THE UNIVERSITY OF QUEENSLAND

AUSTRALIA

Burn Wound Imaging and Deep Learning

By Xinqian Wang

School of Information Technology and Electrical Engineering

The University of Queensland

Submitted for Master of Data Science Capstone 1

19th April 2021

Introduction

I request permission to develop a system for burn wound recognition that helps to classify the categories of burn wounds and outline the edges of the burn wound in hospitals. Currently, burn wound accidents have attracted people's concern due to it ranked as one of the top reasons for depriving people's lives. However, the time-consuming and low-precision drawbacks of traditional burn diagnosis are the main weaknesses which attract many researchers' interests for making improvement. Additionally, Deep Learning technology is utilized in many fields for automatizing some works which were done manually in the past. Therefore, I will put efforts to develop a burn wound diagnosis system by training a deep neural network for improving the efficiency and accuracy of the burn wound diagnosis. The final system will make contributions to save people's lives when they get burned and bring huge convenience to hospitals when they are doing burn wound diagnosis. The final system will take around eight months to finish based on the data science project procedure which corresponds to problem framing, data collecting, data processing, system modeling, results evaluating, and system optimizing.

Background

The background section will include three parts. First, the knowledge of burn wound diagnosis. Second, former researches which related to burn wound but not included Deep Learning technology. Third, the development of Deep Learning technology, especially in Computer Vision field.

Burn injuries are one of the main challenges that human being faces. A burn is a type of injury to the skin, or often tissues, caused by heat, cold, electricity, chemicals, radiation(sunburn), or friction (Burn, 2021). The World Health Organization counted that around 180,000 people dead because of burn annually. The situation is especially severe in developing countries. For example, 17% of children with burns have a temporary disability and 18% have a permanent disability in Bangladesh, Colombia, Egypt and Pakistan (Burns, 2018). The diagnosis of burn severity main consists of four parts. First, the calculation of the percentage of Total Body Surface Area (TBSA). Doctors would decide which treatments are appropriate compared to the percentage of TBSA. For example, Adults with burns greater than 20% TBSA should undergo formal fluid resuscitation. Second, the classification of the burn injury's depth. Generally, superficial burns and deep burns are the two main categories of burn wound. The superficial burns are branch into Frist-degree burns and Second-degree burns. The deep burns are Third-degree burns and Fourth-degree burns. Primarily, the treatments of superficial burns and deep burns are different. The superficial burns heal by re-epithelization and the deep burns heal by skin grafts. Third, the consideration of patients' age. Children need more fluid per TBSA burns. Fourth, the localization of the burn wound position. The burn wounds from different positions of the body need different treatments (Burn Triage and Treatment, 2006). However, the accuracy of burn depth diagnosis by burn specialists would usually achieve an accuracy of lower than 70% in the first several days following a burn, which happened in a stage of proceeding with early excision (Orion Despo et al., 2017). Moreover, not all hospitals have equipped with burn specialists. Non-specialized doctors would do the burn wound diagnosis in the case there is no burn specialists exist. Such cases happening would lead to more misdiagnosis and many patients would get the care they do not need. As a result, many researchers tried to make a concrete model or methods which can automatize the burn wound diagnosis to improve the accuracy and efficiency of the hospitals.

Researchers have developed some techniques for helping the burn depth diagnosis without depending on the Deep Learning technique. Multidimensional scaling is applied to the image data and a K-Nearest Neighbor(K-NN) classifier is used to classify the different depths, which achieved an accuracy of 66.2% (Acha et al., 2013). A pixel color approach is used to make a discrimination between healthy skin and burn wounds (Badea et al., 2016). Mathematic multi-transformation and Support Vector Machine are combined together to make automatic segmentation and burn degree classification (Wantanajittikul et al., 2012). Overall, these researches had provided enough evidence for the feasibility of the Deep Learning implementation because the mathematic approaches they used are just the underlying logic of the Deep Learning technique, which is a successive mathematic transformation involving scaling, kernel space reflecting, etc.

The development of Deep Learning can combine what the former researchers have done and improve their automagical capability. The Deep Learning technique has been well developed since it was introduced in 1986. It has derived other fields such as Computer Vision, Natural Language Processing, Federal Learning, Voice Recognition, etc. For burn wound recognition, The Computer Vision technique is usually involved in the burn wound image. There are four main challenges in Computer Vision, image classification, object localization, semantic segmentation, and instance segmentation, among which burn wound classification and burn wound outlining will involve the utilization of image classification and object localization. Both of the two challenges are conquered by the Convolutional Neural Network (CNN). CNN is one of the representative algorithms of the Deep Learning technique. It has been well-known since the invention of Le-Net in 1998 (Yann LeCun, Leon Bottou, Yoshua Bengio, & Patrick Haffner, 1998). The Le-Net is used to make classification on the images of digits which obtained a good accuracy. Afterward, in the ILSVRC competition 2012, the invention of Alex-Net starts the bloom of the Deep Learning technique research in Computer Vision (Alex Krizhevsky, Ilya Sutskever, Geoffrey E. Hinton, 2012). In 2013, the Ze-Net came out which is only a little improved compared to the Alex-Net but begin a new research field of visualization CNN (Zeiler, Matthew D, & Fergus, Rob, 2014). In 2014, the VGG and Google-Net have shown that the more complicated neural network would probably get better accuracy. In 2015, to prevent the scenario where parameters are vanishing during the training process of a more complicated neural network, the Res-Net has been invented (Kaiming He, Xiangyu Zhang, Shaoqing Ren, & Jian Sun, 2016). In 2017 Dense-Net came out, which has used half number of parameters of Res-Net but achieve a better accuracy (Huang G, 2017). Therefore, it is possible for us to build our system based on one of these Deep Learning Networks.

Related Work

The section will chronologically provide some applications of burn wound recognitions that have involved Deep Learning Neural Networks.

Researchers have done many experiments since the Alex-Net came out. In 2015, A framework that joined the wound segmentation and wound classification is made by implementing an

Autoencoder Network (Changhan Wang et al., 2015). The framework used the Autoencoder Network to produce a binary mask region of the wound and not the wound. The Autoencoder Network is built by many different sizes of CNNs. Then they used the featured product of the Autoencoder Network made to classify by using a Support Vector Machine. The final framework is that it only takes 5 seconds to output the result for each clinic burn picture. In 2016, An automated system called Burn Convolutional Neural Network (B-CNN) is built for classifying the wounds' four different burn degrees (Son Tran, H., Hoang Le, T., & Thanh Nguyen, T., 2016). In 2017, A course work was done by using Fully Convolutional Networks (FCN) to outline the edges of burn wounds (Orion Despo et al., 2017). In 2019, many different state-of-the-art Deep Neural Networks are tested for their accuracy of doing burn wound segmentation and get an accuracy of 84.51% (Chong Jiao, Kehua Su, Weiguo Xie, & Ziqing Ye, 2019). In 2020, Res-Net is used to judge the depth of the burn wounds and get an accuracy of 80% for three types of burns (Yuan Wang et al., 2020). The application of Deep Learning in burn wound research has achieved a mature stage from the accuracy perspective. Moreover, the burn wound research industry is a broad field. The Deep learning algorithm can further undertake more responsibilities for more subtle discriminations in burn wound.

Problems posed by the related researches

In this section, I will explain three difficulties that former researchers have encountered. The common object of these researches is to achieve a robust burn wound recognition system. Therefore, overcoming these three difficulties of scarcity, diversity, and ambiguity is essential.

First, the labeled images of burn wounds are scarce. Most Deep Learning techniques are using supervised learning, which means the more labeled data to do the training, the better accuracy that the model can achieve. For labeling medical images, professional knowledge is commonly needed. Moreover, the process of labeling these images is mainly dependent on the labelers' subjective judgment. Therefore, the public labeled datasets for medical images are usually scarce for the researchers (usually they are from the non-medical domain) to do the experiments. Generally, the model trained by the former researchers is dependent on only a small amount of image data which is not enough for facing the real-world challenge.

Second, the burn wound images acquired by the researchers are usually diverse. All shapes of burn wound images can occur in the real world. For the Deep Learning technique, scientists would usually standardize the image according to their previous understanding of the pictures' distribution. Apart from the ideal assumption, a lack of understanding of burn wound image is a truth due to the scarcity of images. Consequently, it is non-standardized for the burn wound image and it brings challenges to the training of Neural Network. Moreover, one image can usually represent different information for Deep Learning model because of some transformations like rotation, scale, and translation. For instance, it would be hard for the algorithm to recognize the images from Figure 1. are totally the same (Zeiler et al., 2014). Additionally, some background may misguide the algorithm. Some objects that may look like same as the burn wound would appear in the burn wound images' background. For example, used medical gauzes that with the same bloody color and same wound shape could be recognized to wound. Thus, it is hard for the algorithm to handle such diverse burn wound images.

Figure 1. Applied rotation transformation to a burn wound image

Third, some burn wound images may contain underlying information. The image data is only two-dimension. Whereas, some burn wound images would be spatial such as a burn wound from the connection skin part between shoulder and arm as Figure 2. shows. Therefore, the algorithm needs to learn the spatial information from a two-dimension image is a huge challenge. As a result, the algorithm could fail to obtain the spatial information which is important for burn wound diagnosis if no spatial treatment is applied to the image and no such cases are involved in the training process.

Figure 2. The burn wound image (red parts represents the burn area)

Project objective

The project aims to develop a burn wound diagnosis system based on the Deep Learning technique of Neuron Network training. Summarizing the integral goals into three parts, high precision for burn wound classification, high efficiency with getting the output and detailed explanations of the burn wound images' analysis outcome.

For further explanation, the three goals can branch into four subordinate parts: burn wound binary classification, burn severity multi-classification, body part multi-classification, and burn wound segmentation. The five subordinate parts formed the diagnosis process of the algorithm:

- burn wound binary classification: the algorithm makes judgment to the input images and output whether the burn wound exists.
- burn severity multi-classification: the algorithm makes classifications for severities of burn wound image.
- body part multi-classification: the algorithm makes classifications for which part of the body that the burn wound should belong to.
- burn wound segmentation: the algorithm make segmentation to the burn wound.

For burn wound binary classification, the algorithm should be able to achieve a precision that closes to 100%. For the body part multi-classification and burn severity multi-classification, at least 80% precision can achieve. For burn wound segmentation, the algorithm should outline the wound edge most correctly, and healthy skin should never be included.

The burn wound diagnosis system should with high efficiency of less than 30 seconds so that the doctors can have enough time to do the treatment.

The output of the burn wound diagnosis system should be detailed. Otherwise, it is hard for non-specialized doctors to do the left job.

Proposed solution

The solution of the project will correspond to six parts, problem framing, data collecting, data processing, system modeling, results evaluating, and system optimizing. The following parts will give the interpretation for each stage.

Step 1: Problem framing

The design of the burn wound diagnosis system includes four steps according to the data streamflow, data collection, train-test dataset separation, dataset training, and result data analysis.

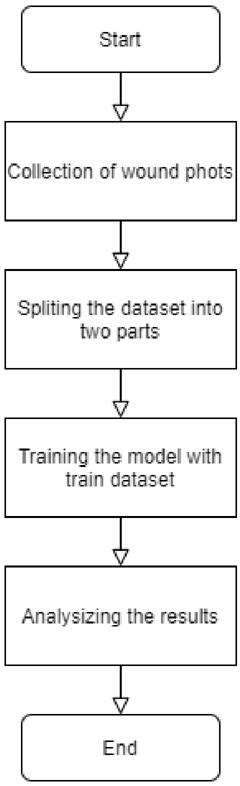

Figure 4. Data streamflow

First, collecting the data from the resource and transform them from image data into matrix data. Then, keep transferring the matrix data into tensor data, which has three dimensions for each image, Height, Width, and Channel. Moreover, the Channel would be RGB Channel, divided by Red, Green, and Blue. Second, Splitting the dataset into the training set and testing set. As a common experience, we would split the whole dataset into 30% and 70%, to be as called test dataset and train dataset. Data augmentation and Cross-Validation are two acceptable methods when the dataset is not big enough for training a robust model. Third, In the training process, we can do some visualization for each convolutional layer. The convolutional layer is a fundamental building block for the Neural Network. Mathematical, it just operates a transformation to the data we input. For visualization, the de-convolutional approach is usually used to get the feature image. The de-convolution is just operating a reverse process to the convolutional layer and switch the output of de-convolution into the image to do visualization. Such a feature map would let researchers know whether the algorithm has learned features correctly (Zeiler M D, 2010). Fourth, a built model needs test data to evaluate the precision. The whole evaluation process should contain two parts. First, considering some common criteria. For example, testing burn wound segmentation with the criteria of mean intersection-over-union (IOU). Second, simulating some adversarial attacks or special input images to test the robustness. For example, from Figure 5., the burn wound image is incomplete and ambiguous, the burn wound diagnosis system should ignore such disturbing signals.

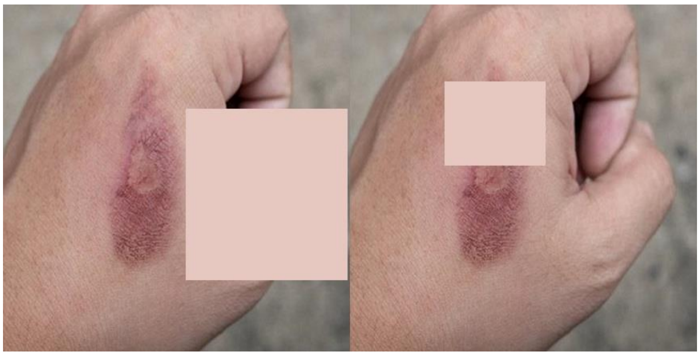

Figure 5. The images with disturbing signals

Step 2: Data collecting

The data source is mainly from Herston Biofabrication Institute (HBI). HBI is the first institute of its kind – advancing knowledge and technology in 3D scanning, 3D modelling and 3D printing of medical devices, bone, cartilage and human tissue. Opened in 2020, the Institute takes a multidisciplinary approach, bringing together clinicians, academics, industry and consumers in its 1,500-sqm, state of the art facility.

Figure 6. The logo of Herston Biofabrication Institute

In addition, some public datasets would also be helpful for training and testing the model. My group and I have already found two datasets. First, the Medetec Medical Image, contains burn images and scalds of varying size and severity. The dataset contains 19 unlabeled images and all of them can be used for research as the website declared. Second, the dataset of Biomedical Image Processing (BIP) Group from the Signal Theory and Communications Department (University of Seville, SPAIN) and Virgen del Rocío Hospital (Seville, SPAIN). The dataset contains 94 images with three different labels of burn depth, Full-thickness, Deep dermal, and Superficial dermal.

Step 3: Data processing

The normal burn wound would not go into standardization as unknown image distribution. Using the Python Library of OpenCV to read the image into the tensor data structure.

Step 4: System modeling

As in Figure 7., the system modeling is going to training two Deep Learning Neuronal Networks for classification and segmentation. Based on the former researches, the Visual Geometry Group (VGG-19) and Fully Convolutional Network (FCN) will be implemented for initialization. VGG-19 classification accounts for the burn wound's categories. FCN segmentation accounts for the burn wound's area calculation. Different area of burn wound with different sorts would make great changes the doctor's treatment.

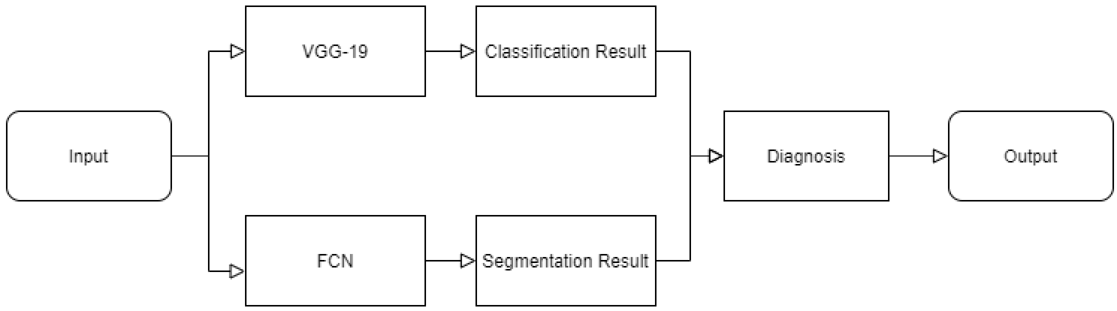

Figure 6. The system modeling

Step 5: Results evaluating

For evaluating the results, six equations are listed. Eq1-4 are the criteria for the VGG-19 classifier and Eq 5-6 are for the FCN segmentation.

Eq.1 provide the Accuracy formulation of the classifier's prediction (Aliyu Abubakar, 2020). Among the equation, TP, TN, FP, and FN are the values from the confusion matrix.

$$\text{Accuracy } = \frac{\text{TP} + \text{TN}}{\text{TP} + \text{TN} + \text{FP} + \text{FN}} \tag{1}$$

Eq.2 provide the Recall (or sensitivity) formulation, which gives the accurate prediction of individual class by the classifier.

$$\text{Recall } = \frac{TP}{TP + FN} \tag{2}$$

Eq.3 provide the Precision formulation, which can measure the classifier's quality. Higher Precision means more irrelevant results are generated by the classifier.

$$\text{Precision } = \frac{TP}{TP + FP} \tag{3}$$

Eq.4 provide the F1-score formulation, which it evaluates the balance between the Precision and the Recall.

$$F1 - \text{ score } = \left(\frac{2}{\text{Recall }^{-1} + \text{ Precision }^{-1}}\right) \tag{4}$$

Eq.5 provide the Pixel Accuracy (PA) formulation, which measures the percentage of image pixels that are classified as burn wound correctly. n_{cl} represents the number of classes and t_i represents the total number of pixels for class i. n_{ii} represents the number of pixels which are classified correctly.

$$Pixel\ Accuracy = \frac{1}{n_{cl}} \sum_{i=1}^{n_{cl}} \frac{n_{ii}}{t_i} \tag{5}$$

Eq.6 provide the Mean Intersection-Over-Union (IOU), which is functioned like F1-score.

$$MeanIOU = \frac{1}{n_{cl}} \sum_{i=1}^{n_{cl}} \frac{n_{ii}}{t_i + \sum_{j=1}^{n_{cl}} n_{ji} - n_{ii}} \tag{6}$$

Step 6: System optimizing

For system optimization, overfitting is the most cases with a lot of researchers want to prevent. Regularization, Dropout layers, Capacity Reduction are three useful methods to increase the model's generalization ability. Regularization is to make punishment to the model if it kept predict same categories. Capacity Reduction is to remove some layers which would force the model to use each layer averagely. Dropout layers is a little same like Capacity Reduction, which make the layers stop participating forward pass.

Resources

In this section, the programming language, programming tool, and related open-source libraries are listed below (Table 1.):

Category	Name	Usage
Programming Language	Python	Build the foundation of the whole system
Programming Tool	Google Colab	Train the Neural Network model
Programming Tool	Jupyter Notebook	Test demos
Open-Source Library	Pytorch	Build the Neural Network structure
Open-Source Library	Pyplot	Make the visualization
Open-Source Library	Numpy	Calculate the array data structure
Open-Source Library	OpenCV	Transfer the image into matrix
Open-Source Library	OS	Create a checkpoint

Table 1.

Proposed schedule

The proposed schedule is shown below (Figure 7.). The schedule firstly implements former researchers' algorithm and builds some Deep Learning Networks to study its performance.

Then, combine with the former knowledge, do design thinking for how to make improvements to these algorithms. Conducting a lot of researches to justify the improvement ideas. After proving the new method is feasible, building the burn wound diagnosis system and making analysis on it. Finally, compile the report.

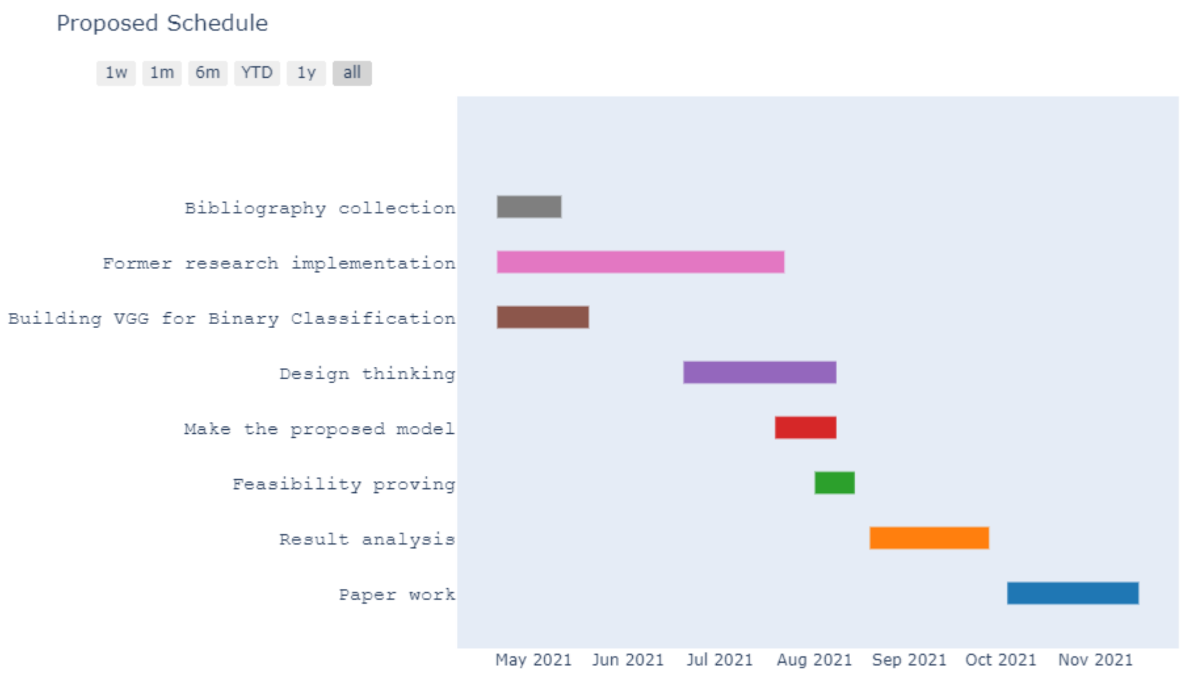

Figure 7. Proposed Schedule Gantt Chart

Researcher experience and qualifications

Being a master of data science student from the University of Queensland in the third semester, the knowledge for handling the difficulties should have been prepared. The knowledge of python programming (CSSE7030) is enough for building the burn wound diagnosis system. The understanding of machine learning (DATA7703) is helpful to learn the Deep Learning techniques. The practice of optimization (MATH7202) is essential to operate on such a programming-based application. The research of statistics (DATA7202) is important to figure out the stochastic data sampling. Additionally, I become a member of the Institute of Electrical and Electronics Engineers (IEEE) of Queensland Section which could help me establishing communications with other scholars of similar fields. In conclusion, I will synthesize the knowledge and communicate with different scholars to keep improving the project.

Cost

The project will not involve any financial cost, this section will introduce the weekly contribution on this project by the degree of hours.

The time allocation for the project in one week is 20 hours as shown in Figure 8., with 5% accounts for one hour. Moreover, time consumption for training the Deep Learning Models will be calculated separately in the following paragraph.

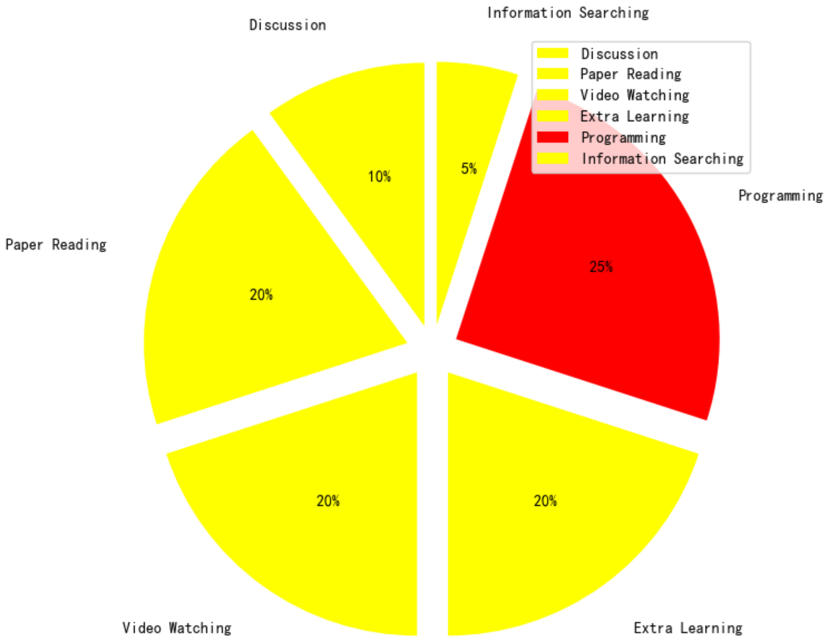

Figure 8. Time Consuming Pie Chart

The time consumption for training the Deep Learning Network could be dependent on the speed of Google Colab and the volume of the data. However, the networks' depth, number of filters, filter sizes, etc. could be more influential to the time consumption (He, Kaiming, & Sun, Jian, 2014). According and consulting to previous studies, the formulation of $O(\sum_{l=1}^{d} n_{l-1} \cdot s_l^2 \cdot n_l \cdot m_l^2)$, with l stands for the index of convolutional layers, m_l stands for the spatial size of the output feature map, d and n_l stands for the depth and width in the lth layer, and s_l stands for the length of the filter in the lth layer. The prediction shows the training process will take almost one hour.

Conclusion

Overall, the burn wound diagnosis system is still in the start-up stage. With less public dataset and the image itself is very private. Many types of research might be happening in a confidential way. Additionally, the lack of research papers to be guidance also make my group team and I try to find other related fields, such as skin cancer detection. However, with the step-by-step solution and schedule which I planned in the proposal, the object should finally be achieved at the end. Considering the huge convenience would bring to the hospitals in developing countries, the complexity of the work is acceptable.

Reference

Abubakar, A., Ugail, H., Smith, K. M., Bukar, A. M., & Elmahmudi, A. (2020). Burns Depth Assessment Using Deep Learning Features. *Journal of Medical and Biological Engineering*, *40*(6), 923–933. https://doi.org/10.1007/s40846-020-00574-z

Acha, B., Serrano, C., Fondon, I., & Gomez-Cia, T. (2013). Burn Depth Analysis Using Multidimensional Scaling Applied to Psychophysical Experiment Data. IEEE Transactions on Medical Imaging, 32(6), 1111–1120. https://doi.org/10.1109/tmi.2013.2254719

Automatic burn area identification in color images. (2016, June 1). Retrieved from https://ieeexplore.ieee.org/document/7528325/;jsessionid=fhXY-bW6Jd3k6qPpuHdcOlBque5gEewfBv-ImP7AXnFmowl_XvmU!-1231716854?tp=&arnumber=7528325

Badea, Mihai-Sorin, Vertan, Constantin, Florea, Corneliu, Florea, Laura, & Badoiu, Silviu. (2016). Automatic burn area identification in color images. 2016 International Conference on Communications (COMM), 2016, 65–68. https://doi.org/10.1109/ICComm.2016.7528325

Burn. (n.d.). In Wikipedia. Retrieved April 16, 2021, from https://en.wikipedia.org/wiki/Burn#cite_note-TBCChp4-6

Burn Triage and Treatment - Thermal Injuries - CHEMM. (2006). Retrieved from https://chemm.nlm.nih.gov/burns.htm#top

Burns. (2018, March 6). Retrieved from https://www.who.int/news-room/fact-sheets/detail/burns

Changhan Wang, Xinchen Yan, Smith, Max, Kochhar, Kanika, Rubin, Marcie, Warren, Stephen M, . . . Honglak Lee. (2015). A unified framework for automatic wound segmentation and analysis with deep convolutional neural networks. 2015 37th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC), 2015, 2415-2418.

He, Kaiming, & Sun, Jian. (2014). Convolutional Neural Networks at Constrained Time Cost.

Huang, G. (2017). CVPR 2017 Open Access Repository. Retrieved from https://openaccess.thecvf.com/content_cvpr_2017/html/Huang_Densely_Connected_Convolutional_CVPR_2017_paper.html

Jiao, C., Su, K., Xie, W., & Ye, Z. (2019). Burn image segmentation based on Mask Regions with Convolutional Neural Network deep learning framework: more accurate and more convenient. Burns & Trauma, 7, 1. https://doi.org/10.1186/s41038-018-0137-9

Kaiming He, Xiangyu Zhang, Shaoqing Ren, & Jian Sun. (2016). Deep Residual Learning for Image Recognition. 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2016, 770–778. https://doi.org/10.1109/CVPR.2016.90

Krizhevsky, A., Sutskever, I., & Hinton, G. E. (2017). ImageNet classification with deep convolutional neural networks. Communications of the ACM, 60(6), 84–90. https://doi.org/10.1145/3065386

Lecun, Y., Bottou, L., Bengio, Y., & Haffner, P. (1998). Gradient-based learning applied to document recognition. Proceedings of the IEEE, 86(11), 2278–2324. https://doi.org/10.1109/5.726791

Orion Despo, Serena Yeung, Jeffrey Jopling, Brian Pridgen, Cliff Sheckter, Sara Silberstein, Li Fei-Fei, Arnold Milstein. (2017). BURNED: Towards Efficient and Accurate Burn Prognosis Using Deep Learning. Retrieved from http://cs231n.stanford.edu/reports/2017/pdfs/507.pdf

Son Tran, Hai, Hoang Le, Thai, & Thanh Nguyen, Thuy. (2016). The Degree of Skin Burns Images Recognition using Convolutional Neural Network. Indian Journal of Science and Technology, 9(45). https://doi.org/10.17485/ijst/2016/v9i45/106772

Wang, Yuan, Ke, Zuo, He, Zhiyou, Chen, Xiang, Zhang, Yu, Xie, Peizhen, Li, Tao, Zhou, Jiao, Li, Fangfang, Yang, Canqun, Zhang, Pihong, Huang, Chun, & Kai, Lu. (2020). Real-time burn depth assessment using artificial networks: a large-scale, multicentre study. Burns, 46(8), 1829–1838. https://doi.org/10.1016/j.burns.2020.07.010

Wantanajittikul, K, Theera-Umpon, Nipon, Auephanwiriyakul, Sansanee, & Koanantakool, T. (2012). Automatic segmentation and degree identification in burn color images. The 4th 2011 Biomedical Engineering International Conference, 169–173. https://doi.org/10.1109/BMEiCon.2012.6172044

Zeiler, Matthew D, & Fergus, Rob. (2014). Visualizing and Understanding Convolutional Networks. Computer Vision – ECCV 2014, 8689(1), 818–833. https://doi.org/10.1007/978-3-319-10590-1_53

Zeiler, M D, Krishnan, D, Taylor, G W, & Fergus, R. (2010). Deconvolutional networks. 2010 IEEE Computer Society Conference on Computer Vision and Pattern Recognition, 2528–2535. https://doi.org/10.1109/CVPR.2010.5539957