**Multiresolution Knowledge Distillation for Anomaly Detection**

Unsupervised representation learning has established itself as a crucial step in the detection and localization of anomalies in images. There are two difficulties with learning such a representation. First, using standard methodologies, the sample size is frequently too small to learn a rich generalizable representation. Second, even though only normal samples were provided for training, the learnt features ought to be able to distinguish between normal and abnormal samples. The generality issue severely reduces confidence for unseen future datasets. Additionally, most approaches have poor or impossible anomaly localization, which necessitates lengthy computations that degrade real-time performance. In addition, many earlier works have unstable training, necessitating unprincipled early stopping to obtain acceptable results. Here, the "distillation" of features from different layers of an expert network that has already been trained on ImageNet into a more basic cloner network is suggested as a solution to both problems. The intermediate activation levels of the expert and cloner networks given the input data is used to locate the anomalies and detect them. As opposed to only using the last layer activation values, it is demonstrated that considering various intermediate clues during distillation results in a better exploitation of the expert's knowledge and a more distinct disparity. On the other hand, interpretability algorithms are included in the unique framework for localization of anomalous regions without the requirement for any particular or intensive training method.

In this work, a cloner network that detects anomalous images in the test set and localizes anomalies in those images with the aid of a pre-trained network is trained using a training dataset made up exclusively of normal images. Cloner networks must have a thorough understanding of the manifold in order to predict how each sample will deviate from the manifold of normal data. It is therefore trained to replicate the extensive behavior of a expert network known as the source network. The aim to transfer the intermediate knowledge of the source network on the normal training data to cloner network as well. By encouraging Cloner network to learn Source network's understanding of normal samples by conforming its intermediate representations in several important levels to Source network's representations, Cloner network is given many intermediate clues from Source network. As a result, by simulating numerous layers, cloner networks are trained in various abstraction levels, leading to a final knowledge of normal data that is more in-depth. The information of the source network is only partially shared with the cloner network when only the last layer is used.

 In order to accelerate the full transfer of information from the Source network to the Cloner network, the concept of knowledge is defined as both the value and direction of all source activation values. Therefore, to represent each aspect, two losses, Lval and Ldir are defined. At each critical layer, Lval, seeks to reduce the Euclidean distance between Cloner's and Source's activation values. Ldir is used to increase the activation vectors' directional similarity. In ReLU networks, where neurons are only triggered after exceeding a zero-value threshold, this is more crucial. This suggests that two activation vectors with the same Euclidean distance from the target vector could exhibit different activation patterns in the neurons that follow them. Cosine similarity is used to define Ldir. Ltotal is calculated and training continues to fully converge, which is the only criterion that can be used to determine when to stop training epochs.

To find samples with anomalies, both the Source and Cloner networks are fed with every test input. Anomalies, or inputs outside of the usual manifold, are a potential surprise for the Cloner network because the Source has only taught the normal point of view to the Cloner. Ltotal gradients are used to locate anomalous areas that are increasing its value. The attribution map is obtained in order to obtain a localization map for the input. Gaussian blur and an opening morphological filter are used to the attribution in order to decrease the inherent sounds in the map of attribution. The localization map is obtained. On normal data, a smaller cloner network, C, is trained to replicate the entire behavior of a source network, S (VGG-16). A total loss function is utilized to formulate the disparity of their intermediate behavior, which is then used to identify anomalies during testing. Using interpretability techniques, pixel-accurate maps of anomaly localization are produced.

 The contributions of the proposed method – The method enabled the pre-trained expert network's knowledge to be transferred to the cloner network in a more thorough manner. Focusing just on the characteristics that separate normal from abnormal helps when knowledge is condensed into a smaller network. Compared to past studies, the proposed method's training procedure is computationally cheap and stable. Based on computing gradients of the discrepancy loss with respect to the input, the method enables real-time and accurate anomaly localization. carrying out an enormous number of diverse tests, surpassing earlier SOTA models by a significant margin on several datasets, while remaining competitive on the remaining ones.

**Understanding Contrastive Learning Requires Incorporating Inductive Biases**

A common method of self-supervised learning called contrastive learning encourages augmentations (views) of the same input to have more comparable representations than augmentations of different inputs. Recent efforts to theoretically explain contrastive learning's effectiveness on downstream classification tasks demonstrate the dependency on the characteristics of augmentations and the importance of contrastive loss of representations. Recent efforts formalize through presumptions that lead to the conclusion that there is little overlap for inputs from different classes but large overlap for augmentation distributions of inputs from the same class. Since methods like SimCLR do not seem to satisfy such assumptions, a recent proposed work provided a more in-depth analysis under milder assumptions, which call for only a small amount of overlap in augmentation distributions, resulting in a dense graph of connections resulting from overlaps within a class. It can be demonstrated once more that a low-dimensional representation is close to ideal in that it ensures that the contrastive loss will linearly differentiate the downstream classes. This work shows that analyses that don't account for inductive biases of the function class and training technique, cannot sufficiently account for the effectiveness of contrastive learning. Extensive studies on the image and text domains illustrate the pervasiveness of this issue; despite having the identical augmentations and contrastive losses, different function classes and algorithms act considerably differently on downstream tasks. Theoretical analysis is provided for the class of linear representations, allowing contrastive learning to function with less complexity by including inductive biases of the function class.

By encouraging representations of "similar pairs" of augmentations, the objective is to learn a representation function that maps augmentations to d-dimensional vectors. Picking two augmentations of the same input is a typical method for selecting a similar pair. The brittleness of transfer and the significance of including inductive biases in transfer limitations are clearly illustrated by a straightforward but instructive example that is presented. Ideal augmentations are those that keep input components that can predict the downstream label but modify input components that are less crucial for the label because contrastive learning aims to make representations invariant to augmentation transformations. In this work, three key phenomena are presented – a) Function class sensitivity - A representation's downstream performance is sensitive to the function class (architecture) and training method used to learn it in addition to its contrastive loss, b) Brittleness of transfer - Despite the augmentations being successful for some function classes, minimizing the contrastive loss to optimality can occasionally have a non-monotonic, detrimental influence on subsequent performance, c) The disjoint augmentations regime -It can be demonstrated that any function-class-agnostic analysis (including those from past work) provably results in vacuous assurances when augmentation distributions for inputs do not overlap.

Contrastive learning with the right function classes can sometimes succeed and non-overlapping augmentations can still be instructive, a fact that is not explained by the existing theory. This study presents straightforward tests and theoretical illustrations that point to gaps in our current understanding. The inductive bias of the employed deep nets, which has thus far mostly been explored in simple architectures (for example, depth 2 or 3), will need to be considered to fill these gaps. The behavior of basic designs like MLPs and the hypercube example is already an unsolved issue. The results demonstrate how function class bias can be accounted for in transfer bounds for linear representations, which is a somewhat challenging task. Additional insights can be gained by extending these findings to more intricate function classes and incorporating training methods. This paper's research was diagnostic in nature, seeking to fill in knowledge gaps.