Identificação de feridas em ratos: uma abordagem de aprendizado de máquina
Wound identification in mice: a machine learning approach

Bruno Uhlmann Marcato [1], José Luis Seixas Junior [2], Rafael Gomes Mantovani[3]

RESUMO

A segmentação semântica tem sido explorada com sucesso em estudos biológicos para lidar com diversas aplicações, como a identificação de feridas. Este estudo explora duas abordagens diferentes de segmentação de imagens para identificar feridas em ratos, especificamente os algoritmos U-Net e Random Forest. Este último foi combinado com recursos extraídos das duas primeiras camadas do VGG16, que foi utilizado como extrator de recursos. Os experimentos foram realizados com um conjunto de dados reais desenvolvido pelo Laboratório de Dor, Neuropatia e Inflamação da Universidade Estadual de Londrina com aprovação do Comitê de Ética Universitária em Pesquisa e Bem-Estar Animal. Os resultados experimentais foram promissores, mostrando que ambas as alternativas podem fornecer previsões precisas para a maioria das imagens em relação às medidas de avaliação FSCore e IoU. Também foram aplicados testes estatísticos, mostrando que a U-Net obteve resultados estatisticamente melhores com FScore médio de 0.72 e IoU de 0.58.
PALAVRAS-CHAVE: Identificação de feridas; Random Forest; Segmentação Semântica; U-Net

ABSTRACT

Semantic segmentation has been successfully explored in biological studies to handle various applications, such as identifying wounds. This study explores two different image segmentation approaches to identify mice wounds, specifically the U-Net and Random Forest algorithms. The latter was combined with features extracted from the first two layers of VGG16, which was used as a feature extractor. Experiments were performed with a real dataset developed by the Pain, Neuropathy, and Inflammation Laboratory at the State University of Londrina with the approval of the University Ethics Committee on Animal Research and Welfare. The experimental results were promising, showing that both alternatives can provide accurate predictions for most images regarding FSCore and IoU evaluation measures. Statistical tests were also applied, showing that U-Net obtained statistically better results with an average FScore of 0.72 and IoU of 0.58.
KEYWORDS: Wound Identification; Random Forest; Semantic Segmentation; U-Net.

INTRODUCTION

Image Segmentation is the process of dividing an image into regions (GONZALEZ; WOODS, 2010), often the first step in the image analysis process. It highlights the portion of the image with relevant information, called the Region of Interest (ROI). However, different portions of images can reference the same object (parts), or the same objects can appear in different regions (plurality), thus having different ROI with the same aspect of meaning. For these portions to be understood as the same object or parts of the same region, the semantic segmentation adds meaning to the pixel (LONG; SHELHAMER; DARRELL, 2015), resulting in a pixel-wise classification problem.

[1] Bolsista no Programa de Iniciação Científica (PIBIC), Federal University of Technology – Paraná (UTFPR), Apucarana - PR, Brazil.; E-mail: marcato.2001@alunos.utfpr.edu.br.
[2] ELTE – Eötvös Loránd University, Faculty of Informatics, Budapest, Hungary E-mail: jlseixasjr@inf.elte.hu.
[3] Federal University of Technology – Paraná (UTFPR), Apucarana - PR, Brazil; E-mail: rafaelmantovani@utfpr.edu.br.

This study explores Semantic Segmentation to solve a mice wound pixel classification problem. Experiments were carried out with two different approaches: i) a Deep Learning (DL) architecture - U-Net, and ii) a traditional Machine Learning (ML) algorithm - Random Forest (RF) using latent representations of a pre-trained Convolutional Neural Network (Visual Geometry Group - VGG (SIMONYAN; ZISSERMAN, 2015)). The image dataset used in the experiments is composed of wounds in mice, generated by the Laboratory of Research in Pain, Neuropathy and Inflammation of the State University of Londrina (UEL).

BACKGROUND

U-NET

The U-Net algorithm is a Convolutional Neural Network (CNN) architecture developed for image segmentation, which has been shown to be efficient in several applications in the medical context (PUNN; AGARWAL, 2022). The network design consists of a symmetric network with descending *pooling* layers to encode low-level information into a high-dimensional representation and ascending convolution (or transposed convolution) and up-sampling layers to reconstruct the segmented image. The technique of concatenating information from the decoding layer with the corresponding encoding layer allows the combination of low and high-level information, which provides a more accurate representation of the original image. U-Net also uses ReLU-like activation functions for all convolutional layers and a sigmoid activation layer at the end to produce the binary output of the segmentation.

RANDOM FOREST

Random Forest (RF) is a traditional ML algorithm that uses multiple Decision Trees (DTs) to generate predictions. Each tree is trained on a random sample of the training data and a random subset of the attributes, creating different data views. The final prediction is made by polling the predictions of all the trees, making the model more robust and less susceptible to overfitting. RF is widely used in classification and regression applications such as image recognition, sentiment analysis, and fraud detection (BIAU; SCORNET, 2016). Semantic image segmentation involves identifying and labeling each pixel in an image based on its semantic content, such as identifying each object or region present in the image. This is important in many areas, such as computer vision, robotics, and medical image processing. RF can be trained to classify each image pixel into different classes, such as object, background, or edge, using texture, color, and shape information. The result is precise image segmentation that can be used in different applications.

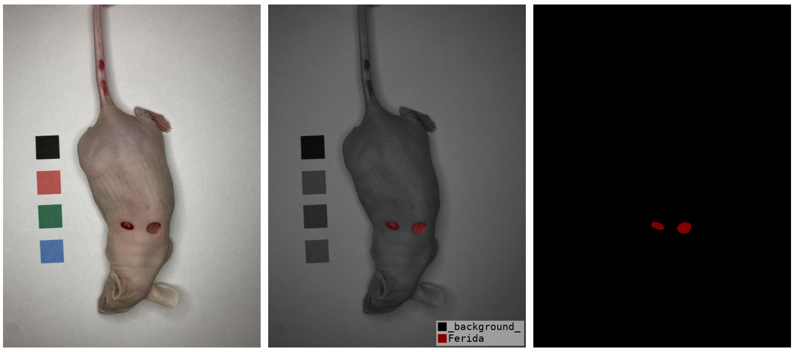
EXPERIMENTAL METHODOLOGY

In this section, we detail the experimental methodology adopted in this article. We present descriptions regarding the image dataset, ML algorithms, and the experimental setup used to generate results and analyze the obtained predictions. The following subsections give additional details.

IMAGE DATASET

The dataset for this article consists of 71 images of mice with wounds on their backs. This is a real dataset developed by the Pain, Neuropathy, and Inflammation Laboratory at the State University of Londrina (UEL) with the approval of the UEL Ethics Committee on Animal Research and Welfare (process number 15654.2019.33). Among all the images, 36 are of mice on day 0 of wound progression, and 35 images are of mice on day 7 of wound progression. Images from day 0 are 1024×768 in size, and images from day 7 are 4032×3024. Assessing the wound progression is outside the scope of this work, as this step requires more information than just detecting the wound region. Hence, this article focuses on pixelwise classification (semantic segmentation) between wounds and non-wounds. Figure 1 shows an example of an image and its corresponding labels and predictions for an animal with a wound progression on day 0. The classes (pixel labels) were defined manually using the `Labelme` tool[1].

Figura 1 – Example of a dataset image instance and its correspondent labels



Fonte: The authors.

DATA AUGMENTATION

Due to the low number of images in the dataset (71), a Data Augmentation (DA) process was required. DA is a technique widely used in Computer Vision to increase the amount and variety of data (YANG et al., 2022). The technique involves applying transformations to the original data and generating new instances. The main advantages of using DA are: (i) models' precision improvement since it generates more significant variability in the data; (ii) cost reduction in data collecting and labeling, given that collecting medical images is challenging and complex work, as well as labeling them, this task is reduced as transformations are applied to the original images and masks simultaneously; and (iii) overfitting prevention, since the model tends to learn specific patterns present in the available training data when there is low variability in the data, losing its generalization power to new data.

A total of five transformations were considered in DA: horizontal flip, vertical flip, normalization, scaling, and resize. With the original images, the dataset with DA has a total of 423 images available

[1] `https://github.com/wkentaro/labelme`

for model induction.

ALGORITHMS

U-Net is one of the most widely used architectures for Semantic Segmentation of medical images (PUNN; AGARWAL, 2022). U-Net was trained with Adam optimizer and a learning rate of $\alpha = 0.0001$. The loss function optimized was the binary cross entropy since the network was used to perform binary classification. This network uses ReLu activation functions between convolutional layers and the sigmoid activation function as a classifier in its last layer. U-Net was trained for 100 epochs and using batch size 2.

In RF algorithm, the default value for the number of trees is between $[100, 500]$ depending on the coding language and correspondent libraries/packages. However, all of them follow the same idea: the higher number of trees, the better the generalization power of the induced model. In our experiments, the RF implementation was provided by the scikit-learn Python library, which uses $t = 100$ trees in the ensemble and '*Gini*' index as the attribute evaluation criterion as default values.

Some initial experiments were done using pixel RGB values for RF algorithm, but the obtained predictions did not identify any pixel as a wound. Thus, an alternative was explored using 64 feature maps extracted from the first two convolutional layers of the VGG16[2] neural network to feed the algorithm, using the weights pre-trained of the `ImageNet` dataset. VGG16 is well known for accurately performing computer vision tasks, including image recognition (OLAITAN et al., 2022). By extracting these feature maps, it is possible to obtain general information about the images that can be used to train machine learning models, such as the RF, with a lower computational cost.

EXPERIMENTAL SETUP

The resampling strategy adopted in experiments was a simple holdout with five repetitions using different seeds. The augmented dataset was split into training, testing, and validation sets, with proportions of 60%, 30%, and 10%, respectively. It is worth mentioning that the validation data was extracted as a part of the training set and not used in the testing set. The results obtained were evaluated by F-Score and the Dice coefficient , an evaluation metric which combines precision and recall to provide an overall measure of the model's accuracy. All the code was written in Python, U-Net was coded using `PyTorch` while RF+VGG used `scikit-learn` and `Keras` libraries. Experiments were executed on a desktop with a Ryzen 5 5600g processor, with 16GB of RAM, and motherboard b550m.

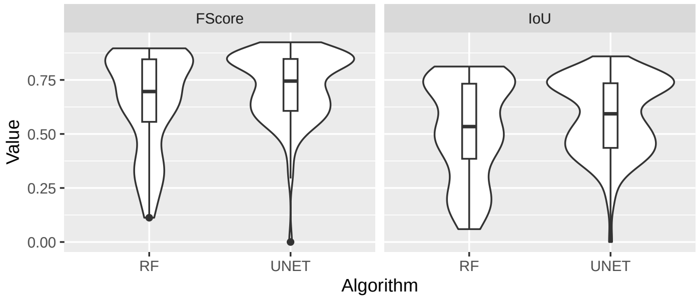
RESULTS

Figure 2 illustrates each algorithm's FScore and Intersection over Union (IoU) values and explicitly that both algorithms present similar result distributions. The U-Net median values for FScore (0.75) and IoU (0.60) were slightly higher than RF ones $(0.70$ and $0.53)$. The figure also shows that most images have FScore values above 0.6. This indicates that both algorithms could identify lesion

[2] Visual Geometry Group, 16 layers.

regions and estimate the shape and wound filling.

Figura 2 – Overall results with Fscore values obtained on test images



Fonte: The authors.

However, due to the nature of the distributions represented by the violins, some images presented difficulty for the algorithms. U-Net could not perform generalization for some images from day 7 with tiny wounds, where the healing process was almost or practically finished. These images have a highly unbalanced pixel distribution, as few pixels compose lesion regions. It may affect the FScore values but not the IoU. IoU only measures hits and misses on the object but not on the background, which is why it is usually used as an image segmentation evaluation metric. RF, on the other hand, has more extended distributions for both metrics. Violins indicate a considerable number of images with lower values of IoU and FScore, more than U-Net. A superficial analysis could identify that this occurred in some images when it fails to fill in the wound region or even incorrectly classifies other parts of the animal as a lesion, such as regions of the paws or tail.

The non-parametric Wilcoxon paired test with α = 0.05 (95% of significance) was applied to assess the statistical significance of these results. The null hypothesis states no difference in performance between the algorithms. The test was applied for both metrics (FScore and IoU), and obtained a p-value = 0.00101, which suggests a difference in terms of performance in favor of the U-Net. Although U-Net is statistically better than RF regarding predictions, the computational cost for training both algorithms is quite different. RF took 20 minutes in total to execute the five repetitions, while U-Net needed 9 hours. That is, the 64 feature maps extracted from the VGG16 and fed into the RF can be a good alternative to speed the process of wound recognition than using a deep architecture, such as U-Net.

CONCLUSIONS

This study investigated U-Net and RF with VGG latent features to solve a semantic segmentation problem of mice wounds. Experiments were carried out with an image dataset generated by the Pain, Neuropathy, and Inflammation Laboratory at the State University of Londrina, composed of 71 images showing wounds in mice. The results were promising: U-Net obtained F-Score and IoU average values of 0.75 and 0.60, respectively. Most images generated accurate predictions of the wound regions, with some exceptions in images where the wounds were tiny (day 7). The RF+VGG alternative was faster regarding the computational cost but statistically lower in performance, with lower average F-Score and IoU (0.70 and 0.53). However, it can be improved with some post-processing techniques to fill regions not wholly identified. Other alternatives for future works include evaluating more and different DL architectures, ML algorithms and automating the entire pipeline.

Acknowledgment

The authors would like to thank the Federal University of Technology – Paraná (UTFPR), campus of Apucarana, and the Tempus Public Foundation through the Stipendium Hungaricum Scholarship Programme, for the financial support.

Code availability

The code repository for the steps of this research is publicly available[3].

Conflict of interest

There is no conflict of interest.

REFERENCES

BIAU, Gérard; SCORNET, Erwan. A random forest guided tour. **TEST**, v. 25, p. 197–227, 2016. DOI: 10.1007/s11749-016-0481-7.

GONZALEZ, R.C.; WOODS, R.E. **Processamento Digital De Imagens**. São Paulo: Pearson, 2010. ISBN 9788576054016.

LONG, Jonathan; SHELHAMER, Evan; DARRELL, Trevor. Fully convolutional networks for semantic segmentation. In: CONFERENCE on Computer Vision and Pattern Recognition (CVPR). Boston, MA, USA: IEEE, jun. 2015. P. 3431–3440. DOI: 10.1109/CVPR.2015.7298965.

OLAITAN, Alashiri et al. Face Recognition Using VGG16 CNN Architecture for Enhanced Security Surveillance—A Survey. In: SINGH, Pradeep Kumar et al. (Ed.). **Futuristic Trends in Networks and Computing Technologies**. Singapore: Springer Nature Singapore, 2022. P. 1111–1125. ISBN 978-981-19-5037-7.

PUNN, Narinder Singh; AGARWAL, Sonali. Modality specific U-Net variants for biomedical image segmentation: a survey. **Artificial Intelligence Review**, v. 55, p. 5845–5889, 7 2022. DOI: 10.1007/s10462-022-10152-1.

SIMONYAN, Karen; ZISSERMAN, Andrew. Very Deep Convolutional Networks for Large-Scale Image Recognition. In: INTERNATIONAL Conference on Learning Representations. San Diego, CA, USA: CoRR, 2015.

YANG, Suorong et al. **Image Data Augmentation for Deep Learning: A Survey**. [S.l.: s.n.], 2022. arXiv: 2204.08610 [cs.CV].

[3] https://github.com/BrunoMarcato/MiceWoundSegmentation