Multi-Task Learning for HEp-2 image analysis:SE-Net50

Luigi D'Antonio, Immacolata Anna Morra, Diletta Vitale

***Abstract*—This paper examines the role of deep learning in HEp-2 image analysis, used for the detection of autoimmune diseases through indirect immunofluorescence. The implemented kind of learning is multi-task learning for single patch intensity classification, single patch pattern classification and single patch segmentation. The used net is U-Net, in an encoder decoder architecture in which the encoder is characterized with a backbone realized with SE-Net.**

I. Introduction

The immune system protects us from diseases and infections by attacking germs that get into our body, such as viruses and bacteria,destroying them. An autoimmune disease occurs when the bodys immune system begins to attack its own antigens. There are more than 80 types of autoimmune diseases. They can affect almost any part of the body. For example, alopecia areata is an autoimmune disease of the skin that causes hair loss. Autoimmune hepatitis affects the liver. In type 1 diabetes, the immune system attacks the pancreas. And in rheumatoid arthritis, the immune system can attack many parts of the body, including the joints, lungs, and eyes.[1] Antinuclear antibodies (ANAs) represent a class of autoantibodies against cellular components, including nuclear, cytoplasmic, cytoskeleton, and cyclin. It can be found in most autoimmune diseases (AIDs), and plays an important role in the diagnosis, classification, treatment, and disease activity monitoring of AIDs , such as Sjogren's syndrome (SS), systemic lupus erythematosus(SLE), and dermatomyositis (DM), etc.It is well-known that the indirect immunofluorescence assay, which detects antibodies directed against various antigens, is one of the most used techniques for investigating autoimmune diseases (AIDs). It consists of two important steps: first, binding of target antigen to specific primary antibodies in the diluted serum sample, and second, recognizing the antigen–antibody conjugates through fluorescein-labeled antihuman antibodies. For example, the indirect immunofluorescence assay (IIFA) on HEp-2 cells is widely used for detection of antinuclear antibodies (ANA).This study analyses six different kind of HEp-2 cells thought images and classifies them.[2]

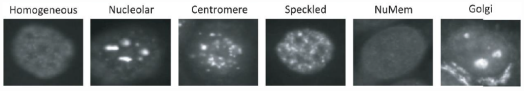

Fig. 1. Hep-2 cells

Acknowledgement: this template was made starting from the *Template for LNT Seminar Report* available on https://www.overleaf.com/

II. The problem

The problem subsists in three main challenges: image segmentation, classification on image intensity, and classification on cell pattern. All these can be solved by machine learning techniques. Image segmentation has been done by hand till now but there are several challenges related to it. Cell pattern analysis requires the coverage of all pixels used for segmentation and classification cannot be performed at one time alone. Therefore, it had to be divided into two phases – extraction of cells from an image and classification of those cells. There are numerous techniques for segmenting images, including: [3]

- Thresholding: In this technique, the intensity of each pixel in the image is given a threshold value, and pixels that are above or below the threshold are divided into distinct classes. Using a seed point in the image as a starting point, this technique iteratively adds nearby pixels that fall into the same class according to certain criteria, such as intensity or hue.
- Watershed: Using this technique, local minima are used to depict basins and the picture is treated as a topographic map. The segments are then generated by filling the basins with various colors to create a gradient representation of the original image.
- Graph-based techniques: In these techniques, the picture is represented as a graph, with pixels acting as nodes and edges joining nearby pixels. The picture is then divided into segments by dividing the graph into linked parts.
- Machine learning-based techniques: In these techniques, a model is trained using a dataset of labeled pictures in order to discover the characteristics and patterns that distinguish the various classes. Then, fresh photos may be segmented using the model. Not to mention, there are also more sophisticated machine learning methods like deep learning.

Due to its capacity to automatically learn and extract characteristics from the data, deep learning in particular has gained popularity as a method for segmenting medical images. In general, the features of the picture and the intended output determine which image segmentation technique should be used. [4] The process of classifying a medical image on the basis of its visual content is known as medical image classification. This kind of machine learning is supervised, so the model is trained on a labeled dataset where the appropriate class for each image is given. Medical picture classification can be done in a variety of ways, including traditionally using hand-crafted feature extraction and support vector machines (SVMs) and

more recently using convolutional neural networks (CNNs). [5] Because of their previously mentioned capacity to automatically learn and extract features from the data, CNNs have grown particularly popular also for what regarding the classification of medical images. Convolutional layers are used to achieve this, applying several filters to the input image to extract features at various scales and locations. Fully connected layers then get these features and utilize them to assign the image to one of the predetermined classes. The lack of readily available annotated training data is one of the major difficulties in medical picture categorization. The amount and diversity of the training dataset may be constrained by the fact that manually annotating images in medical imaging is frequently expensive and time-consuming. Transfer learning, which allows a model that has already been trained on a big dataset to be fine-tuned on a smaller medical imaging dataset, has gained popularity as a solution to this problem. The significant degree of diversity in image appearance, which can be attributed to elements including patient anatomy, imaging modality, and scanner parameters, presents another difficulty in the classification of medical images. Utilizing a broad training dataset that accurately captures the diversity in the test dataset is crucial to overcoming this difficulty. The automatic classification of images based on their visual content, which can help with the diagnosis and treatment of many medical diseases, makes medical image classification a significant task in medical imaging.

III. The solution

To fullfill the task a convolutional network based on Unet architecture, but with the addiction of a Se-Net50 backbone to the encoder, was used. U-Net is a popular deep learning architecture for image segmentation. It was originally developed for biomedical image segmentation, but has since been used in a variety of other domains [6]. The U-Net architecture is named for its U-shaped structure, which consists of an encoder and a decoder. The encoder is responsible for extracting features from the input image, while the decoder maps the features back to the original image resolution and generates a per-pixel prediction of the class label. The U-Net architecture is trained with two auxiliary tasks: label classification and intensity classification. The label classification task involves predicting the class label for each pixel in the image, while the intensity classification task involves predicting the intensity value for each pixel. During training, the U-Net architecture is able to learn features that are useful for both the main segmentation task and the auxiliary tasks. This can improve the overall performance of the model, as the auxiliary tasks provide additional supervision that can help the model learn more discriminative features. The backbone was bone using a SENet50 network. With “Squeeze-and-Excitation” (SE) block that adaptively recalibrates channel-wise feature responses by explicitly modelling interdependencies between channels, SENet is constructed. [7] The squeezing step performs a average pooling at each channel to create a 1x1 squeezed representation of the volume U. Instead for excitation The authors introduce a new parameter called the reduction ratio r, to introduce a first fully connected (FC) layer with a ReLU activation, before the gating network with the sigmoid activation.The reason to do this is to introduce a bottleneck that allows us to reduce the dimensionality at the same time that introduce new non-linearities. Furthermore, we can have better control on the model complexity and aid the generalization property of the network. Having two FC layers will result on having 2 matrices of weights that will be learned by the network during the training in an end-to-end fashion (all of them are backpropagated together with the convolutional kernels). The whole network has the following architecture The authors introduce a new parameter called the reduction ratio r, to introduce a first fully connected (FC) layer with a ReLU activation, before the gating network with the sigmoid activation.[8] The network has the following architecture.

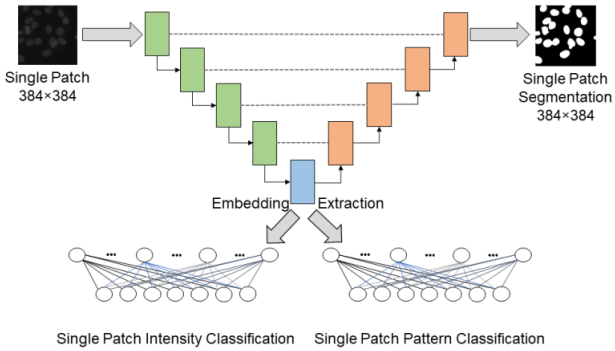

Fig. 2. Network architecture

The architecture is trained using a combination of the main segmentation loss and the auxiliary task losses.A model can be trained to carry out several tasks at once using the multitask loss function. With an optional weighting factor for each loss term, it is described as the total of the losses for each task. The definition of the multitask loss function is: [9]

$$Loss = \sum (w_i * loss_i) \tag{1}$$

where loss i is the loss for task I and w i is the weighting factor for the loss term for task i. In the multitask learning problem, the weighting factors can be used to modify the relative importance of each task.

- Binary cross entropy loss: The binary cross entropy loss is used for binary classification tasks, where the goal is to predict a single binary label (e.g., ”positive” or ”negative”). This loss function is defined as: [10]

  $$Loss = -(y * log(p) + (1 - y) * log(1 - p)) \tag{2}$$

  Where y is the true label (either 0 or 1) and p is the predicted probability of the label being 1. The loss function is minimized when the predicted probability p is close to the true label y. For example, if y is 1, the loss will be minimized when p is close to 1, and if y is 0, the loss will be minimized when p is close to 0.
- Cross entropy loss: The cross entropy loss is a generalization of the binary cross entropy loss to multi-class classification tasks, where the goal is to predict one of

several possible labels. This loss function is defined as: [10]

$$Loss = \sum(y_i * log(p_i)) \quad (3)$$

Where y_i is the true label (either a one-hot encoded vector or a class label) and p_i is the predicted probability of the label being i. The loss function is minimized when the predicted probability is close to the true label y_i. For example, if y_i is 1 and p_i is the predicted probability of the label being i, the loss will be minimized when p_i is close to 1.

- Dice loss: The dice loss is used for semantic segmentation tasks, where the goal is to predict the class label for each pixel in an image. This loss function is defined as: [11]

$$Loss = \ 1 - 2 * (\sum(y_i) * p_i)/(\sum(y_i) + p_i) \quad (4)$$

Where y_i is the true label for pixel i and p_i is the predicted probability of the label being i. The dice loss is commonly used in medical image analysis to evaluate the performance of segmentation models. The loss function is minimized when the predicted probabilities p_i are close to the true labels y_i. For example, if y_i is 1 and p_i is the predicted probability of the label being i, the loss will be minimized when p_i is close to 1.

In this instance, multitask loss was combined with binary cross entropy for segmentation, cross entropy for pattern classification, and binary cross entropy once more for intensity classification. Following that, dice loss was used to segment the data instead, and then the outcomes from the various triads were compared.

To balance the multitask loss gradient normalization is used. GradNorm is a technique for training deep neural networks that aims to improve the stability and generalization performance of the model. It works by normalizing the gradients of the model's parameters during training, in order to ensure that they have a consistent scale and direction. In deep neural networks, the gradients of the model's parameters can have a wide range of values, which can make it difficult for the model to learn effectively. This is because the optimization algorithm, such as Adam, may update the model's parameters based on the scale of the gradients. If the gradients have a large scale, the model's parameters may be updated too aggressively, leading to unstable and suboptimal learning. On the other hand, if the gradients have a small scale, the model's parameters may be updated too slowly, leading to slow convergence.GradNorm addresses this issue by normalizing the gradients of the model's parameters during training, ensuring that they have a consistent scale and direction. This allows the optimization algorithm to update the model's parameters more consistently and effectively, leading to improved stability and generalization performance. To implement GradNorm in a deep neural network, it is necessary to modify the training process to include a gradient normalization step. This can be done by dividing the gradients of the model's parameters by their L2 norm before applying the updates. Overall, GradNorm is a useful technique for improving the stability and generalization performance of deep neural networks, by normalizing the gradients of the model's parameters during training.

IV. Results

Unfortunately until now this are our only acceptable results:

Fig. 3. Segmentation task using Binary Cross Entropy as loss

Fig. 4. Segmentation task using Binary Cross Entropy as loss

It is possible to find the code here https://github.com/dilettavitale/Medical-Imaging-group-3/tree/main

References

[1] Autoimmune Disease https://medlineplus.gov/autoimmunediseases.html
[2] Alireza Ghanadan,Amene Saghazadeh,Issa Jahanzad and Nima Rezaei. Clinical aspects of indirect immunofluorescence for autoimmune diseases https://www.tandfonline.com/doi/abs/10.1586/1744666X.2015.1027152
[3] Image segmentation Wikipedia https://en.wikipedia.org/wiki/Image_segmentation
[4] Dan Luo, Wei Zeng,Jinlong Chen and Wei Tang.Deep Learning for Automatic Image Segmentation in Stomatology and Its Clinical Application https://www.frontiersin.org/articles/10.3389/fmedt.2021.767836/full.
[5] ZhiFei Lai and HuiFang Deng. Medical Image Classification Based on Deep Features Extracted by Deep Model and Statistic Feature Fusion with Multilayer Perceptron https://www.hindawi.com/journals/cin/2018/2061516/
[6] U-Net Wikipedia https://en.wikipedia.org/wiki/U-Net
[7] Sik-Ho Tsang.Review: SENet Squeeze-and-Excitation Network, Winner of ILSVRC 2017 (Image Classification)
[8] Pablo Ruiz.Understanding and visualizing SE-Nets
[9] Alex Kendall. Yarin Gal.Roberto Cipolla. Multi-Task Learning Using Uncertainty to Weigh Losses for Scene Geometry and Semantics. https://openaccess.thecvf.com/content_cvpr_2018/papers/Kendall_Multi-Task_Learning_Using_CVPR_2018_paper.pdf
[10] A Detailed Guide to 7 Loss Functions for Machine Learning Algorithms with Python Code. https://www.analyticsvidhya.com/blog/2019/08/detailed-guide-7-loss-functions-machine-learning-python-code/
[11] Michael Yeung. Evis Sala. Carola-Bibiane Sch¨onlieb. Leonardo Rundo. Unified Focal loss: Generalising Dice and cross entropy-based losses to handle class imbalanced medical image segmentation. https://arxiv.org/pdf/2102.04525.pdf