

# Comparative Study of Lightweight Segmentation Models for Medical Cell Imagery in Resource-Constrained Environments

**Chenyu Liu**<sup>1,\*</sup>

<sup>1</sup>School of Medical and Biological Information Engineering, Northeastern University, Shenyang, 111000, China

\*Corresponding author: 2619955@dundee.ac.uk

## Abstract:

Abstract Precise medical cell image segmentation is an essential prerequisite for cell counting, pathological diagnosis, cancer grading and personalized treatment planning, yet mainstream deep learning segmentation models are hard to deploy in grassroots medical institutions and portable microscopic devices due to their massive parameters and high computational complexity. To address this critical issue, this study evaluated three representative convolutional neural network models for medical nuclear segmentation tasks under GPU-free, resource-constrained conditions, using the 2018 Data Science Bowl dataset with diverse tissue types and imaging modalities. The models included the lightweight Half-UNet as the benchmark, transfer learning-based MobileNetV2-UNet with pre-trained weights, and real-time-oriented Fast-SCNN. Experimental results show that MobileNetV2-UNet achieves the optimal Dice coefficient of 0.9175 with relatively low resource consumption, significantly outperforming the benchmark; Fast-SCNN sacrifices some accuracy (Dice coefficient of 0.7504) but gains superior inference speed (215.8 FPS), which is 3.8 times that of the benchmark. This study provides empirical evidence and technical references for efficient automated medical image analysis on low-computing hardware, highlighting the necessity of flexibly selecting lightweight models according to different clinical demands.

**Keywords:** Medical Image segmentation; Nuclear segmentation; Deep learning Fast-SCNN; MobileNetV2

## 1. Introduction

As one of the core tasks in the intersection of computer vision and medical image analysis, the main task of medical image segmentation is to precisely isolate the tissues or organs of interest from complex medical images such as CT, MRI, pathological sections, etc [1,2]. In pathological diagnosis, especially for cell image analysis with microscope, accurate nuclear segmentation is the prerequisite for subsequent cell counting, morphological feature extraction, cancer grading, and the formulation of personalized treatment plans [3]. In the traditional clinical process, pathologists often need to manually observe and label cells with a microscope. This process is not only time-consuming and laborious, but also highly susceptible to subjective fatigue and experience differences, resulting in inter-observer inconsistency in diagnostic results.

In recent years, with the rapid development of deep learning skills, especially the wide application of convolutional neural networks (CNNs), medical image segmentation skill has made a qualitative leap. The fully convolutional network architecture represented by U-Net has become the “gold standard” in the field of medical segmentation [1], thanks to its unique encoder-decoder structure and the skip connection design that can integrate deep semantics and shallow details. Subsequent studies have given rise to a large number of improved models based on U-Net, such as Attention U-Net and Res-UNet, which have further enhanced the segmentation accuracy [4].

Although the existing deep learning models perform well in segmentation accuracy, their application still faces severe challenges, mainly reflecting in the contradiction between the supply and demand of computing power and resources. On the one hand, mainstream high-precision segmentation models often rely on a large number of parameters and heavy floating-point operation procedures [2]. For instance, the parameter count of a standard U-Net model can be as high as over 30M, while models like DeepLabV3+ have even higher requirements for video memory. This usually demands an expensive high-performance GPU workstation for effective training and real-time inference.

On the other hand, in many actual clinical scenarios, computing resources are extremely limited. For instance, medical centers in remote areas, field hospitals, and embedded systems integrated into portable microscopes or Point-of-Care Testing (POCT) devices are often equipped with only standard CPUs or low-end graphics cards [5]. The deployment of massive deep learning models on such hardware inevitably results in excessive inference latency, failing to meet the requirements for real-time clinical computer-aided diagnosis. Consequently, how to

achieve ‘lightweight’ deployment by drastically reducing computational demands without significantly sacrificing segmentation accuracy has emerged as a critical issue that demands urgent resolution in the current field of medical image analysis [5].

This paper aimed to look into how well lightweight models could work and what their limits were in medical cell segmentation tasks when faced with these challenges. Firstly, an experimental platform was built by this study to simulate situations where resources were limited (for instance, a local CPU environment), so that segmentation models of various sizes could be fairly and systematically compared under the same hardware and data preprocessing standards. Secondly, the Fast - SCNN model [6], which was initially created for real - time situations like self - driving cars, was creatively introduced into the area of medical cell segmentation and then compared side by side with the classic U - Net and the MobileNetV2 - UNet based on transfer learning to see if it could be applied to microscopic image analysis. Finally, by means of multi - dimensional quantitative analysis including accuracy (measured by the Dice coefficient), inference speed (measured by FPS), and the number of model parameters, this study thoroughly showed the strong and weak points of different lightweight strategies on medical images, thus offering firm experimental support for choosing models in clinical applications.

## 2. Related work

### 2.1 The current situation of medical image segmentation

Ever since Ronneberger et al. put forward U - Net in 2015, its symmetrical encoder - decoder framework has turned into the standard for medical image segmentation [1]. The central strength of U - Net resides in its skip connections that directly pass on the spatial detail information which gets lost during the encoding process to the decoding phase, thus attaining precise edge segmentation even with small - sample datasets. In order to further enhance performance, subsequent studies have proposed various variants such as U - Net++, suggested by Zhou et al. which narrows the semantic difference between the encoder and decoder via nested and dense skip connections and Attention U - Net [4], proposed by Oktay et al. that brings in attention gate systems allowing the model to concentrate on target areas. However, most of these enhancements gain in accuracy but lose in computational efficiency and the number of parameters making the inference speed slower so they are not suitable for situations where computation is crucial.

## 2.2 Lightweight neural network

Lightweight neural networks came into being to meet the deployment requirements of mobile and embedded devices, and the fundamental concept was to cut down on redundant parameters by optimizing the operational ways of the network structure. The MobileNet series was a typical instance, for MobileNetV1 put forward Depthwise Separable Convolution, breaking down standard convolution into depthwise and pointwise convolutions which significantly decreased the computational volume [7]. MobileNetV2 further brought in Inverted Residual structures and Linear Bottlenecks, strengthening feature extraction abilities while keeping light - weight characteristics [7]. Embedding MobileNet as a backbone network into the U - Net framework had demonstrated great potential in tasks like skin lesion segmentation and retinal vessel segmentation [8].

On the contrary, Fast - SCNN stands for a distinct lightweight method by discarding the conventional U - shaped framework in preference to a two - branch structure: a shallow “Learning to Downsample” branch which is utilized for swiftly drawing out high - resolution spatial particulars and a deep “Global Feature” branch that is used for grasping low - resolution semantic background [6]. This design was originally aimed at fulfilling the real - time processing demands of high - resolution images in self - driving cars and its capacity to retain edge details seems to be in line with the necessities of cell segmentation. Currently, comprehensive assessments of Fast - SCNN in nucleus segmentation tasks are still rather few so this paper intends to fill this deficiency.

## 3. Methodological

### 3.1 Dataset and preprocessing

The core of this study lies in experimentally validating the effectiveness of different lightweight strategies on medical images. The 2018 Data Science Bowl (DSB2018) nuclei segmentation dataset is utilized [3]. This dataset contains 670 training images covering a variety of tissue types, different magnifications, and multiple imaging modalities. Such high data diversity poses a severe challenge to the model’s generalization ability.

To adapt to the resource-constrained local computing environment and simulate storage bottlenecks potentially encountered in real clinical settings, a series of lightweight preprocessing strategies are implemented. Since the original image dimensions varied, directly inputting them into the model would lead to memory overflow or a surge in computational load; therefore, all input images

and masks are first resized to 256×256 pixels. This size selection strikes a balance between preserving the main morphological features of nuclei and reducing computational overhead. Furthermore, addressing the issue in the original dataset where each nucleus in a single image was annotated as an independent mask file, a mask merging strategy is adopted. Specifically, Max Projection is used to merge all nuclei masks corresponding to the same image into a single-channel binary mask, where a pixel value of 1 represents the foreground nucleus and 0 represents the background. Finally, to prevent overfitting on the small sample size, online data augmentation is performed using the Albumentations library, including random horizontal flips, vertical flips, and brightness/contrast adjustments.

### 3.2 Model architecture

This paper selected three models which represented distinct design philosophies for comparative experiments, namely the Half - UNet as the baseline model, the MobileNetV2 - UNet founded on transfer learning, and the Fast - SCNN aimed at real - time processing.

The first model acts as the baseline: U - Net (Half - UNet). In order to guarantee operation in diverse environments without individual graphics cards, a minimalist version named Half - UNet was built which compresses the standard framework by cutting the quantity of convolutional kernel channels in every layer by half and making the initial channel amount 16 while completely maintaining the 4 - layer downsampling and 4 - layer upsampling structure. The purpose of designing this model was to examine the performance lower limit of the classic U - shaped skip connection structure under severe parameter compression, and its parameter quantity is merely around 0.118M, so it is highly appropriate as a reference baseline for assessing the efficiency of other complicated models.

The second model is MobileNetV2 - UNet founded on transfer learning and has a common asymmetric “encoder - decoder” design for the purpose of boosting performance via transfer learning. Its encoder section was substituted with MobileNetV2 which had been pre - trained on the ImageNet dataset [7], and this backbone made use of inverted residual blocks to extract features effectively and the pre - trained weights endowed the model with robust initial feature recognition abilities making it converge more rapidly. The decoder portion retained the standard U - Net upsampling structure and fused the multi - scale features extracted by MobileNetV2 through skip connections. This design aimed to find out whether the strategy of “strong backbone + lightweight convolution” could considerably enhance segmentation accuracy with only a moderate rise in the number of parameters.

The third comparison model is Fast - SCNN, a real - time segmentation network having a non - U - shaped structure and consisting of three crucial modules, namely the Learning to Downsample module, the Global Feature Extractor, and the Feature Fusion module. The Learning to Downsample module decreased image resolution swiftly while maintaining shallow edge details by means of successive convolutional layers and downsampling with a stride of 2, the Global Feature Extractor carried out MobileNet - style depthwise separable convolution blocks at low resolution to grasp the global semantic information of the image at an extremely low computational cost, and finally, the Feature Fusion module combined the outputs of the two branches straightforwardly and effectively before up - sampling them back to the original image size [6], and this model was introduced to check if a non - U - shaped structure could finish cell segmentation tasks more quickly.

### 3.3 Loss function

In the course of model training, given the extremely unbalanced proportion between foreground (nuclei) and background pixels which was common in cell segmentation tasks, depending solely on the traditional Binary Cross Entropy (BCE) loss frequently made the model excessively prefer the prediction of the background class, thus overlooking small - sized cell structures. In order to deal with this problem, this research developed a combined loss function approach, namely the weighted combination of Binary Cross Entropy Loss (BCE Loss) and Dice Loss.

$$L_{total} = L_{BCE} + L_{Dice} \cdot \quad (1)$$

The BCE Loss was mainly in charge of measuring pixel - level classification precision, guaranteeing smoothness of gradients and stability of convergence during the training process while the Dice Loss directly optimized the set overlap measure, effectively dealing with the class imbalance issue and making the model pay more attention to the completeness of the foreground areas, thus enhancing the overall consistency of the segmentation outcomes [9].

## 4. Results

### 4.1 Experimental setup

All experiments were conducted on a resource-constrained local workstation to simulate a real grassroots medical computing environment. In terms of hardware configuration, an Intel Core i9 processor was selected without an independent GPU acceleration card; inference calculations were performed entirely in CPU mode. The software environment was based on the Windows 11 operating system and the PyTorch 1.13 deep learning framework. Regarding training strategy, the Batch Size was set to 8, the Adam optimizer was selected with an initial learning rate of 0.001, combined with a weight decay of 1e-4 to prevent overfitting. All models were trained for 50 Epochs, and the weights performing best on the validation set were saved for testing.

### 4.2 Evaluation index

This study utilized three fundamental measures for a comprehensive assessment of model performance. Firstly, there was the Dice Coefficient which was the primary means of gauging the degree of overlap between the predicted mask and the ground - truth label and the greater the value, the more precise the segmentation. Secondly, Frames Per Second (FPS) served to quantify the model's inference rate and this measure was determined in the CPU setting by taking into account the average time required to consecutively infer a single 256×256 image one hundred times. Finally, Parameter Count (Params) was employed to gauge how much storage space the model occupied and it also indicated the model's potential for deployment on embedded devices.

### 4.3 Analysis of experimental results

The outcomes of the quantitative comparisons regarding the three models on the DSB2018 test set are presented in Table 1 and by conducting an in - depth examination of the experimental data, substantial disparities in terms of accuracy and efficiency within diverse model architectures could be noticed.

**Table 1. The outputs of the quantitative comparisons**

Model	Params	FPS	BestDice	Time/5Epochs*
U-Net (Baseline)	0.118 M	55.7	0.8713	300.81 s
MobileNetV2-U-Net	6.629 M	21.8	0.9175	416.75 s
Fast-SCNN	0.154 M	215.8	0.7504	146.40 s

\*Note: Training time is based on measured data from the first 5 epochs to reflect relative training efficiency.

First, MobileNetV2 - UNet showed complete supremacy in terms of accuracy by attaining the top Dice score of 0.9175 which was far superior to the baseline U - Net (0.8713) and Fast - SCNN (0.7504). Even though its parameter quantity of 6.629 M was much larger than our specially - designed ultra - lightweight baseline U - Net, because of the pre - trained weights from ImageNet, this model had amazingly powerful feature extraction abilities. It converged rapidly and performed better in generalization when handling complex cell forms, although its inference rate on the CPU was just 21.8 FPS, this was more than enough for pathological diagnostic reporting tasks without the need for real - time feedback.

Second, Fast - SCNN made a significant leap in terms of speed performance, attaining an incredible 215.8 frames per second (FPS), which was almost four times faster than the baseline U - Net and nearly ten times quicker than MobileNetV2 - UNet. And interestingly, it had only 0.154 million parameters, which was roughly equivalent to that of the ultra - lightweight U - Net. Although its Dice score of 0.7504 was somewhat low, showing that it was not as accurate as the traditional U - shaped structure in grasping the fine edges of small nuclei, its extraordinarily high frame rate made it the sole feasible choice for situations highly sensitive to latency, like real - time cell flow video analysis and intraoperative rapid navigation.

Finally, the baseline Half - UNet, despite its channel quantity was slashed down to 16 sharply, still kept a high level of accuracy at 0.8713, which once again confirmed the powerful robustness of the U - Net framework in small - sample medical segmentation tasks and it was a fine balance between precision and model dimensions, yet without the backing of large - scale pre - training data, its maximum attainable accuracy was not easy to outshine that of the MobileNetV2 - UNet based on transfer learning.

## 5. Discussion

### 5.1 The trade-off between precision and efficiency

The experimental data in this research clearly demonstrated the intricate trade - off relationship between “Accuracy - Speed - Resources” in medical image segmentation. MobileNetV2 - UNet stood for an “Accuracy First” approach which showed that through incorporating a pre - trained lightweight backbone network one could attain SOTA-level segmentation performance with an acceptable rise in computational cost [8], and this approach was especially fitting for static image analysis situations where the demands for diagnostic accuracy were extremely high

but there was more leniency regarding processing time. While Fast - SCNN represented a “Speed First” strategy, its distinctive dual - branch structure managed to separate the extraction of spatial details from the processing of semantic information thus attaining real - time processing abilities surpassing 200 FPS on a CPU which had invaluable significance for application scenarios highly sensitive to latency like surgical navigation and flow cytometry.

### 5.2 Future outlook

While CNN - based light - weight models functioned well, when considering the stability of the training procedure, MobileNetV2 - UNet had the most seamless convergence curve because of the excellent initialization offered by transfer learning. On the other hand, U - Net and Fast - SCNN, which were trained from the start, demonstrated greater fluctuations during the initial epochs. Moreover, Fast - SCNN’s reduced accuracy in segmenting densely packed small targets such as nuclei indicated that its global feature branch might have forfeited certain high - frequency information vital for cell borders.

Dealing with the limitations of present research, a number of crucial optimization avenues were worthy of exploration back then as well, one of which was the potential combination of Transformer structures, recent investigations put forward models such as LIT - Unet [10], they showed that incorporating the self - attention system of Transformers was beneficial for grasping long - range connections and also solved the problem of the restricted receptive field in CNNs, so future research might consider inserting lightweight Transformer components into the bottleneck sections of MobileNet or Fast - SCNN to strengthen the ability to model large - scale cell forms.

Secondly, given the privacy - sensitive nature of medical data and the existence of data silos, Federated Learning applications constitute a crucial focus area. The models verified in this paper are light - weight and have minuscule volumes, rendering their transmission overhead virtually non - existent, so that they are very fit for being embedded as client models into Federated Learning frameworks such as FedSegNet [11], allowing for cross - institutional collaborative training by utilizing multi - center data while safeguarding patient privacy.

Finally, in order to further make up for the accuracy loss of Fast - SCNN, future endeavors might utilize the high - precision MobileNetV2 - UNet as a teacher network to direct Fast - SCNN during Knowledge Distillation training [5], with the aim of further enhancing segmentation accuracy while keeping the ultra - high speed.

## 6. Conclusion

This paper made a systematic comparison of how U-Net, MobileNetV2 - UNet, and Fast - SCNN performed in medical cell segmentation tasks within an environment where resources were limited and it was discovered that making rational use of lightweight network structures could be an effective way to deal with the difficulties in deployment and the high computational hurdles in medical artificial intelligence. Specifically speaking, when it came to pathological analysis tasks that required extremely high precision, MobileNetV2 - UNet was the best option as it attained a very high precision of 0.9175 under conditions of low computing power by means of transfer learning and for intra - operative monitoring or flow analysis which pursued real - time feedback. Fast - SCNN showed an invaluable nature thanks to its ultra - high speed of 215.8 FPS.

This study also has certain limitations: the experiments only rely on the DSB2018 dataset, and the generalization of models to datasets with more diverse tissue types or rare disease samples remains untested; Fast-SCNN lacks accuracy in segmenting densely packed small nuclei; and only CNN-based models are evaluated without exploring lightweight Transformer or hybrid architectures. Future research would concentrate on using knowledge distillation methods to enhance the accuracy of real - time models as well as looking into the deployment and application of these lightweight models in federated learning situations.

## References

- [1] Ronneberger O, Fischer P, and Brox T, U-Net: Convolutional Networks for Biomedical Image Segmentation, in Medical Image Computing and Computer-Assisted Intervention (MICCAI), 2015, pp. 234-241.
- [2] Zhou S K et al., A Review of Deep Learning in Medical Imaging: Imaging Traits, Technology Trends, and Types of Goal Dimension, Proceedings of the IEEE, vol. 109, no. 5, pp. 936-968, 2021.
- [3] Caicedo J et al., Data Science Bowl 2018, Kaggle, 2018. [Online]. Available: <https://www.kaggle.com/c/data-science-bowl-2018>.
- [4] Minaee S, Boykov Y, Porikli F, Plaza A, Kehtarnavaz N, and Terzopoulos D, Image Segmentation Using Deep Learning: A Survey, IEEE Transactions on Pattern Analysis and Machine Intelligence, 2021, vol. 44, no. 7, pp. 3523-3542.
- [5] Ma J, Wei Y, Zhang Y, and Wang Y, A Comprehensive Review of Knowledge Distillation for Lightweight Medical Image Segmentation, ResearchGate, 2024.
- [6] Poudel R P, Liwicki S, and Cipolla R, Fast-SCNN: Fast Semantic Segmentation Network, arXiv preprint arXiv:1902.04502, 2019.
- [7] Sandler M, Howard A, Zhu M, Zhmoginov A, and Chen L C, MobileNetV2: Inverted Residuals and Linear Bottlenecks, in Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2018, pp. 4510-4520.
- [8] Saiprasad K M et al., Histopathology Image Segmentation Using MobileNetV2 based U-net Model, IEEE Xplore, 2021.
- [9] Sudre C H, Li W, Vercauteren T, Ourselin S, and Cardoso M J, Generalised Dice Overlap as a Deep Learning Loss Function for Highly Unbalanced Segmentations, in Deep Learning in Medical Image Analysis and Multimodal Learning for Clinical Decision Support, 2017, pp. 240-248.
- [10] Behrouz A, Azar Z T, and Mofrad M R K, LIT-Unet: a lightweight and effective model for medical image segmentation, PubMed, 2024.
- [11] Nguyen D C et al., FedSegNet: A Federated Learning Framework for 3D Medical Image Segmentation, International Journal of Ethical AI Application, 2025.