Computationally efficient processing of in situ underwater digital holograms

Running head: Processing of underwater holograms

Emma Cotter (ecotter@whoi.edu) [1]

Erin Fischell (efischell@whoi.edu)[1]

Andone Lavery (alavery@whoi.edu)[1]

Keywords: digital in-line holography, machine learning, microscopy

[1]Woods Hole Oceanographic Institution, Department of Applied Ocean Physics and Engineering, Falmouth, MA, USA

Abstract

Underwater digital inline holography can provide high-resolution, in situ imagery of marine particles and offers many advantages over alternative measurement approaches. However, processing of holograms requires computationally expensive reconstruction and processing, and computational cost increases with the size of the imaging volume. In this work, a processing pipeline is developed to extract targets from holograms where target distribution is relatively sparse without reconstruction of the full hologram. This is motivated by the desire to efficiently extract quantitative estimates of plankton abundance from a data set (>300,000 holograms) collected in the Northwest Atlantic using a large-volume holographic camera. First, holograms with detectable targets are selected using a transfer learning approach. This was critical as a subset of the holograms were impacted by optical turbulence, which obscured target detection. Then, target diffraction patterns are detected in the hologram. Finally, targets are reconstructed and focused using only a small region of the hologram around the detected diffraction pattern. A search algorithm is employed to select distances for reconstruction, reducing the number of reconstructions required for 1 mm focus precision from 1000 to 31. When compared with full reconstruction techniques, this method detects 99% of particles larger than 0.1 mm^2, a size class which includes most copepods and larger particles of marine snow, and 85% of those targets are sufficiently focused for classification. This approach requires 1% of the processing time required to compute full reconstructions, making processing of long time-series, large imaging volume holographic data sets feasible.

1   Introduction

Marine particles, such as marine snow and plankton, play a critical role in ocean food webs [Anderson et al., 2018], global environmental change [Lombard et al., 2019], and ocean carbon cycling [Koski et al., 2020]. However, quantitative in situ measurements of particle abundance are challenging to obtain and typically require trade-offs between measurement range and resolution. Direct sampling using nets or bottles provides the most precise morphological measurements of samples, but may destroy less robust particles, like marine snow, during sampling and does not provide insight into spatial distribution (i.e., patchiness). High-frequency echosounders can measure abundance over relatively long ranges and provide estimates of abundance and organism type, but typically cannot provide species-level classification or size distributions [Lavery et al., 2007]. Conversely, optical imaging systems such as the video plankton recorder [Davis et al., 2005] and in situ ichthyoplankton imaging system [Cowen and Guigand, 2008] can provide high resolution imagery of plankton in situ, but have comparatively small imaging volumes.

Digital inline holography (DIH) has the potential to address some of these challenges and produce high-resolution imagery over relatively large imaging volumes [Loomis, 2011, Nayak et al., 2021]. A digital in-line holographic microscope (DIHM) records the diffraction patterns of particles within an imaging volume. The recorded diffraction patterns (referred to as a hologram) can then be numerically reconstructed at the depth of the particles to produce a focused image. Particle concentrations measured using a DIHM have been shown to be comparable to established methods, while counting more particles per recording (e.g., larger sample size) [Walcutt et al., 2020]. Recent technological advances have made underwater DIHM systems commercially available [Sun et al., 2007], and as a result they have seen increasing use for marine particle measurements (e.g., Nayak et al. [2018], Greer et al. [2020], Walcutt et al. [2020]).

While the advantages of DIH for measuring marine particles are clear, unfortunately, interpretation of recorded holograms is not straightforward. The depth of a particle within the imaging volume is typically not known a priori, and must be determined based on the depth of the reconstruction where it is in focus. While several methods exist for hologram reconstruction, the angular spectrum method [Ratcliffe, 1956], has been shown to perform best for DIH [Sun et al., 2008, Fonesca et al., 2016]. Reconstruction by this method requires calculation of the two-dimensional Fourier transform of the hologram, propagation of the frequency content to the desired depth within the imaging volume, and calculation of the two-dimensional inverse Fourier transformation to produce the reconstruction [Latychevskaia and Fink, 2015]. Using the fast Fourier Transform (FFT) algorithm, both the two-dimensional Fourier transform and its inverse are $\mathcal{O}(2NM \log_2 NM)$ complexity operations for a hologram measuring M x N pixels [Cooley and Tukey, 1965].

This means that the computational cost of reconstruction increases non-linearly with the number of pixels in the hologram and is directly proportional to the depth resolution of the reconstruction and size of the imaging volume (i.e., number of reconstructions computed). In addition to computational cost, reconstruction also produces significant data volumes - each reconstruction has the same number of pixels as the raw hologram, so reconstruction produces tens or hundreds of times more data than the raw holograms. Finally, individual particles must be located within the reconstruction, and the complexity of this search will scale with the number of reconstructions.

Recent advances in hologram processing have automated this process by using target detection algorithms and focus or sharpness metrics to detect particles and determine the reconstruction depth where they are in focus. Fonesca et al. [2016] evaluated 15 metrics for determining the reconstruction where a particle is in focus and evaluated trade-offs between performance and processing requirements for each metric. Walcutt et al. [2020] developed a processing pipeline that performs high-resolution hologram reconstruction, then computes a sharpness score for each pixel in the resulting reconstructions. Particles are determined to be where there are neighboring pixels with relatively high focus scores. Nayak et al. [2018] use a similar approach, but instead of performing analysis pixel-by-pixel, use a focus metric to determine the most in-focus reconstruction for 0.6 x 0.6 mm sub-regions of the hologram. The focused regions are then assembled into a composite image that contains the focused targets in a two-dimensional plane.

While these methods represent significant advances in DIH capabilities and have enabled analysis of larger data sets than previously possible, they still require computationally expensive high-resolution reconstruction of the hologram prior to target detection and focusing. Most studies to date have used DIHM systems with relatively small imaging volumes (e.g., separation distances of 4 and 1.8 cm in Nayak et al. [2018] and Walcutt et al. [2020], respectively) and data sets (e.g., less than 1 hour of data collected at 15 Hz analyzed in [Greer et al., 2020]). Processing schemes that avoid high-resolution reconstruction would enable analysis of larger (e.g., long time series) data sets and data from larger imaging volume DIHMs.

Guo et al. [2021] addresses the computational cost of hologram reconstruction by employing deep learning to classify particles based on their diffraction patterns in the raw hologram, eliminating the need for reconstruction. When a particle is relatively close to the camera, the diffraction pattern tends to resemble the shape of the target. For small imaging volumes (e.g., 4 cm separation distance in the referenced publication), this means that many particles can be classified without reconstruction. However, when particles are farther from the camera (larger imaging volume) their diffraction patterns no longer resemble the particle, meaning that this approach is not translatable to systems with larger imaging volumes. Further, Guo et al. [2021] do not evaluate the accuracy of target detection based on diffraction patterns in the raw hologram.

Here, we present a processing pipeline to extract and focus targets without reconstruction of the full hologram. First, we use deep learning to select holograms suitable for automatic processing (Section 2.3). Then, we detect the diffraction patterns of targets in the raw hologram (Section 2.4), and reconstruct these targets using only a small window around the diffraction pattern (Section 2.5). A search algorithm is used to reduce the number of reconstructions required to find the focus depth of the target. By reducing the size of the hologram used for reconstruction and the number of reconstructions required, this approach offers a significant reduction in computational cost and data storage requirements compared to existing methods. The processing pipeline is evaluated using data collected by a DIHM with a relatively large imaging volume (approximately 1 m separation distance; 0.88 L imaging volume) deployed in the mesopelagic zone in the Northeast Atlantic. In this work, we do not attempt to estimate particle abundance, rather, our intent is to present methods to rapidly extract targets from large holographic data sets that will enable future analysis and development of automatic classification algorithms.
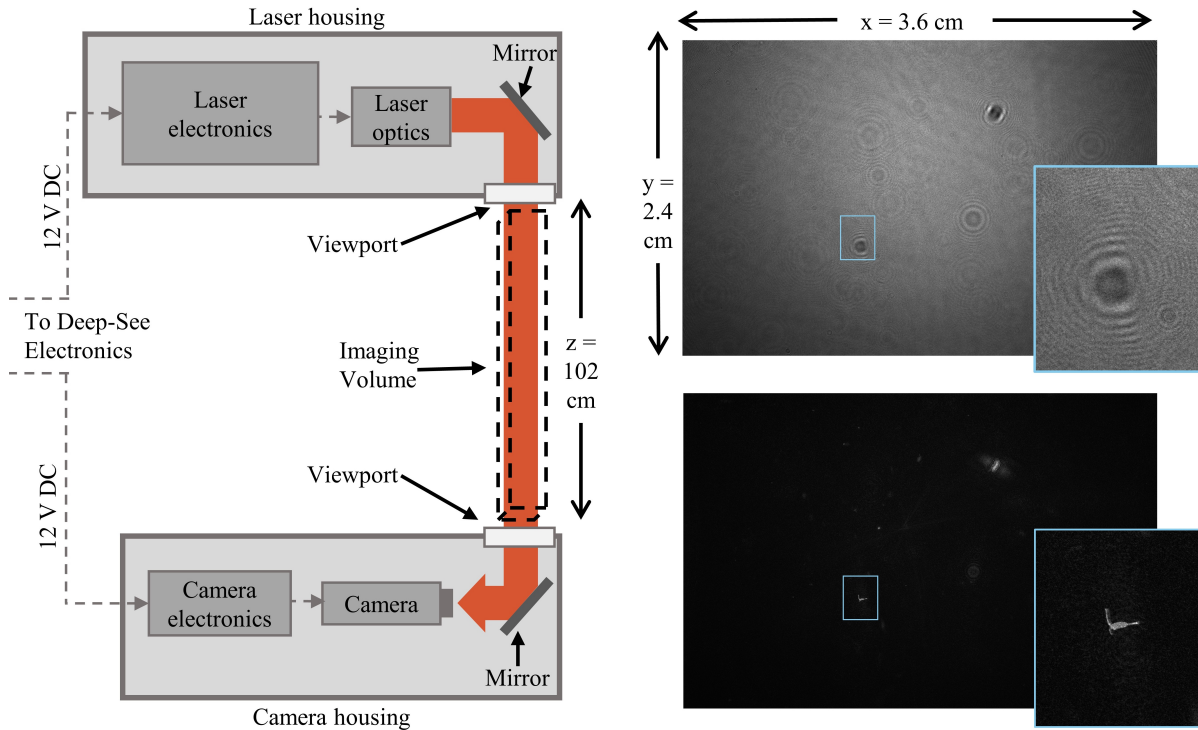
Figure 1: Left: Diagram of the DIHM assembly (not to scale). Right: A raw hologram that is not affected by optical turbulence (top) and its reconstruction at the depth of a particle (bottom). Zoomed-in views of the diffraction pattern and reconstructed particle, a copepod, are shown in the insets. An out-of-focus target (i.e., target is at a different depth than the reconstruction) is visible in the upper right-hand corner of the reconstruction.

2 Materials and Procedures

2.1 Instrumentation and Data

Data were collected using Deep-See, a towed instrumentation platform that integrates active acoustic, optical, and environmental sensors to study the mesopelagic zone (200-1000 m depth) of the ocean [Bassett et al., 2020]. Deep-See includes a custom Seascan Inc. digital in-line holographic microscope (DIHM) with a 16 megapixel camera and a 658 nm wavelength collimated laser light source with a 102.4 cm separation distance (Figure 1). Recorded holograms are 4864 x 3232 pixels. This configuration results in a pixel size of 7.4 μm (hologram dimensions are 3.6 cm x 2.4 cm; imaging volume of 0.88 L). After passing through the imaging volume, the laser travels 23 cm in air within the camera housing before reaching the camera lens. This relatively large imaging volume configuration was selected to maximize detections of zooplankton in sparse environments such as our study region. A diagram of the DIHM and sample data are shown in Figure 1.

Deep-See was deployed 10 times (approximately 100 total hours of holographic data collection) during a 24 July - 7 August 2019 cruise in the Northeast Atlantic off of the New England continental shelf break. During these deployments, vessel speed was approximately two knots. The holographic camera collected data at 1 Hz, meaning that approximately 360,000 holograms were acquired over the course of the cruise. Here, we use data from one of these deployments, conducted on August 4, 2019, to develop and validate automatic processing methods. During this 11-hour deployment, the depth of the platform varied between 0 and 800 m and all data were collected during daytime hours.

3

119   Data processing is implemented in Python and performed on a desktop computer with an i9-
120 9900 Intel(R) Core Processor and an Intel(R) UHD Graphics 630 GPU. We present processing times on this
121 computer for comparison with alternative methods, but note that processing times will depend on computing
122 resources.

2.2   Hologram Reconstruction

124 Where required, hologram reconstruction is performed using the angular spectrum method following the
125 algorithm presented in Latychevskaia and Fink [2015]. For a given hologram, $H(x, y)$, the reconstruction,
126 $U(x, y)$, is calculated as:

$$U(x,y) = FT^{-1}\left(FT(H(x,y)) \times \exp \frac{2\pi i z_o}{\lambda}\sqrt{1 - (\lambda u)^2 - (\lambda v)^2)} \right), \tag{1}$$

127 where FT and FT^{-1} indicate the Fourier transform and inverse Fourier transform, respectively; λ is the
128 laser wavelength, in m; and u and v denote the Fourier domain coordinates. For a given reconstruction, the
129 inverse optical path length, z_o, is:

$$z_o = \int_0^z \frac{1}{n(z)} dl, \tag{2}$$

130 where z is the physical path length from the camera to the reconstruction depth, and $n(z)$ is the index of
131 refraction along the laser path [Loomis, 2011]. For the DIHM system used here, starting at the camera, the
132 laser travels through air ($n_a = 1.0003$) for a distance $L_a = 23$ cm, before travelling through seawater ($n_w = $
133 1.3314) for a distance $L_w = 102.4$ cm. Therefore, within the imaging volume ($23 < z < 125.4$ cm), z_o can
134 be related to z:

$$z_o = L_a/n_a + (z - L_a)/n_w. \tag{3}$$

135 Slight variations in the indices of refraction due to temperature change with depth and geographical location
136 are negligible [Bashkatov and Genina, 2003]. In the following sections, we refer to the physical path length,
137 z, as the reconstruction depth.

2.3   Hologram Selection: Detecting Optical Turbulence

139 Optical turbulence (i.e., variations in the index of refraction due to turbulence-induced fluctuations in tem-
140 perature and salinity) [Korotkova, 2019] created interference in some holograms collected by Deep-See (Figure
141 2), presenting a challenge for automatic target detection. In some cases, it was still possible to classify targets
142 in reconstructions of these holograms, but interference resulted in a large number of false positive target de-
143 tections and reduced the efficacy of focus metrics. Because these false positives result in high CPU loads and
144 inaccurate estimates of particle abundance, we use an automatic classification scheme to detect holograms
145 containing interference from optical turbulence and remove them from the processing pipeline. We note that
146 while optical turbulence is "noise" in this application, these holograms resemble data used to study optical
147 turbulence [Bogucki et al., 1994, Kulikov, 2016], and may contain useful information about temperature and
148 salinity gradients. While a quantitative analysis of trends in optical turbulence is beyond the scope of this
149 work, qualitatively, holograms affected by optical turbulence were more prevalent near the surface (depth
150 less than 100 m).

2.3.1   Optical Turbulence Detection Training Data

152 A training data set containing 328 holograms (50% had visible interference from turbulence) was assembled
153 to train the hologram selection algorithm. The training data included holograms from water depths ranging
154 from 0 to 500 m to ensure that the algorithm was insensitive to ambient light levels and turbulence length

4

scales. Because interference from optical turbulence is straightforward to identify in the raw holograms (Figure 2), human annotation was performed on the raw, unreconstructed holograms.

2.3.2 Optical Turbulence Detection Algorithm

A transfer-learning approach [Weiss et al., 2016] is used for classification. Generally, transfer learning refers to approaches where an existing, trained machine learning model is used as a starting point for a different classification task. This allows for fine-tuning of the model using a relatively small volume of training data, while still leveraging information learned from massive training data sets. Here, the VGG19 deep neural network (pre-trained using the ImageNet database) [Simonyan and Zisserman, 2015] is used to extract features from each hologram. Because the VGG19 network was trained using 224 x 224 pixel color images (3 color planes), and the recorded holograms are 4864 x 3232 gray-scale images (1 color plane), the holograms are re-scaled to 224 x 224 pixels and tiled to create an array with dimensions 244 x 224 x 3 prior to feature extraction. This was necessary to match the dimensions of the VGG19 input layer. The feature extraction segment of the VGG19 network produces a vector of 1000 features for each hologram. Extracted features from the training data set were normalized by subtracting the mean and scaling to unit variance, then used to train a support vector machine (SVM) [Burges, 1998] with a radial basis function (RBF) kernel and $\gamma = 0.001$ (1/number of features) for binary classification.

2.3.3 Optical Turbulence Detection Evaluation

Due to the relatively small training data set, cross-validation was performed using leave-one-out validation. The SVM was trained as described in the previous section using all but one hologram in the training data set, and then used to classify the excluded hologram, before repeating this process for each hologram in the training data set.

To assess classification performance, precision, recall, and accuracy are calculated. Here, precision is defined as the fraction of holograms predicted to contain interference from optical turbulence that were correctly classified; recall is the number of holograms annotated as containing interference that were correctly classified; and accuracy is the fraction of all holograms that were correctly classified.

2.4 Diffraction Pattern Detection

In most holograms in the collected data, particle distribution is relatively sparse (i.e., most of the imaging volume does not contain a particle). Therefore, computation time can be reduced by avoiding reconstruction of regions of the hologram that do not contain a particle (i.e., background). Here, we present an algorithm to detect the diffraction patterns of particles in unreconstructed holograms. We note that this approach may not be appropriate for data sets where particle density is higher (e.g., Nayak et al. [2018]).

2.4.1 Diffraction Pattern Detection Algorithm

The diffraction pattern detection algorithm is outlined in Figure 3. As in Davies et al. [2015], Guo et al. [2021], and [Nayak et al., 2018], the first processing step is background subtraction. Nonuniform light levels, scratches in the camera housing, biofouling, and slight camera misalignment resulted in nonuniformity in the background intensity. Because light levels vary with water depth and time of day and biofouling may accrue over the course of a deployment, the background intensity level is updated for each hologram using the median of the previous 5 recorded holograms rather than using a static background for all data. The background intensity level is then subtracted from each hologram before further processing (Figure 3b). Given the data acquisition rate of 1 Hz and the vessel speed of 2 knots, the imaging volumes of successive
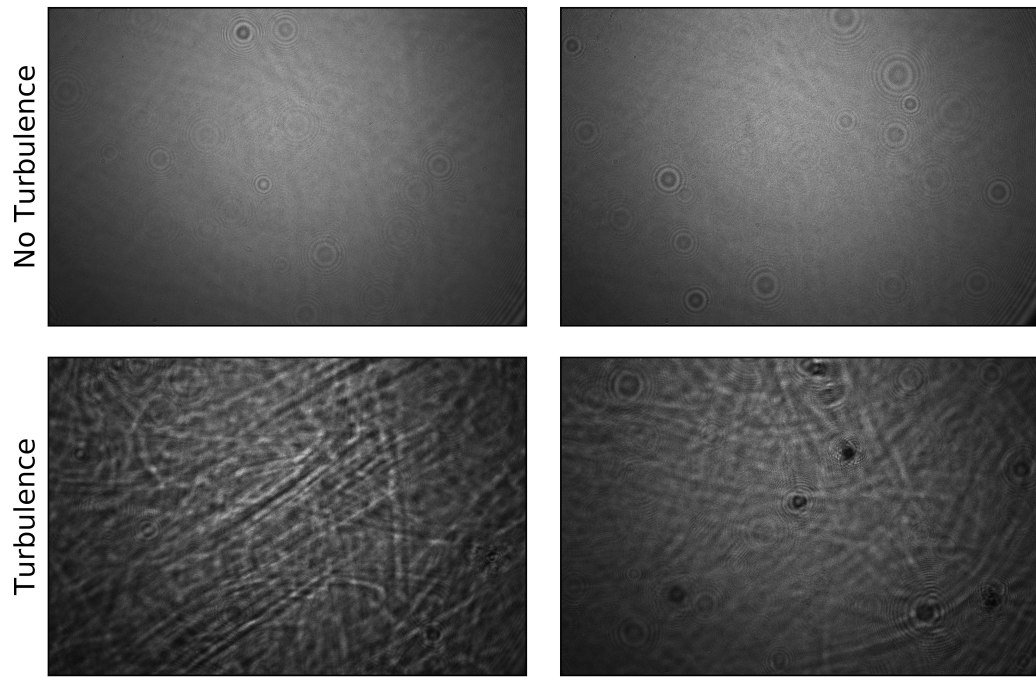
Figure 2: Top: Raw holograms without interference from optical turbulence. Bottom: Raw holograms with interference from optical turbulence. Diffraction patterns of targets are visible in the rightmost example, but interference from turbulence reduces the efficacy of automatic target detection and results in false positive target detections.
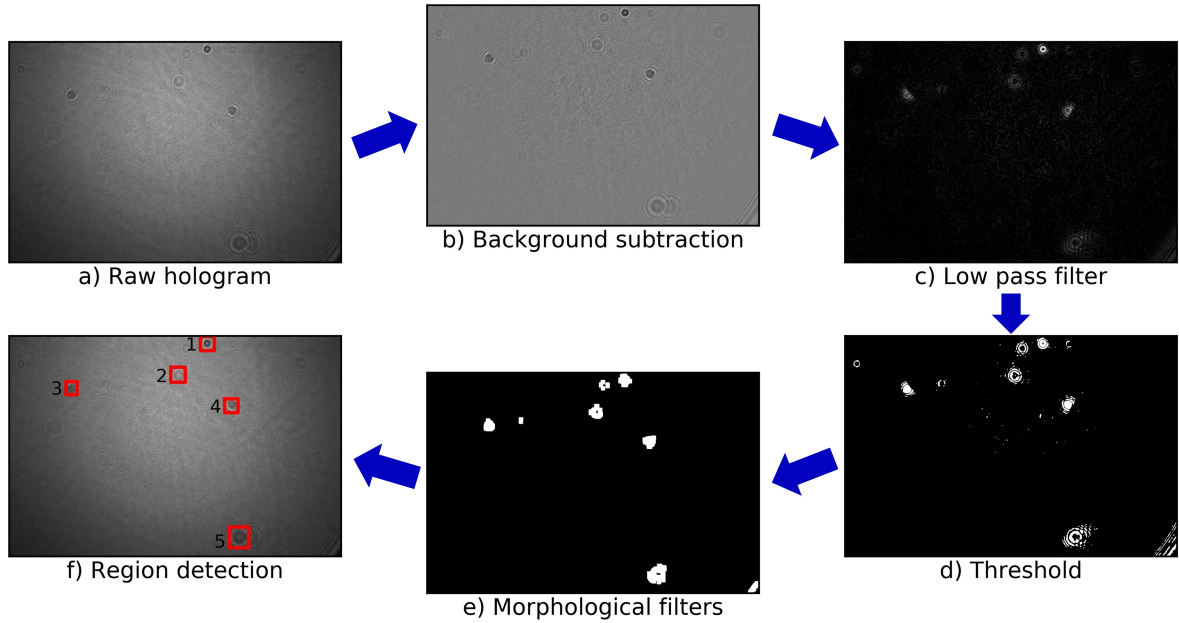
6

Figure 3: Flow chart detailing the diffraction pattern detection algorithm. a) Raw hologram before process-ing, b) hologram after background subtraction, c) background-subtracted hologram after applying low-pass filter retaining only the lowest 5% of spatial frequencies (colors are inverted to highlight diffraction patterns), d) application of a threshold to create a binary image, e) morphological opening to join pixels associated with the same diffraction pattern, and f) diffraction pattern bounding boxes shown on raw hologram. Diffraction patterns are numbered for reference in subsequent figures.

holograms did not overlap, allowing us to use this relatively small window for background subtraction. For a stationary system or a system with a faster acquisition rate, a larger background window may be necessary to avoid targets that persist in multiple holograms.

Next, a series of filters are applied to highlight particle diffraction patterns. First, the two-dimensional Fourier transform of the background-subtracted hologram is computed and a low-pass filter is applied to retain only the lowest 5% of spatial frequencies. The inverse Fourier transform of the filtered frequency-domain data yields a hologram where the diffraction patterns of particles are clearly distinguishable (Figure 3c). To separate particle diffraction patterns from the background, an empirically-tuned intensity threshold is applied to create a binary image (Figure 3d). Then, morphological opening (erosion by a square structuring element of size l_e, followed by dilation by a square structuring element of size l_d) is performed to remove noise and close any holes (Figure 3e). Finally, any regions in the binary image larger than an empirically-tuned area threshold, A_{min}, are detected (Figure 3f). The bounding box of each detected diffrac-tion pattern (smallest box containing the area in the binary image) is retained for reconstruction and focusing of the particle (see Section 2.5).

2.4.2 Diffraction Pattern Detection Evaluation

To evaluate the performance of the diffraction pattern detection algorithm, the positions of detected diffrac-tion patterns were compared to target detection on high-resolution full hologram reconstructions. To do this, we followed a similar approach as in Nayak et al. [2018]: a focus metric is used to determine the depth of

7

Test ID	A_{min} (mm^2)	l_e (pixels)	l_d (pixels)
A	0.5	20	60
B	1	20	60
C	2	20	60
D	1	10	60

Table 1: Diffraction pattern detection parameters: A_{min}, the diffraction pattern area threshold; l_e, the size of the morphological erosion filter; and l_d, the size of the morphological dilation filter

²¹³ any particles in sub-regions of a high-resolution reconstruction of the full hologram. First, the hologram was
²¹⁴ reconstructed with approximately 1 mm depth resolution (1000 reconstructions per hologram). Then, the
²¹⁵ hologram was divided into a 1 mm x 1 mm grid (135 x 135 pixels), and a focus metric was used to determine
²¹⁶ which, if any, reconstruction contained an in focus target within each grid cell. Nayak et al. [2018] determined
²¹⁷ which reconstruction plane was most in-focus based on the region that had the most pixels with intensity
²¹⁸ above an empirically tuned threshold. Here, we avoid tuning and instead employ the standard deviation
²¹⁹ correlation function, f_{sc}, as a focus metric, which Fonesca et al. [2016] found was both unimodal and offers
²²⁰ an reasonable trade-off between computation time and performance. For a given region of a reconstruction,
²²¹ $U(x, y)$:

$$f_{sc} = \frac{1}{MN} \sum_{m=1}^{M-1} \sum_{n=1}^{N-1} \frac{U(m,n)U(m+1,n+1)}{\mu_U} - MN\mu_U^2, \tag{4}$$

²²² where M and N are the dimensions of $U(x, y)$, and μ_U is the average intensity of $U(x, y)$. This metric is
²²³ slightly modified from the definition in Equation 16 of Fonesca et al. [2016] to normalize the reconstruction
²²⁴ by the mean, as we found this improved unimodality in variable light conditions.

²²⁵ For each grid cell, the reconstruction with the highest value of f_{sc} was retained to create a composite,
²²⁶ two-dimensional image containing all focused targets in a single plane. If f_{sc} did not exceed an empirically
²²⁷ tuned threshold ($f_{sc} = 1.5$) for any reconstruction in a given grid cell, that grid cell was assumed to not
²²⁸ contain a target and set to zero in the composite image. This approach requires the assumption that there is
²²⁹ only one particle in each 1 mm x 1 mm grid cell. Particles larger than 1 mm x 1 mm should still be focused
²³⁰ assuming that the focus metric is maximized for each region of the particle.

²³¹ Thresholding and a series of morphological filters were then used to detect targets in the composite
²³² image. First, the composite image was binarized by applying Otsu's method [Otsu, 1979] to each individual
²³³ grid cell. Then, morphological opening was performed to remove noise in the image and close small holes.
²³⁴ Finally, targets in the binarized image were extracted, and those whose longest dimension (major axis) did
²³⁵ not exceed .07 mm (10 pixels) were rejected, as the low resolution of these targets would likely preclude even
²³⁶ coarse human classification.

²³⁷ This ground truth target detection scheme was implemented for 217 randomly selected holograms.
²³⁸ The bounding boxes of ground truth targets were compared to those extracted using the diffraction pattern
²³⁹ detection method described in the previous section. A ground truth target was considered to be detected
²⁴⁰ if more than 75% of the target area was contained within the diffraction pattern bounding box. Target
²⁴¹ detection accuracy is presented as a function of the size of the ground truth target, defined as its area in the
²⁴² binarized composite image, and is analyzed for varying values of A_{min} and l_e, as described in Table 1.

2.5   Target Reconstruction

²⁴⁴ Reconstruction is performed using only a small window around each detected diffraction pattern. Initial
²⁴⁵ analysis of this approach showed that, when a sufficient window around the diffraction pattern bounding
²⁴⁶ box is used, reconstruction using only the region of the hologram around the detected diffraction pattern
²⁴⁷ produces imagery that is nearly as high-resolution as reconstruction using the full hologram, but at much
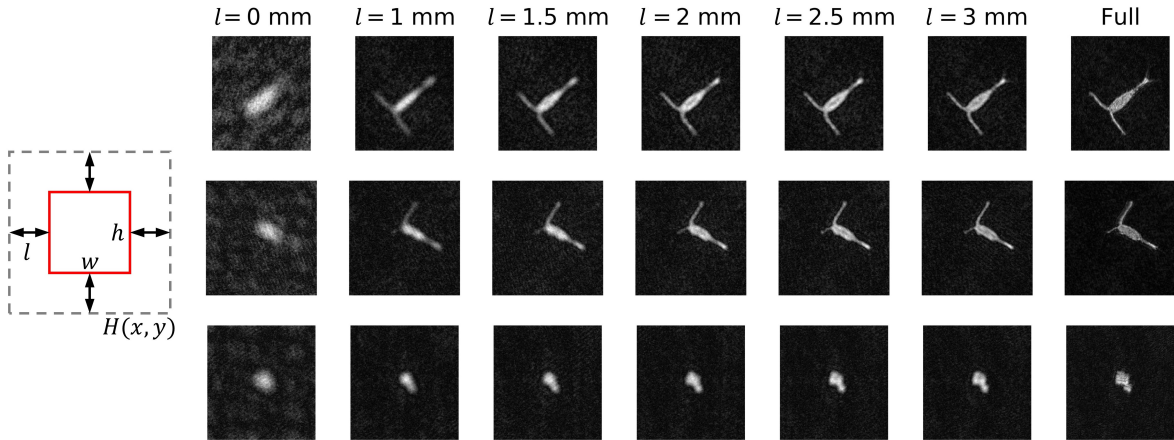
Figure 4: Left: Illustration of target bounding box (red) and reconstruction window (grey). Right: Target reconstruction using varying window sizes for diffraction patterns 1, 2, and 3 detected in the raw hologram shown in Figure 3. Columns indicate the size of the window around the bounding box included in the reconstruction, l. The rightmost column for each target shows reconstruction resolution when the full hologram is reconstructed (e.g., all available information is used).

lower computational cost. We define the window around a diffraction pattern used for reconstruction as follows: if the bounding box of the diffraction pattern is (x, y, w, h), where x and y are the coordinates of the lower left-hand corner of bounding box, respectively, and w and h are the width and height of the bounding box, respectively, the region of the hologram used in reconstruction ($H(x, y)$ in Equation 1) is $(x - l, y - l, w + 2l, h + 2l)$, where l is the window size (Figure 4).

2.5.1 Target Reconstruction Algorithm

Reconstruction and focusing of detected diffraction patterns are performed in parallel using the golden section search algorithm [Woodford and Phillips, 1997] to find the maximum value of f_{sc} (Equation 4). For the DIHM system used here, the minimum and maximum values of z are $z_{min} = 0.230$ m and $z_{max} = 1.254$ m (i.e., any target within the imaging volume must be in focus at some depth, z_f, between z_{min} and z_{max}). Reconstruction of the extracted hologram region is performed at each value of z used in the golden section search, and f_{sc} is calculated for each reconstruction to determine the next reconstruction depth. Assuming $f_{sc}(z)$ is unimodal, this approach reduces the number of reconstructions required to determine z_f with 1 mm precision from 1000 to 31. Figure 5 shows $f_{sc}(z)$ for diffraction pattern 2 in Figure 3, including the points used in the golden section search. We note that this approach could be employed with any unimodal focus metric, but f_{sc} proved to perform well for this data set.

2.5.2 Target Reconstruction Evaluation

The combined reconstruction and focusing algorithm was evaluated using 321 targets identified during human review, including marine snow and various species of plankton. First, the performance of f_{sc} as a focus metric was evaluated using a 1 mm resolution reconstruction of the target (1000 reconstructions per target) with a window size of $l = 2.5$ mm, and the reconstruction with the maximum value of f_{sc} was stored for later analysis. Then, the golden section search reconstruction/focusing algorithm was implemented with 1 mm resolution for the same 321 targets.

9
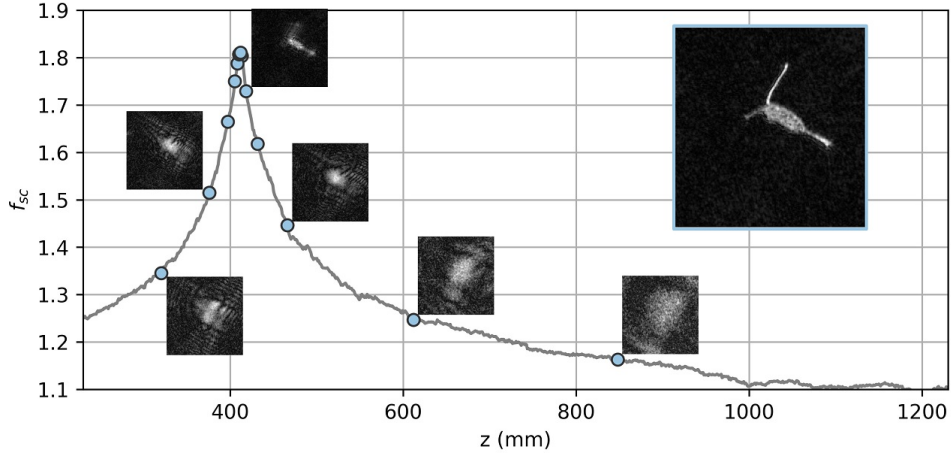
Figure 5: Standard deviation correlation function focus metric (f_{sc}) as a function of distance from the camera for diffraction pattern 2 in Figure 3 (a copepod). The 31 points used in the golden section search are indicated by blue dots, and reconstructions at representative points used by the search algorithm are shown. The large inset shows the reconstruction at the maximum value of f_{sc}. Note that resolution increases close to the focus depth and not all points used by the golden section search are visible in this figure due to overlap.

The automatically focused reconstructions produced by both methods (full reconstruction and golden section search) were reviewed manually to determine whether the reconstructed target was adequately in focus for classification. Each reconstructed target was annotated as 1) in focus, 2) slightly out of focus, but sufficiently focused for coarse classification, or 3) out of focus (classification not possible). We note that the distinction between 1) and 2) is qualitative, and differentiation between these two classes is subject to human bias.

2.6 End-to-End Evaluation

The full hologram processing pipeline (hologram selection, target detection, and target reconstruction/focusing) was used to process 5,663 holograms sampled by Deep-See. These holograms were selected from two regions of the deployment where the depth of Deep-See was relatively constant: approximately 0.5 hours where Deep-See was at 300 m depth, and approximately 1.5 hours where Deep-See was at 450 m depth. First, the hologram selection algorithm was used to determine whether turbulence precluded automatic target detection. Diffraction patterns were then detected using the parameters in Test B in Table 1 and reconstructed following the procedures in Section 2.5. These diffraction pattern detection parameters were selected as they offered an optimal trade-off between precision and recall for targets larger than 0.1 mm^2 (see Section 3.2).

A window size of $l = 2.5$ mm was used for determination of z_f, before calculating a higher-resolution reconstruction at z_f using a window size of $l = 6$ mm. These window sizes were selected empirically and validated through analysis of the automatically focused targets (i.e., a satisfactory fraction of extracted targets was sufficiently focused for classification). A JPEG image containing each reconstructed target was recorded for later analysis. The title of each file contained the timestamp of the hologram from which it originated, the bounding box of the target, and z_f.

The two-dimensional area of each reconstructed target, A_t, was used as a proxy for target size. As in Section 2.4, Otsu's method was used to binarize the extracted bounding box and separate foreground and background pixels (Figure 6). The number of targets detected as a function of A_t and z_f is used to evaluate

10
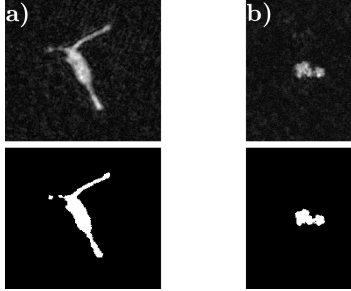
Figure 6: Two representative extracted targets and their binarized images after applying Otsu's method. a) shows a copepod and b) shows a marine snow particle.

biases in the processing pipeline. Finally, the reconstructed targets from 534 randomly selected holograms were manually reviewed and labeled as sufficiently or insufficiently focused for human classification.

3 Assessment

3.1 Hologram Selection

In cross-validation, the combined VGG19-SVM transfer learning approach classified the 328 holograms in the training data set with 94% precision, 91% recall, and 93% accuracy. We note that this classification model would likely not produce the same results for data collected with a different DIHM or in a different environment without re-training with additional data.

3.2 Target Detection

Figure 7 shows the fraction of ground truth targets detected using the diffraction pattern detection method (recall) as a function of target size for each set of diffraction pattern detection parameters (Table 1) and the number of ground truth targets in each size range. Target abundance decreased with target size, and data are only presented for target size ranges where more than two targets were detected. Recall varied with the size of the area threshold, A_{min}, and the size of the structuring element used for morphological erosion, l_e. Precision also varied between tests: 85%, 94%, 95%, and 82% of detected diffraction patterns corresponded to a ground truth target for tests A, B, C, and D, respectively.

For tests A, B, and D, 99% of targets larger than 0.1 mm^2 were detected. For test C (largest value of A_{min}), only 82% of targets larger than 0.1 mm^2 were detected, indicating that the diffraction patterns of many targets in this size range were smaller than 2 mm^2 in the hologram. The superior performance of test D (smallest value of l_e) to test A (smallest value of A_{min}) for targets smaller than 0.1 mm^2 indicates that the erosion filter may have removed the diffraction patterns of some smaller targets. However, the larger erosion filter also reduced the number of false positives, resulting in higher precision for test A than test D. For all tests, a steep roll-off in detection capabilities is observed for targets smaller than 0.075 mm^2. While larval organisms or small particles of marine snow may be smaller than this size threshold, many species of copepods, an abundant organism of particular interest in the mesopelagic [Koski et al., 2020], will be significantly larger than this threshold given body lengths exceeding 1 mm [Conway, 2006]. Further, given the pixel resolution of the DIHM, such particles would have fewer than 1370 pixels (37 pixel square), and likely be difficult to classify.

11
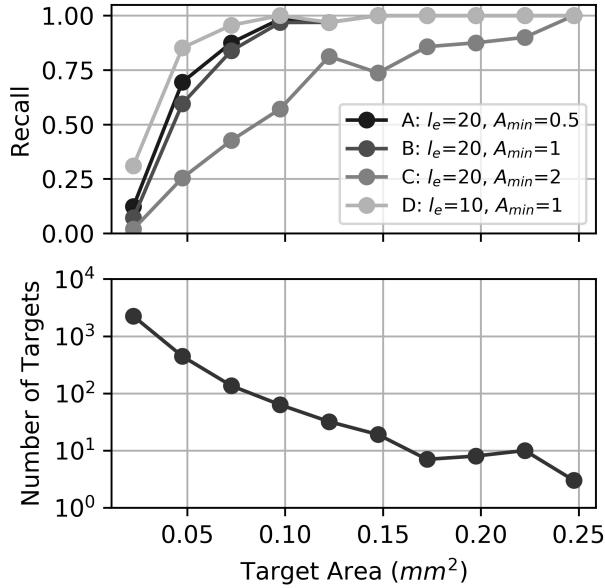
Figure 7: Top: Fraction of targets detected using the diffraction pattern detection method (recall) as a function of target area for each test in Table 1 (l_e is the size of the morphological erosion filter, and A_{min} is the diffraction pattern area threshold); bottom: number of ground truth targets as a function of target area. Note that the y-axis is on a logarithmic scale.

3.3 Target Auto-focusing

Figure 8 shows the percent of targets that were in focus, sufficiently focused for classification, and out-of-focus using the two focusing methods, and Figure 9 shows results of automatic focusing and $f_{sc}(z)$ for several representative targets. The standard deviation correlation function performed relatively well as a focus metric - using 1 mm resolution "full" reconstruction, 89% of targets were either in focus or sufficiently focused at the maximum value of f_{sc}. The golden section search method achieved nearly equivalent performance (88% of targets were in focus or sufficiently focused) while requiring 3% of the number of reconstructions. On average, the focus depths, z_f, determined by the two methods agreed within 5.8 mm, and they agreed within 1 mm for 62% of targets. Agreement between the two methods supports the assumption of unimodality required for the golden section search.

In two cases in the evaluation data, there were two targets at different reconstruction depths within the bounding box. In both cases, there were two peaks in f_{sc} and the two methods focused different targets. Identification of targets with overlapping diffraction patterns is a limitation of any approach that assumes a target is focused where a focus metric is maximized. While the relative sparsity of targets in the data presented here indicates that this will not have a significant impact on results, detection of multiple targets with overlapping diffraction patterns could be addressed through implementation a peak-finding method that allows for multiple targets in the same reconstruction. We note that in the results represented in Figure 8, these two cases were classified as "out of focus."

12
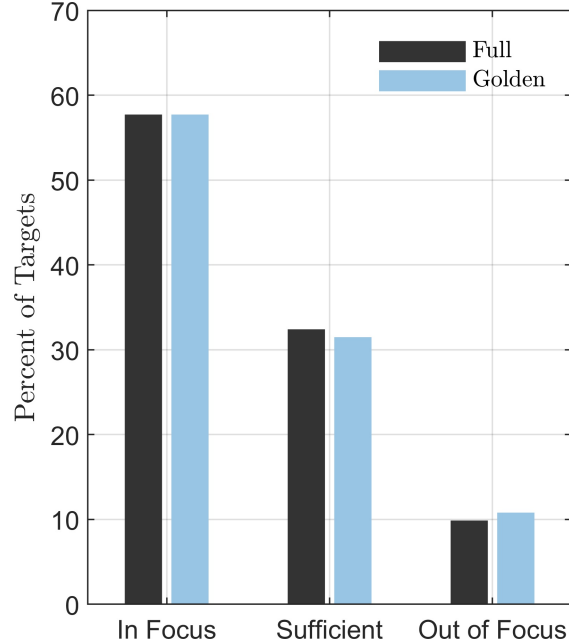
Figure 8: Fraction of targets using each focusing method classified by a human reviewer as in focus, sufficiently focused for classification, and out-of-focus using full reconstructions ("Full") and the golden section search method ("Golden").

3.4 End-to-End Evaluation

Only 0.2% (12 holograms) of the processed data were classified as containing optical turbulence and removed from the processing pipeline. An average of 2.9 target diffraction patterns were detected and reconstructed in each hologram. 83% of human-reviewed reconstructed regions contained a target that was sufficiently focused for human classification. Of the remaining 17% of reconstructed regions, 82% contained an out-of-focus target that could not be classified and 18% did not contain a visible target. The latter case does not necessarily indicate that a target was not present (false positive), because it is possible that the automatically-determined focus depth was sufficiently far from the true focus depth that the target was not visible in the reconstruction.

Figure 10 shows the fraction of targets detected at varying ranges from the camera, p_A, for different particle sizes for all processed holograms. For targets smaller than 0.1 mm^2, the probability of target detection generally decreases with range (though a slight increase is observed between the 250-235 mm and 435-639 mm range bins). This is likely a result of the fact that that the diffraction patterns of smaller targets farther from the camera are lower intensity due to attenuation of the light. In conjunction with the relatively low target detection recall for targets in this size range (Figure 7), this indicates that accurate particle abundance estimates cannot be obtained for targets smaller than 0.1 mm^2 using these detection parameters.

For targets between 0.1 and 0.3 mm^2, more targets are detected towards the center of the imaging volume ($z = 435\text{-}1049$ mm), indicating that while target detection recall is high for particles in this size range (Figure 7), hydrodynamic effects from the camera and laser housings may influence particle position near the edges of the imaging volume. Representative abundance estimates for particles in this size range may require limiting analysis to the center of the imaging volume. This trend is not as clear for larger particles (> 0.3 mm^2), though this may be due to the relatively small number of particles in this size range ($n = 200$).

On average, our processing pipeline took 8 seconds to detect and reconstruct targets from a holo-

13
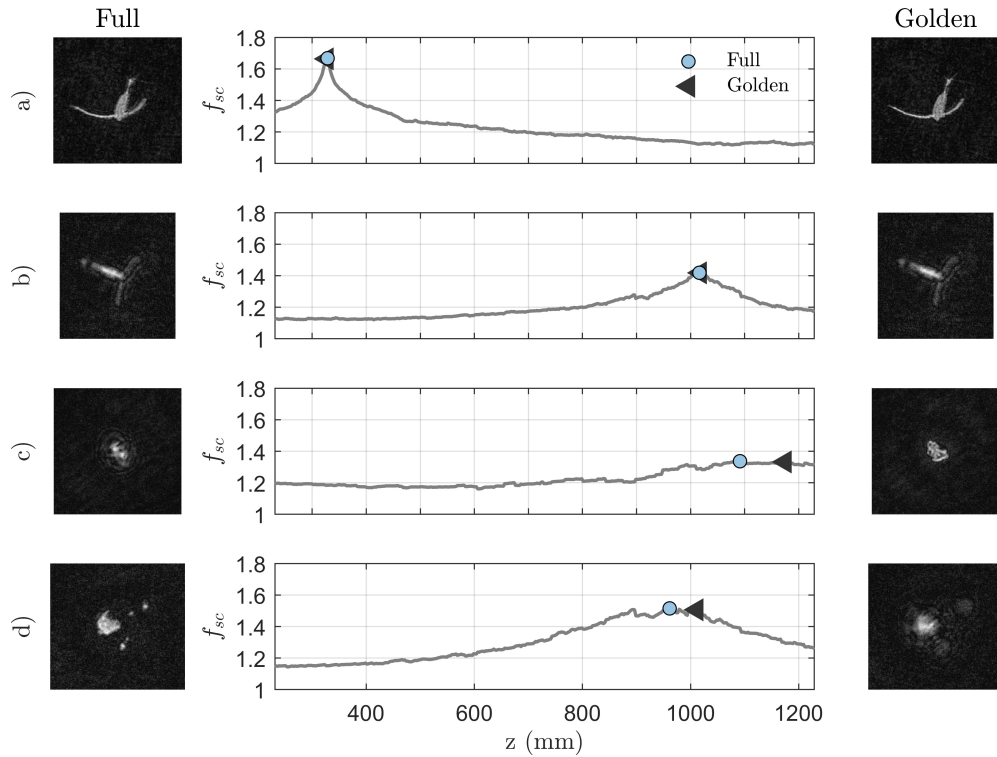
Figure 9: The focus metric, f_{sc} for the full reconstruction and the results of automatic focusing using a 1 mm resolution full reconstruction of the target ("Full") and the golden section search method ("Golden") for four representative targets: a) a copepod which was in focus using both methods, b) a copepod which was classified as sufficiently focused for coarse classification, c) marine snow which was focused using the golden section search method, but not the full reconstruction, and d) a copepod which was focused using the full reconstruction, but not the golden section search method.
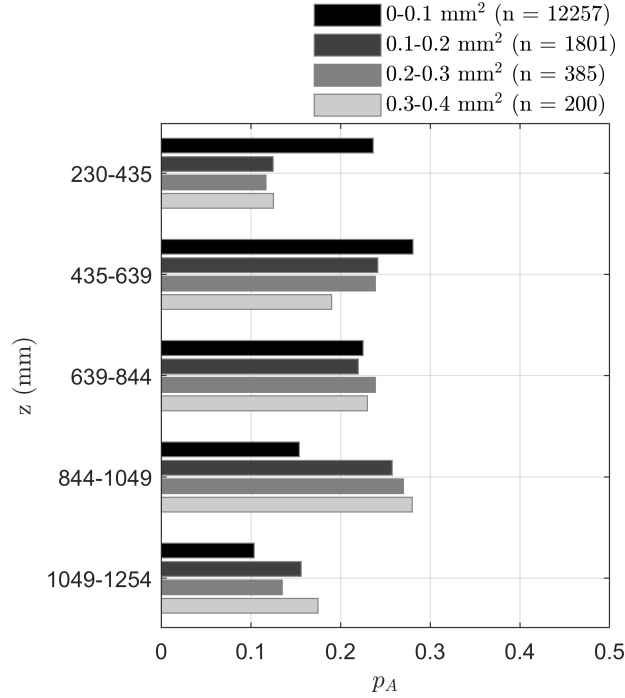
Figure 10: Probability of target detection as a function of particle size and range from the camera.

gram, though computation time varied with the number and size of detected targets. For comparison, calculation of full 1 mm resolution reconstructions (Section 2.4.2) took approximately 16 minutes per hologram, not including time required to write reconstructions to disk or detect targets. We note that reconstruction was performed on the GPU for both methods. The extracted targets from all 5663 processed holograms totaled 500 MB, compared to 6 MB required for a hologram reconstruction at a single depth (6 GB for a full 1 mm resolution reconstruction). All reconstructions and extracted targets were stored as JPEG files.

4   Discussion and Conclusions

This paper addresses the computational challenges of analyzing underwater holograms with the development and demonstration of a processing pipeline to rapidly extract focused targets from underwater digital holograms. This processing pipeline includes selection of holograms for target detection, detection and windowing of diffraction patterns in the raw hologram, then reconstruction of each window.

The classification algorithm for hologram selection focused on excluding images with interference from optical turbulence; these holograms are most likely to result in ineffective target detection/reconstruction, and are therefore not worth the computational expense of reconstruction. We note that the exclusion of these holograms from further processing may result in systematic biases in particle density estimations at strong density interfaces where optical turbulence is persistent, and this should be considered in interpretation of results. The transfer learning approach was found to have 93% accuracy, and has potential in many applications for selection of particular holograms most likely to result in successful reconstruction. For example, one use of DIHM is detection of oil droplets and gas bubbles in water, useful in the area of oil spill mapping and response [White et al., 2016]. Automatic target detection can be used to quantify droplet count and therefore oil density, however, like the data analyzed in this work, not all holograms are suitable for automatic processing. When oil droplet density is too high, the hologram becomes saturated, and reconstruction of these holograms results in artificially low droplet estimates. Preliminary analysis of holographic data of oil droplets collected in a flume has found that a similar transfer-learning based approach

15

is suitable for identification of saturated holograms in a data set collected using a Seascan DIHM in an oil flume.

In the second step of the processing pipeline, the diffraction patterns of targets are detected in the raw holograms. When compared to target detection using full hologram reconstructions, this approach detected over 99% of targets larger than 0.1 mm^2 with 95% precision. Smaller targets can be detected through tuning of diffraction pattern detection parameters, though this increased the number of false positives (decreasing precision). However, smaller targets were less frequently detected at longer range from the camera, likely due to attenuation of the light, and this should be considered when estimating particle density. The minimum detectable target size will be a function of the size of the imaging volume and the resolution of the camera.

Finally, detected diffraction patterns are simultaneously reconstructed and focused using the golden section search algorithm to maximize a focus metric, achieving nearly equivalent performance to application of the same focus metric on a high-resolution reconstruction. This approach leverages the fact that in regions where target density is low, like the mesopelagic, targets occupy a relatively small fraction of the imaging volume. This means that in a full, high-resolution reconstruction, the majority of reconstruction planes do not contain an in-focus target, and it is possible to reconstruct all targets within the imaging volume using a small fraction of the hologram. As such, application of the same method to data collected in coastal areas or other high-particle density regions is unlikely to yield significant computational gains.

These results are significant as they represent a large computational cost saving compared to full reconstruction followed by target detection/classification. The DIHM system used here is capable of collecting data at 10 Hz, or over 216000 holograms during a 6-hour deployment. With a full reconstruction time of 16 minutes per hologram, performing high-resolution full reconstructions of all holograms collected by this kind of holographic camera is infeasible. At the same time, there is a wealth of information in these data, and by selecting holograms suitable for processing, detecting the presence of particles based on their diffraction patterns, and using an optimized reconstruction scheme, large data set information gathering becomes tractable. In this paper, we demonstrated that this more selective reconstruction process provides sufficiently-focused images for classification at a 60 x savings in computational time. We anticipate that these methods are translatable to other holographic data where target distribution is relatively sparse, and have made processing codes available online[2].

The methods in this paper will allow for quantitative processing of long time series, large imaging volume DIHM data sets, which was previously not practical due to the computational requirements of hologram reconstruction. This methodology will enable the development of automatic target classification algorithms, which was not previously feasible given the computational cost of extracting individual targets from the data. Ultimately, this will allow for estimation of zooplankton abundance, studies of organism distribution, and comparison with the acoustic, optical, and genetic sensing capabilities of Deep-See.

5   Acknowledgements

---

[2]https://github.com/emma-d-cotter/Hologram-Processing

The code used to perform the analysis in this paper can be found at: https://github.com/emma-d-cotter/Hologram-Processing.

References

T.R. Anderson, A.P. Martin, R.S. Lampitt, C.N. Trueman, S.A. Henson, and D.J. Mayor. Quantifying carbon fluxes from primary production to mesopelagic fish using a simple food web model. *ICES Journal of Marine Science*, 76(3):690–701, 2018. doi: 10.1093/icesjms/fsx234.

A.N. Bashkatov and E.A. Genina. Water refractive index in dependence on temperature and wavelength: a simple approximation. In *Saratov Fall Meeting 2002: Optical Technologies in Biophysics and Medicine IV*, volume 5068, pages 393–395. International Society for Optics and Photonics, SPIE, 2003. doi: 10.1117/12.518857.

C. Bassett, E. Cotter, T.K. Stanton, and A.C. Lavery. Frequency- and depth-dependent target strength measurements of individual mesopelagic scatterers. *Journal of the Acoustical Society of America*, 148(2): EL153–EL158, 2020. doi: 10.1121/10.0001745.

D. Bogucki, A. Domaradzki, J.R.V. Zaneveld, and T.D. Dickey. Light scattering induced by turbulent flow. In *Ocean Optics XII*, volume 2258, pages 247–255. International Society for Optics and Photonics, SPIE, 1994. doi: 10.1117/12.190068.

C.J.C. Burges. A Tutorial on Support Vector Machines for Pattern Recognition. *Data Mining and Knowledge Discovery*, 2:121–167, 1998.

D.V.P Conway. *Identification of the copepodite developmental stages of twenty-six North Atlantic copepods*. Occasional Publications, Marine Biological Association of the United Kingdom, 2006.

J. Cooley and J. Tukey. An algorithm for the machine calculation of complex fourier series. *Mathematics of Computation*, 19:297–301, 1965. doi: 10.1090/S0025-5718-1965-0178586-1.

R.K. Cowen and C.M. Guigand. In situ ichthyoplankton imaging system (ISIIS): system design and preliminary results. *Limnology and Oceanography: Methods*, 6(2):126–132, 2008. doi: https://doi.org/10.4319/lom.2008.6.126.

E.J. Davies, D. Buscombe, G.W. Graham, and Nimmo-Smith W.A.M. Evaluating unsupervised methods to size and classify suspended particles using digital in-line holography. *Journal of Atmospheric and Oceanic Technology*, 32(6):1241–1256, 2015. doi: 10.1175/JTECH-D-14-00157.1.

C. S. Davis, F. T. Thwaites, S. M. Gallager, and Q. Hu. A three-axis fast-tow digital video plankton recorder for rapid surveys of plankton taxa and hydrography. *Limnology and Oceanography: Methods*, 3(2):59–74, 2005. doi: 10.4319/lom.2005.3.59.

E.S.R. Fonesca, P.T. Fiadeiro, M. Pereira, and A. Pinheiro. Comparative analysis of autofocus functions in digital in-line phase shifting holography. *Applied Optics*, 55(27), 2016. doi: 10.1364/AO.55.00766.

A.T. Greer, J.C. Lehrter, B.M. Binder, A.R. Nayak, R. Barua, A.E. Rice, J.H. Cohen, M.N. McFarland, A. Hagemeyer, N.D. Stockley, K.M. Boswell, I. Shulman, S. deRada, and B. Penta. High-resolution sampling of a broad marine life size spectrum reveals differing size- and composition-based associations with physical oceanographic structure. *Frontiers in Marine Science*, 7(542701), 2020. doi: 10.3389/fmars.2020.542701.

B. Guo, L. Nyman, A.R. Nayak, D. Milmore, M. McFarland, M.S. Twardowski, J.M. Sullivan, J. Yu, and J. Hong. Automated plankton classification from holographic imagery with deep convolutional neural networks. *Limnology and Oceanography: Methods*, 19:21–36, 2021. doi: 10.1002/lom3.10402.

O. Korotkova. Light propagation in a turbulent ocean. *Progress in Optics*, 64, 2019. doi: 10.1016/bs.po.2018.09.001.

476 M. Koski, B. Valencia, R. Newstead, and C. Thiele. The missing piece of the upper mesopelagic carbon
477 budget? Biomass, vertical distribution and feeding of aggregate-associated copepods at the pap site.
478 *Progress in Oceanography*, 181, 2020. doi: 10.1016/j.pocean.2019.102243.

479 V.A. Kulikov. Estimation of turbulent parameters based on the intensity scintillations of the laser beam
480 propagated through a turbulent water layer. *Journal of Applied Physics*, 119(123103), 2016. doi: 10.1063/
481 1.4944725.

482 T. Latychevskaia and H. Fink. Practical algorithms for simulation and reconstruction of digital in-line
483 holograms. *Applied Optics*, 54(9):2424–2434, 2015. doi: 10.1364/AO.54.002424.

484 A.C. Lavery, P.H. Wiebe, T.K. Stanton, G.L. Lawson, M.C. Benfield, and N. Copley. Determining dominant
485 scatterers of sound in mixed zooplankton populations. *The Journal of the Acoustical Society of America*,
486 122(6):3304–3326, 2007. doi: 10.1121/1.2793613.

487 F. Lombard, E. Boss, A.M. Waite, M. Vogt, J. Uitz, L. Stemmann, H.M. Sosik, J. Schulz, J. Romagnan,
488 M. Picheral, J. Pearlman, M.D. Ohman, B. Niehoff, K.O. Möller, P. Miloslavich, A. Lara-Lpez, R. Kudela,
489 R.M. Lopes, R. Kiko, L. Karp-Boss, J.S. Jaffe, M.H. Iversen, J. Irisson, K. Fennel, H. Hauss, L. Guidi,
490 G. Gorsky, S.L.C. Giering, P. Gaube, S. Gallager, G. Dubelaar, R.K. Cowen, F. Carlotti, C. Briseño-Avena,
491 L. Berline, K. Benoit-Bird, N. Bax, S. Batten, S.D. Ayata, L.F. Artigas, and W. Appeltans. Globally
492 consistent quantitative observations of planktonic ecosystems. *Frontiers in Marine Science*, 6:196, 2019.
493 doi: 10.3389/fmars.2019.00196.

494 N.C. Loomis. *Computational Imaging and Automated Identification for Aqueous Environments*. PhD thesis,
495 Massachusetts Institute of Technology and Woods Hole Oceanographic Institution, 2011.

496 A.R. Nayak, M.N. McFarland, J.M. Sullivan, and M.S. Twardowski. Evidence for ubiquitous preferential
497 particle orientation in representative oceanic shear flows. *Limnology and Oceanography*, 63(1):122–143,
498 2018. doi: 10.1002/lno.10618.

499 A.R. Nayak, E. Malkiel, M.N. McFarland, M.S. Twardowski, and J.M. Sullivan. A review of holography in the
500 aquatic sciences: In situ characterization of particles, plankton, and small scale biophysical interactions.
501 *Frontiers in Marine Science*, 7:1256, 2021. doi: 10.3389/fmars.2020.572147.

502 N. Otsu. A threshold selection method from gray-level histograms. *IEEE Transactions on Systems, Man,
503 and Cybernetics*, 9(1):62–66, 1979. doi: 10.1109/TSMC.1979.4310076.

504 J A Ratcliffe. Some aspects of diffraction theory and their application to the ionosphere. *Reports on Progress
505 in Physics*, 19(1):188–267, 1956. doi: 10.1088/0034-4885/19/1/306.

506 K. Simonyan and A. Zisserman. Very deep convolutional networks for large-scale image recognition. In
507 *International Conference on Learning Representations*, 2015.

508 H. Sun, D. C. Hendry, M. A. Player, and J. Watson. In situ underwater electronic holographic camera for
509 studies of plankton. *IEEE Journal of Oceanic Engineering*, 32(2):373–382, 2007. doi: 10.1109/JOE.2007.
510 891891.

511 H. Sun, P.W. Benzie, N. Burns, D.C. Hendry, M.A. Player, and J. Watson. Underwater digital holography
512 for studies of marine plankton. *Philosophical Transactions of the Royal Society A*, 366:1789–1806, 2008.
513 doi: 10.1098/rsta.2007.2187.

514 N.L. Walcutt, B. Knörlein, I. Cetinić, Z. Ljubesic, S. Bosak, T. Sgouros, A.L. Montalbano, A. Neeley,
515 S. Menden-Deuer, and M.M. Omand. Assessment of holographic microscopy for quantifying marine particle
516 size and concentration. *Limnology and Oceanography: Methods*, 18(9):516–530, 2020. doi: 10.1002/lom3.
517 10379.

518 K. Weiss, T.M. Khoshgoftaar, and D. Wang. A survey of transfer learning. *Journal of Big Data*, 3(9), 2016.
519 doi: 10.1186/s40537-016-0043-6.

18

520   H.K. White, R.N. Conmy, I.R. MacDonald, and C.M. Reddy. Methods of oil detection in response to the
521   deepwater horizon oil spill. *Oceanography*, 29:76–87, 2016. doi: 10.5670/oceanog.2016.72.

522   C. Woodford and C. Phillips. *Numerical Methods with Worked Examples*, chapter 7.3, pages 167–170.
523   Springer, 1997.

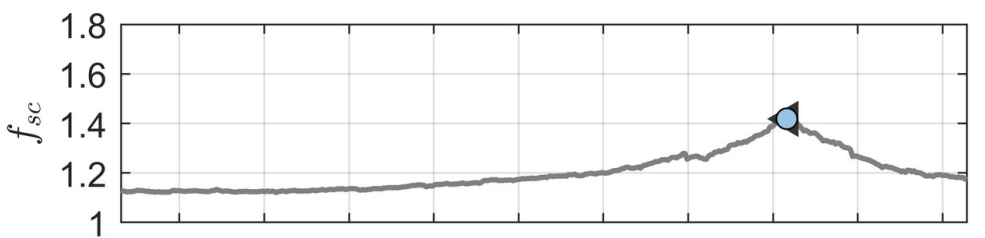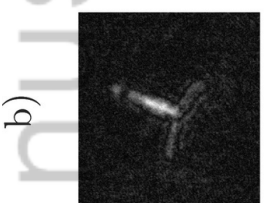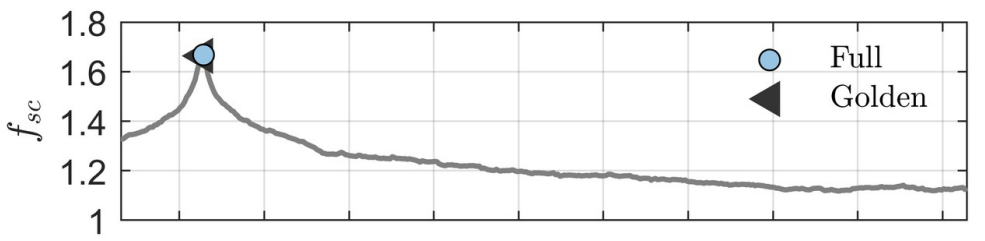Author Manuscript

19

List of Figures

20

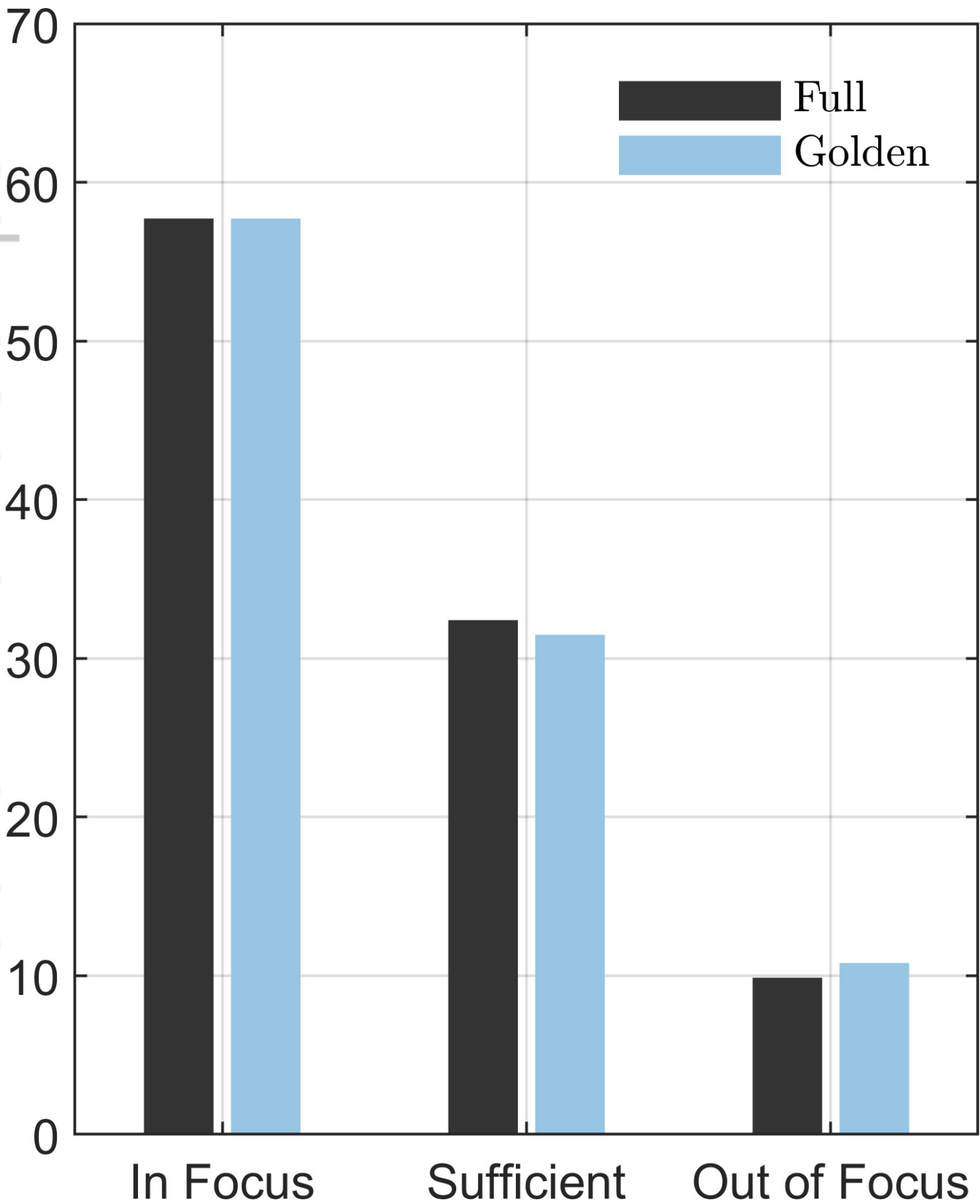LOM3_10438_detectionAccuracy.jpg

LOM3_10438_focusmetric Mockup.jpg

Full   Golden

LOM3_10438_focusresults2.jpg

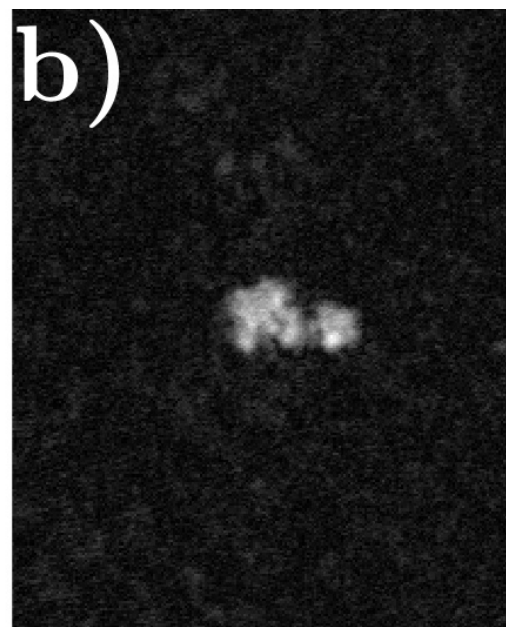LOM3_10438_focusresults.jpg

LOM3_10438_SystemDiagram.jpg

a) Raw hologram

b) Background subtraction
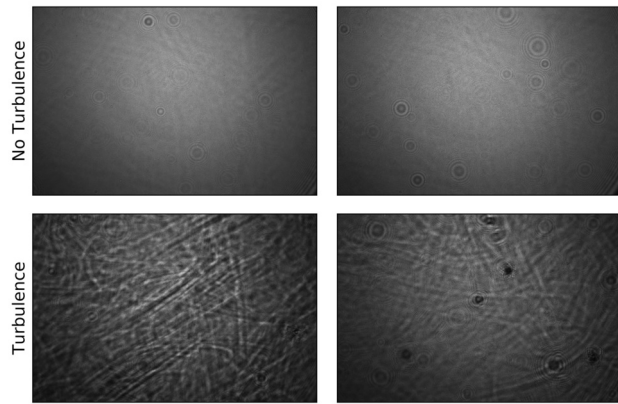
c) Low pass filter

d) Threshold

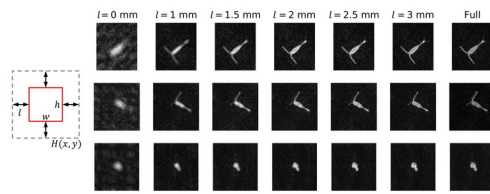e) Morphological filters

f) Region detection

LOM3_10438_Target Detection Mockup.jpg

LOM3_10438_thresholdexamples.jpg

LOM3_10438_turb_examples.jpg

LOM3_10438_WindowSizeFormatted.jpg