Blind Picture Upscaling Ratio Prediction

Todd R. Goodall, Ioannis Katsavounidis, Zhi Li, Anne Aaron, and Alan C. Bovik, *Fellow*, *IEEE*

***Abstract*—Natural scene statistics are well-studied in the context of picture quality assessment and have been used in a wide variety of top-performing picture quality prediction models. Upscaling artifacts have been measured with regards to quality impairment using these kinds of models. However, the assessment and classification of subtle, less discriminable upscaling artifacts remains an unsolved problem. The nearly imperceptible artifacts pertaining to the extent and type of upscaling have not been predicted using NSS-based models.**

We develop an accurate model for predicting the upscaling ratio applied to any natural image. By decomposing an input image frame using an orthogonal filter bank and locally normalizing the resulting responses, we show that the local energy terms can be used to predict the upscaling ratio. In fact, a simple linear regressor can be trained on these energy measurements, hence no hyper-parameter tuning is necessary. We compare the proposed model with other no-reference models using real-world data contained in the Netflix collection.

***Index Terms*—Natural Scene Statistics; Upscaling prediction; Upscaling detection**

I. Introduction

Upscaling increases the pixel count of an image, most commonly by using a bilinear, bicubic, or Lanczos-based interpolation technique. Upscaled pictures often contain a variety of spatial distortions, including periodic artifacts, incidental spatial correlations, and losses of high-frequency energy. Upscaled videos are difficult to detect, especially using human eyes. Video streaming companies, like Netflix, obtain video sources directly from video content producers or distributors, who may introduce upscaling artifacts at one or more stages of their pipeline. These artifacts will ultimately impact the quality of experience of the streaming service end-users. Given the desire to deliver best-of-class video quality, streaming companies would like to be able to determine if video sources have been upscaled upon receiving the sources.

Much prior work in upscaling detection has been forensic research on detecting malign modifications and additions to static images. By exploiting periodicities introduced by upscaling, Mahdian and Saic derive several spatial covariance formulas and use radon transform analysis [1] to produce a predictor of both scaling and rotational transformations. Other methods [2] [3] [4] [5] [6] [7] transform the input signal in some way before applying the Discrete Fourier Transform (DFT) to measure periodicities. A weakness of these DFT-based methods is their ambiguity when handling upscaling ratios outside the range of 1x-2x.

Another common approach to upscaling detection involves measuring high-frequency energy loss. Both Katsavounidis *et al.* [8] and Feng *et al.* [9] make frequency magnitude measurements to determine the extent of this energy loss. The training-free model introduced in [8] measures the "drop-off" or "knee" point that upscaling introduces, which is subsequently used to predict the upscaling ratio. Feng's energy density model extracts 19 energy ratios from the frequency spectrum magnitude, which requires an additional prediction step to obtain the upscaling ratio. The model by Vázquez-Padín *et al.* [10] counts the number of nonzero singular values in the SVD decomposition. One observed weakness of these energy-based techniques is their unpredictable reaction to varying content.

Combining periodicity analysis with DFT magnitude measurements has been shown to reduce the weaknesses related to each approach. Zhu *et. al.* [11] introduced a ranking scheme using a Support Vector Classifier (SVC) to learn the degree of relative upscaling between two image pairs. Phennig and Kirchner [12] provided an in-depth analysis of both classes of techniques, characterizing their weaknesses, and combining the approaches to improve prediction accuracy.

Natural scene statistics have found use in full-reference upscaled picture quality prediction [13]. However, no-reference analysis of upscaling artifacts has not been studied in the context of natural scene statistics models, which are the basis of a variety of powerful perception-based picture quality predictors. No-reference quality prediction models such as the Blind Referenceless Image Spatial QUality Evaluator (BRISQUE) [14] and Naturalness Image Quality Evaluator (NIQE) [15] use simple spatial-domain feature extraction strategies that correlate well with human opinions of multiple picture distortion types. Here, we follow this path by describing a new high-performance blind upscaling prediction model that combines a novel pre-filtering technique with the Mean-Subtracted Contrast-Normalized (MSCN) and "paired product" computations developed in BRISQUE.

The layout of this paper is as follows. Section II describes the proposed model in detail. Section III presents four experiments, where III-A and III-B describe prediction of the upscaling ratio while III-C and III-D describe classification of the upscaling interpolation function. Finally, Section IV presents concluding remarks.

II. Proposed natural scene-based model

As described in [17], Principal Component Analysis (PCA), when applied to images, can find an orthogonal basis of natural image patches. We observed that these derived basis functions change as natural image patches are upscaled, leading us to explore how these changes can provide a useful measurement on upscaling artifacts. Although different filter designs may be applied, we opt for a simple approach learned directly from

Fig. 1. Exemplar pristine image selected from the Berkeley image segmentation database [16].

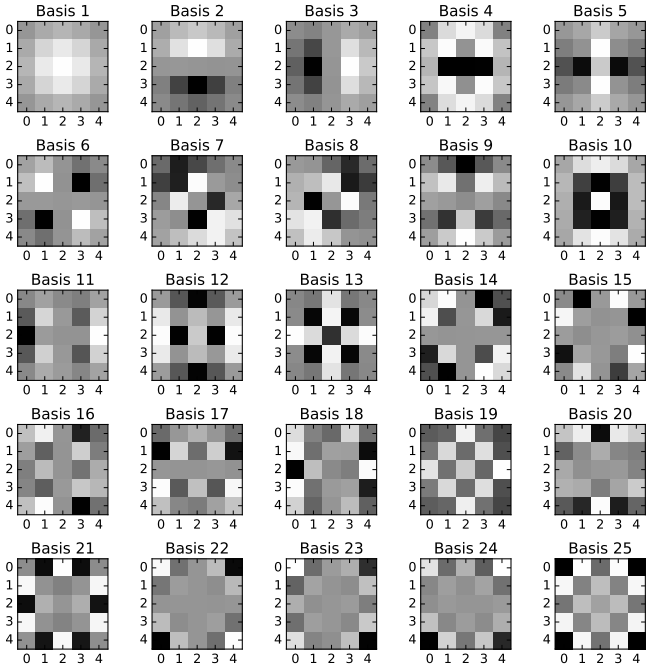

Fig. 2. Basis functions computed using PCA on 5x5 patches. All patches were obtained on pristine images from the Berkeley image segmentation database [16].

natural images, differing from [6] in that the filters used are not specifically optimized for upscaled images.

In this work, we select a corpus of 500 natural luminance images, obtained from the Berkeley image segmentation database [16]. Each image is split into overlapping patches of size 5x5, from which we select 2000 random patches. Each patch is multiplied by a 5x5 Gaussian mask sampled to 2 standard deviations and normalized to unit maximum value to reduce energy at the patch boundaries. Accumulating the weighted patches from each image yields a total of 1 million patches. Given these 5x5 patches, PCA will produce at most 25 orthogonal basis functions, as depicted in Fig. 2, most of which exhibit sinusoidal-like properties.

We use these 25 orthogonal basis functions for image prefiltering. Given an input luminance image, I, a total of 25 response images were produced after filtering with each of these basis functions, yielding $R^{(f)}$ where $f \in \{1, 2, ..., 25\}$. Next, each response image, $R^{(f)}$, undergoes divisive normalization to yield MSCN map $\widehat{R}^{(f)}$ for each f according to

$$\widehat{R}^{(f)}(\mathbf{x}) = \frac{R^{(f)}(\mathbf{x}) - \mu(R^{(f)};\mathbf{x})}{\sigma(R^{(f)};\mathbf{x}) + \epsilon}$$

where

$$\mu(R^{(f)};\mathbf{x}) = \sum_{k=-K}^{K}\sum_{l=-L}^{L} w_{k,l} R^{(f)}_{k,l}(\mathbf{x})$$

and

$$\sigma(R^{(f)};\mathbf{x}) = \sqrt{\sum_{k=-K}^{K}\sum_{l=-L}^{L} w_{k,l}(R^{(f)}_{k,l}(\mathbf{x}) - \mu(R^{(f)};\mathbf{x}))^2},$$

where $K = L = 5$, $\mathbf{x}$ is the pixel location vector, and $w = \{w_{k,l}|k = -K, \cdots, K, l = -L, \cdots, L\}$ is a 2D circularly-symmetric Gaussian weighting function sampled out to 3 standard deviations and normalized to unit volume. Throughout, we fixed the saturation parameter $\epsilon = 1\text{x}10^{-9}$.

The coefficients $\widehat{R}^{(f)}$ are the MSCN versions of the basis filtered responses, like those obtained in BRISQUE. This MSCN transform is inspired by retinal models of divisive normalization in the human visual system. A total of 25 sample standard deviation features, $\sigma^{(f)}_m$, are computed on the 25 $\widehat{R}^{(f)}$ maps. To obtain measurements of local spatial correlations that may exist after normalization, "paired product" coefficient maps are computed for each $\widehat{R}^{(f)}$ according to

$$\begin{aligned}
\text{H}(\widehat{R}^{(f)};i,j) &= \widehat{R}^{(f)}(i,j)\widehat{R}^{(f)}(i,j+1)\\
\text{V}(\widehat{R}^{(f)};i,j) &= \widehat{R}^{(f)}(i,j)\widehat{R}^{(f)}(i+1,j)\\
\text{D1}(\widehat{R}^{(f)};i,j) &= \widehat{R}^{(f)}(i,j)\widehat{R}^{(f)}(i+1,j+1)\\
\text{D2}(\widehat{R}^{(f)};i,j) &= \widehat{R}^{(f)}(i,j)\widehat{R}^{(f)}(i+1,j-1)
\end{aligned}$$

yielding a total of 100 "paired product" maps. The sample standard deviations $pp^{(f)}_H$, $pp^{(f)}_V$, $pp^{(f)}_{D1}$, and $pp^{(f)}_{D2}$ are computed on $\text{H}(\widehat{R}^{(f)})$, $\text{V}(\widehat{R}^{(f)})$, $\text{D1}(\widehat{R}^{(f)})$, and $\text{D2}(\widehat{R}^{(f)})$ respectively. Thus, 25 MSCN features, $\sigma^{(f)}_m$, and 100 local correlation features, $pp^{(f)}_H$, $pp^{(f)}_V$, $pp^{(f)}_{D1}$, and $pp^{(f)}_{D2}$, are computed on each input image, for a total of 125 features.

To observe the behavior of the distributions from which our features are extracted, we plot the histograms of $\widehat{R}^{(6)}$ and $V(\widehat{R}^{(6)})$ in Fig. 3, for the case of the test image in Fig. 1. When upscaling by factors of 1x, 2x, and 3x, a direct relationship appears between the histogram width and upscaling factor, with higher upscaling resulting in narrower histograms.

By measuring correlations between each feature and the upscaling ratio, we can better understand the contribution of each feature to a final prediction. Using the Berkeley dataset,

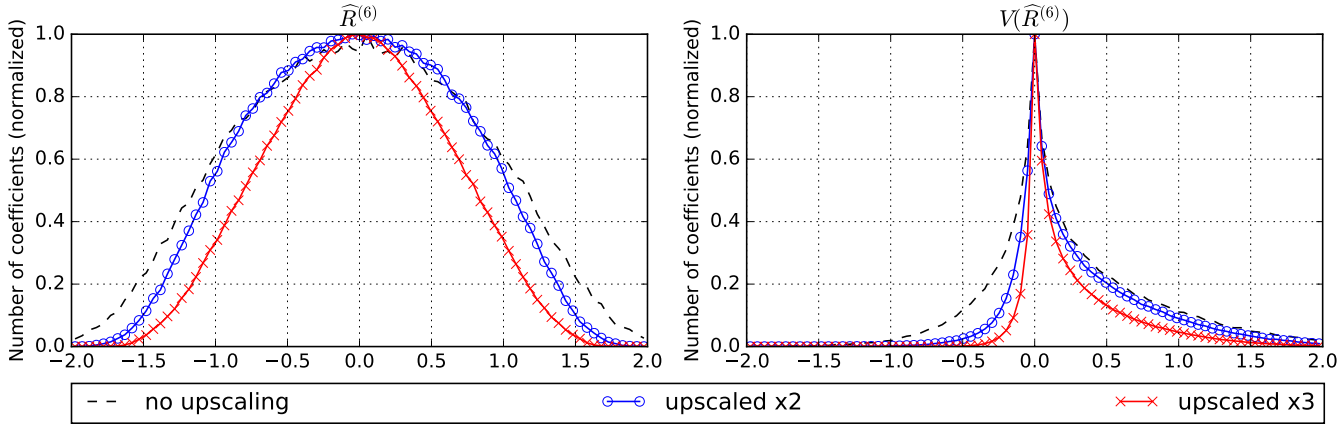

Fig. 3. Histograms of MSCN and vertical paired product for basis filter 6 for different degrees of upscaling. These coefficients were computed using bicubic upscaling of the image in Fig. 1.

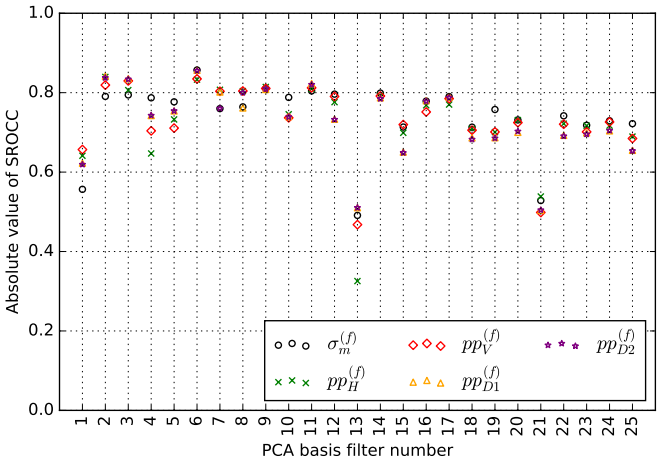

Fig. 4. Absolute value SROCC between each basis function and the upscaling ratio. Images are upscaled using one of bilinear, bicubic, or Lanczos interpolation.

we obtained 1500 images by upscaling the 500 images to upscaling ratios in the continuous range $[1, 3]$ with bilinear, bicubic, and Lanczos upscaling. Next, we observed the correlations between the 125 features and the upscaling ratio. Figure 4 shows the absolute Spearman's Rank-Order Correlation Coefficients (SROCC) between features and upscaling ratio.

From Fig. 4, the highest correlation occurs using Basis 6, which measures responses to a cross-like shape. A low correlation can be observed against the response to the low-pass Basis 1, since the upscaling artifact perturbs high-frequencies. Interestingly, the 5 features extracted from each basis have similar correlations, except $pp_H^{(13)}$.

III. Experiments

A. General Prediction Performance

To compare performance amongst algorithms on a controlled dataset, the Berkeley segmentation dataset was used again. We upscaled 75% of the images in the dataset to have upscaling ratios in the continuous range $[1.25, 3]$, such that each upscaled image was assigned a unique ratio. The remaining 25% of the images were not upscaled. Each image then recieved one of three levels of compression: None, 90%, and 80% quality using the *imagemagick* [18] command line utility, which implements JPEG compression. Introducing both upscaling and compression allows for a more realistic test, since delivery of professional content can include both lossless and compressed images. Note that images in this dataset are likely downscaled, minimizing CFA interpolation artifacts.

For the proposed model, predictions of the upscaling ratio were made using both a linear regressor and a Support Vector Regressor (SVR). We compared performance between these regressors to show that a linear combination of the proposed features yields a competitive predictor. Moreover, comparing models using a linear regressor can provide a basis from which to start tuning more complex models. For the alternative models, the suggested predictors were used. Note that Gallagher directly estimated upscaling without need for a regressor.

The Berkeley dataset was randomized, then partitioned into two sets, with 75% of the dataset for training and 25% for testing. Models were evaluated on the testing data using the Linear Correlation Coefficient (LCC) and Mean-Squared Error (MSE). This process was repeated 1000 times, each time re-randomizing the dataset order before partitioning. The median results of this testing are reported in Table I.

As may be seen, the proposed algorithm achieved top prediction results overall, except for Lanczos interpolation. When performance on all combined categories was measured, the prediction performance of all models was found to suffer. This could perhaps be overcome using a more complex machine learning model, as exemplified by the results obtained using the SVR.

B. Movie and TV Show Upscaling Prediction Performance

Since the Berkeley dataset was used when training the pre-filters, there might be concern that performance on the Berkeley dataset may be inflated owing to some unseen bias (e.g., in the human selection of content). To address this concern, we collected 801 distinct video frames from the Netflix collection, from movie and TV show sequences that were encoded at resolutions of 480p, 720p, 1080p, and 2160p with extremely light compression. Next, each of these frames

TABLE I
MEDIAN PREDICTION PERFORMANCE ACROSS UPSCALING METHODS OVER 1000 TRAIN/TEST TRIALS ON "BERKELEY" DATASET. THE PRESENCE OF '*' INDICATES THAT ALL UPSCALING METHODS ARE PRESENT IN THE TESTING AND TRAINING SETS.

Model	Bilinear		Bicubic		Lanczos		*	
	LCC	MSE	LCC	MSE	LCC	MSE	LCC	MSE
Gallagher	0.624	0.404	0.615	0.431	0.629	0.476	0.420	0.495
Pfennig and Kirchner (SVR)	0.910	0.079	0.860	0.132	0.813	0.188	0.849	0.139
BRISQUE (SVR)	0.956	0.034	0.975	0.021	0.977	0.019	0.966	0.029
Feng *et al.* (SVR)	0.973	0.023	0.982	0.015	**0.994**	**0.005**	0.968	0.027
Proposed (Linear)	0.965	0.030	0.972	0.024	0.981	0.017	0.960	0.035
Proposed (SVR)	**0.981**	**0.016**	**0.985**	**0.013**	0.988	0.012	**0.979**	**0.018**

TABLE II
MEDIAN PREDICTION PERFORMANCE ACROSS UPSCALING METHODS OVER 1000 TRAIN/TEST TRIALS ON "MOVIE AND TV SHOW" IMAGE DATASET. THE PRESENCE OF '*' INDICATES THAT ALL UPSCALING METHODS ARE PRESENT IN THE TESTING AND TRAINING SETS.

Model	Bilinear		Bicubic		Lanczos		*	
	LCC	MSE	LCC	MSE	LCC	MSE	LCC	MSE
Gallagher	0.267	0.477	0.029	0.674	-0.069	0.772	0.416	0.500
Pfennig and Kirchner (SVR)	0.745	0.199	0.460	0.471	0.285	0.623	0.430	0.467
BRISQUE (SVR)	0.952	0.041	0.930	0.058	0.941	0.050	0.928	0.060
Feng *et al.* (SVR)	0.796	0.161	0.877	0.099	0.935	0.055	0.795	0.161
Proposed (Linear)	0.970	0.025	0.961	0.033	0.969	0.026	0.951	0.042
Proposed (SVR)	**0.979**	**0.018**	**0.978**	**0.019**	**0.981**	**0.016**	**0.969**	**0.026**

TABLE III
MEDIAN CLASSIFICATION ACCURACY ACROSS UPSCALING METHODS OVER 1000 TRAIN/TEST TRIALS ON "BERKELEY" DATASET. THE PRESENCE OF '*' INDICATES THAT ALL UPSCALING METHODS ARE PRESENT IN THE TESTING AND TRAINING SETS.

Model	None	JPEG 90%	JPEG 80%	*
BRISQUE (SVC)	0.872	0.816	0.752	0.768
Feng *et al.* (SVC)	0.968	**0.960**	**0.952**	**0.944**
Proposed (LDA)	**0.984**	0.928	0.856	0.880
Proposed (SVC)	0.976	0.912	0.856	0.872

TABLE IV
MEDIAN CLASSIFICATION ACCURACY ACROSS UPSCALING METHODS OVER 1000 TRAIN/TEST TRIALS ON "MOVIE AND TV SHOW" DATASET.

Model	Accuracy
BRISQUE (SVC)	0.776
Feng *et al.* (SVC)	0.672
Proposed (LDA)	**0.935**
Proposed (SVC)	0.915

was subjected to upscaling as before, using bilinear, bicubic, or Lanczos upscaling. This time, JPEG compression was not applied, since, in practice, source inspection of content is applied only to high quality videos.

Using the same 75%/25% training/test split and 1000 trials, we evaluated the prediction performance of each model, as shown in Table II. The proposed algorithm delivered outstanding performance on both the 3 datasets containing only a single type of upscaling and on the dataset with multiple types of upscaling. For this particular use case, the energy-based Feng *et al.* features appear to have significant difficulty for both bicubic and bilinear upscaling techniques.

C. General Classification Performance

Determining the interpolation method used is important for both forensic artifact detection and for reporting source issues. At the same time, study of model classification performance can lead to further insights into the actual artifacts. For instance, if classification accuracy of a model is high, then information specific to each upscaling artifact is captured.

As listed in Table III, several models were used to classify an image as having been upscaled using bilinear, bicubic, or Lanczos interpolation. Decisions were made using Linear Discriminant Analysis (LDA) and Support Vector Classifiers (SVCs) for the same reasons that we used linear regression. Again, a total of 1000 randomized 75%/25% train/test splits were used, and the median results reported in Table III. Feng *et al.* largely outperformed the other models.

D. Movie and TV Show Upscaling Classification Performance

We also measured classification performance on the Netflix video frames as shown in Table IV. Here, Feng *et al.* largely underperformed, indicating that measurements on the frequency magnitude are more ambiguous for the given content. When compared to Table III, more mis-classifications occured for all models. The accuracies across all models are low, implying that classifying the interpolation function is a difficult problem.

IV. CONCLUSION/FUTURE WORK

We proposed a natural scene statistics-based method of predicting the amount of upscaling that has been applied to a picture. We show it to be an accurate and monotonic predictor of upscaling, which can be trained using linear regressors. In addition, the proposed model is a general spatial model that is not necessarily limited to the upscaling artifact. Lastly, the model has only the following tunable parameters: the patch size, the Gaussian mask scale parameter for smoothing the extracted basis filters (controlling bandwidth), a saturation parameter ϵ, and the Gaussian scale parameter for the μ and σ computations. These values can be further explored in future work. It may also be possible to create models that require no training at all, following [15] and [19].

REFERENCES

[1] B. Mahdian and S. Saic, "Blind authentication using periodic properties of interpolation," *IEEE Transactions on Information Forensics and Security*, vol. 3, no. 3, pp. 529–538, 2008.

[2] A. C. Gallagher, "Detection of linear and cubic interpolation in jpeg compressed images," *Canadian Conference on Computer and Robot Vision*, pp. 65–72, 2005.

[3] S. Prasad and K. Ramakrishnan, "On resampling detection and its application to detect image tampering," *IEEE International Conference on Multimedia and Expo*, pp. 1325–1328, 2006.

[4] S.-J. Ryu and H.-K. Lee, "Estimation of linear transformation by analyzing the periodicity of interpolation," *Pattern Recognition Letters*, vol. 36, pp. 89–99, 2014.

[5] A. C. Popescu and H. Farid, "Exposing digital forgeries by detecting traces of resampling," *IEEE Transactions on Signal Processing*, vol. 53, no. 2, pp. 758–767, 2005.

[6] D. Vázquez-Padín and F. Pérez-González, "Prefilter design for forensic resampling estimation," *IEEE International Workshop on Information Forensics and Security*, pp. 1–6, 2011.

[7] M. Kirchner, "Fast and reliable resampling detection by spectral analysis of fixed linear predictor residue," *Proceedings of the 10th ACM workshop on Multimedia and security*, pp. 11–20, 2008.

[8] I. Katsavounidis, A. Aaron, and D. Ronca, "Native resolution detection of video sequences," *Society of Motion Picture & Television Engineers*, 2015.

[9] X. Feng, I. J. Cox, and G. Doerr, "Normalized energy density-based forensic detection of resampled images," *IEEE Transactions on Multimedia*, vol. 14, no. 3, pp. 536–545, 2012.

[10] D. Vázquez-Padín, P. Comesaña, and F. Pérez-González, "An SVD approach to forensic image resampling detection," *EUSIPCO*, pp. 2067–2071, 2015.

[11] N. Zhu, X. Gao, and C. Deng, "Image scaling factor estimation based on normalized energy density and learning to rank," *IEEE International Conference on Security, Pattern Analysis, and Cybernetics*, pp. 197–202, 2014.

[12] S. Pfennig and M. Kirchner, "Spectral methods to determine the exact scaling factor of resampled digital images," *5th International Symposium on Communications Control and Signal Processing*, pp. 1–6, 2012.

[13] H. Yeganeh, M. Rostami, and Z. Wang, "Objective quality assessment of interpolated natural images," *IEEE Transactions on Image Processing*, vol. 24, no. 11, pp. 4651–4663, 2015.

[14] A. Mittal, A. K. Moorthy, and A. C. Bovik, "No-reference image quality assessment in the spatial domain," *IEEE Transactions on Image Processing*, vol. 21, no. 12, pp. 4695–4708, 2012.

[15] A. Mittal, R. Soundararajan, and A. C. Bovik, "Making a completely blind image quality analyzer," *IEEE Signal Processing Letters*, vol. 20, no. 3, pp. 209–212, 2013.

[16] D. Martin, C. Fowlkes, D. Tal, and J. Malik, "A database of human segmented natural images and its application to evaluating segmentation algorithms and measuring ecological statistics," *Proceedings IEEE International Conference on Computer Vision*, vol. 2, pp. 416–423, 2001.

[17] P. J. Hancock, R. J. Baddeley, and L. S. Smith, "The principal components of natural images," *Network: Computation in Neural Systems*, vol. 3, no. 1, pp. 61–70, 1992.

[18] "ImageMagick," http://www.imagemagick.org/script/index.php.

[19] A. Mittal, G. S. Muralidhar, J. Ghosh, and A. C. Bovik, "Blind image quality assessment without human training using latent quality factors," *IEEE Signal Processing Letters*, vol. 19, no. 2, pp. 75–78, 2012.