This CVPR workshop paper is the Open Access version, provided by the Computer Vision Foundation.
Except for this watermark, it is identical to the accepted version;
the final published version of the proceedings is available on IEEE Xplore.

CNT-NeRF: Carbon Nanotube Forest Depth Layer Decomposition in SEM Imagery using Generative Adversarial Networks

Nguyen P. Nguyen[1], Ramakrishna Surya[2], Prasad Calyam[1], Kannappan Palaniappan[1], Matthew Maschmann[2], Filiz Bunyak[1]*
[1]Department of Electrical Engineering and Computer Science
[2]Department of Mechanical and Aerospace Engineering
University of Missouri-Columbia, MO, USA
Email: {npntz3, rst7b}@mail.missouri.edu
{ma8pz, calyamp, pal, maschmannm, bunyak}@missouri.edu

Abstract

Carbon nanotube (CNT) forests are imaged using scanning electron microscopes (SEMs) that project their multilayered 3D structure into a single 2D image. Image analytics, particularly instance segmentation is needed to quantify structural characteristics and to predict correlations between structural morphology and physical properties. The inherent complexity of individual CNT structures is further increased in CNT forests due to density of CNTs, interactions between CNTs, occlusions, and lack of 3D information to resolve correspondences when multiple CNTs from different depths appear to cross in 2D. In this paper, we propose CNT-NeRF, a generative adversarial network (GAN) for simultaneous depth layer decomposition and segmentation of CNT forests in SEM images. The proposed network is trained using a multi-layer, photo-realistic synthetic dataset obtained by transferring the style of real CNT images to physics-based simulation data. Experiments show promising depth layer decomposition and accurate CNT segmentation results not only for the front layer but also for the partially occluded middle and back layers. This achievement is a significant step towards automated, image-based CNT forest structure characterization and physical property prediction.

1. Introduction

Carbon nanotubes (CNTs) [17] are tubular nanostructures with remarkable mechanical, thermal, electrical, and chemical properties [9, 18]. Single-walled CNTs (SWNTs) may exibit metallic or semiconducting traits. Multi-walled CNTs (MWNTs) can be transformed into yarns [41] that are not only electrically conductive and strong, but also retain their flexibility and can even be tied into a knot. The properties of CNTs make them an appealing material for a wide range of applications, including dry spinning conductive, high-strength fibers [19, 41], piezoresistive sensing [23, 24, 30], electrochemical energy storage [5, 8], and thermal interface materials [6, 7].

Production of individual CNTs in isolation remains a challenge and is not practical for device-level integration. CNT forests, with dense populations synthesized on a support substrate, offer a solution by forcing the CNTs to grow vertically. Persistent van der Waals bonds, created by interactions between contacting CNTs, resist mechanical loads and result in an open-cell foam-like morphology. Despite their fascinating properties, CNT forests' characteristics are significantly diminished compared to individual CNTs. For instance, while an individual CNT has an elastic modulus of around 1 TPa, a CNT forest's compressive elastic modulus are frequently on the order of 1-10 MPa [26], akin to natural rubber. Variations in CNT forest morphology created during cooperative synthesis [1,22,31] are thought to be the root cause of the wide range of deformation mechanisms observed in compressed CNT forests [2,3,16,25–27,34,40].

Testing CNT forests' physical properties often necessitates their destruction, limiting further data collection. Imaging and image analysis offer a nondestructive opportunity to indirectly predict CNT forests' properties. CNT forests are imaged using scanning electron microscopy (SEM) or transmission electron microscopy (TEM). Figure 1 shows a CNT forest micropillar imaged at different zoom levels using SEM. Image analytics seeks to quantify structural characteristics such as diameter, orientation, curvature, tortuosity, density, spatial layout etc. to predict correlations between structural morphology and physical properties. This characterization requires robust segmentation of

*Corresponding author.

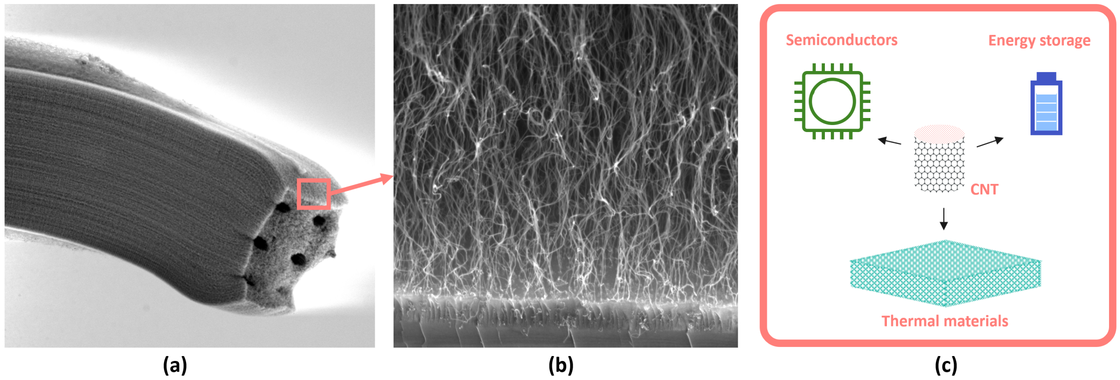

Figure 1. The scanning electron microscope (SEM) was used to capture images of a carbon nanotube (CNT) pillar. The images depict (a) a full view of the pillar, and (b) a magnified side view of the CNT pillar. (c) Various applications of CNTs: semiconductors, energy storage and thermal materials.

CNTs within CNT forests. This is a highly challenging task since the inherent complexity of individual CNT structures is further increased due to density of CNTs in CNT forests, front to back occlusion, lack of 3D information in SEM images to resolve correspondences when multiple CNTs from different depths appear to cross in 2D etc.

Earlier attempts to analyze CNT images mostly relied on traditional image processing techniques. For instance, thresholding was used in [12] to create partial CNT masks that were used to determine the diameters of CNTs, while class-entropy maximization was used in [38] to segment CNT images captured at modest magnification levels (800X-4000X). More recently neural networks-based approaches started to be used for CNT image analysis. In [35], a hybrid approach, combining thresholding and classification using multi-layer neural networks, was used to segment sparse, non-overlapping CNTs in small image patches. In [13, 14], deep learning networks were used to analyze synthetic CNT forest images generated from physics-based simulations to predict mechanical properties. In [29], a self-supervised learning network CNTSegNet was introduced for semantic segmentation of CNTs based on weak labels and orientation histograms computed in Fourier space.

In this paper, we present a neural radiance field (NeRF) [21, 28, 36] inspired deep learning network that aims to reduce image complexity in order to ensure accurate segmentation/tracing of individual CNTs. The proposed generative deep learning network decomposes CNT forest images into their depth layers and generates preliminary segmentation for each layer. The network converts a single 2D image into K 2D images corresponding to different depth layers. This 2.5D representation resolves front-versus-back relationships and occlusions, reduces image complexity for instance segmentation, lowers potential id-switches from one CNT to another during CNT tracing, and ultimately allows for a more comprehensive understanding of the spatial arrangement and intricacies of CNTs in the analyzed images.

2. Methods

2.1. Neural radiance field (NeRF)

Neural radiance field (NeRF) [28] is a recent neural networks-based approach for synthesizing novel views of complex 3D scenes using sparse set of views. The process starts with a virtual camera that casts rays into the scene to sample the 3D coordinates (x, y, z) in the scene with viewing angles (θ, ϕ) in order to project output colors and densities (probability of visibility/transparency). In the next step, a 3D scene dataset is utilized to train the neural network, enabling it to learn the relationships between the 5D input data (3D points and viewing directions) and their corresponding colors and densities [28]. Once trained, the neural network can take in new 3D points and viewing directions as input, and infer a set of colors and densities corresponding to each point in the scene. These inferred properties can then be used to render more realistic and accurate 3D scenes using techniques such as volume rendering or ray tracing.

Use of neural networks as black-box models for inferring properties of 3D scenes represents a significant breakthrough in the field of computer graphics, for applications from scientific imaging to video game/education design. As this technology advances, it has the potential to revolutionize the way we visualize and interact with 3D data, including complex microscopy data.

2.2. NeRF inspired CNT layer synthesis

Scanning electron microscope (SEM) projects the multi-layered 3D structure of CNT forests into a single 2D image. The planar neural radiance field proposed in [21, 36] uses planes instead of rays to represent the camera frustum. This method can be used for single-view image rendering as well as depth estimation. In our application, by sampling a num-

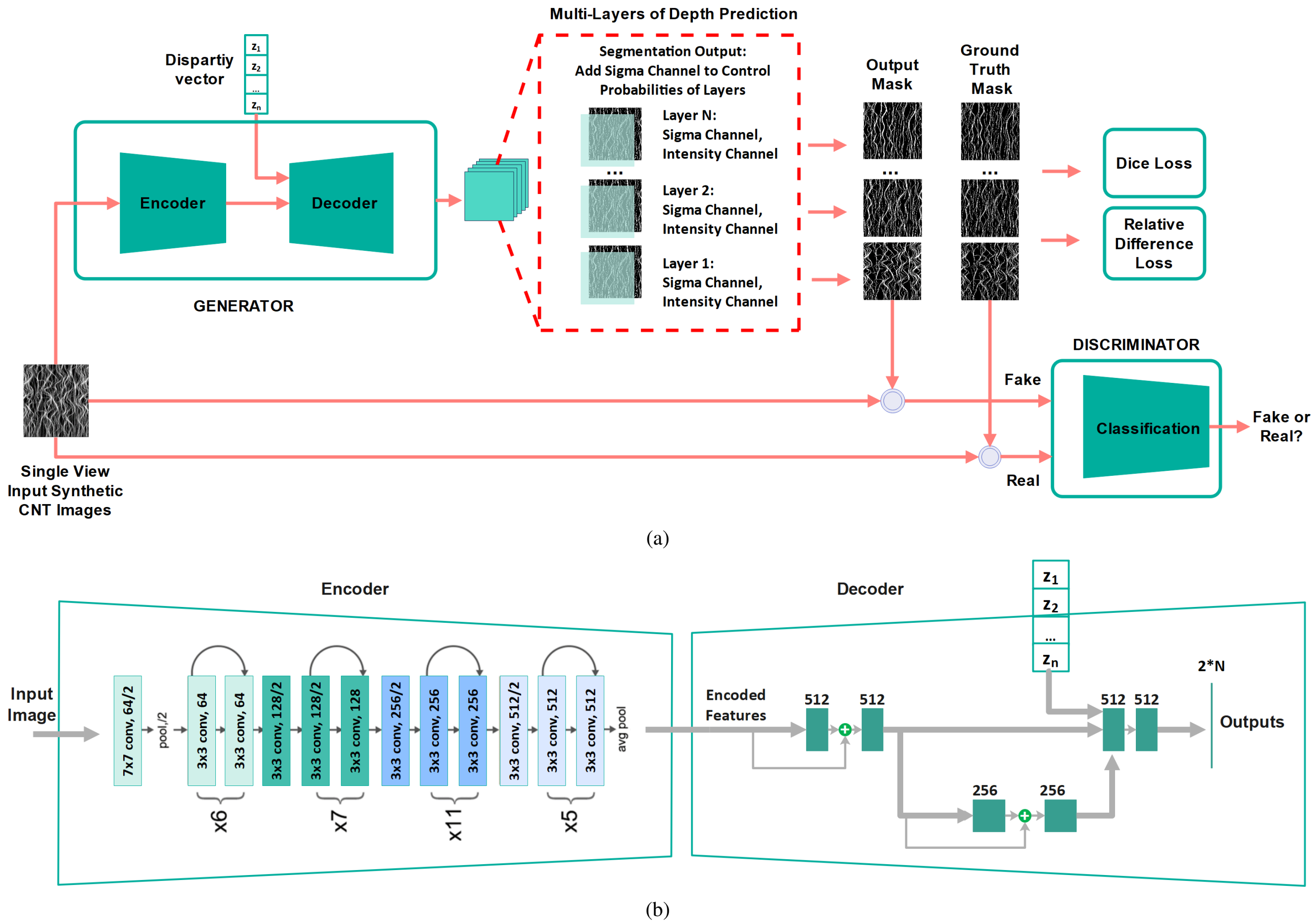

Figure 2. (a) The architecture of the CNT-NeRF network, along with the training process for both the generator and discriminator. (b) The network generator's settings. The design of the discriminator is the same as that of the generator encoder.

ber of planes within the camera frustum at different depth levels, we can efficiently capture the structural properties of complex 3D CNT forests. Our NeRF inspired CNT layer synthesis solution is designed for segmentation and depth estimation based solely on pixel (grayscale) intensity values. We intentionally ignore the camera matrix parameters and rely solely on depth layer information and single-view CNT forest images as inputs. This approach enables us to achieve accurate segmentation and depth estimation results while reducing computational complexity and potential errors caused by incorporating more parameters.

2.3. Network architecture

We assume that the segmentation mask for a CNT forest image consists of K layers partitioning the CNTs according to their depths indicated by similar grayscale intensity levels in the input image. With this assumption, we designed a novel framework for CNT forest depth layer decomposition and segmentation using a generative adversarial network (GAN). Figure 2 illustrates our network architecture and its training approach. The first block of this framework is the generator, which includes an encoder and a decoder. Our network utilizes a ResNet-34 [15] encoder to extract features from the input images. The decoder (using the architecture of Monodepth2 [11]) takes the extracted features from the latent space, combines them with the prior information from the disparity vector, and generates prediction outputs. The disparity vector [21, 33] is a positional encoding, similar to the ones used in transformer models [4, 37]. It maps the layer depth information into a $2 \times L$ dimension embedding space. Given a layer with depth z_i, its encoded disparity vector is computed as [21]:

$$\begin{aligned}\nu(z_i) =& [sin(2^0\pi z_i),\ \cos(2^0\pi z_i),\\ & \sin(2^1\pi z_i),\ \cos(2^1\pi z_i), \ldots\\ & \sin(2^{(L-1)}\pi z_i),\ \cos(2^{(L-1)}\pi z_i)]\end{aligned}$$

where the frequency parameter L was set to 10.

The ResNet-34 [15] encoder is used to extract robust and informative features from the input images, which are then used to generate high-quality predictions in the decoder. The disparity vector provides additional information about the depth of each layer, allowing our network to better distinguish between the adjacent layers and their orders.

For each output layer i, our segmentation network generates two channels: (1) a mask channel $\mathrm{M_i}$; and (2) corresponding sigma channel $\mathrm{Sigma_i}$ which controls visibility. Pixel-wise multiplication of the masks and corresponding sigma channels yield the final segmentation mask. This setting enables effective handling of occlusion issues where retrieval of CNTs in the back layers that may be obstructed by those in the front layers.

The second block of our CNT segmentation network, also using a ResNet-34 model [15]), is a discriminator responsible for classifying the output and ground truth masks as fake or real after they are applied to the input images. This process enables the discriminator to determine authenticity of the generated masks and provides feedback to the generator on how to improve its performance.

By integrating a discriminator into our network architecture, we can improve quality and accuracy of the generated masks. The discriminator provides an additional level of feedback and supervision, which can help the generator to learn the nuances and subtleties of the data more effectively. Below we introduce the discriminator and generator objective functions used to optimize the proposed GAN model.

2.4. Discriminator loss function

Assuming a GAN model with a generator $\mathcal{G}$ and a discriminator $\mathcal{D}$, we feed the input image I and the disparity vector v_d into this GAN model to generate a prediction mask $\mathrm{M_{pred}}$. The discriminator's goal is to classify both ground truth $(\mathrm{M_{GT}})$ and prediction $(\mathrm{M_{pred}})$ masks. To achieve this, the discriminator loss $L_{\mathcal{D}}$ is calculated using cross-entropy loss and is minimized during the training process.

$$\begin{aligned} L_{\mathcal{D}} &= \log \mathcal{D}(\mathrm{M_{GT}} \odot \mathrm{I}) + \log(1 - \mathcal{D}(\mathrm{M_{pred}} \odot \mathrm{I})) \\ &= \log \mathcal{D}(\mathrm{M_{GT}} \odot \mathrm{I}) + \log(1 - \mathcal{D}(\mathcal{G}(\mathrm{I}, \mathrm{v_d}) \odot \mathrm{I})) \end{aligned}$$

where $\odot$ refers to pixel-wise multiplication operation.

2.5. Generator loss function

If the generator is capable of generating perfectly realistic data, the discriminator should classify this data as real with high confidence. To achieve this, the generator's training objective is to minimize the difference between the predicted confidence value and the ground truth value, which represents the true classification of the generated data. The generator loss for classification is defined as:

$$\begin{aligned} L_{\mathcal{G}_{Class}} &= log(1 - \mathcal{D}(\mathrm{M_{pred}} \odot \mathrm{I})) \\ &= log(1 - \mathcal{D}(\mathcal{G}(\mathrm{I}, \mathrm{v_d}) \odot \mathrm{I})) \end{aligned}$$

Segmentation is a challenging task that requires a strong loss function to ensure that the predicted mask closely matches the ground truth mask. In addition to the classification loss above, we utilized two other loss functions, dice loss and scale invariant loss [10].

Given a prediction mask $\mathrm{M_{pred}}$ and a ground truth mask $\mathrm{M_{GT}}$, dice loss is computed using the following equation:

$$\begin{aligned} L_{\mathcal{G}_{Dice}}(\mathrm{M_{pred}}, \mathrm{M_{GT}}) &= 1 - 2 \times \frac{|\,\mathrm{M_{pred}} \cap \mathrm{M_{GT}}\,|}{|\,\mathrm{M_{pred}}\,| + |\,\mathrm{M_{GT}}\,|} \\ &= 1 - 2 \times \frac{|\,\mathcal{G}(\mathrm{I}, \mathrm{v_d}) \cap \mathrm{M_{GT}}\,|}{|\,\mathcal{G}(\mathrm{I}, \mathrm{v_d})\,| + |\,\mathrm{M_{GT}}\,|} \end{aligned}$$

It is common to use dice loss to match the prediction map and its ground truth mask. We have taken this approach a step further by incorporating a scale-invariant loss component, inspired by the method proposed in [10] for depth estimation. This addition aims to enhance the alignment of relative differences between each pair of pixels in the prediction map and the corresponding pair of pixels in the ground truth mask.

Suppose a ground truth mask and associated prediction map contains n pixels and y_i, ${y_i}^*$ denote intensity values at the i^{th} pixel of the associated prediction map and the ground truth mask respectively. We define d_i as the difference (in log scale) between the i^{th} pixel in the prediction map and the corresponding pixel in the ground truth mask:

$$d_i = log(y_i) - log({y_i}^*)$$

Using d_i, the scale-invariant loss computes the relative difference between each pair of pixels i and j in the prediction map and in the ground truth mask as following to minimize their distance.

$$\begin{aligned} &L_{\mathcal{G}_{Dif}}(\mathrm{M_{pred}}, \mathrm{M_{GT}}) \\ &= \frac{1}{2n^2} \sum_{i,j} [(\log y_i - \log y_j) - (\log y_i^* - \log y_j^*)]^2 \\ &= \frac{1}{n} \sum_i d_i^2 - \frac{1}{n^2} \sum_{i,j} d_i d_j \\ &= \frac{1}{n} \sum_i d_i^2 - \frac{1}{n^2} \Big(\sum_i d_i\Big)^2 \end{aligned}$$

The overall generator loss is computed as the weighted sum of the three loss components (classification, dice, and scale invariance) described above:

$$L_{\mathcal{G}} = k_1 \times L_{\mathcal{G}_{Class}} + k_2 \times L_{\mathcal{G}_{Dice}} + k_3 \times L_{\mathcal{G}_{Dif}}$$

where weights k_1,k_2, k_3 were empirically set to 0.1, 0.3, and 0.6 respectively.

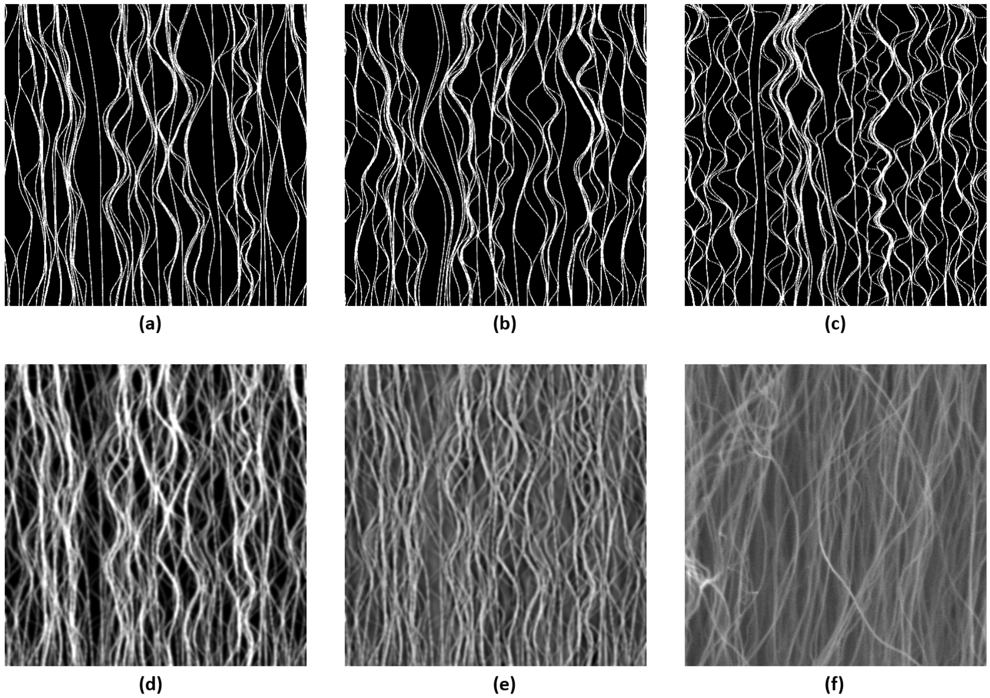

Figure 3. Synthetic and real CNT forest images. (a-c) Physics-based, 2D, single-layer CNT forest simulation [13]; (d) multi-layer synthetic image obtained by stacking single-layer images in (a-c); (e) multi-layer synthetic image in (d) with FDA style transfer [39] from real SEM image in (f).

3. Experimental results

3.1. Datasets

Synthetic images of CNT forests: In order to thoroughly evaluate the performance of the proposed layer segmentation approach, we generated an SEM-style, realistic-looking, synthetic CNT forest image dataset with associated ground truth masks. This synthetic dataset was generated by fusing multiple layers of 2D binary synthetic images obtained using the physics-based simulation technique described in [13]. Prior to fusion using a pixel-wise max operation, the individual layers (as shown in Figure 3a-c) were first smoothed by a Gaussian filter, then multiplied with a global weight according to their depth order in the combined image. The fused images (as shown in Figure 3d) were further improved by style transfer from a real SEM image (as shown in Figure 3f) using the Fourier Domain Adaptation (FDA) method described in [39]. FDA transfers the low-frequency features in Fourier space from the reference image to the target image, resulting in synthetic images with desired realistic styles (as shown in Figure 3e). The synthetic dataset comprised of 133 images of size 512 $\times$ 512 pixels with two versions with and without FDA style transfer. The synthetic dataset was partitioned into a training set of 106 and a test set of 27 images.

SEM images of CNT forests: The carbon nanotube (CNT) forests used in this study were produced using a chemical vapor deposition (CVD) technique as described in [20]. CNT forest images were acquired by utilizing a FEI Quanta scanning electron microscope (SEM) at a pixel dwell time of 10 μs and magnification of 50,000X. In total, 94 image patches of size 512×512 pixels were used to transfer SEM image style to synthetic images.

3.2. Experimental results on synthetic images

To assess the performance of our approach, we trained our network on these two different datasets: multi-layer synthetic images without FDA style transfer, and multi-layer synthetic images with FDA style transfer. Initially, we trained the generator in 125 epochs using both the dice loss and scale-invariant loss. Following this, we proceeded to fine-tune the generator by training it alongside the discriminator for an additional 50 epochs. We conducted a comparison between CNT-NeRF and U-net [32], a widely used architecture for segmentation. Table 1 presents segmentation dice scores for the proposed CNT-NeRF and U-net networks. We used a Resnet-34 [15] backbone encoder in the U-net network and trained it with the same training dataset as the proposed CNT-NeRF network.

As demonstrated in Table 1, CNT-NeRF outperformed

Table 1. Quantitative segmentation evaluation on synthetic data. Dice scores were computed between automated segmentation results for multi-layer synthetic images and ground truth masks corresponding to binary single-layer synthetic images used to generate the multi-layer synthetic images as illustrated in Figure 3.

		TRAINING SET				TEST SET			
		Layer 1 (Front)	Layer 2 (Middle)	Layer 3 (Back)	Avg	Layer 1 (Front)	Layer 2 (Middle)	Layer 3 (Back)	Avg
CNT-NeRF	Synthetic	0.86	0.84	0.81	0.84	0.86	0.82	0.76	0.81
	Synthetic + style transfer	0.87	0.84	0.80	0.83	0.87	0.83	0.78	0.83
U-net	Synthetic	0.85	0.78	0.70	0.78	0.80	0.74	0.67	0.74
	Synthetic + style transfer	0.86	0.82	0.73	0.80	0.78	0.70	0.61	0.70

U-net by achieving higher average dice scores for both datasets, with and without FDA-style transfer. CNT-NeRF has proven to be effective in segmenting the front and middle layers, and particularly accurate in extracting the back layer, even in the presence of occlusions. CNT-NeRF's performance surpassed that of U-net on Layer 1 by 1-to-9%, and on Layer 2 by 2-to-13%. Notably, for the back layer (Layer 3), CNT-NeRF's dice scores consistently surpassed U-net's by a substantial 7-to-17% margin for both training and test sets with and without style transfer. CNT-NeRF's performance remained less affected as the depth of the layers increased, in contrast to U-net, which exhibited a significant decline in layer 3.

Figure 4 (for the synthetic image without style transfer), Figure 5 (for the synthetic images with style transfer) show the depth layer decomposition, along with their corresponding segmentation outputs from synthetic images. In these figures, rows (a) present the synthetic images and their segmentation ground truth masks for the three layers. Row (b) and (d) depict the segmentation results for CNT-NeRF and U-net, respectively, starting from the front (layer 1) and extending to the back (layer 3). Row (c) and (e) highlight the segmentation outcomes (in green) of CNT-NeRF and U-net, overlaid on the segmentation masks (in red) and the synthetic image (in grayscale). The yellow areas signify the intersections where the segmentation outcomes align with the ground truth masks. The residual red regions represent false negatives, while the remaining green areas correspond to false positives of the segmentation tasks. By visual inspection, row (c) of CNT-NeRF displays fewer red/green regions compared to row (e) of U-net. This indicates that CNT-NeRF exhibits a lower rate of mis-prediction (false negatives and false positives) than U-net for synthetic images, both with and without style-transfer.

Additionally, CNT-NeRF exhibited better performance against overfitting, as evidenced by the minimal difference between training and test scores, ranging from 0-to-3%. In contrast, U-net experienced a performance drop of 4% and 10% when transitioning from the training set to the test set. This underscores CNT-NeRF's ability to generalize and maintain its effectiveness across different datasets.

4. Conclusions

In this paper, we proposed CNT-NeRF, a generative adversarial network for simultaneous depth layer decomposition and segmentation of CNT forests in SEM images. CNT-NeRF converts a single 2D image into K $\times$ 2D images. This 2.5D representation aims to reduce image complexity and resolve front-versus-back relationships and occlusions for a robust instance segmentation performance. Training of the network was done using our photo-realistic multi-layer synthetic images as input and associated physics-based binary synthetic layers as target labels. Promising depth layer decomposition and 7-to-15% improved CNT segmentation results were obtained compared to U-net segmentation network. The proposed depth layer decomposition and segmentation process is an important step towards automated and non-destructive characterization of CNT forest physical properties and our ultimate goal of human out of the loop material discovery.

Acknowledgement. This work was partially supported by the National Science Foundation under award number CMMI-2026847. Any opinions, findings, and conclusions or recommendations expressed in this publication are those of the author(s) and do not necessarily reflect the views of the National Science Foundation.

References

[1] Parisa Pour Shahid Saeed Abadi, Matthew R Maschmann, SM Mortuza, Soumik Banerjee, Jeffery W Baur, Samuel Graham, and Baratunde A Cola. Reversible tailoring of mechanical properties of carbon nanotube forests by immersing in solvents. *Carbon*, 69:178–187, 2014. 1

[2] Anna Brieland-Shoultz, Sameh Tawfick, Sei Jin Park, Mostafa Bedewy, Matthew R Maschmann, Jeffery W Baur, and A John Hart. Scaling the stiffness, strength, and toughness of ceramic-coated nanotube foams into the structural regime. *Advanced Functional Materials*, 24(36):5728–5735, 2014. 1

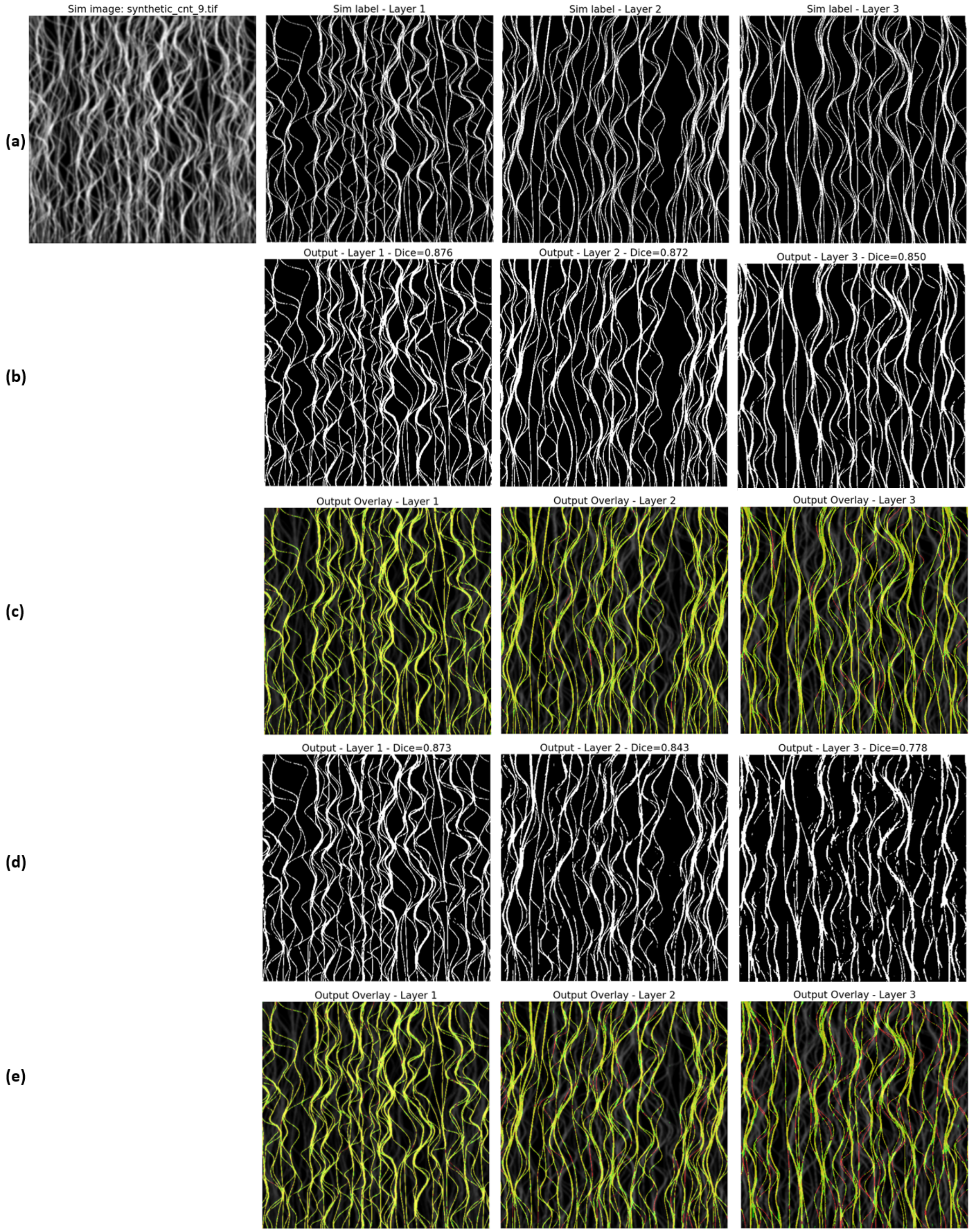

Figure 4. Depth layer decomposition and CNT segmentation results on multi-layer, physics-based, synthetic images (**without FDA style transfer**). Row (a) exhibits the synthetic image, and its segmentation ground truth masks for the three layers, extending from the front (layer 1) to the back (layer 3). Row (b) and (d) illustrate the segmentation outcomes for CNT-NeRF and U-net, respectively. Row (c) and (e) highlight the segmentation results (in green) of CNT-NeRF and U-net superimposed on the segmentation masks (in red) and the synthetic image (in grayscale). The yellow regions represent the intersections between the segmentation outcomes and the segmentation ground truth masks.

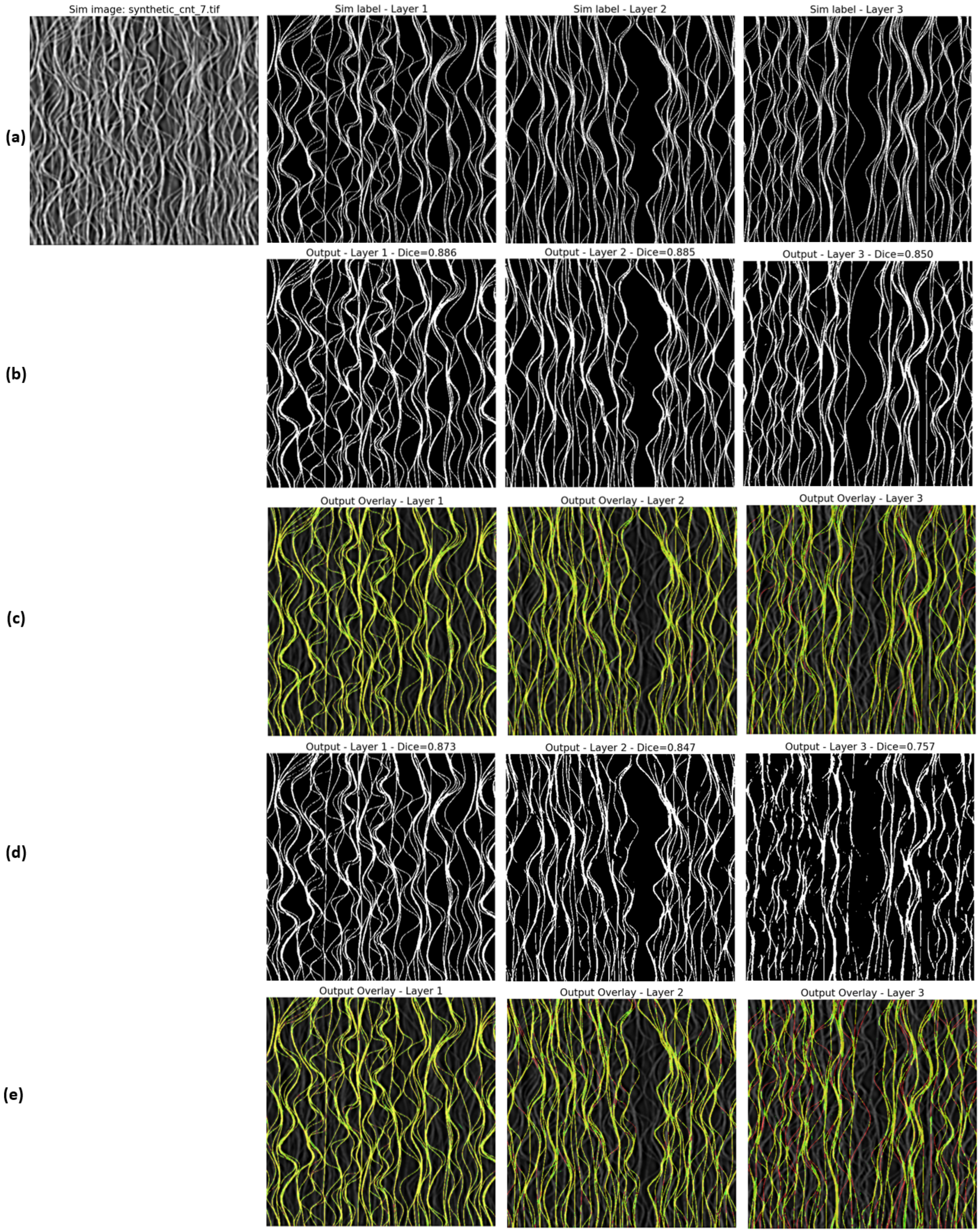

Figure 5. Depth layer decomposition and CNT segmentation results on multi-layer, physics-based, synthetic images (**with style transfer from a real SEM image of CNT forest**). Row (a) shows the synthetic image where the background is derived from a real image, and its segmentation masks for the three layers, spanning from the front (layer 1) to the back (layer 3), , The segmentation outcomes for CNT-NeRF and U-net are depicted in row (b) and (d), respectively. Row (c) and (e) showcase the segmentation outcomes (in green) of CNT-NeRF and U-net overlaid on the segmentation masks (in red) and the synthetic image (in grayscale). The yellow areas illustrate the intersections between the segmentation outcomes and the segmentation ground truth masks.

[3] Anyuan Cao, Pamela L Dickrell, W Gregory Sawyer, Mehrdad N Ghasemi-Nejhad, and Pulickel M Ajayan. Super-compressible foamlike carbon nanotube films. *Science*, 310(5752):1307–1310, 2005. 1

[4] Nicolas Carion, Francisco Massa, Gabriel Synnaeve, Nicolas Usunier, Alexander Kirillov, and Sergey Zagoruyko. End-to-end object detection with transformers. In *Computer Vision–ECCV 2020: 16th European Conference, Glasgow, UK, August 23–28, 2020, Proceedings, Part I 16*, pages 213–229. Springer, 2020. 3

[5] Rachel Carter, Benjamin Davis, Landon Oakes, Matthew R Maschmann, and Cary L Pint. A high areal capacity lithium–sulfur battery cathode prepared by site-selective vapor infiltration of hierarchical carbon nanotube arrays. *Nanoscale*, 9(39):15018–15026, 2017. 1

[6] Baratunde A Cola, Jun Xu, Changrui Cheng, Xianfan Xu, Timothy S Fisher, and Hanping Hu. Photoacoustic characterization of carbon nanotube array thermal interfaces. *Journal of applied physics*, 101(5):054313, 2007. 1

[7] Baratunde A Cola, Xianfan Xu, and Timothy S Fisher. Increased real contact in thermal interfaces: A carbon nanotube/foil material. *Applied physics letters*, 90(9):093513, 2007. 1

[8] Benjamin F Davis, Xingyi Yan, Nitin Muralidharan, Landon Oakes, Cary L Pint, and Matthew R Maschmann. Electrically conductive hierarchical carbon nanotube networks with tunable mechanical response. *ACS Applied Materials & Interfaces*, 8(41):28004–28011, 2016. 1

[9] Michael F. L. De Volder, Sameh H. Tawfick, Ray H. Baughman, and A. John Hart. Carbon Nanotubes: Present and Future Commercial Applications. *Science*, 339(6119):535–539, Feb. 2013. 1

[10] David Eigen, Christian Puhrsch, and Rob Fergus. Depth map prediction from a single image using a multi-scale deep network. *Advances in neural information processing systems*, 27, 2014. 4

[11] Clement Godard, Oisin Mac Aodha, Michael Firman, and Gabriel Brostow. *Digging Into Self-Supervised Monocular Depth Estimation*. IEEE Computer Society, Oct. 2019. 3

[12] C. Gommes, S. Blacher, K. Masenelli-Varlot, Ch. Bossuot, E. McRae, A. Fonseca, J.-B. Nagy, and J.-P. Pirard. Image analysis characterization of multi-walled carbon nanotubes. *Carbon*, 41(13):2561–2572, Jan. 2003. 2

[13] Taher Hajilounezhad, Rina Bao, Kannappan Palaniappan, Filiz Bunyak, Prasad Calyam, and Matthew R Maschmann. Predicting carbon nanotube forest attributes and mechanical properties using simulated images and deep learning. *npj Computational Materials*, 7(1):1–11, 2021. 2, 5

[14] Taher Hajilounezhad, Zakariya A Oraibi, Ramakrishna Surya, Filiz Bunyak, Matthew R Maschmann, Prasad Calyam, and Kannappan Palaniappan. Exploration of carbon nanotube forest synthesis-structure relationships using physics-based simulation and machine learning. In *IEEE Applied Imagery Pattern Recognition Workshop (AIPR)*, pages 1–8, 2019. 2

[15] K. He, X. Zhang, S. Ren, and J. Sun. Deep residual learning for image recognition. In *2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 770–778, 2016. 3, 4, 5

[16] Ryan Hines, Taher Hajilounezhad, Cole Love-Baker, Gordon Koerner, and Matthew R Maschmann. Growth and mechanics of heterogeneous, 3d carbon nanotube forest microstructures formed by sequential selective-area synthesis. *ACS applied materials & interfaces*, 12(15):17893–17900, 2020. 1

[17] Sumio Iijima. Helical microtubules of graphitic carbon. *Nature*, 354(6348):56–58, Nov. 1991. 1

[18] Sumio Iijima. Carbon nanotubes: past, present, and future. *Physica B*, 323(1):1–5, Oct. 2002. 1

[19] Yeonsu Jung, Young Shik Cho, Jae Won Lee, Jun Young Oh, and Chong Rae Park. How can we make carbon nanotube yarn stronger? *Composites Science and Technology*, 166:95–108, 2018. 1

[20] Gordon Koerner, Ramakrishna Surya, Kannappan Palaniappan, Prasad Calyam, Filiz Bunyak, and Matthew R Maschmann. In-situ scanning electron microscope chemical vapor deposition as a platform for nanomanufacturing insights. In *ASME International Mechanical Engineering Congress and Exposition*, volume 85567, page V02BT02A052, 2021. 5

[21] Jiaxin Li, Zijian Feng, Qi She, Henghui Ding, Changhu Wang, and Gim Hee Lee. Mine: Towards continuous depth mpi with nerf for novel view synthesis. *2021 IEEE/CVF International Conference on Computer Vision (ICCV)*, pages 12558–12568, 2021. 2, 3

[22] Matthew R Maschmann. Integrated simulation of active carbon nanotube forest growth and mechanical compression. *Carbon*, 86:26–37, 2015. 1

[23] Matthew R Maschmann, Ben Dickinson, Gregory J Ehlert, and Jeffery W Baur. Force sensitive carbon nanotube arrays for biologically inspired airflow sensing. *Smart Materials and Structures*, 21(9):094024, 2012. 1

[24] Matthew R Maschmann, Gregory J Ehlert, Benjamin T Dickinson, David M Phillips, Cody W Ray, Greg W Reich, and Jeffery W Baur. Bioinspired carbon nanotube fuzzy fiber hair sensor for air-flow detection. *Advanced Materials*, 26(20):3230–3234, 2014. 1

[25] Matthew R Maschmann, Gregory J Ehlert, Sei Jin Park, David Mollenhauer, Benji Maruyama, A John Hart, and Jeffery W Baur. Visualizing strain evolution and coordinated buckling within cnt arrays by in situ digital image correlation. *Advanced Functional Materials*, 22(22):4686–4695, 2012. 1

[26] Matthew R Maschmann, Qiuhong Zhang, Feng Du, Liming Dai, and Jeffery Baur. Length dependent foam-like mechanical response of axially indented vertically oriented carbon nanotube arrays. *Carbon*, 49(2):386–397, 2011. 1

[27] Matthew R Maschmann, Qiuhong Zhang, Robert Wheeler, Feng Du, Liming Dai, and Jeffery Baur. In situ sem observation of column-like and foam-like cnt array nanoindentation. *ACS applied materials & interfaces*, 3(3):648–653, 2011. 1

[28] Ben Mildenhall, Pratul P. Srinivasan, Matthew Tancik, Jonathan T. Barron, Ravi Ramamoorthi, and Ren Ng. Nerf: Representing scenes as neural radiance fields for view synthesis. In *Proceedings of the European Conference on Computer Vision (ECCV)*, 2020. 2

[29] Nguyen P. Nguyen, Ramakrishna Surya, Matthew Maschmann, Prasad Calyam, Kannappan Palaniappan, and Filiz Bunyak. Self-supervised Orientation-Guided Deep Network for Segmentation of Carbon Nanotubes in SEM Imagery. In *Computer Vision – ECCV 2022 Workshops*, pages 412–428. Springer, Cham, Switzerland, Feb. 2023. 2

[30] Myounggu Park, Baratunde A Cola, Thomas Siegmund, Jun Xu, Matthew R Maschmann, Timothy S Fisher, and Hyonny Kim. Effects of a carbon nanotube layer on electrical contact resistance between copper substrates. *Nanotechnology*, 17(9):2294, 2006. 1

[31] Siddhartha Pathak, Nisha Mohan, Elizabeth Decolvenaere, Alan Needleman, Mostafa Bedewy, A John Hart, and Julia R Greer. Local relative density modulates failure and strength in vertically aligned carbon nanotubes. *ACS nano*, 7(10):8593–8604, 2013. 1

[32] Olaf Ronneberger, Philipp Fischer, and Thomas Brox. U-Net: Convolutional Networks for Biomedical Image Segmentation. In *Medical Image Computing and Computer-Assisted Intervention – MICCAI 2015*, pages 234–241. Springer, Cham, Switzerland, Nov 2015. 5

[33] Matthew Tancik, Pratul Srinivasan, Ben Mildenhall, Sara Fridovich-Keil, Nithin Raghavan, Utkarsh Singhal, Ravi Ramamoorthi, Jonathan Barron, and Ren Ng. Fourier features let networks learn high frequency functions in low dimensional domains. *Advances in Neural Information Processing Systems*, 33:7537–7547, 2020. 3

[34] Sameh Tawfick, Zhouzhou Zhao, Matthew Maschmann, Anna Brieland-Shoultz, Michael De Volder, Jeffery W Baur, Wei Lu, and A John Hart. Mechanics of capillary forming of aligned carbon nanotube assemblies. *Langmuir*, 29(17):5190–5198, 2013. 1

[35] María Celeste Ramírez Trujillo, Teresa E. Alarcón, Oscar S. Dalmau, and Adalberto Zamudio Ojeda. Segmentation of carbon nanotube images through an artificial neural network. *Soft Comput.*, 21(3):611–625, Feb. 2017. 2

[36] Richard Tucker and Noah Snavely. Single-view view synthesis with multiplane images. In *CVPR*, 2020. 2

[37] Ashish Vaswani, Noam Shazeer, Niki Parmar, Jakob Uszkoreit, Llion Jones, Aidan N. Gomez, Łukasz Kaiser, and Illia Polosukhin. Attention is All you Need. *Advances in Neural Information Processing Systems*, 30, 2017. 3

[38] Tim Wortmann and Sergej Fatikow. Carbon nanotube detection by scanning electron microscopy. In *In Proc. of the Eleventh IAPR Conference on Machine Vision Applications (MVA'09*, 2009. 2

[39] Yanchao Yang and Stefano Soatto. FDA: Fourier Domain Adaptation for Semantic Segmentation, 2020. [Online; accessed 26. Mar. 2023]. 5

[40] A A Zbib, S Dj Mesarovic, E T Lilleodden, D McClain, J Jiao, and D F Bahr. The coordinated buckling of carbon nanotube turfs under uniform compression. *Nanotechnology*, 19(17):175704, mar 2008. 1

[41] Mei Zhang, Ken R Atkinson, and Ray H Baughman. Multifunctional carbon nanotube yarns by downsizing an ancient technology. *Science*, 306(5700):1358–1361, 2004. 1