Review Feedback

May 31, 2018

Reviewer 1

1. Without more details about the network configuration and training procedure, I am not sure if this work is reproducible.

   Architecture image along with Architecture and Experiment section are enough to reproduce the results. We do believe that training and generating different results on different datasets is a bit complex and could be resolved with more explanation. Due to space limititation we are not able to present a step by step guide in the paper. We have a project page for the paper online, we have code, weights and datasets availabel for sharing. We are waiting for permission from our industrial project partner.

2. More experiment results should be added. For example, this paper should add the evaluation results of "without attention model" in Table 1.

   The proposed model is formed after we added attention and texture encoding layer to the StyleNet's [5] first six convolution layer. We do not use pretrained weights of StyleNet for our experiments. The results of StyleNet[5] on multi-label classification task on both Fashion144k and Fashion550k has been presented by the authors in their other work [28]. We have included the results as baseline in Table 1.

Reviewer 2

3. The details of texture encoding network are absent.

   Texture encoding network is a single layer with hyperparamter codewords. We use codewords of size 32 for all our experiments.