LMSC-UNet: A Lightweight U-Net with Modified Skip Connections for Semantic Segmentation

Shrutika S. Sawant[1][a], Andreas Medgyesy[1][b], Sahana Raghunandan[1][c] and Theresa Götz[1,2][d]

[1]*Fraunhofer Institute of Integrated Circuits, Erlangen, Germany*

[2]*Department of Industrial Engineering and Health, Technical University of Applied Sciences Amberg-Weiden, Germany*
{*shrutika.sawant, andreas.medgyesy, sahana.raghunandan, theresa.goetz*}@*iis.fraunhofer.de*

Keywords: Depthwise Separable Convolutions, Skip Connections, Semantic Segmentation, U-Net.

Abstract: U-Net, an encoder-decoder architecture is the most popular choice in the semantic segmentation field due to its ability to learn rich semantic features while handling enormous amounts of data. However, due to large number of parameters and slow inference, deploying U-Net on devices with limited computational resources such as mobile and embedded devices becomes challenging. To alleviate the above challenge, in this study, we propose an efficient, lightweight, and robust encoder-decoder architecture, LMSC-UNet for semantic segmentation that captures more comprehensive, contextual information and effectively learns rich semantic features. This lightweight architecture considerably reduces the number of trainable parameters, requiring sufficiently less memory space, training, and inference time. Skip connections in original U-Net fuse features from each encoder block to the corresponding decoder block. This simple skip connection reduces the semantic gap to some extent and may limit the segmentation performance. Therefore, we replace the skip connection from the second level of U-Net with a bottleneck residual block (BRB) which helps to enhance the final segmentation map by lessening the semantic gap between the features of decoder with the corresponding features of encoder. Extensive experiments on various segmentation datasets from diverse domains demonstrate the effectiveness of our proposed approach. The experimental results show that the compact model speeds up the inference process, while still maintaining the performance. When compared to the standard U-Net, LMSC-UNet has achieved $7\times$ reduction in Floating Point Operations (FLOPs), and $34\times$ reduction in model size, while maintaining the segmentation accuracy.

1 INTRODUCTION

Semantic segmentation is an essential and fundamental task in computer vision aimed at assigning a label to each pixel in an image. This has been widely studied in various domains, such as remote sensing data analysis (Meng et al., 2022), medical image analysis (Al-Masni and Kim, 2021), autonomous driving (Nawaratne et al., 2020) and so on (Zhao et al., 2024), (Zhang et al., 2021). Since the introduction of deep learning, researchers have taken interest in developing semantic segmentation algorithms based on deep neural networks (Akkus et al., 2017), (Milletari et al., 2016). Fully Convolutional Networks (FCN) is the earliest deep segmentation network (Long et al., 2015). Subsequently, various segmentation networks were introduced based on FCN which showed quite excellent segmentation performance (Badrinarayanan et al., 2017), (Milletari et al., 2016), (Chen et al., 2018). In 2015, an encoder-decoder structure based on FCN was introduced, called as U-Net, for segmenting biomedical images (Ronneberger et al., 2015). By employing skip connections for mapping low-resolution features to high resolution features, U-Net produced more precise segmentation maps. Due to its exceptional segmentation ability, U-Net has gained immense popularity among the research community and has been adopted to various research fields. Furthermore, numerous variants of U-Net have emerged, such as, Dense-UNet (Cao et al., 2020), UNet++ (Zhou et al., 2020), Dual-UNet (Li et al., 2019), MSU-Net (Su et al., 2021), etc.

Despite demonstrating excellent segmentation capability and wide adoption in various domains, standard U-Net suffers from some limitations, especially

[a] https://orcid.org/0000-0002-1532-947X

[b] https://orcid.org/0009-0002-8129-5738

[c] https://orcid.org/0009-0009-8056-3420

[d] https://orcid.org/0000-0001-8751-3404

when deploying it on limited resource devices for real-time inferencing. The main challenge is that U-Net is a large complex model encompassing multiple convolutional layers with many filters and needs an enormous number of trainable parameters. This led to a high computational cost which demands a large amount of memory footprint. As a result, the standard U-Net demands almost 31 million trainable parameters and 54 billion (G) FLOPs to get a segmented output for an image size of 256×256. Furthermore, U-Net's encoder-decoder design does not handle multiscale features compactly, the redundancy in processing image scales often leads to unnecessary computational overhead. To overcome these limitations, compressing U-Net becomes essential that facilitates effective deployment in resource-constrained environments while preserving the performance of the original U-Net model. There exist various methods of U-Net model compression, including model pruning (Sawant et al., 2022a), (Sawant et al., 2022b), (Sawant et al., 2023), quantization (Nam et al., 2024), knowledge distillation (Vaze et al., 2020) and lightweight variants of U-Net (Beheshti and Johnsson, 2020), (El-Assiouti et al., 2023). By incorporating one of the model compression techniques or combining two, the U-Net can be made suitable for resource-constrained environments without significantly compromising segmentation accuracy.

In this study, we propose an efficient, lightweight, and robust encoder-decoder architecture for semantic segmentation that captures more comprehensive contextual information and effectively learns rich semantic features. We adopt U-Net, an encoder-decoder architecture as a backbone. We modify the standard U-Net architecture such that it contains less number of trainable parameters and still yields precise segmentation maps. The main contributions of this study are summarized as: 1) The MobileNetV2 building blocks (MBB) are substituted in the contracting path of the standard U-Net, which significantly reduces the number of trainable parameters. 2) The skip connection from the second level of U-Net is replaced with a bottleneck residual block (BRB) which contains a residual connection and a small feature extraction-dimensionality reduction block. 3) To demonstrate the robustness of the proposed LMSC-UNet, we evaluate our model on two datasets from different domains.

2 PROPOSED FRAMEWORK

In this section, we discuss LMSC-UNet, a lightweight and efficient architecture for semantic segmentation

in detail. This model adopts standard U-Net as its baseline model, which is a fully convolutional neural network with a symmetric encoder-decoder structure. The proposed LMSC-UNet architecture incorporates MBB units in the encoder of the standard U-Net (except first layer) and the flow of features between the second level of encoder to the corresponding decoder is accomplished by a BRB path, as shown in Figure 1. We use the decoder of the proposed model same as in standard U-Net. Compared to the standard U-net, LMSC-UNet has fewer layers with lesser the number of filters in each layer. This shallow architecture of LMSC-UNet is more suitable for deployment on resource-constrained devices than standard U-Net due to its faster training and real-time inference.

As mentioned above, the LMSC-UNet model follows a similar structure as U-Net with few modifications. Except for the first layer in the encoder which has a standard convolution, the remaining layers in the encoder of LMSC-UNet are replaced with MBB units as shown in Figure 2. The MBB (Sandler et al., 2019) unit is built upon the inverted residual layer which is composed of a 1×1 convolution with Relu activation, a 3×3 depthwise separable convolution with Relu activation, a 1×1 convolution with linear activation and a residual connection. The use of MBB offers two noteworthy benefits: firstly, depthwise separable convolutions reduce the number of parameters which in turn decreases the memory consumption and computational cost significantly. And secondly, inverted residual bottleneck preserves more detailed information and facilitates efficient gradient flow during backpropagation, improving training stability and convergence. The use of MBB units in the second and third layer (which are responsible for mid-level feature extractions) of LMSC-UNet, make the encoder more efficient in capturing the contextual details required for segmentation tasks. Overall, the inclusion of MBB units significantly reduces the computational footprint while preserving essential features, making the LMSC-UNet highly efficient for segmenting complex scenes.

As mentioned earlier, we use the same decoder as that of standard U-Net, which produces segmentation maps at full resolution. With skip connections at the corresponding depths of the encoder and decoder, deep semantic features from decoder are combined with shallow, low-level features from encoder, which results in spatially accurate and semantically meaningful segmentation maps. Despite improving the segmentation accuracy, simple skip connections often show some limitations, which can hinder performance in segmenting complex scenes. The key limitation is that directly combining low-level and high-
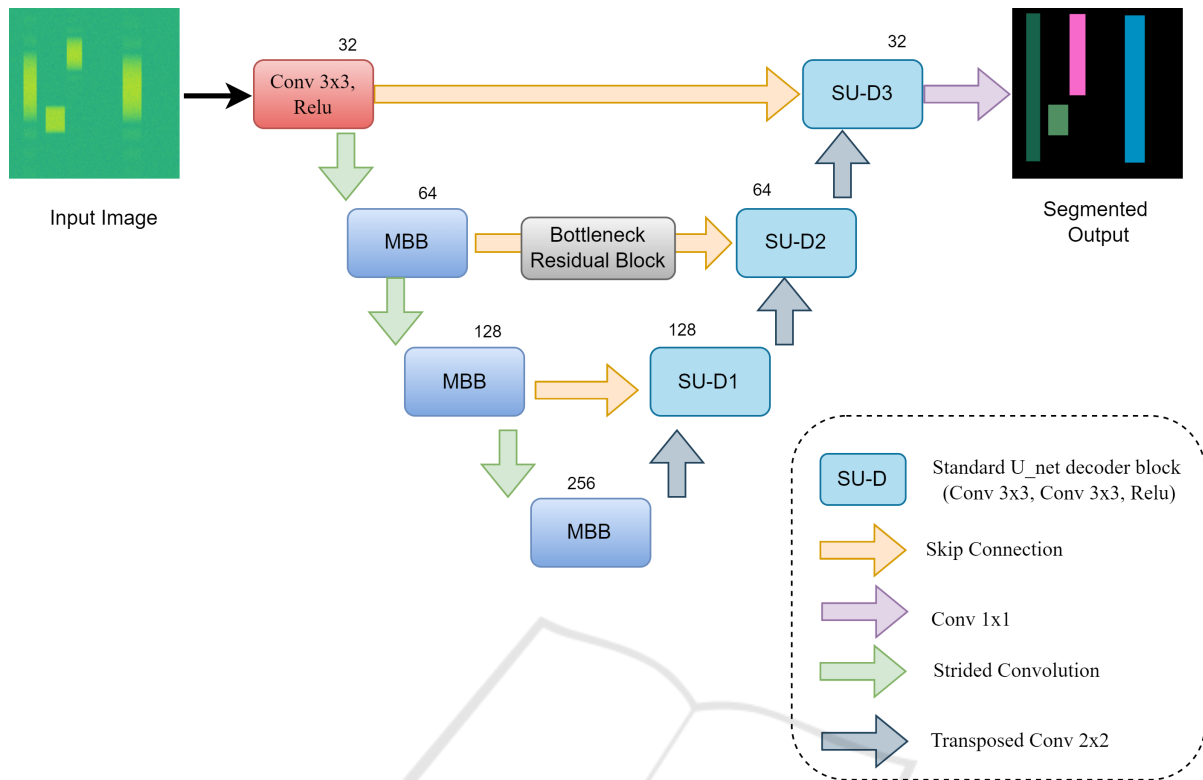
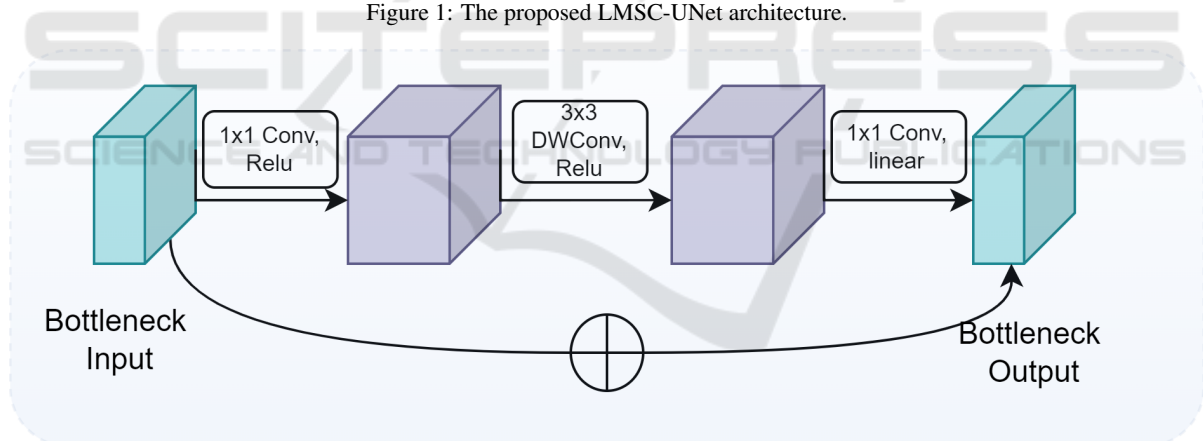Figure 1: The proposed LMSC-UNet architecture.

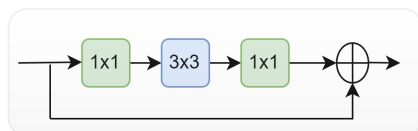Figure 2: The MBB: MobileNetV2 Building Block.

Figure 3: The BRB: Bottleneck Residual Block.

level feature maps may confuse the decoder and degrade segmentation accuracy. To overcome this, we used a BRB (as used in (He et al., 2016)) path in the second level skip connection shown in Figure 1. The BRB is a combination of basic residual block and dimensionality reduction block shown in Figure 3. The

first 1×1 convolution reduces the feature dimensions, whereas the second 1×1 convolution expands them. This not only reduces the redundant information, but also focuses on most relevant features. The intermediate 3×3 convolution block processes the reduced features, enriching the spatial features passed through the skip connection. Since, this arrangement refines the encoder feature maps before passing to the decoder, it certainly reduces the mismatch between encoder-decoder feature spaces. Moreover, we use BRB only in second-level skip connection, which not only preserves intermediate contextual information, but also

avoids overloading the model with additional computations across all levels.

With a smaller number of encoder layers in LMSC-UNet (only three), the model may capture only fewer levels of abstraction and struggle to accurately segment the tiny objects or some intricate boundaries. To mitigate the potential loss of precision, we use a combination of Dice loss and Focal loss.

3 EXPERIMENTAL SETUP

To demonstrate the effectiveness of the proposed LMSC-UNet, we evaluate its performance on two datasets for semantic segmentation from diverse domains and compare with standard U-Net.

3.1 Dataset

We quantitatively analyse the segmentation performance of the LMSC-UNet across two publicly available datasets from two different domains: 1) Biomedical imaging – NuInsSeg (Mahbod et al., 2024) and 2) Radio Frequencies (RF) signal spectrum segmentation dataset (Vagollari et al., 2023). These datasets were selected because they offer diverse and comprehensive cases for image segmentation challenges, which can effectively denote the robustness and generalization capabilities of the proposed model.

NuInsSeg is a dataset of nuclei in Hematoxylin and Eosin (HE)- stained histological images, which has 665 image patches with more than $30,000$ manually segmented nuclei from 31 humans and mouse organs. The dataset comes with binary segmentation masks containing nuclei and non-nuclei regions.

The RF signal dataset comprises a collection of wideband communication signals, which represent recordings of a radio monitoring receiver captured across a wide frequency band. It has 4000 wideband signals that are stored using the Signal Metadata Format (SigMF). To perform the task of detecting, classifying, and localizing all narrowband emissions within a wideband signal, we generated spectrogram images of wideband signals using Short-Time Fourier Transform (STFT). We analyse this dataset to perform the task of signal detection and localization.

3.2 Implementation Details and Evaluation Metrics

The proposed LMSC-UNet is trained and evaluated using Nvidia Tesla V100 (32GB) GPU, Intel Xeon Gold 6134 ("Skylake") CPU with 16 cores, 32 threads @3.2GHz, 96GB RAM based on Pytorch framework (https://pytorch.org/). For model's weight initialization, the He (also known as Kaiming) initialization method is employed. The network parameters are optimized using the Adam optimizer with a batch size of 32, the initial learning rate is set to 0.01 and the learning decay rate is 0.1. Input images from all datasets are down-sampled using linear interpolation with a resolution of 256×256 pixels and fed into the LMSC-UNet. Besides, models are trained for 200 epochs on both NuInsSeg and RF signal dataset. To ensure better performance of the model on the NuInsSeg dataset, a 5-fold cross-validation is applied to enhance the robustness of the results. For the RF signal dataset, we split it into 70% training, 15% validation and 15% test set. We used 133 images as validation set in NuInsSeg dataset and 600 spectrogram images in the test set of RF signal dataset, without data augmentation. Segmentation performance on NuInsSeg dataset is recorded by taking an average of all folds, whereas results on a test set are recorded for RF signal dataset.

Model's segmentation performance is evaluated using segmentation accuracy (ACC.), the mean intersection-over-union (mIoU), F1 score and Dice coefficient (DC), which are commonly used for semantic segmentation tasks. To assess the efficiency and scalability of the model, the number of parameters, model size and MACs (Multiply-Accumulate) are measured as computational metrics.

4 EXPERIMENTAL ANALYSIS

We conducted a series of experiments on the proposed LMSC-UNet for measuring its computational requirements and segmentation performance on above mentioned datasets.

4.1 Computational Results

To deploy the model on a small, embedded platform for real-time inferencing, having a compact or lightweight model is top priority. In this study, we utilized MBB units which contain depthwise separable convolutions to reduce the number of parameters and computations of the model. The computational requirements of LMSC-UNet are measured in terms of number of parameters, FLOPs (as MACs), storage requirements (model size), and compared to the LMSC-UNet without a BRB and the standard U-Net model. LMSC-UNet without a BRB include simple skip connections as used in U-Net. Table 1 reports the comparison between computational requirements

Table 1: Computational requirements of LMSC-UNet and other models. Here, M stands for million and B stands for billion.

Model	Params (M)	MACs (B)	Size in memory (MB)	Inference time (ms/image)
Standard U-Net	31	54.76	118	6.8
LMSC-UNet (without BRB)	0.9	8.1	3.45	2.4
LMSC-UNet (Ours)	0.9	8.2	3.46	2.4

Table 2: Segmentation performance of LMSC-UNet and other models on NuInsSeg dataset.

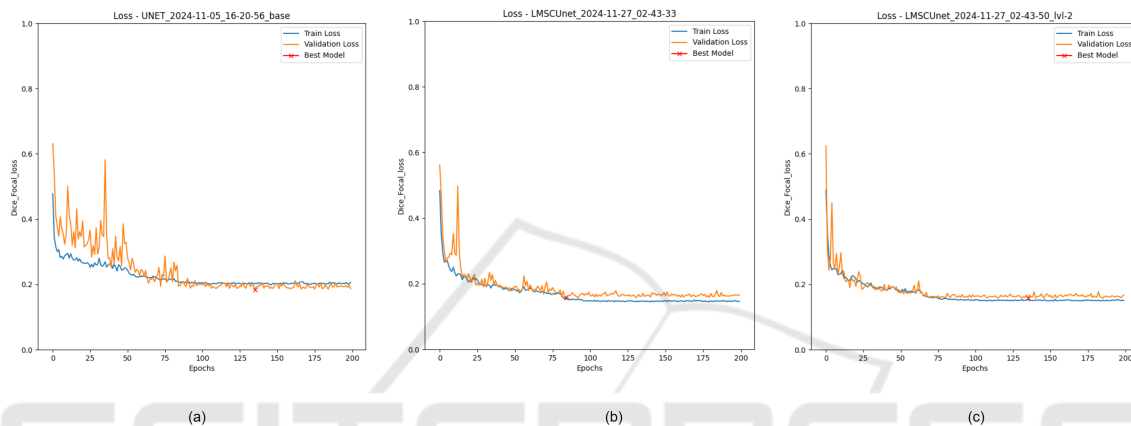| Model	ACC.(%)	mIoU	F1	DC
Standard U-Net	93.57	0.6817	0.8123	0.8094
LMSC-UNet (without BRB)	94.62	0.7102	0.8335	0.8301
LMSC-UNet (Ours)	94.60	0.7123	0.8346	0.8313

Figure 4: Learning curves. (a) Standard U-Net (b) LMSC-UNet without BRB (c) LMSC-UNet (Ours).

of LMSC-UNet and other models. By incorporating MBB units in the encoder of U-Net, the number of parameters of our model is only 0.9M and the model size is only 3.45MB, which is not even a quarter of that of standard U-Net. Meanwhile, LMSC-UNet needs approximately 8.1B MACs, which are 7× fewer compared to the standard U-Net. After adopting a BRB, a bottleneck block in the second level of skip connection, which adds a small computational overhead, the number of parameters and MACs utilization of LMSC-UNet are slightly higher than that of LMSC-UNet without a BRB (unaltered skip connection). Furthermore, LMSC-UNet with and without BRB is almost three times faster in inference than that of standard U-Net.

4.2 Segmentation performance on NuInsSeg Dataset

To verify the improvement in the performance due to adoption of MBB units and BRB in U-Net, the segmentation results of U-Net, LMSC-UNet (without a BRB) and the proposed LMSC-UNet are presented in Table 2. Both LMSC-UNet without BRB and with BRB are seen to achieve comparable performance in

terms of segmentation accuracy, mIoU, F1 score and DC, while being computationally more efficient than the standard U-Net model. More specifically, BRB contributes to the higher segmentation performance of LMSC-UNet in comparison to the standard U-Net. Compared to the standard U-Net, when standard convolution blocks in the encoder are replaced with MBB units, the ACC., mIoU and DC of our lightweight model are enhanced by almost 1%, 2.85% and 2%, respectively. By adopting BRB in the second level of skip connection, the segmentation performance is further improved, which clearly demonstrates the benefits of refining intermediate features before combining with decoder feature maps. Since we have retained unaltered skip connections at other levels, the decoder gets sufficient spatial details and avoids possible accuracy degradation.

For a more comprehensive validation, especially to validate the effect of MBB units and BRB path, the learning curves are also provided in Figure 4. As shown in Figure 4(a), the training process of standard U-Net is unstable, whereas LMSC-UNet without BRB improves the stability of training, however, suffers from little overfitting. The reason behind the overfitting is, MBB units mainly prioritize the param-
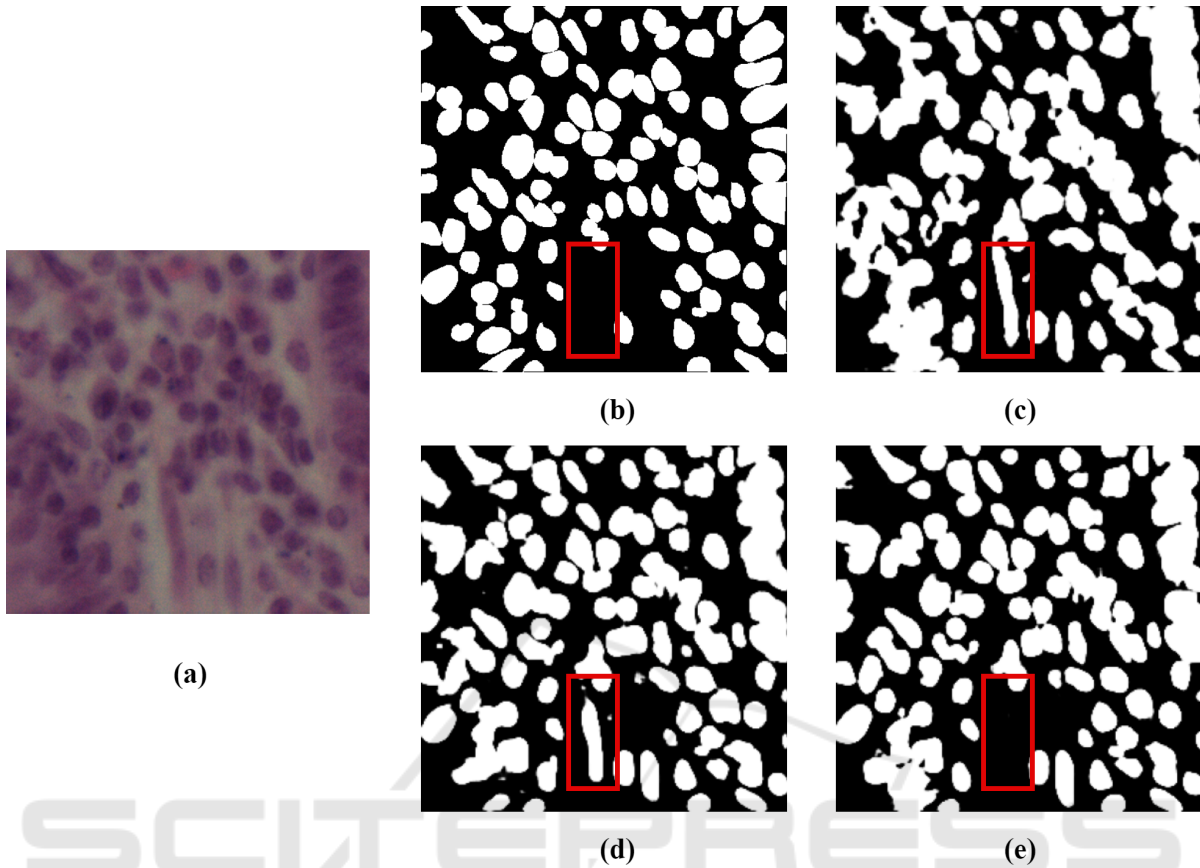
Figure 5: Qualitative results on NuInsSeg dataset. (a) Input image (b) Ground truth (c) Standard U-Net (d) LMSC-UNet without BRB (e) LMSC-UNet (Ours).

Table 3: Ablation study of BRB adopted at different levels of skip connections in LMSC-UNet.

Model	Params (M)	MACs (B)	ACC.(%)	mIoU
LMSC-UNet-1	0.9	8.2	94.56	0.7093
LMSC-UNet-2 (Ours)	0.91	8.2	94.60	0.7123
LMSC-UNet-3	0.93	8.25	94.61	0.7079
LMSC-UNet-12	0.91	8.4	94.39	0.7022
LMSC-UNet-23	0.94	8.4	94.80	0.7071
LMSC-UNet-123	0.95	8.5	94.79	0.7141

eter and computation reductions. However, this strategy may lead to loss of some fine details which affects the segmentation performance. With inclusion of BRB, the proposed model passes only mid-level relevant features to the decoder, which further reduces overfitting and follows a stable training process. The BRB unit not only ensures smooth gradient propagation but also avoids vanishing gradients, leading to better training convergence. Qualitative results on the NuInsSeg dataset are displayed in Figure 5. Red rectangles in the Figure represent the regions with precise and accurate segmentation. Due to the large number of parameters, the standard U-Net may have memo-

rized training data, including noise which resulted in wrong predictions. LMSC-UNet without BRB also produced wrong predictions. On the other hand, the BRB unit in the proposed LMSC-UNet passed only meaningful features and produced segmentation maps which align more closely with the ground truth, as shown in Figure 5(e).

Overall, the combination of MBB and BRB helps the model to extract rich features and ensures that only meaningful and less redundant features are passed to the decoder, enhancing the model's segmentation ability.
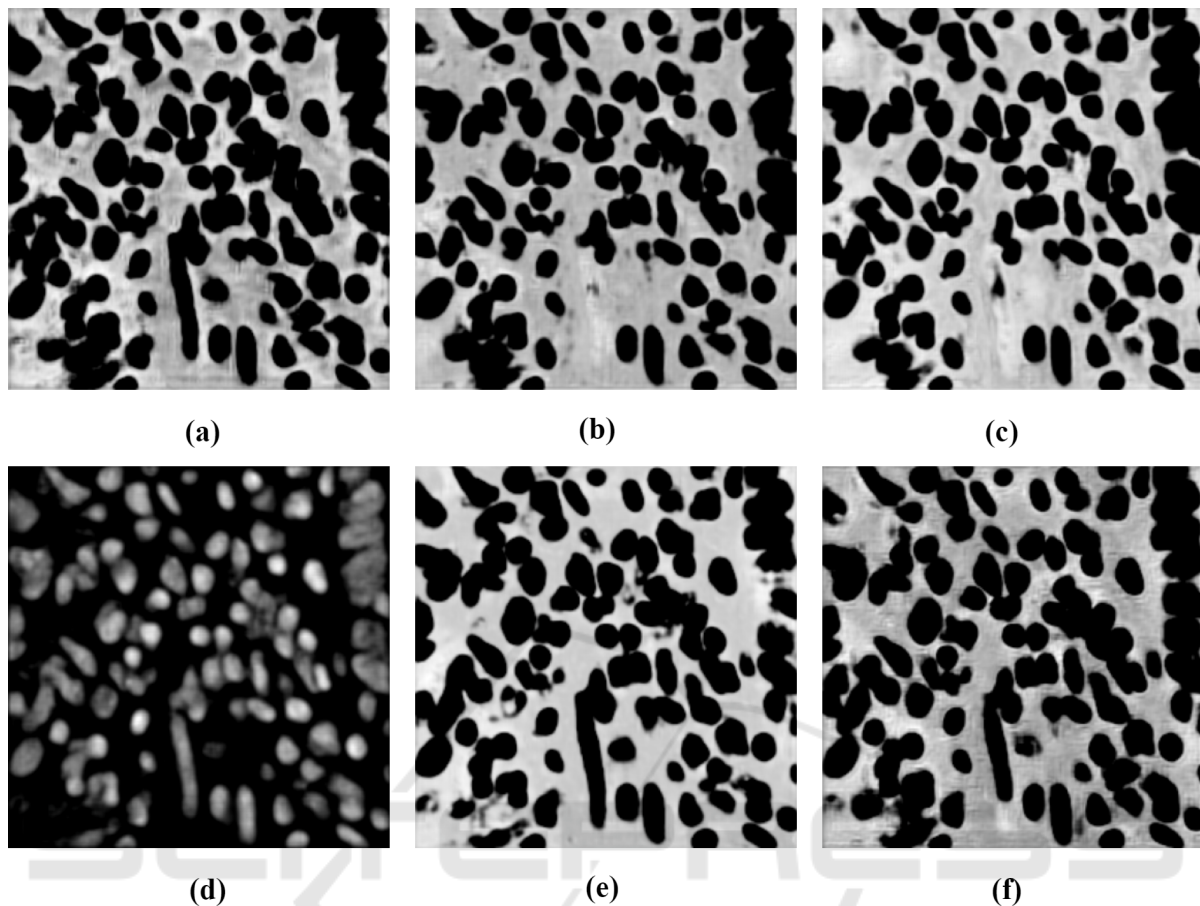
Figure 6: Visualization of feature maps. (a) LMSC-UNet-1 (b) LMSC-UNet-2 (Ours) (c) LMSC-UNet-3 (d) LMSC-UNet-12 (e) LMSC-UNet-23 (f) LMSC-UNet-123.

Table 4: Segmentation performance of LMSC-UNet and other models on RF signal dataset.

Model	ACC.(%)	mIoU	F1	DC
Standard U-Net	98.19	0.9172	0.9547	0.9545
LMSC-UNet (without BRB)	97.29	0.8904	0.9394	0.9392
LMSC-UNet (Ours)	97.27	0.8903	0.9391	0.9390

4.3 Ablation Study

In order to find the level at which BRB can be incorporated in the skip connection that ensures only relevant features will be passed to the decoder, we conducted a series of ablation experiments on inclusion of BRB at various levels of skip connections. Table 3 presents the segmentation performance and corresponding computational demands of different versions of the proposed model. The first level skip connection in the LMSC-UNet is responsible for low-level, fine grained spatial information, second level carries intermediate feature representations, whereas third level carries high-level abstract features to the corresponding decoder. Therefore, refining low-level features may lose fine spatial details which are critical for reconstructing complex object structures, whereas refining high-level abstract features may not contribute much to the final segmentation maps. This has been confirmed by the results recorded in Table 3. In comparison to BRB at other levels of skip connections, the proposed model (BRB at second level) was able to achieve excellent segmentation performance with minimal number of parameters and MACs. Additionally, feature maps extracted from the last decoder block corresponding to each modification is shown in Figure 6. Since the last decoder block plays a critical role in producing the final segmentation map, the feature maps from this layer highlights the richness of semantic details available for segmen-
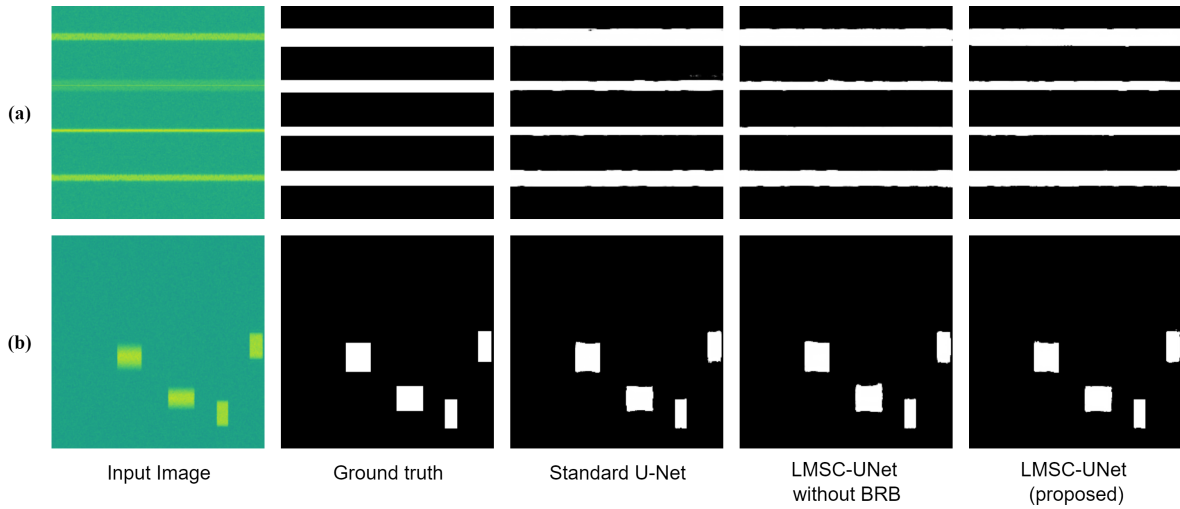
Figure 7: Qualitative results on RF Signal dataset. (a) A spectrogram image 1 (b) A spectrogram image 2.

tation. It is clearly seen that, BRB at the second level of skip connection (our model) passes more relevant features and produces more accurate segmentation.

4.4 Segmentation performance on RF Signal Dataset

To further demonstrate the robustness and generalizability of the proposed model, we evaluate it on RF signal dataset, where input data vary in format, pattern, object orientations than that of histological imaging dataset, NuInsSeg. These experimentations also ensure that the LMSC-UNet is reliable and versatile. Table 4 presents segmentation performance of the proposed model and other models on RF signal dataset. The dataset is quite larger than that of the NuInsSeg dataset. Therefore, it is quite interesting to analyse our lightweight model with a large dataset as well. As shown in Table 4, the segmentation results of LMSC-UNet with and without BRB are marginally lower than that of standard U-Net. The large number of parameters contributes to higher performance of standard U-Net apparently, whereas our lightweight model slightly struggles to maintain the segmentation performance. Albeit showing lower segmentation performance than standard U-Net, our model with and without BRB shows a better trade-off by reaching a mIoU of 0.8903 while requiring less than 1M parameters. These results highlight the generalizability of our lightweight model when dealing with relatively large datasets. Qualitative results on RF signal dataset are displayed in Figure 7. A spectrogram image of a wideband signal containing four signals of long duration, its ground truth and segmented maps generated by different models are shown in Fig-

ure 7(a), whereas Figure 7(b) shows a spectrogram image of a wideband signal containing four signals of short duration, its ground truth and segmented maps generated by different models. The segmented maps generated by standard U-Net, LMSC-UNet with and without BRB look almost similar. Despite achieving lowest segmentation performance on test data, our lightweight LMSC-UNet model accurately detected even signals with short duration.

5 CONCLUSION

In this study, we proposed LMSC-UNet, a lightweight and robust architecture for accurate and precise segmentation. The incorporated MBB units in the encoder not only reduces the number of trainable parameters but also facilitates faster inference by using fewer MACs. Additionally, using BRB in the second level of skip connection, the proposed model ensured that only relevant features are passed to the decoder which further enhanced the segmentation accuracies. We evaluated our LMSC-UNet model on two different datasets from different domains and observed that it obtains comparable segmentation performance by utilizing a smaller number of parameters ($\sim$ 0.9M), significantly fewer MACs and low memory consumption. Overall, the proposed LMSC-UNet model obtained a better trade-off between computational resources and segmentation performance, making it suitable for limited-resource devices and real-time applications.

ACKNOWLEDGEMENTS

REFERENCES

Akkus, Z., Galimzianova, A., Hoogi, A., Rubin, D. L., and Erickson, B. J. (2017). Deep learning for brain MRI segmentation: state of the art and future directions. *J Digit Imag*, 30(4):449–459.

Al-Masni, M. A. and Kim, D.-H. (2021). Cmm-net: Contextual multi-scale multi-level network for efficient biomedical image segmentation. *Scientific Reports*, 11(1):1–18.

Badrinarayanan, V., Kendall, A., and Cipolla, R. (2017). Segnet: A deep convolutional encoder-decoder architecture for image segmentation. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 39(12):2481–2495.

Beheshti, N. and Johnsson, L. (2020). Squeeze u-net: A memory and energy efficient image segmentation network. In *2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops (CVPRW)*, pages 1495–1504.

Cao, Y., Liu, S., Peng, Y., and Li, J. (2020). Denseunet: densely connected unet for electron microscopy image segmentation. *IET Image Processing*, 14(12):2682–2689.

Chen, L.-C., Papandreou, G., Kokkinos, I., Murphy, K., and Yuille, A. L. (2018). Deeplab: Semantic image segmentation with deep convolutional nets, atrous convolution, and fully connected crfs. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 40(4):834–848.

El-Assiouti, H. S., El-Saadawy, H., Al-Berry, M. N., and Tolba, M. F. (2023). Lite-srgan and lite-unet: Toward fast and accurate image super-resolution, segmentation, and localization for plant leaf diseases. *IEEE Access*, 11:67498–67517.

He, K., Zhang, X., Ren, S., and Sun, J. (2016). Deep residual learning for image recognition. In *2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 770–778.

Li, X., Wang, Y., Tang, Q., Fan, Z., and Wu, J. (2019). Dual U-Net for the segmentation of overlapping glioma nuclei. *IEEE Access*, 7:84040–84052.

Long, J., Shelhamer, E., and Darrell, T. (2015). Fully convolutional networks for semantic segmentation. In *IEEE Conf Comp Vision and Pattern Recog*, pages 3431–3440. IEEE.

Mahbod, A., Polak, C., Feldmann, K., Khan, R., Gelles, K., Dorffner, G., Woitek, R., Hatamikia, S., and Ellinger, I. (2024). Nuinsseg: A fully annotated dataset for nuclei instance segmentation in h&e-stained histological images. *Scientific Data*, 11(1):1–7.

Meng, X., Yang, Y., Wang, L., Wang, T., Li, R., and Zhang, C. (2022). Class-guided swin transformer for semantic segmentation of remote sensing imagery. *IEEE Geoscience and Remote Sensing Letters*, 19:1–5.

Milletari, F., Navab, N., and Ahmadi, S.-A. (2016). V-Net: Fully convolutional neural networks for volumetric medical image segmentation. In *Fourth IEEE Int Conf 3D Vision*, pages 565–571, Stanford, CA, USA. IEEE.

Nam, M., Oh, S., and Lee, J. (2024). Quantization of u-net model for self-driving. In *2024 10th International Conference on Applied System Innovation (ICASI)*, pages 1–3.

Nawaratne, R., Alahakoon, D., De Silva, D., and Yu, X. (2020). Spatiotemporal anomaly detection using deep learning for real-time video surveillance. *IEEE Transactions on Industrial Informatics*, 16(1):393–402.

Ronneberger, O., Fischer, P., and Brox, T. (2015). *MICCAI2015*, chapter U-Net: Convolutional networks for biomedical image segmentation. Springer.

Sandler, M., Howard, A., Zhu, M., Zhmoginov, A., and Chen, L.-C. (2019). Mobilenetv2: Inverted residuals and linear bottlenecks.

Sawant, S. S., Bauer, J., Erick, F. X., Ingaleshwar, S., Holzer, N., Ramming, A., Lang, E. W., and Götz, T. (2022a). An optimal-score-based filter pruning for deep convolutional neural networks. *Applied Intelligence*, 52:17557–17579.

Sawant, S. S., Erick, F. X., Göb, S., Holzer, N., Lang, E. W., and Götz, T. (2023). An adaptive binary particle swarm optimization for solving multi-objective convolutional filter pruning problem. *Journal of Supercomputing*, 79:13287–13306.

Sawant, S. S., Wiedmann, M., Göb, S., Holzer, N., Lang, E. W., and Götz, T. (2022b). Compression of deep convolutional neural network using additional importance-weight-based filter pruning approach. *Applied Sciences*, 12(21).

Su, R., Zhang, D., Liu, J., and Cheng, C. (2021). MSU-Net: Multi-scale U-Net for 2D medical image segmentation. *Frontiers in Genetics*, 12:639930.

Vagollari, A., Hirschbeck, M., and Gerstacker, W. (2023). An end-to-end deep learning framework for wideband signal recognition. *IEEE Access*, 11:52899–52922.

Vaze, S., Xie, W., and Namburete, A. I. L. (2020). Low-memory cnns enabling real-time ultrasound segmentation towards mobile deployment. *IEEE Journal of Biomedical and Health Informatics*, 24(4):1059–1069.

Zhang, J., Zhu, H., Wang, P., and Ling, X. (2021). Att squeeze u-net: A lightweight network for forest fire detection and recognition. *IEEE Access*, 9:10858–10870.

Zhao, P., Li, Z., You, Z., Chen, Z., Huang, T., Guo, K., and Li, D. (2024). Se-u-lite: Milling tool wear segmentation based on lightweight u-net model with squeeze-and-excitation module. *IEEE Transactions on Instrumentation and Measurement*, 73:1–8.

Zhou, Z., Siddiquee, M. M. R., Tajbakhsh, N., and Liang, J. (2020). Unet++: redesigning skip connections to exploit multiscale features in image segmentation. *IEEE Trans Med Imaging*, 39:1856–1867.