



ACUERDO n.º 253 DE 2023
18 de Julio

1

Por el cual se aprueba el otorgamiento de la distinción “*Trabajo de Grado Meritorio*” al trabajo presentado por SANTIAGO GÓMEZ HERNÁNDEZ, estudiante del programa de Maestría en Ingeniería de Sistemas e Informática

EL CONSEJO ACADÉMICO DE LA UNIVERSIDAD INDUSTRIAL DE SANTANDER,
en uso de sus atribuciones legales, y

CONSIDERANDO,

- a. Que según lo establecido en la Ley 30 de 1992 y el Estatuto General de la Universidad industrial de Santander, aprobado mediante acuerdo del Consejo Superior mediante Acuerdo n.º 166 del 22 de diciembre de 1993, el Consejo Académico es la máxima autoridad académica de la Universidad.
- b. Que el coordinador de los programas de posgrados de la Escuela de Ingeniería de Sistemas e Informática, solicitó al Consejo Académico otorgar la distinción “*Trabajo de Grado Meritorio*” al trabajo de grado titulado “*Segmentation Of Ischemic Stroke Lesions In Radiological Sequences Using Deep Attention Mechanisms*”, elaborado por el estudiante Santiago Gómez Hernández del programa de Maestría en Ingeniería de Sistemas e Informática y dirigido por el profesor Fabio Martínez Carrillo.
- c. Que los evaluadores del trabajo de grado, profesores, Juan Gabriel Castrillón Guzmán y Julio Ernesto Villalón Reina, recomendaron otorgar la distinción “*Trabajo de Grado Meritorio*” al trabajo referido en el literal b), en consideración al cumplimiento de las disposiciones contenidas en el Artículo 110 del Reglamento General de Posgrados, teniendo en cuenta que en su concepto constituye un aporte significativo en el área, concepto que hace parte integral del presente acuerdo.
- d. Que el Consejo Académico, en sesión del 18 de julio de 2023, aprobó la solicitud referida en el considerando b), del presente acto administrativo.

En mérito de lo anterior,

ACUERDA:

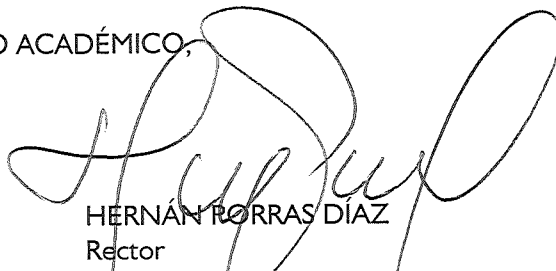
ARTÍCULO 1º. Aprobar la distinción “*Trabajo de Grado Meritorio*” al trabajo titulado “*Segmentation Of Ischemic Stroke Lesions In Radiological Sequences Using Deep Attention Mechanisms*”, elaborado por el estudiante SANTIAGO GÓMEZ HERNÁNDEZ del programa de Maestría en Ingeniería de Sistemas e Informática y dirigido por el profesor Fabio Martínez Carrillo.

ARTÍCULO 2º. Informar sobre el contenido del presente acuerdo a la Dirección de Admisiones y Registro Académico y la Escuela de Ingeniería de Sistemas e Informática, para lo de su competencia.

PUBLÍQUESE, COMUNÍQUESE Y CÚMPLASE

Expedido en Bucaramanga, a los dieciocho (18) días del mes de julio de 2023.

EL PRESIDENTE DEL CONSEJO ACADÉMICO,


HERNÁN PORRAS DÍAZ
Rector


SOFÍA PINZÓN DURÁN

LA SECRETARIA GENERAL,

Bucaramanga, 11 de julio de 2023.

Señores

CONSEJO ACADÉMICO

Universidad Industrial de Santander
Bucaramanga

ASUNTO: Distinción Trabajo de Grado Meritorio para el Trabajo de Investigación Maestría en Ingeniería de Sistemas e Informática.

Cordial saludo:

Por medio de la presente comunicación, me permito enviar la recomendación del jurado evaluador para que se otorgue la distinción **Trabajo de Grado Meritorio** al Trabajo de Investigación titulado “Segmentation Of Ischemic Stroke Lesions In Radiological Sequences Using Deep Attention Mechanisms”; realizado por el estudiante de Maestría en Ingeniería de Sistemas e Informática Santiago Gómez Hernández, con código de estudiante **2218110**. Este Trabajo de investigación cumple con los requisitos establecidos en el Artículo 110 del Acuerdo del Consejo Superior No. 075 de 2013 (Reglamento General de Posgrado). Con respecto al párrafo de este artículo, el estudiante y autor del Trabajo de Investigación inició sus estudios de Maestría en el **primer periodo académico del 2021**, culminó las asignaturas del programa y su Trabajo de Investigación en el primer semestre académico del 2023 siendo este su quinto (5) periodo académico. El estudiante entregó el 27 de febrero de 2023 el informe correspondiente a su Trabajo de Investigación para su respectiva evaluación. El cual recibió la calificación de **APROBADO el 12 de mayo de 2023** por parte del jurado calificador.

Los resultados reportados en el presente Trabajo de Investigación:

- Gómez, S., Florez, S., Mantilla, D., Camacho, P., Tarazona, N. and Martínez, F., 2023, April. An attentional unet with an auxiliary class learning to support acute ischemic stroke segmentation on CT. In Medical Imaging 2023: Image Processing (Vol. 12464, pp. 144-148). SPIE. **Artículo de conferencia internacional.** [Paper](#)
- 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

COORDINACIÓN DE PROGRAMAS DE POSGRADOS
ESCUELA DE INGENIERÍA DE SISTEMAS E INFORMÁTICA

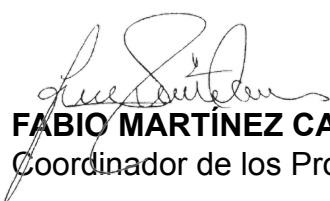
Edificio Facultad de Ingenierías Físico-mecánicas
PBX: (7) 6344000 Ext. 2481 FAX: 6349042 A.A. 678 Bucaramanga, Colombia
Correo-e: maesinfo@uis.edu.co

Express, 9(3), 035026. **Artículo publicado en Revista categoría B según publindex (ISSN 2057-1976).** [Paper](#)

- Garzón, G., Gomez, S., Mantilla, D., & Martínez, F. (2022, July). A deep CT to MRI unpaired translation that preserve ischemic stroke lesions. In 2022 44th Annual International Conference of the IEEE Engineering in Medicine & Biology Society (EMBC) (pp. 2708-2711). IEEE. **Artículo publicado en Revista categoría B según publindex (ISSN 2694-0604).** [Paper](#)

A esta comunicación se adjunta el concepto del jurado calificador.

Atentamente,



FABIO MARTÍNEZ CARRILLO
Coordinador de los Programas de Posgrados

TRABAJO DE GRADO MERITORIO		TRABAJO DE GRADO LAUREADO	
a). Según el concepto del jurado, el trabajo constituya un aporte al campo disciplinar en que se desarrolle el trabajo de grado. <i>Artículo 110 literal a). Reglamento General de Posgrados Acuerdo 075 de 2013 del Consejo Superior.</i> <u>Por favor escribir el aporte científico.</u> El trabajo de grado del estudiante Santiago Gómez Hernández, además de cumplir con los requerimientos necesarios para ser meritorio, resuelve un problema de alta relevancia clínica, la detección áreas afectadas por un infarto cerebral, proponiendo un algoritmo de inteligencia artificial que recopila los aspectos más importantes del estado del arte, proveyendo resultados que superan los reportados en la literatura de manera clara, reproducible y replicable en diferentes bases de datos, impactando de manera positiva el manejo del paciente.		a). Según el concepto del jurado, el trabajo constituya un aporte al campo disciplinar en que se desarrolle el trabajo de grado. <i>Artículo 110 literal a). Reglamento General de Posgrados Acuerdo 075 de 2013 del Consejo Superior.</i> <u>Por favor escribir el aporte científico.</u>	
b). Publicación o aceptación de (1) artículo de su autoría, en revistas indexadas u homologadas por COLCIENCIAS en categoría A o B según la clasificación vigente de PUBLINDEX, de COLCIENCIAS, o en revistas con índice de impacto equivalentes a estas categorías, que contengan expresamente los avances o resultados del trabajo de grado. <i>Artículo 110 literal c). Reglamento General de Posgrados Acuerdo 075 de 2013 del Consejo Superior.</i>	SI ☒ NO ☐	b). Publicación o aceptación de (2) artículo de su autoría, en revistas indexadas u homologadas por COLCIENCIAS en categoría A según la clasificación vigente de PUBLINDEX, de COLCIENCIAS, o en revistas con índice de impacto equivalentes a estas categorías, que contengan expresamente los avances o resultados del trabajo de grado. <i>Artículo 110 literal b). Reglamento General de Posgrados Acuerdo 075 de 2013 del Consejo Superior.</i>	SI ☐ NO ☐
c). Participación con ponencia en, al menos, un (1) evento académico internacional. Esta ponencia debe incluir expresamente los avances o resultados del trabajo de grado. <i>Artículo 110 literal c). Reglamento General de Posgrados Acuerdo 075 de 2013 del Consejo Superior.</i>	SI ☒ NO ☐	c). Participación con ponencia en, al menos, un (1) evento académico internacional. Esta ponencia debe incluir expresamente los avances o resultados del trabajo de grado. <i>Artículo 110 literal c). Reglamento General de Posgrados Acuerdo 075 de 2013 del Consejo Superior.</i>	SI ☐ NO ☐
d). Para acceder a las distinciones de Trabajo de grado laureado y Trabajo meritorio, el estudiante-autor del trabajo de grado debe haber finalizado el programa, en su totalidad, en el tiempo establecido por la universidad para ello. <i>Artículo 139 parágrafo 1. Reglamento General de Posgrados Acuerdo 075 de 2013 del Consejo Superior.</i>	SI ☒ NO ☐	d). Para acceder a las distinciones de Trabajo de grado laureado y Trabajo meritorio, el estudiante-autor del trabajo de grado debe haber finalizado el programa, en su totalidad, en el tiempo establecido por la universidad para ello. <i>Artículo 139 parágrafo 1. Reglamento General de Posgrados Acuerdo 075 de 2013 del Consejo Superior.</i>	SI ☐ NO ☐
Consideraciones, Comentarios y Observaciones:			

NOMBRE: Juan Gabriel Castrillón Guzmán

FIRMA: 

FECHA: 15/05/2023

[illegible]

c). Participación con ponencia en, al menos, un (1) evento académico internacional. Esta ponencia debe incluir expresamente los avances o resultados del trabajo de grado. Artículo 110 literal c). Reglamento General de Posgrados Acuerdo 075 de 2013 del Consejo Superior.	SI ☒ NO ☐	c). Participación con ponencia en, al menos, un (1) evento académico internacional. Esta ponencia debe incluir expresamente los avances o resultados del trabajo de grado. Artículo 110 literal c). Reglamento General de Posgrados Acuerdo 075 de 2013 del Consejo Superior.	SI ☐ NO ☐
d). Para acceder a las distinciones de Trabajo de grado laureado y Trabajo meritorio, el estudiante-autor del trabajo de grado debe haber finalizado el programa, en su totalidad, en el tiempo establecido por la universidad para ello. Artículo 139 parágrafo 1. Reglamento General de Posgrados Acuerdo 075 de 2013 del Consejo Superior.	SI ☒ NO ☐	d). Para acceder a las distinciones de Trabajo de grado laureado y Trabajo meritorio, el estudiante-autor del trabajo de grado debe haber finalizado el programa, en su totalidad, en el tiempo establecido por la universidad para ello. Artículo 139 parágrafo 1. Reglamento General de Posgrados Acuerdo 075 de 2013 del Consejo Superior.	SI ☐ NO ☐
Consideraciones, Comentarios y Observaciones: El trabajo de investigación propuesto supera los alcances académicos en un trabajo de maestría, fundamentados en los productos académicos alcanzados. Además, los objetivos fueron plenamente cumplidos, así como el modelo de segmentación propuesto tiene resultados competitivos en el estado del arte			
NOMBRE: Julio Ernesto Villalón Reina FIRMA:  FECHA: Junio 15 de 2023			

A deep CT to MRI unpaired translation that preserve ischemic stroke lesions

Gustavo Garzón¹, Santiago Gomez¹, Daniel Mantilla² and Fabio Martínez¹

Abstract—Stroke is the second-leading cause of death world around. The immediate attention is key to patient prognosis. Ischemic stroke diagnosis typically involves neuroimaging studies (MRI and CT scans) and clinical protocols to characterize lesions and support decisions about treatment to be administered to the patient. Nowadays, multiparametric MRI images are the standard tool to visualize core and penumbra of ischemic stroke, supporting diagnosis and lesion prognosis. Specially, DWI modality (Diffusion Weighted Imaging) allows to quantify the cellular density of the tissue, and therefore allowing to quantify the lesion aggressiveness, and the recognition of micro-circulation properties. Nevertheless, MRI availability at hospitals is not widespread, and acquisition require special conditions requiring considerable time. Contrary, CT scans commonly have major availability but brain structures are poorly delineated, and even worse, ischemic lesions are only visible at advanced stages of the disease. This work introduces a deep generative strategy that allows ischemic stroke lesion translation over synthetic DWI-MRI images. This encoder-decoder architecture, include U-net modules, hierarchically organized, with inter-level connections that preserve brain structures, while codifying an embedding representation. Then a cyclic loss was here implemented to receive CT inputs and decode DWI-MRI images. To avoid mode collapse, this learning is inversely propagated, *i.e.*, from synthetic DWI-MRI images to original CT-scans. Finally, an embedding projection is recovered to show a proper lesion-slice discrimination, regarding control studies.

Clinical relevance— To recover synthetic DWI-MRI that preserved ischemic lesion using CT scans as an input and following an unpaired image translation setup.

I. INTRODUCTION

Stroke is the second-leading cause of death and the third-leading cause of death and disability combined [6]. Immediate treatment of ischemic stroke events is key in order to preserve as much healthy brain tissue as possible [4]. Among distinct factors, imaging scanner availability is critic in order to diagnose ischemic lesions fast and accurately. Particularly, DWI-MRI imaging properly display ischemic penumbra and ischemic core [3]. Nevertheless, MRI scanner availability is scarce in most clinics, or its access is restricted by many other triage-based events [5]. CT scanners, on the other hand, have major availability and the imaging acquisition time is significantly faster than MRI's. Besides, these sequences has

no contraindication with respect to metallic objects and are less prone to motion artifacts [3].

To tackle scarcity of MRI scanners, currently deep learning strategies have implemented image-to-image translation, but commonly over sophisticated studies that include paired CT and MRI data. For instance, Armanious et al. [1] translated PET to CT sequences by introducing a Cycle-MedGAN framework, in an unsupervised setup. This method minimize textural and perceptual inconsistencies in the translated images. Similarly, Tmenova et al. [12] used a CycleGAN network for generating realistic angiograms using synthetic inputs. In a posterior work, Armanious et al. [2] proposed a method that implements a CasNet generator, a pre-trained VGG-19 feature extractor and a classic PatchGAN discriminator in order to translate PET-to-CT, correction of MRI motion artifacts and PET image denoising. Nevertheless, these strategies depend on paired data and require data registration, which result unrealistic on clinical scenarios.

In image analysis literature, recent methods are dedicated to translation applications from unpaired data. Wolterink et al. [13] introduced an MRI to CT translation strategy employing a CycleGAN architecture for translating MRI to CT images. Nevertheless, such method rely heavily in rigid registration methods and paired images for calculating translation quality. Similarly, Yang et al. [14] proposed a structure-constrained CycleGAN method that employed a modality independent neighborhood descriptor (MIND) to include local image structure into the loss function. This method is also dependent of paired data to evaluate performance. Prokopenko et al. [9] introduced a framework that translates MRI sequences to CT using a DualGAN architecture, a pre-trained VGG-16 model and a Super-Resolution GAN. However, this strategy depends on pre and post processing of image sharpness, and it is conditioned to the existence of CT and MRI ground truth images. These strategies are capable of translating medical images but, nevertheless, have not been validated in an specific domain such as preservation of ischemic lesion over synthetic MRI images.

This work proposed a non-aligned deep representation that allowed to obtain synthetic DWI-MRI sequences from CT scans, focusing on preservation of brain structure and ischemic lesion regions. In the proposed architecture is defined an encoder-decoder framework, from stacked U-net units, that embedded brain information and preserved main morphological structures, among these, the lesion when is present in the studies. The learning of such architecture is carried out from a cycle consistency loss, allowing to

¹ Biomedical Imaging, Vision and Learning Laboratory (BIVL²ab). Universidad Industrial de Santander (UIS), Bucaramanga, Colombia. gustavo.garzon@saber.uis.edu.co, santiago.gomez1@saber.uis.edu.co, famarcas@saber.uis.edu.co

² Clínica FOSCAL, Bucaramanga, Colombia. dmantilla528@unab.edu.co

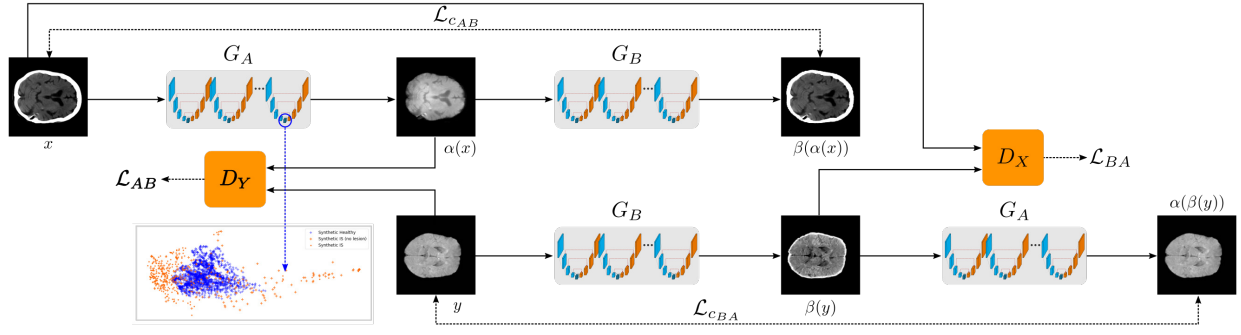


Fig. 1. Pipeline of the proposed approach: input CT image x is translated to the MRI domain $\alpha(x)$ and then translated back to CT domain, obtaining $\beta(\alpha(x))$ and a Cycle consistency loss $\mathcal{L}_{CAB}$ value. For illustration purposes, generators G_A and G_B exhibit our proposed concatenation of n consecutive U-blocks.

generate synthetic sequences in both sides, and avoiding translation bias but also image collapse. From learned representations are captured embedded descriptors that allowed to study discrimination and retrieval of hidden stroke lesions. Also, the recovered synthetic MRI images contained latent ischemic stroke features, which constitute a contribution in clinical scenarios with low capability to carry out MRI studies.

II. PROPOSED APPROACH

Even though simple CT scans are able to detect bleeding in the brain parenchyma, acute ischemic stroke exhibit distinct early signs. On the other hand, DWI-MRI sequences show acute ischemic changes in the brain parenchyma in 80–90% of cases [4]. The proposed approach is able to perform image-to-image translation without depending on paired data, with the possibility to obtain DWI-MRI images from simple CT scans. The pipeline of the proposed approach is illustrated in figure 1

A. Non-aligned translation

This work proposed an encoder-decoder deep representation, adjusted from adversarial and cycle consistency losses. Like a CycleGAN, this net allow to translate images from two different domains in an unpaired setup [15]. Firstly, it is learning a mapping function between domains X_{CT} (CT sequences) and Y_{MRI} (MRI sequences). Such mapping involves two integrated generators: α , and β . Generator α attempts to translate CT to MRI sequences, but this one-way estimation might generate undesirable results such as distorted or anatomically inaccurate synthetic images. Therefore, a generator function β tackles this issue by transforming MRI to CT sequences, and a discriminator function tries to discriminate between synthetic and real MRI sequences. In order to establish these restrictions, a cycle consistency loss enforces that generated images must be transformed into the initial projection space, that is:

$$\mathcal{L}_c(\alpha, \beta) = \mathbb{E}_{X \sim p(X_{CT})} \|\beta(\alpha(X_{CT})) - X_{CT}\|_1 + \mathbb{E}_{Y \sim p(Y_{MRI})} \|\alpha(\beta(Y_{MRI})) - Y_{MRI}\|_1$$

Also, note that $\beta(\alpha(X_{CT})) \approx X_{CT}$ and $\alpha(\beta(Y_{MRI})) \approx Y_{MRI}$ enforce that synthetic images remain in data distributions X_{CT} and Y_{MRI} respectively. It is worthy noting that this work is not aiming for high quality image translations, but for validating preservation of ischemic lesion over synthetic MRI images.

B. Stacked U-net CycleGAN generator

Ischemic lesion in medical images contain fine-grained characteristics, anatomical structures and intensity variations inherent to clinical significant findings, that a standard autoencoder may fail to preserve. On the other hand, U-net architecture [10] overcome autoencoder disadvantages by capturing image context in a classic convolutional contracting path and precisely localizing features in an expanding path. By using skip connections, U-net architecture is able to preserve structural details and texture, and also achieves a significant alleviation of vanishing gradient events.

Inspired by the work of Armanious et al. [2], we here implement image features of U-nets, and boosted it by stacking U-net modules (U-blocks) [11]. This allowed to combine features from distinct resolution scales without losing image resolution at the output layer. Thereafter, we modified CycleGAN generators in such a way that they were replaced by a concatenation of n U-blocks. Such blocks, continuously refined the generated images, improving classic autoencoder limitations.

C. Ischemic lesion embedding space

Measuring the lesion translation capability, from unpaired sequences, is a challenging task. Even for experts, the lesions observed in CT should remain at chronic stages to carry out a proper annotation and characterization. As an alternative to measure such lesion preservation, we built a topological and geometrical space in which synthetic slices are projected as embedding points, and its position will be associated to the presence of ischemic lesion features. The embedding vectors are taken from the bottleneck of encoder-decoder subnet, whose principal components are visualized in a low dimensional space. We hypothesized that a representation of these components might contain characteristics proper from ischemic lesions that, even if they are not clearly visible in

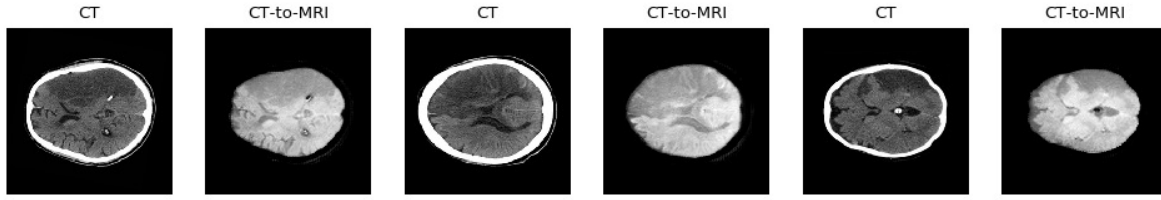


Fig. 2. CT slices with visible ischemic lesion (1st, 3rd and 5th columns) and its corresponding synthetic MRI images (2nd, 4th and 6th columns). It is clearly evidenced that our method qualitatively preserves ischemic lesion positions in translated MRI slices.

the original CT slice, they exist and will differ from features of synthetic MRI slices of control patients.

In such space, the latent features that correspond to each projected point has coded complex visual relationships, which may be related to the existence of an ischemic lesion. In order to address this hypothesis, we located the bottleneck layer in the **last** U-block of generator G_A and used it to calculate an embedding representation. Such approach maps each synthetic MRI slice to a point in the embedding space and, samples tend to group according to the existence of ischemic lesion into each slice.

D. Experimental setup

1) *Datasets*: The proposed framework was validated according to the following datasets. For training were used two unpaired sets of CT and MRI images, coming from two different challenges. In such sense, we employed **ISLES 2017** [7] which consists of MRI Diffusion Weighted Imaging sequences and perfusion maps, with a ground truth segmentation mask drawn over T2/FLAIR images. Complementary, for training we also used **ISLES 2018** [7] dataset which consists of NCCT and CT Perfusion maps, with a ground truth segmentation mask drawn over DWI-MRI images. For both datasets, segmentation masks were drawn after the stroke lesion had stabilised. For validation purposes we considered different subset images. Firstly, we use the **iDB-MRXFDG database** [8] that consists of CT and MRI images from normal adult patients, only employing axial CT and MRI sequences. Images of all three datasets were resized to 256x256 pixels and were converted to grayscale, according to radiological range.

Our proposed unsupervised approach contains a training set comprising CT slices (train A) and MRI slices (train B). Additionally, a testing set composed by a different set of CT slices (test A) and MRI slices (test B). Such slices were distributed following default ISLES 2017 (MRI sequences) and ISLES 2018 (CT sequences) guidelines for training and testing. On testing stage, iDB-MRXFDG slices were mapped alongside test A images in order to differentiate them. As for second experiment, iDB-MRXFDG database was shuffled and training was carried out with 80% of slices, distributed accordingly to train A, train B, test A and test B. As for this testing stage, ISLES 2018 test A subset was mapped alongside, in order to validate ischemic and healthy slices discrimination.

TABLE I
INFLUENCE OF NUMBER OF U-BLOCKS OVER GENERATOR,
DISCRIMINATOR AND RECONSTRUCTION LOSS OVER SYNTHETIC MRI.

U-blocks	Gen. loss	Disc. loss	Rec. loss
1	0.9429	0.6194	0.0407
2	1.0775	0.4809	0.0343
3	8.6516	9.547637e-07	0.1072

2) *Architecture configuration*: Our experiments with baseline CycleGAN and the proposed strategy were designed with the following features: learning rate of 0.0002, three generator iterations for each discriminator and a batchsize of 10. The convergence of baseline translation scheme requires 200 epochs, while the proposed approach converges in 10 epochs. Regarding hidden representation, the synthetic MRI slices were mapped to a 512-dimension embedding space, where its dimensionality was subsequently reduced to the two principal components.

III. EVALUATION AND RESULTS

To evaluate the influence of the amount of U-block units inside generators, we evaluated generator, discriminator and reconstruction losses for $\{1, 2, 3\}$ U-blocks. Discriminator loss directly indicates quality of obtained synthetic images, and reconstruction loss shows image translation performance by measuring how close $\beta(\alpha(x))$ and x are. Both losses obtained the best results for 2 U-blocks as shown in table I.

We must evaluate anatomical accuracy and tissue textures in the image translation process. So, qualitative results of synthetic images must be taken into account for selecting a balanced experimental setup: as shown in figure 3, implementing 2 U-blocks generated synthetic MRI slices with the best trade-off between anatomical accuracy and overall image quality. On the other hand, using only 1 U-block resulted insufficient, producing regions with air in incorrect anatomical positions and wrong intensities. As for 3 U-blocks, it produces an excessive refining that blurs anatomical details and generated wrong bone boundaries. Furthermore, the proposed approach was tested using CT slices with chronic ischemic stroke lesions as input data. The figure 2 shows the recovered synthetic MRI images that clearly preserved ischemic lesion and obtained in average a dice score of 95%.

To validate ischemic lesion from the projection into a embedding space, synthetic MRI slices were projected in a low-dimensional space. Figure 4 shows a representation of

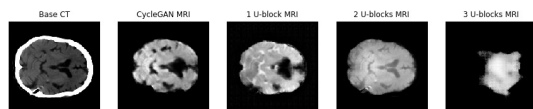


Fig. 3. First column: base CT slice. Synthetic MRI slices generated with: CycleGAN (2nd column), 1 U-block (3rd column), 2 U-block (4th column) and 3 U-block (5th column).

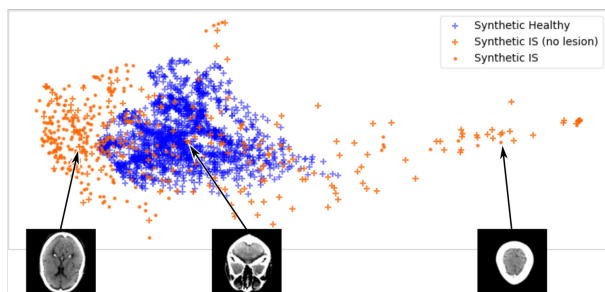


Fig. 4. Embedding space representation for slices containing ischemic lesion (orange dots), without ischemic lesion (orange plus) and slices of healthy patients from iDB-MRXFDG database (blue).

synthetic MRI slices with and without ischemic lesions. Note that right side outliers (orange) correspond to regions where ischemic stroke lesions are infrequent. Further, ischemic and non-ischemic slices follow independent anatomical distributions, which is a promising indicator of preserving lesion and discrimination with respect to control patient slices. Also, notice that healthy patients (blue) form a compact group that is mostly separated from ischemic patients.

As an extra validation, the proposed model was trained with iDB-MRXFDG database as an input for CT images domain, and MRI images domain. This validation followed an anomaly-detection scheme, where, in testing stage, embeddings of synthetic MRI slices of ischemic stroke patients were mapped into the same space. Figure 5 shows a noticeable separation between mappings of ischemic and healthy synthetic slices.

IV. CONCLUSIONS

The proposed approach introduced a novel generative representation that allowed to preserve ischemic lesions in synthetic MRI images with an unpaired and unregistered setup. Evenmore, we introduced a strategy that support acute ischemic stroke diagnosis by discriminating slices of affected

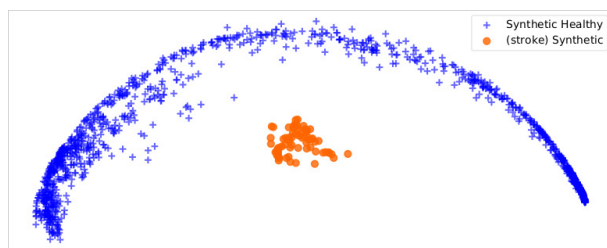


Fig. 5. Embedding space representation for slices of healthy patients from iDB-MRXFDG database (blue) and slices containing ischemic lesions (orange).

brain regions from healthy ones. In this work was explored a embedding space representation that allowed to identify slices that have anomalous structures, but also might be determinant to explore hidden anomalous patterns. Future work will involve a more exhaustive study with a larger dataset that include clinical variables to correlate such information into the embedding space, to obtain a more robust tool for supporting diagnosis of acute ischemic stroke.

ACKNOWLEDGMENT

The authors acknowledge the Vicerrectoría de Investigación y Extensión (VIE) of the Universidad Industrial de Santander for supporting this research work by the project: *Cuantificación de patrones locomotores para el diagnóstico y seguimiento remoto en zonas de difícil acceso*, with SIVIE code 1763.

COMPLIANCE WITH ETHICAL STANDARDS

This research study was conducted retrospectively using human subject data made available at Ischemic Stroke Lesion Segmentation (ISLES) 2018. Ethical approval was not required.

REFERENCES

- [1] Karim Armanious and et al. Unsupervised medical image translation using cycle-medgan. In *EUSIPCO*, pages 1–5. IEEE, 2019.
- [2] Karim Armanious and et al. Medgan: Medical image translation using gans. *Computerized Medical Imaging and Graphics*, 79:101684, 2020.
- [3] Bruce CV Campbell and et al. Imaging selection for acute stroke intervention. *International Journal of Stroke*, 13(6):554–567, 2018.
- [4] Anders et al Christensen. imaging in acute stroke—new options and state of the art. *Frontiers in neurology*, 8:736, 2018.
- [5] Imad Derraz and et al. Flair vascular hyperintensities and functional outcome in nonagenarians with anterior circulation large-vessel ischemic stroke treated with endovascular thrombectomy. *European Radiology*, pages 1–11, 2021.
- [6] Valery L Feigin and et al. Global, regional, and national burden of stroke and its risk factors, 1990–2019: a systematic analysis for the global burden of disease study 2019. *The Lancet Neurology*, 20(10):795–820, 2021.
- [7] Oskar Maier and et al. Isles 2015-a public evaluation benchmark for ischemic stroke lesion segmentation from multispectral mri. *Medical image analysis*, 35:250–269, 2017.
- [8] Inés Mérida and et al. Cermep-idb-mrxfdg: A database of 37 normal adult human brain [18f] fdg pet, t1 and flair mri, and ct images available for research. *EJNMMI research*, 11(1):1–10, 2021.
- [9] Denis Prokopenko and et al. Unpaired synthetic image generation in radiology using gans. In *Workshop on Artificial Intelligence in Radiation Therapy*, pages 94–101. Springer, 2019.
- [10] Olaf Ronneberger and et al. U-net: Convolutional networks for biomedical image segmentation. In *MICCAI*, pages 234–241. Springer, 2015.
- [11] Sohil Shah and et al. Stacked u-nets: a no-frills approach to natural image segmentation. *arXiv preprint arXiv:1804.10343*, 2018.
- [12] Oleksandra Tmenova and et al. Cyclegan for style transfer in x-ray angiography. *International journal of computer assisted radiology and surgery*, 14(10):1785–1794, 2019.
- [13] Jelmer M Wolterink and et al. Deep mr to ct synthesis using unpaired data. In *International workshop on simulation and synthesis in medical imaging*, pages 14–23. Springer, 2017.
- [14] Heran Yang and et al. Unpaired brain mr-to-ct synthesis using a structure-constrained cyclegan. In *Deep Learning in Medical Image Analysis and Multimodal Learning for Clinical Decision Support*, pages 174–182. Springer, 2018.
- [15] Jun-Yan Zhu and et al. Unpaired image-to-image translation using cycle-consistent adversarial networks. In *ICCV*, pages 2223–2232, 2017.

PAPER

A deep supervised cross-attention strategy for ischemic stroke segmentation in MRI studies

To cite this article: Santiago Gómez *et al* 2023 *Biomed. Phys. Eng. Express* **9** 035026

View the [article online](#) for updates and enhancements.

You may also like

- [Attention-enabled 3D boosted convolutional neural networks for semantic CT segmentation using deep supervision](#)
Vasant Kearney, Jason W Chan, Tianqi Wang et al.
- [Thyroid gland delineation in noncontrast-enhanced CTs using deep convolutional neural networks](#)
Xiuxiu He, Bang Jun Guo, Yang Lei et al.
- [Multi-scale and local feature guidance network for corneal nerve fiber segmentation](#)
Wei Tang, Xinjian Chen, Jin Yuan et al.

Biomedical Physics & Engineering Express



PAPER

A deep supervised cross-attention strategy for ischemic stroke segmentation in MRI studies

Santiago Gómez¹, Daniel Mantilla², Edgar Rangel¹ , Andrés Ortiz², Daniela D Vera² and Fabio Martínez¹

¹ Biomedical Imaging and Learning Laboratory (BIVL²ab), Universidad Industrial de Santander, Bucaramanga, 680002, Colombia

² Clínica FOSCAL, Floridablanca, 681004, Colombia

E-mail: famarcar@saber.uis.edu.co

Keywords: attention mechanisms, medical image segmentation, stroke, imbalanced problems

Abstract

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. The main purpose of this study is to introduce an autoencoder architecture that effectively integrates cross-attention mechanisms, together with hierarchical deep supervision to delineate lesions under scenarios of remarked unbalance tissue classes, challenging geometry of the shape, and a variable textural representation. This work introduces a cross-attention deep autoencoder that focuses 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 herein 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 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. Deeply supervised cross-attention autoencoders, trained to pay more attention to lesion tissue, are better at estimating ischemic lesions in MRI studies. The best architectural configuration was achieved by integrating ADC, TTP and Tmax sequences. The contribution of deeply supervised cross-attention autoencoders allows better support the discrimination between healthy and lesion regions, which in consequence results in favorable prognosis and follow-up of patients.

1. Introduction

Strokes are 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, strokes have an associated high morbidity risk. The most common condition is ischemic stroke (80% of all cases), which is caused by blood vessel occlusion. The immediate localization, delineation and quantification of the ischemic lesion, over CT and MRI sequences (according to availability), result fundamental to determine the diagnosis and consequent clinical intervention. This task is however tedious, time-consuming (around 15 minutes each case) and prone to errors, introducing and inherent radiologist bias [3]. In fact, a low inter-rater expert agreement has been reported, being more

consistent with the annotation in FLAIR sequences than in T2 [4].

Specifically, multi-parametric MRI (mp-MRI) is key to detecting early ischemic stroke and assessing the penumbra regions. The mp-MRI includes diffusion-weighted imaging (DWI), the apparent diffusion coefficient (ADC) that approximates the cellular density of the tissues as a magnitude of diffusion of water particles. Also, mp-MRI includes perfusion maps (3D +time volumes) that summarize the responses of a contrast agent. Specifically, time-density curves (TDC) measure wash-in and wash-out process, while the TDA summarizes TDC from a temporal deconvolution algorithm. The AIF parameter measures the arterial input function from five parameter maps: relative Cerebral Blood Volume (rCBV), relative Cerebral Blood Flow (rCBF), Mean Transit Time (MTT), Time to Max (Tmax), and Time to Peak (TTP). TTP, MTT,

and Tmax are commonly used to identify tissue with critical hypoperfusion, while rCBV and rCBF are used to characterize blood flow information.

Deep autoencoder representations have emerged as the most promising tools to support stroke lesion segmentation, discovering, among others, textural differences between affected and healthy regions [5, 6]. A first group of approaches follow 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 [7–12]. In fact, a main issue on 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 have dealt with such imbalance problems using mini-batches of image patches, which are carefully designed to include balanced and stratified information during training [13, 14]. Such strategies, nonetheless, impose local processing, losing global geometry observations, which may be key to properly recover lesions' shape. More recently, these architectures have included attention modules to highlight important local features, but in a global context [5].

This work proposed a deep supervised cross-attention strategy that segments ischemic stroke lesions using a deep autoencoder architecture. The attention modules exploit cross similarities between encoder and decoder observations to improve the decoder representation and preserve 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. 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 maps derived from perfusion-weighted studies.

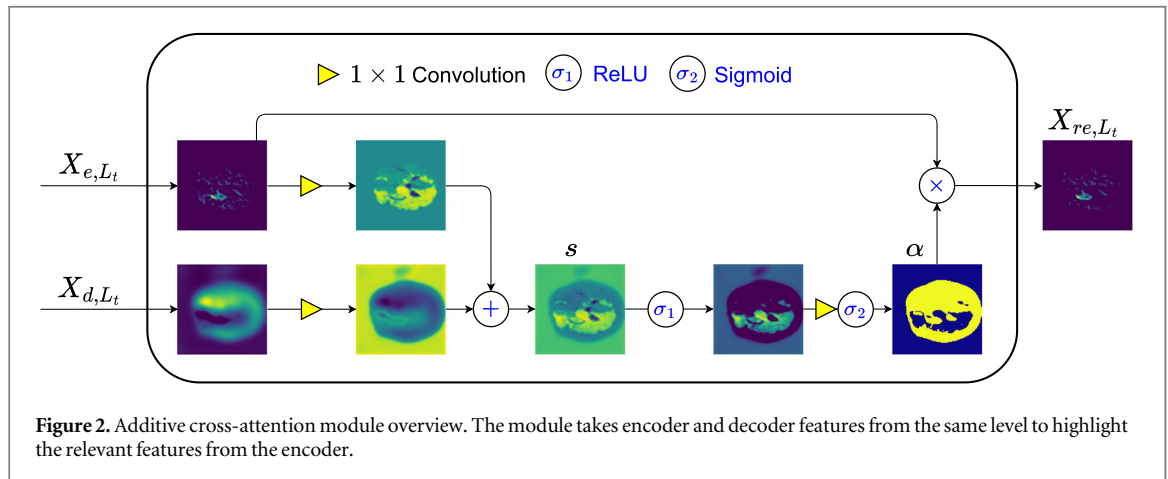
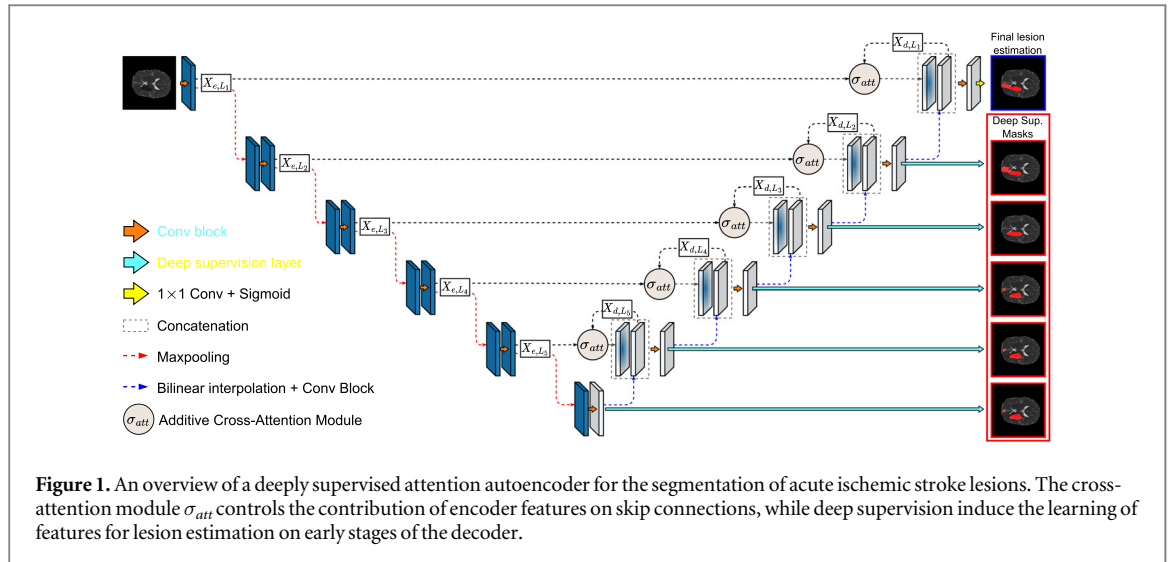
2. Current work

Regarding stroke segmentation support, 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* [7] 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* [8] used a 3D U-Net with dilated convolutions to integrate larger context into the shape lesion modeling. Similarly, Abulnaga *et al* [9] 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 intrinsic lesions characteristics.

Other works have included information from multiple modalities into independent convolutional paths. For instance, Pinto *et al* [10] 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* [11] 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* [12] 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, that may collapse training schemes.

An alternative group of strategies have 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 [13] and a 3D [14] 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 deals with class imbalance, such strategies impose local processing, losing global geometry observations, which may be key to properly recover 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* [5] proposed a U-Net architecture that includes receptive field convolutions, and skip self-attention modules to recover lesion segmentations. Such work deals with the recovering 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* [6] 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.

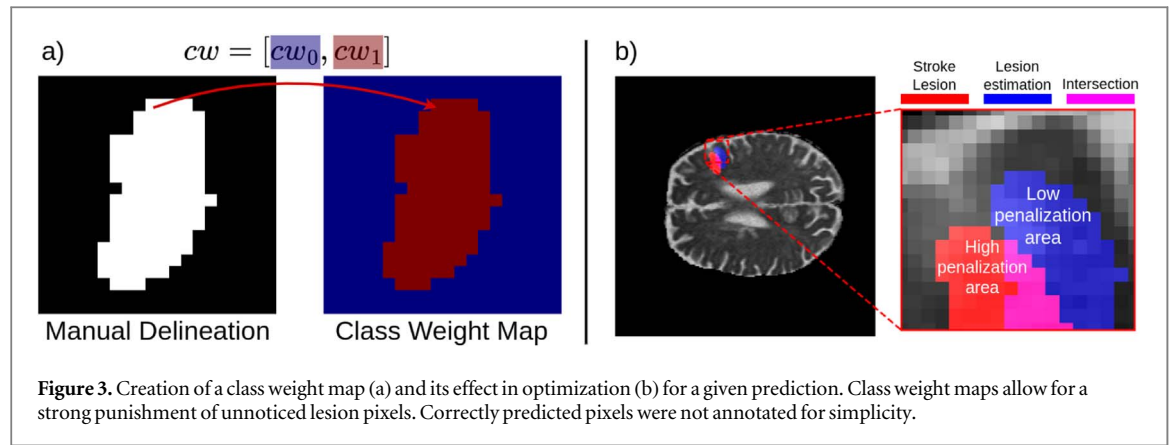
3. Proposed approach

This work introduces 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 a 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 1.

3.1. Additive cross-attention mechanisms

In particular, in stroke segmentation, an average lesion only corresponds to less than 1% of the total volume of

the study (from an average analysis on ISLES2017 training cases was computed a stroke index of 0.56%). To alleviate this effect, we propose the integration of cross-attention modules, at each scale of representation. The attention mechanism 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 cross-attention 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. Particularly, the computation of the attention module follows an additive alignment of query and key features, derived from linear projections of X_{d,L_t} and X_{e,L_t} , using convolution branches [15] (see an illustration in figure 2). This additive alignment can be expressed as: $S = \sigma_1(W_{e,L_t}^T X_{e,L_t} + W_{d,L_t}^T X_{d,L_t})$, where σ_1 is a ReLU activation. Hence, 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. Finally, 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 contribution in the skip connections of classical autoencoders.

3.2. Deep supervision

To enhance ischemic stroke lesion segmentation, we performed a hierarchical deep supervision. 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 (see in figure 1). 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, lesion estimations are used to compute a loss function that considers hierarchical lesion information. The final loss signal is calculated as the weighted sum of each loss term ($\mathcal{L}_i$): $\mathcal{L}_{total} = \sum_{i=1}^k \mathcal{W}_i \mathcal{L}_i$, where k is the number of lesion estimations and $\mathcal{W}_i$ is the corresponding weight for layer L_i that define output segmentation importance at each level. In such cases, the layers near to bottleneck have less contribution due to the spatial details of brain reconstruction at this level. For implementation purposes, the $\mathcal{W}_i$ is a hyperparameter indicating the importance at each level estimation.

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.

3.3. Class weight maps

These special maps affect the loss function score for the classification of each pixel ℓ_{ij} , giving more

importance to lesion pixels. Class weight maps C_{ij} are generated from ground truth segmentations (Y), resulting in weight maps with the same shape as the ground truth segmentation. Each pixel in the weight map is sampled from a class importance vector (cw) based on the class in the ground truth segmentation. These maps are easily integrated with deep supervision models as: $\mathcal{L}(Y, \hat{Y}, C) = 1/(N \cdot M) \sum_{i=1}^N \sum_{j=1}^M C_{ij} \cdot \ell_{ij}(Y, \hat{Y})$, where C_{ij} is the particular weight for the pixel at position (i, j) that distributes a local penalization (see figure 3) of the $\ell_{ij}(Y, \hat{Y})$ loss function at each position and N and M are the height and width of the MRI slice. Particularly, in this work we implement ℓ_{ij} as Focal loss function, defined as: $\mathcal{L} = -(1 - \hat{Y})^\gamma \log(\hat{Y})C$, where γ is a focusing parameter.

The incorporation of class weights maps circumvents common training pitfalls, where all voxels are equally important. More specifically, prevents models from stinking 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 produce more precise estimations.

4. Experimental setup

4.1. 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 ($n = 9$).

4.2. The model

The architecture of the proposed approach was implemented with a convolutional backbone that follows blocks of a 3×3 convolution, a batch normalization and a ReLU activation twice in sequence. 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 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 $\gamma = 2$, and an AdamW optimizer [16]. 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 weights $\mathcal{W}_i$ for each layer were fixed for respective layers as: $\{0.03, 0.045, 0.05, 0.125, 0.25, 0.5\}$. 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. Evaluation and results

An ablation study was conducted to measure the contribution of each special mechanism of the proposed approach. For doing so, we trained standard U-Net architectures with multiple combinations of the proposed approach components. These architectures were compared to the complete architecture that

includes class-weight maps (CW), deep supervision training (DS) and additive cross-attention maps (Att).

In table 1 are summarized the results for all defined configurations, using independent MRI maps. The stroke segmentation was measured with respect to the validation dice scores. The combination of different components is defined in the left part of the table. The ADC (0.45 ± 0.16) achieves the best score without extra requirements of class-weights configuration. It should be noted that whole parameters achieves coherent and robust results with all components of the proposed strategy, without significant differences with respect to other component configurations. The class weights show the major percentage contribution for some parameters, which in general have less textural differences regarding healthy tissue, demonstrating the potential to overcome class imbalance problems. In fact, 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 the 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.

In a second experiment has measured the capability to exploit multi-context information. In such a case, an exhaustive validation was carried out with different architecture components and used as inputs the combinations of the three best modalities from the previous experiment. 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. Table 2 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 score of 0.46 ± 0.14 . Moreover, table 2 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 an increasing score regarding the segmentation of ischemic lesions. The model with all available images gives worse results (0.42 ± 0.22), a fact associated with

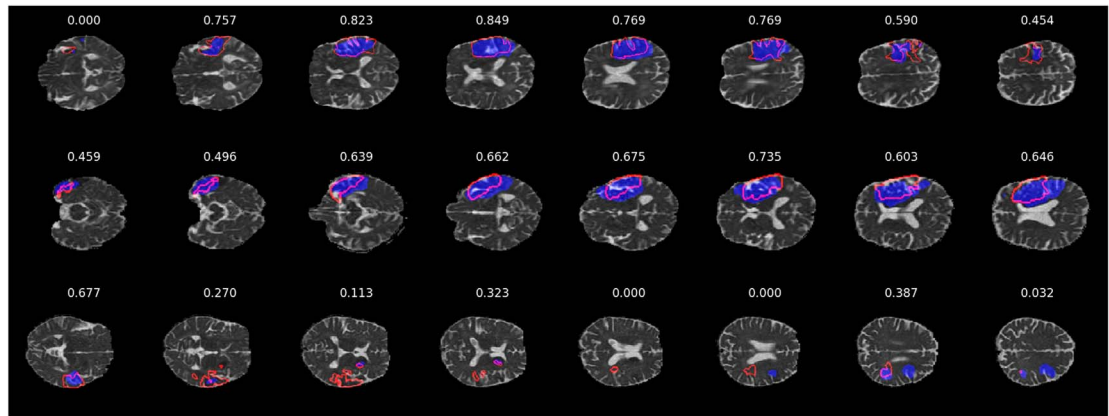


Figure 4. Ischemic stroke lesion predictions for the best multi-parametric attentional model on patients from the validation split. The red contours are the physician manual delineations and the blue areas are the model estimations.

Table 1. Isolated modalities validation results in the ISLES2017 dataset for different configurations of the proposed approach.

Model config			Modalities					
CW	DS	Att	ADC	rCBV	rCBF	TTP	Tmax	MTT
×	×	×	0.45 ± 0.16	0.12 ± 0.15	0.31 ± 0.24	0.30 ± 0.21	0.38 ± 0.20	0.34 ± 0.19
✓	×	×	0.44 ± 0.18	0.18 ± 0.14	0.34 ± 0.24	0.32 ± 0.22	0.36 ± 0.20	0.35 ± 0.20
×	✓	×	0.44 ± 0.18	0.22 ± 0.23	0.29 ± 0.24	0.32 ± 0.22	0.33 ± 0.25	0.36 ± 0.23
×	×	✓	0.42 ± 0.17	0.18 ± 0.18	0.28 ± 0.23	0.31 ± 0.24	0.38 ± 0.19	0.32 ± 0.21
✓	✓	×	0.43 ± 0.18	0.20 ± 0.21	0.32 ± 0.25	0.34 ± 0.20	0.40 ± 0.21	0.36 ± 0.25
✓	×	✓	0.43 ± 0.16	0.18 ± 0.18	0.32 ± 0.22	0.33 ± 0.22	0.39 ± 0.22	0.36 ± 0.23
×	✓	✓	0.45 ± 0.16	0.18 ± 0.17	0.31 ± 0.22	0.33 ± 0.22	0.36 ± 0.20	0.34 ± 0.23
✓	✓	✓	0.44 ± 0.17	0.21 ± 0.22	0.32 ± 0.25	0.36 ± 0.21	0.40 ± 0.23	0.38 ± 0.24

Table 2. Multi-modal validation results in the ISLES2017 dataset for different configurations of the proposed approach.

Model config			Modalities				
CW	DS	Att	ADC + TTP	ADC + Tmax	TTP + Tmax	ADC + TTP + Tmax	All
×	×	×	0.44 ± 0.15	0.43 ± 0.15	0.37 ± 0.19	0.43 ± 0.15	0.40 ± 0.21
✓	×	×	0.44 ± 0.15	0.43 ± 0.16	0.39 ± 0.23	0.44 ± 0.17	0.41 ± 0.22
×	✓	×	0.45 ± 0.19	0.44 ± 0.19	0.37 ± 0.22	0.42 ± 0.20	0.43 ± 0.22
×	×	✓	0.44 ± 0.17	0.42 ± 0.18	0.37 ± 0.21	0.41 ± 0.20	0.42 ± 0.19
✓	✓	×	0.45 ± 0.14	0.45 ± 0.12	0.40 ± 0.21	0.46 ± 0.15	0.42 ± 0.21
✓	×	✓	0.45 ± 0.18	0.43 ± 0.16	0.39 ± 0.22	0.43 ± 0.13	0.42 ± 0.20
×	✓	✓	0.45 ± 0.15	0.42 ± 0.17	0.38 ± 0.21	0.45 ± 0.17	0.43 ± 0.21
✓	✓	✓	0.45 ± 0.13	0.45 ± 0.13	0.41 ± 0.20	0.46 ± 0.14	0.42 ± 0.22

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.

As observational analysis, figure 4 presents a set of segmentation samples (blue areas) obtained with the best multi-parametric model (ADC + TTP + Tmax), and the respective ground truth that correspond to a manual expert annotation (red contour). The first two rows, that correspond to two different patients,

illustrate a 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 at 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,

Table 3. Results for the ISLES2017 hidden test split. All metrics are given in mean $\pm$ standard deviation and best mean values are bolded.

Approach	Dice $\uparrow$	Precision $\uparrow$	Recall $\uparrow$
ISLES2017 Official report [17]			
- Mok <i>et al</i> Ensemble	0.32 $\pm$ 0.23	0.34 $\pm$ 0.27	0.39 $\pm$ 0.27
- Kwon <i>et al</i> Ensemble	0.31 $\pm$ 0.23	0.36 $\pm$ 0.27	0.45 $\pm$ 0.30
- Monteiro <i>et al</i> Single Model	0.30 $\pm$ 0.22	0.34 $\pm$ 0.27	0.51 $\pm$ 0.30
- Lucas <i>et al</i> Single Model	0.29 $\pm$ 0.21	0.34 $\pm$ 0.26	0.51 $\pm$ 0.32
Hu <i>et al</i> [11]			
- Brain SegNet	0.26 $\pm$ 0.22	0.35 $\pm$ 0.28	0.38 $\pm$ 0.29
- Brain SegNet + FL	0.30 $\pm$ 0.22	0.35 $\pm$ 0.27	0.43 $\pm$ 0.27
Pinto <i>et al</i> (2021) [18]			
- Supervised branch	0.30 $\pm$ 0.21	0.30 $\pm$ 0.26	0.55 $\pm$ 0.31
- Unsup + Sup branches	0.38 $\pm$ 0.22	0.41 $\pm$ 0.26	0.53 $\pm$ 0.29
Pinto <i>et al</i> (2018a) [10]			
- Standard model	0.20 $\pm$ 0.19	0.16 $\pm$ 0.20	0.61 $\pm$ 0.28
- 4D-PWI model	0.20 $\pm$ 0.18	0.18 $\pm$ 0.21	0.61 $\pm$ 0.27
- Multi-data model	0.26 $\pm$ 0.21	0.21 $\pm$ 0.20	0.61 $\pm$ 0.28
- Multi-data multi-model	0.29 $\pm$ 0.21	0.23 $\pm$ 0.21	0.66 $\pm$ 0.29
Ours			
- ADC + TTP	0.33 $\pm$ 0.22	0.39 $\pm$ 0.26	0.49 $\pm$ 0.32
- ADC + Tmax	0.34 $\pm$ 0.22	0.34 $\pm$ 0.23	0.57 $\pm$ 0.32
- ADC + TTP + Tmax	0.36 $\pm$ 0.21	0.42 $\pm$ 0.25	0.48 $\pm$ 0.29

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.

A final validation was carried with respect to the baseline 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 3 summarize 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 [11] 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* [10] 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 towards predicting the lesion class (coarse segmentations), with an overestimation of the

lesion. On the other hand, Hu *et al* [11] 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).

6. Discussion

This work introduced a deep supervised cross-attentional deep autoencoder, which follows a training regime focused on ischemic stroke lesions, observed from multi-parametric MRI sequences. The proposed approach achieved a competitive performance regarding state-of-the-art methods on the ISLES2017 dataset. More specifically, our proposed method presented the highest precision (0.42) and the second-best dice score (0.36). In addition, an exhaustive analysis of the contribution of MP-MRI sequences. As expected, the ADC sequence achieves the best performance due to the great anatomical detail it provides, and the fact that manual delineations were performed on it. 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 [19].

To the best of our knowledge, in this work first 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 a fact may be observed in all experiments, where for different map inputs the attention modules achieve an average gain on dice score of 0.33% for the single modalities and 0.15% for the multimodal models. Furthermore, the mean gain in dice score (1.83% for the single modalities and 1.15% for multimodal) supports the importance of the class weights in the training of the attention models. Other works validated on alternative MRI datasets. For instance, Clerigues *et al* used an oversampling strategy to balance the damaged and healthy patches [14]. 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 to the overall lesion estimation geometry, losing anatomical information from the brain. Contrarily, our approach 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 by the mean gain in dice score of 1.25% for the single modalities and 1.25% for the multimodal models.

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* [10], 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* [10] 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 accuracy of 0.42, and 0.48.

In the same line, Hu *et al* [11] 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 representation between encoder and decoder is not significative. 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, to train 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.

According to the results in the ISLES2017 official report [17], the configuration of the multi-parametric attentional model outperforms the top two baseline strategies for single models and ensemble strategies in the dice and precision scores. Regarding the recall, our model ranked third, with a 3% decrease. This fact may be associated with the overestimation of lesions, producing coarse segmentations (relatively low precision and high recall). This fact may be a critic issue to define clinical procedures from estimations of particular studies.

Finally, Pinto *et al* [18] approach combines two models to retrieve stroke segmentations. The first model is a fully supervised approximation that combines a U-Net with an LSTM to leverage all the multi-parametric maps and produce the lesion estimations. The second model uses two Restricted Boltzmann Machines (RBMs) branches to process MRI parameters that describe the blood flow (rCBV and rCBF) and the lesion (ADC, TTP, MTT, and Tmax). From the computed features of both branches is performed an offline feature selection that is the input for the supervised model. Notably, the proposed approach outperformed their supervised version in the dice and precision metrics, while achieving a lower recall. Besides, the proposed approach was better in precision compared to their unsupervised + supervised version. Nevertheless, loses by a 2% in the dice score and by 5% in the recall. Again, the output results of Pinto *et al* may produce major lesion areas that favor overlapping metrics but may be sensitive to defining larger areas of the lesion.

7. Conclusions

This work introduced a new autoencoder architecture that integrates cross-attention mechanism into skip connection, while following a deep hierarchical supervision to segment stroke lesions over MRI scans. The best configuration was achieved from the parameter of ADC, integrated with TTP and Tmax perfusion parametric maps. From such multimodal integration recovered stroke-associated lesions, even considering tiny and variable shapes. Even better a compact architecture with ADC and Tmax sequences showed remarkable shape segmentation results. The proposed approach achieves a competitive performance over the ISLES2017 public challenge. Future works will include the validation in more challenging scenarios, with complete and raw studies, and comparing output shapes with the reference of several radiologist experts. Also, a study of new attention mechanisms will be carried out as an alternative proposal to recover ischemic lesions at different stages (chronic or subacute).

Acknowledgments

The authors thank Ministry of science, technology and innovation of Colombia (MINCIENCIAS) for

supporting this research work by the project ‘Mecanismos computacionales de aprendizaje profundo para soportar tareas de localización, segmentación y pronóstico de lesiones asociadas con accidentes cerebrovasculares isquémicos’, with code 91934.

Data availability statement

The data that support the findings of this study are openly available at the following URL/DOI: <http://www.isles-challenge.org/ISLES2017/>.

ORCID iDs

Edgar Rangel  <https://orcid.org/0000-0002-9228-7226>

Fabio Martínez  <https://orcid.org/0000-0001-7353-049X>

References

- [1] WHO. World Stroke Organization; 2021. Available from: (<https://world-stroke.org/>)
- [2] Roth G A *et al* 2020 Global burden of cardiovascular diseases and risk factors, 1990–2019: update from the GBD 2019 study *J. Am. College Cardiol.* **76** 2982–3021
- [3] Martel A L, Allder S J, Delay G S, Morgan P S and Moody A R 1999 Measurement of infarct volume in stroke patients using adaptive segmentation of diffusion weighted MR images *In Medical Image Computing and Computer-Assisted Intervention—MICCAI’99 Second International Conference (Cambridge, UK, September 19–22)* (Berlin, Heidelberg: Springer) pp 22–31
- [4] Neumann A B *et al* 2009 Interrater agreement for final infarct mri lesion delineation *Stroke* **40** 3768–71
- [5] Liu L, Kurgan L, Wu F X and Wang J 2020 Attention convolutional neural network for accurate segmentation and quantification of lesions in ischemic stroke disease *Med. Image Anal.* **65** 101791
- [6] Wang G, Song T, Dong Q, Cui M, Huang N and Zhang S 2020 Automatic ischemic stroke lesion segmentation from computed tomography perfusion images by image synthesis and attention-based deep neural networks *Med. Image Anal.* **65** 101787
- [7] Kamnitsas K *et al* 2017 Efficient multi-scale 3D CNN with fully connected CRF for accurate brain lesion segmentation *Medical Image Analysis.* **36** 61–78
- [8] Tureckova A and Rodríguez-Sánchez A J 2019 ISLES challenge: U-shaped convolution neural network with dilated convolution for 3D stroke lesion segmentation *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)* **11383 LNCS** 319–27
- [9] Abulnaga S M and Rubin J 2019 Ischemic stroke lesion segmentation in CT perfusion scans using pyramid pooling and focal loss *Brainlesion: Glioma, Multiple Sclerosis, Stroke and Traumatic Brain Injuries* ed A Crimi *et al* (Cham: Springer International Publishing) pp 352–63
- [10] Pinto A *et al* 2018 Enhancing clinical MRI perfusion maps with data-driven maps of complementary nature for lesion outcome prediction *Lect. Notes Comput. Sci.* **2** 107–15
- [11] Hu X *et al* 2020 Brain SegNet: 3D local refinement network for brain lesion segmentation *BMC Medical Imaging.* **20** 1–10
- [12] Dolz J, Ben Ayed I and Desrosiers C 2019 Dense multi-path u-net for ischemic stroke lesion segmentation in multiple image modalities *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)* **11383 LNCS** 271–82
- [13] Clèrigues A, Valverde S, Bernal J, Freixenet J, Oliver A and Lladó X 2019 Acute ischemic stroke lesion core segmentation in CT perfusion images using fully convolutional neural networks *Comput. Biol. Med.* **115** 103487
- [14] Clèrigues A, Valverde S, Bernal J, Freixenet J, Oliver A and Lladó X 2020 Acute and sub-acute stroke lesion segmentation from multimodal MRI *Computer Methods and Programs in Biomedicine.* **194**
- [15] Bahdanau D, Cho K and Bengio Y 2015 Neural machine translation by jointly learning to align and translate *ICLR* **11** 367–73
- [16] Loshchilov I and Hutter F 2017 Decoupled weight decay regularization *arXiv:1711.05101*
- [17] Winzeck S *et al* 2018 ISLES 2016 and 2017-benchmarking ischemic stroke lesion outcome prediction based on multispectral MRI *Front. Neurol.* **9** 2018
- [18] Pinto A *et al* 2021 Combining unsupervised and supervised learning for predicting the final stroke lesion *Med. Image Anal.* **69** 101888
- [19] Copen W A, Schaefer P W and Wu O 2011 Good resource to understand the role of perfusion parametric maps in the assesment of stroke *MR Perfusion Imaging in Acute Ischemic Stroke*

PROCEEDINGS OF SPIE

SPIDigitalLibrary.org/conference-proceedings-of-spie

An attentional unet with an auxiliary class learning to support acute ischemic stroke segmentation on CT

Santiago Gómez, Sebastian Florez, Daniel Mantilla, Paul Camacho, Nick Tarazona, et al.

Santiago Gómez, Sebastian Florez, Daniel Mantilla, Paul Camacho, Nick Tarazona, Fabio Martínez, "An attentional unet with an auxiliary class learning to support acute ischemic stroke segmentation on CT," Proc. SPIE 12464, Medical Imaging 2023: Image Processing, 124640S (3 April 2023); doi: 10.1117/12.2654269

SPIE.

Event: SPIE Medical Imaging, 2023, San Diego, California, United States

An attentional Unet with an auxiliary class learning to support acute ischemic stroke segmentation on CT

Santiago Gómez^a, Sebastian Florez^a, Daniel Mantilla^b, Paul Camacho^b, Nick Tarazona^b, and Fabio Martínez^a

^aBIVL²ab, Universidad Industrial de Santander, Bucaramanga, Santander, Colombia

^bClínica FOSCAL, Floridablanca, Santander, Colombia

ABSTRACT

Computed tomography (CT) is the first-line imaging modality for evaluation of patients suspected of stroke. Specially, such modality is key as screening test between ischemia and hemorrhage strokes. Despite remarkable support of encoder-decoder architectures, the delineation of ischemic lesions remains challenging on CT studies, reporting poor sensitivity, especially in the acute stage. Among others, these nets are affected because of the low scan quality, the challenging stroke geometry, and the variable textural representation. This work introduces a boundary-focused attention U-Net that takes advantage of cross-attention mechanism, that along multiple levels allows to recover stroke segmentation on CT scans. The proposed architecture is enriched with skip connections, that help in the recovering of saliency lesion maps and motivated the preservation of morphology. Besides, an auxiliary class is herein introduced with a weighted special loss function that remark lesion tissue, alleviating the negative impact of class unbalance. The proposed approach was validated on the public ISLES2018 dataset achieving an average dice score of 0.42 and a precision of 0.48.

Keywords: Stroke, Medical Image Segmentation, Computed Tomography, Attention Deep Learning, Lesion Boundaries.

1. INTRODUCTION

Stroke is the second leading cause of mortality worldwide and the major adult disability in developed countries.¹ The blood vessel occlusion (ischemic stroke) is the most prevalent condition (80% of all cases). A timely characterization and localization of the brain lesions is imperative for a good patient outcome. CT scans have major availability in clinical centers, being fundamental the lesion analysis over such studies.² Nonetheless, these CT studies have reported low sensibility to characterize strokes and the expert analysis result tedious, time-consuming (around 15 minutes each case) and prone to errors, introducing an inherent radiologist bias.³

Computational U-Net architectures are today the standard strategies to support ischemic stroke lesion segmentation. For instance, Tureckova *et al.* used a 3D U-Net based model with dilated convolutions,⁴ while Dolz *et al.* proposed an U-Net architecture that take image modalities separately with multiple dilated convolutional encoders.⁵ Nonetheless, these representations remain limited to recover whole context information. In such sense, the strategies may propagate errors through skip connection, transferring whole and equally weighted information from encoder. To deal with this issue, Kuang *et al.* proposed a multi-task learning network to segment early Infarcts and score ASPECTS, simultaneously.⁶ This net implements a multi-level attention gate module (MAGM) that preserves the correlated activations with the early infarct. This net however lost the capability to integrate lesion observations from parametric maps. The stroke segmentation is an unbalanced problem that naturally make learning difficult because of the relatively few information from the lesion in each CT study. For such reasons, it is necessary to explore mechanisms to stand out stroke delineations and boundaries between healthy and lesion tissue. To address this challenge, Clérigues *et al.* proposed a 2D asymmetric residual encoder-decoder framework, following a dynamically weighted loss function and mini-batches of image patches.⁷ Other approaches have integrated MRI-DWI sequences to enrich observations taken from CT-perfusion studies.

Further author information: (Send correspondence Fabio Martínez)

Fabio Martínez. E-mail: famarcar@saber.uis.edu.co, Telephone: (+57) (037) 6344000 Ext 2110

For instance, Liu *et al.*⁸ proposed an adversarial-generation of synthetic MRI-DWI images from CT-studies. This synthetic DWI image is then mapped through a segmentation network to predict the lesion outcome. Similarly, Wang *et al.*⁹ proposed a three network pipeline to generate synthetic DWI sequences from 4D Perfusion CT sequences. These architectures achieve remarkable performance on stroke segmentation but require paired and aligned MRI-DWI and CT sequences, which result unrealistic in clinical scenarios.

Original contribution. This work introduces a boundary-focused U-net framework that recover stroke lesions from CT studies. The proposed approach integrated multiple parametric CT maps and it uses the lesion boundary as an auxiliary class to enforce stroke segmentation. To carefully drive segmentation at different network levels, the strategy replace skip connections with cross-attention mechanism to focus learning on the stroke shape. Besides, a weighted special loss function helps with class imbalance problems.

2. PROPOSED APPROACH

The general description of the proposed net is illustrated in Figure 1a.

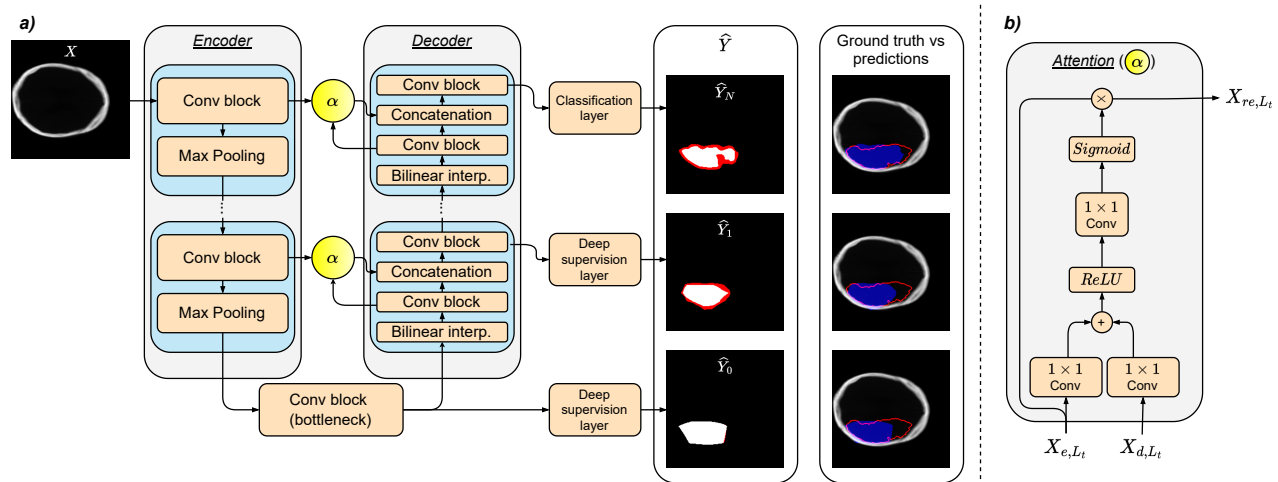


Figure 1. An overview of a boundary-focused attention U-Net for the segmentation of ischemic stroke lesions.

2.1 The U-Net enriched with cross-attention mechanisms

U-net is the most popular deep architecture for the segmentation task.¹⁰ Among others, this architecture has capabilities to capture visual context, using skip connections to integrate earlier encoder representations. Nevertheless, for stroke lesions, expressed as small and poor textural regions with negligible discrimination w.r.t. the background, the skip connection strategy could introduce noisy features and perturb the decoder hidden representation. In this work, we overcome such limitations by introducing additive cross-attention attention modules (see Figure 1b), allowing to control the contribution of the encoder X_e into the decoder X_d representation, at each convolutional level L_t . The refined encoded features at level L_t are computed following the formula: $X_{re} = X_e \cdot \text{Sigmoid}(W_{re}^T \text{ReLU}(W_e^T X_e + W_d^T X_d))$. Consequently, the refined features X_{re} preserve highly correlated features between encoder and decoder branches, at the same level of processing.

2.2 Lesion boundaries as an auxiliary class

A main challenge of stroke segmentation is the close similarity with healthy tissue, consequence of poor sensitivity on CT acquisitions. In this work, we force the architecture to learn lesion segmentation by including an auxiliary class that consists of boundaries of shape. Hence, 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. Regarding the model, we added a classification layer (1×1 conv + softmax) to the output features of the U-net model.

2.3 Refinement multilevel training

To improve the predictions of the boundary-focused attention model, we implemented a refinement training (RT) of the boundary-focused U-Net. Firstly, to minimize error propagation, a deep supervision was carried out on the final layers of each convolutional level, following bilinear interpolation to upscale features, and one classification layer. Besides, to give more importance to the lesion boundary pixels and alleviate the class imbalance problems, class weight maps ($\mathbf{C}$) were added to the computation of the final loss function. In consequence, the lesion estimations and their corresponding class weight maps are used to compute the final loss signal as follows: $\mathcal{L} = \mathbf{WCY}\log(\hat{\mathbf{Y}})$ where $\{\mathbf{Y}, \hat{\mathbf{Y}}, \mathbf{C}\} \in \mathbb{R}^{O \times H \times W}$ and the weights from output model are $\mathbf{W} \in \mathbb{R}^{|W|}$.

EXPERIMENTAL SETUP The proposed approach was validated over 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. Finally, we randomly left out 19 studies from the training partition to validation.

Regarding the model, the proposed symmetric architecture consists of six convolutional levels with convolutional blocks (3×3 , batch normalization, and ReLU activation). At successive encoder levels is applied max-pooling, while for decoder path is applied a bilinear interpolation. Deep supervision layers were placed in the outputs of each decoder level. For training, the architecture was updated over 600 epochs with a batch size of 32, and a Focal loss with an AdamW optimizer. A warm-up strategy was used in the first 90 epochs (learning rate from $5e-3$ and decay $1e-5$), followed by 510 epochs of cosine decay. 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 loss terms were set to $\mathbf{W} = \{0.03, 0.045, 0.05, 0.125, 0.25, 0.5\}$ and the class weights to generate the weight maps to $\mathbf{C} = \{0.7, 2.0, 7.0\}$.

3. EVALUATION AND RESULTS

An ablation study was herein included considering three architecture versions: 1) a standard U-Net (U-Net), 2) a standard U-Net with refinement training (U-Net + RT), 3) a standard U-Net with refinement training and cross-attention modules (U-Net + RT + Att). These architectures were evaluated with and without the auxiliary class. In table 1 is summarized the performance for all versions. As expected, the complete version of the proposed approach achieves the better dice score 0.643 ± 0.185 .

Table 1. Results on the valid set

Model	Dice		Precision		Recall	
	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 + RT	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 + RT + 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

Figure 2 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 lesions of different size 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.

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

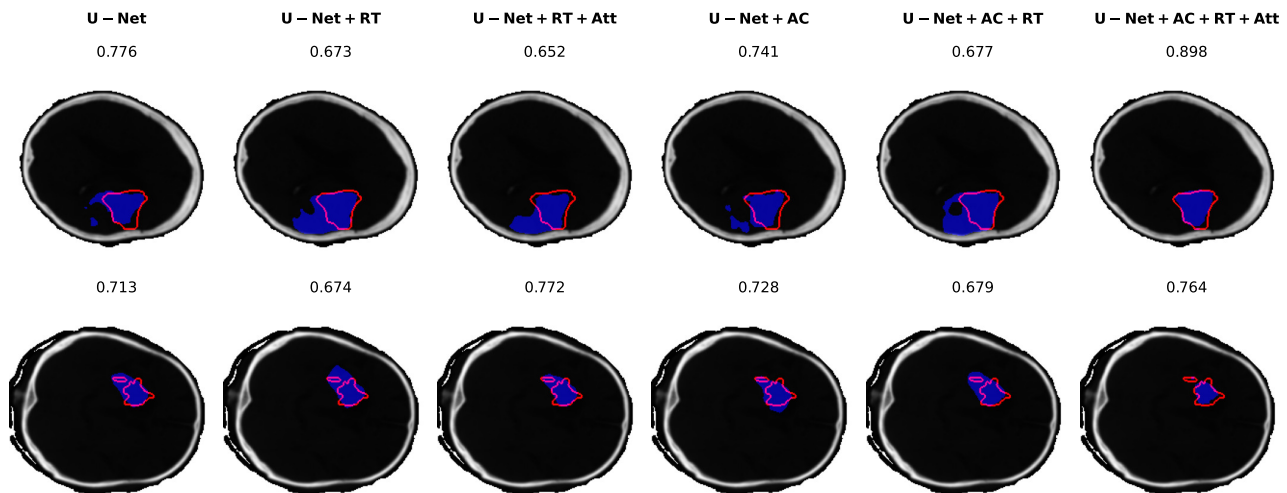


Figure 2. Comparison of the ischemic stroke lesion predictions (blue areas) by the proposed models vs manual expert annotations (red contours). The numbers above each plot are dice scores.

to learn volumetric patterns, while dilated information only exploit large local information, losing context from whole slice.⁴ 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 achieve remarkable results from CT-to-DWI translation schemes, which result unrealistic in clinical routines. Besides, these DWI images are no longer included in the dataset.^{8,9}

4. CONCLUSIONS AND PERSPECTIVES

This work presented an attention Unet strategy that employed an auxiliary contour lesion class to stand out patterns associated with ischemic strokes. The proposed approach is also supported during training by a refinement approach that gives more weight to lesion classes, promotes stable representations and forces intermediate representations to better recover lesions. Further validations over complete CT studies will be carried out to determine advantages and limitations in more suitable and real scenarios.

ACKNOWLEDGMENTS

The authors thank Ministry of science, technology and innovation of Colombia (MINCIENCIAS) for supporting this research work by the project "Mecanismos computacionales de aprendizaje profundo para soportar tareas de localización, segmentación y pronóstico de lesiones asociadas con accidentes cerebrovasculares isquémicos", with code 91934.

REFERENCES

- [1] Roth, G. A. and et al., "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 (2020).
- [2] Campbell, B. C. and Parsons, M. W., "Imaging selection for acute stroke intervention," *International Journal of Stroke* **13**(6), 554–567 (2018).
- [3] Martel, A. L. and et al., "Measurement of infarct volume in stroke patients using adaptive segmentation of diffusion weighted MR images," *Lecture Notes in Computer Science* (1999).
- [4] Tureckova, A. and et al., "ISLES challenge: U-shaped convolution neural network with dilated convolution for 3D stroke lesion segmentation," *Lecture Notes in Computer Science* **11383 LNCS**, 319–327 (2019).

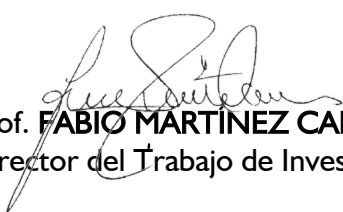
- [5] Dolz, J. and et al., “Dense multi-path u-net for ischemic stroke lesion segmentation in multiple image modalities,” *Lecture Notes in Computer Science* **11383 LNCS**, 271–282 (2019).
- [6] Kuang, H. and et al., “EIS-Net: Segmenting early infarct and scoring ASPECTS simultaneously on non-contrast CT of patients with acute ischemic stroke,” *Medical Image Analysis* **70**(February), 101984 (2021).
- [7] Clèrigues, A. and et al., “Acute ischemic stroke lesion core segmentation in CT perfusion images using fully convolutional neural networks,” *Computers in Biology and Medicine* **115**, 103487 (2019).
- [8] Liu, P., “Stroke lesion segmentation with 2D novel CNN pipeline and novel loss function,” *Lecture Notes in Computer Science* **11383 LNCS**, 253–262 (2019).
- [9] Wang, G. and et al., “Automatic ischemic stroke lesion segmentation from computed tomography perfusion images by image synthesis and attention-based deep neural networks,” *Medical Image Analysis* (2020).
- [10] Ronneberger, O. and et al., “U-Net: Convolutional Networks for Biomedical Image Segmentation,” *Lecture Notes in Computer Science* **9351**(Cvd), 12–20 (2015).

**ACTA DE SUSTENTACIÓN
TRABAJO DE INVESTIGACIÓN DE MAESTRÍA EN INGENIERÍA DE SISTEMAS E
INFORMÁTICA.**

De acuerdo con el artículo 109 del Reglamento General de Posgrado (Acuerdo del Consejo Superior No 075 de 2013), los suscritos miembros del jurado del Trabajo de Investigación de Maestría "Segmentation Of Ischemic Stroke Lesions In Radiological Sequences Using Deep Attention Mechanisms" presentado por el estudiante de Maestría en Ingeniería de Sistemas e Informática **Santiago Gómez Hernández**, código 2218110 una vez evaluada y realizada la defensa oral dan concepto de:

APROBADO ☒ NO APROBADO ☐ APLAZADO ☐

Se envía comunicación escrita de la presente decisión tanto al estudiante como a su director a los 12 días del mes de mayo de 2023.


Prof. **FABIO MARTÍNEZ CARRILLO**
Director del Trabajo de Investigación

(EISI-UIS).


Prof. **JUAN GABRIEL CASTRILLÓN GUZMÁN**
Evaluador.

(Universidad Técnica de Múnich).


Prof. **JULIO ERNESTO VILLALÓN REINA**
Evaluador

(Universidad del Sur de California)