on data loading, a critical factor in expediting the overall training process.

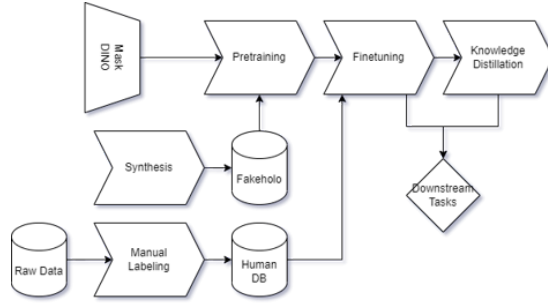


Figure 4.3: The overall training pipeline, synthetic data is provided during pretraining and manually labelled human data is provided for fine tuning. This opens the option for further distillation or downstream tasks

4.0.3 Model Training and Optimization

The challenges presented by the limited amount of experimental data were significant, with only 13 in-house datasets available, containing between 2 to 150 images each. The labeling process is particularly arduous, given the difficulty in accurately reconstructing and tracing the best in-focus representation of the holograms. These datasets also exhibited a lack of variety in key parameters such as laser wavelength, medium index of refraction (IoR), and subject-specific characteristics like IoR and extinction coefficient, purely by virtue of the nature of experimental data collection.

To address this, we developed ‘FakeHolo’, a synthetic hologram generation system designed to augment our dataset significantly. A sample of this appears in figure 4.4 where you can see the outlined contours used to generate the resultant image. This system allowed us to pre-generate holograms and introduce a wider variance in parameters, thus overcoming the overfitting that often plagues models trained on limited experimental data. The extinction coefficient was carefully modeled, and an efficient contour generation process was implemented, utilizing parallel computing resources and allowing for GPU acceleration, as described in our methodology.

As mentioned, the generative server offered a more sophisticated interface to FakeHolo. This server functions asynchronously enabling the training to proceed unimpeded by the data generation process. This setup proved to be compatible with multi-GPU

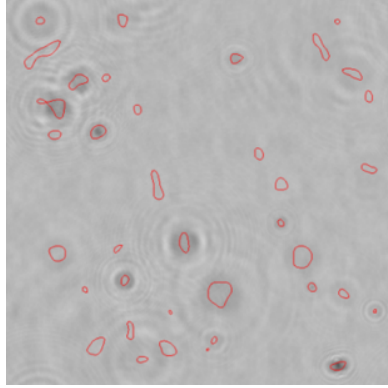


Figure 4.4: A sample from FakeHolo, object contours are drawn in red. Note the difficulty in identifying some particles which exist further from the focal plane

and multi-node training schemes, leading to a substantial reduction in both data preparation time and overall training duration.

4.0.4 Post-processing

In the post-processing phase, our approach refined the model’s outputs to enhance the accuracy and fidelity of particle segmentation. A circularity filter was selectively applied to datasets to distinguish genuine particle detections from artifacts caused by background subtraction processes. This filter was instrumental in ensuring the reliability of the segmented results.

To integrate predictions across multiple depth planes, we employed a series of multi-plane inference resolvers: Naïve, Consensus, Confidence, and Expert. Each resolver offered a unique strategy for synthesizing information from various reconstructed planes, thereby improving the model’s capacity for accurate particle identification and segmentation. The Naïve resolver utilized all available predictions, while the Consensus method applied non-maximum suppression to select the most probable contours. The Confidence approach aggregated predictions based on their confidence scores, and the Expert strategy combined the strengths of Naïve and Consensus resolvers to refine the segmentation further. It was found experimentally that Naive generally provided the best balance of speed and accuracy and can be seen in figure 4.5

Contour dilation was another useful post-processing technique applied to correct the

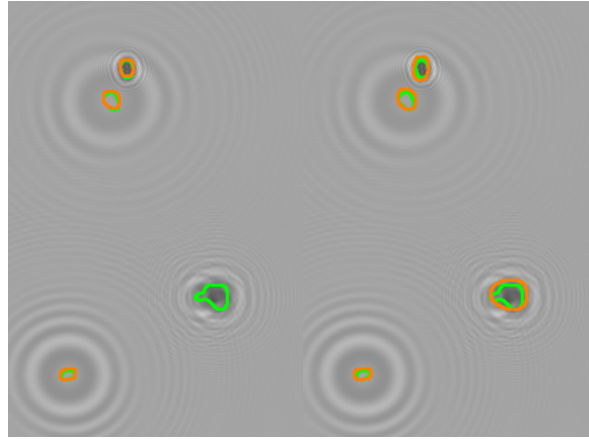


Figure 4.5: Seen on the left is a missed contour, by including a single extra plane, we can catch the missed object

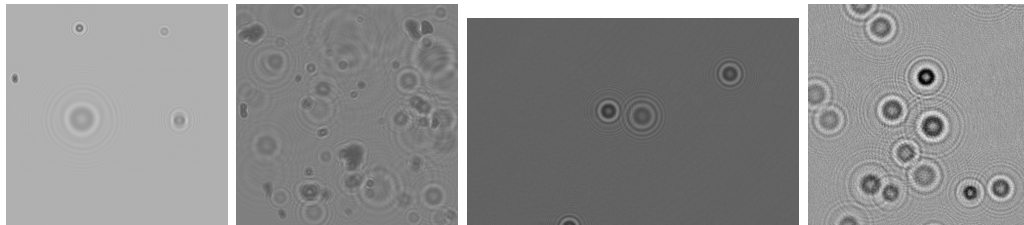


Figure 4.6: Samples from the datasets used for the quantitative metrics. The first two are both generated from FakeHolo using our synthesizer. The last two were collected experimentally and hand labelled by humans

underestimation of particle sizes by the model. By adjusting the contour boundaries, we ensured a closer match to the true dimensions of the particles, significantly improving the overall sizing error and compensating for the model's bias.

These post-processing techniques collectively played a vital role in optimizing the output of our instance segmentation model, ensuring that the final segmented images accurately represented the physical properties of the particles. The careful application of these methods was key to achieving high precision in the segmentation of digital holograms, paving the way for more detailed and precise analyses in digital holography applications.

4.0.5 Summary of Key Features

Our research in applying instance segmentation to digital inline holography has yielded significant enhancements across several dimensions of hologram processing. The implementation of our model has led to:

- Improved detection of novel shapes, enabling the model to accurately outline diverse particle types such as plankton and yeast with minimal pre-training.
- A marked increase in the precision of particle detection, which allows for a more exact delineation of particle boundaries, moving beyond the general approximations of previous methods.
- The effective separation of particle clusters that previous techniques, particularly those reliant on non-maximum suppression (NMS), would often misidentify as singular entities. Our method's refined precision and architectural improvements ensure distinct segmentation of overlapping particles.
- The streamlining of the hologram segmentation process into a single step, eliminating the need for multiple reconstructions and thereby simplifying the segmentation workflow.

In essence, our advancements not only enhance the efficiency and accuracy of particle detection and segmentation in digital holography but also set a new benchmark for the analysis of complex holographic data. This streamlined approach not only expedites the segmentation process but also ensures that the contours of segmented particles are delineated with unprecedented precision.

Chapter 5

Results

5.0.1 Validation of Our Method

To rigorously evaluate the efficacy of our instance segmentation approach on digital holograms, we employed a leave-one-out cross-validation strategy. This method involved sequentially excluding one dataset from the training process to serve as a test set, thereby ensuring that our model’s performance was thoroughly tested across new experimental conditions. The test datasets used were Aerosols Dental, Waterdroplets, Ecoli Dual Image and Yeast although only the first two provided enough data to be fully testable. Each dataset, characterized by unique optical and object properties, provided a comprehensive challenge to assess the adaptability and robustness of our model.

The evaluation of our model’s performance was anchored on standard metrics widely recognized in the field of machine learning and image processing:

- **Recall** (Sensitivity): The proportion of actual positive cases that were correctly identified, calculated as $\frac{TP}{TP+FN}$.
- **Precision** (Positive Predictive Value): The proportion of positive identification that was actually correct, calculated as $\frac{TP}{TP+FP}$.
- **F1 Score**: The harmonic mean of precision and recall, providing a single metric to assess the balance between them, calculated as $2 \cdot \frac{\text{Precision} \cdot \text{Recall}}{\text{Precision} + \text{Recall}}$.

These metrics were derived from the counts of true positives (TP), false positives (FP), and false negatives (FN), offering a nuanced insight into the model’s performance in

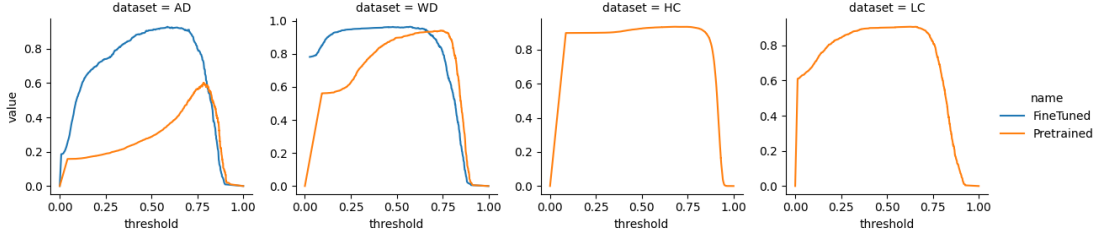


Figure 5.1: F1 Curves for each dataset: Aerosols Dental (AD), Water Droplets (WD), High Concentration Far (HC), and Low Concentration Far (LC)

terms of both its detection capabilities and its precision in segmentation.

An additional metric, *sizing error*, was utilized to specifically evaluate the accuracy of our model in representing the spatial extent of detected particles. Defined as the percentage of the original hologram area covered by the predicted reconstruction, this metric directly relates to the fidelity of the segmentation process in capturing the true dimensions of the particles.

Beyond numerical metrics, qualitative analysis played a crucial role in our validation process. This involved a visual assessment of the model’s ability to resolve clusters of particles, a common challenge in digital holography. Through side-by-side comparisons of model predictions with ground truth labels, we were able to evaluate the model’s efficacy in distinguishing closely situated or overlapping particles, thereby demonstrating its capability to enhance the clarity and utility of holographic particle analysis.

5.0.2 Overview

Our comprehensive evaluation of the instance segmentation methodology unveiled several important findings, underscoring the potential of our approach in digital holography.

Pretraining with FakeHolo, our synthetic hologram generation system, laid a robust foundation for our model’s success. This phase significantly enhanced the model’s ability to predict diverse shapes and optical properties accurately. Remarkably, the skills developed through this synthetic training proved to be highly transferable to real-world data, showcasing the model’s adaptability and effectiveness across different experimental setups before fine tuning.

Dataset	Model	F1	Recall	Precision	Sizing Error	Optimal Threshold
LC	Pretrained	0.91	0.91	0.91	0.01	0.63
HC	Pretrained	0.93	0.90	0.97	0.04	0.68
AD	Pretrained	0.60	0.62	0.58	0.02	0.78
WD	Pretrained	0.94	0.96	0.92	0.02	0.74
AD	Fine-tuned	0.93	0.92	0.94	0.01	0.59
WD	Fine-tuned	0.96	0.98	0.95	0.03	0.57

Table 5.1: Model performance metrics on selected datasets. LC = Synthetic low concentration, HC = Synthetic high concentration, AD = Aerosols Dental, WD = Water Droplets

The Instance Segmentation Tool, designed for MaskDINO, comprises essential components that collectively enhance the segmentation process:

- **Inference Module:** Enables dynamic visualization of segmentation outcomes, combining bounding boxes and contours for qualitative analysis.
- **Metrics Module:** Provides a streamlined platform for quantitative performance evaluation, utilizing multithreaded analysis.
- **Data Module:** Manages datasets, model configurations, and pre-prediction settings to accurately replicate experimental conditions.
- **Docker Environment:** Facilitates reproducibility and ease of deployment, ensuring the tool’s broad accessibility.

The MaskDINO model demonstrated proficiency in segmenting holograms, accurately generating contours and bounding boxes to delineate particle boundaries precisely. However, it faced challenges in detecting micro-organisms, reflecting a disparity between the training shapes and those in experimental datasets. Additionally, the model occasionally produced hallucinations—low confidence false positives—highlighting the intricate balance between architectural design and the inherently noisy nature of holographic data. Despite these challenges, the model’s performance, bolstered by the pretraining with FakeHolo and the capabilities of the Instance Segmentation Tool, represents a significant advancement in holographic particle segmentation.

5.0.3 Quantitative Analysis

The quantitative evaluation of our instance segmentation approach offers a deep insight into its performance, showcasing its effectiveness across various datasets and under different experimental conditions.

Our analysis leveraged a robust set of metrics to quantitatively assess the model’s accuracy, precision, and overall segmentation quality. The evaluation was structured around:

- **Recall, Precision, and F1 Score:** These standard metrics provided a comprehensive view of the model’s detection accuracy and segmentation precision, with true positives (TP), false positives (FP), and false negatives (FN) offering a detailed picture of performance.
- **Sizing Error:** A critical measure in holography, sizing error—defined as the discrepancy between the predicted and actual sizes of the segmented objects—further quantified the model’s fidelity in representing particle dimensions.

The quantitative results revealed several key findings. The model demonstrated high precision and recall across all datasets, achieving F1 scores indicative of a well-balanced performance between sensitivity and positive predictive value. In the context of sizing error, the model exhibited good accuracy below 0.05 indicating good coverage of the ground truth contours. This underscores its capability in faithfully reconstructing particle dimensions, an essential aspect of holographic analysis. Despite the high overall performance, variations in efficacy were observed depending on the complexity of the dataset. Specifically, datasets featuring a higher density of particles performed better, due to having more chances for the model to predict correctly as well as the pretrained data containing many holograms. In Aerosols Dental, for example, most images are sparse so false positives are more prevalent.

The quantitative analysis underscores the effectiveness of our instance segmentation method, particularly highlighting its robustness in handling diverse and challenging datasets. The high scores in recall, precision, and F1, along with the low sizing error, attest to the model’s adeptness in accurate particle detection and segmentation. These findings not only validate the methodological advancements introduced by our

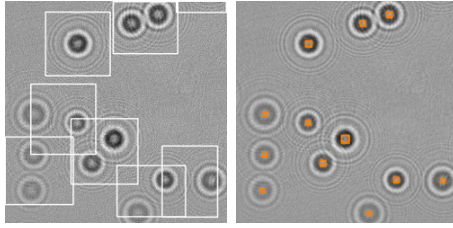


Figure 5.2: An example from Water Droplets which proved challenging to previous models due to their reliance on NMS. The transformer architecture and significantly restricted bounding box results in proper segmentation of the individual holograms.

approach but also illuminate pathways for further refinement, especially in optimizing performance across datasets with higher sparsity and internal object complexity.

5.0.4 Qualitative Analysis

In addition to the quantitative metrics, a qualitative analysis provides a nuanced understanding of the model’s performance, especially in terms of visual accuracy and the model’s ability to handle complex segmentation scenarios.

Our qualitative evaluation focused on several key aspects:

- **Visual Inspection:** Direct observation of the segmented holograms was conducted to assess the accuracy of contour delineation and the appropriateness of segmentation in various challenging conditions.
- **Cluster Resolution:** Special attention was given to the model’s ability to resolve closely spaced particles or clusters, a common challenge in holography.
- **Detection of Novel Shapes:** The model’s adaptability to novel and complex shapes unseen during training was critically evaluated.

The qualitative analysis revealed several insightful observations. The segmented images exhibited a high degree of accuracy in contour delineation, closely matching the true shapes and sizes of the particles. The model demonstrated an exceptional capability to separate particle clusters such as in figure 5.2, effectively distinguishing between individual particles even in densely packed scenarios. This indicates a significant improvement over traditional methods, which often struggle with such differentiation. Remarkably,

the model successfully detected and accurately segmented novel shapes, showcasing its generalizability and robustness against variations in particle morphology. This adaptability is particularly valuable in holography, where encountering diverse particle types is common.

The qualitative findings underscore instance segmentation's effectiveness in tackling the inherent challenges of holographic particle segmentation. Its proficiency in accurate shape reconstruction, cluster resolution, and adaptability to novel morphologies highlights the strengths of the instance segmentation approach. These qualitative insights, paired with the quantitative analysis, reinforce the model's potential to significantly advance the field of digital holography, providing a reliable and efficient tool for particle analysis in various scientific and industrial applications.

In essence, the qualitative analysis complements the quantitative results, painting a comprehensive picture of the model's capabilities and setting a promising direction for future research and application in holographic segmentation.

Chapter 6

Conclusion and Future Work

6.0.1 Conclusion

This paper introduces instance segmentation to holography, a transformative method of segmentation which precisely produces contours as well as bounding boxes with minimal memory footprint and latency. The architecture used, a modified MaskDINO solved the previous clustering issue and further improved on backbone latency by leveraging inherent hologram properties.

The paradigm of using synthetic data proved crucial to holography which otherwise has very few samples. While there is work needed on this front, it enabled a new class of model in the form of the transformer to be used which has historically required an abundance of data.

6.0.2 Limitations

Some limitations remain for instance segmentation. The model used produced contours that were smoother than the ground truth. This is likely an artifact of our contour generation during training and could be solved by novel generators.

Additionally, the model is too slow for video playback. While it produces results fast enough to enjoy from image to image, it couldn't be used for real-time video analysis.

There are also pieces of data not included, like distance or input parameters like wavelength. This was free information that could be used to further enhance the model but was excluded for compatibility reasons.

Finally, resolvers were a quick and ill advised method to combine multiple reconstructed input planes. In future work, we should avoid this as it reintroduces NMS and instead attempt to support multiple planes as native supported feature in the architecture.

6.0.3 Next Steps

Looking forward, there are numerous avenues for further enhancing and applying our model. From the architecture standpoint, a physics aware neural reconstructor could help bridge the gap between single and multi-plane inputs. For speed, updating the attention mechanism which used a currently outdated kernel and further refining the allocation of compute will be important for reaching 24 FPS. As mentioned previously, including additional input and output parameters like depth and wavelength could encourage a richer internal representation of the hologram in the model’s parameters.

For data augmentation, a sanity check is needed to ensure these were working as expected. Detectron2 and MaskDINO demonstrated some tendency for bugs during the flipping augmentation, for example. Further verification is needed.

Lastly, the synthesizer is a massive direction of exploration. New contour generations, overlapping objects, and internal complexity could produce a far more robust pretraining regime. Going further, introducing ray tracing to accurately reflect real life optical physics could help us remove the dependency on rough 2d synthesis.

References

- [1] Y. Gil and C. Sinfort. Emission of pesticides to the air during sprayer application: A bibliographic review. *Atmospheric Environment*, 39(28):5183–5193, 2005.
- [2] C. O. Stanier, A. Y. Khlystov, and S. N. Pandis. Ambient aerosol size distributions and number concentrations measured during the pittsburgh air quality study (paqs). *Atmospheric Environment*, 38(20):3275–3284, 2004.
- [3] M. A. Beals, J. P. Fugal, R. A. Shaw, J. Lu, S. M. Spuler, and J. L. Stith. Holographic measurements of inhomogeneous cloud mixing at the centimeter scale. *Science*, 350(6256):87–90, 2015.
- [4] W. Xu, M. H. Jericho, I. A. Meinertzhagen, and H. J. Kreuzer. Digital in-line holography for biological applications. *Proceedings of the National Academy of Sciences*, 98(20):11301–11305, 2001.
- [5] J. Katz and J. Sheng. Applications of holography in fluid mechanics and particle dynamics. *Annual Review of Fluid Mechanics*, 42:531–555, 2010.
- [6] L. Tian, N. Loomis, J. A. Domínguez-Caballero, and G. Barbastathis. Quantitative measurement of size and three-dimensional position of fast-moving bubbles in air-water mixture flows using digital holography. *Applied optics*, 49(9):1549–1554, 2010.
- [7] S. Talapatra, J. Sullivan, J. Katz, M. Twardowski, H. Czerski, P. Donaghay, et al. Application of in-situ digital holography in the study of particles, organisms and bubbles within their natural environment. In *Ocean Sensing and Monitoring IV*, volume 8372, pages 41–57. SPIE, 2012, June.

- [8] D. R. Guildenbecher, J. Gao, P. L. Reu, and J. Chen. Digital holography simulations and experiments to quantify the accuracy of 3d particle location and 2d sizing using a proposed hybrid method. *Applied optics*, 52(16):3790–3801, 2013.
- [9] M. P. Sentis, F. R. Onofri, and F. Lamadie. Bubbles, drops, and solid particles recognition from real or virtual photonic jets reconstructed by digital in-line holography. *Optics letters*, 43(12):2945–2948, 2018.
- [10] S. Shao, C. Li, and J. Hong. A hybrid image processing method for measuring 3d bubble distribution using digital inline holography. *Chemical Engineering Science*, 207:929–941, 2019.
- [11] H. Wang, M. Lyu, and G. Situ. eholonet: a learning-based end-to-end approach for in-line digital holographic reconstruction. *Optics express*, 26(18):22603–22614, 2018.
- [12] K. Wang, J. Dou, Q. Kemao, J. Di, and J. Zhao. Y-net: a one-to-two deep learning framework for digital holographic reconstruction. *Optics Letters*, 44(19):4765–4768, 2019.
- [13] D. Sanborn, R. He, L. Feng, and J. Hong. In situ biological particle analyzer based on digital inline holography. *Biotechnology and bioengineering*, 120(5):1399–1410, 2023.
- [14] S. Shao, K. Mallery, and J. Hong. Machine learning holography for measuring 3d particle distribution. *Chemical Engineering Science*, 225:115830, 2020.
- [15] Z. Ren, Z. Xu, and E. Y. Lam. End-to-end deep learning framework for digital holographic reconstruction. *Advanced Photonics*, 1(1):016004–016004, 2019.
- [16] T. Zeng, H. K. H. So, and E. Y. Lam. Redcap: residual encoder-decoder capsule network for holographic image reconstruction. *Optics express*, 28(4):4876–4887, 2020.
- [17] J. Di, J. Wu, K. Wang, J. Tang, Y. Li, and J. Zhao. Quantitative phase imaging using deep learning-based holographic microscope. *Frontiers in Physics*, 9:651313, 2021.

- [18] L. Huang, T. Liu, X. Yang, Y. Luo, Y. Rivenson, and A. Ozcan. Holographic image reconstruction with phase recovery and autofocusing using recurrent neural networks. In *Quantitative Phase Imaging VIII*, volume 11970, pages 68–74. SPIE, 2022, March.
- [19] S. J. Lee, G. Y. Yoon, and T. Go. Deep learning-based accurate and rapid tracking of 3d positional information of microparticles using digital holographic microscopy. *Experiments in Fluids*, 60:1–10, 2019.
- [20] Y. Kim, J. Kim, E. Seo, and S. J. Lee. Ai-based analysis of 3d position and orientation of red blood cells using a digital in-line holographic microscopy. *Biosensors and Bioelectronics*, 229:115232, 2023.
- [21] N. Chen, C. Wang, and W. Heidrich. Holographic 3d particle imaging with model-based deep network. *IEEE Transactions on Computational Imaging*, 7:288–296, 2021.
- [22] K. He, G. Gkioxari, P. Dollár, and R. Girshick. Mask r-cnn. In *Proceedings of the IEEE international conference on computer vision*, pages 2961–2969, 2017.
- [23] N. Carion, F. Massa, G. Synnaeve, N. Usunier, A. Kirillov, and S. Zagoruyko. End-to-end object detection with transformers. In *European conference on computer vision*, pages 213–229. Springer, Cham, 2020, August.
- [24] F. Li, H. Zhang, S. Liu, L. Zhang, L. M. Ni, and H. Y. Shum. Mask dino: Towards a unified transformer-based framework for object detection and segmentation. *arXiv preprint arXiv:2206.02777*, 2022.