Proceedings of CONF-MLA 2025 Symposium: Intelligent Systems and Automation: AI Models, IoT, and Robotic Algorithms
DOI: 10.54254/2755-2721/2025.LD28949

A Review of Improvement Strategies for CNNs in Medical Image Segmentation: From Classification Networks to Efficient Segmentation

Bowen Huo

International School of Information Science&Engineering, Dalian University of Technology, Dalian, China
qq13920358345@126.com

Abstract. Convolutional neural networks(CNNs), thanks to its amazing feature learning capacity, has shown a lot of application opportunities on medical images. It is becoming more and more important for disease diagnosis and treatment plan. But for complex medical image processing data,CNNS have problems like too much computational resources are consumed,model is not good at generalizing,and too many labeled data are required. In recent years, in order to solve such problems, various studies by different people have emerged focusing mainly on developing more efficient models of CNNs. This paper mainly reviews how do those improvement techniques help to improve medical image classification and segmentation accuracy, also can speed up model inference process and make the model more robust. Through a deep integration of current research, it is concluded that the improvement of CNN architectures is an effective means to enhance the accuracy and speed of medical image segmentation; and these improvements have clear advantages over different tasks and different types of medical images.

Keywords: Medical image processing, CNN, Image segmentation, Improvement strategies, Deep learning, Model efficiency

1. Introduction

The Medical Image Segmentation is the primary task in medical image. It is exacting with irreplaceable value for medical treatment, disease localization, surgery planning and evaluating therapy in the early stages. But the inherent complexity of medical pictures like low contrast, noise interference, fuzzy boundary, high variable anatomy, different image modality cause big challenge in making sure high accurate and robust segmenting.

Recently, thanks to the wonderful feature learning capability as well as outstanding ability in pattern recognition, CNN can achieve great progresses in analyzing medical images and has a lot of application prospect especially for image classification problems. Can automatically extract multi-level, high-level features from the raw pictures, so that it can greatly simplify the cumbersome process of manual feature extraction required in traditional medical image analysis. And for the like ashwath etc., a new 3 layers self explanatory explanation DCNN is given. It can enhance its multi-region image classification of various kinds, but it still can be interpreted. It is good at customized spot data and the public data set of PAD-UFes-20 skin lesion [1] Furthermore, addressing the urgent need for

© 2026 The Authors. This is an open access article distributed under the terms of the Creative Commons Attribution License 4.0 (https://creativecommons.org/licenses/by/4.0/).

DOI: 10.54254/2755-2721/2025.LD28949

rapid COVID-19 diagnosis, Wang et al. developed COVID-Net, a CNN-based diagnostic assistance system capable of effectively identifying COVID-19 in chest X-rays, significantly aiding clinical decision-making [2]. In Ophthalmology there was one by Gulshan and in that they stated that it would be useful if we use a Deep learning for an automated diagnosis of DR with Grading Scale & Applied Large-Scale Screening using Retinal Fundus Picture [3]. And for those working with Image segmentations: Encoder - Decoder Unet wins on medical images because the special skip connection mix up some of that shallow and space related stuff. And it made cell and lesion segmentation possible [4]

Though CNNs have had great success with image classification, applying them to medical images is still quite challenging. Review starts with the evolution of CNN classification networks into segmentation networks and shows us what were the major principles and structure changes involved. The review hereafter will address new ways to make segments better, help models be more efficient, and assist in making things generalizable. It will give a close examination and comments regarding the equilibrium methods among efficiency as well as improvement techniques.

2. From classification to segmentation: foundational architecture transformation methods

The CNN was first created mainly for picture classification, the whole image is mapped into 1 or more categories. But medical images need to segment each pixel in the picture - so-called pixellevel prediction, which is very different from ordinary classification. To bridge this gap,a lot of changes were made to CNN architecture by researchers and was also able to do heavy work on image pixel.

2.1. Fully convolutional networks: the cornerstone of pixel-level prediction

An important step towards going from object level predictions to pixel level predictions is FCN by Long et al. in 2015 [5]: FCNs main idea is to replace all the fully-connected layers right before the end of classification network with some convolutional layers. The classification network fully-connected layer often flattens this kind of spatial feature map into a 1d-vector, it will lose some spatial info. But, pure fully conv is problematic; it has lots of consecutively pooled encodings which greatly reduces the size and detail of features. And it brings about fuzzy boundaries and bad accurate. In order to overcome this, FCNs had skip connections.

2.2. Evolution of the encoder-decoder framework

Building on the groundwork of fully convolutional networks (FCN)s for predicting pixels, encoder-decoder (ED) architecture became quickly one of the most important paradigms in the area of medical images segmentation. The endgame here is U-net, which was put forth by Ronneberger et al. U-net basic success led to lots more research coming after it, creating many other classic variants trying to make it even better. Among which is U-Net++, it addressed the gap of semantics between encoder-decoder feature maps [6] With more and denser skip-connections, there are now many different info pathways going b/w encoder&decoder: Decoders takes in feature aggregate from all these different depths, this kind of semantic levels stage, so they can have a bit smoother fuse with their features for a lot better segmentation work. And another important one is the ResUnet, which combines Deep Residual Learning to the Unet [7] Using residual module instead of usually convoluted block can solve the problem that gradient will be vanished when training deep network; Therefore, we have more chance to create much deeper and better segmentation models in order for them learn hard feature representation. Along with the popularity of 3D medical imaging technology like CT, MR, etc., the development of 2D-segmentation model is moving towards a necessity for extension to 3D as well. The V-Net developed by Milletari et al is representative of this work [8].

DOI: 10.54254/2755-2721/2025.LD28949

2.3. Limitations of direct classification network transfer

Although the classification network adaptation is quite successful, there always remains a contradiction between designing the classification network to begin with and the basic needs of segmentation tasks. Directly applying classification network in segmentations have two large flaws, it can lose a lot of spatial resolutions and lack overall modeling. First, the spatial resolution loss is a basic flaw in classification network architecture. In order to learn the high-level semantic features which were target location/ shape agnostic, in this case we could use poolig layers or downsample via strided convolutions. This is indeed an increased amount of the receptive field that would decrease the amount of computation we do, but at the cost of fine-graining spatial position information. Architectures like the DeepLab series and FPN [9,10] started to appear due partly from trying to solve this problem that had been around for a long time in segmentation tasks. The 2nd problem point out is no world-scale modeling.

3. Performance enhancement: core improvement mechanisms

To overcome the natural shortcomings of a normal convoluting neural network for medical images and so on, a lot of complicated improvements have been made. In all these mechanisms greatly improves segmentation model performances in many aspects such as makes network see the important feature better, make multi-scale information combine better, let information flow better inside network and use more specific loss function

3.1. Attention mechanisms enhance target perception

Spatial attention SA, “where do I look?” generates a size matched with that of a feature map, which is weighting features placed in various spatial positions. Higher weight means more importance to that region for the task. For instance when it comes to medical image segmentation it can also let model concentrate on lesions, target organs and ignore some unimportant background information. The other way is Channel Attention(CA) which is about “learn what to look for”. It models how different feature channels depend on one another, so it gives higher importance to certain meaningful feature channels and de-emphasizes others [11].

3.2. Multi-scale feature fusion strategy

Since there are various sizes of target in one medical image, it is necessary to add multi-scale detail accurately: Pyramid Scene parsing Network (PSPNet) improve its global and local context perceiving ability with the help of a Pyramid Pooling Module(PMM) at the last layer of CNN which collect the context information in different area [12] In the PPM, it applies several different size average pooling kernel at once to the feature maps created from backbone network and obtain context features at different scale. Using traditional CNN successively adding the receptive field with a pooling layer, thus will lost spatial resoulution by loss some fine detailed info. The Atrous Convolutions and ASPP (Atrous Spatial Pyramid Pooling) Module of the DeepLab Series give an answer to this. They effectively increase the receptive field but keep the feature map resolution so that it is better at representing multiscale objects [13].

3.3. Deep supervision and advanced connection structures

DeepSupervision's main thought was to add some loss terms on the intermediate layer as well, so we have "end-to-end" supervision all the way during network training time. It is not only can improve the vanishing gradient issue of deep networks, it could even enable more discriminative intermediate representations to be learnt. UNet3+ is like that. It uses the full scale skip connection and deep

DOI: 10.54254/2755-2721/2025.LD28949

supervision: links every decoder, so whole group with both, encoder feature map which is at all level and also small decoders features too. And calculate for losses on each one [14]: The Dense Connections, again lots of features in here too: DenseNet is the most information flowing in image classifying task by linking every layer to all following ones [15].

3.4. Loss function innovation

Medical images generally have an extremely unbalanced amount of foreground and background pixels. Traditional cross-Entropy loses it to the large quantity Dice Loss is based on the Dice similarity coefficient and directly optimizes overlap with ground-truth label segments, so it is insensitive to class imbalance and hence very common in medical image segmentation[8]; Focal Loss handles these by modifying regular CrossEntropyloss, which would normally give lower weights to already easy to tell apart stuff(like those many hard background pixes), It makes it have to learn hard to tell apart from foreground or boundaries [16]. Loss functions such as Dice Loss depending on region overlap might be indifferent about the precision of segmentated boundaries. Hausdorff Distance(HD) is the maximal distance that two point sets mismatch, so it is very sensitive to boundary difference in segmentation. Add the hausdorff distance to the loss function will make sure that the differenciate between prediction and real boundary, so it would produce smoother and better result. Just as the example of Karimietal'sHD - Net which has used the Hausdorff distance to improve 3D medical imagesegmentation [17].

4. Efficient segmentation: lightweight and real-time optimization

4.1. Lightweight architecture design

Successfull scalability high performance segmentatoin model it is due to the computing poweer and also memoty requirement is very hgih. So, light weight architecture design which greatly reduces the complexity of models but keep or slightly reduce diagnostic accuracy has become very important for doing medical image segmentation. This is roughly along two dimensions technologically first build the light ones, from whole cloth if you will and then slimming already established high accuracy models Depthwise Separable Convolution(DSC) is an important component which we can build our inherently efficient module.

4.2. Computational efficiency optimization strategies

From algorithm perspective : Doing less unneccessary calculation means getting a better look on the innards of the network's information. It is quite apparent in dc's. Traditional CNNs use successive downsampling(pooing)operations in order to get bigger and better understanding of the whole scene. But it must inevitably lead to a reduction in space, and after all the upsampling recuperation, there will be another burden. Deployment level hardware adaptation is like the final leg to get your model inference running smoothly, it is all about network quantization which has a big impact. when a model that should be training / running using only fp32 is trained / run on an server grade gpu it should work fine. It is become a choke point when we go to the clinic with smaller computation, less electricity and shorter memory Quantization techniques basically resolve it by turning the models FP32 values of weight and activations into small bits of: And we could convert it by Post-Training Quantization(PTQ) or quantization aware training(QAT), where the latter mimics effects of quantization at training time.

DOI: 10.54254/2755-2721/2025.LD28949

4.3. Efficient processing solutions for 3D segmentation

3D medical image segmentation is plagued by computational and memory costs rising cubically, so effective processing solutions are vital. 2.5D compromise strategy is a good practical engineering answer to deal with those two kinds of extreme approach: Pure-3d and pure-2d Only use 3D model(eg.V-Net)can get all spatial context but require much compuatation and gpu memory.pure only use 2D is very effective but ignore the continuity information between every slices. The essence of the 2.5D strategy is that for a given slice, we take an slab/patch around this slice from the 3D volume as our input, rather than just one slice or the whole volume.

5. Medical applications and performance validation

And the improved cnn architechture has proved that it can achieve its theoretical benefit on different clinical situation. Have some specific organ segmentation or detections on lesions, do real-time performance evaluation, it is a tangible thing that we can try out even more methods.

5.1. Organ segmentation task

For cardiac MRI, it is very important to segment the chambers like the atria and ventricles in order to calculate cardiac output (like ejection fraction). It is U-net++, which did best on this. It uses many different levels of sematics with its nested/dense skip connection Like the one saying U-net ++ is better then stanard Unet when using for segmenting complicated boundaries like an atriapulmonary vees junction, where there was less error due to semntatic feature differnces.

5.2. Lesion detection and quantification

Automatic segmentation and nodule detection is very important for the early screening of lung cancer but it has a lot difficulty: Nodules have different sizes, morphologies; they may attach with vessels or pleura which leads to unclear boundary. To solve the above problems,effective methods of multi - scale features fusion improve the probability of detecting small targets and segmenting them. We can also introduce module like ASPP or PPM in which we could take into account context from receptive field on several scales. Minute Nodules: High-resolution Feature Maps have kept the fine details.

5.3. Efficiency-accuracy tradeoff practices

And then real time ultrasound segmentaion, you need to do this fast like video rate on a portable in order to give your clinician immediate feedback. In order to do so, MobileUnet+ modelQuantization has become something of the gold standard. for mobileUNet etc, use DEEP Sep-conv to get light weight en-coder/de-co-ders which save PARAM & COMP at ARCH. Lev Next we do an int8 quantization which will convert all the model floating-point operations to much faster int8 operations leveraging speed at the edge.

6. Conclusion

This article reviews the main improvements on CNN in terms of medical image segmentation. To solve the problems of spatial information lost and computational overburden caused by traditional classification networks on segmentation tasks, people first laid the groundwork with FCN, encoder-decoder architecture around U-Net. To further increase the degree of accuracy attentional methods and several feature-scaling-fusions were integrated for improved comprehension of what parts are important by the model, scalability with various sizes is accomplished. Meanwhile,to satisfy the real-time and deployment requirement of clinical application at the same time,lightweight design such as

DOI: 10.54254/2755-2721/2025.LD28949

deep separable convolution, model quantization,sparse convolutions have become an important part of research. Sum up : today work about how the medical picture would separate, now it is just trying for creating a accurate and meaningful intellgent tool that people could use in clinic, make some perfect combination so they can create something effective.

References

[1] Ashwath V A, Sikha O K, Benitez R. TS-CNN: A three-tier self-interpretable CNN for multi-region medical image classification[J]. IEEE access, 2023, 11: 78402-78418.

[2] Wang L, Lin Z Q, Wong A. COVID-Net: a tailored deep convolutional neural network design for detection of COVID-19 cases from chest X-ray images[J]. Scientific reports, 2020, 10(1): 19549.

[3] Gulshan V, Peng L, Coram M, et al. Development and validation of a deep learning algorithm for detection of diabetic retinopathy in retinal fundus photographs[J]. JAMA, 2016, 316(22): 2402-2410.

[4] Ronneberger O, Fischer P, Brox T. U-net: Convolutional networks for biomedical image segmentation[C]//International Conference on Medical image computing and computer-assisted intervention. Cham: Springer international publishing, 2015: 234-241.

[5] Long J, Shelhamer E, Darrell T. Fully convolutional networks for semantic segmentation[C]//Proceedings of the IEEE conference on computer vision and pattern recognition. 2015: 3431-3440.

[6] Zhou Z, Siddiquee M M R, Tajbakhsh N, et al. UNet++: A nested UNet architecture for medical image segmentation[C]//International Conference on Medical Image Computing and Computer-Assisted Intervention. Cham: Springer International Publishing, 2018: 3-11.

[7] Zhang Z, Liu Q, Wang Y. Road extraction by deep residual U-Net[J]. IEEE Geoscience and Remote Sensing Letters, 2018, 15(5): 749-753.

[8] Milletari F, Navab N, Ahmadi S A. V-net: Fully convolutional neural networks for volumetric medical image segmentation[C]//2016 fourth international conference on 3D vision (3DV). Ieee, 2016: 565-571.

[9] Chen L C, Papandreou G, Kokkinos I, et al. Deeplab: Semantic image segmentation with deep convolutional nets, atrous convolution, and fully connected crfs[J]. IEEE transactions on pattern analysis and machine intelligence, 2017, 40(4): 834-848.

[10] Lin T Y, Dollár P, Girshick R, et al. Feature pyramid networks for object detection[C]//Proceedings of the IEEE conference on computer vision and pattern recognition. 2017: 2117-2125.

[11] Hu J, Shen L, Sun G. Squeeze-and-excitation networks[C]//Proceedings of the IEEE conference on computer vision and pattern recognition. 2018: 7132-7141.

[12] Zhao H, Shi J, Qi X, et al. Pyramid scene parsing network[C]//Proceedings of the IEEE conference on computer vision and pattern recognition. 2017: 2881-2890.

[13] Chen L C, Papandreou G, Schroff F, et al. Rethinking atrous convolution for semantic image segmentation[J]. arXiv preprint arXiv:1706.05587, 2017.

[14] Huang H, Lin L, Tong R, et al. UNet 3+: A full-scale connected UNet for medical image segmentation[C]//ICASSP 2020-2020 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP). IEEE, 2020: 1055-1059.

[15] Huang G, Liu Z, Van Der Maaten L, et al. Densely connected convolutional networks[C]//Proceedings of the IEEE conference on computer vision and pattern recognition. 2017: 4700-4708.

[16] Lin T Y, Goyal P, Girshick R, et al. Focal loss for dense object detection[C]//Proceedings of the IEEE International Conference on Computer Vision. 2017: 2980-2988.

[17] Karimi D, Salcudean S E. Reducing the Hausdorff distance in medical image segmentation with convolutional neural networks[J]. IEEE Transactions on Medical Imaging, 2019, 39(2): 499-513.