Artificial intelligence and Fourier optics: Application of DeepLabV3+ in the recovery of a diffracting aperture in light propagation

C. Camacho-Bello, L. Gutiérrez-Lazcano, and R. M. Ortega-Mendoza

Universidad Politecnica de Tulancingo, Laboratorio de Inteligencia Artificial, Calle Ingenierias No100, Hidalgo, Mexico, 20036.

Received 19 April 2023; accepted 9 August 2023

The combination of Fourier Optics and Artificial Intelligence has driven significant advances in image processing and modeling of optical systems, with the UNet architecture being the main protagonist. However, the DeepLabV3+ network has recently shown promising performance detecting diffracting apertures. In this study, we investigate the effectiveness of DeepLabV3+ in identifying diffracting apertures in light propagation models and compare its performance with that of UNet. The results reveal that DeepLabV3+ outperforms UNet in accuracy and robustness in identifying diffracting apertures, even in the presence of noise and aperture shape variations.

Keywords: Diffraction; optical systems; deep learning.

DOI: https://doi.org/10.31349/RevMexFis.70.011301

1. Introduction

Light propagation and its interaction with various media has been a topic of interest in physics for centuries. As a discipline, optics has focused on studying light and its behavior, providing a deep understanding of phenomena such as refraction, reflection, diffraction, and interference. Among the fundamental concepts in optics, the aperture amplitude transmittance function or diffracting aperture describes how light is transformed as it passes through an aperture, which can be affected by shape and size. This function is essential to understanding how light is propagated and transformed in optical systems [1]. On the other hand, Fourier optics provides a mathematical representation of the amplitude transmittance function, which uses Fourier transforms to analyze and describe the propagation of light in optical systems and provides an accurate description of how light is transmitted changes as it passes through complex optical systems [2]. These theoretical tools have practical applications, from improving precision and efficiency in optical system analysis to simulating light propagation in different media. For example, understanding the amplitude transmittance function in biomedical optics is essential for accurately interpreting optical images of biological tissues [3]. Furthermore, in the design of segmented telescopes, the precise measurement of the piston error, which uses the amplitude transmittance function, is crucial for optimizing the telescope's performance [4]. Fourier Optics has also been used to simulate physical optics, allowing practical performance predictions, including diffraction effects that emerge with fully coherent sources [5]. The aperture amplitude transmittance function and Fourier Optics are fundamental to understanding the formation of images in optical systems because they allow us to describe how light is transformed when passing through a medium, providing a detailed description of how an image is formed, allowing the design and optimization of these systems to improve the precision and efficiency in the analysis of optical systems, which has allowed significant advances in the field.

In other fields, biomedical image segmentation is crucial in modern medicine and research. One of the most significant advances in recent years has been the introduction of the UNet neural network architecture proposed by Olaf Ronneberger in 2015 [6]. Over the years, UNet's impact has extended beyond the biomedical realm, reaching areas such as Fourier Optics. Thanks to the ability of the UNet architecture to extract relevant features and patterns in high-resolution images, it has been possible to improve optical systems significantly, optimizing image reconstruction and detecting key features [7], which makes it ideal for a wide range of applications, such as Holo-UNet, which enables the restoration of holographic images of living cells [8], as well as improvements in computing times [9]. Another architecture based on UNet is PhaseNet, which efficiently recovers the phase of a holographic signal [10]. Similarly, improvements have been made in the phase unwrapping with Doppler optical coherence tomography images in the Fourier domain using Res-UNet [11]. In addition, UNet is implemented to find the diffractive phase that can compensate for chromatic aberrations in the entire visible spectrum [12]. Also, it removes unwanted streaks in thin film fluorescence microscopy [13]. The UNet architecture has proven to be highly effective in various applications, but it has limitations in optimizing and improving optical systems. Therefore, there is a need to propose new architectures that address this problem and overcome these limitations.

In computer vision and semantic segmentation, the DeepLabV3+ network emerges as a benchmark in the field, positioning itself as a tool of great importance within the scientific community. This sophisticated architecture is an extension of its predecessor, DeepLabV3, and introduces distinctive features that enhance its ability to accurately and efficiently identify and segment objects; therefore, it is easy to implement in problems where UNet has been implemented. The versatility and superior performance of DeepLabV3+ have made it an essential tool in various applications, such as detecting objects and people, analyzing traffic, or monitoring crops' health. Its adaptability and efficiency make DeepLabV3+ an ideal candidate for use in areas such as Fourier Optics, where it could significantly improve optical systems. Some possible applications in this field include the reconstruction of holographic images, the analysis of interference patterns, or the optimization of optical lens and filter systems. Implementing the DeepLabV3+ network in Fourier Optics would allow researchers to address more complex challenges while driving the development of innovative technologies across multiple scientific disciplines.

In this article, we propose to investigate the effectiveness of the DeepLabV3+ network in identifying diffracting aperture and to compare its performance with that of the UNet architecture. For this, we present experiments and analyses using data sets of light propagation models with different characteristics, including noise and variations in the shape of the aperture. Through this research, we determine if DeepLabV3+ is a valuable tool to improve the modeling and analysis of optical systems in practical applications.

This research contributes to scientific knowledge by providing a rigorous evaluation of the performance of DeepLabV3+ compared to UNet. This is evidence of the advantages of using DeepLabV3+ for diffracting aperture detection in optical systems. The results obtained in this study can potentially drive future research and applications in the field of Fourier Optics and Artificial Intelligence, as well as in related areas such as microscopy, astronomy, and optical telecommunication.

The main contribution of this article lies in the proposal and validation of a new neural network architecture that can address the limitations of UNet in optical systems. This work has the potential to significantly improve the analysis and design of optical systems, impacting areas such as the characterization of biological tissues, microscopy, and optical communications. By sharing and discussing these results, we aim to enrich knowledge in Fourier Optics and Artificial Intelligence and encourage the development of innovative solutions in the modeling and optimizing of optical systems.

This article is organized as follows: In the "Materials and Methods" section, we concisely describe the DeepLabV3+ and UNet architectures and detail the generation of the data set used in the light propagation model. In addition, we present a brief explanation of the performance measure we used to evaluate both models. In the "Results" section, we compare DeepLabV3+ and UNet in terms of robustness against noise and variations in the diffracting aperture for untrained samples and discuss each approach's main strengths and limitations. Finally, in the "Conclusions" section, we synthesize the contributions of the presented work and highlight the implications of our findings for advancing knowledge and practical applications in the Fourier Optics and Computer Vision field.

2. Materials and methods

2.1. Description of UNet and DeepLabV3+ architectures

In general terms, it is difficult to determine how to compare the two architectures since their performance can vary depending on the context and the specific application. Both architectures have proven effective in semantic segmentation tasks. Still, they have different characteristics and advantages that can make one more suitable than the other in certain situations, such as in our case study. UNet is a convolutional neural network (CNN) architecture specially designed for semantic segmentation tasks in biomedical imaging. The UNet structure comprises two main parts: an encoder and a decoder, which connect by a series of hop connections to improve spatial information retrieval [14]. The encoder is responsible for extracting features from the input image through a series of convolutional and pooling layers, which reduces spatial resolution and increases feature depth. As it progresses through the encoder, the network captures contextual information at multiple levels of abstraction. On the other hand, the decoder uses convolutional and upsampling layers to reconstruct the segmentation of the features extracted by the encoder. Jump connections combine high-resolution (fine detail) and low-resolution (context) information, resulting in more accurate and detailed segmentation, such as shown in Fig. 1.

On the other hand, DeepLabV3+ is also a convolutional neural network architecture aimed at semantic segmentation in images, which introduces an encoder-decoder module in which the encoder uses an Atrous Spatial Pyramid Pooling (ASPP) module [15] to capture contextual information at different spatial resolutions; therefore, improves the network's ability to identify objects of different sizes and shapes. In addition, it incorporates atrous (dilated) convolutions in its architecture, allowing it to expand the reception field of neurons without increasing the number of parameters or calculations required; this feature enables the network to capture

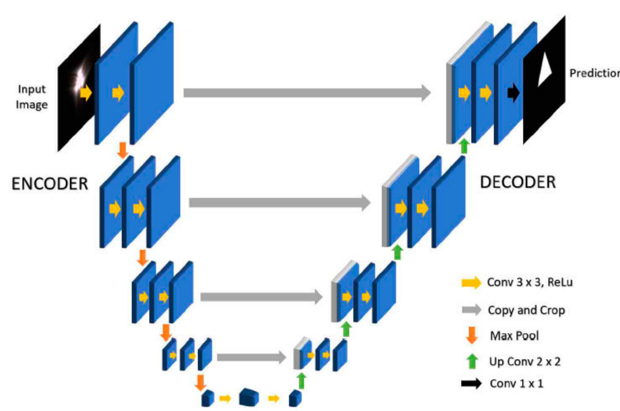

FIGURE 1. UNet structure for recovery of the amplitude transmittance function of the aperture in light propagation.

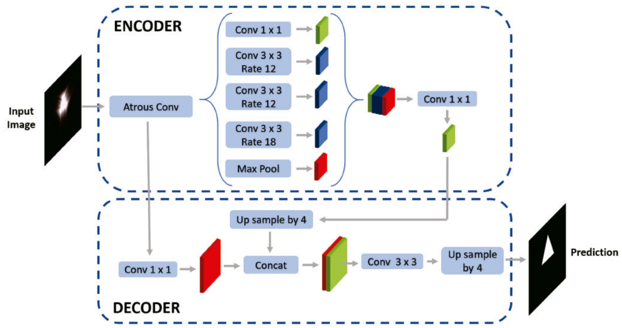

FIGURE 2. DeepLab3+ structure for recovery of the amplitude transmittance function of the aperture.

fine details in the image while maintaining high computational efficiency [16]. Unlike UNet, DeepLabV3+ supports different backbones, accommodating various application and performance requirements. Likewise, DeepLabV3+ uses a "decoding" strategy that reconstructs the segmentation from the features extracted by the encoder, including the ASPP module, resulting in precise and detailed segmentations. The architecture is shown in Fig. 2.

2.2. Data sets used light propagation models

This study prepared various data sets to evaluate and compare the UNet and DeepLabV3+ architectures in identifying diffracting apertures in light propagation models. The dataset consists of $40,000$ images of square and triangular apertures, 20,000 of each type, with dimensions 160×160 pixels, generated from a polychromatic light source at a distance $d = 50$cm to obtain diffraction intensity patterns $I(x, y; d)$. Fig. 3 shows the optical system schematic of light propagation from a diffracting aperture.

The diffracting apertures were subjected to translation, scaling, and rotation transformations to add variability to the models. The image set was divided into training and validation sets in a 90:10 ratio. Figure 4a) shows some images used to train convolutional neural networks. The input image corresponds to an image of the diffraction intensity patterns.

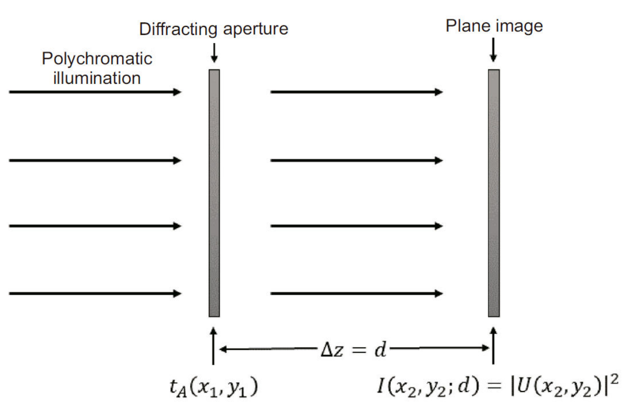

FIGURE 3. Schematic diagram of the simple optical system to generate the dataset.

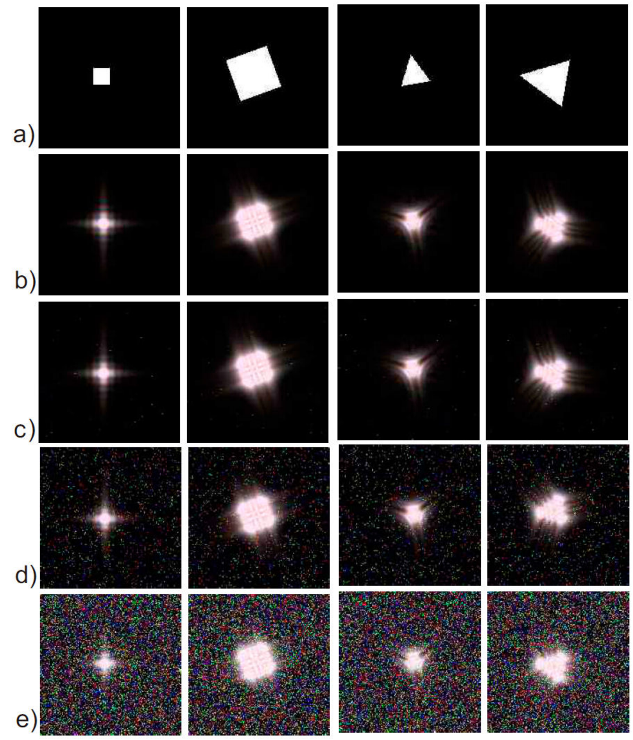

FIGURE 4. Some images for training and validation. a) Output image (diffracting aperture) and b) Input image (diffraction intensity pattern). Degraded patterns with Gaussian noise: c) $\mu = 0$ and $\sigma = 0.2$. d) $\mu = 0$ and $\sigma = 0.5$. e) $\mu = 0$ and $\sigma = 1.0$.

In contrast, the output image is a binary image that labels each pixel as the amplitude transmittance function of the aperture.

The images present significant variability in terms of lighting, contrast, and object shapes, which allows evaluation of the generalization capacity of the networks under different conditions. Also, in the case of robustness in aperture recovery, an additional dataset is generated in which different noise levels are added to the images of the diffraction intensity patterns. In vision systems, the robustness of the algorithms against noise and other disturbances in the input images is crucial. In this sense, the present experiment evaluates the robustness of our proposal against images affected by different levels of noise, as shown in Figs. 4c), 4d), and 4e).

In the design of optical systems, it is essential to guarantee the robustness of the model in unforeseen situations. A critical case is testing the model with significantly different images than those used during training and validation. In some applications, the model uses different environments with varying imaging conditions and variations in diffracting apertures; this situation can be incredibly challenging. Therefore, the work presents an experiment in which we evaluate the robustness of the proposal under extreme conditions of area change in the apertures and the intensity of the diffraction patterns. The study selected the images used for testing to differ significantly from those in the training and validation data set, as shown in Fig. 5.

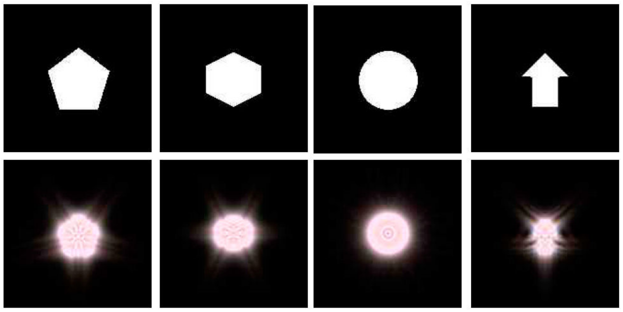

FIGURE 5. Generalization in different conditions: dataset with shapes of the aperture outside the training.

The proposed data sets serve as the basis for training, validating, and comparing the UNet and DeepLabV3+ neural network architectures to identify diffracting apertures in light propagation models and assess their efficiency.

2.3. Performance metrics

Accuracy, F1-score, recall, and precision metrics are standard measures used to assess the performance of deep learning models. S. Zagoruyko and N. Komodakis [17] highlight the importance of using these metrics to evaluate deep learning models. Deep learning models must evaluate their performance using a combination of these metrics to gain a complete understanding of their performance on specific tasks. Since then, these metrics have become the most widely used to evaluate deep learning models, and it is recommended that all models use them to evaluate their performance. Using these metrics is essential to ensure that deep learning models are accurate and reliable in complex tasks. The metrics allow quantifying and comparing the accuracy and robustness of DeepLabV3+ and UNet in identifying diffracting apertures in light propagation models. The metrics used are described below:

- Loss: The metric measures the discrepancy between model predictions and actual values. A lower loss value indicates better model performance.
- Accuracy: The metric evaluates the proportion of successes in the model's predictions compared to the total number of predictions made.
- Intercept Over Union (IoU): The IoU is a metric that measures the overlap between the regions predicted by the model and the actual regions in the images. A higher IoU value indicates a better match between the predicted and existing areas.
- Precision: measures the ratio of true positives to the total positive predictions made by the model.
- Recall: also known as completeness, assesses the ratio of true positives to the total number of positive cases.
- F1 score: measure the combines precision and sensitivity into a single value, providing a balanced assessment of both aspects of model performance.

In addition to these metrics, we calculated GPU and CPU compute time to compare the inference time efficiency of the different architectures. The models were trained on a computer with an Intel Core i7-9700k running at 3.0 GHz, 32 GB of RAM, and an NVIDIA GeForce GTX 1660 SUPER GPU with 6 GB of Video RAM. The computer time measure is especially relevant in practical applications, where processing time can be a critical factor in the implementation and performance of the optical system.

3. Results

3.1. Comparison between UNet and DeepLabV3+

This section compares the UNet and DeepLabV3+ neural network architectures in identifying diffracting apertures in light propagation models. Table I shows a comparative analysis of the performance metrics obtained from both architectures.

The DeepLabV3+ network performed better in Loss and Accuracy than UNet, obtaining a loss value of 0.01023 and an accuracy of 99.543%, compared to values of 0.04628 and 98.146% for UNet, respectively. Furthermore, DeepLabV3+ showed a significant improvement in the IoU (Intersection over Union) metric, reaching a value of 0.89922 compared to UNet's 0.66473. This improvement indicates a better match between the detected and valid regions in the test images.

Regarding Precision, Recall, and F1-Score metrics, DeepLabV3+ outperformed UNet with values of 0.96001, 0.95151, and 0.95574, respectively, versus UNet's values of 0.84931, 0.78093, and 0.81368. These results suggest a greater ability of DeepLabV3+ to identify diffracting apertures and minimize false positives and negatives.

Regarding the computation time, DeepLabV3+ showed an advantage in inference time in GPU and CPU, with times of 0.03733s and 0.06073s, respectively. On the other hand, UNet took 0.04242s on GPU and 0.12059s on CPU.

TABLE I. Comparative analysis with different neural network architectures.

	Loss	Accuracy	IoU	Precision	Recall	F1-Score	Time GPU	Time CPU
UNet	0.04628	0.98146	0.66473	0.84931	0.78093	0.81368	0.04242s	0.12059s
DeepLabV3+	0.01023	0.99543	0.89922	0.96001	0.95151	0.95574	0.03733s	0.06073s

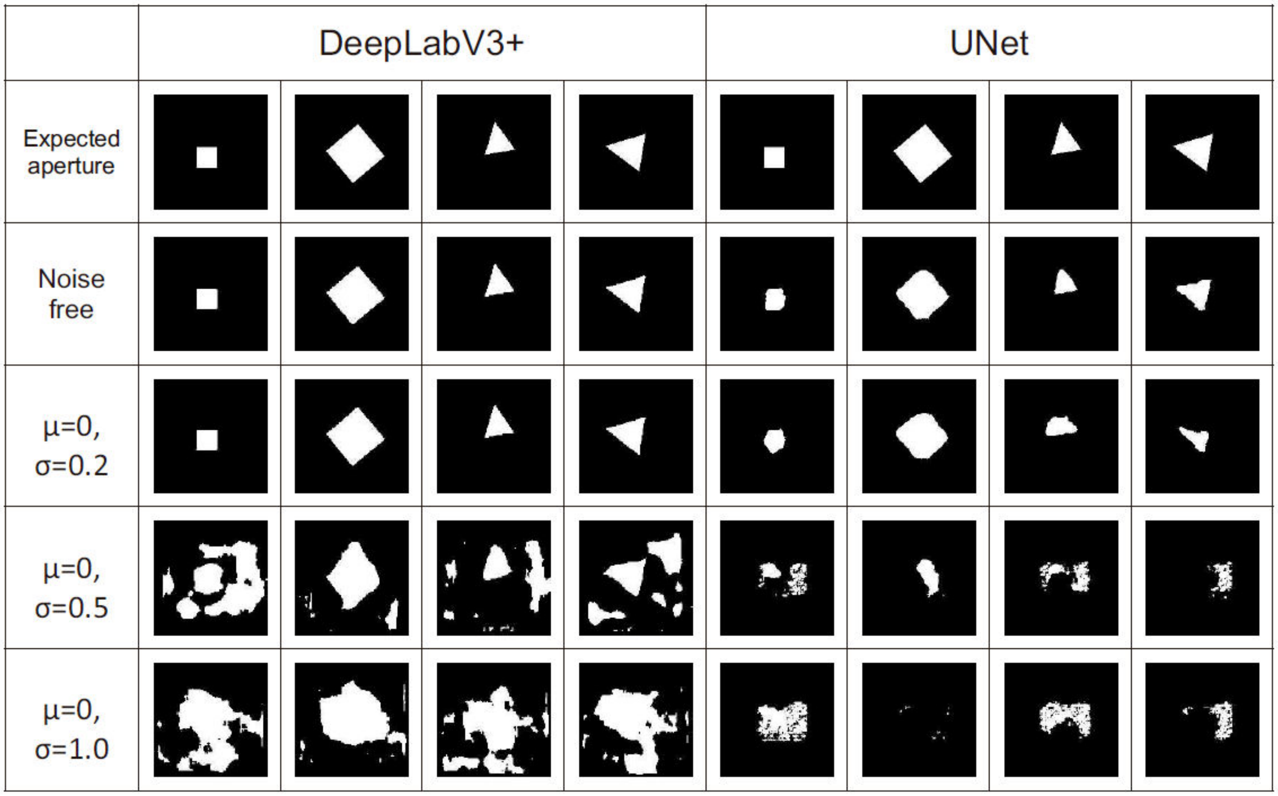

FIGURE 6. Results of the recovery of the degraded diffracting aperture with Gaussian noise from semantic segmentation networks. Note that the models were not trained with images containing noise; nevertheless, DeepLabV3+ recovered the apertures with low noise levels.

The results reveal that the DeepLabV3+ architecture outperforms UNet regarding precision and robustness when identifying diffracting apertures; these findings support the idea that DeepLabV3+ is a valuable tool for improving the modeling and analysis of optical systems in practical applications.

3.2. Evaluation of the robustness against noise and variations in the aperture shape

In this section, we analyze the robustness of the UNet and DeepLabV3+ neural network architectures against noise and variations in the shape of the aperture in light propagation models. To evaluate the robustness against noise, we introduce Gaussian noise in the test images with different intensity levels ($\sigma = 0.2, 0.5$ and 1.0). Table II presents a comparative analysis of the IoU metric under different Gaussian noise levels. Also, Fig. 6 shows the recovery of the validation apertures from the models trained with the training set.

TABLE II. Comparative analysis with the IoU metric under Gaussian noise with $\mu = 0$

	$\sigma = 0.2$	$\sigma = 0.5$	$\sigma = 1.0$
UNet	0.66524	0.21617	0.13995
DeepLabV3+	0.89901	0.21323	0.16438

For $\sigma = 0.2$, UNet obtained an IoU value of 0.66524, while DeepLabV3+ achieved a value of 0.89901, demonstrating a better ability of DeepLabV3+ to identify diffracting apertures in the presence of low noise. By increasing the noise level to $\sigma = 0.5$, both models experienced a decrease in the IoU metric, being 0.21617 for UNet and 0.21323 for DeepLabV3+. Despite the decrease in performance, DeepLabV3+ still holds an advantage over UNet. With even higher noise ($\sigma = 1.0$), UNet showed an IoU value of 0.13995, while DeepLabV3+ reached a value of 0.16438.

The results also demonstrate that both architectures can adapt to variations in the shape of the aperture, with DeepLabV3+ being more robust to extreme changes in imaging conditions. Although the performance of both models is affected by the difference in the images used for the test, DeepLabV3+ shows a minor decrease in its performance compared to UNet.

On the other hand, we focus on exploring the robustness of both models against extreme variations in the shape of the aperture that is outside the training set, a critical aspect of the practical application of these techniques in natural optical systems. Figure 7 shows the results of the IoU metric of 4 shapes (Fig. 5) that cannot be considered in training to provide the aperture recovery capacity of DeepLabV3+.

The DeepLabV3+ architecture demonstrates higher robustness against noise and aperture shape variations com-

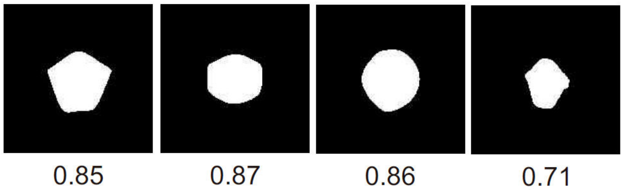

FIGURE 7. Experiment results: obtained aperture image and IoU metric.

pared to UNet and extreme changes in image capture conditions. Its performance is slightly affected by the difference in the images used for the test. The presented findings support the utility of DeepLabV3+ to improve the modeling and analysis of optical systems in practical applications, even under harsh conditions.

3.3. Discussion of results

The comparative study of the UNet, and DeepLabV3+ architectures to identify the diffracting aperture in light propagation models, several contributions can be recognized:

- Learning capacity: The experiments demonstrate that autoencoder architectures could learn optical phenomena and describe the diffracting aperture from diffraction patterns; this indicates their effectiveness in the proposed task, as shown in Table I.

- Computational efficiency: In terms of computational times, Table I shows that the DeepLabV3+ architecture is computationally efficient compared to the other architectures; this can be important for real-time applications.

- Identification accuracy: DeepLabV3+ provided the most accurate identification of the diffracting aperture in light propagation models compared with UNet.

- Generalization ability of trained apertures: The architectures demonstrated generalizability of the diffracting aperture identification task from diffraction patterns to different test data sets under conditions similar to the trained patterns.

The results show that DeepLabV3+ is an innovative solution for calculating the diffracting aperture in light propagation models. However, it also has some limitations that must be considered when using it in applications, such as

- Requires large training data sets: The proposal can learn the relationship between the light intensity distribution in the source plane and the corresponding diffracting aperture from an extensively suitable training data set. The quality and quantity of the training data can influence the accuracy and generalizability, mainly to adapt to different scenarios with patterns not considered in training.

- Limitations on the resolution of the diffracting aperture: The resolution of the diffracting aperture may be limited by the ability of the network to learn and recognize subtle patterns in the light intensity distribution. This can be a limitation in situations requiring a high resolution in the diffracting aperture measurement.

- Limitations on the shape of the diffracting aperture: The ability of DeepLabV3+ to recognize and measure the diffracting aperture may be limited by the shape of the aperture. In particular, it is possible that the network cannot measure diffracting apertures that present extreme shapes or complex interference patterns because it is limited by the number of aperture shapes with which it was trained. Figure 5 shows the difficulty of recovering the arrow aperture. The limitation is probably solved by adding more opening shapes to generalize the model; however, it requires more significant processing and storage capacity.

- Limitations in the memory required in training: DeepLabV3+ architecture is relatively deep and uses many parameters, which require a significant amount of memory to store the gradients and update the weights during training; this may be a limitation on systems with GPUs with limited memory resources. In particular, the present research is limited to the shapes of squares and triangles because we are limited to 6 GB of GPU memory, so we cannot include more shapes to make the model more robust.

- Limitations in the application in real-time: Although DeepLabV3+ presents competitive times (Table I) in real-time applications, it may be limited by the network's processing speed and the CPU or GPU capacity used for its implementation.

4. Conclusions

Artificial Intelligence and Fourier Optics are complementary disciplines used in different ways. However, both fields can lead to more advanced and efficient solutions in light propagation; this can be of great importance in the research and development of optical systems, in the precise measurement of light, and in solving optical problems in practical applications. The comparison demonstrates the effectiveness of artificial intelligence in adapting and describing optical phenomena. DeepLabV3+ is an innovative and efficient solution for the precise and reasonable time calculation of the diffracting aperture in light propagation models. The reconstruction results validate the generalizability and descriptiveness of the diffracting aperture from diffraction patterns. However, it must be considered that the results obtained depend on the selection of the data set used for the training and evaluation of the network. DeepLabV3+ can be used in various applications in optics, where accurate and fast measurement of the diffracting aperture is important to improve image quality and efficiency in image acquisition or processing data. Furthermore, real-time inference of the diffracting aperture is faster, making it suitable for real-time optical systems. The findings of this study provide valuable information for future research and for improving the accuracy of diffracting aperture identification in light propagation models using machine learning techniques. In future studies, the performance of DeepLabV3+ on different types of light propagation images, such as those involving complex optical elements, can be explored.

1. E. Hecht, Optics, 5th ed. (Pearson, 2017).
2. J. W. Goodman, Fourier Optics, 3rd ed. (Roberts and Company Publishers, 2005).
3. L. Wang and H. Wu, Biomedical Optics: Principles and Imaging, 10-15 John Wiley & Sons, New Jersey (2007).
4. D. Yue, Y. He, and Y. Li, Piston error measurement for segmented telescopes with an artificial neural network, Sensors 21 (2021) 3364.
5. U. Flechsig *et al*., Physical optics simulations with PHASE for SwissFEL beamlines, In AIP Conference Proceedings, vol. 1741 (AIP Publishing, 2016).
6. N. Siddique *et al*., U-Net and Its Variants for Medical Image Segmentation: A Review of Theory and Applications, *IEEE Access* **9** (2021) 82031, https://doi.org/10.1109/ACCESS.2021.3086020.
7. T. Zeng, Y. Zhu, and E. Y. Lam, Deep learning for digital holography: a review, *Opt. Express* **29** (2021) 40572, https://doi.org/10.1364/OE.443367.
8. Z. Zhang *et al.*, Holo-UNet: hologram-to-hologram neural network restoration for high fidelity low light quantitative phase imaging of live cells, *Biomed. Opt. Express* **11** (2020) 5478, https://doi.org/10.1364/BOE.395302.
9. J. Wu *et al*., High-speed computer-generated holography using an autoencoder-based deep neural network, *Opt. Lett*. **46** (2021) 2908, https://doi.org/10.1364/OL.425485.
10. T. Zhang *et al.*, Rapid and robust two-dimensional phase unwrapping via deep learning, *Opt. Express* **27** (2019) 23173, https://doi.org/10.1364/OE.27.023173.
11. C.Wu *et al*., Phase unwrapping based on a residual en-decoder network for phase images in Fourier domain Doppler optical coherence tomography, *Biomed. Opt. Express* **11** (2020) 1760, https://doi.org/10.1364/BOE.386101.
12. X. Dun *et al*., Learned rotationally symmetric diffractive achromat for full-spectrum computational imaging, *Optica* **7** (2020) 913, https://doi.org/10.1364/OPTICA.394413.
13. Z. Wei *et al*., Elimination of stripe artifacts in light sheet fluorescence microscopy using an attention-based residual neural network, *Biomed. Opt. Express* **13** (2022) 1292, https://doi.org/10.1364/BOE.448838.
14. O. Ronneberger *et al*., Medical Image Computing and Computer-Assisted Intervention - MICCAI 2015 (Springer International Publishing, Cham, 2015) pp. 234-241, https://doi.org/10.1007/978-3-319-24574-4_28.
15. L.-C. Chen *et al*., Rethinking atrous convolution for semantic image segmentation, arXiv preprint arXiv:1706.05587 (2017)
16. L.-C. Chen *et al*., Encoder-Decoder with Atrous Separable Convolution for Semantic Image Segmentation, In Computer Vision - ECCV 2018: 15th European Conference, Munich, Germany, September 8-14, 2018, Proceedings, Part VII (Springer-Verlag, Berlin, Heidelberg, 2018) p. 833, https://doi.org/10.1007/978-3-030-01234-2_49.
17. K. He and J. Sun, Convolutional neural networks at constrained time cost, In 2015 IEEE Conference on Computer Vision and Pattern Recognition (CVPR) (2015) pp. 5353, https://doi.org/10.1109/CVPR.2015.7299173.