Deep Self-Learning From Noisy Labels

Abraham Jose

1 Summary

The author proposes a learning framework to work on the real noisy dataset. The iterative learning approach uses several class prototypes and learning the network jointly using the corrected class label and the original noisy label iteratively. The cosine similarity is used for constructing the similarity matrix from the generated deep features for prototypes from a pool of randomly sampled image set for label correction and loss is generated from both the corrected label and the noisy label. They tested their model on real noisy datasets like Clothing1M(14 classes/50K images from web) and Food101-N(101 classes/310K images), which contains many annotation noises and was able to state-of-the-art results.

2 Good points

The author was able to explain how prototypes and the selection of prototype matters and shows why multi-prototypes are necessary to learn the noisy labels in the dataset. The selection criteria ensures that the chosen p prototypes belongs to same classes but not close to each other which increases the representative ability of the prototypes. The clustering method in the prototyping is not sensitive towards the clustering, which ensures robust performance. The end-to-end training framework without using an accessorial network or adding extra supervision on a real noisy dataset makes the method effective and efficient.

3 Weak points

If the deep features described by $G(x)$ is not able to extract the features which will help us to distinguish between the noisy label and to correct the label(in label correction stage), the algorithm will not perform as expected. The prototypes can only be randomly sampled from a certain class(for huge dataset) and there is a trade-off between the number of samples in the prototype and the computation required for the similarity matrix, which affects the performance. The ablation study on α and its progression will help us understand the effect on the noisy label training and the effects, which is not presented in the paper.

4 Questions

Often times, for the feature extractors for a certain class is associated with our object of interest(car) but the decision is made because of car and the road that the network see. How much this association will affect the label correction procedure? The effect is not trivial in case of large varying dataset like ImageNet dataset.

How can these noise removal techniques can be leveraged for a CNN used for detection of objects?

5 Ideas

Incorporating Region Proposal networks followed by the prototype and clustering method ensures that this technique can be expanded for the object detection tasks.