

U-Net with Dilated Convolutions and Channel Attention for Cell Segmentation

MSc Research Project
Data Analytics

Abhishek Minz
Student ID: x21173184

School of Computing
National College of Ireland

Supervisor: Vladimir Milosavljevic

National College of Ireland
Project Submission Sheet
School of Computing



Student Name:	Abhishek Minz
Student ID:	x21173184
Programme:	Data Analytics
Year:	2023
Module:	MSc Research Project
Supervisor:	Vladimir Milosavljevic
Submission Due Date:	14/08/2023
Project Title:	U-Net with Dilated Convolutions and Channel Attention for Cell Segmentation
Word Count:	3631
Page Count:	13

I hereby certify that the information contained in this (my submission) is information pertaining to research I conducted for this project. All information other than my own contribution will be fully referenced and listed in the relevant bibliography section at the rear of the project.

ALL internet material must be referenced in the bibliography section. Students are required to use the Referencing Standard specified in the report template. To use other author's written or electronic work is illegal (plagiarism) and may result in disciplinary action.

Signature:	Abhishek Minz
Date:	18th September 2023

PLEASE READ THE FOLLOWING INSTRUCTIONS AND CHECKLIST:

Attach a completed copy of this sheet to each project (including multiple copies).	<input type="checkbox"/>
Attach a Moodle submission receipt of the online project submission , to each project (including multiple copies).	<input type="checkbox"/>
You must ensure that you retain a HARD COPY of the project , both for your own reference and in case a project is lost or mislaid. It is not sufficient to keep a copy on computer.	<input type="checkbox"/>

Assignments that are submitted to the Programme Coordinator office must be placed into the assignment box located outside the office.

Office Use Only	
Signature:	
Date:	
Penalty Applied (if applicable):	

U-Net with Dilated Convolutions and Channel Attention for Cell Segmentation

Abhishek Minz
x21173184

Abstract

The process of manual cell segmentation is a time-intensive and tedious task, frequently influenced by the subjective judgment of the operator. Consequently, the application of deep learning-based semantic segmentation for automating cell segmentation has proven highly beneficial in the analysis of microscopic images. The U-Net architecture, a widely used fully Convolutional Neural Network, achieves state-of-the-art results in biomedical image segmentation. However, despite its excellent performance, it does encounter certain limitations such as loss of spatial information and inconsistent segmentation of images with large variations in artifact size. To address these issues a novel convolutional neural network is proposed by modifying the backbone and skip connection of the U-Net. Non-linear filters have been used in place of simple concatenation of features to reduce the semantic gap and dilated convolutions have been used to extract multi-scale features with channel attention module to better segment images with a wider color spectrum. The proposed model outperforms the baseline U-Net by 1.77% with a mean IoU of 87.92 % and produces better segmentation masks of cellular images.

1 Introduction

Traditional image processing techniques like histogram equalization, thresholding, morphological operations, etc. used to be standard approaches in computer vision but recently a paradigm shift has been observed with deep learning becoming the state of the art Minaee et al. (2022). After the publication of AlexNet Krizhevsky et al. (2012) there has been a tremendous increase in Deep Learning based solutions for computer vision tasks, one of the first such applications had been on biomedical images. Since then Deep Learning, particularly Convolutional Neural Networks(CNNs) has been used extensively in the field of biomedical image analysis such as classification (classifying malignant cells), object detection (lesion detection), and segmentation (cell segmentation), these solutions have helped in drug discovery, early detection and diagnosis of diseases, and have often surpassed human-level performance.

Manual segmentation of cells is a time-intensive, tedious task, is often affected by operator subjectivity, and is thus error-prone, because of this automated cell segmentation using deep learning based semantic segmentation has been very useful in microscopic image analysis. There are some additional challenges associated with microscopic images, they have various staining protocols (hematoxylin and eosin staining, etc), imaging modalities (brightfield microscopy, fluorescence microscopy, etc.) have variations in cell size,

intercell occlusion, etc. and are often low resolution and noisy. Due to the extraneous nature of such images (Figure 1) transfer learning usually performs poorly and training from scratch becomes necessary Alzubaidi et al. (2021).

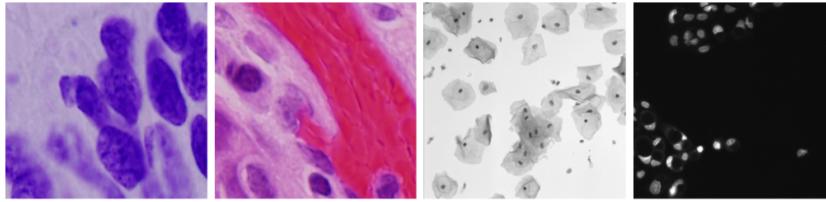


Figure 1: Kaggle Data Science Bowl, Goodman et al. (2018)

The task of segmentation and classification are structurally different since unlike image classification, semantic segmentation has the added complexity of classifying each pixel present in the image, Encoder-Decoder based architectures have been very popular for this with U-Net still being the state of the art in biomedical image segmentation with only slight modifications. The U-Net as proposed by Ronneberger et al. (2015) is a fully Convolutional Neural Network architecture having a contracting path the encoder, and an identical expansive path the decoder with skip connections between them to concatenate features. Despite its excellent performance in biomedical image segmentation, the U-Net has some limitations such as loss of spatial information and difficulty in handling images that are noisy and/or have variations in size Ibtehaz and Rahman (2020). This paper will tackle the problem of cell segmentation in microscopic images by proposing a novel convolutional neural network(CNN) by modifying the backbone and skip connections of the classical U-Net architecture.

The following study will focus on the research question ”**How well can a U-Net with dilated convolutions and attention mechanism perform in cell segmentation?**”. In order to address the research question a CNN architecture is proposed with multi-scale dilated residual blocks and channel attention module. Additionally, the simple concatenation of feature maps from the encoder and decoder is replaced with non-linear filters to reduce the semantic gap and improve segmentation accuracy. To evaluate the effectiveness of the proposed model experiments were conducted on the publically available Kaggle Data Science Bowl 2018 cell microscopy dataset. The main contributions of the paper are as follows:

1. Proposed a novel CNN network with dilated convolutions and channel attention
2. Improved segmentation accuracy in comparison to the baseline U-Net
3. Fewer parameters than the baseline U-Net

The rest of the paper is organized as follows. Related work in semantic segmentation using U-Net is discussed in section 2, the methodology followed for the experimentation is discussed in section 3, and the Design Specification of the proposed model in section 4. Implementation, evaluation, and discussion of the same are discussed in section 5, section 6, and section 7 respectively, lastly the conclusion and future work is discussed in section 8.

2 Related Work

Although the classical U-Net architecture produces excellent results for biomedical image segmentation, despite being state of the art it has some limitations. Many modifications have been done by researchers on the backbone and skip connection of the U-Net to improve segmentation results.

2.1 Semantic Gap

The addition of skip connections in an encoder-decoder architecture is what differentiated the U-Net from other architectures such as VGG-16 when it was first published. Although the skip connection makes it possible to recover fine-grained details and make pixel-level semantic segmentation possible there has been numerous evidence of loss of information because of the encoder-decoder architecture (repeated down-sampling and up-sampling of input features), to address this issue various papers have been proposed such as by Gudhe et al. (2021) where they proposed non-linear multi-level residual blocks in the skip connection to reduce the semantic gap and by Mubashar et al. (2022) where the authors used UNET++ (Zhou et al. (2018)) like dense skip connections for the same.

2.2 Multi-Scale Features

Objects in images can have different shapes and sizes, traditionally this has been addressed by using models such as Google’s Inception Convolutional Neural Network (Szegedy et al. (2015)) where multiple kernels of different sizes are used parallelly to get feature maps of different sizes. Similar work has also been done for biomedical image segmentation to increase the receptive field of the convolution operation by using kernels of different sizes and more recently by using dilated convolutions. Morelli et al. (2021) inserted an additional residual block at the end of the encoding path of the U-Net with a 5x5 kernel instead of a 3x3 kernel to increase the receptive field, Alam et al. (2023) used inception-like blocks with dilated convolutions to increase the receptive field and obtain feature maps of different sizes.

2.3 Attention Mechanism

Microscopic images can be very noisy and have obscure boundaries therefore it is important to build a model robust to such outliers. Attention mechanisms, a popular concept in the field of natural language processing have also been used in computer vision. Attention in the context of computer vision is used primarily in two ways, spatial attention (focuses on where the informative part is located), and channel attention (focuses on the important color channel given an input image). Amer et al. (2022) replaced the skip connections of the U-Net with a spatial attention module to improve segmentation accuracy. Combining both spatial and channel attention has also been done such as by Wang et al. (2023) where a modified U-Net was used for semantic segmentation of remote sensing images.

Although the above papers address one or two of the limitations of the classical U-Net there has not been much work where the three have been addressed together, additionally, research on multi-scale dilated convolutions for feature fusion and its integration with attention mechanisms is fairly recent, and deserves further work. Taking this into consideration, a modified U-Net architecture is proposed for cell segmentation.

3 Methodology

The Kaggle Data Science Bowl Goodman et al. (2018) dataset was chosen for the study, the dataset contains 670 microscopic images and their corresponding pixel-level masks, these images and masks were used in a supervised learning fashion for training the neural networks.



Figure 2: Image and its corresponding Masks

3.1 Data Pre-Processing

Since the raw data had separate masks of each individual cell in an image (shown in Figure 2), the first step was to merge these masks to form a single mask. After obtaining 670 images and corresponding 670 masks they were sent to the data transformation pipeline.

3.2 Data Transformation

The next step was data augmentation and image resizing. In order to avoid overfitting and help the model generalize better Albumentations¹ Python library was used for pixel-level data augmentation techniques such as rotation, horizontal flip, vertical flip, and grid distortion to get 3,350 images and masks. To maintain consistency all the images and masks present in the dataset were then resized to 256x256 pixels, 10% of this augmented data was randomly held back as the test set, and 15% of the remaining data was used as the validation set for guiding the neural network training.

3.3 Model Building

An implementation of the U-Net found in literature and the proposed model was trained with images of size 256x256 pixels for 150 epochs with early stopping criteria. The loss function used was the dice coefficient, the metric chosen for evaluating the models was intersection over union and the optimizer used was 'adam' with a batch size of 32.

3.4 Evaluation

Ablation studies were conducted to better understand the architectures. Finally, a comparison of the proposed and the baseline model was done quantitatively using mean Intersection over union and qualitatively using the segmentation masks produced by the two models.

¹Albumentation Python Library <https://albumentations.ai/>

4 Design Specification

This section describes the model-building process and the functionality of the modules utilized in the proposed model. Furthermore, the choice of the baseline model, loss function, and evaluation metric is also discussed.

4.1 U-Net

In semantic segmentation tasks, state-of-the-art models vary with the dataset and no one model performs well across all datasets. The U-Net architecture is by far the most popular in biomedical image segmentation and most biomedical image segmentation benchmark datasets including the Kaggle Data Science Bowl 2018 have a state-of-the-art solution which is a variation of the classical U-Net architecture taking these into consideration the U-Net model was chosen as the baseline for comparison.

4.2 Proposed Model

Since the classical U-Net architecture performs well with a limited number of data and can be adapted to have different modules and attention mechanisms the proposed model was based on it and was designed with similar principles in mind.

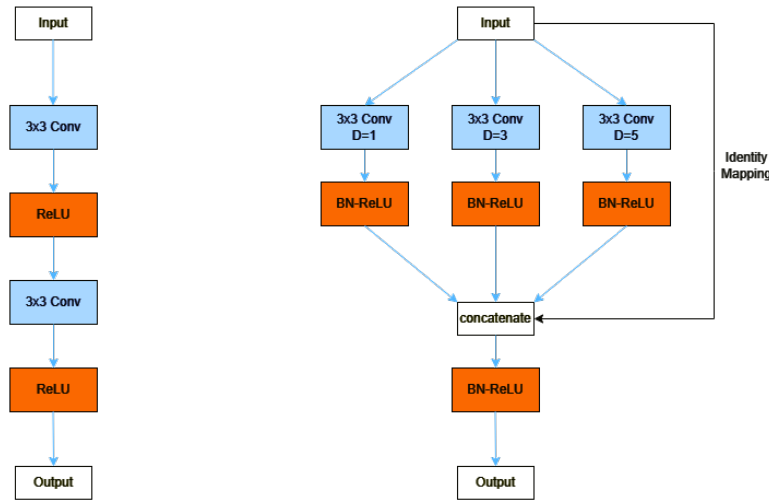


Figure 3: Original U-Net block (left), Multiscale Dilated Residual(MDR) block(right)

Taking inspiration from the MLDR block of Gudhe et al. (2021) the convolutional blocks have been replaced with multiple dilated convolutional blocks having a dilation rate of 1, 3, and 5 with a single level. The number of kernels is divided equally among each convolutional operation and concatenated to get feature maps of the same size as the input feature map coming from the residual connection for element-wise concatenation(identity mapping). The architecture for the same is presented in Figure 3. Different dilation rates would help increase the receptive field and extract spatial information of different scales without drastically increasing the number of parameters Yu and Koltun (2016) and residual connections will make the learning process easier. A sequence of batch

normalization and ReLU activation functions is also used after the convolution operations as suggested by He et al. (2016) to help the model converge faster.

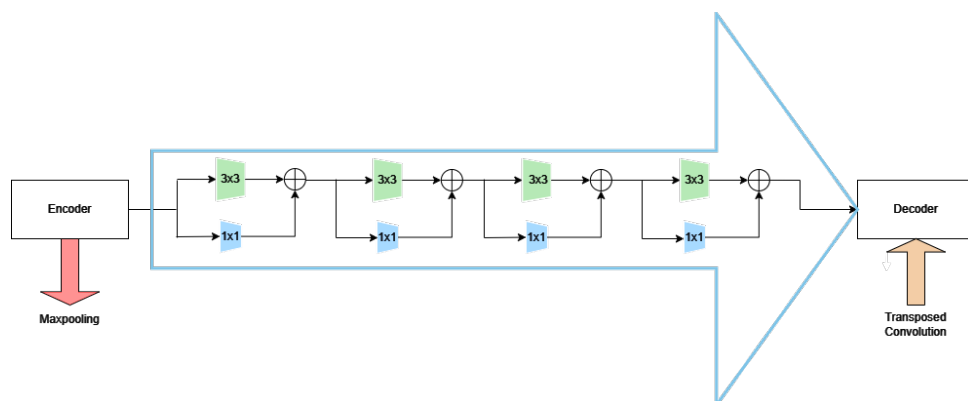


Figure 4: Res Path

As discussed in the literature review, the skip connections of the encoder-decoder architecture of the classical U-Net suffer from a semantic gap, to address this issue the skip connections have been replaced with Res Path as proposed by Ibtehaz and Rahman (2020). The Res Path (shown in Figure 4) consists of a sequence of non-linear filters (3x3 kernel with residual connections), the number of filters is decreased in number as we go deeper in the network (the topmost Res Path has four filters, the second three filters and so on) to accommodate for the lower semantic gap.

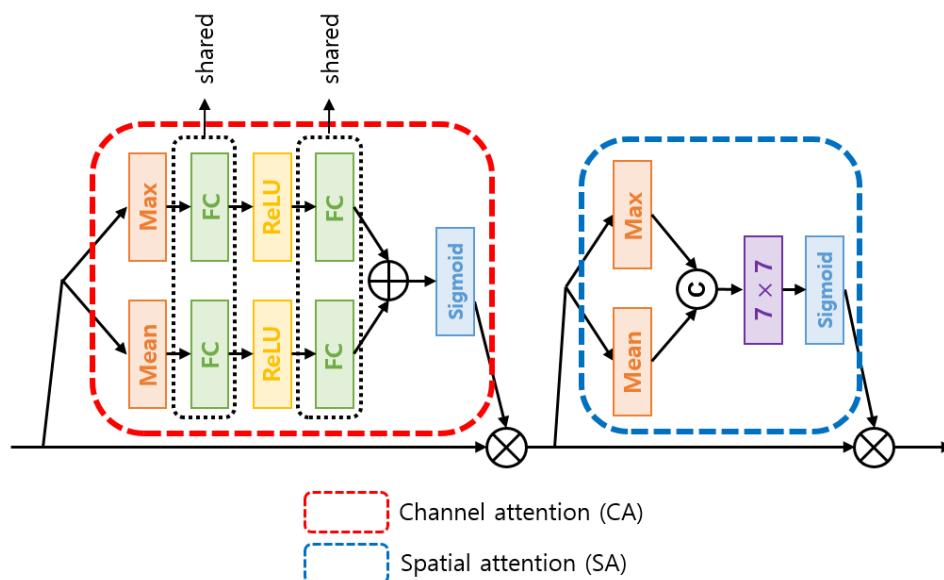


Figure 5: Convolutional Block Attention Module

To better segment histopathological images (larger color space) present in the dataset the channel attention module as proposed in Convolutional Block Attention Module by Woo et al. (2018) (Figure 5) has been added at the end of the encoder path (bottleneck) of the model. The module would apply soft attention weights to each channel to help focus

on the channel with the regions of interest. This is done by average pooling and max pooling of the input feature map of size $H \times W \times C$ to get a more distinctive 1-dimensional channel feature of size $C \times 1 \times 1$ which is multiplied with the input feature map and finally added again to get a more refined and highlighted map.

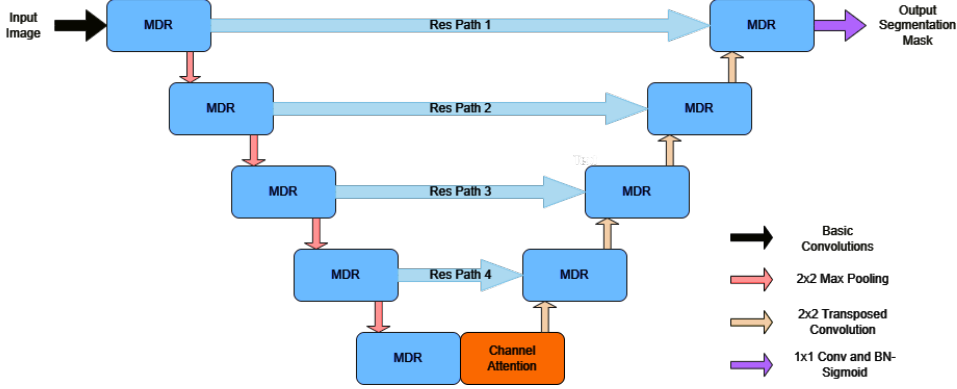


Figure 6: Architecture of the modified U-Net

The architecture of the proposed model with multi-scale dilated residual blocks, res path, and channel attention module is shown in Figure 6. It consists of 5 encoder blocks (kernel size 16, 32, 64, 128, 256), 4 decoder blocks (kernel size 128, 64, 32, 16), and finally a 1x1 convolution at the end of the decoder path with a sigmoid activation function for binary segmentation. A 2x2 max-pooling operation is done after each encoder block and a 2x2 transposed convolution operation with stride 2 is done after each decoder block before the concatenation of feature maps of the encoder and decoder.

4.3 Loss Function

The Dice coefficient, also known as the Sørensen–Dice coefficient is a common metric used to measure the similarity between two sets. Because of its robustness to class imbalance and ease of interpretation (values ranging from 0 to 1), it has also been widely used as a loss function in image segmentation Bertels et al. (2019) and can be formulated as $DSC = \frac{2|X \cap Y|}{|X| + |Y|}$ where X and Y represent the ground truth and predicted mask respectively. In binary classification, it can be framed as $DSC = \frac{2TP}{2TP + FP + FN}$

4.4 Evaluation Metric

Since the images in the dataset predominantly consist of ground truth pixels representing the background, with only a minority of pixels corresponding to the actual cells, accuracy won't be a suitable metric for pixel-level semantic segmentation because the inaccuracies in the minority class will get overshadowed by the accuracy in the majority class Bertels et al. (2019). For this reason, the intersection over union is chosen for the evaluation of the models, mathematically it can be formulated as $Jaccard(X, Y) = \frac{|X \cap Y|}{|X \cup Y|}$ where X and Y represent the ground truth and predicted mask respectively. In binary classification, it can be framed as $JaccardIndex = \frac{TP}{TP + FP + FN}$



Figure 7: Loss Curves

5 Implementation

In this section, the implementation of the proposed model and the U-Net architecture found in the literature for cell segmentation is discussed. The number of kernels in the first five encoder blocks of the two models was set as 16, 32, 64, 128, 256 and the number of kernels in the decoder block as 128, 64, 32, 16, 1. Adam optimizer was used to minimize the loss function(dice coefficient) and both the models were trained for 150 epochs with a batch size of 32. The Intersection over Union (IoU) was calculated after every epoch on the training and validation set to monitor the performance of the models. Moreover, the same seed value was used for train test split, and model training to make sure the same set of images were used for training and testing. After looking at the loss curves of the vanilla U-Net and the proposed model early stopping criteria of 20 and 10 were used respectively to help the model converge without overfitting(different patience values were used and patience values 20 and 10 gave the best results), the same was done for the models in the ablation study. The development environment used Python 3 and Keras v2.11.0 for model building with TensorFlow v2.13.0 as the backend, since the training of the models was done on a cloud GPU the operations were performed deterministically² for the sake of reproducibility.

6 Evaluation

The learning curves of the baseline U-Net and the proposed model is presented in Figure 7. It can be observed that the baseline model has a steeper learning curve in comparison to the proposed model, which can be attributed to the higher learning capacity of the proposed model, therefore, requiring a longer time to learn the needed information.

6.1 Ablation Study

In order to verify the contribution of each module(MDR Block, Res Path, CBAM, and Channel Attention module of CBAM) they were used separately with a vanilla U-Net. The models created using only the MDR blocks, Res Path, CBAM, and channel attention module(taken out from the CBAM) resulted in higher accuracy than the vanilla U-Net,

²https://keras.io/examples/keras_recipes/reproducibility_recipes/

moreover the U-Net with CBAM(Woo et al. (2018)) resulted in higher accuracy than a U-Net with only the channel attention module, this was expected since the spatial and channel attention modules of the CBAM are supposed to be complementary to each other. Looking at the results the channel attention module in the proposed model was replaced with CBAM(channel and spatial Woo et al. (2018)) but no significant change was seen in comparison to the proposed model(MDR block, Res Path, and channel attention). Ablation experiments were done to better understand the network and the results of it are presented in Table 1.

Table 1: Ablation experiments

Channel Attention	Res Path	MDR	mIoU
✓	✗	✗	86.88
✓	✗	✓	87.34
✓	✓	✗	87.82
✓	✓	✓	87.92

It can be seen in Table 1 that the highest accuracy was achieved by a combination of channel attention, res path, and MDR blocks(proposed model), while the U-Net model with only the channel attention module had the lowest accuracy.

6.2 Results

Although having a fewer number of parameters the proposed model had a better mean intersection over union on the test dataset than the baseline U-Net. The proposed model achieved a mean IoU of 87.92 which was 1.77% higher than that of the baseline model.

Table 2: Comparison of model performance

	U-Net	Proposed
mIoU Score(%)	86.39	87.92
Parameters	1,941,105	1,355,904

For qualitative assessment, the segmentation masks of the vanilla U-Net and the proposed model on a few images from the test set with different cell sizes and color space were compared (Figure 8). Looking at the first image it can be seen that they both had an equally good overlap of segmentation masks and ground truth masks, but the vanilla U-Net missed some cells in the second image. The difference becomes even larger when the images have a larger color space where the proposed model outperformed the vanilla U-Net with the latter completely missing the cells in the fourth image.

7 Discussion

The experimental results suggest that the proposed model has better and more reliable segmentation masks in comparison to the baseline model, the benefit of the channel attention module can also be seen in Figure 8, especially in images with a larger colorspace

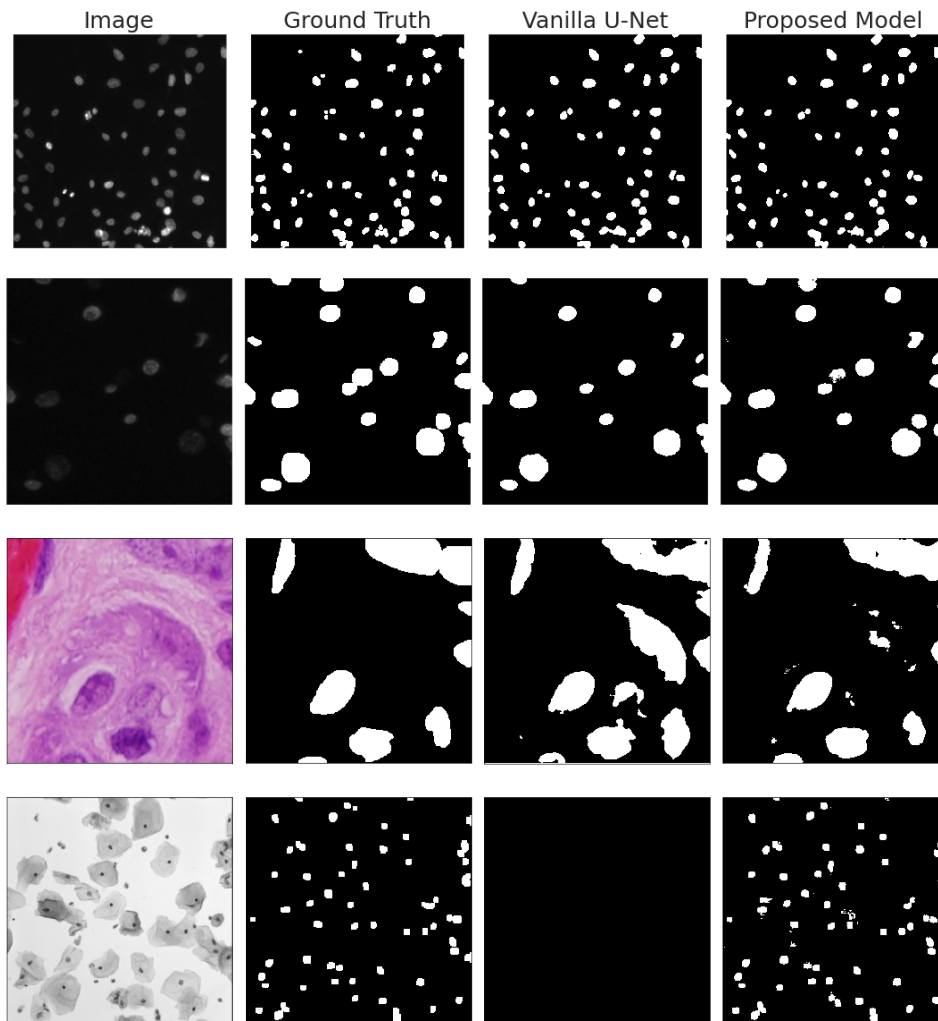


Figure 8: Segmentation comparison between U-Net and the proposed model

where the vanilla U-Net performs poorly and sometimes even has zero intersection of segmentation masks with ground truth. In the third image, the vanilla U-Net could not distinguish cells from staining artifacts (Taqi et al. (2018)) and segmented the regions (stains) as cells, meanwhile, the proposed model had the ability to give soft attention to different color channels and better segment the cells.

In spite of the relative improvement in results the model suffers from certain drawbacks, firstly it could not accommodate both the MDR block and spatial attention module, since the spatial attention module of CBAM applies a convolution filter with a kernel size of 7x7 on the input feature maps it might not have integrated well with the feature maps produced by dilation. Secondly, although using dilated convolutions and having a lesser number of parameters compared to the convolutional blocks of the vanilla U-Net the MDR block is computationally expensive and the addition of it in the architecture gave only a slight improvement in segmentation accuracy when compared to a U-Net with channel attention module and Res Path. A possible explanation for the higher computational cost might be that previous authors who used similar multi-scale dilated residual blocks such as Gudhe et al. (2021) performed a grid search for finding optimal values of dilation rate, due to time and resource limitations this could not be performed in the study, a more probable reason for higher computation cost can be the choice of activation functions, since the choice of activation function can bias the neural network when training on a certain type of data (e.g., linear functions will be like linear regression models) therefore trying activation functions other than ReLu in the MDR block can improve training time.

The Kaggle Data Science Bowl 2018 dataset has images of cell nuclei acquired using different imaging modalities (brightfield, fluorescence, etc.), since a minority of the images in the dataset are histopathological images (images with a larger color spectrum) and the majority of the images are grayscale images, the overall dataset is imbalanced, therefore to further improve upon the study an exploration of artifact oversampling (generating more augmented data) of under-represented samples (images with wider color spectrums), as suggested by Morelli et al. (2021), could be conducted to ascertain whether enhancements can be made to the model's segmentation accuracy.

8 Conclusion and Future Work

Aiming at the problem that the U-Net model has certain limitations such as semantic gap, lack of attention modules, and multiscale feature fusion, a novel cell segmentation model was proposed using the U-Net backbone architecture, multiscale dilated residual blocks, non-linear filters in the skip connection, and channel attention module. The proposed model addressed all three limitations which had not been done before, consequently, the effect of Convolutional Block Attention Module (CBAM) a channel and spatial attention module on features produced by dilated convolutions was also done. The experimental results proved that the proposed model can perform better semantic segmentation of cells while having fewer parameters.

Although achieving the research objectives, the work has certain limitations and deserves further study. Despite having fewer parameters than the vanilla U-Net the computational cost of the proposed model was much higher and therefore is not suitable for deployment in a commercial setting, further investigation needs to be done to address this issue. Implementation of a spatial attention block along with the channel attention

block in the proposed model can also be of future research interest, for this spatial attention blocks used in the skip connections can be considered such as the one proposed by Alshomrani et al. (2023) instead of the CBAM module where the spatial and channel attention modules are primarily used in the bottleneck(end of encoder path) of the U-Net. Furthermore, the study can be extended by assessing the model’s performance on a distinct dataset to verify its robustness.

Acknowledgement

I want to convey my deepest gratitude to Professor Vladimir Milosavljevic for his unwavering support and direction throughout the research process.

References

- Alam, M. S., Wang, D., Liao, Q. and Sowmya, A. (2023). A multi-scale context aware attention model for medical image segmentation, *IEEE Journal of Biomedical and Health Informatics* **27**(8): 3731–3739.
- Alshomrani, S., Arif, M. and Al Ghamdi, M. A. (2023). Saa-unet: Spatial attention and attention gate unet for covid-19 pneumonia segmentation from computed tomography, *Diagnostics* **13**(9).
URL: <https://www.mdpi.com/2075-4418/13/9/1658>
- Alzubaidi, L., Al-Amidie, M., Al-Asadi, A., Humaidi, A. J., Al-Shamma, O., Fadhel, M. A., Zhang, J., Santamaría, J. and Ye, D. (2021). Novel transfer learning approach for medical imaging with limited labeled data, *Cancers (Basel)* **13**(7): 1590.
- Amer, A., Lambrou, T. and Ye, X. (2022). Mda-unet: A multi-scale dilated attention u-net for medical image segmentation, *Applied Sciences* **12**(7).
URL: <https://www.mdpi.com/2076-3417/12/7/3676>
- Bertels, J., Eelbode, T., Berman, M., Vandermeulen, D., Maes, F., Bisschops, R. and Blaschko, M. B. (2019). Optimizing the dice score and jaccard index for medical image segmentation: Theory and practice, in D. Shen, T. Liu, T. M. Peters, L. H. Staib, C. Essert, S. Zhou, P.-T. Yap and A. Khan (eds), *Medical Image Computing and Computer Assisted Intervention – MICCAI 2019*, Springer International Publishing, Cham, pp. 92–100.
- Goodman, A., Carpenter, A., Park, E., jlefman nvidia, BoozAllen, J., Kyle, Maggie, Nilofer, Sedivec, P. and Cukierski, W. (2018). 2018 data science bowl.
URL: <https://kaggle.com/competitions/data-science-bowl-2018>
- Gudhe, N. R., Behravan, H., Sudah, M., Okuma, H., Vanninen, R., Kosma, V.-M. and Mannermaa, A. (2021). Multi-level dilated residual network for biomedical image segmentation, *Scientific Reports* **11**(1): 2045–2322.
- He, K., Zhang, X., Ren, S. and Sun, J. (2016). Identity mappings in deep residual networks, in B. Leibe, J. Matas, N. Sebe and M. Welling (eds), *Computer Vision – ECCV 2016*, Springer International Publishing, Cham, pp. 630–645.

- Ibtehaz, N. and Rahman, M. S. (2020). Multiresunet : Rethinking the u-net architecture for multimodal biomedical image segmentation, *Neural Networks* **121**: 74–87.
URL: <https://www.sciencedirect.com/science/article/pii/S0893608019302503>
- Krizhevsky, A., Sutskever, I. and Hinton, G. E. (2012). Imagenet classification with deep convolutional neural networks, in F. Pereira, C. Burges, L. Bottou and K. Weinberger (eds), *Advances in Neural Information Processing Systems*, Vol. 25, Curran Associates, Inc.
- Minaee, S., Boykov, Y., Porikli, F., Plaza, A., Kehtarnavaz, N. and Terzopoulos, D. (2022). Image segmentation using deep learning: A survey, *IEEE Transactions on Pattern Analysis and Machine Intelligence* **44**(7): 3523–3542.
- Morelli, R., Clissa, L., Amici, R., Cerri, M., Hitrec, T., Luppi, M., Rinaldi, L., Squarcio, F. and Zoccoli, A. (2021). Automating cell counting in fluorescent microscopy through deep learning with c-resunet, *Scientific Reports* **11**(1): 2045–2322.
- Mubashar, M., Ali, H., Grönlund, C. and Azmat, S. (2022). R2u++: a multiscale recurrent residual u-net with dense skip connections for medical image segmentation, *Neural Computing and Applications* **34**(20): 1433–3058.
- Ronneberger, O., Fischer, P. and Brox, T. (2015). U-net: Convolutional networks for biomedical image segmentation, in N. Navab, J. Hornegger, W. M. Wells and A. F. Frangi (eds), *Medical Image Computing and Computer-Assisted Intervention – MICCAI 2015*, Springer International Publishing, Cham, pp. 234–241.
- Szegedy, C., Liu, W., Jia, Y., Sermanet, P., Reed, S., Anguelov, D., Erhan, D., Vanhoucke, V. and Rabinovich, A. (2015). Going deeper with convolutions, *2015 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, IEEE Computer Society, Los Alamitos, CA, USA, pp. 1–9.
URL: <https://doi.ieeecomputersociety.org/10.1109/CVPR.2015.7298594>
- Taqi, Syed, A., Sami, S. A., Sami, L. B. and Zaki, S. A. (2018). A review of artifacts in histopathology, *Journal of oral and maxillofacial pathology* **22**(2): 1590.
- Wang, X., Hu, Z., Shi, S., Hou, M., Xu, L. and Zhang, X. (2023). deep learning method for optimizing semantic segmentation accuracy of remote sensing images based on improved unet, *Scientific Reports* **13**(1): 2045–2322.
- Woo, S., Park, J., Lee, J.-Y. and Kweon, I. S. (2018). Cbam: Convolutional block attention module, in V. Ferrari, M. Hebert, C. Sminchisescu and Y. Weiss (eds), *Computer Vision – ECCV 2018*, Springer International Publishing, Cham, pp. 3–19.
- Yu, F. and Koltun, V. (2016). Multi-scale context aggregation by dilated convolutions.
- Zhou, Z., Rahman Siddiquee, M. M., Tajbakhsh, N. and Liang, J. (2018). Unet++: A nested u-net architecture for medical image segmentation, in D. Stoyanov, Z. Taylor, G. Carneiro, T. Syeda-Mahmood, A. Martel, L. Maier-Hein, J. M. R. Tavares, A. Bradley, J. P. Papa, V. Belagiannis, J. C. Nascimento, Z. Lu, S. Conjeti, M. Moradi, H. Greenspan and A. Madabhushi (eds), *Deep Learning in Medical Image Analysis and Multimodal Learning for Clinical Decision Support*, Springer International Publishing, Cham, pp. 3–11.