Fast and Efficient Image Quality Enhancement via Desubpixel Convolutional Neural Networks

Thang Vu[✉], Cao V. Nguyen, Trung X. Pham, Tung M. Luu, and Chang D. Yoo

Department of Electrical Engineering, Korea Advanced Institute of Science and Technology (KAIST), Daejeon, South Korea
{thangvubk,nguyenvancao,trungpx,tungluu2203,cd_yoo}@kaist.ac.kr

Abstract. This paper considers a convolutional neural network for image quality enhancement referred to as the fast and efficient quality enhancement (FEQE) that can be trained for either image super-resolution or image enhancement to provide accurate yet visually pleasing images on mobile devices by addressing the following three main issues. First, the considered FEQE performs majority of its computation in a low-resolution space. Second, the number of channels used in the convolutional layers is small which allows FEQE to be very deep. Third, the FEQE performs downsampling referred to as desubpixel that does not lead to loss of information. Experimental results on a number of standard benchmark datasets show significant improvements in image fidelity and reduction in processing time of the proposed FEQE compared to the recent state-of-the-art methods. In the PIRM 2018 challenge, the proposed FEQE placed first on the image super-resolution task for mobile devices. The code is available at https://github.com/thangvubk/FEQE.git.

Keywords: Image super-resolution · Image enhancement Mobile devices

1 Introduction

Image transformation is a classical problem which includes image super-resolution and image enhancement, where an input image is transformed into an output image with the desired resolution, color, or style [1–3]. For example, given a low-quality image, a transformation may be introduced to produce a enhanced-quality image that is as similar as possible to the desired high-quality image in terms of resolution and/or color rendition.

Recent example-based methods based on deep convolutional neural networks (CNN) have made great strides in image quality enhancement. However, most of the methods are focused on improving only the qualitative measure such as peak signal-to-noise ratio and mean-opinion score without any consideration to
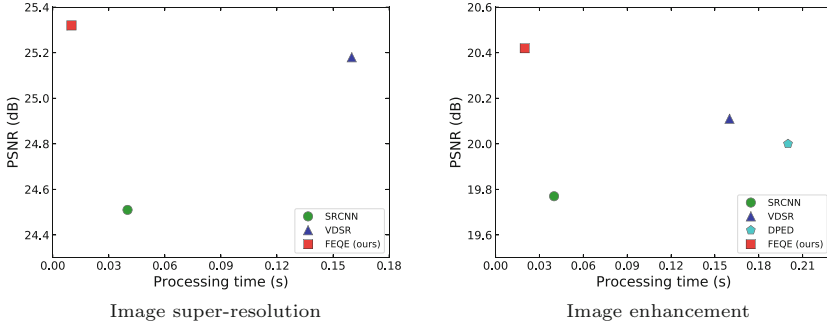
Fig. 1. Comparison in PSNR and processing time of proposed FEQE with recent state-of-the-art methods on image super-resolution and enhancement.

execution time. As a results, their computational requirements are enormous even for high-end desktops, not to mention mobile devices.

To address this problem, this paper considers a CNN referred to as Fast and Efficient image Quality Enhancement (FEQE) for both image super-resolution and enhancement for mobile devices. Preliminary results are illustrated in Fig. 1. To achieve the shown performance, the FEQE is designed to produce highest possible image quality under a certain memory and computational constraint. To reduce the computational complexity, an input image is downsampled and then upsampled at the very first and very last layers respectively, keeping the convolution operation mostly in the low-resolution space. However, downsampling generally leads to loss in information such that the operation is irreversible. To address this problem, FEQE provides an effective way to perform downsampling without losing information such that the operation becomes reversible, which is referred to as desubpixel.

The proposed desubpixel systematically rearranges spatial features into channels, keeping the feature values intact, hence providing sufficient information for inferences in the following convolutional layers. To improve prediction accuracy with restricted resources, the FEQE is designed to be deep as possible but with small channel-depth. As investigated in [4], with the same number of parameters, a deeper network provides considerably higher capacity compared to that of a shallow network. Experimental results show that the proposed FEQE achieves significant improvements in both accuracy and runtime compared to recent state-of-the-art methods.

The rest of this paper is organized as follows. Section 2 reviews various image super-resolution and enhancement methods. Section 3 presents and explains the effectiveness of the proposed method. Section 4 reports experiment results on standard benchmark datasets. Finally, Sect. 5 summarizes and concludes the paper.

2   Related Work

2.1   Image Super-Resolution

Image super-resolution has received substantial attention for its applications, ranging from surveillance imaging [5,6], medical imaging [7,8] and object recognition [9,10]. Conventional methods are based on interpolation such as bilinear, bicubic, and Lancroz [11], which are simple but usually produce overly-smoothed reconstructions. To mitigate this problem, example-based methods using hand-crafted features have been proposed, ranging from dictionary learning [12–15], neighborhood learning [16–18], to regression tree [19,20].

Recent advances in deep learning have made great strides in super-resolution [1,21–23]. Dong *et al.* [1,24] first introduced SRCNN for learning the low- to high-quality mapping in an end-to-end manner. Although SRCNN is only a three-convolutional-layer network, it outperforms previous hand-crafted-feature-based methods. In [25], Shi *et al.* propose subpixel modules, providing efficient upsampling method for reconstructing high-quality images. It turns out super-resolution also benefits from very deep networks as in many other applications. The 5-layer FSRCNN [26], 20-layer VDSR [21], and 52-layer DRRN [27] show significant improvements in terms of accuracy. Lim *et al.* [23] propose a very deep modified ResNet [28] to achieve state-of-the-art PSNR performance. Although their improvements in terms of accuracy are undeniable, the computational requirements leave a lot to be desired especially for use in mobile devices.

2.2   Image Enhancement

Image enhancement aims to improve image quality in terms of colors, brightness, and contrasts. Earlier methods are mainly based on histogram equalization and gamma correction. Although these methods are simple and fast, their performance are limited by the fact that individual pixels are enhanced without consideration to contextual information. More advanced methods are based on the retinex theory [29], and these methods estimate and normalize illumination to obtain the enhanced image [30–32]. In [30], Zhang *et al.* utilize mutual conversion between RBG and HSV color space in obtaining the desired illuminations with a guided filter before obtaining the target enhanced image. Meanwhile, Fu *et al.* [31] consider a novel retinex-based image enhancement using illumination adjustment.

Recently, various CNN-based methods have been demonstrated to be conducive for image enhancement [2,33,34]. Showing that multi-scale retinex is equivalent to CNN with different Gaussion kernels, Shen *et al.* [33] propose MSR-net to learn an end-to-end mapping between a low-light and a bright image. Ignatov *et al.* [2] propose DPED to produce DLSR- from mobile-quality images by using deep residual networks trained with a composite loss function of content, color, and texture. Despite showing improvements in enhancing image quality, these method exposed limitations in processing time since the computational operations are performed in the high-resolution space.

2.3   Convolutional Network for Mobile Devices

There has been considerable interest in building small and fast networks to perform various computer vision tasks for mobile devices [35–39]. To accelerate the networks, recent methods often simplify the convolution operation. One such method simultaneously performs spatial convolution in each channel and linear projection across channels [35]. In [36], Iandola *et al.* introduce Squeezenet which achieves Alexnet-level accuracy with 50 times fewer parameters by leveraging small filter sizes. Meanwhile, Howard *et al.* [39] propose Mobilenet built from depthwise separable convolutions, showing promising results on various vision problems on mobile devices.
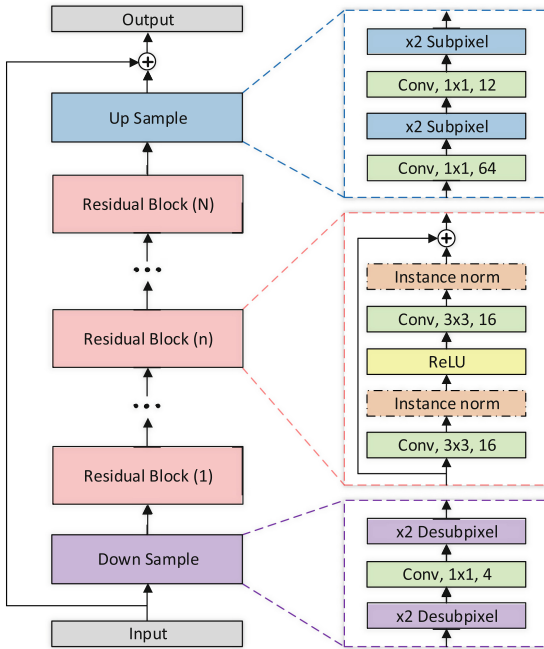

Fig. 2. Architecture of the proposed FEQE

Beside speed, a small network is essential for mobile devices, which is generally obtained by shrinking, compressing, and quantizing the network. Molchanov *et al.* [38] propose a pruning mechanism for resource efficient inferences by discarding the least important neurons. In [37], Kim *et al.* perform network decomposition which provides light networks for mobile applications. Another approach is distillation [40], where a large network is trained to teach a small network.

3   Proposed Method

3.1   Network Architecture

Overview. Figure 2 presents the architecture of the proposed FEQE. Here, a low-quality image is first downsampled with a factor of 4 using two proposed $\times 2$ desubpixel modules. An 1×1 convolutional layer is used to adjust the number of channels into the desired value. After downsampling into a low-resolution space, the features are fed into a series of N residual blocks, each of which consists of two 3×3 convolutional layers followed by instance normalization and ReLU activations. It is noted that the instance normalization layers are used for image enhancement task only to normalize the contrast variation among samples. The output of the N-th residual block is upsampled using two $\times 2$ subpixel models before summing with the low-quality input image to produce a predicted high-quality output image.

The proposed FEQE is a fast and efficient method for the following three reasons. First, the considered FEQE performs the majority of its computation in the low-resolution space. As illustrated in Fig. 3, the computational complexity of FEQE is much lower than that of resolution-unchanged or progressive encoder-decoder networks. Second, the number of channels used in the residual blocks is small which allows FEQE to be very deep. The reason is a convolutional layer requires $C^2 K^2$ parameters to map from a C-channel input to a C-channel output

Fixed-resolution network

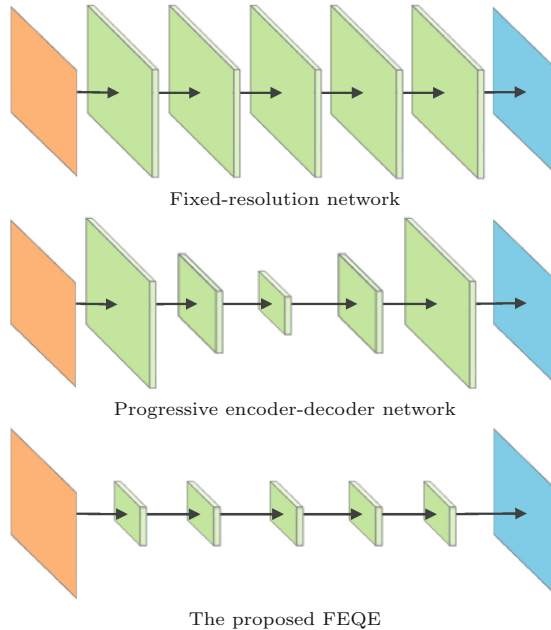Progressive encoder-decoder network

The proposed FEQE

Fig. 3. Feature dimensions of the proposed FEQE in comparison with other networks.

using a kernel size of $K \times K$. Therefore, using the same number of parameters, reducing the number of channels n times can allow the network to be deeper by a factor of n^2. Third, the FEQE performs downsampling referred to as desubpixel that does not lead to loss of information. The details of the desubpixel modules are presented in the following subsections.

Desubpixel Downsample. Downsampling generally leads to loss of information which is more severe when performed early in the network. Inspired by the subpixel upsampling in [25], a reversible downsampling module referred to as desubpixel performs downsampling such that its input can be recovered as shown in Fig. 4. The proposed desubpixel module systematically rearranges the spatial features into channels to reduce spatial dimensions without losing information. Let $\mathcal{U}$ and $\mathcal{D}$ respectively denote subpixel-upsampling and desubpixel-downsampling function. A concatenation of downsampling and upsampling operation leads to the identity transform such that:

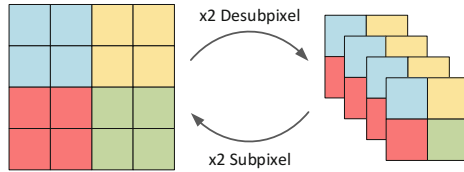$$\mathcal{U}(\mathcal{D}(X)) = X. \tag{1}$$

Fig. 4. Subpixel and the proposed desubpixel

Figure 5 illustrates the effectiveness of the proposed desubpixel over other common downsampling methods that includes convolution with stride 2, max-pooling, and bilinear interpolation. It is assumed the contribution that a neuron makes in a network is proportional to the number of links to the next layer. For the 3×3 convolution with stride 2, the number of times a neuron in a particular layer is filtered varies from 1, 2 and 4. Here, a neuron indicated by the darkest shade of blue is filtered 4 times while a neuron indicated by the lightest shade of blue is filtered only once. Downsampling requires the stride to be at least 2, and this leads to non-uniform contribution of neurons. In other words, certain neurons of high importance may not be given a chance to contribute adequately. In the 2×2 max-pooling, only one out of four neurons in a 2×2 block is connected to the next layer. Although the pooling filter selects the most prominent neuron, the accuracy degradation in the reconstruction is inevitable as a result of pruning. In the bilinear interpolation, every 2×2 neurons are represented by their weighted sum, which can be thought of as a "relaxed" max-pooling. As in other two, the bilinear interpolation is irreversible. The proposed $\times 2$ desubpixel permits for all neurons an equal opportunity to contribute. The desubpixel allows
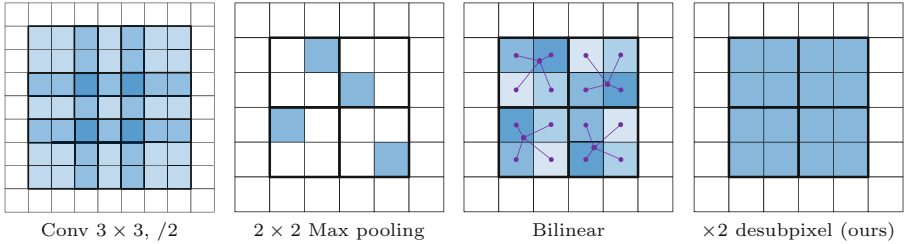
Fig. 5. Impact of each neuron in current layer to next layer in considered down-sampling methods. In each subfigure, darker colors indicate higher impacts. Different from the others, the proposed desubpixel provides an uniform neural impact. (Color figure online)

arbitrary integer downsampling provided that the spatial dimension of the input is desirable for channel rearrangement. Incorporating subpixel with desubpixel, the FEQE can be applied for various image-to-image mapping tasks such as image generation, segmentation, style transfer, and image quality enhancement. In this paper, the proposed method is applied for image super-resolution and enhancement.

3.2   Loss Functions

Given a low-quality image I^{LQ}, a network attempts to predict an enhanced image that is as similar as possible to the desired high-quality image I^{HQ}. Mathematically, the network G parameterized by θ is trained to minimize the loss between $G_\theta(I^{LQ})$ and I^{HQ} as follows:

$$\theta = \arg\min_\theta \mathcal{L}\left(G_\theta(I^{LQ}), I^{HQ}\right). \tag{2}$$

The final loss function composes of mean-squared error (MSE) loss $\mathcal{L}_M$ and a VGG loss $\mathcal{L}_V$ with trade-off parameters of α_M and α_V, respectively:

$$\mathcal{L} = \alpha_M \mathcal{L}_M + \alpha_V \mathcal{L}_V. \tag{3}$$

The MSE loss is the most common objective function in enhancing the fidelity of the reconstructed images:

$$\mathcal{L}_M = \sum_i \left\| I_i^{HQ} - G(I_i^{LQ}) \right\|_2^2. \tag{4}$$

Meanwhile, the VGG loss aims to produce images with high perceptual quality:

$$\mathcal{L}_V = \sum_i \left\| \phi(I_i^{HQ}) - \phi(G(I_i^{LQ})) \right\|_2^2. \tag{5}$$

where ϕ denotes the feature maps obtained from the forth convolutional layer before the fifth max-pooling layer of a pre-trained 19-layer VGG network [41].

4   Experiments

4.1   Single Image Super-Resolution

Datasets. The proposed FEQE is trained using DIV2K [42] dataset which composes of 800 high-quality images (2K resolution). For testing, four standards benchmark datasets are used, including: Set5 [16], Set14 [15], B100 [43], Urban100 [44].

Table 1. Comparison in PSNR and SSIM of the proposed method with different settings of channels C and residual blocks N. Bold indicates better results.

Setting	PSNR	SSIM
$C16\ N20$	**28.75**	**0.9652**
$C32\ N5$	28.70	0.9650

Evaluation Metrics. The image super-resolution performance is measured on Y (luminance) channel in YCbCr space. Following existing super-resolution studies, the conventional peak-signal-noise-ratio (PSNR) and structural similarity (SSIM) index are used in the experiments. However, these metrics do not always objectively reflect image quality. Additionally, a recently-proposed perceptual metric is considered [45]:

$$\text{Perceptual index} = \frac{(10 - \text{NRQM}) + \text{NIQE}}{2}, \tag{6}$$

where NRQM and NIQE are the quality metrics proposed by Ma *et al.* [46] and Mittal *et al.* [47], respectively. The lower perceptual indexes (PI) indicate better perceptual quality.

Training Details. The images are normalized to a mean of 0 and a standard deviation of 127.5. At training time, to enhance computational efficiency, the images are further cropped into patches of size 196×196. It is noted that the proposed network is fully convolutional; thus, it can take arbitrary size input at test time.

The final network for image super-resolution task has 20 residual blocks. We train the network using Adam optimizer [48] with setting $\beta_1 = 0.9$, $\beta_2 = 0.999$, and $\epsilon = 10^{-8}$. Batch size is set to 8. We pre-train the network with MSE loss and a downsampling factor of 2, before training with the full loss function and a downsampling factor of 4. The numbers of iterations for pre-training and training phases are set to 5×10^5. In the default FEQE setting, the trade-off parameters are set to $\alpha_M = 1$ and $\alpha_V = 10^{-4}$. Besides, in the setting referred to as FEQE-P such that PSNR is maximized, the trade-off parameters are changed to $\alpha_M = 1$ and $\alpha_V = 0$.

Our models are implemented using Tensorflow [49] deep learning framework. The experiments are run on a single Titan X GPU, and it takes about 5 hours for the networks to converge.

Ablation Study. The effectiveness of the FEQE is demonstrated using an ablation analysis. In the first part, we measure the performance of FEQE with different downsampling methods, including convolution with stride 2, 2×2 max pooling, bicubic interpolation, and the proposed desubpixel. For the fair comparison in computational complexity, a kernel size of 1×1 is used in the convolution method. The remaining parts of the network and training hyper-parameters are kept unchanged. Figure 6 illustrates the comparison results in terms of PSNR and SSIM. The convolution method performs the worst, followed by the bicubic and pooling methods. The proposed desubpixel achieves the best performance in both PSNR and SSIM.

In the second part of the ablation study, the performance with different network settings is presented. Table 1 shows that with the same number of
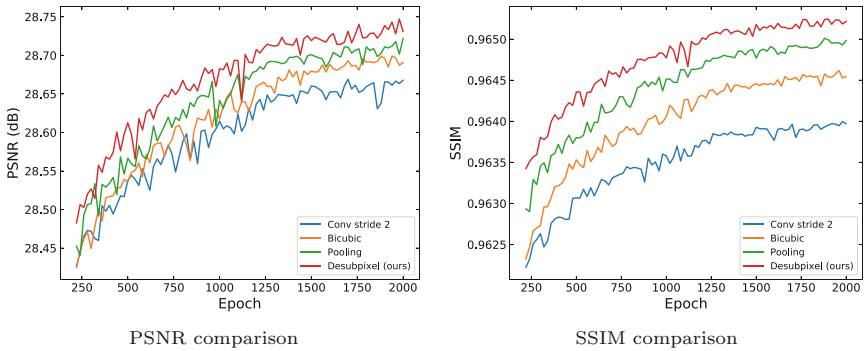

PSNR comparison   SSIM comparison

Fig. 6. Comparison in PSNR and SSIM of the proposed FEQE with different downsampling methods.

Table 2. Comparison in PSNR/SSIM/PI of the proposed FEQE with other image super-resolution methods. Bold indicates the best results.

Method	Set5	Set14	B100	Urban100
Bicubic	28.42/0.8096/7.32	26.08/0.7047/6.97	25.96/0.6691/6.94	23.14/0.6587/6.88
SRCNN	30.47/0.8610/6.79	27.57/0.7528/6.03	26.89/0.7108/6.04	24.51/0.7232/5.94
VDSR	**31.53**/**0.8840**/6.45	**28.42**/**0.7830**/5.77	27.29/0.7262/5.70	25.18/0.7534/5.54
FEQE-P (ours)	**31.53**/0.8824/6.03	28.21/0.7714/5.77	**27.32**/**0.7273**/5.79	**25.32**/**0.7583**/5.57
FEQE (ours)	31.32/0.8754/**5.94**	28.09/0.7660/**5.40**	27.23/0.7229/**5.64**	25.26/0.7547/**5.50**

Table 3. Comparison in computational complexity of the proposed FEQE with other image super-resolution methods. Bold indicates the best results.

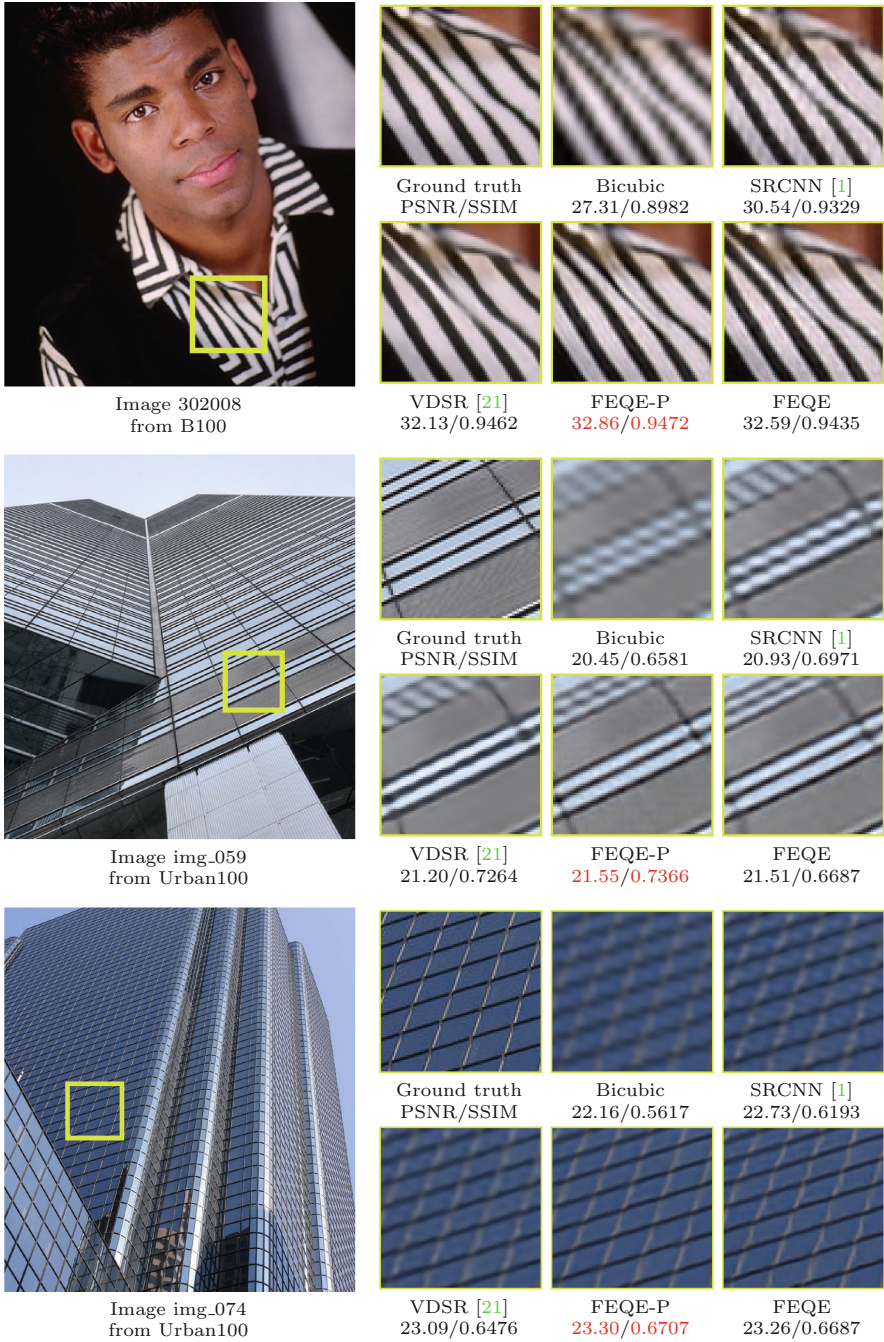| Method	# parameters	# FLOPs	Time (s)
SRCNN	$\mathbf{69 \times 10^3}$	128×10^9	0.04
VDSR	668×10^3	1231×10^9	0.16
FEQE (ours)	96×10^3	$\mathbf{11 \times 10^9}$	**0.01**

Fig. 7. Qualitative comparison of FEQE with other image super resolution methods. RED indicates the best results. (Color figure online)

parameters, the deep network with a small number of convolution channels performs much better than the shallow one.

Comparison with State-of-the-art SISR Methods. The proposed FEQE is compared with recent state-of-the-art SISR methods including SRCNN [1] and VDSR [21], which are conducive for mobile devices. The PSNR, SSIM, and PI results of referred methods are obtained from source codes provided by the authors. For fair processing-time comparison, we implement and measure all the network architectures using the same hardware and Tensorflow deep learning framework. Table 2 summaries the results of the considered methods over 4 benchmark datasets. The proposed FEQE-P archives better overall PSNR and SSIM meanwhile FEQE outperforms the others in terms of perceptual quality. Here, the computational complexities and running time of the considered methods are reported in Table 3. The processing time is averaged over 100 HD-resolution ($1280 \times 720p$) images. The proposed FEQE is the fastest since most of the computation is performed in the low-resolution feature space. In particular, FEQE is 16 times faster than VDSR while achieving better quantitative performance. The proposed FEQE is also visually compared to the others. As shown in Fig. 7, the proposed methods provides more plausible visualization with sharper edges and textures.

4.2   Image Enhancement

Training Details. We demonstrate that the proposed FEQE is also conducive for image enhancement. In this task, the DPED dataset [2] is used for training and testing. The low- and high-quality images are taken from a iphone 3GS and Canon 70D DSLR, respectively. The provided training images are in patches of 100×100. The considered quality metrics are PSNR and SSIM on RGB channels. The instance normalization layers are injected into the residual blocks, and the number of the residual blocks is changed to 14. The other training procedures are similar to those of the super-resolution task.

Effectiveness of Instance Normalization. In image enhancement, the contrasts usually vary among low- to high-quality mappings, which should be addressed using Instance Normalization layers. Figure 8 show that without Instance Normalization, the predicted image exhibits unpleasing visualization for the unwanted color spillages. When the contrasts are normalized the enhanced image is much more plausible looking and no color spillages are observed.

Comparison with State-of-the-art Methods. The proposed FEQE is compared with recent state-of-the-art methods including SRCNN [1], VDSR [21], and DPED [2]. Although SRCNN and VDSR are originally SISR methods, they are related to end-to-end image-to-image mapping, which is relevant for image enhancement. We re-implemented SRCNN and VDSR and train the network
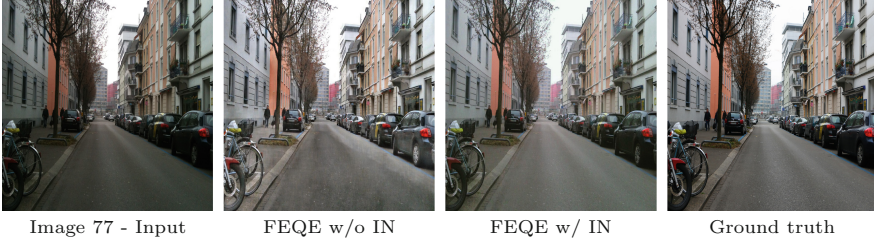
| Image 77 - Input | FEQE w/o IN | FEQE w/ IN | Ground truth |

Fig. 8. Visual comparison of FEQE with and without instance normalization.

Table 4. Quantitative comparison of the proposed FEQE with other image enhancement methods. Bold indicates the best results.

Method	PSNR	SSIM	Time
SRCNN	19.77	0.8823	0.04
VDSR	20.11	0.8837	0.16
DPED	20.00	**0.9192**	0.20
FEQE (ours)	**20.42**	0.9181	**0.02**

| Image 101 - Input | SRCNN [1] | VDSR [21] |
| DPED [2] | FEQE (ours) | Ground truth |

Fig. 9. Qualitative comparison of FEQE with other image enhancement methods

with our loss function. The experimental results of DPED method are reproduced from publicly available source codes provided by the authors. Table 4 shows that the proposed FEQE not only achieves better performance in terms of PSNR but also is the fastest method. The qualitative results are illustrated in Fig. 9. Here, SRCNN and VDSR expose limitations in enhancing image quality for unpleasing color spillages. Our proposed FEQE are competitive with DPED and exhibits significant improvement in terms of brightness and vivid colors compare to the other methods.

4.3   PIRM 2018 Challenge

The Perceptual Image Restoration and Manipulation(PIRM) 2018 challenge aims to produce images that are visually appealing to human observers. The authors participated in the Perceptual Image Enhancement on Smartphones challenge which requires light, fast, and efficient solutions. The challenge composes of two conventional computer vision tasks: image super-resolution and image enhancement. The evaluation metric is based on PSNR, multi-scale SSIM, and processing time of the solution (in subscript s) and the baseline (in subscript b) as follows:

$$\text{Score} = \alpha(\text{PSNR}_s - \text{PSNR}_b) + \beta(\text{SSIM}_s - \text{SSIM}_b) + \gamma \min\left(4, \frac{\text{Time}_b}{\text{Time}_s}\right). \quad (7)$$

Here, α, β and γ are the trade-off parameters. There are three evaluation scores corresponding to three combinations of trade-off parameters. Score A is giving preference to the solution with the highest fidelity (PSNR), score B is aimed at the solution providing the best visual results (SSIM), and score C is targeted at the best balance between the speed and quantitative performance. The details are provided in [50]. Table 5 summaries the challenge results for the image super-resolution tasks. The proposed method wins the first place for achieving the best overall score.

Table 5. Comparison of the proposed FEQE with other top-ranking methods in the PIRM 2018 super-resolution challenge on mobile devices. Bold indicates the best scores.

Method	PSNR (dB)	SSIM	CPU (ms)	GPU (ms)	Razer Phone (ms)	Score A	Score B	Score C
FEQE (ours)	28.21	0.9636	701	48	936	**13.21**	**15.15**	**14.14**
Method 1	28.14	0.963	343	34	812	12.86	14.83	13.87
Method 2	28.19	0.9633	773	112	1101	13.08	15.02	14.04
Method 3	28.13	0.9636	767	70	1198	12.88	15.05	13.97
Method 4	28.13	0.9632	654	56	1414	12.84	14.92	13.91

In the image enhancement task, we used super-resolution-based loss function and instance normalization without applying heavily-optimized techniques for

image enhancement. Although our method exposed some limitations in qualitative results compared to the competitors, our quantitative results is in top-ranking teams and FEQE is the fastest network measured on smartphones.

4.4   Limitation

Since the proposed FEQE is designed under the resource constrain of mobile devices, its limitation on challenging samples is inevitable. The limitation is visually presented in Fig. 10. In image super-resolution, FEQE introduces antifacts for difficulties of distinguishing cross or vertical line patterns in bicubic input. In image enhancement, since the input is in poor light condition, FEQE fails to enhance vivid colors.
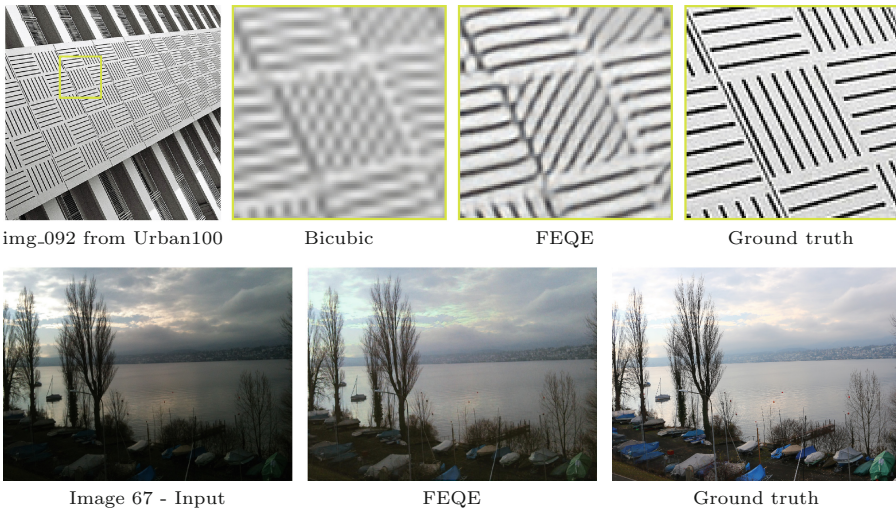

| img_092 from Urban100 | Bicubic | FEQE | Ground truth |

| Image 67 - Input | FEQE | Ground truth |

Fig. 10. Limitation of FEQE in challenging examples of image super-resolution (first row) and image enhancement (second row)

5   Conclusion

A Fast and Efficient image Quality Enhancement referred to as FEQE for image super-resolution and enhancement on mobile devices is introduced. To accelerate the inference time, the proposed FEQE performs most of the computational operations in a low-resolution space. The low-resolution features are obtained by the proposed desubpixel which provides an effective way to downsample the high-resolution images. In desubpixel, the spatial features are systematically rearranged into channels, keeping the feature values intact, hence providing sufficient information for the following convolutional layers. To improve the fidelity of the reconstruction, convolutional architecture is designed to be deep

with small channel-depth. Experimental results on standard benchmark datasets show significant achievements in terms of image quality and running time of the proposed FEQE over recent state-of-the-art image super-resolution and enhancement methods.

References

1. Dong, C., Loy, C.C., He, K., Tang, X.: Image super-resolution using deep convolutional networks. IEEE Trans. Pattern Anal. Mach. Intell. **38**(2), 295–307 (2016)
2. Ignatov, A., Kobyshev, N., Timofte, R., Vanhoey, K., Van Gool, L.: DSLR-quality photos on mobile devices with deep convolutional networks. In: ICCV (2017)
3. Gatys, L.A., Ecker, A.S., Bethge, M.: Image style transfer using convolutional neural networks. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 2414–2423 (2016)
4. Pascanu, R., Montufar, G., Bengio, Y.: On the number of response regions of deep feed forward networks with piece-wise linear activations. arXiv preprint arXiv:1312.6098 (2013)
5. Zou, W.W., Yuen, P.C.: Very low resolution face recognition problem. IEEE Trans. Image Process. **21**(1), 327–340 (2012)
6. Jiang, J., Ma, J., Chen, C., Jiang, X., Wang, Z.: Noise robust face image super-resolution through smooth sparse representation. IEEE Trans. Cybern. **47**(11), 3991–4002 (2017)
7. Shi, W., et al.: Cardiac image super-resolution with global correspondence using multi-atlas patchmatch. In: Mori, K., Sakuma, I., Sato, Y., Barillot, C., Navab, N. (eds.) MICCAI 2013. LNCS, vol. 8151, pp. 9–16. Springer, Heidelberg (2013). https://doi.org/10.1007/978-3-642-40760-4_2
8. Ning, L., et al.: A joint compressed-sensing and super-resolution approach for very high-resolution diffusion imaging. NeuroImage **125**, 386–400 (2016)
9. Sajjadi, M.S., Schölkopf, B., Hirsch, M.: Enhancenet: single image super-resolution through automated texture synthesis. In: 2017 IEEE International Conference on Computer Vision (ICCV), pp. 4501–4510. IEEE (2017)
10. Zhang, Y., Li, K., Li, K., Wang, L., Zhong, B., Fu, Y.: Image super-resolution using very deep residual channel attention networks. arXiv preprint arXiv:1807.02758 (2018)
11. Duchon, C.E.: Lanczos filtering in one and two dimensions. J. Appl. Meteorol. **18**(8), 1016–1022 (1979)
12. Wang, S., Zhang, L., Liang, Y., Pan, Q.: Semi-coupled dictionary learning with applications to image super-resolution and photo-sketch synthesis. In: 2012 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), pp. 2216–2223. IEEE (2012)
13. Yang, J., Wang, Z., Lin, Z., Cohen, S., Huang, T.: Coupled dictionary training for image super-resolution. IEEE Trans. Image Process. **21**(8), 3467–3478 (2012)
14. Yang, J., Wright, J., Huang, T.S., Ma, Y.: Image super-resolution via sparse representation. IEEE Trans. Image Process. **19**(11), 2861–2873 (2010)
15. Zeyde, R., Elad, M., Protter, M.: On single image scale-up using sparse-representations. In: Boissonnat, J.-D., et al. (eds.) Curves and Surfaces 2010. LNCS, vol. 6920, pp. 711–730. Springer, Heidelberg (2012). https://doi.org/10.1007/978-3-642-27413-8_47

16. Bevilacqua, M., Roumy, A., Guillemot, C., Alberi-Morel, M.L.: Low-complexity single-image super-resolution based on nonnegative neighbor embedding (2012)
17. Timofte, R., De Smet, V., Van Gool, L.: Anchored neighborhood regression for fast example-based super-resolution. In: Proceedings of the IEEE International Conference on Computer Vision, pp. 1920–1927 (2013)
18. Timofte, R., De Smet, V., Van Gool, L.: A+: adjusted anchored neighborhood regression for fast super-resolution. In: Cremers, D., Reid, I., Saito, H., Yang, M.-H. (eds.) ACCV 2014. LNCS, vol. 9006, pp. 111–126. Springer, Cham (2015). https://doi.org/10.1007/978-3-319-16817-3_8
19. Salvador, J., Perez-Pellitero, E.: Naive bayes super-resolution forest. In: Proceedings of the IEEE International Conference on Computer Vision, pp. 325–333 (2015)
20. Schulter, S., Leistner, C., Bischof, H.: Fast and accurate image upscaling with super-resolution forests. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 3791–3799 (2015)
21. Kim, J., Lee, J.K., Lee, K.M.: Accurate image super-resolution using very deep convolutional networks. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 1646–1654 (2016)
22. Kim, J., Lee, J.K., Lee, K.M.: Deeply-recursive convolutional network for image super-resolution. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 1637–1645 (2016)
23. Lim, B., Son, S., Kim, H., Nah, S., Lee, K.M.: Enhanced deep residual networks for single image super-resolution. In: The IEEE Conference on Computer Vision and Pattern Recognition (CVPR) Workshops, vol. 1, p. 4 (2017)
24. Dong, C., Loy, C.C., He, K., Tang, X.: Learning a deep convolutional network for image super-resolution. In: Fleet, D., Pajdla, T., Schiele, B., Tuytelaars, T. (eds.) ECCV 2014. LNCS, vol. 8692, pp. 184–199. Springer, Cham (2014). https://doi.org/10.1007/978-3-319-10593-2_13
25. Shi, W., et al.: Real-time single image and video super-resolution using an efficient sub-pixel convolutional neural network. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 1874–1883 (2016)
26. Dong, C., Loy, C.C., Tang, X.: Accelerating the super-resolution convolutional neural network. In: Leibe, B., Matas, J., Sebe, N., Welling, M. (eds.) ECCV 2016. LNCS, vol. 9906, pp. 391–407. Springer, Cham (2016). https://doi.org/10.1007/978-3-319-46475-6_25
27. Tai, Y., Yang, J., Liu, X.: Image super-resolution via deep recursive residual network. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, vol. 1, p. 5 (2017)
28. He, K., Zhang, X., Ren, S., Sun, J.: Deep residual learning for image recognition. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 770–778 (2016)
29. Land, E.H., McCann, J.J.: Lightness and retinex theory. Josa **61**(1), 1–11 (1971)
30. Zhang, S., Tang, G.J., Liu, X.H., Luo, S.H., Wang, D.D.: Retinex based low-light image enhancement using guided filtering and variational framework. Optoelectron. Lett. **14**(2), 156–160 (2018)
31. Fu, X., Sun, Y., LiWang, M., Huang, Y., Zhang, X.P., Ding, X.: A novel retinex based approach for image enhancement with illumination adjustment. In: 2014 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), pp. 1190–1194. IEEE (2014)
32. Li, D., Zhang, Y., Wen, P., Bai, L.: A retinex algorithm for image enhancement based on recursive bilateral filtering. In: 2015 11th International Conference on Computational Intelligence and Security (CIS), pp. 154–157. IEEE (2015)

33. Shen, L., Yue, Z., Feng, F., Chen, Q., Liu, S., Ma, J.: MSR-net: Low-light image enhancement using deep convolutional network. arXiv preprint arXiv:1711.02488 (2017)
34. Tao, F., Yang, X., Wu, W., Liu, K., Zhou, Z., Liu, Y.: Retinex-based image enhancement framework by using region covariance filter. Soft Comput. **22**(5), 1399–1420 (2018)
35. Wang, M., Liu, B., Foroosh, H.: Factorized convolutional neural networks. In: ICCV Workshops, pp. 545–553 (2017)
36. Iandola, F.N., Han, S., Moskewicz, M.W., Ashraf, K., Dally, W.J., Keutzer, K.: Squeezenet: alexnet-level accuracy with 50x fewer parameters and 0.5 mb model size. arXiv preprint arXiv:1602.07360 (2016)
37. Kim, Y.D., Park, E., Yoo, S., Choi, T., Yang, L., Shin, D.: Compression of deep convolutional neural networks for fast and low power mobile applications. arXiv preprint arXiv:1511.06530 (2015)
38. Molchanov, P., Tyree, S., Karras, T., Aila, T., Kautz, J.: Pruning convolutional neural networks for resource efficient inference. arXiv preprint arXiv:1611.06440 (2016)
39. Howard, A.G., et al.: Mobilenets: efficient convolutional neural networks for mobile vision applications. arXiv preprint arXiv:1704.04861 (2017)
40. Hinton, G., Vinyals, O., Dean, J.: Distilling the knowledge in a neural network. arXiv preprint arXiv:1503.02531 (2015)
41. Simonyan, K., Zisserman, A.: Very deep convolutional networks for large-scale image recognition. arXiv preprint arXiv:1409.1556 (2014)
42. Agustsson, E., Timofte, R.: NTIRE 2017 challenge on single image super-resolution: dataset and study. In: CVPRW, vol. 3, p. 2 (2017)
43. Martin, D., Fowlkes, C., Tal, D., Malik, J.: A database of human segmented natural images and its application to evaluating segmentation algorithms and measuring ecological statistics. In: ICCV, vol. 2, pp. 416–423. IEEE (2001)
44. Huang, J.B., Singh, A., Ahuja, N.: Single image super-resolution from transformed self-exemplars. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 5197–5206 (2015)
45. Blau, Y., Michaeli, T.: The perception-distortion tradeoff. In: CVPR (2018)
46. Ma, C., Yang, C.Y., Yang, X., Yang, M.H.: Learning a no-reference quality metric for single-image super-resolution. Comput. Vis. Image Underst. **158**, 1–16 (2017)
47. Mittal, A., Soundararajan, R., Bovik, A.C.: Making a "completely blind" image quality analyzer. IEEE Sig. Process. Lett. **20**(3), 209–212 (2013)
48. Kingma, D.P., Ba, J.: Adam: a method for stochastic optimization. In: ICLR (2014)
49. Abadi, M., et al.: TensorFlow: large-scale machine learning on heterogeneous systems (2015). Software available from www.tensorflow.org
50. Ignatov, A., Timofte, R., et al.: PIRM challenge on perceptual image enhancement on smartphones: report. In: European Conference on Computer Vision Workshops (2018)