# Report on Learning Deep Features for One-Class Classification

Valliappan

December 9, 2018

# 1 Introduction

## 1.1 Problem Statement

- The paper address the problems of similar to one shot learning, where they just have information/data of one class.
- Instances of only a single object class are available during training.
- In the context of this paper, all other classes except the class given for training is called alien classes.
- This can be an example of abnormality classification.

## 1.2 Conventional Approach

- The feature extraction network from a pre-trained model is used. Only a new classification network is trained/fine tuned in the case of low training samples.

## 1.3 Proposed Approach

- Strategy used in paper uses deep-features extracted from a pre-trained model $g_s$, where training is carried out on a different dataset (complement of the one class instance), to perform one class classification.
- The paper proposes a fine tuning framework $g_l$ which produces deep features specialized for one class classification.
- Classification network for "one-class" is defined as $h_c$
- Based on the output of the classification networks $h_c$ output two loss function is defined.

## 1.4 Key Idea

- The two loss function compactness and descriptiveness loss.

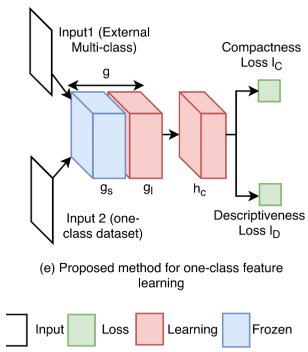


Figure 1: Proposed one-class classification network

# 2 Network Architecture

## 2.1 Notation

- $g_s$ - pre-trained weigths
- $g_l$ - fine tuning network
- $h_c$ - classification network
- $l_C$ - Compactness loss
- $l_D$ - Descriptiveness loss
- $R$ - Reference network (Alexnet)
- $r$ - Reference Dataset (Imagenet)
- $S$ - Secondary Network (similar to R)
- $t$ - Target Dataset
- $W$ - Model weights initialised with pretrained weights

## 2.2 Loss function

**Compactness Loss:** A desired quality of a feature is to have a similar feature representation for different images of the same class. Hence, a collection of features extracted from a set of images of a given class will be compactly placed in the feature space. This quality is desired even in features used for multi-class classification. In such cases, compactness is quantified using the intra-class distance; a compact representation would have a lower intra-class distance.

**Descriptiveness loss:** The given feature should produce distinct representations for images of different classes. Ideally, each class will have a distinct feature representation from each other. Descriptiveness in the feature is also a desired quality in multi-class classification. There, a descriptive feature would have large inter-class distance.

To classify the one class instance properly, it is important to satisfy both these character collectively. The optimisation problem statement is as follows:

$$loss = \max_t (D(g(t)) + \lambda C(g(t))) \tag{1}$$

where $t$ is the training data, $g$ is feature representation and $\lambda$ is a positive constant.

In this work, the variance of the feature distribution approximate it to the variance of each feature batch. This quantity as the compactness loss ($l_C$). On the other hand, descriptiveness of the learned feature cannot be assessed using a single class training data. However, if there exists a reference dataset with multiple classes, even with random object classes unrelated to the problem at hand, it can beused to assess the descriptiveness of the engineered feature. The learned feature to perform classification on an external multi-class dataset, and consider classification loss there as an indicator of the descriptiveness of the learned feature. So, the paper use cross-entropy loss calculated in this fashion as the descriptiveness loss ($l_D$)

So with this formulation, the optimization function can be writen as :

$$loss = \min_g (l_D(r) + \lambda l_C(t)) \tag{2}$$

Here, the r is the reference data and the optimization works with the hyper parameter as $g_t$ the learning network.

## 2.3 Architecture

## 2.4 Training

- 

## 2.5 Testing

- 

# 3 Experiments and Results