machine learning & knowledge extraction

MDPI

Article

Semantic Image Segmentation Using Scant Pixel Annotations

Adithi D. Chakravarthy [1,*], Dilanga Abeyrathna [1], Mahadevan Subramaniam [1], Parvathi Chundi [1] and Venkataramana Gadhamshetty [2]

[1] Computer Science Department, University of Nebraska at Omaha, Omaha, NE 68182, USA; dabeyrathna@unomaha.edu (D.A.); msubramaniam@unomaha.edu (M.S.); pchundi@unomaha.edu (P.C.)
[2] 2-Dimensional Materials for Biofilm Engineering Science and Technology (2DBEST) Center, South Dakota School of Mines & Technology, Rapid City, SD 57701, USA; venkata.gadhamshetty@sdsmt.edu
* Correspondence: achakravarthy@unomaha.edu

Abstract: The success of deep networks for the semantic segmentation of images is limited by the availability of annotated training data. The manual annotation of images for segmentation is a tedious and time-consuming task that often requires sophisticated users with significant domain expertise to create high-quality annotations over hundreds of images. In this paper, we propose the segmentation with scant pixel annotations (*SSPA*) approach to generate high-performing segmentation models using a scant set of expert annotated images. The models are generated by training them on images with automatically generated pseudo-labels along with a scant set of expert annotated images selected using an entropy-based algorithm. For each chosen image, experts are directed to assign labels to a particular group of pixels, while a set of replacement rules that leverage the patterns learned by the model is used to automatically assign labels to the remaining pixels. The *SSPA* approach integrates active learning and semi-supervised learning with pseudo-labels, where expert annotations are not essential but generated on demand. Extensive experiments on bio-medical and biofilm datasets show that the SSPA approach achieves state-of-the-art performance with less than 5% cumulative annotation of the pixels of the training data by the experts.

Keywords: image processing; image segmentation; machine vision; neural networks; semi-supervised learning

Citation: Chakravarthy, A.D.; Abeyrathna, D.; Subramaniam, M.; Chundi, P.; Gadhamshetty, V. Semantic Image Segmentation Using Scant Pixel Annotations. *Mach. Learn. Knowl. Extr.* **2022**, *4*, 621–640. https://doi.org/10.3390/make4030029

Academic Editor: Salim Lahmiri

Received: 24 May 2022
Accepted: 25 June 2022
Published: 1 July 2022

1. Introduction

Semantic image segmentation is the task of assigning to each pixel the class of its enclosing object or region as its label, thereby creating a segmentation mask. Due to its wide applicability, this task has received extensive attention from experts in several areas, such as autonomous driving, robot navigation, scene understanding, and medical imaging. Owing to its huge success, deep learning has become the de-facto choice for semantic image segmentation. Recent approaches have used convolutional neural networks (CNNs) [1,2] and fully convolutional networks (FCNs) [3–5] for this task and achieved promising results. Several recent surveys [6–11] describe the successes of semantic image segmentation and directions for future research.

Typically, large volumes of labeled data are needed to train deep CNNs for image analysis tasks, such as classification, object detection, and semantic image segmentation. This is especially so for semantic image segmentation, where each pixel in each training image has to be labeled or annotated in order to infer the labels of the individual pixels of a given test image. The availability of densely annotated images in sufficient numbers is problematic, particularly in domains such as material science, engineering, and medicine, where annotating images is time consuming and requires significant user expertise. For instance, while reading retinal images to identify unhealthy areas, it is common for graders (with ophthalmology training) to discuss each image at length to carefully resolve several confounding and subtle image attributes [12–14]. Labeling cells, cell clusters, and microbial byproducts in biofilms take up to two days per image on average [15–17]. Therefore, it is

highly beneficial to develop high-performance deep segmentation networks that can train with scantly annotated training data.

In this paper, we propose a novel approach for semantic segmentation of images that can work with datasets with scant expert annotations. Our approach, segmentation with scant pixel annotations (SSPA) combines active learning and semi-supervised learning approaches to build segmentation models where segmentation masks are generated using automatic pseudo-labeling as well as by using expert manual annotations on a selective small set of images. The proposed SSPA approach employs a marker-based watershed algorithm based on image morphology to automatically generate pseudo-segmentation masks for the full training dataset. The performance of a segmentation model generated using these pseudo-masks is analyzed, and a sample of images to be annotated by experts is selected. These images with expert-generated masks along with images with pseudo-masks are used to train the next model. The process is iterated to successively generate a sequence of models until either the performance improvement plateaus or no further refinements are possible. The approach uses top-k (bottom-k) image entropy over pixel prediction confidence probabilities to identify the sample images at each iteration.

Despite the careful selection of image samples in each iteration to limit the overall manual annotation effort, annotating each image in its entirety can be tedious for experts. This is especially so for the high information density of the segmentation task, where each and every pixel needs to be classified. The proposed *SSPA* approach reduces annotation effort by using pixels as the unit of annotation instead of annotating entire images or image patches. For each image that is selected for manual annotation, only pixels within a specified uncertainty range are marked for expert annotation. For the rest of the image, a set of replacement rules that leverage useful patterns learned by the segmentation model are used to automatically assign labels. The segmentation mask for an image includes these labels along with expert annotations, and is used to generate the next model. Using such directed annotations enables the *SSPA* approach to develop high-performance models with minimal annotation effort. The results of the *SSPA* approach are validated on biomedical and biofilm datasets and achieves high segmentation accuracy with less than 1% annotation effort. We also evaluated our method on a benchmark dataset for melanoma segmentation and achieved state-of-the-art performance with less than 1% annotation effort. The approach is general purpose and is equally applicable for the segmentation of other image datasets.

The rest of this paper is organized as follows. Section 2 discusses related work on semantic segmentation with scantly annotated data. The *SSPA* approach is detailed in Section 3. In Section 4, the datasets, network architecture, setup and evaluation metrics are presented. Experimental results and conclusions are discussed in Sections 5 and 6, respectively.

2. Related Work

Semantic segmentation [4] is one of the challenging image analyses tasks that has been studied earlier using image processing algorithms and more recently using deep learning networks; see [6,10,11,18] for detailed surveys. Several image processing algorithms based on methods including clustering, texture and color filtering, normalized cuts, superpixels, graph and edge-based region merging, have been developed to perform segmentation by grouping similar pixels and partitioning a given image into visually distinguishable regions [6]. More recent supervised segmentation approaches based on [4] use fully connected networks (FCNs) to output spatial maps instead of classification scores by replacing the fully connected layers with convolutional layers. These spatial maps are then upsampled using deconvolutions to generate pixel-level label outputs. Other decoder variants to transform a classification network to a segmentation network include the SegNet [19] and the U-Net [20].

Currently, deep learning-based approaches are perhaps the de facto choice for semantic segmentation. Recently, Sehar and Naseem [11], reviewed most of the popular learning

algorithms (∼120) for semantic segmentation tasks, and concluded the overwhelming success of deep learning compared to the classical learning algorithms. However, as pointed out by the authors, the need for large volumes of training data is a well-known problem in developing segmentation models using deep networks. Two main directions that were explored earlier for addressing this problem are the use of limited dense annotations (scant annotations) and the use of noisy image-level annotations (weakly supervised annotations). The approach proposed in this paper is based on the use of scant annotations to address manual labeling at scale. Active learning and semi-supervised learning are two popular methods in developing segmentation models using scant annotations and are described below.

2.1. Active Learning for Segmentation

In the iterative active learning approach, a limited number of unlabeled images are selected in each iteration for annotation by experts. The annotated images are merged with training data and used to develop the next segmentation model, and the process continues until the model performance plateaus on a given validation set. Active learning approaches can be broadly categorized based on the criteria used to select images for annotation and the unit (images, patches, and pixels) of annotation. For instance, in [21], FCNs are used to identify uncertain images as candidates, and similar candidates are pruned leaving the rest for annotation. In [22], the drop-out method from [23] is used to identify candidates and then discriminatory features of the latent space of the segmentation network are used to obtain a diverse sample. In [24], active learning is modeled as an optimization problem maximizing Fisher information (a sample has higher Fisher information if it generates larger gradients with respect to the model parameters) over samples. In [25], sample selection is modeled as a Boolean knapsack problem, where the objective is to select a sample that maximizes uncertainty while keeping annotation costs below a threshold. The approach in [21] uses 50% of the training data from the MICCAI Gland challenge (85 training, 80 test) and lymph node (37 training, 37 test) datasets; [22] uses 27% of the training data from MR images dataset (25 training, 11 test); [24] uses around 1% of the training data from an MR dataset with 51 images; and [25] uses 50% of the training data from 1,247 CT scans (934 training, 313 test) and 20% annotation cost. Each of these works produces a model with the same performance as those obtained by using the entire training data.

The unit of annotation for most active learning approaches used for segmentation is the whole image. Though the approach in [25] chooses samples with least annotation cost, it requires experts to annotate the whole image. An exception to these are [24,26,27], where 2D patches are used as the unit of annotation. While active learning using pixel-level annotations (as used by SSPA approach) is rare, some recent works show how pixel-level annotations can be cost effective and produce high-performing segmentation models [28]. Pixel-level annotations require experts to be directed to the target pixels along with the surrounding context, and such support is provided by software prototypes, including those such as the PIXELPICK described in [28]. There are several domain-specific auto-annotators exist for medical images and authors have also developed a domain-specific auto-annotator for biofilms that will be released soon to that community.

2.2. Semi-Supervised Segmentation with Pseudo-Labels

Semi-supervised segmentation approaches usually augment manually labeled training data by generating pseudo-labels for the unlabeled data and using these to generate segmentation models. As an exception, the approach in [29] uses K-means along with graph cuts to generate pseudo-labels and use these to train a segmentation model, which is then used to produce refined pseudo-labels, and the process is repeated until the model performance converges. Such approaches do not use any labeled data for training. A more typical approach in [30] first generates a segmentation model by training on a set of scant expert annotations, and the model is then used to assign pseudo-labels to unlabeled training data. The final model is obtained by training it on the expert-labeled data along

with pseudo-labeled data until the performance converges. For a more comprehensive discussion on semi-supervised approaches, please see [10,18].

2.3. Proposed SSPA Approach

The *SSPA* approach seamlessly integrates active learning and semi-supervised learning approaches with pseudo-labels to produce high-performing segmentation models with cost-effective expert annotations. Similar to the semi-supervised approach in [29], the SSPA does not require any expert annotation to produce the base model. It uses an image processing algorithm based on the watershed transform [31] to generate pseudo-labels. The base model generated using these pseudo-labels is then successively refined using active learning. However, unlike the prior active learning approaches used for segmentation, we employ image entropy instead of image similarity to select top-k high entropy or low entropy images for expert annotation. Further, unlike most of the earlier active learning approaches for segmentation (with the exception of [28]), our unit of annotation is a pixel, targeting uncertain pixels only while other pixels are labeled based on the behavior learned by the models (please see Section 3 for more details).

Our preliminary work reported as a short paper in [32], explored the viability of using pseudo-labels in place of expert annotations for semantic segmentation. In that paper, we considered datasets where expert annotations are available for the entire dataset and built a benchmark segmentation model using fully supervised learning. We then compared models built using a mixture of pseudo-labels and expert annotated labels with the benchmark model to show that the viability of pseudo-labels for building segmentation models. Requiring the experts to annotate all of the training data a priori and building a fully supervised segmentation model makes our prior work very different from the proposed approach. Further, having the experts deeply annotate each pixel in each image in all of the training data makes our prior approach impractical for several domains, where significant expertise is needed to annotate each image.

In contrast, in the *SSPA* approach, expert annotations are obtained on demand only for the training samples identified in each active learning step. Further, the unit of annotation is a pixel, and the process is terminated when the model performance plateaus or no further refinements are possibly similar to [29]. The *SSPA* approach outperforms state-of-the-art results in multiple datasets including those used in [32].

The SSPA uses the watershed algorithm to generate pseudo-segmentation masks. This algorithm [31,33–35] treats an image as a topographic surface with its pixel intensities capturing the height of the surface at each point in the image. The image is partitioned into basins and watershed lines by flooding the surface from minima. The watershed lines are drawn to prevent the merging of water from different sources. The variant of watershed algorithm used in this paper, the marker-controlled watershed algorithm (MC-WS) [36], automatically determines the regional minima and achieves better performance than the regular one. MC-WS uses morphological operations [37] and distance transforms [38] of binarized images to identify object markers that are used as regional minima.

In Petit et al. [39], the authors proposed a ConvNets-based strategy to perform segmentation on medical images. They attempted to reduce the annotation effort by using a partial set of noisy labels such as scribbles, bounding boxes, etc. Their approach extracts and eliminates ambiguous pixel labels to avoid the error propagation due to these incorrect and noisy labels. Their architecture consists of two stages. In the first stage, ambiguity maps are produced by using K FCNs that perform binary classification for each of the K classes. Each classifier is given the input of pixels only true positive and true negative to the given class and the rest are ignored. In the second stage, the model trained at the first stage is used to predict labels for missing classes, using a curriculum strategy [40]. The authors stated that only 30% of training data surpassed the baseline trained with complete ground-truth annotations. Even though this approach allows recovering the scores obtained without incorrect/incomplete labels, it relies on the use of a perfectly labeled sub-dataset (100% clean labels). This approach was further extended to an approach called INERRANT [41]

to achieve better confidence estimation for the initial pseudo-label generation, by assigning a dedicated confidence network to maximize the number of correct labels collected during the pseudo-labeling stage.

Pan et al. [42] proposed a label-efficient hybrid supervised framework for medical image segmentation, where the annotation effort is reduced by mixing a large quantity of weakly annotated labels with a handful of strongly annotated data. Mainly two techniques, namely dynamic instance indicator (DII) and dynamic co-regularization (DCR), are used to extract the semantic clues while reducing the error propagation due to strongly annotated labels. Specifically, DII adjusts the weights for weakly annotated instances based on the gradient directions available in strongly annotated instances, and DCR handles the collaborative training and consistency regularization. The authors stated that the proposed framework shows competitive performance only with 10% of strongly annotated labels, compared to the 100% strongly supervised baseline model. Unlike SSPA, their approach assumes the existence of strongly annotated data to begin with. Without using directed expert annotation as done in SSPA, it is highly unlikely that a handful of strongly annotated samples chosen initially will cover all the variations of the data, and hence we argue that the involvement of experts in a directed manner guided by model predictions is important, especially in sensitive segmentation application domains, such as medical and material science.

Zhou et al. [43] recently proposed a watershed transform-based iterative weakly supervised approach for segmentation. This approach first generates weak segmentation annotations through image-level class activation maps, which are then refined by watershed segmentation. Using these weak annotations, a fully supervised model is trained iteratively. However, this approach carries many downsides, such as no control over initial segmentation error propagation in the iterative training, requires many manual parameterization during weak annotation generation, and lack of grasping fuzzy, low-contrast and complex boundaries of the objects [44,45]. Segmentation error propagation through iterations can adversely impact model performance, especially in areas requiring sophisticated domain expertise. In such cases, it may be best to seek expert help in generating segmentation ground truth to manage boundary complexities of the objects and mitigate the error propagation of weakly supervision. Our experiments show that the SSPA models outperform the watershed-based iterative weakly supervised approach.

3. The SSPA Algorithm

The inputs to the SSPA algorithm (Figure 1) are a set of images, $U = \{u_1, \ldots, u_N\}$ and an optional set of corresponding ground truth, $GT = \{g_1, \ldots, g_M\}$, binary labels, where $M \leq N$. The SSPA employs an iterative algorithm that uses a sequence of training sets, $T = (T_1, \ldots, T_k)$ to build a sequence of models, $(M_1, \ldots, M_k)$. Model M_i is generated at the i^{th} iteration using training set T_i. Let L_i be the set of pixel-level binary labels for each training sample in T_i. Each T_i comprises image–label pairs $< t_p, l_p >$, where the t_p refers to a training sample from U, and l_p refers to the corresponding pseudo labels from L_i, distinct from GT. We apply M_i to each training sample in T_i to obtain a set of confidence predictions C_i. For each image t_p in T_i, C_i contains an element e_p with a pair of values, $(conf, label)$ for each pixel in t_p, where $conf$ denotes the prediction confidence value that the pixel belongs to class 1 and $label$ denotes the pixel label assigned by M_i. Min-max normalization of raw $conf$ values of all pixels in e_p is used to normalize them to a $[0...1]$ interval to construct $(\overline{conf}, label)$ pairs for each e_p. We use $\overline{C_i}$ to refer to C_i containing the normalized value, label pairs for each image in T_i. The entropy of a sample image is found using its corresponding normalized prediction confidence values. The mean entropy μ_i of T_i (and of M_i) is calculated as the mean of the entropy of all images in T_i.

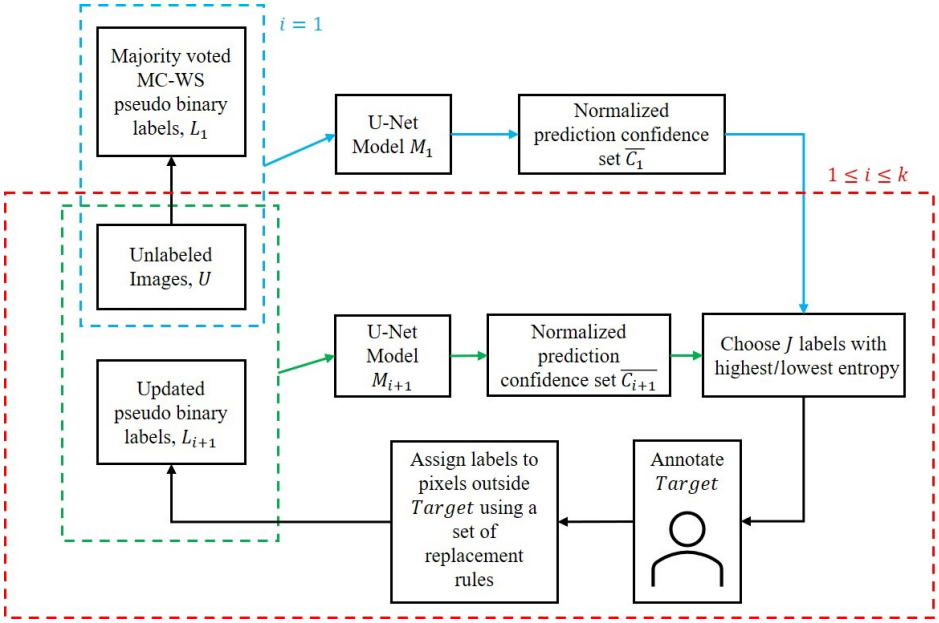

Figure 1. SSPA architecture. In this figure, k represents the termination condition described in Algorithm 1.

The SSPA algorithm can be divided into three main steps as described below.

Initial Model and Initial Pseudo-label Set Generation: The marker controlled watershed (MC-WS) algorithm was employed to avoid the over-segmentation caused due to noise and irregularities that typically occur in the use of the watershed transform. The MC-WS floods the topographic image surface from a predefined set of markers, thereby preventing over-segmentation. To apply MC-WS on each image, an approximate estimate of the foreground objects in the image was first found using binarization. White noise and small holes in the image were removed using morphological opening and closing, respectively. To extract the sure foreground region of the image, distance transform was then used to apply a threshold. Then, to extract the sure background region of the image, dilation was applied on the image. The boundaries of the foreground objects were computed as the difference between the sure foreground and sure background regions. Marker labeling was implemented by labeling all sure regions with positive integers and labeling all unknown (or boundary) regions with a 0. Finally, watershed was applied on the maker image to modify the boundary region to obtain the watershed segmentation mask or binary label of the image.

Using MC-WS, we created an ensemble of three watershed segmentation modules, $\{ws_1, ws_2, ws_3\}$ and applied it to the set U to generate labels, $\{L_{ws_1}, L_{ws_2}, L_{ws_3}\}$. Use majority voting to determine the initial set of pseudo binary labels, L_1. Train a segmentation network on pair $< U, L_1 >$ to obtain initial model M_1 (refer to lines 23–27 in Algorithm 1). We use model M_1 to generate the prediction confidence values C_1 and normalized prediction confidence set $\overline{C_1}$ for U. Let μ_1 be the mean entropy of M_1.

Algorithm 1 Segmentation with scant pixel annotations.

Input:
$U = \{u_1, \ldots, u_N\}$
$GT = \{g_1, \ldots, g_M\}, M \leq N$ ▷ Optional

1: **procedure** SSPA (T_i, M_i, L_i, $\overline{C_i}$)
2: Choose J images with highest entropy (HE)/lowest entropy (LE) from T_i
3: **for each** $t_p \in J$ **do**
4: Let $e_p \in \overline{C_i}(t_p)$
5: $Target = e_p \in (0.5 \pm \delta)$
6: $Target \leftarrow ExpertAnnotation$
7: Let each p_{xy} be a pixel outside *Target* in t_p
8: **if** $\mu_i > \mu_{i-1}$ **then**
9: $L_{i+1}(p_{xy}) \leftarrow L_{i-1}(p_{xy})$
10: **else**
11: $threshold = \mu(\overline{C_i(t_p)})$
12: $(\overline{C_i(p_{xy})} > threshold) \leftarrow 1$
13: $(\overline{C_i(p_{xy})} < threshold) \leftarrow 0$
14: **end if**
15: **end for**
16: $L_{i+1} =$ Replace J in L_i
17: **return** $M_{i+1}, \overline{C_{i+1}}$
18: **end procedure**

19: **procedure** WATERSHED(U)
20: $WS = \{L_{ws_1}, L_{ws_2}, L_{ws_3}\}$ ensemble
21: $L_1 \leftarrow MajorityVoting(L_{ws_1}, L_{ws_2}, L_{ws_3})$
22: **return** $M_1, \overline{C_1}$
23: **end procedure**

24: **repeat**
25: SSPA ($T_i, M_i, L_i, \overline{C_i}$)
26: **until** $\mu_{i+1} > \mu_i$

Segmentation with Scant Pixel Annotations: Let $i \geq 1$. The pseudo label set L_i, the corresponding model M_i and the normalized prediction confidence values $\overline{C_i}$ generate the pseudo binary label set L_{i+1} from M_i as follows. First, choose J images from T_i with highest entropy (HE) or the lowest entropy (LE) values. Let t_p be one such image chosen and $L_i(t_p)$ be its label in L_i. We construct the training label L_{i+1} for t_p as follows. Consider all pixels in t_p with prediction confidence values between $(0.5 \pm \delta)$ in $e_p \in \overline{C_i}$ (pixels whose predictions from model M_i are in the uncertainty range), for expert annotation (lines 8–9). The value of the parameter δ is assigned empirically for each dataset. Let *Target* be the set of all pixels in t_p that are marked for expert annotation. One way to obtain expert annotation for pixel labels is to manually label each pixel in *Target*. If GT is available, we copy the pixel label for each pixel from the ground-truth label of t_p into $L_{i+1}(t_p)$.

Now consider the pixels that are not in *Target*. The pixel-level labels for these pixels in t_p can be decided using either the previous model M_{i-1} or the current model M_i (lines 11–17). Let p_{xy} be a pixel in x^{th} row and y^{th} column of t_p. If $\mu_i > \mu_{i-1}$, then the label for p_{xy} in $L_{i+1}(t_p)$ is the same as that in $L_{i-1}(t_p)$. Else, label for p_{xy} in $L_{i+1}(t_p)$ is the same as assigning a class 0 or 1 to $\overline{C_i(t_p)}$ based on a *threshold*. *Threshold* is calculated as the mean prediction confidence value of t_p. Generate the next set of training labels, L_{i+1} by replacing J labels in L_i (line 16). Train a segmentation network on pair $< U, L_{i+1} >$ to obtain next model M_{i+1}.

Termination condition: At each iteration i, record the mean entropy of M_i, μ_i. The algorithm terminates when the mean entropy of M_{i+1}, μ_{i+1} is higher than the mean entropy of M_i (lines 29–31). The decrease in model performance indicates that the model is unlearning useful patterns or features during training at the $(i+1)$th iteration. In the presence of GT labels, we also record evaluation metrics such as intersection over union (IoU) and Dice score. However, mean entropy as an evaluation metric takes precedence over IoU and Dice score, even in the presence of GT labels. Refer to Section 4.4 for a detailed discussion of evaluation metrics. Select M_i with the best evaluation metrics as the best model to obtain binary labels, L_i using the least expert intervention.

Note that the SSPA uses two parameters—the uncertainty range threshold δ and the number of images J selected for expert annotation in each iteration. Model prediction values around 0.5 lead to most uncertainty and a range around this value determined by δ is well suited for many datasets. The value of J can be set based on the improvement of model performance across iterations. The approach can be adapted to other datasets by setting these parameters appropriately.

4. Evaluation of the SSPA Approach

In this section, we describe the study we conducted for understanding the effectiveness of the SSPA approach. We describe the datasets used in the study, the CNN architecture used to build semantic segmentation models, the hardware setup used for running the experiments, and the evaluation metrics used to measure the performance of the models.

4.1. Datasets

Three datasets, electron microscope (EM) and melanoma datasets from the bio-medical domain along with a dataset from the biofilm domain, were used to study the effectiveness of the SSPA approach. The ground-truth labels GT were available for all three datasets. However, note that the GT labels are not required and were used only for evaluation.

4.1.1. EM Dataset

The electron microscope (EM) dataset (Figure 2A,B) is a set of 30 grayscale images from a serial section transmission electron microscopy dataset of the Drosophila first instar larva ventral nerve cord [46]. This dataset was collected to compare and rank different approaches for the automatic segmentation of neural structures to that performed by an expert neuro-anatomist. It was published as a part of the *IEEE ISBI 2012 challenge on 2D segmentation* to determine the boundary map (or binary label) of each grayscale image, where "1" or white indicates a pixel inside a cell, and "0" indicates a pixel at the boundary between cross-sections. A binary label was considered equivalent to a segmentation of the image. The ground-truth binary labels for the training images were provided as part of the challenge.

Figure 2. Sample of an unlabeled image and GT label of the (**A**,**B**) EM dataset (**C**,**D**) melanoma dataset and (**E**,**F**) biofilm dataset.

4.1.2. Melanoma Dataset

The melanoma dataset (Figure 2C,D) contains 43 color images of malignant melanomas obtained with consumer-level cameras from the Dermatology Information System [47]. The dataset is a part of a skin cancer detection project that used dermatological photographs taken with standard consumer-grade cameras to analyze and assess the risk of melanoma

in patients. The dataset was used to develop methods for the automated diagnosis of melanoma from dermoscopic images. Each image containing a single lesion of interest was manually segmented to create a binary label for differentiating pixels of the lesion from those of surrounding skin. A "1" or white pixels indicated a pixel inside a lesion, and "0" or black indicated a pixel of the surrounding skin.

4.1.3. Biofilm Dataset

The biofilm dataset (Figure 2E,F) consists of scanning electron microscope (SEM) images of *Desulfovibrio alaskensis* G20 (DA-G20, a sulfate reducing bacteria (SRB), and their biofilm grown on bare mild steel that was used as a working electrode (working specimen) in microbiologically influenced corrosion (MIC) experiments [48]. The details of the growth procedures and biocorrosion tests were discussed in [49]. Owing to its high ductility, weldability, and low cost, mild steel remains a popular choice of metal in civil infrastructure, transportation and oil and gas industry applications, and routine applications. However, under aqueous conditions, mild steel is susceptible to microbially infused corrosion caused by microorganisms, including SRB [50]. The goal of segmentation of the biofilm dataset is to identify the shape and size of each bacterial cell or a cluster of cells to detect metal corrosion. A "1" or white indicates a pixel in a bacterial cell, and "0" indicates a pixel at the boundary between bacteria.

4.2. Network Architecture

The U-Net [20] is a convolutional network architecture for pixel-based image segmentation. It consists of an encoder (contracting path) and decoder (expansive path) designed specifically to perform segmentation tasks on medical images. The contracting path is a stack of convolutional and max-pooling layers that encodes high-level semantic information at each layer into feature representations at different levels. The decoder projects the discriminative features learned by the encoder onto the pixel space by recovering spatial information at each layer using transposed convolutions. Bottleneck layers combine high resolution features from the encoder and upsampled features from the decoder by concatenation, resulting in a symmetrical network in contrast to traditional FCNs. The U-Net architecture accepts a set of pairs—unlabeled images and their corresponding binary masks (labels)—as its input for training a segmentation model. The unlabeled images and their masks are both 512×512 pixels in height and width.

4.3. Experimental Setup

Training for the U-Net models was implemented using Keras with a Tensorflow backend as the deep learning framework on an Ubuntu workstation with 12-Core Intel iO-9920x and 128GB RAM. We randomly selected 20% of the dataset as the validation set and the remaining as the training set during each fold. We used a learning rate of 0.1 and compiled the models using Adam optimizer [51] for 25 epochs, using the binary cross-entropy loss function. An early-stop mechanism was used to prevent over-fitting.

4.4. Evaluation Metrics

When an image is classified using a semantic segmentation model, each pixel in the image is assigned two values (model confidence, pixel label). Recall from Section 3 that the min-max normalization of the raw model confidence prediction values of all pixels in each image is performed to normalize them to $[0...1]$. We use the confidence prediction values (p_i) to compute the entropy of an image. A low entropy value indicates low uncertainty about the pixel labels assigned by the model. Entropy S is given by Equation (1) as follows:

$$S = -\sum_{i=0}^{n-1} p_i(log_b p_i) \tag{1}$$

Following the standard practice [7,52], intersection over union (IoU) and Dice similarity score are used to evaluate the segmentation performance. As given in Equation (2), IoU computes the area of overlapping between the predicted model and the GT label divided by the area of union between the predicted label and the GT label. Similarly, the Dice score is calculated as twice the overlap between the predicted label and the GT label, divided by the sum of the number of pixels in both labels as given in Equation (3). Both the IoU and Dice score are positively correlated. However, the IoU metric tends to penalize single instances of incorrect classification more than the Dice score.

$$IoU = \frac{TP}{TP + FP + FN} \tag{2}$$

$$Dice = \frac{2TP}{2TP + FP + FN} \tag{3}$$

5. Experimental Results and Discussion

For each dataset, the *initial model and initial pseudo-label set generation* step was applied to obtain the first set of pseudo labels L_1 for each image in the dataset. Model M_1 was constructed using a training set and the label set L_1. Models $M_2, \ldots, M_k$ were constructed iteratively by following the segmentation with scant pixel annotation step using two different values of J and both high entropy and low entropy pixel label replacement strategies. The SSPA approach was terminated when the mean entropy of a model constructed in an iterative step increases from the previous step. Since we have access to ground-truth labels for all datasets, we used it to construct the ground-truth model M_0 and to benchmark the evaluation results from the models built using the SSPA approach. In addition to studying the prediction accuracy of the models constructed from the SSPA approach, we also observed the behavior of the SSPA approach's pixel annotation strategies using heatmaps and confidence values of the pixels labels assigned by the models $M_1, \ldots, M_k$.

We now discuss the results of applying the SSPA approach to each of the three datasets. Below, we use HE (LE) to denote the strategy of choosing J images from the training set T_i with the HE (LE) values and then selectively replace the uncertain pixel labels identified by the SSPA approach in the training set T_{i+1}. For each data set, we considered $J = 1$, the minimum value as well as J values corresponding to 10% of the training data. Models obtained using $J = 1$ values under-performed in all cases and are discussed in the paper. We also calculated the percentage of pixels replaced as the ratio of the total number of pixels labels replaced over all the J images to the total number of the pixels in the training set.

5.1. EM Dataset

For this dataset, experiments were conducted using $J = 3$ (10% of the training data) in each iteration. Models obtained using LE pixel label replacements outperformed others.

Pixel label replacements in HE images. Our results in Figure 3 indicate that the performance of models with HE pixel label replacements did not improve, despite increased expert annotation efforts. In Figure 3, for each model and J value combination on the x-axis, three types of information are depicted—the IoU (black bar), Dice scores (blue bar), and the percentage of pixels replaced (the red trend-line). Models M_2 and M_3 had the same Dice (0.874) and IoU (0.776) scores. Model M_2 was obtained by replacing 1.18% of the pixel labels from the 3 images in the output of model M_1. Model M_3 was obtained by replacing 2.46% of the pixel labels from the 3 images in the output of model M_2. Since the mean entropy of M_3 (2.976) was higher than that of M_2 (2.845), the algorithm terminated. The models obtained using HE pixel label replacements achieved similar IoU values but had lower Dice scores in comparison to the benchmark model M_0, which had mean entropy of 1.986, IoU of 0.823, and Dice score of 0.903.

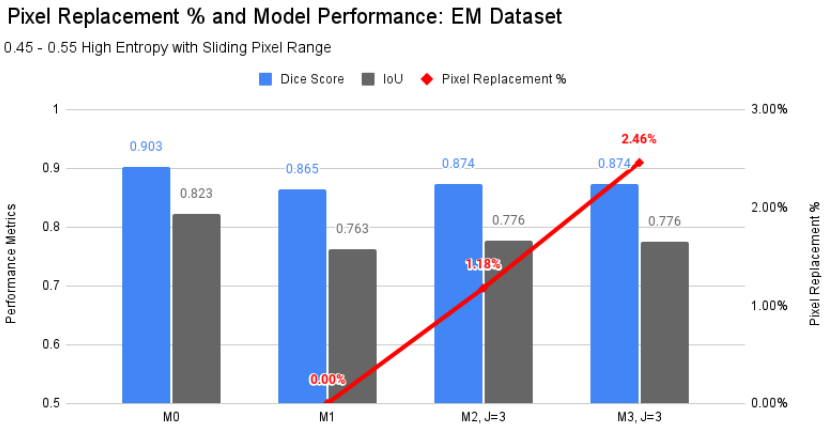

Figure 3. Performance on the EM dataset using the SSPA approach with HE pixel label replacements.

Pixel label replacements in LE images. Figure 4 shows the results obtained using LE pixel label replacements. For this experiment, we randomly chose one image as a test image and trained M_1 using the remaining 29 images. Next, we generated models M_2 and M_3 using $J = 3$. Model M_2 was obtained by replacing 0.86% of the pixel labels from the 3 LE images in the output of model M_1. Model M_3 was obtained by replacing 1.69% of the pixel labels from the 3 LE images in the output of model M_2. Since the mean entropy of M_3 (2.447) was higher than that of M_2 (2.441), the algorithm terminated. Model M_2 with IoU value 0.818 and Dice score 0.9 performs comparably with M_0, having mean entropy 1.953, IoU 0.82, and Dice score of 0.9. The M_0 values for LE slightly differ from those for HE since they are computed using 29 instead of 30 images. We also studied the entropy distribution of models generated using the LE pixel label replacement strategy. The entropy distribution of M_0 had high variability, while M_2 had the least variability and the best performance.

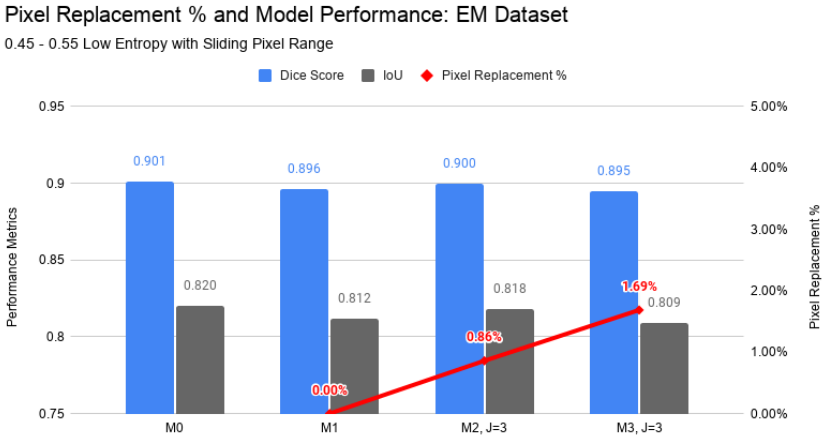

Figure 4. Performance on the EM dataset using the SSPA approach with LE pixel label replacements.

We studied the model prediction entropy distribution of models generated using LE pixel label replacement strategy. The results are displayed in Figure 5. Here, the x-axis plots the model and the y-axis plots the entropy values. Each box plot in the figure shows the entropy value distribution of images for each model. As can be seen from the figure, the entropy distribution of M_0 had high variability, while M_2 had the least variability and the best performance. Although the median entropy values across models M_1, M_2, and M_3 were higher than the upper quartile of M_0, the SSPA approach seemed to have reduced the variability in the entropy values to obtain better performance. Further, we studied pixel oscillation to understand the effectiveness of the pixel replacement using the SSPA approach.

We define oscillating pixels as the pixels with normalized prediction confidence values in the target range [0.45,0.55] ($\delta = 0.05$), which result in inverse prediction confidence values after being replaced by GT labels in the model input. Oscillating pixels can be problematic since they represent the unlearning of useful patterns in the input. We observed that 50.32% of the 0.86% replaced pixels in M_2 oscillated, whereas 58.07% of the 1.69% replaced pixels in M_3 oscillated. We conjecture that there is a correlation that smaller number of oscillating pixels lead to better performance in models.

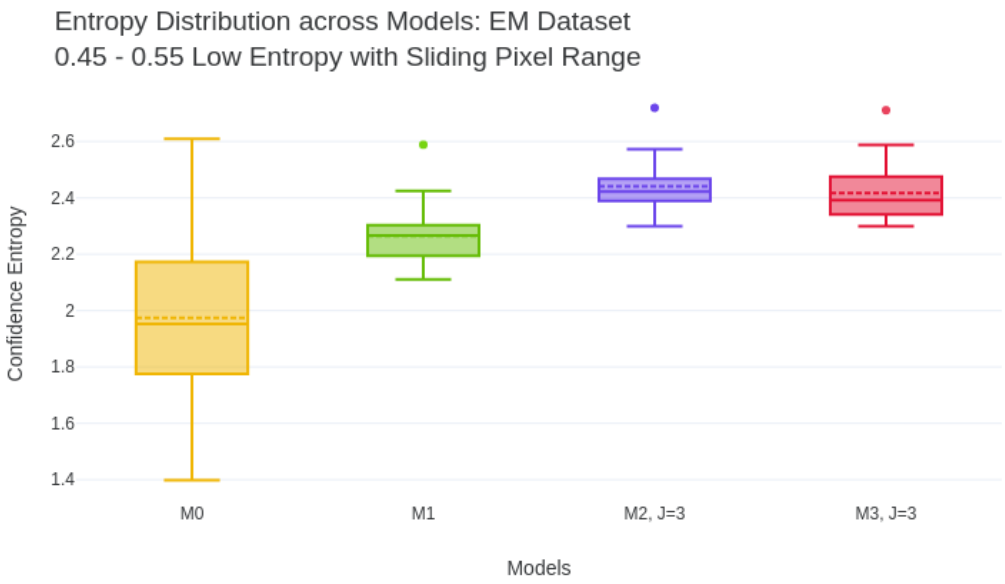

Figure 5. Entropy distribution across models.

5.2. Melanoma Dataset

For this dataset, experiments were conducted using $J = 5$ (10% of the training data) in each iteration. Two test images were randomly chosen for evaluation from the dataset and rest of the data were used to generate three successive models using the SSPA approach. Models obtained using HE pixel label replacements outperformed others.

Pixel label replacements for HE images. Figure 6 shows the results obtained using the HE pixel label replacements. As depicted in the figure, model M_2 was obtained by replacing only 0.8% of the pixel labels from the 5 HE images in the output of model M_1. Model M_3 was obtained by replacing 4.05% of the pixel labels from the 5 HE images in the output of model M_2. Since the mean entropy of M_3 (1.689) was higher than that of M_2 (1.432), the algorithm terminated. Model M_2 with IoU value 0.824 and Dice Score 0.973 outperformed the benchmark model M_0, having IoU value 0.764, and Dice score of 0.962. The mean entropy of M_0, 2.232, was higher than that of M_2.

To visually track and assess the change in entropy induced by pixel label replacements, we constructed a heatmap for each image in the training set using the normalized prediction confidence values. The heatmap of a sample image from Figure 7A demonstrated that the confidence predictions of M_2 were consistent with the GT labels shown in Figure 2. A more detailed view of the pixels (in red) to be annotated by experts can be seen in Figure 7B,C, at two different scales. The target regions in the range [0.45,0.55] ($\delta = 0.05$) occurring mostly around the boundaries of the lesion in Figure 7B,C, showed the highest uncertainty.

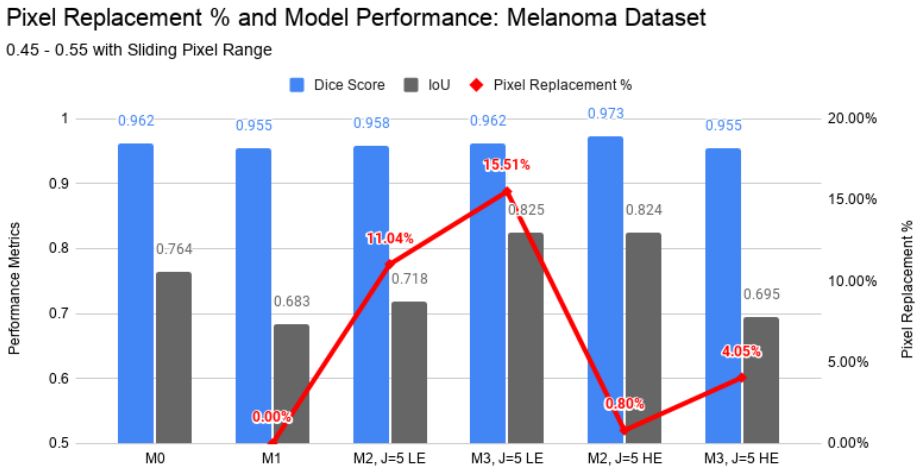

Figure 6. Performance on the melanoma dataset using the SSPA approach with LE and HE pixel label replacements.

(A) (B) (C)

Figure 7. (**A**) Heatmap of confidence entropy values, (**B**) view of pseudo label before replacement with pixels (in red) to be annotated by expert, (**C**) enlarged view of (**B**).

Pixel label replacements for LE images. On the other hand, M_2 with LE pixel label replacements (also depicted in Figure 6) resulted in a mean entropy 2.064, Dice score 0.958, and IoU 0.718, comparable to those of M_0. However, M_2 was generated using 11.04% pixel label replacements. Recall that model M_2 was generated using only 0.8% pixel label replacements and had a much lower mean entropy value of 1.432 in the HE case. Since the mean entropy of M_3, 2.669, was higher than that of M_2, the SSPA, approach terminated. The performance of M_3 with LE pixel label replacements was comparable to that M_3 with HE pixel label replacements but required 15.51% pixels to be replaced in comparison to 4.05%.

5.3. Biofilm Dataset

For this dataset, experiments were conducted using $J = 8$ (10% of the training data) in each iteration. Three test images were randomly chosen for evaluation from the dataset, and rest of the data were used to generate three successive models using the SSPA approach. Models obtained using HE pixel label replacements outperformed others.

Pixel label replacements for HE images. Figure 8 shows the results obtained using HE pixel label replacements. As shown in the figure, model M_2 was obtained by replacing only 0.85% of the pixel labels from the 8 HE images in the output of model M_1. Model M_3 was obtained by replacing 3.73% of the pixel labels from the 8 HE images in the output of model M_2. The performance of M_1 and M_2 was similar with IoU values around 0.691 and Dice scores around 0.815. The mean entropy of M_2 decreased to 1.778 from 2.587 in M_1. The mean entropy of M_3 increased to 1.991, and the algorithm terminated. The benchmark model M_0 had mean entropy 2.822, IoU 0.609, and Dice score 0.754.

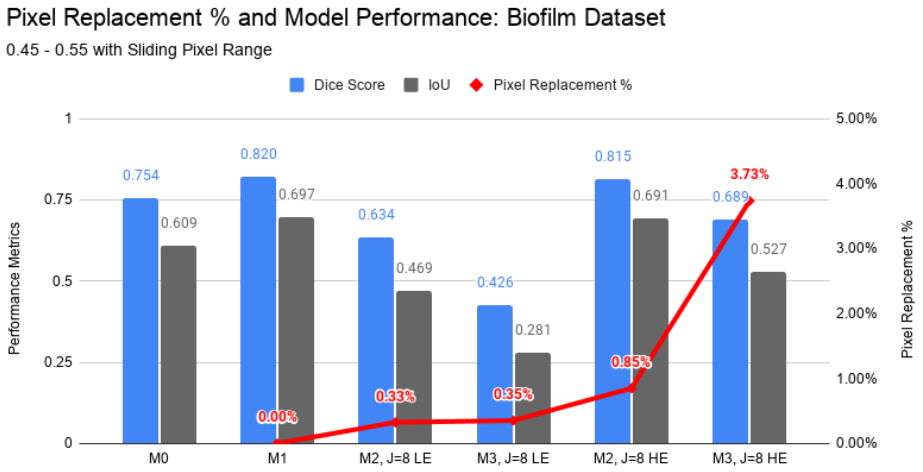

Figure 8. Performance on the biofilm dataset using the SSPA approach with HE pixel label replacements.

The heatmap in Figure 9A shows that the target regions for pixel label replacements of M_2 were found within the bacterial cells, contrary to the melanoma datasets where the boundaries of objects showed the highest uncertainty. The uncertainty of the model within the bacterial cells is likely due to the unique nature of having to segment the biofilm dataset. Similar to the EM dataset, the goal of segmenting the biofilm dataset was to determine the boundary map of the bacterial cells. Figure 9B shows a more explicit view of the pixels (in red) to be annotated by the experts.

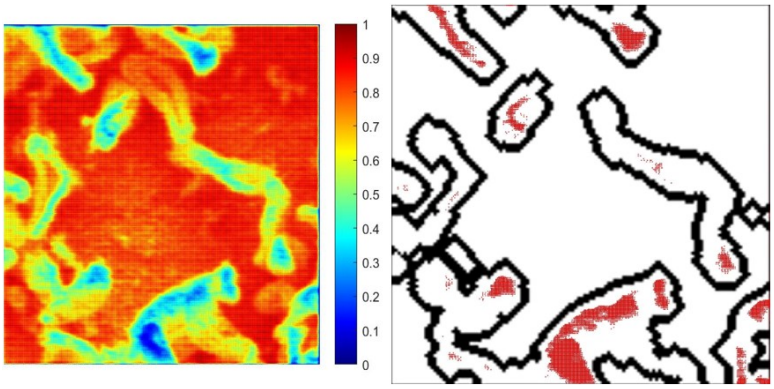

Figure 9. (**A**) Heatmap of confidence entropy values, (**B**) view of the pseudo label before replacement with pixels (in red) to be annotated by expert.

Figure 10 illustrates the entropy distribution of HE pixel label replacement models with the y–axis representing the distribution of entropy values for each model in the x–axis. All models display a normal distribution with a few outliers. The best model, M_2 had the lowest median, lower than the lower quartile of M_0 and M_1. The entropy distribution of both M_2 and M_3 showed decreasing variability, which illustrates the positive effect of pixel level replacements on model variability. We also observed that M_3 had a higher rate of oscillation than M_2, further validating the correlation between improved performance and lower oscillation.

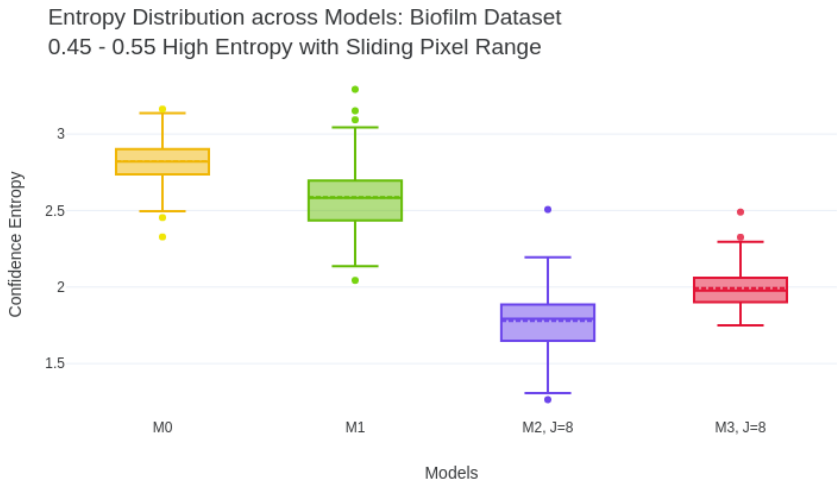

Figure 10. Entropy distribution across models.

Pixel label replacements for LE images. On the other hand, model M_2 with LE pixel label replacements (also depicted in Figure 8) recorded a mean entropy 1.802, IoU 0.469 and Dice score 0.634. M_2 was generated using 0.33% pixel label replacements. The pixel replacements required by M_3 was 0.35%. Since the mean entropy of M_3, 2.541 was higher than that of M_2, the algorithm terminated here.

5.4. Comparing SSPA with Other Methods

We also investigated the effectiveness of the SSPA approach by comparing its performance with the state-of-the-art fully supervised and weakly supervised segmentation methods. We trained two fully supervised encoder–decoder architectures, one using U-Net and another using DeepLabV3+ with Resnet101 [53,54]. The DeepLabV3+ model has an encoding phase which uses atrous spatial pyramid pooling (ASPP) and a decoding phase to give a better segmentation results along object boundaries. To compare our method with weakly supervised segmentation methods, we trained two U-Net models using grab-cut [55] and MC-WS methods, respectively, both of which generate initial pseudo-labels.

The segmentation results of SSPA and other fully supervised (FS) and weakly supervised (WS) methods are summarized in Table 1. FS models were trained using full pixel level (P) expert labels, whereas WS models were trained using complete image-level (I) labels. The SSPA approach performs approximately equally or better on all datasets in comparison to these fully supervised and weakly supervised methods. This is despite the minimal annotation effort needed for the SSPA in comparison to the other methods.

For the EM dataset, SSPA+LE performs equally to the two fully supervised U-Net methods (Dice scores for both are around 90.0% and IOU scores are around 82%). For this dataset, the performance of the SSPA+HE is lower in comparison to these two FS methods. The SSPA+HE and SSPA+LE perform better than both the weakly supervised methods in both the Dice and IOU measures. For the melanoma dataset, the SSPA+HE outperforms all of the FS and WS with respect to both IOU (82.4%) and Dice score (97.3%) values. For the biofilm dataset, the SSPA+HE outperforms both the FS methods as well as WS: grab-cut+UNET method. The performance of SSPA+HE and WS:MC-WS+U-NET is approximately the same. Further, we observed that SSPA+LE shows better IOU and Dice-score results on EM dataset compared to SSPA+HE version, whereas the SSPA+HE version performs well with the melanoma and biofilm datasets. Further, the performance improvement of the SSPA method on the biofilm dataset is approximately 9% compared to the supervised approaches.

The best performing models were obtained by replacing less than 5% of the pixels summed across all iterations for all datasets (2.55% for EM with LE, 4.85% for melanoma

with HE, and 4.58% for biofilms with HE). These percentages include the pixels to be replaced to generate one more model after the best performing model in order for the active learning process to terminate.

Table 1. Evaluation of the effect of SSPA approach on segmentation quality using both fully supervised and weakly supervised approaches. Best result is shown in bold. For IOU and Dice score metrics, the higher the number, the better (↑). *I* represents the image-level labels and *P* represents pixel-level labels; percentages are cumulative expert labels across all the iterations of active learning.

Method	Expert Labels	Dataset [IOU / Dice] (%) ↑		
		EM	***Melanoma***	***Biofilm***
FS: U-Net	*P* (100%)	**82.3 / 90.3**	76.4 / 96.2	60.9 / 75.4
FS: DeepLabV3+	*P* (100%)	81.6 / 89.1	76.0 / 86.2	61.6 / 74.4
WS: Grab-Cut + U-Net	*I* (100%)	76.1 / 85.2	65.5 / 92.7	56.4 / 62.0
WS: MC-WS + U-Net	*I* (100%)	76.3 / 86.5	68.3 / 95.5	**69.7 / 82.0**
SSPA + LE	*P* (<26%)	81.8 / 90.0	71.8 / 95.8	46.9 / 63.4
SSPA + HE	*P* (<5%)	77.6 / 87.4	**82.4 / 97.3**	69.1 / 81.5

5.5. Discussion

The SSPA is a novel approach for generating high-performing deep learning models for image semantic segmentation using scant expert image annotations. Automated methods generating pseudo-labels are integrated with an iterative active learning approach to selectively perform manual annotation and improve model performance. We used an ensemble of MC-WS segmentation modules to generate pseudo-labels. We also considered other popular choices, such as grab-cut [55] to generate pseudo-labels and chose MC-WS based on its relative superior performance. Pseudo-labeling approaches other than MC-WS may perform better for other applications, and these can be easily incorporated into the SSPA approach. Note that using a method that generates high-quality pseudo-labels is beneficial to the SSPA, but it is not essential to its success. In the SSPA approach, the pixel replacement effort required by the expert is inversely proportional to the initial pseudo-label quality. In the worst-case scenario, a low-quality initial pseudo-label set has to be compensated by the extra labeling effort from the experts. In the SSPA, images that need expert attention are chosen based on their model prediction entropy values. We employ entropy as the uncertainty sampling measure for the active learning process, over marginal, ratio, and least confidence sampling techniques. Entropy-based sampling is well known and has been shown to be well suited for selecting candidates for classification and segmentation tasks [56]. In the SSPA, we compute entropy value for each image and use these values to identify the top-k images whose certain pixels have to be manually annotated by experts. A high entropy (HE) value for an image indicates an image where most pixel predictions are uncertain (probability in the range $0.5 \pm \delta$) in that image. If an image with HE value is selected as one of the top-k images for annotation by experts, then pixels with prediction values around 0.5 are labeled by the experts in order to reduce the uncertainty of predictions.

Alternatively, a low entropy (LE) value for an image indicates that most of the pixel predictions are made with high confidence. If an image with LE entropy value is selected as one of the top-k images for annotation by experts, then this means that there are sufficient pixels with uncertain predictions (probability in the range $0.5 \pm \delta$) in that image, and these need to be labeled by experts to improve the performance of the model. Table 1 illustrates the experiments conducted on both high entropy and low entropy and the best-performed strategy (HE or LE) for each dataset. The uncertainty range threshold δ is one of the two parameters to the SSPA that was empirically determined to be 0.05 for our experiments. The

parameter value may be varied based on different datasets based on expert assessments of model predictions.

From the above experimental results, we can conclude that the best model constructed from the SSPA approach achieved high prediction accuracy with a mix of over 94% pseudo pixel labels generated from the MC-WS ensemble that were iteratively improved using select expert annotations. The terminating condition we employed also worked well in practice by stopping the constructing of new models when mean entropy increases with increased expert annotations. The additional methods—heatmaps and oscillating pixels—were valuable in understanding the behavior of the SSPA approach. They provided insights on which pixels are hard for a model to learn and how the scant annotations provided in each iteration contributed to the mean entropy of model outputs and the accuracy of the models. From these methods, we observed that the SSPA method may not always assign the same label as the expert to a pixel consistently. Therefore, in the final model, certain pixels may be assigned incorrect labels, though they were assigned correct labels in earlier models.

The SSPA is a general purpose segmentation method that should be applicable to several datasets. The segmentation performance of the SSPA method evaluated through IOU and Dice scores does not depend on the percentage of the pixels to be relabeled. The percentage of pixels to relabeled is related to the manual labeling effort. No specific threshold values are used to identify images with HE and LE values. Top-k HE (or LE) images are chosen for annotation. Similarly, pixels with most uncertain predictions (probability value $0.5 \pm \delta$) are examined by the experts and labeled. Two parameters that need to be chosen in order to apply the SSPA are (1) the number of images to be analyzed in each iteration (the value j), and (2) the uncertainty range delta for pixels. For each dataset, experiments can be run based on on both LE and HE values, and the resulting models can be compared and chosen.

6. Conclusions and Future Work

The SSPA, a novel approach for generating high-performing deep learning models for semantic segmentation, was presented. The SSPA approach seamlessly combines psuedo-segmentation masks are automatically generated using image processing methods with active learning to generate a sequence of segmentation models using minimal manual annotations by experts. The segmentation model output by the SSPA approach is shown to achieve high-quality segmentation while using a fragment of expert annotations in comparison to supervised learning methods. An ensemble of marker-based watershed segmentation (MC-WS) algorithm modules were used to first generate pseudo-segmentation masks because of their superior performance, which were then iteratively refined by experts to generate a series of segmentation models. In each iteration, the model prediction entropy values of images were used to select top-k high entropy (low entropy) images to be given to experts for annotation. The experts were directed to pixels with the most model prediction uncertainty to manually assign labels and use them with other images to generate the next model. The segmentation model generated when no more improvements are feasible is finally output by the SSPA.

Our experiments verify that our approach achieves superior results on all the datasets considered. Using the SSPA approach with pixel level replacements, we recorded a Dice score of 0.9 and IoU of 0.818 on the EM dataset with only 0.86% expert annotation. We recorded a Dice score of 0.973 and an IoU of 0.824 on the melanoma dataset with just 0.8% expert annotation. We also obtained a Dice score of 0.815 and an IoU of 0.691 with 0.85% expert annotation using only 50% of the original biofilm dataset from [32]. The SSPA approach was effective in determining the boundaries between cross sections or bacterial cells in the EM and biofilm datasets with scant annotations. The approach was also effective in determining the entire lesion of interest in the melanoma datasets with scant annotations. The SSPA approach is a general purpose segmentation approach, and our results suggest that the SSPA approach can be effective for both boundary and semantic

segmentation tasks across different applications. The SSPA approach is parameterized based on the uncertainty range of pixels that are targets of directed expert annotations and the number of images that are selected in each iteration, which can be empirically determined for the application at hand. In addition, the SSPA network architecture is modular and can be easily adapted to incorporate a variety of psuedo-segmentation mask generation algorithms and for selecting images in each iteration of active learning. The manual effort involved in directed pixel-level annotations can be further alleviated by the design of a software workbench. For this purpose, in the future we plan to design and implement a workbench to improve the pixel replacement task by the expert and provide a convenient interface as well.

Author Contributions: Conceptualization, A.D.C., M.S. and P.C.; Methodology, A.D.C., M.S. and P.C.; Funding acquisition, P.C. and V.G.; Formal analysis, M.S.; Investigation, A.D.C. and D.A.; Project administration and supervision, M.S. and P.C.; Software, A.D.C.; Writing—original draft, A.D.C. and M.S.; Writing—review and editing, M.S., P.C. and D.A. All authors have read and agreed to the published version of the manuscript.

Funding: D. Abeyrathna, P. Chundi, M. Subramaniam are partially supported by NSF EPSCoR RII Track 2 FEC #1920954. V. R. Gadhamshetty acknowledges funding support from NSF RII Track-1 award (1849206) and NSF CAREER Award (1454102).

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: EM dataset: Allen AM, Neville MC, Birtles S, Croset V, Treiber C, Waddell S, Goodwin SF. 2019. https://www.overleaf.com/project/62ba61f82518660b1708ff3f, accessed on 23 May 2022. A single-cell transcriptomic atlas of the adult Drosophila ventral nerve cord. NCBI Gene Expression Omnibus. GSE141807. Melanoma dataset: Tschandl P. 2018. Harvard Dataverse. https://doi.org/10.7910/DVN/DBW86T, accessed on 23 May 2022.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Farabet, C.; Couprie, C.; Najman, L.; Lecun, Y. Learning hierarchical features for scene labeling. *IEEE Trans. Pattern Anal. Mach. Intell.* **2013**, *35*, 1915–1929. [CrossRef] [PubMed]
2. Hariharan, B.; Arbeláez, P.; Girshick, R.; Malik, J. Simultaneous detection and segmentation. In Proceedings of the European Conference on Computer Vision, Zurich, Switzerland, 6–12 September 2014; Springer: Berlin/Heidelberg, Germany, 2014; pp. 297–312.
3. Chen, L.C.; Papandreou, G.; Kokkinos, I.; Murphy, K.; Yuille, A.L. Deeplab: Semantic image segmentation with deep convolutional nets, atrous convolution, and fully connected crfs. *IEEE Trans. Pattern Anal. Mach. Intell.* **2017**, *40*, 834–848. [CrossRef] [PubMed]
4. Long, J.; Shelhamer, E.; Darrell, T. Fully convolutional networks for semantic segmentation. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Boston, MA, USA, 7–12 June 2015; pp. 3431–3440.
5. Dai, J.; He, K.; Sun, J. Boxsup: Exploiting bounding boxes to supervise convolutional networks for semantic segmentation. In Proceedings of the IEEE International Conference on Computer Vision, Santiago, Chile, 7–13 December 2015; pp. 1635–1643.
6. Zhu, H.; Meng, F.; Cai, J.; Lu, S. Beyond pixels: A comprehensive survey from bottom-up to semantic image segmentation and cosegmentation. *J. Vis. Commun. Image Represent.* **2016**, *34*, 12–27. [CrossRef]
7. Garcia-Garcia, A.; Orts-Escolano, S.; Oprea, S.; Villena-Martinez, V.; Martinez-Gonzalez, P.; Garcia-Rodriguez, J. A Survey on Deep Learning Techniques for Image and Video Semantic Segmentation; *Appl. Soft Comput.* **2018**. *70*, 1568–4946. [CrossRef]
8. Zhao, B.; Feng, J.; Wu, X.; Yan, S. A survey on deep learning-based fine-grained object classification and semantic segmentation. *Int. J. Autom. Comput.* **2017**, *14*, 119–135. [CrossRef]
9. Thoma, M. A survey of semantic segmentation. *arXiv* **2016**, arXiv:1602.06541.
10. Lateef, F.; Ruichek, Y. Survey on semantic segmentation using deep learning techniques. *Neurocomputing* **2019**, *338*, 321–348. [CrossRef]
11. Sehar, U.; Naseem, M.L. How deep learning is empowering semantic segmentation. *Multimed. Tools Appl.* **2022**, 1573–7721. [CrossRef]
12. Chakravarthy, A.D.; Bonthu, S.; Chen, Z.; Zhu, Q. Predictive Models with Resampling: A Comparative Study of Machine Learning Algorithms and their Performances on Handling Imbalanced Datasets. In Proceedings of the 2019 18th IEEE International Conference on Machine Learning and Applications (ICMLA), Boca Raton, FL, USA, 16–19 December 2019; pp. 1492–1495. [CrossRef]

13. Abeyrathna, D.; Subramaniam, M.; Chundi, P.; Hasanreisoglu, M.; Halim, M.S.; Ozdal, P.C.; Nguyen, Q. Directed Fine Tuning Using Feature Clustering for Instance Segmentation of Toxoplasmosis Fundus Images. In Proceedings of the 2020 IEEE 20th International Conference on Bioinformatics and Bioengineering (BIBE), Cincinnati, OH, USA, 26–28 October 2020; pp. 767–772. [CrossRef]
14. Halim, S.M. (Byers Eye Institute at Stanford University, Palo Alto, CA, USA). Personal communication, 2020.
15. Abeyrathna, D.; Life, T.; Rauniyar, S.; Ragi, S.; Sani, R.; Chundi, P. Segmentation of Bacterial Cells in Biofilms Using an Overlapped Ellipse Fitting Technique. In Proceedings of the 2021 IEEE International Conference on Bioinformatics and Biomedicine (BIBM), Houston, TX, USA, 9–12 December 2021; pp. 3548–3554. [CrossRef]
16. Bommanapally, V.; Ashaduzzman, M.; Malshe, M.; Chundi, P.; Subramaniam, M. Self-supervised Learning Approach to Detect Corrosion Products in Biofilm images. In Proceedings of the 2021 IEEE International Conference on Bioinformatics and Biomedicine (BIBM), Houston, TX, USA, 9–12 December 2021; pp. 3555–3561. [CrossRef]
17. Kalimuthu, J. (Civil and Environmental Engineering Department, South Dakota School of Mines Technology, Rapid City, SD, USA). Personal Communication, 2022.
18. Tajbakhsh, N.; Jeyaseelan, L.; Li, Q.; Chiang, J.N.; Wu, Z.; Ding, X. Embracing imperfect datasets: A review of deep learning solutions for medical image segmentation. *Med. Image Anal.* **2020**, *63*, 101693. [CrossRef]
19. Badrinarayanan, V.; Kendall, A.; Cipolla, R. SegNet: A Deep Convolutional Encoder-Decoder Architecture for Image Segmentation. *IEEE Trans. Pattern Anal. Mach. Intell.* **2017**, *39*, 2481–2495. [CrossRef]
20. Ronneberger, O.; Fischer, P.; Brox, T. U-net: Convolutional networks for biomedical image segmentation. In Proceedings of the International Conference on Medical Image Computing and Computer-Assisted Intervention, Munich, Germany, 5–9 October 2015; Springer: Berlin/Heidelberg, Germany, 2015; pp. 234–241.
21. Yang, L.; Zhang, Y.; Chen, J.; Zhang, S.; Chen, D.Z. Suggestive Annotation: A Deep Active Learning Framework for Biomedical Image Segmentation. In Proceedings of the MICCAI, Quebec City, QC, Canada, 11–13 September 2017.
22. Ozdemir, F.; Peng, Z.; Tanner, C.; Fuernstahl, P.; Goksel, O. Active learning for segmentation by optimizing content information for maximal entropy. In *Deep Learning in Medical Image Analysis and Multimodal Learning for Clinical Decision Support*; Springer: Berlin/Heidelberg, Germany, 2018; pp. 183–191.
23. Gal, Y.; Ghahramani, Z. Dropout as a bayesian approximation: Representing model uncertainty in deep learning. In Proceedings of the International Conference on Machine Learning, New York, NY, USA, 19–24 June 2016; pp. 1050–1059.
24. Sourati, J.; Gholipour, A.; Dy, J.G.; Kurugol, S.; Warfield, S.K. Active deep learning with fisher information for patch-wise semantic segmentation. In *Deep Learning in Medical Image Analysis and Multimodal Learning for Clinical Decision Support*; Springer: Berlin/Heidelberg, Germany, 2018; pp. 83–91.
25. Kuo, W.; Häne, C.; Yuh, E.; Mukherjee, P.; Malik, J. Cost-sensitive active learning for intracranial hemorrhage detection. In Proceedings of the International Conference on Medical Image Computing and Computer-Assisted Intervention, Granada, Spain, 16–20 September 2018; Springer: Berlin/Heidelberg, Germany, 2018; pp. 715–723.
26. Zheng, H.; Yang, L.; Chen, J.; Han, J.; Zhang, Y.; Liang, P.; Zhao, Z.; Wang, C.; Chen, D.Z. Biomedical image segmentation via representative annotation. In Proceedings of the AAAI Conference on Artificial Intelligence, Honolulu, HI, USA, 27 January–1 February 2019; Volume 33, pp. 5901–5908.
27. Sourati, J.; Gholipour, A.; Dy, J.G.; Tomas-Fernandez, X.; Kurugol, S.; Warfield, S.K. Intelligent labeling based on fisher information for medical image segmentation using deep learning. *IEEE Trans. Med. Imaging* **2019**, *38*, 2642–2653. [CrossRef] [PubMed]
28. Shin, G.; Xie, W.; Albanie, S. All You Need Are a Few Pixels: Semantic Segmentation With PixelPick. In Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV) Workshops, Montreal, BC, Canada, 11–17 October 2021; pp. 1687–1697.
29. Zhang, L.; Gopalakrishnan, V.; Lu, L.; Summers, R.M.; Moss, J.; Yao, J. Self-learning to detect and segment cysts in lung ct images without manual annotation. In Proceedings of the 2018 IEEE 15th International Symposium on Biomedical Imaging (ISBI 2018), Washington, DC, USA, 4–7 April 2018; pp. 1100–1103.
30. Bai, W.; Oktay, O.; Sinclair, M.; Suzuki, H.; Rajchl, M.; Tarroni, G.; Glocker, B.; King, A.; Matthews, P.M.; Rueckert, D. Semi-supervised learning for network-based cardiac MR image segmentation. In Proceedings of the International Conference on Medical Image Computing and Computer-Assisted Intervention, Quebec City, QC, Canada, 11–13 September 2017; Springer: Berlin/Heidelberg, Germany, 2017; pp. 253–260.
31. Vincent, L.; Soille, P. Watersheds in digital spaces: An efficient algorithm based on immersion simulations. *IEEE Comput. Archit. Lett.* **1991**, *13*, 583–598. [CrossRef]
32. Chakravarthy, A.D.; Chundi, P.; Subramaniam, M.; Ragi, S.; Gadhamshetty, V.R. A Thrifty Annotation Generation Approach for Semantic Segmentation of Biofilms. In Proceedings of the 2020 IEEE 20th International Conference on Bioinformatics and Bioengineering (BIBE), Cincinnati, OH, USA, 26–28 October 2020; pp. 602–607.
33. Grau, V.; Mewes, A.U.; Alcañiz, M.; Kikinis, R.; Warfield, S.K. Improved watershed transform for medical image segmentation using prior information. *IEEE Trans. Med. Imaging* **2004**, *23*, 447–458. [CrossRef] [PubMed]
34. Grau, V.; Kikinis, R.; Alcañiz, M.; Warfield, S.K. Cortical gray matter segmentation using an improved watershed transform. In Proceedings of the Annual International Conference of the IEEE Engineering in Medicine and Biology—Proceedings, Cancun, Mexico, 17–21 September 2003; Volume 1, pp. 618–621. [CrossRef]

35. Ng, H.P.; Ong, S.H.; Foong, K.W.C.; Goh, P.S.; Nowinski, W.L. Medical Image Segmentation Using K-Means Clustering and Improved Watershed Algorithm. In Proceedings of the 2006 IEEE Southwest Symposium on Image Analysis and Interpretation, Hangzhou, China, 2–4 November 2001; pp. 61–65. [CrossRef]
36. Beucher, S.; Meyer, F. The Morphological Approach to Segmentation: The Watershed Transformation. Avalilable online: https://www.researchgate.net/profile/Serge-Beucher/publication/230837870_The_Morphological_Approach_to_Segmentation_The_Watershed_Transformation/links/00b7d5319b26f3ffa2000000/The-Morphological-Approach-to-Segmentation-The-Watershed-Transformation.pdf (accessed on 23 May 2022).
37. Salembier, P. Morphological multiscale segmentation for image coding. *Signal Process.* **1994**, *38*, 359–386. [CrossRef]
38. Malpica, N.; De Solórzano, C.O.; Vaquero, J.J.; Santos, A.; Vallcorba, I.; García-Sagredo, J.M.; Del Pozo, F. Applying watershed algorithms to the segmentation of clustered nuclei. *Cytometry: J. Int. Soc. Anal. Cytol.* **1997**, *28*, 289–297. [CrossRef]
39. Petit, O.; Thome, N.; Charnoz, A.; Hostettler, A.; Soler, L. Handling missing annotations for semantic segmentation with deep convnets. In *Deep Learning in Medical Image Analysis and Multimodal Learning for Clinical Decision Support*; Springer: Berlin/Heidelberg, Germany, 2018; pp. 20–28.
40. Bengio, Y.; Louradour, J.; Collobert, R.; Weston, J. Curriculum learning. In Proceedings of the 26th Annual International Conference on Machine Learning, Montreal, QC, Canada, 14–18 June 2009; pp. 41–48.
41. Petit, O.; Thome, N.; Soler, L. Iterative confidence relabeling with deep ConvNets for organ segmentation with partial labels. *Comput. Med. Imaging Graph.* **2021**, *91*, 101938. [CrossRef] [PubMed]
42. Pan, J.; Bi, Q.; Yang, Y.; Zhu, P.; Bian, C. Label-efficient Hybrid-supervised Learning for Medical Image Segmentation. *arXiv* **2022**, arXiv:2203.05956.
43. Zhou, H.; Song, K.; Zhang, X.; Gui, W.; Qian, Q. WAILS: Watershed Algorithm With Image-Level Supervision for Weakly Supervised Semantic Segmentation. *IEEE Access* **2019**, *7*, 42745–42756. [CrossRef]
44. Zhou, S.; Nie, D.; Adeli, E.; Yin, J.; Lian, J.; Shen, D. High-resolution encoder–decoder networks for low-contrast medical image segmentation. *IEEE Trans. Image Process.* **2019**, *29*, 461–475. [CrossRef]
45. Ning, Z.; Zhong, S.; Feng, Q.; Chen, W.; Zhang, Y. SMU-Net: Saliency-Guided Morphology-Aware U-Net for Breast Lesion Segmentation in Ultrasound Image. *IEEE Trans. Med. Imaging* **2022**, *41*, 476–490. [CrossRef]
46. Arganda-Carreras, I.; Turaga, S.C.; Berger, D.R.; Cireşan, D.; Giusti, A.; Gambardella, L.M.; Schmidhuber, J.; Laptev, D.; Dwivedi, S.; Buhmann, J.M.; et al. Crowdsourcing the creation of image segmentation algorithms for connectomics. *Front. Neuroanat.* **2015**, *9*, 142. [CrossRef] [PubMed]
47. Amelard, R.; Wong, A.; Clausi, D.A. Extracting morphological high-level intuitive features (HLIF) for enhancing skin lesion classification. In Proceedings of the 2012 Annual International Conference of the IEEE Engineering in Medicine and Biology Society, San Diego, CA, USA, 28 August–1 September 2012; pp. 4458–4461.
48. Chilkoor, G.; Jawaharraj, K.; Vemuri, B.; Kutana, A.; Tripathi, M.; Kota, D.; Arif, T.; Filleter, T.; Dalton, A.B.; Yakobson, B.I.; et al. Hexagonal boron nitride for sulfur corrosion inhibition. *ACS Nano* **2020**, *14*, 14809–14819. [CrossRef] [PubMed]
49. Chilkoor, G.; Sarder, R.; Islam, J.; ArunKumar, K.; Ratnayake, I.; Star, S.; Jasthi, B.K.; Sereda, G.; Koratkar, N.; Meyyappan, M.; et al. Maleic anhydride-functionalized graphene nanofillers render epoxy coatings highly resistant to corrosion and microbial attack. *Carbon* **2020**, *159*, 586–597. [CrossRef]
50. Chilkoor, G.; Shrestha, N.; Kutana, A.; Tripathi, M.; Robles, F.C.; Yakobson, B.I.; Meyyappan, M.; Dalton, A.B.; Ajayan, P.M.; Rahman, M.M.; et al. Atomic Layers of Graphene for Microbial Corrosion Prevention. *ACS Nano* **2020**, *15*, 447–454. [CrossRef] [PubMed]
51. Kingma, D.P.; Ba, J. Adam: A method for stochastic optimization. *arXiv* **2014**, arXiv:1412.6980.
52. Minaee, S.; Boykov, Y.Y.; Porikli, F.; Plaza, A.J.; Kehtarnavaz, N.; Terzopoulos, D. Image segmentation using deep learning: A survey. *IEEE Trans. Pattern Anal. Mach. Intell.* **2021**, *44*, 3523–3542. [CrossRef]
53. Chen, L.C.; Zhu, Y.; Papandreou, G.; Schroff, F.; Adam, H. Encoder-decoder with atrous separable convolution for semantic image segmentation. In Proceedings of the European Conference on Computer Vision (ECCV), Munich, Germany, 8–14 September 2018; pp. 801–818.
54. He, K.; Zhang, X.; Ren, S.; Sun, J. Deep residual learning for image recognition. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Las Vegas, NV, USA, 27–30 June 2016; pp. 770–778.
55. Rother, C.; Kolmogorov, V.; Blake, A. "GrabCut" interactive foreground extraction using iterated graph cuts. *ACM Trans. Graph. (TOG)* **2004**, *23*, 309–314. [CrossRef]
56. Monarch, R.M. *Human-in-the-Loop Machine Learning: Active Learning and Annotation for Human-Centered AI*; Simon and Schuster: Manning Publications: New York, NY, USA, 2021; pp. 1–456.