# NIMA: Neural Image Assessment

Hossein Talebi and Peyman Milanfar, *Fellow, IEEE*

***Abstract*— Automatically learned quality assessment for images has recently become a hot topic due to its usefulness in a wide variety of applications, such as evaluating image capture pipelines, storage techniques, and sharing media. Despite the subjective nature of this problem, most existing methods only predict the *mean* opinion score provided by data sets, such as AVA and TID2013. Our approach differs from others in that we predict the *distribution* of human opinion scores using a convolutional neural network. Our architecture also has the advantage of being significantly simpler than other methods with comparable performance. Our proposed approach relies on the success (and retraining) of proven, state-of-the-art deep object recognition networks. Our resulting network can be used to not only score images reliably and with high correlation to human perception, but also to assist with adaptation and optimization of photo editing/enhancement algorithms in a photographic pipeline. All this is done without need for a "golden" reference image, consequently allowing for single-image, semantic- and perceptually-aware, *no-reference* quality assessment.**

***Index Terms*— Image quality assessment, no-reference quality assessment, deep learning.**

## I. Introduction

QUANTIFICATION of image quality and aesthetics have been a long-standing problem in image processing and computer vision. While technical quality assessment deals with measuring low-level degradations such as noise, blur, compression artifacts, etc., aesthetic assessment quantifies semantic level characteristics associated with emotions and beauty in images. In general, image quality assessment can be categorized into full-reference and no-reference approaches. While availability of a reference image is assumed in the former (metrics such as PSNR, SSIM [3], etc.), typically blind (no-reference) approaches rely on a statistical model of distortions to predict image quality. The main goal of both categories is to predict a quality score that correlates well with human perception. Yet, the subjective nature of image quality remains the fundamental issue. Recently, more complex models such as deep convolutional neural networks (CNNs) have been used to address this problem [4]–[11]. Emergence of labeled data from human ratings has encouraged these efforts [1], [2], [12]–[14]. In a typical deep CNN approach, weights are initialized by training on classification related datasets (e.g. ImageNet [15]), and then fine tuned on annotated data for perceptual quality assessment tasks.

Manuscript received September 15, 2017; revised March 6, 2018; accepted April 14, 2018. Date of publication April 30, 2018; date of current version May 16, 2018. The associate editor coordinating the review of this manuscript and approving it for publication was Dr. Alin M. Achim. *(Corresponding author: Hossein Talebi.)*

The authors are with the Google Research, Mountain View, CA 94043 USA (e-mail: htalebi@google.com; milanfar@google.com).

Color versions of one or more of the figures in this paper are available online at http://ieeexplore.ieee.org.

Digital Object Identifier 10.1109/TIP.2018.2831899

### A. Related Work

Machine learning has shown promising success in predicting technical quality of images [4]–[7]. Kang et. al. [5] show that extracting high level features using CNNs can result in state-of-the-art blind quality assessment performance. It appears that replacing hand-crafted features with an end-to-end feature learning system is the main advantage of using CNNs for pixel-level quality assessment tasks [5], [6]. The proposed method in [5] is a shallow network with one convolutional layer and two fully-connected layers, and input patches are of size $32 \times 32$. Bosse *et al.* [6] use a deep CNN with 12 layers to improve on image quality predictions of [5]. Given the small input size ($32 \times 32$ patch), both methods require score aggregation across the whole image. Bianco *et al.* in [7] propose a deep quality predictor based on AlexNet [15]. Multiple CNN features are extracted from image crops of size $227 \times 227$, and then regressed to the human scores.

Success of CNNs on object recognition tasks has significantly benefited the research on aesthetic assessment. This seems natural, as semantic level qualities are directly related to image content. Recent CNN-based methods [8]–[11], [16] show a significant performance improvement compared to earlier works based on hand-crafted features [1]. Murray *et al.* [1] is the benchmark on aesthetic assessment. They introduce the AVA dataset and propose a technique to use manually designed features for style classification. Later, Lu *et al.* [8], [17] show that deep CNNs are well suited to the aesthetic assessment task. Their double-column CNN [17] consists of four convolutional and two fully-connected layers, and its inputs are the resized image and cropped windows of size $224 \times 224$. Predictions from these global and local image views are aggregated to an overall score by a fully-connected layer. Similar to Murray *et al.* [1], in [17] images are also categorized to low and high aesthetics based on mean human ratings. A regression loss and an AlexNet inspired architecture is used in [9] to predict the mean scores. In a similar approach to [9], Jin *et al.* [11] fine-tune a VGG network [18] to learn the human ratings of the AVA dataset. They use a regression framework to predict the histogram of ratings. A recent method by Zeng *et al.* [19] retrains AlexNet and ResNet CNNs to predict quality of photos. More recently, [10] uses an adaptive spatial pooling to allow for feeding multiple scales of the input image with fixed size aspect ratios to their CNN. This work presents a multi-net (each network a pre-trained VGG) approach which extracts features at multiple scales, and uses a scene aware aggregation layer to combine predictions of the sub-networks. Similarly, Ma *et al.* [20] propose a layout-aware framework in which a saliency map is used to select patches

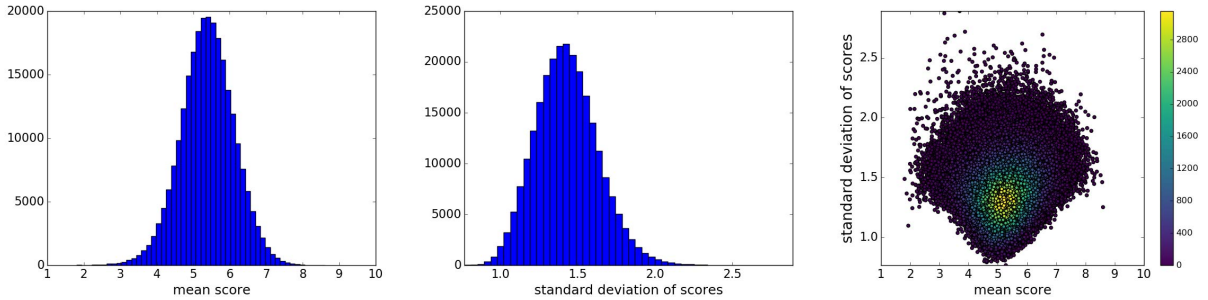
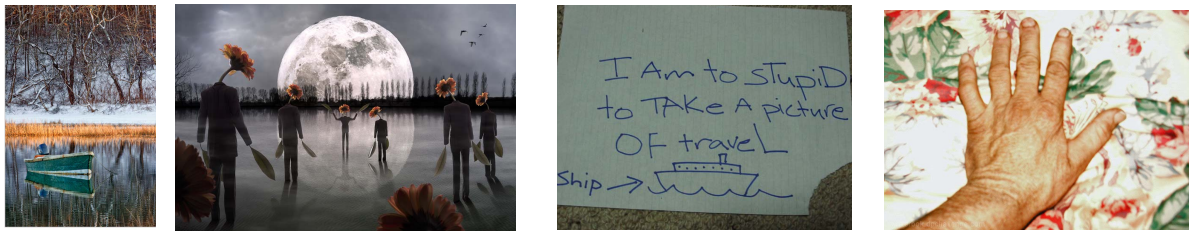



Fig. 1. Histograms of ratings from AVA dataset [1]. Left: Histogram of mean scores. Middle: Histogram of standard deviations. Right: Joint histogram of the mean and standard deviation.




(a) (b) (c) (d)

Fig. 2. Some example images from AVA dataset [1] with quality score $\mu(\pm\sigma)$, where $\mu$ and $\sigma$ represent mean and standard deviation of score, respectively. (a) high aesthetics and low unconventionality (challenge name: "*Best of 2007*", $\mu = 6.36$, $\sigma = 1.04$), (b) high aesthetics and high unconventionality (challenge name: "*Extreme super moon*", $\mu = 7.84$, $\sigma = 2.08$), (c) low aesthetics and high unconventionality (challenge name: "*Travel*", $\mu = 2.62$, $\sigma = 2.15$), (d) low aesthetics and low unconventionality (challenge name: "*Pieces of the human form*", $\mu = 3.12$, $\sigma = 1.28$).

with highest impact on predicted aesthetic score. Overall, none of these methods reported correlation of their predictions with respect to ground truth ratings. Recently, Kong *et al.* in [14] proposed a method to aesthetically rank photos by training on AVA with a rank-based loss function. They trained an AlexNet-based CNN to learn the difference of the aesthetic scores from two input images, and as a result, indirectly optimize for rank correlation. To the best of our knowledge, [14] is the only work that performed a correlation evaluation against AVA ratings.

### B. Our Contributions

In this work, we introduce a novel approach to predict both technical and aesthetic qualities of images. We show that models with the same CNN architecture, trained on different datasets, lead to state-of-the-art performance for both tasks. Since we aim for predictions with higher correlation with human ratings, instead of classifying images to low/high score or regressing to the mean score, the distribution of ratings are predicted as a histogram. To this end, we use the squared EMD (earth mover's distance) loss proposed in [21], which shows a performance boost in classification with ordered classes. Our experiments show that this approach also leads to more accurate prediction of the mean score. Also, as shown in aesthetic assessment case [1], non-conventionality of images is directly related to score standard deviations. Our proposed paradigm allows for predicting this metric as well.

It has recently been shown that perceptual quality predictors can be used as learning loss to train image enhancement models [22], [23]. Similarly, image quality predictors can be used to adjust parameters of enhancement techniques [24]. In this work we use our quality assessment technique to effectively tune parameters of image denoising and tone enhancement operators to produce perceptually superior results.

This paper begins with reviewing three widely used datasets for quality assessment. Then, our proposed method is explained in more detail. Finally, performance of this work is quantified and compared to the existing methods.

### C. A Large-Scale Database for Aesthetic Visual Analysis (AVA) [1]

The AVA dataset contains about 255,000 images, rated based on aesthetic qualities by amateur photographers.[1] Each photo is scored by an average of 200 people in response to photography contests. Each image is associated to a single challenge theme, with nearly 900 different contests in the AVA. The image ratings range from 1 to 10, with 10 being the highest aesthetic score associated to an image. Histograms of AVA ratings are shown in Fig. 1. As can be seen, mean ratings are concentrated around the overall mean score ($\approx$5.5). Also, ratings of roughly half of the photos in AVA dataset have a standard deviation greater than 1.4. As pointed out in [1], presumably images with high score variance tend to be subject to interpretation, whereas images with low score variance seem to represent conventional styles or subject matter. A few examples with ratings associated with different levels of aesthetic quality and unconventionality are illustrated in Fig. 2. It seems that aesthetic quality of a photograph can be represented by the mean score, and unconventionality of it closely correlates

[1] AVA images are obtained from www.dpchallenge.com, which is an on-line community for amateur photographers.

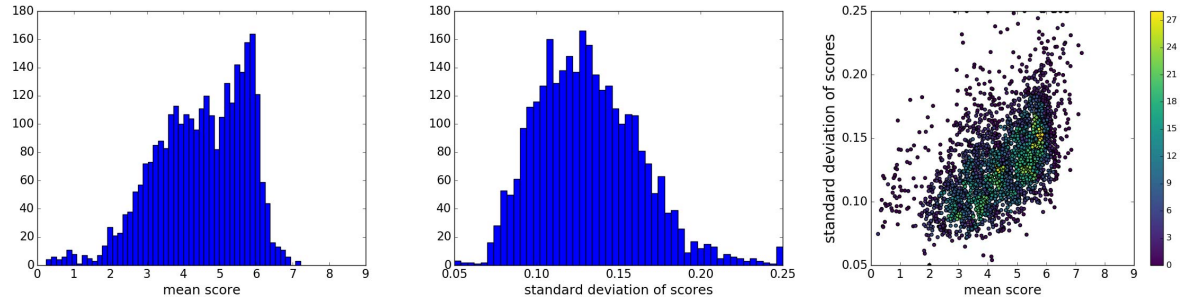



Fig. 3. Histograms of ratings from TID2013 dataset [2]. Left: Histogram of mean scores. Middle: Histogram of standard deviations. Right: Joint histogram of the mean and standard deviation.


Fig. 4. JPEG artifact example images from TID2013 dataset [2] with quality score $\mu(\pm\sigma)$, where $\mu$ and $\sigma$ represent mean and standard deviation of score, respectively. Clean image and 5 levels of JPEG compression artifacts are shown here. (a) clean image, (b) compression artifact level 1, $\mu = 5.73$, $\sigma = 0.15$, (c) compression artifact level 2, $\mu = 5.47$, $\sigma = 0.11$, (d) compression artifact level 3, $\mu = 4.86$, $\sigma = 0.11$, (e) compression artifact level 4, $\mu = 3.0$, $\sigma = 0.11$, (f) compression artifact level 5, $\mu = 1.66$, $\sigma = 0.16$.

to the score deviation. Given the distribution of AVA scores, typically, training a model on AVA data results in predictions with small deviations around the overall mean (5.5).

It is worth mentioning that the joint histogram in Fig. 1 shows higher deviations for very low/high ratings (compared to the overall mean 5.5, and mean standard deviation 1.43). In other words, divergence of opinion is more consistent in AVA images with extreme aesthetic qualities. As discussed in [1], distribution of ratings with mean value between 2 and 8 can be closely approximated by Gaussian functions, and highly skewed ratings can be modeled by Gamma distributions.

### *D. Tampere Image Database 2013 (TID2013) [2]*

TID2013 is curated for evaluation of full-reference perceptual image quality. It contains 3000 images, from 25 reference (clean) images (Kodak images [25]), 24 types of distortions with 5 levels for each distortion. This leads to 120 distorted images for each reference image; including different types of distortions such as compression artifacts, noise, blur and color artifacts.

Human ratings of TID2013 images are collected through a forced choice experiment, where observers select a better image between two distorted choices. Set up of the experiment allows raters to view the reference image while making a decision. In each experiment, every distorted image is used in 9 random pairwise comparisons. The selected image gets one point, and other image gets zero points. At the end of the experiment, sum of the points is used as the quality score associated with an image (this leads to scores ranging from 0 to 9). To obtain the overall mean scores, total of 985 experiments are carried out.

Mean and standard deviation of TID2013 ratings are shown in Fig. 3. As can be seen in Fig. 3(c), the mean and score deviation values are weakly correlated. A few images from TID2013 are illustrated in Fig. 4 and Fig. 5. All five levels of JPEG compression artifacts and the respective ratings are


Fig. 5. Some example images from TID2013 dataset [2] with quality score $\mu(\pm\sigma)$, where $\mu$ and $\sigma$ represent mean and standard deviation of score, respectively. Clean image and 5 levels of contrast change distortions are shown here. (a) clean image, (b) contrast change distortion of level 1, $\mu = 5.67$, $\sigma = 0.10$, (c) contrast change distortion of level 2, $\mu = 6.80$, $\sigma = 0.18$, (d) contrast change distortion of level 3, $\mu = 4.83$, $\sigma = 0.16$, (e) contrast change distortion of level 4, $\mu = 6.69$, $\sigma = 0.29$, (f) contrast change distortion of level 5, $\mu = 3.88$, $\sigma = 0.18$.

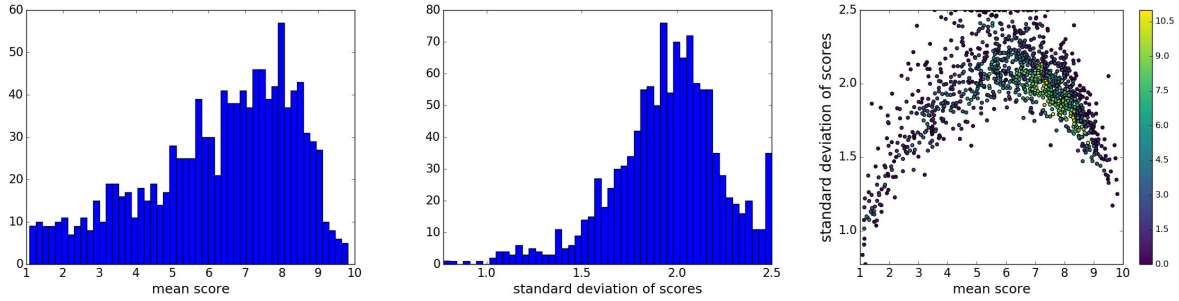



Fig. 6. Histograms of ratings from LIVE dataset [26]. Left: Histogram of mean scores. Middle: Histogram of standard deviations. Right: Joint histogram of the mean and standard deviation. Note that LIVE scores are scaled to [1,10].

illustrated in Fig. 4. Evidently higher distortion level leads to lower mean score[2]. Effect of contrast compression/stretching distortion on the human ratings is demonstrated in Fig. 5. Interestingly, stretch of contrast (Fig. 5(c) and Fig. 5(e)) leads to relatively higher perceptual quality.

### *E. LIVE In the Wild Image Quality Challenge Database [26]*

LIVE dataset contains 1162 photos captured by mobile devices. Each image is rated by an average of 175 unique subjects. Mean and standard deviation of LIVE ratings are shown in Fig. 6. As can be seen in the joint histogram, images that are rated near overall mean score show higher standard deviation. A few images from LIVE dataset are illustrated in Fig. 7. It is worth noting that in this paper, LIVE scores are scaled to [1, 10].

Unlike AVA, which includes distribution of ratings for each image, TID2013 and LIVE only provide mean and standard deviation of the opinion scores. Since our proposed method requires training on score probabilities, the score distributions are approximated through maximum entropy optimization [27].

The rest of the paper is organized as follows. In Section II, a detailed explanation of the proposed method is described. Next, in Section III, applications of our algorithm in ranking photos and image enhancement are exemplified. We also provide details of our implementation. Finally, this paper is concluded in SectionIV.

## II. Proposed Method

Our proposed quality and aesthetic predictor stands on image classifier architectures. More explicitly, we explore a

[2]This is a quite consistent trend for most of the other distortions too (namely noise, blur and color distortions). However, in case of the contrast change (Fig. 5), this trend is not obvious. This is due to the order of contrast compression/stretching from level 1 to level 5)

Fig. 7. Some example images from LIVE dataset [26] with quality score $\mu(\pm\sigma)$, where $\mu$ and $\sigma$ represent mean and standard deviation of score, respectively. Note that LIVE scores are scaled to [1, 10]. (a) 9.99 (±1.22). (b) 9.35 (±1.49). (c) 8.29 (±1.99). (d) 3.50 (±1.69). (e) 2.33 (±1.51). (f) 1.95 (±1.39).

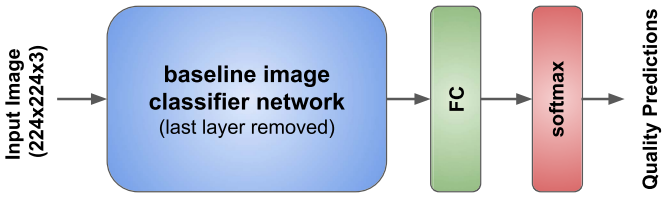



Fig. 8. Modified baseline image classifier network used in our framework. Last layer of classifier network is replaced by a fully-connected layer to output 10 classes of quality scores. Baseline network weights are initialized by training on ImageNet dataset [15], and the added fully-connected weights are initialized randomly.

few different classifier architectures such as VGG16 [18], Inception-v2 [28], and MobileNet [29] for image quality assessment task. VGG16 consists of 13 convolutional and 3 fully-connected layers. Small convolution filters of size $3 \times 3$ are used in the deep VGG16 architecture [18]. Inception-v2 [28] is based on the Inception module [30] which allows for parallel use of convolution and pooling operations. Also, in the Inception architecture, traditional fully-connected layers are replaced by average pooling, which leads to a significant reduction in number of parameters. MobileNet [29] is an efficient deep CNN, mainly designed for mobile vision applications. In this architecture, dense convolutional filters are replaced by separable filters. This simplification results in smaller and faster CNN models.

We replaced the last layer of the baseline CNN with a fully-connected layer with 10 neurons followed by soft-max activations (shown in Fig. 8). Baseline CNN weights are initialized by training on the ImageNet dataset [15], and then an end-to-end training on quality assessment is performed. In this paper, we discuss performance of the proposed model with various baseline CNNs.

In training, input images are rescaled to $256 \times 256$, and then a crop of size $224 \times 224$ is randomly extracted. This lessens potential over-fitting issues, especially when training on relatively small datasets (e.g. TID2013). It is worth noting that we also tried training with random crops without rescaling. However, results were not compelling. This is due to the inevitable change in image composition. Another random data augmentation in our training process is horizontal flipping of the image crops.

Our goal is to predict the distribution of ratings for a given image. Ground truth distribution of human ratings of a given image can be expressed as an empirical probability mass function $\mathbf{p} = [p_{s_1}, \ldots, p_{s_N}]$ with $s_1 \leq s_i \leq s_N$, where $s_i$ denotes the $i$th score bucket, and $N$ denotes the total number of score buckets. In both AVA and TID2013 datasets $N = 10$, in AVA, $s_1 = 1$ and $s_N = 10$, and in TID $s_1 = 0$ and $s_N = 9$. Since $\sum_{i=1}^{N} p_{s_i} = 1$, $p_{s_i}$ represents the probability of a quality score falling in the $i$th bucket. Given the distribution of ratings as $\mathbf{p}$, mean quality score is defined as $\mu = \sum_{i=1}^{N} s_i \times p_{s_i}$, and standard deviation of the score is computed as $\sigma = (\sum_{i=1}^{N} (s_i - \mu)^2 \times p_{s_i})^{1/2}$. As discussed in the previous section, one can qualitatively compare images by mean and standard deviation of scores.

Each example in the dataset consists of an image and its ground truth (user) ratings $\mathbf{p}$. Our objective is to find the probability mass function $\widehat{\mathbf{p}}$ that is an accurate estimate of $\mathbf{p}$. Next, our training loss function is discussed.

TABLE I

PERFORMANCE OF THE PROPOSED METHOD WITH VARIOUS ARCHITECTURES IN PREDICTING AVA QUALITY RATINGS [1] COMPARED TO THE STATE-OF-THE-ART. REPORTED ACCURACY VALUES ARE BASED ON CLASSIFICATION OF PHOTOS TO TWO CLASSES (COLUMN 2). LCC (LINEAR CORRELATION COEFFICIENT) AND SRCC (SPEARMAN'S RANK CORRELATION COEFFICIENT) ARE COMPUTED BETWEEN PREDICTED AND GROUND TRUTH MEAN SCORES (COLUMN 3 AND 4) AND STANDARD DEVIATION OF SCORES (COLUMN 5 AND 6). EMD MEASURES CLOSENESS OF THE PREDICTED AND GROUND TRUTH RATING DISTRIBUTIONS WITH $r = 1$ IN EQ. 1. THE ACCURACY, LCC, AND SROC VALUES ARE IN $\pm 0.3$, $\pm 0.005$, AND $\pm 0.004$ WITHIN 95% CONFIDENCE, RESPECTIVELY

| *Model* | *Accuracy (2 classes)* | *LCC (mean)* | *SRCC (mean)* | *LCC (std.dev)* | *SRCC (std.dev)* | *EMD* |
|---|---|---|---|---|---|---|
| Murray et al. [1] | 66.70% | – | – | – | – | – |
| Kao et al. [9] | 71.42% | – | – | – | – | – |
| Lu et al. [36] | 74.46% | – | – | – | – | – |
| Lu et al. [17] | 75.42% | – | – | – | – | – |
| Kao et al. [37] | 76.58% | – | – | – | – | – |
| Wang et al. [38] | 76.80% | – | – | – | – | – |
| Mai et al. [10] | 77.10% | – | – | – | – | – |
| Kong et al. [14] | 77.33% | – | 0.558 | – | – | – |
| Ma et al. [20] | 81.70% | – | – | – | – | – |
| NIMA(MobileNet) | 80.36% | 0.518 | 0.510 | 0.152 | 0.137 | 0.081 |
| NIMA(VGG16) | 80.60% | 0.610 | 0.592 | 0.205 | 0.202 | 0.052 |
| NIMA(Inception-v2) | 81.51% | 0.636 | 0.612 | 0.233 | 0.218 | 0.050 |

### A. Loss Function

Soft-max cross-entropy is widely used as training loss in classification tasks. This loss can be represented as $\sum_{i=1}^{N} -p_{s_i} \log(\widehat{p}_{s_i})$ (where $\widehat{p}_{s_i}$ denotes estimated probability of $i$th score bucket) to maximize predicted probability of the correct labels. However, in the case of ordered-classes (e.g. aesthetic and quality estimation), cross-entropy loss lacks the inter-class relationships between score buckets. One might argue that ordered-classes can be represented by a real number, and consequently, can be learned through a regression framework. Yet, it has been shown that for ordered classes, the classification frameworks can outperform regression models [21], [31]. Hou *et al.* [21] show that training on datasets with intrinsic ordering between classes can benefit from EMD-based losses. These loss functions penalize misclassifications according to class distances.

For image quality ratings, classes are inherently ordered as $s_1 < \cdots < s_N$, and $r-$norm distance between classes is defined as $\|s_i - s_j\|_r$, where $1 \leq i, j \leq N$. EMD is defined as the minimum cost to move the mass of one distribution to another. Given the ground truth and estimated probability mass functions $\mathbf{p}$ and $\widehat{\mathbf{p}}$, with $N$ ordered classes of distance $\|s_i - s_j\|_r$, the normalized Earth Mover's Distance can be expressed as [32]:

$$\mathrm{EMD}(\mathbf{p}, \widehat{\mathbf{p}}) = \left( \frac{1}{N} \sum_{k=1}^{N} |\mathrm{CDF}_{\mathbf{p}}(k) - \mathrm{CDF}_{\widehat{\mathbf{p}}}(k)|^r \right)^{1/r} \quad (1)$$

where $\mathrm{CDF}_{\mathbf{p}}(k)$ is the cumulative distribution function as $\sum_{i=1}^{k} \mathbf{p}_{s_i}$. It is worth noting that this closed-form solution requires both distributions to have equal mass as $\sum_{i=1}^{N} \mathbf{p}_{s_i} = \sum_{i=1}^{N} \widehat{\mathbf{p}}_{s_i}$. As shown in Fig. 8, our predicted quality probabilities are fed to a soft-max function to guarantee that $\sum_{i=1}^{N} \widehat{\mathbf{p}}_{s_i} = 1$. Similar to [21], in our training framework, $r$ is set as 2 to penalize the Euclidean distance between the CDFs. $r = 2$ allows easier optimization when working with gradient descent.

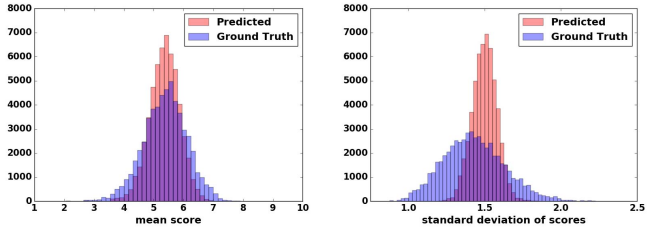



Fig. 9. Histograms of the ground truth and predicted scores using NIMA(Inception-v2) applied on our AVA test set. Left: histograms of mean scores. Right: histograms of standard deviations.

## III. EXPERIMENTAL RESULTS

We train two separate models for aesthetics and technical quality assessment on AVA, TID2013, and LIVE. For each case, we split each dataset into train and test sets, such that 20% of the data is used for testing. In this section, performance of the proposed models on the test sets are discussed and compared to the existing methods. Then, applications of the proposed technique in photo ranking and image enhancement are explored. Before moving forward, details of our implementation are explained.

The CNNs presented in this paper are implemented using TensorFlow [33], [34]. The baseline CNN weights are initialized by training on ImageNet [15], and the last fully-connected layer is randomly initialized. The weight and bias momentums are set to 0.9, and a dropout rate of 0.75 is applied on the last layer of the baseline network. The learning rate of the baseline CNN layers and the last fully-connected layers are set as $3 \times 10^{-7}$ and $3 \times 10^{-6}$, respectively. We observed that setting a low learning rate on baseline CNN layers results in easier and faster optimization when using stochastic gradient descent. Also, after every 10 epochs of training, an exponential decay with decay factor 0.95 is applied to all learning rates.

### A. Performance Comparisons

Accuracy, correlation and EMD values of our evaluations on the aesthetic assessment model on AVA are presented

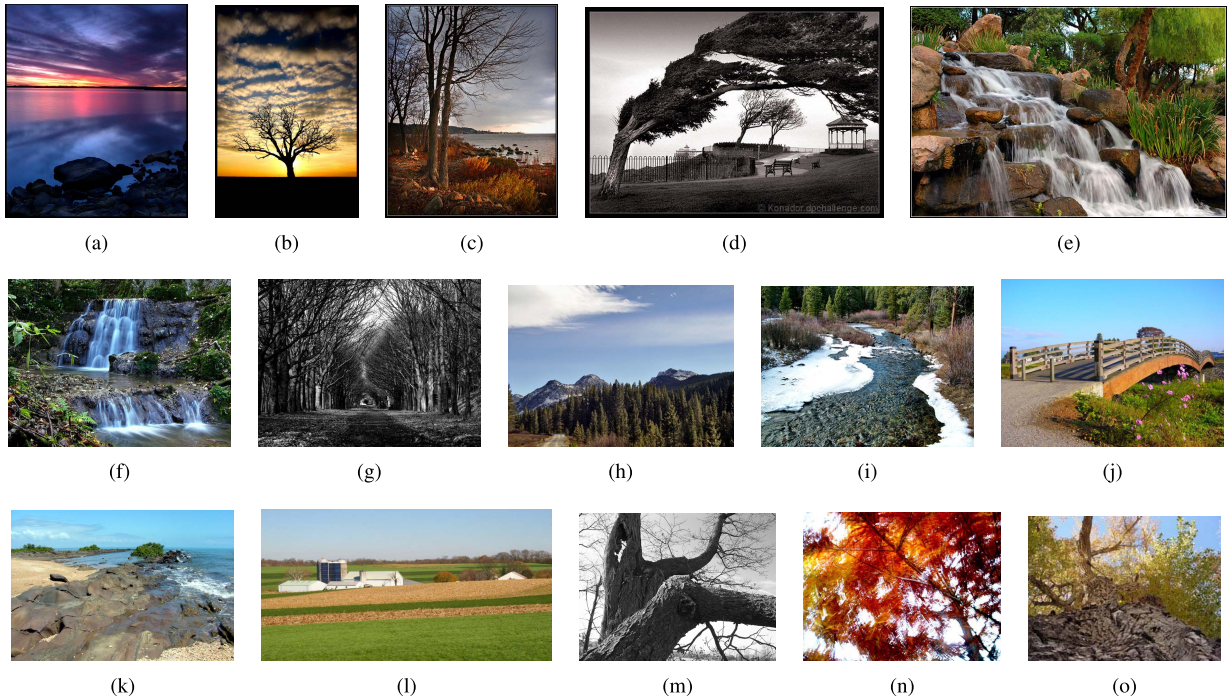


Fig. 10. Ranking some examples labelled with "*landscape*" tag from AVA dataset [1] using our proposed aesthetic assessment model NIMA(VGG16). Predicted (and ground truth) scores are shown below each image. (a) 6.38 (7.16). (b) 6.24 (6.79). (c) 6.22 (6.64). (d) 6.16 (6.93). (e) 5.92 (6.23). (f) 5.71 (5.78). (g) 5.61 (5.54). (h) 5.28 (5.32). (i) 5.11 (5.23). (j) 5.03 (5.35). (k) 4.90 (4.91). (l) 4.83 (4.89). (m) 4.77 (4.55). (n) 4.48 (3.95). (o) 3.55 (3.53).

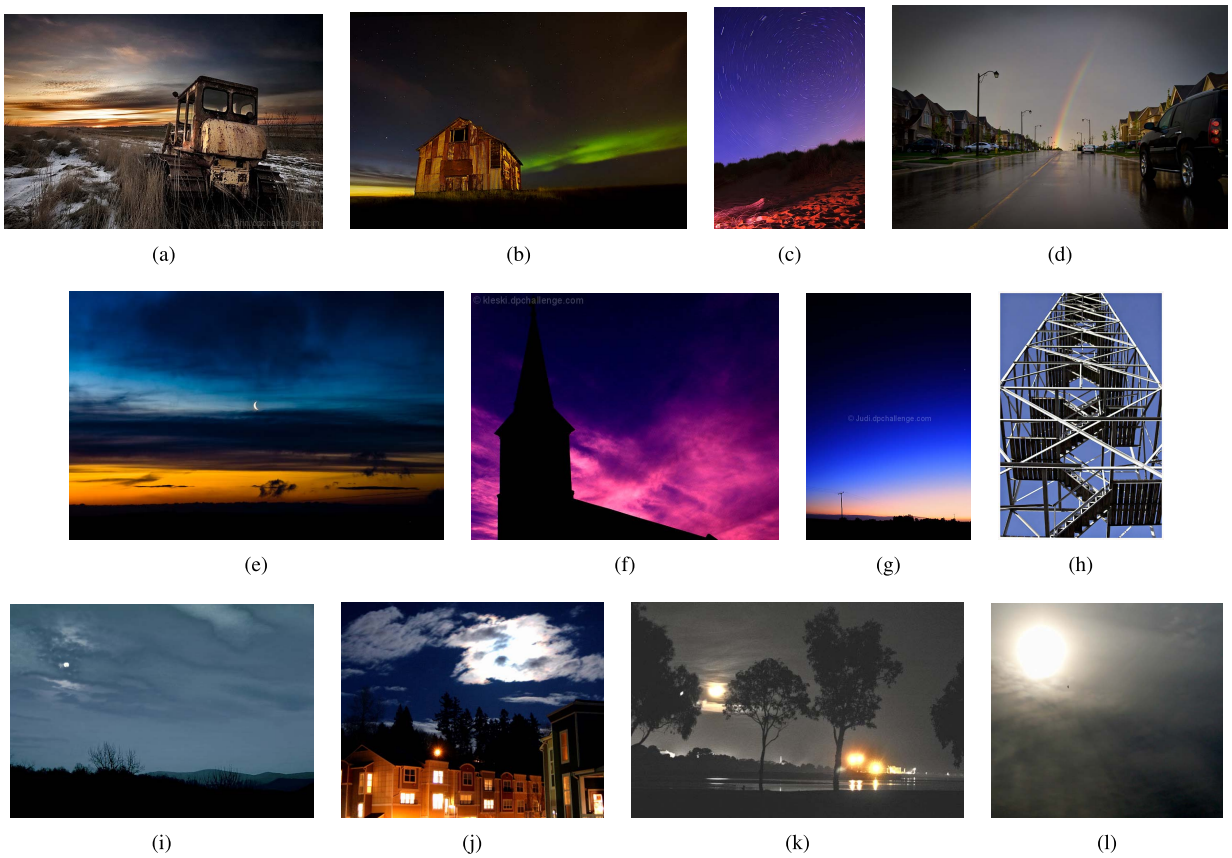


Fig. 11. Ranking some examples labelled with "*sky*" tag from AVA dataset [1] using our proposed aesthetic assessment model NIMA(Inception-v2). Predicted (and ground truth) scores are shown below each image. (a) 6.88 (7.40). (b) 6.63 (6.89). (c) 6.29 (6.59). (d) 5.86 (6.16). (e) 5.77 (5.52). (f) 5.51 (5.47). (g) 5.46 (5.38). (h) 5.24 (4.74). (i) 4.96 (4.83). (j) 4.90 (4.71). (k) 4.60 (4.59). (l) 4.53 (5.05).

in Table I. Most methods in Table I are designed to perform binary classification on the aesthetic scores, and as a result, only accuracy evaluations of two-class quality categorization are reported. In this binary classification, predicted

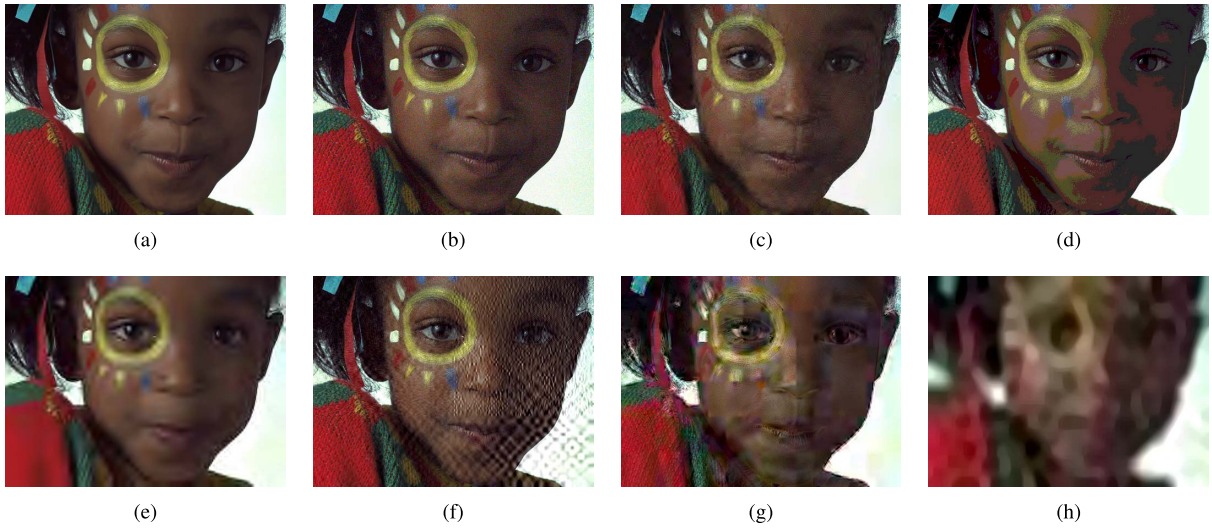



Fig. 12. Ranking some examples from TID2013 dataset [2] using our proposed quality assessment model NIMA(VGG16). Predicted (and ground truth) scores are shown below each image. (a) 5.31 (5.93). (b) 4.35 (4.64). (c) 4.00 (3.91). (d) 3.56 (3.61). (e) 3.05 (3.26). (f) 2.87 (2.86). (g) 2.33 (2.44). (h) 1.67 (0.73).

mean scores are compared to 5 as cut-off score. Images with predicted scores above the cut-off score are categorized as high quality. In two-class aesthetic categorization task, results from [20], and NIMA(Inception-v2) show the highest accuracy. Also, in terms of rank correlation, NIMA(VGG16) and NIMA(Inception-v2) outperform [14]. *NIMA is much cheaper*: [20] applies multiple VGG16 nets on image patches to generate a single quality score, whereas computational complexity of NIMA(Inception-v2) is roughly one pass of Inception-v2 (see Table V).

Our technical quality assessment model on TID2013 is compared to other existing methods in Table II. While most of these methods regress to the mean opinion score, our proposed technique predicts the distribution of ratings, as well as mean opinion score. Correlation between ground truth and results of NIMA(VGG16) are close to the state-of-the-art results in [35] and [7]. It is worth highlighting that Bianco *et al.* [7] feed multiple image crops to a deep CNN, whereas our method takes only the rescaled image.

The predicted distributions of AVA scores are presented in Fig. 9. We used NIMA(Inception-v2) model to predict the ground truth scores from our AVA test set. As can be seen, distribution of the ground truth mean scores is closely predicted by NIMA. However, predicting distribution of the ground truth standard deviations is a more challenging task. As we discussed previously, unconventionality of subject matter or style has a direct impact on score standard deviations.

TABLE II

PERFORMANCE OF THE PROPOSED METHOD WITH VARIOUS ARCHITECTURES IN PREDICTING TID2013 QUALITY RATINGS [2] COMPARED TO THE STATE-OF-THE-ART. LCC (LINEAR CORRELATION COEFFICIENT) AND SRCC (SPEARMAN'S RANK CORRELATION COEFFICIENT) ARE COMPUTED BETWEEN PREDICTED AND GROUND TRUTH MEAN SCORES (COLUMN 2 AND 3) AND STANDARD DEVIATION OF SCORES (COLUMN 4 AND 5). EMD MEASURES CLOSENESS OF THE PREDICTED AND GROUND TRUTH RATING DISTRIBUTIONS WITH $r = 1$ IN EQ. 1. THE LCC, AND SROC VALUES ARE IN $\pm 0.005$, AND $\pm 0.004$ WITHIN 95% CONFIDENCE, RESPECTIVELY

| *Model* | *LCC (mean)* | *SRCC (mean)* | *LCC (std.dev)* | *SRCC (std.dev)* | *EMD* |
|---|---|---|---|---|---|
| Kim et al. [16] | 0.80 | 0.80 | – | – | – |
| Moorthy et al. [39] | 0.89 | 0.88 | – | – | – |
| Mittal et al. [40] | 0.92 | 0.89 | – | – | – |
| Saad et al. [41] | 0.91 | 0.88 | – | – | – |
| Kottayil et al. [42] | 0.89 | 0.88 | – | – | – |
| Xu et al. [35] | 0.96 | 0.95 | – | – | – |
| Bianco et al. [7] | 0.96 | 0.96 | – | – | – |
| NIMA(MobileNet) | 0.782 | 0.698 | 0.209 | 0.181 | 0.105 |
| NIMA(VGG16) | 0.941 | 0.944 | 0.538 | 0.557 | 0.054 |
| NIMA(Inception-v2) | 0.827 | 0.750 | 0.470 | 0.468 | 0.064 |

### B. Cross Dataset Evaluation

As a cross validation test, performance of our trained models are measured on other datasets. These results are presented in Table III and Table IV. We test NIMA(Inception-v2) model trained on AVA, TID2013 [2] and LIVE [26] across all three test sets. As can be seen, on average, training on AVA dataset shows the best performance. For instance, training on AVA and testing on LIVE results in 0.552 and 0.543 linear and rank correlations, respectively. However, training on LIVE and testing on AVA leads to 0.238 and 0.2 linear and rank correlation coefficients. We believe this observation shows that NIMA models trained on AVA can generalize to other test examples more effectively, whereas training on TID2013 results in poor performance on LIVE and AVA test sets. It is worth mentioning that AVA dataset contains roughly 250 times more examples (in comparison to the LIVE dataset), which allows training NIMA models without any significant overfitting.

### C. Photo Ranking

Predicted mean scores can be used to rank photos, aesthetically. Some test photos from AVA dataset are ranked in Fig. 10

Fig. 13. Predicted aesthetic score (NIMA(VGG16)) for various parameter settings of multi-layer Laplacian technique [43]. Predicted aesthetic scores are shown below each image. (a) Input (5.52). (b) contrast compression (4.79). (c) boosting details (5.73). (d) increasing brightness (5.52). (e) increasing shadows (5.95).

TABLE III

LCC (LINEAR CORRELATION COEFFICIENT) OF NIMA(INCEPTION-V2) MODEL FOR TRAINING AND TESTING ON VARIOUS DATASETS

| Train Dataset | Test Dataset | | | Average |
|---|---|---|---|---|
| | LIVE [26] | TID2013 [2] | AVA [1] | |
| LIVE [26] | 0.698 | 0.537 | 0.238 | 0.491 |
| TID2013 [2] | 0.178 | 0.827 | 0.101 | 0.369 |
| AVA [1] | 0.552 | 0.514 | 0.636 | **0.567** |

TABLE IV

SRCC (SPEARMAN'S RANK CORRELATION COEFFICIENT) OF NIMA(INCEPTION-V2) MODEL FOR TRAINING AND TESTING ON VARIOUS DATASETS

| Train Dataset | Test Dataset | | | Average |
|---|---|---|---|---|
| | LIVE [26] | TID2013 [2] | AVA [1] | |
| LIVE [26] | 0.637 | 0.327 | 0.200 | 0.388 |
| TID2013 [2] | 0.155 | 0.750 | 0.087 | 0.331 |
| AVA [1] | 0.543 | 0.432 | 0.612 | **0.529** |

and Fig. 11. Predicted NIMA scores and ground truth AVA scores are shown below each image. Results in Fig. 10 suggest that in addition to image content, other factors such as tone, contrast and composition of photos are important aesthetic qualities. Also, as shown in Fig. 11, besides image semantics, framing and color palette are key qualities in these photos. These aesthetic attributes are closely predicted by our trained models on AVA.

Predicted mean scores are used to qualitatively rank photos in Fig. 12. These images are part of our TID2013 test set, which contain various types and levels of distortions. Comparing ground truth and predicted scores indicates that our trained model on TID2013 accurately ranks the test images.

### *D. Image Enhancement*

Quality and aesthetic scores can be used to perceptually tune image enhancement operators. In other words, maximizing

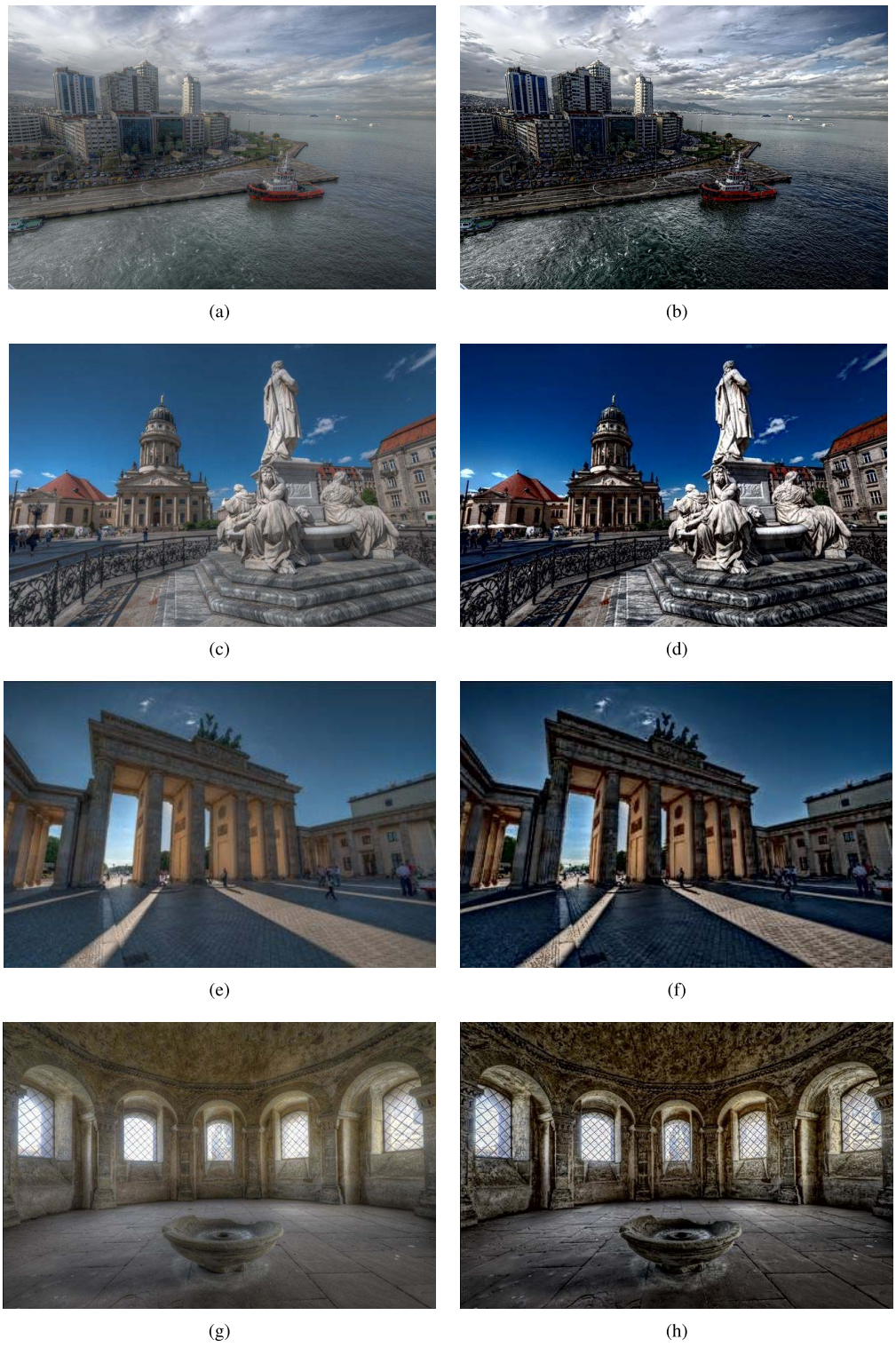


Fig. 14. Tone enhancement by multi-layer Laplacian technique [43] along with our proposed aesthetic assessment model NIMA(VGG16). Predicted aesthetic scores are shown below each image. (Input photos are downloaded from www.farbspiel-photo.com). (a) Input (5.80). (b) Enhanced (6.12). (c) Input (5.52). (d) Enhanced (6.13). (e) Input (4.87). (f) Enhanced (5.57). (g) Input (5.59). (h) Enhanced (5.98).

NIMA score as a prior can increase the likelihood of enhancing perceptual quality of an image. Typically, parameters of enhancement operators such as image denoising and contrast enhancement are selected by extensive experiments under various photographic conditions. Perceptual tuning could be quite expensive and time consuming, especially when human opinion is required. In this section, our proposed models are used to tune a tone enhancement method [43], and an image denoiser [44]. A more detailed treatment is presented in [23].

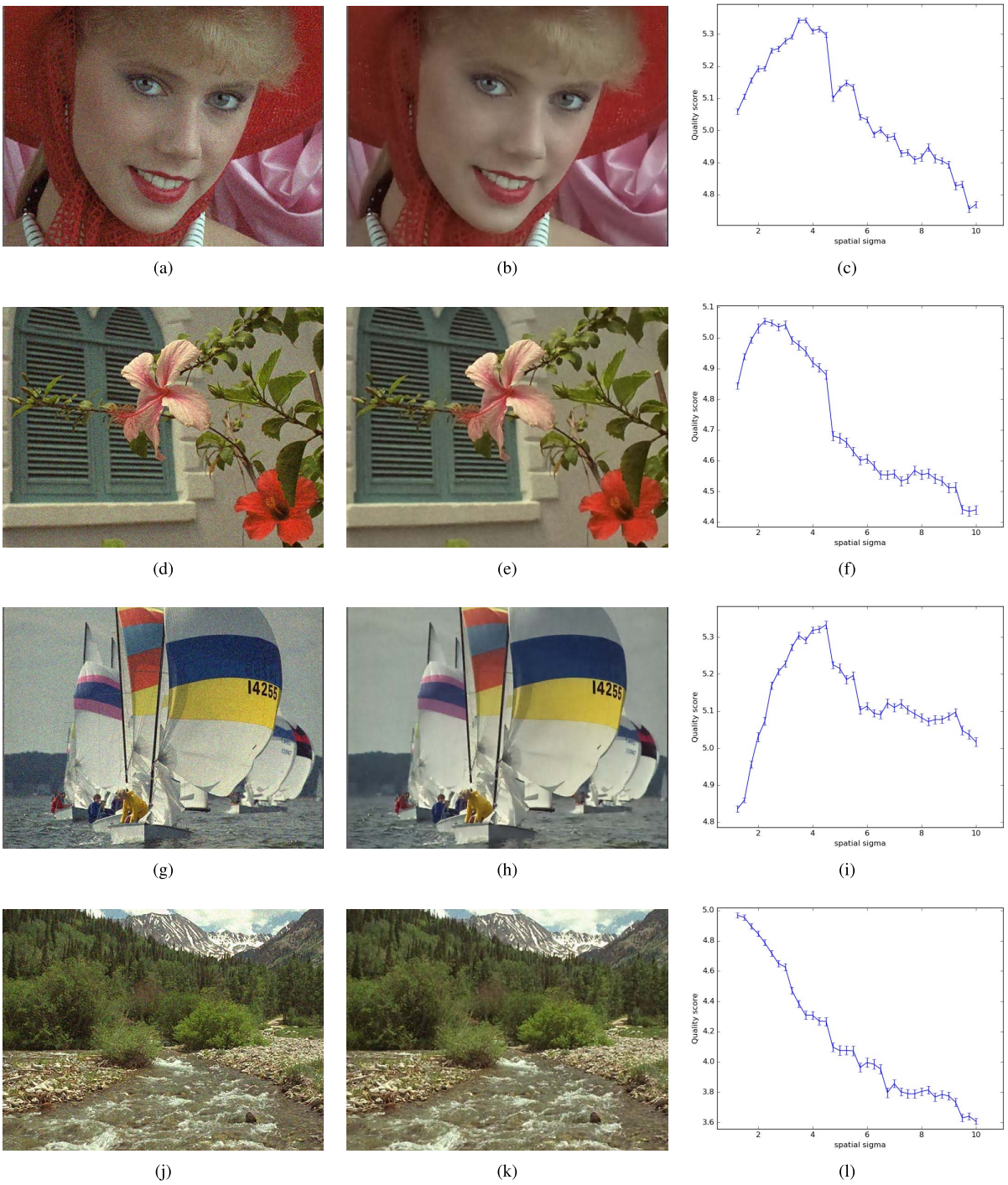



Fig. 15. Tuning spatial parameter of Turbo denoising [44] by using our proposed quality assessment model NIMA(VGG16). Standard deviation of the additive white Gaussian noise is set as 30. Denoised results are shown for maximum quality score. (a) Noisy Input. (b) Optimized (denoising parameter=3.75). (c) Quality score vs. denoising parameter. (d) Noisy Input. (e) Optimized (denoising parameter=2.25). (f) Quality score vs. denoising parameter. (g) Noisy Input. (h) Optimized (denoising parameter=4.50). (i) Quality score vs. denoising parameter. (j) Noisy Input. (k) Optimized (denoising parameter=1.25). (l) Quality score vs. denoising parameter.

The multi-layer Laplacian technique [43] enhances local and global contrast of images. Parameters of this method control the amount of detail, shadow, and brightness of an image. Fig. 13 shows a few examples of the multi-layer Laplacian with different sets of parameters. We observed that the predicted aesthetic ratings from training on the AVA dataset can be improved by contrast adjustments. Consequently, our model is able to guide the multi-layer Laplacian filter to find aesthetically near-optimal settings of its parameters. Examples of this type of image editing are represented in Fig. 14, where a combination of detail, shadow and brightness change is applied on each image. In each example, 6 levels of detail boost, 11 levels of shadow change, and 11 levels of brightness change account for a total of 726 variations.

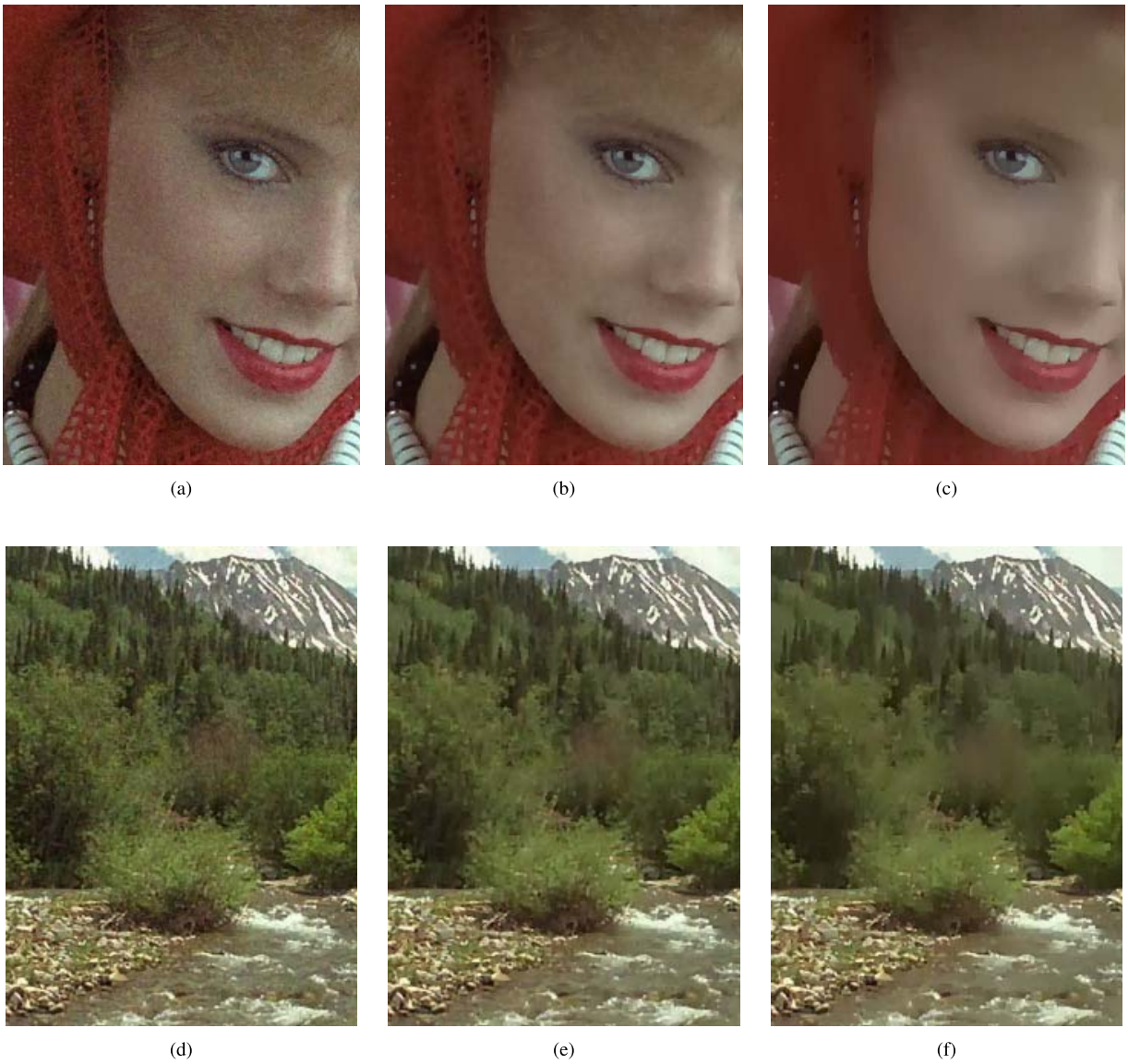


Fig. 16. Effect of Turbo denoising [44] on our predicted quality scores. Input noisy images are shown in Fig. 15. (a) denoising parameter=1.25, score=5.06. (b) denoising parameter=3.0, score=5.15. (c) denoising parameter=9.75, score=4.76. (d) denoising parameter=1.25, score=4.97. (e) denoising parameter=3.0, score=4.62. (f) denoising parameter=9.75, score=3.64.

The aesthetic assessment model tends to prefer high contrast images with boosted details. This is consistent with the ground truth results from AVA illustrated in Fig. 10.

Turbo denoising [44] is a technique which uses the domain transform [45] as its core filter. Performance of Turbo denoising depends on spatial and range smoothing parameters, and consequently, proper tuning of these parameters can effectively boost performance of the denoiser. We observed that varying the spatial smoothing parameter makes the most significant perceptual difference, and as a result, we use our quality assessment model trained on TID2013 dataset to tune this denoiser. Application of our no-reference quality metric as a prior in image denoising is similar to the work of Zhu and Milanfar [46], [47]. Our results are shown in Fig. 15. Additive white Gaussian noise with standard deviation 30 is added to the clean image, and Turbo denoising with various spatial parameters is used to denoise the noisy image. To reduce the score deviation, 50 random crops are extracted from denoised image. These scores are averaged to obtain the plots illustrated in Fig. 15. As can be seen, although the same amount of noise is added to each image, maximum quality scores correspond to different denoising parameters in each example. For relatively smooth images such as (a) and (g), optimal spatial parameter of Turbo denoising is higher (which implies stronger smoothing) than the textured image in (j). This is probably due to the relatively high signal-to-noise ratio of (j). In other words, the quality assessment model tends to respect textures and avoid over-smoothing of details. Effect of the denoising parameter can be visually inspected in Fig. 16. While the denoised result in Fig. 16 (a) is under-smoothed, (c), (e) and (f) show undesirable over-smoothing effects. The predicted quality scores validate this perceptual observation.

TABLE V

COMPARISON OF THE PROPOSED QUALITY ASSESSMENT TECHNIQUE WITH VARIOUS CNN ARCHITECTURES. AVERAGE TIMINGS ARE REPORTED IN ms FOR XEON INTEL CPU @ 3.5 GHz, AND NVIDIA QUADRO K620 GPU. TIMINGS ARE REPORTED FOR APPLYING NIMA MODELS ON IMAGES OF SIZE $224 \times 224 \times 3$

| *Model* | *Million Parameters* | *Billion Flops* | *CPU Timing (ms)* | *GPU Timing (ms)* |
|---|---|---|---|---|
| NIMA(MobileNet) | 3.22 | 1.29 | 30.45 | 20.23 |
| NIMA(Inception-v2) | 10.16 | 4.37 | 70.49 | 39.11 |
| NIMA(VGG16) | 134.30 | 31.62 | 150.34 | 85.76 |

### *E. Computational Costs*

Computational complexity of NIMA models are compared in Table V. Our inference TensorFlow implementation is tested on an Intel Xeon CPU @ 3.5 GHz with 32 GB memory and 12 cores, and NVIDIA Quadro K620 GPU. Timings of one pass of NIMA models on an image of size $224 \times 224 \times 3$ are reported in Table V. Evidently, NIMA(MobileNet) is significantly lighter and faster than other models. This comes at the expense of a slight performance drop (shown in Table I and Table II).

## IV. CONCLUSION

In this work we introduced a CNN-based image assessment method, which can be trained on both aesthetic and pixel-level quality datasets. Our models effectively predict the distribution of quality ratings, rather than just the mean scores. This leads to a more accurate quality prediction with higher correlation to the ground truth ratings. We trained two models for high level aesthetics and low level technical qualities, and utilized them to steer parameters of a few image enhancement operators. Our experiments suggest that these models are capable of guiding denoising and tone enhancement to produce perceptually superior results.

As part of our future work, we will exploit the trained models on other image enhancement applications. Our current experimental setup requires the enhancement operator to be evaluated multiple times. This limits real-time application of the proposed method. One might argue that in case of an enhancement operator with well-defined derivatives, using NIMA as the loss function is a more efficient approach.

## ACKNOWLEDGMENT

The authors would like to thank Dr. Pascal Getreuer for valuable discussions and helpful advice on approximation of score distributions.

## REFERENCES

[1] N. Murray, L. Marchesotti, and F. Perronnin, "AVA: A large-scale database for aesthetic visual analysis," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2012, pp. 2408–2415.

[2] N. Ponomarenko *et al.*, "Color image database TID2013: Peculiarities and preliminary results," in *Proc. 4th Eur. Workshop Vis. Inf. Process. (EUVIP)*, 2013, pp. 106–111.

[3] Z. Wang, A. C. Bovik, H. R. Sheikh, and E. P. Simoncelli, "Image quality assessment: From error visibility to structural similarity," *IEEE Trans. Image Process.*, vol. 13, no. 4, pp. 600–612, Apr. 2004.

[4] W. Xue, L. Zhang, and X. Mou, "Learning without human scores for blind image quality assessment," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Jun. 2013, pp. 995–1002.

[5] L. Kang, P. Ye, Y. Li, and D. Doermann, "Convolutional neural networks for no-reference image quality assessment," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Jun. 2014, pp. 1733–1740.

[6] S. Bosse, D. Maniry, T. Wiegand, and W. Samek, "A deep neural network for image quality assessment," in *Proc. IEEE Int. Conf. Image Process. (ICIP)*, Sep. 2016, pp. 3773–3777.

[7] S. Bianco, L. Celona, P. Napoletano, and R. Schettini. (2016). "On the use of deep learning for blind image quality assessment." [Online]. Available: https://arxiv.org/abs/1602.05531

[8] X. Lu, Z. Lin, X. Shen, R. Mech, and J. Z. Wang, "Deep multi-patch aggregation network for image style, aesthetics, and quality estimation," in *Proc. IEEE Int. Conf. Comput. Vis.*, Jun. 2015, pp. 990–998.

[9] Y. Kao, C. Wang, and K. Huang, "Visual aesthetic quality assessment with a regression model," in *Proc. IEEE Int. Conf. Image Process. (ICIP)*, Sep. 2015, pp. 1583–1587.

[10] L. Mai, H. Jin, and F. Liu, "Composition-preserving deep photo aesthetics assessment," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Jun. 2016, pp. 497–506.

[11] B. Jin, M. V. O. Segovia, and S. Süsstrunk, "Image aesthetic predictors based on weighted CNNs," in *Proc. IEEE Int. Conf. Image Process. (ICIP)*, Sep. 2016, pp. 2291–2295.

[12] H. R. Sheikh, Z. Wang, L. Cormack, and A. C. Bovik. (2005). *Live Image Quality Assessment Database Release 2*. [Online]. Available: http://live.ece.utexas.edu/research/quality

[13] E. C. Larson and D. M. Chandler, "Most apparent distortion: Full-reference image quality assessment and the role of strategy," *J. Electron. Imag.*, vol. 19, no. 1, p. 011006, 2010.

[14] S. Kong, X. Shen, Z. Lin, R. Mech, and C. Fowlkes, "Photo aesthetics ranking network with attributes and content adaptation," in *Proc. Eur. Conf. Comput. Vis.*, 2016, pp. 662–679.

[15] A. Krizhevsky, I. Sutskever, and G. E. Hinton, "ImageNet classification with deep convolutional neural networks," in *Proc. Adv. Neural Inf. Process. Syst.*, 2012, pp. 1097–1105.

[16] J. Kim, H. Zeng, D. Ghadiyaram, S. Lee, L. Zhang, and A. C. Bovik, "Deep convolutional neural models for picture-quality prediction: Challenges and solutions to data-driven image quality assessment," *IEEE Signal Process. Mag.*, vol. 34, no. 6, pp. 130–141, Nov. 2017.

[17] X. Lu, Z. L. Lin, H. Jin, J. Yang, and J. Z. Wang, "Rating image aesthetics using deep learning," *IEEE Trans. Multimedia*, vol. 17, no. 11, pp. 2021–2034, Nov. 2015.

[18] K. Simonyan and A. Zisserman. (2014). "Very deep convolutional networks for large-scale image recognition." [Online]. Available: https://arxiv.org/abs/1409.1556

[19] H. Zeng, L. Zhang, and A. C. Bovik. (2017). "A probabilistic quality representation approach to deep blind image quality prediction." [Online]. Available: https://arxiv.org/abs/1708.08190

[20] S. Ma, J. Liu, and C. W. Chen, "A-lamp: Adaptive layout-aware multi-patch deep convolutional neural network for photo aesthetic assessment," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jul. 2017, pp. 722–731.

[21] L. Hou, C.-P. Yu, and D. Samaras. (2016). "Squared earth mover's distance-based loss for training deep neural networks." [Online]. Available: https://arxiv.org/abs/1611.05916

[22] R. Zhang, P. Isola, A. A. Efros, E. Shechtman, and O. Wang, "The unreasonable effectiveness of deep features as a perceptual metric," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2018.

[23] H. Talebi and P. Milanfar, "Learned perceptual image enhancement," in *Proc. IEEE Int. Conf. Comput. Photograph. (ICCP)*, May 2018.

[24] K. Gu, G. T. Zhai, and M. Lin, "The analysis of image contrast: From quality assessment to automatic enhancement," *IEEE Trans. Cybern.*, vol. 46, no. 1, pp. 284–297, Jan. 2015.

[25] Kodak. *Kodak Lossless True Color Image Suite*. Accessed: Jan. 2017. [Online]. Available: http://r0k.us/graphics/kodak/

[26] D. Ghadiyaram and A. C. Bovik, "Massive online crowdsourced study of subjective and objective picture quality," *IEEE Trans. Image Process.*, vol. 25, no. 1, pp. 372–387, Jan. 2016.

[27] T. M. Cover and J. A. Thomas, *Elements of Information Theory*. Hoboken, NJ, USA: Wiley, 2012.

[28] C. Szegedy, V. Vanhoucke, S. Ioffe, J. Shlens, and Z. Wojna, "Rethinking the inception architecture for computer vision," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Jun. 2016, pp. 2818–2826.

[29] A. G. Howard *et al.*, (2017). "MobileNets: Efficient convolutional neural networks for mobile vision applications." [Online]. Available: https://arxiv.org/abs/1704.04861
[30] S. Ioffe and C. Szegedy, "Batch normalization: Accelerating deep network training by reducing internal covariate shift," in *Proc. Int. Conf. Mach. Learn.*, 2015, pp. 448–456.
[31] P. Golik, P. Doetsch, and H. Ney, "Cross-entropy vs. squared error training: A theoretical and experimental comparison," in *Proc. Interspeech*, 2013, pp. 1756–1760.
[32] E. Levina and P. Bickel, "The earth mover's distance is the mallows distance: Some insights from statistics," in *Proc. 8th IEEE Int. Conf. Comput. Vis. (ICCV)*, vol. 2. Jul. 2001, pp. 251–256.
[33] M. Abadi *et al.* (2016). "TensorFlow: Large-scale machine learning on heterogeneous distributed systems." [Online]. Available: https://arxiv.org/abs/1603.04467
[34] M. Abadi *et al.*, "TensorFlow: A system for large-scale machine learning," in *Proc. 12th USENIX Symp. Operat. Syst. Design Implement. (OSDI)*, Savannah, GA, USA, 2016, pp. 265–283.
[35] J. Xu, P. Ye, Q. Li, H. Du, Y. Liu, and D. Doermann, "Blind image quality assessment based on high order statistics aggregation," *IEEE Trans. Image Process.*, vol. 25, no. 9, pp. 4444–4457, Sep. 2016.
[36] X. Lu, Z. Lin, H. Jin, J. Yang, and J. Z. Wang, "Rapid: Rating pictorial aesthetics using deep learning," in *Proc. 22nd ACM Int. Conf. Multimedia*, 2014, pp. 457–466.
[37] Y. Kao, R. He, and K. Huang. (2016). "Visual aesthetic quality assessment with multi-task deep learning." [Online]. Available: https://arxiv.org/abs/1604.04970
[38] Z. Wang, S. Chang, F. Dolcos, D. Beck, D. Liu, and T. S. Huang. (2016). "Brain-inspired deep networks for image aesthetics assessment." [Online]. Available: https://arxiv.org/abs/1601.04155
[39] A. K. Moorthy and A. C. Bovik, "Blind image quality assessment: From natural scene statistics to perceptual quality," *IEEE Trans. Image Process.*, vol. 20, no. 12, pp. 3350–3364, Dec. 2011.
[40] A. Mittal, A. K. Moorthy, and A. C. Bovik, "No-reference image quality assessment in the spatial domain," *IEEE Trans. Image Process.*, vol. 21, no. 12, pp. 4695–4708, Dec. 2012.
[41] M. A. Saad, A. C. Bovik, and C. Charrier, "Blind image quality assessment: A natural scene statistics approach in the DCT domain," *IEEE Trans. Image Process.*, vol. 21, no. 8, pp. 3339–3352, Aug. 2012.
[42] N. K. Kottayil, I. Cheng, F. Dufaux, and A. Basu, "A color intensity invariant low-level feature optimization framework for image quality assessment," *Signal, Image Video Process.*, vol. 10, no. 6, pp. 1169–1176, 2016.
[43] H. Talebi and P. Milanfar, "Fast multilayer Laplacian enhancement," *IEEE Trans. Comput. Imag.*, vol. 2, no. 4, pp. 496–509, Dec. 2016.
[44] T.-S. Wong and P. Milanfar, "Turbo denoising for mobile photographic applications," in *Proc. IEEE Int. Conf. Image Process. (ICIP)*, Sep. 2016, pp. 988–992.
[45] E. S. L. Gastal and M. M. Oliveira, "Domain transform for edge-aware image and video processing," *ACM Trans. Graph.*, vol. 30, no. 4, p. 69, 2011.
[46] X. Zhu and P. Milanfar, "A no-reference sharpness metric sensitive to blur and noise," in *Proc. Int. Workshop Qual. Multimedia Exper. (QoMEx)*, Jul. 2009, pp. 64–69.
[47] X. Zhu and P. Milanfar, "Automatic parameter selection for denoising algorithms using a no-reference measure of image content," *IEEE Trans. Image Process.*, vol. 19, no. 12, pp. 3116–3132, Dec. 2010.

**Hossein Talebi** received the B.S. and M.S. degrees in electrical engineering from the Isfahan University of Technology, Iran, and the Ph.D. degree in electrical engineering from the University of California at Santa Cruz, Santa Cruz, CA, USA. Since 2015, he has been with Google Research, Mountain View, CA, USA, where he involved in computational imaging, image processing, and machine learning problems.

**Peyman Milanfar** (F'10) received the bachelor's degree in electrical engineering and mathematics from the University of California (UC) at Berkeley, Berkeley, and the M.S. and Ph.D. degrees in electrical engineering from the Massachusetts Institute of Technology. He was a Professor of electrical engineering at UC Santa Cruz from 1999 to 2014. He was an Associate Dean for research with the School of Engineering from 2010 to 2012. From 2012 to 2014, he was on leave at Google-x, where he helped to develop the imaging pipeline for Google Glass. He currently leads the Computational Imaging Team in Google Research. He holds 12 U.S. patents, several of which are commercially licensed. He founded MotionDSP, which was acquired by Cubic Inc. (NYSE:CUB). He has been a keynote speaker at numerous technical conferences, including the Picture Coding Symposium, SIAM Imaging Sciences, SPIE, and the International Conference on Multimedia. Along with his students, he received several best paper awards from the IEEE Signal Processing Society. He is a Distinguished Lecturer of the IEEE Signal Processing Society for contributions to inverse problems and super-resolution in imaging.