Proceedings of the 2023 International Conference on Machine Learning and Automation
DOI: 10.54254/2755-2721/32/20230179

Investigation of medical image segmentation techniques and analysis of key applications

Hao Dong

College of Letters and Science, University of California, Los Angeles, 90024, United States

haodong@g.ucla.edu

Abstract. This research examines the application of the UNet convolutional neural network model, specifically for semantic segmentation tasks in the field of medical imaging, juxtaposing its efficacy with Fully Convolutional Networks (FCNs). The primary focus of this comparative analysis rests on the performance of the UNet model on the dataset employed for this study. Surpassing our initial expectations, the UNet model demonstrated remarkable performance superiority over the FCN model on the curated dataset, thereby suggesting its potential applicability and utility for analogous tasks within the realm of medical imaging. In a surprising turn of events, our trials revealed that data augmentation techniques did not usher in a notable enhancement in segmentation accuracy. This observation was especially striking given the substantial size of the dataset employed for the experiments, encompassing as many as 1000 images. This outcome suggests that the merits of data augmentation may not always come to the fore when dealing with considerably large datasets. This intriguing discovery prompts further exploration and investigation to uncover the underlying reasons behind this observed phenomenon. Moreover, it brings to light an open-ended research query - the quest for alternative methodologies that could potentially amplify segmentation accuracy when operating on large scale datasets in the sphere of medical imaging. As the field continues to evolve and mature, it is these open questions that will continue to push the boundaries of what is possible in medical image analysis.

Keywords: semantic segmentation, UNet, fully convolutional networks, medical imaging, data augmentation.

1. Introduction

Semantic segmentation, a critical task that partitions an image into meaningful regions by assigning a corresponding label to each pixel, has garnered considerable attention due to its extensive applications. These include autonomous driving, image comprehension, augmented reality, and notably, medical imaging. Currently, semantic segmentation is integral to medical image segmentation, supporting processes like organ segmentation, tumor delineation, lesion segmentation, and anatomical structure localization. Medical image semantic segmentation plays a pivotal role in the domain of medical imaging and computer-aided diagnosis. It involves segmenting different anatomical structures or regions of interest within medical images, such as MRI, CT scans, or ultrasound images, to aid in accurate diagnosis and treatment planning. Through semantic segmentation, precise classification and segmentation of different tissue structures or pathological regions within an image are possible, thereby enabling exact localization and segmentation.

Furthermore, semantic segmentation automates the processing and analysis of images, vastly enhancing diagnostic efficiency and accuracy. This contrasts with traditional diagnostic methods, which often require manual analysis and interpretation by physicians, a process that can be time-consuming and subjective. Additionally, semantic segmentation visually represents the segmentation results, allowing physicians and patients to intuitively comprehend the distribution and morphology of pathological regions or organs. This not only facilitates effective communication but also improves patient understanding and acceptance of disease diagnosis and treatment.

Finally, semantic segmentation can be coupled with other medical imaging analysis techniques like computer-aided diagnosis and surgical navigation. This enables automation of analysis and decision-making assistance in cases such as cataract and knee joint diseases, thereby enhancing diagnostic accuracy, surgical outcomes, and overall patient care.

2. Research on key technologies of medical semantic segmentation

2.1. U-Net

To address localization in visual tasks like biomedical image processing, Ciresan et al. developed a sliding-window approach [1]. Their network predicted class labels for each pixel by using local patches as input. This approach overcame limited training data availability and enabled accurate pixel-level classification in biomedical tasks. As shown in Figure 1.

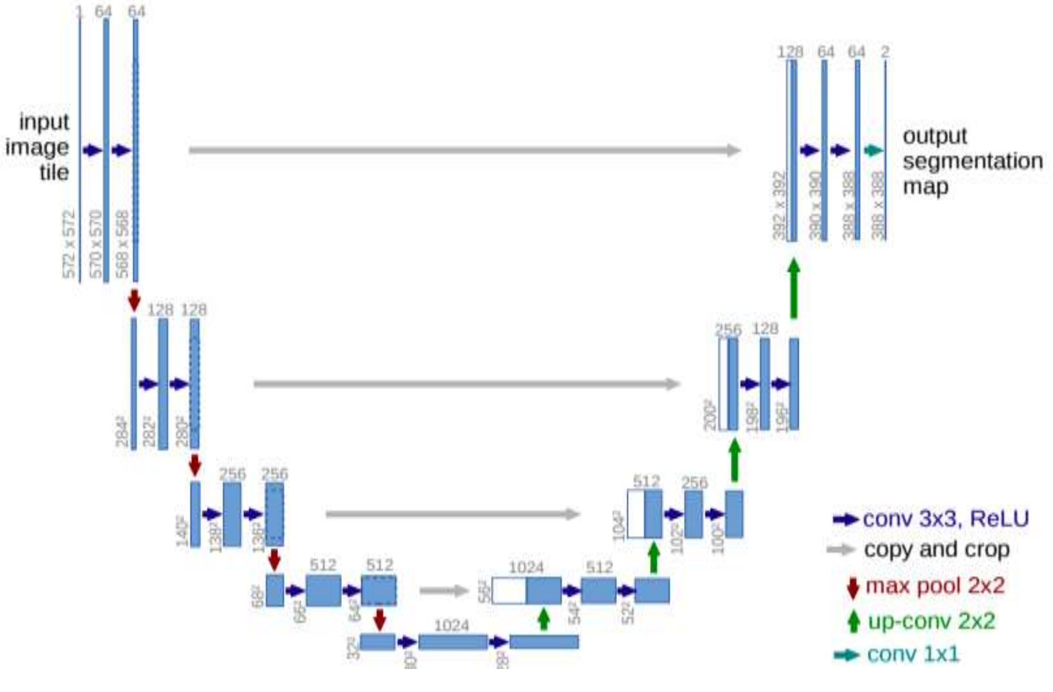

Figure 1. Unet network Structure [2].

The networks comprise a contracting path and an expansive path. To achieve downsampling and feature extraction, the contracting path utilizes a series of downsampling operations, which involve iteratively applying two unpadded 3x3 convolutions followed by ReLU activation. Additionally, 2x2 max pooling with stride 2 is employed in this process. At each downsampling step, the network doubles the number of feature channels. Conversely, the expansive path in U-Net focuses on upsampling, reconstructing the image segmentation map with higher resolution. This is achieved by applying a 2x2 convolution that reduces the number of feature channels. To ensure detailed information from earlier layers is preserved, the up-convolved feature map is concatenated with the corresponding cropped feature map from the contracting path. This fusion of information facilitates the precise localization of objects in the segmentation process.

To further refine the feature map, two 3x3 convolutions with Rectified Linear Unit (ReLU) activation functions are applied to the concatenated feature map. These convolutions help enhance the discriminative power of the network and capture more intricate details present in the input image. One challenge encountered during the U-Net architecture's implementation is the loss of border pixels due to convolution operations. To address this issue, an Overlap-tile strategy is employed, which extrapolates the missing pixels by mirroring. This technique ensures that the segmentation map covers the entire input image, even at the borders. Finally, the U-Net architecture culminates in a 1x1 convolutional layer, which transforms each 64-component feature vector into the desired number of classes. This final layer plays a crucial role in generating the segmentation map, where each pixel is assigned to a specific class, enabling precise object localization and delineation. In conclusion, the U-Net architecture has become a cornerstone in image segmentation tasks, showcasing its effectiveness in various domains. Its unique combination of contracting and expansive paths, coupled with the concatenation and convolution operations, allows for accurate and detailed segmentation of objects in images. With its robust design and impressive performance, U-Net continues to be a prominent choice for researchers and practitioners in the field of computer vision.

2.2. Common network structures such as FCN

Long et al. aimed to developed a more precise and context-aware segmentation method [3]. The structure of Fully Convolutional Networks (FCNs) incorporates position attention and channel attention modules to capture semantic dependencies in both spatial and channel dimensions, resulting in improved feature representation and accurate segmentation results. FCNs consist of multiple layers of convolutional operations, which extract hierarchical features from input images. The position attention module aggregates features at each position, taking into account the relationship between similar features regardless of their spatial distance. On the other hand, the channel attention module emphasizes interdependent channel maps by integrating associated features from all channel maps. As shown in Figure 2.

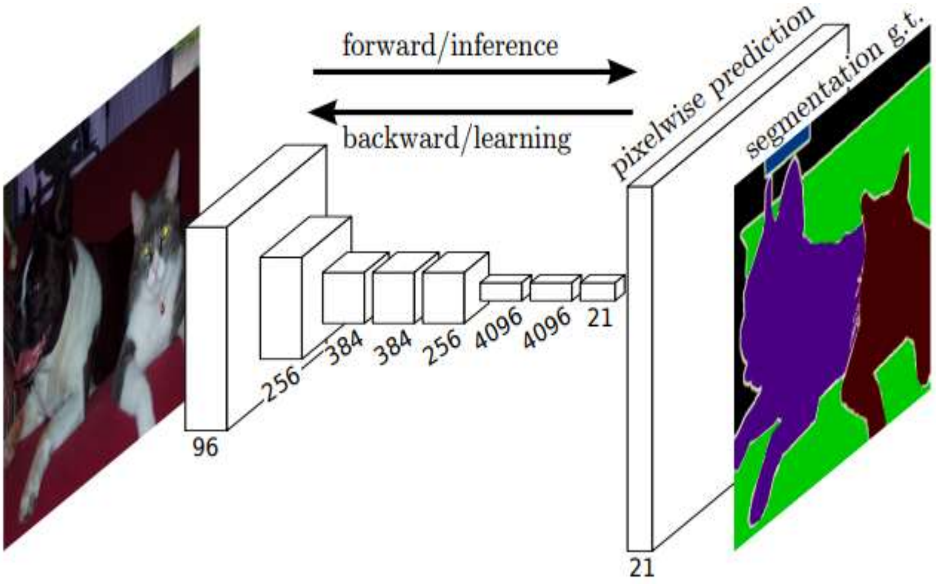

Figure 2. Fully convolutional networks [4].

Additionally, Fully Convolutional Networks (FCNs) provide flexibility in handling input image sizes. FCNs incorporate deconvolution layers to upsample the feature map from the last convolutional layer, aligning it with the original image size. This enables pixel-level predictions while preserving spatial information.

2.3. Data enhancement and regularization

Data enhancement is a method to generate new training samples by performing a series of random transformations on the original training data. These transformations can include image rotation, scaling, translation, flipping and other operations, and for text data, lexical replacement, insertion or deletion can also be performed. By applying these random transformations, data enhancement expands the diversity of the training data and helps the model better learn the invariance and generalization of the data. Regularization is a technique used to control model complexity and reduce overfitting. Common regularization methods include L1 regularization and L2 regularization, which penalize the weight of the model by adding regularization terms to the model's loss function. Regularization prevents the model from over-relying on the details and noise of the training data, thereby improving its performance on previously unseen test data.

3. Analysis of typical medical semantic segmentation applications

3.1. Cataract image segmentation

To improve the performance of models on cataracts segmentation task, Grammatikopoulou et al. focus on four prominent architectures: UNet, DeepLabV3+, UPerNet, and HRNetV2 [5]. Three distinct tasks are considered, with an emphasis on different groups of instrument classes. It is used to analyze the impact of simultaneous instrument classification

The study's findings indicate that when dealing with a small number of classes and comparable pixel representation, all four networks perform similarly and successfully tackle the class imbalance. However, HRNetV2 and UPerNet exhibit superior performance compared to DeepLabV3+ and UNet, specifically in simultaneous anatomy segmentation and instrument classification with moreclasses. This advantage can be attributed to the larger receptive fields and enhanced capability of HRNetV2 and UPerNet in segmenting finer features. DeepLabV3+, UPerNet, and HRNetV2 all have a better performance in instrument segmentation and classification across all tasks than Une while UNet has a better performance in segmenting large areas. Notably, achieving spatially consistent instrument classification remains a significant challenge, as different parts of the same instrument may be classified as different types. Additionally, accurately segmenting instruments, particularly when they are inserted into anatomical structures, poses a persistent challenge. Quellec et, al proposed a solution for real-time segmentation and categorization of surgical tasks in ophthalmology using video recordings [6]. The goal was to provide timely information and recommendations to surgeons, particularly those with less experience. The proposed system utilized the content-based video retrieval paradigm and employed analogy reasoning to analyze the ongoing surgery.

The videos were segmented into idle phases, which indicated periods of no clinically-relevant motions, and action phases, representing active surgical tasks. Whenever an idle phase was detected, the preceding action phase was categorized, and the subsequent action phase was predicted using a conditional random field. To evaluate the system's performance, a dataset comprising 186 cataract surgeries performed by ten different surgeons was used. The dataset was manually annotated with up to ten possibly overlapping surgical tasks per surgery. The proposed system achieved an average recognition performance, measured by the Az metric, of 0.832 ±0.070. This experiment demonstrates the effectiveness of the proposed solution in accurately segmenting and categorizing surgical tasks during ophthalmic surgeries. The system's high recognition performance indicates its potential to provide valuable real-time information and recommendations to surgeons, aiding in the improvement of surgical outcomes, especially for less experienced practitioners.

Fox et. al. researchers present a novel application of the Mask R-CNN deep-learning segmentation method for the automatic detection and localization of surgical tools in ophthalmic cataract surgery videos [7]. Their approach involved annotating datasets for multi-class instance segmentation, and they achieved promising results with a mean average precision of 61% for instance segmentation. Furthermore, the approach performed well in bounding box detection and binary segmentation tasks. To enhance the robustness of their model, they conducted an in-depth analysis of the segmentation performance for each instrument class and experimented with various data augmentation techniques. This research significantly contributes to the field of content-based video analysis in ophthalmology, offering valuable resources for training and further investigation of medical research questions in ophthalmic surgery. The successful application of deep learning in surgical tool detection and localization opens up new possibilities for improving surgical workflow and enhancing surgical outcomes in ophthalmic procedures.

3.2. Knee tissue image segmentation

Kessler et al. use a cGAN with U-Net generator and PatchGAN discriminator to segment knee joint tissues in MR images [8]. A cGAN model was successfully trained with a small dataset of only eight subjects, yielding promising results. The segmentation accuracy of bone structures surpassed expectations, with a Dice similarity coefficient (DSC) exceeding 0.95, indicating high precision. Cartilage and muscle tissues also displayed good segmentation performance, achieving a DSC above 0.83. However, the specific area of cruciate ligament segmentations revealed room for improvement, as the DSC was around 0.66. Overall, despite the limited training data, the cGAN model showcased its potential for accurate segmentation, while identifying the need for further enhancements in the cruciate ligament segmentations. It is worth noting that these segmentation results were achieved despite the limited training dataset, which consisted of only eight subjects.

To summarize, this study successfully demonstrated the application of a cGAN with a U-Net generator for knee joint tissue segmentation on MR images. While achieving high segmentation performance for several tissues, further improvements are necessary for cruciate ligament segmentations. Nevertheless, this study lays the groundwork for future technical developments and the utilization of cGANs in segmentation tasks, offering potential benefits for evaluating joint health in osteoarthritis.

Zhou, Z et al. developed new segmentation method for knee joint tissue segmentation which combining CNN, 3D fully connected CRF, and 3D simplex deformable modelling [9]. The evaluation of the method demonstrated excellent performance, with Dice coefficients exceeding 0.9 for four tissue types and mean coefficients between 0.8 and 0.9 for seven other tissue types. Joint effusion and Baker's cyst achieved mean Dice coefficients between 0.7 and 0.8. Most musculoskeletal tissues showed an average symmetric surface distance lower than 1 mm, indicating high accuracy. The method exhibited rapid segmentation, making it promising for musculoskeletal imaging applications. The method achieved accurate segmentation of knee joint tissues, offering potential benefits for clinical use and further research.

Khan, S et al. presented a deep learning framework for achieving precise knee tissue segmentation [10]. The framework merges encoder-decoder segmentation with low-rank tensor-reconstructed segmentation network. To model tissue boundaries and utilize superimposed regions, trimap generation is employed. The method achieves a segmentation dice score of 0.8925 on Osteoarthritis Initiative datasets, with cartilage segments exceeding a dice score of 0.9. The framework's performance highlights significant advancements in knee tissue segmentation, offering promising prospects for improved diagnosis and treatment of musculoskeletal disorders.

4. Experiments and analysis

4.1. Dataset description and preprocessing

Kvasir-SEG is an open-access dataset specifically developed to address the challenging task of pixel-wise image segmentation in medical image analysis. It comprises 1000 gastrointestinal polyp images and their corresponding segmentation masks and all of the data has been verified by a professional gastroenterologist. The Kvasir-SEG dataset has 1000 polyp images and their segmentation masks sourced from the Kvasir Dataset v2.

4.2. Experimental setups and evaluation index selection

The original dataset used in this study was divided into two subsets: testing data with 200 images and training data with 800 images. The testing data served as an independent evaluation set to measure the performance of the U-net models with different types of data augmentation.

To establish a control group, the training data without undergoing any data augmentation was used to train the U-net model and FCN through cross-validation. Both validation pixel accuracy and validation IoU were used as the evaluation metrics. Then we will select the method that has a better performace to test the impact of data enhancement. This control group allowed for a direct evaluation of the impact of data augmentation on the performance of the U-net models. Data augmentation techniques, including cropping, brightness adjustment, perspective transformation, and combinations of these techniques, were exclusively applied to the training data subset. This process involved generating augmented versions of the original training images. The augmented training data was then used for training the U-net models using cross-validation as the experimental groups. Once the U-net model was trained using the training and validation data, it was applied to the test data, and evaluation metrics, such as IoU, were calculated to measure the performance of the models. To evaluate the significance of data augmentation on the accuracy, Two Sample T-tests were employed. These tests were used to compare the performance of models trained with and without data augmentation.

4.3. Experimental results and performance comparison

In the comparison between Unet and FCN for segmentation performance, Unet demonstrated superior results. It achieved a pixel accuracy of 0.913 and an IoU (Intersection over Union) score of 0.5, indicating its excellent segmentation capabilities. On the other hand, FCN attained a pixel accuracy of 0.847 and an IoU of 0.218, indicating comparatively lower performance. These findings suggest that Unet outperforms FCN, particularly in terms of IoU, which serves as a crucial evaluation metric for segmentation tasks. Unet's higher IoU score highlights its ability to accurately delineate object boundaries, making it a more reliable choice for this dataset. As shown in Table 1.

Table 1. Results and Analysis.

	Validation Pixel Accuracy	Validation IoU
U-net	0.913	0.5
FCN	0.847	0.218

In the comparison between Unet and FCN for segmentation performance, Unet demonstrated superior results. It achieved a pixel accuracy of 0.913 and an IoU (Intersection over Union) score of 0.5, indicating its excellent segmentation capabilities. On the other hand, FCN attained a pixel accuracy of 0.847 and an IoU of 0.218, indicating comparatively lower performance. These findings suggest that Unet outperforms FCN, particularly in terms of IoU, which serves as a crucial evaluation metric for segmentation tasks. Unet's higher IoU score highlights its ability to accurately delineate object boundaries, making it a more reliable choice for this dataset. As shown in Table 1.

Table 2. Comparison of Segmentation Accuracy Across Different Data Augmentation Techniques.

	Mean	Sample Size	P_value
Control Group	0.51148	5	
Cropping	0.47562	5	0.6941
Brightness	0.39240	5	0.2622
Perspective	0.41924	5	0.3424
Cropping, Brightness and Perspective	0.47836	5	0.7424

In the data augmentation experiment, we analyzed a sample of size 5 for each group. Notably, the control group, which did not undergo validation, exhibited the highest average IoU score, indicating the best overall performance. Upon calculating the p-values between the experimental groups and the control group, we observed that all p-values exceeded 0.025. This suggests that, at a significance level of 0.05, data augmentation does not have a significant impact on the segmentation accuracy. Consequently, the experiment results imply that the use of data augmentation techniques does not lead to substantial improvements in the accuracy of segmentation tasks based on the evaluation of the average IoU scores. As shown in Table 2.

5. Discussion and challenges

5.1. Limitations and challenges of medical semantic segmentation

Annotated medical images with pixel-level segmentation masks are scarce and time-consuming to create. Besides, different body parts or different diseases require different kinds of training sets. The limited availability of such data hinders the training and evaluation of segmentation models. Segmentation that can be applied to medicine requires extreme precision. However, Medical images often contain intricate structures and fine-grained details that can be challenging to segment accurately. Ambiguous boundaries and subtle variations in textures and appearances make it difficult for segmentation models to precisely classify different regions.

5.2. Possible improvements and future development directions

The transformation function and loss function in the model involve several hyperparameters, such as cropping size, rotating angle, and weight. These parameters were optimized specifically for the dataset at hand but might require manual adjustment when building a new model. Alternatively, embedding these parameters into the neural network could enable automatic optimization. In the future, adversarial learning techniques, like generative adversarial networks (GANs), can be explored to enhance the robustness of the UNet model in medical image segmentation. Through adversarial learning, the U-net models can be trained to generate more accurate and visually realistic segmentations, thereby improving their overall performance.

6. Conclusion

The chosen UNet model, a variant of the convolutional neural network, has demonstrated superior efficiency in medical semantic segmentation as compared to FCN, according to this image dataset. Interestingly, in an experiment involving data augmentation on a dataset consisting of 1000 images, there was no observed increase in accuracy. This could potentially suggest that in the context of a fairly substantial dataset, such as one with 1000 images, data augmentation might not contribute significantly towards improving segmentation accuracy. It is imperative to further investigate and experiment to understand the underlying reasons for this lack of impact, and to explore alternative strategies that may enhance segmentation performance for datasets of this magnitude.

References

[1] Ciresan, D. C., Gambardella, L. M., Giusti, A., & Schmidhuber, J. (2012). Deep neural networks segment neuronal membranes in electron microscopy images. In Proceedings of the NIPS (pp. 2852-2860).

[2] Long, J., Evan, S., & Trevor, D. (2015). Fully convolutional networks for semantic segmentation. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR).

[3] Grammatikopoulou, M., Flouty, E., Kadkhodamohammadi, A., Quellec, G., Chow, A., Nehme, J., Luengo, I., & Stoyanov, D. (2021). CaDIS: Cataract dataset for surgical RGB-image segmentation. Medical Image Analysis, 71, 102053.

[4] Quellec, G., Lamard, M., Cochener, B., & Cazuguel, G. (2014). Real-time segmentation and recognition of surgical tasks in cataract surgery videos. IEEE Transactions on Medical Imaging, 33(12), 2352-2360. doi: 10.1109/TMI.2014.2340473.

[5] Fox, M., Taschwer, M., & Schoeffmann, K. (2020). Pixel-based tool segmentation in cataract surgery videos with Mask R-CNN. In IEEE 33rd International Symposium on Computer-Based Medical Systems (CBMS) (pp. 565-568). Rochester, MN, USA.

[6] Sharma, A., et al. (2019). The optimisation of deep neural networks for segmenting multiple knee joint tissues from MRIs. Journal of Medical Systems, 43(7), 210.

[7] Zhou, Z., Zhao, G., Kijowski, R., & Liu, F. (2018). Deep convolutional neural network for segmentation of knee joint anatomy. Magnetic Resonance in Medicine, 80(6), 2759-2770.

[8] Khan, S., Azam, B., Yao, Y., & Chen, W. (2022). Deep collaborative network with alpha matte for precise knee tissue segmentation from MRI. Computer Methods and Programs in Biomedicine, 222, 106963.

[9] Jha, D., Smedsrud, P. H., Riegler, M. A., Halvorsen, P., Lange, T. d., Johansen, D., & Johansen, H. D. (2020). Kvasir-seg: A segmented polyp dataset. In International Conference on Multimedia Modeling.

[10] Siddique, N., Paheding, S., Elkin, C. P., et al. (2021). U-net and its variants for medical image segmentation: A review of theory and applications. IEEE Access, 9, 82031-82057.