

Received 9 July 2024, accepted 28 July 2024, date of publication 1 August 2024, date of current version 13 August 2024.

Digital Object Identifier 10.1109/ACCESS.2024.3436509

## RESEARCH ARTICLE

# CNN-Modified Encoders in U-Net for Nuclei Segmentation and Quantification of Fluorescent Images

AISHWARYA SHRESTHA<sup>1</sup>, XUEFENG BAO<sup>2</sup>, (Member, IEEE), QINGSU CHENG<sup>2</sup>,  
AND SUSAN MCROY<sup>1</sup>

<sup>1</sup>Department of Computer Sciences, University of Wisconsin–Milwaukee, Milwaukee, WI 53211, USA

<sup>2</sup>Department of Biomedical Engineering, University of Wisconsin–Milwaukee, Milwaukee, WI 53211, USA

Corresponding author: Qingsu Cheng (chengq@uwm.edu)

This work was supported in part by the US Department of Agriculture (USDA) under Grant 2024-67019-42619; in part by the National Science Foundation (NSF) through a subcontract from the University of Nevada, Reno, under Grant 2344476; and in part by the National Aeronautics and Space Administration (NASA) under Grant 80NSSC23K0989.

**ABSTRACT** This research introduces an advanced approach to automate the segmentation and quantification of nuclei in fluorescent images through deep learning techniques. Overcoming inherent challenges such as variations in pixel intensities, noisy boundaries, and overlapping edges, our devised pipeline integrates the U-Net architecture with state-of-the-art CNN models, such as EfficientNet. This fusion maintains the efficiency of U-Net while harnessing the superior capabilities of EfficientNet. Crucially, we exclusively utilize high-quality confocal images generated in-house for model training, purposefully avoiding the pitfalls associated with publicly available synthetic data of lower quality. Our training dataset encompasses over 3000 nuclei boundaries, which are meticulously annotated manually to ensure precision and accuracy in the learning process. Additionally, post-processing is implemented to refine segmentation results, providing morphological quantification for each segmented nucleus. Through comprehensive evaluation, our model achieves notable performance metrics, attaining an F1-score of 87% and an Intersection over Union (IoU) value of 80%. Furthermore, its robustness is demonstrated across diverse datasets sourced from various origins, indicative of its broad applicability in automating nucleus extraction and quantification from fluorescent images. This innovative methodology holds significant promise for advancing research efforts across multiple domains by facilitating a deeper understanding of underlying biological processes through automated analysis of fluorescent imagery.

**INDEX TERMS** Deep learning, UNet architecture, fluorescent image processing, nuclei segmentation, mammary epithelial cells.

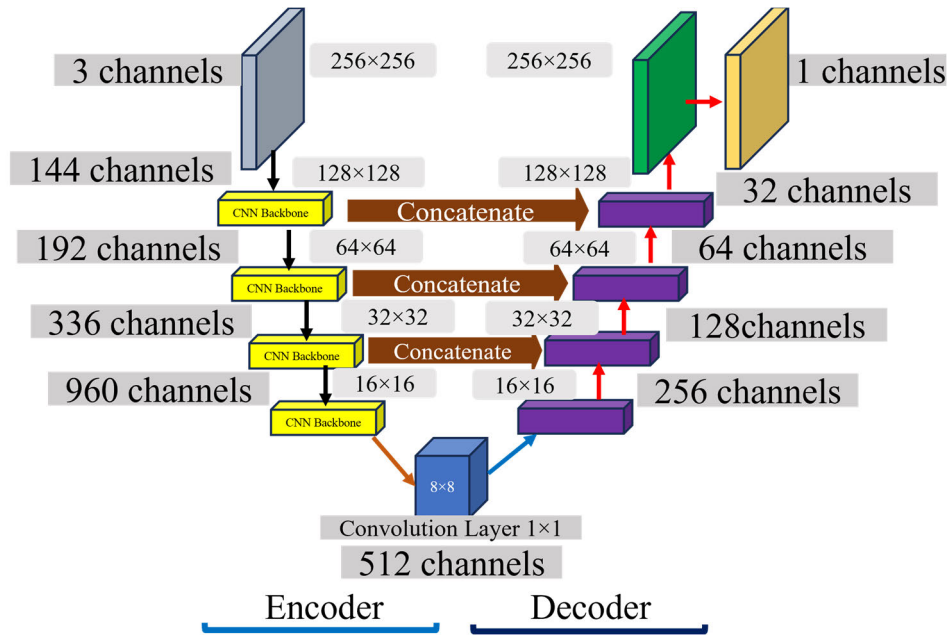
## I. INTRODUCTION

Biological image analysis typically initiates with object segmentation, a critical phase involving the extraction of regions or objects of interest within the images. These regions encapsulate vital information such as molecular activities, organelle details, cell phenotypes, tissue structures, and anatomical features. The precision and efficacy of biomedical image segmentation are paramount, as they form the basis for subsequent quantitative analyses, visualization projections, and the

elucidation of complex biological events [1]. A meticulously performed segmentation process not only facilitates accurate identification and delineation of relevant objects but also enables researchers to quantitatively analyze their characteristics, track dynamic changes, and unveil underlying patterns or relationships within the biological system.

At a fundamental level, cell segmentation enables quantitative analyses of cellular and molecular data, as well as the classification and clustering of morphological phenotypes in individual cells and organoids [2]. Analyzing this foundational information allows for the extraction of mid-level insights such as cellular migration, division, and tissue

The associate editor coordinating the review of this manuscript and approving it for publication was Humaira Nisar<sup>1</sup>.



**FIGURE 1.** Illustration of deep learning frameworks, replaces the UNet encoder with ResNet-50, EfficientNet-B5, EfficientNet-B7, and Inception-ResNet-v2 models, is shown.

development under environmental stimuli in real-time. Moreover, the combination of low- and mid-level data enables the identification and highlighting of apoptotic/necrotic cells, diseased cells, damaged tissue, and aberrant genetic and proteomic expressions, facilitating disease diagnosis and assessment of therapeutic effects [3]. However, manual examination and analysis of these images pose significant challenges, as the tasks are labor-intensive and error-prone, which may lead to missed diagnoses. Additionally, semantic segmentation of cell nuclei proves challenging due to the unconventional morphologies in diseased environments, noise in images, large dynamic range within a single picture, and overlapping nuclei. Deep learning emerges as a powerful solution to address these shortcomings and enhance the efficiency of automated cell segmentation [4].

Deep learning algorithms offer precise mappings and enhanced performance by automating processes, making them invaluable tools. Leveraging these benefits, this paper delves into utilizing a deep-learning approach for segmenting and identifying cell nuclei in fluorescent images. Among the prominent frameworks in computer vision, U-Net stands out as a widely adopted architecture that was introduced by Ronneberger et al. [5]. U-Net features a distinctive “U-shaped” design. Its architecture comprises an encoder which is responsible for capturing features through a contracting path, and a decoder which facilitates a symmetric expanding path. This configuration enables U-Net to effectively capture both local and global information at the pixel level. Notably, U-Net’s efficiency surpasses that of fully convolutional networks (FCN), as evidenced by previous studies [6].

In this project, we capitalized on the remarkable efficiency of the encoder-decoder-based U-Net architecture, which produces favorable segmentation results even when working with a limited training dataset, thanks to the incorporation of data augmentation techniques. Expanding upon this foundation, we integrated compound scaled CNN models as the encoder during the feature extraction phase, coupled with the U-Net decoder to reconstruct intricate segmentation maps. Specifically, we employed the EfficientNet-B5 model as the encoder, departing from the conventional use of traditional convolutional layers. The design of EfficientNet systematically adjusts the network’s width, depth, and image resolution, resulting in enhanced performance for U-Net in generating finely detailed segmentation maps upon replacing its feature extraction encoder pathway.

EfficientNet models have demonstrated state-of-the-art accuracy while maintaining smaller sizes and faster processing speeds compared to existing convolutional neural networks [7]. The proposed network architecture utilizes the first four layers in the contracting path and the last four layers in the expansion path, with the fifth layer serving to connect the two. As we traverse deeper into the encoder, the image dimensions gradually diminish. Beginning with images sized at  $256 \times 256 \times 3$ , we progressively downscale them to  $8 \times 8 \times 256$ , effectively capturing the broader image context. The model’s architecture, including channel count, resolution, and bandwidth, is depicted in Figure 1. By systematically scaling the network’s width, depth, and image resolution, EfficientNet optimizes performance. Consequently, when we replace the feature extraction encoder path in U-Net

with CNN modules, it excels in generating finely detailed segmentation maps.

As a result of this modification, the U-Net model was significantly improved, boasting enhanced capabilities in both its encoder and decoder components. Subsequently, we conducted an evaluation of various deep-learning models to replace the U-Net backbones for nuclei segmentation. This evaluation utilized fluorescent microscopy images that had been meticulously annotated at the pixel level. Additionally, in order to tackle the challenge of segmenting overlapping nuclei regions, we introduced a post-processing step [8]. This step effectively resolves touching boundaries, assigns unique labels to each identified cell, and provides morphological measurements for individual nuclei.

Therefore, we evaluated the performance of different deep learning models to replace UNet backbones for nuclei segmentation by using fluorescent microscopy images that were manually annotated at the pixel level. We used the following pre-trained models, ResNet-50, EfficientNet-B5, EfficientNet-B7, Inception-ResNet-v2, Vgg19, Densenet 121, and Mobilenet as the backbone to replace the decoder. DAPI images will be input images to test the efficiency.

## II. METHOD

The code is available on the GitHub webpage: <https://github.com/aishstha/Segment-and-quantify-cells>

### A. SELECTION OF PRETRAINED BACKBONES FOR UNet's ENCODER

This project aimed to assess the performance of different deep learning models in replacing the UNet backbones for nuclei segmentation, utilizing manually annotated fluorescent microscopy images at the pixel level. The pre-trained models to be evaluated include:

- ResNet-50 [9], [10] is a convolutional neural network comprising 50 layers and is renowned for its success in reducing performance errors and enhancing efficiency in object detection, segmentation, and localization [9].
- EfficientNet-B5 [11] and EfficientNet-B7 [11] are engineered to operate within fixed resource constraints, yet they can be scaled up for improved accuracy given additional resources. These models are particularly suitable for smaller-scale tasks and research endeavors operating within limited budgetary constraints. EfficientNet models have consistently achieved state-of-the-art accuracy while maintaining smaller sizes and faster processing speeds than existing convolutional neural networks [7].
- Inception-ResNet-v2 [12], [13] is another residual neural network model that facilitates the direct flow of information from earlier layers to subsequent layers. Trained on over a million images from ImageNet [13], Inception-ResNet-v2 is trained on more than a million images from ImageNet [14] and has proven to be effective on various image types. Inception-ResNet-v2 has

demonstrated effectiveness across various image types as potential replacements for the decoder.

- Vgg19 [15], [16] includes 16 convolutional layers and 3 fully connected layers. It is known for its simplicity and effectiveness in image classification tasks. The architecture comprises several sets of convolutional layers followed by max-pooling layers, with fully connected layers at the end.
- Densenet 121 [17], [18] consists of 121 layers and is known for its dense connectivity pattern. In DenseNet, each layer receives feature maps from all preceding layers as input, and its own feature maps are passed on to all subsequent layers. This dense connectivity facilitates feature reuse, enhances gradient flow, and encourages feature propagation throughout the network.
- Mobilenet [19], [20] consists of a series of depthwise separable convolutional layers followed by pointwise convolutional layers and optional additional layers such as batch normalization and ReLU activation.

### B. DATA COLLECTION

The dataset utilized in this study originates from a prior publication by Cheng et al. [21] The 3D data were captured using a Zeiss LSM 710 imaging system equipped with a Zeiss Apochromat 40X/1.1 water immersion objective lens. Excitation filters were set at 405 and 561 nm, while emission filters were configured to detect signals between 420–480 and 597–700 nm, respectively. Laser intensity was maintained at 1%, and a twin-gate main beam splitter with two wheels, each containing 10 filter positions (resulting in 100 possible combinations), was employed to separate excitation and emission beams. The pinhole aperture was set at “1”, and digital gain was adjusted to approximately  $\frac{3}{4}$  of the maximum gain, ensuring a dynamic range of pixel values between 500–2000 (12 bits). Leveraging the acquired raw data, we introduce a novel dataset termed the ‘Breast Mammary Gland Dataset (BMGD)’, featuring manually annotated masks that provide precise delineation of objects of interest. This dataset serves as the training dataset for deep learning frameworks and comprises 547 images containing over 3000 cell nuclei.

The dataset from the Kaggle Data Science Bowl 2018 [11] competition comprises annotated histopathological images showcasing nuclei with varying shapes, cell types, magnifications, and modalities, including fluorescence or bright fields. Originally, the dataset contained 670 images contributed by diverse research laboratories in various fields of cell biology. To ensure consistency and precision, we manually curated 546 fluorescent images with distinctive characteristics from this pool.

Additionally, the Broad Bioimage Benchmark Collection includes the BBBC022 dataset [22], which comprises 20 plates. Each plate contains 384 wells, with each well featuring 9 fields of view, resulting in a total of 69,120 fields of view. Each field was captured in five channels (detection wavelengths), with each channel stored as a

separate grayscale image file. The images were acquired at a magnification of 20X, corresponding to  $0.656 \mu\text{m}$  per pixel.

### C. DATA PRE-PROCESSING

To begin the pre-processing phase, we utilized a Python script to extract 2D image slices from the 3D dataset, employing a threshold intensity level set above 2000. We opted for 2D images to streamline analysis, as the features of interest were primarily concentrated in a select few slices. Subsequently, we manually filtered out noisy data to ensure precise data curation. Given that the raw images collected vary in size and cannot be directly inputted into our deep learning model, we standardized their dimensions across the dataset by resizing them to  $256 \times 256$  pixels. This resizing step was necessary to accommodate GPU memory limitations and ensure uniform input size for our model.

### D. DATA ANNOTATION

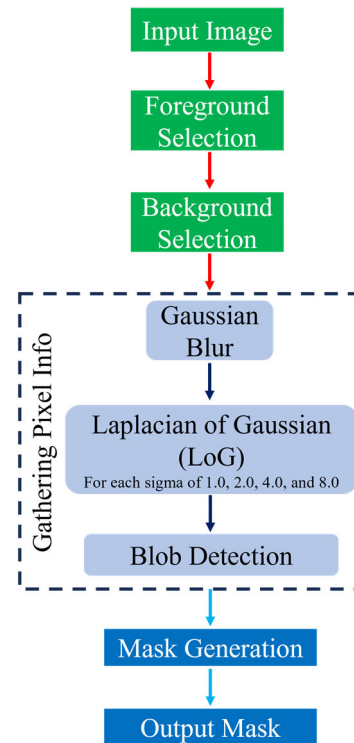
The annotation process comprises several sequential steps. Initially, we undertook manual pixel-based delineation of nuclear boundaries, marking them as foreground by highlighting them in red. Additionally, we designated background examples, which were colored in blue. This facilitates the labeling of a few pixels per class (foreground and background), following which we utilized a random forest classifier integrated into Labkit to generate the mask. We then manually verified the segmented mask against the original image. In cases where noise was detected, Laplacian of Gaussian filters are applied to enhance edge detection. Upon finalizing segmentation, we rescaled image values to a range of 0-1 and converted the mask type to an 8-bit image format. The data annotation process is depicted in **Figure 2**.

### E. DATA AUGMENTATION

To mitigate the risk of overfitting, we implemented augmentation techniques on the images. Our approach leveraged the augmentation library developed by Buslaev et al [23] to accomplish this task. We integrated a range of traditional 2D image processing techniques, including horizontal flipping, random cropping, elastic transformations, shift-scale rotations, and random adjustments to brightness and contrast. These techniques were applied in various combinations to diversify the dataset and enhance the robustness of our model. The data augmentation was performed on the fly, and some of the examples are shown in **Figure 3**.

### F. POST PROCESSING

As a post processing step, we implemented histogram-based segmentation including marker-based watersheds to separate the nuclei that are close to each other. **Figure 4** shows the workflow carried out in this step. For the quantification of nuclei, we labeled each cell with numbering and calculated various metrics by labeling each nucleus, such as area, equivalent diameter, mean intensity, perimeter, small diameter, and large diameter.



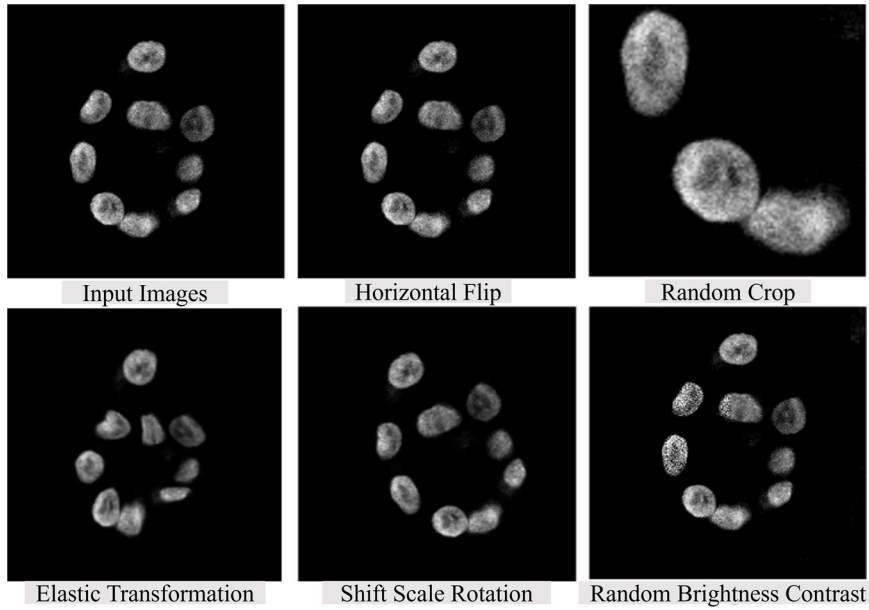
**FIGURE 2.** The flow of data annotation is shown.

### G. TESTING BED AND EXPERIMENTAL SETUP

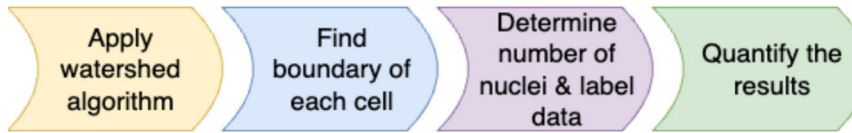
The deep learning framework was implemented using Keras with TensorFlow backend, while data processing was conducted using Python 3.10 alongside popular machine learning libraries such as Keras, NumPy, and pandas. To facilitate model training and evaluation, we divided the dataset using the [train\_test\_split] function from the scikit-learn library. This split allocated 80% of the data for training the proposed deep learning model, with 10% reserved for evaluation and another 10% designated for testing, as detailed in **Table 1**. The training was executed over 100 epochs, with each epoch consuming approximately 2-4 hours to complete. Optimal results were achieved with a batch size of 4, surpassing performance obtained with both 2 and 8, hence this value was selected. Adam was chosen as the optimizer, with a learning rate (LR) of  $10e-4$  during training to maximize loss reduction and tolerate sparse gradients of nucleus foreground pixels. The entire model training process was conducted on a server with specifications outlined in **Table 2**.

**TABLE 1.** Subdivision of dataset is shown.

| Set        | Number of Images | Percentage |
|------------|------------------|------------|
| Train      | 437              | 80 %       |
| Validation | 55               | 10 %       |
| Test       | 55               | 10 %       |
| Total      | 547              | 100 %      |



**FIGURE 3.** Example of transformations applied for data augmentation.



**FIGURE 4.** Pipeline of post-processing method is shown.

**TABLE 2.** Training specifications are shown.

| Parameter     | Percentage        |
|---------------|-------------------|
| Optimizer     | Adam              |
| Callbacks     | ReduceLROnPlateau |
| Learning rate | 10e-4             |
| Batch size    | 4                 |
| Epochs        | 100               |
| Loss function | Combined loss     |

#### H. EVALUATION METRICS

The efficacy of the nuclei segmentation models is assessed through two metrics: mean F1-score (Dice coefficient) and mean IoU (Jaccard Index). These metrics offer distinct advantages in evaluating the accuracy and overlap of segmented regions. By combining precision and recall, the F1 score provides a balanced measure of model performance and is particularly useful for tasks with imbalanced data or varying region sizes. Additionally, the IoU metric quantifies the overlap between predicted and ground truth regions, offering insights into segmentation accuracy. This combined approach ensures a comprehensive assessment of segmentation quality, considering both precision and recall while accounting for potential data imbalances and region size variations.

The F1 score is computed using the following equation:

$$F1 = \frac{TP}{TP + \frac{1}{2}(FP + FN)}$$

where TP, FP, and FN denote true positives, false positives, and false negatives classified by the model. The F1 score ranges from 0 to 1, with a value of 1 indicating all predicted segments align perfectly with the ground truth. Generally, an F1 score larger than 0.7 indicates good segmentation accuracy and the training model is considered a well-performing model for image segmentation tasks.

The Intersection-over-Union (IoU) metric is computed by dividing the area of intersection between the predicted and ground truth masks by the area of their union. The mean IoU is then calculated by dividing the sum of IoU scores for all instances by the total number of instances. The detailed equations for these computations are as follows:

$$IoU = \frac{|A \cap B|}{|A \cup B|}$$

where  $c$  is the number of classes.

The Dataset in this project also includes binary class segmentation, therefore, the sigmoid activation function is used in the final layer, which is computed as follows:

$$\sigma(x) = \frac{1}{1 + e^{-x}}$$

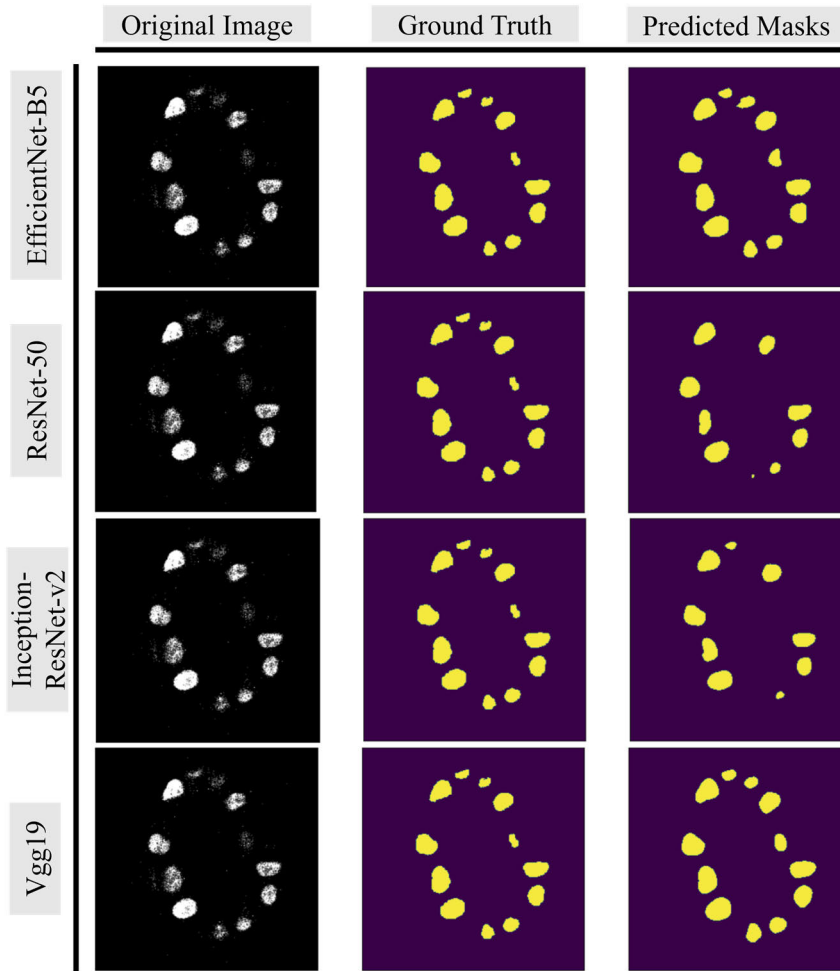


FIGURE 5. Illustration of predicted image in different CNN models is shown.

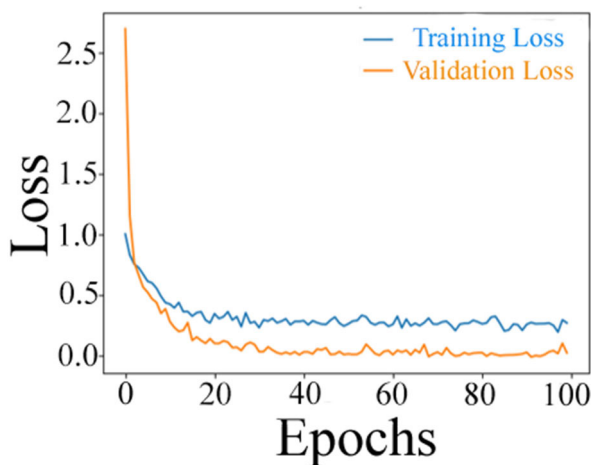


FIGURE 6. The loss per epoch is shown.

The loss function is used by models to learn the trainable parameters, such as weights and biases. Due to the sizable imbalance of overlapping and background regions, we trained the model using a combination of dice loss and binary focal

loss. The purpose of combining these loss functions was to create a composite loss that considers both the similarity between the predicted and ground truth and the difficulty of classifying the pixels. Dice loss uses the dice coefficient to measure the overlapping of the pixels of the predicted labels with the ground truth label. Binary focal loss generalizes binary cross-entropy by introducing a hyperparameter  $\gamma$ , called the focusing parameter, that allows hard-to-classify examples to be penalized more heavily relative to easy-to-classify examples [24]. The formulas for Dice loss, Focal loss, and Total loss are illustrated in the following equations:

$$\text{Dice} = \frac{2 \times |X \cap Y|}{|X| + |Y|}$$

$$\text{Focal Loss } (p_t) = -a_t (1 - p_t) \log(p_t)$$

$$\text{Total Loss} = \text{Dice Loss} + (1 \times \text{Binary Focal Loss})$$

### III. RESULTS AND DISCUSSION

In our feature extraction experiments, we tested several encoders including EfficientNet, ResNet-50, InceptionResNetV2, Vgg19, Densenet121, and MobileNet paired with

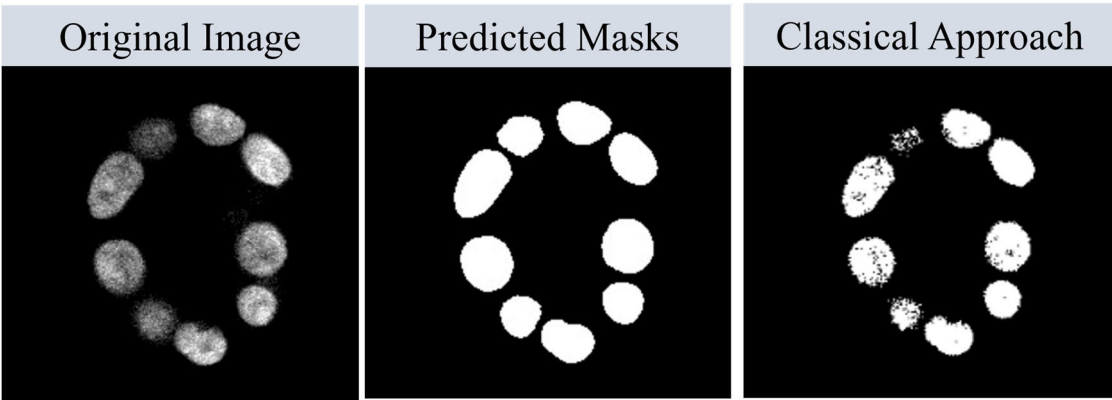


FIGURE 7. The results are shown compared to the original images.

the U-Net decoder, as outlined in Table 3. Figure 5 represents a comparison of the predicted masks that were generated using various CNN models from the same original image. The comparative analysis presents results for the same predicted image across different CNN frameworks employed as U-Net’s encoder. Our findings revealed that using EfficientNet-B5 as the backbone of the nuclei segmentation network yielded the most promising results, achieving an F1 score of 87.1% and a mean IoU score of 80%. This indicates that the combination of Efficient-B5 with U-Net produced superior outcomes. Additionally, we noted a significant enhancement in the F1-score, which rose from 0.85485 to 0.871 when altering the data split from 70-30% to 80-20%. Moreover, our proposed network demonstrated effective generalization, evidenced by a validation loss of 0.02456. Given the notable performance of EfficientNet-B5, further investigation into the performance of various EfficientNet frameworks was deemed prudent. Our findings, presented in Table 4, illustrate that EfficientNet-B5 offers the best cost-performance efficiency. While newer versions such as EfficientNet-B6 and EfficientNet-B7 showed comparable performance. We believe the better performance of EfficientNet-B5 is because the use of multiple MBCov blocks with different receptive field sizes in EfficientNet-B5 for multi-scale feature extraction is particularly advantageous in segmentation tasks where objects of different

TABLE 3. Performance of different CNN encoders in U-Net on BMDG dataset is listed.

| Encoder             | F1-Score | IoU Score | Validation Loss |
|---------------------|----------|-----------|-----------------|
| EfficientNetB5      | 0.87114  | 0.80898   | 0.02456         |
| Resnet50            | 0.77501  | 0.65102   | 0.16675         |
| Inception ResNet V2 | 0.84259  | 0.74521   | 0.13852         |
| Vgg19               | 0.45499  | 0.37303   | 1.09673         |
| Densenet121         | 0.82262  | 0.71279   | 0.14649         |
| Mobilenet           | 0.77661  | 0.64324   | 0.15554         |

TABLE 4. Comparison of EfficientNet frameworks is shown.

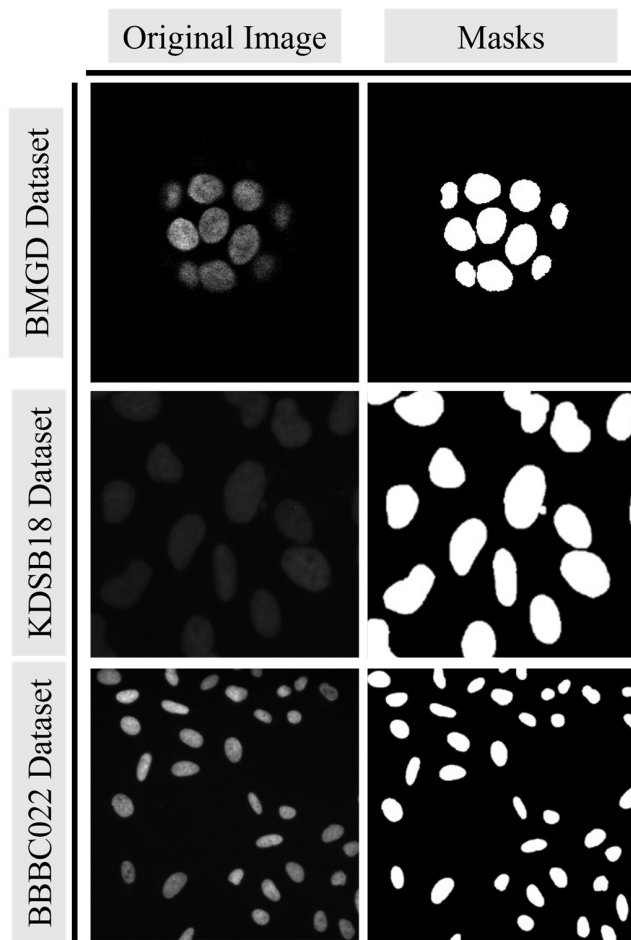
| Encoder        | F1-Score | IoU Score | Total Train Params |
|----------------|----------|-----------|--------------------|
| EfficientNetB0 | 0.84532  | 0.79948   | 10,071,501         |
| EfficientNetB1 | 0.8499   | 0.7896    | 14,225,747         |
| EfficientNetB2 | 0.85753  | 0.79099   | 12,577,137         |
| EfficientNetB3 | 0.86522  | 0.81435   | 17,867,833         |
| EfficientNetB4 | 0.86953  | 0.81842   | 25,607,833         |
| EfficientNetB5 | 0.87114  | 0.80898   | 37,293,953         |
| EfficientNetB6 | 0.85056  | 0.7598    | 50,679,593         |
| EfficientNetB7 | 0.85334  | 0.81177   | 74,735,393         |

sizes need to be accurately delineated. Since our dataset is small, EfficientNet-B7’s greater capacity might not be fully exploited, which would result in declining results.

In Figure 6, we present the loss per epoch, where it’s worth noting that augmentation was not applied to the validation dataset, resulting in a lower validation loss. A low validation loss indicates that the model’s predictions on unseen data closely match the actual target values, underscoring the robust performance of the developed deep learning framework in capturing underlying patterns in the dataset.

Furthermore, Figure 7 illustrates that our deep learning-based model outperforms a classical computer vision-based segmentation approach, the Otsu algorithm [25]. This superiority is particularly noteworthy as the Otsu algorithm struggles to identify an optimal threshold in cases where there is no clear bimodal distribution in the intensity histogram. This observation highlights the shortcomings of traditional computer vision methods when faced with images lacking clear distinctions between background and foreground values. In such scenarios, the newly developed model excels, yielding more precise and nuanced segmentation results.

Additionally, we implemented histogram-based segmentation, including marker-based watersheds, as a post-processing step to separate nuclei that are in close proximity.



**FIGURE 8.** The efficiency of deep learning model on segmenting unseen datasets is shown.

For the quantification of nuclei, we assigned each cell a numerical label and calculated various metrics for each nucleus, such as area, equivalent diameter, mean intensity, perimeter, small diameter, and large diameter.

Finally, we evaluated the effectiveness of the developed deep learning algorithms by testing them on public datasets, namely KDSB18 and BBBC022. These datasets represent new data that our algorithm has never encountered or been trained with. Our results demonstrated successful segmentation, as illustrated in **Figure 8**.

#### IV. CONCLUSION

In this study, we have established a pipeline for the automatic segmentation of nuclei in DAPI-stained fluorescent images, aimed at assisting medical experts in their diagnostic processes through automation. Even with expert oversight, errors can occur, underscoring the necessity for fully automated pipelines.

In conclusion, our study successfully addresses the challenges associated with semantic segmentation of cell nuclei, including issues such as unconventional morphology, noise, and overlapping instances. We employ a U-Net

encoder-decoder-based approach, leveraging a pre-trained EfficientNet-B5 as the network backbone for pixel-level nuclei segmentation. Experimental results demonstrate the effectiveness of our proposed method in extracting nuclei cells, achieving an F1-score of 87% and an IoU of 80%. Compared to state-of-the-art networks like ResNet-50, InceptionResNetV2, Vgg19, Densenet121, and Mobilenet, our model exhibits highly accurate segmentation.

Furthermore, this investigation offers quantified outcomes for the analysis of individual cell nuclei. The manual segmentation conducted on this dataset serves as a valuable resource for training segmentation algorithms, reducing the labor-intensive manual tracing typically required by other researchers.

#### ACKNOWLEDGMENT

The authors would like to thank Dr. Bahram Parvin from the University of Nevada Reno for the inspiration of this project and unconditional help.

#### REFERENCES

- [1] X. Xu, Q. Lu, L. Yang, S. Hu, D. Chen, Y. Hu, and Y. Shi, "Quantization of fully convolutional networks for accurate biomedical image segmentation," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit.*, Jun. 2018, pp. 8300–8308, doi: [10.1109/CVPR.2018.00866](https://doi.org/10.1109/CVPR.2018.00866).
- [2] P. Mergenthaler, S. Hariharan, J. M. Pemberton, C. Lourenco, L. Z. Penn, and D. W. Andrews, "Rapid 3D phenotypic analysis of neurons and organoids using data-driven cell segmentation-free machine learning," *PLOS Comput. Biol.*, vol. 17, no. 2, Feb. 2021, Art. no. e1008630, doi: [10.1371/journal.pcbi.1008630](https://doi.org/10.1371/journal.pcbi.1008630).
- [3] F. Shi, J. Wang, J. Shi, Z. Wu, Q. Wang, Z. Tang, K. He, Y. Shi, and D. Shen, "Review of artificial intelligence techniques in imaging data acquisition, segmentation, and diagnosis for COVID-19," *IEEE Rev. Biomed. Eng.*, vol. 14, pp. 4–15, 2021, doi: [10.1109/RBME.2020.2987975](https://doi.org/10.1109/RBME.2020.2987975).
- [4] S. Minaee, Y. Boykov, F. Porikli, A. Plaza, N. Kehtarnavaz, and D. Terzopoulos, "Image segmentation using deep learning: A survey," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 44, no. 7, pp. 3523–3542, Jul. 2022, doi: [10.1109/TPAMI.2021.3059968](https://doi.org/10.1109/TPAMI.2021.3059968).
- [5] O. Ronneberger, P. Fischer, and T. Brox, "U-Net: Convolutional networks for biomedical image segmentation," in *Proc. Med. Image Comput. Comput.-Assist. Intervent. (MICCAI)*, N. Navab, J. Hornegger, W. M. Wells, and A. F. Frangi, Eds. Cham, Ed., Switzerland: Springer, 2015, pp. 234–241.
- [6] S. Piao and J. Liu, "Accuracy improvement of UNet based on dilated convolution," *J. Phys., Conf. Ser.*, vol. 1345, no. 5, Nov. 2019, Art. no. 052066, doi: [10.1088/1742-6596/1345/5/052066](https://doi.org/10.1088/1742-6596/1345/5/052066).
- [7] M. Tan and Q. Le, "EfficientNet: Rethinking model scaling for convolutional neural networks," in *Proc. 36th Int. Conf. Mach. Learn.*, vol. 97, 2019, pp. 6105–6114. [Online]. Available: <https://proceedings.mlr.press/v97/tan19a.html>
- [8] H. P. Ng, S. H. Ong, K. W. C. Foong, P. S. Goh, and W. L. Nowinski, "Medical image segmentation using K-Means clustering and improved watershed algorithm," in *Proc. IEEE Southwest Symp. Image Anal. Interpretation*, Mar. 2006, pp. 61–65, doi: [10.1109/SSIAI.2006.1633722](https://doi.org/10.1109/SSIAI.2006.1633722).
- [9] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2016, pp. 770–778, doi: [10.1109/CVPR.2016.90](https://doi.org/10.1109/CVPR.2016.90).
- [10] M. B. Hossain, S. M. H. S. Iqbal, M. M. Islam, M. N. Akhtar, and I. H. Sarker, "Transfer learning with fine-tuned deep CNN ResNet50 model for classifying COVID-19 from chest X-ray images," *Informat. Med. Unlocked*, vol. 30, Mar. 2022, Art. no. 100916, doi: [10.1016/j.imu.2022.100916](https://doi.org/10.1016/j.imu.2022.100916).
- [11] J. C. Caicedo, A. Goodman, K. W. Karhohs, B. A. Cimini, J. Ackerman, M. Haghighi, C. Heng, T. Becker, M. Doan, C. McQuinn, M. Rohban, S. Singh, and A. E. Carpenter, "Nucleus segmentation across imaging experiments: The 2018 data science bowl," *Nature Methods*, vol. 16, no. 12, pp. 1247–1253, Dec. 2019, doi: [10.1038/s41592-019-0612-7](https://doi.org/10.1038/s41592-019-0612-7).

- [12] P. M. V. Kumar, A. Venaik, P. Shanmugaraja, P. J. Augustine, and M. Madijagan, "Design of inception ResNet V2 for detecting malarial infection using the cell image captured from microscopic slide," *Imag. Sci. J.*, vol. 72, no. 5, pp. 657–668, Jul. 2024, doi: [10.1080/13682199.2023.2219878](https://doi.org/10.1080/13682199.2023.2219878).
- [13] B. Das, A. Saha, and S. Mukhopadhyay, "Rain removal from a single image using refined inception ResNet V2," *Circuits, Syst., Signal Process.*, vol. 42, no. 6, pp. 3485–3508, Jun. 2023, doi: [10.1007/s00034-022-02279-x](https://doi.org/10.1007/s00034-022-02279-x).
- [14] C. Szegedy, S. Ioffe, V. Vanhoucke, and A. Alemi, "Inception-v4, Inception-ResNet and the impact of residual connections on learning," in *Proc. AAAI Conf. Artif. Intell.*, 2017, vol. 31, no. 1, pp. 4278–4284, doi: [10.1609/aaai.v31i1.11231](https://doi.org/10.1609/aaai.v31i1.11231).
- [15] F. Wu, P. Wang, H. Yang, J. Wu, Y. Liu, Y. Yang, Z. Zuo, T. Wu, and J. Li, "Research on predicting hematoma expansion in spontaneous intracerebral hemorrhage based on deep features of the VGG-19 network," *Postgraduate Med. J.*, vol. 100, no. 1186, pp. 592–602, Mar. 2024, doi: [10.1093/postmj/qgae037](https://doi.org/10.1093/postmj/qgae037).
- [16] P. K. Sahoo, M. K. Panda, U. Panigrahi, G. Panda, P. Jain, M. S. Islam, and M. T. Islam, "An improved VGG-19 network induced enhanced feature pooling for precise moving object detection in complex video scenes," *IEEE Access*, vol. 12, pp. 45847–45864, 2024, doi: [10.1109/access.2024.3381612](https://doi.org/10.1109/access.2024.3381612).
- [17] G. Mohandass, G. H. Krishnan, D. Selvaraj, and C. Sridhathan, "Lung cancer classification using optimized attention-based convolutional neural network with DenseNet-201 transfer learning model on CT image," *Biomed. Signal Process. Control*, vol. 95, Sep. 2024, Art. no. 106330, doi: [10.1016/j.bspc.2024.106330](https://doi.org/10.1016/j.bspc.2024.106330).
- [18] Y. Hou, Z. Wu, X. Cai, and T. Zhu, "The application of improved densenet algorithm in accurate image recognition," *Sci. Rep.*, vol. 14, no. 1, Apr. 2024, doi: [10.1038/s41598-024-58421-z](https://doi.org/10.1038/s41598-024-58421-z).
- [19] M. P. Singh and N. Sinha, "A novel image classification approach via MobileNet model for security assessment of power system," (in English), *Electric Power Compon. Syst.*, pp. 1–14, Mar. 2024, doi: [10.1080/15325008.2024.2329336](https://doi.org/10.1080/15325008.2024.2329336).
- [20] X. Sang, T. Ruan, C. Li, H. Li, R. Yang, and Z. Liu, "A real-time and high-performance MobileNet accelerator based on adaptive dataflow scheduling for image classification," *J. Real-Time Image Process.*, vol. 21, no. 1, Feb. 2024, doi: [10.1007/s11554-023-01378-5](https://doi.org/10.1007/s11554-023-01378-5).
- [21] Q. Cheng, C. C. Bilgin, G. Fontenay, H. Chang, M. Henderson, J. Han, and B. Parvin, "Stiffness of the microenvironment upregulates ERBB2 expression in 3D cultures of MCF10A within the range of mammary density," *Sci. Rep.*, vol. 6, no. 1, p. 28987, Jul. 2016, doi: [10.1038/srep28987](https://doi.org/10.1038/srep28987).
- [22] V. Ljosa, K. L. Sokolnicki, and A. E. Carpenter, "Annotated high-throughput microscopy image sets for validation," (in English), *Nature Methods*, vol. 9, p. 637, Jul. 2012, doi: [10.1038/nmeth.2083](https://doi.org/10.1038/nmeth.2083).
- [23] A. Buslaev, V. I. Iglovikov, E. Khvedchenya, A. Parinov, M. Druzhinin, and A. A. Kalinin, "Albumentations: Fast and flexible image augmentations," *Information*, vol. 11, no. 2, p. 125, Feb. 2020, doi: [10.3390/info11020125](https://doi.org/10.3390/info11020125).
- [24] F. Xing and L. Yang, "Robust nucleus/cell detection and segmentation in digital pathology and microscopy images: A comprehensive review," *IEEE Rev. Biomed. Eng.*, vol. 9, pp. 234–263, 2016, doi: [10.1109/RBME.2016.2515127](https://doi.org/10.1109/RBME.2016.2515127).
- [25] J. Cao, L. Chen, M. Wang, and Y. Tian, "Implementing a parallel image edge detection algorithm based on the otsu-canny operator on the Hadoop platform," *Comput. Intell. Neurosci.*, vol. 2018, May 2018, Art. no. 3598284, doi: [10.1155/2018/3598284](https://doi.org/10.1155/2018/3598284).



**AISHWARYA SHRESTHA** was born in Nepal. She received the M.S. degree in computer science from the University of Wisconsin–Milwaukee, in December 2023. From 2022 to 2023, she was a Research Assistant with the Dr. Qingsu Cheng's Laboratory focusing on fluorescent image segmentations. She is currently a full-time Software Engineer with Northwestern Mutual, Milwaukee, WI, USA.



**XUEFENG BAO** (Member, IEEE) received the B.Sc. degree in physics from Xi'an Jiaotong University, Xi'an, China, in 2010, and the Ph.D. degree in mechanical engineering from the University of Pittsburgh, Pittsburgh, PA, USA, in 2019. He was a Postdoctoral Researcher with the School of Medicine and the Advanced Platform Technology Center, Case Western Reserve University; and U.S. Department of Veterans Affairs Rehabilitation Research Center. He is currently a Visiting Assistant Professor with the Department of Biomedical Engineering, University of Wisconsin–Milwaukee, Milwaukee, WI, USA. His research interests include medical robotics, human motion control, and bio-signal processing.



**QINGSU CHENG** received the B.S. degree in bioengineering from Nanjing University of Technology, Nanjing, China, in 2006, the M.S. degree in biochemical engineering from Sichuan University, Sichuan, China, in 2010, and the Ph.D. degree in biomedical engineering from the University of South Carolina, Columbia, SC, USA, in 2013. From 2013 to 2016, he was a Postdoctoral Researcher with Lawrence Berkeley National Laboratory, and from 2016 to 2022, he was a Research Scientist with the University of Nevada, Reno. He is currently an Assistant Professor with the University of Wisconsin–Milwaukee. His research interest includes the development of AI-assisted algorithms to segment biomolecular endpoints.



**SUSAN MCROY** received the B.S. degree in computer and communication science from The University of Michigan, Ann Harbor, MI, USA, in 1985, and the M.S. and Ph.D. degrees in computer science from the University of Toronto, in 1987 and 1993, respectively. After that, she became a Professor with the University of Wisconsin–Milwaukee. Her work aims to address the difficulty people have in obtaining useful health information from text documents. She hopes to improve overall health by creating and implementing more accurate and easy-to-use forms by using natural language processing, information retrieval, and artificial intelligence.

...