Multimedia Tools and Applications
https://doi.org/10.1007/s11042-023-17220-w

Impact of attention mechanisms for organ segmentation in chest x-ray images over U-Net model

Tomás de la Sotta[1] · Violeta Chang[2] · Benjamín Pizarro[1,3] · Héctor Henriquez[4,5] · Nicolás Alvear[1] · Jose M. Saavedra[4]

Received: 4 May 2023 / Revised: 8 August 2023 / Accepted: 21 September 2023

Abstract

Chest x-ray images are one of the most commonly performed imaging tests, providing crucial clinical information about structures like the heart, lungs, ribs, bones, and blood vessels. In this context, image segmentation is a critical stage as it aims to separate significant parts of an image. However, manual image segmentation presents serious difficulties that can be tackled using deep learning-based methods, such as U-Net. Furthermore, attention mechanisms have attracted the interest of the machine-learning community and can bring improvements in medical image segmentation. This research aims to assess the impact of attention-based mechanisms for segmenting different organs in chest X-ray images, like heart, lungs, clavicles and ribs, over the well-known U-Net architecture as a baseline. We study five U-Net encoder variations, replacing the U-Net encoder with ResNet 18, 34 and 50, Swin Transformer and a simple residual structure comprised of a single ResNet-50 prior to each U-Net encoder layer. In the original U-Net, the skip layers are identity layers connecting each encoder block with its corresponding decoder one. Here, we replace these layers with different attention mechanisms: spatial attention, full spatial attention, double spatial attention, multiple spatial attention, spatial cross-attention and Swin spatial cross-attention. Each encoder variation and attention mechanism was evaluated from scratch and on a pre-trained scenario, independently for lungs, ribs, heart and clavicles. ResNet-UNet-18 achieves up to 0.96 and 0.53 of average overlapping with hand-segmented masks of lungs and clavicles, respectively. The best encoder for rib and heart segmentation was Residual U-Net, with 0.88 and 0.83 overlapping, respectively. Furthermore, for attention mechanisms, the most suitable were selected according to overlapping with hand-segmented masks as Spatial Attention U-Net for lungs (0.96), Three-Head Attention for ribs (0.88), Full Spatial Attention for the heart (0.82) and Spatial Cross-Attention for clavicles (0.54). Encoder variations and attention mechanism have a positive impact over a U-Net for segmentation of lungs, ribs, heart and clavicles from scratch, without transfer learning. Moreover, the encoder and the attention mechanism are not universal for segmenting different organs in chest X-ray images. While organs share the same backbone architecture (lungs and clavicles), the most appropriate attention mechanisms for each organ are all different, achieving up to 0.99 of overlapping in lung segmentation.

✉ Jose M. Saavedra
jmsaavedrar@miuandes.cl

Extended author information available on the last page of the article

Published online: 31 October 2023

Keywords Attention mechanism · Chest x-ray segmentation · U-Net · Encoder variation

1 Introduction

Medical imaging is an area of medicine that uses images to represent anatomical structures and functional and physiological properties of the body [1]. Medical images include radiography, fluoroscopy, computed tomography (CT), magnetic resonance imaging (MRI), and nuclear imaging techniques like SPECT and PET, among others [2].

Radiology is a technology-based specialty of medicine that uses medical imaging to diagnose and facilitate treatments, usually through non-invasive procedures [3]. Radiology has been enormously successful, evolving from an optional assistant to a crucial component of multidisciplinary patient care and medical decision-making [4].

X-ray imaging is considered the oldest medical imaging modality developed after Roetgen's discovery at the end of the 19th century [5]. X-rays are ionizing radiation with a wavelength ranging from 0.01 to 10 nm [6, 7]. The underlying basis for the medical application of X-rays depends on the differential attenuation of X-rays produced when interacting with the human body. The transmitted X-ray beam is detected to produce a 2-D image depicting the patient's anatomy. Dense organs present more absorption levels than lesser dense ones. Thus, high-density structures, like bones, appear white in the X-ray image, while soft organs, like lungs, appear gray. In this way, the X-ray procedure results in a more straightforward and lower-cost technology that provides images directly, without any reconstruction, compared to other techniques such as CT, MRI, or PET [8].

Chest x-rays are one of the most commonly performed imaging tests, providing crucial clinical information rapidly, cheaply, and with low radiation exposure. Chest x-ray images allow clinicians to inspect structures like the heart, lungs, ribs, bones, or even blood vessels [9]. As one of the most cost-effective images modalities, its utility in chest radiology is broad in several clinical scenarios such as chest pain, dyspnea, fever, fractures, hemoptysis, pneumoperitoneum, foreign body, trauma, pneumonia, fibrosis, pulmonary nodules, among others [10, 11].

Despite all the benefits of chest x-rays, it still requires an expert to interpret the produced images [12]. Experts perceive abnormalities and understand what to attend or ignore [13]. In this process, image segmentation is a critical stage and has become widely used in other tasks related to medical imaging [14].

Segmentation aims to separate significant parts of an image, such as organs or tumors. In clinical practice, this allows calculating cardiac volume in MRI, measuring lung nodules with high precision, extracting radiomics features from a lung tumor, or measuring the extent of pneumonia on chest x-ray, which correlates with disease severity, among others [15, 16].

However, manual image segmentation presents serious difficulties, like time-consuming, inter and intra-observer variability, and the requirement of highly-trained operators. For this reason, support tools are required to help clinicians and radiologists to manage the high volume and complexity of these diagnostic images, to prevent errors, and improve diagnostic performance [17, 18].

Computational advances, especially those related to artificial intelligence, have enabled it to generate new solutions for the aforementioned adversities. In this sense, the significant advances in deep learning allow for finding more precise and valuable characteristics and patterns that provide clues or conclusions about the context of medical images [19].

Pioneer computer-based segmentation methods for medical images were based on low-level features like pixel-wise intensities, and mid-level features like local patterns without analyzing the semantic content of the underlying images [20]. Close to a semantic approach are the ATLAS-based models that account for anatomy information [21]. However, with the explosion of deep learning, different computer vision tasks, including segmentation, have achieved outstanding performance [22].

In the medical context, U-Net [23] is still the most popular deep-learning architecture proposed for segmenting medical images. It generalizes a previous segmentation model named FCN (Fully convolutional network) that combines features from high-resolution layers with those from deeper ones [24].

In recent years, attention mechanisms have attracted the interest of the machine learning community, especially after their positive impact on natural language processing [25]. In visual perception, there is a phenomenon called perceptual grouping, where various elements in a complex display are perceived as going together in one perceptual experience [26]. The interrelation of different visual components (a.k.a. visual structure) may be learned from experience, particularly during the first months of our lives. Regarding the relevance of perceptual grouping for a visual system, it also can bring improvements in medical image segmentation.

Therefore, this research seeks to contribute to the medical area by presenting a study on the impact of attention-based mechanisms for segmenting different organs in chest x-ray images like heart, lungs, clavicles and ribs. Our methodology takes the well-known U-Net architecture as a baseline, which is then modified by different encoder architectures and attention strategies. Our study shows the positive impact of encoder variations and attention mechanism over a U-Net for segmentation of lungs, ribs, heart and clavicles from scratch, without transfer learning. Moreover, the encoder and the attention mechanism are not universal for segmenting different organs in chest x-ray images. While there are organs that share the same backbone architecture (lungs and clavicles), the most appropriate attention mechanisms for each organ turn out to be all different, achieving up to 99% of overlapping in lung segmentation.

In this context, our main contributions are:

- A robust and comprehensive experimental framework for evaluating the impact of backbone architecture and attention mechanism on the original U-net architecture for segmentation of lungs, heart, clavicle and ribs.
- A deep study of U-net encoder variations for segmentation of lungs, heart, clavicle and ribs.
- The introduction of attention mechanisms, by means of replacing layers, over a U-net network for segmentation of lungs, heart, clavicle and ribs.

This paper is organized as follows. First, Section 2 reviews state of the art on chest x-ray image segmentation. Section 3 presents our methodology, including a detailed description of the involved datasets. Section 4 describes the experimental protocols in detail. In Section 5, our experimental results are presented and discussed in Section 6. Finally, Section 7 draws the conclusions and future works.

2 Related work

Although there is extensive research on the segmentation of x-ray images, in this section, we will focus on reviewing the most recent and relevant work related to the segmentation of lungs, heart, clavicle and ribs. We are interested in presenting the state-of-the-art in this field with models based on U-net architecture and/or attention mechanisms.

For lung segmentation in chest x-ray images, Rashid et al. [27] proposed a U-net-based architecture that combines a convolutional neural network architecture with mathematical morphology operators. The method was tested on JSRT and Montgomery County datasets (see Section 3.1 for details), achieving a Dice coefficient of 0.951 and 0.954, respectively. In addition, Mittal et al. [28] introduced LF-SegNet. This modified U-net-based approach incorporates a normalization mechanism and improves the up-sampling strategy for lung field segmentation in chest x-ray images. The proposed network was trained and tested on publicly available standard datasets (JSRT for training and JSRT and Montgomery County for testing), obtaining a Jaccard index of 0.951. In 2020, Yahyatabar et al. [29] introduced a model in which the information flow increases throughout the network by dense connectivity between various layers with significantly fewer parameters while keeping the segmentation robust. The model was evaluated on the previous datasets achieving a Dice coefficient of 0.974 and 0.973, respectively. On the other hand, Liu et al. [30] proposed an automatic lung segmentation method based on U-Net architecture. This method uses the pre-trained Efficientnet-b4 as the encoder, residual block, and LeakyReLU to optimize the decoder. Their algorithm was tested on the same datasets, and the Dice coefficients were 0.979 and 0.977, respectively.

Pal et al. [31] proposed UWnet, a modified version of the U-Net model which implements attention gates. The authors introduced an intermediate layer to bridge the encoder and decoder pathways. The intermediate layer is a series of fully connected convolutional layers generated from the upsampling of the final encoder layer connected to the corresponding up-sampled and down-sampled blocks via skip connections. A downsampling layer further connects the intermediate layer to the decoder pathway. The model was tested on a manually annotated subset of 200 images from the NIH chest x-ray dataset, achieving an F1-score of 0.957 and 0.809 for lung and heart segmentation, respectively.

Wang et al. [32] proposed MDU-Net, a multitask dense connection U-Net model, to segment the complete ribs and clavicles in chest x-ray images. The authors evaluated the proposed network on 88 own chest x-ray images, achieving a Dice coefficient value of 0.937, and 0.850 in the clavicle and rib segmentation, respectively.

Novikov et al. [33] introduced the InvertedNet model combining delayed subsampling, exponential linear units, highly restrictive regularization and many high-resolution, low-level abstract features for heart, lung and clavicle segmentation in chest x-ray images. The proposed strategy was tested on the JSRT dataset achieving 0.972, 0.902 and 0.935 for lung, clavicle and heart segmentation.

Recently, Ullah et al. [34] presented a novel dual encoderâ€“decoder architecture to effectively segment the anatomical structures in chest x-ray images based on a VGG19 encoder. The proposed method segments the more prominent structures (lungs and heart) and the smaller structures (clavicles) in chest x-rays. The proposed method incorporates Attention Gating Modules (AGMs) to allow the model to focus on the regions of interest while maintaining the spatial resolution and improving the quality of the feature maps. The proposed method was tested on three datasets. First, on the JSRT dataset achieving a Dice coefficient value of 0.868, 0.907, and 0.955 for clavicle, heart and lung segmentation. Second, on the Montgomery County dataset, achieving a Dice coefficient value of 0.977 for lung segmentation. Furthermore, finally, on the Shenzhen dataset (see Section 3.1 for details), achieving a Dice coefficient of 0.9568 for lung segmentation.

Table 1 reports a summary of the SOTA models for chest x-ray segmentation. To the best of our knowledge, there is no comprehensive report of exploration studies about the impact

Table 1 Summary of the performance of different SOTA models for chest x-ray segmentation

Model	Metric	Lungs	Ribs	Heart	Clavicles
U-Net [27]	Dice	0.951(J) 0.954(M)	–	–	–
LF-SegNet [28]	Jaccard	0.951(M)	–	–	–
InvertedNet [33]	Jaccard	0.972(J)	–	0.902(J)	0.935(P)
Dense-Unet [29]	Dice	0.974(J) 0.973(M)	–	–	–
MDU-Net [32]	Dice	–	0.937(P)	–	0.850(P)
EfficientNet-b4 [30]	Dice	0.979(J) 0.977(M)	–	–	–
UWnet [31]	F1-score	0.957(N)	–	0.809(N)	–
VGG19+AGM [34]	Dice	0.955(J) 0.977(M)	–	0.907(J)	0.868(J)

Results are shown in terms of Dice Coefficient, Jaccard Index and F1-score. (J) stands for JSRT dataset, (M) stands for Montgomery County dataset, (N) stands for HIH Chest X-ray dataset, and (P) for private dataset

of attention mechanism on U-net architecture for segmentation of lungs, heart, clavicle and ribs.

3 Material and methods

We start this section with a description of the involved datasets. Then, we present the proposed attention-based architectures for chest x-ray image segmentation. Finally, at the end of this section, we will describe the evaluation metrics used in this work.

3.1 Datasets

We used four chest x-ray segmentation datasets for the experiments: the JSRT dataset for clavicles and heart segmentation, the Montgomery County dataset and Shenzhen dataset for lung segmentation, and VinDr-RibCXR for rib segmentation. In the following lines, we describe the details for each dataset.

3.1.1 Japanese society of radiological technology dataset (JSRT)

This dataset includes 247 12-bit grayscale raw frontal chest x-ray images from the Japanese Society of Radiological Technology [35]. It provides basic patient report information (age and gender), nodule type diagnosis, nodule center coordinates and a basic nodule location map. It comprises 154 nodule-presenting images, including 100 malignant and 54 benign, and 93 no-nodule cases. The SCR dataset [36] extends the JSRT dataset adding full-size segmentation masks for the images within the original set. It includes the independent, left and right masks for clavicles and lungs and the heart segmentation mask for the JSRT images.

3.1.2 Montgomery county dataset (MC)

This dataset includes 138 12-bit grayscale frontal chest x-ray images from Montgomery County's Tuberculosis screening program [37]. There are 58 images with the presence of tuberculosis and 80 images free of tuberculosis. The dataset includes basic patient reports and lung segmentation masks, including age and gender.

3.1.3 Shenzhen dataset (SH)

This dataset includes 662 frontal grayscale chest x-rays from Shenzhen Nr. 3 Hospital in China [38]. It comprises 336 cases with tuberculosis and 326 tuberculosis-free cases, including pediatric anteroposterior (AP) images. As the Montgomery Dataset, it provides basic patient reports (age and gender) and lung segmentation masks.

3.1.4 VinDr-RibCXR dataset (VD)

This dataset is a private on-demand image set for automatically segmenting and labeling individual ribs from chest x-ray (CXR) scans [39]. The VinDr-RibCXR contains 245 CXRs with corresponding ground truth annotations provided by human experts. Each image was assigned to an expert, who manually segmented and annotated each of the 20 ribs, denoted as L1â†’L10 (left ribs) and R1â†’R10 (right ribs). All scans have been de-identified to protect patient privacy.

3.2 U-Net architecture

Image segmentation is one of the traditional computer vision tasks. It decomposes an image I into meaningful parts (a.k.a. semantic parts). Thus, a segmentation model receives an image as input and returns a disjoint set of regions S as output, where each region $R_i \in S$ represents a semantic part of I. Here, the term *semantic* means a part of an image with a relevant meaning for the underlying application. A typical representation of the segmented parts of an input image is by a multimask image, where each mask represents a different kind of object in the image. Figure 1 depicts a scheme of the image segmentation process.

U-Net [23] is the most popular deep-learning segmentation model for medical images. It consists of a two-block neural architecture; the first block is called *the encoder*, responsible for computing relevant features from the input image to facilitate the segmentation. Like most neural models, the encoder is a multiresolution architecture, starting from a high-resolution representation with poor semantic features to lower resolutions with more discriminant features. In addition, the U-Net encoder is commonly used as backbone for a diversity of computer-vision tasks.

Different from a traditional classification task, where features are aggregated into coarse levels, in the segmentation problem, we need to classify at the finest levels, almost at a pixelwise level. Therefore, the model needs to take the features produced by the encoder to generate segmentation masks from coarse to the finest levels. To this end, U-Net uses a second block called *the decoder*, that combines information from deeper layers (decoder's layers)

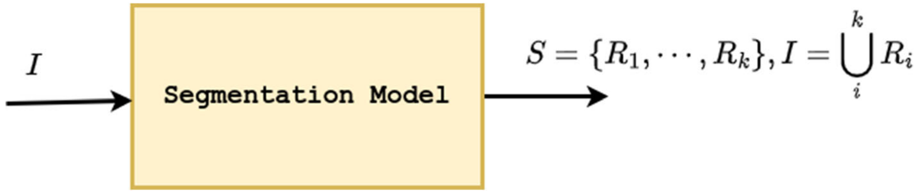

Fig. 1 General scheme of a segmentation task. A segmentation model receives an image I as input and returns a disjoint set of regions S as output, where each region $R_i \in S$ represents a semantic part of I

with high-resolution levels (encoder's layers), allowing the model to generate high-resolution segmentation masks. Figure 2 shows a general scheme of the U-Net architecture.

U-Net has been widely studied in the context of medical images. For instance, Liu et al. [30] proposed an improved U-Net for lung segmentation. They improved U-Net using the pre-training EfficientNet-b4 as the encoder and residual blocks and LeakyReLU activations in the decoder.

3.3 Attention mechanism

Attention is a critical component of visual systems. This is close to the *Gestalt Principles* [40] that make some objects be perceived as a whole, facilitating visual perception. This phenomenon is also called *Perceptual Grouping* and suggests that different receptive fields of a scene influence the others. In computer vision, perceptual grouping can be implemented by attention mechanisms like the popular Transformer modules [25], which is a crucial component in natural language processing.

Figure 3 illustrates the self-attention mechanism for visual recognition tasks using the Transformer strategy. As described by Vaswani et al. [25], a transformer module received a sequence of image embeddings. These embeddings can be obtained from the feature map produced by a convolutional neural network [41]. In addition, we can compute positional encoding that is added to the visual embeddings. The result is then used as input to the attention block. The attention block requires three representations computed by a linear transformation from the input embeddings. These representations are denoted as Q, K and V, where the letters come from **Query**, **Key** and **Value**, respectively. When these three representations come from the same source, the attention is called **self-attention**, and when Q comes from a source different from that from where K and V are produced, the attention

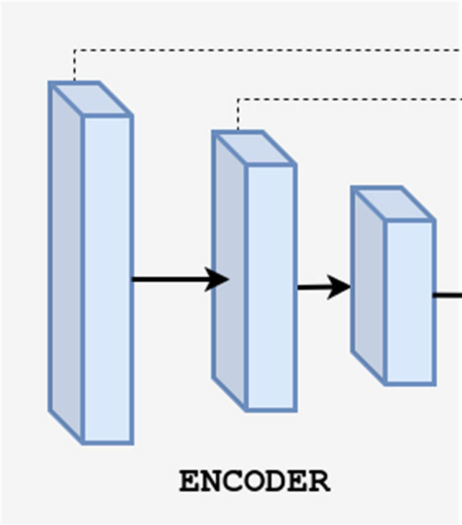

Fig. 2 Scheme of the U-Net architecture. The U-Net consists of a two-block neural architecture: an encoder for computing relevant features from the input image, and a decoder for combining information from deeper layers with high-resolution levels

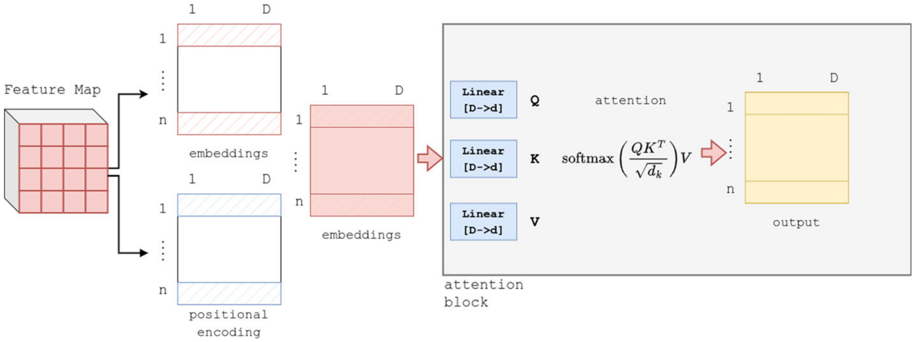

Fig. 3 Attention mechanism using a transformer strategy. A transformer module receives image embeddings with positional encoding, extracted from the feature map produced by a CNN. The attention block requires three representations (Q, K and V.) computed by a linear transformation from the input embeddings

is called **cross-attention**. In any case, after obtaining Q, K and V, the attention is computed by the Equation 1.

$$attention(Q, K, V) = \text{softmax}(\frac{QK^T}{\sqrt{d_k}})V \tag{1}$$

where d_k is a scalar factor.

Even though attention is a popular mechanism in natural language processing, the effectiveness of medical image segmentation has yet to be deeply studied. Therefore in this work, we evaluate the impact of attention mechanisms on the segmentation of different organs appearing in chest x-ray images, including lungs, heart, ribs and clavicles. We incorporate attention modules in a U-Net architecture under different settings, as described later in this section.

To improve diversification in the attention mechanism, Vaswani et al. [25] proposed a multi-head attention module that applies multiple single attention over the input sequence of embedding independently. All the attentional outputs are combined by concatenation, which is passed through a linear transformation.

In medical image segmentation, Cao and Zhao [42] proposed a three-input channel-wise attention mechanism for lung segmentation, combining features from both the encoder and the decoder. Figure 4 depicts the attention module proposed by Cao and Zhao. However, this proposal does not leverage the interrelation between different receptive fields as modern attention, like Transformers, does.

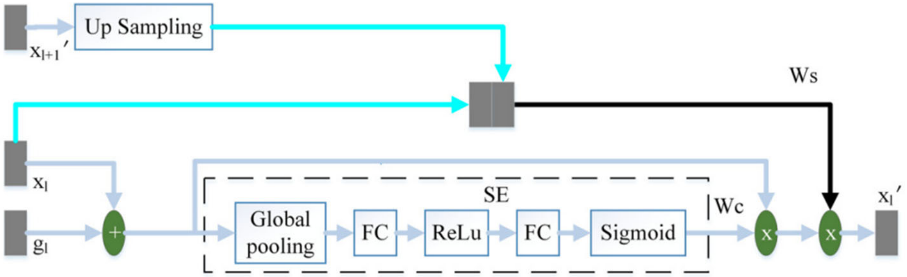

Fig. 4 A three-input attention proposed by Cao and Zhao. A three-input channel-wise attention mechanism for lung segmentation, combining features from both the encoder and the decoder

3.4 Proposed U-Net-based attentional architectures

The proposed attention mechanisms are constructed over a simplification of the architecture proposed by Cao and Zhao [42] that we call Tree-Head Attention U-Net. Figure 5 shows the general scheme of this baseline.

In our baseline, the attention diagram shown in Fig. 4 implements the circular gray components. However, we will describe other different ways to implement these components.

3.4.1 Spatial attention

Spatial attention is based on Transformer attention, which is independently applied by channels. Our implementation of this mechanism receives information from the encoder (E), the attention module (A), and the decoder (D) and computes the output as follows:

$$y = SE(s \times {D^c}^T @ A^c) \times E^c \tag{2}$$

where SE is an Squeeze and Excitation module, c indicates a particular channel, @ is matrix multiplication, and s is the scalar defined by Vaswani et al. [25]. Figure 6 illustrates the Spatial Attention mechanism.

3.4.2 Full spatial attention

This is a variation of the Spatial Attention where the following equation defines the output:

$$y = SE(s \times {D^c}^T @ A^c @ E^c) \times E^c \tag{3}$$

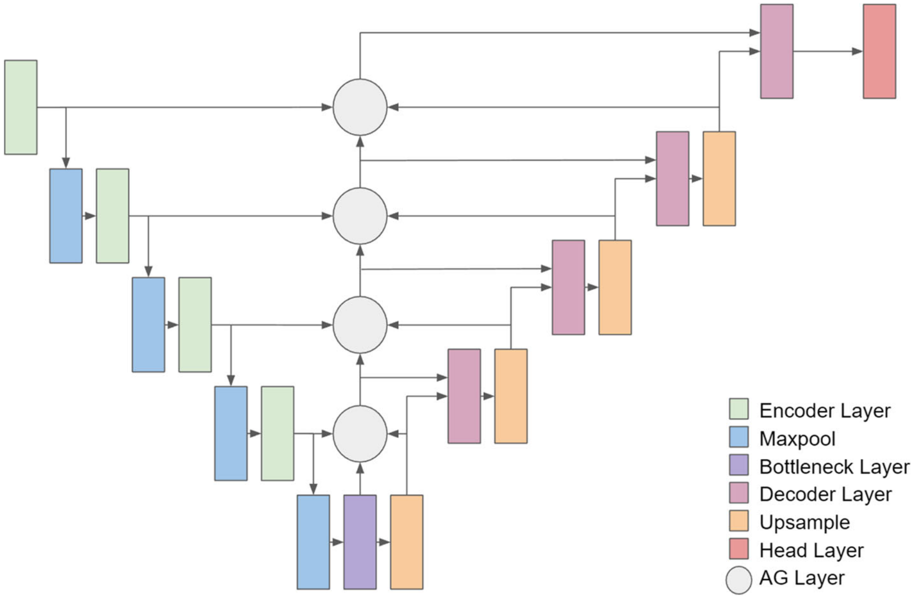

Fig. 5 A general scheme of our proposal three-head attention mechanism. A basic U-Net model based on Cao And Zhao proposal [42] where the gray component (AG Layer) implements an attention mechanism receiving information from three sources: the encoder, the previous attention module, and the decoder

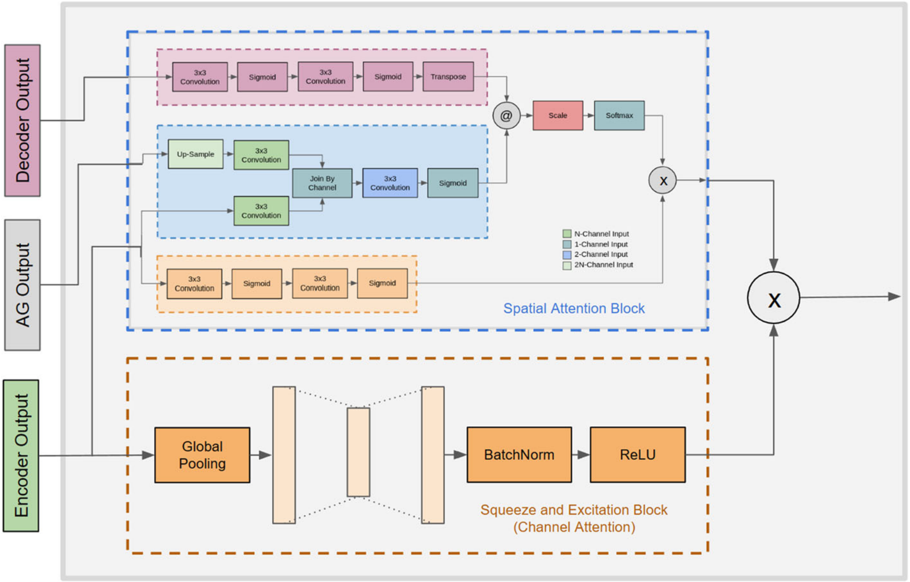

Fig. 6 A graphical scheme of Spatial Attention Layer. This attention mechanism receives information from the encoder (green), the attention module (gray), and the decoder (pink) and computes the output independently by channels

Unlike simple Spatial Attention, Full Spatial Attention uses a double matrix multiplication between E, A, and D. Figure 7 shows this attention strategy.

3.4.3 Double spatial attention U-Net

This is another extension of simple Spatial Attention, where a simple self-attention module follows the attention mechanism.

3.4.4 Multiple spatial attention U-Net

In this case, we use multiple spatial attention modules where the outputs are concatenated to compute the final result. Figure 8 illustrates this component.

3.4.5 Spatial cross-attention U-Net

This changes the three-head spatial attention module to a two-head attention module, like a classical cross-attention mechanism between the encoder and the decoder.

3.4.6 Swin spatial cross-attention U-Net

Unlike the previous attention strategies, this applies the traditional cross attentions mechanism between the encoder and the decoder. Therefore, the attention component (gray circle) is modified in this case, taking out the input from the attention components. Then, the attention uses the Swin block proposed by Liu et al. [43].

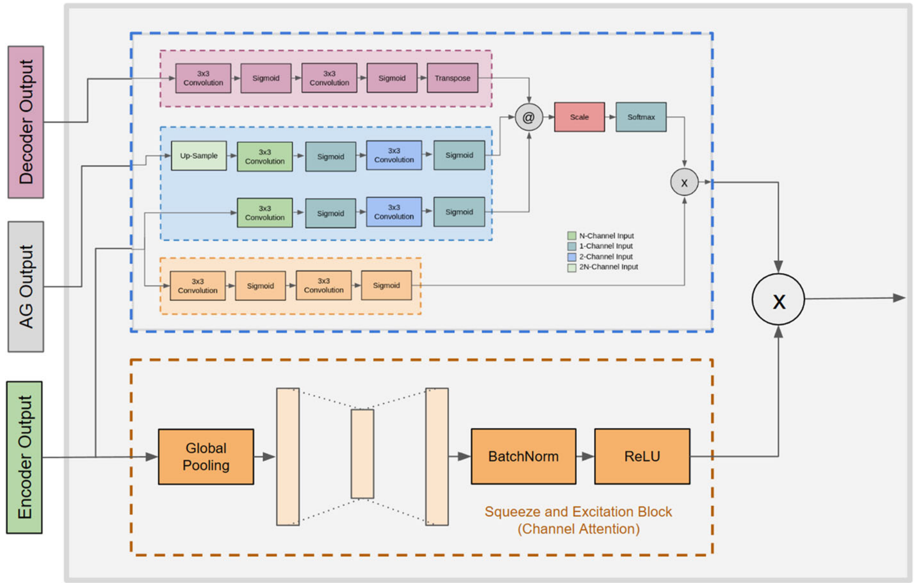

Fig. 7 A Full Spatial Attention Layer graphical scheme. This attention mechanism receives information from the encoder (green), the attention module (gray), and the decoder (pink) and computes the output using a double matrix multiplication between the inputs

3.5 Evaluation metric

Given a hand-made segmentation, various methods exist to evaluate the segmentation quality. The idea is to measure the difference between the automatic segmentation S against the manually segmented image G by computing some evaluation metrics. These metrics can be

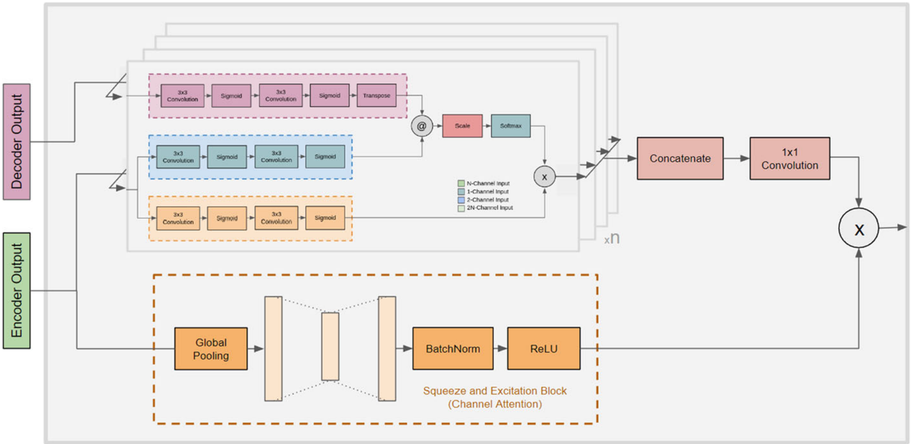

Fig. 8 Graphical scheme for Multiple Spatial Attention Module. This attention mechanism receives information from the encoder (green) and the decoder (pink) in multiple spatial attention modules. A concatenation of the outputs generates the final result

based on spatial overlap measures (e.g., Dice coefficient [44]) and on distance measures (e.g., Hausdorff distance [45]). Our evaluation metric is similar to the one used in previous works related to organ segmentation, that is, the Dice coefficient, to compare our results to the ones of the state-of-the-art methods.

The Dice coefficient value is calculated as

$$D = \frac{2||S \cap G||}{||S|| + ||G||}$$

S represents an automatic segmentation method, and G represents the hand-segmented masks (gold-standard). Therefore, the Dice coefficient values vary in the [0, 1] range, where 0 indicates no spatial overlap between S and G, and 1 indicates complete overlap.

4 Experimental protocol

Our research evaluates the impact of adding attention mechanisms in a U-Net architecture for organ segmentation in chest x-ray images. We study four organs in chest x-rays: lungs, ribs, heart and clavicles. Each model has been trained three times, and we have reported the average Dice coefficient.

4.1 Evaluated architectures

4.1.1 The backbone

We study five U-Net variations, replacing the U-Net encoder with ResNet 18, 34 and 50, Swin-Tranformer Base Model and a simple residual structure comprised by a single ResNet-50 [46], prior to each U-Net encoder layer. We do not evaluate bigger ResNets (e.g. ResNet-101) because of the huge amount of data they require and the good performance achieved by smaller backbones as showed in in Table 2.

4.1.2 Attention module

In the original U-Net, the skip layers are identity layers connecting each encoder block with its corresponding decoder one. We replace these layers with the strategies described in Section 3.4 for these experiments.

Table 2 Encoder variations from scratch

	Lungs	Ribs	Heart	Clavicles
U-Net	0.930	0.876	0.798	0.481
Residual U-Net	0.934	**0.883**	**0.831**	0.511
ResNet-UNet-18	**0.956**	0.864	0.812	**0.525**
ResNet-UNet-34	0.953	0.862	0.815	0.515
ResNet-UNet-50	0.949	0.859	0.815	0.522
Swin-UNet	0.933	0.782	0.768	0.469

Encoder variations from scratch. We present the mean Dice coefficient value for each organ, obtained using the JSRT dataset for clavicles and heart segmentation, the Montgomery County dataset for lung segmentation, and VinDr-RibCXR for rib segmentation

4.2 Dataset splitting

We separated the images from datasets into two disjoint groups: training and testing images. The testing images were not used for tasks other than for the final assessment of the models. We used the MC and the SH datasets (MC-SH from now) for lung segmentation, with 610 images for training and 68 images for testing. The heart and clavicle models used the JSRT dataset, particularly 222 images for training and 24 for testing. Finally, we used the VD dataset for rib segmentation, considering 221 images for training and 24 for testing.

4.3 Training

For training, we used PyTorch and a single RTX3080 GPU. The optimizer is the RMSProp with a base learning rate of $1e-5$. In addition, we use Cross-Entropy and Dice losses.

5 Experimental results

5.1 The backbone

Each encoder variation was trained over a randomly initialized set of weights (referred to in the future *from scratch*) and on a pre-trained scheme using MC-SH referred to in the future as *MC-SH pre-trained*. The following sections show the results given by each training variation.

As shown in Table 2, from scratch, lung segmentation is the only task in which all the evaluated encoder variations outperform the U-net base model. It is observed that the ResNet-UNet-18 model is the most competitive for lung segmentation achieving a Dice coefficient of 0.956. For heart and clavicle segmentation, most of the evaluated encoder variations outperform the base U-Net model. In the case of heart segmentation, the U-Net base model alone is not outperformed by Swin-UNet, with the U-Net residual model being the most competitive. The clavicle segmentation is the most challenging task for the evaluated models, with ResNet-UNet-18 being the most competitive model reaching 0.525 average Dice coefficient, outperforming the base U-Net model that only reaches 0.481 for this task. For the rib segmentation case, the Residual U-Net network is the only one that outperforms the U-Net base model. It is important to note that the Residual U-Net encoder variation consistently outperforms the base U-Net model in all four tasks evaluated (lung, rib, heart and clavicle segmentation). The ResNet encoder variations outperform the base U-net model in three out of four tasks (no ribs). In contrast, the Swin-UNet model outperforms the base U-Net model only in lung segmentation and by a minimal margin.

Table 3 shows the results of using transfer learning to pre-train the models using the MC-SH dataset (for lung segmentation) since it is the largest dataset used in this work. In the table, it can be observed that, although transfer learning does not have a significant impact except in the case of clavicle segmentation, the most competitive models for each of the tasks (rib, clavicle and heart segmentation) remain similar to those presented in Table 2. In this sense, for lung and clavicle segmentation, the most suitable encoder backbone turns out to be ResNet-UNet-18. At the same time, Residual U-Net is the encoder backbone for heart and rib segmentation.

Table 3 Encoder variations over MC-SH pre-trained weights

	Ribs	Heart	Clavicles
U-Net	0.870	0.805	0.521
Residual U-Net	**0.872**	**0.818**	0.524
ResNet-UNet-18	0.865	0.815	**0.532**
ResNet-UNet-34	0.857	0.813	0.524
ResNet-UNet-50	0.856	0.801	0.520
Swin-UNet	0.788	0.801	0.480

5.2 Attention module variations

We use the same procedure for the attention module variations as encoder variation evaluation from scratch and MC-SH pre-trained.

Table 4 shows the positive impact of using attention mechanisms on the base U-Net model for lung, heart, rib and clavicle segmentation from scratch. In this regard, it is also observed that there is no standard attention module that improves the segmentation capacity of all organs at the same time. Thus, the full-spatial-attention module is the most suitable for the lungs and heart. In contrast, the three-head-attention module best suits ribs and the spatial-cross attention module for clavicles.

Additionally, Table 5 shows the impact of transfer learning with attention modules on the base U-net model for rib, heart and clavicle segmentation when the models are pre-trained on the MC-SH dataset. It is interesting to note that transfer learning has no significant impact on the segmentation of heart and ribs, in the latter case, even a factor that decreases the segmentation quality, going from 0.880 to 0.875. However, it is also important to note the positive impact of using transfer learning combined with attentional modules for clavicle segmentation on the base U-Net model.

6 Discussion

This paper presents a robust experimental framework for evaluating the impact of backbone architecture and attention mechanism on the original U-net architecture for segmentation of lungs, heart, clavicle and ribs, achieving up to 99% overlapping against hand segmented

Table 4 Attention module variations from scratch

	Lungs	Ribs	Heart	Clavicles
U-Net	0.930	0.876	0.798	0.481
Three-Head Attention U-Net	0.948	**0.880**	0.812	0.485
Spatial Attention U-Net	**0.959**	0.870	0.807	0.475
Double Spatial Attention U-Net	0.946	0.863	0.803	0.493
Full Spatial Attention U-Net	**0.959**	0.872	**0.817**	0.489
Swin Spatial Attention U-Net	0.926	0.878	0.808	0.513
Spatial Cross-Attention U-Net	0.800	0.851	0.810	**0.538**
Multiple Spatial Cross-Attention U-Net	0.820	0.853	0.775	0.415

We present the mean Dice coefficient value for each organ, obtained using the JSRT dataset for clavicles and heart segmentation, the Montgomery County dataset for lung segmentation, and VinDr-RibCXR for rib segmentation

Table 5 Attention module variations over MC-SH pre-trained weights

	Ribs	Heart	Clavicles
U-Net	0.870	0.805	0.521
Three-Head Attention U-Net	0.873	0.799	0.538
Spatial Attention U-Net	**0.875**	0.809	**0.545**
Double Spatial Attention U-Net	0.858	0.799	0.481
Full Spatial Attention U-Net	0.865	**0.820**	0.503
Swin Spatial Attention U-Net	0.868	0.800	0.473
Spatial Cross-Attention U-Net	0.862	0.818	0.517
Multiple Spatial Cross-Attention U-Net	0.828	0.770	0.494

We present the mean Dice coefficient value for each organ, obtained using the JSRT dataset for clavicles and heart segmentation, the Montgomery County dataset for lung segmentation, and VinDr-RibCXR for rib segmentation

masks, for lung segmentation in particular. It is essential to realize that with this work, we are aimed to have a baseline comparison to understand the impact of varying the encoder and attentional mechanisms on the original U-net model.

First, we studied the impact of varying the encoder on the original U-net for segmentation of lungs, heart, ribs and clavicles in chest x-ray images. Our experiments showed that the more accurate results were achieved using U-net with different encoder backbones without transfer learning, getting up to 0.991, 0.957, 0.872, and 0.536 of Dice coefficient to segment lungs, ribs, heart, and clavicles, respectively. Figure 9 shows an image gallery with some segmentation results using the best approach presented in the previous section (Table 2), considering lung, rib, heart, and clavicle segmentation using encoder variations. We used

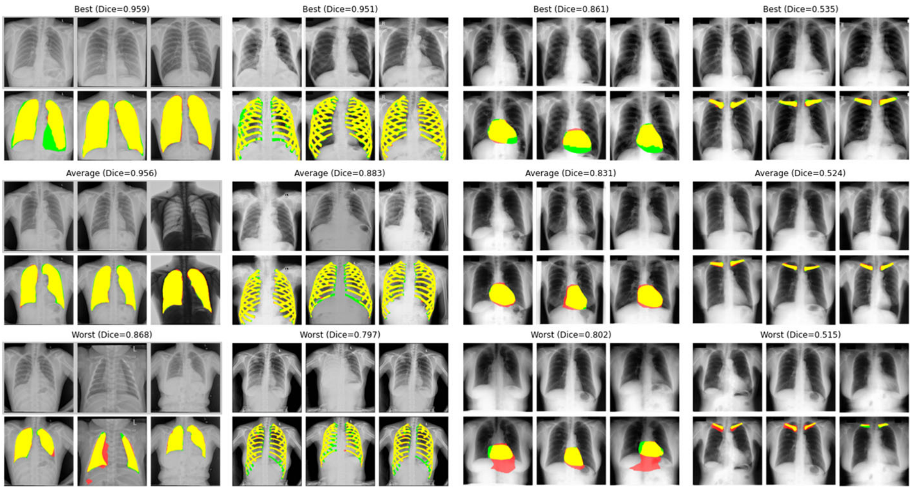

Fig. 9 Results of lung, ribs, heart and clavicle segmentation using encoder variations. Upper row shows representatives for the best results, the middle row for average results, and the last row for the worst results. For each organ (lung, ribs, heart and clavicles), we present the original (first row), and the result obtained using the best model (second row). The green color represents the gold-standard, red presents the best model segmentation result from scratch, and yellow is the overlap between hand-segmented and automatic model masks

ResNet-UNet-18 for the lungs and clavicles and residual U-net for the ribs and the heart. For each segmentation (lung, rib, heart, and clavicle), we present the best, average, and worst results regarding the Dice coefficient as an evaluation metric.

Second, we assessed the impact of attention strategies fairly, using the original U-net model as a baseline and avoiding other strategies that can disturb our results, such as data augmentation. In this sense, our experiments showed that the more accurate results were achieved using U-net with attention modules with transfer learning for segmentation of lungs and ribs, getting up 0.996 and 0.958 of Dice coefficient, respectively. In the case of segmentation of the heart and clavicles, the more accurate results were achieved using attention modules over U-net architecture with transfer learning, getting up 0.858 and 0.547 of the Dice coefficient, respectively. Figure 10 shows an image gallery with some segmentation results using the best approach from scratch presented in the previous section (Table 4), considering lung, rib, heart, and clavicle segmentation using attention mechanisms from scratch. In this sense, we used Spatial Attention U-Net for the lungs, Three-Head Attention U-Net for the ribs, Full Spatial Attention U-Net for the heart, and Spatial Cross-Attention U-Net for the clavicles. For each segmentation (lung, rib, heart, and clavicle), we present the best, average, and worst results regarding the Dice coefficient as an evaluation metric.

As described in Tables 4 and 5, the three-head attention models (from second to fifth row) present superior results with respect to the cross-attention models (last three rows). This fact, could be due to the channel-wise attention included by the first models, which produces a kind of multiattention guided by each channel. We also observe that adding an extra attention module after the Double Spatial Attention model, does not improve the results. This fact is observed for all the experiments. Finally, we show the benefits of including transformer-based three-head attention, replacing the pseudo-attention proposed by the Three-Head Attention U-Net. We achieved segmentation improvements for lungs, heart and clavicles.

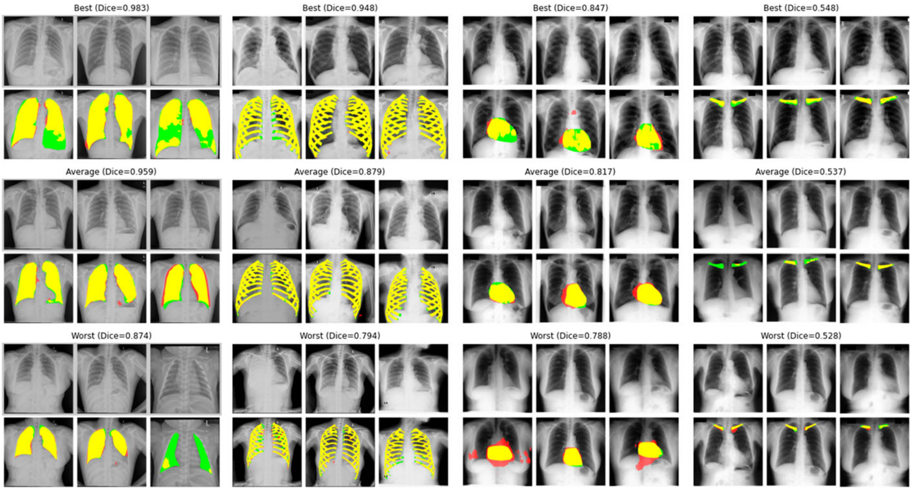

Fig. 10 Results of lung, ribs, heart and clavicle segmentation using attention mechanisms. Upper row shows representatives for the best results, the middle row for average results, and the last row for the worst results (second row). For each organ (lung, ribs, heart and clavicles), we present the original (first row), and the result obtained using the best model (second row). The green color represents the gold-standard, red presents the best model segmentation result from scratch, and yellow is the overlap between hand-segmented and automatic model masks

From the statistical viewpoint, our experiments showed that the distribution of values for the Dice coefficient achieved by the best encoder for clavicle segmentation (with $\sigma = 0.005$) presents lower Dice coefficients but provides a lower variance also. However, the rib segmentation (with $\sigma = 0.039$) provides higher Dice coefficients with a higher variance. The same phenomenon is observed when analyzing the impact of using attentional strategies, as the distribution of Dice coefficients has a lower variance in the case of clavicles (with $\sigma = 0.005$). However, it is also the organ with lower Dice coefficient values. On the opposite, the case of lung segmentation where the variance is higher (with $\sigma = 0.033$) but reaches Dice coefficient values of up to 0.996 in the best case.

Furthermore, by comparing Tables 2 and 4 (together with Figs. 9 and 10), it can be seen that both the encoder and the attention mechanism are not universal for the segmentation of different organs in chest X-ray images. While some organs share the same backbone architecture (ResNet-UNet-18 for lungs and clavicles), and the transfer learning has no positive impact on the results, that is not true for the case of attention mechanisms. In this last scenario, the most appropriate attention mechanism for lungs and clavicles is the shared (as for the encoder variation), Spatial Attention U-Net; however, the best results using this attention mechanism for segmenting lungs were achieved from scratch, while for clavicle segmentation there was transfer learning. According to our experiments, it turns out to be different attentional mechanisms and transfer learning used for heart and ribs segmentation: Three-head attention U-Net from scratch for ribs and Full Spatial Attention U-Net for the heart.

Additionally, according to the experimental framework, the attention mechanisms positively impact lung and clavicle segmentation, using the original U-net architecture from scratch and compared to the best encoder variation of the U-net, also from scratch. On the other hand, the quality of heart and rib segmentation is not positively affected by the use of attention mechanisms over the original U-Net architecture from scratch, being that just varying the encoder used in the U-Net outperforms, in terms of Dice coefficient, the use of attention mechanisms.

Several studies have explored the role of deep learning in chest segmentation in different image modalities, including x-rays. However, more studies have focused on MRI and CT rather than x-rays [47, 48], showing that the availability of annotated datasets significantly impacts researching specific radiology problems. It is not a novelty that this impact is also reflected in research on segmentation methods on chest x-ray images. More curated data is needed to develop reliable solutions for potential implementation in real clinical settings. In this scenario, the challenge is to develop models that work with fewer data taking advantage of strategies such as transfer learning, data augmentation, and modern algorithms.

It is essential to realize that the chest is a complex body region that includes the heart, ribs, clavicles, lungs, vessels, and trachea. However, research studies ignored most of them, focusing on the segmentation of lungs, lung nodules and pathological findings in tuberculosis [49]. In this way, our research expands the knowledge of the performance of deep learning in the segmentation of the heart, clavicles and ribs.

The most used architecture for segmentation in biomedical images is the U-Net; however, it has limitations and novel architectures are needed. Recently, and as it is investigated in our work, ResNets have been jointly developed with U-Net for improving segmentation in medical images. As shown from the tables above, ResNets used as the encoder improved the Dice coefficient in lung segmentation.

Our work also contributes to chest x-ray image segmentation due to the extensive study on the role and application of attention mechanisms in U-Net structures. Previous works have also explored the impact of attention mechanisms with promising results (lungs 0.98, heart 0.91 and clavicles 0.87 of Dice coefficient). The difference with our research work is that we used different public datasets, evaluated a variety of organs in chest x-ray images, and assessed the impact of attention mechanisms fairly with a standard backbone and with no other influential factors such as a strong data augmentation or hyper-parameter fine-tuning, achieving mean Dice coefficient 0.96 for lung, 0.82 for heart, 0.88 for ribs, and 0.55 for clavicle segmentation. A limitation of this and previous works is that the results achieved are not comparable between pieces of work due to the lack of gold-standard datasets. In this way, future standardized datasets are needed to establish state-of-the-art results and architectures and objectively validate the performance of the models. This present work is an initial kick-off to uniform the study of the impact of some factors (aka encoder variation and attention mechanisms) while trying to take advantage of the limited gold-standard datasets.

7 Summary and conclusions

We have evaluated the impact of attention mechanisms and encoder variations over the original U-Net architecture for segmenting lungs, heart, clavicles and ribs in chest x-ray images. Our methodology takes the original U-Net model as a baseline, which is then modified by different encoder architectures and attention strategies. Our results show the positive impact of encoder variations and attention mechanism over a U-Net for segmentation of lungs, ribs, heart and clavicles from scratch, without transfer learning. Moreover, the encoder and the attention mechanism are not universal for segmenting different organs in chest x-ray images. While there are organs that share the same backbone architecture (lungs and clavicles), the most appropriate attention mechanisms for each organ turn out to be all different, achieving up to 99% of overlapping in lung segmentation. In addition, we noted that pretraining the models with a lung segmentation dataset only showed marginal improvements. An interesting remark about the low performance of clavicles concerning the other organs suggests a more profound study that goes beyond this work. In conclusion, segmentation is a critical step in chest x-ray image analysis. Our work contributes to the field due to the extensive research on using ResNet-Unet architecture and implementing attentional modules. Furthermore, we used all the available datasets to train and test the models and explored other chest areas such as ribs, heart and clavicles. Finally, efficient segmentation could facilitate acquisition techniques surveillance, reducing patient recitation and increasing DL classifications accuracy in pathology detection. However, more research is needed to develop a reliable system that could be applied to clinical practice. For the last, more extensive datasets, more diverse and with high-quality annotations, are needed. Meanwhile, the field will require appropriate data augmentation, transfer learning and efficient models, for example, semi or self-supervised learning and attentional mechanisms, to overcome these limitations.

Appendix A: Training plots

Figures 11, 12, 13 and 14 show training graphics for lungs, heart, clavicles and ribs, respectively.

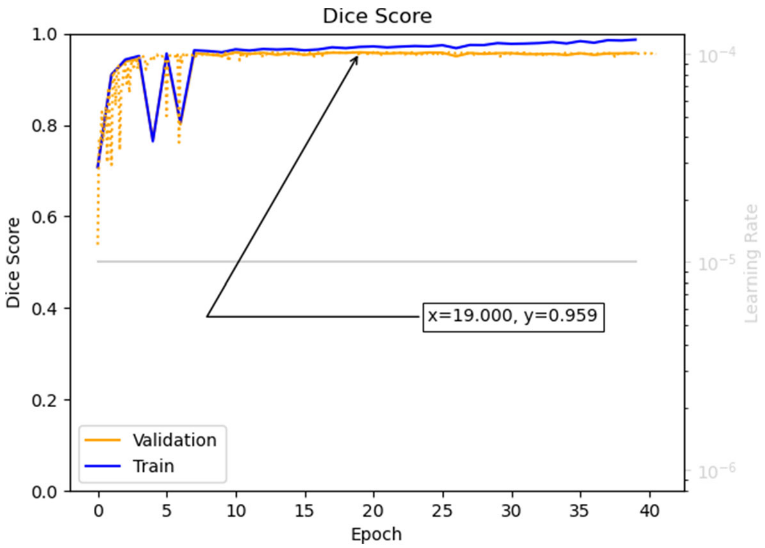

Fig. 11 Lung Spatial Attention model training Dice

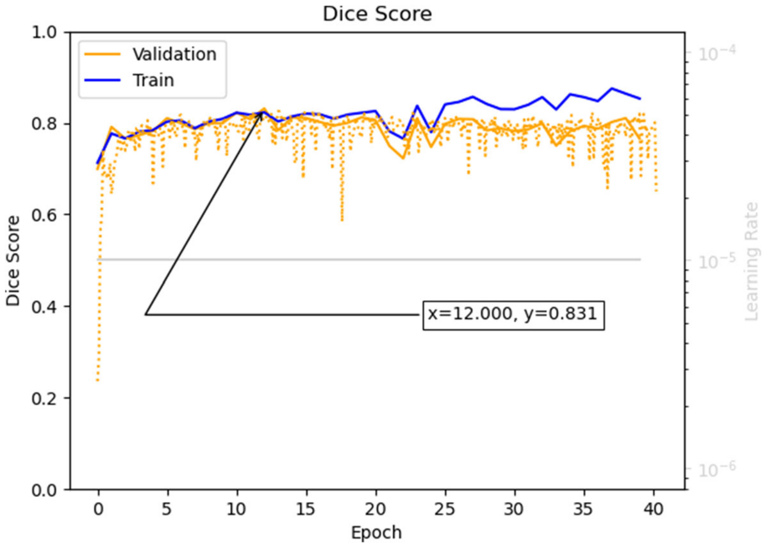

Fig. 12 Heart Residual UNet training Dice

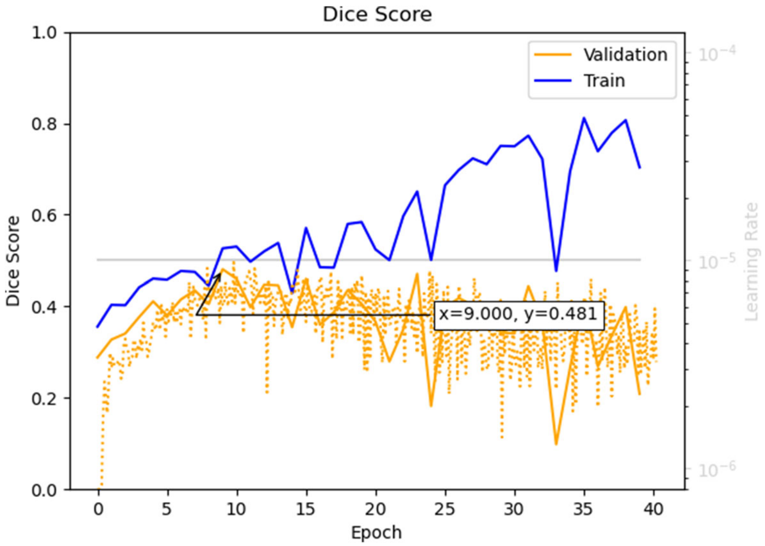

Fig. 13 Clavicle Spatial Attention model training Dice

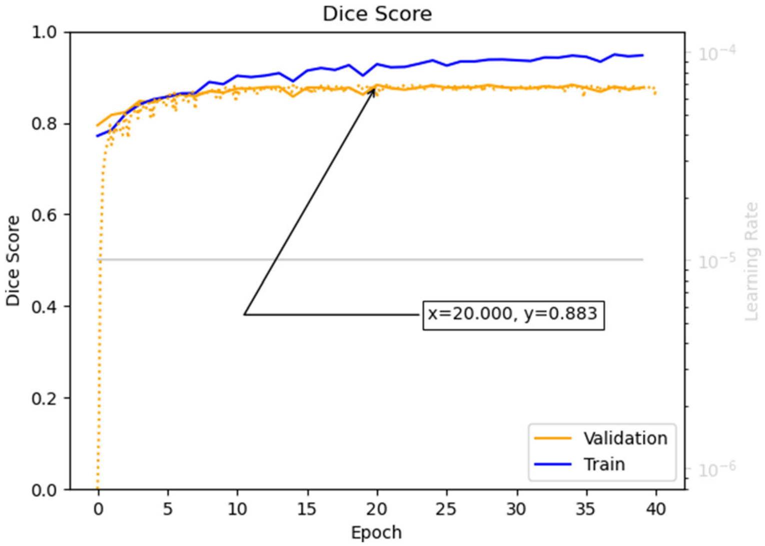

Fig. 14 Rib Residual UNet model training Dice

Data Availability The authors declare that the data supporting the findings of this study are available within the paper.

Declarations

Conflict of interest/Competing interests The authors have no relevant financial or non-financial interests to disclose. The authors have no competing interests to declare that are relevant to the content of this article. All authors certify that they have no affiliations with or involvement in any organization or entity with any financial interest or non-financial interest in the subject matter or materials discussed in this manuscript. The authors have no financial or proprietary interests in any material discussed in this article.

References

1. Ganguly D, Chakraborty S, Balitanas M, Kim Th (2010) Communications in Computer and Information Science. Medical Imaging: A review, vol 78, pp 504–516, Springer, Berlin, Heidelberg
2. Zdora M (2021) Principles of X-ray Imaging. Springer Theses, Springer, Cham, pp 11–57
3. Hong S, Yoon D, Lim K, Moon J, Yoon S, Seo Y, Yun E (2019) Radiological clinical practice guidelines published in the last decade: a bibliometric analysis. Journal of the Belgian Society of Radiology, 103(1)
4. Brady AP, Beets-Tan RG, Brkljačič B, Catalano C, Rockall A, Fuchsjäger M (2022) The role of radiologist in the changing world of healthcare: a white paper of the european society of radiology ESR. Insights Imaging 13(1):9167391
5. Jan J (2005) Medical Image Processing, Reconstruction and Restoration: Concepts and methods. CRC Press, Boca Raton, Fla
6. Röntgen W (1896) On a new kind of rays. Sci, 3(59):227–231
7. Seibert J (1997) The AAPM/RSNA physics tutorial for residents x-ray generators. Radiographics 17(6):1533–1557
8. Ou X, Chen X, Xu X, Xie L, Chen X, Hong Z, Bai H, Liu X, Chen Q, Li L, Yang H (2021) Recent development in x-ray imaging technology: Future and challenges. Research, 2021–9892152
9. Sussmann A, Ko J (2010) Understanding chest radiographic anatomy with MDCT reformations. Clin Radiol 65(2):155–166
10. Broder J (2011) Chapter 5 - imaging the chest: The chest radiograph. Diagnostic Imaging for the Emergency Physician. W.B. Saunders, Philadelphia, US, pp 185–296
11. Goodman M, Huber N, Johannigman J, Pritts T (2010) Omission of routine chest x-ray after chest tube removal is safe in selected trauma patients. Am J Surg, 199(2):199–203
12. Waite S, Grigorian A, Alexander R, Macknik S, Carrasco M, Heeger D, Martinez-Conde S (2019) Analysis of perceptual expertise in radiology - current knowledge and a new perspective. Front Hum Neurosci 13:2019–00213
13. Gunderman R, Williamson K, Fraley R, Steele J (2001) Expertise: Implications for radiological education. Acad Radiol 8(12):1252–1256
14. Zhang B, Rahmatullah B, Wang S, Zhang G, Wang H, Ebrahim N (2021) A bibliometric of publication trends in medical image segmentation: quantitative and qualitative analysis. J Appl Clin Med Physicss 22(10):45–65
15. Sander J, de Vos B, Išgum I (2020) Automatic segmentation with detection of local segmentation failures in cardiac MRI. Sci Rep 10(1):2020–21769
16. Tunguturi M, Singu S (2022) Affectation index and severity degree by covid-19 in chest x-ray images using artificial intelligence. Int J Innovations Eng Res & Technol, 9(9):37–43
17. Sun H, Ren G, Teng X, Song L, Li K, Yang J, Hu X, Zhan Y, Wan SB, Wong M, Chan K, Tsang H, Xu L, Wu T, Kong F-M, Wang Y, Qin J, Chan W, Ying M, Cai J (2023) Artificial intelligence-assisted multistrategy image enhancement of chest x-rays for covid-19 classification. Quant Imaging Med Surg 13(1):394–416
18. Satia I, Bashagha S, Bibi A, Ahmed R, Mellor S, Zaman F (2013) Assessing the accuracy and certainty in interpreting chest x-rays in the medical division. Clin Med 13(4):349–352
19. Cheng P, Montagnon E, Yamashita R, Pan I, Cadrin-Chênevert A, Perdigón F, Chartrand G, Kadoury S, Tang A (2021) Deep learning: An update for radiologists. Radiographics 41(5):1427–1445
20. Zhou S (2015) Medical Image Recognition, Segmentation and Parsing: Machine learning and multiple object approaches. Academic Press Inc, Cambridge, Massachusetts

21. Iglesias J, Sabuncu M (2015) Multi-atlas segmentation of biomedical images: A survey. Med Image Anal 24(1):205–219
22. Sharma N, Ray A, Shukla K, Sharma S, Pradhan S, Srivastva A, Aggarwal L (2010) Automated medical image segmentation techniques. J Med Phys, 35(1):3–14
23. Ronneberger O, Fischer P, Brox T (2015) U-net: Convolutional networks for biomedical image segmentation. In: Proceedings of medical image computing and computer-assisted intervention (MICCAI), pp 234–241
24. Long J, Shelhamer E, Darrell T (2015) Fully convolutional networks for semantic segmentation. In: Proceedings of the international conference on computer vision and pattern recognition (CVPR), pp 3431–3440
25. Vaswani A, Shazeer N, Parmar N, Uszkoreit J, Jones L, Gomez A, Kaiser L, Polosukhin I (2017) Attention is all you need. In: Proceedings of the 31st international conference on neural information processing systems, pp 6000–6010
26. Palmer S (1999) Vision Science?: Photons to phenomenology. MIT Press, Cambridge, Massachusetts
27. Rashid R, Akram MU, Hassan T (2018) Fully convolutional neural network for lungs segmentation from chest x-rays. In: Proceedings of the international conference image analysis and recognition, pp 71–80
28. Mittal A, Hooda R, Sofat S (2018) Lf-segnet: A fully convolutional encoder-decoder network for segmenting lung fields from chest radiographs. Wireless Pers Commun 101(1):511–529
29. Yahyatabar M, Jouvet P, Cheriet F (2020) Dense-Unet: a light model for lung fields segmentation in chest x-ray images. In: Proceedings of the 42nd annual international conference of the IEEE engineering in medicine & biology society (EMBC), pp 1242–1245
30. Liu W, Luo J, Yang Y, Wang W, JD, Yu L (2022) Automatic lung segmentation in chest x-ray images using improved u-net. Scientific Reports 12(8649)
31. Pal D, Balakrishna P, Roy S (2022) Attention UW-Net: A fully connected model for automatic segmentation and annotation of chest x-ray. Comput Biol Med 150:2022–106083
32. Wang W, Feng H, Bu Q, Cui L, Xie Y, Zhang A, Feng J, Zhu Z, Chen Z (2020) MDU-Net: a convolutional network for clavicle and rib segmentation from a chest radiograph. Journal of Healthcare Engineering 2020:1–9
33. Novikov A, Lenis D, Major D, Hladuvka J, Wimmer M, Buhler K (2018) Fully convolutional architectures for multiclass segmentation in chest radiographs. IEEE Trans Med Imaging 37(8):1865–1876
34. Ullah I, Ali F, Shah B, El-Sappagh S, Abuhmed T, Park S (2023) A deep learning based dual encoder-decoder framework for anatomical structure segmentation in chest x-ray images. Sci Rep 13:2023–791
35. Shiraishi J, Katsuragawa S, Ikezoe J, Matsumoto T, Kobayashi T, Komatsu K-i, Matsui M, Fujita H, Kodera Y, Doi K (2000) Development of a digital image database for chest radiographs with and without a lung nodule. Am J Roentgenol, 174(1):71–74
36. van Ginneken B, Stegmann M, Loog M (2006) Segmentation of anatomical structures in chest radiographs using supervised methods: a comparative study on a public database. Med Image Anal 10(1):19–40
37. Jaeger S, Candemir S, Antani S, Wang Y, Lu P, Thoma G (2014) Two public chest x-ray datasets for computer-aided screening of pulmonary diseases. Quant Imaging Med Surg 4(6):475–477
38. Jaeger S, Karargyris A, Candemir S, Folio L, Siegelman J, Callaghan F, Xue Z, Palaniappan K, Singh RK, Antani S, Thoma G, Wang Y-X, Lu P-X, McDonald CJ (2014) Automatic tuberculosis screening using chest radiographs. IEEE Trans Med Imaging 33(2):233–245
39. Nguyen H, Le T, Pham H, Nguyen H (2021) Vindr-ribcxr: A benchmark dataset for automatic segmentation and labeling of individual ribs on chest x-rays. In: Proceedings of the medical imaging with deep learning
40. Koffka K (1935) Principles Of Gestalt Psychology. Routledge, London, UK
41. Dosovitskiy A, Beyer L, Kolesnikov A, Weissenborn D, Zhai X, Unterthiner T, Dehghani M, Minderer M, Heigold G, Gelly S, Uszkoreit J, Houlsby N (2021) An image is worth 16x16 words: Transformers for image recognition at scale. In: Proceedings of the 9th international conference on learning representations ICLR
42. Cao F, Zhao H (2021) Automatic lung segmentation algorithm on chest x-ray images based on fusion variational auto-encoder and three-terminal attention mechanism. Symmetry 13(5):814
43. Liu Z, Lin Y, Cao Y, Hu H, Wei Y, Zhang Z, Lin S, Guo B (2021) Swin transformer: Hierarchical vision transformer using shifted windows. In: Proceedings of the international conference on computer vision (ICCV), pp 9992–10002
44. Dice L (1945) Measures of the amount of ecologic association between species. Ecol 26(3):297–302
45. Rote G (1991) Computing the minimum hausdorff distance between two point sets on a line under translation. Inf Process Lett 38(3):123–127
46. He K, Zhang X, Ren S, Sun J (2016) Deep residual learning for image recognition. In: Proceedings of the IEEE Conference on computer vision and pattern recognition, pp 770–778

47. Astley J, Wild J, Tahir B (2021) Deep learning in structural and functional lung image analysis. Br J Radiol 95:20201107
48. Agrawal T, Choudhary P (2022) Segmentation and classification on chest radiography: a systematic survey. Vis Comput 39:1–39
49. Chavan M, Varadarajan V, Gite S, Kotecha K (2022) Deep neural network for lung image segmentation on chest x-ray. Technologies, 10(5)

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Authors and Affiliations

Tomás de la Sotta[1] · Violeta Chang[2] · Benjamín Pizarro[1,3] · Héctor Henriquez[4,5] · Nicolás Alvear[1] · Jose M. Saavedra[4]

Tomás de la Sotta
tomas@retinarx.cl

Violeta Chang
violeta.chang@usach.cl

Benjamín Pizarro
benjamin@retinarx.cl

Héctor Henriquez
hhenriquez@miuandes.cl

Nicolás Alvear
nicolas@retinarx.cl

[1] RetinaRX, Quimera 231, Valparaíso, V, Chile

[2] Departamento de Ingeniería Informática, Universidad de Santiago de Chile, Av. Victor Jara 3659, Estación Central, RM, Chile

[3] Facultad de Medicina, Universidad de Chile, Av. Independencia 1027, Independencia, Santiago, 646 RM, Chile

[4] Universidad de los Andes, Monseñor Álvaro del Portillo 12455, Santiago, RM, Chile

[5] Servicio de Radiología, Clínica Santa María, Av. Santa María 0500, Providencia, Santiago, RM, 650, Chile