A Comprehensive Review on Deep Learning Architectures for Image Segmentation

Madhurima Mukherjee
Adobe, India
Bengaluru, India
mukherjeemadhurima4@gmail.com

***Abstract*— At a micro level, image segmentation is an important computer vision task that assists in applications in various fields such as medical diagnostics, self-driving cars, and robotics. The introduction of deep learning in the scene led to great improvements in the performance of segmentation models in terms of accuracy, efficiency, and size. This paper presents a comprehensive literature review of recent advancements pertaining to the deep learning-based architectures for the image segmentation task. Such models include the well-known U-Net and Mask R-CNN models, as well as more recent models including Attention U-Net, segmentation using Generative Adversarial Networks (GANs) and transformer-based models including Vision Transformer (ViT) for segmentation tasks. We also include the analysis of hybrid architectures that utilize both CNN and transformer components for better performance in segmentation tasks. Other important issues such as data shortage and over-difficulty of annotation tasks are also considered, as well as the influence of data augmentation and few apologies on these problems. In addition to that, we also perform a comparative analysis of these architectures and evaluate and compare their performance on several benchmark datasets including their merits and demerits and usage. Finally, we identify current challenges and suggest promising future research directions which will include the development of more efficient models for real-time deployment and improvements in model interpretability and explainability.**

***Index Terms*—Attention U-Net, DeepLab, Image Segmentation, Mask R-CNN, U-Net**

I. INTRODUCTION

IMAGE segmentation is the process in computer vision where an image is split into multiple parts so that it is easier to analyze them separately. These parts are usually referred to as segments in imagery and frequently respond to distinct objects or other forms of interest in the image. Areas like medical image evaluation, autonomous cars, industrial and satellite imaging have been among the diverse areas of segmentation applications complementing. For example, in medicine, segmentation algorithms allow the detection of tumors, organs and other relevant spaces in medical scans, which are vital in diagnosis and treatment processes. Likewise, when a vehicle is in autonomous driving mode, it is important to accurately segment different targets in order for the self-driving car to interpret the environment and respond to pedestrians, other vehicles, and traffic signs.

As with many challenges in computer graphic, the first attempts addressing the problem of image segmentation employed manual labor and a number of heuristic operations, including thresholding, edge detection and region growing. Still, these techniques faced challenges during application in the real world since they dealt with complicated scenes with a lot of noise. Challenges were overcome with the advent of deep learning, more specifically with CNNs that allowed to automatically learn features directly from data automatically in an end-to-end manner. This evolution led to a marked enhancement in the precision and the efficiency of segmentation models.

One of the first works on deep learning-based segmentation is the Fully Convolutional Network (FCN) [1], where convolutional layers are put in place of the fully connected ones so pixel-level prediction becomes applicable. Following this trend, The U-Net [2], Mask R-CNN [3], and Attention U-Net [4] architectures advanced very much in terms of quality of the segmentation. These models are new but, however, have certain weaknesses, such as small object segmentation, class biasness, and larger number of annotated datasets. Recently developed models, for example, transformer-based ones, like Vision Transformer (ViT) [5] for segmentation seek to minimize such problems. ViT, developed for image classification and further working with segmentation tasks, achieved state-of-the-art performance due to long-range dependencies across an image. Moreover, the hybrid CNN-transformer models [6] have appeared that capitalize on both approaches for segmentation accuracy improvement. These hybrid models present a viable answer to the problems of global context representation and local feature extraction.

To tackle the challenge of the few labeled instances, data augmentation techniques, and few-shot learning approaches [7] are becoming popular in the segmentation based on deep learning. In addition to these approaches, recurrent and active learning strategies also assist to alleviate the problem of lack of annotated data and its big impact, especially for niche areas of applications like medical imaging.

In this paper, we analyze different architectures of deep learning for image segmentation in a greater depth. We start from a brief overview of classical methods, from where we go on with the advancement of deep learning algorithms. We then perform a thorough analysis of the performance of these models against benchmark datasets with the aim of exploring the pros and the cons of the models. Finally, we analyze the gaps in the literature and the problems that exist in the context,

such as lack of data, time and resources, as well as how underlying models can be interpreted and recommend the areas for further research.

II. Historical Context of Image Segmentation

Segmentation techniques used to analyze specific regions of interest in an image to obtain meaningful information have improved greatly through the decades. Deploying classical approaches such as edge detection, thresholding and region-based techniques were prevailing methods that preceded other models. Although these approaches provided the basic groundworks there were significant shortcomings that required new developments in the discipline.

A. Early Techniques

- Edge Detection: Edge detection was the first of the techniques that was employed in image segmentation, the main aim being to detect regions within images with distinct changes in intensity excluding any gradual transition between them. And in these, the Canny edge detector introduced by John F. Canny in 1986 became one more and more favored as it singled out edges with a lot of noise. The Canny detector devised a multi-stage process involving image smoothing, gradient finding, no maximum suppression, and hysteresis for edge tracking. But edge detection methods had difficulties in high noise and texture altered environments which high edges were vague or lost behind image variations like lighting or object texture variation [8]. These methods were employed with caution due to the sensitivity of the parameters employed although they did not prove effective in the analysis of complex realistic images particularly when the objects edges were indistinct.
- Thresholding: Apart from the other approaches mentioned above, another early approach was the thresholding technique which was based on dividing the image pixels into discrete categories depending on their intensities. A single global intensity value was used during simple global thresholding to segment the image into foreground and background regions. A strength of the techniques based on thresholding is that they are practically fast and efficient, however, some weaknesses are common especially in the cases of segmentation: over-segmentation or under-segmentation. Over-segmentation happened when the object was divided into too many parts which an algorithm considers separate and distinct, while under-segmentation was when too many regions were grouped into one because the algorithm considered too many separate regions to be the same. These problems arose because the thresholding region, either local or global used a specific intensity value or level without considering other imposing factors, shapes or texture of the object which affected the end result in complex images [9]
- Region-Based Methods: Region-based segmentation techniques attempted to address some of these limitations by grouping together some pixels depending on the similarity in the properties of those pixels. These methods worked by observing that there were regions have certain characteristics (such as uniform color) and these regions can be used to grow and subdivide the image into smaller regions or every area into smaller regions such as growing and region splitting. Even though these techniques were found to generate more coherent segments than the edge-based ones, they do have limitations with regards to the effects of texture and lighting variations as well as noise.

B. Emergence of Machine Learning

With the increased need for accuracy and compactness of segmentation approaches, the possibility of machine learning approaches was considered. After all, such algorithms can study the data and find their way in the most intricate and variable outlines.

- Support Vector Machines (SVMs): Among the first trends in the implementation of the machine learning paradigm for the image segmentation task was the application of Support Vector Machines (SVMs). SVMs are a kind of supervised models set to scan the dataset`s features and separate classes by the hyperplane which creates the largest margin between the classes. For example, in image segmentation, SVMs were employed in the determination of pixels as either foreground and background pixels or to separate other geometrically different areas of the same picture. Though the use of SVMs represented a better onto others more classical approaches, some problems such as the problem of working with wide and high dimensional data sets that needed more attention were present, but later the solutions to these challenges came through deep learning methods. [10].
- Random Forests: Within the tasks of image segmentation, random forests gained acceptance as a machine learning technique. Random forests are seed based learning techniques that log information from multiple decision trees in a bid to create a more accurate and robust model. For segmentation tasks, random forests were utilized on learned features to determine how all perform different pixel values in spatial context. While random forests did outperform SVMs (especially when it comes to large datasets), the method relied, as all other techniques of the time, on the hand-engineered features and used pre-defined decision trees in order to segment the data.

C. The Shift to Deep Learning

The emergence of deep learning represents the most significant leap in the field of image segmentation. Deep learning models, and especially Convolutional Neural Networks (CNNs), enabled researchers to perform automatic learning of hierarchical features from the data without the need for defining them first. CNN networks, which were modeled after the human brain, contained multiple layers of convolutions that facilitated the automatic recognition of local and global structures or patterns in images. This capacity to extract feature representation from data enabled CNNs to succeed over other previous techniques in almost all tasks, including image segmentation.

- Fully Convolutional Networks (FCNs): The emergence of Fully Convolutional Networks or FCNs in 2015 became a game changer in image segmentation. FCN architectures did not have any fully connected layers like traditional CNN's geared towards image classification but instead, embed – one or several – convolutional layers, allowing the network to produce a pixel-by-pixel or dense prediction map for every unit in the image. With the introduction of FCNs, it was established that simultaneous extraction and segmentation of features (end to end training) is better than the traditional methods with predefined features [11]. This was really a profound success as it showed that deep learning could perform complex segmentation tasks and that leverage improvements in accuracy and robustness. The FCN model was very effective in segmentation of images down to the pixel level as it allowed for the segmentation of more complex image regions accurately where this type of operations was needed.
- Evolution of Segmentation Architectures: Following the success of FCNs, more complex deep networks were being researched for segmentation purposes and these include U-Net, Mask R-CNN and DeepLab which improved the performance of previous models. The integration of novel techniques like skip connections, multi-scale feature extraction, and instance segmentation enhanced in some cases not just the performance but also the efficiency of image segmentation models. For instance, U-Net was able to solve the problem of losing spatial information especially in medical imaging by proposing an encoder-decoder structure with skip connections. The need for distinguishing between objects in images was solved by Mask R-CNN through the integration of the Faster R-CNN framework and a segmentation mask prediction branch.

In conclusion, the evolution of image segmentation techniques is also a reflection of the algorithms' advancement as well as an increasing segmentation task's difficulty. It is worth noting that segmentation studies were commenced with edge detection and thresholding methods but machine learning and deep learning have made a huge change in the field today. The development of CNN-based architectures has improved accuracy and efficiency in segmentation, paving the way for advanced applications in areas like medical imaging, autonomous driving, and remote sensing.

III. Current Deep Learning Architectures for Image Segmentation

A. U-Net

U-Net has proved to be one of the most popular architectures in medical image segmentation, mostly as a result of its skip connection architecture. It is worth to note that U-Net's architecture has been proven to be quite effective in preserving important spatial details necessary for the segmentation of a given structure or object like a Tumor or an organ. The ability of this model to work with a small number of annotated samples has contributed to its widespread use in the medical domain, where the process of annotating the data is tedious and time-consuming [12]. In recent years It has been integrated to other models like 3D U-Net and U-Net++ in order to extend its uses, especially in areas that require the analysis of 3D images [13].

B. Mask R-CNN

Mask R-CNN is a sophisticated solution for image segmentation tasks that extends the Faster R-CNN architecture with the additional concept of instance segmentation, in which individual instances of an object in an image are independently segmented, using a separate branch for segmentation tasks. Autonomous driving is one of the main areas that benefits from such segmentation, since it can be used to distinguish between pedestrians, vehicles, and other obstacles. The downside is that the fusion of additional neural networks (masks) into the final structure at times leads to longer inference times making it unsuitable for real time applications despite Mask R-CNN maintaining accuracy [14].

C. DeepLab

The method DeepLab uses relies on the so-called atrous convolutional properties of the network that increase its receptive field and allow the model to handle multi-scale features without spatial resolution loss. This feature makes DeepLab appropriate for understanding the context and contents of complex scenes and images of cities or urban environments in general, and for remote sensing in particular. Furthermore, the authors of DeepLab claim that the combination of their model with Conditional Random Fields (CRFs) improves the discrimination of segmentation borders and makes the model more efficient for tasks that require both local and global features extraction [15]. DeepLab modifications, and in particular one of its derivatives DeepLabv3+, have achieved high performances in a large number of benchmark datasets [16].

D. Attention U-Net

Attention mechanisms have also emerged recently in segmentation tasks as in the Attention U-Net model which places emphasis on the most relevant features present in an image on an as needed basis. In medical areas, it helps the model focus on the essential regions of the image in order to improve performance in segmentation for cases where minor variations of features are of diagnostic importance [4].

E. PSPNet

The Pyramid Scene Parsing Network (PSPNet) presents a pyramid pooling module that captures global contextual information, hence solving the problem of scene parsing of complicated scenes. This enables PSPNet to operate optimally in areas where many objects of different sizes and types are present. It also ensures a good feature fusion as the model integrates both local and global properties, thus requiring fewer limitations when recognizing the whole scene context and broadening its usability across different tasks [17].

F. ReSeg

In the case of ReSeg, it applies recursive structures for video object segmentation, which is a fundamental task in areas such as self-driving cars and video monitoring. ReSeg

applies temporal context effectively to constrain motion dynamics and perform better on tasks that are sensitive to the time domain [18].

IV. Comparative Analysis of Deep Learning Models

A comparative analysis of deep learning architectures shows trade-offs between accuracy, computational demands, and suitability for specific tasks.

TABLE I
KEY FEATURES AND PERFORMANCE COMPARISON OF SEGMENTATION MODELS

Model	Architecture Type	Key Features	Strengths	Weaknesses	Primary Applications	Computational Efficiency
U-Net	Encoder-Decoder with Skip Connections	Hierarchical feature extraction with skip connections, effective with small datasets	High accuracy with limited annotated data, fast for medical imaging	Computationally intensive, especially with larger images	Medical image segmentation (e.g., MRI, CT scans), biological image analysis	Requires high computational resources for large datasets and 3D images
Mask R-CNN	Extension of Faster R-CNN, with segmentation branch	Instance segmentation, combines object detection and segmentation	Accurate object differentiation, excellent for real-time object segmentation	Slower inference time, complex for real-time applications	Object segmentation (e.g., autonomous driving, robotics, video surveillance)	Computationally expensive, slower inference times for real-time tasks
DeepLab	Convolutional Neural Network (CNN) with Atrous Convolution	Multi-scale feature extraction, Conditional Random Fields (CRFs) for boundary refinement	Effective for complex, high-resolution scenes, robust at multi-scale feature extraction	High computational cost due to CRF refinement, slower for real-time applications	Urban scene segmentation, remote sensing, image analysis	Computationally expensive, CRFs increase inference time
Attention U-Net	U-Net with Attention Mechanism	Focuses on relevant features using attention mechanisms, improves performance with small, subtle features	Improves segmentation accuracy, particularly for small structures in medical images	Higher computational load due to attention mechanism, not suitable for fast inference	Medical image segmentation, focusing on precise small feature segmentation	Higher computational overhead compared to standard U-Net
PSPNet	Pyramid Pooling Module (PPM)	Multi-scale global context aggregation, effective for complex scenes	Excellent at handling diverse scenes and objects, strong global context modeling	Increased computational complexity, longer training and inference times	Scene parsing, urban analysis, autonomous driving, satellite imagery	Computationally intensive, longer inference times for large datasets

1) **U-Net**
In medical applications, U-Net is said to be extremely efficient but is hampered by the high computational power required however this limits its implementation in those areas which require quick processing. Such extensions as U-Net++ did enhance its flexibility however in regards to greater complexity [12,13].

2) **Mask R-CNN**
This model is quite good in separating objects in cluttered backgrounds because of which it has good applications in the areas with such requirements as object detection/segmentation. Nevertheless, its increased inference times pose some challenges in its applications in robotics where speed is critical [14].

3) **DeepLab**
DeepLab can capture multi-scale features due to its convolution, plus, CRFs enhance its real-time capabilities but this adds extra computational burden. Its use in urban scene analysis and remote sensing shows its range of application in different fields [15, 16].

4) **Attention U-Net**
Segmentations performed with applications that require special focus on the detail with attention networks are reported to be more accurate, the attention mechanism in itself poses extra computational demand and may therefore, affect efficiency [2].

5) **PSPNet**
The pyramid pooling module of PSPNet allows for complete understanding of the scene, however this adds complexity which may result in greater training and inference times. Such a trade-off is particularly obvious in situations that need both local and global context such as that of autonomous navigation [17].

V. Critical Insights and Future Directions

A. Data and Annotation Challenges

The ongoing frustration regarding the availability of sufficiently large and annotated data sets for deep learning model training is especially pronounced within the subfield of medical imaging. In particular, the annotation activity is labor intensive and expensive. Self-supervised and semi-supervised learning approaches are emerging to mitigate the problem of unlabeled data by reducing the reliance on labelled data [19].

B. Computational Demands

The inability to deploy deep learning models in time-sensitive applications due to their considerable computational demands, usually in terms of GPUs/TPUs, is one of the barriers towards robotics applications. However, recent developments in edge computing, combined with neural architecture search, could resolve these challenges by enabling more efficient processing on mobile platforms which is critical for augmented reality (AR) applications [20].

C. Generalization Challenges

Many of models trained under these circumstances perform reasonably well within the confines of lab-based benchmarks, but when applied in a real-world context, where room for degrees of freedom exists, they do not translate effectively. Furthermore, it is evident that domain adaptation and transfer learning should render the model less prone to these issues, particularly for instance, industrial automation, robotics, and any areas where segmentation models will have different deployment contexts [21]. These challenges must be solved if segmentation technology is to be adapted to different use cases in real world scenarios.

V. Conclusion

Deep learning approaches have greatly improved image segmentation, and featured changes in many applications. While networks such as U-Net, Mask R-CNN, and DeepLab have shown great promise in their applications, there are still issues with data scarcity, efficiency, and generalization that must be resolved. The emphasis of subsequent studies must be on improving the flexibility, clarity and processing

requirements of the models to increase the practical relevance of segmentation approaches. As segmentation technologies advance, these advances will enable greater precision and application across a wide spectrum of domains, including healthcare and unmanned systems.

REFERENCES

[1] J. Long, E. Shelhamer, and T. Darrell, "Fully Convolutional Networks for Semantic Segmentation," *arXiv preprint* arXiv:1411.4038, 2014. [Online]. Available: https://doi.org/10.48550/arXiv.1411.4038.

[2] 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*, N. Navab, J. Hornegger, W. Wells, and A. Frangi, Eds., vol. 9351, Cham, Switzerland: Springer, 2015, pp. 234–241. [Online]. Available: https://doi.org/10.1007/978-3-319-24574-4_28.

[3] K. He, G. Gkioxari, P. Dollár, and R. Girshick, "Mask R-CNN," in *2017 IEEE International Conference on Computer Vision (ICCV)*, Venice, Italy, 2017, pp. 2980–2988, doi: 10.1109/ICCV.2017.322.

[4] O. Oktay, J. Schlemper, L. L. Folgoc, M. J. Lee, M. P. Heinrich, K. Misawa, K. Mori, S. G. McDonagh, N. Y. Hammerla, B. Kainz, B. Glocker, and D. Rueckert, "Attention U-Net: Learning Where to Look for the Pancreas," *arXiv preprint* arXiv:1804.03999, 2018. [Online]. Available: https://doi.org/10.48550/arXiv.1804.03999.

[5] T.-Y. Lin *et al.*, "Microsoft COCO: Common Objects in Context," in *Computer Vision – ECCV 2014*, D. Fleet, T. Pajdla, B. Schiele, and T. Tuytelaars, Eds., vol. 8693, Cham, Switzerland: Springer, 2014, pp. 740–755. [Online]. Available: https://doi.org/10.1007/978-3-319-10602-1_48.

[6] M. Cordts *et al.*, "The Cityscapes Dataset for Semantic Urban Scene Understanding," in *2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, Las Vegas, NV, USA, 2016, pp. 3213–3223. [Online]. Available: https://doi.org/10.1109/CVPR.2016.350.

[7] V. Badrinarayanan, A. Kendall, and R. Cipolla, "SegNet: A Deep Convolutional Encoder-Decoder Architecture for Image Segmentation," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 39, no. 12, pp. 2481–2495, Dec. 2017, doi: 10.1109/TPAMI.2016.2644615.

[8] J. Canny, "A Computational Approach to Edge Detection," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. PAMI-8, no. 6, pp. 679–698, Nov. 1986, doi: 10.1109/TPAMI.1986.4767851.

[9] R. C. Gonzalez and R. E. Woods, *Digital Image Processing*, 4th ed. New Delhi, India: Pearson Education, 2019, ISBN: 9789353062989. [Online]. Available: https://dl.icdst.org/pdfs/files4/01c56e081202b62bd7d3b4f8545775fb.pdf.

[10] Y. Gani, E. Ustinova, H. Ajakan, P. Germain, H. Larochelle, F. Laviolette, M. Marchand, and V. Lempitsky, "Domain-Adversarial Training of Neural Networks," *arXiv preprint* arXiv:1505.07818, 2015. [Online]. Available: https://doi.org/10.48550/arXiv.1505.07818.

[11] E. Shelhamer, J. Long, and T. Darrell, "Fully Convolutional Networks for Semantic Segmentation," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 39, no. 4, pp. 640–651, Apr. 2017, doi: 10.1109/TPAMI.2016.2572683.

[12] O. Ronneberger, P. Fischer, and T. Brox, "U-Net: Convolutional Networks for Biomedical Image Segmentation," *arXiv preprint* arXiv:1505.04597, 2015. [Online]. Available: https://www.semanticscholar.org/paper/U-Net%3A-Convolutional-Networks-for-Biomedical-Image-Ronneberger-Fischer/6364fdaa0a0eccd823a779fcdd489173f938e91a?utm_source=direct_link.

[13] Z. Zhou, M. M. Rahman Siddiquee, N. Tajbakhsh, and J. Liang, "UNet++: A Nested U-Net Architecture for Medical Image Segmentation," in *Deep Learning in Medical Image Analysis and Multimodal Learning for Clinical Decision Support (DLMIA ML-CDS 2018)*, D. Stoyanov *et al.*, Eds., Lecture Notes in Computer Science, vol. 11045, Cham, Switzerland: Springer, 2018, pp. 3–11. [Online]. Available: https://doi.org/10.1007/978-3-030-00889-5_1.

[14] K. He, G. Gkioxari, P. Dollár, and R. Girshick, "Mask R-CNN," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 42, no. 2, pp. 386–397, Feb. 2020, doi: 10.1109/TPAMI.2018.2844175.

[15] 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 Trans. Pattern Anal. Mach. Intell.*, vol. 40, no. 4, pp. 834–848, Apr. 2018, doi: 10.1109/TPAMI.2017.2699184.

[16] L.-C. Chen, Y. Zhu, G. Papandreou, F. Schroff, and H. Adam, "Encoder-Decoder with Atrous Separable Convolution for Semantic Image Segmentation," in *Computer Vision – ECCV 2018*, V. Ferrari, M. Hebert, C. Sminchisescu, and Y. Weiss, Eds., Lecture Notes in Computer Science, vol. 11211, Cham, Switzerland: Springer, 2018, pp. 833–851. [Online]. Available: https://doi.org/10.1007/978-3-030-01234-2_49.

[17] H. Zhao, J. Shi, X. Qi, X. Wang, and J. Jia, "Pyramid Scene Parsing Network," in *Proc. 2017 IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Honolulu, HI, USA, 2017, pp. 6230–6239, doi: 10.1109/CVPR.2017.660.

[18] F. Perazzi, A. Khoreva, R. Benenson, B. Schiele, and A. Sorkine-Hornung, "Learning Video Object Segmentation from Static Images," in *Proc. 2017 IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Honolulu, HI, USA, 2017, pp. 3491–3500, doi: 10.1109/CVPR.2017.372.

[19] M. Caron, I. Misra, J. Mairal, P. Goyal, P. Bojanowski, and A. Joulin, "Unsupervised Learning of Visual Features by Contrasting Cluster Assignments," *arXiv preprint* arXiv:2006.09882, 2020. [Online]. Available: https://doi.org/10.48550/arXiv.2006.09882.

[20] M. Tan and Q. V. Le, "EfficientNet: Rethinking Model Scaling for Convolutional Neural Networks," *arXiv preprint* arXiv:1905.11946, 2019. [Online]. Available: https://doi.org/10.48550/arXiv.1905.11946

[21] B. Sun and K. Saenko, "Deep CORAL: Correlation Alignment for Deep Domain Adaptation," in *Computer Vision – ECCV 2016 Workshops*, G. Hua and H. Jégou, Eds., Lecture Notes in Computer Science, vol. 9915, Cham, Switzerland: Springer, 2016, pp. 443–450. [Online]. Available: https://doi.org/10.1007/978-3-319-49409-8_35

Madhurima Mukherjee was born in Kolkata, India. She received the B.Tech degree in electronics and communication engineering from Vellore Institute of Technology, Chennai, India, in 2019. She also currently completed a PG Level Advanced Certification in AI/MLOps at the Indian Institute of Science, Bengaluru, India, in July 2024.

She is currently a Senior Software Engineer / Technical Consultant at Adobe, Bengaluru, India, where she develops and implements frontend and mobile app solutions, enhancing user engagement through personalized content sections. Her work involves the use of Adobe Target, Adobe Experience Manager, and Adobe Analytics to design custom solutions, as well as collaborating with cross-functional teams for cohesive project execution. She also collaborates with Adobe's ML team to integrate AI features, expanding Adobe's recommendation engine through advanced model proposals. Previously, she was a Software Development Intern at Siemens Industry Software, where she created WPF applications and a Patient Health Monitoring System using React. She has also completed summer internships at Vodafone, Bharat Sanchar Nigam Limited, and Siemens Private Ltd, gaining experience in network performance analysis, electronics, and automation. Her current research interests include deep learning architectures for image segmentation, bias mitigation in facial recognition systems, and recommendation system optimization.

Ms. Mukherjee is a member of AI Product Hive and an active participant in Adobe Gen AI Initiatives, where she contributed to Adobe's inaugural Gen AI Hackathon as a finalist.