Machine Learning Engineer Nanodegree

Capstone Proposal

Tae Jun Moon

October 31st, 2018

Proposal

Using Deep Learning and Transfer Learning to train an image classifier with few data

Domain Background

One of the most popular application of machine learning is image classification. Image classification refers to the task of extracting information classes from an image. There are two types of classification depending on method approach, supervised and unsupervised. Supervised classification shows high performance when given data is labeled correct.

Supervised learning is a method of machine learning to derive a function from training data. The training data typically contains the attributes for the input object as a vector and indicates what the desired result is for the vector. Supervised classification uses labeled images to train a classifier. Regression is a method of outputting consecutive values among functions deduced from the training data. Classification is a method of indicating a certain kind of value of a given input vector. The goal of supervised learning technique is to accurately guess the value to be predicted for the test data from the training data.

For image classification, Convolutional Neural Network is mostly used today. There were lot of algorithms that people used for image classification before CNN became popular. People used to create features from images and then feed those features into some classification algorithm like SVM. Some algorithm also used the pixel level values of images as a feature vector too.

Concept of CNN was first proposed in 1998, but it was later popularized by AlexNet in 2012 after winning the imagenet classification competition by outstanding result. CNN can be thought of automatic feature extractor from the image. We lose a lot of spatial interaction between pixels if we use a algorithm with pixel vector. A CNN effectively uses adjacent pixel information to effectively downsample the image first by convolution and then uses a prediction layer at the end.

With developed image classifying networks available and increase of training image data, it has become easy to train a image classifier that can classify images fairly well. From AlexNet to today's Inception-ResNet, CNN can classify images better than human when it is trained by datasets provided with enough data with correct labeling.

Problem Statement

Problem with CNN is that when training data is provided less, they perform poorly. CNN usually have millions of parameters. For AlexNet, it has 62,378,344 parameters. Generally, number of training data should be large when the parameter is large.

To overcome this shortcoming, transfer learning technique is used for CNN. Instead of training the network from scratch, it uses pretrained weights (well-initialized parameters) to train the model. This method can be effective especially when solving problems similar to transferring models.

For this project we will use deep learning and transfer learning technique to train a image classifier, but in a case where pretrained model is not similar to training data.

Datasets and Inputs

Tiny-imagenet is a relatively small dataset compare to imagenet dataset.

Dataset	ImageNet	Tiny-ImageNet
Number of classes	1,000	200
Number of training images	50,000 per class	500 per class
Image size	(256,256,3)	(64,64,3)

Tiny-ImageNet consist of 200 classes with 500 training images each. Each images are 64x64 pixels with RGB channel. Inputs for this dataset will be image and label.

Solution Statement

Our goal is to train a model that can classify well without further collecting training image data. When we are dealing with dataset with less training data, SVM would be a good model to use. However, since we are dealing with images, it is better to use CNN then SVM even with though we have less training data.

We will use ResNet to train our model. ResNet won the 1st place in the ILSVRC 2015 classification competition with top-5 error rate of 3.57%. Since AlexNet was first proposed, the layers of the CNN architecture are getting deeper and deeper. AlexNet has only five layers, while VGGNet has 19 layers and GoogleNet has 22 layers. However, as the deeper the layer goes, the gradient vanishing occurs. The core idea of ResNet is the residual block to make a shortcut so that the gradient can flow well.

Transfer learning is a way to speed up learning and improve predictability when creating new models using existing models. Transfer learning can be useful when we are dealing with dataset with less training data. In this project we will use transfer learning method to train a classifier.

Benchmark Model

We will first use AlexNet as our benchmark model to observe the performance of AlexNet when it is trained on dataset with few images. Then we will use ResNet to compare the performance difference between AlexNet and ResNet.

Evaluation Metrics

We will use accuracy as our evaluation metrics of the model using validation images provided in tiny-imagenet dataset.

Project Design

1. Train Benchmark model.
2. Train ResNet18 from scratch.
3. Train ResNet18 with pretrained weight.
4. Preprocess the image by resizing the image from 64x64 to 224x224.

   (Use cv2 library)
5. Train ResNet18 with preprocessed images with pretrained weight.
6. If the performance was not good enough, augment the image by random cropping, flipping, and rotating.
7. Finetune some layers of ResNet18 to improve the performance.

Dataset is provided with Train/Validation/Test, however test images are not labeled. Thus we will split validation images by 50:50 for test images.

Reference

https://towardsdatascience.com/transfer-learning-946518f95666

https://tiny-imagenet.herokuapp.com/

https://medium.com/@14prakash/understanding-and-implementing-architectures-of-resnet-and-resnext-for-state-of-the-art-image-cf51669e1624