Using temporal GAN to translate the current CTP scan to follow-up MRI, for predicting final AIS lesions

Paper Review

Mahdi Ranjbar, Samuel Pariente

May 31, 2024

1 Introduction

Acute ischemic stroke (AIS) is one of the leading causes of death and long-term disability worldwide [2]. It occurs when a blood clot blocks an artery that supplies blood to the brain tissue [3]. It is critical to know what brain regions are affected by an ischemic stroke, as this enables doctors to make more effective decisions about stroke patient therapy. The treatment of AIS is a very time-sensitive task since failure to restore blood flow can lead to irreversible brain damage [4]. Computed Tomography Perfusion (CTP) is preferred over other imaging modalities in the early stages of ischemic stroke assessment because of its widespread availability, speed of image acquisition [7]. However, the low contrast of baseline CTP images makes it difficult to determine AIS lesions precisely, while follow-up MRI images do.

To deal with this problem, Soltanpour et al. [14] propose a novel method using Temporal Generative Adversarial Network (TGAN). This approach encodes baseline CTP frames with a series of encoders and uses a decoder to forecast high-resolution follow-up MRIs. TGAN includes a discriminator that competes with the generator to identify whether its input MRI is real or fake. Additionally, it incorporates a segmentator capable of identifying AIS lesions in the synthesized MRI images.

2 Literature Review

The current clinical standard for segmenting AIS lesions is done by simple thresholding CTP maps. A developed software processes CTP source data to generate predefined perfusion maps [8]. Such approaches have the drawback of only modeling a single univariate threshold between affected and non-affected tissue. Since AIS lesions are varying widely over time, these simple algorithms are not suitable for accurately predicting the regions at risk [5].

Recently, deep learning has been increasingly used for AIS segmentation with better performance. For example, Soltanpour et al. [12] used a modified version of U-Net [11] called MultiRes U-Net [6] that has been redesigned to be robust in detecting objects of various scales and irregular appearances. Indeed, this can serve as an appropriate solution for addressing the major challenges of AIS lesion segmentation: the wide range of appearances, scales, and outer borders.

Additionally, Wang et al. [15] proposed a novel framework based on synthesized pseudo Diffusion-MRI from perfusion parameter maps to obtain better image quality for more accurate segmentation. This was a solution to low image contrast and resolution of CTP maps.

The present paper draws inspiration from Liu's framework [9] by replacing U-Net with MultiRes U-Net. It uses GANs to generate Diffusion-MRI from raw CTP time series images instead of CTP maps, and segments AIS lesion regions on the generated images (Appx D).

3 Main Contributions

The present paper sets out to combine previous promising ideas into one cohesive approach. The key contributions are as follows:
(1) it uses raw CTP time series images as input, as opposed to relying on pre-defined CTP maps, which have already lost substantial information and are prone to errors.
(2) it introduces a straightforward frame selection method for CTP images, aimed at reducing unnecessary computational overhead.
(3) it takes advantage of the MultiRes U-Net architecture to first translate CTP images into Diffusion-MRI representations within an adversarial learning framework, followed by the segmentation process.

4 Methodology

The proposed AIS prediction method consists of four main sections as follows.

4.1 Pre-processing

To prevent overfitting with the small dataset, training data is augmented for better generalization. The images are rotated to an angle that is randomly chosen from a normal distribution of $\mathcal{N}(0°, 20°)$, and normalized to zero mean and unit standard deviation. Then a Gaussian noise $\mathcal{N}(0, 0.03)$ is added to the images.

4.2 Frame Selection from Raw CTP Images

To optimize computational efficiency, only 10 frames of CTP images are selected rather than the entire sequences. The frame with the highest contrast agent volume along with 3 preceding and 4 succeeding frames are selected, since they contain information related to the AIS lesion [10]. Then to provide some information on the structure of brain tissue without highlighting blood flow or arteries, the first and last frames are added (Appx E).

4.3 Temporal Image Translation

There is a generator based on MultiRes U-Net that consists of 10 encoders to encode 10 selected frames and one decoder to synthesize their corresponding follow-up MRI image. Meanwhile the discriminator competes with the generator to identify whether its input MRI is real or generated (Appx F).

4.4 AIS Lesion Segmentation

Finally, there is a segmentator network again based on MultiRes U-Net to segment AIS lesion from generated MRI image.

5 Results

The study uses a local dataset including 55 patients, with 12 patients excluded due to a lack of valid follow-up images. Using Leave-one-person-out cross-validation (LOOCV), the proposed method achieves an average Dice score of 57.73%, with statistical significance ($p < 0.05$). A comparison with SOTA approaches using MRI-generated images reveals that the proposed model outperforms in terms of Dice score and precision by 57.73% and 53.38%, respectively. However, based on the Jaccard score, the proposed model achieves 52.94%, which is 2.79% lower than the best-performing alternative model. Similarly, concerning recall, the proposed model achieves 51.91%, showing a 5.7% deficit compared to the best-performing alternative model.

6 Strengths Analysis

Using MultiRes U-Net is a thoughtful decision, considering the wide range of sizes of AIS lesions. The paper effectively backs this choice with demonstrated results.

Despite lacking details in the paper, using the 10-frames selection method stands out effectively as a simple yet efficient way to use CTP times series as input data.

Furthermore, given the constraints of a small and fixed local training set, opting for an end-to-end training approach seems to be a good choice.

7 Weaknesses Critique

While the authors introduce a straightforward method for CTP frame selection to reduce complexity, this study lacks a comparative analysis with other frame selection methods, such as the one proposed by Wang et al. [15], which appears to be more sophisticated, necessitating an assessment of its impact (Appx C).

The role of the loss function in this study is crucial, as many previous works have introduced novel frameworks by modifying the loss function. Although the paper is a conference proceeding (Appx A), it lacks technical details about the training procedure.

Furthermore, the study relies on a small local dataset and does not detail how it replicates other previous methods. This raises questions about the fairness of the comparison with other models, especially considering that these models are not open-source and may differ in input data modality.

The lack of technical information and dataset availability makes reproducibility of this study impossible. This hinders the verification of reported results and the opportunity for others to build upon this work.

Moreover, although the input CTP images are 3D, the study does not justify the reason for choosing 2D CNNs instead of 3D CNNs which seems to be more logical choice. One may assume that this selection is influenced by the same author's earlier findings [13] suggesting that multiple 2D U-Nets outperform a 3D U-Net. However, it is crucial to acknowledge that the modality of the input data is not consistent.

Additionally, using LOOCV method for validation would result in training 43 models in their case, which is both computationally intensive and prone to high variation in performance among models. Previous studies have typically used 4 to 10 cross-validation folds instead.

As a few minor critiques, we can mention that the methodology is not well written. Specifically, the main parts are confused with extra detailed information and figures. Besides, there's a discrepancy in the reported Dice coefficient: 57.73% in the result section versus 56.73% in the reported result table (Appx H). Finally, an inconsistency is noted in Figure 6 regarding the generator architecture; despite the claim of using MultiRes U-Net, yet is uses copy and crop—which is used in U-Net—instead of CNN shortcut connections between corresponding layers.

8 Recommendations for Improvement

Given that the framework operates end-to-end, the input data significantly influences the segmentation output. Thus, as discussed in [15], using an auxiliary model to extract high-level features from raw CTP images and incorporating them into the input could be a viable approach. Alternatively, a simpler strategy could involve adding CTP perfusion parameter maps (CBF, CBV, MTT, Tmax) to CTP raw images as input, as each of CTP maps adds valuable information and helps in identifying stroke lesions.

Moreover, by using cross-attention, a multimodal fusion approach, clinical information such as age, sex, etc., can be integrated into the learning process. This can lead to enhance the model's predictive outcomes.

Finally, regarding the paper itself, authors can incorporate some visual graphs to depict a clear comparison of the results.

References

[1] URL: https://spie.org/conferences-and-exhibitions/authors-and-presenters/spie-proceedings-to-journal-pathway#_=_.

[2] Claire L. Allen and Ulvi Bayraktutan. "Risk factors for ischaemic stroke". In: *Int. J. Stroke* 3.2 (2008), pp. 105–116.

[3] Valery L. Feigin et al. "Stroke epidemiology: a review of population-based studies of incidence, prevalence, and case-fatality in the late 20th century". In: *Lancet Neurol.* 2.1 (2003), pp. 43–53.

[4] Mayank Goyal et al. "Endovascular thrombectomy after large-vessel ischaemic stroke: a meta-analysis of individual patient data from five randomised trials". In: *The Lancet* 387 (2016), pp. 1723–1731.

[5] Jan W. Hoving et al. "Volumetric and spatial accuracy of computed tomography perfusion estimated ischemic core volume in patients with acute ischemic stroke". In: *Stroke* 49.10 (2018), pp. 2368–2375.

[6] Nabil Ibtehaz and M Sohel Rahman. "Multiresunet: Rethinking the U-Net architecture for multimodal biomedical image segmentation". In: *Neural Network* 121 (2020), pp. 74–87.

[7] Angelos A. Konstas et al. "Theoretic basis and technical implementations of CT perfusion in acute ischemic stroke, part 1: theoretic basis". In: *Am. J. Neuroradiol.* 30.4 (2009), pp. 662–668.

[8] B. Laughlin et al. "Rapid automated CT perfusion in clinical practice". In: *Practical Neurology* (2019), pp. 41–55.

[9] Pengbo Liu. "Stroke Lesion Segmentation with 2D Novel CNN Pipeline and Novel Loss Function". In: *International MICCAI Brainlesion Workshop*. Cham: Springer, 2018.

[10] K. Murayama et al. "Preliminary study of time maximum intensity projection computed tomography imaging for the detection of early ischemic change in patient with acute ischemic stroke". In: *Medicine (Baltimore)* 97 (2018), e9906.

[11] Olaf Ronneberger, Philipp Fischer, and Thomas Brox. "U-Net: Convolutional Networks for Biomedical Image Segmentation". In: *CoRR* abs/1505.04597 (2015). arXiv: 1505.04597. URL: http://arxiv.org/abs/1505.04597.

[12] Mohsen Soltanpour et al. "Improvement of automatic ischemic stroke lesion segmentation in CT perfusion maps using a learned deep neural network". In: *Computers in Biology and Medicine* 137 (2021), p. 104849.

[13] Mohsen Soltanpour et al. "Ischemic Stroke Lesion Prediction in CT Perfusion Scans Using Multiple Parallel U-Nets Following by a Pixel-Level Classifier". In: *2019 IEEE 19th International Conference on Bioinformatics and Bioengineering (BIBE)*. 2019, pp. 957–963. DOI: 10.1109/BIBE.2019.00179.

[14] Mohsen Soltanpour et al. "Using temporal GAN to translate the current CTP scan to follow-up MRI, for predicting final acute ischemic stroke lesions". In: *Medical Imaging 2023: Computer-Aided Diagnosis*. Ed. by Khan M. Iftekharuddin and Weijie Chen. Vol. 12465. International Society for Optics and Photonics. SPIE, 2023, p. 124653D. DOI: 10.1117/12.2654069. URL: https://doi.org/10.1117/12.2654069.

[15] Guotai Wang et al. "Automatic ischemic stroke lesion segmentation from computed tomography perfusion images by image synthesis and attention-based deep neural networks". In: *Medical Image Analysis* 65 (Oct. 2020), p. 101787. ISSN: 1361-8415. DOI: 10.1016/j.media.2020.101787. URL: http://dx.doi.org/10.1016/j.media.2020.101787.

Appendices

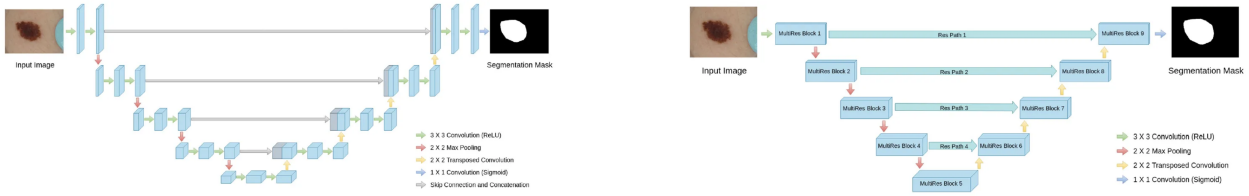
Appendices Outline

- Appendix A, Conference proceedings papers
- Appendix B, UNet vs MultiResUNet Architecture
- Appendix C, Feature Extraction from Raw Spatiotemporal CTP Images
- Appendix D, Proposed Framework
- Appendix E, Proposed Frame Selection
- Appendix F, Proposed Generator Framework
- Appendix G, Proposed Generator Architecture
- Appendix H, Mismatch Between Reported Dice Score

A Conference proceedings papers

Conference proceedings provide a vehicle for rapid reporting of ideas, techniques, and results. It is not uncommon for these reports to be somewhat incomplete and inconclusive. The purposes of proceedings papers range from snapshots of recent or continuing work to the reporting of a completed work or project. [1]

B UNet vs MultiResUNet Architecture



(a) Original UNet

(b) MultiResUNet

Figure 1: MultiResUNet is a modified version of original UNet such that for the sequence of two convolutional layers at each level, they are replaced by the MultiRes block.

C Feature Extraction from Raw Spatiotemporal CTP Images

In the study by Wang et al. [15], a Convolutional Neural Network (CNN) is used to extract important features from raw spatiotemporal CTP images. Identifying the beginning and end points of perfusion is crucial for this feature extraction process. The authors devise a method to detect these points as follows:

They first define a curve representing the accumulated intensity over time, denoted as $q(t) = \sum_{x,y,z} I(x, y, z, t)$. Here, T_s and T_e represent the estimated start and end time points of perfusion, respectively. These points are determined based on the following rules:

$$T_s = \min \left\{ t \mid 0 \leq t < T - K, \sum_{k=0}^{K-1} \mathcal{H}\left(q'(t+k)\right) = K \right\}$$

$$T_e = \max \left\{ t \mid K \leq t < T, \sum_{k=0}^{K-1} \mathcal{H}\left(q'(t-k)\right) = 0 \right\}$$

Here, $\mathcal{H}(\cdot)$ represents the Heaviside function, which yields 0 for negative inputs and 1 for positive inputs. $q'(t)$ denotes the first derivative of $q(t)$, and K is a positive integer empirically set to 5. Therefore, T_s is defined as the earliest time point where the first derivative of $q(t)$ remains positive for the subsequent K consecutive time points, while T_e is defined as the latest time point where the first derivative of $q(t)$ remains negative for the preceding K consecutive time points. Figure 2 illustrates the curve of $q(t)$ with T_s and T_e in two different scenarios.

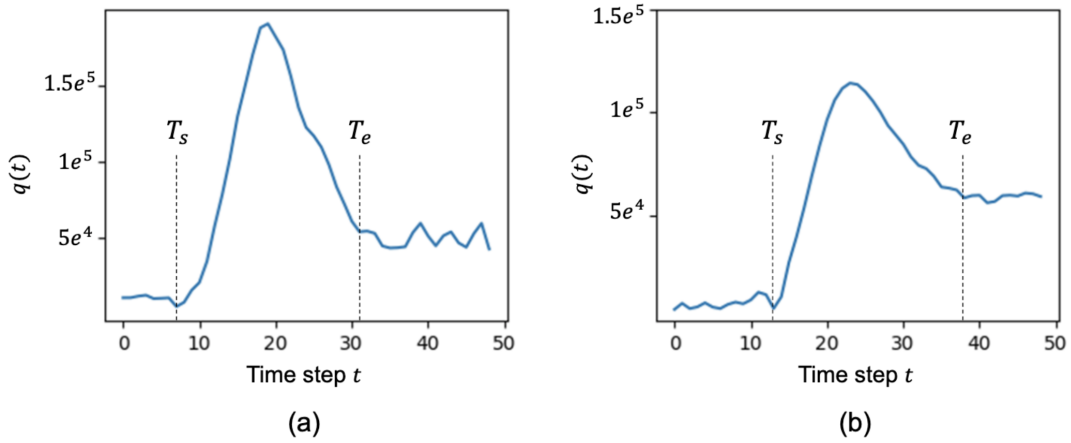


Figure 2: Illustration of start time (Ts) and end time (Te) detection of the perfusion stage.

D Proposed Framework

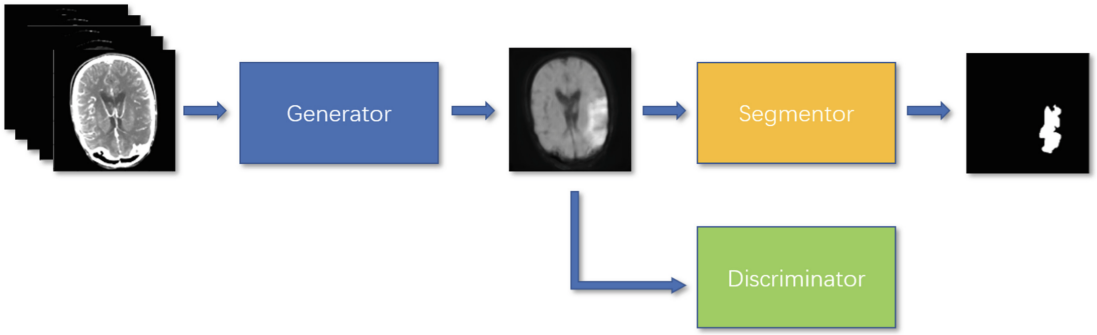


Figure 3: Generator is used to synthesize the MRI images from the CT data. Discrimintator is used to perform the classification of the generated MRI image and the true MRI image. Segmentator is finally used to segment the lesion of brain on the generated DWI image.

E Proposed Frame Selection

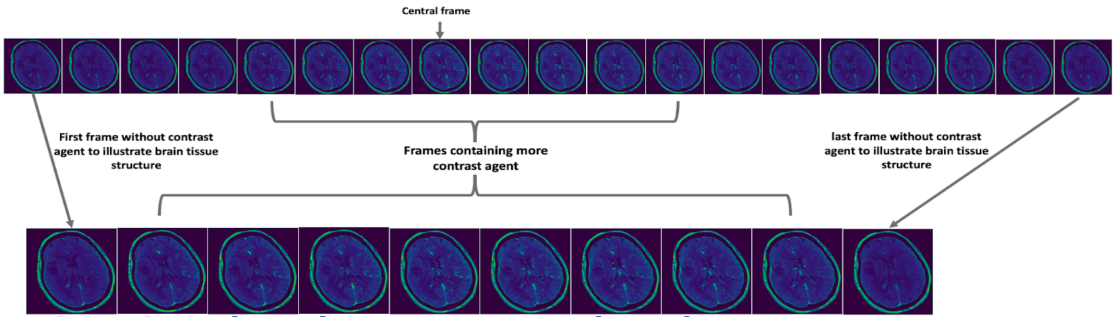


Figure 4: 10 frames selection for the temporal generator.

F Proposed Generator Framework

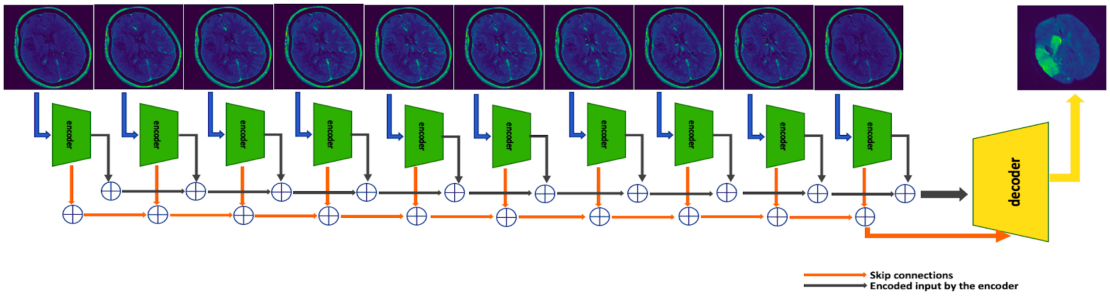


Figure 5: Temporal generator architecture.

G Proposed Generator Architecture

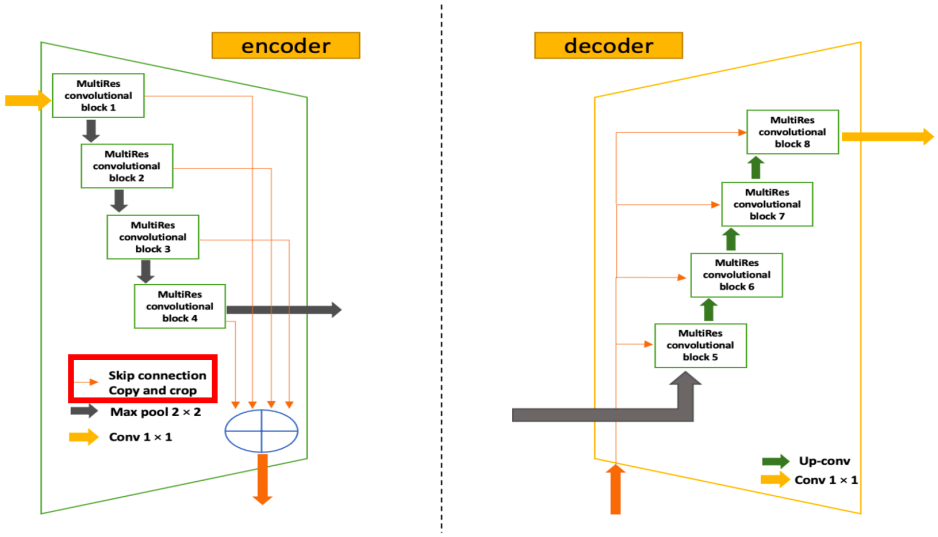


Figure 6: MultiRes encoder and decoder

H Mismatch Between Reported Dice Score

Results:

To evaluate our proposed AIS lesion predictor we have used our local dataset including 55 patients. We had to exclude 12 patients because they did not have follow-up scans or had CT scans as the follow-up image. In our experiments, an average dice coefficient of 57.73%, with a significant $p<0.05$, is achieved with Leave-one-person-out cross-validation (LOOCV). Figure 4 shows some randomly selected images. The first row shows the CTP images, and the second and third rows show the generated and actual MRI images, respectively. On the generated MRI images, the red line indicates the predicted lesion by our model, and on the actual MRI, the red line indicates the actual lesions. Moreover, table 1 shows a comparison between our proposed method and some previous methods for predicting AIS lesions. They were selected for comparison due to their similarity in approach because they also used generated MRI rather than baseline CTP to predict AIS lesions. According to Table 1, our proposed TGAN can improve the accuracy of AIS lesion prediction by synthesizing MRI data more efficiently.

Table 1. Evaluation results were obtained by using our proposed method and some previously proposed ones.

Model	Dice Score (%)	Jaccard (%)	Precision (%)	Recall (%)
C²MA-Net [6]	53.35 ±19	54.81 ±16	49.18 ±13	**57.61 ± 15**
Pengbo Liu [7]	49.14 ±23	50.87 ±12	48.19 ±21	51.88 ±20
Mobarakol Islam et al. [8]	54.05 ± 22	**55.73 ± 17**	50.73 ± 18	54.42 ± 21
TGAN (original U-Net)	51.35 ±21	50.59 ±21	47.14 ±13	53.16 ±23
TGAN (MultiRes U-Net)	**56.73 ± 17**	52.94± 16	**53.38 ± 20**	51.91 ±24

Figure 7: The reported dice score in the result section does not correspond to the one in reported result table as shown with surrounded red rectangle.