2019 Eleventh International Conference on Quality of Multimedia Experience (QoMEX)

SUR-Net: Predicting the Satisfied User Ratio Curve for Image Compression with Deep Learning

Chunling Fan[1,2], Hanhe Lin[3], Vlad Hosu[3], Yun Zhang[2], Qingshan Jiang[2], Raouf Hamzaoui[4], and Dietmar Saupe[3]

[1]Shenzhen College of Advanced Technology, University of Chinese Academy of Sciences, China
[2]Shenzhen Institutes of Advanced Technology, Chinese Academy of Sciences, China
[3]Department of Computer and Information Science, University of Konstanz, Germany
[4]School of Engineering and Sustainable Development, De Montfort University, UK

***Abstract*—The Satisfied User Ratio (SUR) curve for a lossy image compression scheme, e.g., JPEG, characterizes the probability distribution of the Just Noticeable Difference (JND) level, the smallest distortion level that can be perceived by a subject. We propose the first deep learning approach to predict such SUR curves. Instead of the direct approach of regressing the SUR curve itself for a given reference image, our model is trained on pairs of images, original and compressed. Relying on a Siamese Convolutional Neural Network (CNN), feature pooling, a fully connected regression-head, and transfer learning, we achieved a good prediction performance. Experiments on the MCL-JCI dataset showed a mean Bhattacharyya distance between the predicted and the original JND distributions of only 0.072.**

***Index Terms*—Satisfied User Ratio, Just Noticeable Difference, Convolutional Neural Network, Deep Learning**

I. Introduction

Image compression is typically used to meet constraints on transmission bandwidth and storage space. The compressed image quality is quantitatively determined by encoding parameters, e.g., Quality Factor (QF) in JPEG compression. When images are compressed, artifacts such as blocking and ringing appear at low bit rates. Viewers' visual experience may be degraded because of these artifacts. The Satisfied User Ratio (SUR) is the fraction of viewers that do not perceive any distortion when comparing the original image to its compressed version. The SUR is important in many real-world applications. Different applications need to satisfy different percentages of users. E.g., at least 75% of customers may need to be satisfied in entertainment applications whereas the percentage may be different in other applications.

Determining the SUR for compressed images is a challenging task. The conventional method consists of three steps. First, we collect a number of pristine images and artificially distort them multiple times with increasing distortion levels, using the image compression scheme. Next, for each of the sequences of distorted images, we ask a group of subjects to identify the smallest distortion level that they can perceive. People cannot notice the distortion until it reaches a certain minimum level. This Just Noticeable Difference (JND) level is different from one person to another due to their physiological and visual attention mechanisms. Finally, we obtain the overall SUR for each image by statistical analysis such as a Gaussian model. Following this procedure, several subjective quality studies have resulted in JND-based image and video databases, e.g., MCL-JCI [1], MCL-JCV [2], VideoSet [3], and SIAT-JSSI [4]. However, subjective visual quality assessment studies are time-consuming and expensive. In contrast, objective, i.e., algorithmic, SUR estimation can work in real-time and at no extra cost.

978-1-5386-8212-8/19/$31.00 ©2019 IEEE

In recent years, deep learning has made tremendous progress in computer vision tasks such as image classification [5] [6], object detection [7] [8], and Image Quality Assessment (IQA) [9] [10]. Instead of carefully designing handcrafted features, deep learning-based methods automatically discover representations from raw image data that are most suitable for the specific tasks, hence improve the performance significantly. Inspired by these works, we propose a novel deep learning approach to predict the SUR curve for compressed images.

The main contributions of our work are as follows:

1) We propose a deep learning architecture which can predict the SUR of compressed images automatically. To the best of our knowledge, this is the first work of its kind.
2) We model the prediction of the SUR as a regression problem. A key technical aspect of our model is the use of a new type of full-reference IQA model for a different purpose than quality assessment, in this case predicting points on an SUR curve.
3) We improve the performance of our model by using transfer learning from a similar prediction task. First, we train the proposed architecture independently on a compressed image quality assessment task and then fine-tune it as our SUR-Net.

II. Related works

Existing JND research can be classified into subjective quality assessment studies and mathematical modeling. Jin *et al.* [1] conducted a subjective test on JND for JPEG compressed images and built a JND-based image dataset called MCL-JCI. They found that humans can distinguish only a few discriminative quality levels (5 to 7) for an image. A staircase quality function for each image was then generated using a Gaussian mixture model from the JND samples. Wang *et*

al. [2] conducted subjective tests on JND for compressed videos using H.264/AVC coding. They built a JND-based video dataset called MCL-JCV. They collected JND samples from 50 subjects and generated a staircase quality function for each video. Wang *et al.* [3] built a large-scale JND-based video dataset called VideoSet. They adopted a binary search procedure for locating the JND. They found the first three JND levels and generated an SUR curve for each video sequence. These generated JND-based datasets can be used as benchmarks for future research.

Mathematical modeling focus on JND and SUR prediction. Huang *et al.* [11] proposed a Support Vector Regression (SVR)-based model to predict the mean JND value of HEVC encoded videos. Wang *et al.* [12] first extracted a group of handcrafted features from videos, including quality degradation features and spatial-temporal randomness features. Then they used the features to train an SVR-based model to predict the SUR. Wang *et al.* [13] extended the framework in [12] to predict the second and third JND points. However, the success of this approach highly depends on the ability to design suitable features.

III. Definitions

We consider a lossy image compression scheme that produces a monotonically increasing distortion level as a function of a discrete encoding parameter. In JPEG, for example, the distortion may be the mean squared error and the parameter may be $n = 101 - \text{QF}$, where $\text{QF} \in \{1, \ldots, 100\}$ is the JPEG quality factor. Thus, $n = 1$ gives the smallest and $n = 100$ the largest distortion level.

Definition 1 (First JND level). The first JND level for a given image is a random variable whose value is the smallest distortion level that can be perceived by an observer.

Definition 2 (SUR function and curve). The SUR function is the complementary cumulative distribution function of the first JND level. The graph of this function is called the SUR curve.

The SUR function SUR gives the proportion of the population for which the first JND level is greater than a given value. That is, $\text{SUR}(x) = Pr(\text{JND} > x)$ where JND is the first JND level random variable.

Since the range of the first JND level is the finite set $\{1, 2, \ldots, N\}$, where N is the number of distortion levels, the SUR function is a monotonically decreasing step function.

The SUR curve can be used to determine the highest distortion level for which a given proportion of the population is satisfied (in the sense that it cannot perceive it). If we set this ratio to 0.75, as suggested in [12], we can define a first JND level for the whole population as follows.

Definition 3 (75% JND). The 75% JND is the largest value of the first JND level for which the SUR function is greater than or equal to 0.75.

Finding the first JND level is time-consuming and expensive as it requires subjective quality assessment tests to compare the original image with its distorted versions. In Wang *et al.* [3], a robust binary search algorithm is proposed to speed up the procedure.

In [1], it was assumed that the JND distribution for an image is a Gaussian mixture with N components. For simplicity, let us assume that the first JND is normally distributed ($N = 1$) with mean μ and variance σ^2 (as in [2] for the case of video coding). Then the SUR function is

$$\bar{\Phi}(x|\mu, \sigma^2) = 1 - \int_{-\infty}^{x} \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(s-\mu)^2}{2\sigma^2}} ds \qquad (1)$$

where only the two parameters μ and σ^2 need to be determined. Fig. 1 shows an SUR curve and the 75% JND under the normality assumption.

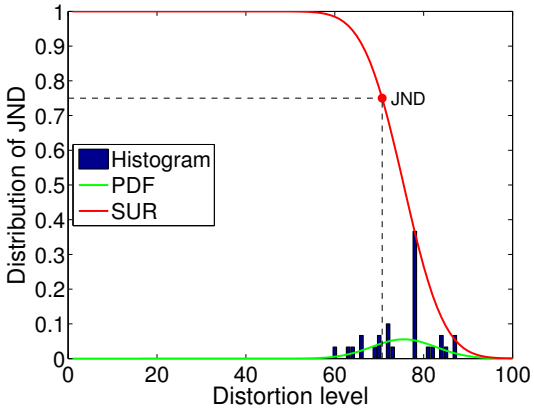


Fig. 1: Example of SUR curve and 75% JND. The data is from the first source image in the MCL-JCI dataset [1].

IV. Deep learning for SUR prediction

Let $I_1[0], I_2[0], \ldots, I_K[0]$ be a large training set of K pristine reference images. For each pristine image $I_k[0]$, $k \in \{1, \ldots, K\}$, we associate the N distorted images $I_k[n]$, $n = 1, \ldots, N$ corresponding to the N distortion levels $n = 1, \ldots, N$.

Let SUR_k denote the SUR function of image $I_k[0]$. Our objective is to find a deep regression model f_θ parameterized by θ such that:

$$f_\theta(I_k[0], I_k[n]) \approx \text{SUR}_k(n), \quad k = 1, \ldots, K, \; n = 1, \ldots, N.$$

The architecture of the proposed deep regression model is illustrated in Fig. 2. A pair of images, i.e., pristine and distorted, are fed into a Siamese network that uses an InceptionV3 [5] CNN body with shared weights. The network body is truncated, such that the global average pooling layer and the final fully-connected layer are removed. Each branch of the Siamese network yields feature maps with a depth of 2048. For each feature map, we apply min, max, and average pooling in the same size as the feature maps, yielding 2048-dimensional global feature vectors $\mathbf{f}_{\min}$, $\mathbf{f}_{\max}$, and $\mathbf{f}_{\text{avg}}$. Then we calculate $\Delta\mathbf{f}_{\min}$, $\Delta\mathbf{f}_{\max}$, and $\Delta\mathbf{f}_{\text{avg}}$, corresponding to feature vector differences between the distorted images $I_k[n]$ and the pristine image $I_k[0]$, e.g., $\Delta\mathbf{f}_{\min} = \mathbf{f}_{\min}(I_k[n]) - \mathbf{f}_{\min}(I_k[0])$. By concatenating global feature vectors $\mathbf{f}_{\min}(I_k[0])$, $\mathbf{f}_{\max}(I_k[0])$,

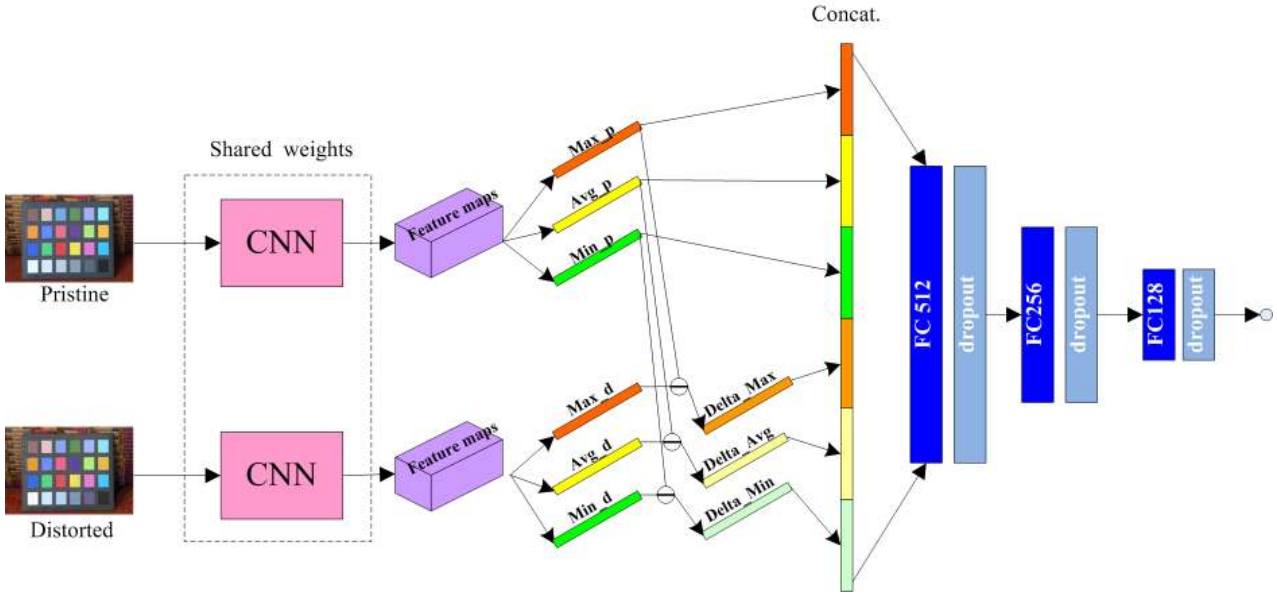


Fig. 2: Proposed architecture for SUR prediction.

$\mathbf{f}_{\text{avg}}(I_k[0])$ from the pristine image and feature vector differences $\Delta\mathbf{f}_{\min}$, $\Delta\mathbf{f}_{\max}$, $\Delta\mathbf{f}_{\text{avg}}$, we obtain a 12,288-dimensional vector. The latter is passed to three fully-connected (FC) layers with 512, 256, and 128 neurons, respectively, where each FC layer is followed by a dropout layer to avoid over-fitting. The output layer is linear with one neuron to predict values on the SUR curve. Given the training data:

$$\{(I_k[0], I_k[n], \text{SUR}_k(n)) \mid k = 1, \ldots, K,\ n = 1, \ldots, N\},$$

our objective is to minimize the Mean Absolute Error (MAE) loss function:

$$L = \frac{1}{KN} \sum_{k=1}^{K} \sum_{n=1}^{N} \left| f_\theta(I_k[0], I_k[n]) - \text{SUR}_k(n) \right|. \quad (2)$$

V. Prediction of the SUR curve and the JND

For any source image $I[0]$, together with its distorted versions $I[1], \ldots, I[n]$, a sequence of predicted satisfied user ratios $\text{SUR}(1), \ldots, \text{SUR}(N)$ is obtained from the network. Assuming that the JND is normally distributed, we estimate the mean μ and variance σ^2 by least squares fitting:

$$(\hat{\mu}, \hat{\sigma}^2) = \arg\min_{\mu, \sigma^2} \sum_{n=1}^{N} \left| \bar{\Phi}(n|\mu, \sigma^2) - \text{SUR}(n) \right|^2. \quad (3)$$

The fitted SUR curve is given by $\bar{\Phi}(x|\hat{\mu}, \hat{\sigma}^2)$ as in Equation (1).

VI. Experimental results and analysis

A. Setup

In our experiments, we used the MCL-JCI dataset [1] to evaluate the performance of the proposed method. The dataset contains 50 pristine images with a resolution of 1920×1080. Each pristine image was encoded 100 times by a JPEG encoder with QF decreasing from 100 to 1, corresponding to distortion levels 1 to 100. Thus, there are 5,050 images in total.

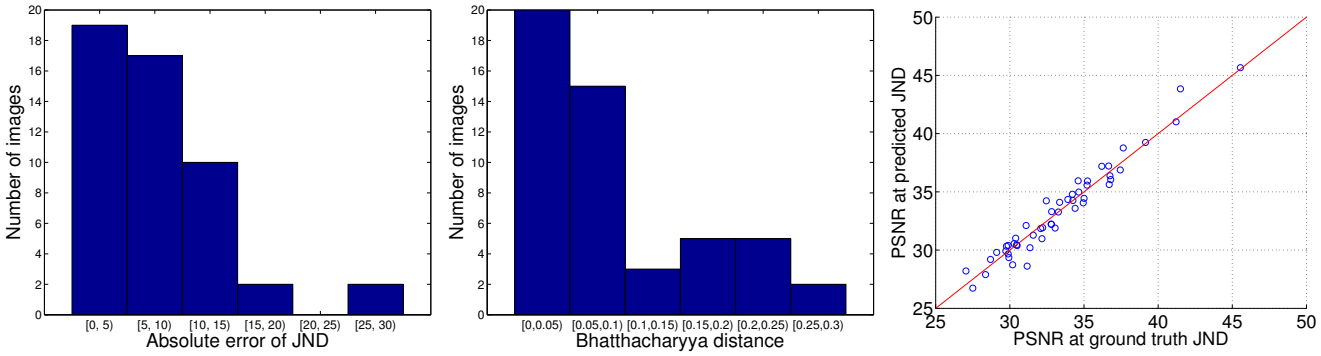
The annotation provided for the image sequences in MCL-JCI and for each of the $M = 30$ participants of the study [1] is the QF value corresponding to the first JND level (and also those of the second, third, etc.). For each source image $I_k[0]$ $(k = 1, \ldots, 50)$ in the MCL-JCI dataset, we computed the empirical SUR as an estimate of $\text{SUR}_k(n)$ at each distortion level $n = 1, \ldots, 100$ by m_n/M where m_n is the number of participants in the study whose first JND level was larger than n. We then used (3) to fit the probabilistic SUR model (1) to these empirical values and obtained the ground truth for μ and σ. Finally, we sampled the fitted SUR model to derive the target values $\text{SUR}_k(n), k = 1, \ldots, 50, n = 1, \ldots, 100$ for the deep learning algorithm.

k-fold ($k = 10$) cross validation was used to evaluate the performance. Specifically, the dataset was divided into 10 subsets, each containing five pristine images and all 500 distorted versions of them. Each time, one subset was kept as a test set, and the remaining nine subsets were used for training and validation. The overall result was the average of 10 test results.

The Adam optimizer [14] was used to train SUR-Net with the default parameters $\beta_1 = 0.9$, $\beta_2 = 0.999$, and a custom learning rate α. We set $\alpha = 10^{-5}$ and trained for 30 epochs. In the training process, we monitored the MAE loss on the validation set and saved the best performing model. Our implementation used the Python Keras library with Tensorflow as a backend [15] and ran on two NVIDIA Titan Xp GPUs, where the batch size was set to 16.

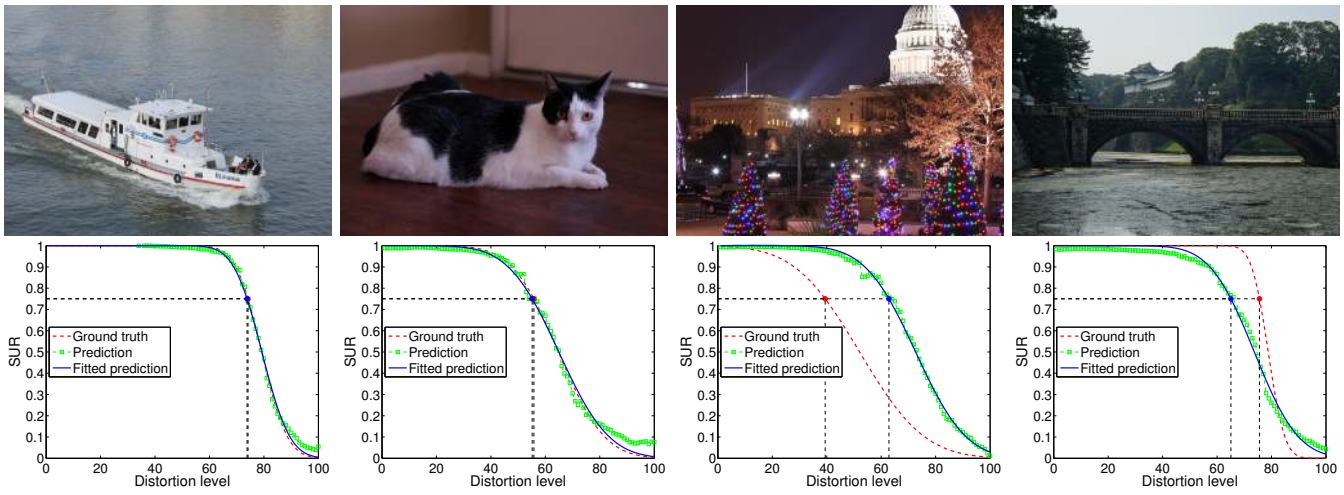
B. Strategies to address over-fitting

The MCL-JCI dataset is relatively small and training our model from scratch may be prone to over-fitting. Therefore,



(a) (b) (c)

Fig. 3: (a) Histogram of JND error. (b) Histogram of the Bhattacharyya distance. (c) PSNR comparison between the ground truth JND and predicted JND. The PLCC is 0.9755 and SROCC is 0.9619.



(a) Best prediction. (image 33) Bhattacharyya distance = 0.0017, $|\Delta\text{JND}| = 0.35$, $|\Delta\text{PSNR}| = 0.00$ dB.

(b) Second best prediction. (image 40) Bhattacharyya distance = 0.0018, $|\Delta\text{JND}| = 0.64$, $|\Delta\text{PSNR}| = 0.09$ dB.

(c) Worst prediction. (image 37) Bhattacharyya distance = 0.2039, $|\Delta\text{JND}| = 23.40$, $|\Delta\text{PSNR}| = 2.54$ dB.

(d) Second worst prediction. (image 23) Bhattacharyya distance = 0.2442, $|\Delta\text{JND}| = 10.48$, $|\Delta\text{PSNR}| = 1.16$ dB.

Fig. 4: Best two and worst two prediction results according to Bhattacharyya distance.

we applied transfer learning and data augmentation.

We downloaded 110,000 pristine images (100,000 for training and 10,000 for validation) from Pixabay [16]. Each image was compressed using a random quality factor of the JPEG encoder in Matlab R2018a. Then we used the full-reference IQA metric from [17] to compute the objective quality scores for all distorted images. SUR-Net was first trained to predict these objective quality scores with input given by a distorted image and its reference image (the pristine image). After five epochs, we fine-tuned the pretrained network on the MCL-JCI dataset to predict the SUR.

For data augmentation, each image of the MCL-JCI dataset was split into four non-overlapping patches with a resolution of 960×540. We also cropped one patch of the same resolution from the center of the image. The SURs for the patches were set to be equal to those of their source images. With this data augmentation, we had 25,250 annotated patches.

After training the network with this training set, SUR values can be predicted. To predict the SUR of an entire image, predictions for its five corresponding patches were generated by the network and averaged.

C. Results and analysis

Three metrics were used to evaluate the performance of SUR-Net: MAE of the 75% JND, MAE of the Peak Signal-to-Noise Ratio (PSNR) at the 75% JND, and Bhattacharyya distance [18] between the predicted and ground truth JND (Gaussian) distributions.

Table I presents the detailed results for each image sequence. Fig. 3 shows the statistics. There are 19 images with a very small 75% JND error (less than 5) and 36 images (72%) with a 75% JND error less than 10 (Fig. 3(a)). 35 images (70%) had a Bhattacharyya distance smaller than 0.1 (Fig. 3(b)). Fig. 3(c) compares the PSNR at ground truth JND and

TABLE I: Normal distribution model for the first JND levels in the 50 image sequences of the MCL-JCI dataset. Shown are the mean μ and standard deviations σ, for both ground truth and SUR-Net, together with the 75% JND values and PSNR at the 75% JND value. The Bhattacharyya distance is a measure for the divergence between the predicted and ground truth distributions, $\Delta\text{JND} = \widehat{\text{JND}} - \text{JND}$, and $\Delta\text{PSNR} = \widehat{\text{PSNR}} - \text{PSNR}$.

Image	Ground truth				SUR-Net				Bhattacharyya	$\lvert\Delta\text{JND}\rvert$	$\lvert\Delta\text{PSNR}\rvert$
k	μ	σ	JND	PSNR	$\hat{\mu}$	$\hat{\sigma}$	$\widehat{\text{JND}}$	$\widehat{\text{PSNR}}$	distance		
1	75.50	7.18	70.66	32.79	84.54	14.55	74.72	32.24	0.1930	4.06	0.55
2	65.40	14.47	55.64	41.50	49.30	29.01	29.73	43.84	0.1740	25.91	2.33
3	70.93	13.76	61.65	32.81	76.70	12.16	68.50	32.21	0.0285	6.85	0.60
4	75.43	9.12	69.28	29.77	71.86	13.39	62.82	30.33	0.0482	6.46	0.56
5	71.53	8.62	65.72	32.82	71.69	15.41	61.29	33.30	0.0800	4.43	0.48
6	71.87	12.16	63.67	34.41	79.05	11.08	71.58	33.59	0.0499	7.91	0.82
7	60.30	17.27	48.65	31.36	71.61	14.69	61.70	30.20	0.0687	13.05	1.16
8	73.73	6.43	69.40	29.12	70.33	12.71	61.76	29.81	0.1225	7.64	0.69
9	80.63	5.89	76.66	28.70	79.15	8.69	73.29	29.20	0.0419	3.37	0.50
10	75.50	8.83	69.54	37.65	67.90	16.40	56.83	38.77	0.1320	12.71	1.12
11	65.87	10.30	58.92	35.01	74.18	12.18	65.96	34.44	0.0749	7.04	0.56
12	48.90	13.73	39.64	34.95	66.33	22.71	51.01	34.05	0.1685	11.37	0.89
13	76.57	6.79	71.99	36.20	72.85	14.16	63.30	37.19	0.1386	8.69	0.99
14	75.90	9.50	69.50	33.91	74.24	14.92	64.18	34.35	0.0516	5.32	0.43
15	74.33	12.29	66.05	27.50	82.22	10.42	75.19	26.73	0.0667	9.15	0.77
16	75.60	11.68	67.73	31.58	77.82	10.16	70.97	31.28	0.0099	3.24	0.30
17	81.17	8.94	75.13	29.73	79.24	9.21	73.03	29.93	0.0058	2.10	0.20
18	79.17	7.08	74.39	34.22	76.80	11.21	69.24	34.79	0.0591	5.16	0.58
19	75.17	9.20	68.96	30.39	72.47	13.53	63.35	31.02	0.0430	5.61	0.62
20	60.10	10.91	52.74	33.06	72.53	10.38	65.53	31.89	0.1710	12.78	1.16
21	64.63	11.57	56.83	30.19	76.27	8.35	70.64	28.74	0.1924	13.81	1.46
22	73.43	13.62	64.25	29.94	77.95	11.04	70.50	29.35	0.0275	6.25	0.58
23	78.80	4.79	75.57	27.04	73.68	12.74	65.09	28.21	0.2442	10.48	1.16
24	76.83	7.43	71.82	33.36	74.84	15.35	64.49	34.10	0.1250	7.34	0.74
25	75.63	9.37	69.31	29.89	78.83	11.46	71.10	29.67	0.0217	1.78	0.22
26	63.60	10.90	56.25	34.25	66.58	16.09	55.73	34.25	0.0429	0.52	0.00
27	82.20	7.80	76.94	30.29	81.69	9.76	75.11	30.56	0.0129	1.83	0.27
28	70.83	10.81	63.54	41.21	74.65	12.61	66.14	41.01	0.0192	2.60	0.20
29	74.67	7.25	69.78	36.75	80.55	9.35	74.24	36.38	0.0778	4.46	0.37
30	78.00	9.03	71.91	36.66	76.03	12.52	67.59	37.22	0.0302	4.32	0.56
31	74.83	10.39	67.83	34.65	74.42	13.97	64.99	34.98	0.0218	2.84	0.34
32	74.30	9.16	68.12	32.17	82.84	8.05	77.41	30.98	0.1268	9.29	1.19
33	79.17	7.43	74.15	33.27	79.25	8.07	73.80	33.27	0.0017	0.35	0.00
34	69.60	9.01	63.52	32.21	74.53	11.25	66.94	31.93	0.0415	3.42	0.29
35	76.20	10.29	69.26	32.09	78.12	10.09	71.31	31.85	0.0045	2.05	0.23
36	77.80	7.35	72.84	30.46	78.49	8.56	72.72	30.46	0.0067	0.12	0.00
37	52.07	18.73	39.43	31.17	72.64	14.55	62.83	28.62	0.2039	23.40	2.54
38	78.20	9.44	71.83	30.49	80.31	10.47	73.25	30.39	0.0083	1.42	0.11
39	73.80	12.27	65.53	35.24	70.64	17.23	59.02	35.93	0.0339	6.50	0.69
40	64.80	13.35	55.80	39.15	64.95	14.52	55.16	39.24	0.0018	0.64	0.09
41	79.43	9.82	72.81	28.36	83.33	9.02	77.25	27.90	0.0232	4.44	0.46
42	78.67	8.69	72.80	31.10	72.41	14.13	62.87	32.11	0.0925	9.93	1.01
43	67.67	12.21	59.43	36.80	75.32	11.77	67.39	36.05	0.0513	7.96	0.75
44	82.83	5.98	78.80	29.88	81.40	9.30	75.13	30.40	0.0515	3.67	0.52
45	53.67	15.83	42.99	45.55	54.22	19.29	41.21	45.66	0.0099	1.78	0.11
46	84.07	8.03	78.65	32.46	74.71	13.10	65.87	34.23	0.1505	12.78	1.77
47	69.10	14.89	59.06	36.70	77.19	11.32	69.56	35.63	0.0654	10.50	1.07
48	78.97	9.92	72.28	35.20	77.34	12.39	68.99	35.58	0.0149	3.29	0.37
49	76.73	11.00	69.32	37.45	81.78	8.69	75.92	36.88	0.0462	6.61	0.57
50	81.87	7.58	76.75	34.61	75.50	11.68	67.62	35.95	0.0975	9.13	1.34
								Overall	**0.0715**	**6.73**	**0.6869**

predicted JND. The Pearson Linear Correlation Coefficient (PLCC) and Spearman Rank Order Correlation Coefficient (SROCC) were 0.9755 and 0.9619, respectively. Overall, the mean values of the Bhattacharyya distance, $|\Delta\text{JND}|$, and $|\Delta\text{PSNR}|$ were 0.0715, 6.73, and 0.687, respectively.

Fig. 4 shows the best and worst predictions, sorted by Bhattacharyya distance. The red dotted lines are ground truth SUR curves, the green dotted lines show predictions, and the blue curves are fitted SUR curves. The red and blue points show the ground truth and predicted 75% JND, respectively. The best prediction result was for image 33, with a 75% JND error of 0.35, a Bhattacharyya distance of 0.0017, and a PSNR difference at the 75% JND of less than 10^{-2} dB. The fitted SUR curve is very close to the ground truth.

The prediction results for a few images were not as good. This may be due to the fact that the size and diversity of the training set were too small for the deep learning algorithm. We expect that this problem can be overcome by training on a large-scale JND dataset.

Transfer learning significantly improved performance measures. Disregarding transfer learning, hence only using the pretrained weights from ImageNet [19], increased the mean of the Bhattacharyya distance from 0.0715 to 0.272, the $|\Delta\text{JND}|$ from 6.73 to 18.1, and $|\Delta\text{PSNR}|$ from 0.687 to 1.63.

D. Discussion and limitations

To the best of our knowledge, SUR-Net is the first deep learning-based approach to predict the SUR for compressed images. We predict the SUR of each distorted image rather than the SUR curve of the source image, which is given by N SUR points corresponding to N distorted versions. Thereby, the size of the training set was increased by a factor of N, which helps to train an effective deep learning model. Cropping patches from the source images is a typical way in data augmentation for deep learning. We cropped five patches from each source image and set their SUR values equal to that of the source images.

We conclude the discussion by mentioning limitations and open problems. It remains a topic for future work to study the effect of the patch size, number of patches, and the SUR value for each patch. In this contribution, we used the normality assumption about the first JND point on the QF scale. However, an empirical test (β_2 test [20]) showed that only 29 of the 50 source images passed the normality test. Thus, other models for the distribution of the JND may be more suitable than the Gaussian.

VII. Conclusion

We proposed a deep learning approach to predict the SUR curve for compressed images. First, pairs of images (a reference and a distorted one) were fed into a Siamese CNN. Second, features were extracted by feature pooling and concatenation. Finally, connected layers were added to learn a regression from image pairs to SUR values. The proposed approach can be easily generalized to predict the SUR curves for images compressed with other coders. As in [12], we provided results for the 75% JND. Results for other satisfied user ratios can be obtained in a similar way. Given a target percentage of satisfied users, the predicted SUR curve can be used to determine the JPEG quality factor QF that provides a compressed image, which is indistinguishable from the original for these users, thereby saving bit rate without the need for subjective visual quality tests.

Acknowledgment

This work was supported in part by the NSFC under Grant 61871372, Guangdong NSF for Distinguished Young Scholar under Grant 2016A030306022, Guangdong Provincial Science and Technology Development under Grant 2017B010110014, Shenzhen International Collaborative Research Project under Grant GJHZ20170314155404913, Shenzhen Science and Technology Program under Grant JCYJ20170811160212033, Guangdong International Science and Technology Cooperative Research Project under Grant 2018A050506063, Membership of Youth Innovation Promotion Association, CAS under Grant 2018392, and Shenzhen Discipline Construction Project for Urban Computing and Data Intelligence. This work was also funded by the Deutsche Forschungsgemeinschaft (DFG, German Research Foundation) Projektnummer 251654672 TRR 161 (Project A05).

References

[1] L. Jin, J. Y. Lin, S. Hu, H. Wang, P. Wang, I. Katsavounidis, A. Aaron, and C.-C. J. Kuo, "Statistical study on perceived JPEG image quality via MCL-JCI dataset construction and analysis," *Electronic Imaging*, vol. 2016, no. 13, pp. 1–9, 2016.

[2] H. Wang, W. Gan, S. Hu, J. Y. Lin, L. Jin, L. Song, P. Wang, I. Katsavounidis, A. Aaron, and C.-C. J. Kuo, "MCL-JCV: a JND-based H. 264/AVC video quality assessment dataset," in *IEEE International Conference on Image Processing (ICIP)*, 2016, pp. 1509–1513.

[3] H. Wang, I. Katsavounidis, J. Zhou, J. Park, S. Lei, X. Zhou, M.-O. Pun, X. Jin, R. Wang, X. Wang *et al.*, "VideoSet: A large-scale compressed video quality dataset based on JND measurement," *Journal of Visual Communication and Image Representation*, vol. 46, pp. 292–302, 2017.

[4] C. Fan, Y. Zhang, R. Hamzaoui, and Q. Jiang, "Interactive subjective study on picture-level just noticeable difference of compressed stereoscopic images," in *IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP)*, 2019.

[5] C. Szegedy, V. Vanhoucke, S. Ioffe, J. Shlens, and Z. Wojna, "Rethinking the inception architecture for computer vision," in *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2016, pp. 2818–2826.

[6] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2016, pp. 770–778.

[7] J. Redmon, S. Divvala, R. Girshick, and A. Farhadi, "You only look once: Unified, real-time object detection," in *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2016, pp. 779–788.

[8] T.-Y. Lin, P. Goyal, R. Girshick, K. He, and P. Dollar, "Focal loss for dense object detection," in *IEEE International Conference on Computer Vision (ICCV)*, 2017, pp. 2999–3007.

[9] S. Bosse, D. Maniry, K.-R. Müller, T. Wiegand, and W. Samek, "Deep neural networks for no-reference and full-reference image quality assessment," *IEEE Transactions on Image Processing*, vol. 27, no. 1, pp. 206–219, 2018.

[10] O. Wiedemann, V. Hosu, H. Lin, and D. Saupe, "Disregarding the big picture: Towards local image quality assessment," in *International Conference on Quality of Multimedia Experience (QoMEX)*, 2018.

[11] Q. Huang, H. Wang, S. C. Lim, H. Y. Kim, S. Y. Jeong, and C.-C. J. Kuo, "Measure and prediction of HEVC perceptually lossy/lossless boundary QP values," in *Data Compression Conference (DCC)*, 2017, pp. 42–51.

[12] H. Wang, I. Katsavounidis, Q. Huang, X. Zhou, and C.-C. J. Kuo, "Prediction of satisfied user ratio for compressed video," in *IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, 2018, pp. 6747–6751.

[13] H. Wang, X. Zhang, C. Yang, and C.-C. J. Kuo, "Analysis and prediction of JND-based video quality model," in *Picture Coding Symposium (PCS)*, 2018, pp. 278–282.

[14] D. P. Kingma and J. Ba, "Adam: A method for stochastic optimization," *arXiv preprint arXiv:1412.6980*, 2014.

[15] F. Chollet *et al.*, "Keras," https://keras.io, 2015.

[16] https://pixabay.com/.

[17] H. Z. Nafchi, A. Shahkolaei, R. Hedjam, and M. Cheriet, "Mean deviation similarity index: Efficient and reliable full-reference image quality evaluator," *IEEE Access*, vol. 4, pp. 5579–5590, 2016.

[18] A. Bhattacharyya, "On a measure of divergence between two statistical populations defined by their probability distributions," *Bulletin of the Calcutta Mathematical Society*, vol. 35, pp. 99–109, 1943.

[19] J. Deng, W. Dong, R. Socher, L.-J. Li, K. Li, and L. Fei-Fei, "ImageNet: A large-scale hierarchical image database," in *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2009, pp. 248–255.

[20] Recommendation ITU-R BT.500-11, "Methodology for the subjective assessment of the quality of television pictures," 2002.