Group 9: Review

October 19, 2019

1 Summary

In this report, the authors run experiments to compare the efficacy of feature extraction from deep CNNs and classification using Logistic Regression, Support Vector Machines and Linear Discriminant Analysis with feature extraction using scattering nets and using the aforementioned models for hand written digit recognition (on the MNIST dataset). They compare the final accuracy achieved by each model ensemble and discuss the differences in performance.

In addition, the authors follow similar methodologies to identify forgeries on a dataset of Raphael's paintings. An analysis similar to the previous one is performed.

2 Strengths

The analysis can generally be considered sound with the authors making sure that they cover possible ground in terms of considering different models with different strengths for classification, different CNNs with different properties for feature extraction. The authors generally make good use of existing research to argue performance differences between the various models considered.

The figures showing architectures of scattering net, ResNet help the reader's understanding of the content. The visualizations of low dimensional feature vectors helps with their arguments regarding performance differences between models.

3 Weaknesses

In general while the authors have made efforts to be as thorough as is practically possible for a graduate level course report, there are some weaknesses that are apparent. In particular, the report has numerous grammatical errors and use of complicated but incoherent sentences that sometimes affect the fluidity of the report. Other technical limitations are too large to list in this section and will be addressed in the section pertaining to technical evaluation of the writing.

4 Evaluation on Clarity and quality of writing

Overall: 3/5 (average)

While the conclusions of the report are easy to observe, occasionally in the report, one notices incoherent grammar and typos, such as in section 2.3: "After determine the considerable direction, thresholds of different classes can be dissolved," in section 1: "... and the second one is a convenlutional neural network with ..." and the use of the word "database" when the authors are referring to a "dataset."

Apart from this, while presenting and arguing about the results, there is uncertain language which does not motivate confidence in the results and just geneerally reduces the quality of writing. For example, in section 2.5, "...features extracted by neural networks (...) are adjusted according to the image and its label which can be somehow considered as supervised feature extraction". Not only would removing "somehow" from sentences like these help with the conviction of the statement, but also it would be best to shorten sentences like these that do not add much power to the final conclusion.

5 Evaluation on Technical Quality

Overall: 3.5/5 (good but with errors)

Generally speaking, the report is technically sound with few apparent flaws. There are numerous, thorough experiments that are run which inspire great confidence in the final conclusion. However, there are also a few things that should be noted:

- While loading MNIST dataset, it is not exactly clear if the input dataset is being correctly normalized and resized before feeding into pretrained networks. In the file `load_data.py`, in the funnction `get_dataloader()`, we see that the normalization used on the dataset is `(0.5,), (0.25,)`. However, for pre-trained networks trained on ImageNet, there is a fixed set of summary statistics that should be used to normalize the input data. Also the input images should be resized at `224 x 224 x 3` which is the size of the ImageNet dataset images.
- In section 2.6, paragraph 1, the authors try to provide an intuitive explanation of why random forests do not perform well on the classification problem with extracted features. The authors, in my opinion, make a few inaccurate statements:
  - "Note that random forest randomly selects on pixel (variable) of the image as its filter". This should instead say that random forest selects a subset of features from an input feature vector/tensor. It is true that one can make it such that at each iteration the random forest algorithm will choose just 1 feature, but that is not apparent in the implementation in `classification.py`.
  - Also, it is not true that once an "important" feature is not chosen while growing a particular random tree, that would greatly affect the model as a

whole. While there would be some effect, the idea of a random forest is to grow a large number of random trees with a large amount of variance across those trees. Presumably, at least one random tree in the forest will be able to capture the pattern.

– "But other classification methods would somehow take the weights of pixels, which measure the impact of the pixels, into consideration". It is an unsubstantiated claim that a random forest will not be able to do this. A random forest classifier would implicitly assign weights to certain features by the number of random trees that vote in the favor of the output (in a classification setting, which is what we are talking about here)

6 Overall rating

Overall: 3.5/5 (average)

Overall, the report has thorough experiments but the quality is marred by unsubstantiated claims and grammatical errors.

7 Confidence on assessment

Overall: 3

I have read the report carefully and perused the attached code. However, I didn't run the code myself but I feel I have a good understanding of the approach.