SARNet: A Self-Attention Embedded Residual Network for Multiclass Classification of Chronic Wounds

Chen Qian*, Boyin Yang† and Jiyun Li‡
School of Computer Science and Technology, Donghua University
Shanghai, China
Email: *chen.qian@dhu.edu.cn, †2212620@mail.dhu.edu.cn, ‡jyli@dhu.edu.cn

***Abstract*—Nowadays, chronic wounds have become an increasingly heavy healthcare burden. Therefore, wound classification is the most crucial task in wound diagnosis, which directly affects whether the treatment plan is optimal. This paper proposes a self-attention embedded residual network, or SARNet for short, which takes wound images as input and categorizes them into six types, i.e., burn wounds, surgical wounds, venous lower limb ulcers, pressure ulcers, diabetic foot ulcers, and normal skin. The classification accuracy of SARNet satisfactorily exceeds 80% mainly because its residual structure enhances the feature representation, and its built-in self-attention mechanism enables the global reference.**

***Keywords*—chronic wound classification; residual network; polarized self-attention; multiclass classification.**

I. Introduction

A chronic wound is defined as a breach in skin continuity that fails to achieve an anatomically and functionally intact state through an orderly and timely sequence of repair processes [1]. Traditionally, a patient's wounds are manually analyzed, identified, and documented by the clinicians. However, the number of patients with chronic wounds is tremendous. Thus, performing large-scale wound classification by human means results in a massive medicare burden. Fortunately, with the rapid development of artificial intelligence (AI), we can now use computer vision technologies to classify wounds from only images.

Machine learning plays a vital role in computer vision with the aim of extracting critical information from images [3]. In the healthcare realm, it is frequently used to improve the quality and recognize the crucial features of medical imaging. However, conventional machine learning has certain limitations. For example, it requires that human experts remove unnecessary features before training, which is a time-consuming and laborious mission unsuitable for large-scale projects. Therefore, as an alternative, deep learning has received increasing attention [4]. By identifying and learning meaningful features from the totality of features by itself, deep learning can solve more complex problems with little human intervention.

Currently, most deep learning algorithms for chronic wound assessment are towards a binary classification, i.e., normal skin or physically harmed [5]. However, the sad fact is that such algorithms are useless in a real-world clinical diagnosis. Another sad fact is that the rest can barely provide a satisfactory classification in a multiclass fashion [6]. Thereby, we put forward a novel deep neural network that shows outstanding performance in the multiclass classification of chronic wounds. Based on the research achievement, we have successfully commercialized the proposed AI model and helped hundreds of patients to date.

DOI reference number: 10.18293/SEKE2023-069

II. Related Work

In this paper, we only consider the most typical categories of chronic wounds [7]:

- Burn wounds (BW): Nearly 11 million people worldwide annually are severely burned, which requires long-time medical treatment.
- Surgical wounds (SW): Annually, roughly 4.5% of people worldwide undergo surgery that inflicts a wound.
- Venous lower limb ulcers (VLU): About 0.15% to 0.3% of people worldwide have a VLU.
- Pressure ulcers (PU): Each year, nearly 2.5 million people are suffered from PU.
- Diabetic foot ulcers (DFU): Approximately 34% of diabetic patients are threatened by DFU during their lifetimes, whilst more than 50% of patients with DFU become infected.

Most existing algorithms of wound classification aim to differentiate a wound from normal skin, among which the DFU diagnosis constitutes the majority. Because severe DFU usually leads to limb amputation, identifying DFU is in great demand so that treatment can come in time. Veredas et al. [8] proposed a hybrid system for automatic region segmentation and tissue recognition in an uncontrolled environment, which can detect the wound using color and texture features extracted by a multilayer neural network. Wannous et al. [9] implement color and texture region descriptors to perform a 3D-wound assessment. Wang et al. [10] used a cascaded two-level classifier to determine the boundaries of DFU. As image-based machine learning becomes even more sophisticated, more and more end-to-end models are adopted for better wound diagnosis.

However, the binary chronic wound classification is still considered ineffective in the real world. Hence, the multiclass

classification of chronic wounds has received increasing attention over the past several years. For example, Abubakar et al. [11] proposed a machine learning approach to distinguish BW or PU from normal skin, in which the image features are extracted by deep architecture, e.g., VGG-face, ResNet-101 or ResNet-152, and fed to an SVM classifier. Rostami et al. [6] put forward an integrated end-to-end DCNN classifier to divide the wound into multiple categories, including SW, DFU, and VLU. A total of 538 images of the natural wound are used in the experiment, which results in mean classification accuracy values of 94.28% for binary classification and 87.7% for ternary classification. Sarp et al. [12] performed a quaternary wound classification using a classifier model generated through interpretable artificial intelligence (XIA) and transfer learning, leading to an F1 mean score of 0.76.

Unfortunately, existing wound classification methods are generally unreliable. For example, SVM is frequently used to extract wound features [13]. Despite the experimental results that show accuracy improvement, they are barely convincing due to the small size of the evaluation set. Moreover, such experiments usually require specific lighting conditions (shading), markers, and skin colors. Otherwise, the model performs inadequately. Some other models show a stark contrast between binary and multi-class classifications, e.g., the accuracy drops rapidly from 97% to 72% [6]. For all these reasons, in this paper, we present SARNet, a self-attention embedded residual network using multi-branch topology, to classify the most common six types of chronic wounds precisely.

III. Our Work

Currently, most research on chronic wound classification refers to a single type of wound, whereas the rest achieves low accuracy involving multiple types. We hereby present a self-attention embedded residual network, or SARNet for short, which is structured using a multi-branch topology. Notably, the branches adopt different convolution kernels, respectively, in order to obtain different receptive fields, thus thoroughly extracting features at various depths.

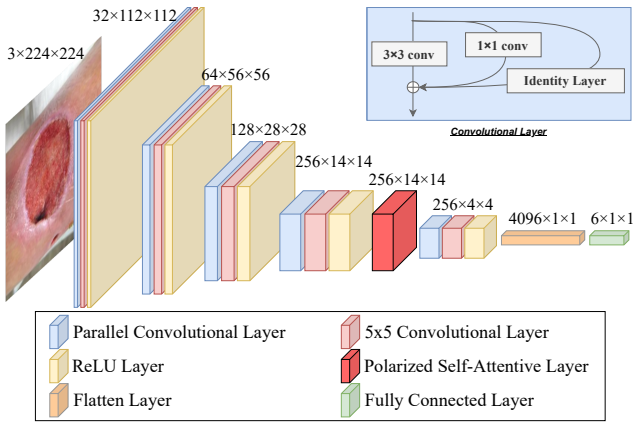


Fig. 1. An overview of SARNet.

As shown in Figure 1, the input of SARNet is a three-channel image of a chronic wound with a size of 224×224. Such an image is dealt with several parallel convolutional layers and a self-attention layer. Each parallel convolutional layer consists of a branch using 3×3 convolutional kernel, another one using 1×1 kernel, and the third in the form of an identity layer that keeps the output unchanged before and after the transformation. Downsampling is performed by convolution with stride two at the beginning of the first layer. The self-attention layer is placed between the fourth and fifth layers, which brings the advantage of maintaining excellent performance on fine-grained pixel-level tasks. The final stage is a fully connected layer resulting in a senary classification of wounds, viz. BW, SW, VLU, PU, DFU and normal skin.

A. Multi-branch Convolutional Neural Network

After VGG achieved a top-1 accuracy of over 70% on ImageNet classification [14], many innovations have emerged in making the network more complex to achieve high performance. For example, ResNet proposed a simplified dual-branch architecture that implicitly integrated many shallower models with the aim of training a multi-branch model to avoid the vanishing gradients problem [15]. Although complex neural architectures can generate networks with higher performance, the cost of computing resources or workforce becomes enormous. Moreover, some models are too sophisticated to be trained using ordinary GPUs, let alone the usage in practice. In spite of the inconvenience of implementation, complex models may reduce the parallelism and hence slow down the inference [16].

Our model is constructed on the basis of a simple VGG architecture. Since the wound surface usually appears as circular or irregular shapes, we adapt the multi-branch structure originally proposed in RepVGG [15] to enhance the representation of the network model. Each branch applies a specific receptive field and captures more relevant image features accordingly. Apparently, the residual branches are the key to SARNet architecture, which divides the training process into three paths. Each path contains downsampling and BatchNorm (BN) layers. The role of the BN layer is to normalize the data, which stabilizes the distribution of input data and thus accelerates the overall learning speed of the model.

The formulas used in the convolutional and BN layers are expressed as follows:

$$Conv(x) = \sum_i w_i x_i + b \tag{1}$$

$$BN(x) = \gamma \times \frac{(x - mean)}{\sqrt{var}} + \beta \tag{2}$$

where w is the weight of the convolutional kernel, x is the input, b is the bias, β and γ are the learnable parameters, $mean$ is the mean value and $\sqrt{var}$ is the variance. Equation 3 is obtained by substituting Eq. 1 into Eq. 2, which shows the convolutional layer with bias vectors obtained by fusing the BN with the previous convolutional layer. Using a multi-branch convolutional neural network with a parallel structure

can improve the accuracy of the model during training and avoid the problem of vanishing gradients.

$$BN(Conv(x)) = \frac{\gamma \times W(x)}{\sqrt{var}} + \frac{\gamma \times (b - mean)}{\sqrt{var}} + \beta \quad (3)$$

B. Polarized Self-attentive Mechanism

Since convolution can only use local rather than global information to calculate the target pixel, this may introduce

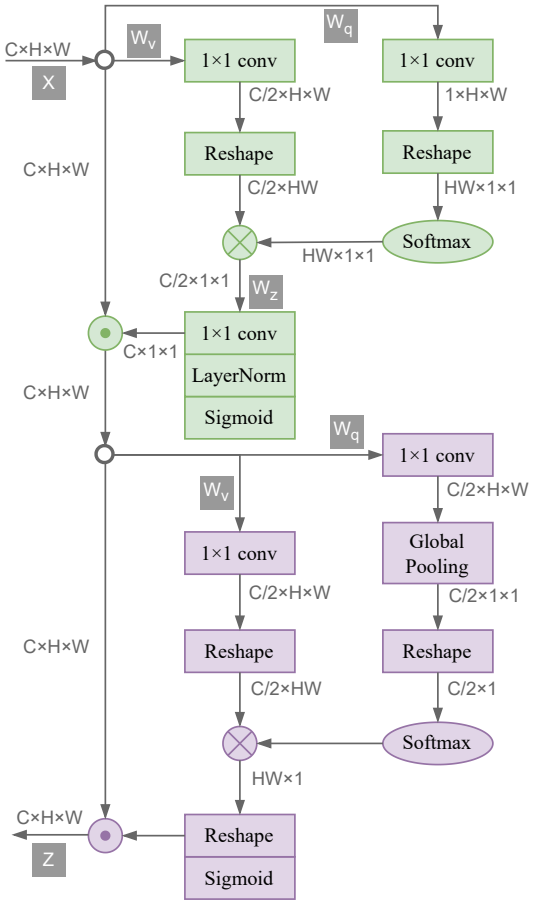


Fig. 2. The structure of PSA module.

some bias. Suppose we treat each pixel in the feature map as a random variable and calculate the pairwise covariance between all pixels, i.e., the similarity between two random variables. In that case, we can enhance or weaken the value of each predicted pixel based on its similarity with other pixels in the image. Since the wound figure may appear at any position in the image (some wound figures may occupy a small proportion of the image with much irrelevant background), adding an attention mechanism between the fourth and fifth convolution layers can achieve global reference in the model training and prediction process, thereby enhancing the model classification performance.

In this paper, we use the *polarized self-attentive* (PSA) mechanism, which is realized as a combination of two branches. One performs a channel-wise self-attention mechanism, while the other is spatial-wise. Eventually, the results of the two branches are fused to generate the output of the structure, as demonstrated in Figure 2. To reduce the information loss caused by dimension reduction, PSA uses the polarized filtering mechanism, which maintains the size of $[H, W]$ in the spatial dimension and uses the size of $C/2$ in the channel dimension. In addition, a non-linear function that regresses the output distribution at a fine granularity is used to enhance the information. In other words, the *Softmax* function is used to increase the attention range on the smallest tensor in the attention module, and then the *Sigmoid* function is used for dynamic mapping. Equations 4 and 5 calculate the weight of the channel branch and spatial branch, respectively.

$$A^{ch}(X) = F_{SG}[W_{z|\theta_1}((\sigma_1(W_v(X)) \times F_{SM}(\theta_2(W_q(X)))))] \quad (4)$$

$$A^{sp}(X) = F_{SG}[\sigma_3(F_{SM}(\sigma_1(F_{GP}(W_q(X)))) \times \theta_2(W_v(X)))] \quad (5)$$

As a result, the fusion of the channel and spatial branches can be calculated as follows:

$$\begin{aligned} PSA_s(x) &= Z^{sp}(Z^{ch}) \\ &= A^{sp}(A^{ch}(X) \odot^{ch} X) \odot^{sp} A^{ch}(X) \odot^{ch} X \end{aligned} \quad (6)$$

IV. Experiment

A. Dataset

The dataset used in this paper is from the Kaggle big data competition platform. It contains 1777 wound images, including six categories: 323 images of BW, 209 images of SW, 447 images of VLU, 373 images of PU, 325 images of DFU, and 100 images of normal skin. These data are divided into a training set and a test set in a 7:3 ratio.

B. Data Preprocessing

All wound images are preprocessed via two steps. The first step is to augment the data by horizontal flipping, rotation, and random cropping, whereas the second is to process the wound images using *mixup* data enhancement [17]. Mixup is a data augmentation principle independent of data and a form of neighborhood risk minimization. It uses modeling between different categories to achieve data augmentation. First, two samples are randomly selected from the training samples for simple random weighted summation, and the labels of the samples are also correspondingly weighted and summed. Then, the predicted result and the label after weighted summation are used to calculate the loss and update the parameters in reverse differentiation. Mixup extends the training distribution by combining prior knowledge that linear interpolating feature vectors should lead to linear interpolation of relevant labels. In addition, the mixup method can reduce the considerable memory loss and sample sensitivity in the network and can reduce the memory of incorrect labels.

C. Experimental Results

To thoroughly investigate the classification performance, we use accuracy, precision, recall and F1-score to evaluate our SARNet.

We achieved an accuracy of 80.87% in the senary classification of the chronic wound using the SARNet. Heretofore, the highest record of senary classification is 75.64%, which is achieved by a VGG16 network using the AZH dataset [5].

So, We improved the accuracy by 5.23%. In addition, we also performed six quinary classifications on this dataset. The accuracy results are illustrated in Table I. It is worth noting that B, S, V, P, D, and N are the abbreviations of BW, SW, VLU, PU, DFU, and normal skin, respectively.

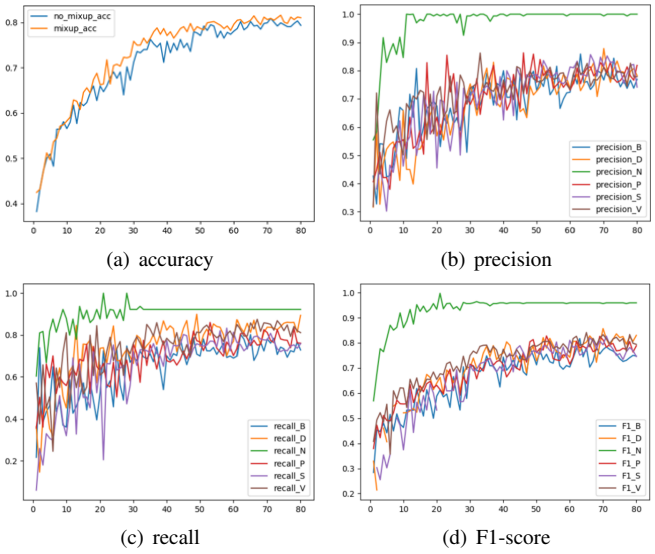
TABLE I
ACCURACY OF MULTI-CLASS CLASSIFICATIONS OF CHRONIC WOUNDS

Num of Classes	Classes	Test Accuracy
5 classes	BDNPS	84.24%
	BDNPV	82.78%
	BDNSV	83.45%
	BDPSV	77.94%
	BNPSV	83.99%
	DNPSV	84.71%
6 classes	BDNPSV	80.87%

In addition, we compared the effects of mixup and attention mechanism on the senary classification, as shown in TableII, which provides ample evidence of their importance.

TABLE II
THE IMPACT OF MIXUP AND ATTENTION MECHANISM ON ACCURACY

Classifier	Test Accuracy
SARNet without mixup and attention	69.34%
SARNet without mixup	77.24%
SARNet without attention	71.92%
SARNet	80.87%



(a) accuracy (b) precision (c) recall (d) F1-score

Fig. 3. The performance of SARNet.

To further demonstrate the effectiveness of this network model for six classifications in the chronic wound image dataset, we calculated the model's accuracy, precision, recall, and F1-score, as shown in Figure 3. It can be seen that the precision, recall, and F1-score are also around 80% for all wound images, except for normal skin.

V. CONCLUSION

A high-performance classifier is urgently needed to classify wounds with less financial and time costs. This paper presents SARNet, which is constructed based on a multi-branch topology that enables various convolutional kernels referring to receptive fields at different levels of granularity for thorough feature extraction. Additionally, SARNet is introduced by a polarized self-attentive mechanism to capture long-distance dependencies for better global reference. The experiment shows SARNet outperforms all other existing approaches.

REFERENCES

[1] V. Falanga, R. R. Isseroff, A. M. Soulika, M. Romanelli, D. Margolis, S. Kapp, M. Granick, and K. Harding, "Chronic wounds," *Nature Reviews Disease Primers*, vol. 8, no. 1, p. 50, 2022.

[2] C. K. Sen, "Human wounds and its burden: an updated compendium of estimates," pp. 39–48, 2019.

[3] A. I. Khan and S. Al-Habsi, "Machine learning in computer vision," *Procedia Computer Science*, vol. 167, pp. 1444–1451, 2020.

[4] A. Esteva, K. Chou, S. Yeung, N. Naik, A. Madani, A. Mottaghi, Y. Liu, E. Topol, J. Dean, and R. Socher, "Deep learning-enabled medical computer vision," *NPJ digital medicine*, vol. 4, no. 1, p. 5, 2021.

[5] D. Anisuzzaman, C. Wang, B. Rostami, S. Gopalakrishnan, J. Niezgoda, and Z. Yu, "Image-based artificial intelligence in wound assessment: A systematic review," *Advances in Wound Care*, vol. 11, no. 12, pp. 687–709, 2022.

[6] B. Rostami, D. Anisuzzaman, C. Wang, S. Gopalakrishnan, J. Niezgoda, and Z. Yu, "Multiclass wound image classification using an ensemble deep cnn-based classifier," *Computers in Biology and Medicine*, vol. 134, p. 104536, 2021.

[7] R. G. Frykberg and J. Banks, "Challenges in the treatment of chronic wounds," *Advances in wound care*, vol. 4, no. 9, pp. 560–582, 2015.

[8] F. Veredas, H. Mesa, and L. Morente, "Binary tissue classification on wound images with neural networks and bayesian classifiers," *IEEE transactions on medical imaging*, vol. 29, no. 2, pp. 410–427, 2009.

[9] H. Wannous, Y. Lucas, and S. Treuillet, "Enhanced assessment of the wound-healing process by accurate multiview tissue classification," *IEEE transactions on Medical Imaging*, vol. 30, no. 2, pp. 315–326, 2010.

[10] L. Wang, P. C. Pedersen, E. Agu, D. M. Strong, and B. Tulu, "Area determination of diabetic foot ulcer images using a cascaded two-stage svm-based classification," *IEEE Transactions on Biomedical Engineering*, vol. 64, no. 9, pp. 2098–2109, 2016.

[11] A. Abubakar, H. Ugail, and A. M. Bukar, "Can machine learning be used to discriminate between burns and pressure ulcer?" in *Intelligent Systems and Applications: Proceedings of the 2019 Intelligent Systems Conference (IntelliSys) Volume 2*. Springer, 2020, pp. 870–880.

[12] S. Sarp, M. Kuzlu, E. Wilson, U. Cali, and O. Guler, "A highly transparent and explainable artificial intelligence tool for chronic wound classification: Xai-cwc," 2021.

[13] M. Goyal, N. D. Reeves, S. Rajbhandari, N. Ahmad, C. Wang, and M. H. Yap, "Recognition of ischaemia and infection in diabetic foot ulcers: Dataset and techniques," *Computers in biology and medicine*, vol. 117, p. 103616, 2020.

[14] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2016, pp. 770–778.

[15] X. Ding, X. Zhang, N. Ma, J. Han, G. Ding, and J. Sun, "RepVGG: Making VGG-style ConvNets great again," in *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, 2021, pp. 13 733–13 742.

[16] N. Ma, X. Zhang, H.-T. Zheng, and J. Sun, "ShuffleNet v2: Practical guidelines for efficient CNN architecture design," in *Proceedings of the European conference on computer vision (ECCV)*, 2018, pp. 116–131.

[17] H. Zhang, M. Cisse, Y. N. Dauphin, and D. Lopez-Paz, "mixup: Beyond empirical risk minimization," *arXiv preprint arXiv:1710.09412*, 2017.