Perceptual Quality Enhancement with Multi-scale Deep Learning for Video Transmission: A QoE Perspective

Chaoyi Han[1,2], Yiping Duan[1,2], Xiaoming Tao[1,2(✉)], Rundong Gao[1], and Jianhua Lu[1,2]

[1] Department of Electronic Engineering, Tsinghua University, Beijing 100084, China
{hancy16,gaord}@mails.tsinghua.edu.cn
{yipingduan,taoxm,lhh-dee}@tsinghua.edu.cn
[2] Beijing National Research Center for Information Science and Technology (BNRist), Beijing 100084, China

Abstract. With the development of mobile Internet technologies, wireless communication is facing huge challenges under the explosive growth of multimedia data, e.g. video conferences, online education. This makes it difficult to guarantee the communication quality where communication resources (bandwidth, channel, etc.) are limited. In this paper, we propose an image enhancement method to transform blurred images into images with high perceptual quality. The proposed method serves as a post-processing part for communication systems and is incorporated into the receiver. Specifically, we learn the prior of high quality images using a collected dataset. We train a neural network to accomplish this task and adopt a multi-scale perceptual loss as the objective, which is more consistent with the quality of experience (QoE). To validate the proposed method, we train our model on a large dataset with both blurred images and high quality images. Experimental results show that, using a pre-collected dataset with high quality images, the proposed approach can effectively restore the blurred images.

Keywords: Quality of experience (QoE) · Perceptual image enhancement · Wireless communications

1 Introduction

With the rapid development of mobile Internet technologies, multimedia data such as images and videos are becoming the mainstream. Cisco Visual Network reported that Global IP video traffic will grow four-fold from 2017 to 2022, with a compound annual growth rate (CAGR) of 29%. In particular, live Internet video has the potential to drive large amounts of traffic as it is replacing traditional broadcast services. Nowadays, live video already accounts for 5% of Internet video traffic and will reach 17% by 2022 [1]. Correspondingly, the global

S. Shi et al. (Eds.): AICON 2020, LNICST 356, pp. 26–40, 2021.
https://doi.org/10.1007/978-3-030-69066-3_3

average broadband speed will double from 2017 to 2022, from 39.0 Mbps to 75.4 Mbps, which exhibits a notable mismatch. When transmitting images/videos with limited bandwidth, image compression must be applied to significantly save the encoding bit rate [3,4]. However, commonly used image compression methods usually have artifacts such as blocking and ringing, which may servery degrade the quality of user experience (QoE). Moreover, these artifacts may reduce the accuracy of subsequent classification, recognition and other high-level tasks. Therefore, it is necessary to study the quality enhancement methods to make the blurred image high-definition [2]. The quality enhancement module follows the decoder to improve the degraded image quality that caused by limited bandwidth, channel and other transmission issues. The proposed approach can not only help to improve the quality of experience (QoE) significantly but can also be used to relieve the pressure on communication bandwidth.

Recently, there has been increasing interest in enhancing the quality of the compressed images/videos [5,6]. The quality enhancement of image aims to restore the original undistorted image as much as possible from the degraded image, and at the same time improving the perceptual quality [7,15]. According to different research methods, it can be roughly classified into two types, namely model-based degraded image quality enhancement and data-based degraded image quality enhancement. The whole quality enhancement process of the model-based method includes three parts: modeling the degradation process, estimating the degradation degree and inferring the reconstructed signal. Since restoring the original data from degraded data is an ill-conditioned problem, it is generally necessary to add certain prior constraints to solve it. In actual application, it is necessary to know the degradation process in advance, and then estimate the degradation degree so as to select the corresponding model for restoration and solution. Data-based enhancement of degraded images is expected to use real low-quality images and high-quality image data through parameterized methods [9,10]. According to whether the training data is paired, it is divided into: 1. The same image pair in the training data with low-quality and high-quality image pairs, so the whole process can be trained by supervised learning; 2. Degraded images and high-quality images that not corresponding to each other, then optimization can be performed by minimizing the distance between the reconstructed data and the distribution of high-quality images, and the consistency of the image content before and after the restoration enhancement can be ensured by the cyclic mapping method.

In this paper, we proposed a perceptual quality enhancement method with multi-scale nerual network for video transmission toward QoE. Specifically, we train an encoder-decoder model to exploit the relationship between the blurred images and the high quality image for each scale. Moreover, we present multiscale perceptual loss function that mimics conventional coarse-to-fine approaches. Experiments on the benchmark dataset show that using the loss function of the feature domain for training, the neural network has improved the subjective perceptual quality of the restored image, and even achieved better results in

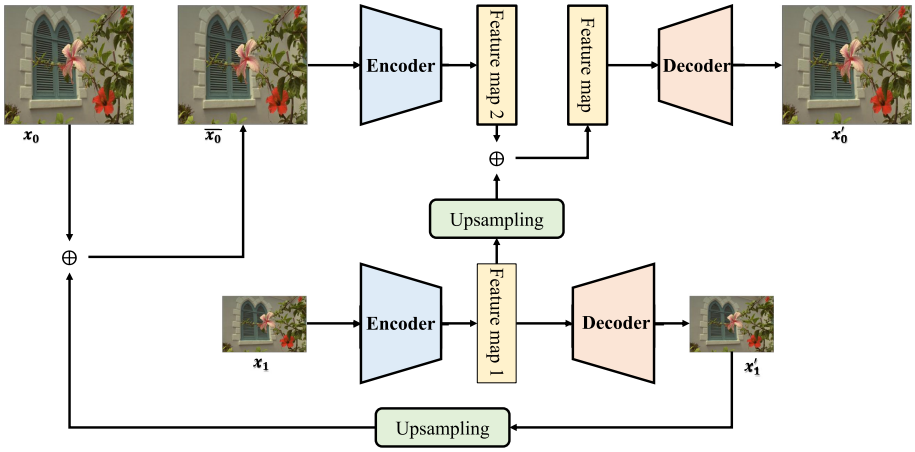

Fig. 1. The framework of the proposed approach (scale $s = 2$). x_0 is the original scale and x_1 is the downsampling scale with the stride $ds = 2$. $\bigoplus$ represents the average weighted sum.

automated evaluation metrics. The corresponding framework is shown in Fig. 1. The contributions can be summarized as follows,

- We established a QoE-oriented image quality enhancement framework. A novel optimization objectives combing mean square error and feature-level error to preserve the fidelity from pixel-level and semantic-level.
- We develop a multi-scale deep learning method to learn the relationships between the blurred image and the high quality image. The multi-scale and multi-feature learning are used to improve the performance of the proposed approach.

The rest of this paper is organized as follows. Section 2 reviews the related work about image enhancement and Sect. 3 presents the proposed perceptual quality enhancement model including the network architecture and multi-scale end-to-end optimization. Section 4 presents the experimental results on a benchmark dataset and Sect. 5 concludes this paper.

2 Related Work

Recently, extensive works have focused on enhancing the quality of images after compression [11–14]. In general, the input is a blurred image and the output is a high-definition image. Through a deep neural network, a complex nonlinear relationship between the blurred image and the high-definition image is established to improve the image quality.

The method based on deep learning needs to solve three extremely important basic problems when applied to a specific field: high-quality data, a network structure that can effectively extract features, and a loss function that can effectively evaluate the results. In terms of data, early researchers mainly used some

simple methods such as downsampling to generate fuzzy images on the existing clear images to build a data set [15]. However, this simple way of generating blurred images makes the training data not in line with the real-world data distribution, which greatly restricts the effectiveness of deep learning. Therefore, some researchers have built a data set closer to the real world. This data set is called the GoPro dataset [12], which meets the data needs of training neural networks. In terms of network structure, because image deblurring is a pixel-dense task, the network is required to generate output at each pixel, and this is similar to some classic computer vision tasks in terms of output, such as image semantic segmentation tasks [16]. Therefore, researchers have migrated the classic network U-shaped network in the field of image semantic segmentation to the field of image deblurring and the U-shaped network has become almost the only network basic framework in the field of image deblurring [17–19]. In terms of loss function, researchers generally use a pixel-by-pixel two-norm loss function. However, in recent years, research work has shown that the pixel-by-pixel two-norm loss function cannot effectively describe the subjective perceptual quality of the image. This phenomenon is called the "perceptual gap", that is: higher PSNR (Peak Signal to Noise Ratio) is not necessarily more in line with human subjective perception [20–22].

Kim et al. [23] tried to solve non-uniform blind image deblurring problem. In this paper, in contrastive the restrictive assumption that the underlying scene is static and the blur is caused by only camera shake, authors address the deblurring problem of general dynamic scenes which contain multiple moving objects as well as camera shake. [11] to use deep learning technique to solve image deblurring problem. In this paper, authors address the problem of estimating and removing non-uniform blur from a single blurry image. They propose a deep learning approach to predicting the probabilistic distribution of motion blur at the patch level using a convolution neural network (CNN). In [12], researchers present a deep learning framework that mimics conventional coarse-to-fine approaches, which restores sharp images in an end-to-end manner through a multi-scale convolution neural network. To tackle the above problems, in [24], researchers present a deep hierarchical multi-patch network inspired by Spatial Pyramid Matching to deal with blurry images via a fine-to-coarse hierarchical representation. To deal with the performance saturation w.r.t depth, they propose a stacked version of their multi-patch model.

3 Perceptual Quality Enhancement Model

3.1 Network Architecture

The framework of the proposed approach with multi-scale structure is show in Fig. 1. For each scale, we adopt the encoder-decoder structure in Fig. 2. This structure accepts a blurred image as input, and outputs a high-definition image with same size as the input image. The encoder is responsible for extracting the original image features for processing, and the decoder is responsible for restoring a high-definition image based on the extracted features.

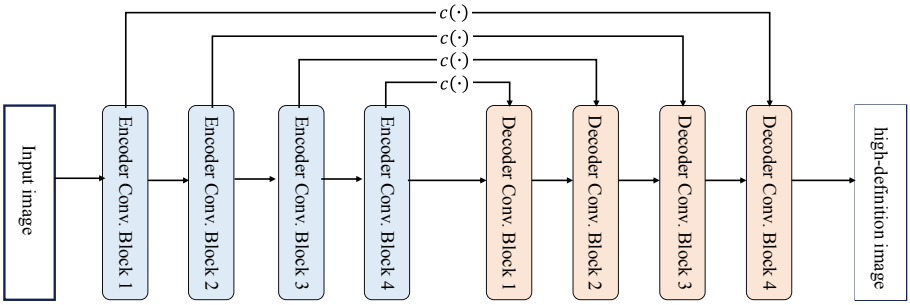

Fig. 2. The network architecture of each scale. The network includes 4 encoder convolutional block and 4 decoder convolutional block. Encoder convolution block 1 is the inverse process of decoder deconvolution block 1, and so on.

As shown in Fig. 1, the encoder-decoder structure consists of two basic convolutional blocks stacked: encoding convolutional block and decoding convolutional block. The internal structure of the two convolutional blocks is shown in Fig. 2. The encoding convolutional block is first composed of a convolution layer (the convolution kernel size is 3×3) followed by two residual convolution blocks [25]. The decoding convolutional block is first convolved by two residuals followed by a deconvolution layer [26] (the deconvolution kernel size is 4×4). The residual convolution block contains a two-layer convolution layer with ReLu activation function [27]. With such basic components, we can build an encoder-decoder structure: the encoding convolutional blocks are stacked to become an encoder, and the decoding convolutional blocks are stacked to become a decoder, and the two modules are symmetrical. The structure first reduces the size of the feature map through the multi-layer convolutional neural network on the encoder side, and increases the number of feature channels at the same time. The encoder composed of a multi-layer convolutional neural network extracts the image semantic features necessary for the deblurring task from the original input image, and then such features are input to the decoder, and the multi-layer convolutional network on the decoder side is gradually upsampled and increased The feature map size is reduced while the number of feature channels is reduced, and the processed image semantic features are decoded to generate a clear image with the same size as the input after deblurring operation. $\oplus$ splices the feature maps of the encoder and decoder as the input of the corresponding decoding convolution block. By this way, the decoder can make full use of different Hierarchical information including the low-level and high-level features (Fig. 3).

The specific structure parameters of the network we adopted are as follows. The encoder is composed of encoding convolutional block. The first encoding convolutional block converts the image from a 3-channel (RGB) original image to a 32-channel feature map with the same size. Subsequently, each of the three encoding convolutional blocks doubles the number of input feature channels, and at the same time reduces the size of the feature map by one time. Because of

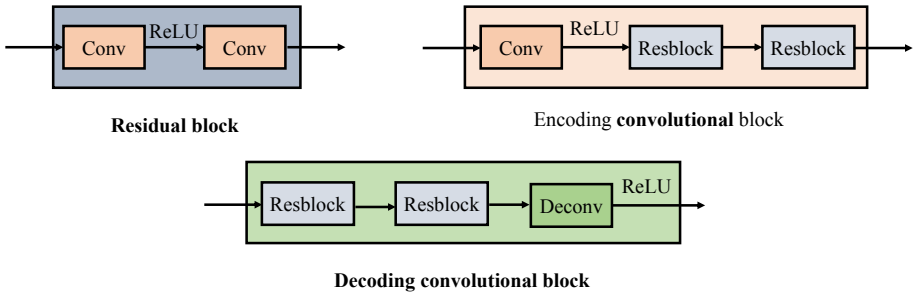

Fig. 3. The left is the residual block. The middle is the encoder convolution. The right is the decoder decovolution.

the symmetry of the encoder-decoder structure, the decoder is also composed of four decoded convolutional blocks. Each of the first three decoded convolutional blocks reduces the number of feature channels of the input feature map by a factor of two, while the feature map size doubled. The last decoder convolution block converts the input feature map into a 3-channel restored image as the final output.

3.2 Perceptual Quality Optimization

The current convolutional neural network based image quality enhancement employ PSNR or SSIM as the optimization target. However, the PSNR of the image does not consider the quality of experience (QoE) of users. Perceptual quality is an objective measure of QoE. Therefore, we define the optimization target from pixel-level and semantic-level as,

$$L_{loss} = L_{pixel} + L_{feature} \tag{1}$$

where m and n represents the height and width of the image, respectively. S is the real clear image and O is the clear image learned by the network. In (1), the optimization target includes two parts of the loss function in the pixel domain and the feature domain. The loss function of the pixel domain is defined in (2). It directly calculates the Euclidean distance of each pixel between the real clear image and the clear image learned by the network.

$$L_{pixel} = \frac{1}{m \times n} \sum_{i=1}^{m} \sum_{j=1}^{n} (S(i,j) - O(i,j))^2 \tag{2}$$

where n and m respectively represent the length and width of the image, S represents the real clear image, and O represents the clear image generated by the network.

Some researchers began to search for new loss functions to guide the neural network to produce images that fit the human eye's perceptual quality. In this

paper, we use neural networks trained on large-scale data sets to extract features from RGB three-channel images, or to convert images from pixel domain to feature domain, similar to the human visual perception system Refine the image in the same way. In this way, training in the feature domain will guide the neural network to output images that are more in line with the perceived quality of the human eye. The loss function of the feature domain is defined in (3). It calculates the Euclidean distance of each element between the feature representation of the real clear image and the feature representation of the clear image generated by the network.

$$L_{feature} = \frac{1}{m \times n} \sum_{i=1}^{m} \sum_{j=1}^{n} (f(S(i,j)) - f(O(i,j)))^2 \tag{3}$$

We choose the pre-trained VGG16 [6] on ImageNet dataset as the feature extraction function. The network is a neural network structure developed by the Google DeepMind research team and the Oxford University Computer Vision Laboratory. The neural network is formed by stacking a 3×3 convolutional layer and a 2×2 maximum pooling layer. In order to use multiple feature maps, the feature loss can be written as,

$$L_{feature} = \frac{1}{m \times n} \sum_{i=1}^{m} \sum_{j=1}^{n} \sum_{k=1}^{C} \alpha_k (f(S(i,j)) - f(O(i,j)))^2 \tag{4}$$

where α_k is the weights and C is the number of the feature maps.

3.3 Multi-scale Deep Learning Model

Intuitively, images always contain different features at different scales. The image will show more texture details at large resolutions and the overall structure of the image will be more compact at a small resolution. Therefore, under large and small resolutions, different levels of information can be captured effectively. In this way, multi-scale algorithms can fully extract the features of different levels of the image and increase the accuracy of image feature description.

The multi-scale structure based on the encoder-decoder is shown in Fig. 1. The encoder-decoder network is a fully convolutional network, and the convolutional layer has no dimensional assumptions on the input of the image. Figure 1 shows the two-scale network architectures. We can see that the two encoder-decoders are completely the same in structure. The overall network structure is from bottom to top, and the resolution of the processed image is from small to large. In addition, in order to accelerate convergence and strengthen the interaction between different scales, residual connections are also introduced in the intermediate feature maps of the encoder and decoder of the two scales.

$$L_{feature} = \frac{1}{m \times n} \sum_{i=1}^{m} \sum_{j=1}^{n} \sum_{k=1}^{C} \alpha_k (f(S(i,j)) - f(O(i,j)))^2 \tag{5}$$

4 Experiments

Recently, a researcher proposed that the blurred frames of long exposure time can be approximated by aggregating several consecutive short exposure time frames in the video recorded by high-speed cameras, and released a public data set called GoPRO data through this method [12]. The researchers used the professional sports camera GoPRo Hero 4 Black to record. When the camera continuously receives light during the exposure process, a blurred image is generated. The fuzzy generation process can be modeled as,

$$B = g(\frac{1}{T}\int_{t}^{T} s(t)dt) \approx g(\frac{1}{M}\sum_{i=1}^{M} s[i]) \tag{6}$$

where $s(t)$ represents the clear image corresponding to time t, and T represents the exposure time. Correspondingly, M represents the number of sampled video frames, and $s[i]$ represents the $i-th$ clear image signal sampled during the exposure time. g represents the camera response function, which converts the signal received by the camera into an image signal that we can observe.

The researchers used the GoPro data set to shoot $240FPS$ videos, and then gathered 7 to 13 consecutive frames to obtain different degrees of blurred images, using the middle frame in the blurred frame segment as the target clear image. The training data of the GoPRo dataset can simulate complex camera shake and object movement scenarios, which fits the real application scene very well, and the amount of training data is larger than the previous dataset, which can greatly satisfy the neural network's training data. Therefore, this dataset has also become the most important evaluation benchmark dataset for image deblurring methods based on deep learning. The data set contains 3214 clear-fuzzy image pairs. We use 2103 data for training, and the remaining 1111 data are used as the validation set.

4.1 Settings

In the training phase, the adaptive momentum estimation optimization algorithm (Adam) [8] is used to optimize the neural network, and the Adam algorithm is widely used in deep learning training. The model is trained for a total of 3000 rounds, the initial learning rate is set to 0.0001, and after each 1000 rounds of training, it is reduced to 0.3 times the current learning rate. According to the experimental results, 3000 rounds of training can make the model fully converge. In each iteration, we sample two blurred images and randomly crop the image area of 256×256 size as a batch (and in the test, the input is the original size, that is, 720×1280), because the original input is 720×1280 resolution is very large. If the original resolution of 720×1280 is used as input in the training phase, the video memory requirement cannot be met, and the experimental results show that a 256×256 cropped image block contains enough information to make the neural network Learn the mapping relationship from blurred images to clear images. The input image is divided by 255 and normalized to the

range $[0, 1]$, then 0.5 is subtracted, and normalized to the range $[-0.5, -0.5]$. All trainable parameters of the model are initialized using Xavier [9] initialization method. All experiments are performed on an NVIDIA Titan X.

4.2 Performance of the Proposed Approach

Objective Evaluation. This section compares several previous algorithms on the benchmark data set. Because the training data of the GoPro data set contains general camera shake blur and object movement blur, the model trained on such a data set has the ability to deal with non-uniform blur, so it is compared with many traditional algorithms under the assumption of uniform blur. it's meaningless. The algorithm proposed by Whyte [12] can be used as a representative of the classic non-uniform deblurring algorithm. In addition, the comparison algorithm also includes the algorithm proposed by Nah [3], the algorithm proposed by Tao [10] and the algorithm proposed by Sun [2]. Nah introduced multi-scale structure into the field of image deblurring for the first time. On the basis of Nah, a recurrent neural network was added, and Sun used a convolutional neural network to estimate the local fuzzy kernel, and then applied random fields and traditional deconvolution algorithms to restore the entire clear image.

Table 1. Performance of different approaches.

Approaches	Sun [11]	Nah [12]	Tao [31]	Ours
PSNR	24.64	29.08	30.10	30.43
SSIM	0.8429	0.9135	0.9323	0.9031
Time	20 min	3.09 s	1.6 s	0.26 s

The automated evaluation indicators are shown in Table 1. It can be seen from Table 1 that this model is not inferior to the comparison algorithm in the automated evaluation index (the PSNR is even better than the comparison algorithm). In addition, one advantage of this model compared to the comparison algorithm is that it takes less time to process a picture (0.26 s vs 1.6 s), which makes this model more practical in some scenarios that are extremely sensitive to time loss (such as video Stream processing, etc.).

Subjective Evaluation. The visual results of the proposed approach is shown in Fig. 4 with three images randomly sampled from the test dataset. The first row is the original blur image. The second row is the quality enhancement result by Whyte et al. It can be observed that Whyte algorithm failed to restore the photos with good sharpness, and the visual effect is very poor, such as the first picture. The black streaks seriously affect the sensory quality. Sun's algorithm is also unable to restore effective clear images. It can be seen that the third row has almost no effect on removing blur compared to the first row, because Sun's

model is trained on artificially synthesized data sets, and Compared with real fuzzy data, artificially synthesized data is too simple to represent the real-world data distribution well. Nah's algorithm can output a certain quality of clear images. Nah's algorithm can output a certain quality of clear images, but there is still a considerable degree of blur compared to the proposed model, because the number of layers of the proposed model is deeper. Tao's model can produce clearer restored images. The results of the proposed model are similar to those of Tao. But the overall look and feel of the results of the proposed model is sharper, such as the text in the blue box area in the second row. In the output result of the proposed model, the text in this area is clearer than the result of Tao. Moreover, the proposed approach takes less time to process an image, which makes this model more suitable for actual production scenarios, because in actual production and life, the application scenarios for image deblurring are often some Scenarios that require rapid response, such as monitoring equipment or video stream processing. In these scenarios, the image processing speed and image quality are equally important. Compared with Tao's model, the delay of this model is reduced by more than 1 s.

4.3 Ablation Study

Performance of Different Scales. In order to analyze the effect of multi-scale structure, models of different scale levels are trained. The evaluation results of these models are shown in Table 2. Considering the trade-off between the limitations of video memory and training time and the performance improvement brought by increasing the scale level, the network is only stacked to three scales at most.

Table 2. Performance of different scales.

Scale-level	1	2	3
PSNR	30.37	30.45	30.43
SSIM	0.9018	0.9031	0.9030

It can be seen from Table 2 that the introduction of multi-scale has improved the performance of the neural network. Among the three scale parameter settings, the neural network with a scale parameter of 2 has the best performance. Compared with a basic encoder-decoder network, That is, for a network with a scale level of 1, the PSNR increased by 0.08 dB, and the SSIM increased by 0.0013. However, the network performance of scale level 3 has not been further improved, but has slightly decreased. This phenomenon is more difficult to explain. In the specific experiment, at the beginning, we guessed that the stacking of scales caused the network to deepen and the trainability problem, so we tried to use the multi-scale loss function proposed in the Nah algorithm, that

Fig. 4. Visualization of image quality enhancement by different methods. The first row is the original blurred image. The second row is the result of Whyte's method. The third row is the result of Sun's method. The fourth row is the result of Nah's method. The fifth row is the result of Tao's method. The sixth row is the result of our proposed method.

Fig. 5. Visualization of image quality enhancement with different losses. The first column is the original blurred image. The second column is the result with MSE loss. The third column is the result with MSE loss and perceptual loss.

is, the output result of each scale is subjected to the loss function The calculation (ground truth is directly obtained by downsampling the clear image of the original resolution), but the result of the training strategy training in this way has been described in the previous section. The training loss of the model drops very quickly, but in the validation set The PSNR and SSIM indicators are far lower than those of models trained without multi-scale loss function. Therefore, we believe that the introduction of a multi-scale loss function may cause the loss function of the model to be dominated by the low-resolution image deblurring results, that is to say, it is overfitted to the low-resolution output results. This is obviously not the result we expected, so the multi-scale loss function is not used.

Table 3. Performance of different losses.

Loss function	MSE	MSE+VGG9	MSE+VGG23	MSE+VGG_9_23
PSNR	30.16	30.30	30.21	30.45
SSIM	0.8991	0.9009	0.9007	0.9031

Performance of Different Losses. In order to analyze the impact of introducing feature loss, a comparative experiment with and without feature loss is carried out. Table 3 shows the impact of the introduction of the feature loss function on the performance of the neural network during the training phase. In the experiment, the feature maps of the $9th$ and $23rd$ layers of the VGG network are used for perceptual loss, and the weighted coefficients are 0.002 and 0.0015, respectively. It can be seen from the results in Table 3 that the introduction of feature loss, whether it is a separate $9th$ layer, a separate $23rd$ layer, and the combination of $9th$ and $23rd$ layers, has brought a great improvement in performance. It can be seen that the feature loss of adding two layers at the same time is the most obvious for the model performance improvement. Figure 5 shows three examples of visual results. It can be seen that the networks generated by the two training strategies (introducing feature loss vs without introducing feature loss) can remove blur better, and the deblurring generated by the neural network that introduces the feature loss function The image will become clearer and more realistic at the edges (such as the edge of the clothes of the person on the left) and texture (such as the front window glass of the car in the picture).

5 Conclusions

For the purpose of image deblurring, this paper adds a simple loss function to the original pixel-wise two-norm loss function and performs multi-scale expansion on the network structure to improve the subjective perception quality and objective quality of the restored image. We established a QoE-oriented image quality enhancement framework and adopted a novel optimization objective that combines mean square error and feature-level error to preserve the fidelity from pixel-level and semantic-level. We developed a multi-scale deep learning method to learn the relationships between the blurred image and the high quality image. The multi-scale and multi-feature learning are used to improve the performance of the proposed approach. Experiments on the benchmark data set for image deblurring show that when using feature domain loss function for training, the multi-scale image processing network has improved the subjective perceptual quality of restored images, and achieved better results in objective quality.

Acknowledgement. This work was supported by the National Key R&D Program of China (2018YFB1 800804) and the National Natural Science Foundation of China (NSFC 61925105, 61801260).

References

1. LNCS Homepage. https://www.cisco.com/c/en/us/solutions/collateral/executive-perspectives/annual-internet-report/white-paper-c11-741490.html. Accessed 4 July 2020
2. Fergus, R., Singh, B., Hertzmann, A., et al.: Removing camera shake from a single photograph. ACM Trans. Graph. **25**(3), 787–794 (2006)
3. Kim, Y., Cho, S., Lee, J., Jeong, S., Choi, J., Do, J.: Towards the perceptual quality enhancement of low bit-rate compressed images. In: 2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops (CVPRW), Seattle, WA, USA, pp. 565–569. IEEE (2020)
4. Lee, J., et al.: FBRNN: feedback recurrent neural network for extreme image super-resolution. In: 2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops (CVPRW), Seattle, WA, USA, pp. 565–569. IEEE (2020)
5. Ding, D., Wang, W., Tong, J., Gao, X., Liu, Z., Fang, Y.: Biprediction-based video quality enhancement via learning. IEEE Trans. Cybern. (2020, in press)
6. Guan, Z., Xing, Q., Xu, M., Yang, R., Li, T., Wang, Z.: MFQE 2.0: a new approach for multi-frame quality enhancement on compressed video. IEEE Trans. Pattern Anal. Mach. Intell. (2020, in press)
7. Singh, G., Mittal, A.: Various image enhancement techniques a critiacal review. Int. J. Innov. Sci. Res. **10**(2), 267–274 (2014)
8. Rani, S., Jindal, S., Kaur, B.: A brief review on image restoration techniques. Int. J. Comput. Appl. **150**(12), 30–33 (2016)
9. Chen, Y., Wang, Y., Kao, M., Chuang, Y.: Deep photo enhancer: unpaired learning for image enhancement from photographs with GANs. In: 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), Salt Lake City, UT, USA, pp. 6306–6314. IEEE (2018)
10. Park, J., Lee, J., Yoo, D., Kweon, I.: Distort-and-recover: color enhancement using deep reinforcement learning. In: 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), Seattle, WA, USA, pp. 5928–5936 569. IEEE (2018)
11. Sun, J., Cao, W., Xu, Z., Ponce, J.: Learning a convolutional neural network for non-uniform motion blur removal. In: 2015 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Boston, MA, USA, pp. 769–777. IEEE (2015)
12. Nah, S., Kim, T., Lee, K.: Deep multi-scale convolutional neural network for dynamic scene deblurring. In: 2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Honolulu, HI, USA, pp. 257–265. IEEE (2017)
13. Johnson, J., Alahi, A., Fei-Fei, L.: Perceptual losses for real-time style transfer and super-resolution. In: Leibe, B., Matas, J., Sebe, N., Welling, M. (eds.) ECCV 2016. LNCS, vol. 9906, pp. 694–711. Springer, Cham (2016). https://doi.org/10.1007/978-3-319-46475-6_43
14. Ledig, C., et al.: Photo-realistic single image super-resolution using a generative adversarial network. In: 2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Honolulu, HI, USA, pp. 257–265. IEEE (2017)
15. Dong, W., Zhang, L., Shi, G., Wu, X.: Image deblurring and super-resolution by adaptive sparse domain selection and adaptive regularization. IEEE Trans. Image Process. **20**(7), 1838–1857 (2011)
16. Han, C., Duan, Y., Tao, X., Lu, J.: Dense convolutional networks for semantic segmentation. IEEE Access **7**, 43369–43382 (2019)

17. Whyte, O., Sivic, J., Zisserman, A., Ponce, J.: Non-uniform deblurring for shaken images. In: 2010 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), San Francisco, CA, USA, pp. 491–498. IEEE (2010)
18. Jia, C., Zhang, X., Zhang, J., Wang, S., Ma, S.: Deep convolutional network based image quality enhancement for low bit rate image compression. In: 2016 Visual Communications and Image Processing (VCIP), Chengdu, China, pp. 1–4. IEEE (2016)
19. Yu, J., Gao, X., Tao, D., Li, X., Zhang, K.: A unified learning framework for single image super-resolution. IEEE Trans. Neural Netw. Learn. Syst. **25**(4), 780–792 (2014)
20. Hameed, A., Dai, R., Balas, B.: A decision-tree-based perceptual video quality prediction model and its application in FEC for wireless multimedia communications. IEEE Trans. Multimedia **18**(4), 764–774 (2016)
21. Ahar, A., Barri, A., Schelken, P.: From sparse coding significance to perceptual quality: a new approach for image quality assessment. IEEE Trans. Image Process. **27**(2), 879–893 (2018)
22. Tao, X., Duan, Y., Xu, M., Meng, Z., Lu, J.: Learning QoE of mobile video transmission with deep neural network: a data-driven approach. IEEE J. Sel. Areas Commun. **37**(6), 1337–1348 (2019)
23. Kim, T., Ahn, B., Lee, K.: Dynamic scene deblurring. In: 2013 IEEE International Conference on Computer Vision, Sydney, NSW, Australia, pp. 3160–3167. IEEE (2013)
24. Zhang, H., Dai, Y., Li, H., Koniusz, P.: Deep stacked hierarchical multi-patch network for image deblurring. In: 2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), Long Beach, CA, USA, USA, pp. 5971–5979. IEEE (2019)
25. He, K., Zhang, X., Ren, S., Sun, J.: Deep residual learning for image recognition. In: 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Las Vegas, NV, USA, pp. 770–778. IEEE (2016)
26. Zeiler, M., Taylor, G., Fergus, R.: Adaptive deconvolutional networks for mid and high level feature learning. In: 2011 International Conference on Computer Vision, Barcelona, Spain, pp. 2018–2025. IEEE (2011)
27. Glorot, X., Bordes, A., Bengio, Y.: Deep sparse rectifier neural networks. J. Mach. Learn. Res. (11), 315–323 (2011)
28. Sergey, L., Christian, S.: Batch Normalization: Accelerating Deep Network Training by Reducing Internal Covariate Shift (2015). http://arxiv.org/abs/1502.03167
29. Ba, J., Kiros, J., Hinton, G.: Layer Normalization (2016). https://arxiv.org/abs/1607.06450
30. Wu, Y., He, K.: Group normalization. Int. J. Comput. Vision **128**(3), 742–755 (2020)
31. Tao, X., Gao, H., Shen, X., Wang, J., Jia, J.: Scale-recurrent network for deep image deblurring. In: 2018 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Salt Lake City, UT, USA, pp. 8174–8182. IEEE (2018)