Explaining the Effect of Data Augmentation on Image Classification Tasks

Jerry Tang*, Manasi Sharma, Ruohan Zhang

Stanford Department of Computer Science, Stanford Vision and Learning Lab, *only CS231N author

Background and Introduction

- Deep learning and image classification research exploded in the past decade and reached sufficiency for many applications
- Explainability in deep learning is a research challenge, putting barrier to intuition for the “black box” that is neural networks
- Data augmentation is proven to be useful to improve models but lacks understanding to why it works
- Saliency mapping produces heat map that represents the attention of models given a query image, and is a potential tool for explainability research
- The project is part of a greater study for a SVL research project looking to optimize data augmentation

Problem Statement

- There is no research on the explainability aspect of data augmentation, and we see saliency mapping as a potential solution that can ultimately optimize data augmentation
- This project will train with different augmented datasets and a baseline model. Saliency mappings will be queried from all models. Qualitative and quantitative analysis on data augmentation will focus on both the classification accuracy, resulting changes in saliency mappings, and the relationship between the two

Dataset and Data Augmentation

- MNIST is the dataset, but only use 100 training examples because the dataset is too easy to achieve high accuracy
- Five augmentation techniques applied to this subset: scaling, random cropping, random erasing, distortion, and blurring
- 10 augmented image for each of original= 1000 for each method. Add original to each set = 1100 images, 6 total train sets

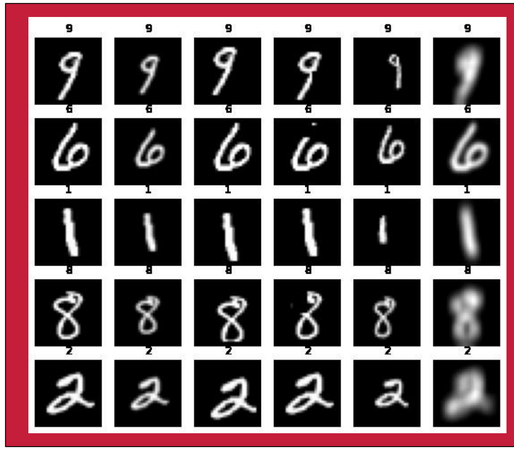

Figure 1: Dataset Examples

- Same five images from each of six dataset
- Columns from left to right: scaling, random cropping, random erasing, distortion, and blurring
- Augmentation parameters are assigned to at least preserve eligibility by human eyes

Methodology

- 30 epochs with 10 steps each trained = 6 unique CNNs
- Uniform CNN architecture: 2 conv layers + Linear & Softmax
- Test for validation accuracy every epochs, pick the model with highest validation accuracy as the final model for that specific data augmentation trained model, and obtain testing accuracy
- Generate saliency mapping with and without laying over the query image
- Compute Pearson’s Correlation Coefficient (CC) **(Eqn. 1)**
- Compute Kullback-Leibler Divergence (KL) **(Eqn. 2)**
- Compute Shannon entropy **(Eqn. 3)**

$$CC(P,Q) = \frac{cov(P,Q)}{\sigma(P) \times \sigma(Q)} \quad (1)$$

$$KL(P,Q) = \Sigma_i Q_i log(\epsilon + \frac{Q}{\epsilon + P}) \quad (2)$$

$$H(X) = -\Sigma_{x \in X} p(x) log(p(x)) \quad (3)$$

Dataset	BestVal	Test
Scaling	0.71	0.67
RandCrop	0.88	0.87
Distortion	0.85	0.79
Blur	0.74	0.73
Erasing	0.79	0.79
Orig	0.81	0.78

Table 1. Training Results

Augmentation Results Evaluation

- Training on augmented data converges slower: more complexity
- Distortion converges significantly slower than others
- Random cropping result in consistently the most accurate model
- Distortion and random erasing also performs pretty well

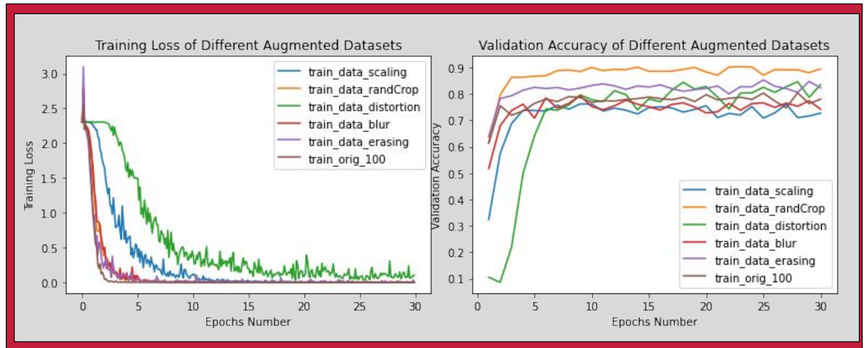

Figure 2: Training loss and validation accuracy over 30 epochs

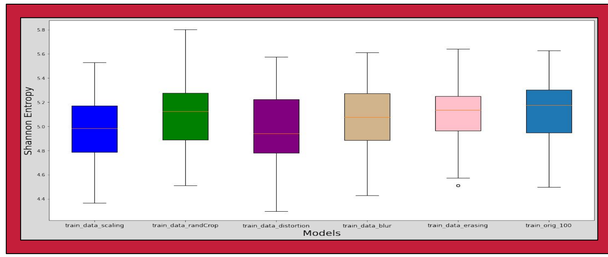

Figure 3: Shannon entropy

- No significant variation between models

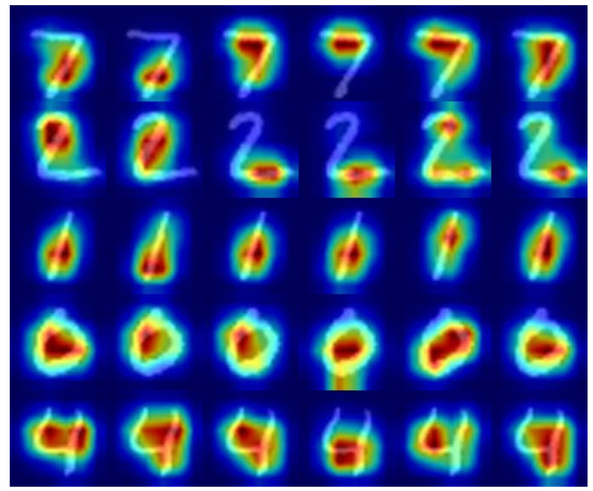

Figure 4:

- Saliency maps overlapping their query images.
- No background bias
- Heat map is local and connected
- Clear variation between models

Saliency Mappings Evaluation

- Data augmentation clearly alters saliency mappings, but mass annotation needed to make a qualitative conclusion **(Fig. 4)**
- CC and KL show preservation/loss of information compared to baseline saliency mapping from linear and probabilistic statistical methods **(Fig. 5)**
- Augmentation strength and parameters have high impact on CC and KL scores. MNIST is too easy of a dataset
- CC/KL have no correlation with val/test accuracy
- Shannon entropy **(Fig. 3)** measures randomness and information content. Disappointingly showed no significant difference across augmentation methods and with the baseline model

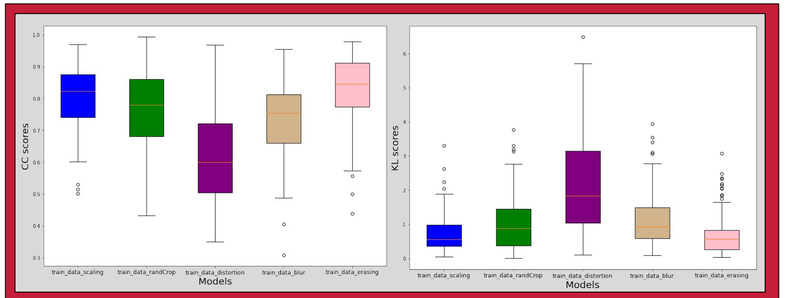

Figure 5: CC and KL scores respect to baseline saliency maps

Conclusion and Future Research

- Saliency maps indicates strength of augmentation, but is not correlated to better classification accuracy
- Still difficult to use saliency maps to optimize data augmentation
- Future work should use more sophisticated images and more radical data augmentation techniques