Image Enhancement using Entwined Auto-Encoder CNN with Nested Convolution Kernels

Andrey Shcherbinin[†,*], Ekaterina Safronova[†], Valery Anisimovskiy[†], Ilya Kurilin[†]

[†] Media Processing Team, Samsung R&D Institute Russia

Abstract

We present a novel method for image enhancement aimed at restoring or hallucinating fine-grained natural image details while retaining well-detailed areas intact. To that end, we employ convolutional neural network trained using aligned patches from pairs of high- and low-quality images depicting the same scenery. Our training procedure includes our novel modulated retention loss which makes the learning concentrate on image areas requiring improvement, while retaining the rest. To address the problem of large-scale consistency of fine-grained details (for example, integrity of long hair strands), we propose the use of nested convolution kernels, which allows leveraging fractal self-similarity of feature maps produced from the input image. Our experiments show clear improvement of subjective quality of fine-grained details (human hair, garment fabric) in image areas which suffered from detail degradation. Objective quality measurements (using non-reference image quality metrics) show competitive performance of our method compared to the state-of-the-art image enhancement methods.

Keywords: image enhancement, machine learning, computer vision, convolutional neural network

1. Introduction

Photographic image enhancement is a long-standing task in image processing and computer vision. The need for image enhancement stems from various sources of photographic image deficiency: image sensor noise, insufficient resolution of camera optics, low light conditions, limited dynamic range of imaging device, etc. Vast body of work has been produced over decades to improve various aspects of photographic images, like contrast, brightness, sharpness, dynamic range, noise level, resolution, etc. Lately, in the wake of growing popularity of convolutional neural networks (CNNs), CNN-based methods of image enhancement achieved remarkable progress in improving image quality [1], [2].

However, many image enhancement methods (including CNN-based ones) focus on improving global image characteristics, like brightness or dynamic range [3], [4], [5]. On the other hand, many methods aimed at improving fine-grained image details often tend to make already crisp image details even crisper, thereby introducing undesirable artefacts, while being unable to create lost image details in over-smoothed areas.

In contrast, CNN-based image super-resolution approaches, being one the most actively developed and successful family of image enhancement methods, are often able to restore lost fine-grained details, while preserving well-detailed regions. These approaches, however, provide improved detalization by increasing image resolution, thereby increasing storage size [1], [2].

In our work, we try to overcome many of the aforementioned shortcomings. We propose CNN-based method of fine-grained image detail enhancement, based on supervised training, capable of improving high-resolution images (e.g. 4K) without changing image size by restoring or hallucinating consistent image details even in smooth and noisy areas, while

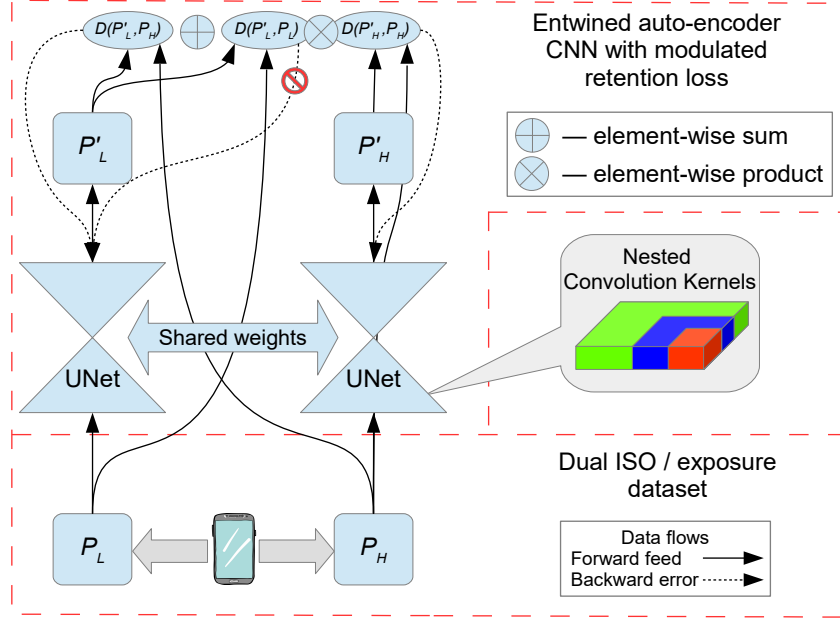[*]s.andrey@samsung.com, katyasafit@gmail.com, avn@mccme.ru, i.kurilin@samsung.com

Figure 1. An overview of our method.

maintaining large-scale integrity of created details. Our contributions, by introducing which we achieve such results, are the following:

- **Dual ISO/exposure training dataset collection procedure** provides an easy and routine way of creation of the dataset consisting of pixel-aligned pairs of images of the same scene with significantly different quality of fine-grained details while having largely the same global image characteristics.
- **Entwined auto-encoder CNN with modulated retention loss** is tailored so as to drive the training process towards both improving low-quality image regions as well as retaining high-quality regions.
- **Nested convolution kernels** allow us to enlarge receptive field of the CNN without increasing trainable weights amount, while, at the same time, providing consistency of large-scale details.

2. **Related Work**

Image enhancement and, in particular, image detail enhancement is a well-studied area with long history of developments. Unsharp masking [6] is well-known and widely-used method aimed at enhancing image edges. More recently, Aubry et al. [7] proposed local Laplacian operator for enhancing details. Such approaches are based on heuristic rules and are typically unable to create fine-grained details in smooth image areas (where the details may have been smoothed out by, for example, preceding denoising operation) and often suffer from excessive sharpening of already sharp edges.

Lately, CNN-based methods have been widely adopted for various image processing tasks, including image enhancement. Ignatov et al. [5] developed the method of mapping from mobile phone camera photos to DSLR-quality photos. Their method is built upon Generative Adversarial Networks (GANs) [8] and the assumption that the overall perceptual image quality can be decomposed into three independent parts: color quality, texture quality and content quality, leading to the corresponding loss function decomposition. They have
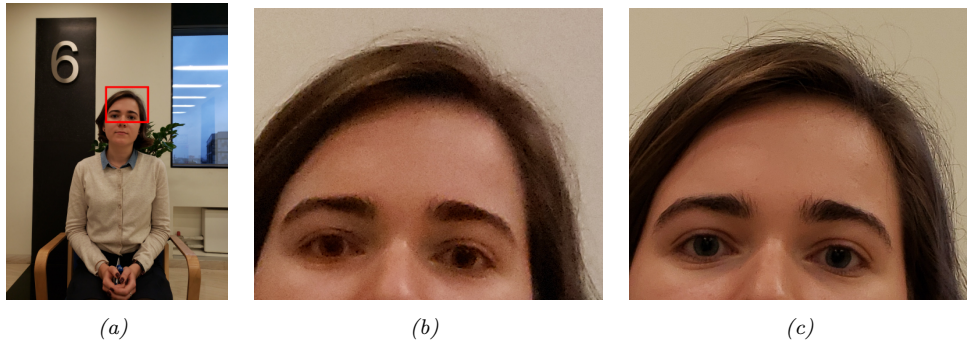
Figure 2. Example of paired images: (a) a scene for each photo; (b) crop from low quality image; (c) crop from high quality image.

also collected the large-scale DPED dataset taken synchronously by a DSLR camera and 3 low-end cameras of smartphones in a wide variety of conditions. Since the training image pairs of this dataset cannot be densely matched, their method is not able to benefit from dense pixel-aligned regression, which limits amount of detail improvement provided by the method. Further development of this approach [9] leverages the idea of cycle-consistent GANs [10] to allow for training on unpaired low- and high-quality images, thereby making this method weakly-supervised (in contrast to fully-supervised one in [5]). Though showing impressive overall image quality improvement (in terms of brightness, contrast, saturation and color quality), this method still lacks fine-grained detail enhancement strength. Another cycle-consistent GAN-based method [4] is using UNet [11] augmented with global features, Wasserstein GAN [12] with adaptive weighting scheme and individual batch normalization to improve the stability of GAN training for their application. Although this approach allows to improve overall quality of images in terms of sharpness and color rendition, it is still not able to recover smeared or smoothed-out image details.

While unsupervised or weakly-supervised methods provide a strong advantage of easy training data collection, fully-supervised methods, trained on densely pixel-aligned pairs of low- and high-quality images, are still the best-performing in terms of fine-grain detailization. One of the most popular directions of such methods related to image detail enhancement is CNN-based image super-resolution. An example of such an approach is the method of [3], which first downsamples low-quality image with a factor of 4, then processes the resulting low-resolution features by a series of residual blocks before upsampling back to the original resolution. Usage of desubpixel modules proposed in [3] allows them to achieve good resulting image quality, while performing most of the processing in low resolution helps to keep computational complexity low. Other examples are [1] and [2], which won NTIRE 2017 and NTIRE 2018 Challenges on Single Image Super-Resolution, respectively. They achieve their results by careful design of residual blocks of their CNN architectures trained within the same fully-supervised super-resolution scenario.

3. Method

In this section, we first describe our dual ISO/exposure dataset and procedure of collection. Then we introduce proposed Siamese enhancement CNN with the total training loss being composed of differences between input and output patches. Next we present a special weight structure coined nested convolution kernels used in the network encoder.
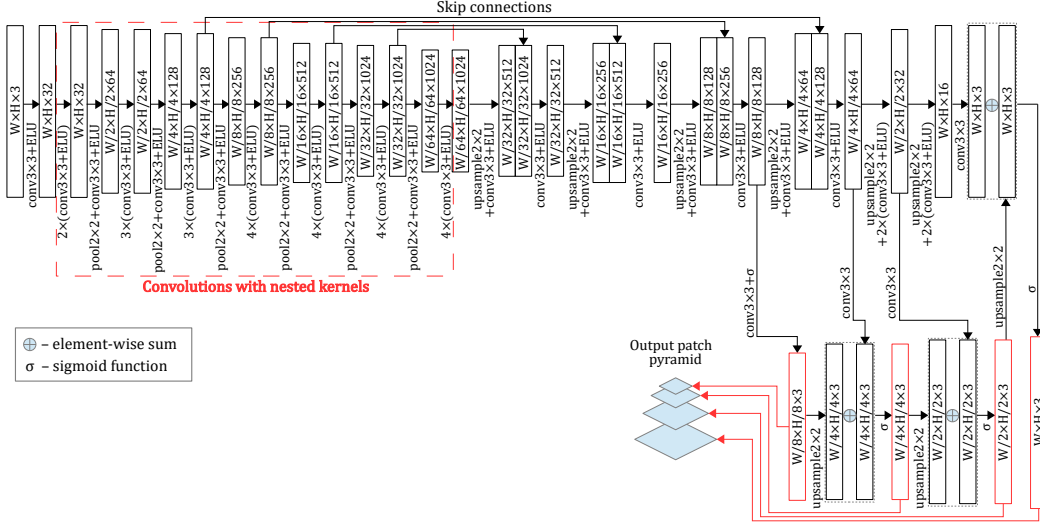
Figure 3. The proposed UNet-style model architecture.

3.1. Dual ISO/exposure dataset

For supervised image enhancement methods, training dataset plays a crucial role in achieving good results. In supervised super-resolution methods, widely used approach is to use a dataset containing high quality images (e.g. DIV2K dataset [13]) for ground-truth target and their artificially degraded versions as low quality input to the model. Bicubic downsampling is often employed as an artificial degradation operator. We argue that this approach may lead to learning the inverse of the degradation operator, which is not always a desirable outcome, especially when the method is going to be applied to the real-world photographs with more complex degradation. In view of this point, we propose the Dual ISO/exposure dataset for image enhancement.

The overall structure of our solution is shown in Figure 1. In the first step, using burst photo capture Android application, the dataset is readily collected. A pair of photos of the same scene is captured having two quality settings: (a) a high quality image is taken with low ISO speed and long exposure time; (b) high ISO speed and short exposure time were set for taking a low quality image. We have assembled 370 pairs of photographs at 4K resolution capturing people at different distances, both indoors and outdoors. Figure 2 shows an example of a pair of images in the dataset.

3.2. Entwined auto-encoder architecture

The proposed model consists of the two Siamese UNet-style CNNs, as illustrated in Figure 1. Each of the low and high quality image patches is independently processed by the same enhancement network having UNet-style architecture (autoencoder with bottleneck and skip connections) which is shown in Figure 3. The output of the enhancement CNN is a 4-level pyramid of the output patch. The encoder of the network is using nested convolution kernels, except for the first convolution. ELU [14] activation is used in the CNN.
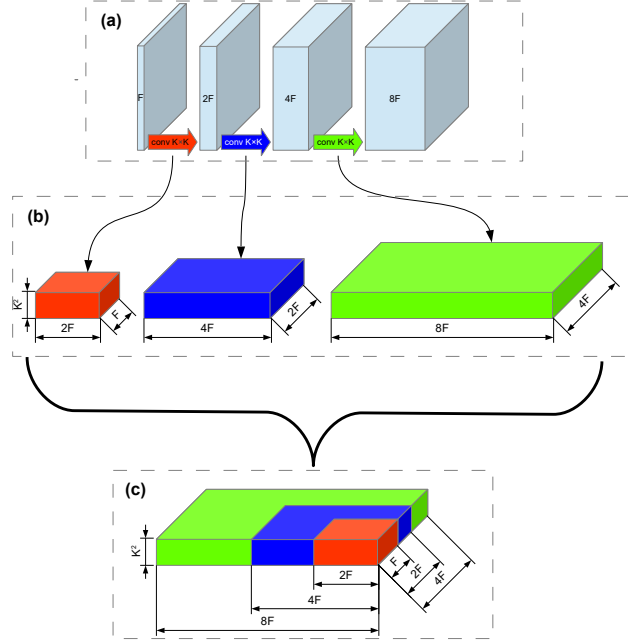
Figure 4. Illustration of the nested convolution kernels concept. (a) Consecutive feature maps produced by $K \times K$ convolutions. (b) Convolution kernels in conventional approach. (c) Nested convolution kernels. (First two dimensions $K \times K$ merged into one of K^2).

3.3. **Nested kernels of CNN convolutions**

Previous CNN approaches are using convolution kernels which are updated independently during training, which often results in large model capacity, making it prone to overfitting thereby being less efficient in processing unfamiliar data.

Instead of maintaining separate independent convolution kernels, our approach allocates a single pool of kernel weights having the size of the largest convolution in the encoder of the enhancement CNN. Each convolution (except for the first one, which is separate) is assigned a subset of allocated kernel weights.

Figure 4 shows a sequence of convolutions example: three consecutive $K \times K$ convolutions successively convert $H \times W \times F$ feature tensor to $H \times W \times 2F$, $H \times W \times 4F$ and $H \times W \times 8F$ tensors. The largest convolution has kernel weights tensor with dimensions $K \times K \times 8F \times 4F$ which is allocated by our method and then distributed to all three convolutions: the first one receives $K \times K \times 2F \times F$ subtensor, the second one receives $K \times K \times 4F \times 2F$ subtensor, the third one receives the whole $K \times K \times 8F \times 4F$ tensor. During the training, the kernel weights of the convolutions are updated according to this assignment. This allows us to drastically enlarge CNN receptive field without increase in CNN weights amount thereby providing larger visual context for image processing without the risk of overfitting.

Another advantage of applying nested convolution layers is that they provide long-range fine-grained detail consistency. Each nested kernel, while processing its scale, provides fractal self-similarity clues for the enclosing kernel processing subsequent scale. Since natural images contain many elements, which possess fractal self-similarity property (like the leaves in Figure 5), these are highly efficient clues for natural image understanding. This way, the same set of kernel weights is made suitable for processing different scales of image.
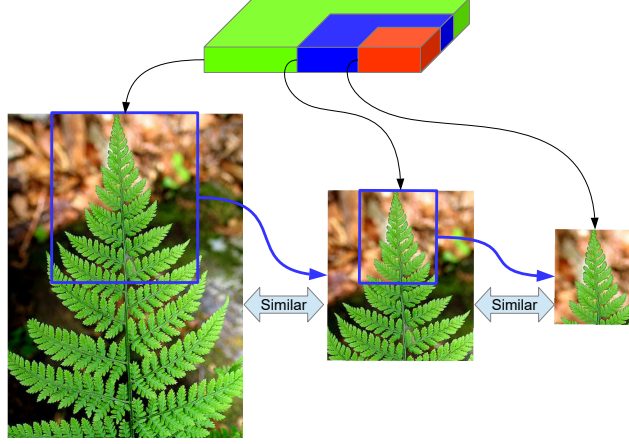
Figure 5. Illustration of nested kernels providing fractal self-similarity clues with an example of fractal self-similarity in the nature.

Thus, due to nested convolution kernels approach, the enhancement CNN shown in Figure 3 contains relatively small number of trainable weights: 25.1M. Without nested convolution kernels, the same architecture contains 53.5M trainable weights, while having the same receptive field.

3.4. **Loss function**

Input pair of aligned image patches, P_L and P_H, are separately processed by the Siamese enhancement CNN to produce output patches, P'_L and P'_H. Many previous learning-based image enhancement approaches are based on direct image-to-image regression which is formed by the difference between the output patch P'_L and the input high quality patch P_H: $\mathcal{L}_R = \frac{1}{N} \sum_{i,j,k} \left[D(P'_L, P_H) \right] (i,j,k)$, where $D(\cdot, \cdot)$ is a difference operator, N is the size of the patch, i, j, k are the spatial and channel indices. However, such approach is prone to creating artefacts in good quality image areas.

So, in addition to the regression loss, modulated retention loss is proposed, which is defined as $\mathcal{L}_M = \frac{1}{N} \sum_{i,j,k} \left[D(P'_L, P_L) \bigodot D(P'_H, P_H) \right] (i,j,k)$, where $\bigodot$ is element-wise product, and helps to preserve already crisp image details. A total loss function for each level l of the output patch pyramid is a sum of regression and modulated retention differences: $\mathcal{L}^l_T = \mathcal{L}^l_R + \mathcal{L}^l_M$. The error back-propagation from $D(P'_L, P_L)$ factor of the modulated retention difference is blocked, so that the loss gradient w.r.t. the weights $\boldsymbol{\Theta}$ of the Siamese enhancement CNN is calculated as the following:

$$\frac{\partial \mathcal{L}^l_T}{\partial \boldsymbol{\Theta}} = \frac{\partial D(P'^l_L, P^l_H)}{\partial \boldsymbol{\Theta}} + D(P'^l_L, P^l_L) \cdot \frac{\partial D(P'^l_H, P^l_H)}{\partial \boldsymbol{\Theta}}. \tag{3.1}$$

For patch difference operator $D(\cdot, \cdot)$, both L_1 and VGG-based perceptive difference are used. L_1 is calculated in YCbCr (in the ITU-R BT.601 standard) color space. Loss function $\mathcal{L}^l_T$ is calculated at each level of the output patch pyramid, so the total loss is computed as follows:

$$\mathcal{L} = \sum_{l=1}^{4} \frac{1}{N_l} \sum_{i,j,k} \left[\left| P_L^{'(l)} - P_H^{(l)} \right| + \left| P_L^{'(l)} - P_L^{(l)} \right| \cdot \left| P_H^{'(l)} - P_H^{(l)} \right| \right] (i,j,k) + \frac{1}{N_f} \sum_{i,j,k} \left[|\varphi(P_L') - \varphi(P_H)| \right] (i,j,k), \tag{3.2}$$

where φ is VGG-19 4-th layer transform, l is image pyramid level index, N_l is the size of the output at l-th level of the pyramid, N_f is the number of elements of VGG 4-th layer features, i, j, k are the spatial and channel indices. The combination of regression and modulated retention losses provides selective enhancement of poor quality image areas without impairing good quality areas by restoring or hallucinating perceptually-motivated fine-grained image details.

4. **Results**

In this section we provide details of the implementation of our method, describe the testing dataset, evaluation metrics and compare the performance of our method with prior image enhancement methods.

4.1. **Implementation details**

The architecture of the enhancement CNN is described in detail in the section 3.2. The model was implemented in MXNet framework [15] and trained for 2300 epochs on Nvidia GeForce GTX 1080 Ti GPU with a batch size of 8 and crop size of 320×320 using Adam [16] with the learning rate of $\lambda = 10^{-4}$ and $\beta_1 = 0.9$, $\beta_2 = 0.999$, $\epsilon = 10^{-6}$. The number of epochs was chosen so that the training converges after this number.

4.2. **Dataset**

We have carried out several experiments to evaluate the proposed method using our Dual ISO/exposure dataset for training and validation. Since the task is quite specific, it is hard to find universally recognized dataset that comply with the requirements: it must include low quality high resolution photographs capturing people at different distances, while low quality occurs due to camera settings and scene illumination. Therefore we have selected images with higher than 1K resolution from Helen Facial Feature Dataset [17], consisting of portrait images gathered from Flickr, for evaluation, as it is the most representative subset from the point of view of the task being solved (improvement of human facial details).

4.3. **Metrics**

Since our method peforms image detail re-drawing (so as the re-drawn details look natural) rather than reconstruction of real true details, it cannot be evaluated by full-reference image quality metrics (the re-drawn details cannot perfectly match original true details). So, we use well-established no-reference image quality assessment metrics: CPBD [18] (Cumulative Probability of Blur Detection), BIQAA [19] (Blind Image Quality Assessment through Anisotropy), WaDIQaM-NR [20] (Weighted Average Deep Image QuAlity Measure for No-Reference) and BRISQUE [21] (Blind/Referenceless Image Spatial QUality Evaluator). CPDB is a non-trained metric assessing only image blurriness (without taking into account human perception), BIQAA is another non-trained method measuring the average

Table 1. Results on the subset of Helen Facial Feature Dataset [17]. Metrics: CPDB [18], BIQAA [19], WaDIQaM [20], BRISQUE [21]. The best results are highlighted by bold font.

Methods	Non-trained metrics (higher is better)		Trained metrics (lower is better)	
	CPDB	BIQAA	WaDIQaM	BRISQUE
Input	0.55	0.0015	44.6	28.1
FEQE [3]	0.58	0.0023	**39.4**	26.3
DPE [4]	0.58	0.0005	47.9	23.9
DPED [5]	0.56	0.0010	42.0	34.3
EDSR [1]	**0.60**	0.0023	42.9	24.4
WDSR [2]	0.56	0.0015	45.4	29.1
Ours	0.58	0.0023	39.7	19.8
Ours (w/o nested convolutions)	0.53	0.0025	39.8	18.5
Ours (w/o nested convolutions, only $\mathcal{L}_R$)	0.57	**0.0026**	40.5	**16.4**

anisotropy of the image which can be used as a quality metric, WaDIQaM-NR is CNN-based metric trained on human-labeled image quality dataset, BRISQUE is a support vector machine regressor, which uses natural scene statistics as features, trained on LIVE IQA database [22].

4.4. **Enhancement performance evaluation**

Quantitative results. Table 1 shows the results of our method in comparison with the previous works on image enhancement.

Since our goal was to enhance image quality by restoring fine-grained details and keeping already detailed areas intact, we pay more attention to metrics that assess image quality with regard to subjective opinion of human, i.e. WaDIQaM and BRISQUE. From Table 1 one can see that our method outperforms all previous approaches in terms of BRISQUE and most of them in terms of WaDIQaM. CPDB and BIQAA measure such characteristics as image blurriness and average anisotropy, and Table 1 shows that our method improves these properties in comparison with the input images and outperforms most of the previous approaches.

Ablation study. Since our solution consists of several contributions, we performed three experiments and obtained the following models: (1) without neither nested convolutions nor the modulated retention part of the loss function (Model 1), (2) without nested convolutions (Model 2), and (3) the full model (Model 3). Figure 6 shows the results of applying all the three trained models, each row of it presents detail enrichment of different image objects. The first two rows consist of crops from the one image captured at the same resolution and by the similar device as the training set while the second two rows involves the whole photographs from Helen Facial Feature Dataset. Considering the examples of image areas capturing hair, the first and the third rows in Figure 6, one can see that Model 1 (Fig. 6c), as well as Model 2 (Fig. 6d), emphasises already sharp details making them too crisp but doesn't create details in smooth areas. At the same time, Model 3 (Fig. 6e) provides more consistent spatially extended fine-grained details than both Model 1 and Model 2, creating them even in such noisy regions that don't contain any initial details in the input image. As for the second row presenting the example of image area with sweater fabric, one can argue that Model 1 makes strips on the sweater more discernible, Model 2 intensifies the fabric strips even more but creates artefacts in the middle of them, while Model 3 doesn't produce such artefacts and makes the strips distinguishable enough. The last row of the figure presents the example of enhancement result in image areas with human skin. One can see that Model 1 emphasises existing details in this case too, Model 2 does it more
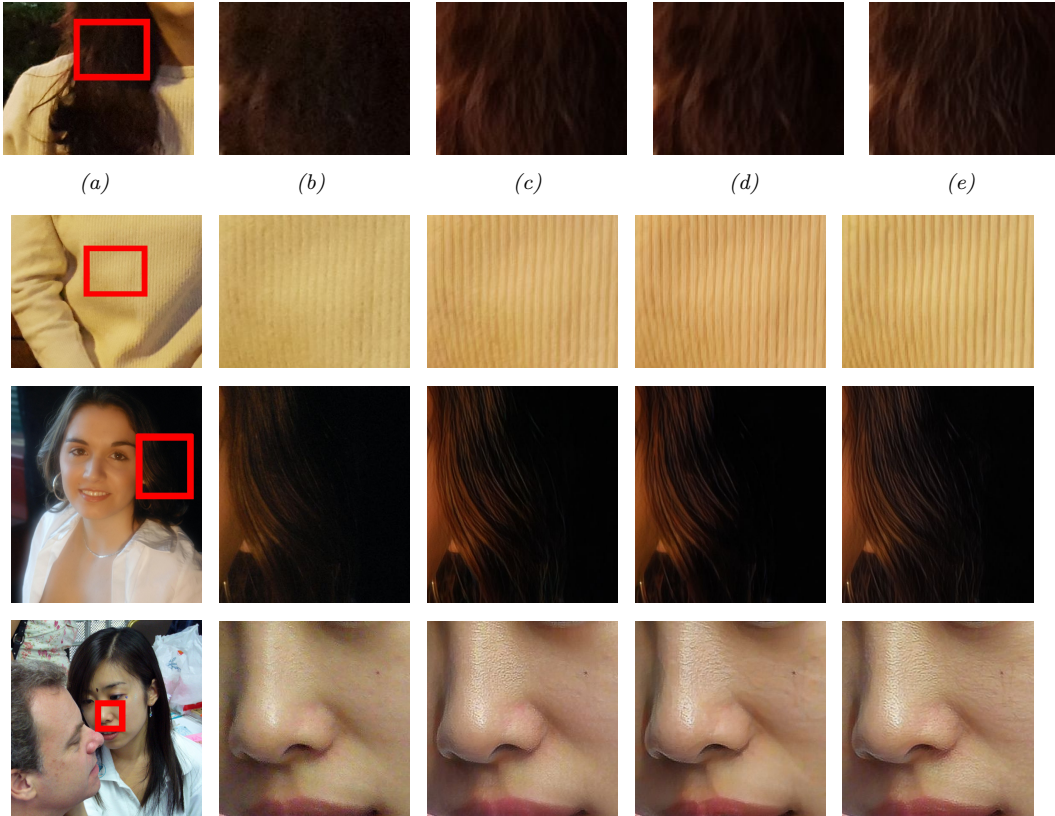
Figure 6. Examples of ablation study experiment results. (a) Input images. (b) Crops from the input images. (c) Crops from the outputs of our model without nested convolutions and modulated retention loss. (d) Crops from the outputs of our model without nested convolutions. (e) Crops from the outputs of our full model. (Better viewed on 4K display with diagonal size of at least 28 inch).

carefully smoothing a lot of details, and Model 3 creates rather realistic and sharp skin texture. According to the results, we see the quality progress from Model 1 to Model 3.

In Table 1 we provide the results of our method obtained by all the three models. We can interpret the results in the following way: the most sophisticated trainable metric, WaDIQaM, which measures overall image quality as weighted average sum of quality of image patches estimated by a trained CNN, provides results which correspond to the qualitative results. So we can argue that the most of improvement comes from the contributions: nested convolutions and the modulated retention part of the loss function. However, we do not see such clear progress of quality estimation with other metrics, that measure average quality across a whole image. These metrics are tailored for measuring specific properties of images, that's why their results are not so consistent. The two of them, CPDB and BIQAA, are fully hand-crafted and not completely correlated with human perception, while the third, BRISQUE, is based on a support vector machine regressor, which is much more simple model than CNN-based WaDIQaM.

Qualitative results. Table 2 and Table 3 show the outputs for several images from the subset of Helen Facial Feature Dataset predicted by previous approaches and our model. In this figure one can see that our model produces richer image detalisation than the previous approaches: e.g. skin, hair, moustache, eyelashes.

Table 2. Comparison of image enhancement models' outputs: FEQE [3], DPE [4], DPED [5] and our model. (Better viewed on 4K display with diagonal size of at least 28 inch).
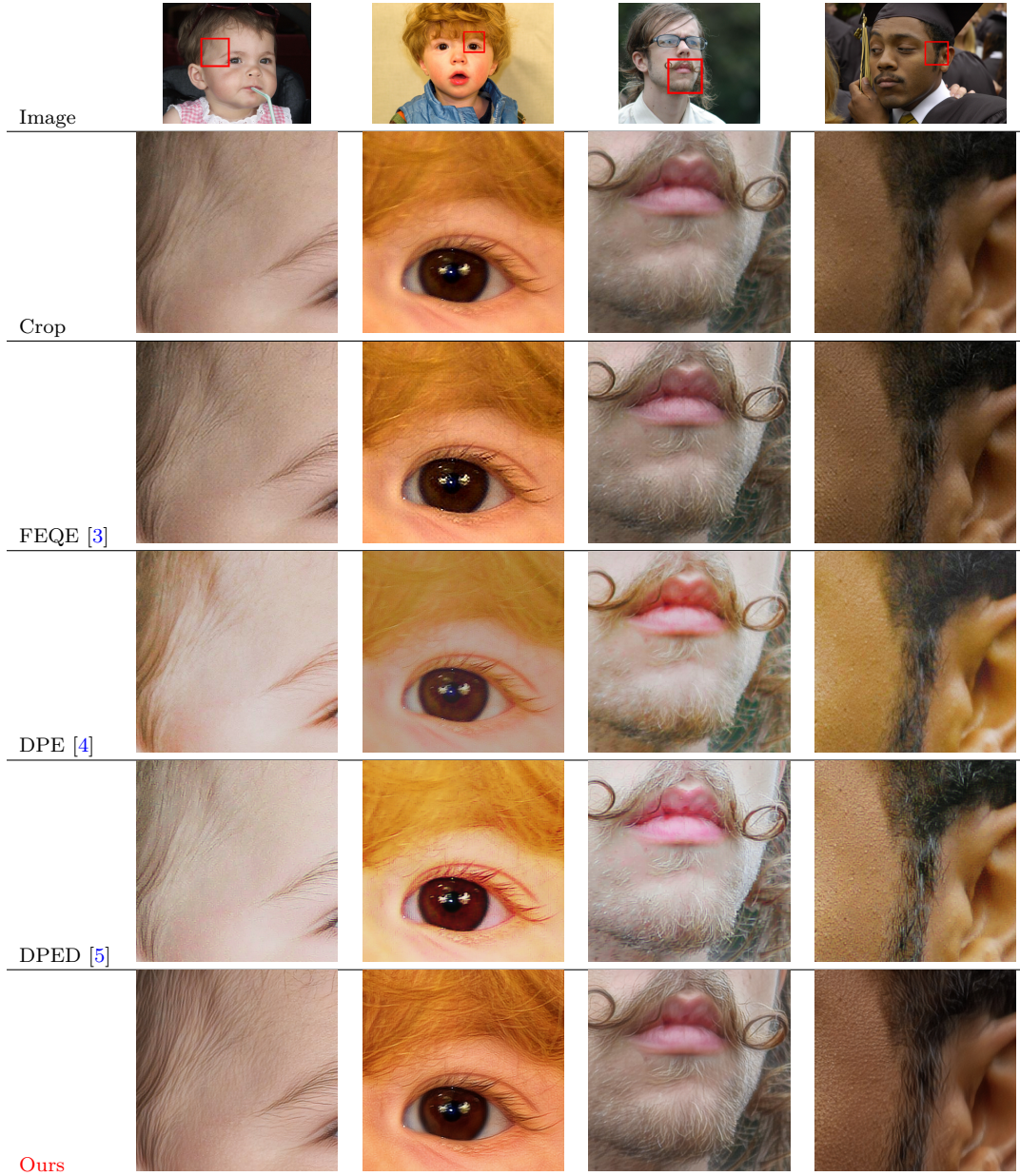
5. **Conclusions**

We have presented entwined auto-encoder CNN for image enhancement. While incorporating some of the ideas from the previous state-of-the-art methods, such as UNet-structured networks and regression loss, we propose significant novel improvements. The model is trained on abundant and diverse dataset of well-aligned low/high quality image pairs easily

Table 3. Comparison of image enhancement models' outputs: EDSR [1], WDSR [2] and our model. (Better viewed on 4K display with diagonal size of at least 28 inch).

and routinely collected, that produces highly-generalizable image enhancement. Furthermore, training with entwined auto-encoder and perceptive loss provides selective enhancement of poor quality areas without impairing good quality areas. Finally, nested convolution kernels in the encoder provide larger visual context for image processing without the risk of overfitting as well as long-range detail consistency. These improvements allow us to outperform the previous results across the set of metrics and show better qualitative results. However, the important limitation of our method is that it works best for images which possess fractal self-similarity property, otherwise the usage of nested convolution kernels may lead to peculiar artefacts.

References

[1] B. Lim, S. Son, H. Kim, S. Nah, and K. Mu Lee. "Enhanced deep residual networks for single image super-resolution". In: *Proceedings of the IEEE conference on computer vision and pattern recognition workshops*. 2017, pp. 136–144.

[2] J. Yu, Y. Fan, J. Yang, N. Xu, Z. Wang, X. Wang, and T. Huang. "Wide activation for efficient and accurate image super-resolution". In: *arXiv preprint arXiv:1808.08718* (2018).

[3] T. Vu, C. Van Nguyen, T. X. Pham, T. M. Luu, and C. D. Yoo. "Fast and efficient image quality enhancement via desubpixel convolutional neural networks". In: *Proceedings of the European Conference on Computer Vision (ECCV) Workshops*. 2018, pp. 0–0.

[4] Y.-S. Chen, Y.-C. Wang, M.-H. Kao, and Y.-Y. Chuang. "Deep photo enhancer: Unpaired learning for image enhancement from photographs with gans". In: *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*. 2018, pp. 6306–6314.

[5] A. Ignatov, N. Kobyshev, R. Timofte, K. Vanhoey, and L. Van Gool. "Dslr-quality photos on mobile devices with deep convolutional networks". In: *Proceedings of the IEEE International Conference on Computer Vision*. 2017, pp. 3277–3285.

[6]   M. Trentacoste, R. Mantiuk, W. Heidrich, and F. Dufrot. "Unsharp masking, countershading and halos: Enhancements or artifacts?" In: *Computer Graphics Forum.* Vol. 31. 2pt3. Wiley Online Library. 2012, pp. 555–564.

[7]   M. Aubry, S. Paris, S. W. Hasinoff, J. Kautz, and F. Durand. "Fast local laplacian filters: Theory and applications". In: *ACM Transactions on Graphics (TOG)* 33.5 (2014), pp. 1–14.

[8]   I. J. Goodfellow, J. Pouget-Abadie, M. Mirza, B. Xu, D. Warde-Farley, S. Ozair, A. Courville, and Y. Bengio. "Generative adversarial networks". In: *arXiv preprint arXiv:1406.2661* (2014).

[9]   A. Ignatov, N. Kobyshev, R. Timofte, K. Vanhoey, and L. Van Gool. "Wespe: weakly supervised photo enhancer for digital cameras". In: *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops.* 2018, pp. 691–700.

[10] J.-Y. Zhu, T. Park, P. Isola, and A. A. Efros. "Unpaired image-to-image translation using cycle-consistent adversarial networks". In: *Proceedings of the IEEE international conference on computer vision.* 2017, pp. 2223–2232.

[11] O. Ronneberger, P. Fischer, and T. Brox. "U-net: Convolutional networks for biomedical image segmentation". In: *International Conference on Medical image computing and computer-assisted intervention.* Springer. 2015, pp. 234–241.

[12] M. Arjovsky, S. Chintala, and L. Bottou. "Wasserstein generative adversarial networks". In: *International conference on machine learning.* PMLR. 2017, pp. 214–223.

[13] R. Timofte, E. Agustsson, L. Van Gool, M.-H. Yang, and L. Zhang. "Ntire 2017 challenge on single image super-resolution: Methods and results". In: *Proceedings of the IEEE conference on computer vision and pattern recognition workshops.* 2017, pp. 114–125.

[14] D.-A. Clevert, T. Unterthiner, and S. Hochreiter. "Fast and accurate deep network learning by exponential linear units (elus)". In: *arXiv preprint arXiv:1511.07289* (2015).

[15] T. Chen, M. Li, Y. Li, M. Lin, N. Wang, M. Wang, T. Xiao, B. Xu, C. Zhang, and Z. Zhang. "Mxnet: A flexible and efficient machine learning library for heterogeneous distributed systems". In: *arXiv preprint arXiv:1512.01274* (2015).

[16] D. P. Kingma and J. Ba. "Adam: A method for stochastic optimization". In: *arXiv preprint arXiv:1412.6980* (2014).

[17] V. Le, J. Brandt, Z. Lin, L. Bourdev, and T. S. Huang. "Interactive facial feature localization". In: *European conference on computer vision.* Springer. 2012, pp. 679–692.

[18] N. D. Narvekar and L. J. Karam. "A no-reference image blur metric based on the cumulative probability of blur detection (CPBD)". In: *IEEE Transactions on Image Processing* 20.9 (2011), pp. 2678–2683.

[19] S. Gabarda and G. Cristóbal. "Blind image quality assessment through anisotropy". In: *JOSA A* 24.12 (2007), B42–B51.

[20] S. Bosse, D. Maniry, K.-R. Müller, T. Wiegand, and W. Samek. "Deep neural networks for no-reference and full-reference image quality assessment". In: *IEEE Transactions on image processing* 27.1 (2017), pp. 206–219.

[21] A. Mittal, A. K. Moorthy, and A. C. Bovik. "No-reference image quality assessment in the spatial domain". In: *IEEE Transactions on image processing* 21.12 (2012), pp. 4695–4708.

[22] H. R. Sheikh, M. F. Sabir, and A. C. Bovik. "A statistical evaluation of recent full reference image quality assessment algorithms". In: *IEEE Transactions on image processing* 15.11 (2006), pp. 3440–3451.