

Semantic Segmentation on FDFD-generated Wideband Radar Images of Potential Shooters

Mahshid Asri, Ann Morgenthaler, and Carey M. Rappaport

Department of Electrical Engineering
Northeastern University, Boston, MA 02115, USA
asri.m@northeastern.edu, a.morgenthaler@northeastern.edu, rappaport@coe.neu.edu

Abstract – This paper presents a deep learning model for fast and accurate radar detection and pixel-level localization of large concealed metallic weapons on pedestrians walking along a sidewalk. The considered radar is stationary, with a multi-beam antenna operating at 30 GHz with 6 GHz bandwidth. A large modeled data set has been generated by running 2155 2D-FDFD simulations of torso cross sections of persons walking toward the radar in various scenarios.

Index Terms – Concealed object, deep learning, millimeter-wave radar, object detection, semantic segmentation, U-Net.

I. INTRODUCTION

Developing a solution to address threats to soft targets and crowded spaces such as schools, stadiums, hospitals, train stations and places of worship is a complex challenge. Soft targets are civilian sites where a lot of unarmed people gather and can be vulnerable to active shooters. SENTRY (Soft-target Engineering to Neutralize the Threat Reality) is a DHS Center of Excellence that addresses the challenges of protecting soft targets and crowded places (STCPs). The ideal solution is a semi-autonomous procedure that keeps the human in the loop and takes advantage of an integration of sensor data to create a cost effective pipeline to support and help decision makers and first responders to detect, deter, and mitigate threats. Such system can be helpful in preventing a scenario like Boston marathon bombing from happening.

This paper discusses the role of radar in SENTRY Advanced Sensor Concept projects, and provides a deep learning-based real-time solution for detecting potential threats and shooters among large crowds in one specific scenario. A radar system has been designed to monitor the pedestrians walking along the sidewalk for large metallic weapons such as guns and knives. A large data set of radar images has been generated using Finite Difference Frequency Domain (FDFD) simulations and a 2D U-Net model has been trained to perform the image

segmentation task and predict pixel-level masks of concealed weapons from the FDFD-generated radar image.

Image segmentation is an important computer vision task that has many key applications in scene understanding [1, 2], robotic perception [3], video surveillance [4], medical image analysis [5, 6], augmented reality [7], image compression [8], human-computer interface [9, 10], satellite imagery analysis [11], self-driving vehicles [12, 13], and pedestrian detection [14, 15]. The goal of image segmentation task is to provide a pixel-level label of the image and predict the category of each pixel. Instance segmentation is an extension of semantic segmentation which tries to detect and delineate each object in the image [16, 17]. The categories in this paper are human body, anomaly or metallic weapon, and the background image.

In recent years, deep learning has been widely used in various applications due to its capability in handling large amounts of data accurately. Fully convolutional models (FCMs) [18], CNN with graphical models [19–21], encoder-decoder based models [22–25], multiscale and pyramid network based models [26], R-CNN based models [27, 28], Dilated convolutional models [29–32], RNN models [33–37], attention-based models [38–40], generative models and adversarial training [41–43], and CNN models with active contour [44–46] are among famous deep learning models used for the image segmentation task.

Encoder-decoders are a family of deep learning models that can learn to map data points from one domain to another using a two-stage network. The first stage is used to capture the context of the image and the second stage provides object localization. These models are useful in several image to image translations such as image deblurring or super resolution, and image segmentation [16]. The U-Net model developed by Ronneberger et al. [47] is a famous encoder-decoder model that is mainly used for analyzing medical images. Similar to medical images of biological samples, CT and MRI scans, radar images are very different from the typical RGB images that most deep learning models are trained

on. Because of this similarity, U-Net models are good candidates for analyzing radar images.

Multiple studies have used deep-learning to detect pedestrians [48]. In most of these studies, data sets contain RGB images captured by cameras. In [49], a random forest method has been used to detect pedestrians, using surveillance camera information. In [50], two-stream deep convolutional neural networks have been trained to learn multispectral human-related features under different illumination conditions. There also have been various studies where radar data has been used to detect pedestrians [51, 52]. In [53], Fast-RCNN has been applied to NuScenes dataset [54] to detect humans on input images. In [55], radar cross sections along with RGB camera images have been used for pedestrian detection. There also have been some studies where authors have used deep learning models to detect people as well as weapons. For example in [56] researchers have developed single-shot and multi-shot prediction networks to detect anomalies like laptops, phones, and knives based on the captured data from a commercial radar working at 77 GHz.

Many of these studies have focused on detection of human subjects and only a few have looked at detecting body-worn anomalies. Only some are using radar signals as the input and all of them are providing a bounding box of the detected human body and anomaly. This paper takes advantage of image segmentation to find body-worn anomalies which is by definition a deeper task than classification (detection). The trained model is capable of detecting anomalies on pedestrians as well as providing a pixel-level mask that shows the location and general shape and size of the anomaly.

This paper presents the results of training a 2D U-Net model on FDFD-generated radar scattering of cross sections of human torsos parallel to the ground to detect potential threats among pedestrians walking along sidewalks and localize the concealed weapon. It is worth mentioning that the simulated radar images in this paper have been generated by a minimum radar imaging system that uses only three receivers and one transmitter. Because of the location of radar with respect to passengers, only parts of the human torsos are visible on reconstructed radar images, which makes it harder to perform the image segmentation task.

This research proposes a novel approach in detecting body-worn anomalies using radar, a technique not widely explored in existing research and introduces a deep learning model that performs pixel-level segmentation on radar images. These images are generated using a minimal radar system, consisting of just three receivers and one transmitter, which captures only partial views of human torsos. This partial view significantly complicates the image segmentation task, making

the challenge of localizing concealed weapons more difficult. By overcoming this limitation, this paper offers a groundbreaking solution for real-time, high-accuracy threat detection in crowded, soft-target environments. This approach enables more precise identification and localization of potential threats, enhancing decision-making and threat mitigation capabilities in public safety scenarios.

The remainder of the paper is structured as follows. Section II discusses the details of FDFD simulations used to generate the radar image data set, and section III goes over the process of creating the radar data set from FDFD simulations. Section IV talks about the training process, evaluation metrics, and illustrates the segmentation results. Finally section V provides a brief summary of the paper.

II. FDFD SIMULATIONS

The data set used for training the U-Net model has been generated by running 2155 FDFD simulations of human torso in various scenarios. The considered imaging system is a 30 ± 3 GHz TX/RX radar designed to monitor pedestrians walking along a sidewalk for large, concealed metal objects. The TX is modeled by a uniform aperture of length 37.5 cm and there are also three collinear RX apertures, each of length 50 cm so this system can illuminate and image only the 40 cm width of a torso at 30 m range. The TX/RX antenna system gathers scattering signals for a given pedestrian as he moves toward the stationary radar and the resulting TX/RX signals processed using the inverse synthetic aperture radar (ISAR) technique to generate torso images with bright spots in the images corresponding to areas with strong reflectivity. Figure 1 illustrates the imaging setup.

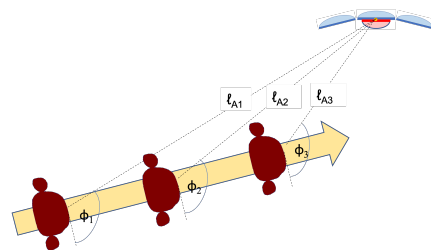


Fig. 1. Imaging setup, showing various instances of a torso cross section moving towards a fixed radar antenna.

The radar is installed on a light pole and takes sixteen snapshots of the pedestrian walking towards it at sixteen frequency steps between 27 and 33 GHz. Three receive focal points are considered in each simulation: $x = 0$ m which is approximately the middle of the torso, $x = \pm 0.25$ m which correspond to approximate location of left and right arms.

Each of the 768 simulations (16 positions $\times$ 16 frequencies $\times$ 3 receive focal points) used to create these images is performed by analytically generating the TX aperture (source) fields in a computational region surrounding the torso, computing the scattered electromagnetic fields in that region using the 2D FDFD algorithm assuming no z variation ($d/dz = 0$), and then finding the RX fields far outside the computational region using the Stratton-Chu near-to-far field transform [57, 58]. This near-to-far field transform requires computing both electric and magnetic transverse scattered fields and the 2D scalar Green's function on a rectangular bounding box enclosing the torso. The bounding box is centered in the FDFD computational region and is about 70% of its size; the four computational boundaries are each terminated by a PML [59].

Accurate FDFD simulations of scattering from human torsos at frequencies above 30 GHz requires significant computational storage, even for 2D calculations. The grid size for FDFD is defined by equation 1:

$$h = \lambda_{skin}/ppw, \quad (1)$$

where the number of points per wavelength (ppw) is typically greater than about 16. The grid size is determined by the shortest wavelength in the computational grid, which in these simulations is the wavelength of skin (the total electric and magnetic fields within metal regions are always zero and so are not directly computed). For skin with complex dielectric of $20 + j 16$ at 16 ppw and 30 GHz [60], the grid size will be $h = 0.13$ mm, which is computationally expensive to simulate.

In this radar system, the goal is to distinguish torsos with metal objects (e.g. guns) attached from non-threatening (non-metal containing) torsos. Modeling skin as metal is therefore counterproductive. But because skin is quite lossy, radar fields attenuate quickly as they penetrate the torso (within about a wavelength). Using the coarser grid spacing of 16 ppw in air (rather than 16 ppw in skin) does introduce a small amount of error in the FDFD simulations because within the skin region, the coarse grid spacing is now about 3.3 ppw. However, skin and metal responses are distinguishable and the coarse field plots closely resemble the fine (expensive) field plots for a range of test cases.

III. DATA SET

FDFD simulations were performed by MATLAB and were run on CPUs. On average each simulation took about 15 hours. In total, 53 torso shapes of various sizes (including both male and female) were simulated. The data set includes 198 cases of rotated torsos at various angles as well as some cases with two bodies in one image. Figure 2 shows a simulated torso shape as well as the metallic anomaly attached to it.



Fig. 2. Cross section of simulated torso with arms and metallic anomaly (yellow).

The data set includes torsos with and without metallic anomalies. The anomalies are randomly located at different parts of torso contours. The data set also includes various cases with multiple objects attached to the torso(s). Although the attached anomalies in most simulations were circular, other shapes such as rectangle, square, and barrel guns (a rectangle with a half circle attached to it) were considered in generating the data set. Table 1 illustrates the dimensions of each anomaly.

Since three receive focal points have been considered for the radar system, three reconstructed radar images can be produced by applying the appropriate relative phase for the three antenna apertures for rudimentary SAR processing of simulated files. Considering the ground truth file of Fig. 2, when radar is focusing on $x = -0.25$ m, radar image of Fig. 3 (a) is generated where the left arm shows high intensity. And when radar illuminates $x = 0$ m, and $x = 0.25$ m, reconstructed images with brighter response for torso (Fig. 3 (b)), and right arm (Fig. 3 (c)) are achieved.

By combining these three images in a complex form, Fig. 3 (d) will be generated. Since these images are in grey scale format and only have amplitude, each of them can be assumed as a channel in an RGB image. Here, the complex combined reconstructed image, the reconstructed image illuminating torso and right arm have been considered the red, green and blue channels in a colored image. Figure 4 (a) shows the resulting RGB image. Since the radar is located at subject's right side, the left arm is sometimes blocked by the presence of

Table 1: Canonical shape anomalies representing large weapon cross sections

Anomaly's Shape	Dimensions
Circle	Radius = 12 mm
Small circle	Radius = 10 mm
Rectangle	5×13 mm
Square	15×15 mm
Barrel gun	Rectangle (13×30 mm) Half circle (radius = 17 mm)



Fig. 3. Reconstructions of torso and arm cross section for three focusing cases and their combination: (a) focus on left arm, (b) focus on middle of torso, (c) focus on right arm and (d) complex combination of three images.



Fig. 4. The processed radar image of Fig. 2 along with its corresponding masks fed into the deep learning model for training: (a) RGB image, (b) body mask, (c) anomaly mask.

torso, making the reconstructed image of the left arm weaker and less reliable. By considering the complex combined image instead of the left arm image, the significance of torso and right arm are intensified while reducing the emphasis of the left arm in inputs.

The typical torso and anomaly image of Fig. 2 is used to produce the body and anomaly ground truth masks. For training the deep learning model, the RGB image is considered the input and both body and anomaly masks are considered labels (outputs). Figure 4 illustrates an example data point used in training the U-Net model.

The 2155 FDFD simulations which correspond to 8620 grey-scale radar images, are divided into training, validation and test sets. The training set includes 1775 FDFD simulations (7100 grey-scale images), and the validation and testing sets each include 190 FDFD simulations (760 grey-scale images).

IV. TRAINING AND EVALUATION

The evaluation metrics used to examine the predictability of the trained model are the F1-score, and the Intersection over Union (IoU). The F1-score can be calculated using equation (2).

$$F1 = 2 \times \frac{\text{precision} \times \text{recall}}{(\text{precision} + \text{recall})}, \quad (2)$$

where $\text{precision} = TP/(TP + FP)$, $\text{recall} = TP/(TP + FN)$, and TP, FP and FN are true positive, false positive, and false negative cases respectively. Additionally the IoU metric can be obtained using equation (3).

$$IoU = \frac{\text{area of overlap}}{\text{area of union}} = \frac{TP}{TP + FP + FN}, \quad (3)$$

U-Net is one of the popular techniques used in semantic segmentation and various extensions of it have been used in medical image analysis. In this paper, efficientnetb3 [61] has been used as the backbone of U-Net

Table 2: Model's performance

Metric	Training	Validation	Testing
F1 score	96%	86%	86%
IoU	93%	81%	81%

with about 17 million trainable parameters. Categorical focal loss has been used as the model's loss function.

The reconstructed radar images and the RGB image are 128×256 in size. The U-Net model is trained for 70 epochs with learning rate of 0.001. Each epoch took about 20 minutes to run on V100 GPU. Table 2 shows the final F1 and IoU score on training, validation and testing sets.

Figure 5 illustrates some of the test cases and the predicted torso and anomaly masks. The first column shows the input RGB images. The second and fourth columns correspond to the ground truth masks considered for each FDFD simulation and finally the third and fifth columns show the predicted masks for the human torso and the attached metallic anomaly. As Fig. 5 shows, the predicted and ground truth masks are very similar and the model can accurately estimate the location and shape of the torso and the body-worn anomaly based on the available RGB image. In case the subject is not carrying a metallic weapon, the predicted anomaly mask will be an empty image (Fig. 5 (e)). Since most anomalies in the data set are circular metals, the predicted anomalies tend to have curved edges. The images show that the trained model is capable of predicting the correct body and anomaly masks for both male (Fig. 5 (a)) and female (Fig. 5 (b)) torso shapes, various torso sizes (Fig. 5 (c)). Moreover, the model can predict multiple anomalies and multiple torsos in the image (Fig. 5 (d)).

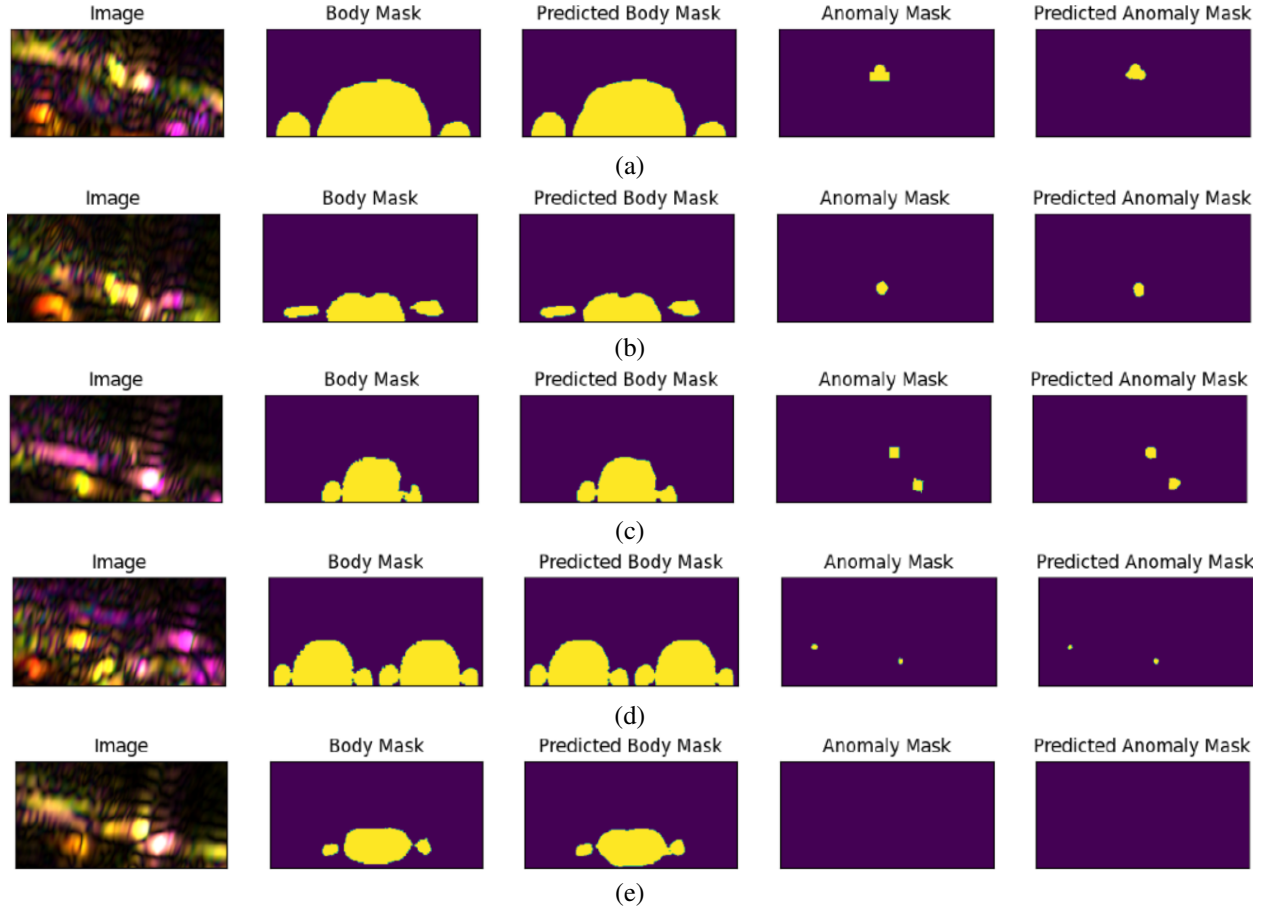


Fig. 5. Segmentation results on five typical test cases showing the complex combined reconstructed image in the first column, and in the successive columns, the true body geometry, AI predicted body geometry, the true and the AI predicted anomaly: (a) large male torso and a barrel anomaly, (b) female torso and a large circular anomaly, (c) small female torso and multiple anomalies, (d) two small male torsos and multiple anomalies, (e) small male torso with no anomaly.

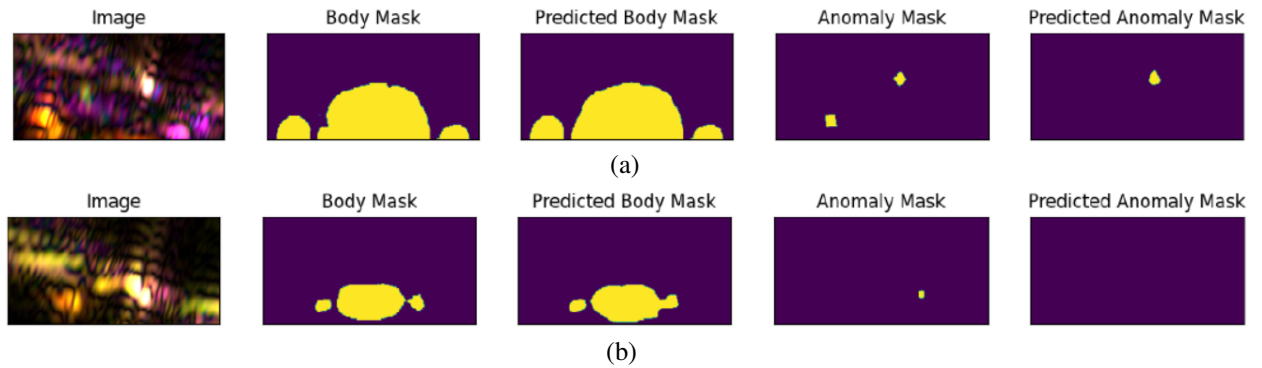


Fig. 6. Examples of undetected anomalies from the test set: (a) left object is partially inside the body mask, (b) object located between the torso and right arm.

Figure 6 shows two examples where the model could not detect the anomalies correctly. In general there were two common features among undetected cases. If any of

the anomalies are partially inside the body (similar to the left anomaly on Fig. 6 (a)), the model is less likely to detect it. Also, if the anomaly is located between one of

the arms and the torso (similar to the case shown in Fig. 6 (b)), the model is more prone to miss it. Out of 190 test cases, 51 cases had two or three anomalies present. The trained model was able to detect 39 cases with two or three anomalies correctly but could not detect all the anomalies on the remaining 12 cases. However, in all these 12 cases, the model was able to detect at least one anomaly on the image and did not miss the potential shooter. The remaining 139 cases included zero or one anomaly. Out of these 139 cases, the model predicted the masks correctly for 126 cases corresponding to 90%.

V. SUMMARY AND CONCLUSION

This paper presented the results of training a U-Net model with the backbone of efficientnetb3 for detection and localization of large concealed metallic objects on pedestrians. A large data set was generated by running FDFD simulations of a person walking toward a 6 GHz bandwidth radar operating at 30 GHz. The body and anomaly mask prediction results are very promising and the model can be used to detect and localize anomalies accurately in real time.

ACKNOWLEDGMENT

This work is supported by the U.S. Department of Homeland Security, Science and Technology Directorate, Office of University Programs, under Grant Award 22STES00001-02-00. The views and conclusions contained herein are those of the authors and should not be interpreted as necessarily representing the official policies, either expressed or implied, of the DHS (Corresponding author: Mahshid Asri)

REFERENCES

- [1] Z. Guo, Y. Huang, X. Hu, H. Wei, and B. Zhao, "A survey on deep learning based approaches for scene understanding in autonomous driving," *Electronics*, vol. 10, no. 4, p. 471, 2021.
- [2] K. Muhammad, T. Hussain, H. Ullah, J. Del Ser, M. Rezaei, N. Kumar, M. Hijji, P. Bellavista, and V. H. C. de Albuquerque, "Vision-based semantic segmentation in scene understanding for autonomous driving: Recent achievements, challenges, and outlooks," *IEEE Transactions on Intelligent Transportation Systems*, vol. 23, no. 12, pp. 22694-22715, 2022.
- [3] Q. M. Rahman, P. Corke, and F. Dayoub, "Run-time monitoring of machine learning for robotic perception: A survey of emerging trends," *IEEE Access*, vol. 9, pp. 20067-20075, 2021.
- [4] X. Cao, S. Gao, L. Chen, and Y. Wang, "Ship recognition method combined with image segmentation and deep learning feature extraction in video surveillance," *Multimedia Tools and Applications*, vol. 79, pp. 9177-9192, 2020.
- [5] Z. Wang, L. Wei, L. Wang, Y. Gao, W. Chen, and D. Shen, "Hierarchical vertex regression-based segmentation of head and neck CT images for radiotherapy planning," *IEEE Transactions on Image Processing*, vol. 27, no. 2, pp. 923-937, 2017.
- [6] X. Zhu, H.-I. Suk, S.-W. Lee, and D. Shen, "Subspace regularized sparse multitask learning for multiclass neurodegenerative disease identification," *IEEE Transactions on Biomedical Engineering*, vol. 63, no. 3, pp. 607-618, 2015.
- [7] L. Tanzi, P. Piazzolla, F. Porpiglia, and E. Vezzetti, "Real-time deep learning semantic segmentation during intra-operative surgery for 3D augmented reality assistance," *International Journal of Computer Assisted Radiology and Surgery*, vol. 16, no. 9, pp. 1435-1445, 2021.
- [8] M. Akbari, J. Liang, and J. Han, "DSSLIC: Deep semantic segmentation-based layered image compression," in *ICASSP 2019 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, pp. 2042-2046, 2019.
- [9] N. Sebe, M. S. Lew, and T. S. Huang, "The state-of-the-art in human-computer interaction," in *Computer Vision in Human-Computer Interaction: ECCV 2004 Workshop on HCI, Prague, Czech Republic, May 16, 2004. Proceedings*, pp. 1-6, 2004.
- [10] M. Lew, E. M. Bakker, N. Sebe, and T. S. Huang, "Human-computer intelligent interaction: A survey," in *Human-Computer Interaction: IEEE International Workshop, HCI 2007 Rio de Janeiro, Brazil, October 20, 2007 Proceedings 4*, pp. 1-5, 2007.
- [11] P. Yu, A. K. Qin, and D. A. Clausi, "Unsupervised polarimetric SAR image segmentation and classification using region growing with edge penalty," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 50, no. 4, pp. 1302-1317, 2011.
- [12] B. Li, S. Liu, W. Xu, and W. Qiu, "Real-time object detection and semantic segmentation for autonomous driving," in *MIPPR 2017: Automatic Target Recognition and Navigation*, vol. 10608, pp. 167-174, 2018.
- [13] Y.-H. Tseng and S.-S. Jan, "Combination of computer vision detection and segmentation for autonomous driving," in *2018 IEEE/ION Position, Location and Navigation Symposium (PLANS)*, pp. 1047-1052, 2018.
- [14] F. Flohr and D. Gavrilu, "PedCut: An iterative framework for pedestrian segmentation combining shape models and multiple data cues," in *BMVC*, 2013.
- [15] G. Brazil, X. Yin, and X. Liu, "Illuminating pedestrians via simultaneous detection & segmentation,"

- in *Proceedings of the IEEE International Conference on Computer Vision*, pp. 4950-4959, 2017.
- [16] S. Minaee, Y. Boykov, F. Porikli, A. Plaza, N. Kehtarnavaz, and D. Terzopoulos, "Image segmentation using deep learning: A survey," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 44, no. 7, pp. 3523-3542, 2021.
 - [17] M. Asri, R. Chowdhury, A. Care, D. F. Lamptey, A. Morgenthaler, O. Camps, and C. M. Rappaport, "Pixel-wise localization of concealed objects on millimeter-wave radar images using deep learning," *IEEE Transactions on Radar Systems*, vol. 2, pp. 1027-1035, 2024.
 - [18] 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*, pp. 3431-3440, 2015.
 - [19] 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.
 - [20] G. Lin, C. Shen, A. Van Den Hengel, and I. Reid, "Efficient piecewise training of deep structured models for semantic segmentation," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pp. 3194-3203, 2016.
 - [21] Z. Liu, X. Li, P. Luo, C.-C. Loy, and X. Tang, "Semantic image segmentation via deep parsing network," in *Proceedings of the IEEE International Conference on Computer Vision*, pp. 1377-1385, 2015.
 - [22] H. Noh, S. Hong, and B. Han, "Learning deconvolution network for semantic segmentation," in *Proceedings of the IEEE International Conference on Computer Vision*, pp. 1520-1528, 2015.
 - [23] 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*, vol. 39, no. 12, pp. 2481-2495, 2017.
 - [24] Y. Yuan, X. Chen, X. Chen, and J. Wang, "Segmentation transformer: Object-contextual representations for semantic segmentation," *arXiv preprint arXiv:1909.11065*, 2019.
 - [25] X. Xia and B. Kulis, "W-net: A deep model for fully unsupervised image segmentation," *arXiv preprint arXiv:1711.08506*, 2017.
 - [26] T.-Y. Lin, P. Dollár, R. Girshick, K. He, B. Hariharan, and S. Belongie, "Feature pyramid networks for object detection," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pp. 2117-2125, 2017.
 - [27] S. Ren, K. He, R. Girshick, and J. Sun, "Faster r-cnn: Towards real-time object detection with region proposal networks," *Advances in Neural Information Processing Systems*, vol. 28, 2015.
 - [28] K. He, G. Gkioxari, P. Dollár, and R. Girshick, "Mask r-cnn," in *Proceedings of the IEEE International Conference on Computer Vision*, pp. 2961-2969, 2017.
 - [29] L.-C. Chen, G. Papandreou, I. Kokkinos, K. Murphy, and A. L. Yuille, "Deeplab: Semantic image segmentation with deep convolutional nets, atrous convolution, and fully connected crfs," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 40, no. 4, pp. 834-848, 2017.
 - [30] F. Yu and V. Koltun, "Multi-scale context aggregation by dilated convolutions," *arXiv preprint arXiv:1511.07122*, 2015.
 - [31] P. Wang, P. Chen, Y. Yuan, D. Liu, Z. Huang, X. Hou, and G. Cottrell, "Understanding convolution for semantic segmentation," in *2018 IEEE Winter Conference on Applications of Computer Vision (WACV)*, pp. 1451-1460, 2018.
 - [32] A. Paszke, A. Chaurasia, S. Kim, and E. Culurciello, "Enet: A deep neural network architecture for real-time semantic segmentation," *arXiv preprint arXiv:1606.02147*, 2016.
 - [33] F. Visin, M. Ciccone, A. Romero, K. Kastner, K. Cho, Y. Bengio, M. Matteucci, and A. Courville, "Reseg: A recurrent neural network-based model for semantic segmentation," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops*, pp. 41-48, 2016.
 - [34] W. Byeon, T. M. Breuel, F. Raue, and M. Liwicki, "Scene labeling with lstm recurrent neural networks," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pp. 3547-3555, 2015.
 - [35] X. Liang, X. Shen, J. Feng, L. Lin, and S. Yan, "Semantic object parsing with graph lstm," in *Computer Vision-ECCV 2016: 14th European Conference, Amsterdam, The Netherlands, October 11-14, 2016, Proceedings, Part I 14*, pp. 125-143, 2016.
 - [36] Y. Xiang and D. Fox, "DA-RNN: Semantic mapping with data associated recurrent neural networks," *arXiv preprint arXiv:1703.03098*, 2017.
 - [37] R. Hu, M. Rohrbach, and T. Darrell, "Segmentation from natural language expressions," in *Computer Vision-ECCV 2016: 14th European Conference, Amsterdam, The Netherlands, October 11-14, 2016, Proceedings, Part I 14*, pp. 108-124, 2016.
 - [38] L.-C. Chen, Y. Yang, J. Wang, W. Xu, and A. L. Yuille, "Attention to scale: Scale-aware

- semantic image segmentation,” in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pp. 3640-3649, 2016.
- [39] Q. Huang, C. Xia, C. Wu, S. Li, Y. Wang, Y. Song, and C.-C. J. Kuo, “Semantic segmentation with reverse attention,” *arXiv preprint arXiv:1707.06426*, 2017.
- [40] H. Li, P. Xiong, J. An, and L. Wang, “Pyramid attention network for semantic segmentation,” *arXiv preprint arXiv:1805.10180*, 2018.
- [41] P. Luc, C. Couprie, S. Chintala, and J. Verbeek, “Semantic segmentation using adversarial networks,” *arXiv preprint arXiv:1611.08408*, 2016.
- [42] W.-C. Hung, Y.-H. Tsai, Y.-T. Liou, Y.-Y. Lin, and M.-H. Yang, “Adversarial learning for semi-supervised semantic segmentation,” *arXiv preprint arXiv:1802.07934*, 2018.
- [43] N. Souly, C. Spampinato, and M. Shah, “Semi supervised semantic segmentation using generative adversarial network,” in *Proceedings of the IEEE International Conference on Computer Vision*, pp. 5688-5696, 2017.
- [44] P. Marquez-Neila, L. Baumela, and L. Alvarez, “A morphological approach to curvature-based evolution of curves and surfaces,” *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 36, no. 1, pp. 2-17, 2013.
- [45] T. H. N. Le, K. G. Quach, K. Luu, C. N. Duong, and M. Savvides, “Reformulating level sets as deep recurrent neural network approach to semantic segmentation,” *IEEE Transactions on Image Processing*, vol. 27, no. 5, pp. 2393-2407, 2018.
- [46] C. Rupprecht, E. Huaroc, M. Baust, and N. Navab, “Deep active contours,” *arXiv preprint arXiv:1607.05074*, 2016.
- [47] O. Ronneberger, P. Fischer, and T. Brox, “U-net: Convolutional networks for biomedical image segmentation,” in *Medical Image Computing and Computer-Assisted Intervention—MICCAI 2015: 18th International Conference, Munich, Germany, October 5-9, 2015, Proceedings, Part III* 18, pp. 234-241, Springer, 2015.
- [48] K. Li, Y. Zhuang, J. Lai, and Y. Zeng, “PFYOLOv4: An improved small object pedestrian detection algorithm,” *IEEE Access*, vol. 11, pp. 17197-17206, 2023.
- [49] S. Kim, S. Kwak, and B. C. Ko, “Fast pedestrian detection in surveillance video based on soft target training of shallow random forest,” *IEEE Access*, vol. 7, pp. 12415-12426, 2019.
- [50] D. Guan, Y. Cao, J. Yang, Y. Cao, and M. Y. Yang, “Fusion of multispectral data through illumination-aware deep neural networks for pedestrian detection,” *Information Fusion*, vol. 50, pp. 148-157, 2019.
- [51] J. V. B. Severino, A. Zimmer, T. Brandmeier, and R. Z. Freire, “Pedestrian recognition using micro Doppler effects of radar signals based on machine learning and multi-objective optimization,” *Expert Systems with Applications*, vol. 136, pp. 304-315, 2019.
- [52] J. Castanheira, F. Curado, A. Tomé, and E. Gonçalves, “Machine learning methods for radar-based people detection and tracking,” in *Progress in Artificial Intelligence: 19th EPIA Conference on Artificial Intelligence, EPIA 2019, Vila Real, Portugal, September 3-6, 2019, Proceedings, Part I* 19, pp. 412-423, 2019.
- [53] R. Nabati and H. Qi, “Rrpn: Radar region proposal network for object detection in autonomous vehicles,” in *2019 IEEE International Conference on Image Processing (ICIP)*, pp. 3093-3097, 2019.
- [54] H. Caesar, V. Bankiti, A. H. Lang, S. Vora, V. E. Liong, Q. Xu, A. Krishnan, Y. Pan, G. Baldan, and O. Beijbom, “NuScenes: A multimodal dataset for autonomous driving,” in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 11621-11631, 2020.
- [55] H. Li, R. Liu, S. Wang, W. Jiang, and C. X. Lu, “Pedestrian liveness detection based on mmwave radar and camera fusion,” in *2022 19th Annual IEEE International Conference on Sensing, Communication, and Networking (SECON)*, pp. 262-270, 2022.
- [56] X. Gao, H. Liu, S. Roy, G. Xing, A. Alansari, and Y. Luo, “Learning to detect open carry and concealed object with 77 GHz radar,” *IEEE Journal of Selected Topics in Signal Processing*, vol. 16, no. 4, pp. 791-803, 2022.
- [57] J. A. Stratton, *Electromagnetic Theory*. Hoboken, NJ: John Wiley & Sons, 2007.
- [58] J. A. Kong, *Electromagnetic Wave Theory*. Hoboken, NJ: John Wiley & Sons, 1986.
- [59] J.-P. Berenger, “A perfectly matched layer for the absorption of electromagnetic waves,” *Journal of Computational Physics*, vol. 114, no. 2, pp. 185-200, 1994.
- [60] S. Gabriel, R. Lau, and C. Gabriel, “The dielectric properties of biological tissues: II. Measurements in the frequency range 10 Hz to 20 GHz,” *Physics in Medicine & Biology*, vol. 41, no. 11, p. 2251, 1996.
- [61] M. Tan and Q. Le, “Efficientnet: Rethinking model scaling for convolutional neural networks,” in *International Conference on Machine Learning*, pp. 6105-6114, PMLR, 2019.



Mahshid Asri received the B.S. in electrical engineering from Iran University of Science and Technology, Tehran, Iran, in 2017 and the MS and Ph.D. in electrical engineering from Northeastern University, Boston, MA, USA, in 2020 and 2023, respectively. She is currently working as a postdoctoral research associate at Northeastern University, Boston, MA, USA. Her work is focused on radar image processing and object characterization, and applying deep learning techniques to radar images for anomaly detection. She completed an internship at Metalenz in 2022.



Ann Morgenthaler received the B.A. in physics from Harvard University, Cambridge, MA, and the E.E. degree in electrical engineering and computer science from Massachusetts Institute of Technology (MIT), Cambridge.

She is an author of the textbook “Electromagnetic Waves” with David Staelin and Jin Au Kong. Her research interests include fast forward modeling techniques for ground-penetrating radar scattering from realistic objects in lossy media under rough ground surfaces.



Carey M. Rappaport (Fellow, IEEE) received the SB degree in mathematics and the SB, SM, and EE degrees in electrical engineering from the Massachusetts Institute of Technology (MIT), Cambridge, MA, USA, in 1982, and the Ph.D. degree in electrical engineering from MIT, in June 1987.

He worked as a Teaching and Research Assistant with MIT from 1981 to 1987, and during the summers at COMSAT Labs, Clarksburg, MD, USA, and The Aerospace Corporation, El Segundo, CA, USA. He joined the Faculty of Northeastern University, Boston, MA, USA, in 1987. He has been a Professor of electrical and computer engineering since July 2000. In 2011, he was appointed as the College of Engineering Distinguished Professor. During fall 1995, he was a Visiting Professor of electrical engineering with the Electromagnetics Institute of the Technical University of Denmark, Lyngby, Denmark, as a part of the W. Fulbright International Scholar Program. During the second half of 2005, he was a Visiting Research Scientist with the Commonwealth Scientific Industrial and Research Organization (CSIRO) in Epping, Australia. He has consulted for CACI, Alion Science and Technology, Inc., Geo-Centers, Inc., PPG, Inc., and several municipalities on wave propagation and modeling, and microwave heating and safety. He was PI of an ARO-sponsored Multi-disciplinary University Research Initiative on Humanitarian Demining, Co-PI of the NSF-sponsored Engineering Research Center for Subsurface Sensing and Imaging Systems (CenSSIS), and Co-Principal Investigator and Deputy Director of the DHS-Sponsored Awareness and Localization of Explosive Related Threats (ALERT) Center of Excellence. He has authored over 430 technical journal and conference paper in the areas of microwave antenna design, electromagnetic wave propagation and scattering computation, and bioelectromagnetics. He holds two reflector antenna patents, two biomedical device patents, and four subsurface sensing device patents.