Learning Robust representations by projecting superficial statistics out

Abraham Jose
ID :5068109, CAP6614
abraham@knights.ucf.edu

1 Summary

Deep neural networks depend heavily on superficial statistics of the training data and are liable to break under distribution shift, for example changes to the background or texture of an image can break a classifier. The author proposes the gray-level co-occurrence matrix(GLCM) which is used to extract the patterns that our prior knowledge suggests are superficial;they are sensitive to the texture but unable to capture the gestalt of an image. Using the GLCM representation, we can make the model more robust and generalize the model in two different ways using NGLCM and HEX. This will make the classifier generalize to previously unseen domain.

2 Strengths of the proposal

1. The author presents two different ways to make the deep neural network generalize better. Either we can built on the reverse gradient method that pushes our model to learn representations from which the GLCM representation is not predictable or built on the independence introduced by projecting the model's representation onto the subspace orthogonal to GLCM representation's.
2. They proposed methods are parameter-free, architecture-agnostics designed to discard superficial information available in the data-set.
3. They also introduces two synthetic data-sets for Domain adaption/Domain generalization which is correlated with semantic information.

3 Weaknesses of the proposal

1. The NGLCM cannot be completely free of semantic information of an image. In MNIST performance drops if NGLCM is used, because it tries to learn the semantic information from the image as well rather than domain specific noise.
2. Training performance of HEX fluctuates dramatically during training, which makes the model less stable.

4 Results

The authors found that NGLCM performs slightly poorly compared to MLP in domain and digit classification while performing experiments on a number of data-set like Office dataset,MNIST, PACS etc.. Hex method refers to adding another MLP with feature extracted by the traditional GLCM methods which perform well with the Office data. We can further see a trend here, where HEX performs very well on smaller angles in the MNIST-rotation data.

5 Discussion

The paper assumes that by making the model less dependant on superficial aspects/noise in image, we can force the model to rely more on difference that makes a difference, ie the semantic data available in the image once you remove the domain specific noise. The goal of the paper was to build a network that has the capacity to extract textural information from an image without extracting semantic information.