



Vol 17, N° 2

<https://revistas.usb.edu.co/index.php/IJPR>

ISSN 2011-2084

E-ISSN 2011-7922

 OPEN ACCESS

Manuscript received: 30-10-2023

Revised: 19-03-2024

Accepted: 05-09-2024

***Corresponding author:**

Fabio Martínez

Email: famarcar@saber.uis.edu.co

Copyright: ©2024. International Journal of Psychological Research provides open access to all its contents under the terms of the license [creative commons Attribution-NonCommercial-NoDerivatives 4.0 International \(CC BY-NC-ND 4.0\)](https://creativecommons.org/licenses/by-nc-nd/4.0/)

Declaration of data availability: All relevant data are within the article, as well as the information support files.

Conflict of interests: The authors have declared that there is no conflict of interest.

How to Cite:

Florez, S., Gómez, S., García, J., & Martínez, F. (2024). A Deep Cascade Architecture for Stroke Lesion Segmentation and Synthetic Parametric Map Generation over CT Studies. *International Journal of Psychological Research*, 17(2), 47-53. <https://doi.org/10.21500/20112084.7013>



A Deep Cascade Architecture for Stroke Lesion Segmentation and Synthetic Parametric Map Generation over CT Studies

Arquitectura profunda en cascada para la segmentación de lesiones de accidente cerebrovascular y la generación de mapas paramétricos sintéticos sobre estudios de TC

Sebastian Florez¹ , Santiago Gómez¹ , Julian Garcia¹ ,
Fabio Martínez^{1,*} .

¹Biomedical Imaging, Vision and Learning Laboratory (BIVL^{2ab}), Universidad Industrial de Santander (UIS).

Abstract.

Stroke, the second leading cause of death globally, necessitates prompt diagnosis for effective prognosis. CT imaging has limitations, especially in identifying acute lesions. This work introduces a novel deep representation that uses multimodal inputs from CT studies and perfusion parametric maps, to retrieve stroke lesions. The architecture follows an autoencoder representation that forces attention on the geometry of stroke through additive cross-attention modules. Besides, a cascade train is herein proposed to generate synthetic perfusion maps that complement multimodal inputs, refining stroke lesion segmentation at each stage of processing and supporting the observational expert analysis. The proposed approach was validated on the ISLES 2018 dataset with 92 studies; the method outperforms classical techniques with a Dice score of .66 and a precision of .67.

Resumen.

El accidente cerebrovascular (ACV), segunda causa de muerte en el mundo, requiere un diagnóstico temprano para un pronóstico favorable. Las imágenes de TC tienen limitaciones, especialmente en la identificación de lesiones agudas. Este trabajo introduce una novedosa representación profunda que utiliza datos multimodales TC y mapas paramétricos de perfusión para segmentar lesiones de ACV. La arquitectura sigue una representación autocodificadora que fuerza la atención sobre la geometría del ACV a través de módulos aditivos de atención cruzada. Además, se propone un entrenamiento en cascada para generar mapas de perfusión sintéticos que complementen las entradas multimodales, refinando la segmentación de las lesiones en cada etapa del procesamiento y apoyando el análisis observacional del experto. El enfoque propuesto fue validado en el conjunto de datos ISLES 2018 con 92 estudios; el método supera a las técnicas clásicas con una puntuación Dice de .66 y una precisión de .67.

Keywords.

Computed Tomography, Stroke Segmentation, Parametric Map, Attention.

Palabras Clave.

Tomografía computarizada, segmentación de accidentes cerebrovasculares, mapa paramétrico, atención.

1. Introduction

Stroke is the disease with the second highest mortality rate in the world and the first cause of disability in developed countries (Roth et al., 2020). Stroke lesions are related to the interruption of blood flow, causing irreversible damage by tissue death on the brain cells (Rekik et al., 2012). The early localization, measurement, and characterization of stroke lesions are critical to determine if a patient is a candidate for a reperfusion therapy. Computerized tomography (CT) is the most common study for stroke triage and diagnosis, due to the high availability in clinical centers and the relatively fast protocols of attention. Nonetheless, CT studies have low contrast, and the use in early stages is limited to triage analysis to distinguish between ischemic and hemorrhagic, or other brain affectations (Von Kummer et al., 1994). To complement such analysis, it is demanding to use perfusion protocols, obtaining complementary maps (CTP, CT perfusion) that describe the behavior of a fluid around the lesion. From this bank of multimodal CT and CTP studies, the expert should localize, and delineate lesions to characterize lesions and determine an attention protocol. This procedure is a tedious task that takes approximately 15 minutes per case (Martel et al., 1999), and a reported low concordance between experts (Neumann et al., 2009).

Computational strategies have recently been proposed to deal with automatic stroke segmentation from CT and CTP studies. Recently, these strategies have been based on deep autoencoders that learn to discriminate between healthy and hypoperfused tissue, guided by expert annotations, allowing to support the segmentation (Liu et al., 2020; Wang et al., 2020). For instance, Tureckova and Rodríguez-Sánchez (2018) built a Unet with dilated convolutions to enlarge the receptive field, including a major computation context, and allowing the characterization of lesions with different sizes. Likewise, Dolz et al. (2018) built a multi-input autoencoder to model the intrinsic features from each modality with a dedicated convolutional branch. Moreover, Clerigues et al. (2019) separated the images into small patches, to build a balanced dataset with the same number of lesions and healthy patches. Subsequently, the balanced patches dataset was used to train an asymmetric autoencoder. Alternative strategies have used generative frameworks to segment ischemic lesions from synthetic diffusion-weighted MRI modalities, which have greater sensitivity in the acute setting. For instance, Liu (2019) built a GAN strategy to generate synthetic diffusion-weighted imaging (DWI) sequences from CTP maps which were used afterward to estimate the lesions using a Unet.

This work introduces a novel deep cascade strategy that includes multimodal inputs (CT and CTP maps) to segment ischemic stroke lesions. For doing so, a deep autoencoder is herein adjusted embedded multimodal

information and decoded lesion annotations, following a supervised scheme. From the architecture, multiple additive-cross attention mechanisms are included between the encoder and decoder to force representation to stand out hypoperfused marks and regions associated with stroke lesions. From the learning scheme, this work introduced a cascade scheme to refine lesion estimations in two stages. In the first stage, the architecture is optimized to generate synthetic CTP maps that complement information and help with stroke localization. Afterward, in the second stage, the synthetic input together with CT and CTP studies estimates a refined Stroke lesion.

2. Proposed Approach

We propose a cross-attentional autoencoder to segment ischemic stroke lesions from CT and CTP parametric maps. The proposed architecture is trained under a cascade framework that monitors the contribution of all decoder levels, and a weighted loss function to deal with the natural class imbalance of stroke segmentation. In a first cascade stage, synthetic maps are generated to guide and refine stroke segmentation in further stages. The general description of our proposed framework is described in Figure 1.

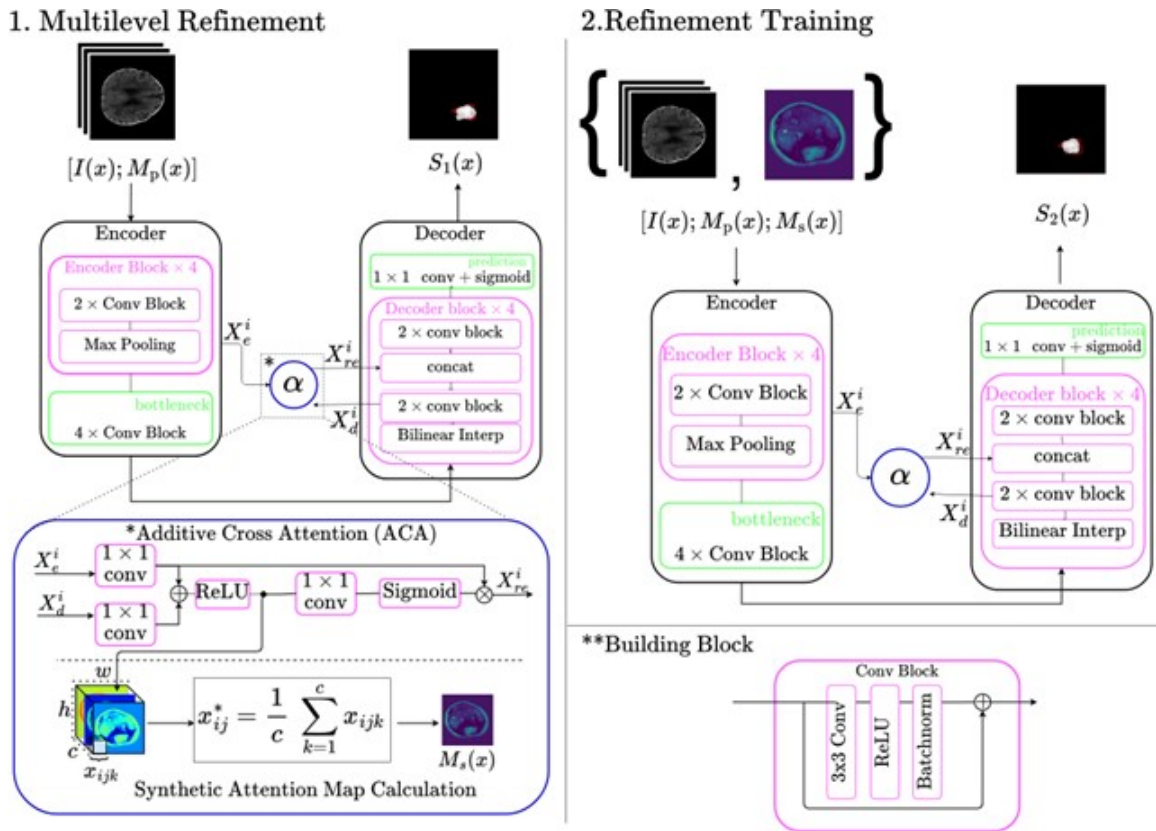
2.1 Additive Cross-Attentional Autoencoder

Nowadays, autoencoders are the most used architecture to deal with stroke lesion segmentation. Specifically, Unet (Ronneberger et al., 2015) autoencoder includes skip connections to integrate deep features from the encoder to the decoder, in the same level of processing. In such cases, the architecture has been proven to be ideal in problems that require the recovery of structural and global shape information. Nevertheless, these connections could integrate uncorrelated features into the decoder representation, harming the decoder representation and slowing the learning process. Hence, we implement a special autoencoder that includes cross-attention mechanisms (ATT) as skip connections, preserving highly correlated features, and following an additive alignment (Gómez et al., 2023).

In general, the cross-attentional architecture is a convolutional neural network (CNN) that receives a CT image ($I(x)$) and CTP parametric maps ($M_p(x)$) in a simple concatenation scheme ($\{I(x); M_p(x)\}$). The encoder maps the sequences into a low-dimensional latent space representation ($h \rightarrow S(x)$) that retains the most important visual stroke patterns. Next, the latent vector is decoded into a probability map ($S(x)$) that highlights potential hypoperfused regions. In every level of the decoder, the convolutional activations (X_d^l) are complemented by the responses from the encoder (X_e^l) on the same processing level (l) with additive cross-attention modules. To compute the similarity between encoder and decoder features, both features are linearly pro-

Figure 1

Deep Cascade Cross-attentional Architecture to Segment Acute Ischemic Stroke Lesions from CTP Sequences



jected, summed, and activated with a ReLU (σ_1) function to consider only positive similarities. Following this, a single attention map is calculated using a 1×1 convolutional layer and a sigmoid activation (σ_2) function. Thereafter, refined characteristics are computed from the encoder (X_{re}) with a Hadamard product (\odot) between the features of the encoder and the attention map. The refinement of encoder features is described by: $X_{re}^l = X_e^l \odot \sigma_2(W_{re}^T \sigma_1(W_e^T X_e^l + W_d^T X_d^l))$.

2.2 Multilevel Refinement

To deal with the natural class imbalance of stroke segmentation and promote the learning of deep features in the attentional autoencoder, the cross-attentional architecture is adjusted with multilevel refinement (MR) in the decoder. The multilevel refinement consists of monitoring the contribution of the decoder blocks with a weighted loss between a low-scale prediction and the ground truth. To measure this contribution, deep supervision layers were placed in the decoder blocks from all levels. This deep supervision consists of applying a bilinear interpolation on output features, followed by 1×1 conv and sigmoid activation to estimate a low-scale version of the stroke lesion. Deep supervision of the representation allows to minimize the propagation of errors and avoid the vanishing of the gradient (Li et

al., 2022). The loss function is a binary cross-entropy, weighted by class weight maps (C) that give more importance to lesion pixels. These weight class maps are built from the manual delineations of the ischemic lesions as a reference and assign a specific weight to the pixels of the same class. Consequently the estimations are used alongside their corresponding weight maps to calculate the final loss as: $L_{total} = W^l C Y \log(\hat{Y}^l)$ where $\{Y, \hat{Y}^l, C\} \in \mathbb{R}^{O \times H \times W}$, Y is the expert’s manual delineation, \hat{Y}^l is the model’s prediction at each level l , and W^l is a positive number evidencing the importance of the l th layer segmentation.

2.3 Cascade Training

Typically, stroke is assessed from multiple CT sequences. In particular, perfusion studies record the blood flow through the brain parenchyma, potentially indicating brain damage due to stroke. From these CTP studies are computed parametric maps that summarize kinematic information, allowing to approximate indexes related to the damaged tissue. To exploit the non-local information from these CTP sequences, we conducted a cascade training to i) learn attention maps that highlight the ischemic stroke lesion visual patterns and ii) segment the acute ischemic stroke lesions. In the first

Table 1

Dice Score for Different Configurations of the Proposed Method with all the Present Modalities on the ISLES2018 Dataset

Configuration		Modalities				
MR	ATT	NCCT	CBV	CBF	TMax	MTT
×	×	.20 ± .26	.51 ± .30	.57 ± .26	.50 ± .25	.56 ± .30
✓	×	.26 ± .24	.51 ± .27	.58 ± .22	.55 ± .26	.57 ± .26
✓	✓	.24 ± .19	.54 ± .28	.57 ± .23	.55 ± .28	.55 ± .21

Note. MR is for multilevel refinement and ATT is for attention mechanisms.

phase, the proposed attentional autoencoder is trained to segment the ischemic stroke lesions from the CTP inputs. During training the attention maps learn to highlight regions that resemble ischemic patterns. Afterwards, from the additive-cross attention module in the first level, synthetic maps $Ms(x)$ are obtained by averaging the positive similarities across channels. Finally, the proposed architecture is adjusted from the concatenation of the CTP inputs and the synthetic attention map ($[I(x); Mp(x); Ms(x)]$). These maps emphasize the most important parts of the image and a possible indication of the shape of the lesion, guiding the network when segmenting if they are added from the start.

3. Experimental Setup

The proposed methodology was validated on the *Ischemic Stroke Lesion Segmentation 2018* (ISLES2018) public dataset. This dataset contains 156 studies of patients diagnosed with acute ischemic stroke, and it is partitioned in 94 training studies and 62 testing studies. Each study contains a plain CT image, the raw 4D perfusion study, and four parametric maps, namely cerebral blood flow (CBF), cerebral blood volume (CBV), mean transit time (MTT), and time to maximum (TMax). The segmentation of the ischemic lesions was performed manually by expert radiologists over DWI sequences, acquired after the CTP. The delineations are only available for the training studies.

The cross-attentional autoencoder is made of convolutional blocks with two 2D convolutional layers that include batch normalization and a ReLU activation. In total, the encoder and decoder both have six levels of processing with 32, 64, 128, 256, 512, and 1024 filters, respectively. In each encoder level, max-pooling layers, with a reduction factor of 2 were included after every convolutional block to reduce the spatial dimensionality. Bilinear interpolations were added pre-convolutional block to expand the spatial dimensionality by a factor of 2 in the decoder.

For training, 19 studies were randomly excluded from the training set to use as validation. All images were resized to 224×224 . The training was done for 600 epochs with a binary cross-entropy loss, and an AdamW optimizer with an initial learning rate of $3e-2$, and weight

decay of $1e-5$. Furthermore, data augmentation such as random brightness and contrast, flips, rotations, random elastic transformations, and random grid and optical distortions were applied to the slices. The weights for the segmentation outputs were set to $W = \{.03, .045, .05, .125, .25, .5\}$ and the weights used to construct the class weight maps were .7 for the lesion and .3 for the background.

4. Evaluation and Results

Firstly, we carry out an ablation study to measure the contribution of each modality included in this study, and the contribution of each component of the proposed approach. Hence, in the first experiment, we measured the performance of unimodal models to exploit the information available on every modality of the dataset. Table 1 shows the dice score for all the selected configurations on every modality on the dataset.

The best configuration was achieved from the architecture with MR (.58), and using CBF maps. Is noteworthy an improvement on the dice score of .026 and .022 for the configurations MR and MR + ATT compared with the standard autoencoder. There is also a considerable difference in the models that use CT together with the parametric maps. This fact is due to the low CT sensitivity. As the image shows, the best overlap with the expert's segmentation is obtained when using the parametric maps in comparison to the CT one which failed in both localization and characterization of the lesion. In this case, multimodal information provides better approximations when generating the segmentations.

In a second experiment, we exploit the multimodal nature of the adjusted autoencoder. For doing so, the architecture was adjusted using multi-context information, i.e., following an early fusion of the CT and parametric maps at the input of the autoencoder. Table 2 shows the obtained results for every chosen parametric map group and configuration of the model. In such cases, there exists an exhaustive validation regarding the components of the architecture, as well as, the combination of potential maps that observe lesion findings.

As observed, the proposed cascade methodology has the capability to fuse and take advantage of lesion findings observed from different studies. In fact, the models that include multicontext information are superior

Table 2

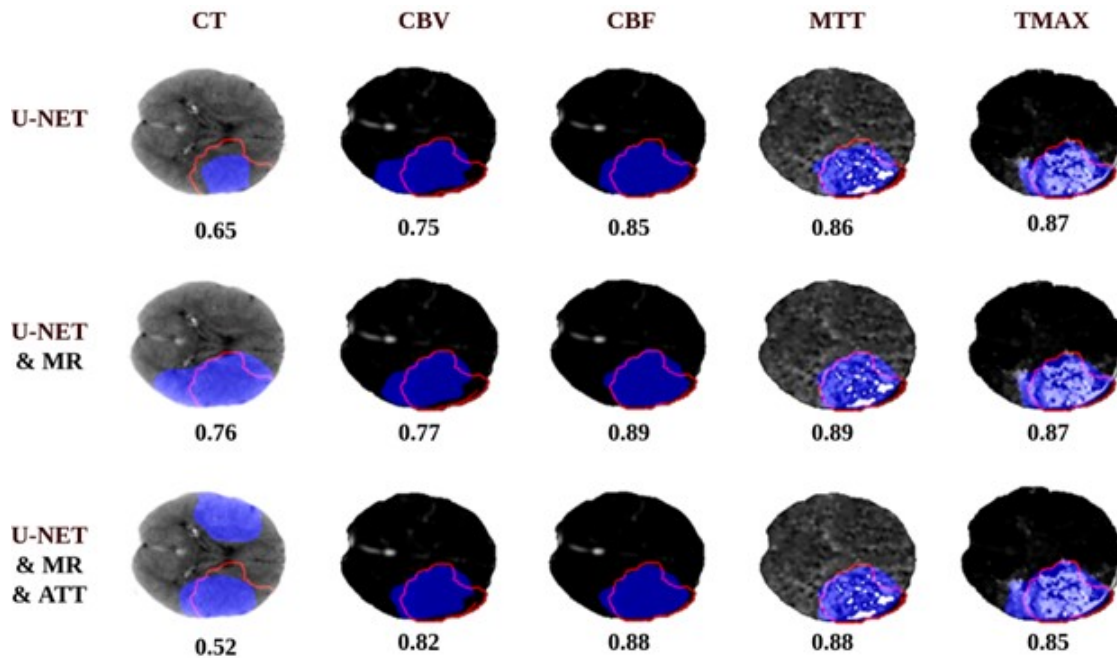
Dice Score for Different Configurations of the Proposed Method with Different Combinations of Modalities from the ISLES2018 Dataset

Configuration		Modalities				
MR	ATT	TMax+CVB	TMax+CBF	CVB+CBF	TMax+CVB+CBF	All
×	×	.58 ± .25	.58 ± .24	.53 ± .26	.60 ± .24	.62 ± .23
✓	×	.61 ± .23	.62 ± .20	.54 ± .27	.63 ± .21	.63 ± .22
✓	✓	.60 ± .23	.62 ± .21	.55 ± .25	.61 ± .25	.63 ± .16

Note. MR is for multilevel refinement and ATT is for attention mechanisms.

Figure 2

Examples of the Realized Segmentation for Every Modality of the First Table



Note. The red line shows the shape of the lesion indicated by the expert and the blue one the output from the network.

to unimodal models. The best performance (.63) is obtained by two configurations, the one that uses all parametric maps with the original CT and the one that uses a combination of Tmax, CBV and CBF. Nevertheless, the mean dice scores for the experiments with all inputs are more consistent than the combination that only uses three of them. An example of the output segmentations obtained is presented on Figure 3. As this image illustrates, the MR and attention mechanisms provide a greater degree of local delineation, preserving particular structures of the lesion and obtaining a great overlap in the masks.

Finally, Figure 4 shows the achieved results with and without cascaded refinement stages. Also, it is observed in the last row the output synthetic maps, generated by the autoencoder in the first phase, which is then used to complement inputs in a refinement stage. As observed, the use of a cascade strategy allows to deal with the chal-

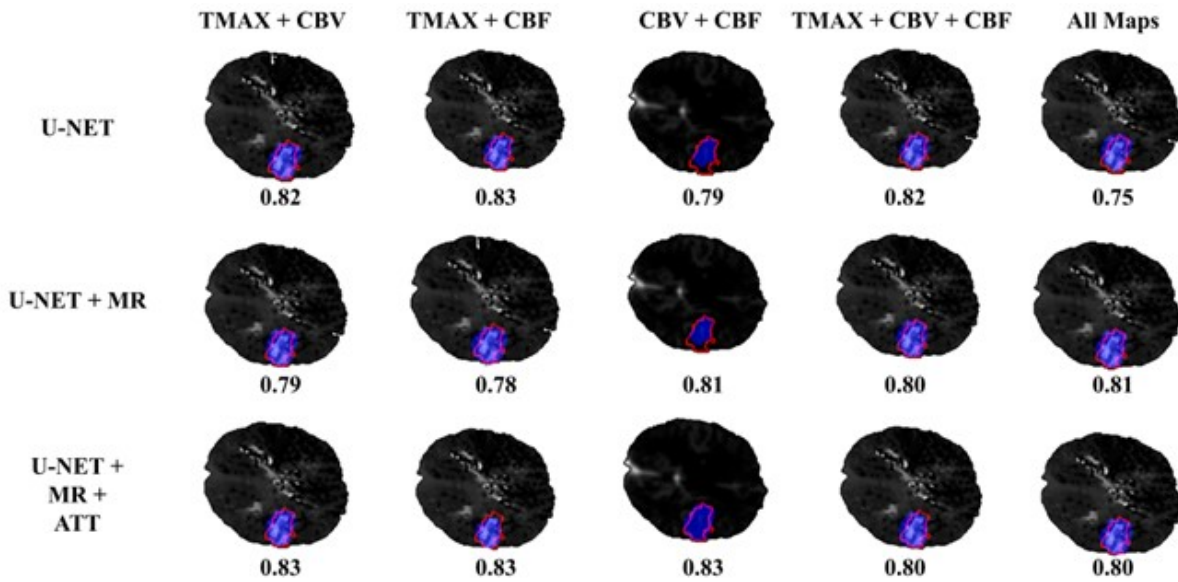
lenging problem of stroke lesion segmentation, allowing among others to obtain a better overlap score, but at least, allowing to localize tiny lesions, which are indicative of early stages of the disease. Additionally, these synthetic maps can potentially be used in clinical practice to support observational analysis, for instance, in centers that only have available CT studies.

5. Conclusions and Perspectives

This work introduced a cascade training scheme that adjusts a deep attention autoencoder to segment stroke lesions. The introduced methodology has the ability to manage multi-context information, allowing an early fusion of parametric maps (CTP) and CT studies. The introduced cascade training strategy in the first stage generates a synthetic map that complements studies to observe localizations, while in the second stage achieves

Figure 3

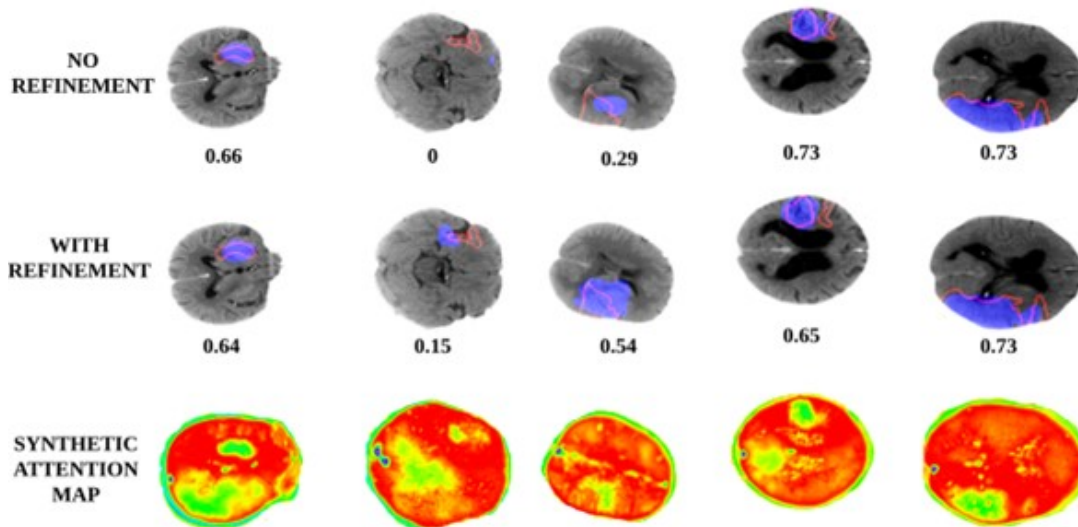
Examples of the Realized Segmentation for Every Modality of the First Table



Note. The red line shows the shape of the lesion indicated by the expert and the blue one the output from the network.

Figure 4

Examples of the Realized Segmentation for Every Modality of the First Table



Note. The red line shows the shape of the lesion indicated by the expert and the blue one the output from the network.

a refinement of the lesion. The reported results evidence a better support of cascade strategy, that together with the multilevel refinement and the attention mechanisms, achieve the localization and delineation of stroke lesions. The proposed approach was validated on a public dataset outperforming results, but also showing the challenges

on the stroke segmentation from raw CT studies. Future works include alternative mechanisms to learn from multiple modalities. Also, the validation on extra datasets that include additional studies, but also multiple expert radiologists to avoid the bias from an expert.

6. Acknowledgment

To the Ministry of Science, Technology and Innovation of Colombia by the project: *Caracterización de movimientos anormales del parkinson desde patrones oculomotores, de marcha y enfoques multimodales basados en visión computacional*, with code 92694.

References

- Clerigues, A., Valverde, S., Bernal, J., Freixenet, J., Oliver, A., & Lladó, X. (2019). Acute ischemic stroke lesion core segmentation in CT perfusion images using fully convolutional neural networks. *Computers in Biology and Medicine*, *115*, 103487. <https://doi.org/10.1016/j.compbiomed.2019.103487>
- Dolz, J., Ayed, I., & Desrosiers, C. (2018). Dense multi-path u-net for ischemic stroke lesion segmentation in multiple image modalities. *International MICCAI Brainlesion Workshop*, 271–282. https://doi.org/10.1007/978-3-030-11723-8_27
- Gómez, S., Mantilla, D., Rangel, E., Ortiz, A., Vera, D., D., & Martínez, F. (2023). A deep supervised cross-attention strategy for ischemic stroke segmentation in MRI studies. *Biomedical Physics & Engineering Express*, *9*(3), 035026. <https://doi.org/10.1088/2057-1976/acc853>
- Li, R., Wang, X., Huang, G., Yang, W., Zhang, K., Gu, X., Tran, S. N., Garg, S., Alty, J., & Bai, Q. (2022). A comprehensive review on deep supervision: Theories and applications. *arXiv Preprint arXiv:2207.02376*. <https://doi.org/10.48550/arxiv.2207.02376>
- Liu, L., Kurgan, L., Wu, F. X., & Wang, J. (2020). Attention convolutional neural network for accurate segmentation and quantification of lesions in ischemic stroke disease. *Medical Image Analysis*, *65*. <https://doi.org/10.1016/j.media.2020.101791>
- Liu, P. (2019). Stroke lesion segmentation with 2D novel CNN pipeline and novel loss function. *Brainlesion: Glioma, Multiple Sclerosis, Stroke and Traumatic Brain Injuries: 4th International Workshop, BrainLes 2018, Held in Conjunction with MICCAI 2018, Granada, Spain, September 16, 2018, Revised Selected Papers, Part I 4*, 253–262.
- Martel, A. L., Allder, S. J., Delay, G. S., Morgan, P. S., & Moody, A. R. (1999). Measurement of infarct volume in stroke patients using adaptive segmentation of diffusion weighted MR images. *Lecture Notes in Computer Science (Including Subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*. https://doi.org/10.1007/10704282_3
- Neumann, A. B., Jonsdottir, K. Y., Mouridsen, K., Hjort, N., Gyldensted, C., Bizzi, A., Fiehler, J., Gasparotti, R., Gillard, J., Hermier, M., Kucinski, T., Larsson, E. M., Sørensen, L., & Østergaard, L. (2009). Interrater agreement for final infarct mri lesion delineation. *Stroke*, *40*(12), 3768–3771. <https://doi.org/10.1161/STROKEAHA.108.545368>
- Rekik, I., Allasonnière, S., Carpenter, T. K., & Wardlaw, J. M. (2012). Medical image analysis methods in MR/CT-imaged acute-subacute ischemic stroke lesion: Segmentation, prediction and insights into dynamic evolution simulation models. A critical appraisal. *NeuroImage: Clinical*, *1*(1), 164–178.
- Ronneberger, O., Fischer, P., & Brox, T. (2015). U-net: convolutional networks for biomedical image segmentation. *Medical Image Computing and Computer-Assisted Intervention MICCAI 2015: 18th International Conference, Munich, Germany, October 5-9, 2015, Proceedings, Part III 18*, 234–241.
- Roth, G. A., Mensah, G. A., Johnson, C. O., Addolorato, G., Ammirati, E., Baddour, L. M., Barengo, N., Beaton, A. Z., Benjamin, E. J., & Benziger, C. P. (2020). Global burden of cardiovascular diseases and risk factors, 1990–2019: Update from the GBD 2019 study. *Journal of the American College of Cardiology*, *76*(25), 2982–3021.
- Tureckova, A., & Rodríguez-Sánchez, A. J. (2018). ISLES challenge: U-shaped convolution neural network with dilated convolution for 3D stroke lesion segmentation. *International MICCAI Brainlesion Workshop*, 319–327.
- Von Kummer, R., Meyding-Lamade, U., Forsting, M., Rosin, L., Rieke, K., Hacke, W., Sartor, K., & Tomsick, T. A. (1994). Sensitivity and prognostic value of early CT in occlusion of the middle cerebral artery trunk. *American Journal of Neuroradiology*, *15*(1), 9–18.
- Wang, G., Song, T., Dong, Q., Cui, M., Huang, N., & Zhang, S. (2020). Automatic ischemic stroke lesion segmentation from computed tomography perfusion images by image synthesis and attention-based deep neural networks. *Medical Image Analysis*, *65*, 101787. <https://doi.org/10.1016/j.media.2020.101787>