Image Classification Transfer Learning

Domain Background

Computer vision is one of the fields where machine learning has made great strides in this decade. It involves gaining high-level understanding from digital images or videos. It can be thought of as an application of artificial intelligence to automate the tasks that the human visual system can do. Computer vision itself has many sub-domains depending on the objective one is trying to achieve. Some of these are scene reconstruction, event detection, object recognition, motion estimation, image restoration,[1] etc. Object recognition can also be called object or image classification and the one we shall focus on here is with many types of images with labels and hence is a case of multi-label classification.

Image classification is important as it helps in other areas of computer vision like object detection and localization used in applications like self driving cars. Large amounts of effort have gone into research of this area and huge datasets have been collected for benchmark to measure the progress. The breakthrough happened with the introduction of Convolution Neural Networks(CNNs) for image classification task and LeNet architecture[2] was applied to classify MNIST dataset. More bigger datasets like ImageNet[3] came after that and the architectures applied on this benchmark dataset changed the course of AI. Some state of the art CNN architectures used on ImageNet are AlexNet[4], ZF Net[5], VGG Net[6], GoogLeNet[7], ResNet[8], etc. The history shows the importance and efforts put in the task of image classification.

Problem Statement

Given a dataset of images with multiple categories, train a model to classify the images into their respective categories. The problem is best resolved by applying CNN but finding a good architecture is the challenge. The problem is quantifiable and measurable in the sense that we can we can attribute performance of the model trained in terms of the correct predictions of class it can make.

Datasets and Inputs

CIFAR-100[9] is a labeled subset of the 80 million tiny images dataset[10] and was collected by Alex Krizhevsky, Vinod Nair, and Geoffrey Hinton. The dataset consists of 60000 32x32 colour images of which 50000 are training images and 10000 are test images. CIFAR-100 has 100 classes with 600 images per class and 20 superclasses.
We would use the CIFAR-100 dataset as our input to the model. The accuracy of CIFAR-100 tends to be lower we have less images per class to learn from. CIFAR-100 has two labels per

image. It has 100 fine labels and 20 coarse labels which are superclasses of the classes. We would use both the labels in our project. Alex Krizhevsky allows use of the above datasets provided [11] has been cited and we are in compliance.

Solution Statement

Different CNN architectures have been designed and used on the dataset to get high range of accuracies. Rodrigob[12] has listed the best accuracies obtained for CIFAR-100 dataset along with links to the respective research papers. The best result[13] obtained so far has correct prediction rate of 75.72%. Finding the suitable architecture of CNN for image classification is a very difficult task and years of research have gone into it. Hence we would be using a technique called transfer learning on some of the state of the art CNN architectures pre-trained on ImageNet dataset.

Benchmark Model

We would use a simple CNN similar to the LeNet architecture to benchmark the accuracy obtained by our approach. Although the accuracy would be very low for the simple benchmark model we have chosen, but it would be still better than a random guessing accuracy of 1% for fine labels and 5% for coarse labels as we have 100 classes and 20 superclasses.

Evaluation Metrics

Accuracy is the evaluation metrics we shall use on both benchmark model and solution model. Accuracy in case of multi-label classification can obtained by dividing the number of correct predictions by the total number of input examples. We shall compare accuracies for both the training set and the test set images. This will help us to monitor the model for overfitting. We would treat fine labels and coarse labels separately and define two different accuracies.

Project Design

Data exploration

First of all, we would download the CIFAR-100 dataset and do data exploration about the format of the images available in the downloaded files.

Preprocessing

Then we would preprocess the downloaded images to convert them to a format suitable for suitable for visualization and also for input to our model.

Data visualization and analysis

We would visualize random samples of images and also analyse the coarse and fine labels associated with the images.

Benchmark Model

We would create a simple CNN architecture with just a few layers similar to LeNet and keep the accuracies obtained as a reference.

Pre-trained Models

Obtain some pre-trained models of some SoTA CNN architectures and pick one to perform transfer learning.

Analysis of Transfer Learning

Transfer learning may involve fine tuning, feature detector, freezing or learning from scratch depending on the size and similarity between the pre-trained and target dataset. We would analyse and pick the suitable approach based on CIFAR-100 and ImageNet.

Train the model

Depending on the transfer learning technique chosen, we would apply it on the pre-trained model and train our solution on the CIFAR-100 dataset.

Visualisations

If possible, plot some metrics to get a better visual understanding of the results.

Conclusion

Based on the performance obtained, discuss the pros and cons of the method and compare the results with other implementations on the same dataset.

References:

1. https://en.wikipedia.org/wiki/Computer_vision
2. LeCun, Y., Bottou, L., Bengio, Y., and Haffner, P. (1998d). Gradient-based learning applied to document recognition. Proceedings of the IEEE, 86(11), 2278–2324.
3. Deng, Jia, et al. "Imagenet: A large-scale hierarchical image database." Computer Vision and Pattern Recognition, 2009. CVPR 2009. IEEE Conference on. IEEE, 2009.
4. Krizhevsky, Alex, Ilya Sutskever, and Geoffrey E. Hinton. "Imagenet classification with deep convolutional neural networks." Advances in neural information processing systems. 2012.

5. Zeiler, Matthew D., and Rob Fergus. "Visualizing and understanding convolutional networks." European Conference on Computer Vision. Springer International Publishing, 2014.
6. Simonyan, Karen, and Andrew Zisserman. "Very deep convolutional networks for large-scale image recognition." arXiv preprint arXiv:1409.1556(2014).
7. Szegedy, Christian, et al. "Going deeper with convolutions." Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. 2015.
8. He, Kaiming, et al. "Deep residual learning for image recognition." arXiv preprint arXiv:1512.03385 (2015).
9. https://www.cs.toronto.edu/~kriz/cifar.html
10. http://groups.csail.mit.edu/vision/TinyImages
11. Learning Multiple Layers of Features from Tiny Images, Alex Krizhevsky, 2009.
12. http://rodrigob.github.io/are_we_there_yet/build/classification_datasets_results.html#43494641522d313030
13. Clevert, D., Hochreiter, S., & Unterthiner, T. (2015). Fast and Accurate Deep Network Learning by Exponential Linear Units (ELUs). *CoRR, abs/1511.07289*.