A Study of Multi-Task Learning Using a VoVNet-OSA Block Enhanced U-Net on the Med++ MNIST Dataset

Syuan-Hao,Li[1] *, Jie-Xin,Li[1] **, Jing-Yuan,Ting[1]

[1]*Department of Computer Science and Engineering, National Taitung University, Taitung, Taiwan*
**hibana2077@gmail.com*
***john489142@ gmail.com*

Keywords: Multi-task Learning, VoVNet-OSA Block Enhanced UNet, Med++ MNIST Dataset, Image Classification

Abstract

In this study, we propose a multi-task learning framework using a modified VoVNet-OSA Block Enhanced UNet, named VovUnet_Var, for image segmentation and classification on the Med++ MNIST dataset. VovUnet_Var features downsampling (DownOsa) and upsampling (UpOsa) blocks, with a classification head. The architecture sequentially downscales the input image to capture hierarchical features and uses adaptive average pooling and a fully connected layer for classification. For segmentation, upsampling layers restore spatial dimensions, producing segmentation masks. This model effectively handles complex medical imaging tasks, providing a robust solution for simultaneous image segmentation and classification.

1 Introduction

This introductory section provides a brief overview of U-Net. It then goes on to our study.

U-Net [1], a convolutional network with data augmentation, can efficiently segment biomedical images from few images, outperforming prior methods and winning the ISBI cell tracking challenge 2015. [2] proposes a simple yet effective end-to-end depth-wise encoder-decoder fully convolutional network architecture, named Sharp U-Net, to work around the drawback of merging semantically different low- and high-level convolutional features, which causes both blurred feature maps and over- and under-segmented target regions. [3] integrates the Inception-Res module and densely connecting convolutional module into the U-net architecture. AdaResU-Net [4], which integrates the U-Net architecture with a residual learning framework and adaptive convolutional neural network, enhances medical image segmentation performance. It achieves superior results compared to U-Net while using 30% fewer parameters.

Based on previous studies [2, 3, 4], we propose a new multi-task learning framework called VovUnet_Var for image segmentation and classification. VovUnet_Var builds on the VoVNet architecture with enhanced One-Shot Aggregation (OSA) blocks and integrates downsampling (DownOsa) and upsampling (UpOsa) blocks along with a classification head. This method captures hierarchical features through downsampling and generates precise segmentation masks through upsampling. We evaluated VovUnet_Var on the Med++ MNIST dataset and found that it performs exceptionally well in complex medical imaging tasks, offering a robust solution for simultaneous segmentation and classification.

2. Methodology

In this study, we employed the Med++ MNIST dataset, where each image is resized to 224x224 pixels. We utilized a batch size of 16 for training. The model optimization was performed using the Adam optimizer with a learning rate set to 1e-4. To adjust the learning rate dynamically, we applied the CosineAnnealingLR scheduler with T_max set to 100. The total number of training epochs was set to 10.

For the loss functions, we used the CrossEntropyLoss for the classification task and the Mean Squared Error (MSE) Loss for the segmentation task. The training process involved the following steps: the model was set to training mode, and for each batch, the input images and corresponding targets were loaded onto the device. The targets were then reshaped as needed. The optimizer's gradients were reset before the forward pass. The model produced outputs for both segmentation and classification tasks, and the respective losses were computed. The combined loss (classification loss plus segmentation loss) was used to update the model weights through backpropagation. Accuracy and loss metrics were tracked and reported at the end of each epoch to monitor training progress.

3 Results

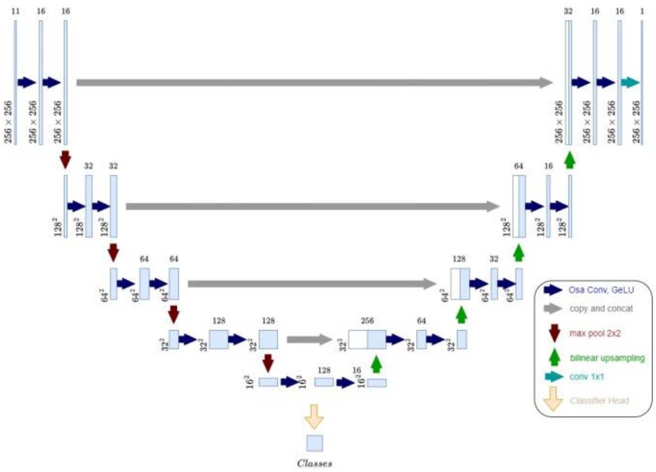
Our experiments evaluated multiple models on the Med++ MNIST dataset, which includes various sub-datasets such as BloodMNIST, BreastMNIST, DermaMNIST, OrganAMNIST, OrganCMNIST, OrganSMNIST, PathMNIST, PneumoniaMNIST, and RetinaMNIST. The primary focus was on the performance of our modified VoVNet-OSA Block Enhanced UNet (VovUnet_Var) compared to other models.

The VovUnet_Var achieved outstanding performance across most datasets. Specifically, it obtained the highest accuracy on

BloodMNIST (98.77%), BreastMNIST (88.46%), DermaMNIST (77.06%), OrganAMNIST (96.20%), OrganCMNIST (93.40%), OrganSMNIST (81.82%), PathMNIST (95.00%), PneumoniaMNIST (95.19%), and RetinaMNIST (57.25%).

In comparison, the UNet_VAR also performed well, with significant results on BloodMNIST (97.95%), BreastMNIST (85.26%), DermaMNIST (74.81%), OrganAMNIST (95.89%), OrganCMNIST (92.55%), OrganSMNIST (81.38%), PathMNIST (94.87%), PneumoniaMNIST (94.87%), and RetinaMNIST (58.25%).

Other notable performances include DenseNet201 on BloodMNIST (98.25%) and PneumoniaMNIST (91.83%), and ResNet18 on RetinaMNIST (55.00%).



a

Fig. 1 (a) Architecture diagram of VovUnet_Var.

4 Discussion

The results indicate that VovUnet_Var outperforms other models across most Med++ MNIST sub-datasets, showcasing its robustness in multi-task learning for medical image analysis. The VoVNet-OSA blocks enhance feature extraction, leading to higher accuracy in tasks such as BloodMNIST and PneumoniaMNIST, demonstrating its potential for critical diagnostic applications.

Despite the strong performance, the accuracy on RetinaMNIST was relatively lower, suggesting a need for further fine-tuning or model adjustments for specific medical images. This highlights that while VovUnet_Var is versatile, some datasets may still benefit from tailored approaches.

Overall, VovUnet_Var proves to be a powerful tool for both classification and segmentation tasks, with potential for further optimization to enhance its effectiveness across all types of medical images.

5 Conclusion

The VovUnet_Var model excels in multi-task learning for medical image analysis, achieving top accuracy across most Med++ MNIST sub-datasets. Its enhanced feature extraction capabilities make it robust for both classification and segmentation tasks. The model shows great potential in critical diagnostics, though further fine-tuning is needed for specific datasets like RetinaMNIST. Overall, VovUnet_Var proves to be a powerful and versatile tool, with future work aimed at optimizing performance across all medical image types.

6 References

[1] O. Ronneberger, P. Fischer and T. Brox. "U-Net: Convolutional Networks for Biomedical Image Segmentation." ArXiv, abs/1505.04597 (2015). https://doi.org/10.1007/978-3-319-24574-4_28.
[2] H. Zunair and A. Hamza. "Sharp U-Net: Depthwise Convolutional Network for Biomedical Image Segmentation." Computers in biology and medicine, 136 (2021): 104699 . https://doi.org/10.1016/j.compbiomed.2021.104699.
[3] Ziang Zhang, Chengdong Wu, S. Coleman and D. Kerr. "DENSE-INception U-net for medical image segmentation." Computer methods and programs in biomedicine, 192 (2020): 105395 . https://doi.org/10.1016/j.cmpb.2020.105395.
[4] Maria G. Baldeon Calisto and S. Lai-Yuen. "AdaResU-Net: Multiobjective adaptive convolutional neural network for medical image segmentation." Neurocomputing, 392 (2020): 325-340. https://doi.org/10.1016/J.NEUCOM.2019.01.110.