Image Denoising via Generative Adversarial Networks with Detail Loss

Songkui Chen
College of Computer Science and Software Engineering, Shenzhen University, Shenzhen, 518060, P.R. China
chensongkui2017@email.szu.edu.cn

Daming Shi
College of Computer Science and Software Engineering, Shenzhen University, Shenzhen, 518060, P.R. China
dshi@szu.edu.cn

Muhammad Sadiq
College of Computer Science and Software Engineering, Shenzhen University, Shenzhen, 518060, P.R. China
sadiq_paec@yahoo.com

Meilu Zhu
College of Computer Science and Software Engineering, Shenzhen University, Shenzhen, 518060, P.R. China
zhumeilu2016@email.szu.edu.cn

ABSTRACT
Image denoising is a challenging task which aims to remove additional noise and preserve all useful information. Many existing image denoising algorithms focus on improving the typical object measure, peak signal-to-noise ratio (PSNR), and take the mean square error (MSE) as their loss function to train their networks. Although these algorithms can effectively improve the PSNR on the benchmark dataset, their denoised images often lose some important details or become over-smooth in some texture-rich regions. In order to solve this problem, we introduce Generative Adversarial Networks (GAN) and perceptual loss from single image super-resolution (SISR) field into our image denoising work. The GAN and perceptual loss can help our network to better focus on the recovering of details during denoising. To understand easily, we use the term Detail Loss to represent the whole loss which includes the MSE and the perceptual loss. Besides, we propose a new convolutional neural network which achieves state-of-the-art result on PSNR. Our experimental results show that our method outperforms the current state-of-the-art methods on preserving the details during denoising. Compared with the current state-of-the-art methods, the denoised images by our method are clearer, sharper and more realistic on details.

CCS Concepts
• Networks→Layering.

Keywords
Image Denoising; Generative Adversarial Networks; Detail Loss.

1. INTRODUCTION
Image denoising is an important pre-processing step in many computer vision tasks. The task of image denoising is to recover a noise-free image from its noisy observation by removing the noise. For better solving this problem, many approaches have been proposed. Nowadays, with deep learning techniques becoming popular, many algorithms based on deep learning have been proposed [9][10] and most of them perform better than traditional approaches. There are some categories of image denoising algorithms based on deep learning, such as algorithms based on Multi-layer Perceptron Networks (MLP) and algorithms based on convolutional neural networks (CNN) and so on.

ICISS 2019, March 16–19, 2019, Tokyo, Japan

ACM ISBN 978-1-4503-6103-3/19/03…$15.00
https://doi.org/10.1145/3322645.3322656

Multi-layer Perceptron Networks (MLP) [1][7] can approximate arbitrary functions and have strong nonlinearities, so some researchers proposed their image denoising approaches based on MLP. And their experimental results showed that their approaches perform better than traditional methods. However, the MLP connects its layers using full connection, so its number of parameters exponentially grows with the depth of the network. The training will be more difficult for deep network.

Denoising algorithms based on convolutional neural networks (CNN) are most popular algorithms nowadays [11][3]. Because CNN has a structure of local receptive field, which is very conducive to perceive images like human eyes, CNN is very suitable for extracting and learning the features of images. In addition, the numbers of parameters are very less than the MLP. So the algorithms based on CNN perform better than those algorithms based on MLP. What's more, after skip connection [5] and batch normalization [8] been proposed, problems such as gradient vanishing or gradient exploding caused by deeping the CNN can be effectively alleviated.

2. RELATED WORK
Image denoising. In order to improve the PSNR, reconstruction of images by minimizing the MSE is famous. However, MSE minimization usually result in losing some details or becoming over-smooth in some texture-rich regions of the denoised images. To alleviate this problem, [12] proposed a cascade architecture which connects their image denoising network to a high-level vision network, such as a network for image classification. Their denoising network was trained by jointly minimizing the image reconstruction loss MSE and the high-level vision loss. With the guidance of the loss from image classification network, the denoising network is able to generate more visually appealing outputs. However, the denoised image can still be further improved in the visual quality.

Recovering details. In SISR field, to recover more natural and realistic textures in the recovered high-resolution (HR) images, [9] introduced the GAN into their SISR work and proposed perceptual loss which includes content loss term and adversarial loss term. As the results, their network can better recover the important high-frequency details in the recovered HR images.

Because the preservation of detail on image denoising task is also a tricky problem now, we try to introduce their perceptual loss to our image denoising work. The perceptual loss can well measure the difference of details between the denoised image and the ground truth.

Generative Adversarial Networks. Ian Goodfellow in 2014 proposed a new training framework named as Generative Adversarial Networks (GAN) [6]. Generally, the GAN includes two networks one is generator network and 2nd one is discriminator network. The generator network is responsible for learning the distribution of real dataset and generating samples. The generative samples should be so similar to the real samples that the discriminator network cannot discriminate the generative samples from the real samples. In contrast, the discriminator network is trained to discriminate the generative samples from the real samples correctly. The two networks continuously improve their performance by iteratively adversarial training with each other. The GAN have been applied to many tasks, such as super-resolution, image inpainting, semantic segmentation and so on. Recently, GAN is applied for image denoising [3] which learns the distribution of real noise from the their noisy images and then to generate many couples of images for their training.

3. METHODOLOGY

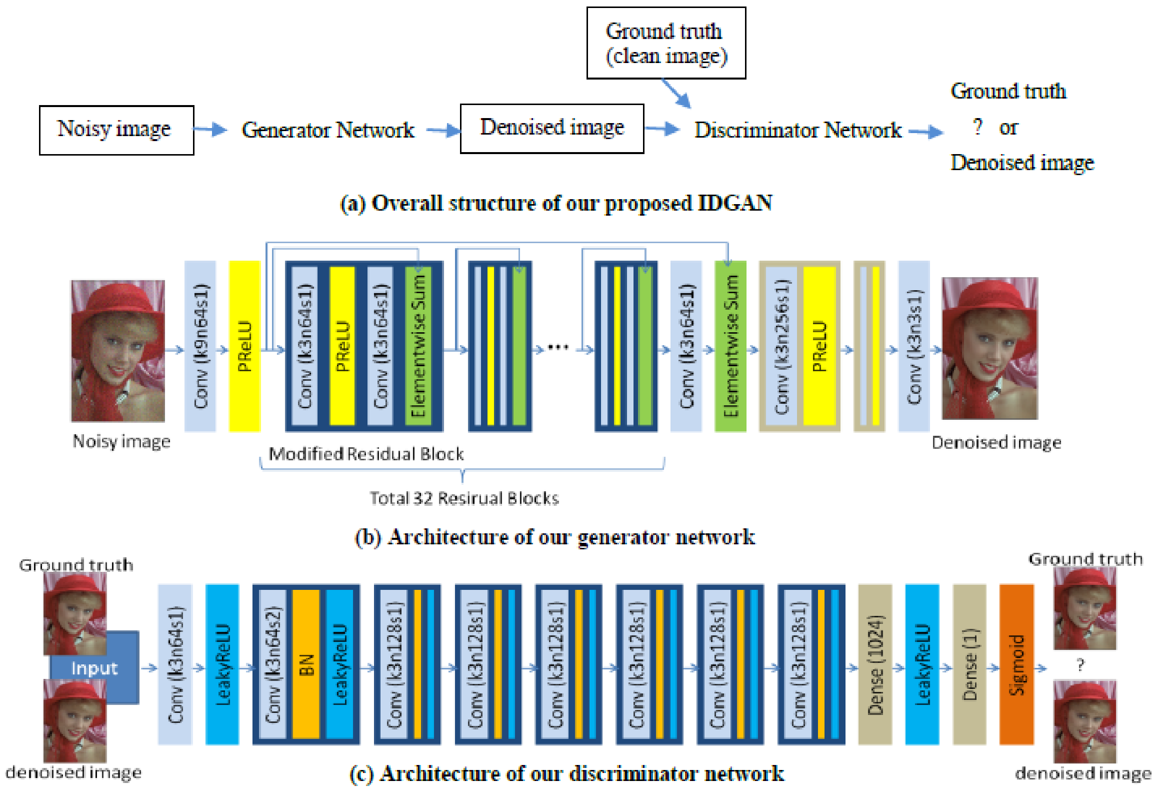

Figure 1. The overall structure of our proposed model and the architectures of our generator network and discriminator network. Each convolution layer is indicated with corresponding kernel size (k), number of feature maps (n) and stride (s).

Image denoising is a classic topic in low-level computer vision tasks. Following the degradation model $y = x + v$, image denoising targets at recovering a noise-free image x from its noisy observation y by reducing the noise v.

Our goal is to learn a mapping function G that can estimate for a given noisy image its corresponding noise-free image. And the recovered noise-free image should preserve the details well. We represent the mapping function G as a feed-forward CNN parametrized by θ_G. The mapping relationship is

$$\hat{x} = G(y) \tag{1}$$

In order to recovering an ideal $\hat{\mathbf{x}}$, we mainly make two innovations. Firstly, we propose a new CNN which achieves state-of-the-art result on PSNR. Besides, inspired by [9], for better preserving the important details, we introduce GAN and perceptual loss into our image denoising work. We name the whole loss, including the common loss MSE and the introduced perceptual loss, as Detail Loss.

3.1 Adversarial Network Architecture

In this paper, different from [3] which uses the GAN to learn the distribution of noise for data augment, we use the GAN to directly learn the process of image denoising. We take the noisy image as the input of our generator network, and take the ground truth, i.e. the clean image, as the real images. The generator network is directly responsible for the denoising and its output is the corresponding denoised image. And the discriminator network is responsible for discriminating the denoised image from the ground truth. Through the iteratively adversarial training between the two networks, the generator network can well recognize the noise model and remove the noise from noisy image. Besides, compared to the MSE which encourages finding solutions by minimizing pixel-wise error measurements, the GAN encourages perceptually superior solutions residing in the subspace, the manifold, of natural images [9]. So compared to the denoised images of other methods without using GAN, the denoised image from our generator network will be higher in the visual quality.

3.1.1 Generator Network

We design our generator network based on the architecture of [16]. and make some changes. Firstly, because the batch normalization layer gets rid of range flexibility from networks by normalizing the features, which is not good for the learning capacity of denoising network, we remove the batch normalization layers on our generator network. Secondly, because the batch normalization layers consume the same amount of memory as the preceding convolutional layers, we increase the number of residual block from 16 to 32 while keeping the GPU memory almost unchanged. Thirdly, we remove the PixelShuffler layers because our denoising work doesn't need to increase the resolution of input image. The architecture of our generator network is showed in Figure 1. The adversarial loss of generator network is l_{gen} which is included in the perceptual loss and will be described in the Section 3.2.1.

3.1.2 Discriminator Network

The architecture of our discriminator network is consistent with their discriminator architecture [16] and is showed in Figure 1. The adversarial loss of discriminator network is l_{dis}, as follows:

$$l_{dis} = -\log\big(D(x)\big) - \log(1 - D(\hat{x})) \tag{2}$$

Where the function $D(\cdot)$ represents the discriminator network parametrized by θ_D. And $\hat{x}$, x are the denoised image generated by generator network and the corresponding ground truth, respectively.

3.2 Detail Loss

3.2.1 MSE

This is the most widely used optimization target for image denoising on which many state-of-the-art approaches rely. By minimizing this loss, network can find a solution that the denoised image can be high on PSNR. This pixel-wise MSE loss is calculated as:

$$l_{MSE} = \frac{1}{WH}\sum_{i=1}^{W}\sum_{j=1}^{H}(x_{i,j} - \hat{x}_{i,j})^2 \tag{3}$$

Where the W and H are weight and height of the image respectively.

However, while achieving particularly high PSNR, solutions of MSE optimization problems often lose some important details or are over-smooth in some texture-rich regions. Instead of relying on pixel-wise losses, we optimize our generator network relying on the new loss Detail Loss. The composition of this loss is as (4).

$$l_{detail} = \alpha_{MSE} l_{MSE} + l_{per} \tag{4}$$

Where the l_{detail} represents the detail loss, the l_{per} represents the perceptual loss, the l_{MSE} represents the MSE loss and the α_{MSE} represents the weight of the l_{MSE}.

3.2.2 Perceptual Loss

The perceptual loss includes two parts: content loss l_{con} and adversarial loss l_{gen}, as (5). The adversarial loss comes from the GAN, as (6). And the content loss is computed by calculating the Euclidean distances between feature maps taken from a pretrained 19-layer VGG network which takes the denoised image and the ground truth as input respectively, as (7). Because the feature maps of these deeper layers focus purely on the content [9], minimization of this content loss can help to preserve the details during the denoising process.

$$l_{per} = \alpha_{con} l_{con} + \alpha_{gen} l_{gen} \tag{5}$$

Where the α_{con} and α_{gen} are the respective weight of the two losses.

$$l_{gen} = -\log(D(\hat{x})) \tag{6}$$

Where the function $D(\cdot)$ represents the discriminator network and $\hat{x}$ represents the denoised image.

$$l_{con} = \frac{1}{W_{5,4}H_{5,4}}\sum_{i=1}^{W_{5,4}}\sum_{j=1}^{H_{5,4}}(\varphi_{5,4}(x)_{i,j} - \varphi_{5,4}(\hat{x})_{i,j})^2 \tag{7}$$

Where the function $\varphi_{5,4}(\cdot)$ indicates the feature maps obtained by the 4th convolution (after activation) before the 5th maxpooling layer within the VGG19 network. And the $H_{5,4}$ and $W_{5,4}$ describe the dimensions of the feature maps in height and weight respectively. Why we choose this layer is that [9] proved that using the feature maps extracted from this layer can better recover texture details.

4. EXPERIMENTS

4.1 Datasets

Our generator network takes RGB images as input, and as output gives back its denoised images. We add independent and identically distributed Gaussian noise with zero mean to the clean images as the noisy input images during training. We use the DIV2K dataset [2] as our training dataset. The DIV2K dataset is a high-quality (2K resolution) color image dataset for image restoration tasks. The DIV2K dataset consists of 800 training images, 100 validation images, and 100 test images. Because the 100 test images are not released, we use the 800 training images and 100 validation images, totally 900 images, as our clean images for training. We compare our denoising network with other state-of-the-art color image denoising approaches on various noise levels: σ = 25, 35 and 50. We evaluate the denoising performance over the widely used Kodak [18] dataset which consists of 24 color images.

4.2 Training Details

We train all networks on a NVIDIA TITAN XP GPU. For each mini-batch we crop 16 random 100 × 100 sub images of the high-resolution training dataset. We scale the range of the input noisy images to [0,1] and scale the ground truths, i.e. the clean images, to [-1,1]. The MSE loss is thus calculated on images of intensity range [-1,1]. For optimization we use Adam with β_1 = 0.9. We have two training phases. In the first phase, we just pre-train our generator network, and we use the MSE as the loss function to guide the optimization. The weight of the MSE loss is 1. We name this generator network as **IDResNet**. The IDResNet is trained with an initial learning rate of 10^{-4} and 4×10^5 update iterations. The learning rate is divided by 10 after every 2×10^5 iterations. In the second phase, we employ the trained MSE-based IDResNet network as initialization for the generator network when training the actual GAN to avoid undesired local optima. And we optimize the generator network by minimizing the detail loss. The weights of the content loss term and the adversarial loss term are 0.0061 and 0.001, respectively. And the weight of the MSE loss term is 0.001. We name this network as **IDGAN**. The IDGAN is trained with an initial learning rate of 10^{-4} and 2×10^5 update iterations.

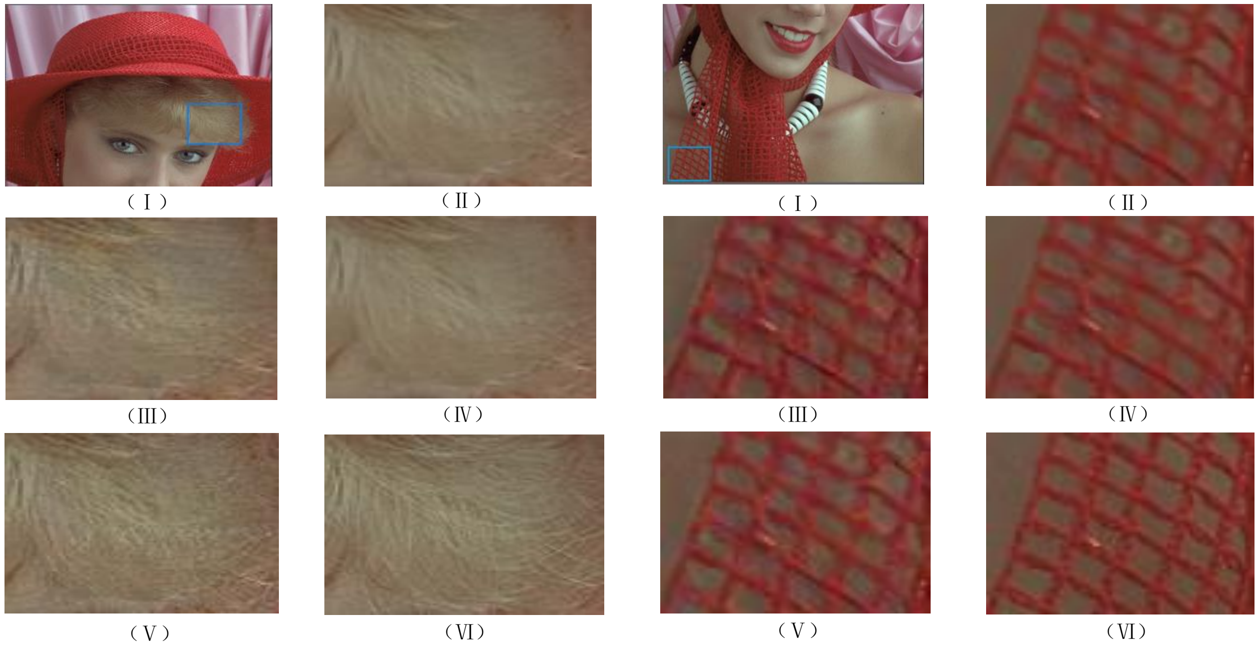

Figure 2. Image denoising examples from Kodak dataset. We show （Ⅰ） the ground truth image and the zoom-in regions of: （Ⅱ）the denoised image by DnCNN; （Ⅲ）the denoised image by DEEPDENOISING;（Ⅳ）the denoised image by our IDResNet network; （Ⅴ）the denoised image by our IDGAN network and （Ⅵ）the ground truth.

The learning rate is divided by 10 after every 10^5 iterations. We alternate updates to the generator network and discriminator network, which is equivalent to k = 1 as used in [6]. Our generator network has 32 identical residual blocks. Our implementation is based on Tensorflow. It is worth noting that we can apply the trained generator model to images of arbitrary size as it is fully convolutional. We train a different denoising network for each noise level in our experiment.

4.3 Evaluation on Kodak Dataset

Table 1. The average denoising results (PSNR) of different methods on Kodak dataset. The best result is shown in bold.

		Average
σ = 25	**CBM3D**	31.81
	MCWNNM	31.35
	DnCNN	32.35
	DEEPDENOISING	32.39
	IDResNet	**32.52**
	IDGAN	30.58
σ = 35	**CBM3D**	30.04
	MCWNNM	29.70
	DnCNN	30.76
	DEEPDENOISING	30.83
	IDResNet	**30.93**
	IDGAN	29.41
σ = 50	**CBM3D**	28.62
	MCWNNM	28.02
	DnCNN	29.16
	DEEPDENOISING	29.27
	IDResNet	**29.45**
	IDGAN	29.19

We compare our denoising networks with several state-of-the-art color image denoising networks. Table 1 shows the average PSNR results on Kodak dataset for CBM3D [4], MCWNNM [16], DnCNN [17], DEEPDENOISING [12], and our proposed IDResNet and IDGAN on various noise levels: σ = 25, 35 and 50. From the Table 1, it is clear that our proposed IDResNet network outperforms all the compared approaches quantitatively across different noise levels. Note that the data of other methods in Table 1 is from [14] which obtained data by implementing the codes from the respective authors' websites.

Meanwhile, we can see from the Table 1 that our IDGAN network doesn't perform good on the PSNR than our IDResNet network, even lower than other comparing approaches. However, it is worthy to emphasize that the goal of our IDGAN network is not to achieve higher PSNR, but is to better preserve the details on the denoised image, and thus improving the visual quality. From [15] we can observe that the typical metric for image denoising, PSNR, sometimes correlate poorly with human assessment of visual quality. And as the visual comparison illustrated in Figure 2, although it doesn't perform good than our IDResNet network and other comparing approaches on the PSNR, it performs better on preserving the details. In the Figure 2, we can easily observe that on the details the denoised images by our IDGAN are clearer and sharper. Although the denoised images by DEEPDENOISING are almost as sharp as our IDGAN are on the details, their details look like high-frequency artifacts. In contrast, the details by our IDGAN not only are clearer, but also are more realistic.

5. CONCLUSION AND DISCUSSION

In this paper, we propose a new algorithm focusing on better preserving the details during the image denoising process. Based on the adversarial network architecture, our generator network is optimized by minimizing a newly proposed loss, detail loss. The loss can well measure the difference of details between the denoised image and the ground truth, so the minimization of this loss can guide the generator network to better preserve the details during the denoising. We verify the effectiveness of our algorithm on the benchmark dataset, Kodak dataset. The experimental results show that our IDResNet network outperforms all the compared approaches quantitatively on the PSNR. And although our IDGAN network doesn't perform best on the PSNR, it performs best on preserving the details. In the future, how to improve the PSNR while preserving well the details during denoising is the work we plant to do.

6. REFERENCES

[1] Harold Christopher Burger, Christian J. Schuler, and Stefan Harmeling. 2012. Image denoising with multi-layer perceptrons, part 1: comparison with existing algorithms and with bounds. 1–38.

[2] Mark J. Cartledge. 2004. Affective theological praxis: Understanding the direct object of practical theology. International Journal of Practical Theology 8, 1: 34–52.

[3] Jingwen Chen, Jiawei Chen, Hongyang Chao, and Ming Yang. 2018. Image Blind Denoising With Generative Adversarial Network Based Noise Modeling. Cvpr2018: 3155–3164.

[4] Kostadin Dabov, Alessandro Foi, Vladimir Katkovnik, and Karen Egiazarian. 2006. Color image denoising via sparse 3D collaborative filtering with grouping constraint in luminance-chrominance space. Proceedings - International Conference on Image Processing, ICIP 1: 3–6.

[5] Quan Gan, Qipeng Guo, Zheng Zhang, and Kyunghyun Cho. 2015. First Step toward Model-Free, Anonymous Object Tracking with Recurrent Neural Networks. (ed.), Oxford, U.K., Pergamon Press PLC, 1989, Section 3, p.111–120. (ISBN 0–08–036148–X): 1–9.

[6] Ian J Goodfellow, Jean Pouget-abadie, Mehdi Mirza, Bing Xu, and David Warde-farley. 2014. Generative-Adversarial-Nets. Nips: 1–9.

[7] J. W. Grizzle and A. Isidori. 1989. Block noninteracting control with stability via static state feedback. Mathematics of Control, Signals, and Systems 2, 4: 315–341.

[8] Sergey Ioffe and Christian Szegedy. 2015. Batch Normalization: Accelerating Deep Network Training by Reducing Internal Covariate Shift.

[9] Christian Ledig, Lucas Theis, Ferenc Huszár, Jose Caballero, Andrew Cunningham, Alejandro Acosta, Andrew Aitken, Alykhan Tejani, Johannes Totz, Zehan Wang, and Wenzhe Shi. 2017. Photo-realistic single image super-resolution using a generative adversarial network. Proceedings - 30th IEEE Conference on Computer Vision and Pattern Recognition, CVPR 2017 2017–Janua: 105–114.

[10] Bee Lim, Sanghyun Son, Heewon Kim, Seungjun Nah, and Kyoung Mu Lee. 2017. Enhanced Deep Residual Networks for Single Image Super-Resolution. IEEE Computer Society Conference on Computer Vision and Pattern Recognition Workshops 2017–July: 1132–1140.

[11] Kyung Tae Lim, Seung Chan Lee, Yimeng Gao, Kee Pyo Kim, Guangqi Song, Su Yeon An, Kenjiro Adachi, Yu Jin Jang, Jonghun Kim, Kyoung Jin Oh, Tae Hwan Kwak, Seon In Hwang, Jueng Soo You, Kinarm Ko, Seung Hoi Koo, Amar Deep Sharma, Jong Hoon Kim, Lijian Hui, Tobias Cantz, Hans R. Schöler, and Dong Wook Han. 2016. Small Molecules Facilitate Single Factor-Mediated Hepatic Reprogramming. Cell Reports 15,4: 814–829.

[12] Ding Liu, Bihan Wen, Xianming Liu, Zhangyang Wang, and Thomas S. Huang. 2017. When Image Denoising Meets High-Level Vision Tasks: A Deep Learning Approach. 1.

[13] Mehdi S.M. Sajjadi, Bernhard Scholkopf, and Michael Hirsch. 2017. EnhanceNet: Single Image Super-Resolution Through Automated Texture Synthesis. Proceedings of the IEEE International Conference on Computer Vision 2017–Octob: 4501–4510.

[14] Christian Szegedy, Wei Liu, Yangqing Jia, Pierre Sermanet, Scott Reed, Dragomir Anguelov, Dumitru Erhan, Vincent Vanhoucke, and Andrew Rabinovich. 2015. Going deeper with convolutions. Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition 07–12–June: 1–9.

[15] D Wu, H Wang, B Wu, H Ni, S Huang, Y Xiong, P Wang, Q Han, Z Niu, I Tangring, and S M Wang. 2008. Low threshold current density 1.3 μm metamorphic InGaAs/GaAs quantum well laser diodes. Electronics Letters 44, 7: 7–8.

[16] Jun Xu, Lei Zhang, David Zhang, and Xiangchu Feng. 2017. Multi-channel Weighted Nuclear Norm Minimization for Real Color Image Denoising. Proceedings of the IEEE International Conference on Computer Vision 2017–Octob: 1105–1113.

[17] Kai Zhang, Wangmeng Zuo, Yunjin Chen, Deyu Meng, and Lei Zhang. 2017. Beyond a Gaussian denoiser: Residual learning of deep CNN for image denoising. IEEE Transactions on Image Processing 26, 7: 3142–3155.

[18] http://r0k.us/graphics/kodak/