BRB U-Net: Bottleneck Residual Blocks in U-Net for Light-Weight Semantic Segmentation

Aruna Kumari Kakumani
VNR Vignana Jyothi Institute of Engineering and Technology,
Hyderabad, Telangana, India, 500090

L. Padma Sree
VNR Vignana Jyothi Institute of Engineering and Technology,
Hyderabad, Telangana, India, 500090

ABSTRACT
Cell is the fundamental entity of all living organisms. Understanding cell behaviour is improtant to study the biological processes in living organisms. In this work semantic segmentation of cells in microscopy images is studied. Specifically a novel deep learning architechture, BRB U-Net is proposed for the semantic segmentation of cells in microscopy**.** Bottleneck residual blocks are incorporated in U-Net architechture to achieve a light weight semantic segmentation model. The proposed method is evaluated with Phc-C2DH-U373 dataset of cell tracking challenge and achieves 0.9430 and 0.8383 dice similarity coefficient and intersection over union respectively. BRB U-Net achieved 7.68 times less number of parameters and model size is 7.35 times lesser than U-Net.

Keywords
Deep Learning, Semantic Segmentation, Microscopy, U-Net, Cells, MobileNetV2

1. INTRODUCTION
Analysis of biological cells plays critical role in understanding the cells behavior in different settings like response of cells to new drugs for drug discovery[1], real time observation of biological process like angiogenesis[2], studying cell proliferation to observe severity of tumors[3], understanding dynamics of cells in wound healing[4], understanding cell growth in embryos known as embryogenesis[5]and also to study different phases of cell cycle like mitosis[6]. One of the ways to analyze these cells is by capturing the cell images using time-lapse microscopy. Time lapse microscopy is a method of capturing microscopy images of cells periodically after certain intervals of time so that one can study the spatiotemporal dynamics of cells. The research in microscopy image analysis is popular from many years. Generally, analysis of microscopy involves two tasks: Cell segmentation in which individual cells are detected and localized; Cell tracking in which each cell is tracked in timelapse microscopy images. Traditionally these images are analysed manually by human experts, which takes long hours of human curation, may subject to inter human variability, and the same expert may analyze differently at different times due to complex nature of cell images in microscopy. The complexities of these images are topological changes of cells, cells arecrowded, sometimes cells touch each other, low contrast of foreground and background, cells moving in and out of the frame to name a few. Due to these challenges involved and laborious manual analysis, automatic or semi-automatic analysis of microscopyimages are essential. In line with that, there are many image processing techniques proposed by researchers around the globe. Currently due to extraordinary results of deep learning methods, lot of research is still going on for much efficient and accurate deep learning models to address the complex issues involved in the analysis of cells in microscopy images.

U-Net[7] is the most celebrated deep learning architecture for the segmentation of biomedical images. U-Net consists of a contracting path, bottleneck, and expanding path. Innumerable methods have been proposed which follow the U-Net style: UNet++[8], ResUNet-a[9],Inception UNet v1and Inception UNet v2[10], Attention UNet[11], Squeeze U-Net [12] to name a few. U-Net consists of a encoder CNN (contracting path or down sampling path), bottleneck, decoder CNN (expansive path or upsampling path) and skip connections. Contracting path mainly extracts the contextual features from the image. Bottleneck is used to produce only the important features that can reconstruct the segmentation map. Upsampling path does the localization of features extracted. Skip connections are included between the corresponding layers to retain the spatial information which would have lost during downsampling path. In UNet++ the skip pathways are reengineered consisting of dense convolution blocks between the contracting and expansive path sub-networks which resulted in improved segmentation accuracy. ResUNet-a is an encoder/decoder like network in which traditional convolutions are replaced with ResNet units with several parallel atrous convolutions. In addition, pyramid scene parsing pooling is included for enhancing the performance of the network. Inception UNet-v1 and Inception Unet-v2 are encoder/decoder type networks designed by combining inception blocks[13] with convolutional layers for obtaining improved performance. Attention U-Net is an encoder/decoder type architecture which utilizes a novel attention gate in the decoder path. Attention gate helps the network to learn the salient features thus improving the network performance. Squeeze U-Net use fire modules[14] and transposed fire modules in the encoding and expanding paths respectively in order to reduce the model size to generate memory and power efficient segmentation model so that it can be used for real-time applications.

In Usiigachi[15] mask regional convolutional neural networks(Mask RCNN) [16]is used for the segmentation of cell instances. In DeepLabv3[17]atrous convolution in cascade or in parallel is used to obtain multiscale object segmentation by employing different atrous rates. Also, atrous spatial pyramid pooling module is combined with image level features to improve the segmentation performance.

Inspired by U-Net and MobileNetV2[18], we design a novel deep learning architecture for segmentation of cells in time-lapse microscopy images.

The contributions of this work are listed below:

1. A novel deep learning architecture for segmentation of cells in time-lapse microscopy images with a smaller number of parameters is designed.
2. Comparative analysis of the proposed architecture with U-Net architecture for time-lapse microscopy images is performed.
3. Quantitative Performance of the proposed deep learning framework : Dice similarity coefficient, Intersection over Union(IoU) is demonstrated.

The organization of the remaining article is as follows. Section 2 depicts the methodology used in this work. Section 3 illustrates the results obtained. Section 4 mentions the conclusion which is followed by the references.

2. MATERIALS AND METHODS

2.1 Reference data and Data Augmentation

The dataset for the current research is Phc-C2DH-U373 dataset and is taken from ISBI cell tracking challenge[19].This dataset has two time-lapse image sequences of Glioblastoma-astrocytoma U373 cells consisting of 114 frames each, thus it consists of a total of 228 frames of 2D phase contrast microscopy images of mouse stem cells. To increase the dataset size so that a deep learning model is more generalizable, the following data augmentation techniques are employed: performed horizontal flip, vertical flip, image shear by 30%, image zoom by 5%, image width shift by 5%, image height shift by 5%, and image rotation by 90 degrees thus increasing the dataset size by seven times. Following data augmentation, image normalization is done by dividing each pixel value by 255, which helps in faster convergence. The whole dataset is partitioned into 85% train data, 10% validation data and 5% test data.

2.2 Methodology

The proposed architecture namely BRB U-Net is inspired by U-Net [7] and bottleneck residual blocks of MobileNetV2 [18]. The number of parameters is reduced drastically with the use of bottleneck residual blocks producing a lightweight model which can be used for real time applications or specifically for mobile devices.

2.2.1 Deep Learning Architecture

The proposedarchitecture as shown in Fig. 1, is a U-Net type consisting of encoder (contracting path), bridge and decoder (expansive path). The novel idea in the proposed architecture is the use of basic building blocks of MobileNetV2namelybottleneck depth separable convolution with residuals or bottleneck residual blocks. In this work a bottleneck residual blocks with ReLU activation functionare employed. It is based on an inverted residual structure where the residual connections are between the bottleneck layers. Lightweight depthwise convolutions are used in the intermediate expansion layer as a source of non-linearity to filter features. İn this building block, first the channels are expanded using a greater number of 1x1 kernels followed by depthwise convolution with kernel size 3x3. This is followed by 1x1 convolution to squeeze (compress) the channels. The squeeze operation is concatenated with the input of the block. This block concatenates the narrow layers while the channels in between are wide.

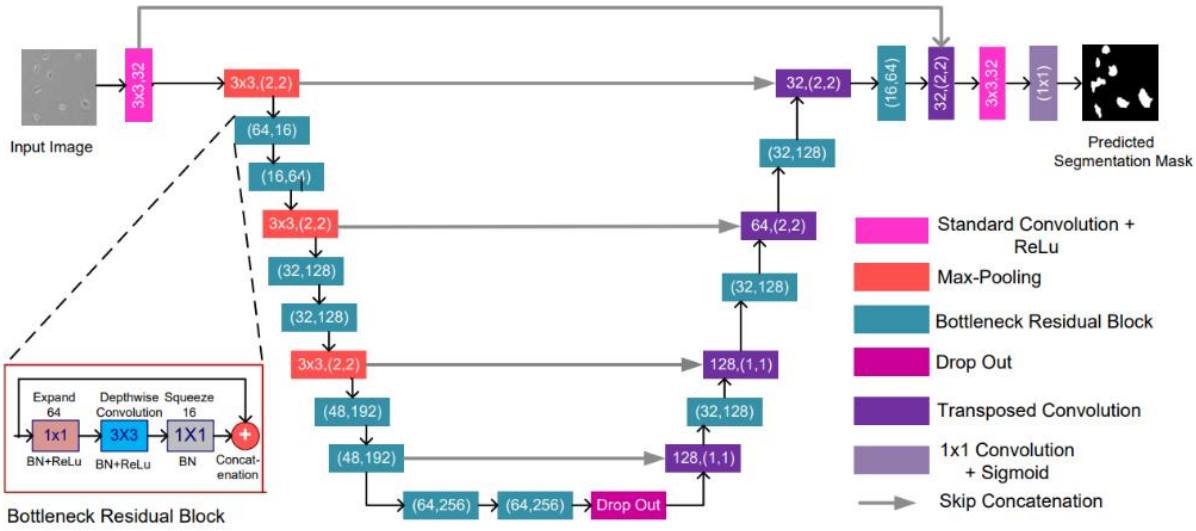


Fig 1: BRB U-Net Architechture

The original input image is resized to 256 x 256 and is given to the network. The first operation is standard 3x3 convolution followed by ReLU activation function. Then max pooling is done by kernel size of 3x3, stride 2x2 and same padding. Maxpooling helps in extracting the prominent features and helps in dimensionality reduction. This results in image size which is reduced by half of the input. This is followed by two bottleneck depth separable convolution with residual blocks of parameters expand consisting of 64 channels and squeeze consisting of 16 channels. This is followed by maxpooling which further reduces image size by half namely 64x64. Next two bottleneck depth separable convolution with residual blocks with parameters expand consisting of 128 channels and squeeze consisting of 32 channels are used.The expand and squeeze channels are doubled from the previous layers to extract more features. This operation is followed by maxpooling. This operation is followed by two bottleneck depth separable convolution with residual blocks of 192 channels in expand and 48 channels in squeeze followed by dropout. Dropout is used to reduce overfitting. In the up-sampling path, initially up-sampling operation is done by transposed convolutions of 128 filters, kernel size 3 and stride 1x1. This is followed by two bottleneck depth separable convolution with residual blocks.The decoder consists of a series of transposed convolution and one bottleneck depth separable convolution with residual block as shown in the fig.

1. Skip concatenation is done from output of max pooling of the encoder to transposed convolution block of the decoder. Skip concatenation helps to preserve the spatial features of the encoder. The final layer consists of 3x3 convolution followed by ReLU and 1x1 convolution followed by sigmoid activation function so that each pixel is either labelled cell or background.

2.2.2 Deep Learning Training

The network is trained with Phc-C2DH-U373 datasets. Data augmentation is performed to increase the dataset size by seven times. The entire dataset is divided into 85% training data, 10% validation data and 5% test data. Input image pixel values are divided by 255 so that pixels are normalized to lie between 0 to 1. The weight initialization for the filters is done by Glorot_uniform method. Batch size of 16 is applied. Adam optimizer[20]is used with learning rate of 0.0001, decayed every 6 epochs of using exponential rate of 0.6.The network is trained for 60epochs. The Loss function is the sum of binary cross entropy and Dice Loss. The deep learning framework is developed in TensorFlow in Python. The networkis trained using Google colaboratory. Hyperparameters used in the proposed BRB U-net is shown in Table 1.

Table 1. Hyperparameters used in the proposed architecture

Hyperparameter	Input
Epochs in training a CNN	60
Batch Size	16
Kernel size	1x1, 3x3, 3x3 depthwise
Padding	Same
Learning Rate	0.0001
Loss function	Binary Cross Entropy + Dice Loss
Optimizer	Adam
Metric	Accuracy, Dice coefficient
Train-validation-test split ratio	8.5:1: .5
Activation function	ReLU, Sigmoid

3. RESULTS

3.1 Evaluation Metrics

The segmentation output of the proposed model is evaluated using the following two metrics.

Dice Similarity Coefficient (*DSC*) and Intersection over Union (*IoU*) as shown in equations (1) and (2).

$$DSC = \frac{2\,X\,TP}{2\,X\,TP + FP + FN} \quad (1)$$

$$IoU = \frac{TP}{TP + FP + FN} \quad (2)$$

where, where TP, FP, FN, and TN are true positive, false positive, false negative, and true negative metrics, respectively.

3.2 Training Curves

The training Dice Coefficient and Loss and Accuracy curves are shown in Fig. 2(a), Fig.2(b) and Fig. 2(c) respectively. These curves show the training performance of the proposed model. As can be seen, the proposed method depicted in red line converges faster compared to the other method.

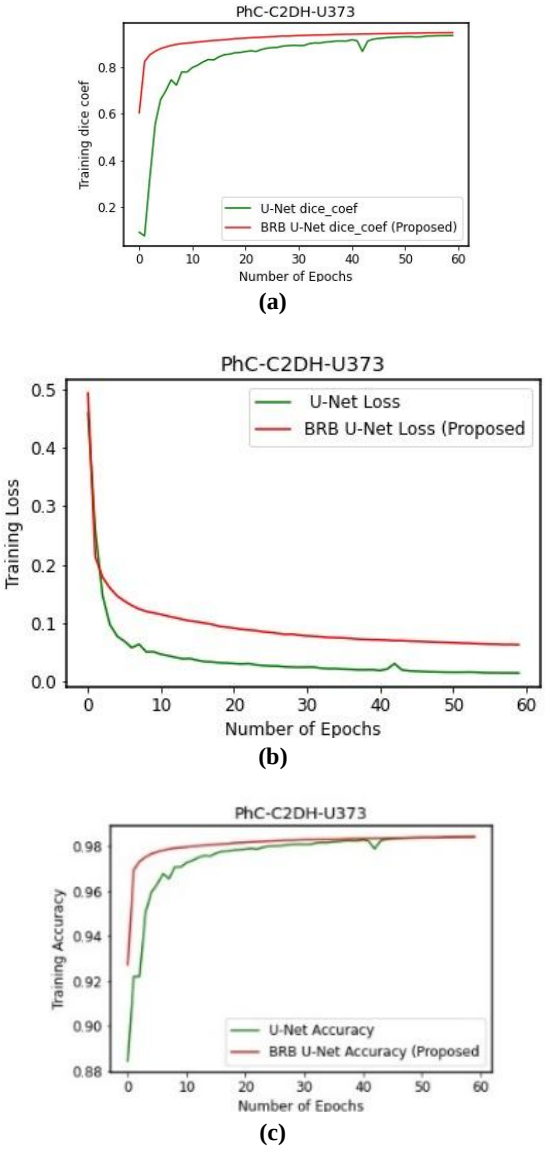


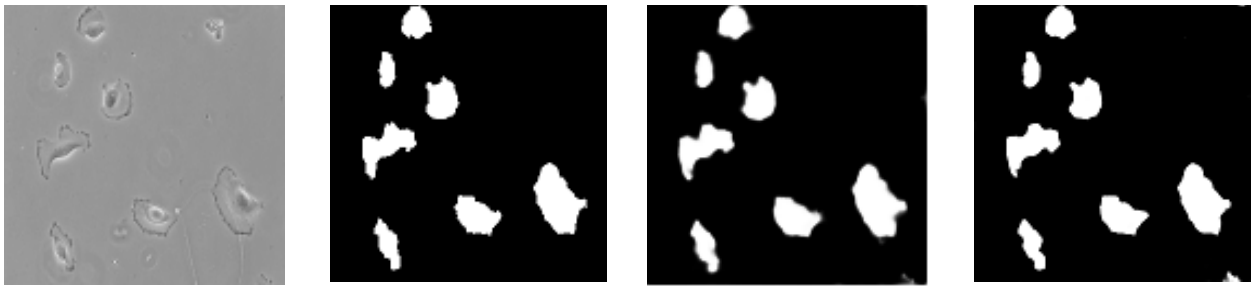
Fig 2: Training Dice Coefficient, Loss, and Accuracy curves

Quantitative Results

The quantitative results of the above metrics for the segmentation of Phc-C2DH-U373 is shown in the Table 2.

Table 2. Performance Segmentation Metrics

Dataset	Method	Quality Metrics			
		DSC	**IoU**	**Parameters**	**Model size**
Phc-C2DH-U373	U-Net	0.9446	0.8426	7759521	89MB
	BRB U-Net (Proposed Method)	0.9430	0.8383	**1009153**	**12.1MB**



(a) Original Image **(b) Ground Truth** **(c) U-Net** **(d) BRB U-Net**

Fig 3: Qualitative Results

The quantitative evaluation metrics used are DSC and IoU. Also,the number of parameters and the model size are compared. The results indicate that the DSC and IoU of the proposed method i.*e.*, BRB U-Net is almost the same as U-Net, but the number of parameters and the model size of BRB U-Net is much smaller than U-Net. Hence BRB U-Net is recommended for real time and memory efficient applications.

3.3 Qualitative Results

The qualitative results are shown in Fig. 3. The proposed BRB U-Net method in Fig. 3(d) has performed well as compared to the U-Net method Fig. 3(c) and is closer to the ground truth Fig. 3(b). The input image to both the methods is shown in Fig. 3(a).

4. CONCLUSIONS

In this article, a novel deep learning architecture known as BRB U-Net is proposed. This architecture utilizes bottleneck residual blocks in U-Net for achieving light weight segmentation model for semantic segmentation of microscopy images. This architecture may be trained for any semantic segmentation problem in general. BRB U-Net gave Dice Similarity Coefficient and Intersection over Union of0.9430 and 0.8383 respectively. BRB U-Net has 7.68 times lesser parameters than U-Net and model size is 7.35 times lesser than U-Net. This is due to the use of bottleneck residual blocks in U-Net. Hence BRB U-Net can be used for real time applications, in embedded devices where memory efficiency is needed and mobile devices. This method may be used for semantic segmentation in general.

5. REFERENCES

[1] J. Boyd, M. Fennell, and A. Carpenter, “Harnessing the power of microscopy images to accelerate drug discovery: what are the possibilities?,” *Expert Opin. Drug Discov.*, vol. 15, no. 6, pp. 639–642, 2020, doi: 10.1080/17460441.2020.1743675.

[2] E. M. Gabriel, D. T. Fisher, S. Evans, K. Takabe, and J. J. Skitzki, “Intravital microscopy in the study of the tumor microenvironment: From bench to human application,” *Oncotarget*, vol. 9, no. 28, pp. 20165–20178, 2018, doi: 10.18632/oncotarget.24957.

[3] G. A. Romar, T. S. Kupper, and S. J. Divito, “Research techniques made simple: Techniques to assess cell proliferation,” *J. Invest. Dermatol.*, vol. 136, no. 1, pp. e1–e7, 2016, doi: 10.1016/j.jid.2015.11.020.

[4] J. E. N. Jonkman *et al.*, “An introduction to the wound healing assay using live-cell microscopy,” *Cell Adhes. Migr.*, vol. 8, no. 5, pp. 440–451, 2014, doi: 10.4161/cam.36224.

[5] S. Iyer, S. Mukherjee, and M. Kumar, “Watching the embryo: Evolution of the microscope for the study of embryogenesis,” *BioEssays*, vol. 43, no. 6, pp. 1–17, 2021, doi: 10.1002/bies.202000238.

[6] Y. T. Su, Y. Lu, J. Liu, M. Chen, and A. A. Liu, “Spatio-Temporal Mitosis Detection in Time-Lapse Phase-Contrast Microscopy Image Sequences: A Benchmark,” *IEEE Trans. Med. Imaging*, vol. 40, no. 5, pp. 1319–1328, 2021, doi: 10.1109/TMI.2021.3052854.

[7] T. B. Olaf Ronneberger, Philip Fischer, “unet for biomedical image segmentation,” in Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics), 2015, vol. 9351, no. Cvd, p. 234241. doi: 10.1007/978-3-319-24574-4.

[8] Z. Zhou, M. M. R. Siddiquee, N. Tajbakhsh, and J. Liang, *UNet++: A Nested U-Net Architecture*, vol. 11045 LNCS. Springer International Publishing, 2018. doi: 10.1007/978-3-030-00889-5.

[9] F. I. Diakogiannis, F. Waldner, P. Caccetta, and C. Wu, “ResUNet-a: A deep learning framework for semantic segmentation of remotely sensed data,” *ISPRS J. Photogramm. Remote Sens.*, vol. 162, pp. 94–114, 2020, doi: 10.1016/j.isprsjprs.2020.01.013.

[10] I. Delibasoglu and M. Cetin, “Improved U-Nets with inception blocks for building detection,” *J. Appl. Remote*

Sens., vol. 14, no. 04, pp. 1–15, 2020, doi: 10.1117/1.jrs.14.044512.

[11] O. Oktay *et al.*, "Attention U-Net: Learning Where to Look for the Pancreas," no. Midl, 2018, [Online]Available: http://arxiv.org/abs/1804.03999

[12] N. Beheshti, "Beheshti_Squeeze_U-Net_A_Memory_and_Energy_Efficient_Image_Segmentation_Network_CVPRW_2020.

[13] C. Szegedy et al., "Going deeper with convolutions," Proc. IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recognit., vol. 07-12-June, pp. 1–9, 2015, doi: 10.1109/CVPR.2015.7298594.

[14] F. N. Iandola, S. Han, M. W. Moskewicz, K. Ashraf, W. J. Dally, and K. Keutzer, "SqueezeNet: AlexNet-level accuracy with 50x fewer parameters and <0.5MB model size," pp. 1–13, 2016, [Online]. Available: http://arxiv.org/abs/1602.07360

[15] H. F. Tsai, J. Gajda, T. F. W. Sloan, A. Rares, and A. Q. Shen, "Usiigaci: Instance-aware cell tracking in stain-free phase contrast microscopy enabled by machine learning," *SoftwareX*, vol. 9, pp. 230–237, 2019, doi: 10.1016/j.softx.2019.02.007.

[16] K. He, G. Gkioxari, P. Dollár, and R. Girshick, "Mask R-CNN," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 42, no. 2, pp. 386–397, 2020, doi: 10.1109/TPAMI.2018.2844175.

[17] L.-C. Chen, G. Papandreou, F. Schroff, and H. Adam, "Rethinking Atrous Convolution for Semantic Image Segmentation," 2017, [Online]. Available: http://arxiv.org/abs/1706.05587

[18] M. Sandler, A. Howard, M. Zhu, A. Zhmoginov, and L. C. Chen, "MobileNetV2: Inverted Residuals and Linear Bottlenecks," *Proc. IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recognit.*, pp. 4510–4520, 2018, doi: 10.1109/CVPR.2018.00474.

[19] Cell Tracking Challenge, http://celltrackingchallenge.net/

[20] D. P. Kingma and J. L. Ba, "Adam: A method for stochastic optimization," *3rd Int. Conf. Learn. Represent. ICLR 2015 - Conf. Track Proc.*, pp. 1–15, 2015.