Faster R-CNN Approach for Diabetic Foot Ulcer Detection

Artur Leandro da Costa Oliveira[a], André Britto de Carvalho[b] and Daniel Oliveira Dantas[c]
Departamento de Computação, Universidade Federal de Sergipe, São Cristóvão, SE, Brazil

Keywords: Deep Learning, Convolutional Neural Networks.

Abstract: Diabetic Foot Ulcer (DFU) is one of the major health concerns about Diabetes. These injuries impair the patient's quality of life, bring high costs to public health, and can even lead to limb amputations. The use of automatic tools for detection can assists specialists in the prevention and treatment of the disease. Some methods to address this problem based on machine learning have recently been presented. This article proposes the use of deep learning techniques to assist the treatment of DFUs, more specifically, the detection of ulcers through photos taken from the patient's feet. We propose an improvement of the original Faster R-CNN using data augmentation techniques and changes in parameter settings. We used a training dataset with 2000 images of DFUs annotated by specialists. The training was validated using the Monte Carlo cross-validation technique. Our proposal achieved a mean average precision of 91.4%, a F1-score of 94.8%, and an average detection speed of 332ms which outperformed traditional detector implementations.

1 INTRODUCTION

Diabetes is a serious complication with a high long-term impact on the population. The incidence of diabetes has grown globally in the last decades causing high health costs. It is among the top 10 causes of death in adults (Saeedi et al., 2019). Diabetic Foot Ulcer (DFU) is one of the major complications of Diabetes. The patients have a probability of 12-25% of developing DFU during their lifetime. This rate can reach 19-34% depending on the data used (Armstrong et al., 2017). Such ulcers have become a major problem in public health because of the increase in morbidities, decreased quality of life, and because the treatment is expensive. Due to inadequate conduct in the treatment of foot ulcers, there is a delay in the improvement of the injury and the possibility of lower limb amputation (Leung, 2007).

In the early stages of the DFU, it is important to quickly detect and to keep track of the disease. To make a diagnostic, specialists take into account different evaluation criteria, such as the medical history of the patient, examination of the diabetic foot, and additional tests like CT scans, MRI, and X-Ray (Goyal et al., 2018a). The use of computer vision techniques can lead to an improvement in the diagnosis of the disease and in the agility of the entire clinical process. Image processing is used in the medical field in several types of systems and has been successful in different medical applications. These systems are used in treatment planning, surgery, and biological images. Databases can have two, three, or more dimensions. These dimensions carry a vast amount of information that can be used in the clinical area or application research (Bankman, 2008). Initially, low-level pixel processing methods (edge detection, line detection filters, and region growth) and mathematical models were used to solve specific problems in the medical field. In the late 90s, supervised learning techniques, where training data is used to develop a system, started to become popular. A crucial step in the design of such systems is the extraction of discriminant features from the images. Later, the use of deep learning techniques arises, allowing computers to learn the features that optimally represent the data of the problem at hand (Litjens et al., 2017).

Generally, from a computer vision and medical image perspective, three different tasks are performed to detect anomalies in medical images: classification, localization, and segmentation (Goyal et al., 2018b). Classification is to recognize the type of the anomaly. Localization is to point out the region of the anomaly. Segmentation is to define precise limits of the anomaly. To solve these tasks, convolutional neural network (CNN) based object detectors have been used, such as the faster region-based

[a] https://orcid.org/0000-0002-0165-6699
[b] https://orcid.org/0000-0002-6498-9706
[c] https://orcid.org/0000-0002-0142-891X

Oliveira, A., Britto de Carvalho, A. and Dantas, D.
Faster R-CNN Approach for Diabetic Foot Ulcer Detection.
DOI: 10.5220/0010255506770684
In *Proceedings of the 16th International Joint Conference on Computer Vision, Imaging and Computer Graphics Theory and Applications (VISIGRAPP 2021) - Volume 4: VISAPP*, pages 677-684
ISBN: 978-989-758-488-6

Figure 1: Samples of the DFU Dataset.

convolutional network (Faster R-CNN) (Ren et al., 2015), region-based fully convolutional networks (R-FCN) (Dai et al., 2016), single shot multibox detector (SSD) (Liu et al., 2016), and you only look once (YOLO) (Redmon et al., 2016). These methods are accurate enough to be used in consumer applications (Huang et al., 2017) and are usually evaluated by their mean average precision (mAP), but other metrics can also be used, such as specificity, sensitivity, execution time, and memory usage.

Studies that assist the treatment of DFU using computational methods are rarely found in the literature. Liu et al. (Liu et al., 2015) and Saminathan et al. (Saminathan et al., 2020) proposed automatic methods that use the temperature characteristic in infrared images to perform DFU detection. Their papers show good results in detecting DFU in images of feet. However, they present difficulties in terms of different symmetries and positioning of feet in images. Deformed feet and/or amputated limbs can also impair detection by these methods. Goyal et al. (Goyal et al., 2017) proposes to segment DFU lesions using Fully Convolutional Networks (FCNs). Its results demonstrate a high accuracy which can help in the detection and treatment of the disease. Goyal proposes a convolutional neural network architecture called DFUNet to improve the classification of DFU images (Goyal et al., 2018a). Its good performance in classifying parts of skin with DFU allows it to be used also for classifying other skin diseases. The DFUNet obtains a better performance compared to GoogLeNet. GoogleNet is a convolutional neural network also known as Inception and was responsible for achieving the state-of-the-art in detection and classification in the ImageNet Large-Scale Visual Recognition Challenge 2014 (ILSVRC14) (Szegedy et al., 2015).

Goyal et al. (Goyal et al., 2018b) proposes a real-time detection tool of DFUs for mobile devices. The usage of such a tool on a mobile device assists specialists in quickly detecting and diagnosing the disease. The major challenge of automatic methods for DFU is to optimize metrics such as specificity, sensitivity, execution time, and memory usage. An improvement in these metrics allows greater reliability in the use of this type of application for the treatment of DFUs. In contrast to traditional machine learning, deep learning methods have demonstrated superiority in object localization and segmentation of DFUs, which suggests that the robust fully automated detection of DFUs may be viable (Goyal et al., 2018b).

This work proposes a tool for detecting foot ulcers in individuals with diabetes based on the Faster R-CNN object detection (Ren et al., 2015). The purpose of this work is to help the prevention and treatment of the disease. The main task of the tool is to locate areas of interest in the image and classify them as ulcers or not. This work was motivated by the Diabetic Foot Ulcers Grand Challenge 2020 (DFUC 2020) (Cassidy et al., 2020), challenge that aims to improve the accuracy of DFU detection in real environments. Our main contribution is the improvement of the Faster R-CNN for DFU detection. In our experiments, we achieved better mAP, F1-score, and detection speed in comparison to the state-of-the-art detectors. Our strategy reduced the number of false positives, which lead to an improvement in precision.

The rest of the work is organized as follows: Section 2 summarizes basics concepts needed to understand the work. Section 3 describes the methodology used to create the tool. Section 4 presents the experiments and the results, and Section 5 presents the final considerations.

2 BACKGROUND

Created to approach the problem of object detection by region proposal, the Faster R-CNN (Ren et al., 2015) is an evolution of the Fast R-CNN (Girshick, 2015). Unlike its predecessor, the Faster R-CNN consists of two modules. The first module is the region proposal network (RPN), a deep convolutional neural network. The second module is the Fast R-CNN detector. Both the RPN and the object classifier share convolutional layers. The region proposal network is intended to guide the detection, determining the best regions among different scales and proportions. Basically, the RPN tells the classification module where to look. The classification module, composed of a deep

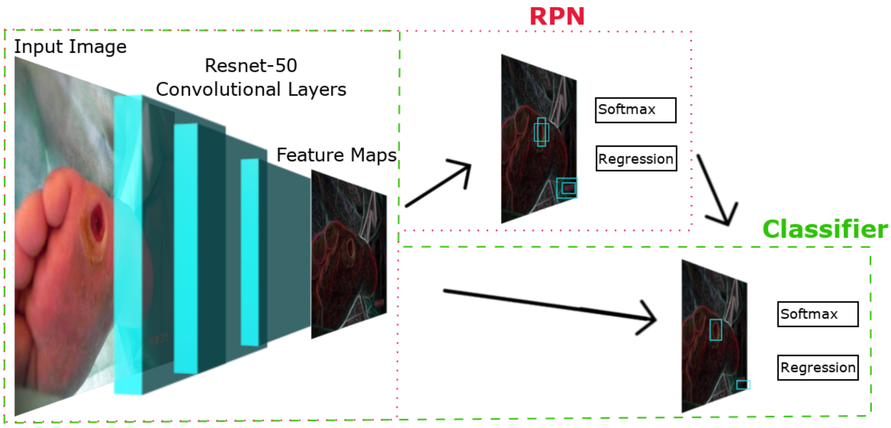


Figure 2: Faster R-CNN DFU Architecture.

convolutional network, receives different regions of the image and classifies them.

The motivation for using the Faster R-CNN is due to its high accuracy in object detection. This network achieved the best accuracy in PASCAL VOC (Everingham et al., 2007; Everingham et al., 2015) both in 2007 and in 2012 and was the basis for the winners of Imagenet detection and localization in ILSVRC 2015 and COCO detection in the COCO 2015 competition (Ren et al., 2015). Regardless of the CNN used for detection, Faster R-CNN is still superior to other detection methods in terms of accuracy. Regarding the detection speed, the SSD method comes out ahead (Huang et al., 2017). Despite being slower, the Faster R-CNN guarantees an adequate speed to be used in this work.

Deep convolutional neural networks require a large set of training data to avoid overfitting, but large sets are often difficult to obtain. One approach to avoid overfitting is to use regularization techniques such as Dropout (Srivastava et al., 2014) and Batch Normalization (BN) (Ioffe and Szegedy, 2015). Another regularization technique is data augmentation, which consists of creating new examples from the training base (Lemley et al., 2017). It increases the training base by using various transformations in the image: translation, rotation, flipping, cropping, addition of noise etc.

Knowledge transfer is often used, and is shown to be successful by several machine learning works (Pan and Yang, 2009). Traditional machine learning techniques learn from scratch, whereas transfer learning train a previously trained model with new data. Using models and weights trained in generic bases such as ImageNet and MS-COCO for detection in the medical field helps improving the performance of the convolutional network (Goyal et al., 2017).

3 METHODOLOGY

This section details the image dataset and the detector used to solve the problem. Also exposes the adaptations in the detector, parameter settings, and other functions to improve the performance in the detection of foot ulcers in patients with diabetes.

3.1 The DFU Dataset

The image dataset used in this work is part of the Diabetic Foot Ulcers Grand Challenge 2020 challenge (DFUC 2020) (Cassidy et al., 2020). There are 2000 images for training, 200 for validation, and in the end, 2000 images were released as test dataset. The images were collected over the years at Lancashire Teaching Hospital (LTH). These images are close-ups of feet with ulcers from patients with diabetes. Figure 1 shows image examples of the dataset. All images have 640×480 pixels. The images were acquired without flash as primary light source, and instead, room lights were used to ensure consistent colors. The ulcers were marked on the images as a rectangular region of interest (ROI) by specialists who used a specific software for this task (Cassidy et al., 2020).

3.2 Implementation Details

In this work, we propose an adapted version of the Faster R-CNN architecture for DFU detection, called Faster R-CNN DFU. Figure 2 describes the entire architecture of our approach. The RPN and the Classifier are the two main modules that share a common set of convolutional layers. The feature maps extracted by the convolution layers serve as input for the RPN and the Classifier. The RPN outputs a set of rectangular object proposals, each one with an objectness score, which also serves as input for the Classifier. Each rectangular object is classified into a set of predefined labels, each one with a score. Our adaption of this architecture aims to improve the precision of the ROIs, enhance the detection of different sizes of ulcers, minimize the detection of false positives, and speed up the detection time. We also propose a variant of the detector which can be used for general kind of problems, called Faster R-CNN FP. Its focus is on reducing false positives and improving detection performance.

The original Faster R-CNN implementation used ZF (Zeiler and Fergus, 2014) and VGG (Simonyan and Zisserman, 2014) as part of the RPN and of the classifier. However, Ren et al. (Ren et al., 2018) have experimentally proved that the pre-trained ResNet-50 model achieves a better performance when compared to other popular CNNs such as VGG and Inception (Zeiler and Fergus, 2014; Szegedy et al., 2015; Szegedy et al., 2016). Therefore, in this work, ResNet-50 was chosen as the deep convolutional neural network for the Faster R-CNN DFU and Faster R-CNN FP.

In the Fast R-CNN detector (Girshick, 2015) the negative ROI (Region of Interest) samples that are sent for classification are those that have an IoU (Intersection over Union) in the range of [0.1, 0.5). The IoU is an evaluation metric, also known as the Jaccard Index, given by Equation 1:

$$IoU = \frac{|A \cap B|}{|A \cup B|} \tag{1}$$

where A and B are respectively the detected and ground truth bounding boxes. An IoU greater than or equal to 0.1 causes the classifier not to be trained with regions of the image without ulcers, which can favor the appearance of false positives. We changed the interval to [0.0, 0.5), so that true negatives, regions without ulcers, are also used in training. This range is used by Ren et al. (Ren et al., 2015) and is shown to improve the accuracy of the detector. We used this strategy in both approaches in order to minimize false positives.

Detailed analysis of the dataset reveals a wide variety in the size of the ulcers. The original Faster R-CNN implementation uses 9 standard anchors, defined by all the possible combinations of the sizes 128×128, 256×256, and 512×512 with the aspect ratios 1:1, 1:2, and 2:1. Due to this fact, the network fails to detect very small lesions. In our implementation of the Faster R-CNN DFU, we added the 64×64 anchor size to the set of standard anchor scales, maintaining the original aspect ratios. Therefore, a total of 12 anchors were used, which improved the accuracy in the detection of small lesions (Ren et al., 2015; Sun et al., 2018).

One of the great advantages of the Faster R-CNN is due to the use of shared convolutional layers with the RPN, which significantly reduces the region proposal cost. RPN suggests regions of the image for the classifier. The number of regions suggested in the standard implementation is 300. But Fan et al. (Fan et al., 2016) verified that a decrease of this number, besides improving the response time, can also improve precision. In the training of our approaches, the value of 100 ROI suggestions was used.

The Faster R-CNN FP approach is an improvement on the Faster R-CNN. It uses the ResNet-50 as CNN, an IoU sample range of $[0.0, 0.5)$ for the negative ROIs, and 100 ROI suggestions. The Faster R-CNN DFU approach uses the same Faster R-CNN FP configurations and improves it with specific strategies for DFU detection. The main strategy was to use 12 different anchors for the detection of a greater variety of ulcer formats. Our algorithms were implemented using the Tensorflow API (Abadi et al., 2016), which provides an open-source framework that assists in the implementation of several detection models. The code is written in Python and is publicly available [1].

3.3 Training

To augment the training and validation datasets, horizontal and vertical flips, rotations by 180°, and Gaussian blur, to emulate the blur caused by cell phone cameras, were used. The Dropout and BN regularization techniques were used in the neural network. We used the weights of ResNet-50 (He et al., 2016), pre-trained with the image database ImageNet (Russakovsky et al., 2015). This dataset contains millions of images with annotations of different classes of objects. The regularization techniques showed to improve detection in our experiments.

We randomly divided the whole dataset of 2000 images provided by the challenge into 1600 images (80%) for the training set and 400 images (20%) for the test set. During training, we used the Monte Carlo cross-validation methodology (Xu and Liang, 2001), which randomly partitions the training set into 85% for training and 15% for validation. At each new training iteration, new images are selected for training and validation. Faster R-CNN requires the scaling of the training images based on the smallest side

[1] https://github.com/ArturLeandro/dfu_faster_rcnn

of the image. The 640×480 size was maintained for the training and validation images.

We used 100 epochs to perform the training of our algorithm. This number of epochs is enough for the loss function to converge to its lowest value. At each epoch, 1000 images were selected to train the RPN and the classifier. The learning rate used was 0.00001 in the first 60 epochs and 0.000001 for the others. The loss function implementation follows the same definitions of multi-task loss minimization proposed by Ren et al. (Ren et al., 2015).

4 EXPERIMENTS AND RESULTS

We tested four different approaches to detect DFUs in our experiments. First of all, the Faster R-CNN DFU detector with all the implementation details described in the methodology section. The second is the Faster R-CNN FP, our implementation that reduces false positives. The third one is the standard SSD300 (Liu et al., 2016) approach with the convolutional network VGG. And the fourth is the standard Faster R-CNN (Ren et al., 2015) approach with the pre-trained convolutional network ResNet-50. The objective is to identify positive and negative points in our strategies compared to the standard implementations of the detectors. We used a total of 100 epochs for training the SSD300 and the three versions of Faster R-CNN detectors. The experiments were done by detecting ulcers in the 400 images of the test set. The machine used in the experiments has a CPU Intel i3-8100 @ 3.6GHz, GPU NVIDIA GeForce GTX 1050 Ti SC 4GB, and 16GB DDR4 of RAM.

Table 1 shows the mean average precision (mAP) and F1-Score of each detector. The mAP is a metric widely used in detection works and is given by the area under the precision/recall (PR) curve of the detector. This metric needs an *overlap criterion* that specifies the minimum value of the intersection over union (IoU) to be considered a correct detection. The value of 0.5 was chosen for this criterion as it is a value widely used in the literature. The F1-score is a metric defined by the harmonic mean of precision and recall. The precision, recall, and F1-score can be calculated using the followings expressions:

$$\text{precision} = \frac{\text{TP}}{\text{TP}+\text{FP}} \tag{2}$$

$$\text{recall} = \frac{\text{TP}}{\text{TP}+\text{FN}} \tag{3}$$

$$F_1 = 2 * \frac{\text{precision} * \text{recall}}{\text{precision} + \text{recall}} \tag{4}$$

where TP represents the number of true positive detections, FP the false positives, and FN the false negatives.

Table 1 shows the results obtained by the tested techniques. The proposed Faster R-CNN DFU outperformed the other techniques. It has the best mAP and F1-score for the DFU dataset, which indicates that the regions found by our approach are closer to the regions of ulcers marked by specialists. Because of the high value of the F1-score, many true positive regions are detected and a very low number of false positives are detected. Figure 3 shows the approximation of all detection techniques regions with the ground-truth boxes. The returned values of classification accuracy are labeled on top of the region mark. It is possible to notice in the images the improvement in the detection of DFUs with the two proposed techniques. Unlike the standard version of Faster R-CNN, the Faster R-CNN FP decreases the false positive detection, and the Faster R-CNN DFU, besides increasing the precision of the detection, is also successful in detecting small ulcers. SSD300 has good results, but it fails to find DFUs and does not achieve a good precision.

Data augmentation and the changes proposed to decrease false positives increased the mAP in 10.7 percentage points, and the F1-Score in 18.5 percentage points. Figure 4 (a) evidences this improvement by showing the ROC curve of all detectors. The number of false positives decreases considerably after using the proposed techniques, particularly when compared to the results of Faster R-CNN. Likewise, as shown in Figure 4 (b), precision and recall both remain at high values, increasing the area under the curve. A high recall is related to a low number of

Table 1: Performance of DFU detection techniques on the DFU Dataset. Proposed techniques are denoted with *.

Technique	mAP	F1-score
Faster R-CNN DFU*	91.4	94.8
Faster R-CNN FP*	86.5	91.9
Faster R-CNN	80.7	76.3
SSD300	52.7	65.7

Table 2: Detection average speed (DAS) in milliseconds and model size in megabytes of DFU detection neural networks. Proposed techniques are denoted with *.

Technique	DAS (ms)	Size (MB)
SSD300	48	92.9
Faster R-CNN DFU*	332	111.1
Faster R-CNN FP*	362	111.1
Faster R-CNN	807	111.1

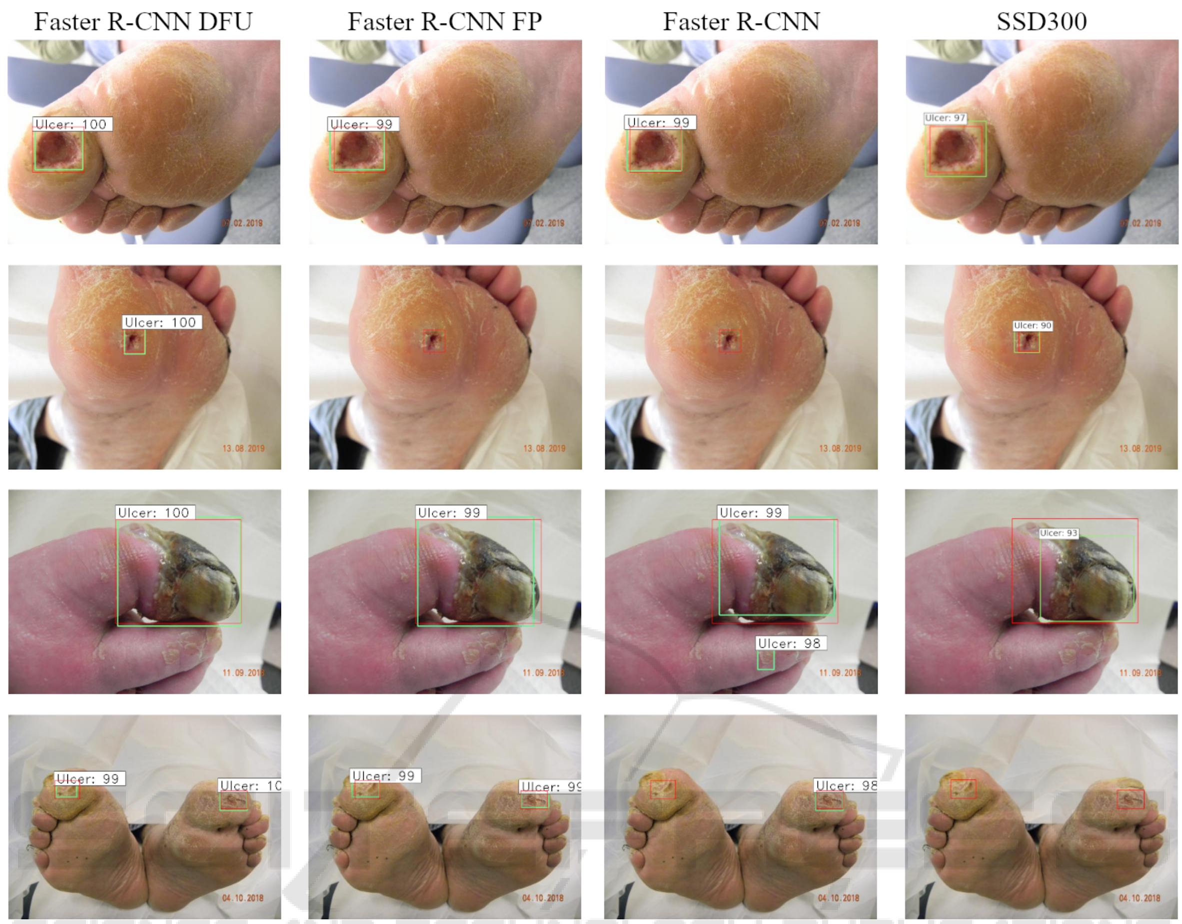


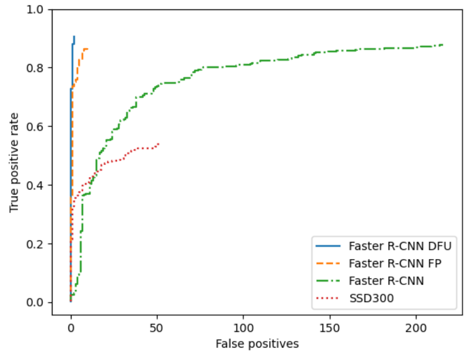
Figure 3: Detection results from the DFU detection techniques. In green are the detected regions and in red the ground-truth boxes.

false negatives, which is usually desirable in medical context.

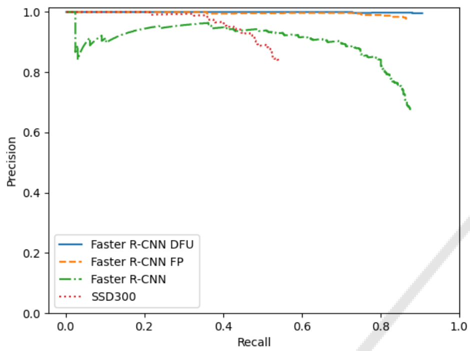
Table 2 shows the results of the detection average speed, and model size of each detector. The SSD300 obtained the best average speed and the smallest size of the model in comparison to the other detectors. This is mainly due to the simpler architecture to generate anchor boxes (Liu et al., 2016). However, its precision is lower than the other techniques. The average detection time of our two proposals is smaller than the standard Faster R-CNN implementation due to the use of a smaller number of ROI suggestions. The size of all Faster R-CNN variants is the same, as all of them are based on the Resnet-50 CNN. Their sizes are slightly larger than SSD300. Our approaches can be used on devices that have limited resources due to the small size of the model and a lower process time consumption.

5 CONCLUSIONS

In this work, we propose an automatic approach to detect DFUs using deep learning techniques. We have implemented an extended version of the Faster R-CNN approach. We have adopted several strategies to achieve high precision in detecting ulcers, to decrease the number of false positives, and to speed up the detection time. We changed the numbers of regions, the anchor scales, used data augmentation in the dataset, and adopted a CNN that has better detection results than previous approaches. Finally, we carried out experiments with the chosen detectors, training each one with 100 epochs. Results showed that our strategies improve the mAP and F1-score when compared to standard detector implementations known as the state-of-the-art. Better mAP, F1-score and detection speeds have been achieved which allows not only for better detection of the DFUs, but also a better confidence to use the Faster R-CNN DFU in real applications.



(a)



(b)

Figure 4: Comparisons of ROC curves for different experimental settings for DFU detection.

REFERENCES

Abadi, M., Barham, P., Chen, J., Chen, Z., Davis, A., Dean, J., Devin, M., Ghemawat, S., Irving, G., Isard, M., et al. (2016). Tensorflow: A system for large-scale machine learning. In *12th {USENIX} symposium on operating systems design and implementation ({OSDI} 16)*, pages 265–283.

Armstrong, D. G., Boulton, A. J., and Bus, S. A. (2017). Diabetic foot ulcers and their recurrence. *New England Journal of Medicine*, 376(24):2367–2375.

Bankman, I. (2008). *Handbook of medical image processing and analysis*. Elsevier.

Cassidy, B., Reeves, N. D., Joseph, P., Gillespie, D., O'Shea, C., Rajbhandari, S., Maiya, A. G., Frank, E., Boulton, A., Armstrong, D., et al. (2020). Dfuc2020: Analysis towards diabetic foot ulcer detection. *arXiv preprint arXiv:2004.11853*.

Dai, J., Li, Y., He, K., and Sun, J. (2016). R-fcn: Object detection via region-based fully convolutional networks. In *Advances in neural information processing systems*, pages 379–387.

Everingham, M., Eslami, S. A., Van Gool, L., Williams, C. K., Winn, J., and Zisserman, A. (2015). The pascal visual object classes challenge: A retrospective. *International journal of computer vision*, 111(1):98–136.

Everingham, M., Van Gool, L., Williams, C. K., Winn, J., and Zisserman, A. (2007). The pascal visual object classes challenge 2007 (voc2007) results.

Fan, Q., Brown, L., and Smith, J. (2016). A closer look at faster r-cnn for vehicle detection. In *2016 IEEE intelligent vehicles symposium (IV)*, pages 124–129. IEEE.

Girshick, R. (2015). Fast R-CNN. In *Proceedings of the IEEE international conference on computer vision*, pages 1440–1448.

Goyal, M., Reeves, N. D., Davison, A. K., Rajbhandari, S., Spragg, J., and Yap, M. H. (2018a). Dfunet: Convolutional neural networks for diabetic foot ulcer classification. *IEEE Transactions on Emerging Topics in Computational Intelligence*.

Goyal, M., Reeves, N. D., Rajbhandari, S., and Yap, M. H. (2018b). Robust methods for real-time diabetic foot ulcer detection and localization on mobile devices. *IEEE journal of biomedical and health informatics*, 23(4):1730–1741.

Goyal, M., Yap, M. H., Reeves, N. D., Rajbhandari, S., and Spragg, J. (2017). Fully convolutional networks for diabetic foot ulcer segmentation. In *2017 IEEE international conference on systems, man, and cybernetics (SMC)*, pages 618–623. IEEE.

He, K., Zhang, X., Ren, S., and Sun, J. (2016). Deep residual learning for image recognition. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 770–778.

Huang, J., Rathod, V., Sun, C., Zhu, M., Korattikara, A., Fathi, A., Fischer, I., Wojna, Z., Song, Y., Guadarrama, S., et al. (2017). Speed/accuracy trade-offs for modern convolutional object detectors. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 7310–7311.

Ioffe, S. and Szegedy, C. (2015). Batch normalization: Accelerating deep network training by reducing internal covariate shift. *arXiv preprint arXiv:1502.03167*.

Lemley, J., Bazrafkan, S., and Corcoran, P. (2017). Smart augmentation learning an optimal data augmentation strategy. *Ieee Access*, 5:5858–5869.

Leung, P. (2007). Diabetic foot ulcers—a comprehensive review. *The Surgeon*, 5(4):219–231.

Litjens, G., Kooi, T., Bejnordi, B. E., Setio, A. A. A., Ciompi, F., Ghafoorian, M., Van Der Laak, J. A., Van Ginneken, B., and Sánchez, C. I. (2017). A survey on deep learning in medical image analysis. *Medical image analysis*, 42:60–88.

Liu, C., van Netten, J. J., Van Baal, J. G., Bus, S. A., and van Der Heijden, F. (2015). Automatic detection of diabetic foot complications with infrared thermography by asymmetric analysis. *Journal of biomedical optics*, 20(2):026003.

Liu, W., Anguelov, D., Erhan, D., Szegedy, C., Reed, S., Fu, C.-Y., and Berg, A. C. (2016). Ssd: Single shot multibox detector. In *European conference on computer vision*, pages 21–37. Springer.

Pan, S. J. and Yang, Q. (2009). A survey on transfer learning. *IEEE Transactions on knowledge and data engineering*, 22(10):1345–1359.

Redmon, J., Divvala, S., Girshick, R., and Farhadi, A. (2016). You only look once: Unified, real-time object detection. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 779–788.

Ren, S., He, K., Girshick, R., and Sun, J. (2015). Faster R-CNN: Towards real-time object detection with region proposal networks. In *Advances in neural information processing systems*, pages 91–99.

Ren, Y., Zhu, C., and Xiao, S. (2018). Object detection based on fast/faster rcnn employing fully convolutional architectures. *Mathematical Problems in Engineering*, 2018.

Russakovsky, O., Deng, J., Su, H., Krause, J., Satheesh, S., Ma, S., Huang, Z., Karpathy, A., Khosla, A., Bernstein, M., et al. (2015). Imagenet large scale visual recognition challenge. *International journal of computer vision*, 115(3):211–252.

Saeedi, P., Petersohn, I., Salpea, P., Malanda, B., Karuranga, S., Unwin, N., Colagiuri, S., Guariguata, L., Motala, A. A., Ogurtsova, K., et al. (2019). Global and regional diabetes prevalence estimates for 2019 and projections for 2030 and 2045: Results from the international diabetes federation diabetes atlas. *Diabetes research and clinical practice*, 157:107843.

Saminathan, J., Sasikala, M., Narayanamurthy, V., Rajesh, K., and Arvind, R. (2020). Computer aided detection of diabetic foot ulcer using asymmetry analysis of texture and temperature features. *Infrared Physics & Technology*, 105:103219.

Simonyan, K. and Zisserman, A. (2014). Very deep convolutional networks for large-scale image recognition. *arXiv preprint arXiv:1409.1556*.

Srivastava, N., Hinton, G., Krizhevsky, A., Sutskever, I., and Salakhutdinov, R. (2014). Dropout: a simple way to prevent neural networks from overfitting. *The journal of machine learning research*, 15(1):1929–1958.

Sun, X., Wu, P., and Hoi, S. C. (2018). Face detection using deep learning: An improved faster rcnn approach. *Neurocomputing*, 299:42–50.

Szegedy, C., Liu, W., Jia, Y., Sermanet, P., Reed, S., Anguelov, D., Erhan, D., Vanhoucke, V., and Rabinovich, A. (2015). Going deeper with convolutions. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 1–9.

Szegedy, C., Vanhoucke, V., Ioffe, S., Shlens, J., and Wojna, Z. (2016). Rethinking the inception architecture for computer vision. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 2818–2826.

Xu, Q.-S. and Liang, Y.-Z. (2001). Monte carlo cross validation. *Chemometrics and Intelligent Laboratory Systems*, 56(1):1–11.

Zeiler, M. D. and Fergus, R. (2014). Visualizing and understanding convolutional networks. In *European conference on computer vision*, pages 818–833. Springer.