

Automatic Head Overcoat Thickness Measure with NASNet-Large-Decoder Net

Youshan Zhang, Brian D. Davison
Lehigh University
{yoz217, bdd3}@lehigh.edu

Vivien W. Talghader, Zhiyu Chen, Zhiyong Xiao, Gary J. Kunkel
Seagate Technology

{vivien.w.talghader, zhiyu.chen, zhiyong.x.xiao, gary.j.kunkel}@seagate.com

Abstract

Transmission electron microscopy (TEM) is one of the primary tools to show microstructural characterization of materials as well as film thickness. However, manual determination of film thickness from TEM images is time-consuming as well as subjective, especially when the films in question are very thin and the need for measurement precision is very high. Such is the case for head overcoat (HOC) thickness measurements in the magnetic hard disk drive industry. It is therefore necessary to develop software to automatically measure HOC thickness. In this paper, for the first time, we propose a HOC layer segmentation method using NASNet-Large as an encoder and then followed by a decoder architecture, which is one of the most commonly used architectures in deep learning for image segmentation. To further improve segmentation results, we are the first to propose a post-processing layer to remove irrelevant portions in the segmentation result. To measure the thickness of the segmented HOC layer, we propose a regressive convolutional neural network (RCNN) model as well as orthogonal thickness calculation methods. Experimental results demonstrate a higher dice score for our model which has lower mean squared error and outperforms current state-of-the-art manual measurement.

1. Introduction

Transmission electron microscopy (TEM) is a microscopy technique in which a beam of electrons is transmitted through a specimen to form an image. TEM is one of the premier tools to show microstructural characterization of materials [6]. Differences in material density show up as contrast in a TEM image.

The hard disk drive industry uses TEM to measure head overcoat (HOC) thickness on magnetic recording heads. In a bright-field TEM image, HOC appears as a white

layer sandwiched between two darker layers, a magnetic metal layer that is part of the recording head and a capping layer deposited as part of TEM sample preparation. This is because the HOC is comprised predominantly of Diamond-like Carbon (DLC) which has relatively low density whereas the other two layers are of much higher density.

Current state-of-the-art HOC thickness is sub-50Å and its accurate and precise measurement is critical for deposition process feedback since HOC thickness has a direct impact on magnetic recording performance. There is currently no automatic method to calculate the thickness of HOC from a TEM image, and manual calculation is time-consuming and highly subjective. Therefore, there is an industrial need to develop an efficient algorithm that can automatically measure HOC thickness.

There are several issues in need of resolution to automatically measure the HOC layer thickness. We do not have segmentation masks already defined for HOC, and there is no standard rule to define the thickness of the HOC layer—the boundary between the HOC layer and the sandwiching layers can be fuzzy, making it challenging to consistently discern, whether by the same operator or different operators. However, even if we have a clear mask of the HOC layer, it is still difficult to determine the HOC thickness due to the curvature and slope of the masked area. Hence, our challenges are: 1) how to correctly generate a segmentation that separates the HOC from the surrounding material; and, 2) how to accurately measure the thickness of the HOC layer.

Many traditional techniques have been proposed for segmentation, such as threshold methods [21, 20, 24], region-based methods [5, 15], genetic methods [8], level set methods [22, 13, 12] and artificial neural networks [17, 2, 1].

The threshold method is one of the most common and straightforward segmentation methods. It is a region segmentation technique which divides gray values into two or more gray intervals, and chooses one or more appropriate

thresholds to judge whether the region meets the threshold requirement according to the difference between the target and the background, and separates the background and the target to produce a binary image. Threshold processing has two forms: global threshold and adaptive threshold. The global threshold only sets one threshold, and the adaptive threshold sets multiple thresholds. The target and background regions are segmented by determining the threshold at the peak and valley of the gray histogram [10, 16]. Level set methods are also widely used in the segmentation task. The basic idea of the level set method for image segmentation is the continuous evolution of curve motion. The boundary of the image is searched until the target contour is found and then the moving curve is stopped. Curves are moved along every three-dimensional section of images to slice different levels of the three-dimensional surface. The level of the obtained closed curves of each layer change over time, and finally get a corresponding shape extraction contour [22]. In our tasks, the brightness is not consistent from image to image. Therefore, the performance of traditional threshold-based and level-set based methods are not good.

Deep neural networks have also been applied in segmentation and they supersede many traditional image segmentation approaches. Garcia et al. presents an overview of deep learning-based segmentation methods [7]. There are several models to address the segmentation task. Fully convolutional networks (FCN) [14] is one of the earliest approaches for deep learning-based image segmentation which performs end-to-end segmentation. FCN is a type of convolutional network for dense prediction that does not require an additional fully connected layer. This method leads to the possibility of segmenting images of any size effectively, and it is much faster than the patch classification method (which divides images into several patches). Almost all of the other more advanced methods follow this architecture. However, there are several limitations of the FCN model, such as the inherent spatial invariance causes the model to fail to take into account useful global context information and its efficiency in high-resolution scenarios is worse and not available for real-time segmentation. Another difficulty of using CNN networks in segmentation is the existence of pooling layers. The pooling layer not only enlarges the sensing field of the upper convolution layer but also aggregates the background and discards part of the location information. However, the semantic segmentation method needs to adjust the category map accurately, so it needs to retain the location information abandoned in the pooling layer.

Later the encoder-decoder architecture becomes widely used in segmentation. Segnet [3] and U-net [18] are representative encoder-decoder architectures. This architecture first selects a classification network such as VGG-16, and then removes its full connection layer to produce a low-resolution image representation or feature mapping. This

part of the segmentation network is usually called an encoder. A decoder is a complementary part of the network that learns how to decode or map these low-resolution images to the prediction at the pixel level. The difference between different encoder-decoder architectures is the design of the decoder. U-net uses a structure called dilated convolutions and removes the pooling layer structure. Chen et al. [4] proposed the Deeplab model, which used dilated convolutions and fully connected conditional random field to implement the atrous spatial pyramid pooling (ASPP), which is an atrous version of SPP [9] and can account for different object scales and improve the accuracy.

After getting the segmented HOC layer from a segmentation network, we need to predict the thickness of the HOC layer. Thickness measurement is a regression problem, and deep networks have been developed to solve such tasks. Kang et al. [11] proposed a convolutional neural network for estimating image quality, but its architecture is simple, consisting of max pooling and two dense layers. Zhang et al. [25] proposed a regressive neural network for electricity prediction, but the input is a vector, not images. Rothe et al. [19] proposed a regressive VGG-16 network for age prediction. However, a complex neural network may not solve the problem due to overfitting, since the input to the neural network here is the segmented HOC layer. Therefore, it is necessary to design a robust regression convolutional neural network model for predicting the thickness of the HOC layer.

Our contributions are three-fold:

1. We are the first to define a NASNet-Large-decoder network to segment TEM head overcoat images;
2. To remove irrelevant small segments, we are the first to propose a post-processing layer which is able to filter out false segmented sections in prediction images; and,
3. We propose an orthogonal calculation method and a regressive convolutional neural network to predict the thickness of the segmented head overcoat layer. Experimental results demonstrate that the proposed model achieves the highest dice and IoU score versus state-of-the-art methods.

This paper is organized as follows: in Sec. 2, the NASNet-Large-decoder network, orthogonal distance calculation and regressive networks are summarized; we present the segmentation results and the predicted thickness in Sec. 3; in Sec. 4, we discuss the advantages and disadvantage of proposed model and conclude in Sec. 5.

2. Methods

In this section, we first introduce the network to segment the TEM HOC images and then describe the regressive net-

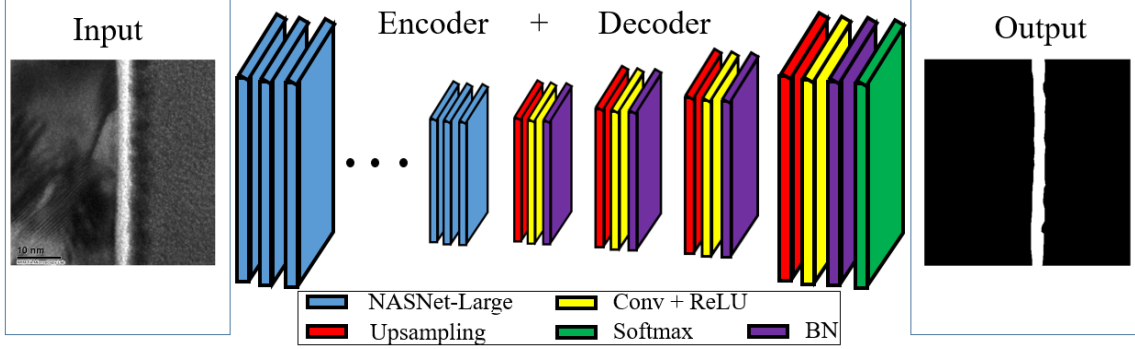


Figure 1. The architecture of our NASNet-Large segmentation network. The Encoder consists of the first 414 layers from NASNet-Large model. There are four blocks in the decoder, and each block contains Upsampling, Conv+ReLU, and BN layers. (Convolution (Conv), Batch normalization (BN), Rectified linear units (ReLU).) The final decoder is fed into a softmax layer for HOC layer prediction.

work and concept of an orthogonal distance to calculate the thickness of segmented HOC images.

2.1. NASNet-Large segmentation network

The NASNet-Large segmentation network contains an encoder and a decoder, which is followed by a classification layer. The architecture is shown in Fig. 1. There are two significant differences in our model comparing with Segnet, which employs the pre-trained VGG16 network for the encoder. Our Nasnet-Large-decoder net uses the first 414 layers of Nasnet-Large net (which is a network well-trained for ImageNet classification) as the encoder to decompose images [26]. We choose the first 414 layers because the size of the final layer is close to the size of the original image, so it will not lose much information. If we choose the last layer as the feature extraction layer, it will destroy significant structural information of objects since the last layer will be more suitable for classification, rather than segmentation. We do not use the pre-trained weights but retrain the net using new data to fit Nasnet-Large in our experiment since our dataset is significantly different from ImageNet. In addition, the decoder is different and there are no pooling indices in our model since Nasnet-Large net can produce detailed information for the decoder.

An appropriate decoder can upsample its input feature map using the max-pooling layer. The decoding technique is illustrated in Fig. 1. There are four blocks in the decoder. Each block begins with upsampling which can expand the feature map, followed by convolution and rectified linear units. A batch normalization layer is then applied to each of these maps. The first decoder, which is closest to the last encoder can produce a multi-channel feature map. This is similar to Segnet, which can generate a different number of sizes and channels as their encoder inputs. The final output of the last decoder layer is fed to a trainable soft-max classifier which produces a K channel image of probabilities where K is the number of classes (two in our problem). The predicted segmentation corresponds to the class with

maximum probability at each pixel.

2.2. Orthogonal calculation of HOC thickness

The current measurement technique for HOC layer thickness is a manually labeled image with three lines as shown in Fig. 3. From the lengths of the three lines, we calculate the mean thickness and standard deviation of the HOC layer. However, there are two obvious shortcomings of this method. One is that it only measures three lines, not all lines of the HOC layer. Secondly, the three lines are horizontal, which may not measure the orthogonal distance of HOC layer. If there is a slope in the HOC layer, the horizontal three-line mean will produce a larger distance than the real thickness. Therefore, to calculate thickness of the segmented HOC layer from Sec. 2.1, and capture the physical meaning of the measurement, we also propose to calculate an orthogonal distance. We need to give a formal definition of thickness of the HOC layer. As shown in Fig. 4, ten black lines sample the thickness of the HOC layer, and are perpendicular to the middle regression line. There are more than one hundred orthogonal lines in the segmented HOC mask (i.e., one at each pixel); we only show ten of them. The thickness of HOC is defined as the mean of all lines which are perpendicular to middle regression line. The standard deviation (SD) represents the variation in thickness across all orthogonal lines.

2.3. Regressive thickness prediction net

However, orthogonal calculation of thickness heavily depends on the boundary of segmented mask. To alleviate this issue, we propose a regressive CNN (RCNN) model, which is described in Fig. 2. The input is the segmented mask from Sec. 2.1, which contains all information about the HOC layer within the images instead of simple boundary of the image, and the output is mean thickness of HOC layer. There are two major blocks in the RCNN model. Each of them consists of three convolutional layers and is

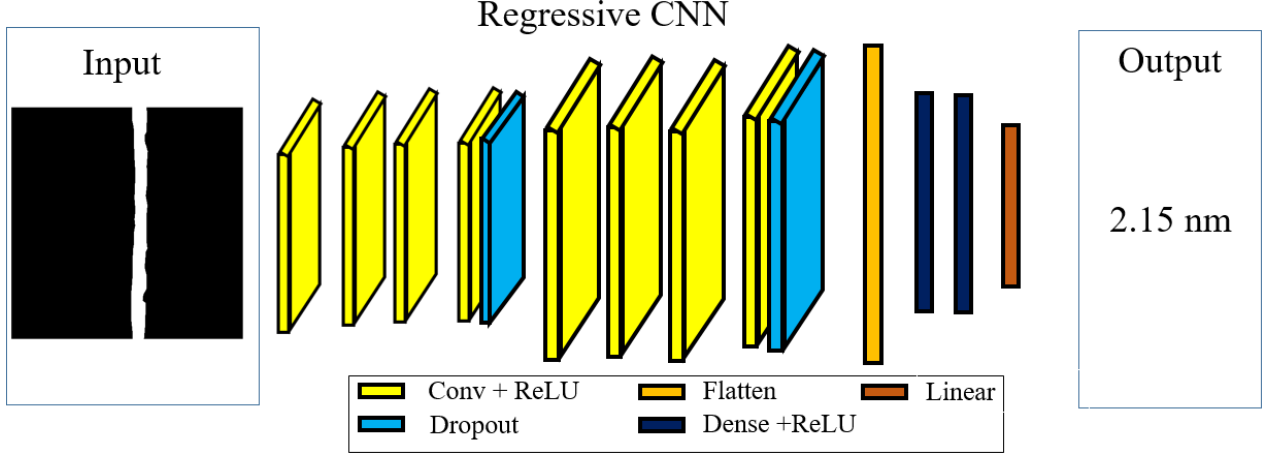


Figure 2. The architecture of the thickness regression network. The input is from the NASNet-Large segmentation network. The thickness regression network primarily consists of two blocks, and each block includes Conv + ReLU and Dropout layers. It is then followed by a Flatten layer. Finally, it ends with a linear activation layer. The output is the thickness of the HOC layer.

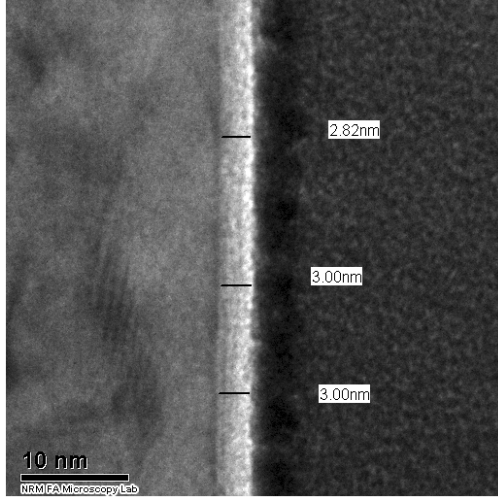


Figure 3. The existing three-line thickness measurement of the HOC layer. However, the three-line approach is manually labeled, and consists of horizontal lines, which will overestimate the thickness if within a slope or curve.

followed by one dropout layer. After that a flatten layer expands all features in its front dropout layer into a vector. Two dense and activation layers follow. The regression network ends with a linear activation to predict the thickness of the HOC layer. During the training stage of the RCNN model, the labeled thickness is from the proposed orthogonal calculation method.

There are two advantages of our RCNN model. First, the RCNN model is not complex and easy to train. If the architecture is too complex, it may lead to an overfitting problem. One reason is the input is the segmented HOC layer, which does not contain too much information. Hence, a complex neural network is unnecessary. Another advantage of the

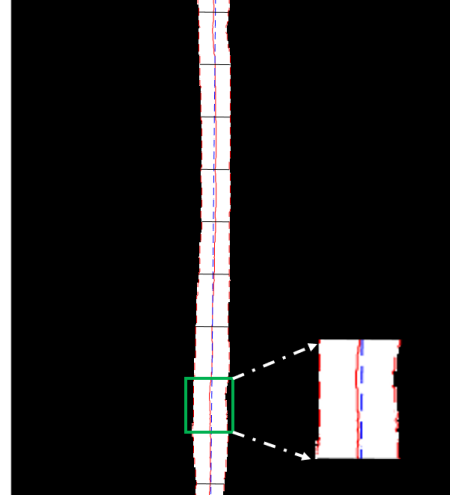


Figure 4. The orthogonal distance of the HOC layer. Two red dashed lines are the boundaries of the HOC layer; the middle red line consists of midpoints of boundary points, and the blue dashed line is the regression line. Several black lines are the orthogonal distance of the HOC layer. Lower right shows magnified green box.

RCNN model is that it can give a prediction with less error. In the training stage, the RCNN model is fed with the segmented mask and outputs orthogonal calculation distance. The loss function is mean squared error (MSE), which is defined in Eq. 1.

$$\text{MSE} = \frac{1}{n} \sum_{i=1}^n (Y_i - \hat{Y}_i)^2, \quad (1)$$

where Y_i is the prediction value, and \hat{Y}_i is the orthogonal distance measurement as shown in Sec. 2.2. Therefore, we minimize the error between orthogonal distance using the

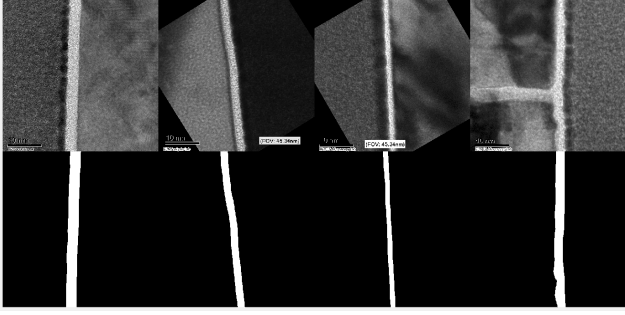


Figure 5. Four TEM HOC images and their corresponding HOC segmentations.

predicted HOC mask and the orthogonal thickness. This further drives the prediction mean thickness to be close to the orthogonal thickness.

Unfortunately, a deep network is often difficult to trust given its opacity. Domain experts prefer a clear measurement. However, the proposed RCNN model can be validated, if needed, by the orthogonal calculation.

3. Results

3.1. Datasets

Our modeling was trained with data, that are collected from three different dates, differing in the composition of the HOC material. Due to a limited sample size of 364 images, we use five-fold cross-validation to estimate the performance of our model. We normalize pixel values of the image into a scale between 0 and 1. Since we do not have ground truth for the mask of HOC layer, the TEM image was first marked by an expert using thickness measurement software (an internally developed package based on Zhang et al. [23]) to manually segment the mask and output rough thickness. Fig. 5 shows four TEM images with their corresponding segmentation labels. The middle two examples in Fig. 5 cause difficulty with the HOC layer segmentation; there is a FOV (field of view) sticker in the third image, and there is an extra horizontal layer in the fourth image. Due to these reasons, the traditional threshold-based methods failed in segmenting these three images.

Our experiments are implemented within Keras, and our network was trained on an NVIDIA TITAN XP equipped with 12 GB of memory in order to exploit its computational speed. The network parameters were set to: Batch size: 4, Step size: 5, Number of epochs: 1000.

3.2. Segmentation results visualization

Fig. 6 compares the predicted segmented HOC image with the ground truth image in the test dataset. The prediction image is close to the real mask, which qualitatively demonstrates the high performance of our model.

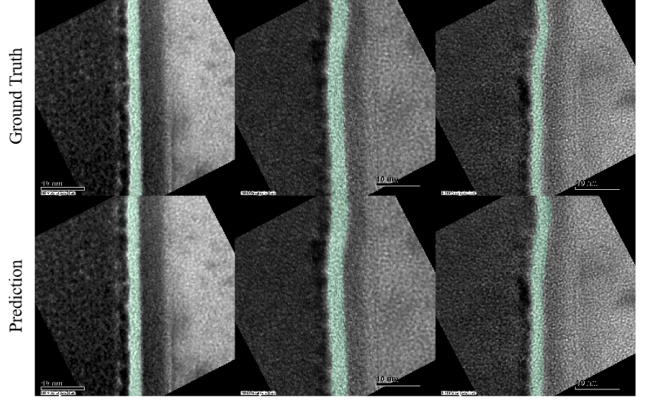


Figure 6. Ground truth vs. prediction results. The first row is the ground truth mask overlay with the raw image, and the second row is the prediction result overlaid with the raw image.

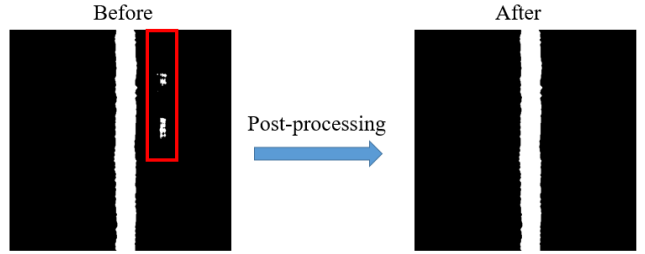


Figure 7. The post-processing of prediction result. On the left is the prediction result from the proposed network, and on the right is the result after using our post-processing layer. The red box outlines an irrelevant feature.

3.3. Post-processing layer

However, there are some areas in the image which are not the true HOC layer in the prediction result. We then propose a post-processing layer, which can filter the irrelevant parts of the prediction. The first step of post-processing layer is to classify the predicted image into several parts¹, and we then simply select the largest areas (the HOC area) as the final segmented HOC layer. As shown in the left image in Fig. 7, the red box highlights erroneous predictions of the image (false negatives). After post-processing, the red box is removed. Therefore, the prediction result can be improved if we filter out such irrelevant parts.

3.4. Metrics

To evaluate the performance of our NASNet-Large segmentation network, we use the widely-used Dice coefficient index to indicate the goodness of the segmentation results. Furthermore, we also report the IoU score. The two metrics

¹The major step of post-processing step is using the *connectedComponentsWithStats* function in OpenCV.

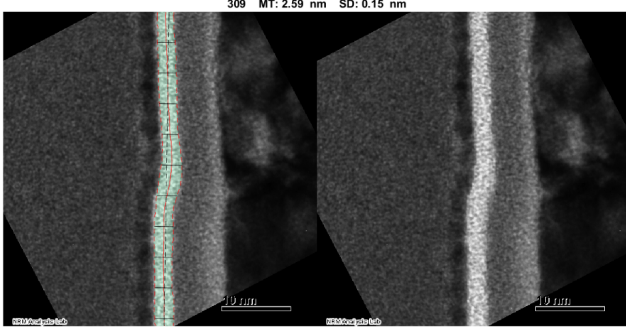


Figure 8. Sample output image. Left: the predicted result of our model; right: the original image. The segmented mask is perfectly overlapped with the original image (MT: mean thickness, SD: standard deviation).

are defined in the following formulas:

$$Dice = 2 \times \frac{|A \cap B|}{|A| + |B|}, \quad IoU = \frac{|A \cap B|}{|A \cup B|},$$

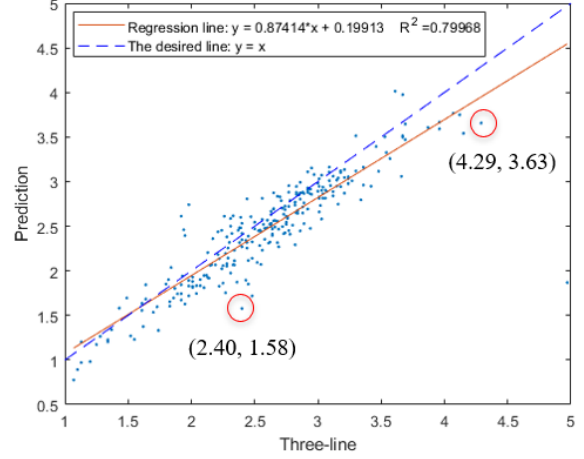
where A is ground truth mask, and B is the prediction mask. We also compare our results with performance of state-of-the-art methods.

Table 1. Five-fold cross-validation segmentation results

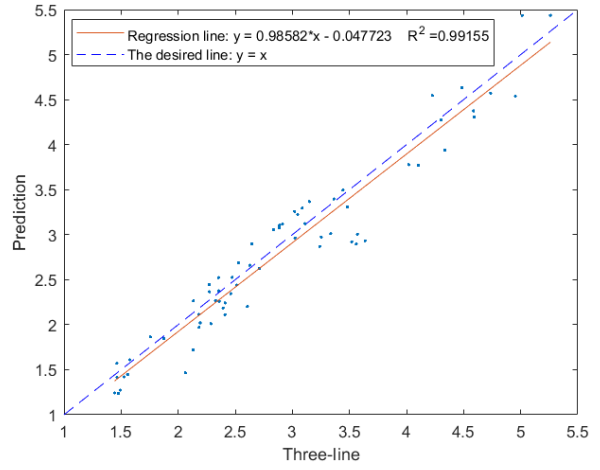
Method	IoU	Dice score
U-Net	0.80	0.85
Deep-Lab	0.83	0.89
Segnet	0.78	0.83
NasnetLarge-net	0.86	0.92
NasnetLarge-net-post	0.89	0.94

Tab. 1 lists the comparison results of different models (the metrics are only reported on HOC layer and not the background area). Our NASNet-Large-Decoder has a higher IoU and Dice score than other models. We also observe that the method including post-processing outputs the highest scores, which illustrates that the post-processing layer is useful in our segmentation task.

We use MSE to evaluate the performance of RCNN model and traditional orthogonal distance calculation as defined in Eq. 1. As shown in Tab. 2, the MSE of RCNN model is less than the other three methods, which implies that RCNN model gives a closer prediction to orthogonal measurement than Segnet RCNN and NASNet-Large RCNN. Differing from original Segnet and NASNet-Large model, we replace the final softmax layer with a regressive layer to form the Segnet RCNN and NASNet-Large RCNN model. Hence, these two models are able to predict the thickness. Fig. 8 shows the final output of our thickness measurement tool. The left image is the segmented HOC layer overlay with original image, and it also shows orthogonal distance. We also add the file name, mean thickness



(a) Training with 0.0929 MSE



(b) Test with 0.0719 MSE

Figure 9. Comparison of the existing three-line method and our predication in training and test datasets.

and standard deviation of the thickness on the top of the image.

Table 2. Thickness prediction results comparison

Methods	MSE
Regressive CNN	0.0089
Three-line measurement	0.1524
Segnet RCNN	0.1792
NASNet-Large RCNN	0.2023

4. Discussion

In this section, we first analyze predicted HOC thickness and then discuss the merits and demerits of our model.

4.1. Statistical analysis

As shown in Fig. 9, we compare the three-line measurements with orthogonal distance from predicted HOC mask

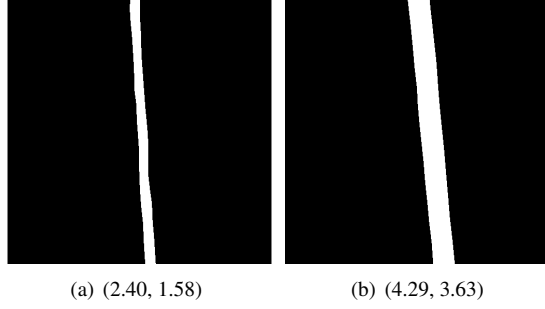


Figure 10. Two outlier images, which correspond to the highlighted points in Fig. 9(a). Labels are (X,Y) pairs. X-value is the thickness from three-line measurement and y-value is the prediction value from orthogonal distance.

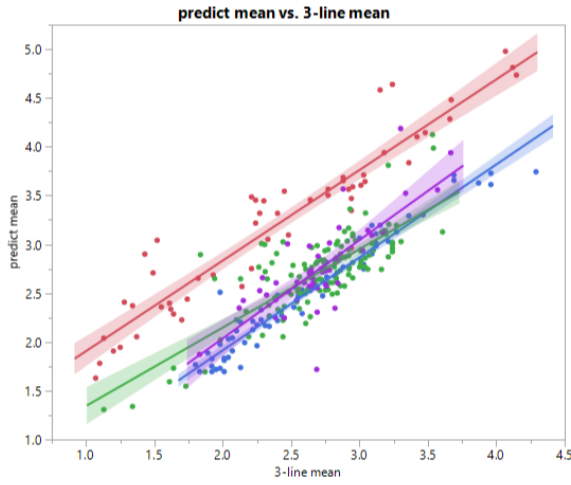
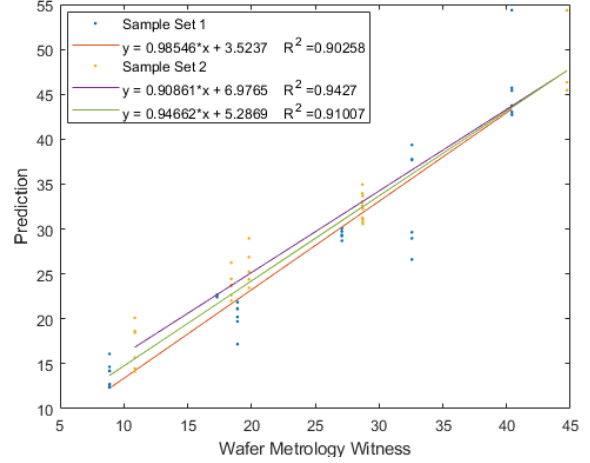
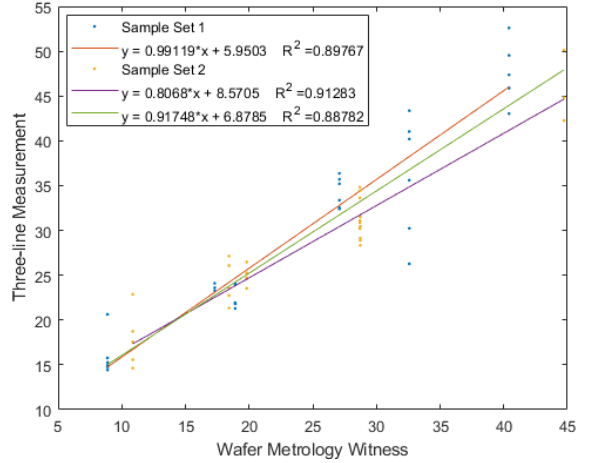


Figure 11. Linear regression fit of predicted thickness and the three-line thickness (Different color represents data are collected from different dates).

using linear regression in training and test datasets, respectively. The blue dashed lines in the figure indicate the desired line if prediction results are the same as the human three-line measurements. The regression slope from training and test are 0.79968 and 0.99155, respectively, which demonstrate a strong linear correlation between the human measurement and deep learning method. Especially notable is the test dataset, in which the prediction is almost the same as human measurement. We also calculate the mean squared error (MSE) in Eq. 1 to indicate the error between deep learning prediction and human measurement. Although there are some outliers in Fig. 9(a), eg., (2.40, 1.58) and (4.29, 3.63) (the x-value is three-line result and y-value is prediction results from our model), the overall performance of the orthogonal measurement method is closer to human measurement. The three-line method has a higher value than deep learning prediction in these two off-line points. We further check the prediction mask images of these two points as shown in Fig. 10. We observe that there



(a) Deep learning prediction



(b) Human measurement

Figure 12. TEM human 3-line measurement and TEM deep learning prediction compared against measurements made with witness wafer metrology. Deep learning prediction gives more accurate results than human three-line measurement.

is a slope in these two prediction HOC layer. Remember, the three-line is measured using horizontal lines, which will lead to a higher value than the real distance if there is a slope in the boundary of mask images. However, we calculate orthogonal distance, which represents the real thickness of the HOC layer. Therefore, we can conclude that our method can be applied in thickness measurements, and it will lead to better results than human three-line measurement.

Fig. 11 describes overall performance in the training and test in terms of three different data collection dates using the mean thickness and standard deviation. We further compare the prediction thickness from our model and the three-line approach with measurements made on witness wafers included in the same HOC depositions as the product heads that were measured with TEM. The witness wafers were measured using metrology calibrated using an

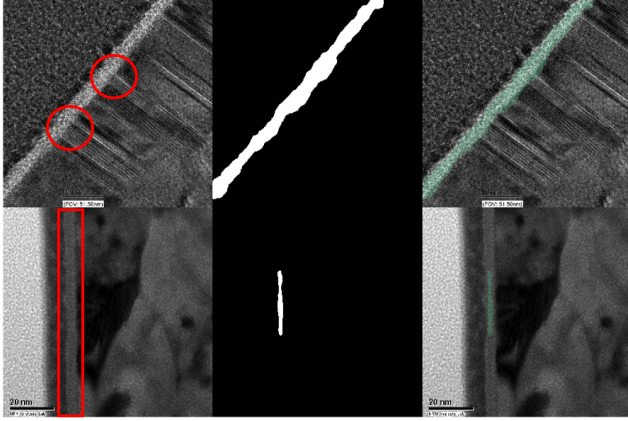


Figure 13. Two failures in NASNet-Large-Decoder model. The first row of image is caused by the fuzzy boundary and the second row of image is caused by the unusually dark HOC layer.

amalgamation of techniques, such as AFM (Atomic Force Microscopy), ellipsometry, and XRR (X-Ray Reflectometry).

From Fig. 12, we can find that the deep learning-based model has slightly worse results in sample set #1 since the slope is smaller than the three-line measurement (0.98546 vs. 0.99119). However, the performance in sample set #2 is significantly higher than the three-line measurement (0.90861 vs. 0.8068). Also, the overall prediction of our model is better than human measurement (as shown in the green lines in Fig. 12). Therefore, we can conclude that our model has a better performance than traditional manual measurement.

Although the RCNN model gives more accurate results than three-line measurement (as shown in Tab. 2), we did not use it to predict thickness of test datasets. The crucial reason is that customers require a comparison image for thickness measure as shown in Fig. 8. But RCNN model cannot output images with actual calculation of distance (i.e., cannot output an image as in the left of Fig. 8). Hence it is not used on test datasets. However, RCNN model is a back-up model if the orthogonal distance failed.

One of the distinct advantages of our model is that it achieves a higher Dice and IoU score. And there are two reasons: the designed NASNet-Large segmentation net is suitable for TEM image segmentation, and the post-processing layer filters out the unnecessary parts in the image, which improves the segmented results.

Although our model achieves a 0.94 Dice score, it still fails in some cases. As shown in Fig. 13, we observe that segmentation results are worse in these two situations. There are two reasons; one is that there is no similar image in the training dataset, which leads to worse performance. Second is that our model is not robust enough to deal with some unusual images. Comparing with normal training im-

ages as shown in Fig. 5, two worse raw images (in Fig. 13) are either too light (first row) or too dark (second row), and it is even difficult for humans to distinguish the HOC layer and background. Also, our model has limited sample size (less than 400 images), and so while in the validation and test datasets the MSE is small, our model could not cover all the cases, e.g., in Fig. 13. Therefore, including more samples in training would help to get a more robust model. In addition, a limitation of the NASNet-Large-Decoder network is that it needs a large memory to train the model. Therefore, designing a more efficient architecture is a direction for future work.

5. Conclusion

In this paper, we are the first to present a TEM HOC layer segmentation using the NASNet-Large-decoder architecture, and we get an accurate segmentation with 0.94 Dice score. A post-processing layer is employed to remove the unnecessary part in the prediction map. We then propose an orthogonal thickness calculation method as well as a regressive convolutional neural network (RCNN) model to measure the thickness of the segmented HOC layer. These advances enable manufacturers to replace a manual method with a much more accurate automated process.

Our model can be further improved with a more robust encoder and decoder model, to enable it to be applied in a wider variety of cases. The RCNN model can also be enhanced by predicting both the mean thickness and the standard deviation of the mean thickness.

References

- [1] M. Awad. An unsupervised artificial neural network method for satellite image segmentation. *Int. Arab J. Inf. Technol.*, 7(2):199–205, 2010. 1
- [2] M. Awad, K. Chehdi, and A. Nasri. Multicomponent image segmentation using a genetic algorithm and artificial neural network. *IEEE Geoscience and remote sensing letters*, 4(4):571–575, 2007. 1
- [3] V. Badrinarayanan, A. Kendall, and R. Cipolla. Segnet: A deep convolutional encoder-decoder architecture for image segmentation. *IEEE transactions on pattern analysis and machine intelligence*, 39(12):2481–2495, 2017. 2
- [4] L.-C. Chen, G. Papandreou, I. Kokkinos, K. Murphy, and A. L. Yuille. Semantic image segmentation with deep convolutional nets and fully connected crfs. *arXiv preprint arXiv:1412.7062*, 2014. 2
- [5] J. Freixenet, X. Muñoz, D. Raba, J. Martí, and X. Cufí. Yet another survey on image segmentation: Region and boundary information integration. In *European Conference on Computer Vision*, pages 408–422. Springer, 2002. 1
- [6] B. Fultz and J. M. Howe. *Transmission electron microscopy and diffractometry of materials*. Springer Science & Business Media, 2012. 1

- [7] A. Garcia-Garcia, S. Orts-Escolano, S. Oprea, V. Villena-Martinez, and J. Garcia-Rodriguez. A review on deep learning techniques applied to semantic segmentation. *arXiv preprint arXiv:1704.06857*, 2017. 2
- [8] K. Hammouche, M. Diaf, and P. Siarry. A multilevel automatic thresholding method based on a genetic algorithm for a fast image segmentation. *Computer Vision and Image Understanding*, 109(2):163–175, 2008. 1
- [9] K. He, X. Zhang, S. Ren, and J. Sun. Spatial pyramid pooling in deep convolutional networks for visual recognition. *IEEE transactions on pattern analysis and machine intelligence*, 37(9):1904–1916, 2015. 2
- [10] S. Hu, E. A. Hoffman, and J. M. Reinhardt. Automatic lung segmentation for accurate quantitation of volumetric x-ray ct images. *IEEE transactions on medical imaging*, 20(6):490–498, 2001. 2
- [11] L. Kang, P. Ye, Y. Li, and D. Doermann. Convolutional neural networks for no-reference image quality assessment. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 1733–1740, 2014. 2
- [12] C. Li, R. Huang, Z. Ding, J. C. Gatenby, D. N. Metaxas, and J. C. Gore. A level set method for image segmentation in the presence of intensity inhomogeneities with application to mri. *IEEE transactions on image processing*, 20(7):2007–2016, 2011. 1
- [13] C. Li, C. Xu, C. Gui, and M. D. Fox. Distance regularized level set evolution and its application to image segmentation. *IEEE transactions on image processing*, 19(12):3243–3254, 2010. 1
- [14] J. Long, E. Shelhamer, and T. Darrell. Fully convolutional networks for semantic segmentation. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 3431–3440, 2015. 2
- [15] J. Ning, L. Zhang, D. Zhang, and C. Wu. Interactive image segmentation by maximal similarity based region merging. *Pattern Recognition*, 43(2):445–456, 2010. 1
- [16] J. Pu, J. Roos, A. Y. Chin, S. Napel, G. D. Rubin, and D. S. Paik. Adaptive border marching algorithm: automatic lung segmentation on chest ct images. *Computerized Medical Imaging and Graphics*, 32(6):452–462, 2008. 2
- [17] W. E. Reddick, J. O. Glass, E. N. Cook, T. D. Elkin, and R. J. Deaton. Automated segmentation and classification of multispectral magnetic resonance images of brain using artificial neural networks. *IEEE Transactions on medical imaging*, 16(6):911–918, 1997. 1
- [18] O. Ronneberger, P. Fischer, and T. Brox. U-net: Convolutional networks for biomedical image segmentation. In *International Conference on Medical image computing and computer-assisted intervention*, pages 234–241. Springer, 2015. 2
- [19] R. Rothe, R. Timofte, and L. Van Gool. Dex: Deep expectation of apparent age from a single image. In *Proceedings of the IEEE International Conference on Computer Vision Workshops*, pages 10–15, 2015. 2
- [20] W.-B. Tao, J.-W. Tian, and J. Liu. Image segmentation by three-level thresholding based on maximum fuzzy entropy and genetic algorithm. *Pattern Recognition Letters*, 24(16):3069–3078, 2003. 1
- [21] O. J. Tobias and R. Seara. Image segmentation by histogram thresholding using fuzzy sets. *IEEE transactions on Image Processing*, 11(12):1457–1465, 2002. 1
- [22] L. A. Vese and T. F. Chan. A multiphase level set framework for image segmentation using the mumford and shah model. *International journal of computer vision*, 50(3):271–293, 2002. 1, 2
- [23] C. Zhang, K. Loken, Z. Chen, Z. Xiao, and G. Kunkel. Mask editor: an image annotation tool for image segmentation tasks. *arXiv preprint arXiv:1809.06461*, 2018. 5
- [24] J. Zhang and J. Hu. Image segmentation based on 2d otsu method with histogram analysis. In *2008 International Conference on Computer Science and Software Engineering*, volume 6, pages 105–108. IEEE, 2008. 1
- [25] Y. Zhang and Q. Li. A regressive convolution neural network and support vector regression model for electricity consumption forecasting. In *Future of Information and Communication Conference*, pages 33–45. Springer, 2019. 2
- [26] B. Zoph, V. Vasudevan, J. Shlens, and Q. V. Le. Learning transferable architectures for scalable image recognition. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 8697–8710, 2018. 3