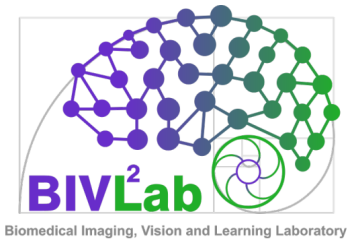

SEGMENTATION OF ISCHEMIC STROKE LESIONS IN RADIOLOGICAL SEQUENCES USING DEEP ATTENTION MECHANISMS

SANTIAGO GÓMEZ HERNÁNDEZ

UNIVERSIDAD INDUSTRIAL DE SANTANDER
FACULTAD DE INGENIERÍA FISICOMECÁNICAS
ESCUELA DE INGENIERÍA DE SISTEMAS E INFORMÁTICA
BUCARAMANGA

2023

SEGMENTATION OF ISCHEMIC STROKE LESIONS IN RADIOLOGICAL SEQUENCES USING DEEP ATTENTION MECHANISMS

SANTIAGO GÓMEZ HERNÁNDEZ

Research work in partial fulfillment of the requirements for the degree of:
Magíster en Ingeniería de Sistemas

Advisor:
Fabio Martínez Carrillo
Ph.D in Systems and Computer Engineering

UNIVERSIDAD INDUSTRIAL DE SANTANDER
FACULTAD DE INGENIERÍA FISICOMECÁNICAS
ESCUELA DE INGENIERÍA DE SISTEMAS E INFORMÁTICA
BUCARAMANGA
2023

ACKNOWLEDGEMENTS

The author expresses his acknowledgement:

I would like to express my gratitude for professor Fabio Martínez for his great guidance, patience, and dedication. I would also like to thank him for his interest in the formation of integral and professional skills, without him it would not have been possible to carry out this work. Also, thanks to all my colleagues of the BIVL^{2}ab research group for making the workplace an ideal place for collaboration and communication. Moreover, thanks to Dr. Daniel Mantilla and his team for the time spent in teaching me about stroke.

Finally, a special thanks to my grandparents Jaime and Omaira, my sister Valentina, my mother Jennie and uncle Jimmy for their unwavering support throughout this process. Also, thanks to my girlfriend Camila for supporting me in the good and difficult moments. Their continued trust and support was of great help to continue every day.

CONTENTS

page

LIST OF FIGURES

page

LIST OF TABLES

page

LIST OF APPENDICES

page

ABSTRACT

TITLE: SEGMENTATION OF ISCHEMIC STROKE LESIONS IN RADIOLOGICAL SEQUENCES USING DEEP ATTENTION MECHANISMS [*]

AUTHOR: SANTIAGO GÓMEZ HERNÁNDEZ [**]

KEYWORDS: ISCHEMIC STROKE SEGMENTATION, ATTENTION MECHANISMS, CT, MRI, IMBALANCED PROBLEMS.

DESCRIPTION: The key component of stroke diagnosis is the localization and delineation of brain lesions, especially from MRI studies. Nonetheless, this manual delineation is time-consuming and biased by expert opinion. This work introduces an autoencoder architecture that effectively integrates cross-attention mechanisms, together with hierarchical deep supervision to delineate lesions under scenarios of remarked imbalance tissue classes, challenging geometry of the shape, and a variable textural representation. Firstly, a cross-attention deep autoencoder is herein proposed to focus on the lesion shape through a set of convolutional saliency maps, forcing skip connections to preserve the morphology of affected tissue. Moreover, a deep supervision training scheme was adapted to induce the learning of hierarchical lesion details. Besides, a special weighted loss function remarks lesion tissue, alleviating the negative impact of class imbalance. The proposed model was effectively trained and validated over MRI studies. Interestingly, the proposed work was also adapted to segment stroke lesions over CT sequences. Taking into account the low contrast of radiological findings over CT studies, the approach includes an auxiliary learning task to pay attention of lesion boundaries. Regarding MRI studies, the proposed approach was validated on the public ISLES2017 dataset outperforming state-of-the-art results, achieving a dice score of 0.36 and a precision of 0.42. The best architectural configuration was achieved by integrating ADC, TTP and Tmax sequences. With respect to model performance over CT sequences, An evaluation over the ISLES2018 dataset showed competitive results of 0.42 in dice score and 0.48 in precision. The contribution of deeply supervised cross-attention autoencoders allows a better support to the discrimination between healthy and lesion regions, which consequently results in favorable prognosis and follow-up of patients.

---

[*] Research work

[**] Facultad de Ingeniería fisicomecánicas. Escuela de Ingeniería de Sistemas e Informática. Advisor: Fabio Martínez Carrillo, Ph.D.

RESUMEN

TÍTULO: SEGMENTACIÓN DE ACCIDENTES CEREBROVASCULARES ISQUÉMICOS EN SECUENCIAS RADIOLÓGICAS UTILIZANDO MECANISMOS DE ATENCIÓN CON APRENDIZAJE PROFUNDO [*]

AUTOR: SANTIAGO GÓMEZ HERNÁNDEZ [**]

PALABRAS CLAVE: SEGMENTACIÓN DE ACCIDENTE CEREBROVASCULAR ISQUÉMICO, MECANISMOS DE ATENCIÓN, TC, RM, PROBLEMAS DESBALANCEADOS.

DESCRIPCIÓN: La localización y delimitación de las lesiones cerebrales es un componente clave del diagnóstico de los accidentes cerebrovasculares, especialmente a partir de estudios de RM. Sin embargo, esta delineación manual requiere mucho tiempo y está sesgada por la opinión de los expertos. Este trabajo presenta una arquitectura de autocodificador que integra eficazmente mecanismos de atención cruzada, junto con supervisión profunda jerárquica para delinear lesiones en escenarios con gran desbalance de clases, geometría desafiante de la forma y una representación textural variable. En primer lugar, se propone un autocodificador profundo de atención cruzada para centrarse en la forma de la lesión a través de un conjunto de mapas convolucionales de saliencia, forzando las conexiones de salto para preservar la morfología del tejido afectado. Además, se adaptó un esquema de entrenamiento de supervisión profunda para inducir el aprendizaje de detalles jerárquicos de la lesión. Adicionalmente, una función de pérdida ponderada especial destaca el tejido de la lesión, aliviando el impacto negativo del desbalance de clases. El modelo propuesto se entrenó y validó eficazmente en estudios de RM. Interesantemente, el trabajo propuesto también se adaptó para segmentar lesiones de accidente cerebrovascular en secuencias de TC. Teniendo en cuenta el bajo contraste de los hallazgos radiológicos en los estudios de TC, el enfoque incluye una tarea de aprendizaje auxiliar para prestar atención a los bordes de la lesión. En cuanto a los estudios de RM, el enfoque propuesto se validó en el conjunto de datos público ISLES2017 superando los resultados del estado del arte, alcanzando un dice score de 0.36 y una precisión de 0.42. La mejor configuración arquitectónica se consiguió integrando las secuencias ADC, TTP y Tmax. Con respecto al rendimiento del modelo sobre secuencias TC, una evaluación sobre el conjunto de datos ISLES2018 mostró

---

[*] Trabajo de investigación

[**] Facultad de Ingeniería Fisicomecánicas. Escuela de Ingeniería de Sistemas e Informática. Director: Fabio Martínez Carrillo, Ph.D.

resultados competitivos de 0.42 en dice score y 0.48 en precisión. La contribución de los autocodificadores de atención cruzada profundamente supervisados permite un mejor apoyo en la discriminación entre las regiones sanas y lesionadas, lo que en consecuencia se traduce en un pronóstico y seguimiento favorables de los pacientes.

INTRODUCTION

Stroke is the second leading cause of mortality worldwide, being responsible for the major adult disability in developed countries. A dramatic projection estimates that one of four people over 25 years will suffer a stroke [1] [2]. Even worse, stroke has an associated high morbidity risk. Particularly in Colombia, stroke is one of the first five death causes, reporting during 2019, 32 deaths per 100.000 people, which represented 15.882 deaths associated to this disease [3].

Ischemic stroke is the most common condition (80% of all cases), caused by blood vessel occlusion. The immediate localization, delineation and quantification of the ischemic lesion, over CT and MRI sequences (according to availability), is fundamental to determining the diagnosis and consequent clinical intervention. This task is however tedious, time-consuming (around 15 minutes in each case) and prone to errors, introducing an inherent radiologist bias [4]. In fact, a low inter-rater expert agreement has been reported, being more consistent with the annotation in FLAIR sequences than in T2 [5].

Deep autoencoder representations have emerged as the most promising tools to support stroke lesion segmentation, discovering, among others, textural differences between affected and healthy

---

[1] WHO. *World Stroke Organization*. 2021.

[2] Gregory A Roth et al. "Global burden of cardiovascular diseases and risk factors, 1990–2019: update from the GBD 2019 study". In: *Journal of the American College of Cardiology* 76.25 (2020), pp. 2982–3021.

[3] Ministerio de Salud y Protección Social de Colombia. *Enfermedad cerebrovascular, otra comorbilidad priorizada contra el covid-19*. `https://www.minsalud.gov.co/Paginas/Enfermedad-cerebrovascular,-otra-comorbilidad-priorizada-contra-el-covid-19.aspx`. Accessed: 2023-01-19.

[4] Anne L. Martel et al. "Measurement of infarct volume in stroke patients using adaptive segmentation of diffusion weighted MR images". In: *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*. 1999. DOI: `10.1007/10704282_3`.

[5] Anders B. Neumann et al. "Interrater agreement for final infarct mri lesion delineation". In: *Stroke* 40.12 (2009), pp. 3768–3771. DOI: `10.1161/STROKEAHA.108.545368`.

regions [6] [7]. The first group of approaches capture multi-context information at different scales. These approaches include several image modalities, but are designed under rigid architectures, requiring an excessive hyper-parameter grid searching to recover latent lesions representation [8] [9] [10] [11] [12] [13]. In fact, a central issue in stroke segmentation is the predominant class imbalance of the lesion region w.r.t healthy tissue, which may collapse traditional training schemes. In consequence, an alternative group of multi-context strategies has dealt with such imbalance problems using mini-batches of image patches, which are carefully designed to include balanced and strat-

---

[6] Liangliang Liu et al. "Attention convolutional neural network for accurate segmentation and quantification of lesions in ischemic stroke disease". In: *Medical Image Analysis* 65 (2020). DOI: 10.1016/j.media.2020.101791.

[7] 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 (2020), p. 101787. DOI: 10.1016/j.media.2020.101787. arXiv: 2007.03294.

[8] Konstantinos Kamnitsas et al. "Efficient multi-scale 3D CNN with fully connected CRF for accurate brain lesion segmentation". In: *Medical Image Analysis* 36 (2017), pp. 61–78. DOI: 10.1016/j.media.2016.10.004. arXiv: 1603.05959.

[9] Alzbeta Tureckova and Antonio J. Rodríguez-Sánchez. "ISLES challenge: U-shaped convolution neural network with dilated convolution for 3D stroke lesion segmentation". In: *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)* 11383 LNCS (2019), pp. 319–327. DOI: 10.1007/978-3-030-11723-8_32.

[10] S Mazdak Abulnaga and Jonathan Rubin. "Ischemic Stroke Lesion Segmentation in CT Perfusion Scans Using Pyramid Pooling and Focal Loss". In: *Brainlesion: Glioma, Multiple Sclerosis, Stroke and Traumatic Brain Injuries*. Ed. by Alessandro Crimi et al. Cham: Springer International Publishing, 2019, pp. 352–363. DOI: https://doi.org/10.1007/978-3-030-11723-8_36.

[11] Adriano Pinto et al. "Enhancing Clinical MRI Perfusion Maps with Data-Driven Maps of Complementary Nature for Lesion Outcome Prediction". In: *Lecture Notes in Computer Science* 2 (2018), pp. 107–115.

[12] Xiaojun Hu et al. "Brain SegNet: 3D local refinement network for brain lesion segmentation". In: *BMC Medical Imaging* 20.1 (2020), pp. 1–10. DOI: 10.1186/s12880-020-0409-2.

[13] Jose Dolz, Ismail Ben Ayed, and Christian Desrosiers. "Dense multi-path u-net for ischemic stroke lesion segmentation in multiple image modalities". In: *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)* 11383 LNCS (2019), pp. 271–282. DOI: 10.1007/978-3-030-11723-8_27. arXiv: 1810.07003.

ified information during training [14] [15]. Such strategies, nonetheless, impose local processing, losing global geometry observations, which may be key to recovering lesions' shape properly. More recently, these architectures have included attention modules to highlight important local features, but in a global context[6].

This work proposed a deep supervised cross-attention strategy that segments ischemic stroke lesions using an autoencoder architecture. The attention modules exploit cross similarities between encoder and decoder observations to improve the decoder representation, preserving lesion patterns. Moreover, a strong penalization for lesion misclassifications is achieved with a hierarchical deep supervision and class weight maps on each level of the decoder. This architecture properly recovered radiological findings from MRI studies, complementing observations with maps computed from perfusion studies. Additionally, the proposed approach was adapted to deal with CT images that typically exhibit poor contrast and low sensitivity with respect to ischemia. In such a case, the proposed methodology was complemented with an auxiliary task to guide lesion boundaries. Regarding the MRI studies, the proposed approach was validated on the ISLES2017 public dataset, achieving state-of-the-art performance on automatic stroke segmentation from Apparent Diffusion Coefficient (ADC) sequences and the maps derived from perfusion-weighted studies. Furthermore, a validation with CT was carried out on the ISLES2018 public dataset, showing competitive stroke shape retrieval capabilities.

---

[14] Albert Clèrigues et al. "Acute ischemic stroke lesion core segmentation in CT perfusion images using fully convolutional neural networks". In: *Computers in Biology and Medicine* 115 (2019), p. 103487. DOI: 10.1016/j.compbiomed.2019.103487.

[15] Albert Clèrigues et al. "Acute and sub-acute stroke lesion segmentation from multimodal MRI". in: *Computer Methods and Programs in Biomedicine* 194 (2020). DOI: 10.1016/j.cmpb.2020.105521. arXiv: 1810.13304.

1. FUNDAMENTALS AND PREVIOUS WORK

1.1. Ischemic stroke and radiological findings

A stroke or brain attack suddenly affects the blood vessels of the brain. according to the cause, the stroke is categorized as ischemic or hemorrhagic. The ischemic stroke is caused by a clot or other blockage within an artery of the brain, being the most common lesion (80% of all strokes). This blockage causes a blood flow interruption, constraining oxygen and glucose in the brain, with a consequent neurological deficit. These deficits generate a chain of events at the cellular level, which may result in cell death. During the stroke, the brain uses collateral vessels to supply the blood and nutrients to affected tissue, an alternative that only operates in a relatively short time period [16].

Hence, a fast diagnostic is crucial to define endovascular therapy. The lesions associated to stroke are typically observed from Computed Tomography (CT) and Magnetic Resonance (MR), depending on availability. Over such medical image sequences are supported the lesion localization and quantification (an example is illustrated in Figure 3). In fact, a main challenge over such radiological findings is to determine the salvageable tissue (penumbra) and the infarct core. The main modalities to observe strokes are summarized thereafter.

Non-contrast Computed Tomography (NCCT) is usually the first modality used in the clinical workflow for stroke patients. This technique is useful to discriminate between ischemia and hemorrhage, as well as, to detect early ischemic changes [17]. This modality is commonly used because of its high availability and quick acquisition, a critical issue for diagnosing stroke. In the first hours subtle hypoattenuation in the affected area is a predictive factor for stroke.

---

[16] David S. Liebeskind. "Collateral circulation". In: *Stroke* 34.9 (2003), pp. 2279–2284. DOI: 10.1161/01.STR.0000086465.41263.06.

[17] R. Von Kummer et al. "Sensitivity and prognostic value of early CT in occlusion of the middle cerebral artery trunk". In: *American Journal of Neuroradiology* 15.1 (1994), pp. 9–18.

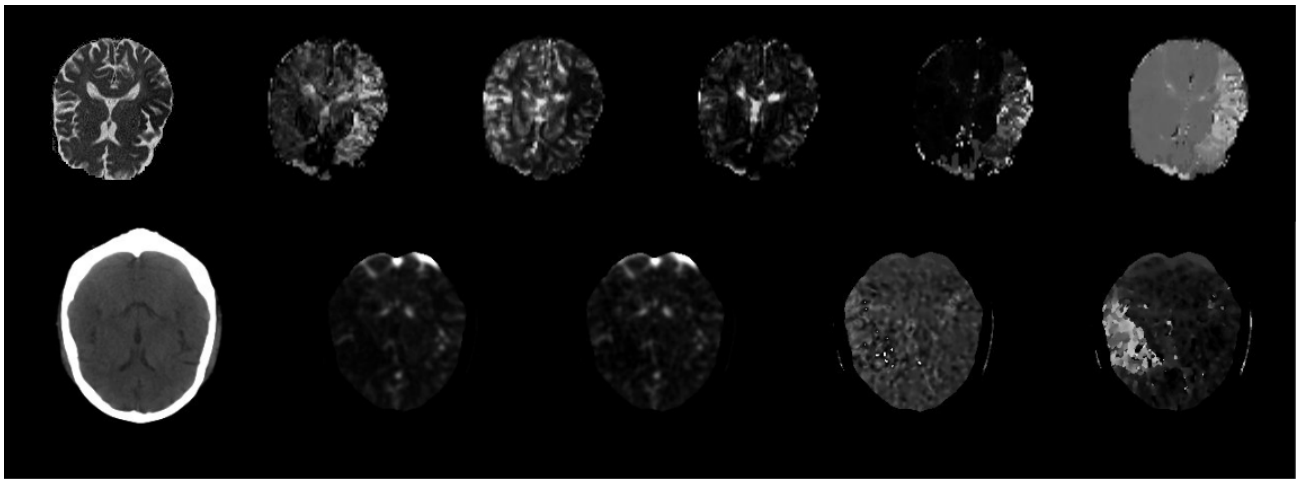

Figure 1. Imaging modalities used to analyze ischemic stroke patients. In the first row: ADC, rCBF, rCBV, Tmax, TTP. In the second row: NCCT, CBF, CBV, MTT, Tmax

This sign is due to increased vasogenic edema that occurs in the first hours after the onset of hypoxia and neuronal damage.

Magnetic resonance imaging (MRI) provides highly detailed anatomical information of intracranial, cervical and spinal structures. This imaging technique has significantly higher sensitivity and specificity in the diagnosis of acute ischemic infarction at the first few hours after the patient report symptoms. Complementary, the parametric diffusion MRI provides the most accurate assessment of the ischemic core. Nevertheless, this modality is expensive, time-consuming and less available than CT. Moreover, monitoring unstable patients in MRI is more difficult and images tend to be more susceptible to patient movement.

Perfusion from CT and MRI is a dynamic contrast alternative that allows obtaining penumbra images in patients with stroke. This phenomenon can be captured from CT or MRI sequences, providing a visual, and quantitative assessment of the penumbra area and ischemic core. CT perfusion is the most commonly used in patient evaluation [18]. CT perfusion is performed

[18] Jeremy J Heit, Greg Zaharchuk, and Max Wintermark. "Advanced neuroimaging of acute ischemic stroke: penumbra and collateral assessment". In: *Neuroimaging Clinics* 28.4 (2018), pp. 585–597.

by applying iodinated contrast, followed by the capture of temporal images. Hence, commercial computational alternatives offer the quantification of dynamic maps that summarize behavior along the sequence. The most common maps describe density changes of the brain parenchyma in images such as cerebral blood volume (CBV), cerebral blood flow (CBF), mean transit time (MTT), and time-to-maximum of the residue function (Tmax). In these images, the infarct core is represented by a region with severely reduced CBV or CBF, and the penumbra area is represented by regions with prolonged MTT or Tmax. On the other hand, MRI perfusion is also known as dynamic susceptibility imaging, and its availability is more limited than CT. This technique is performed by applying gadolinium followed by serial imaging. The advantages of MRI over CT are the absence of ionizing radiation and greater anatomical definition. Nonetheless, MRI is more expensive and performing the study requires a longer time.

1.2. Medical image segmentation using deep learning

In recent years, computer-aided diagnosis systems have made remarkable efforts to integrate segmentation tasks into the clinical workflow from a semiautomatic or automatic point of view. Figure 2 illustrates examples of stroke segmentation over MRI sequences. The U-net, a deep learning standard for segmentation, has provided impressive results in diverse medical image domains [19]. This architecture is an encoder-decoder U-shaped network that consists of two paths, one contracting and one expanding. The contracting path is used to capture the context of the image, through a stack of convolutional layers. On the other hand, the expanding path aims to produce a segmentation from the low-dimensional features. Moreover, to achieve more precise segmentations, U-net networks leverage skip connections from the encoder to the decoder on feature maps of the same spatial dimension. This architecture is based on

---

[19] Olaf Ronneberger, Philipp Fischer, and Thomas Brox. "U-Net: Convolutional Networks for Biomedical Image Segmentation". In: *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)* 9351.Cvd (2015), pp. 12–20. DOI: 10.1007/978-3-319-24574-4.

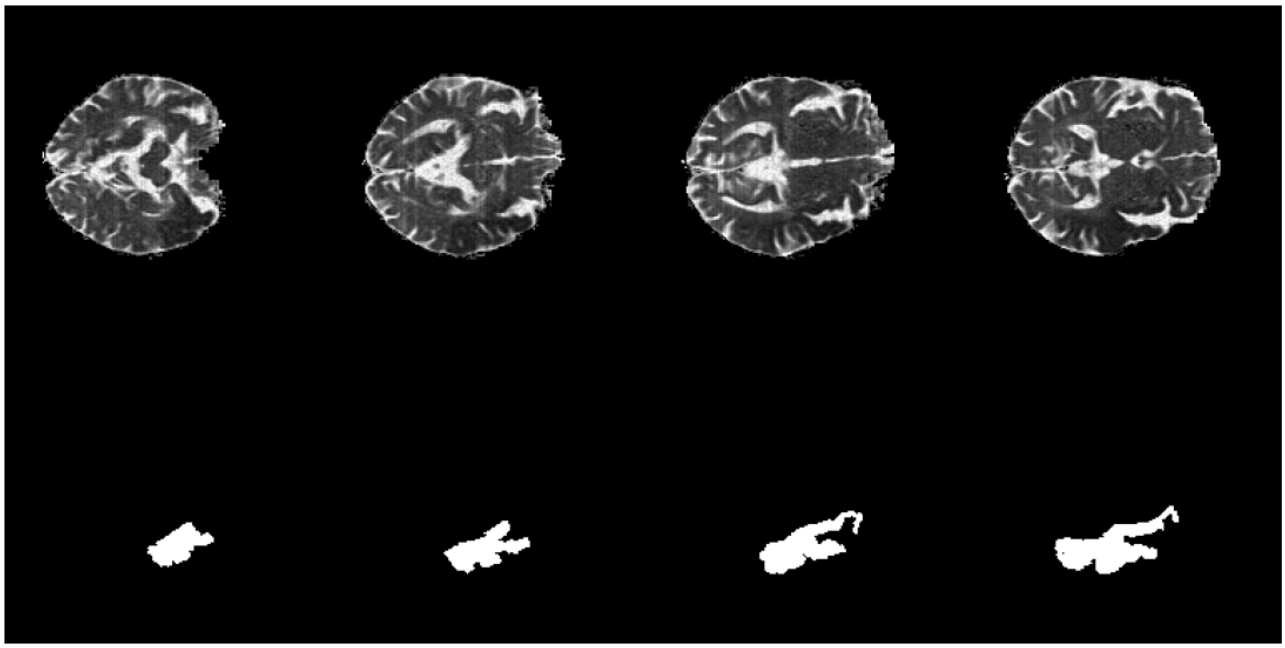

Figure 2. Semantic segments (second row) that were manually delineated from ADC images (first row). The white color represent the underperfused tissue.

a convolutional representation that summarises features from activation maps, hierarchically organized at each layer. These specialized networks have benefits such as sparse interactions that produce fewer connections; parameter sharing; a locally equivariant representation to make them locally invariant to translations and allow them to handle inputs of varying lengths. Nonetheless, the convolution's inductive bias implies locality during the modeling. Therefore, it is not possible to exploit the non-local relationships present in an image. For this reason, numerous works have nowadays adopted attention modules, which may capture non-local patterns, which are crucial to model complex shapes, like stroke lesions.

1.2.1. Attention-based architectures. Attention provides the capability of weighting the most relevant non-local features of input sequences. The first differentiable attention model

was first introduced by Bahdanau et al. [20] in the context of Neural Machine Translation using sequence-to-sequence (Seq2Seq) recurrent networks. Nowadays, attention is included in other tasks (computer vision and speech processing) [21] [22] and types of neural networks (CNNs and GNNs) [23] [24].

In brief, attention can be thought as a process that mimics the retrieval of information from a typical database scheme. We issue a query (q) to a database and get the value (v) which most likely fits our needs by its key (k). The implementation of this process in a neural network requires a function $sim(q, k)$ [20] [25] [26] that measures similarity (s) between the set of source and target sequences, *i.e.* queries and keys respectively. Subsequently, we compute the scores (a) using a non-linear function, typically a softmax, over the set of similarity scores (s). Finally, a weighted sum of the values v with their corresponding scores (a) is calculated. The process described is represented in Figure 3.

---

[20] Dzmitry Bahdanau, KyungHyun Cho, and Yoshua Bengio. "Neural Machine Translation by jointly learning to align and translate". In: *ICLR* 11.3 (2015), pp. 367–373. arXiv: 1409.0473v7.

[21] Xiao Yang. "An Overview of the Attention Mechanisms in Computer Vision". In: *Journal of Physics: Conference Series* 1693.1 (2020). DOI: 10.1088/1742-6596/1693/1/012173.

[22] Priyabrata Karmakar, Shyh Wei Teng, and Guojun Lu. "Thank you for Attention: A survey on Attention-based Artificial Neural Networks for Automatic Speech Recognition". In: (2021), pp. 1–11. arXiv: 2102.07259.

[23] Zhu Baozhou et al. "An Attention Module for Convolutional Neural Networks". In: (2021). arXiv: 2108.08205v1.

[24] Petar Veličković et al. *Graph Attention Networks*. Tech. rep. 2018, p. 12. arXiv: 1710.10903v3.

[25] Minh-Thang Luong, Hieu Pham, and Christopher D Manning. *Effective Approaches to Attention-based Neural Machine Translation*. Tech. rep. 2015, pp. 17–21.

[26] Ashish Vaswani et al. *Attention Is All You Need*. Tech. rep. 2017, p. 15. arXiv: 1706.03762v5.

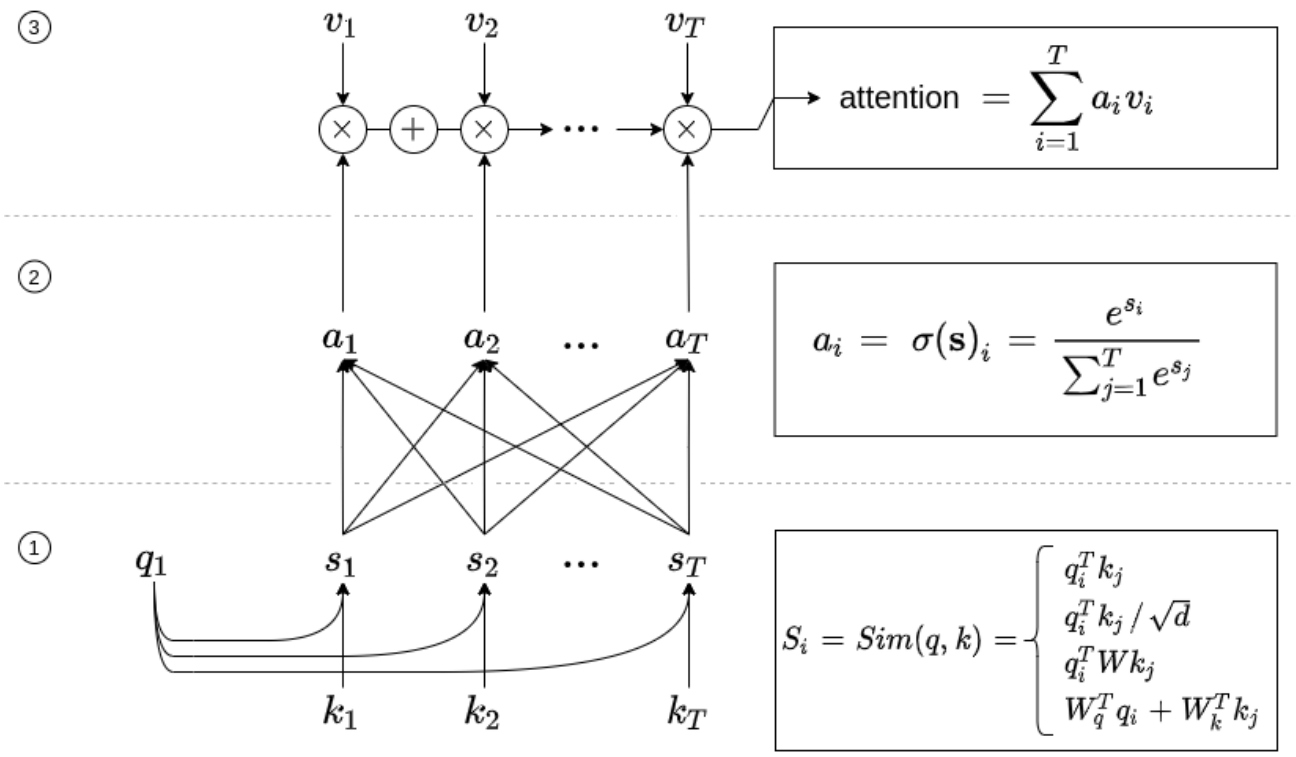

Figure 3. Attention coefficients computation flow. Similarity can take any of the functions listed in part 1.

1.3. Current work

Regarding stroke segmentation, some approaches have taken advantage of stroke observations from multiple parameters in perfusion MRI studies, modeled from multiple scales, hierarchically organized and included into convolutional backbones. These representations use different tissue dynamic responses to embedded shape descriptors. For instance, Kamnitsas *et al.* [8] proposed a 3D model with a dual pathway that processes FLAIR and T2 images at normal and low resolution, simultaneously. This scheme incorporates local and larger contextual information to deal with lesion characterization from local details to coarse shape moments. Nevertheless, this strategy is biased by the adequate size for the low-resolution sequences, to achieve complementary multi-scale shape description. Additionally, Tureckova *et al.* [9] used a 3D U-Net with dilated

convolutions to integrate larger context into the shape lesion modeling. Similarly, Abulnaga *et al.* [10] implemented a focal loss function to learn more difficult stroke shape variations, together with a pyramid pooling scheme that leverages global and local contextual information. These approaches strongly depend on initial setup configurations, requiring an appropriate choice of hyperparameters, such as dilation rate, to avoid losing sight of the characteristics of the intrinsic lesions.

Other works have included information from multiple modalities into independent convolutional paths. For instance, Pinto *et al.* [11] proposed a late data fusion approach from the outputs of two U-Nets that process a stack of MRI diffusion, together with 3D MRI perfusion and raw 4D PWI sequences. Likewise, Hu *et al.* [12] proposed Brain SegNet (a 3D U-Net with residual connections) and a curriculum training strategy to merge multi-modality information, refining fine-scale features at different scales in a multistage training. Dolz *et al.* [13] proposed a special U-Net to take image modalities separately with multiple encoders, learning specialized intrinsic characteristics of each modality. The encoders are densely connected and integrate inception modules with two convolutional blocks with different dilation rates. Although these approaches incorporate information from multiple medical sequences, there is no study of the contribution of each modality in the final lesion prediction. Besides, these approaches lack mechanisms to deal with the natural scarcity of annotated lesion regions regarding healthy regions, which may collapse training schemes.

An alternative group of strategies has focused on making predictions over patches to deal with the class imbalance problem. For instance, Clerigues *et al.* proposed two frameworks composed of a 2D [14] and a 3D [15] asymmetric residual encoder–decoder networks to generate lesion predictions, at a patch level. These networks were trained with a dynamically weighted loss function and mini-batches of image patches, which were carefully designed to include balanced stratified information. Despite the fact that training using image patches deal with class imbalance, such strategies impose local processing, losing global geometry observations, which may be key to properly recovering lesions' shape. Additionally, the individual patch lesion estimations are

totally independent, which may be undesired for big lesions and to achieve coherent segmentation from neighborhood regions. More recently, these architectures have included attention modules to highlight important local features, but in a global context. For instance, Liu *et al.* [6] proposed a U-Net architecture that includes receptive field convolutions, and skip self-attention modules to recover lesion segmentations. Such work deals with the recovery of lesion patterns from weighted self-attention maps. Nevertheless, the implemented learning is dependent only on encoder local observations, being the self-representation limited to take attention regarding shape segmentation. Also, Wang *et al.* [7] proposed a generator scheme to learn a robust stroke lesion segmentation, from a translation (from CT to MRI) pretext task. This strategy, however, requires aligned CT-MRI data which results unrealistic in clinical scenarios.

2. Research Problem

Stroke is a multifactorial disease, whose timely management is determinant to reduce mortality and avoid irreversible neurological consequences. The quantification of lesions, over radiological images, is a fundamental task to determine treatment and patient intervention. Nonetheless, this task is tedious with reported high expert variability and also a bias associated with image source and machine sensibility. Recently, deep learning strategies have supported the segmentation and quantification of such stroke lesions. However, the segmentation problem remains open due to the high variability of lesion geometry and the remarked class imbalance of lesion tissue with the rest of the study. Besides, much of the reported strategies focus on fully convolutional representations that lose non-local lesion patterns that may be key to discriminating among healthy and damaged brain tissue, with similar textural patterns. More recent approaches have considered multi-context features, allowing to produce segmentations with major overlapping with respect to expert annotations. These architectures have shown promising results but their respective design remains inflexible, given in much of the cases the same importance to whole recovered learned features. The design of attention modules may cope with such disadvantages, providing the architectures the capability to learn and filter dependencies between extracted features. Nonetheless, some reported approaches remain to focus on learning dependencies in specific parts of the network.

Research Question

How to design a deep attention mechanism to segment ischemic stroke lesions over radiological image sequences?

3. OBJECTIVES

General Objective

- To propose a deep learning strategy with attention mechanisms for segmenting ischemic stroke lesions in radiological sequences.

Specific Objectives

- To select a radiological image modality dataset that reports expert annotations of ischemic stroke lesions.
- To develop an encoder-decoder deep representation model that learns correspondences between images and segmentation inputs.
- To integrate attention modules on encoder-decoder deep representation to capture non-local dependencies of ischemic annotated lesions.
- To validate the segmentations produced using metrics that measure the performance of the model.

4. PROPOSED APPROACH

This work introduces an encoder-decoder representation that includes a set of additive cross-attention modules to enrich the segmentation of stroke lesions through the decoder path. Besides, the proposed encoder-decoder scheme is trained from deep multi-scale supervision to carefully focus on the proper recovery of lesion shapes along the different structural levels of the decoder. The general description of the proposed net is illustrated in Figure 4. The complete content of this section has been accepted in the "SPIE Medical Imaging" international conference [27] and it is under review for the "Biomedical Physics & Engineering Express" journal [28].

4.1. Additive cross-attention mechanisms

In this work an encoder-decoder architecture is enhanced from cross-attention modules (σ_{att}) that recover stroke lesions, dealing even with typical small samples. A detailed illustration of the attention module is illustrated in Figure 5. In fact, in stroke segmentation, an average lesion only corresponds to less than 1% of the total volume of the study. For instance, from an average analysis of ISLES2017 training cases was computed a stroke index of 0.56%.

To deal with the size of the lesion and the reported variability, the cross-attention modules properly regularize and select encoder outputs X_e that will be mapped into the decoder representation X_d, at the same level of detail L_i. These modules allows passing the most aligned encoder features (from encoder (X_{e,L_t}) with respect to decoder features in the same scale (X_{d,L_t})) to avoid disrupting the decoder representation. Each attention module follows an additive align-

---

[27] Santiago Gómez et al. "An attentional Unet with an auxiliary class learning to support acute ischemic segmentation on CT". in: *SPIE Medical Imaging*. 2023 (in press).

[28] Santiago Gómez et al. "A deep supervised cross-attention strategy for ischemic stroke segmentation in MRI studies". In: *Biomedical Physics & Engineering Express* (in press).

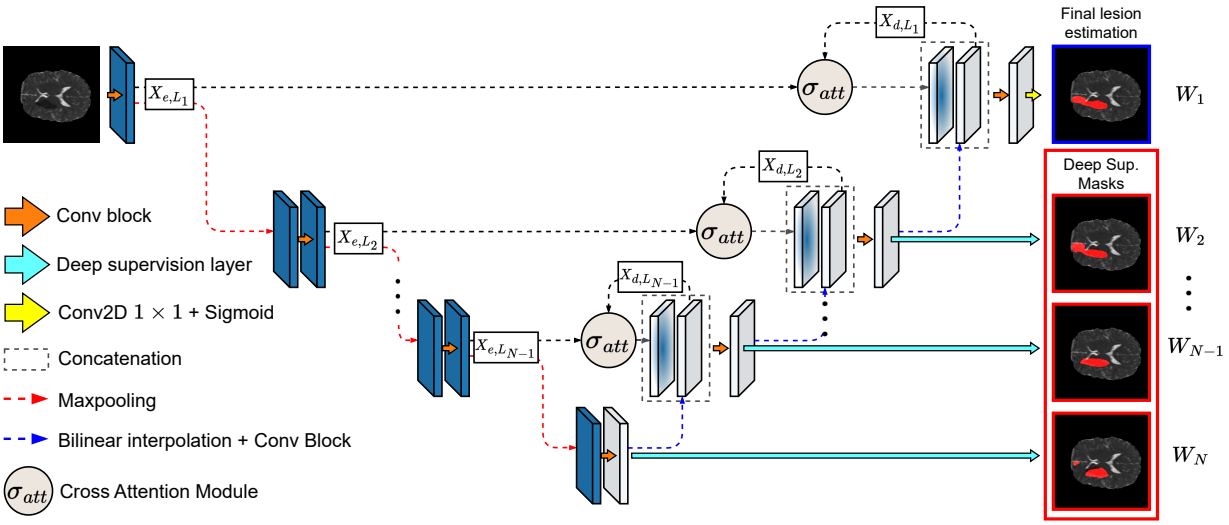

Figure 4. An overview of a deeply supervised attention autoencoder for the segmentation of acute ischemic stroke lesions. The cross-attention module σ_{att} controls the contribution of encoder features on skip connections, while deep supervision induce the learning of features for lesion estimation on early stages of the decoder.

ment, expressed as:

$$S = \sigma_1(W_{e,L_t}^T X_{e,L_t} + W_{d,L_t}^T X_{d,L_t})$$

Particularly, the additive alignment of query and key features are derived from linear projections of X_{d,L_t} and X_{e,L_t}, using convolution branches [20]. In this module, σ_1 is a ReLU activation. As illustrated in Figure 5 a pool across positive operation is implemented by following a 1×1 convolutional layer, as $X_{re,L_t} = \sigma_2(W_S^T \cdot S) \cdot X_{e,L_t}$, where σ_2 is a sigmoid activation. The attention map highlights main brain structures associated to stroke lesions. Hence, the attention map is broadcast across all feature maps from X_e to produce refined encoder features X_{re} through an element-wise multiplication.

The refined features X_{re} preserve highly correlated features between encoder and decoder branches, at the same level of processing. Hence, cross-attention modules promote a better representation and segmentation of ischemic stroke lesions by decreasing noisy signals contribu-

tion in the skip connections of classical autoencoders.

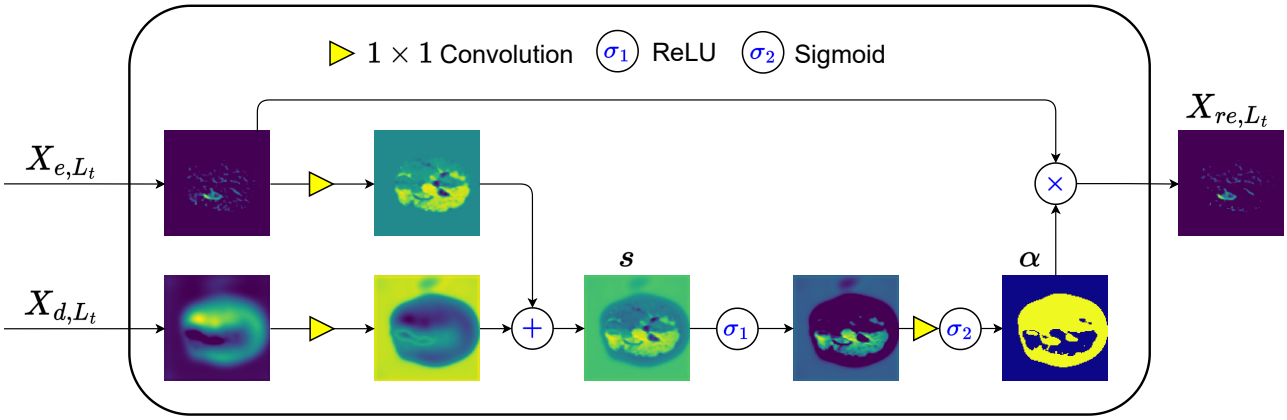

Figure 5. Additive cross-attention module overview. The module takes encoder and decoder features from the same level to highlight the relevant features from the encoder.

4.2. Deep supervision

In this work, we performed a hierarchical deep supervision to enhance ischemic stroke lesion segmentation (see an illustration in Figure 4). This kind of supervision adds companion objective functions, defined as replicas of the loss function calculated from coarser lesion estimations. Stroke lesion estimations at coarse levels, *i.e.,* in layers close to embedding representation, are computed from deep supervision layers on the decoder. These layers are made of one bilinear interpolation and a cross-feature estimation from one 1×1 convolution with sigmoid activation to produce the probability map for the lesion estimation. During training, all the lesion estimations are leveraged to compute a loss function that considers hierarchical lesion information. Therefore, the final loss signal is computed as the weighted sum of all the loss terms ($\mathcal{L}_i$), as: $\mathcal{L}_{total} = \sum_{i=1}^{k} \mathcal{W}_i \mathcal{L}_i$, where k is the number of lesion estimations and $\mathcal{W}_i$ is the corresponding weight for layer L_i.

This training constrains the decoder representations to produce predictions close to the manually delineated masks on early stages, allowing the model to focus on refining details on deeper representations. As a result, the error propagation from low to high dimensional levels is

minimized.

4.3. Class weight maps

Complementary, to overcome the strong class imbalance and increase precision, we proposed a careful training of the proposed architecture, following class weight maps. Class weights maps are built with the ischemic lesions manual delineations as reference, hence, they have the same spatial dimension. Pixels of the same class are assigned a specific weight to highlight its importance. An illustration of the class weight maps creation and its effect in the optimization process is in Figure 6.

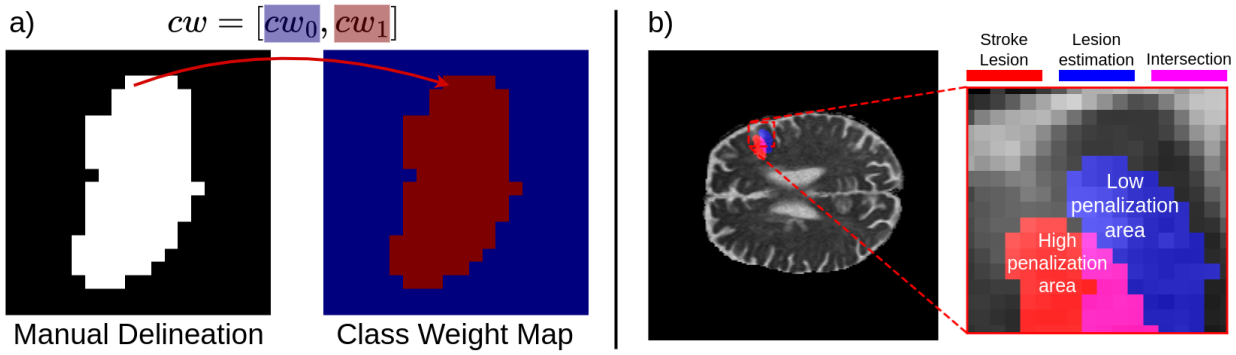

Figure 6. Creation of a class weight map (a) and its effect in optimization (b) for a given prediction. Class weight maps allow for a strong punishment of unnoticed lesion pixels. Correctly predicted pixels were not annotated for simplicity.

These special maps affect the loss function score for the classification of each pixel ℓ_{ij}, giving more importance to lesion pixels. These maps are easily integrated with deep supervision models as:

$$\mathcal{L}(y, \hat{y}, w) = 1/(H \cdot W) \sum_{i=1}^{H} \sum_{j=1}^{W} w_{ij} \cdot \ell_{ij}$$

where w_{ij} is the particular weight for the pixel at position (i, j) that distributes a local penalization of the ℓ_{ij} loss function, at each position. The incorporation of class weights maps

circumvents common training pitfalls, where all voxels are equally important. More specifically, prevents models from stucking in local minimums where labeling all pixels as non-lesion yields low loss values ($< 1\%$ of pixels are lesion). As a consequence, the model learns meaningful patterns from the lesion samples and produces more precise estimations.

4.4. Lesion boundaries as an auxiliary class for stroke segmentation in CT studies

A main advantage of the introduced scheme to segment stroke lesions is the capability and adaptability to be operated over other medical image sequences. For instance, the proposed architecture could be trained over CT scans from acute ischemic stroke patients (few hours from onset) and try to infer stroke lesions. Nonetheless, These CT images have low contrast and therefore exists a close similarity between healthy and lession tissues. To force a stroke lesion learning over such poor sensitivity sequences, we include an auxiliary class that consists of boundaries of the shape. As a result, the proposed approach focus on lesion shape but also on learning discrimination among healthy and affected boundary tissue. We built on the intuition that having a dedicated contours class will promote the learning of finer masks. Hence, classification and deep supervision layers were modified to produce three probability maps that resemble the lesion borders, lesion and background classes. To achieve this, the convolution layer was changed to produce three feature maps and the sigmoid activation was replaced with a softmax activation. Figure 7 depicts the adaptation of the proposed strategy for segmenting ischemic stroke on CT studies.

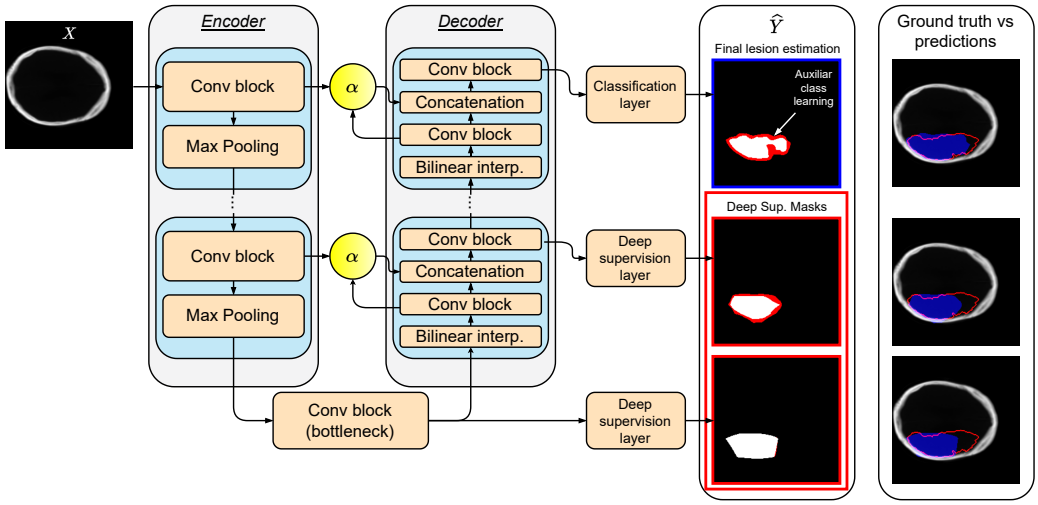

Figure 7. An overview of a boundary-focused attention autoencoder for the segmentation of ischemic stroke lesions on CT images. The auxiliar class learning happens on the deep supervision and prediciton layers to control the shape of the final prediction.

5. EXPERIMENTAL SETUP

5.1. The Data

The proposed approach was validated over *the Ischemic Stroke Lesion Segmentation challenge dataset from 2017 (ISLES2017)*. This dataset contains 75 studies from patients diagnosed with acute ischemic stroke. The dataset comes pre-partitioned in training (n=43) and hidden testing (n=32) splits. For each study, there is one MRI diffusion map (ADC), five MRI perfusion maps (rCBF, rCBV, MTT, TTP, Tmax), and one 4D PWI. All MRIs were already co-registered and skull-stripped. The manually delineated masks were only available for the studies on the training partition. The volumes were normalized following a min-max normalization. The slices from the normalized volumes were extracted and resized to 224×224. Finally, we randomly left out some studies from the training partition to form a validation partition (n=9).

5.2. The Model

The architecture of the proposed approach was implemented with a convolutional backbone that follows blocks with a convolution layer (kernel of 3×3), batch normalization, and a ReLU activation, twice at each block. The encoder and decoder have six processing levels. At each level of the encoder, one convolutional block is applied, followed by a max-pooling operation to shrink the features spatial dimension by a factor of 2. The encoder's output is then a semantic hidden representation 64 times smaller than the original input image. Each decoder level applies 1 convolutional block over the concatenation of filtered encoder features and upsampled features from the same decoder level. The upsampling operation increases the feature size by a factor of 2, following a bilinear interpolation. Deep supervision layers were placed at the outputs of each decoder level, except for the last decoder output which is projected into a single map from a 1×1 convolution and activated with a sigmoid function.

Regarding the training, the proposed approach was adjusted during 600 epochs with a batch

size of 32 and a Focal loss with an AdamW optimizer [29]. The initial learning rate was set to 5e-3 and the weight decay to 1e-5. A linear warm-up strategy was used in the first 90 epochs and a cosine decay was applied for the next 510 epochs to slowly decrease the learning rate. The class weights were set to 0.3 for the non-lesion (w_0) class and 0.7 for the lesion class (w_1). Also, train-time augmentations were applied to the slices by adding random brightness and contrast, random flips along the horizontal and vertical axes, random rotations between -7 and 7 degrees, random elastic transformations and random grid and optical distortions. For the deep supervised models, each of the six outputs was compared to the full resolution manual annotation. The corresponding loss terms were weighted with $\mathcal{W} = \{0.03, 0.045, 0.05, 0.125, 0.25, 0.5\}$. Thus, the total loss computation had a lower contribution from the coarser estimations and greater importance from the finer ones. The code of the proposed strategy will be available here.

5.3. An extended evaluation over CT sequences

The validation of proposed approach was also validated over CT studies to validate the capabilities of the scheme with an auxiliary boundary task and acting over low-contrast sequences. For doing so, the validation was carried out on *the Ischemic Stroke Lesion Segmentation* challenge dataset from 2018 (*ISLES2018*). This dataset contains 156 studies from acute ischemic stroke patients (94 for training and 62 for test). For each study, there is one CT image, four CT perfusion maps (CBF, CBV, MTT, Tmax), and one CTP source data. Manually delineated masks were only available for the training studies, and all modalities were already co-registered. All volumes were normalized and resized to 224×224 pixels. We randomly left out 19 studies from the training partition to validation.

The same data augmentation strategy and optimization schedule adopted for MRI studies were also used for the architectural design implemented for the CT studies. Particularly, for the clas-

[29] Ilya Loshchilov and Frank Hutter. "Decoupled weight decay regularization". In: *arXiv preprint arXiv:1711.05101* (2017).

sification layers and the class weight maps, we included an extra class to learn and remark lesion shape geometry. The class weights to generate the weight maps were set to $\mathbf{C} = \{0.7, 2.0, 7.0\}$. Also, a cross-entropy function was used instead of a focal loss for the multiclass strategy.

6. EVALUATION AND RESULTS

Evaluation multiparametric MRI studies. The proposed approach was principally validated to recover stroke segmentations from multiparametric MRI studies, over the ISLES2017 dataset. Firstly, an ablation study was conducted to measure the contribution of each special mechanism of the proposed approach. For doing so, we train architectures over: 1) a standard U-Net, 2) a strategy that includes class-weight maps (U-Net + CW) and 3) a strategy that include class-weight maps but also follows deep supervision. These architectures were compared with the complete architecture that includes class-weight maps, deep supervision training and cross-attention maps (U-Net + CW + DS + Att).

In Figure 8 are summarized the results for all defined configurations, using independent MRI maps. The bars represent the validation dice scores. The stacking of bars above the baseline shows how the added methodological components improve the ability to segment ischemic lesions. For components without an improvement, the respective score is illustrated with dotted lines (U-Net in ADC). Additionally, the thick bars positioned on the right are the zoomed versions to better observe the contribution of each component. As expected, the complete version of the proposed approach achieves better scores in the modalities rCBV (0.21 ± 0.22), MTT (0.36 ± 0.21), TTP (0.40 ± 0.23), and Tmax (0.38 ± 0.24). This showcases the capability of the proposed components to improve the learned representation. More specifically, the class weights show the major percentage contribution (an average increase of 1.46% on all single experiments), demonstrating the potential to overcome class imbalance problems. The remaining modalities reached their best lesion estimation configuration with simpler approaches. For instance, CBF (0.34 ± 0.24) only needed the class weights (U-Net + CW), and surprisingly ADC achieved the best score (0.45 ± 0.16). In such a case, the model reaches the maximum ADC representation with basic architectural components, expressing the lesion as the changes of magnitude on diffusion. This can be associated with the fact that expert annotations were performed on this modality. Moreover, this may suggest that ADC should be supported by an architecture

that also integrates other modalities to better support the stroke characterization. In total, an average gain of 3.34% over all the modalities was achieved with the complete version of the proposed approach w.r.t. to the baseline model. These gains represent a substantial difference considering that lesion size is tiny.

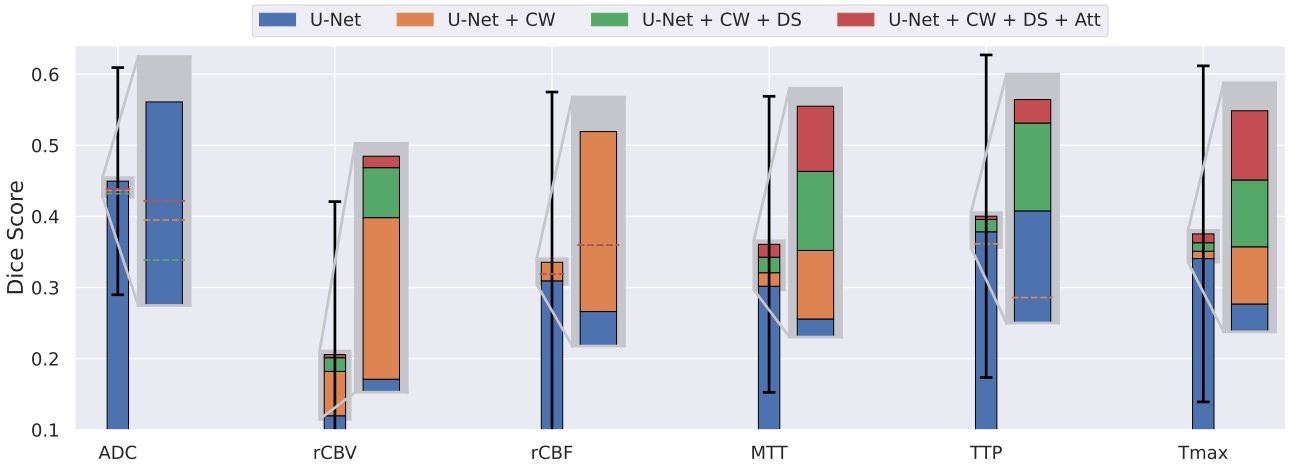

Figure 8. Isolated modalities validation results in the ISLES2017 dataset for different configurations of the proposed approach.

In a second experiment was measured the capability to exploit multi-context information. In such a case, an exhaustive validation was carried out with the proposed components and the three best modalities from the previous experiment as inputs. Also, in this experiment was considered a model that integrates all the modalities available in the dataset (All). The input of the network was made of the concatenation of medical images along the channel axis. Figure 9 summarizes the achieved results for each considered multimodal integration. In general, from such multimodal representation is observed a marked increase in the capability to segment ischemic lesions. The best combination resulted from combining the three top modalities (ADC + TTP + Tmax), with a reported dice score of 0.46 ± 0.14. Moreover, figure 9 shows that the framework clearly benefits from using multiple image modalities as input, which could be an indicator that it is leveraging multi-context information. For instance, multimodal input approaches that include ADC leverage all the proposed approach components to enrich the deep representation. In fact, these always achieve a higher score regarding the segmentation of

ischemic lesions. The model with all available images gives worse results (0.42 ± 0.22), a fact associated with noise propagation from whole modalities. The integration of all modalities may force to background bias, losing standing-out features from more promising modalities. Similar to the first experiment, the complete configuration increased the dice scores by an average of 2.21% w.r.t. baseline. The proposed components still contribute to a better ischemic stroke characterization with multiple sources of information.

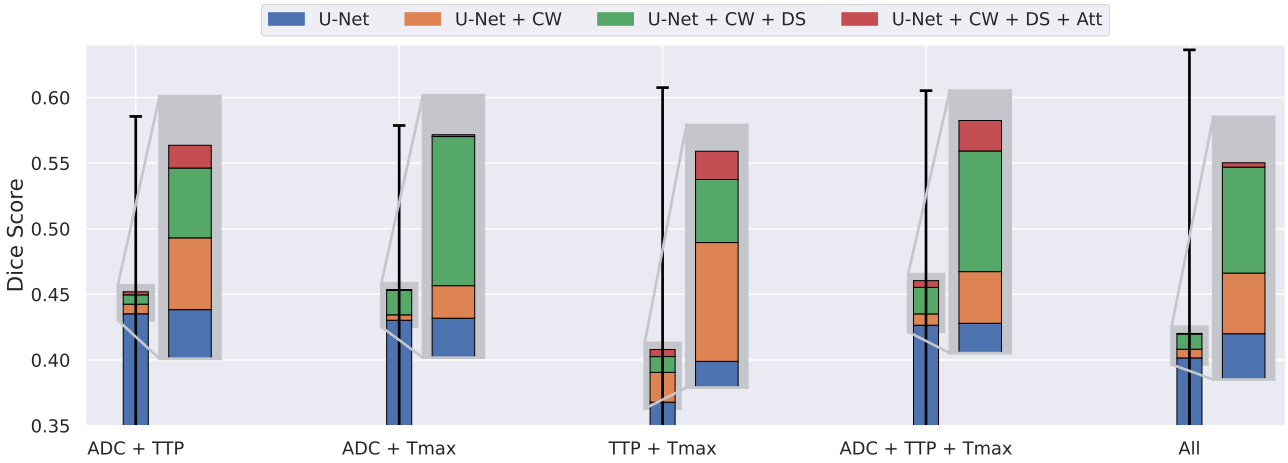

Figure 9. Multi-modal validation results in the ISLES2017 dataset for different configurations of the proposed approach.

As observational analysis, Figure 10 presents a set of segmentation samples (blue areas) obtained with the best multi-parametric model (ADC + TTP + Tmax), and the respective ground truth that corresponds to a manual expert annotation (red contour). The first two rows, which correspond to two different patients, illustrate the good capability of the model to localize damaged tissue in all slices of the medical volume, as well as to overlap boundaries delineated for experts. In the first row, the model prediction achieves an average dice score of 0.72, while in the second row is observed a stroke lesion segmentation that achieves an average dice score of 0.61. These results are coherent at different slices, even in tiny lesion regions, support that may be related with cross attention schemes and also with the learning at different scales. In the third row are illustrated the slices that correspond to a segmentation that achieved an average score of 0.28. The lesion in this study is more challenging, showing a slight difference

between healthy and stroke tissue. Nevertheless, the borders are poorly defined and in most cases, the size of the predicted lesion is larger than that delineated by the experts. Despite such disagreement with respect to the expert, the proposed approach properly localizes the lesion, a key issue in clinical routine.

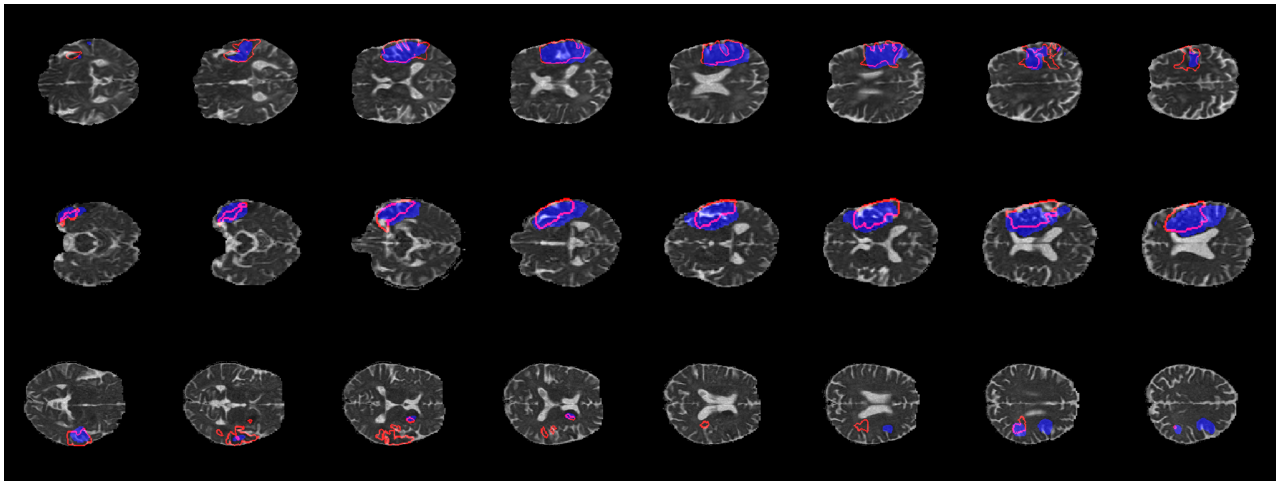

Figure 10. Ischemic stroke lesion predictions for the best attentional model on patients from the validation split of ISLES2017. The red contours are the physician manual delineations and the blue areas are the model estimations.

A baseline validation was carried out with respect to the reported in the test set on the official competition website. In such evaluation, the proposed approach outperforms the state-of-the-art, achieving the best scores in the ISLES2017 challenge with three different configurations. Table 1 summarizes the performance of the proposed approach, as well as, the other works reported in the literature. In fact, the proposed models outperformed the best existing approach [12] by 3%, 4%, and 6% on the dice score metric, respectively. More specifically, the best model used ADC, TTP and Tmax images, while the second used ADC + Tmax and the third ADC + TTP. Regarding the baseline, Pinto *et al.* [11] achieves an average dice score of 0.29 with a recall of 0.66 but a precision of only 0.23. These results suggest that in general such models are biased toward predicting the lesion class (coarse segmentations), with an overestimation of the lesion. On the other hand, Hu *et al.* [12] achieved a dice score of 0.30, a precision of 0.35, and a recall of 0.43. Significantly, the proposed approach was able to outperform their model in all

the metrics by 6% (dice), 7% (precision), 5% (recall).

Table 1. Results for the ISLES2017 hidden test split

Approach	Dice	Precision	Recall
Pinto *et al.*[30]			
- Standard model	0.20 ± 0.19	0.16 ± 0.20	0.61 ± 0.28
- 4D-PWI model	0.20 ± 0.18	0.18 ± 0.21	0.61 ± 0.27
- Multi-data model	0.26 ± 0.21	0.21 ± 0.20	0.61 ± 0.28
- Multi-data multi-model	0.29 ± 0.21	0.23 ± 0.21	$\mathbf{0.66 \pm 0.29}$
Hu *et al.*[31]			
- Brain SegNet	0.26 ± 0.22	0.35 ± 0.28	0.38 ± 0.29
- Brain SegNet + FL	0.30 ± 0.22	0.35 ± 0.27	0.43 ± 0.27
Ours			
- ADC + TTP	0.33 ± 0.22	0.39 ± 0.26	0.49 ± 0.32
- ADC + Tmax	0.34 ± 0.22	0.34 ± 0.23	0.57 ± 0.32
- ADC + TTP + Tmax	$\mathbf{0.36 \pm 0.21}$	$\mathbf{0.42 \pm 0.25}$	0.48 ± 0.29

Extended validation over CT sequences. To consider the validation of the proposed approach over CT studies, the proposed approach includes an auxiliary classification task. This extended version of the approach was validated over CT studies using the ISLES2018 dataset. Firstly, three architecture versions were evaluated with and without the auxiliary class: 1) a standard U-Net (U-Net), 2) a standard U-Net with class weights and deep supervision (U-Net + CW + DS), and 3) a standard U-Net with class weights, deep supervision and cross-attention modules (U-Net + CW + DS + Att). In table 2 is summarized the performance for all versions. As expected, the complete version of the proposed approach achieves the best dice score (0.643 ± 0.185).

Moreover, Figure 11 illustrates several segmentation outputs for the proposed approach (blue areas) and the expert reference (red contours). At first row samples, the complete strategy achieves the best score, evidencing a remarkable capability to properly define the boundaries of the ischemic lesion. In the second row, it is illustrated a challenging study with two ischemic

Table 2. Results for the ISLES2018 validation split

	Dice		Precision		Recall	
Model	**Binary**	**Auxiliary Class**	**Binary**	**Auxiliary Class**	**Binary**	**Auxiliary Class**
U-Net	0.624 ± 0.223	**0.635 ± 0.216**	**0.606 ± 0.234**	0.610 ± 0.241	0.736 ± 0.203	**0.732 ± 0.204**
U-Net + CW + DS	0.625 ± 0.215	**0.642 ± 0.199**	0.577 ± 0.246	**0.598 ± 0.220**	**0.790 ± 0.175**	0.756 ± 0.199
U-Net + CW + DS + Att	0.632 ± 0.219	**0.643 ± 0.185**	**0.650 ± 0.264**	0.630 ± 0.211	0.695 ± 0.227	**0.731 ± 0.188**

lesions of different sizes and geometry. The proposed approach nonetheless achieves a proper overlapping without an overestimation of the lesion, a key issue to define treatments and estimate patient prognosis.

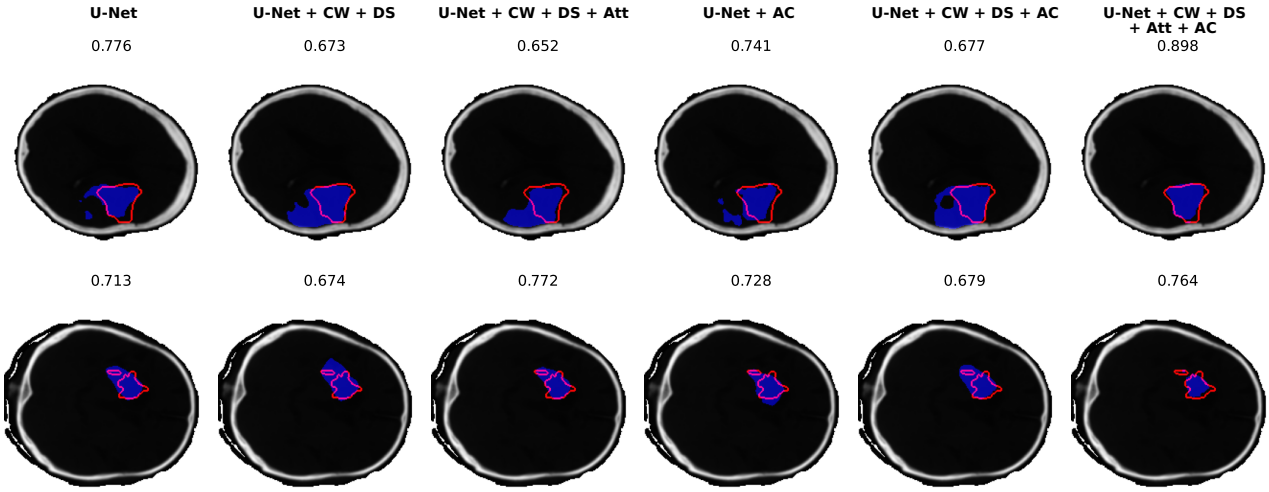

Figure 11. Comparison of the ischemic stroke lesion predictions (blue areas) by the proposed models vs manual expert annotations (red contours). The numbers above each plot are dice scores.

Finally, a baseline comparison was herein reported for our top model in the test set (results obtained from official competition website). In such evaluation, the model estimations obtained a competitive dice score of 0.42 ± 0.31. In comparison, Tureckova *et al.* implemented a 3D U-Net with dilated convolutions achieving a lower mean dice score of 0.37 (precision = 0.37 and recall = 0.44), a limitation associated to scarce information to learn volumetric patterns, while dilated information only exploit large local information, losing context from whole slice [9]. The network proposed by Clèrigues *et al.* used a patch based approach and achieved a dice of 0.49 ± 0.31, precision of 0.51 ± 0.36 and recall of 0.57 ± 0.35. This network loses contextual

information, considering independent patch observations, which may be a limitation to translate the strategy in real scenarios with complete studies and taking into account multiple lesions in one volume. Other approaches [32] [7] achieve remarkable results from CT-to-DWI translation schemes. However, these approaches require the existence of CT and MRI studies captured one after the other, which is difficult to achieve in clinical practice. Besides, these DWI images are no longer included in the dataset.

[32] Pengbo Liu. "Stroke lesion segmentation with 2D novel CNN pipeline and novel loss function". In: *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)* 11383 LNCS (2019), pp. 253–262. DOI: 10.1007/978-3-030-11723-8_25.

7. DISCUSSION

This work introduced a deep supervised cross-attentional deep autoencoder, which follows a training regime focused on ischemic stroke lesions, observed principally from multi-parametric MRI sequences. The proposed approach outperformed state-of-the-art methods on the ISLES2017 dataset, achieving a dice score of 0.36 and precision of 0.42. An exhaustive analysis of the contribution of MP-MRI sequences evidences that anatomical detail provided by ADC sequences brings better performance on the proposed approach. Likewise, the next modalities that gave the best information were the parametric maps TTP (0.36) and Tmax (0.38). Contrarily, the rCBV parametric map obtained the worst results with a dice score of 0.21, which may be correlated with the poor sensitivity of this modality [33].

To the best of our knowledge, this work introduced the cross-attention modules for stroke segmentation problems. The implemented cross-attention modules take into account partial encoder representation into the decoder, hence, diminishing the perturbation effect of the skip connections. Such fact may be observed in all experiments, where for different map inputs the attention modules achieve an average gain on dice score of 0.8% for the single modalities and 0.2% for the multimodal models w.r.t. to the U-Net + CW + DS version. Furthermore, the mean gain in dice score (1.5% for the single modalities and 0.8% for multimodal) supports the importance of the class weights in the training of the attention models. Other works validate different MRI datasets, for instance, [15] used an oversampling strategy to balance the damaged and healthy patches. This fact may cause models to overfit to the lesion class. This approach also considers the patches as multiple independent parts, which can be harmful for the overall lesion estimation geometry, losing anatomical information from the brain. Contrarily, our approach

---

[33] William A. Copen, Pamela W. Schaefer, and Ona Wu. *MR Perfusion Imaging in Acute Ischemic Stroke*. Good resource to understand the role of perfusion parametric maps in the assesment of stroke. May 2011. DOI: 10.1016/j.nic.2011.02.007.

does not affect the data distribution. Instead, it affects the loss function value by giving more importance to the underrepresented class. Similarly, carrying deep supervision of the decoder representation results crucial to further enhance the lesion estimation, as demonstrated with the mean gain in dice score of 1.11% for the single modalities and 1.16% for the multimodal w.r.t. the U-Net + CW baseline.

The proposed strategy achieves a robust lesion representation integrating MRI and perfusion maps, which compactly is convolved to recover stroke information. Other strategies, such as the proposed by Pinto *et al.* [??], have reported the analysis of stroke lesions using perfusion maps and raw 3D+t information. Such work reported a CNN architecture to model multiple parametric maps (Standard model), the analysis of 4-dimensional perfusion images (4D-PWI model), as well as the fusion of two previous inputs in a single branch (Multi-data model). Also, Pinto *et al.* [??] considers a model with late fusion of two independent branches (multiparametric and 4D-PWI) using a GRU layer (Multi-data multi-model) to enforce greater spatial context in the prediction of ischemic stroke lesion outcome.

Despite the fact that the 3D+t (4D-PWI) strategy considers more information, it only achieved a 0.20 dice score. Furthermore, they were able to obtain a dice score of 0.26 by combining the 3D+t data with the multi-parametric maps (Multi-data). Their respective best model obtained a dice score of 0.29 by processing the two data sources independently and fusing them with specialized GRU layers to learn spatiotemporal context (Multi-data multi-model). These results suggest that in general such models are biased towards predicting the lesion class (coarse segmentation), with an overestimation of the lesion. In such a case, the model may be biased to predict large ischemic lesions, *i.e.,* patients that will not receive any treatment, such as the mechanical thrombectomy. Contrary, the proposed approach outperforms their best model on the dice and the precision metrics by 7% and 19%, while achieving a balanced trade-off between precision and recall of 0.42, and 0.48. In the same line, Hu *et al.* [??] proposed a 3D residual Segnet with refinement modules that consider multi-level convolutional features, recursively from the top convolutional block to the bottom. Such work implements a three-stage curriculum learning

regime to perform a voxel-wise segmentation of brain lesions (Brain SegNet). Their framework comes in two variants, one with (Brain SegNet + FL) and one without (Brain SegNet) focal loss. Their best version achieved a dice score of 0.30, as well as a more subtle difference between precision (0.35) and recall (0.43). Nonetheless, this feature fusion may be corrupted with noise information if the representation between encoder and decoder is not significant. This problem is better handled by attention modules, which are explicitly designed to let through signals when they are strongly correlated. Moreover, the models that include volumetric regions may exploit information of new dimensional axis and take advantage of lesion geometry. However, training a robust 3D model is required a large amount of data, with strong dependencies of study resolution (mm among slices). Significantly, our approach was able to outperform their model in all the metrics by 6% (dice), 7% (precision), 5% (recall). Such improvement may be associated with the implemented attentional 2D models. Showing a better performance than the 3D Segnet from Hu *et al.*, with a difference of 6% in dice score.

Extension over CT sequences. The cross-attention mechanisms were also validated in stroke segmentation over CT studies, using the ISLES2018 dataset. In such case, it was demanding the use of an auxiliary contour lesion class to force representation to discriminate boundaries between healthy and ischemic tissue patterns. An ablation study reported the contribution of proposed components (U-Net + CW + DS and U-Net + CW + DS + Att), regarding the U-net baseline. In such case, the attention modules report a an average gain of 0.4% and 5.25% from dice score and precision, supporting a major trade-off between precision and recall.

8. CONCLUSIONS AND FUTURE WORK

This work introduced a new autoencoder architecture that integrates a cross-attention mechanism into skip connections while following deep hierarchical supervision to principally segment stroke lesions over MRI scans. Regarding MRI scans, the best input configuration was achieved from the ADC, integrated with TTP and Tmax perfusion parametric maps. From such multimodal integration, the proposed approach properly delineates stroke lesions, showing robustness even to recover tiny and variable shapes. The proposed approach outperforms the state-of-the-art over the ISLES2017 public challenge. Additionally, the proposed cross-attention mechanisms were also adapted to recover segmentations from challenging CT scan inputs. For doing so, an auxiliary class was introduced to force geometry learning over poor contrast images. The proposed approach evidenced competitive results over the public ISLES2018 dataset, showing a remarkable performance for multimodal CT inputs. The attention mechanisms and deep supervision during learning showed to be effective to filter the most relevant features associated with stroke lesions, avoiding the bias associated with a natural imbalance between healthy and affected tissue. The evaluation from two image domains evidences the generalization properties of the proposed approach, which can be further analyzed and adapted to achieve more efficacy on the stroke segmentation task. Future works will include the validation in more challenging scenarios, with complete and raw MRI and CT studies, that allow exploring capabilities of the approaches to search and localize lesions. Also, a more rigorous study will be carried out involving several radiologist experts to analyze possible bias in annotations during learning and validation. Besides, a study of new attention mechanisms will be carried out as an alternative proposal to recover ischemic lesions at different stages (chronic or subacute).

BIBLIOGRAPHY

Abulnaga, S Mazdak and Jonathan Rubin. "Ischemic Stroke Lesion Segmentation in CT Perfusion Scans Using Pyramid Pooling and Focal Loss". In: *Brainlesion: Glioma, Multiple Sclerosis, Stroke and Traumatic Brain Injuries*. Ed. by Alessandro Crimi et al. Cham: Springer International Publishing, 2019, pp. 352–363. DOI: https://doi.org/10.1007/978-3-030-11723-8_36 (cit. on pp. 13, 21).

Bahdanau, Dzmitry, KyungHyun Cho, and Yoshua Bengio. "Neural Machine Translation by jointly learning to align and translate". In: *ICLR* 11.3 (2015), pp. 367–373. arXiv: 1409.0473v7 (cit. on pp. 19, 26).

Baozhou, Zhu et al. "An Attention Module for Convolutional Neural Networks". In: (2021). arXiv: 2108.08205v1 (cit. on p. 19).

Clèrigues, Albert et al. "Acute and sub-acute stroke lesion segmentation from multimodal MRI". In: *Computer Methods and Programs in Biomedicine* 194 (2020). DOI: 10.1016/j.cmpb.2020.105521. arXiv: 1810.13304 (cit. on pp. 14, 21, 41).

— "Acute ischemic stroke lesion core segmentation in CT perfusion images using fully convolutional neural networks". In: *Computers in Biology and Medicine* 115 (2019), p. 103487. DOI: 10.1016/j.compbiomed.2019.103487 (cit. on pp. 14, 21).

Copen, William A., Pamela W. Schaefer, and Ona Wu. *MR Perfusion Imaging in Acute Ischemic Stroke*. Good resource to understand the role of perfusion parametric maps in the assesment of stroke. May 2011. DOI: 10.1016/j.nic.2011.02.007 (cit. on p. 41).

Dolz, Jose, Ismail Ben Ayed, and Christian Desrosiers. "Dense multi-path u-net for ischemic stroke lesion segmentation in multiple image modalities". In: *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)* 11383 LNCS (2019), pp. 271–282. DOI: 10.1007/978-3-030-11723-8_27. arXiv: 1810.07003 (cit. on pp. 13, 21).

Gómez, Santiago et al. "A deep supervised cross-attention strategy for ischemic stroke segmentation in MRI studies". In: *Biomedical Physics & Engineering Express* (in press) (cit. on p. 25).

Gómez, Santiago et al. "An attentional Unet with an auxiliary class learning to support acute ischemic segmentation on CT". In: *SPIE Medical Imaging*. 2023 (in press) (cit. on p. 25).

Heit, Jeremy J, Greg Zaharchuk, and Max Wintermark. "Advanced neuroimaging of acute ischemic stroke: penumbra and collateral assessment". In: *Neuroimaging Clinics* 28.4 (2018), pp. 585–597 (cit. on p. 16).

Hu, Xiaojun et al. "Brain SegNet: 3D local refinement network for brain lesion segmentation". In: *BMC Medical Imaging* 20.1 (2020), pp. 1–10. DOI: 10.1186/s12880-020-0409-2 (cit. on pp. 13, 21, 37, 38, 42).

Kamnitsas, Konstantinos et al. "Efficient multi-scale 3D CNN with fully connected CRF for accurate brain lesion segmentation". In: *Medical Image Analysis* 36 (2017), pp. 61–78. DOI: 10.1016/j.media.2016.10.004. arXiv: 1603.05959 (cit. on pp. 13, 20).

Karmakar, Priyabrata, Shyh Wei Teng, and Guojun Lu. "Thank you for Attention: A survey on Attention-based Artificial Neural Networks for Automatic Speech Recognition". In: (2021), pp. 1–11. arXiv: 2102.07259 (cit. on p. 19).

Liebeskind, David S. "Collateral circulation". In: *Stroke* 34.9 (2003), pp. 2279–2284. DOI: 10.1161/01.STR.0000086465.41263.06 (cit. on p. 15).

Liu, Liangliang et al. "Attention convolutional neural network for accurate segmentation and quantification of lesions in ischemic stroke disease". In: *Medical Image Analysis* 65 (2020). DOI: 10.1016/j.media.2020.101791 (cit. on pp. 13, 14, 22).

Liu, Pengbo. "Stroke lesion segmentation with 2D novel CNN pipeline and novel loss function". In: *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)* 11383 LNCS (2019), pp. 253–262. DOI: 10.1007/978-3-030-11723-8_25 (cit. on p. 40).

Loshchilov, Ilya and Frank Hutter. "Decoupled weight decay regularization". In: *arXiv preprint arXiv:1711.05101* (2017) (cit. on p. 32).

Luong, Minh-Thang, Hieu Pham, and Christopher D Manning. *Effective Approaches to Attention-based Neural Machine Translation*. Tech. rep. 2015, pp. 17–21 (cit. on p. 19).

Martel, Anne L. et al. "Measurement of infarct volume in stroke patients using adaptive segmentation of diffusion weighted MR images". In: *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*. 1999. DOI: 10.1007/10704282_3 (cit. on p. 12).

Ministerio de Salud y Protección Social de Colombia. *Enfermedad cerebrovascular, otra comorbilidad priorizada contra el covid-19*. https://www.minsalud.gov.co/Paginas/Enfermedad-cerebrovascular,-otra-comorbilidad-priorizada-contra-el-covid-19.aspx. Accessed: 2023-01-19 (cit. on p. 12).

Neumann, Anders B. et al. "Interrater agreement for final infarct mri lesion delineation". In: *Stroke* 40.12 (2009), pp. 3768–3771. DOI: 10.1161/STROKEAHA.108.545368 (cit. on p. 12).

Pinto, Adriano et al. “Enhancing Clinical MRI Perfusion Maps with Data-Driven Maps of Complementary Nature for Lesion Outcome Prediction”. In: *Lecture Notes in Computer Science* 2 (2018), pp. 107–115 (cit. on pp. 13, 21, 37, 38, 42).

Ronneberger, Olaf, Philipp Fischer, and Thomas Brox. “U-Net: Convolutional Networks for Biomedical Image Segmentation”. In: *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)* 9351.Cvd (2015), pp. 12–20. DOI: 10.1007/978-3-319-24574-4 (cit. on p. 17).

Roth, Gregory A et al. “Global burden of cardiovascular diseases and risk factors, 1990–2019: update from the GBD 2019 study”. In: *Journal of the American College of Cardiology* 76.25 (2020), pp. 2982–3021 (cit. on p. 12).

Tureckova, Alzbeta and Antonio J. Rodríguez-Sánchez. “ISLES challenge: U-shaped convolution neural network with dilated convolution for 3D stroke lesion segmentation”. In: *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)* 11383 LNCS (2019), pp. 319–327. DOI: 10.1007/978-3-030-11723-8_32 (cit. on pp. 13, 20, 39).

Vaswani, Ashish et al. *Attention Is All You Need*. Tech. rep. 2017, p. 15. arXiv: 1706.03762v5 (cit. on p. 19).

Veličković, Petar et al. *Graph Attention Networks*. Tech. rep. 2018, p. 12. arXiv: 1710.10903v3 (cit. on p. 19).

Von Kummer, R. et al. “Sensitivity and prognostic value of early CT in occlusion of the middle cerebral artery trunk”. In: *American Journal of Neuroradiology* 15.1 (1994), pp. 9–18 (cit. on p. 15).

Wang, Guotai 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 (2020), p. 101787. DOI: 10.1016/j.media.2020.101787. arXiv: 2007.03294 (cit. on pp. 13, 22, 40).

WHO. *World Stroke Organization*. 2021 (cit. on p. 12).

Yang, Xiao. "An Overview of the Attention Mechanisms in Computer Vision". In: *Journal of Physics: Conference Series* 1693.1 (2020). DOI: 10.1088/1742-6596/1693/1/012173 (cit. on p. 19).

APPENDICES

Anexo A. Academic Products

Journals

- S. Gómez, D. Mantilla, E. Rangel, A. Ortiz, D. D Vera, F. Martínez. "A deep supervised cross-attention strategy for ischemic stroke segmentation in MRI studies". Biomedical Physics Engineering Express. 2022.
  Status: Under review.

Conference papers

- G. Garzón, S. Gomez, D. Mantilla, F. Martínez. "A deep CT to MRI unpaired translation that preserve ischemic stroke lesions". 44th Annual International Conference of the IEEE Engineering in Medicine & Biology Society (EMBC). 2022.
  Status: Published.

- S. Gómez, S. Florez, D. Mantilla, P. Camacho, N. Tarazona, F. Martínez. "An attentional unet with an auxiliary class learning to support acute ischemic segmentation on CT". SPIE Medical Imaging. 2023.
  Status: Accepted.

- S. Gómez, B. Valenzuela, D. Mantilla, A. Ortiz, D. D Vera, J. Garcia, F. Martínez. "A multimodal and multi-task representation for ischemic stroke lesion segmentation over DWI sequences". 26th International Conference on Medical Image Computing and Computer Assisted Intervention (MICCAI). 2023.
  Status: Under preparation.

Anexo B. APIS: A paired CT-MRI dataset for ischemic stroke segmentation challenge

During the development of this work, a dataset with paired CT and MRI images of patients diagnosed with ischemic stroke was constructed in collaboration with the BIVL^{2}ab research group and FOSCAL clinic. A challenge around this dataset was accepted for the annual conference International Symposium on Biomedical Imaging (ISBI) of 2023. All the information related to the challenge can be found in the official website in this link: https://bivl2ab.uis.edu.co/challenges/apis

- S. Gómez, D. Mantilla, G. Garzon, E. Rangel, A. Ortiz, D. D Vera, F. Martínez. "APIS: A paired CT-MRI dataset for ischemic stroke segmentation challenge". 20th IEEE International Symposium on Biomedical Imaging (ISBI). 2023.
  Status: Accepted.

Informed Consent for the APIS dataset

APIS Dataset Informed Consent

You request to download *APIS: A Paired CT-MRI Dataset for Ischemic Stroke Segmentation*. This dataset contains non-contrast computed tomographies (NCCT) and apparent diffusion coefficient (ADC) images from patients diagnosed with ischemic stroke. By doing so, you agree to the following terms of use:

Users of this data must abide by the Data Usage Policy and the Creative Commons Attribution 4.0 International License (https://creativecommons.org/licenses/by/4.0/) under which it has been published. Attribution should include the reference to the following paper:

“Santiago Gómez, Sebastian Florez, Daniel Mantilla, Paul Camacho, Nick Tarazona, and Fabio Martínez "An attentional unet with an auxiliary class learning to support acute ischemic segmentation on CT", Proc. SPIE 12033, Medical Imaging 2023: Image Processing, 120322O”

DATA USAGE POLICY

You agree to reference the recommended bibliographic citation(s) in any publication that employs these resources. These references may be updated till 2024, and you agree to receive emails indicating those updates.

This dataset requires the data consumer to comply with access regulations imposed both by law and by the data repository, and to conform to codes of conduct that are generally accepted in higher education and scientific research for the exchange of knowledge and information.

By continuing past this point to the data retrieval process, you signify your agreement to comply with the requirements stated below:

PRIVACY OF RESEARCH SUBJECTS

The Dataset is considered anonymized for every subject as established in Resolution Nº 008430 of October 4, 1993, which establishes the scientific, technical and administrative norms for health research in Colombia. The results of the processing of such information may be disclosed for scientific purposes, through presentations at conferences or publications in scientific journals, with the required protection of subject identities. However, by linking this dataset to other, not de-identified datasets it may be theoretically possible to identify single RESEARCH SUBJECTS of this dataset.

Any intentional identification of a RESEARCH SUBJECT (whether an individual or an organization) or unauthorized disclosure of his or her confidential information violates the PROMISE OF CONFIDENTIALITY given to the providers of the information. Therefore, users of data agree:

- To not use the Dataset, either alone or in concert with any other information, to make any effort to identify or contact individuals who are or may be the sources of the information in the Dataset.
- If User inadvertently receives identifiable information or otherwise identifies a subject, User will promptly notify the BIVL^{2}ab team and follow their instructions. (famarcar@saber.uis.edu.co)
- To follow relevant institutional policies and applicable federal, state, and local laws and regulations (if any) concerning the completion of IRB or ethics review or approval that may be required for the Project.
- To make no use of the identity of any RESEARCH SUBJECT discovered inadvertently, and to advise the BIVL^{2}ab team of any such discovery.

REDISTRIBUTION OF DATA

You agree not to redistribute data or other materials without the written agreement of the BIVL^{2}ab team. When sharing data or other materials in these approved ways, you must include all accompanying files with the data, including terms of use.

PUBLICATION

The User is encouraged to make the results of the Project publicly available with appropriate citations. The User is requested to provide a courtesy copy of all published manuscripts and publications via email (famarcar@saber.uis.edu.co)

NO WARRANTIES.

Any Dataset delivered pursuant to this License are provided as-is. We make no representations and extend no warranties of any kind, either express or implied, concerning the Dataset. We provide no express or implied warranties of merchantability or fitness for a particular purpose, or that the use of the Dataset will not infringe any patent, copyright, trademark, or other proprietary rights of a third party.

ASSUMPTION OF RISK.

Except to the extent prohibited by law, User assumes all responsibility for damages which may arise from its use, storage, disclosure, or disposal of the Dataset. UIS and FOSCAL will not be liable to User for any loss, claim, or demand made by User, or

made against the User by any third party, arising from the use of the Dataset by User, except to the extent permitted by law when caused by the gross negligence or willful misconduct of UIS or FOSCAL. No indemnification or shifting of the costs of defense for any loss, claim, damage, or liability is intended or provided by either party under this License. Nothing contained herein affects any right UIS or FOSCAL may have to seek remedies for the User's violation of this License.

ASSIGNMENT.
Except as otherwise provided in this License, User shall not assign or transfer any rights or obligations hereunder or any part hereof. Any such purported assignment will be void.

RELATIONSHIP.
Nothing in this License shall be considered to create a partnership, employment relationship, work made for hire, or any other type of joint venture.

By signing this form, you agree to these terms of use.

_____

Signature

Full name:	
Institution:	
Country:	