Article

Feet Segmentation for Regional Analgesia Monitoring using Convolutional RFF and Layer-Wise Weighted CAM Interpretability

Juan Carlos Aguirre-Arango[1] *, Andrés Marino Álvarez-Meza[1] and German Castellanos-Dominguez[1]

[1]Signal Processing and Recognition Group, Universidad Nacional de Colombia, Manizales 170003, Colombia; jucaguirrear@unal.edu.co (J.A.-A.); amalvarezme@unal.edu.co (A.A.-M.); cgcastellanosd@unal.edu.co (G.C.-D.)

Abstract: The administration of regional neuraxial analgesia for pain relief during labor is widely recognized as a safe and effective method involving medication delivery into the epidural or subarachnoid space in the lower back. This study proposes an innovative semantic image segmentation methodology emphasizing enhanced interpretability using convolutional Random Fourier Features and layer-wise weighted class-activation maps tailored explicitly for foot segmentation in regional analgesia monitoring. Namely, our contribution is twofold: i) a novel Random Fourier Features layer is introduced to deal with image data to enhance three well-known architectures (FCN, UNet, and ResUNet); ii) three novel quantitive measures are presented to evaluate the interpretability of a given deep learning model devoted to segmentation tasks. Our approach is rigorously evaluated on a demanding dataset of foot thermal images from pregnant women who received epidural anesthesia. Its small size and considerable variability characterize the dataset. Our validation results demonstrate that the proposed methodology not only achieves competitive foot segmentation performance but also significantly enhances the explainability of the process, rendering it well-suited for applications such as epidural insertion during labor.

Keywords: Infrared Thermal Segmentation, Regional Neuraxial Analgesia, Deep Learning, Random Fourier Features, Class Activation Maps

1. Introduction

The use of regional neuraxial analgesia for pain relief during labor is widely acknowledged as a safe method [1]. It involves the administration of medication into the epidural or subarachnoid space in the lower back. This procedure blocks pain signals from the uterus and cervix to the brain. This method is considered safe and effective for most women and is associated with lower rates of complications than other forms of pain relief [1,2]. Electrophysiological testing measures nerve fiber reactions to painful stimuli with electromyography, excitatory or inhibitory reflexes, evoked potentials, electroencephalography, and magnetoencephalography [3]. In addition, imaging techniques objectively measure relevant bodily function patterns (such as blood flow, oxygen use, and sugar metabolism) using positron emission tomography (PET), single-photon emission computed tomography (SPECT), and functional magnetic resonance imaging (fMRI) [4].

Nonetheless, imaging techniques can be costly and are generally prohibited in obstetric patients, limiting their use. A cost-effective alternative approach is utilizing thermographic skin images to measure body temperature and predict the distribution and efficacy of epidural anesthesia [5]. This approach is achieved by identifying areas of cold sensation [6]. The use of thermal imaging provides an objective and non-invasive solution to assess warm modifications resulting from blood flow redistribution after catheter placement [7]. However, an adequate assessment requires temperature measurements from the patient's foot soles at various times after catheter placement to accurately characterize early thermal modifications [8,9]. Regarding this, semantic segmentation of feet in infrared thermal

 doi:10.20944/preprints202305.0670.v1

images in obstetric environments is challenging due to various factors. Firstly, thermal images possess inherent characteristics such as low contrast, blurred edges, and uneven intensity distribution, making it difficult to identify objects accurately [10,11]. The second challenge is the high variability of foot position in clinical settings. Additionally, the specialized equipment required for collecting these images and the limited willingness of mothers to participate in research studies resulted in a need for more available samples and the challenge of acquiring annotated data, which is crucial for developing effective segmentation techniques.

Semantic segmentation is crucial in medical image analysis, with deep learning widely used. Fully Convolutional Networks (FCN) [12] is a popular approach that uses Convolutional layers for pixel-wise classification but produces coarse Region of Interest (ROI) and poor boundary definitions for medical images [13]. Likewise, U-Net [14] consists of encoders and decoders that handle objects of varying scales but have difficulty dealing with opaque or unclear goal masks [15]. U-Net++ [16] extends U-Net with nested skip connections for highly accurate segmentation but with increased complexity and overfitting risk. Besides, SegNet [17] is an encoder-decoder architecture that handles objects of different scales but cannot handle fine details. Mask R-CNN [18] extends Faster R-CNN [19] for instance segmentation with high accuracy but requires a large amount of training data and has high computational complexity. On the other hand, PSPNet uses a pyramid pooling module for multi-scale contextual information and increased accuracy but with high computational complexity and a tendency to produce fragmented segmentation maps for small objects [20].

Specifically for semantic segmentation of feet from infrared thermal images, most works were developed in the context of diabetic foot disorders. In [21] authors combine RGB, infrared, and depth images to perform plantar foot segmentation based on a U-Net architecture together with RANdom SAmple Consensus (RANSAC) [22], which relies too much on depth information. The authors in [23] use a similar approach to integrating thermal and RGB images to be fed into a U-Net model. Their experiments show that RGB images help in more complex cases. In [24], the authors compare multiple models on thermal images, including U-Net, Segnet, FCN, and prior shape active contour-based methodology, proving Segnet outperforms them all. Alike, in [25], authors compare multiple infrared thermographic feet segmentation models using transfer learning and removal algorithms based on morphological operations on U-Net, FCN, and Segnet, showing that Segnet outperforms the rest of the models but with high computational cost.

On the other hand, Visual Transformers (VIT) [26] have revolutionized self-attention mechanisms to identify long-range image dependencies. Several recent works have leveraged VIT capabilities to enhance global image representation. For instance, in [27], a U-Net architecture fused with a VIT-based transformer significantly improves model performance. However, this approach requires a pre-trained model and many iterations. Similarly, in [28], a pure U-Net-like transformer is proposed to capture long-range dependencies. Another recent work [29] suggests parallel branches, one based on transformers to capture long-range dependencies and the other on CNN to conserve high resolution. The authors of [30] propose a squeeze-and-expansion transformer that combines local and global information to handle diverse representations effectively. This method has unlimited practical receptive fields, even at high feature resolutions. However, it relies on a large dataset and has higher computational costs than conventional methods. To address the data-hungry nature of transformer-based models, the work in [31] proposes a semi-supervised cross-teaching approach between CNN and Transformers. The most recent work in this field, Meta Segment Anything [32], relies on an extensive natural database (around 1B images) for general segmentation. However, medical and natural images have noticeable differences, including color and blurriness. It is also pertinent to note that accepting ambiguity can incorporate regions that may not be part of the regions of interest. Specifically, transformers for semantic segmentation of feet still need to be tested for low-size databases.

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 9 May 2023 doi:10.20944/preprints202305.0670.v1

This study presents a cutting-edge Convolutional Random Fourier Features (CRFFg) technique for foot segmentation in thermal images, leveraging layer-wise weighted class activation maps. Our proposed data-driven method integrates Random Fourier Features within a convolutional framework, enabling weight updates through gradient descent. To assess the efficacy of our approach, we benchmark it against three widely-used architectures: U-Net [14], FCN [12], and ResUNet [33]. We enhance these architectures by incorporating CRFFg at the skip connections, bolsters representation, and facilitate the fusion of low-level semantics from the decoder to the encoder. Moreover, we introduce a layer-wise strategy for quantitatively analyzing Class Activation Maps (CAMs) for semantic segmentation tasks [34]. Our experimental findings showcase the competitive performance of our models and the accurate quantitative assessment of CAMs. The proposed CRFFg method offers a promising solution for foot segmentation in thermal images, tailored explicitly for regional analgesia monitoring. Additionally, layer-wise weighted class activation maps contribute to a more comprehensive understanding of feature representations within neural networks.

The paper is organized as follows: Section 2 describes the materials and methods used in the study. Sections 3 and 4 present the experimental setup and results, respectively, followed by Section 5, which provides the concluding remarks.

2. Material and Methods

2.1. Deep Learning for Semantic Segmentation

Provided an image set, $\{\boldsymbol{I}_n \in \mathbb{R}^{H\times\tilde{W}\times C} : n \in N\}$, we will call a label mask the corresponding matrix $\boldsymbol{M}_n$ that encodes the membership of each n-th image pixel to a particular class, where H is height, $\tilde{W}$ is width, and C holds the color channels of the image set. For simplicity, $C = 1$ is assumed. As regards the semantic segmentation task under consideration, each mask is binary, $\boldsymbol{M} \in \{0,1\}^{H\times\tilde{W}}$, representing either the background or the foreground.

An estimate for matrix mask $\hat{\boldsymbol{M}} \in [0,1]^{H\times\tilde{W}}$ can be obtained through deep learning models for semantic segmentation, stacking convolutional layers as follows:

$$\hat{\boldsymbol{M}} = (\varphi_L \circ \cdots \circ \varphi_L)(\boldsymbol{I}) \tag{1}$$

where $\varphi_l : \mathbb{R}^{H_{l-1}\times\tilde{W}_{l-1}\times D_{l-1}} \to \mathbb{R}^{H_l\times\tilde{W}_l\times D_l}$ denotes a function composition for the l-th layer ($l \in L$), which comprises learnable parameters represented by $\boldsymbol{W}_l \in \mathbb{R}^{\tilde{k}_l\times\tilde{k}_l\times D_{l-1}\times D_l}$ and $\boldsymbol{b}_l \in \mathbb{R}^{D_l}$ ($\tilde{k}_l$ holds the l-th convolutional kernel size). Of note, the feature map $\boldsymbol{F}_l = \varphi_l(\boldsymbol{F}_{l-1}) = \varsigma_l(\boldsymbol{W}_l \otimes \boldsymbol{F}_{l-1} + \boldsymbol{b}_l) \in \mathbb{R}^{H_l\times\tilde{W}_l\times D_l}$, is comprised of D_l distinct features extracted, $\varsigma_l(\cdot)$ is a nonlinear activiation function, and $\otimes$ stands for image-based convolution. Essentially, the function composition in Eq. 1 transforms the input feature map from the previous layer, $(l-1)$, into the output feature map for the current layer, l, by employing the learnable parameters $\boldsymbol{W}_l$ and $\boldsymbol{b}_l$. The resulting $\boldsymbol{F}_l$ captures the salient information within the l-th network layer.

The parameter set $\Theta = \{\boldsymbol{W}_l, \boldsymbol{b}_l : l \in L\}$ is estimated within the following optimizing framework [35]:

$$\Theta^* = \arg\min_{\Theta} \mathbb{E}\{\mathcal{L}\{\boldsymbol{M}_n, \hat{\boldsymbol{M}}_n|\Theta\} : \forall n \in N\}, \tag{2}$$

where $\mathcal{L} : \{0,1\}^{H\times\tilde{W}} \times [0,1]^{H\times\tilde{W}} \to \mathbb{R}$ in Eq. 2 is a given loss function and notation $\mathbb{E}\{\cdot\}$ stands for the expectation operator.

2.2. Convolutional Random Fourier Features Gradient - CRFFg

Random Fourier Features establish a finite-dimensional, explicit mapping that approximates shift-invariant kernels $k(\cdot)$ as described in Rahimi et al. (2009) [36]. This explicit mapping, denoted by $z : \mathbb{R}^{\tilde{Q}} \to \mathbb{R}^Q$, serves to transform the input space into a finite-dimensional space $\mathcal{H} \subset \mathbb{R}^Q$, where the inner product can be obtained as:

$$k(\boldsymbol{x} - \boldsymbol{x}') = \langle \phi(\boldsymbol{x}), \phi(\boldsymbol{x}') \rangle_{\mathcal{H}} \approx z(\boldsymbol{x})^\top z(\boldsymbol{x}'). \tag{3}$$

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 9 May 2023 doi:10.20944/preprints202305.0670.v1

The mapping $\boldsymbol{z}$ in Eq. 3 is defined through Bochner's theorem [37]:

$$k(\boldsymbol{x}-\boldsymbol{x}') = \int_{\mathbb{R}^{\tilde{Q}}} p(\boldsymbol{\omega}) \exp(i\boldsymbol{\omega}^\top(\boldsymbol{x}-\boldsymbol{x}'))d\boldsymbol{\omega} = \mathbb{E}_{\boldsymbol{\omega}}\{\exp(i\boldsymbol{\omega}^\top(\boldsymbol{x}-\boldsymbol{x}'))\}, \tag{4}$$

where $\boldsymbol{x}, \boldsymbol{x}' \in \mathbb{R}^{\tilde{Q}}$, $p(\boldsymbol{w})$ is the probability density function of $\boldsymbol{w} \in \mathbb{R}^{\tilde{Q}}$ that defines the type of kernel. Specifically, the Gaussian kernel, favored for its universal approximating properties and mathematical tractability [38], is achieved from Eq. 4 by setting $p(\boldsymbol{w}) = \mathcal{N}(0, \sigma^2\hat{\boldsymbol{I}})$; $\sigma \in \mathbb{R}^+$ is a length-scale and $\hat{\boldsymbol{I}}$ is an identity matrix of proper size.

As both the kernel and the probability are real values, the imaginary component can be disregarded by employing the Euler equation. This leads to the use of a cosine function rather than an exponential, ensuring the following relationship:

$$z(\boldsymbol{x}) = \sqrt{\frac{2}{Q}}\left[\cos(\boldsymbol{\omega}_1^\top\boldsymbol{x}+b_1), \ldots, \cos(\boldsymbol{\omega}_Q^\top\boldsymbol{x}+b_Q)\right]^\top, \tag{5}$$

where $\boldsymbol{\omega}_q \in \mathbb{R}^{\tilde{Q}}$, $b_q \in \mathbb{R}$, and $q \in Q$.

We aim to extend the kernel-based mapping depicted in Eq. 5 for application to spatial data, such as images, by utilizing the power of convolutional operations. These operations have garnered significant attention for their efficacy in processing grid data [39]. Convolutional operations exhibit two crucial properties—translation equivariance and locality—that render them particularly suitable for handling spatial data [39]. In order to integrate these properties into the Random Fourier Features framework, we adapt the z mapping to operate within local regions of the grid input space. This results in the computation of the feature map $\mathbf{F}_l \in \mathbb{R}^{H_l \times \tilde{W}_l \times Q_l}$, where the mapping is defined as $z : \mathbb{R}^{H_{l-1} \times \tilde{W}_{l-1} \times D_{l-1}} \rightarrow \mathbb{R}^{H_l \times \tilde{W}_l \times Q_l}$, yielding:

$$\boldsymbol{F}_l = z(\boldsymbol{F}_{l-1}) = \cos\left(\frac{\boldsymbol{W}_l}{\Delta_l} \otimes \boldsymbol{F}_{l-1} + \boldsymbol{b}_l\right), \tag{6}$$

where $\Delta_l \in \mathbb{R}^+$ is a scale parameter. The parameters $\boldsymbol{W}_l \in \mathbb{R}^{\tilde{k}_l \times \tilde{k}_l \times D_{l-1} \times Q_l}$ and $\boldsymbol{b}_l \in \mathbb{R}^{Q_l}$ are initialized as in Eqs. 4 and 5, and updated through gradient descent under a back-propagation-based optimization of Eq. 2 [35]. Consequently, we refer to the layers in Eq. 6 as Convolutional Random Fourier Features Gradient (CRFFg).

2.3. Layer-Wise Weighted Class Activation Maps for Semantic Segmentation

Class Activation Maps (CAMs) are a powerful tool to enhance the interpretability of outcomes derived from deep learning models. They achieve this by emphasizing the critical image regions in determining the model's predicted output. To evaluate the contribution of these regions to a specific class $r \in \{0, 1\}$, a linear combination of feature maps from a designated convolutional neural network layer l can be employed [40]. Here, given an input image $\boldsymbol{I}$ and a target class r, the salient input spatial information coded by the l-th layer into a trained deep learning semantic segmentation model with parameter set Θ^*, as in Eq. 2, is gathered through the Layer-CAM algorithm, yielding [41]:

$$\boldsymbol{S}_l^r = (\Lambda \circ ReLU)\left(\sum_{d \in D_l} \boldsymbol{\alpha}_l^{rd} \odot \boldsymbol{F}_l^{rd}\right) \tag{7}$$

where $\boldsymbol{S}_l^r \in \mathbb{R}^{H \times \tilde{W}}$ holds the Layer-CAM for class r at layer l, $\Lambda : \mathbb{R}^{H_l \times \tilde{W}_l} \rightarrow \mathbb{R}^{H \times \tilde{W}}$ is the up-sampling operator, $ReLU(x) = \max(0, x)$ is the Rectified Linear activation function, and $\odot$ stands for Hadamard product. Besides, $\boldsymbol{F}_l^{rd} \in \mathbf{R}^{H_l \times \tilde{W}_l}$ collects the d-th feature map and $\boldsymbol{\alpha}_l^{rd} \in \mathbf{R}^{H_l \times \tilde{W}_l}$ is a weighting matrix holding elements:

$$\alpha_l^{rd}[i,j] = ReLU\left(\partial y^r / \partial F_l^{rd}[i,j]\right), \tag{8}$$

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 9 May 2023 doi:10.20944/preprints202305.0670.v1

with $\alpha_l^{rd}[i,j] \in \boldsymbol{\alpha}_l^{rd}$ and $F_l^{rd}[i,j] \in \boldsymbol{F}_l^{rd}$. y^r is the score for class r that is computed using the approach in [42] adopted for the semantic segmentation tasks, as follows:

$$y^r = \mathbb{E}\{\tilde{F}_L[i,j] : \forall i,j | M[i,j] = r\} \tag{9}$$

where $\tilde{F}_L[i,j] \in \tilde{\boldsymbol{F}}_L$ holds the feature map elements for layer L in Eq. 1 fixing a linear activation function.

As previously mentioned, the use of CAM-based representations enhances the explainability of deep learning models for segmentation tasks. To evaluate the interpretability of CAMs for a given model, we propose the following semantic segmentation measures, where higher scores indicate better interpretability:

– **CAM-Dice ($D^{'}$)**: a version of the Dice measure that quantifies mask thickness and how the extracted CAM is densely filled:

$$D_r^{'} = \mathbb{E}_l\left\{\mathbb{E}_n\left\{2\frac{\mathbf{1}^\top(\tilde{\boldsymbol{M}}_n^r \odot \boldsymbol{S}_{nl}^r)\mathbf{1}}{\mathbf{1}^\top\tilde{\boldsymbol{M}}_n^r\mathbf{1} + \mathbf{1}^\top\boldsymbol{S}_{nl}^r\mathbf{1}} : \forall n \in N\right\} : \forall l \in L\right\}, \quad D_r^{'} \in [0,1], \tag{10}$$

where $\boldsymbol{S}_{nl}^r$ holds the Layer-CAM for image n with respect to layer l (see Eq. 7). Additionally, $\tilde{\boldsymbol{M}}_n^r \in \{0,1\}^{H\times\tilde{W}}$ collects a binary mask that identifies the pixel locations associated with the class r.

– **CAM-based Cumulative Relevance (ρ_r)**: It involves computing the cumulative contribution from each CAM representation to detect class r within the segmented region of interest. This can be expressed as follows:

$$\rho_r = \mathbb{E}_l\left\{\mathbb{E}_n\left\{\frac{\mathbf{1}^\top(\tilde{\boldsymbol{M}}_n^r \odot \boldsymbol{S}_{nl}^r)\mathbf{1}}{\mathbf{1}^\top\boldsymbol{S}_{nl}^r\mathbf{1}} : \forall n \in N\right\} \forall l \in L\right\}, \quad \rho_r \in [0,1]. \tag{11}$$

– **Mask-based Cumulative Relevance (ϱ_r)**: It assesses the relevance averaged across the class pixel set related to the target mask of interest. Then, each class-based cumulative relevance is computed as follows:

$$\varrho_r = \mathbb{E}_l\left\{\mathbb{E}_n\left\{\frac{\mathbf{1}^\top(\tilde{\boldsymbol{M}}_n^r \odot \boldsymbol{S}_{nl}^r)\mathbf{1}}{\mathbf{1}^\top\tilde{\boldsymbol{M}}_n^r\mathbf{1}} : \forall n \in N\right\} \forall l \in L\right\}, \varrho_r \in \mathbb{R}^+. \tag{12}$$

The normalized Mask-based Cumulative Relevance can be computed as:

$$\varrho_r^{'} = \frac{\varrho^{'}{}_r}{\max\limits_{r\in\{0,1\}} \varrho_c}, \quad \varrho_r^{'} \in [0,1]. \tag{13}$$

As demonstrated, the proposed measures enable the weighting of each layer's contribution to a given class across the model by adjusting the normalization term related to the target mask, the estimated CAM, or both pixel-based salient activations.

3. Experimental Set-Up

The proposed deep learning model for semantic segmentation enhances foot thermal images' interpretability, achieving competitive segmentation performance. To this end, we evaluate the impact of incorporating a convolutional representation of CRFFg and layer-wise weighted CAM into three well-known deep-learning architectures. The proposed methodology is evaluated using the pipeline shown in Figure 1), including the following testing stages:

i) Foot Infrared Thermal Data Acquisition and Preprocessing.
ii) Architecture Set-Up of tested Deep models for foot segmentation. Three DL architectures are contrasted using our CRFFg: U-Net, Fully Convolutional Network (FCN), and ResUNet.

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 9 May 2023 doi:10.20944/preprints202305.0670.v1

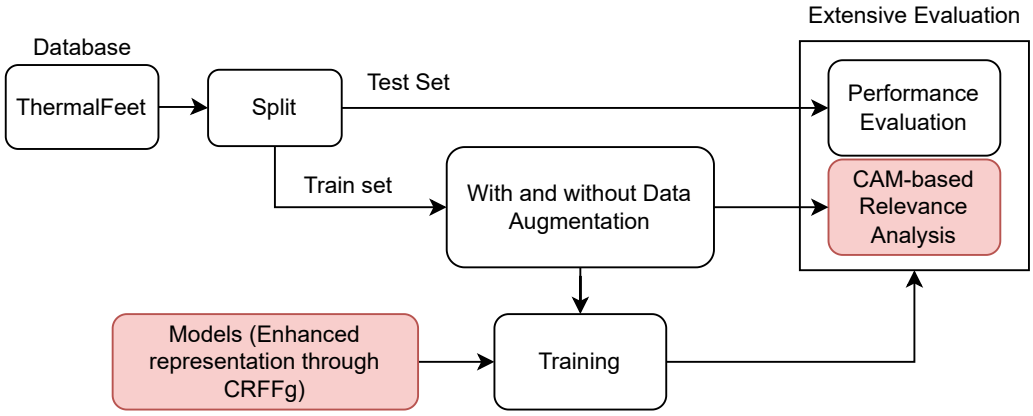


Figure 1. Foot segmentation from thermal images using our CRFFg-based deep learning enhancement holding layer-wise weighted CAM interpretability.

iii) Assessment of semantic segmentation accuracy. In this study, we examine how data augmentation affects the performance of tested deep learning algorithms.
iv) Relevance-maps extraction from our Layer-Wise weighted CAMs to provide interpretability.

3.1. Protocol for Infrared Thermal Data Acquisition: ThermalFeet Dataset

The protocol for data acquisition was designed by the physician staff at "SES Hospital Universitario de Caldas" to standardize the data collection of infrared thermal images acquired from pregnant women who underwent epidural anesthesia during labor. This protocol is in accordance with the occupational risks associated with assisting local anesthetics via epidural neuraxial as specified by the hospital's administration, following previously implemented protocols [9,43–46].

Patient monitoring includes the necessary equipment for taking vital signs and a metal stretcher with foam cushion and plastic exterior covered only with a white sheet. The continuous monitoring device is placed 1.5 meters from the stretcher in the same room, as shown in Fig. 2. Before the epidural procedure, anesthesiologists assess each patient clinically and provide written and verbal information about the trial before obtaining her written consent. The patient's body temperature, heart rate, oxygen saturation, and non-invasive blood pressure are monitored every five minutes. Skin temperature values are recorded during the procedure. Sensitivity responses are evaluated using superficial touch and cold tests with cotton wool soaked in water applied to the previously determined dermatomes. The temperature test records the verbal response as Yes or No for superficial touch and Cold or No Cold.

The protocol timeline for acquiring infrared thermal images is as follows: Initially, the woman is asked to be in a supine position before the first thermal image (T0) is captured once the first dose of the analgesic mixture is administered. A single thermal picture is taken at the placement of the operated catheter (0.45mm; Perifix, Braun®) positioned within the space selected for injecting epidural anesthesia in the cervical region (at L2 to L3 or L3 to L4), measuring a few millimeters.

Within the next 25 minutes, one thermographic recording of the lower extremity is taken every five minutes (T1-T5). The catheter remains in the epidural space taped to the skin so that one image is captured every five minutes until six pictures have been collected. Though the clinical protocol demands images of both feet taken in a fixed corporal position, this condition is barely achievable due to the difficulty of labor procedures and contractions.

The data was collected under two different hardware specifications: i) A set of 196 images captured from 22 pregnant women during labour using a FLIR A320 infrared camera with a resolution of 640×480 and a spectral range within 7.5 to $13\mu m$. ii) A set of 128 images with improved sensitivity and flexibility taken using a FLIR E95 thermal

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 9 May 2023 doi:10.20944/preprints202305.0670.v1

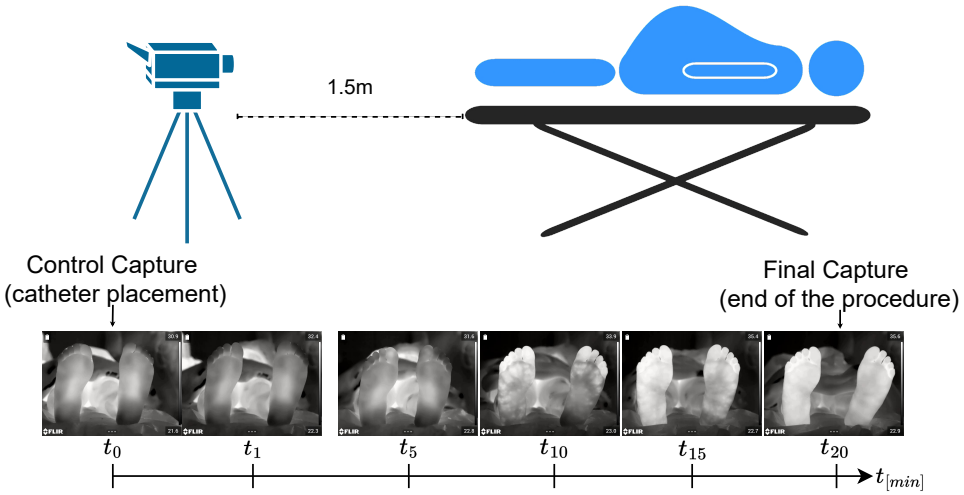


Figure 2. Regional analgesia monitoring protocol using local anesthetics via epidural neuraxial and thermal images.

camera, having a resolution of 640 × 480 and spectral range within 7.5 to 14μm. In this study, 166 thermal images are selected from both sets as fulfilling the quality criteria of validation, as detailed in [25]. The dataset is publicly available at [1].

3.2. Set-Up of compared Deep Learning Architectures

The following deep learning architectures are contrasted and enhanced using our CRRg approach:

- **Fully Convolutional Network (FCN)** [12]: This architecture is based on the VGG (Very Deep Convolutional Network) [47] model to recognize large-scale images. By using only convolutional layers, FCN models can deliver a segmentation map with pixel-level accuracy while reducing the computational burden.
- **U-Net** [14]: This architecture unfolds into two parts: The encoder consists of convolutional layers to reduce the spatial image dimensions. The decoder holds layers to upsample the encoded features back to the original image size.
- **ResUNet** [33]: This model extends the U-Net architecture by incorporating residual connections to improve performance. Deep learning training is improved by residual connections, which allow gradients to flow directly through the network.

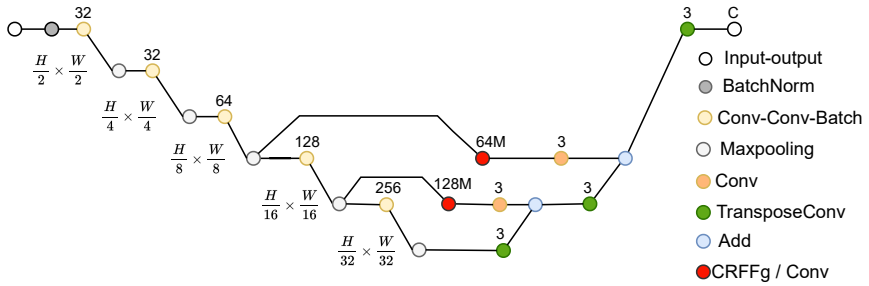
We estimate the effectiveness of incorporating the CRFFg layer for comparison purposes in FCN, U-Net, and ResUNet architectures. However, each evaluated CRFFg layer arrangement differs from another in the semantic segmentation features that feed the decoder, as detailed in [48–50]. Then, the CRFFg layer is placed at skip connections to overcome this issue to enhance the feature fusion between encoders and decoders (see Fig. 6).

To evaluate the performance difference with the proposed CFFg-layer strategy, we utilize a standard convolutional layer featuring an identical number of filters and a ReLU activation function at the same position within the architecture. In particular, we analyze the influence of the CFFg layer dimension on segmentation performance, testing two multiplication values (one and three). Besides, to study the impact of CRFFg, we set the hyperparameters of all models the same. The selected optimizer is Adam, with Keras default parameter values, and dice-based loss is employed in Eq. 2, as follows:

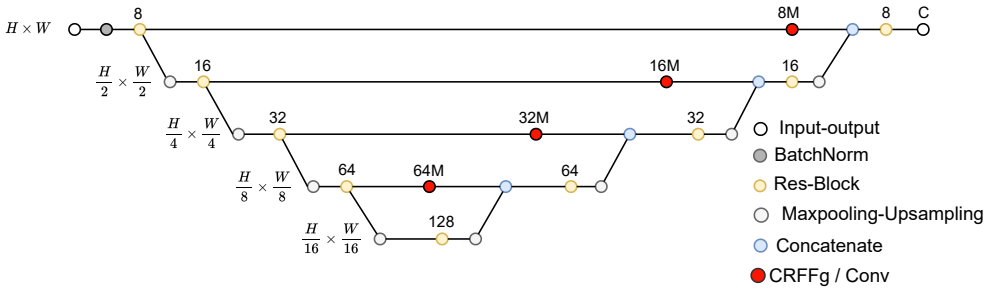
$$\mathcal{L}_{Dice}(\boldsymbol{M}_n, \hat{\boldsymbol{M}}_n) = 2\frac{\mathbf{1}^\top(\boldsymbol{M}_n \odot \hat{\boldsymbol{M}}_n)\mathbf{1} + \epsilon}{\mathbf{1}^\top \boldsymbol{M}_n \mathbf{1} + \mathbf{1}^\top \hat{\boldsymbol{M}}_n \mathbf{1} + \epsilon} \tag{14}$$

[1] https://gcpds-image-segmentation.readthedocs.io/en/latest/notebooks/02-datasets.html

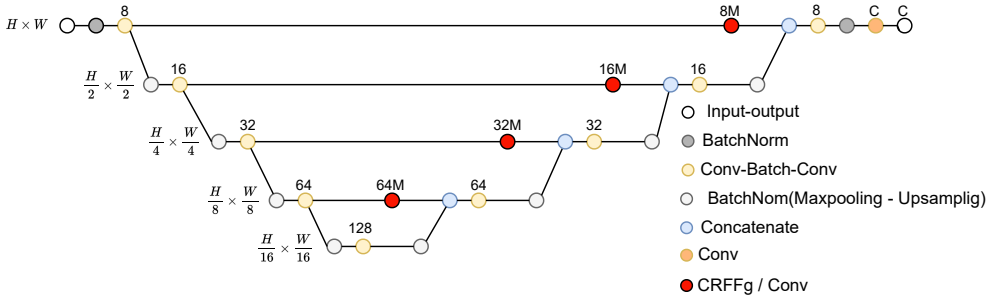
Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 9 May 2023 doi:10.20944/preprints202305.0670.v1



(a) FCN.



(b) ResUNet.



(c) U-Net.

Figure 3. Tested segmantic segmentation architectuhres. Our CRFFg approach enhances the data representation (see red dots).

where ϵ=1 avoids numerical instability. All experiments are carried out in Python 3.8, with the Tensorflow 2.4.1 API, on a Google Colaboratory environment (code repository: https://github.com/aguirrejuan/Foot-segmentation-CRFFg, accessed on 25 April 2023).

3.3. Training Details and Quantitative Assessment

With the aim to prevent overfitting and improve the generalization of trained models, the data augmentation procedure is performed on each image with horizontal flip enabled since feet are mostly symmetrical on the horizontal axis, specifically left-right and right-left on each foot. Hence, vertical overturn is disabled to prevent unrealistic upside-down foot representations. In the augmentation procedure, the images are rotated seven times within a range of -15 to 15 degrees, translated by 10% right to left, and zoomed in and out by 15%, as described in [51].

Morevoer, the following metrics are used to measure segmentation performance [**?**]:

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 9 May 2023 doi:10.20944/preprints202305.0670.v1

$$D = \frac{2|\boldsymbol{M} \cap \hat{\boldsymbol{M}}|}{|\boldsymbol{M}| + |\hat{\boldsymbol{M}}|} = \frac{2T_P}{2T_P + F_P + F_N} \tag{15a}$$

$$J = \frac{|\boldsymbol{M} \cap \hat{\boldsymbol{M}}|}{|\boldsymbol{M} \cup \hat{\boldsymbol{M}}|} = \frac{T_P}{F_N + F_P + T_P} \tag{15b}$$

$$S_e = \frac{T_P}{T_P + F_N} \tag{15c}$$

$$S_p = \frac{T_N}{T_N + F_P} \tag{15d}$$

where T_P, F_N, and F_P represent the true positive, false negative, and false positive predictions, respectively, for comparing the actual and estimated label masks $\boldsymbol{M}_n$ and $\hat{\boldsymbol{M}}_n$ for a given input image $\boldsymbol{I}_n$. In addition, the introduced layer-wise, weighted CAM-based interpretability measures are computed for CAM-Dice, CAM-based Cumulative Relevance, and Mask-based Cumulative Relevance (see Eqs. 10-13).

As for the validation strategy, we selected the hold-out cross-validation strategy with the following partitions: 80% of the samples for training, 10% for validation, and 10% for testing.

4. Results and Discussion

4.1. Method Comparison Results of Semantic Segmentation Performance

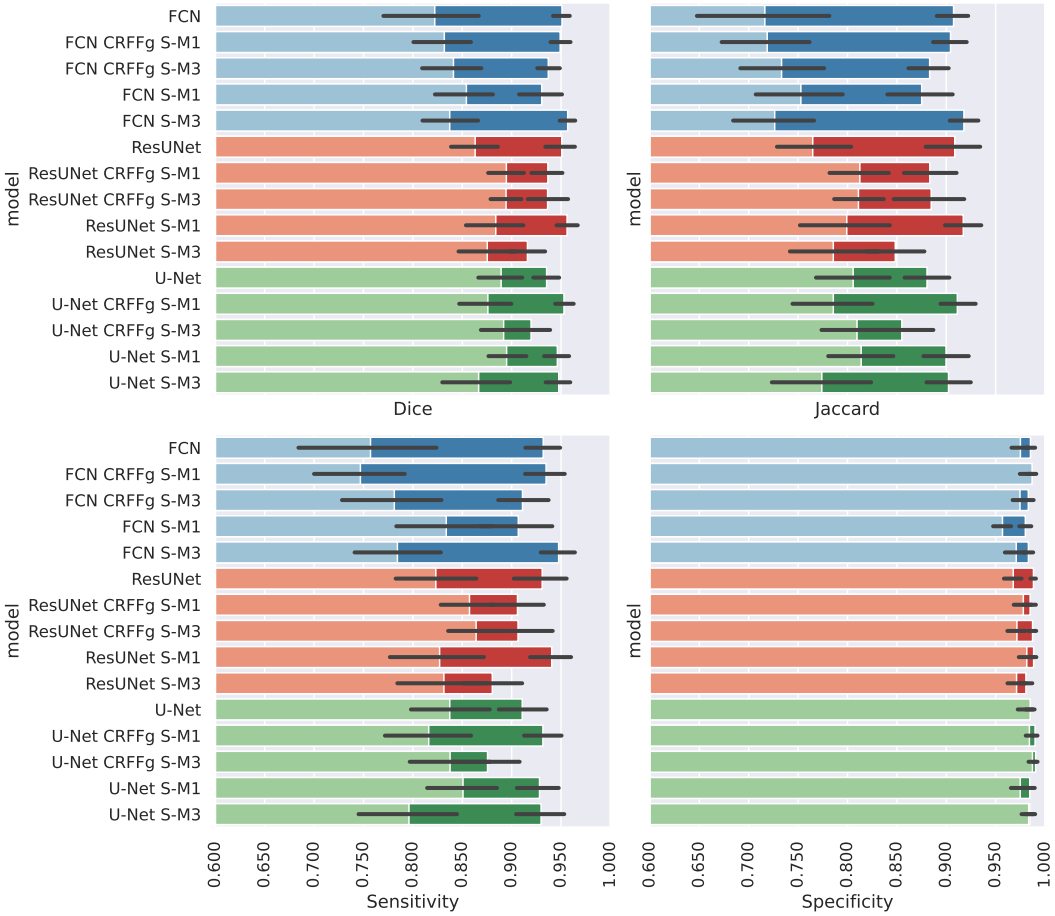
Figure 4a displays the values of semantic segmentation performance for the ThermalFeet dataset achieved by each compared deep learning architecture: FCN (colored in blue), ResNet (red), U-Net (green). For interpretation purposes, the results are presented for the evaluation measures separately. As seen, the specificity estimates are very close to the maximal value and show the lowest variability. This result can be explained by the relatively small feet sizes compared with the background, making their correct detection and segmentation more difficult. On the contrary, sensitivity assessments are of less value and have much more variability, accounting for the diversity in the regions of interest (i.e., size, shape, and location). Due to the changing behavior of thermal patterns and the limited datasets available, learners have difficulty obtaining an accurate model.

Regarding overlapping between estimated thermal masks, the Dice value is acceptable but with higher variance values for FCN, implying that other tested models segment complex shapes more accurately. As expected, the Jaccard index mean values resemble the Dice assessments, although with increased variance, which highlights the mismatch between the ground truth and the predicted mask even more.

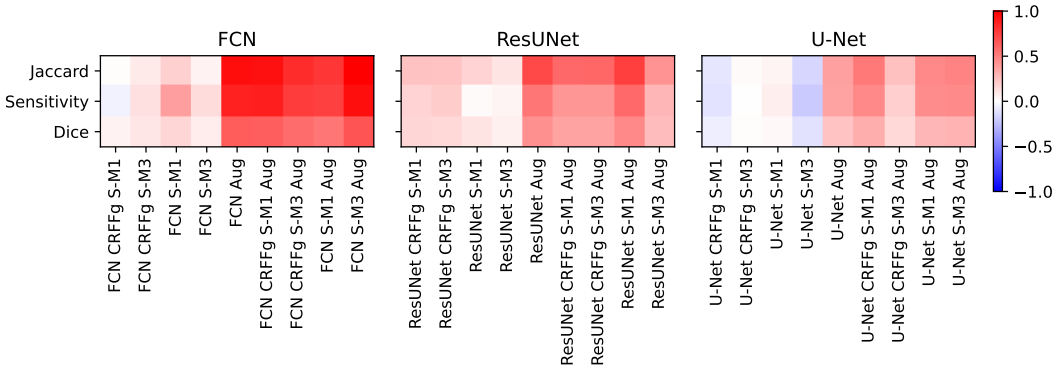
A comparison between the segmentation metric value achieved by the baseline architecture (without any modifications) and the value estimated for every evaluated semantic segmentation strategy is presented in Figure 4b. Note that specificity is removed because its estimates are obtained with minimal variations.

As seen, the performance improvement depends on the learner model size (also called algorithm complexity). Namely, the baseline architecture of FCN holds 1,197,375 parameters, baseline ResUnet – 643,549, and baseline Unet – 494,093. Thus, the FCN model contains the largest tuning parameter set and achieves the poorest performance, but it benefits the most from the evaluated architectures. As data augmentation is also applied, this finding becomes more evident. It may be pointed out that adding new data decreases model overfitting inherent to massive model sizes. Likewise, the following ResUnet model takes advantage of the enhanced architecture strategy using our CRFFg and improves performance. It increases more by generating new data points, however, to a lesser extent. Lastly, the learner with the lowest parameter set gets almost no benefits or is negatively affected by the strategies considered for architecture enhancement. Still, the strategies taken into account combined with expanded training data sizes can be improved, though very modestly.

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 9 May 2023 doi:10.20944/preprints202305.0670.v1



(a) Segmentation performance results on ThermalFeet database. The three types of architecture used in this study (FCN, U-Net, ResUNet) are differentiated by color. The type of variation in the architecture is indicated by the marker used



(b) The improvement of each strategy, normalized with respect to the baseline performance of each architecture

Figure 4. Results of the comparison between methods. The segmentation performance of ThermalFeet is evaluated using baseline models FCN, UNet, and ResUNet, and compared to our proposal that incorporates CRFFg-based enhancements.

4.2. Results of Assessing the Proposed CAM-based Relevance Analysis Measures

We aim to evaluate the tested deep learning models for assessing the contribution of CAM-based representations to interpretability. To this end, we plot the pairwise relationship between the essential explanation elements (background and foreground) and the above-proposed meeasure for assessing the CAM-based relevance of performed image

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 9 May 2023 doi:10.20944/preprints202305.0670.v1

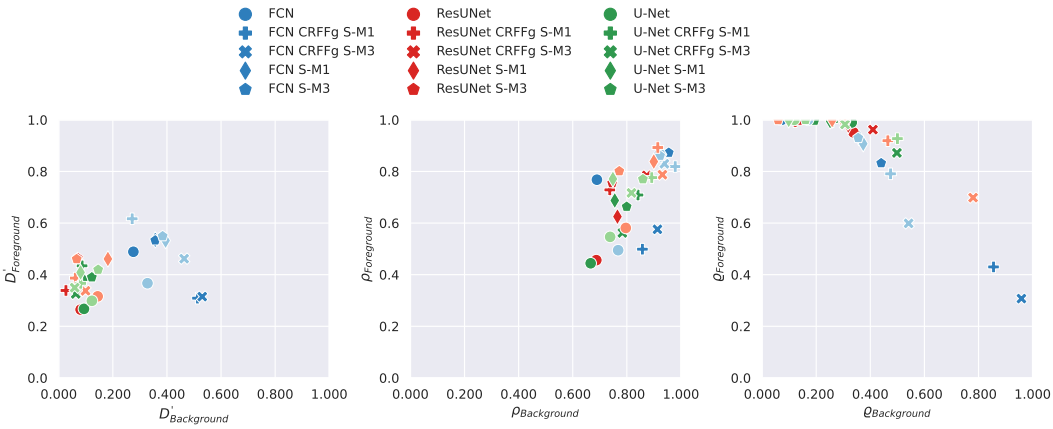


Figure 5. Results of Interpretability Measures on ThermalFeet. The three types of architecture used in this study (FCN, U-Net, ResUNet) are differentiated by color. The type of variation in the architecture is indicated by the marker used.

segmentation masks. Figure 5 displays the scatter plots obtained by each segmentation learner. CAMs extracted by the learner contribute more to the interpretability of regions of interest if the measure value tends toward the top-right corner. Moreover, we focus on the contribution of CAM representations to segmenting between background and foreground, utilizing the patient's feet as critical identification features.

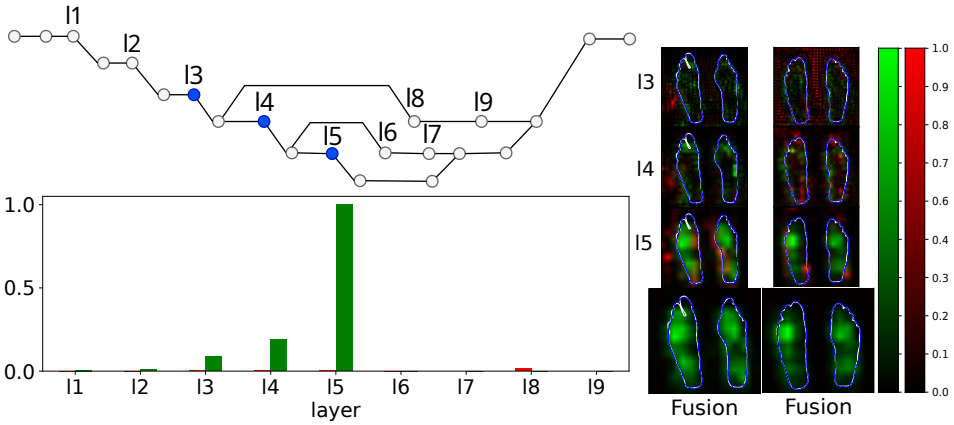
The findings from the modified CAM-Dice results can be split into two groups (refer to the left plot in Figure 5). One group involves ResUnet and UNet architectures, and the other showcases the better performance, featuring FCN architectures. It is also important to mention that the data augmentation strategy does not significantly boost interpretability as much as it enhances segmentation performance measures. Looking at the CAM-based Cumulative Relevance (refer to the middle plot in Figure 5), it is apparent that models with refined representations at skip connections surpass the baseline models. Even though there is no substantial difference between models with these enhancements, most models are situated in the top-right corner. This position suggests that the primary relevance is focused on the area of interest. Significantly, relevance seems to accumulate more in the background than in the foreground, which is logical, considering the relative sizes of both areas. In Figure 5, the Mask-based Cumulative Relevance plot on the right side demonstrates that most models tend to exhibit high-foreground-low-background relevance. This pattern leads to a bias favoring the foreground class, as reflected in the more robust activation of CAMs for the foreground class. However, it is interesting that models employing CRFFg perform better in separating classes situated towards the top-right corner, suggesting superior capabilities in differentiating foreground and background classes.

Figure 6 displays examples of CAMs extracted by the best models per architecture under the Mask-based Cumulative Relevance for feet (colored in green) and background (red color), respectively. As seen, the higher weight is located at the last part of the decoder, where the higher values of semantic information are found. Besides, the weights for the background class are also less than for the foreground class, showing that the models emphasize the latter while preserving the relevance weights for the former.

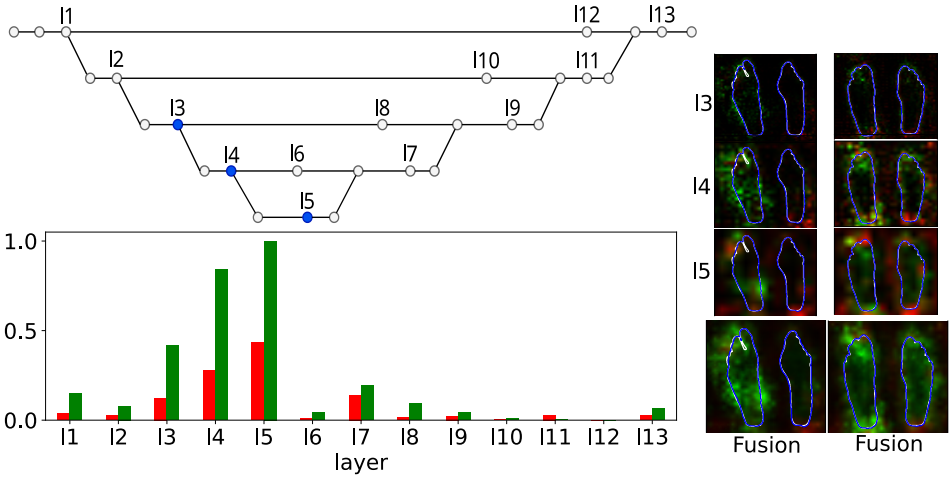
In particular, FCN CRFFg S-M3 is the best FCN model, as shown in Figure 6a, and extracts most of the weights in three layers (i.e., l3, l4, and l5), meaning that other layers do not contribute to the class foreground. On the other hand, this architecture leads to CAMs with lower values for background class (see examples on the right). This behavior can be explained because the FCN architecture holds an extensive receptive field. Hence, the FCN CRFFg S-M3 model enables capturing more global information crucial for segmentation and concentrating weights in a few layers.

In the case of ResUNet, ResUNet CRFFg S-M3 performs the most efficiently, as shown in Figure 6b. Since the receptive field decreases, the ResUNet architecture distributes the

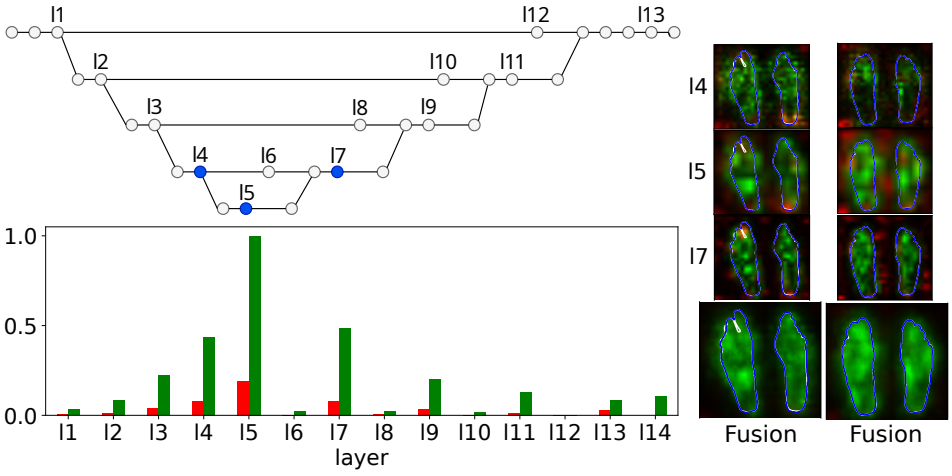
Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 9 May 2023 doi:10.20944/preprints202305.0670.v1



(a) FCN CRFFg S-M3 without data Augmentation



(b) ResUNet CRFFg S-M3 without data Augmentation



(c) U-Net CRFFg S-M3 with data Augmentation

Figure 6. Salient relevance analysis results. Best models concerning the Mask-based Cumulative Relevance ϱ_r measure are presented for FCN, UNet, and ResUNet with our CRFFg-based enhancement.

contribution more evenly among the extracted CAM representations. However, the more significant values remain in the l3, l4, and l5 layers. There is also activation of weights for the background class that can be explained, firstly, since the CRFFg configuration helps capture complex non-linear dependencies. Secondly, the local receptive field allows class separation.

 doi:10.20944/preprints202305.0670.v1

Lastly, the CRFFg S-M3 model is the most effective for the U-Net architecture, with a performance similar to the outperforming ResUNet architecture, as shown in Figure 6c. However, several differences in the Fusion CAMs extracted by U-Net CRFFg S-M3 show high activation within the feet, suggesting that this model is not only sensitive to the foreground class. Also, it captures more global features from feet.

5. Concluding Remarks

We introduce an innovative semantic segmentation approach that enhances interpretability by incorporating convolutional random Fourier features and layer-wise weighted class activation. This methodology has been tested on a unique dataset of thermal foot images from pregnant women who have received epidural anesthesia, which is small but exhibits considerable variability. Our strategy is two-pronged. Firstly, we introduce a novel Random Fourier Features layer known as CRFFg for handling image data, aiming to enhance three renowned architectures - FCN, UNet, and ResUNet. Secondly, we introduce three new quantitative measures to assess the interpretability of any deep learning model used for segmentation tasks. Our validation results indicate that the proposed approach boosts explainability and maintains competitive foot segmentation performance. In addition, the dataset used is tailored explicitly for epidural insertion during childbirth, reinforcing the practical relevance of our methodology.

There are, however, several observations worth mentioning:

Data acquisition tailored for Epidural. Epidural anesthesia involves the delivery of medicines that numb body parts to relieve pain, and the acquisition of data is usually performed under uncontrolled conditions with strong maternal artifacts. Moreover, it is impossible to fix a timeline for data collection. In addition, a timeline for gathering data cannot be set correctly. To the extent of our knowledge, this is the first time a protocol has been presented to regulate the data collection of infrared thermal images acquired from pregnant women who underwent epidural anesthesia during labor. As a result, data were assembled under real-world conditions that contained 196 thermal images fulfilling validation quality criteria.

Deep learning models for image semantic segmentation. Combined with machine learning, thermal imaging has proven helpful for performing semantic segmentation as a powerful method of dense prediction to adverse lighting conditions, providing better performance compared to their traditional counterparts. State-of-the-art medical image segmentation models include variants of U-Net models. A major reason for their success is that they employ skip connections, combining deep, semantic, and coarse-grained feature maps from the decoder subnetwork with shallow, low-level, fine-grained feature maps from the encoder subnetwork. They recover fine-grained details of target objects despite complex backgrounds [52]. Nevertheless, the collected image data from epidural anesthesia is insufficient for training the most commonly-known deep learners, which may result in overfitness to the training set. We address this issue by employing data augmentation addresses that artificially increase training data inputs to feed three tested architectures of deep learning models (FCN, U-Net, ResUNet), thus improving segmentation accuracy results. As seen in Figure 4b, the segmentation accuracy gain depends on the learner model complexity used: The fewer parameters the learner holds, the more the effectivity of data augmentation. Thus, the UNet learner with the lowest parameter set gets almost no benefit.

Strategies for enhancing the performance of deep learning-based segmentation. Three deep-learning architectures are explored to increase the interpretability of semantic segmentation results at competitive accuracy, ranked in decreased order of computational complexity as follows: FCN, ResUNet, and U-Net. Regarding the accuracy of semantic models, the data augmentation yields a sensibility metric value dependent on the model complexity: the more parameters the architecture holds, the higher the segmentation

accuracy improvement. Thus, FCN benefits more from artificial data than ResUNet and U-Net. In the same way, both overlapping metrics (Jaccard and Dice) depend on the complexity of models. By contrast, the specificity reaches very high values regardless of trained deep learning because the background texture's homogeneity saturates most captured thermal images. Nonetheless, the proposed modifications to architectures are not a solid argument for influencing their performed accuracy of semantic segmentation. In terms of enhancing explainability, the weak influence of data augmentation is the first finding to be drawn, as seen in the scatterplots of Fig.5. All tested models produce more significant CAM activations from layers with a wider receptive field. Moreover, the CRFFg layer also improves the representation of the foreground and background. It is also important to note the metrics developed for assessing the explainability of CAM representations, allowing scalability to larger image sets without visual inspection.

In terms of future research, the authors intend to integrate attention mechanisms for semantic segmentation into the CRFFg-based representation that has been introduced in [53]. Besides, we propose to include variational autoencoders within our framework to prevent overfitting and enhance data interpretability [54].

Author Contributions: Conceptualization, J.C.A.-A., A.A.-M. and G.C.-D.; methodology, J.C.A.-A., A.A.-M. and G.C.-D.; software, J.C.A.-A.; validation, J.C.A.-A., A.A.-M. and G.C.-D.; formal analysis, J.C.A.-A. and G.C.-D.; investigation, J.C.A.-A., A.A.-M. and G.C.-D.; resources, A.A.-M. and G.C.-D.; data curation, J.C.A.-A.; writing–original draft preparation, J.C.A.-A., A.A.-M. and G.C.-D.; writing–review and editing, A.A.-M. and G.C.-D.; visualization, J.C.A.-A.; supervision, A.A.-M. and G.C.-D.; project administration, A.A.-M.; funding acquisition, A.A.-M. and G.C.-D. All authors have read and agreed to the published version of the manuscript.

Funding: Under grants provided by the projects: "Prototipo de visión por computador para la identificación de problemas fitosanitarios en cultivos de plátano en el departamento de Caldas" (Hermes 51175) funded by Universidad Nacional de Colombia, and "Desarrollo de una herramienta de visión por computador para el análisis de plantas orientado al fortalecimiento de la seguridad alimentaria" (Hermes 54339) funded by Universidad Nacional de Colombia and Universidad de Caldas.

Institutional Review Board Statement: This study uses anonymized public datasets with institutional review board statement as presented in https://gcpds-image-segmentation.readthedocs.io/en/latest/notebooks/02-datasets.html

Informed Consent Statement: This study uses anonymized public datasets as presented in https://gcpds-image-segmentation.readthedocs.io/en/latest/notebooks/02-datasets.html

Data Availability Statement: Dataset is publicly available at: https://gcpds-image-segmentation.readthedocs.io/en/latest/notebooks/02-datasets.html (accessed on April 2023).

Conflicts of Interest: The authors declare that this research was conducted without any commercial or financial relationships that could be construed as a potential conflict of interest.

References

1. Brown, D.T.; Wildsmith, J.A.W.; Covino, B.G.; Scott, D.B. Effect of Baricity on Spinal Anaesthesia with Amethocaine. *BJA: British Journal of Anaesthesia* **1980**, *52*, 589–596, [https://academic.oup.com/bja/article-pdf/52/6/589/822181/52-6-589.pdf]. https://doi.org/10.1093/bja/52.6.589.
2. McCombe, K.; Bogod, D. Regional anaesthesia: risk, consent and complications. *Anaesthesia* **2021**, *76*, 18–26, [https://associationofanaestheti publications.onlinelibrary.wiley.com/doi/pdf/10.1111/anae.15246]. https://doi.org/https://doi.org/10.1111/anae.15246.
3. Chae, Y.; Park, H.J.; Lee, I.S. Pain modalities in the body and brain: Current knowledge and future perspectives. *Neuroscience & Biobehavioral Reviews* **2022**, *139*, 104744. https://doi.org/https://doi.org/10.1016/j.neubiorev.2022.104744.
4. Curatolo, M.; Petersen-Felix, S.; Arendt-Nielsen, L. Assessment of regional analgesia in clinical practice and research. *British Medical Bulletin* **2005**, *71*, 61–76, [https://academic.oup.com/bmb/article-pdf/71/1/61/25152159/ldh035.pdf]. https://doi.org/10.1093/bmb/ldh035.
5. Bruins, A.; Kistemaker, K.; Boom, A.; Klaessens, J.; Verdaasdonk, R.; Boer, C. Thermographic skin temperature measurement compared with cold sensation in predicting the efficacy and distribution of epidural anesthesia. *Journal of clinical monitoring and computing* **2018**, *32*, 335–341. https://doi.org/10.1007/s10877-017-0026-y.

6. Bruins, A.A.; Kistemaker, K.R.J.; Boom, A.; Klaessens, J.; Verdaasdonk, R.; Boer, C. Thermographic skin temperature measurement compared with cold sensation in predicting the efficacy and distribution of epidural anesthesia. *Journal of Clinical Monitoring and Computing* **2018**, *32*, 335 – 341.
7. Haren, F.; Kadic, L.; Driessen, J. Skin temperature measured by infrared thermography after ultrasound-guided blockade of the sciatic nerve. *Acta anaesthesiologica Scandinavica* **2013**, *57*. https://doi.org/10.1111/aas.12170.
8. Stevens, M.F.; Werdehausen, R.; Hermanns, H.; Lipfert, P. Skin temperature during regional anesthesia of the lower extremity. *Anesthesia and analgesia* **2006**, *102*, 1247—1251. https://doi.org/10.1213/01.ane.0000198627.16144.77.
9. Werdehausen, R.; Braun, S.; Hermanns, H.; Freynhagen, R.; Lipfert, P.; Stevens, M.F. Uniform Distribution of Skin-Temperature Increase After Different Regional-Anesthesia Techniques of the Lower Extremity. *Regional Anesthesia and Pain Medicine* **2007**, *32*, 73–78. https://doi.org/https://doi.org/10.1016/j.rapm.2006.07.009.
10. Zhang, L.; Nan, Q.; Bian, S.; Liu, T.; Xu, Z. Real-time segmentation method of billet infrared image based on multi-scale feature fusion. *Scientific Reports* **2022**, *12*, 6879.
11. Kütük, Z.; Algan, G. Semantic Segmentation for Thermal Images: A Comparative Survey, 2022. https://doi.org/10.48550/ARXIV.2205.13278.
12. Long, J.; Shelhamer, E.; Darrell, T. Fully Convolutional Networks for Semantic Segmentation. *CoRR* **2014**, *abs/1411.4038*, [1411.4038].
13. Bi, L.; Kim, J.; Kumar, A.; Fulham, M.J.; Feng, D. Stacked fully convolutional networks with multi-channel learning: application to medical image segmentation. *The Visual Computer* **2017**, *33*, 1061 – 1071.
14. Ronneberger, O.; Fischer, P.; Brox, T. U-Net: Convolutional Networks for Biomedical Image Segmentation. *CoRR* **2015**, *abs/1505.04597*, [1505.04597].
15. Kumar, V.; Webb, J.M.; Gregory, A.; Denis, M.; Meixner, D.D.; Bayat, M.; Whaley, D.H.; Fatemi, M.; Alizad, A. Automated and real-time segmentation of suspicious breast masses using convolutional neural network. *PloS one* **2018**, *13*, e0195816.
16. Zhou, Z.; Siddiquee, M.M.R.; Tajbakhsh, N.; Liang, J. UNet++: A Nested U-Net Architecture for Medical Image Segmentation. *CoRR* **2018**, *abs/1807.10165*, [1807.10165].
17. Badrinarayanan, V.; Kendall, A.; Cipolla, R. SegNet: A Deep Convolutional Encoder-Decoder Architecture for Image Segmentation, 2016, [arXiv:cs.CV/1511.00561].
18. He, K.; Gkioxari, G.; Dollár, P.; Girshick, R. Mask R-CNN, 2018, [arXiv:cs.CV/1703.06870].
19. Ren, S.; He, K.; Girshick, R.; Sun, J. Faster R-CNN: Towards Real-Time Object Detection with Region Proposal Networks, 2016, [arXiv:cs.CV/1506.01497].
20. Zhao, H.; Shi, J.; Qi, X.; Wang, X.; Jia, J. Pyramid Scene Parsing Network, 2017, [arXiv:cs.CV/1612.01105].
21. Arteaga-Marrero, N.; Hernández, A.; Villa, E.; González-Pérez, S.; Luque, C.; Ruiz-Alzola, J. Segmentation Approaches for Diabetic Foot Disorders. *Sensors 2021, Vol. 21, Page 934* **2021**, *21*, 934. https://doi.org/10.3390/S21030934.
22. Fischler, M.A.; Bolles, R.C. Random Sample Consensus: A Paradigm for Model Fitting with Applications to Image Analysis and Automated Cartography. *Commun. ACM* **1981**, *24*, 381–395. https://doi.org/10.1145/358669.358692.
23. Bouallal, D.; Bougrine, A.; Douzi, H.; Harba, R.; Canals, R.; Vilcahuaman, L.; Arbanil, H. Segmentation of plantar foot thermal images: Application to diabetic foot diagnosis. *International Conference on Systems, Signals, and Image Processing* **2020**, *2020-July*, 116–121. https://doi.org/10.1109/IWSSIP48289.2020.9145167.
24. Bougrine, A.; Harba, R.; Canals, R.; Ledee, R.; Jabloun, M. On the segmentation of plantar foot thermal images with deep learning. *European Signal Processing Conference* **2019**, *2019-September*. https://doi.org/10.23919/EUSIPCO.2019.8902691.
25. Mejia-Zuluaga, R.; Aguirre-Arango, J.C.; Collazos-Huertas, D.; Daza-Castillo, J.; Valencia-Marulanda, N.; Calderón-Marulanda, M.; Aguirre-Ospina, Ó.; Alvarez-Meza, A.; Castellanos-Dominguez, G. Deep Learning Semantic Segmentation of Feet Using Infrared Thermal Images. In Proceedings of the Advances in Artificial Intelligence – IBERAMIA 2022; Bicharra Garcia, A.C.; Ferro, M.; Rodríguez Ribón, J.C., Eds.; Springer International Publishing: Cham, 2022; pp. 342–352.
26. Dosovitskiy, A.; Beyer, L.; Kolesnikov, A.; Weissenborn, D.; Zhai, X.; Unterthiner, T.; Dehghani, M.; Minderer, M.; Heigold, G.; Gelly, S.; et al. An Image is Worth 16x16 Words: Transformers for Image Recognition at Scale. *CoRR* **2020**, *abs/2010.11929*, [2010.11929].
27. Chen, J.; Lu, Y.; Yu, Q.; Luo, X.; Adeli, E.; Wang, Y.; Lu, L.; Yuille, A.L.; Zhou, Y. TransUNet: Transformers Make Strong Encoders for Medical Image Segmentation **2021**.
28. Cao, H.; Wang, Y.; Chen, J.; Jiang, D.; Zhang, X.; Tian, Q.; Wang, M. Swin-Unet: Unet-like Pure Transformer for Medical Image Segmentation **2021**. pp. 205–218. https://doi.org/10.1007/978-3-031-25066-8_9.
29. Zhang, Y.; Liu, H.; Hu, Q. TransFuse: Fusing Transformers and CNNs for Medical Image Segmentation. *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)* **2021**, *12901 LNCS*, 14–24. https://doi.org/10.1007/978-3-030-87193-2_2.
30. Li, S.; Sui, X.; Luo, X.; Xu, X.; Liu, Y.; Goh, R. Medical Image Segmentation Using Squeeze-and-Expansion Transformers. *IJCAI International Joint Conference on Artificial Intelligence* **2021**, pp. 807–815. https://doi.org/10.24963/ijcai.2021/112.
31. Luo, X.; Hu, M.; Song, T.; Wang, G.; Zhang, S. Semi-Supervised Medical Image Segmentation via Cross Teaching between CNN and Transformer **2022**. pp. 820–833.
32. Kirillov, A.; Mintun, E.; Ravi, N.; Mao, H.; Rolland, C.; Gustafson, L.; Xiao, T.; Whitehead, S.; Berg, A.C.; Lo, W.Y.; et al. Segment Anything, 2023, [arXiv:cs.CV/2304.02643].

33. Anas, E.M.A.; Nouranian, S.; Mahdavi, S.S.; Spadinger, I.; Morris, W.J.; Salcudean, S.E.; Mousavi, P.; Abolmaesumi, P. Clinical Target-Volume Delineation in Prostate Brachytherapy Using Residual Neural Networks. In Proceedings of the Medical Image Computing and Computer Assisted Intervention - MICCAI 2017; Descoteaux, M.; Maier-Hein, L.; Franz, A.; Jannin, P.; Collins, D.L.; Duchesne, S., Eds.; Springer International Publishing: Cham, 2017; pp. 365–373.
34. Zhou, B.; Khosla, A.; Lapedriza, A.; Oliva, A.; Torralba, A. Learning Deep Features for Discriminative Localization, 2015, [arXiv:cs.CV/1512.04150].
35. Zhang, A.; Lipton, Z.C.; Li, M.; Smola, A.J. Dive into deep learning. *arXiv preprint arXiv:2106.11342* **2021**.
36. Rahimi, A.; Recht, B. Random features for large-scale kernel machines. 2009.
37. Rudin, W... *Fourier analysis on groups*; Interscience tracts in pure and applied mathematics ;, Interscience: New York, New York, 1976; p. 19.
38. Álvarez-Meza, A.M.; Cárdenas-Peña, D.; Castellanos-Dominguez, G. Unsupervised Kernel Function Building Using Maximization of Information Potential Variability. In Proceedings of the Progress in Pattern Recognition, Image Analysis, Computer Vision, and Applications; Bayro-Corrochano, E.; Hancock, E., Eds.; Springer International Publishing: Cham, 2014; pp. 335–342.
39. Bronstein, M.M.; Bruna, J.; Cohen, T.; Velickovic, P. Geometric Deep Learning: Grids, Groups, Graphs, Geodesics, and Gauges. *CoRR* **2021**, *abs/2104.13478*, [2104.13478].
40. Zhou, B.; Khosla, A.; Lapedriza, À.; Oliva, A.; Torralba, A. Learning Deep Features for Discriminative Localization. *CoRR* **2015**, *abs/1512.04150*, [1512.04150].
41. Jiang, P.T.; Zhang, C.B.; Hou, Q.; Cheng, M.M.; Wei, Y. LayerCAM: Exploring hierarchical class activation maps for localization. *IEEE Transactions on Image Processing* **2021**, *30*, 5875–5888. https://doi.org/10.1109/TIP.2021.3089943.
42. Jimenez-Castaño, C.A.; Álvarez-Meza, A.M.; Aguirre-Ospina, O.D.; Cárdenas-Peña, D.A.; Orozco-Gutiérrez, Á.A. Random fourier features-based deep learning improvement with class activation interpretability for nerve structure segmentation. *Sensors* **2021**, *21*, 7741.
43. Galvin, E.M.; Niehof, S.; Medina, H.J.; Zijlstra, F.J.; van Bommel, J.; Klein, J.; Verbrugge, S.J.C. Thermographic temperature measurement compared with pinprick and cold sensation in predicting the effectiveness of regional blocks. *Anesthesia and analgesia* **2006**, *102*, 598—604. https://doi.org/10.1213/01.ane.0000189556.49429.16.
44. Chestnut, D.H.; Wong, C.A.; Tsen, L.C.; Kee, W.M.D.N.; Beilin, Y.; Mhyre, J. Chestnut's Obstetric Anesthesia: Principles and Practice E-Book: Expert Consult-Online and Print **2014**.
45. ASGHAR, S.; LUNDSTRØM, L.H.; BJERREGAARD, L.S.; LANGE, K.H.W. Ultrasound-guided lateral infraclavicular block evaluated by infrared thermography and distal skin temperature. *Acta Anaesthesiologica Scandinavica* **2014**, *58*, 867–874, [https://onlinelibrary.wiley.com/doi/pdf/10.1111/aas.12351]. https://doi.org/https://doi.org/10.1111/aas.12351.
46. Lange, K.H.; Jansen, T.; Asghar, S.; Kristensen, P.; Skjønnemand, M.; Nørgaard, P. Skin temperature measured by infrared thermography after specific ultrasound-guided blocking of the musculocutaneous, radial, ulnar, and median nerves in the upper extremity. *British journal of anaesthesia* **2011**, *106 6*, 887–95.
47. Simonyan, K.; Zisserman, A. Very Deep Convolutional Networks for Large-Scale Image Recognition, 2014. https://doi.org/10.48550/ARXIV.1409.1556.
48. Ibtehaz, N.; Rahman, M.S. MultiResUNet: Rethinking the U-Net architecture for multimodal biomedical image segmentation. *Neural Networks* **2020**, *121*, 74–87. https://doi.org/10.1016/J.NEUNET.2019.08.025.
49. Zhou, Z.; Siddiquee, M.M.R.; Tajbakhsh, N.; Liang, J. UNet++: A Nested U-Net Architecture for Medical Image Segmentation. *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)* **2018**, *11045 LNCS*, 3–11. https://doi.org/10.1007/978-3-030-00889-5_1.
50. Wang, X.; Wang, L.; Zhong, X.; Bai, C.; Huang, X.; Zhao, R.; Xia, M. PaI-Net: A modified U-Net of reducing semantic gap for surgical instrument segmentation. *IET Image Processing* **2021**, *15*, 2959–2969. https://doi.org/10.1049/IPR2.12283.
51. Arteaga-Marrero, N.; Hernández, A.; Villa, E.; González-Pérez, S.; Luque, C.; Ruiz-Alzola, J. Segmentation Approaches for Diabetic Foot Disorders. *Sensors 2021, Vol. 21, Page 934* **2021**, *21*, 934. https://doi.org/10.3390/S21030934.
52. Zhou, Z.; Siddiquee, M.M.R.; Tajbakhsh, N.; Liang, J. UNet++: Redesigning Skip Connections to Exploit Multiscale Features in Image Segmentation. *IEEE Transactions on Medical Imaging* **2020**, *39*, 1856–1867. https://doi.org/10.1109/TMI.2019.2959609.
53. Peng, H.; Pappas, N.; Yogatama, D.; Schwartz, R.; Smith, N.A.; Kong, L. Random feature attention. *arXiv preprint arXiv:2103.02143* **2021**.
54. Nguyen, T.P.; Pham, T.T.; Nguyen, T.; Le, H.; Nguyen, D.; Lam, H.; Nguyen, P.; Fowler, J.; Tran, M.T.; Le, N. EmbryosFormer: Deformable Transformer and Collaborative Encoding-Decoding for Embryos Stage Development Classification. In Proceedings of the Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision, 2023, pp. 1981–1990.