No-Reference Image Quality Assessment Using Convolutional Neural Networks

Niranjan, Rudrapatna Nataraja (nnataraj@stanford.edu)

Motivation

Digital images and videos can be found everywhere today. It is one of the primary modes of communication and entertainment, thus it is very important to ensure high quality image is delivered to the end users. Image quality assessment (IQA) is an important area of research because of its applications in:

- Monitoring Quality of Service (QoS) in internet streaming applications
- To identify level of image degradation which can affect image recognition accuracy
- In medical imaging to help decide compression ratio without loss of information

vs

Which image is visually better?

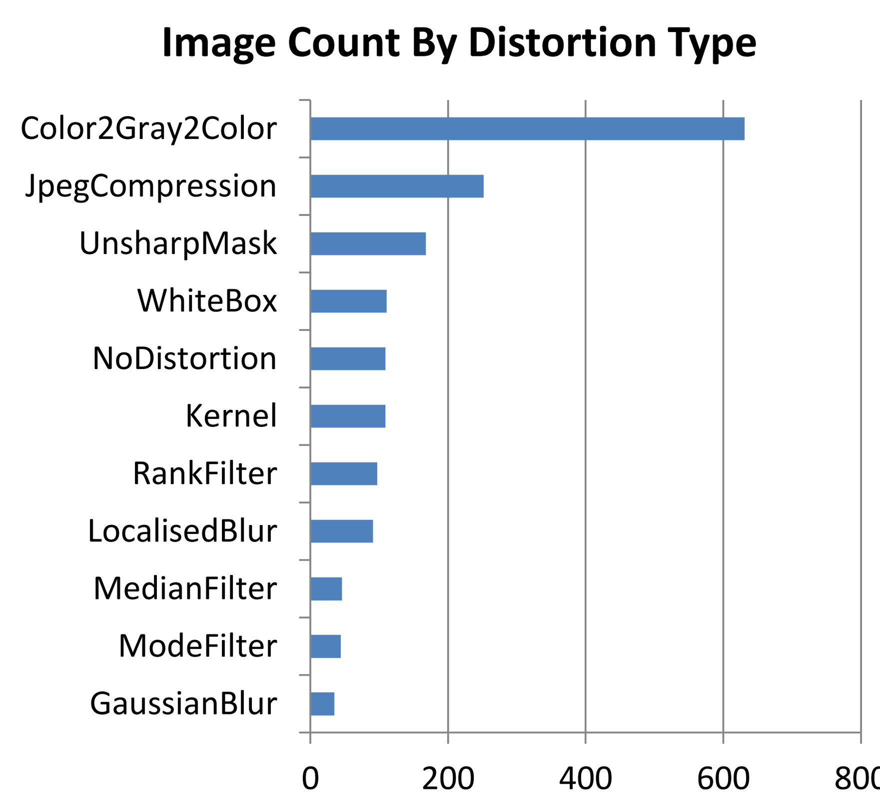
Problem Definition

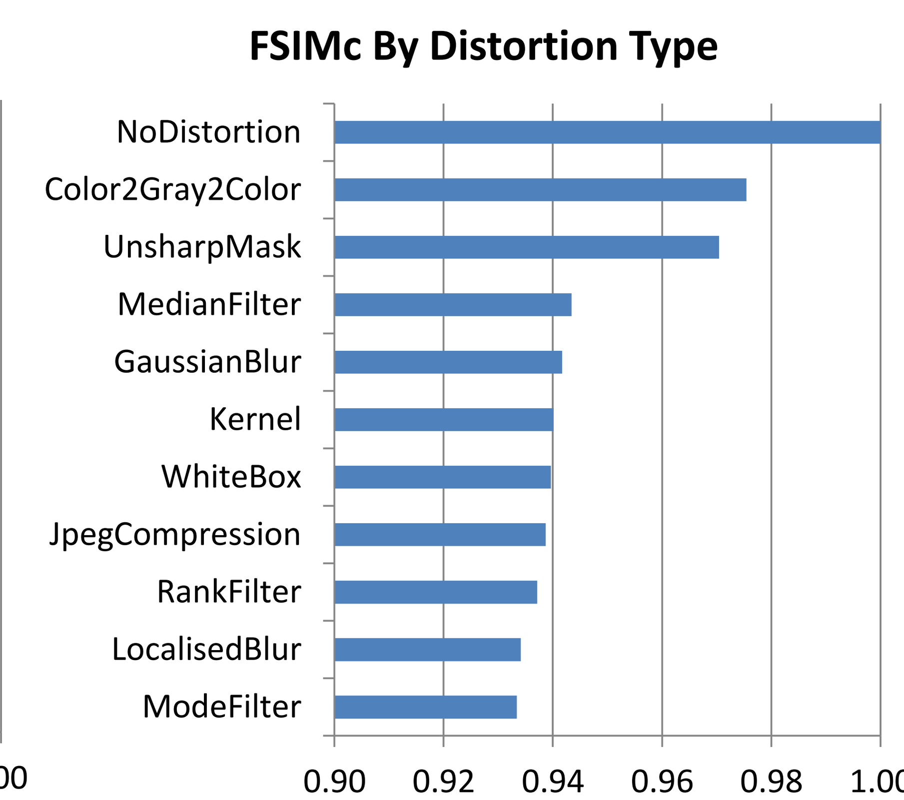
Human visual system can easily distinguish between good quality images versus bad ones even when a reference image is not available. Feature Similarity Index (FSIM) tries to capture the quality of an image which is a close approximation to human perceived quality. This serves as a motivation for us to be able to predict FSIM color (FSIMc) scores in the absence of reference images.

Dataset

Apply distortions to images and calculate FSIMc scores.

Source	ImageNet
Size	128000
Images Per Class	200
Number of Classes	10 (FSIMc >0.9 - 1.0)
Image Size	256x256x3
Image Patches	64 - 32x32x3
Distortions Applied	Filter, Blur, Compression, Sharpness, White Box, Color Conversion



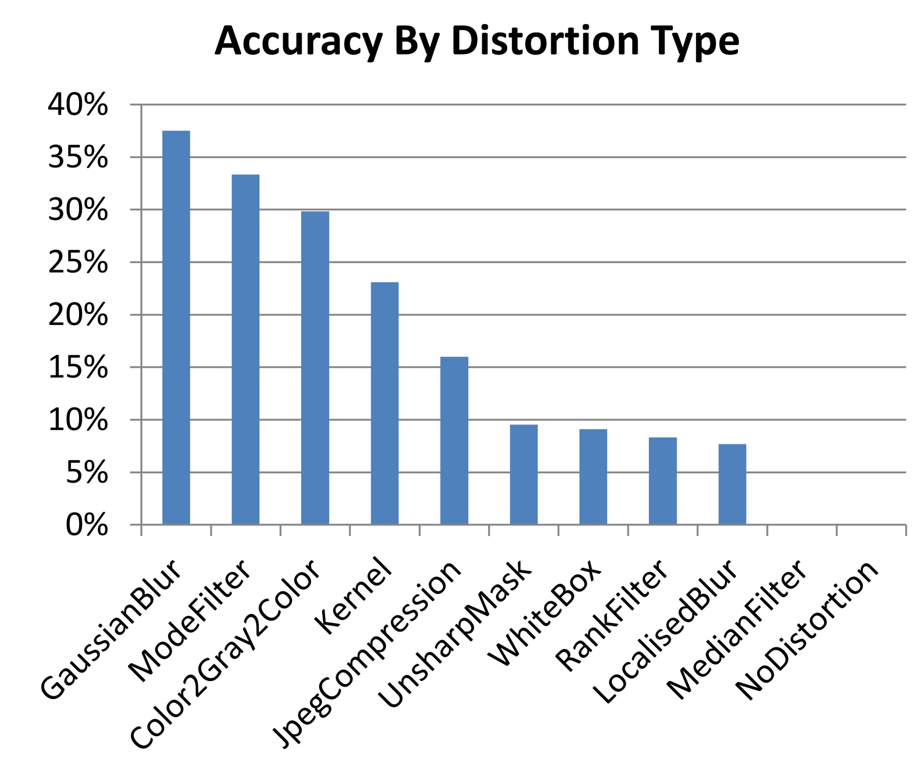


Algorithm

A multi-layered convolutional neural network (CNN) is used for feature extraction and a sotfmax layer is used for classification. The classification process involves passing the 32x32 input image through a series of convolutional layers, pooling layers and fully connected layers. Mode and Mean is calculated for each image and used as the final predicated value. Patches are mainly used to reduce the training time.

Results

We experimented with different architectures which included 2, 4 and 6 convolutional layers with Relu activation and pooling and batch normalization in between every 2 convolutional layers. Also we added 0.7 dropout probability for the last but one fully connected layer for regularization effect. The best architecture was using 6 convolutional layers with dropout. The model had difficulty predicting correct label for localized distortions.



The accuracy values are reported for the whole image instead of the patches because our goal is to evaluate the input image.

Correlation with FSIMc	0.36
Accuracy based on Mode	0.19
Accuracy based on Mean	0.34

Future Work

The performance of the model was better for distortions applied to the entire image and this makes sense because we use the same score for the entire image. Future work involves exploring options to distinguish between localized distortions versus global distortions. Another area to explore is to pass the entire image as input instead of using patches. This model can be further extended to video evaluation as well.

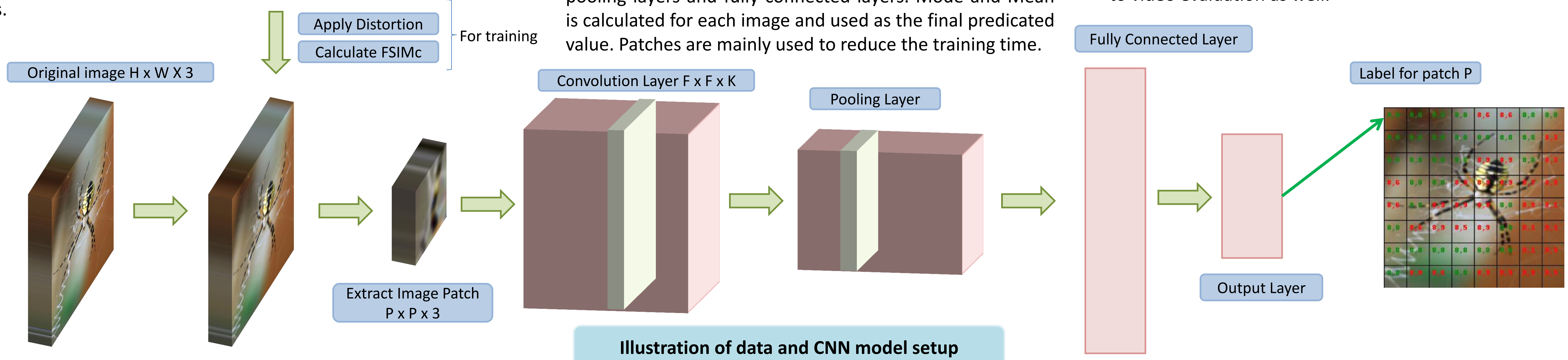


Illustration of data and CNN model setup