Label Embedding Network can learn label representation (label embedding) during the training process of deep networks. With the proposed method, the label embedding is adaptively and automatically learned through back propagation. The original one-hot represented loss function is converted into a new loss function with soft distributions, such that the originally unrelated labels have continuous interactions with each other during the training process. As a result, the trained model can achieve substantially higher accuracy and with faster convergence speed. Experimental results based on competitive tasks demonstrate the effectiveness of the proposed method, and the learned label embedding is reasonable and interpretable. The proposed method achieves comparable or even better results than the state-of-the-art systems.

The contributions of this work are as follows:  
**Learning label embedding and compressed embedding**: We propose the Label Embedding Network that can learn label representation for soft training of deep networks. Furthermore, some large-scale tasks have a massive number of labels, and a naive version of label embedding network will suffer from intractable memory cost problem. We propose a solution to automatically learn compressed label embedding, such that the memory cost is substantially reduced.

**Interpretable and reusable**: The learned label embeddings are reasonable and interpretable, such that we can find meaningful similarities among the labels. The proposed method can learn interpretable label embeddings on both image processing tasks and natural language processing tasks. In addition, the learned label embeddings can be directly adapted for training a new model with improved accuracy and convergence speed.

**General-purpose solution and competitive results**: The proposed method can be widely applied to various models, including ResNet models. We conducted experiments on computer vision tasks including CIFAR-10. Results suggest that the proposed method achieves significantly better accuracy than the existing methods ResNet.

Error rate curve for CIFAR-10. 20 times experiments (the light color curves) are conducted for credible results both on the baseline and our proposed model. The average results are shown as deep color curves:

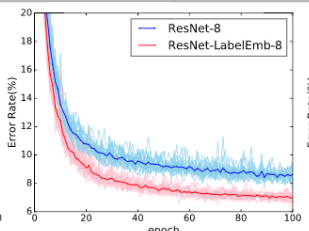
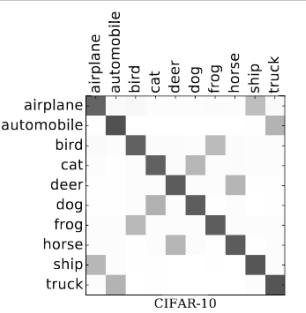


Fig:- ResNet8 error rate

### Heatmaps generated by the label embeddings:



### Results of Label Embedding:

