MLP Coursework 3: Project Interim Report

s1, s2, s3

Abstract

The abstract should be 100–200 words long, providing a concise summary of the contents of your report. **still needs to be written.**

1. Introduction and Motivation

During the last six years there has been an increase in popularity of connectionist based approaches (specifically, variations on deep neural network architectures) to solving vision based pattern recognition problems (Dinsmore, 2014). This surge in popularity has been the result of numerous advancements, including an increase in the amount of available training data and compute power, as is demonstrated in (Krizhevsky et al., 2012) (Szegedy et al., 2015) (Simonyan & Zisserman, 2014). Although deep neural networks work well with large amounts of training data (lot), the performance of these models typically drops off in situations where only small amounts of training data is available [INSERT REFERENCE TO LEARNING CURVE HERE]. This poses a problem to smaller businesses and organisations that may not have the appropriate amount of data to utilise these emergent technologies. Such a problem motivates the investigation presented within this (and the forthcoming) report that examines the application of fruitful techniques (specifically, transfer learning) to boost the performance of deep neural network architectures using small datasets.

In order to address this problem, there exists a number of data manipulation (non-machine learning) methods that have been proposed already, such as (Hu et al., 2018) and other more naive approaches such as data augmentation (Krizhevsky et al., 2012). However, addressing this problem using novel techniques within the domain of deep neural networks has only recently been explored. Within this report we aim to utilise such techniques that have already been initially established within the literature, such as transfer learning [1] (Oquab et al., 2014).

In order to formalise the problem presented within this report (using small data on deep neural network architectures), an investigation into the effects of dramatically reducing the training set size is presented. The utilised methodology includes performing an observation of the proposed network accuracy as the training sample size is decreased (4). Furthermore, this procedure is applied to two comparable datasets, one with subtle differences between classes and the other with apparent differences between classes. We hypothesise that classes with subtle differences between them will typically score lower on a shallow network than those with obvious differences between them (2.4). In short, a comparison of the proposed datasets (3) may provide a means of identifying relationships between aspects of the proposed datasets and the employed network architecture.

Within the remainder of this paper a set of research questions and associated hypotheses is initially presented 2. After which, an overview of the selected datasets and the task is documented (3). Subsequently, the methodology employed to address the aforementioned research questions and hypotheses are outlined (4) and experimental results are documented (5). The experimental results are then draw upon to derive a set of initial conclusions (6). Finally, details of any associated risks, backup plans and further work are provided (7).

2. Research Questions

Within this section two sets of research questions are presented. Firstly, a set of research questions that are addressed within this report is provided (2.1). Secondly, a set of future research questions to be addressed within the concluding report is offered (2.2). Thereafter, the aims and objectives of the proposed research questions are probed (2.3). Finally, a set of hypotheses is given (2.4).

2.1. Interim Research Questions

Using the methodology outlined within 4, the following research questions are investigated within this interim report:

1. How do differences in similarity between datasets affect the performance (generalisation, accuracy and error) of the proposed convolutional neural network architecture (4.1)?
2. How does reducing the size of a training dataset affect the performance of fairly standard convolutional neural network architectures?

Within the first research question, it is assumed that visual similarity is easily identifiable by humans. Throughout our research, images that only contain variation of facial expression are considered to have maximal similarity. Conversely, images of distinctively different objects are considered to have minimal similarity. Although a similarity metric between instances of data (and datasets) is not pro-

[1] Transfer learning has already been applied to our specific use case, however this particular technique has other use cases that are not necessarily explored within this paper.

posed within this report, this topic is touched upon within our discussion of potential future work (7).

Furthermore, the proposed neural network architecture (4.1) is assumed to be fairly standard. Therefore we assume a degree of generalisation to similar problem domains in any further research. However, more work may need to be undertaken in order to validate this (7).

It is commonly thought that smaller datasets lead to poor generalisation (lot) (Krizhevsky et al., 2012). However, the second research question links into our future research questions (2.2) associated with using techniques [2] to improve the performance of small datasets on neural network architectures. For this reason, it may be important that an investigation is conducted in order to establish a more conclusive understanding of the proposed datasets (3) on our employed architecture (4.1).

2.2. Future Research Questions

Within the last section, research questions associated with this interim report were presented. The future research questions we intend to address within the concluding report are outlined below.

1. How does the application of transfer learning affect the performance of the proposed neural network architecture? [3]
2. How does the application of transfer learning affect the performance of the proposed neural network architecture when the size of the dataset used to tune the network is greatly reduced?
3. If time permits, how does one shot learning (the use of siamese network architectures (Bromley et al., 1994)(osl)) perform on small datasets within classification tasks?
4. If time permits, how can deep feature extraction be used in order to improve the performance of small datasets on deep neural network architectures?

2.3. Aims and Objectives

The core objective of the concluding report is to investigate connectionist based methodologies for improving classification performance on vision based tasks using small datasets. Initially, this investigation will be addressed by obtaining a set of baseline classification accuracies using a shallow convolutional neural network architecture. Baseline accuracies will be obtained for training dataset sizes of: 100%, 75%, 50%, 25%, 10%, 1%.

As an optional objective, the interim report (in conjunction with the concluding report) aims to investigate how subtlety between different classes (given the same sized dataset) affects the performance of the proposed network architecture (4.1).

[2] Such as transfer learning and one shot learning.

[3] This will be tested using the same classification experiments performed during the interim report.

2.4. Hypotheses

To conclude this section regarding the intended research questions, a set of hypotheses is provided:

H.1 Datasets with subtle differences between classes will perform worse on classification tasks than datasets with obvious differences between classes using the proposed convolution neural network architecture (4.1).

H.2 Reducing the size of the training dataset will result in worse generalisation using the proposed network architecture (4.1).

H.3 Reducing the size of the training dataset will result in an overfitting of the network architecture to the subsampled dataset.

H.4 The application of transfer learning using a pre-trained model (any ImageNet variant) will result in improved model performance.

3. Data Set and Task

TODO: chaining paragraph linking 2 -> 3.

- https://www.kaggle.com/zalando-research/fashionmnist (Clothes database): took five classes with 6000 examples in each - 30000 examples in total.
- https://grail.cs.washington.edu/projects/deepexpr/ferg-db.html (Facial expression database): had to subsample the dataset, remove a lot of data and sample 30000 examples using a balanced set of classes (6000 examples in each).

3.1. Preprocessing

For the aforementioned datasets, each image is scaled up or scaled down to the same dimensionality. This scaling allows for the input vector to be the same size regardless of employed dataset. Images from the faces dataset were scaled down from 512x512 to 64x64 (x8 down) and images from the clothes database were scaled up from 32x32 to 64x64 (x2 up). Furthermore, the proposed network architecture was trained using the original examples in addition to examples where data augmentation had been applied to each instance.

3.2. Evaluation

Each experiment was evaluated by measuring the loss function (the model error), but primarily the evaluation was done through measuring the accuracy whilst employing cross-validation. This cross-validation was implemented within the SKLearn library as a function called: $\text{train}_{t}est_{s}plit$.

4. Methodology

In the first phrase of our project, the interim research questions (2.1) are first examined to create a baseline system for further work in transfer learning. Both of the two image databases (3) are subjected to pre-processing (3.1) before using the result as inputs into the proposed neural network architecture **??**. In the facial expression dataset, original png files with 256x256 pixels are downscaled to 28x28 pixels so as to be comparable with the default pixels in clothes dataset. To understand the effect of the subtle and obvious feature differences between classes on performances (prediction accuracy, error) for distinctive tasks, 30k images from both the clothes dataset and the facial dataset are evaluated on convolutional neural network respectively by performing multi-class classification tasks.

Is this part of our methodology? Or does this belong to the data side of things?

4.1. Proposed Neural Network Architecture

The proposed architecture consists of three convolutional layers with one max-pooling ReLU layer in between. The final layer is then flattened to produce one numerical output with categorical cross-entropy as a loss function (maybe add one more softmax layer before flattening to increase stability as suggested by MLP lecture?). For the optimiser, Adam or RMSprop would be used. Weight and bias is also initialised using (gloro-bengio ini. ?, random ?).

After inspecting the results from first experiment. Same task is performed on much smaller dataset to investigate the discrepancy of size of dataset on classification performance. (1000 dataset maybe?). By implementing the two experiments, baseline systems could be set up to investigate possible strategies to perform prediction/classification task given very small dataset which is the main goal of our project.

In the second phrase of the project, two different transfer learning methods will be studied to examine potential methods to improve performances given very small dataset which is frequent in real-world scenario. Firstly, we transfer a very large pre-trained network VGG16 on our aforementioned baseline system with pre-trained weights on small dataset. Since VGG16 trains on 200 types of general objects. The generality of the model might be beneficial to train on common objects (clothes dataset). Apart from transferring model to domain-specific dataset (clothes dataset). We also transfer the model to dataset with unrelated and subtle differences between classes in the dataset (facial dataset), to test the effectiveness of pre-trained model on task that shares little similarity with the pre-trained model.

Besides transferring pre-trained model, we also wish to investigate the effect of one-shot learning on small dataset. To demonstrate a basic version of one-shot learning we will implement Siamese network on either one of the dataset(clothes/facial expression) with the help of existing models and our modification to these models , due to time constrain and taking potential difficulty of implementing one-shot learning architecture from scratch. As a backup plan, we will abandon this experiment and focus more on transferring models methods.

- Input Layer (are we going to pre-process the input data, such that the input layer is the same for both datasets? i.e: make the images the same dimensionality?)
- Three convolutional layers.
- Do we want to use batch norm, drop out, etc?
- Softmax output layer for multi-class classification.
- Probably a good idea to include a diagram of the architecture.

5. Baseline Experiments

TODO

5.1. Further Experiments

TODO

6. Interim Conclusions

TODO

7. Future Work

TODO

7.1. Backup Plans

TODO

References

Andrew ng: Why âĂŸdeep learningâĂŹ is a mandate for humans, not just machines. https://www.wired.com/brandlab/2015/05/andrew-ng-deep-learning-mandate-humans-not-just-machines/. Accessed: 2015-05-01.

One shot learning with siamese networks in pytorch. https://hackernoon.com/one-shot-learning-with-siamese-networks-in-pytorch-8ddaab10340e. Accessed: 2017-07-15.

Bromley, Jane, Guyon, Isabelle, LeCun, Yann, Säckinger, Eduard, and Shah, Roopak. Signature verification using a" siamese" time delay neural network. In *Advances in Neural Information Processing Systems*, pp. 737–744, 1994.

Dinsmore, John. *The symbolic and connectionist paradigms: closing the gap*. Psychology Press, 2014.

Hu, Guosheng, Peng, Xiaojiang, Yang, Yongxin, Hospedales, Timothy M, and Verbeek, Jakob. Frankenstein: Learning deep face representations using small

data. *IEEE Transactions on Image Processing*, 27(1): 293–303, 2018.

Krizhevsky, Alex, Sutskever, Ilya, and Hinton, Geoffrey E. Imagenet classification with deep convolutional neural networks. In *Advances in neural information processing systems*, pp. 1097–1105, 2012.

Oquab, Maxime, Bottou, Leon, Laptev, Ivan, and Sivic, Josef. Learning and transferring mid-level image representations using convolutional neural networks. In *Computer Vision and Pattern Recognition (CVPR), 2014 IEEE Conference on*, pp. 1717–1724. IEEE, 2014.

Simonyan, Karen and Zisserman, Andrew. Very deep convolutional networks for large-scale image recognition. *arXiv preprint arXiv:1409.1556*, 2014.

Szegedy, Christian, Liu, Wei, Jia, Yangqing, Sermanet, Pierre, Reed, Scott, Anguelov, Dragomir, Erhan, Dumitru, Vanhoucke, Vincent, and Rabinovich, Andrew. Going deeper with convolutions. In *The IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, June 2015.