mathematics

MDPI

Article

Ambiance Preservation Augmenting for Semantic Segmentation of Pediatric Burn Skin Lesions

Laura Florea [1], Corneliu Florea [1,*], Constantin Vertan [1] and Silviu Bădoiu [2]

[1] Image Processing and Analysis Laboratory, Faculty of Electronics, Telecommunications and Information Technology, National University of Science and Technology Politehnica Bucharest, 060042 București, Romania; laura.florea@upb.ro (L.F.); constantin.vertan@upb.ro (C.V.)
[2] Department of Anatomy and Embryology, Faculty of Medicine, Carol Davila University of Medicine and Pharmacy, 050474 București, Romania; silviu.badoiu@umfcd.ro
* Correspondence: corneliu.florea@upb.ro

Abstract: Burn injuries pose a significant threat to human life, with high morbidity and mortality rates. Accurate diagnosis, including the assessment of burn area and depth, is essential for effective treatment and can sometimes be lifesaving. However, access to specialized medical professionals is often limited, particularly in remote or underserved regions. To address this challenge and alleviate the burden on healthcare providers, researchers are investigating automated diagnostic tools. The severity of the burn and the affected body surface area are critical factors in diagnosis. From a computer vision perspective, this requires semantic segmentation of burn images to assess the affected area and determine burn severity. In collaboration with medical personnel, we have gathered a dataset of in situ images from a local children's hospital annotated by specialist burn surgeons. However, due to the limited amount of data, we propose a two-step augmentation approach: training with synthetic burn images and controlling the encoder by ambiance preservation. The latter is a technique that forces the encoder to represent closely the embeddings of images that are similar and is a key contribution of this paper. The method is evaluated on the BAMSI database, demonstrating that the proposed augmentations lead to better performance compared with strong baselines and other potential algorithmic improvements.

Keywords: burn assessment; image semantic segmentation; G-Cascade; ambiance preservation; synthetic burn images

MSC: 68T45

check for updates

Academic Editors: Zhi-Ri Tang and Ruihan Hu

Received: 26 January 2025
Revised: 15 February 2025
Accepted: 24 February 2025
Published: 25 February 2025

Citation: Florea, L.; Florea, C.; Vertan, C.; Bădoiu, S. Ambiance Preservation Augmenting for Semantic Segmentation of Pediatric Burn Skin Lesions. *Mathematics* **2025**, *13*, 758. https://doi.org/10.3390/math13050758

Copyright: © 2025 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (https://creativecommons.org/licenses/by/4.0/).

1. Introduction

The human skin is the body's largest organ and spans roughly 1.5–2 square meters in an average adult [1]. It is composed of two primary layers: the epidermis and the dermis. Embedded within these layers are structures like hair follicles, sweat glands, and sebaceous glands. These structures provide sources of proliferating epithelial cells (keratinocytes) that migrate into the clot and wound bed, playing an essential role in the healing process. When the skin's protective barrier is breached, it becomes susceptible to invasion by harmful microorganisms, potentially leading to infections and, in severe cases, sepsis. Burn injuries further complicate the repair process due to substantial fluid loss through the wound, which can commence just hours after the trauma [2]. Even minor burns can result in lasting functional and aesthetic repercussions for the patient.

In a recent review, Żwierełło et al. [3] pointed out that burn injuries depend on various factors, such as depth, cause, and the percentage of body surface area affected. These

factors are also criteria for classification and determination of the burn's severity. Burns are primarily classified into "partial-thickness" and "full-thickness". Partial-thickness burns are further divided into *superficial* and *deep* categories. Superficial partial-thickness burns affect the epidermis and the outer part of the dermis, preserving most appendage structures. They usually heal by themselves within 10–14 days with minimal scarring. On the other hand, deep partial-thickness burns penetrate deeper into the dermis, causing more substantial damage to appendage structures. These burns typically require 3–6 weeks to heal and have a high chance of hypertrophic scarring. Full-thickness burns destroy all layers of the skin and generally need surgical intervention for proper healing [4].

In the treatment of thermal injuries, prompt and appropriate methods are not only vital for saving the patient's life but also play a significant role in reducing hospital stays and accelerating recovery. However, one of the challenges in managing such injuries is the rapid and dynamic change in the appearance of burned areas, which complicates the process of providing timely and accurate diagnoses.

An automated system capable of localizing burn areas and assessing their severity would provide significant support to medical personnel. Additionally, such tools could address the challenges posed by geographical limitations, ensuring that patients in remote or underserved regions have access to reliable diagnostic assistance. By enabling informed health decisions and reducing the risk of complications, these systems can enhance the quality of care. A solution that analyzes images of burn injuries to determine the affected area and severity would not only streamline the diagnostic process but also serve as a valuable resource for improving patient outcomes. In this paper, we present a software solution designed to assist medical personnel by analyzing images of burned skin. This tool aims to support healthcare professionals in making faster and more precise diagnostic decisions, ultimately improving patient outcomes.

One of the main challenges faced was the limited amount of data available for traditional, purely supervised training. Specifically, in the context of pediatric burns, the difficulty in collecting clinical data and the shortage of medical specialists to annotate these burn images have resulted in a restricted amount of annotated data. To address these limitations, we sought additional image sources. We identified some options:

- Kaggle collection of cropped images of burns, where each image focuses solely on the burn area. The background has been removed and replaced with a plain white background.
- Healthy skin images that were collected for other purposes (for example, skin detection of hand gesture recognition).

We propose combining these two options to generate synthetic images of burned skin. These synthetic images can be utilized independently or in conjunction with annotated real burn ones through a technique that we call "ambiance preservation". Our focus is on the technical aspects of semantic segmentation for pediatric burned skin. The key contribution of this work lies in the introduction of augmentation stages that leverage this supplementary data to improve segmentation accuracy.

The structure of the paper is as follows: the next subsection reviews recent advancements in automatic burn image segmentation. Section 2 introduces the databases utilized, describes the algorithms and deep learning modules employed, and concludes with implementation details. Section 3 covers the evaluation of the proposed method, including ablation experiments. This is followed by discussions in Section 4, where the method's limitations are analyzed. Finally, the paper concludes with Section 5 summarizing the key findings.

Prior Works

The current procedure for assessing the degree of severity for burns is based, mainly, on visual inspection from surgeons. To support this process, automated methods utilizing image processing and pattern recognition techniques have been developed. Researchers have applied a wide range of approaches to classify burn images and to automatically predict the severity of burn damage.

In the early stages of developing solutions for automatic burn degree assessment, image analysis and computer vision techniques were used to classify patches from images with skin burns. For example, Kolesnik and Fexa [5] used a Support Vector Machine over modified inter-histogram distances to recognize burn degrees. Acha et al. [6] used thresholding on a weighted combination of hue and saturation to separate wounds from healthy skin. Wantanajittikul et al. [7] encoded the depth and color of the burn to classify the lesions.

More recently, Rangel-Olvera and Rosas-Romero used sparse dictionaries over color and texture information to reach 95% sensitivity in a database of nearly 100 images. The same authors [8] expanded the database to more than 200 images, prepared it for segmentation and proposed a solution based on k-means clustering, followed by linear discriminant analysis and Support Vector Machine classification. Kuan et al. [9] evaluated several traditional classification algorithms for identifying varying burn depths using an image mining approach applied to a dataset of 48 images. We emphasize that all these solutions were tested on very small datasets.

Building upon the development of deep learning, more sophisticated classification methods have been proposed. As an intermediate step, solutions that combine traditional image processing techniques (such as CLAHE) with neural network learning found some success in burn image retrieval [10]. In the same line, Chauhan and Goyal [11] proposed a method that utilizes CNN-extracted features from images, which are then input into Support Vector Machines to evaluate burn severity across various body parts. Abubakar and Ugail [12] utilized Support Vector Machines over features extracted from a pre-trained ResNet-101 for binary classification between normal and burned skin. Sevik et al. [13] evaluated multiple algorithms for a binary burn/background segmentation task using color and texture features on a dataset of 6750 image crops, each with a resolution of 64×64 pixels.

Hybrid methods are still considered useful due to insufficient databases for training deep learning models, but deep learning models are becoming increasingly common. For instance, using a modified ResNet-50Xt with attention, Yanev et al. [14] classified burn image patches from an original set of 94 images with varying degrees of burn.

Due to limited data, extraction of patches and classification has been approached more frequently, as emphasized by the solutions mentioned until now. Yet, this paper focuses on semantic segmentation. In this area, Khan et al. [15] applied the classical Otsu thresholding method for burn area segmentation, combined with a deep convolutional neural network for burn depth classification. Jiao et al. [16] utilized a convolutional neural network to separate and automatically grade burned skin from images. Boissin et al. [17] provided an overview of using deep learning for assessing burned skin through semantic segmentation, with a focus on listing available tools rather than diving into technical details. Karthik et al. [18] used a dataset of 104 images to train a recurrent neural network for classifying 1st, 2nd, and 3rd-degree burns. Liu et al. [19] gathered 1150 images and tested several architectures, finding HRNetV2 [20] as the most efficient in determining the area and the depth of burns. Xu et al. [21] built a two-step segmentation process where the body parts are identified and separated first, followed by the burn area and depth estimation using a ConvNetXt ("modernized" ConvNet based on visual transformer

findings). Rozo et al. [22] advanced the field by matching images of burned skin with images of the healed area to predict the healing process.

Precise segmentation of burned areas remains a demanding task, requiring substantial technical advancements and the development of robust frameworks, even though there is a strong motivation due to high morbidity and mortality rates [3,16]. This challenge arises from the variability of burned skin, which frequently blends with both healthy skin and the background. In our previous work [23], we introduced a semi-supervised approach that leveraged domain adaptation and soft pseudo-labeling to mitigate the limitations of labeled data. Specifically, we sought to enhance the model's generalization ability by utilizing both labeled and unlabeled images in the training process. To handle burn images without annotations, we adopted a pseudo-labeling strategy [24], wherein initially trained models generated pseudo-labels for unlabeled images. These pseudo-labels were subsequently refined and used for further training, thereby enabling the model to improve iteratively. The loss function played a crucial role in our approach. For labeled images, we employed cross-entropy loss to guide the learning process. For unlabeled images, we encouraged the model to produce confident predictions by defining the loss as the difference between certainty and the predicted logits. This confidence-based approach helped the model adapt more effectively to the unlabeled data, promoting better segmentation results even in the absence of ground truth annotations. While our method proved effective in improving segmentation performance, it was constrained by the lack of diverse backgrounds in the training data. The model's exposure to a limited range of skin tones, lighting conditions, and burn severities led to challenges when generalizing to new, unseen cases.

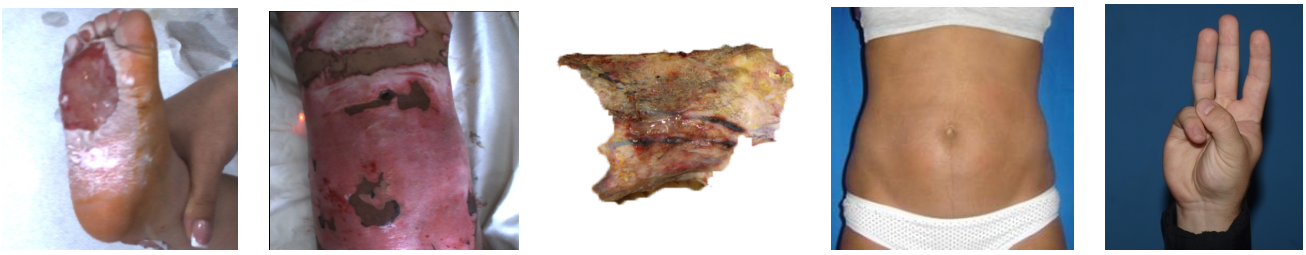
Although not yet applied to burn skin images, alternative solutions that do not rely on accurately annotated data have been explored in other areas of computer vision, particularly in settings where labeled data are scarce or noisy. These methods focus on utilizing imperfect labels while identifying and selecting the most reliable samples for training, helping models learn effectively despite label errors. A notable, very recent advancement in this area is the Progressive Sample Selection with Contrastive Learning (PSSCL) framework [25]. This method introduces a dynamic approach for sample selection, where a small, clean set of data is initially identified. A key component of the PSSCL framework is the use of contrastive loss, which plays a crucial role in distinguishing between similar and dissimilar samples.

Unlike the PSSCL approach, our method takes a different route by starting with a large set of synthetic images. These images are not perfect representations of real burns, but they provide a diverse and plentiful dataset that helps the model learn general patterns and features. By using these synthetic data early on, we expose the model to a wider range of burn types, lighting conditions, and backgrounds than we would with a limited real-world dataset. An advantage of using synthetic images is that they allow one to pre-train the model in a supervised manner, even in the absence of extensive real-world labeled data. Even if the synthetic dataset is not entirely accurate, it serves as a useful starting point for feature extraction and segmentation learning. In the final phase of training, we fine-tune the model using a smaller set of high-quality, well-annotated real burn images. This step allows the model to correct biases introduced by the synthetic data and adapt to the complexities of real-world cases. The fine-tuning process ensures that the model aligns better with real medical imaging conditions, improving its ability to generalize. Since the additional synthetic data are different from the real-life cases, the technique used falls into the broad category of domain adaptation techniques.

2. Materials and Methods

2.1. Databases

Images from several databases are used in this work. They will be detailed in the next paragraphs, and examples from each of them can be seen in Figure 1.



(**a**) (**b**) (**c**) (**d**) (**e**)

Figure 1. Image examples from the used databases: (**a**,**b**) Pediatric burns from BAMSI database, (**c**) Kaggle burns, and (**d**,**e**) healthy skin images (Abdominal Skin Dataset and, respectively, hand gesture recognition).

BAMSI (Burn Assessment by MultiSpectral Imaging) database [26] was collected at the Pediatric Burns Unit of the "Grigore Alexandrescu" Children Clinical Emergency Hospital in Bucharest, Romania. The images have been captured by medical staff. The database is divided into two sections. The first one contains 627 images annotated by experts (plastic surgeons). As described in our previous work on the topic [23], the second set consists of 968 images without annotations. However, the techniques proposed here demonstrate superior performance compared with the previous unsupervised learning-based approach. As a result, this unlabeled set is not used in the current work.

The labels of the annotated section are provided as full-image masks. The original resolution of the images is 1664 × 1248 pixels, but both annotations and operations are performed at a resolution of 320 × 240 pixels. The unannotated subset follows a similar structure but does not have labels.

The dataset includes images of burn injuries with varying severity from 55 pediatric patients aged between 8 months and 17 years. The average age is 4 years, with a median of 2 years and a standard deviation of 4 years. The images were taken between 1 and 62 days after the burn incident, with most captured within the first 10 days. Written consent was obtained from all parents or guardians for the use of this anonymized dataset for research purposes only under the supervision of the hospital's ethical committee.

The burns from the annotated imagesof burn images are categorized into the following grades:

- Grade I: They are superficial burns involving partial damage to the epidermis. These appear as erythema and cause pain, but they heal completely without surgical intervention. This grade corresponds to "superficial burns".
- Grade IIA: Burns affecting the full epidermis and the superficial dermis. They are areas of erythema, blisters with serous-citrin fluid, and pain. Healing occurs through non-excisional debridement and proper rebalancing, although some discoloration may remain.
- Grade IIB: Burns involving the entire epidermis, the superficial dermis, and part of the deep dermis. Clinically, they appear bright pink–red and cause pain. Healing involves both excisional and non-excisional debridement, often leaving scars and discoloration. Grades IIA and IIB together are considered "partial thickness" burns, with depth being the difference.

- Grade III: Burns penetrating the entire epidermis and dermis, reaching the hypodermis or deeper structures such as fascia, tendons, or bone. These burns appear as firm white or brown eschars, are painless, and lack capillary pulses. Surgical excision and grafting are mandatory. This corresponds to "full-thickness burns".
- Grade IV: Burns that extend beyond the dermis into underlying tissues like muscles or bones. These areas are insensate due to destroyed nerve endings and require surgical intervention and grafting. Grades III and IV are classified as "full-thickness burns".

Kaggle burns dataset is an unmarked and isolated collection of burn images, publicly available (available at https://www.kaggle.com/datasets/brendarangelolvera/human-skin-burns, accessed on 12 November 2024). It consists of burn images on a white background. The dataset description is limited, noting that the images were taken in hospitals and cropped to ensure the privacy of all patients. It can be associated with the work of Rangel-Olvera and Rosas-Romero [8], but the collection available is more comprehensive than the one described in the paper. It contains 1088 burn crops having the resolution 1134×1134 pixels.

Healthy skin collection comprises two databases that accompany the work of Topiwala et al. [27]. The first subset is the Abdominal Skin Dataset, which includes images sourced online through Google Images. The images have been manually segmented. To ensure diversity and minimize racial biases in segmentation algorithms, the selection process prioritized representation across different ethnic groups. The dataset includes 700 images of darker-skinned individuals, encompassing African, Indian, and Hispanic groups, and 700 images of lighter-skinned individuals, including Caucasian and Asian groups. Additionally, 400 images were chosen to represent individuals with higher body mass indices, evenly distributed between light and dark skin categories. Variations among individuals, such as hair and tattoos, as well as external factors like shadows, were carefully considered during dataset preparation. Each image has a resolution of 227×227 pixels.

The second subset is the Hand Gesture Recognition dataset [28]. It contains 899 images at resolution varying from 174×131 up to 640×480 and uncontrolled background and lighting conditions.

Both datasets are valuable for our purposes because they include uncontrolled background and lighting conditions, which help increase variability. In the BAMSI images, the background is typically associated with a hospital environment, but the conditions vary. Given the limited amount of data, there is a risk that deep models may overfit information from the background.

2.2. Segmentation Model

We propose a novel approach for developing an automatic segmentation deep learning model designed to function as an assistive tool for pediatric burn assessment. The schematic representation is illustrated in Figure 2. The trained model is designed to predict both the affected area and the burn degree.

The segmentation task utilizes a deep learning framework built on the Cascaded Graph Convolutional Attention Decoder (G-Cascade) architecture [29]. Like many segmentation models inspired by the U-Net [30], this architecture adopts a similar encoder-decoder structure. However, recent models frequently utilize the more powerful visual transformer paradigm. A notable example is the TransUnet model [31], which serves as a straightforward adaptation of U-Net using transformers. In contrast, the G-Cascade model [29] takes a significant step forward by progressively refining multi-stage feature maps produced by a hierarchical transformer in the encoder and incorporating an efficient graph convolution block in the decoder for enhanced performance.

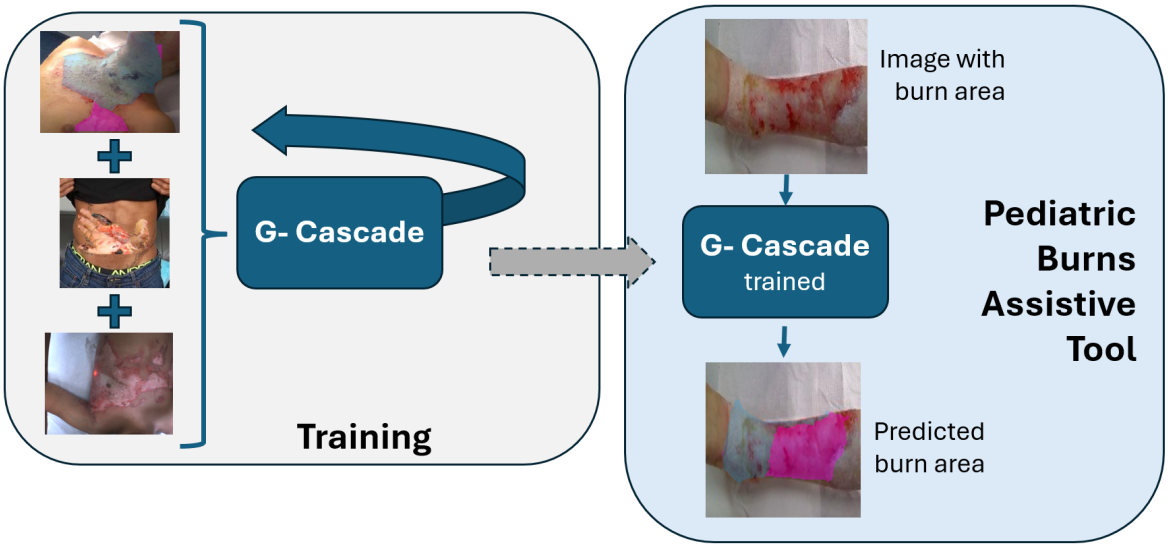


Figure 2. The method schematic. The proposed method assumes training of deep model (G-Cascade) using, sequentially, three types of data (unlabeled, synthetic, and labeled). The trained model is further used to assist medical personnel by predicting the burned area and the degree of burn in images with burns.

The encoder in the G-Cascade framework is built upon the Pyramid Vision Transformer (PVT) architecture [32], which is visually depicted in Figure 3. It employs a self-attention mechanism to capture long-range dependencies within the data. The architecture is rooted in the principles of the Visual Transformer (ViT) family [33], treating each image as a sequence of fixed-length tokens (or patches) that are fed into multiple Transformer layers for processing and classification. The key element is the spatial-reduction attention (SRA) layer that replaces the classical multi-head attention by reducing the spatial scale of key (K) and Values (V):

$$\begin{aligned} SRA(Q,K,V) &= Concat(head_0,\ldots,head_{N_i})W^O \\ head_j &= Attention(QW_j^Q, SR(K)W_j^K, SR(V)W_j^V) \end{aligned} \tag{1}$$

where W^O, W_j^Q, W_j^K, and W_j^V are the learnable projection matrix from the attention block; N_i is the head number of the attention layer in Stage i. The PVT returns four feature maps that are flattened and used as embeddings ($\mathbf{x}_1, \mathbf{x}_2, \mathbf{x}_3, \mathbf{x}_4$). In this work, these embeddings are extracted and used in a pre-training procedure, which we refer to as *ambiance preservation*.

The decoder follows the encoder and has to preserve and refine these feature maps ($\mathbf{x}_1, \mathbf{x}_2, \mathbf{x}_3, \mathbf{x}_4$). It also aims to maintain long-range information through the use of graph convolution blocks within global receptive fields. The main components of the G-Cascade model's decoder [29] are illustrated in Figure 4.

The G-Cascade decoder model integrates three main components: efficient up-convolution blocks (UCBs) for feature upsampling, graph convolutional attention modules (GCAMs) for enhancing feature maps, and segmentation heads (SegHeads) to produce the final segmentation output. The model employs four GCAMs corresponding to the four stages of pyramid features extracted by the encoder. Each of the GCAM contains the same blocks, the difference being the working resolution (size).

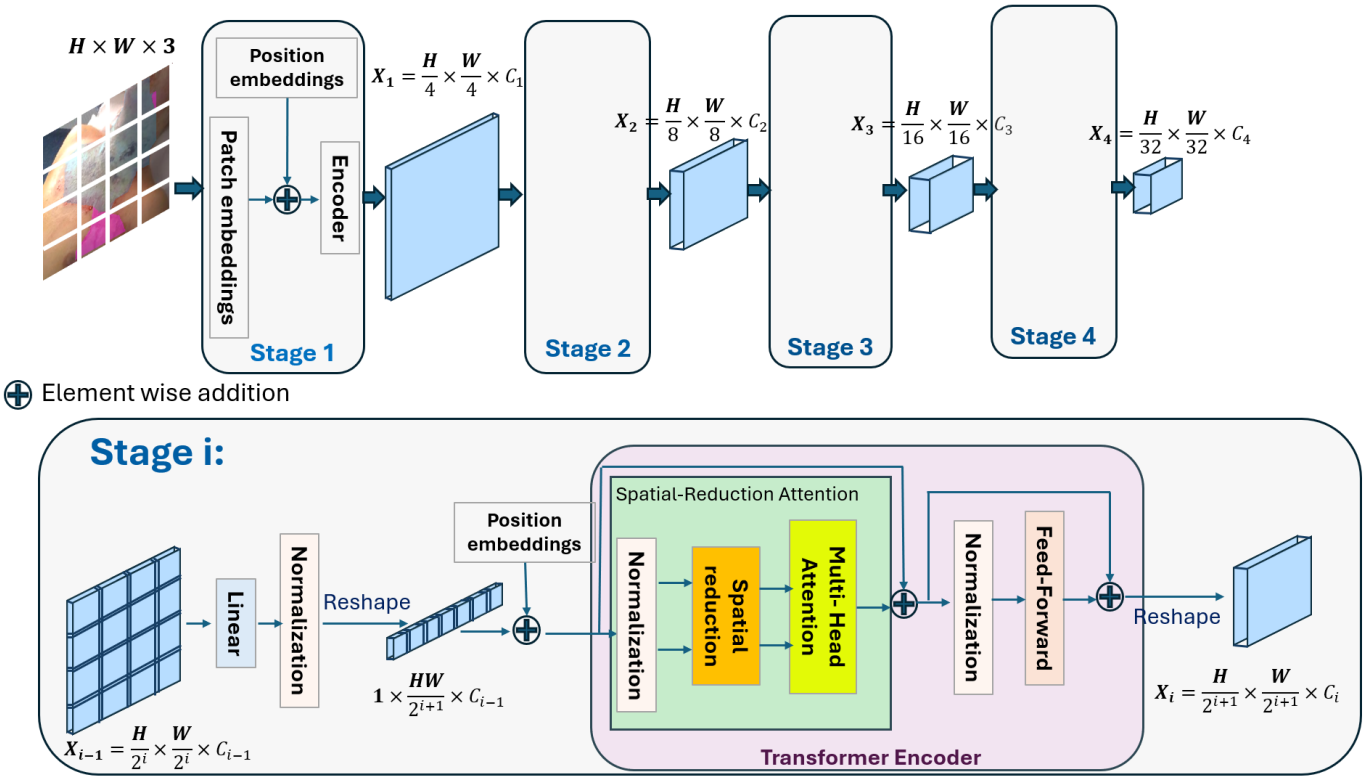


Figure 3. The Pyramid Vision Transformer (PVT) [32], which is used as encoder. The model contains four stages, each made from a patch projection layer and an encoder transformer.

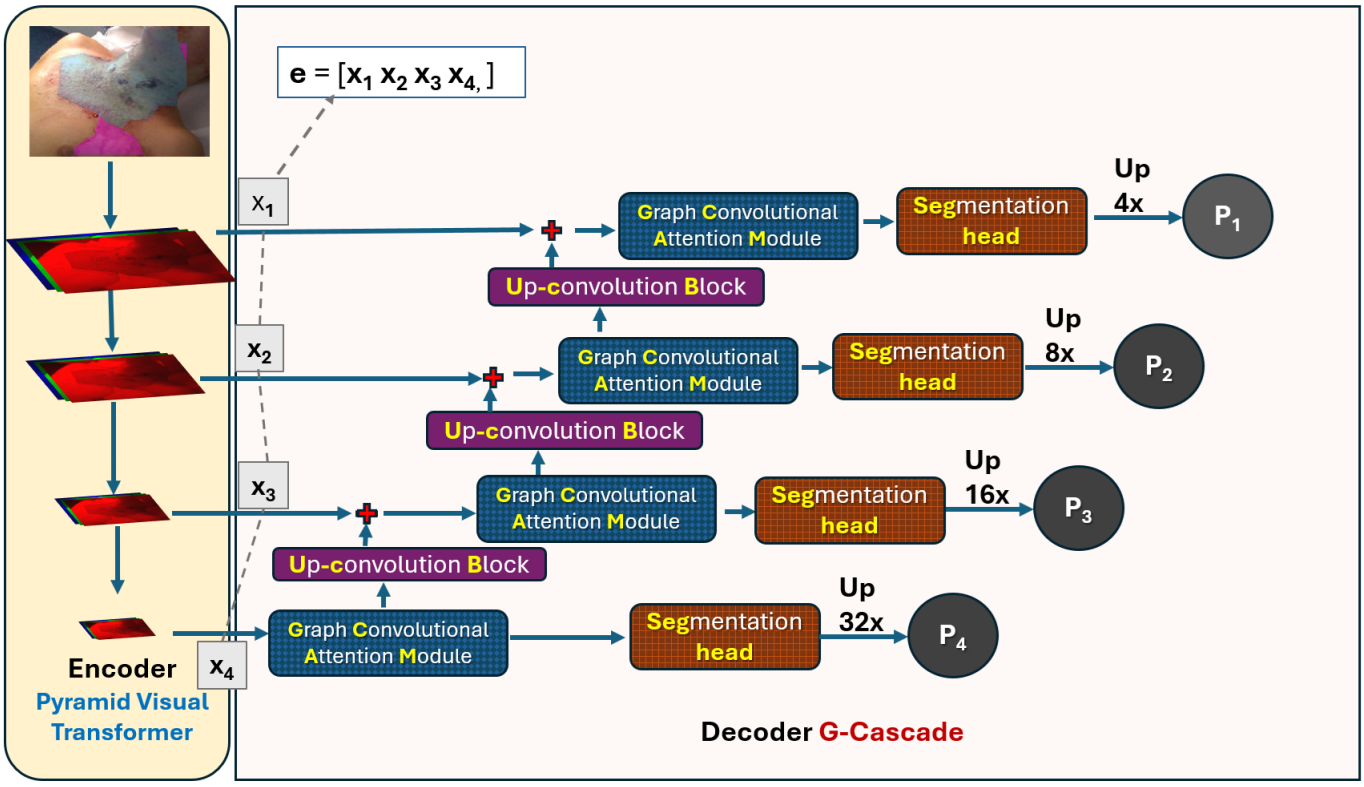


Figure 4. The G-cascade model [29] using a PVT encoder [32]. The embeddings are captured at the exit of the encoder and concatenated to form the feature descriptor **e**.

To merge multi-scale features, the upsampled features from the previous decoder block are combined with features from skip connections through addition or concatenation.

These combined features are then processed by the GCAM module, which enriches the semantic information. The output from each GCAM is subsequently passed to a prediction head, and the outputs of all four prediction heads are aggregated to generate the final segmentation result.

The critical components of the G-Cascade decoder are the UCB, GCAMs, and Seg-Heads. Each GCAM is composed of a Graph Convolution Block (GCB) and a Spatial Attention (SPA) block, as introduced by Chen et al. [34]. The spatial attention module is designed to extract local contextual information from the input feature map $\mathbf{x}$:

$$GCAM(\mathbf{x}) = SPA(GCB(\mathbf{x})) \tag{2}$$

The Graph Convolution Block (GCB) used in the G-Cascade architecture [29] includes a graph convolution layer followed by two 1×1 convolution layers, each paired with batch normalization and ReLU activation layers; it is developed from the Visual Graph Neural Network [35]. It contains a graph convolution layer denoted by $GConv(\cdot)$ and two 1×1 convolution layers $C(\cdot)$ each followed by a batch normalization layer $BN(\cdot)$ and a ReLU activation layer $ReLU(\cdot)$:

$$GCB(x) = ReLU(BN(C(GConv(ReLU(BN(C(\mathbf{x}))))))) \tag{3}$$

The Spatial Attention (SPA) block highlights significant regions within a feature map, computed as follows:

$$SPA(\mathbf{x}) = Sigmoid(Conv([\mathrm{C}_{\max}(\mathbf{x}), C_{avg}(\mathbf{x})])) \circledast \mathbf{x} \tag{4}$$

In Equation (4), $Sigmoid()$ denotes the Sigmoid activation function, C_{max} and C_{avg} represent the channel-wise maximum and average pooling operations, Conv is a 7×7 convolutional layer with padding, and $\circledast$ is the Hadamard product.

The up-convolution block (UCB) is designed to restore spatial resolution. It achieves this by upsampling features from the current layer to align with the dimensions of the next skip connection. It includes an UpSampling operation Up (having a scale factor of 2), a 3×3 depth-wise convolution (DWC), batch normalization (BN), a ReLU activation, and a 1×1 convolution (Conv), mathematically expressed as

$$UCB(\mathbf{x}) = Conv(ReLU(BN(DWC(Up(\mathbf{x}))))) \tag{5}$$

The Segmentation Head (SegHead) takes the refined feature maps from the four stages of the decoder as input and generates four segmentation maps. Each segmentation map is produced by applying a 1×1 convolution over the feature maps, with the number of output channels corresponding to the target classes for classification.

The final segmentation output is obtained by aggregating the four segmentation maps, $\mathbf{p}_1$, $\mathbf{p}_2$, $\mathbf{p}_3$, and $\mathbf{p}_4$, generated by the prediction heads for the four decoder stages. This aggregation is performed by summing the outputs:

$$out_{map} = \mathbf{p}_1 + \mathbf{p}_2 + \mathbf{p}_3 + \mathbf{p}_4 \tag{6}$$

In the supervised learning stages, the optimization loss is formulated as a combination of the weighted cross-entropy (WCE) loss and the Dice Score:

$$\mathcal{L} = \alpha \mathcal{L}_{WCE} + \beta \mathcal{L}_{DICE} \tag{7}$$

Cross-entropy loss weights were assigned to linearly increase from the background to Grade IV burns (background: $w_1 = 1, \ldots,$ Grade IV burns: $w_5 = 5$). Compared with our previous work [23], the skin is no longer detected by an off-the shelf Segment Anything Model. The problem was alleviated by the introduction of pre-training for synthetic burn images.

2.3. Synthesize Burn Skin Images

Given the limited number of examples in the original burn image database, BAMSI, we sought to augment it with additional data. To achieve this, we synthesized new images derived from real ones sourced from the internet. As previously mentioned, irregularly shaped patches of burned skin are available online in the Kaggle burns dataset. A significant challenge for our synthetic images lies in the need for a highly varied background. While such backgrounds are typical in hospital images, they introduce considerable confusion for the deep segmentation model when relying solely on the annotated images.

To address this, we overlaid burn patches onto healthy skin, carefully selecting healthy skin images to exhibit high variability. This variability is essential for enabling the model to effectively distinguish burned skin patches from the background. For this purpose, we utilized examples from two databases (Hand Gesture Recognition and Abdomen) to represent images with healthy skin. The process of generating a synthetic burned skin image is detailed below and illustrated in Figure 5:

1. **Input**: Database with burned skin patches and database with healthy skin (including skin map label).
2. For each healthy skin image, identify the largest inscribed rectangle within the skin area.
3. For each burned skin patch, create the burned skin map by applying a threshold to the patch.
4. Randomly select a burned skin patch and, respectively, a healthy skin image.
5. Rotate by an arbitrary angle and introduce a slight warp on the burned skin patch. Apply the same transform also on the burned skin map.
6. Compute the bounding box of the skin patch. If the skin patch does not fit inside the healthy skin inscribed box, scale down the burned patch.
7. Overlay the transformed burned skin patch with healthy skin.
8. Form the burn segmentation map by placing the map associated with the burn onto the same location. The burn will be considered from grade IIA, and the rest is background.

One particular aspect that requires further clarification is how to determine the largest rectangle inscribed in a binary shape, which in our case, is the healthy skin map. This process is more complex than determining the bounding (circumscribed) box. The procedure is iterative and follows these steps:

- Find all possible rectangles: compute all possible rectangles for the given shape.
- Check each rectangle: For each possible rectangle, verify if it lies entirely within the binary shape.
- Store the largest rectangle: Keep track of the rectangle with the maximum area that is fully inscribed.
- If a speed-up is necessary, downscale the binary map with a factor, and after determining the inscribed rectangle, scale up its coordinates with the same factor.

It should be noted that for our application, the resulting image needs to be at the resolution required by the deep segmentation model (i.e., 227×227). For this resolution, the computation of the inscribed box is fast enough.

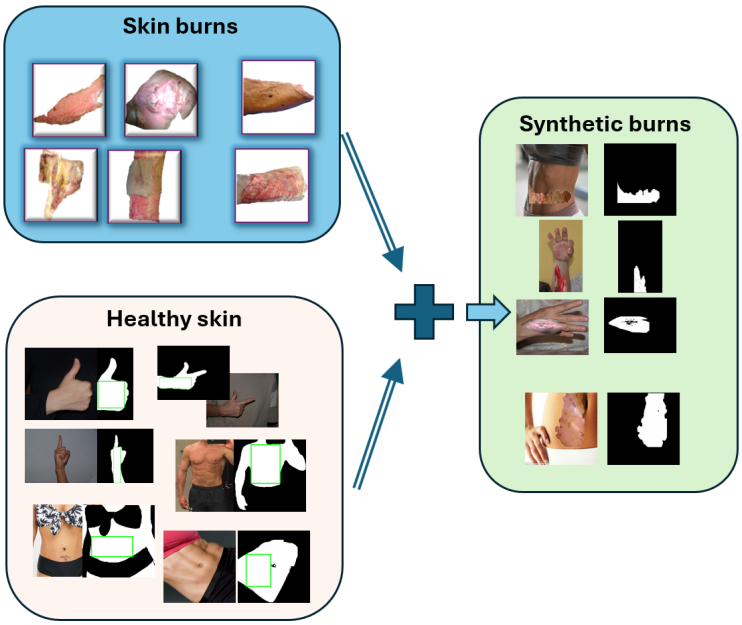


Figure 5. The process to create synthetic burn skin images. First, given healthy skin images with annotations, the inscribed box is computed. Then, a random burn image and a random healthy skin image are selected and combined to create a synthetic burned skin image. The resulting image has a segmentation label map generated from the location where the burn was overlaid.

2.4. Ambiance Preservation

The proposed *ambiance preservation* stage uses labeled data. The data here are either with images from the BAMSI database, which have original annotations, or synthetic burn skin images that have one class label that marks the position where the burn patch has been overlaid.

This stage focuses solely on the encoder, aiming to force it to generate embeddings in a space that maintains the "ambiance" (or structural organization) of the original image space. This implies that similar images should produce similar embeddings, while dissimilar images should result in distinct embeddings. The structure that combines both the class and location is referred to as "Ambiance" in this context. The purpose of the loss function is to enforce that the embeddings retain this ambiance.

Now the problem is to find a suitable definition of "similar" and "dissimilar". Taking into account that the purpose is semantic segmentation, two images should be "similar" if they have a *patch from the same class* and in the *related positions*. In parallel, images that have burns from different degrees in the same position or the same burn but in different positions should be counted as "dissimilar".

The next step is to encode these principles into a loss function. To do that, we draw inspiration from the MarginMix [36] concept. In the MarginMix, the loss function is computed as the difference between the distance from the current embedding to the centroid of its class and, respectively, to the other classes. The formal definition is

$$\mathcal{L}_{\mathcal{E}} = \sum_{i=1}^{N} \left(\mathcal{D}\left(\frac{\mathbf{e}_i}{\|\mathbf{e}_i\|_2}, \frac{\mathbf{c}^c}{\|\mathbf{c}^c\|_2} \right) - \frac{1}{C-1} \sum_{j=1, j\neq c}^{C} \mathcal{D}\left(\frac{\mathbf{e}_i}{\|\mathbf{e}_i\|_2}, \frac{\mathbf{c}^j}{\|\mathbf{c}^j\|_2} \right) \right) \tag{8}$$

where an instance $\mathbf{I}_i$ from class c ($y_i = c$) generates the embedding $\mathbf{e}_i$; $\mathbf{c}^c$ is the centroid of class c from the total of C classes. $\mu_i^c \in [0;1]$ shows the measure in which data $\mathbf{I}_i$ belongs to class c and is similar to the membership values used in classical fuzzy clustering. They are formally defined as

$$\mathbf{c}^c = \frac{\sum_{i=1}^{N} \mu_i^c \mathbf{e}_i}{\sum_{i=1}^{N} \mu_i^c}; \quad \mu_i^c = \begin{cases} 1, & y_i = c \\ 0, & y_i \neq c \end{cases} \tag{9}$$

This concept was used in the context of image classification, and the idea is that all data are potentially from all classes, but with different degrees of memberships. Yet the concept of "class", alone, is less relevant for semantic segmentation as it ignores spatial localization.

In the case of semantic segmentation, we define the expected distance from two embeddings as follows:

$$\mathcal{D}(\mathbf{e}_i, \mathbf{e}_j) = aDC(Y_i, Y_j) \cdot \left(\frac{\mathbf{e}_i}{\|\mathbf{e}_i\|_2} - \frac{\mathbf{e}_j}{\|\mathbf{e}_j\|_2} \right)^2 \tag{10}$$

where $\mathbf{e}_i$ is the embedding produced for image I_i, while $aDC(I_i, I_j)$ is an adapted version of the Dice score (or Dice coefficient) between the labels (segmentation maps) Y_i and Y_j of images I_i and, respectively, I_j. The adaption refers to the fact that the background is excluded. It is formally defined as two times the ratio between intersection over union:

$$aDC(Y_i, Y_j) = \frac{1}{C} \sum_{c=1}^{C} \frac{2 \cdot |Y_i^c \cap Y_j^c|}{|Y_i^c| + |Y_j^c|} \tag{11}$$

where summation ignores $c = 0$, which is the background class. Y_i^c selects, from the segmentation label of I_i, only pixels marked as being in class c, and Y_j^c is its counterpart from image I_j. Visually, the process is detailed in Figure 6. We recall that the Dice score is 1 if the maps are identical and 0 if there is no spatial overlap between class patches.

Subsequently, the ambiance preservation loss function $\mathcal{L}_{AP}$, for a set (i.e., batch) of C examples, is the sum of distances between all points:

$$\mathcal{L}_{AP} = \sum_{i=1}^{C} \sum_{j=1, j>i}^{C} \mathcal{D}(\mathbf{e}_i, \mathbf{e}_j) \tag{12}$$

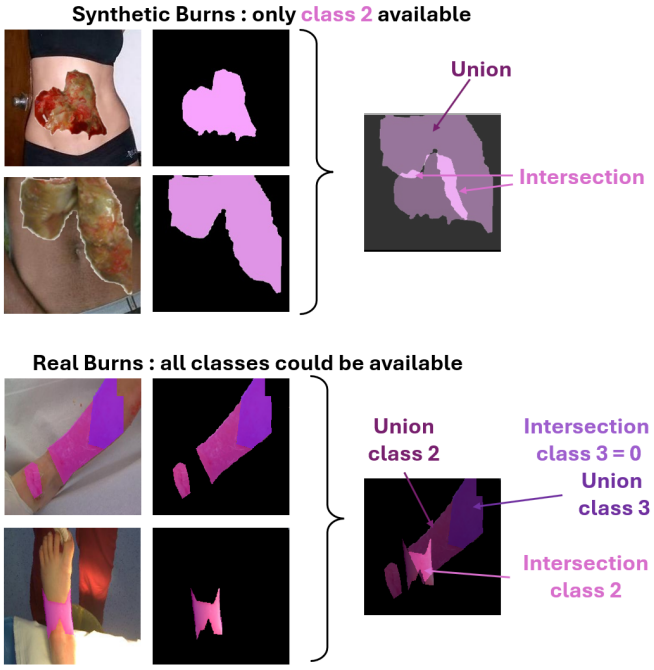


Figure 6. Computation of the adaptive Dice Coefficient, used as a weighting factor when computing the Ambiance preservation loss. The background label (marked with black) is not taken into account.

Only the burn classes are considered. The Dice Coefficient uses the ratio between intersection over union for each class. These are marked for two examples: in the upper part, there is a pair of synthetic burns, which are considered to have the burn grade IIA (class 2). In the lower part, there is an example with real burn images, which may contain all classes. In this example, both contain class 2, and thus, the term for class 2 is non-zero, while only one contains class 3 (grade IIB), and thus, the intersection is null.

2.5. Implementation

The training process takes into account the fact that deep networks require large amounts of training data to learn useful values in all weights. For many applications, it is custom to use pre-trained weights, as they offer a much better starting point than initialization from random values. In our case, we divide the training process into three stages. The first two stages may be seen as a pre-training, while the final stage acts as a fine-tuning. In more detail, the three stages, which are also illustrated in Figure 7, are as follows:

1. Training on the synthetic burn skin images. The synthetic burn skin image set is much larger than the one with real images (BAMSI), and it contains more variability with respect to the background and to the healthy skin. We recall that all synthetic burns have been artificially labeled with burn grade IIA, which is class 2.
2. Ambiance preservation learning. In this stage, only the encoder is used, and the procedure aims to ensure that the embeddings represent the relationships between the input images. Two images with the same burn degree at overlapping positions should have similar embeddings, as highlighted in the previous section.
3. The last stage is purely supervised learning, where only the annotated real images from BAMSI are used. The large number of epochs is required by the nature of the model which is based on visual transformer.

For all stages, the learning rate started at the same value. For the first two stages, it remained at the original value for the entire duration, while for the last, the supervised learning, it decreased once, at the half of the epochs.

Since synthetic burn skin images do not perfectly replicate real-life conditions, they are used exclusively for pre-training. This allows the model to learn more effectively general patterns and useful features than when starting with random initialization. However, the third stage, which involves fine-tuning, relies entirely on real-life images accurately annotated by surgeons. This high-quality, clean data help correct inaccuracies the model may have acquired during pre-training. The advantages of this approach will be further demonstrated in the ablation study.

The code is in Python, version 3.10.12 and was developed using PyTorch, version 2.3.10, available at https://pytorch.org/, supported by PyTorch Foundation. for deep learning. All models were trained on a single NVIDIA RTX A4000 GPU (NVIDIA, Santa Clara, CA, USA) with 16 GB of memory. The Adam optimizer was used for training, with a learning rate of 0.0001 and a cosine scheduler. For the third stage, it was reduced by a factor of 10 at half of the epochs. The batch size was set to 1 image for stages 1 and 3 and to 8 images for stage 2. All backbone models were pre-trained on ImageNet. The G-Cascade model comprises 1.78 million parameters and requires 0.342 FLOPs.

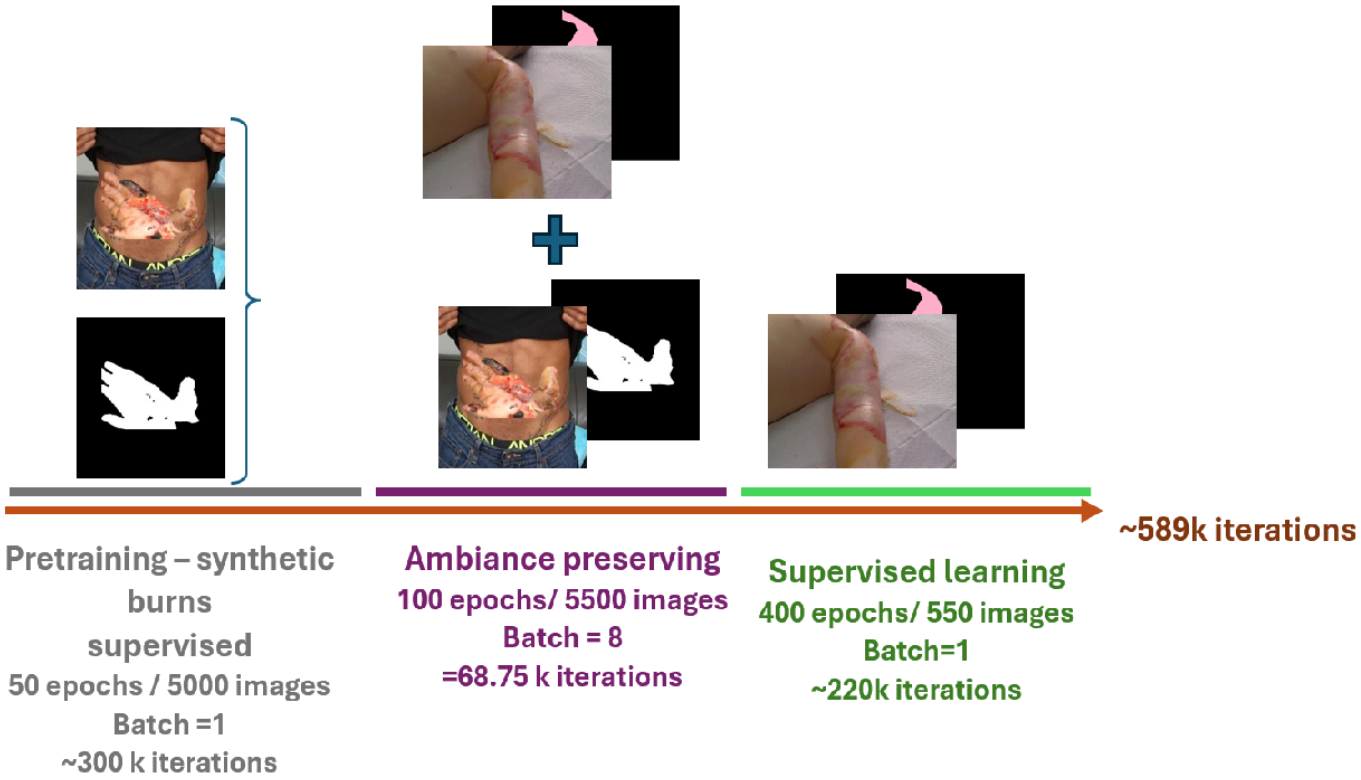


Figure 7. The training process has been split into three stages: (pre)training in supervised manner with synthetic burns, partially supervised training using the Ambiance preservation technique over the encoder embeddings over all images, and supervised training of the annotated images only.

The division between training and testing is fixed and has been set at database formation. Given the sets, the order of images inside an epoch is randomly chosen without fixed seed. Thus, consecutive training processes present a different image order, yet the results are very similar (less than 0.5% variance in accuracy), emphasizing that convergence is stable. Each image goes through a chain of random transformation which includes random horizontal and vertical flipping and random rotation with a degree up to 30 degrees. Training images from BAMSI are randomly cropped and zoomed: the 227 × 227 image is taken from the original images of 320 × 240 by first extrapolating the original image to a random resolution between the two, followed by a crop from a random location that ensures output image fits entirely in original one. Testing images and synthetic images are only rescaled to the 227 × 227 resolution. The results presented are obtained by averaging 3 runs of the experiments.

Performance evaluation was conducted using the average accuracy for burn degree predictions and the Dice score. The average accuracy (AvgAcc) implies computing accuracy on each class and averaging over all burn degree classes.

3. Results

Table 1 presents the results of the segmentation process along with a comparison against standard baselines. From the convolutional networks family used for semantic segmentation we selected HRNetV2 [19,20], UNet3+ [37] and DeepLab-v3 [38]. We have also reimplemented and tested the models used in previous solutions for burned skin segmentation, namely [16,22]. From the transformer family, we tested against TransUNet [31].

Table 1. Results of burn skin semantic segmentation on BAMSI dataset. The best results are shown in bold. "AmbPres" stands for the here proposed Ambiance preservation. CE denotes cross-entropy.

Strategy	Architectures/Loss	AvgAccuracy [%]	Dice Score [%]
Supervised	MaskRCNN − CE [16]	36.73	35.90
	UNet − CE [22]	32.26	36.01
	UNet3+ − CE	34.14	37.69
	UNet3+ − CE+Dice	35.45	45.85
	DeepLab-v3 − CE	36.37	37.15
	DeepLab-v3 − CE + Dice	37.15	43.58
	HRNetV2 − CE + Dice [20]	39.34	41.20
	TransUnet − CE + Dice [31]	48.78	45.24
	G-Cascade − CE [23]	39.72	33.81
	G-Cascade − Dice [23]	43.28	34.94
	G-Cascade − CE + Dice [23]	66.72	48.14
Domain Adaptat.	G-Cascade + Pseudolab − 800 ep [23]	65.24	45.17
	G-Cascade + ColorTransfer − 800 ep	66.25	48.71
	G-Cascade + ColorTransfer − 200 ep [23]	67.07	49.85
	G-Cascade + AmbPres − proposed	**70.11**	**51.15**

The improvement with respect to the baseline (denoted by "G-Cascade − CE + Dice" and reported in our previous work [23]) is due to the use of additional different data, therefore using a technique from the domain adaptation. To offer a better perspective on the method, we also compare it with other variants that use domain adaptation solutions.

In our previous approach [23], we noted that our database contains both labeled and unlabeled images; therefore, a semi-supervised approach was also feasible. The typical strategy of pseudolabels [24] for burn images without annotations was utilized; it required employing cross-entropy as the loss function, and for non-annotated images, the model is encouraged to produce confident predictions, with the loss being defined as the difference between certainty and the predicted logits. We also compare with our previous proposal [23], which used a statistical description of colors existing in burns and self annotated images that were initially without annotations; compared with standard pseudolabels, that version allowed inclusion in the loss function of the Dice Coefficient.

Regarding the results, two key observations can be made. First, adopting a dual-loss approach proves advantageous: combining cross-entropy (CE) with the Dice score achieves the best performance, whereas relying on a single loss function reduces effectiveness. Second, pre-trained Vision Transformers (ViTs) outperform convolutional networks, with the more sophisticated G-Cascade architecture delivering better results than the TransUNet model within the ViT family.

Overall, the proposed method leads to the best performance by a noticeable margin against both purely supervised reference and our previous solution [23]. The addition of synthetic burns and Ambiance preservation allowed better separation of background and better discrimination between burned skin and healthy skin. However these aspects are developed into next subsection, dedicated to ablation studies.

The confusion matrix of the proposed solution may be seen in Table 2, and visual examples of original images, ground truth annotation, previous method [23], and predicted masks with the proposed solution may be seen in Figure 8. The masks have been overlaid onto the original image.



Figure 8. Example of images with pediatric skin burns (original images—first column), ground truth (second column) and predictions with the solution from [23] (third column) and the proposed solution (forth row). Annotations and predictions showing in this example are color-coded by pink (grade IIA), magenta (grade III), and cyan (grade IV).

Table 2. Confusion matrix in [%], at pixel level, over the burn areas of the best solution: G-Cascade with pre-training on synthetic burns and Ambiance preservation. The true and predicted labels are marked with bold.

		Predicted				
		Grade I	**Grade IIA**	**Grade IIB**	**Grade III**	**Grade IV**
True	**Grade I**	85.12	5.56	4.17	2.27	2.88
	Grade IIA	13.98	64.01	18.88	3.13	0
	Grade IIB	17.88	10.33	66.98	4.81	0
	Grade III	11.03	3.27	24.45	59.71	1.54
	Grade IV	3.48	2.99	2.07	16.7	74.76

Ablation and Parameter Influence

The main contribution of this paper lies in the two-stage augmentation process: pre-training with synthetic burns and encoder tuning using the Ambiance preservation technique. To demonstrate their individual impact, we conducted ablation tests, with the results presented in Table 3. As observed, both augmentation techniques positively influence performance. Pre-training with synthetic burns provides a greater contribution, likely due to the increased background variability it introduces and the additional perspectives of

normal skin it offers. These factors enhance the main model's generalization capabilities with respect to background complexity and skin variability.

Table 3. Ablation experiments. The baseline and removal or addition of each of the augmentation stages have been evaluated.

Synthetic Burns	Amb. Preserv.	AvgAccuracy [%]	Dice Score [%]
		66.72	48.14
✓		68.65	50.01
	✓	67.86	49.21
✓	✓	70.11	51.15

Furthermore, both augmentation techniques rely on several parameters, and we evaluate their influence here. Regarding pre-training on synthetic burns, the key factor was the number of images generated. The performance variation relative to the number of images is presented in Table 4. In this experiment, the Ambiance preservation augmentation was excluded, but supervised training was conducted. As observed, a small number of images does not lead to noticeable improvement, and increasing the dataset beyond 5000 images yields minimal additional benefits.

Table 4. Impact of the number of synthetic burned skin images used in pre-training over the final accuracy.

No. Syntehtic Burns Img	AvgAccuracy [%]	Dice Score [%]
0	66.72	48.14
500	66.65	48.24
2000	67.16	48.84
5000	68.65	50.01
10,000	68.73	50.15

Regarding the Ambiance preservation encoder training, we tested the validity of applying the MarginMix approach, which forces two embeddings to be similar if they represent burns of the same severity, regardless of their location. The evaluation revealed that this approach had no significant impact, with performance remaining virtually identical to the case where it was not applied. Consequently, we concluded that burn localization plays a crucial role, and using the Dice Coefficient for encoding leads to improved results.

4. Discussion

In this paper, we introduced a novel framework for semantic segmentation. The approached theme refers to the identification of the area and the degree of skin burns. The main limitation of the problem is a typical one for medical image analysis: the finite amount of annotated data.

The limitation in data arises because, for patients, the primary focus is on care and treatment. Only once their condition stabilizes does the opportunity to record images for further investigation and study arise. As a result, databases in medical image analysis are significantly smaller compared with those used in general image analysis. The second limitation is with respect to annotation, and this may be related to the busy schedule of the doctors, who also prioritize patients' treatment in front of medical image annotation. The consequences are visible as the annotated part of our dataset is less than 700 labeled images.

To compensate for the limited amount of annotated data, we devise two strategies. The first is referred to as Ambience Preservation, as it ensures that the embeddings generated by the encoder maintain the topology based on the structure and localization of burn areas.

We collectively name the class and the location as the "ambiance" of the data. The metric used to control the distance between embeddings is based on Euclidean distance. This method helped, yet the impact is somehow limited. There are several potential reasons for its limitation.

One such limiting reason refers to the very high dimensionality of the embeddings; the dimensions of the vectors are as follows: $\mathbf{x}_1 = 512$, $\mathbf{x}_2 = 320$, $\mathbf{x}_3 = 128$, $\mathbf{x}_4 = 64$, which leads to the total dimension of $\mathbf{e} = 1024$. The size encourages spreading the correction over all dimensions, and this may be problematic. The second reason for limitation lies in the fact that only the encoder is tuned while the decoder is left untouched. A third potential reason is about the diversity of the burns from the same degree: they have individual cues that made the doctor categorize them in those specific degree, while the holistic structure may be different from one case to another. For example, two burn areas with 4th degree, pixelwise, are different. Forcing them together may be suboptimal.

The second strategy refers to creating synthetic images that contain burn patches surrounded by healthy skin areas. This was achieved by combining a public set that contains only burned skin areas (no person, no background) and two databases with healthy skin. While overall, this synthetic database is larger than our real burn images database, the variability here is also limited. The healthy skin is limited to hand images or abdomen areas. At the same time, in many real burn cases, other parts of the body, such as the legs or the face, are wounded.

Another limitation of using synthetic burns refers to the fact that the degree of each burn patch is not known, and therefore, we labeled them into a unique degree, which, probably, in many cases, is incorrect. This potential error forced us to use this stage at the beginning of the training. We needed the supervised stage to come later and correct all the potential inaccuracies introduced here.

However, our experiments showed that both strategies are beneficial, and ablation studies argue convincingly that the performance improves when they are used. Further experiments showed some insights, and these insights can help other works to build on this foundation.

5. Conclusions

This paper addresses burn skin segmentation in color images, a step toward assistive diagnosis. Due to limited annotated data, we introduce two augmentation strategies. Our key innovation, Ambiance preservation, clusters segmentation embeddings for images with similar burn severity and location, measured via the Dice coefficient. Additionally, by pre-training with synthetic images made by overlaying synthetic burns onto healthy skin in random positions, we increase robustness against variations in healthy skin and background textures. Results show our method outperforms a purely supervised baseline and other augmentation techniques, improving segmentation accuracy and reliability.

Author Contributions: Conceptualization, L.F. and C.F.; methodology, L.F. and C.F.; software, C.F.; validation, L.F. and C.V.; resources, S.B.; data curation, S.B.; writing—original draft preparation, C.F.; writing—review and editing, L.F. and C.F.; supervision, C.V.; funding acquisition, L.F. All authors have read and agreed to the published version of the manuscript.

Funding: This work was supported by grant 68/11.10.2023 from the National Program for Research of the National Association of Technical Universities—GNAC ARUT 2023.

Data Availability Statement: The datasets presented in this article are not readily available because of the patients' confidential agreement (patients agreed for the data to be used only for academic research). Requests to access the datasets should be directed to Laura Florea.

Conflicts of Interest: The authors declare no conflicts of interest. The funders had no role in the design of the study; in the collection, analysis, or interpretation of data; in the writing of the manuscript; or in the decision to publish the results.

Abbreviations

The following abbreviations are used in this manuscript:

BAMSI	Burn Assessment by MultiSpectral Imaging
ViT	Visual Transformer
PVT	Pyramid Visual Transformer
UCB	Up-convolution Blocks
GCAM	Graph Convolutional Attention Module
SegHeads	segmentation heads
SPA	Spatial Attention
GCB	Graph Convolution Block
DWC	Depth Wise Convolution
BN	Batch Normalization
CE	Cross-Entropy

References

1. Richardson, M. Understanding the structure and function of the skin. *Nurs. Times* **2003**, *99*, 46–48. [PubMed]
2. Rowan, M.P.; Cancio, L.C.; Elster, E.A.; Burmeister, D.M.; Rose, L.F.; Natesan, S.; Chan, R.K.; Christy, R.J.; Chung, K.K. Burn wound healing and treatment: Review and advancements. *Crit. Care* **2015**, *19*, 243. [CrossRef] [PubMed]
3. Żwierełło, W.; Piorun, K.; Skórka-Majewicz, M.; Maruszewska, A.; Antoniewski, J.; Gutowska, I. Burns: Classification, pathophysiology, and treatment: A review. *Int. J. Mol. Sci.* **2023**, *24*, 3749. [CrossRef] [PubMed]
4. Shpichka, A.; Butnaru, D.; Bezrukov, E.A.; Sukhanov, R.B.; Atala, A.; Burdukovskii, V.; Zhang, Y.; Timashev, P. Skin tissue regeneration for burn injury. *Stem Cell Res. Ther.* **2019**, *10*, 94. [CrossRef]
5. Kolesnik, M.; Fexa, A. Multi-dimensional color histograms for segmentation of wounds in images. In Proceedings of the Image Analysis and Recognition: Second International Conference, ICIAR, Toronto, ON, Canada, 28–30 September 2005; Springer: Berlin/Heidelberg, Germany, 2005; pp. 1014–1022.
6. Acha, B.; Serrano, C.; Fondón, I.; Gómez-Cía, T. Burn depth analysis using multidimensional scaling applied to psychophysical experiment data. *IEEE Trans. Med. Imaging* **2013**, *32*, 1111–1120. [CrossRef]
7. Wantanajittikul, K.; Auephanwiriyakul, S.; Theera-Umpon, N.; Koanantakool, T. Automatic segmentation and degree identification in burn color images. In Proceedings of the Biomedical Engineering International Conference, Macau, China, 28–30 May 2012; IEEE: Piscataway, NJ, USA, 2012; pp. 169–173.
8. Rangel-Olvera, B.; Rosas-Romero, R. Detection and classification of skin burns on color images using multi-resolution clustering and the classification of reduced feature subsets. *Multimed. Tools Appl.* **2024**, *83*, 54925–54949. [CrossRef]
9. Kuan, P.; Chua, S.; Safawi, E.; Wang, H.; Tiong, W. A comparative study of the classification of skin burn depth in human. *J. Telecommun. Electron. Comput. Eng. (JTEC)* **2017**, *9*, 15–23.
10. Rangaiah, P.K.B.; Pradeep kumar, B.P.; Augustine, R. Improving burn diagnosis in medical image retrieval from grafting burn samples using B-coefficients and the clahe algorithm. *Biomed. Signal Process. Control* **2025**, *99*, 106814. [CrossRef]
11. Chauhan, J.; Goyal, P. BPBSAM: Body part-specific burn severity assessment model. *Burns* **2020**, *46*, 1407–1423. [CrossRef] [PubMed]
12. Abubakar, A.; Ugail, H. Discrimination of human skin burns using machine learning. In Proceedings of the Intelligent Computing: Computing Conference, Chongqing, China, 6–8 December 2019; Springer: Berlin/Heidelberg, Germany, 2019; pp. 641–647.
13. Şevik, U.; Karakullukçu, E.; Berber, T.; Akbaş, Y.; Türkyılmaz, S. Automatic classification of skin burn colour images using texture-based feature extraction. *IET Image Process.* **2019**, *13*, 2018–2028. [CrossRef]
14. Yadav, D.; Aljrees, T.; Kumar, D.; Kumar, A.; Singh, K.U.; Singh, T. Spatial attention-based residual network for human burn identification and classification. *Sci. Rep.* **2023**, *13*, 12516. [CrossRef]
15. Khan, F.A.; Butt, A.U.R.; Asif, M.; Ahmad, W.; Nawaz, M.; Jamjoom, M.; Alabdulkreem, E. Computer-aided diagnosis for burnt skin images using deep convolutional neural network. *Multimed. Tools Appl.* **2020**, *79*, 34545–34568. [CrossRef]
16. Jiao, C.; Su, K.; Xie, W.; Ye, Z. Burn image segmentation based on Mask Regions with Convolutional Neural Network deep learning framework: More accurate and more convenient. *Burn. Trauma* **2019**, *7*, 6. [CrossRef] [PubMed]

17. Boissin, C.; Laflamme, L.; Fransén, J.; Lundin, M.; Huss, F.; Wallis, L.; Allorto, N.; Lundin, J. Development and evaluation of deep learning algorithms for assessment of acute burns and the need for surgery. *Sci. Rep.* **2023**, *13*, 1794. [CrossRef] [PubMed]
18. Karthik, J.; Nath, G.S.; Veena, A. Deep learning-based approach for skin burn detection with multi-level classification. In *Advances in Computing and Network Communications: Proceedings of CoCoNet 2020*; Springer: Berlin/Heidelberg, Germany, 2021; Volume 2, pp. 31–40.
19. Liu, H.; Yue, K.; Cheng, S.; Li, W.; Fu, Z. A framework for automatic burn image segmentation and burn depth diagnosis using deep learning. *Comput. Math. Methods Med.* **2021**, *2021*, 5514224. [CrossRef]
20. Sun, K.; Zhao, Y.; Jiang, B.; Cheng, T.; Xiao, B.; Liu, D.; Mu, Y.; Wang, X.; Liu, W.; Wang, J. High-resolution representations for labeling pixels and regions. *arXiv* **2019**, arXiv:1904.04514.
21. Xu, X.; Bu, Q.; Xie, J.; Li, H.; Xu, F.; Li, J. On-site burn severity assessment using smartphone-captured color burn wound images. *Comput. Biol. Med.* **2024**, *182*, 109171. [CrossRef]
22. Rozo, A.; Miskovic, V.; Rose, T.; Keersebilck, E.; Iorio, C.; Varon, C. A deep learning image-to-image translation approach for a more accessible estimator of the healing time of burns. *IEEE Trans. Biomed. Eng.* **2023**, *70*, 2886–2894. [CrossRef]
23. Florea, C.; Florea, L.; Vertan, C.; Badoiu, S.C. BurnSafe : Automatic Assistive Tool for Burn Severity Assessment by Semantic Segmentation. In Proceedings of the 12th International Workshop on Assistive Computer Vision and Robotics, European Conference on Computer Vision, Marco, China, 29 September–4 October 2024; Springer: Berlin/Heidelberg, Germany, 2024.
24. Zou, Y.; Zhang, Z.; Zhang, H.; Li, C.L.; Bian, X.; Huang, J.B.; Pfister, T. PseudoSeg: Designing Pseudo Labels for Semantic Segmentation. In Proceedings of the International Conference on Learning Representations—ICLR, Addis Ababa, Ethiopia, 26–30 April 2020.
25. Zhang, Q.; Zhu, Y.; Cordeiro, F.R.; Chen, Q. PSSCL: A progressive sample selection framework with contrastive loss designed for noisy labels. *Pattern Recognit.* **2025**, *161*, 111284. [CrossRef]
26. Badea, M.S.; Vertan, C.; Florea, C.; Florea, L.; Bădoiu, S. Severe burns assessment by joint color-thermal imagery and ensemble methods. In Proceedings of the 2016 IEEE 18th International Conference on e-Health Networking, Applications and Services (Healthcom), Munich, Germany, 14–17 September 2016; IEEE: Piscataway, NJ, USA, 2016; pp. 1–5.
27. Topiwala, A.; Al-Zogbi, L.; Fleiter, T.; Krieger, A. Adaptation and Evaluation of Deep Leaning Techniques for Skin Segmentation on Novel Abdominal Dataset. In Proceedings of the 2019 IEEE 19th International Conference on Bioinformatics and Bioengineering (BIBE), Athens, Greece, 28–30 October 2019; IEEE: Piscataway, NJ, USA, 2019; pp. 752–759.
28. Kawulok, M.; Kawulok, J.; Nalepa, J.; Smolka, B. Self-adaptive algorithm for segmenting skin regions. *EURASIP J. Adv. Signal Process.* **2014**, *2014*, 170. [CrossRef]
29. Rahman, M.M.; Marculescu, R. G-CASCADE: Efficient cascaded graph convolutional decoding for 2D medical image segmentation. In Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision, Waikoloa, HI, USA, 3–8 January 2024; pp. 7728–7737.
30. Ronneberger, O.; Fischer, P.; Brox, T. U-net: Convolutional networks for biomedical image segmentation. In Proceedings of the International Conference on Medical Image Computing and Computer-Assisted Intervention—MICCAI, Munich, Germany, 5–9 October 2015; Springer: Berlin/Heidelberg, Germany, 2015; pp. 234–241.
31. 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. *arXiv* **2021**, arXiv:2102.04306.
32. Wang, W.; Xie, E.; Li, X.; Fan, D.P.; Song, K.; Liang, D.; Lu, T.; Luo, P.; Shao, L. Pyramid vision transformer: A versatile backbone for dense prediction without convolutions. In Proceedings of the International Conference Computer Vision—ICCV, Montreal, QC, Canada, 10–17 October 2021; pp. 568–578.
33. 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. In Proceedings of the International Conference Learning Representations—ICLR, Vienna, Austria, 3–7 May 2021.
34. Chen, L.; Zhang, H.; Xiao, J.; Nie, L.; Shao, J.; Liu, W.; Chua, T.S. Sca-cnn: Spatial and channel-wise attention in convolutional networks for image captioning. In Proceedings of the IEEE Conference Computer Vision and Pattern Recognition, Honolulu, HI, USA, 21–26 July 2017; pp. 5659–5667.
35. Han, K.; Wang, Y.; Guo, J.; Tang, Y.; Wu, E. Vision gnn: An image is worth graph of nodes. *Adv. Neural Inf. Process. Syst.* **2022**, *35*, 8291–8303.
36. Florea, C.; Badea, M.; Florea, L.; Racoviteanu, A.; Vertan, C. Margin-mix: Semi-supervised learning for face expression recognition. In *Proceedings of the Computer Vision—ECCV 2020: 16th European Conference, Glasgow, UK, 23–28 August 2020, Proceedings, Part XXIII 16*; Springer: Berlin/Heidelberg, Germany, 2020; pp. 1–17.

37. Huang, H.; Lin, L.; Tong, R.; Hu, H.; Zhang, Q.; Iwamoto, Y.; Han, X.; Chen, Y.W.; Wu, J. Unet 3+: A full-scale connected unet for medical image segmentation. In Proceedings of the IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), Barcelona, Spain, 4–8 May 2020; IEEE: Piscataway, NJ, USA, pp. 1055–1059.
38. Chen, L.C.; Papandreou, G.; Kokkinos, I.; Murphy, K.; Yuille, A.L. Deeplab: Semantic image segmentation with deep convolutional nets, atrous convolution, and fully connected crfs. *IEEE Trans. Pattern Anal. Mach. Intell.* **2017**, *40*, 834–848. [CrossRef]

Disclaimer/Publisher's Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.