Lilian Sao de Rivera

Individual Report

Final Project 05/01/2020

Prof. Amir Jafari PhD

1.Introduction

This project aims to explore the use of neural networks in computer vision and by doing so, participate in the field of medicine and maybe with time and a lot of work help to create these types of scenarios in the places where help is needed.

The problem was presented by a Kaggle competition on November 2019. The aim of the competition was to detect the type of intercranial tumor in MRI images. There are five type of tumors: Intraparenchymal, Intraventicular, Subarahnoid, Subdural, and Epidural. The aim of the solution is to detect if an image presents a tumor and what type of tumor it is. There are several problems associated with the images. The first one is that the tumors are spread inside the brain in very particular ways, thus sometimes is difficult to distinguish what type of case we have at hand. Other problem is that the tumors are blood clots that are difficult to separate from the brain matter. The images that we are using are monochrome images that need to be modified to clearly identify the different densities in the brain.

2. Description of your individual work. Provide some background information on the development of the algorithm and include necessary equations and figures.

I was in charge of designing the architecture of the problem, as computer science major my main goal is to get the main components design and the computational structure ready to perfume in the teams that I work for. I also was in charge of the research and the implementation of the STN. My background allows me to learn new frameworks and software, so It was the perfect job me.

Model 1, Model 3 , STN, Labels_pr, Model3_testdataset

View_Transformation3

FPModel1

3. Describe the portion of the work that you did on the project in detail. It can be figures, codes, explanation, pre-processing, training, etc.

I created the templates for the training, testing and models, I fully developed the Model I, the data loaders and developed the presentation in Shiny. This was not necessary, but I always find rewarding to learn an implement new tools in my leaning experiences. Having Shiny as a information tool made me realize how important are dashboards to look for clues in the information and by implementing them I can design new controls and metrics of information that I want to see or would like to see. Sometimes only to fulfill my curiosity.

4. Results. Describe the results of your experiments, using figures and tables wherever possible. Include all results (including all figures and tables) in the main body of the report, not in appendices. Provide an

explanation of each figure and table that you include. Your discussions in this section will be the most important part of the report.

- Batches of 15 or 10 images yield better results.
- The prototype shows two models are a good approach
- The use of the function loss in the validation phase as a tool to choose the best model might not best the best option
- The accuracy and other metrics are necessary to evaluate the model.
- The number of images use in “Any” category shows a good metric for sampling the other classes for the second phase.
- It is important to include feature enhancing in the model
- The prototype allowed us to build a technical working infrastructure to continue with the second phase which includes:
    - Data Augmentation
    - Sampling
    - Fine Tuning the models
    - Include Inception V3.

Results

From the graph below we can see that the first model shows declining leaning process, whereas the validation phase shows a different behavior. These results are from the first phase where the sample is very small. The behavior of model II is mor stable in the validation, which means that having two models for the process shows stability to the process.

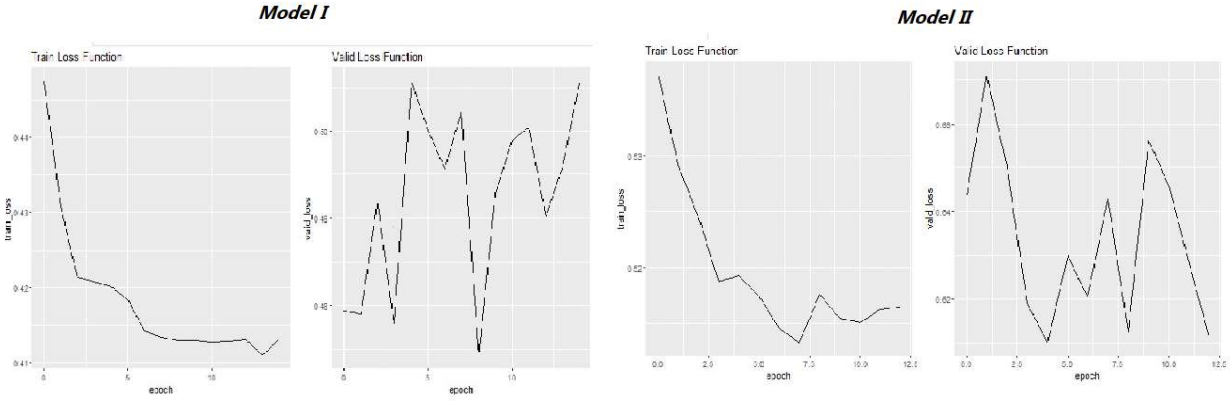


Figure 1 Result Val Loss Train Loss

The AUC in model I in the first phase shows an a good result for the “any” levels, whereas the other labels are very close to the chance lines. You can see these trends in the first graph below. Fig.8. But the AUC on the left shows improvement. That means that what we learn in the first model makes a huge difference in the second model.

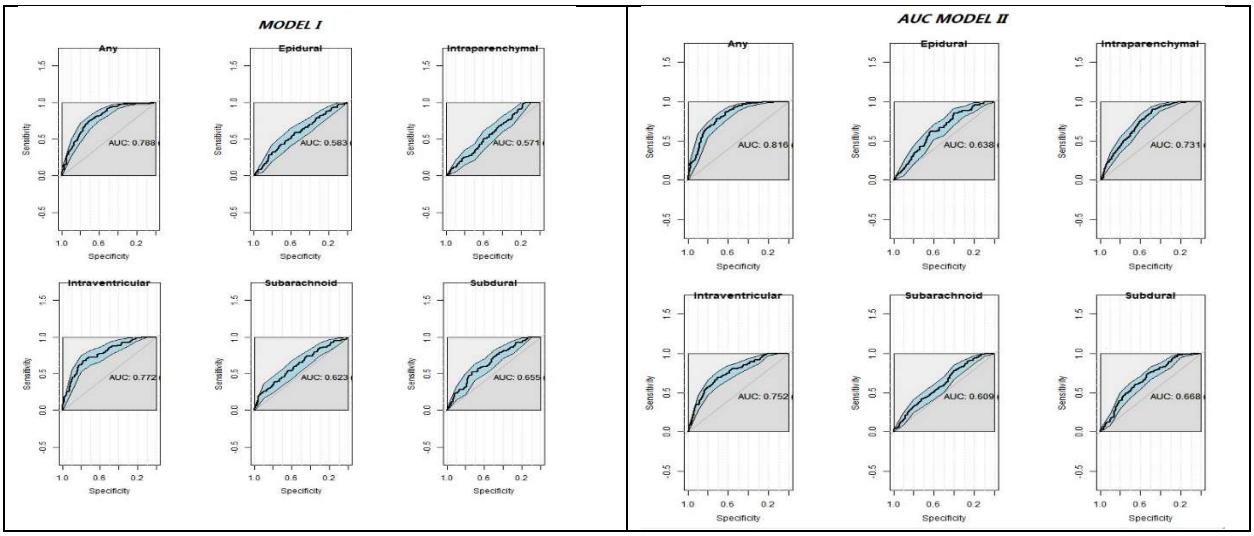


Figure 2 AUC For Model I and Model II in Phase I

The results for phase two shows a better AUC in general in model I and Model II

For Model I we got the following results

Test set: Loss: 0.382323

Accuracy : 13.741497

AUC : 0.723737

f1 : 0.177195

For Model II we got the following results

Test set: Loss: 0.468402

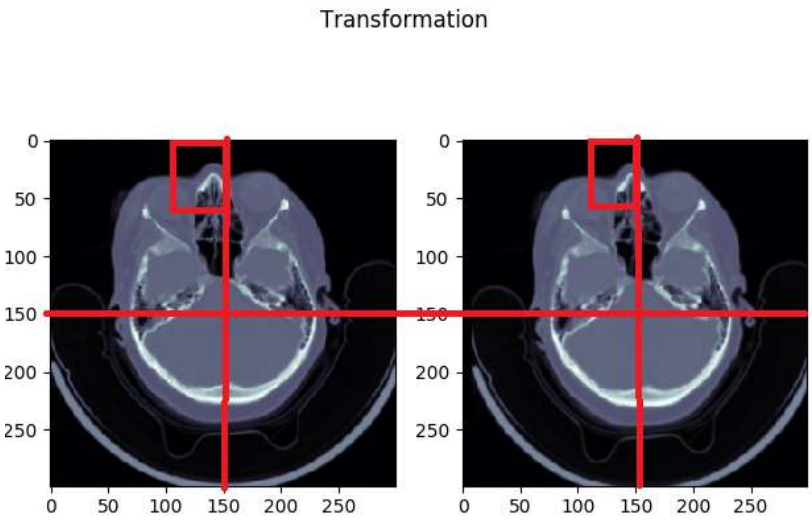
Accuracy : 16.244898

****AUC : 0.748952**

****f1 : 0.425590**

We can see in the second model AUC general of 0.74 whereas an F1 of 0.42 for 19 epochs with three thousand images per class. What we are observing in this experiment is that our model can improve greatly with more time to process and more images.

Another success that we obtain from this experiment was the transformation, the images were transformed during the process to a better position. We can see from the figure below that image on the left shows the nose bone to the left , whereas in the image of the right is aligned to the center. The transformed image on the right shows a shape more centered in the grid, the canonical position. This result was obtained in the second phase.



This was my most rewarding experience doing the project, to center this image by using the SNT Network, I invest most of my time learning and reading the papers to archieve this result.

Conclusions

- The second phase exhibited a significant improvement in terms of valid_loss, train_loss, validation loss and accuracy.
- AUC and F1 score were included in the second phase with an average of 70% and 17%d. F1 still does not yield a good score, however the images can be increased and more epochs can be included in the training.
- The substitution of VGG16 for Incenption_V3 led to this significant improvement of the results.
- The inclusion of AUC and F1 score made evident the behavior of the training process.
- Each epoch lasted about half an hour for the fist model and twenty minutes for the second model, which shows that we need to budget more computer processing time for more images.

Recommendations

- More time in training is necessary and more images which will require more time, so the planning for the training is important.
- Feature enhance only included using Hounsfield scale for the images, using brain window for channel two, and epidural window for channel 3, however we have seen studies posted of competition of Kaggle that using an RGB color space for each window might help the model to enhance the features. This would be an interesting experiment to try.
- The study from A. Esteva, B, Kuperl, et al. (2017) shows three stages of evaluation, we only use two, thus trying this approach will be also an interesting experiment to try on our dataset.

Lessons Learned

- Design NN requires a great deal of research
- It is better to sit and read before you start designing something
- You should take time to try on a small set and then continue with the second phase

- Computer Vision is a lot of work, but very interesting, I feel like a just scratch the surface but at the same time I am confident that I can continue on my own but It will require a year of experimentation to really understand the concepts and deliver something meaningfull

6. Calculate the percentage of the code that you found or copied from the internet. For example, if you used 50 lines of code from the internet and then you modified 10

Data Loaders

100 Lines and added 100

Reference : https://stanford.edu/~shervine/blog/pytorch-how-to-generate-data-parallel

SNL

I added the function that gets theta, and added the additional layers

https://pytorch.org/tutorials/intermediate/spatial_transformer_tutorial.html

All the other networks and and tricks I got it from blogs but I didn't really use the code just as examples on how to create neworks, that are commands of pytorch and its standard use

The percentage that I use of code I would say that was approximately 25% percent between the dataloders and the SNL. All the other work are standar loops.