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/o3G9VQ8FtTk

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.

Data

Training/Dev/Test data (2 Datasets)

Constant levels of 6 distortion types → **7 classes** (inc no distortion): 0, 1, ..., 0, 0

Random levels of Gaussian Blur and White Noise → **3 classes** (inc no distortion): 0, 0, 1

Images downloaded from ImageNet modified with specific distortions.
200k images for each dataset

Evaluation data

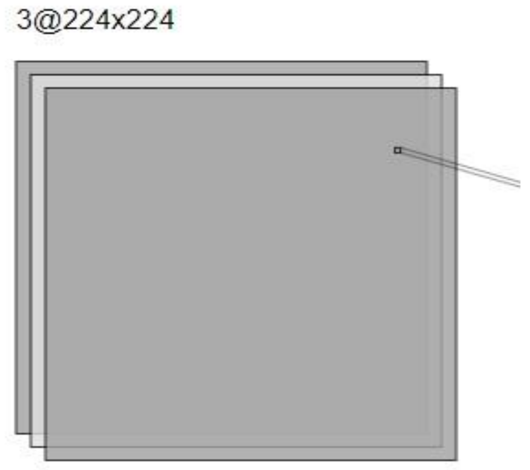
The University of Texas at Austin
LIVE
Laboratory for Image & Video Engineering

One **reference IQA dataset** with Gaussian Blur and White Noise distorted images

Features

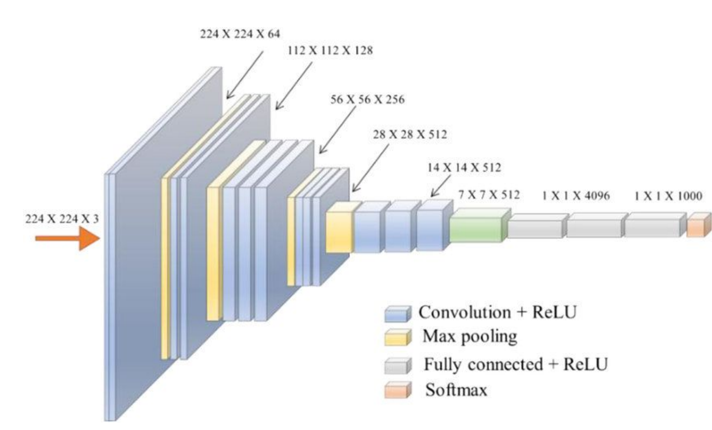
The size of each image was normalized to 224 x 224 pixels. This was done by rescaling the image so that the shorter side becomes 224 pixels and cropping out the central 224 x 224 segment. Each input had a depth of 3 which corresponds with RGB color. We did not normalize the pixel values, since it showed little or effect on our performances. We utilized image processing filters from JH Labs to generate the distortions.

3@224x224



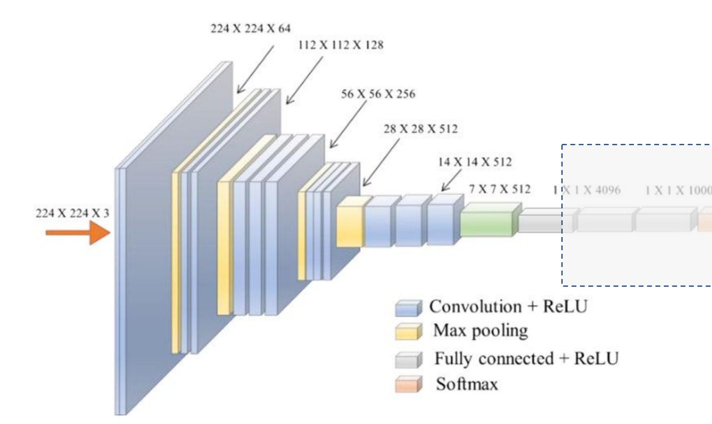
Models

1. VGG16 trained from scratch



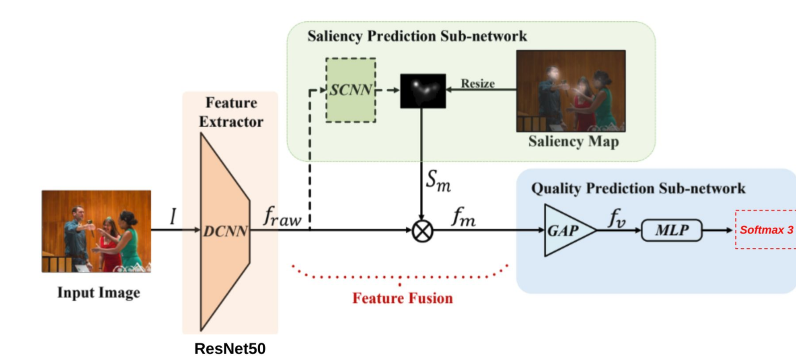
We reused the architecture of VGG16 with similar architecture to our transfer learning model. Trained all 112.7 million parameters

2. VGG16 fine tuned



Removed last 8 layers.
Added 1 Conv, 1 Max Pool, 1 FC
Ended with Softmax(3)
Retrained 105.1 million parameters

3. SGDNet fine tuned



SGDNet is based on ResNet50, and pretrained on IQA
Removed last 4 layers, added Softmax(3)
Retrained 1,539 parameters

Results

Training on our first dataset resulted in clear **overfitting** to our generated 6 distortions

Model	Validation Acc.	Test Acc.	LIVE Acc.
Softmax	0.243	0.243	-
VGG16 scratch	0.989	0.988	0.484
VGG16 transfer learning	0.991	0.992	0.170
SGDNet transfer learning	0.872	0.887	0.102

Data Mismatch!

And **incapacity to generalize** to real world images

Generalization highly improved after training on our second dataset

Model	Validation Acc.	Test Acc.	LIVE Acc.
Softmax	0.450	0.455	-
VGG16 scratch	0.864	0.856	0.661
VGG16 transfer learning	0.852	0.850	0.40
SGDNet transfer learning	0.869	0.837	0.724

And **error analysis** showed that low distortion levels were the main remaining pain point :

Discussions

- **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

[1] Imagenet. http://www.image-net.org/.
[2] Jh labs java image processing. http://www.jhlabs.com/ip/distorting.html.
[3] Alex Krizhevsky, Ilya Sutskever, and Geoffrey E Hinton. Imagenet classification with deep convolutional neural networks. In F. Pereira, C. J. C. Burges, L. Bottou, and K. Q. Weinberger,editors, *Advances in Neural Information Processing Systems 25*, pages 1097–1105. Curran Associates, Inc., 2012.
[4] Weisi Lin Yongtao Wang Sheng Yang, Jiang Qiuping. Sgdnet: An end-to-end saliency-guided deep neural network for no-reference image quality assessment. 10 2019