Distortion Classifier Using Deep Convolutional Neural Networks
Mike Hsieh - Roland Duffau - Alex He
{mhsieh33, rduffau, yuzehe}@stanford.edu
CS230 - Winter 2020
YouTube Link: https://youtu.be/o3G9VQ8FlTk

Problem statement
Given an input image, output the corresponding distortion, if any, found in that image. We trained three CNN distortion classifiers which can be useful as part of an image restoration process. Our models are unique in that they identify distortion type rather than quantifying the distortion level as commonly found in Image Quality Assessment (IQA) models. Our models when trained on a dataset with varying distortion levels generalized well to the reference LIVE dataset.

Features
The size of each image was normalized to 224 x 224 pixels. This was done by rescaling the image so that the shorter side becomes 224 pixels and cropping out the central 224 x 224 segment. Each input had a depth of 3 which corresponds with RGB color. We did not normalize the pixel values, since it showed little or no effect on our performances. We utilized image processing filters from JH Labs to generate the distortions.

Results
Training on our first dataset resulted in clear overfitting to our generated 6 distortions. Generalization highly improved after training on our second dataset. And error analysis showed that low distortion levels were the main remaining pain point:

<table>
<thead>
<tr>
<th>Model</th>
<th>Validation Acc</th>
<th>Test Acc</th>
<th>LIVE Acc</th>
</tr>
</thead>
<tbody>
<tr>
<td>Softmax</td>
<td>0.243</td>
<td>0.243</td>
<td></td>
</tr>
<tr>
<td>VGG16 scratch</td>
<td>0.989</td>
<td>0.983</td>
<td>0.964</td>
</tr>
<tr>
<td>VGG16 transfer learning</td>
<td>0.991</td>
<td>0.992</td>
<td>0.170</td>
</tr>
<tr>
<td>SGDNet transfer learning</td>
<td>0.872</td>
<td>0.887</td>
<td>0.102</td>
</tr>
</tbody>
</table>

Discussion
• Baseline: it is evident that the softmax model does not classify the distortions well. Our VGG16 scratch model performed similarly well to the fine-tuned VGG16 network on our first dataset. SGDNet transferred model performed worst amongst the three models, suggesting an inadequation of our first dataset vs IQA.
• Generalization: if VGG16 scratch model gives the best performance on IQA dataset, all models (particularly the pre-trained ones) generalize very poorly to real world distribution. Confirmation of a data mismatch.
• Retrain on new distribution: Test accuracies for all three models on external dataset improved significantly once we switched to two distortion types with varying distortion severity, more similar to IQA distributions.
• Computing the Precision, Recall and F1 Score suggested that this weak performance may have been due to imbalanced data (reference images are less represented). But it is also possible that the reference images themselves had innate blurriness or noise, hard to identify even by human eyes.
• Deeper error analysis showed that quality of predictions was correlated to the level of distortion. The higher the distortion severity, the better was the prediction. Wrong predictions occurred mostly on images with low distortion level, which tend to be less differentiable (both from reference images and inter distortion types).

Future
Real world images may present multiple types of distortion at a time. To that extend, the multiclass classifier we built may not be enough. As a next step, we’d be keen to extend our model to a multi-label classifier, being able to predict properties of a sample that are not mutually exclusive.
Images with only slight distortions require disproportionately more attention in order to achieve better accuracy. It may be helpful to fine-tune further with more images with only slight distortions, or create dedicated classes to improve their classification.
Finally, we need to test more extensively on external datasets. A better supervised dataset with human-labeled images may also be helpful.

References