Multimodal Segmentation of Stroke Lesions in Brain MR Images Using U-Net Architecture and Input-Level Fusion

Lillian Ekem, Alana Gonzales, and Hae Sol Moon
24 November 2020

Abstract

Successful delineation of lesions in acute ischemic strokes is essential for increasing the likelihood of positive patient outcomes. Automated stroke lesion segmentation is a useful biomedical tool for predicting patient outcomes and response to treatment. Multimodality imaging techniques are increasingly being introduced into algorithm development. The use of various modalities on the same target provides unique and complementary information. Moreover, multimodal imaging may be beneficial in cases of limited image quality, such as the presence of motion artifacts. This report describes the development and training of a convolutional neural network (CNN) that can be used to estimate the acute location and volume of stroke lesions using multiple MRI modalities. This CNN model was first used to automatically segment lesions on three different MRI modalities: diffusion-weighted imaging (DWI), fluid attenuated inversion recovery (FLAIR) imaging, and apparent fusion diffusion coefficient imaging (ADC). The performance of these single modality networks was compared to that of a multimodal fusion network including all three. The performance of these models was also analyzed after the introduction of Gaussian blur to simulate motion artifacts. Another motivation of applying Gaussian blur was that it might serve as a normalization tool for possible noise discrepancies between images of different modality and possible mis-registration that could have occured during pre-processing. Results showed that the network performed stroke lesion segmentation with the highest accuracy when multimodality imaging was used with no Gaussian blurring. The validation accuracy in this case was 82.5%. When a 3x3 and 7x7 Gaussian blur kernel was applied, the validation accuracies for multimodal imaging were 78.1% and 79.0%, respectively, indicating that the network successfully handled the simulation of motion artifacts.

Introduction

Ischemic stroke, the most common stroke type, is caused by artery blockage that reduces blood flow and oxygen to the brain. The resulting damaged and dead brain cells form a core of brain lesions that are surrounded by an outer region of potentially salvageable tissue. Over time, the stroke lesions can grow to encompass the salvageable tissue, worsening patient prognosis. The ability to quickly and accurately determine lesion volume and location, at both acute and chronic timepoints is essential to making informed health decisions. MR imaging is utilized to visualize, diagnose and select treatment courses. Currently, lesion localization on MR images is done manually by trained professionals. Manual lesion segmentation is not optimal for ischemic stroke lesion diagnosis as it takes time and introduces operator bias. The development of automatic segmentation methods has the potential to accurately and efficiently localize stroke lesions in a reproducible manner. Accuracy achieved by automatic segmentation methods can be further improved with the use of multimodal imaging.

Clinically, active changes in stroke lesion volume are assessed with multimodal MR imaging. In clinical practice, diffusion weighted images (DWI) are used to provide anatomical location and extent of acute lesions. In the weeks following a stroke, fluid attenuated inversion recovery (FLAIR) images are used to delineate chronic lesions. In FLAIR images, however, white matter (WM) lesions can make it difficult to distinguish true stroke lesions. WM lesions, common in healthy elderly adults, present as hyperintense regions in contrast with surrounding white matter. As a result, DWI imaging is commonly required to inform analysis of FLAIR imaging. Automatic segmentation methods that fuse these imaging modalities can reduce this information uncertainty and improve performance. Multimodal automated stroke lesion segmentation is necessary to assist in the prediction of patient outcomes and response to treatment. We have constructed a convolutional neural network (CNN) that was trained to estimate the acute location and volume of lesions on single- and multimodality networks.

Diffusion-weighted MRI (DW-MRI) is a specific method of magnetic resonance imaging that uses the diffusion of water molecules in tissues to generate contrast in MR images [1]. FLAIR MRI, on the other hand, is an MRI sequence that uses inversion recovery pulse sequences to null the signal for certain fluids in the tissue, such as cerebral spinal fluid of the brain [2]. In these modes of MR imaging, as well as ADC imaging, motion artifacts can introduce blur into the image when it is collected by the physical layer of the system [3]. To analyze this aspect of the physical layer, we modeled MRI image formation with motion artifacts by applying Gaussian blur to the images before passing them through the network. After introducing Gaussian blur, we observed whether the model could still successfully segment the stroke lesions.

Related Work

Multimodal imaging is now commonly used in medical imaging to provide more information about the image target than a single image modality could provide [4]. Our multimodal network, for example, uses three modes of MRI imaging to automatically segment stroke lesions: DWI, FLAIR, and ADC. Previous, related work has shown that using deep-learning-based methods for multimodal medical image segmentation tasks can improve the performance of the segmentation when compared to single modality imaging [4]. A review conducted in 2019 assessed several different deep learning network architectures used for multimodal segmentation tasks and specifically focused on their image fusion strategies [4]. The review concluded that image fusion strategy plays an important role in achieving accurate segmentation results and that a layer-level fusion strategy tends to outperform both decision-level and input-level fusion strategies [4]. However, we used input-level fusion because of its simplicity, which is also why input-level fusion is more commonly used in the literature [4]. Additionally, van Garderen et al. proposed a patch-based U-net segmentation method with dropout for multimodal MR images [5].

Methods

Data Set

The dataset used for training and validation consists of images from 24 stroke patients who were scanned with FLAIR, DW and ADC MRI with a diffusion coefficient of 1000 s/mm^2 . A total of 72 images was used. The dataset was obtained from Dr. Wayne Feng in the Duke University School of Medicine, Department of Neurology.

Pre-processing

The images were registered with Advanced Normalization Tools (ANTs) [6]. One FLAIR image was selected as a reference and all other FLAIR images in the dataset were registered to the reference image. The DWI and ADC images were then registered to the registered FLAIR images of the corresponding patient to a size of 512 x 448 in order to achieve spatial correlation throughout the entire dataset.

CNN Network

The CNN network uses U-net architecture as this type of architecture is pervasively used to work with fewer training images while sustaining precise segmentations. In the network, each slice of images for FLAIR, DWI, and ADC were concatenated into three-dimensional data as separate channels. Thus, the size of input after the fusion was 512 x 448 x 3. The input was then convoluted with two-dimensional kernels (7 x 7 x 64) in the first layer with RELU activation and were downsampled with a max-pooling layer. The tensors went through another layer with two convolution kernels followed by max-pooling. Then the tensors went through two additional convolution kernels. The number of filters of each layer was increased by 2. The tensors went through transpose convolution layers, upsampled and finally, went through a sigmoid activation layer. Each down layer was fused with the corresponding up layer with the same number of filters using concatenation. The number of kernels and kernel size were optimized experimentally, and a kernel size of 7 was determined to result in the best performance. The network was trained with NVIDIA Quadro RTX 8000.

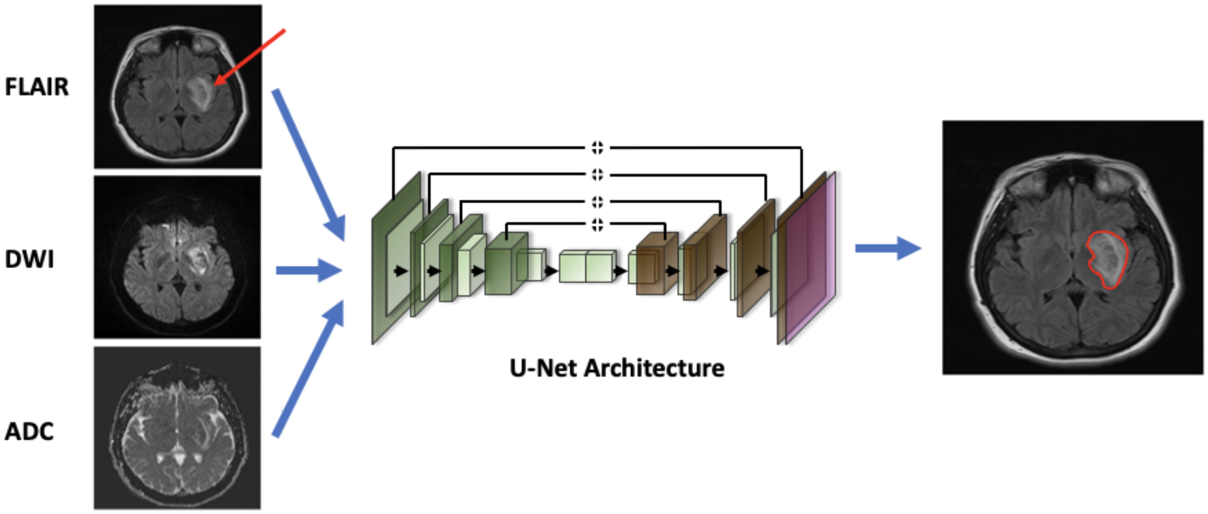

Figure 1. Schematic of the U-net convolutional neural network.

Physical Layer

A physical layer was introduced by the convolution of the images in the dataset with a blur kernel. Motion artifacts were simulated by using a convolutional filter to introduce Gaussian blur in the MR images. Kernel sizes of (3,3) and (7,7) were evaluated for their ability to impact model performance. In addition to simulating motion artifacts, introducing a Gaussian blur could also compensate for possible mis-registration of images that occurs when there are different levels of noise in images of different MR modalities. If this is the case, introducing a Gaussian blur could be advantageous for CNN network training.

Validation

To perform quantitative analysis of the physical layer, we compared the segmentation results with and without the motion artifact simulation. We plotted the training and validation Dice coefficients of the two models and compared their final accuracy and loss values to assess the performance of the model when blur was introduced and when it was not.

Results

Qualitative Evaluation of Multi-Modality Segmentation

When overlaying the ground truth segmentation ROIs with the multimodality segmentation mask, we can see qualitatively that the proposed method delineates the stroke lesions quite similarly to the ground truth ROIs (**Figure 2**). However, when the lesions are very small, the network is less likely to perform accurate segmentation (**Figure 2D**)

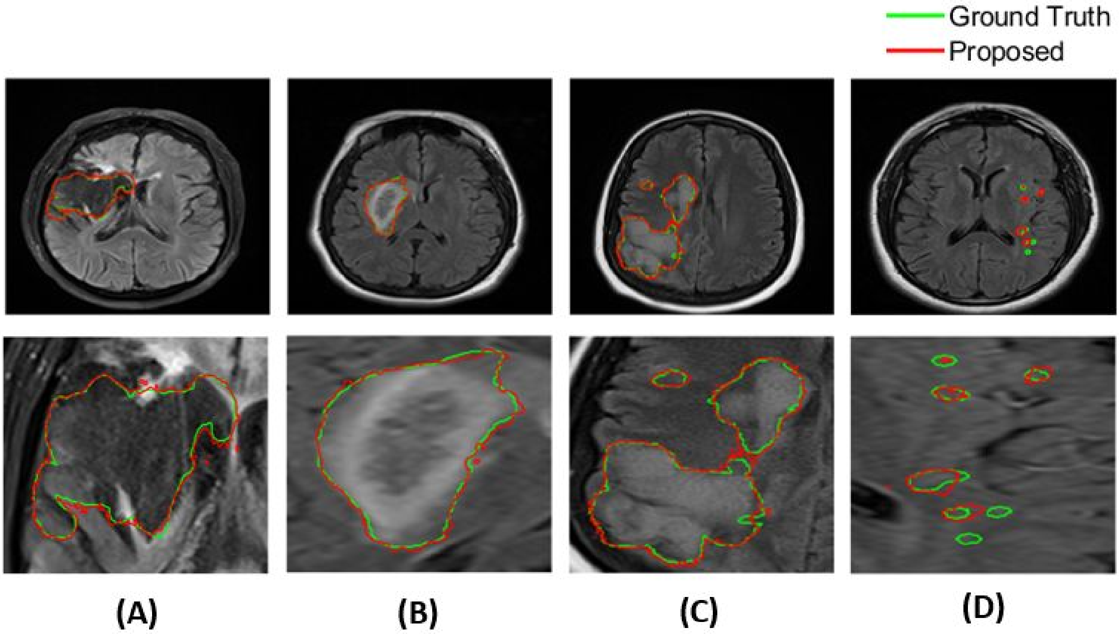

Figure 2. Top row shows comparison between ground truth ROIs and segmentation results from the proposed multi-modality segmentation method. Bottom row shows zoomed in version of the visualization. (A)-(C) Shows reliable segmentation from the proposed method. (D) The network failed to segment some of the small lesions.

Comparing Multi and Single-Modality Segmentation

A CNN was trained with three different modalities: DWI, FLAIR, and ADC. All three modalities were fused using an input-level fusion strategy to develop a multi-image modality network. The binary cross entropy loss and Dice metrics without Gaussian blurring are reported in **Figure 3**. Without the use of a blur kernel, the final validation Dice metrics of networks training on DWI, FLAIR, and ADC images were 79.8%, 68.5%, and 62.6%, respectively. As expected, diffusion weighted imaging provided the best performance. FLAIR images often have white matter hyperintensities that are difficult to differentiate from stroke lesions. When tested, the multimodal network was able to achieve an accuracy of 82.5%. Network performance improved slightly with the inclusion of all three modalities.

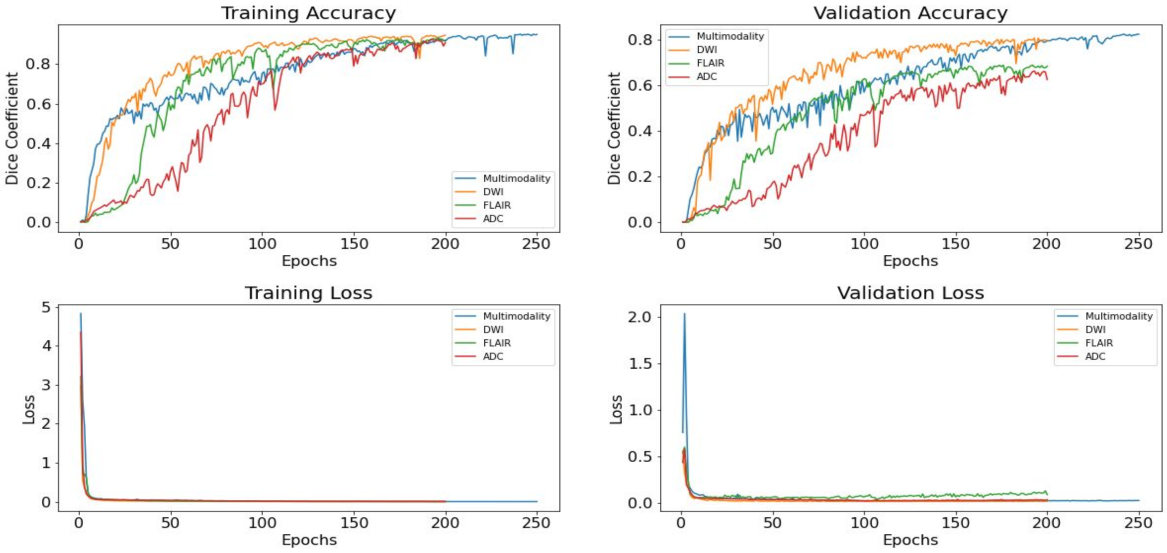

Figure 3. Binary cross-entropy loss and Dice metrics for single-modality and multimodality networks without the use of a Gaussian blur kernel.

Comparing Multi and Single-Modality Segmentation: Physical Layer (Gaussian Blurring)
The effects of motion artifacts were attenuated with the application of Gaussian blurs with kernel sizes of (3,3) and (7,7). The binary cross-entropy loss and Dice metrics with the blurring are reported in **Figure 4** and **Figure 5**. With the use of a 3x3 blur kernel, the final validation Dice metrics of networks training on DWI, FLAIR, and ADC images were 80.3%, 68.3%, and 63.7%. respectively. When tested, the multimodal network was able to achieve a validation accuracy of 78.1% on the 3x3 blurred dataset. With the use of a 7x7 blur kernel, the final validation Dice metrics of networks training on DWI, FLAIR, and ADC images were 81.6%, 68.0%, and 57.2%, respectively. When tested, the multimodal network was able to achieve a validation accuracy of 79.0% on the 7x7 blurred dataset.

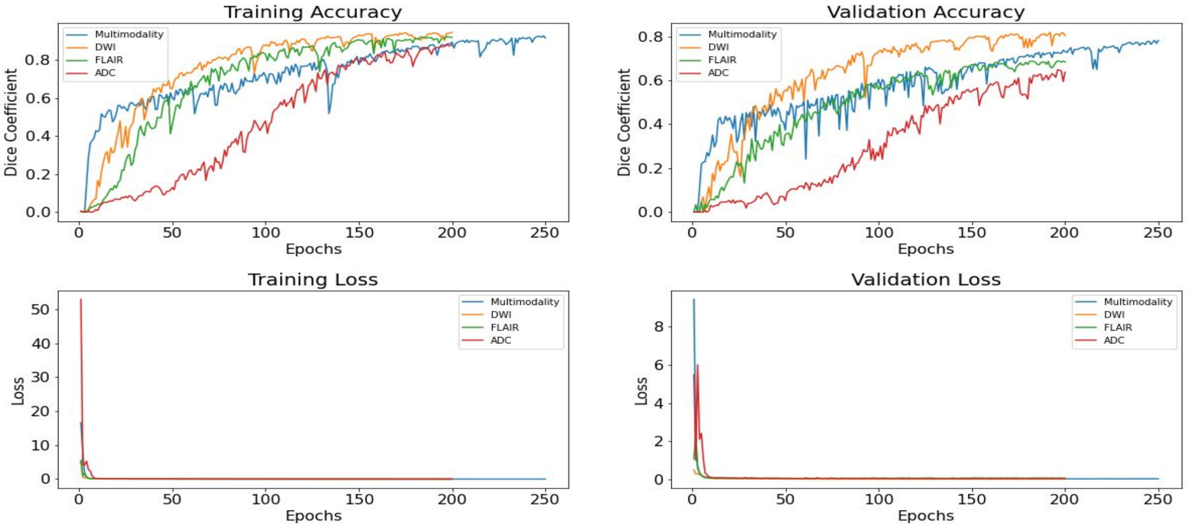

Figure 4. Binary cross-entropy loss and Dice metrics for single-modality and multimodality networks with the use of a 3x3 Gaussian blur kernel.

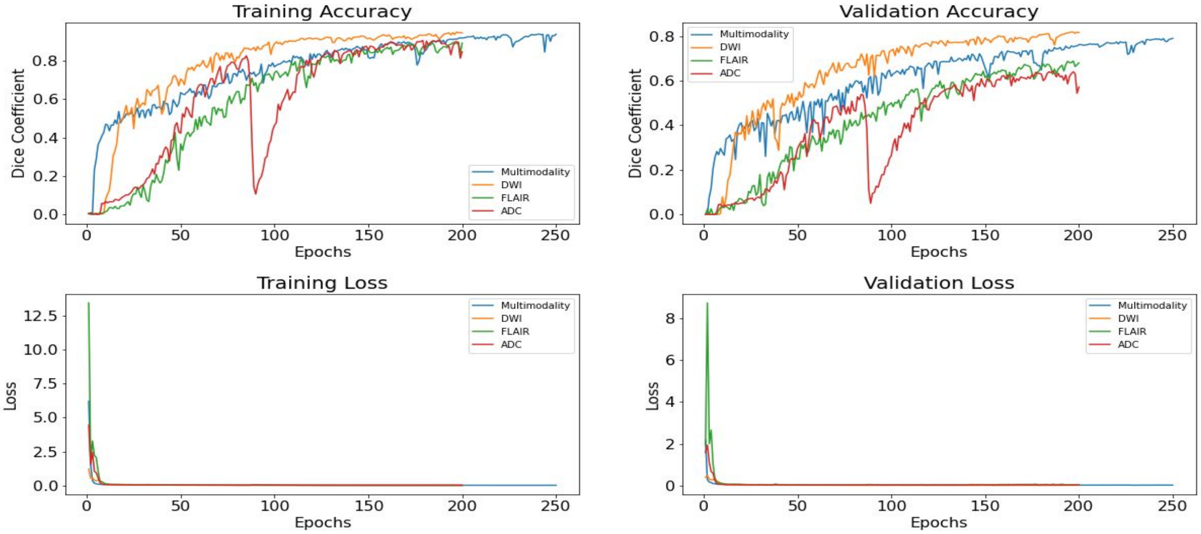

Figure 5. Binary cross-entropy loss and Dice metrics for single-modality and multimodality networks with the use of a 7x7 Gaussian blur kernel.

Discussion

Seventy-two FLAIR, DWI, and ADC images from 24 stroke patients were preprocessed and run through a CNN with U-net architecture (**Figure 1**). Using the dataset, the CNN was trained to estimate the location and volume of stroke lesions by precise segmentation. Multimodality segmentation with no Gaussian blur yielded the best results, as the training data was segmented with a Dice coefficient of 95.0% and a loss of 0.002. Similarly, the validation data was segmented with a Dice coefficient of 82.5% and a loss of 0.022. After optimizing the model, we introduced Gaussian blur into the dataset using a Gaussian convolution kernel to simulate the presence of motion artifacts in the physical layer of the model. After introducing the Gaussian blur, the model was re-trained using the blurred data. Single-modality DWI segmentation yielded the best results with the blurred data. The images blurred with a 3x3 Gaussian kernel were segmented with a validation Dice coefficient of 80.3% and a loss of 0.013, while the images blurred with a 7x7 Gaussian kernel were segmented with a validation Dice coefficient of 81.6% and a loss of 0.015. Based on all of the results, we conclude that multimodality segmentation with no blur yielded the best results, but the network handled the simulation of motion artifacts with reasonable accuracy.

Although multi-modality segmentation was expected to achieve a much higher accuracy than a single-modality segmentation, the Dice coefficient of these two methods were fairly similar. A possible explanation for similar accuracy could be a limited dataset. Only 16 images were trained and tested with 8 images which may not have been enough for multi-modality segmentation to outperform single modality by much.

In the future, segmentation of stroke lesions will be improved with expansion of the data set. Current methods were limited by both hardware and memory. In addition, for the multimodality network, a different fusion approach can be applied such as a modified input-layer fusion network as with HyperDense-Net [7], or a decision-level fusion network as with EMMA [8].

References:

[1] Bihan, Denis Le. “Diffusion MRI: what water tells us about the brain.” EMBO Molecular Medicine 6.5 (2014): 569-573.

[2] Bakshi, R et al. “Fluid-attenuated inversion recovery magnetic resonance imaging detects cortical and juxtacortical multiple sclerosis lesions.” Archives of Neurology 58.5 (2001): 742-748.

[3] Debnath, Arunabha, et al. “Deblurring and Denoising of Magnetic Resonance Images using Blind Deconvolution Method.” International Journal of Computer Applications 81.10 (2013).

[4] Zhou, Tongxue, Su Ruan, and Stéphane Canu. "A review: Deep learning for medical image segmentation using multi-modality fusion." *Array* 3 (2019): 100004.

[5] van Garderen, Karin, Marion Smits, and Stefan Klein. "Training CNNs for Multimodal Glioma Segmentation with Missing MR Modalities." (2018).

[6] Avants, Brian B., Nick Tustison, and Gang Song. "Advanced normalization tools (ANTS)." *Insight j* 2.365 (2009): 1-35.

[7] Dolz, Jose, et al. "HyperDense-Net: a hyper-densely connected CNN for multi-modal image segmentation." IEEE transactions on medical imaging 38.5 (2018): 1116-1126.

[8] Kamnitsas, Konstantinos, et al. "Ensembles of multiple models and architectures for robust brain tumour segmentation." International MICCAI Brainlesion Workshop. Springer, Cham, 2017.