

Few Shot Learning (FSL) for Image classification

April 7, 2021

1 Problem statement

Few Shot Learning (FSL) is a sub-area of machine learning (ML) that can be used in Computer Vision tasks, such as image classification. Unlike traditional ML algorithms that work on large datasets, FSL is designed to classify images (recognise objects) with few data samples. In essence, it is clear that highly accurate image classifying models can be built from training sets with thousands of positive and negative samples. However, not all classes are created equal. Some have few recorded instances for various reasons - a rare bone disease might have few X-rays available for study, or an enemy's military armoured vehicle might have few recorded sightings.

These cases cannot be solved in the same way as conventional deep learning, and at its surface seems to be an almost impossible task for a machine to be able to recognise such classes any better than a random guess.

2 Intuition of FSL

It is obvious that we are, for example, able to recognise our parents' faces as we have seen them countless times from different angles and in different conditions. However, consider also that we are able to glance at someone's IC and tell immediately if the person in question is indeed the owner of his/her identification. The human mind somehow has the ability to one-shot learn, let alone FSL. If deep learning attempts to emulate the workings of the human brain and the connection between its neurons to identify objects,

maybe there exists some way to adapt the current deep learning techniques to objects with limited samples.

Returning to our example and taking a closer look, we realise that the brain is able to accurately recognise a face with a single positive sample because it is, in a sense, 'cheating'. It has likely seen millions of other faces, or even other objects, and trained itself to differentiate (or recognise the similarity between) any 2 objects. We can say that the brain has **learnt to learn**. This is the key finding that will allow us to train machines in FSL.

Meta learning, or learning to learn, means that a model learns the reason behind classification decisions, rather than a rote approach in the case of conventional deep learning. In this case, the reason our brains can tell if an unknown object belongs to some class is due to the similarity between the Query image (the person's actual face) and the Support Set (the person's IC).

A model of this example would be a *1-way 1-shot* classifier. However, a support set usually contains images from multiple classes. The model compares the query with each support image and outputs a similarity value for each image pair, usually between 0 and 1. The model's prediction would be the class of the support image that produced the highest similarity value when paired with the query. A more apt example would be a theft victim picking out the thief from a list of n suspects, with each suspect having k photos. This model would be a *n-way k-shot* classifier.

3 How FSL Works

In traditional deep learning, the machine is tasked with recognising a preset list of classes (objects). Beyond that, it is relatively useless. For example, a model that classifies lions would be trained on a large dataset with images of lions (positive sample) and not lions (negative sample). The positive samples are labelled with 1, and 0 for negative samples. From there, the model uses an optimisation algorithm (e.g. gradient descent) to minimise some loss function that measures the error between the model's output is from the samples' label. From this very simplified explanation, we see that the model would only be able to accurately tell if an image was a lion or not given a sufficiently large training set. An image from any other class would simply result in a 0 output (i.e. a support set would be useless to the model). Moreover, if our dataset was small, the model would not be accurate in any

way.

However, if we change the way we label our samples such that each pair of image would be a positive sample if they belong to the same class, and a negative sample otherwise, we are able to alter the model's goal to detecting the similarity of 2 images. This eliminates the need to train our model with samples of the rare class. We can use other classes which we have large datasets of to meta learn. Thus, when given an unseen query image from the rare class, we can use the one/few samples that we have and ask our model if they are the same object; this would yield a useful output.

In summary, FSL consists of the following: 1. Take the pair of images of a sample and iteratively pass them through a convolutional neural network for feature extraction 2. Take the absolute value of the difference between the 2 feature vectors 3. Process the resultant vector via dense layers to produce a scalar value 4. Apply the sigmoid function to obtain the model's prediction as a similarity value between 0 and 1 5. Compute the loss function by comparing the similarity value with the sample's label 6. Calculate the gradients of the loss function with respect to the dense layers' parameters and perform gradient descent to update its parameters 7. Further propagate the gradient to update the parameters of the neural network 8. Repeat using all training data 9. Test the model using a query image and a support set. Note that all the images involved do not belong to any of the classes found in the training data

3.1 Optimisation Function

3.2 Loss Function

3.3 Input and Forward Propagation

3.4 Backpropagation & Gradient Descent