ISSN 2616-5775 Vol. 7, Issue 4: 110-117, DOI: 10.25236/AJCIS.2024.070415

Comparative Study on the Application of Deep Learning Algorithm in Burn Depth Diagnosis

Bohan Wei[1,*]

[1]Shanghai Datong High School, Shanghai, China
[*]Corresponding author

Abstract: *As one of the main causes of casualties in the world, fires, burns and scalds result in serious social and economic burdens in the world, and the annual incidence of burns in China exceeds 1%. However, the accuracy of surgeons in judging burn depth is quite low at present, with less than 50% accuracy of inexperienced doctors and only 64% to 76% accuracy of experienced doctors. For these reasons, there is great potential according to the studies on burn diagnosis based on deep learning algorithm. Therefore, in this study, a variety of convolutional neural networks were used to compare the classification of burn degree, and the processed datasets were used for training. By testing a series of neural networks, the best result was the precision of 88.75% when using the pre-trained VGG16 network. This result shows that the deep learning algorithm has high accuracy and application potential in the field of burn diagnosis.*

Keywords: *Burn; Convolutional neural networks; ResNet; VGGNet; DenseNet*

1. Introduction

In terms of fatal injuries, scalds rank sixth in the world. According to statistics, the incidence of scalds in China exceeds 1% every year, which brings a heavy burden to society and economy. For scald patients, timely and accurate assessment of scald degree is regarded as the key premise, which can provide accurate fluid infusion and determine the next treatment plan, which directly affects the wound treatment and rehabilitation of patients, and even relates to the life of patients [1]. However, even experienced surgeons have an accuracy rate of only about 50% when evaluating the depth of burns; for surgeons with less experience, the accuracy rate is lower. The determination of burn depth is an important link in disease diagnosis. In addition, the cost of burn ward is also expensive. In the absence of a scald specialist, an on-line scald diagnostic tool is necessary to initially determine the depth of the scald. The subject of this research project is scald diagnosis based on skin imaging, so as to achieve rapid and preliminary diagnosis of scald degree.

Ning Wei et al. demonstrated a scheme to diagnose burn degree by using non-visible light imaging technology, which cannot be realized on ordinary mobile phones [2]; Oriole put forward an idea of objectively diagnosing burn depth by laser Doppler imaging, which has not been applied in practice and can not be used on general mobile phones [3]; Han Xuhui et al. put forward a lightweight model BI-YOLOv5 algorithm, which is used to classify skin burns. Under the condition of detecting and distinguishing different burn categories and environmental interference, burn detection has high accuracy and efficiency, and realizes high accuracy burn detection [4]; Liu Hao used image segmentation and classification technology based on deep learning to build a complete technical framework of burn diagnosis process to achieve accurate and efficient burn image segmentation and burn depth classification diagnosis [5]; Ji Xiaofeng introduced the diagnostic techniques and methods of burn wound depth [6]; Yadav et al. have continuously studied the judgment of burn degree by machine learning and compared many different models [7-8].

After previous studies, we found that the existing burn diagnosis technology mainly depends on expert experience and expensive detection equipment, and can not be carried out on ordinary mobile phones without using external equipment. However, through this study, we can help hospitals achieve rapid graded shunt treatment of burns, thus saving medical resources. Suppose we can accurately diagnose the severity of burns through images. In order to achieve this goal, we use image recognition, target detection and target segmentation methods to detect the burn degree of the target.

I have a plan to launch a mobile phone application in the future, which can quickly detect the severity of users' injuries after burns. This application will help many patients quickly identify their burn degree. Meanwhile, I will continue to expand the dataset, update and improve the model, and open source all the code so that scholars and latecomers can reuse it.

2. Materials and Methods

2.1. Grade of Burn

Burns can be divided into three major levels and four minor levels. These three main levels are: first-degree burns, second-degree burns and third-degree burns; The four minor grades are subdivided into superficial second-degree burns and deep second-degree burns on the basis of second-degree burns. First-degree burns only affect the epidermis, causing local dryness, burning pain, slight swelling and redness of the skin, but without blisters, and usually take 3 to 5 days to recover. Second-degree burns can be divided into superficial second-degree burns and deep second-degree burns. Superficial second-degree burns involve the superficial layer of epidermis and dermis, resulting in thin-walled blisters, flushing at the bottom and obvious edema on the skin. If it is not infected, it usually takes 2 weeks to recover, and the pain is severe, but no scars are left. Deep second-degree burns involve deep dermis and residual skin accessories, resulting in pale or pale skin, small blisters with thick blister walls and mild pain. It usually takes 3 to 4 weeks to heal, and it is easy to infect and form scars. The depth of third-degree burn reaches the whole layer of skin and some subcutaneous tissues, muscles and bones, which leads to pale skin and scab formation. There is edema under scab, no pain, no self-recovery, and scars will be left after recovery.

The labeling of the dataset in this study is completely based on the above description, but this study does not distinguish the subdivided superficial second-degree burns from deep second-degree burns for the time being.

2.2. Assumption stage

As of August 2022, YOLOv7 model is the latest model in YOLO model series. According to YOLOv7 paper published by Wang, Bochkovskiy and Liao in 2022, this model is the fastest and most accurate detector in real-time object detection at present. By improving performance, YOLOv7 lays an important benchmark for subsequent research. In the post-processing of the model, non-maximum suppression (NMS) is used to get the final prediction results.

YOLOv7 studies the re-parameterization of model structure, and analyzes the re-parameterization strategy suitable for each layer structure in different networks by using the concept of gradient propagation path. They proposed a re-parameterization method for programming model structure. In addition, when using the dynamic label allocation strategy, the multi-output layer model will encounter a new problem in training, that is, how to better allocate dynamic targets for different branches. In order to solve this problem, the author proposes a boot label allocation strategy called "coarse to fine", which can better allocate dynamic targets for different branches.

Structural Reconfiguration refers to creating a series of structures (usually used for training), and then mapping their parameters to another set of parameters (usually used for reasoning), so as to transform this series of structures into another series of structures equally.

In addition, YOLOv7 is very suitable for beginners, and only simple settings can be used for model training. Therefore, in the initial stage of this study, I collected some public datasets Skin Burn Dataset from the network (kaggle), and tried to design it with YOLOv7 framework to verify the feasibility of the assumption. In the following article, I will call this model Model Zero.

The skin burn dataset contains 1437 images captured through the network, which cover different degrees of skin burns. The dataset is labeled in YOLO format, where grades 0, 1 and 2 represent first-degree burns, second-degree burns and third-degree burns, respectively. It is understood that the number of first-degree burns and second-degree burns in Skin Burn dataset far exceeds the number of third-degree burns. Specifically, there were 611 samples with first-degree burns, 587 samples with second-degree burns and only 239 samples with third-degree burns.

This program written by me aims to divide the open dataset on the network into training set, verification set and test set. It can randomly divide automatic datasets into YOLO dataset formats.

Next, you can complete the partition of the dataset by performing the following operations. Simply put, after executing the following command, the script mentioned earlier will read data from the dataset located under the ./data folder, and randomly use 80% of the data as the training set, 10% as the validation set (val), and another 10% as the test set (test). The partitioned dataset will be exported to the./work/yolov7/dataset folder.

After the dataset is segmented, the model training can be started. Because of the complexity of YOLOv7 model, the demand for computing power is high. Therefore, I chose a server equipped with NVIDIA RTX A4000 image processing unit, 6-core Intel ® Xeon ® Processor E5-2680 v4 central processing unit and 30.1 GB RAM for model training. To run YOLOv7, I installed Ubuntu 18.04, Python 3.7.10, Docker 20.10.10, CUDA 11.2, PyTorch 1.10, and TensorFlow 2.7. 0 on the server.

Subsequently, this project trains the model on the above server with the super parameters img-size=1000, batch-size=6 and epochs=600.

For model zero, this project uses Precision and Recall (as shown in the following formula) as indicators to evaluate the model. Among them, it represents a true class set, a false positive class set and a false negative class set.

$$Precision = \frac{TP}{(TP + FP)}$$

$$Recall = \frac{TP}{(TP + FN)}$$

2.3. Model

2.3.1. Models adopted at this stage

This study compared the performance of the following models on the same dataset.

Model A is DenseNet 121, a member of the DenseNet model family.

Model B, Model C and Model D are VGGNet16, VGGNet16 (pre-training) and VG-GNet19, respectively.

Model E, Model F, Model G, Model H and Model I are ResNet18, ResNet34, ResNet50, ResNet101 and ResNet152, respectively.

Theoretically speaking, on the premise of sufficient data, the more complex the model often means higher accuracy; however, when the data available for training is limited, with the increase of model complexity, the phenomenon of over-fitting often occurs.

2.3.2. Datasets used at this stage

After obtaining the original data, I began to process it to provide the model for learning.

Some of this data is collected from the Internet, accounting for about 50% of the total data. The rest of the data are desensitized images obtained from hospitals through cooperation with the instructor, Teacher Wang. After obtaining these data, I also carried out the work of data annotation.

The following figure shows the data processing process. We need to cut the raw data (see the left side of the data processing in Figure 1) into 4 blocks (see the right side of the data processing in Figure 1). In this process, one original data can usually be made into multiple samples, which can improve the reuse rate of data.

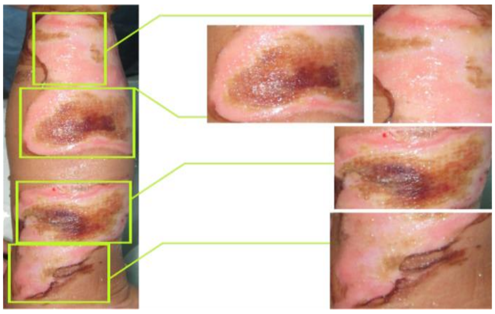

Figure 1: Data processing process

After that, this project classifies the data. According to my knowledge reserve, all the data are divided into three categories: first-degree burn, second-degree burn and third-degree burn.

2.3.3. Comparison and Tracking of Models

In this study, WandB (https://wandb.ai) and other platforms were used to record experimental data such as model performance.

WandB is an online model training tracking platform, which is used to track the super parameters, system indicators and predictions set by researchers for models, so that researchers can compare models in real time.

2.3.4. Model training

After determining the model to use and the dataset to use, I began to train the model on the server. For the research in this stage, the server mentioned above is still used for training in this stage.

In order to record and compare the experimental data of model training, we must first create projects on WandB, and then initialize WandB in Python. (As shown in the following code)

In this study, the Receiver Operating Characteristic Curve (ROC curve, as shown in Fig. 2) was used to judge the prediction performance (accuracy) of the model.

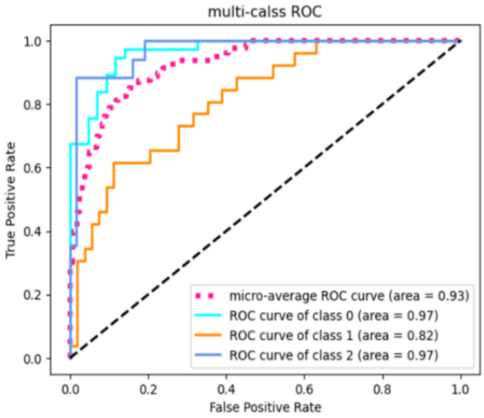

Figure 2: Receiver operating characteristic curve

3. Results and Discussion

3.1. Results and discussions in the assumption stage

As shown in the Precision-Recall curve in Fig. 3, all curves show a concave trend, and the Precision decreases with the increase of Recall. The orange curve (second-degree burn) has the largest area, and the green curve (third-degree burn) has the smallest area.

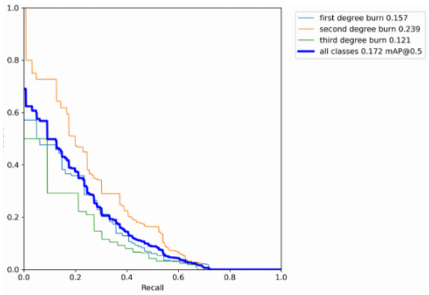

Figure 3: Precision-Recall curve

As shown in the F1-Confidence graph in Fig. 4, all curves show an upward trend, and the blue bold curve (the curve of all categories of F1 Score about Confidence) achieves a maximum value of 0.27 when the Confidence is 0.121.

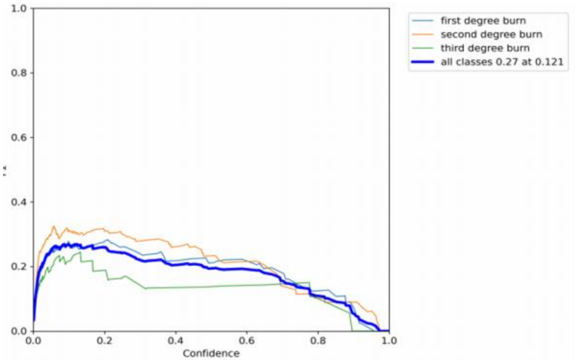

Figure 4: F1-Confidence curve

As shown in the target anchor frame of the training set in Fig. 5, the loss value of the target anchor frame of the training set shows a downward trend and converges.

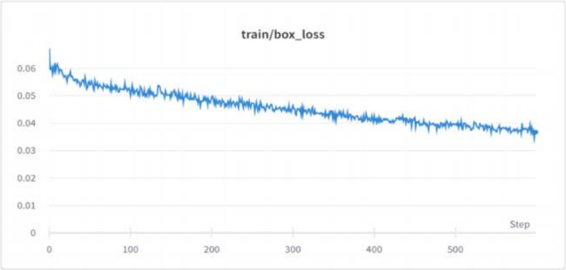

Figure 5: Target anchor frame of training set

As shown in the target detection of the training set in Fig. 6, the loss value of the target detection of the training set shows a downward trend and converges.

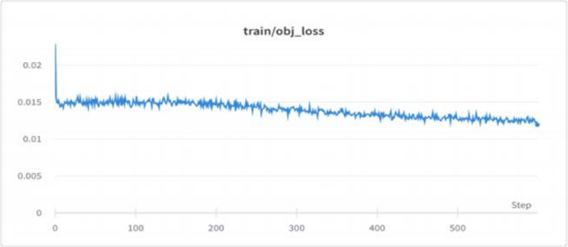

Figure 6: Target detection of training set

As shown in the classification of the training set in Fig. 7, the loss value of the classification of the training set shows a downward trend and converges around 0.005.

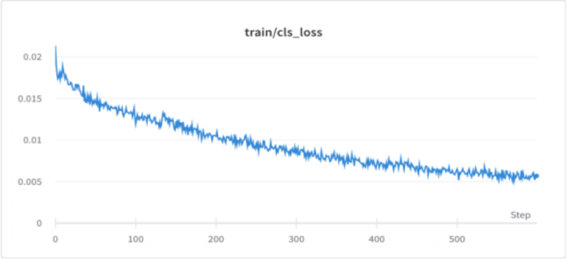

Figure 7: Classification of training sets

According to the obfuscation matrix heat diagram (Fig. 8 obfuscation matrix heat diagram), the three values on the diagonal are 0.26, 0.21 and 0.17 respectively, indicating that the model is not good for the three grades.

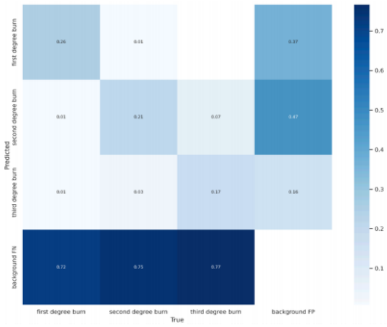

Figure 8: Confusion matrix thermodynamic diagram

According to the above experimental results, it can be known that YOLOv7 (No.0 model) does not perform well on the original network dataset. Considering the above index values and datasets, compared with other burn grades, this model is slightly better in the recognition of second-degree burns. Although the performance of the overall model zero is not good, looking at the confusion matrix heat diagram in Fig. 3-6, it can be found that the performance of the model zero among first-degree burns, second-degree burns and third-degree burns is still considerable, except for the high classification error rate between FN rows and FP columns in the background and other columns and rows.

According to my inference, the poor performance of model zero on this dataset may be due to the following factors: the poor quality of the dataset (there are problems such as image deformation, watermarking, data classification and labeling errors, etc.); The number distribution of datasets in several categories (first-degree burns, second-degree burns, third-degree burns) is uneven, etc.

3.2. Results and discussion in the in-depth study stage

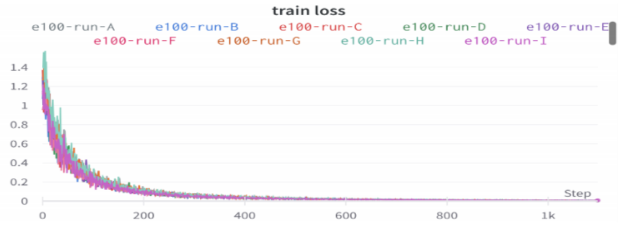

Figure 9: Train Loss

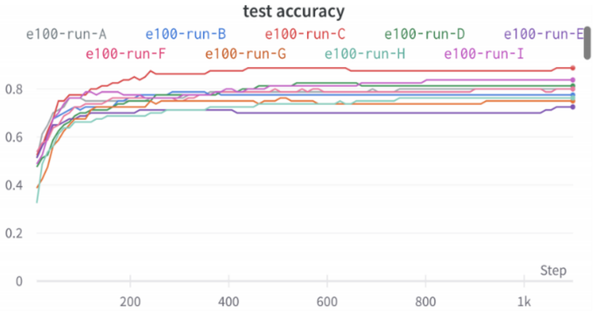

Figure 10: Test Accuracy

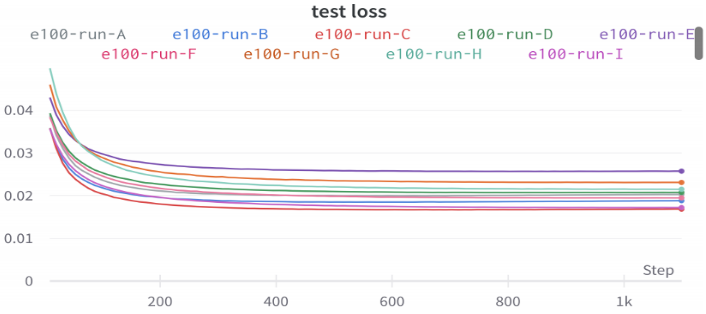

Figure 11: Test Loss

Because WandB is used to record the experimental data in this study, the performance of different models can be clearly observed.

According to the above test results (Fig. 9 test results, Fig. 10 test results and Fig. 11 test results), it can be seen that VGG family models (gray, blue and red curves) perform better on the existing datasets, among which model C (e100-run-C, VGG16 (pre-training)) performs best on the current datasets, and its test accuracy reaches 0.8875, namely 88.75%; Its test loss reaches 0.01686. This accuracy is actually higher than I expected. I think this may be due to the advantages of VGG's simplified network structure for the relatively small amount of data and the processing of datasets.

However, by further observing ROC curves, it is not difficult to find that the performance of this model on second-degree burns (class 1) is slightly worse than that on first-degree burns (class 0) and third-degree burns (class 2), which may be result from the fact that the characteristics of second-degree burns are more than those of first-degree burns and third-degree burns, and second-degree burns can be subdivided into superficial second-degree burns and deep second-degree burns.

4. Conclusions

4.1. Conclusion at this stage

This study is based on using pictures to quickly judge the degree of burn, so as to help patients carry out graded shunt treatment. Deep learning is used to train the model, so as to achieve this goal. According to the current dataset and the model evaluation index after training, the research goal is completely achievable, and the core task has been successfully completed. If an interactive system can be designed to assist patients in painting burn areas, then convolutional neural networks can be used to achieve relatively high accuracy of burn grade diagnosis. Then, we can develop application software on this basis for the majority of patients to use.

Compared with the previous stage, the quality of the dataset at this stage has been greatly improved, and the improvement in model performance brought by this is also obvious. This is mainly because the datasets used today have been carefully processed and well and accurately marked.

Based on the training results shown above, it is not difficult to find that the performance of relatively simple neural networks (Model C, Model D, Model F) is better than that of relatively complex neural networks (Model A, Model G, Model H). This kind of anomaly is caused by the small amount of data, so with the increase of data scale, it is more likely that relatively complex neural networks will train models with higher performance (accuracy). This also means that the performance (accuracy) of the current model still has a lot of room for improvement.

4.2. Outlook

In the future, I hope to collect more data about second-degree burns and subdivide them into superficial second-degree burns and deep second-degree burns to improve the accuracy of the model. In this way, even on the same dataset, the simplified network model can get better prediction results. Of course, while collecting more second-degree burn data, researchers also need to ensure the number of first-degree burn and third-degree burn data, because the imbalance of datasets is an important reason for the poor performance of the model in the hypothetical stage of research. After obtaining further data, we will continue to try various models and constantly adjust super parameters, and strive to train the best performance models and parameters. At the same time, I have planned to start developing an application that can be installed on mobile phones and quickly diagnose the degree of burns by taking photos and specifying the burn areas in the photos. This plan is already on the agenda, and I will develop such an application in the near future.

References

[1] Jiao C, Su K, Xie W, et al. Burn image segmentation based on mask regions with convolutional neural network deep learning framework: more accurate and more convenient[J/OL]. Burns Trauma, 2019, 7. DOI: 10.1186/s41038-018-0137-9.

[2] Ning W, Qi F, Wang J. Terahertz technology utilized to achieve the assessment of burn injuries[J/OL]. China Medical Devices, 2018, 33: 14-16. DOI: 10.3969/ j.issn.1674-1633.2018.07.003.

[3] Huang Y, Qiu L, Mei A L, et al. Meta-analysis on the diagnostic value of laser doppler imaging for

burn depth [J/OL]. Chinese Journal of Burns, 2017, 33: 301 – 308[2023-10-14]. https: //pubmed. ncbi.nlm. nih.gov/28651422/. DOI: 10.3760/cma.j.issn.1009-2587.2017.05.009.

[4] Han X, Liu Y, He G, et al. Skin burn wound classification algorithm based onmulti-scale feature fusion[J/OL]. Electronic Measurement Technology, 2022, 45: 114 – 118. DOI: 10.19651/j.cnki. emt. 2209382.

[5] Yadav D P. A method for human burn diagnosis using machine learning and slic superpixels based segmentation [J/OL]. IOP Conference Series: Materials Science and Engineering, 2021, 1116: 012186. DOI: 10.1088/1757-899x/1116/1/012186.

[6] Yadav D P, Sharma A, Singh M, et al. Feature extraction based machine learning for human burn diagnosis from burn images[J/OL]. IEEE Journal of Translational Engineering in Health and Medicine, 2019, 7: 1-7. https://dx.doi.o rg/10.1109%2FJTEHM.2019.2923628. DOI: 10.1109/jtehm. 2019. 2923628.

[7] Rowland R A, Ponticorvo A, BALDADO M L, et al. Burn wound clas-sification model using spatial frequency-domain imaging and machine learning [J/OL]. Journal of Biomedical Optics, 2019, 24. DOI: 10.1117/1.jbo.24.5.056007.

[8] Rowland R A, Ponticorvo A, BALDADO M L, et al. A simple burn wound severity assessment classifier based on spatial frequency domain imaging (sfdi) and machine learning[J/OL]. International Society for Optics and Photonics, 2019, 10851: 1085109. DOI: 10.1117/12.2510670.