
BOOSTING MEDICAL IMAGE SEGMENTATION PERFORMANCE WITH ADAPTIVE CONVOLUTION LAYER

A PREPRINT

Seyed M.R. Modaresi
LIPN-UMR-CNRS 7030,
Sorbonne University Paris Nord,
Paris, France
modaresi@lipn.univ-paris13.fr

Aomar Osmani
LIPN-UMR-CNRS 7030,
Sorbonne University Paris Nord,
Paris, France
ao@lipn.univ-paris13.fr

Mohammadreza Razzazi
Computer Engineering Department,
Amirkabir University of Technology,
Tehran, Iran
razzazi@aut.ac.ir

Abdelghani Chibani
Laboratory of Images, Signals and Intelligent Systems
Université Paris-Est Créteil,
Paris, France
achibani@gmail.com

April 18, 2024

ABSTRACT

Medical image segmentation plays a vital role in various clinical applications, enabling accurate delineation and analysis of anatomical structures or pathological regions. Traditional convolutional neural networks (CNNs) have achieved remarkable success in this field. However, they often rely on fixed kernel sizes, which can limit their performance and adaptability in medical images where features exhibit diverse scales and configurations due to variability in equipment, target sizes, and expert interpretations. In this paper, we propose an adaptive layer placed ahead of leading deep-learning models such as UTransNet, which dynamically adjusts the kernel size based on the local context of the input image. By adaptively capturing and fusing features at multiple scales, our approach enhances the network's ability to handle diverse anatomical structures and subtle image details, even for recently performing architectures that internally implement intra-scale modules, such as UTransnet. Extensive experiments are conducted on benchmark medical image datasets to evaluate the effectiveness of our proposal. It consistently outperforms traditional CNNs with fixed kernel sizes with a similar number of parameters, achieving superior segmentation Accuracy, Dice, and IoU in popular datasets such as SegPC2021 and ISIC2018. The model and data are published in the open-source repository, ensuring transparency and reproducibility of our promising results.

Keywords Medical Image Segmentation · Deep Learning · Adaptive Convolution

1 Introduction

As a trending subject in the field of image processing and computer vision [Asgari Taghanaki et al., 2021], Medical image segmentation (MIS) involves extracting the boundaries of desired targets, such as tumors, in medical images and determining their class [Luo et al., 2022]. The accurate segmentation of medical images, such as Computed Tomography (CT), Positron-Emission Tomography (PET), and Magnetic Resonance Imaging (MRI), plays a vital role in the diagnosis and treatment of various diseases and assists physicians in patient management, including staging, assessment, and prognosis of the treatment response [Li et al., 2019, Liu et al., 2021, Tian et al., 2021].

Deep learning-based MIS has gained considerable traction in recent years [Chen et al., 2023, Houssein et al., 2021, Devunooru et al., 2021, Asgari Taghanaki et al., 2021, Malhotra et al., 2022, Luo et al., 2022, Kaviani et al., 2022]. A myriad of models has been introduced in the literature for various MIS tasks and clinical outcomes, encompassing multi-

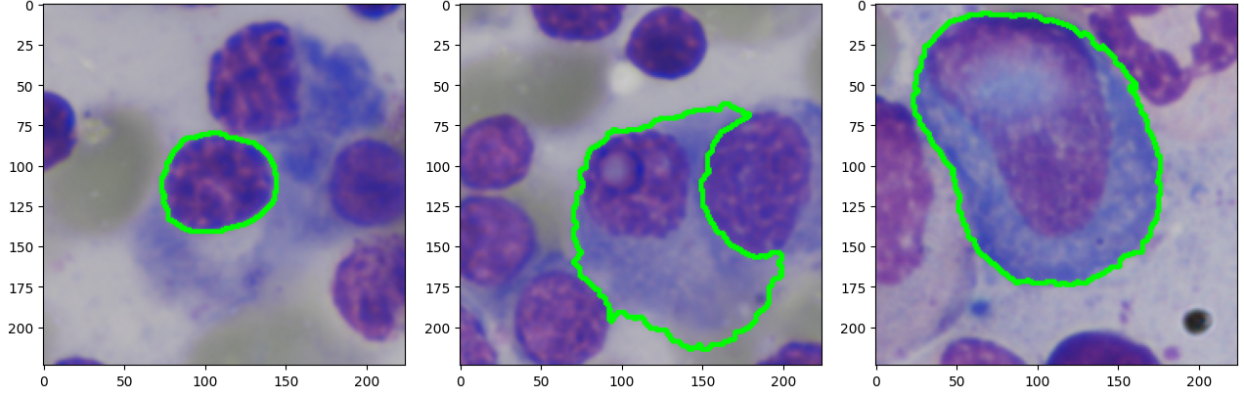


Figure 1: Illustration of diverse scales observed in medical images sourced from the SegPC-2021 dataset. The green contours highlight the regions containing cancerous lesions.

organ detection, tissue mass detection, tumor or nodule segmentation and classification, cell counting, multiple diagnoses, prognosis, and the prediction of treatment outcomes for various chronic diseases like cancers or neurodegenerative diseases [Kumar et al., 2022, Devunooru et al., 2021, Roy et al., 2023, Simpson et al., 2019, Antonelli et al., 2022, Roth et al., 2016, Song et al., 2022]. In the context of cancer diagnosis, for instance, deep models such as [Melekoodappattu et al., 2022, Eali et al., 2022, Iqbal and Qureshi, 2022, Isensee et al., 2021, Singh and Alam, 2022] have shown improved performance in segmenting tumors or nodules. Despite the significant progress, implementing deep learning in MIS continues to pose challenges since medical images often contain noise, artifacts, adhesions, and other distortions that can negatively affect the performance of deep learning models, particularly when discerning tumoral tissue boundaries and surroundings [Kumar et al., 2022, Malhotra et al., 2022, Tian et al., 2021].

CNN models are widely used for analyzing and processing medical images. The fundamental block of CNN is the convolution layer, which contains tiny adjustable weight grids (known as kernels) that convolved on the input image. This work involves moving the kernel over the image, and at every stop, it performs a mathematical operation (dot product) using the weights of filters and the pixel values of the image to get a single output. This process provides Local Receptive Fields for every pixel, assisting the model in concentrating on nearby characteristics. As we go deeper, the model starts recognizing more intricate and broader patterns, enhancing its understanding of the image [Liu et al., 2021].

Typically, kernels of a predetermined size are employed, often configured as a 3×3 grid. This can be likened to a windowing technique where the kernel moves over the image in small segments, processing each segment to create a corresponding output in a new feature map. This process involves the kernel striding across the image, analyzing one small region at a time. A number of research works highlight that the size of these receptive fields plays a crucial role in enhancing the efficiency of the models [Ibtehaz and Rahman, 2020, Milletari et al., 2016, Wang et al., 2021a, Qiu et al., 2018]. Consequently, sticking to a kernel of a fixed size might not work well for every image. For instance, larger input images or targets might require bigger receptive fields. As demonstrated in figure 1, having varying scales in different images suggests the usefulness of having diverse receptive fields to accommodate all of them [Ibtehaz and Rahman, 2020].

To address these challenges, this paper focuses on developing an adaptive layer to be placed ahead of leading deep models, such as UTransNet architecture for MIS. The proposed model aims to overcome the limitations imposed by fixed kernel sizes in traditional CNNs in the first layer, which may not adequately handle diverse anatomical structures, irregular shapes, and varying feature scales in medical images. It dynamically adjusts the kernel size based on the local context of the input image. This adaptive mechanism enables the network to capture relevant multiple scale features from the original input image before feeding to the inner network, leading to improved segmentation accuracy and robustness.

In the experiments, we conduct an extensive evaluation of this architecture using benchmark medical image datasets (ISIC2018 and SegPC2021 described in section 4.2). By comparing the results with traditional CNNs that employ fixed kernel sizes, we demonstrate the superior performance and generalizability of our adaptive approach. To ensure transparency, reproducibility, and consistency in implementation, our method and used datasets, including the state-of-the-art deep models, are published in our open-source repository¹.

¹https://github.com/modaresimr/adaptive_mis

The rest of the paper is organized as follows: Section 2 provides an overview of the related work in medical image segmentation and multi-scale CNN architectures. Section 3 describes the methodology and architecture of the proposed architecture in detail. Section 4 presents the experimental setup, including the datasets, evaluation metrics, the results and analysis, followed by the conclusion and directions for future research in Section 5.

2 Related Works

The well-known U-Net model proposed by [Ronneberger et al. \[2015\]](#) gained significant attention and is an influential architecture in the field of deep learning. Similar to the Autoencoder models, the U-Net model contains the encoder (contracting) and the decoder (expanding) paths. The unique feature of U-Net is the incorporation of skip connections that enable the flow of information from the encoder to the decoder at different scales, facilitating the preservation of spatial details and improving the localization accuracy of the segmentation results [\[Ronneberger et al., 2015, Azad et al., 2022\]](#). Numerous extensions of U-Net have been proposed to improve recognition quality in medical tasks. [Azad et al. \[2022\]](#) provide a comprehensive survey on U-Net and categorize the U-Net extensions into *Skip Connection Enhancements*, *Backbone Design Enhancements*, *Bottleneck Enhancements*, *Transformers*, *Rich Representation Enhancements*, and *Probabilistic Design*. The MISSFormer model [\[Huang et al., 2023\]](#) redesigns the U-Net architecture by incorporating a position-free and hierarchical U-shaped transformer. It utilizes the Enhanced multi-scale Transformer module to bridge the gap between the encoder and decoder feature maps. It has a slightly higher performance in the Synapse dataset [\[Azad et al., 2022\]](#).

The recent UTransNet [\[Wang et al., 2022\]](#) proposes replacing simple skip connections in U-Net with a multi-scale channel-wise module to solve the semantic gaps for an accurate MIS. It also includes an attention mechanism and transformer sub-module. The attention mechanism implicitly learns to suppress irrelevant regions while emphasizing the regions of interest, while the transformer aids in capturing long-range dependencies and addresses the limitation of local receptive fields. The transformer sub-module tokenizes feature maps in each stage within the appropriate patch sizes.

In addition to altering skip connections, transformers, and attention mechanisms, alternative backbones are commonly used to improve U-Net performance. ResNet [\[Cicek et al., 2016\]](#) is also a common backbone for the U-Net architecture which addresses the issues of stacking many layers in deep neural networks that causes vanishing gradient problem. The Google inception module, widely utilized for extracting features across multiple scales, was initially introduced in InceptionV1, where kernels of different sizes were concatenated in parallel [\[Szegedy et al., 2015, 2016, 2017, Zhang et al., 2021\]](#). This architecture underwent further refinements in subsequent versions, with InceptionV2 replacing the 5×5 convolution with two stacked 3×3 convolutions and InceptionV4 breaking down square convolutional kernels into two vectors to reduce computational operations while increasing the receptive field [\[Szegedy et al., 2016, 2017, Al Shoura et al., 2023\]](#). However, this led to a limitation where larger kernels could not be broken down, resulting in fewer selectable features. MultiRes blocks, which employ a series of convolutional layers with residual connections, have been utilized to provide features at different scales, although with limited efficacy for small images and fuzzy objects [\[Ibtehaz and Rahman, 2020, Hossain et al., 2023, Lou et al., 2021\]](#). To overcome these limitations, the dual-channel UNet (DC-UNet) was proposed to incorporate more different-scale features at the cost of increased network parameters and GPU memory consumption [\[Lou et al., 2021, Ansari et al., 2022\]](#). [Gridach \[2021\]](#), [Jiang et al. \[2019\]](#), [Lou et al. \[2022\]](#), [Yang et al. \[2020\]](#), [Fu et al. \[2023\]](#), [Wang et al. \[2021b\]](#), [Zhan et al. \[2023\]](#) consider fixed numbers of features for each multi-scale dilated atrous receptive field in parallel to not increase the computational complexity, which is another way of representing information at various scales. This strategy increases the receptive field of the layer without adding additional network parameters. A convolutional filter in a CNN can be decomposed as a linear combination of pre-fixed bases particularly Fourier-Bessel bases [\[Qiu et al., 2018\]](#). [Wang et al. \[2021a\]](#) combine the idea of adaptive convolutional kernel and combination of pre-fixed bases and replace all convolution filters with adaptive atoms which are shown slightly better in Image Classification tasks particularly when dealing with intra-image variance, while it is not yet applied for image segmentation tasks. In summary, recently many approaches are proposed to improve segmentation performance that are mainly concentrated on the improving skip connections, including attention mechanism and transformer and improving the backbone such as including dynamic convolution. The UTransNet [\[Wang et al., 2022\]](#) performs better compare to other approaches such as TransUNet [\[Chen et al., 2021\]](#), Residual U-Net [\[Zhang et al., 2018\]](#), MultiResUNet [\[Ibtehaz and Rahman, 2020\]](#), U-Net++ [\[Zhou et al., 2018\]](#), Att-UNet [\[Oktay et al., 2018\]](#) and original U-Net [\[Ronneberger et al., 2015\]](#) over several datasets such as ISIC 2018, and SegPC 2021 [\[Azad et al., 2022\]](#).

Although employing a dynamic receptive field offers theoretical advantages, it is still challenging to implement in practice. On one side, certain approaches necessitate extensive computational resources and memory to calculate the dynamic receptive field; on the other, they may not yield any enhancement in performance. After a comprehensive study

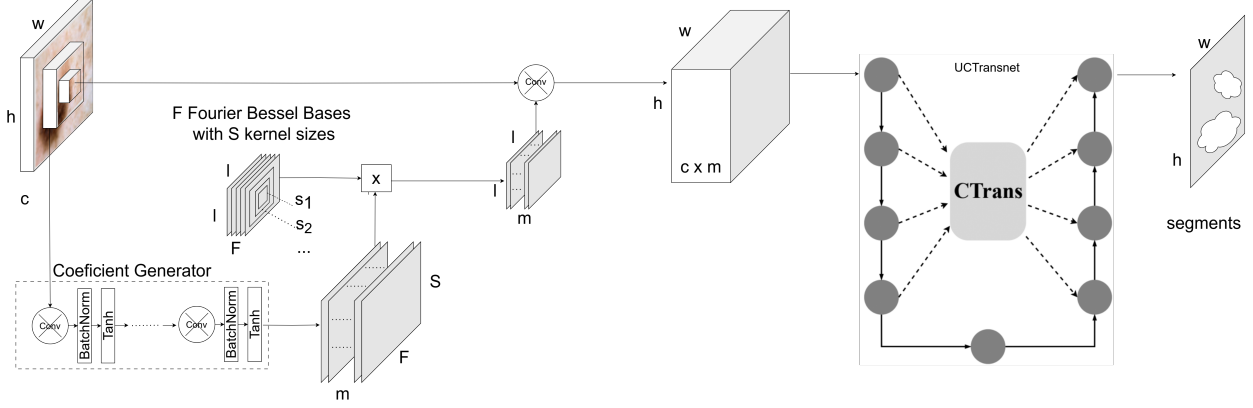


Figure 2: Adaptive Convolution Layer added to the leading UTransNet architecture. The coefficient generator network generates the weights for Fourier-Bessel bases with different sizes for each pixel and channel. It results a fixed kernel to be convolved for that pixel.

of various adaptive methods, the rest of this paper will present a novel dynamic receptive field layer that can be placed ahead of leading models. It improves the recognition performance while maintaining a similar number of parameters.

3 Proposed Approach

In this section, we present an adaptive convolution layer to be incorporated into leading models. Though it is not exclusive to UTransNet, we showcase its integration at the top of this architecture, naming it AdaptUTransNet. The AdaptUTransNet dynamically adjusts the kernel size based on the local context of the input image, enabling the network to capture relevant features at multiple scales from the first layer, leading to improved segmentation accuracy, robustness, and better consider diverse anatomical structures, irregular shapes, and varying feature scales in medical images. Additionally, leveraging the benefits of the UTransNet architecture, which integrates U-Net and transformer models, AdaptUTransNet combines spatial information with self-attention mechanisms to extract meaningful features from medical images.

In MIS, the presence of diverse segments necessitates the adoption of a flexible approach. By adjusting the kernel size based on the specific context of each pixel, we can have a higher level of discernment when extracting features within the inner layers of deep networks. By liberating the network from strict reliance on the hyperparameter associated with the dataset for determining the kernel size, we empower it to adapt and optimize its performance according to the unique characteristics of the input data. Therefore, the proposed approach is to add the adaptive multi-size kernel convolution layer to the best deep learning model for MIS. We leverage the fact that convolution layers can be mathematically expressed as a linear combination of predetermined bases [Qiu et al., 2018]. By employing a limited number of Fourier-Bessel bases, we substantially reduce the number of parameters. Notably, this reduction in parameters does not compromise the accuracy of image classification tasks [Qiu et al., 2018, Wang et al., 2021c], while it has not yet been explored for MIS. Additionally, using Fourier-Bessel bases improve the recognition of the structural information of the input image and effectively mitigates the impact of high-frequency noise and addresses the computational complexities associated with employing multiple kernel sizes within the convolution layer.

Formally, the symbol $[p]$ is employed to denote the spatial coordinates on a feature map Z . These coordinates are applicable to a range of dimensional structures, extending from one-dimension to more complex, higher-dimensional forms. The notation $\mathcal{N}_{Z[p]}^\delta$ is used to represent the receptive field surrounding the feature vector $Z[p]$, where the distance is within a δ . The size of this receptive field can vary, ranging from being quite small to encompassing the entirety of the input feature. Considering this receptive field, the \mathcal{T} is used to represent transformed inner receptive field. For instance, in the dilated receptive field, \mathcal{T} selects a subset of input features at specified intervals. Conversely, in the case of a traditional convolution layer, \mathcal{T} encompasses all pixels, incorporating them without any alterations. Then, traditionally, this inner receptive field is convolved with a set of kernels, denoted as K . The convolution operation can be mathematically expressed as $Z'[p] = K * \mathcal{T}(\mathcal{N}_{Z[p]}^\delta)$, where Z' denotes the resultant feature map post-convolution, and $*$ operation sums the element-wise product. i.e., this can be expressed as $K * N = \sum_q (K \circ N)_q$, where \circ signifies Hadamard element-wise multiplication and the sum is taken over all elements of the product. These kernels are learned end-to-end by the network during the training process, typically via backpropagation and gradient descent.

The inclusion of these kernels results in the addition of extra parameters to the network. As expected, using larger kernels further increases the total parameter count. Therefore, it will be more challenging to dynamically select the most suitable kernel size for each spatial coordinate.

Based on Qiu et al. [2018], a convolution kernel can be decomposed as a combination of Fourier-Bessel bases. Therefore, instead of using a learnable kernel, we can learn the weights (W) for the pre-fixed Fourier-Bessel bases with different sizes (FS). Therefore, $Z'[p] = FS \times W * \mathcal{T}(\mathcal{N}_{Z[p]}^\delta)$. For adaptively changing the receptive field, we use another inner network to learn these weights (W) based on the receptive field $\mathcal{N}_{Z[p]}^\delta$. This inner network, called the coefficient generator network, is trained end-to-end with the backpropagation and gradient descent. In the process of selecting appropriate weights for kernels of varying sizes, we stack several layers of smaller kernels to control the complexity, as suggested by Simonyan and Zisserman [2015]. This approach ensures that the output comprehensively covers the entire receptive field. For example, by stacking a minimum of four layers of 3×3 kernels, we can achieve 9×9 receptive field. This multi layer network convolved through the receptive field $\mathcal{N}_{Z[p]}^\delta$ with a smaller kernel size, and the output of this network is the local weights ($W(\mathcal{N}_{Z[p]}^\delta)$) for the pre-fixed Fourier-Bessel bases with different sizes (FS). This inner network remains fixed across all receptive fields. Therefore, this approach maintains the translation invariance characteristic of convolutional networks while taking into account the local context.

Building upon the specific attributes of medical images, we present our framework to enhance the accuracy of the best segmentation deep network. The inclusion of our adaptive layer does lead to an increase in the number of parameters, but this increase constitutes only a minor fraction of the entire parameters of the network (section 4.4). A graphical representation of the network architecture is provided in figure 2. Given set F of Fourier-Bessel bases with $|S|$ different sizes denoted as FS , for every pixel and channel another receptive field (local neighbors of the corresponding pixel which is greater than S) will be convolved in the coefficient generator network using smaller kernels. This network will be trained end to end in the training phase of the whole network. It produces W with size $|F| \times |S| \times m$ weights, where m is the number of intermediate channels. Matrix multiplication of Fourier-Bessel bases (FS) and these weights (W), generates a kernel for the given pixel and channel. The adaptive layer convolves the input image with the explained kernel, resulting in m intermediate features for each pixel and channel. Then, these intermediate features will be fed to the leading segmentation technique UTransNet [Wang et al., 2022], which replaces the simple skip connections in U-Net with a multi-scale channel-wise module. By using this combination, we can improve the performance of the model.

4 Experiments

In the experiments, we conduct an extensive evaluation of the adaptive layer added ahead of UTransNet, AttUNet and well-known U-Net architecture using benchmark medical image datasets. By comparing the results with traditional CNNs that employ fixed kernel sizes, we demonstrate the superior performance and generalizability of our adaptive approach. Experiments are conducted on various public testbeds, including the Multiple Myeloma Plasma Cell Segmentation (SegPC) 2021 [Gupta et al., 2023, 2021] and the International Skin Imaging Collaboration (ISIC) 2018 datasets [Codella et al., 2019], which will be explained in details in the next sub section. Then, after explaining the details of implementation, we present the experimental results.

4.1 Environment Setup

In order to foster transparency and repeatability of our work, All the codes, datasets, and documentation are freely accessible on our GitHub repository: https://github.com/modaresimr/adaptive_mis. All experiments are run on an NVIDIA DGX-1 machine featuring a Tesla V10032 GPU, Intel Xeon E5-2698v4 CPUs, and 512 GB of RAM. However, we use only a part of these resources.

4.2 Datasets

Experiments are conducted on various public testbeds, including the Multiple Myeloma Plasma Cell Segmentation (SegPC) 2021 [Gupta et al., 2023, 2021] and the International Skin Imaging Collaboration (ISIC) 2018 datasets [Codella et al., 2019]. SegPC contains a collection of 775 microscopic 2D images from the bone marrow samples of Multiple Myeloma patients. It significantly helped hematologists in making more accurate diagnoses and facilitates cancer screening. The dataset from ISIC 2018 boasts a large collection of 2,594 RGB dermoscopy images. Robust segmentation of these images plays a crucial role in medical diagnosis and is challenging due to inconsistent lighting conditions, varying lesion sizes, texture disparities, and differences in color and positioning. Moreover, the presence of unrelated elements like air bubbles, hair strands, or ruler markers further add to the complexity [Hasan et al., 2020, Codella et al.,

2019, Gupta et al., 2023, 2021, Azad et al., 2022]. Both datasets are illustrated in figures 3 and 4 with the segmentation result of our approach.

Similar to the work by Azad et al. [2022], we allocated 70% of images for training, 10% for validation, and the remaining 20% for testing, and our research focused on the segmentation of Cytoplasm components in SegPC 2021 and segmentation of cancer lesions in ISIS 2018 datasets.

4.3 Hyperparameters and Implementation Details

Our pipeline infers the segments from raw image data. All images underwent a sizing down operation to a standard size of 224×224 pixels. The pipeline is composed of an adaptive convolution layer with the kernel size of 3, 5, 7, and 9. We also select six Fourier Bessel bases similar to Wang et al. [2021a]. For the coefficient generator network, we have used six intermediate features (m), which, is responsible for encoding the weights of the prefixed bases. This network will be trained end to end during the global training process. We maintain early stopping with a patience of 20 epochs during the training. For better comparison, we make the other hyperparameters similar to the ones used in [Azad et al., 2022], such as the batch size of 16, epochs limit of 100, Adam optimizer with a learning rate of 0.0001, and the average of cross-entropy loss and dice loss for the loss function. The entire implementation, along with hyperparameters, is accessible and verifiable through our publicly available open-source repository.

4.4 Model Complexity

An essential factor in the assessment of models is the computational complexity. The number of trainable parameters of these components is listed in table 1. Therefore, although the training complexity is similar (the differences are less than 2%), its performance is better regarding to tables 2 and 3.

Table 1: Number of million parameters in each model, including the adaptive variant. This indicates that although our method is effective in determining the ideal dynamic kernel size, it keeps the number of parameters almost the same as those of the original model.

Methods	Normal	with Adaptive Layer
U-Net	19.487	19.850
Att-UNet	34.879	35.242
UCTransNet	66.431	66.794

4.5 Evaluation Metrics

We used a series of performance metrics for a comprehensive analysis of our model’s effectiveness. Precision and Recall served as the primary metrics. Precision assesses the model’s ability to correctly predict positive instances, while Recall measures the completeness of these positive predictions. Although Accuracy gives a general idea of the model’s overall performance, it’s crucial to interpret it alongside the other metrics due to potential data imbalance. We utilized the Intersection over Union (IoU) to measure the overlap between the predicted segmentation and the actual one.

As for the Dice Coefficient, it was used as an alternative to the F1 score due to its increased relevance in medical imaging. It places double emphasis on true positives which is the harmonic mean of precision and recall. It effectively gauges the spatial overlap of the predictions, which is particularly useful in biomedical image segmentation tasks.

4.6 Experimental Results

4.6.1 Cell (SegPC 2021)

We have showcased the visual segmentation results from the SegPC 2021 dataset in figure 3. Our adaptive multi-size-kernel representation effectively demonstrates its aptitude to generate more accurate segmentation for cells of diverse scales and backgrounds. A detailed comparison of results for the SegPC 2021 dataset is provided in table 2, further highlighting the effectiveness of our methodology. They demonstrate that incorporating our adaptive layer into the first layer of the model enhances the extraction of structural information, leading to a better virtual representation. In ??, we have plotted the training and validation loss curves for the SegPC 2021 dataset. These curves indicate the robust performance of the model, as it is neither underfitting nor overfitting.

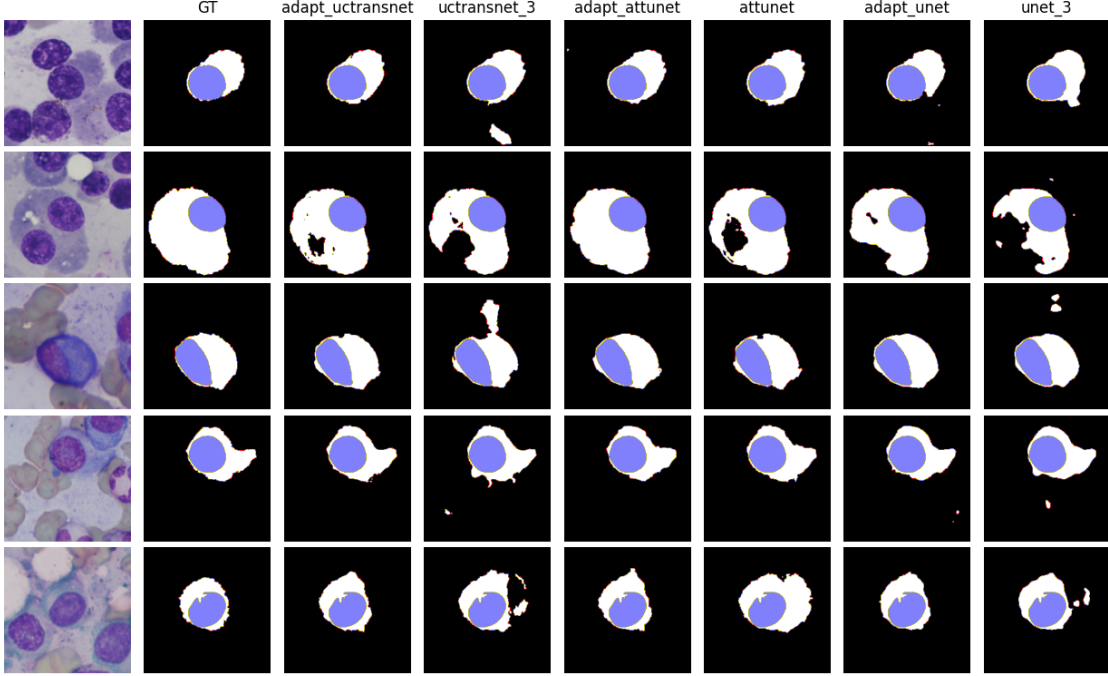


Figure 3: Visual comparisons of different methods for cytoplasm segmentation (depicted as the white region) on the SegPC 2021 dataset. The blue region denotes the Nucleus area of a cell. The initial column displays the input image, while the second column presents the ground truth. Following these, the subsequent columns feature the models along with their adaptive versions. As is evident, models incorporating the adaptive layer more accurately recognize the shape of the cytoplasm, and this improvement is particularly greater in larger segments.

Table 2: Comparison of results for the SegPC 2021 dataset. Each experiment is repeated five times. We have added our adaptive layer to two leading models (AttUNet, UCTransnet), and traditional UNet. As it is shown, even for architectures that internal include multi scale module, such as UCTransnet, adding our adaptive layer on top of that network improves not only the performance of these models but also their consistency (as shown by the standard deviation). For a more comprehensive comparison, other deep models, such as Missformer, ResUNet, and MultiResUNet, are included at the second part of the table. Their models with our adaptive layer are accessible in our repository.

Model	Accuracy	Dice	IoU
adapt_UCTransnet	98.66±0.01	92.11±0.02	91.96±0.02
UCTransnet	98.61±0.04	91.85±0.23	91.71±0.22
adapt_AttUNet	98.71±0.01	92.41±0.03	92.25±0.03
AttUNet	98.65±0.02	92.10±0.08	91.95±0.08
adapt_UNet	98.22±0.01	89.58±0.13	89.60±0.11
UNet	98.07±0.05	88.69±0.35	88.80±0.30
Missformer	98.35±0.04	90.38±0.16	90.32±0.15
ResUNet	97.74±0.04	86.70±0.15	87.04±0.14
MultiResUNet	96.15±0.41	80.29±1.70	81.46±1.41

4.6.2 Skin Cancer (ISIC 2018) dataset

The segmentation result of our approach is illustrated in figure 4. Skin lesions typically manifest within the texture and seldom adhere to a definite shape or geometric pattern. This unpredictable behavior might explain why transformer-based networks may not yield substantial benefits for texture-related patterns [Azad et al., 2022]. Yet again, the adaptive multi-size-kernel representation capability of our methodology demonstrates its proficiency. Compared to other approaches, it is remarkably effective at localizing abnormal regions, which is clearly illustrated in the segmentation results shown in figure 4. This calls for a deeper exploration of the robustness and applicability of our approach. In ??, we have plotted the training and validation loss curves for the ISIC 2018 dataset. The model is performing well, as indicated by the minimal gap (less than 0.003) between the training and validation losses and their stable behavior.

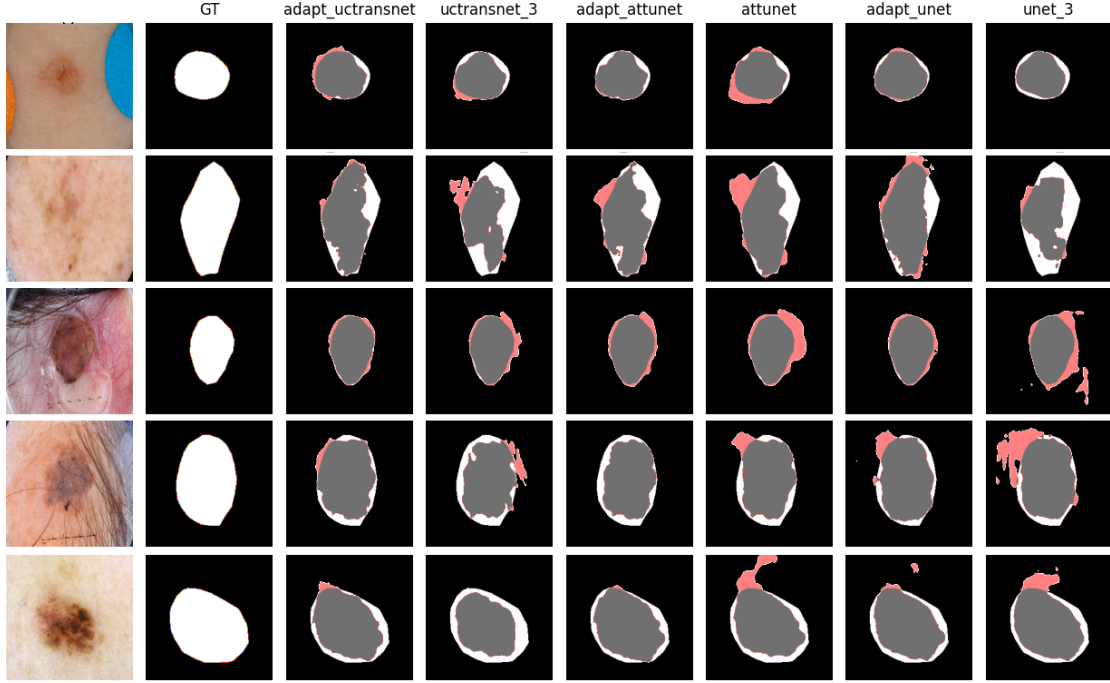


Figure 4: Segmentation output of various deep model in ISIC 2018 dataset. The white region represents the ground truth that remains undetected (False Negative (FN)), while the gray region represents the detected ground truth (True Positive (TP)), and red denotes the False Positive (FP). The columns orders are similar to figure 3. Once again, our model is more effective in identifying target regions, particularly noticeable in larger ones where traditional models with fixed kernels face difficulties in detecting intra-size features.

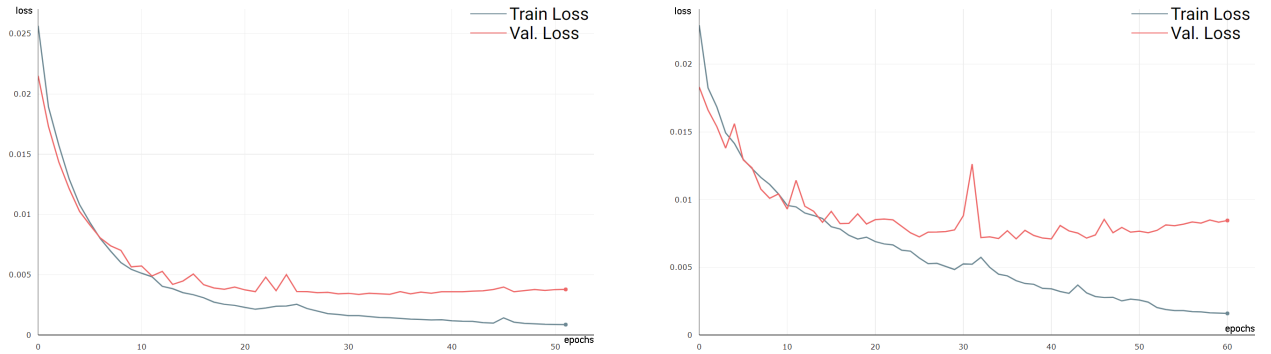


Figure 5: The train loss and validation loss of SegPC 2021 dataset (left) and ISIC 2018 dataset (right). They indicate that the models are neither overfitting nor underfitting.

4.7 Discussion

The conducted experiments substantiate the efficacy of integrating an adaptive layer at the initial stage of deep networks, enhancing their resilience to diverse scales. This layer augments the network’s capability to discern structural information across varying sizes, while maintaining a comparable parameter count. The experiments demonstrate that the inputs including larger segments are better recognized by the proposed method. The noteworthy aspect of these experiments was the enhancement of all existing models through the integration of the adaptive layer, without necessitating any modifications to their structure. This improvement was observed even in models that inherently feature a multi-scale module (such as UTransnet). The accuracy of our experiment and number of parameters aligns with the recent survey by [Azad et al. \[2022\]](#), further validating the credibility of our experimental results.

Table 3: Comparison of results for the ISIC 2018 dataset. It shows the effectiveness of our methodology. Similar to table 2, Each experiment is repeated five times and we have added our adaptive layer to two leading models (AttUNet, UTransnet), and traditional UNet. This approach improves the performance of even the models with internal multi-scale module. For a more comprehensive comparison, other deep models, such as Missformer, ResUNet, and MultiResUNet, are included at the second part of the table. Their models with our adaptive layer are accessible in our repository.

Model	Accuracy	Dice	IoU
adapt_UTransnet	95.64±0.13	89.31±0.18	87.68±0.16
UTransnet	95.54±0.07	89.04±0.27	87.40±0.25
adapt_AttUNet	95.57±0.16	88.96±0.28	87.36±0.32
AttUNet	95.44±0.15	88.66±0.25	87.04±0.29
adapt_UNet	94.80±0.21	87.09±0.29	85.41±0.36
UNet	94.43±0.25	86.18±0.39	84.49±0.46
Missformer	95.25±0.18	88.38±0.42	86.69±0.43
ResUNet	94.35±0.09	85.84±0.07	84.19±0.09
MultiResUNet	92.83±0.63	84.01±1.02	81.80±1.15

5 Conclusion

In this research, we delved into recent advancements in the field of computer vision, focusing on the dynamic modification of the receptive field, a strategy bearing resemblance to the previously described windowing approach. We introduced a novel adaptive layer designed for integration into MIS, aiming to enhance their overall performance. We have shown the effectiveness of our adaptive layer approach by including a dynamic layer on the top of the best segmentation deep network. This approach improves the recognition performance by dynamically changing the receptive field, resulting better identification of structural information and various size targets, and reducing high frequency noises while keeping the number of parameters nearly unchanged. These promising results highlight the potential of our approach for broader applications in medical imaging. Future research could explore adding the adaptive layer to the end of the network, potentially improving output processing, with initial signs suggesting further gains. We also posit that the integration of an adaptive layer ahead of a lower-parameter inner network might achieve, or even surpass, the performance achieved with that higher-parameter inner network, that needs deeper investigation in future studies.

Funding Declaration

This research did not receive any specific grant from public, commercial, or not-for-profit sectors.

Declaration of competing interest

The authors declare that they have no conflict of interest.

Data Availability Statement

The datasets, codes, and framework used in this study are accessible online through our repository at https://github.com/modaresimr/adaptive_mis.

References

- Saeid Asgari Taghanaki, Kumar Abhishek, Joseph Paul Cohen, Julien Cohen-Adad, and Ghassan Hamarneh. Deep semantic segmentation of natural and medical images: a review. *Artificial Intelligence Review*, 54(1):137–178, 1 2021. ISSN 0269-2821. doi:[10.1007/s10462-020-09854-1](https://doi.org/10.1007/s10462-020-09854-1). URL <https://link.springer.com/10.1007/s10462-020-09854-1>.
- Shuai Luo, Yujie Li, Pengxiang Gao, Yichuan Wang, and Seiichi Serikawa. Meta-seg: A survey of meta-learning for image segmentation. *Pattern Recognition*, 126:108586, 2022. ISSN 00313203. doi:[10.1016/j.patcog.2022.108586](https://doi.org/10.1016/j.patcog.2022.108586). URL <https://doi.org/10.1016/j.patcog.2022.108586>.

- Ruijiang Li, Lei Xing, Sandy Napel, and Daniel L Rubin. *Radiomics and Radiogenomics: Technical Basis and Clinical Applications*. CRC Press, Boca Raton, FL, 7 2019. ISBN 9781351208277. doi:[10.1201/9781351208277](https://doi.org/10.1201/9781351208277). URL <https://www.taylorfrancis.com/books/9781351208260>.
- Xiangbin Liu, Liping Song, Shuai Liu, and Yudong Zhang. A Review of Deep-Learning-Based Medical Image Segmentation Methods. *Sustainability*, 13(3):1224, 1 2021. ISSN 2071-1050. doi:[10.3390/su13031224](https://doi.org/10.3390/su13031224). URL <https://www.mdpi.com/2071-1050/13/3/1224>.
- Jie Tian, Di Dong, Zhenyu Liu, and Jingwei Wei. *Radiomics and Its Clinical Application: Artificial Intelligence and Medical Big Data*. Academic Press, an imprint of Elsevier, 2021. ISBN 9780128181010. doi:[10.1016/C2018-0-02044-7](https://doi.org/10.1016/C2018-0-02044-7). URL <https://linkinghub.elsevier.com/retrieve/pii/C20180020447>.
- Gongping Chen, Yuming Liu, Jiang Qian, Jianxun Zhang, Xiaotao Yin, Liang Cui, and Yu Dai. DSEU-net: A novel deep supervision SEU-net for medical ultrasound image segmentation. *Expert Systems with Applications*, 223(March): 119939, 2023. ISSN 09574174. doi:[10.1016/j.eswa.2023.119939](https://doi.org/10.1016/j.eswa.2023.119939). URL <https://doi.org/10.1016/j.eswa.2023.119939>.
- Essam H. Houssein, Marwa M. Emam, Abdelmgeid A. Ali, and Ponnuthurai Nagaratnam Suganthan. Deep and machine learning techniques for medical imaging-based breast cancer: A comprehensive review. *Expert Systems with Applications*, 167(October 2020):114161, 2021. ISSN 09574174. doi:[10.1016/j.eswa.2020.114161](https://doi.org/10.1016/j.eswa.2020.114161). URL <https://doi.org/10.1016/j.eswa.2020.114161>.
- Sindhu Devunooru, Abeer Alsadoon, P. W. C. Chandana, and Azam Beg. Deep learning neural networks for medical image segmentation of brain tumours for diagnosis: a recent review and taxonomy. *Journal of Ambient Intelligence and Humanized Computing*, 12(1):455–483, 1 2021. ISSN 1868-5137. doi:[10.1007/s12652-020-01998-w](https://doi.org/10.1007/s12652-020-01998-w). URL <https://doi.org/10.1007/s12652-020-01998-w>. URL <https://link.springer.com/10.1007/s12652-020-01998-w>.
- Priyanka Malhotra, Sheifali Gupta, Deepika Koundal, Atef Zaguia, and Wegayehu Enbeyle. Deep Neural Networks for Medical Image Segmentation. *Journal of Healthcare Engineering*, 2022:1–15, 3 2022. ISSN 2040-2309. doi:[10.1155/2022/9580991](https://doi.org/10.1155/2022/9580991). URL <https://www.hindawi.com/journals/jhe/2022/9580991/>.
- Sara Kaviani, Ki Jin Han, and Insoo Sohn. Adversarial attacks and defenses on AI in medical imaging informatics: A survey. *Expert Systems with Applications*, 198(March):116815, 2022. ISSN 09574174. doi:[10.1016/j.eswa.2022.116815](https://doi.org/10.1016/j.eswa.2022.116815). URL <https://doi.org/10.1016/j.eswa.2022.116815>.
- Yogesh Kumar, Apeksha Koul, Ruchi Singla, and Muhammad Fazal Ijaz. Artificial intelligence in disease diagnosis: a systematic literature review, synthesizing framework and future research agenda. *Journal of Ambient Intelligence and Humanized Computing*, 2022. ISSN 18685145. doi:[10.1007/s12652-021-03612-z](https://doi.org/10.1007/s12652-021-03612-z).
- Santanu Roy, Devikalyan Das, Shyam Lal, and Jyoti Kini. Novel edge detection method for nuclei segmentation of liver cancer histopathology images. *Journal of Ambient Intelligence and Humanized Computing*, 14(1):479–496, 2023. ISSN 18685145. doi:[10.1007/s12652-021-03308-4](https://doi.org/10.1007/s12652-021-03308-4).
- Amber L. Simpson, Michela Antonelli, Spyridon Bakas, Michel Bilello, Keyvan Farahani, Bram van Ginneken, Annette Kopp-Schneider, Bennett A. Landman, Geert Litjens, Bjoern Menze, Olaf Ronneberger, Ronald M. Summers, Patrick Bilic, Patrick F. Christ, Richard K. G. Do, Marc Gollub, Jennifer Golia-Pernicka, Stephan H. Heckers, William R. Jarnagin, Maureen K. McHugo, Sandy Napel, Eugene Vorontsov, Lena Maier-Hein, and M. Jorge Cardoso. A large annotated medical image dataset for the development and evaluation of segmentation algorithms. *arXiv*, 2 2019. URL <http://arxiv.org/abs/1902.09063>.
- Michela Antonelli, Annika Reinke, Spyridon Bakas, Keyvan Farahani, Annette Kopp-Schneider, Bennett A. Landman, Geert Litjens, Bjoern Menze, Olaf Ronneberger, Ronald M. Summers, Bram van Ginneken, Michel Bilello, Patrick Bilic, Patrick F. Christ, Richard K. G. Do, Marc J. Gollub, Stephan H. Heckers, Henkjan Huisman, William R. Jarnagin, Maureen K. McHugo, Sandy Napel, Jennifer S. Golia Pernicka, Kawal Rhode, Catalina Tobon-Gomez, Eugene Vorontsov, James A. Meakin, Sebastien Ourselin, Manuel Wiesenfarth, Pablo Arbeláez, Byeonguk Bae, Sihong Chen, Laura Daza, Jianjiang Feng, Baochun He, Fabian Isensee, Yuanfeng Ji, Fucang Jia, Ildoo Kim, Klaus Maier-Hein, Dorit Merhof, Akshay Pai, Beomhee Park, Mathias Perslev, Ramin Rezaiifar, Oliver Rippel, Ignacio Sarasua, Wei Shen, Jaemin Son, Christian Wachinger, Liansheng Wang, Yan Wang, Yingda Xia, Daguang Xu, Zhanwei Xu, Yefeng Zheng, James L. Simpson, Lena Maier-Hein, and M. Jorge Cardoso. The Medical Segmentation Decathlon. *Nature Communications*, 13(1):4128, 7 2022. ISSN 2041-1723. doi:[10.1038/s41467-022-30695-9](https://doi.org/10.1038/s41467-022-30695-9). URL <https://www.nature.com/articles/s41467-022-30695-9>.
- Holger R Roth, Amal Farag, Evrim B Turkbey, Le Lu, Jiamin Liu, and Ronald M Summers. NIH Pancreas-CT Dataset, 2016. URL <https://wiki.cancerimagingarchive.net/x/eILXAQ>.
- Jiahuan Song, Xinjian Chen, Qianlong Zhu, Fei Shi, Dehui Xiang, Zhongyue Chen, Ying Fan, Lingjiao Pan, and Weifang Zhu. Global and Local Feature Reconstruction for Medical Image Segmentation. *IEEE Transactions on Medical Imaging*, 41(9):2273–2284, 2022. ISSN 1558254X. doi:[10.1109/TMI.2022.3162111](https://doi.org/10.1109/TMI.2022.3162111).

- Jayesh George Melekoodappattu, Anto Sahaya Dhas, Binil Kumar Kandathil, and K S Adarsh. Breast cancer detection in mammogram: combining modified CNN and texture feature based approach. *Journal of Ambient Intelligence and Humanized Computing*, pages 1–10, 1 2022. ISSN 1868-5137. doi:[10.1007/s12652-022-03713-3](https://doi.org/10.1007/s12652-022-03713-3). URL <https://link.springer.com/10.1007/s12652-022-03713-3>.
- Stephen Neal Joshua Eali, Debnath Bhattacharyya, Thirupathi Rao Nakka, and Seng Phil Hong. A Novel Approach in Bio-Medical Image Segmentation for Analyzing Brain Cancer Images with U-NET Semantic Segmentation and TPLD Models Using SVM. *Traitement du Signal*, 39(2):419–430, 2022. ISSN 19585608. doi:[10.18280/ts.390203](https://doi.org/10.18280/ts.390203).
- Saeed Iqbal and Adnan N. Qureshi. A Heteromorphous Deep CNN Framework for Medical Image Segmentation Using Local Binary Pattern. *IEEE Access*, 10:63466–63480, 2022. ISSN 21693536. doi:[10.1109/ACCESS.2022.3183331](https://doi.org/10.1109/ACCESS.2022.3183331).
- Fabian Isensee, Paul F. Jaeger, Simon A.A. Kohl, Jens Petersen, and Klaus H. Maier-Hein. nnU-Net: a self-configuring method for deep learning-based biomedical image segmentation. *Nature Methods*, 18(2):203–211, 2021. ISSN 15487105. doi:[10.1038/s41592-020-01008-z](https://doi.org/10.1038/s41592-020-01008-z). URL <http://dx.doi.org/10.1038/s41592-020-01008-z>.
- Laxman Singh and Altaf Alam. An efficient hybrid methodology for an early detection of breast cancer in digital mammograms. *Journal of Ambient Intelligence and Humanized Computing*, 13(5), 5 2022. ISSN 1868-5137. doi:[10.1007/s12652-022-03895-w](https://doi.org/10.1007/s12652-022-03895-w). URL <https://doi.org/10.1007/s12652-022-03895-w><https://link.springer.com/10.1007/s12652-022-03895-w>.
- Nabil Ibtihaz and M. Sohel Rahman. MultiResUNet: Rethinking the U-Net architecture for multimodal biomedical image segmentation. *Neural Networks*, 121:74–87, 2020. ISSN 18792782. doi:[10.1016/j.neunet.2019.08.025](https://doi.org/10.1016/j.neunet.2019.08.025). URL <https://arxiv.org/pdf/1902.04049.pdf>.
- Fausto Milletari, Nassir Navab, and Seyed-ahmad Ahmad Ahmadi. V-Net: Fully convolutional neural networks for volumetric medical image segmentation. *Proceedings - 2016 4th International Conference on 3D Vision, 3DV 2016*, pages 565–571, 6 2016. doi:[10.1109/3DV.2016.79](https://doi.org/10.1109/3DV.2016.79). URL <https://arxiv.org/pdf/1606.04797><http://arxiv.org/abs/1606.04797>.
- Ze Wang, Zichen Miao, Jun Hu, and Qiang Qiu. Adaptive Convolutions with Per-pixel Dynamic Filter Atom. In *2021 IEEE/CVF International Conference on Computer Vision (ICCV)*, pages 12282–12291. IEEE, 10 2021a. ISBN 978-1-6654-2812-5. doi:[10.1109/ICCV48922.2021.01208](https://doi.org/10.1109/ICCV48922.2021.01208). URL https://openaccess.thecvf.com/content/ICCV2021/papers/Wang_Adaptive_Convolutions_With_Per-Pixel_Dynamic_Filter_Atom_ICCV_2021_paper.pdfhttps://openaccess.thecvf.com/content/ICCV2021/supplemental/Wang_Adaptive_Convolutions_With_ICCV_2021_supplemental.pdf.
- Qiang Qiu, Xiuyuan Cheng, Robert Calderbank, and Guillermo Sapiro. DCFNet: Deep Neural Network with Decomposed Convolutional Filters. *35th International Conference on Machine Learning, ICML 2018*, 9:6687–6696, 2018. URL <https://arxiv.org/pdf/1802.04145.pdf>.
- Olaf Ronneberger, Philipp Fischer, and Thomas Brox. U-Net: Convolutional Networks for Biomedical Image Segmentation. In *IEEE Access*, volume 9, pages 234–241. IEEE Computer Society, 5 2015. doi:[10.1007/978-3-319-24574-4_28](https://doi.org/10.1007/978-3-319-24574-4_28). URL http://link.springer.com/10.1007/978-3-319-24574-4_28<http://arxiv.org/abs/1505.04597>.
- Reza Azad, Ehsan Khodapanah Aghdam, Amelie Rauland, Yiwei Jia, Atlas Haddadi Avval, Afshin Bozorgpour, Sanaz Karimijafarbigloo, Joseph Paul Cohen, Ehsan Adeli, and Dorit Merhof. Medical Image Segmentation Review: The success of U-Net. *arXiv*, pages 1–38, 2022. URL <http://arxiv.org/abs/2211.14830>.
- Xiaohong Huang, Zhifang Deng, Dandan Li, Xueguang Yuan, and Ying Fu. MISSFormer: An Effective Transformer for 2D Medical Image Segmentation. *IEEE Transactions on Medical Imaging*, 42(5):1484–1494, 5 2023. ISSN 0278-0062. doi:[10.1109/TMI.2022.3230943](https://doi.org/10.1109/TMI.2022.3230943). URL <https://arxiv.org/pdf/2109.07162.pdf><https://ieeexplore.ieee.org/document/9994763/>.
- Haonan Wang, Peng Cao, Jiaqi Wang, and Osmar R. Zaiane. UCTransNet: Rethinking the Skip Connections in U-Net from a Channel-Wise Perspective with Transformer. *Proceedings of the 36th AAAI Conference on Artificial Intelligence, AAAI 2022*, 36:2441–2449, 2022. ISSN 2159-5399. doi:[10.1609/aaai.v36i3.20144](https://doi.org/10.1609/aaai.v36i3.20144).
- Ozgun Cicek, Ahmed Abdulkadir, Soeren S. Lienkamp, Thomas Brox, and Olaf Ronneberger. 3D U-net: Learning dense volumetric segmentation from sparse annotation. *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, 9901 LNCS:424–432, 2016. ISSN 16113349. doi:[10.1007/978-3-319-46723-8_49](https://doi.org/10.1007/978-3-319-46723-8_49). URL <https://arxiv.org/pdf/1606.06650.pdf>.
- Christian Szegedy, Wei Liu, Yangqing Jia, Pierre Sermanet, Scott Reed, Dragomir Anguelov, Dumitru Erhan, Vincent Vanhoucke, and Andrew Rabinovich. Going deeper with convolutions. In *2015 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 1–9. IEEE, 6 2015. ISBN 978-1-4673-6964-0. doi:[10.1109/CVPR.2015.7298594](https://doi.org/10.1109/CVPR.2015.7298594). URL <https://www.cs.unc.edu/~wliu/papers/GoogLeNet.pdf><https://arxiv.org/pdf/1404.5997v2.pdf>.

- [//onlinelibrary.wiley.com/doi/10.1002/jctb.4820](http://onlinelibrary.wiley.com/doi/10.1002/jctb.4820)<http://ieeexplore.ieee.org/document/7298594/>.
- Christian Szegedy, Vincent Vanhoucke, Sergey Ioffe, Jon Shlens, and Zbigniew Wojna. Rethinking the Inception Architecture for Computer Vision. *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, 2016-Decem:2818–2826, 2016. ISSN 10636919. doi:10.1109/CVPR.2016.308. URL https://www.cv-foundation.org/openaccess/content_cvpr_2016/papers/Szegedy_Rethinking_the_Inception_CVPR_2016_paper.pdf.
- Christian Szegedy, Sergey Ioffe, Vincent Vanhoucke, and Alexander Alemi. Inception-v4, Inception-ResNet and the Impact of Residual Connections on Learning. *Proceedings of the AAAI Conference on Artificial Intelligence*, 31(1):11–24, 2 2017. ISSN 2374-3468. doi:10.1609/aaai.v31i1.11231. URL <http://arxiv.org/abs/1512.00567><https://ojs.aaai.org/index.php/AAAI/article/view/11231>.
- Yan Zhang, Yao Lu, Wankun Chen, Yankang Chang, Haiming Gu, and Bin Yu. MSMA Net: A multi-scale mesh aggregation network for brain tumor segmentation. *Applied Soft Computing*, 110:107733, 2021. ISSN 15684946. doi:10.1016/j.asoc.2021.107733. URL <https://doi.org/10.1016/j.asoc.2021.107733>.
- Tariq Al Shoura, Henry Leung, and Bhashyam Balaji. An Adaptive Kernels Layer for Deep Neural Networks Based on Spectral Analysis for Image Applications. *Sensors*, 23(3), 2023. ISSN 14248220. doi:10.3390/s23031527.
- Md Shafayet Hossain, Sakib Mahmud, Amith Khandakar, Nasser Al-Emadi, Farhana Ahmed Chowdhury, Zaid Bin Mahbub, Mamun Bin Ibne Reaz, and Muhammad E.H. Chowdhury. MultiResUNet3+: A Full-Scale Connected Multi-Residual UNet Model to Denoise Electrooculogram and Electromyogram Artifacts from Corrupted Electroencephalogram Signals. *Bioengineering*, 10(5), 2023. ISSN 23065354. doi:10.3390/bioengineering10050579.
- Ange Lou, Shuyue Guan, and Murray H. Loew. DC-UNet: rethinking the U-Net architecture with dual channel efficient CNN for medical image segmentation. In Bennett A. Landman and Ivana Išgum, editors, *Medical Imaging 2021: Image Processing*, page 98. SPIE, 2 2021. ISBN 9781510640214. doi:10.1117/12.2582338. URL <https://arxiv.org/ftp/arxiv/papers/2006/2006.00414.pdf><https://www.spiedigitallibrary.org/conference-proceedings-of-spie/11596/2582338/DC-UNet--rethinking-the-U-Net-architecture-with-dual/10.1117/12.2582338.full>.
- Mohammed Yusuf Ansari, Yin Yang, Shidin Balakrishnan, Julien Abinshed, Abdulla Al-Ansari, Mohamed Warfa, Omran Almokdad, Ali Barah, Ahmed Omer, Ajay Vikram Singh, Pramod Kumar Meher, Jolly Bhadra, Osama Halabi, Mohammad Farid Azampour, Nassir Navab, Thomas Wendler, and Sarada Prasad Dakua. A lightweight neural network with multiscale feature enhancement for liver CT segmentation. *Scientific Reports*, 12(1):1–12, 2022. ISSN 20452322. doi:10.1038/s41598-022-16828-6. URL <https://doi.org/10.1038/s41598-022-16828-6>.
- Mourad Gridach. PyDiNet: Pyramid Dilated Network for medical image segmentation. *Neural Networks*, 140:274–281, 2021. ISSN 18792782. doi:10.1016/j.neunet.2021.03.023. URL <https://doi.org/10.1016/j.neunet.2021.03.023>.
- Yun Jiang, Ning Tan, Tingting Peng, and Hai Zhang. Retinal Vessels Segmentation Based on Dilated Multi-Scale Convolutional Neural Network. *IEEE Access*, 7:76342–76352, 2019. ISSN 21693536. doi:10.1109/ACCESS.2019.2922365.
- Meng Lou, Jie Meng, Yunliang Qi, Xiaorong Li, and Yide Ma. MCRNet: Multi-level context refinement network for semantic segmentation in breast ultrasound imaging. *Neurocomputing*, 470:154–169, 2022. ISSN 18728286. doi:10.1016/j.neucom.2021.10.102. URL <https://doi.org/10.1016/j.neucom.2021.10.102>.
- Tiejun Yang, Yudan Zhou, Lei Li, and Chunhua Zhu. DCU-Net: Multi-scale U-Net for brain tumor segmentation. *Journal of X-Ray Science and Technology*, 28(4):709–726, 2020. ISSN 08953996. doi:10.3233/xst-200650. URL <https://sci.bban.top/pdf/10.3233/XST-200650.pdf#view=FitH>.
- Zhaojin Fu, Jinjiang Li, and Zhen Hua. MSA-Net: Multiscale spatial attention network for medical image segmentation. *Alexandria Engineering Journal*, 70:453–473, 2023. ISSN 11100168. doi:10.1016/j.aej.2023.02.039. URL <https://doi.org/10.1016/j.aej.2023.02.039>.
- Guangjie Wang, Hui Liu, Xianpeng Yi, Jinjun Zhou, and Lin Zhang. ARMS Net: Overlapping chromosome segmentation based on Adaptive Receptive field Multi-Scale network. *Biomedical Signal Processing and Control*, 68(January): 102811, 2021b. ISSN 17468108. doi:10.1016/j.bspc.2021.102811. URL <https://doi.org/10.1016/j.bspc.2021.102811>.
- Bangcheng Zhan, Enmin Song, Hong Liu, Zhenyu Gong, Guangzhi Ma, and Chih Cheng Hung. CFNet: A medical image segmentation method using the multi-view attention mechanism and adaptive fusion strategy. *Biomedical Signal Processing and Control*, 79(P1):104112, 2023. ISSN 17468108. doi:10.1016/j.bspc.2022.104112. URL <https://doi.org/10.1016/j.bspc.2022.104112>.

- Jieneng Chen, Yongyi Lu, Qihang Yu, Xiangde Luo, Ehsan Adeli, Yan Wang, Le Lu, Alan L. Yuille, and Yuyin Zhou. TransUNet: Transformers Make Strong Encoders for Medical Image Segmentation. *arXiv*, pages 1–13, 2 2021. URL <http://arxiv.org/abs/2102.04306>.
- Zhengxin Zhang, Qingjie Liu, and Yunhong Wang. Road Extraction by Deep Residual U-Net. *IEEE Geoscience and Remote Sensing Letters*, 15(5):749–753, 2018. ISSN 15580571. doi:[10.1109/LGRS.2018.2802944](https://doi.org/10.1109/LGRS.2018.2802944). URL <https://arxiv.org/pdf/1711.10684.pdf>.
- Zongwei Zhou, Md Mahfuzur Rahman Siddiquee, Nima Tajbakhsh, and Jianming Liang. Unet++: A nested u-net architecture for medical image segmentation. *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, 11045 LNCS:3–11, 2018. ISSN 16113349. doi:[10.1007/978-3-030-00889-5_1](https://doi.org/10.1007/978-3-030-00889-5_1). URL <https://arxiv.org/pdf/1807.10165.pdf>.
- Ozan Oktay, Jo Schlemper, Loic Le Folgoc, Matthew Lee, Mattias Heinrich, Kazunari Misawa, Kensaku Mori, Steven McDonagh, Nils Y Hammerla, Bernhard Kainz, Ben Glocker, and Daniel Rueckert. Attention U-Net: Learning Where to Look for the Pancreas. *arXiv*, 4 2018. URL <http://arxiv.org/abs/1804.03999>.
- Aiguo Wang, Shenghui Zhao, Chundi Zheng, Jing Yang, Guilin Chen, and Chih Yung Chang. Activities of Daily Living Recognition with Binary Environment Sensors Using Deep Learning: A Comparative Study. *IEEE Sensors Journal*, 21(4):5423–5433, 2021c. ISSN 15581748. doi:[10.1109/JSEN.2020.3035062](https://doi.org/10.1109/JSEN.2020.3035062).
- Karen Simonyan and Andrew Zisserman. Very deep convolutional networks for large-scale image recognition. *3rd International Conference on Learning Representations, ICLR 2015 - Conference Track Proceedings*, pages 1–14, 2015. URL <https://arxiv.org/pdf/1409.1556.pdf>.
- Anubha Gupta, Shiv Gehlot, Shubham Goswami, Sachin Motwani, Ritu Gupta, Avaro Garca Faura, Dejan Stepec, Tomaz Martincic, Reza Azad, Dorit Merhof, Afshin Bozorgpour, Babak Azad, Alaa Sulaiman, Deepanshu Pandey, Pradyumna Gupta, Sumit Bhattacharya, Aman Sinha, Rohit Agarwal, Xinyun Qiu, Yucheng Zhang, Ming Fan, Yoonbeom Park, Daehong Lee, Joon Sik Park, Kwangyeol Lee, and Jaehyung Ye. SegPC-2021: A challenge & dataset on segmentation of Multiple Myeloma plasma cells from microscopic images. *Medical Image Analysis*, 83 (October 2022):102677, 2023. ISSN 13618423. doi:[10.1016/j.media.2022.102677](https://doi.org/10.1016/j.media.2022.102677). URL <https://doi.org/10.1016/j.media.2022.102677>.
- Anubha Gupta, Ritu Gupta, Shiv Gehlot, and Shubham Goswami. Segpc-2021: Segmentation of multiple myeloma plasma cells in microscopic images. *IEEE Dataport*, 1(1):1, 2021. doi:[10.21227/7np1-2q42](https://doi.org/10.21227/7np1-2q42).
- Noel Codella, Veronica Rotemberg, Philipp Tschandl, M. Emre Celebi, Stephen Dusza, David Gutman, Brian Helba, Aadi Kallou, Konstantinos Liopyris, Michael Marchetti, Harald Kittler, and Allan Halpern. Skin Lesion Analysis Toward Melanoma Detection 2018: A Challenge Hosted by the International Skin Imaging Collaboration (ISIC). pages 1–12, 2019. URL <http://arxiv.org/abs/1902.03368>.
- Md Kamrul Hasan, Lavsén Dahal, Prasad N. Samarakoon, Fakrul Islam Tushar, and Robert Martí. DSNet: Automatic dermoscopic skin lesion segmentation. *Computers in Biology and Medicine*, 120(March):103738, 2020. ISSN 18790534. doi:[10.1016/j.compbiomed.2020.103738](https://doi.org/10.1016/j.compbiomed.2020.103738). URL <https://doi.org/10.1016/j.compbiomed.2020.103738>.