Attention-based U-Net for Image Demoiréing

Tomasz M. Lehmann

Warsaw University of Technology, Warsaw, Poland

tomasz.lehmann.dokt@pw.edu.pl

Abstract. Image demoiréing is a particular example of a picture restoration problem. Moiré is an interference pattern generated by overlaying similar but slightly offset templates.

In this paper, we present a *deep learning* based algorithm to reduce moiré disruptions. The proposed solution contains an explanation of the *cross-sampling* procedure – the training dataset management method which was optimized according to limited computing resources.

Suggested neural network architecture is based on *Attention U-Net* structure. It is an exceptionally effective model which was not proposed before in image demoiréing systems. The greatest improvement of this model in comparison to *U-Net* network is the implementation of *attention gates*. These additional computing operations make the algorithm more focused on target structures.

We also examined three MSE and SSIM based loss functions. The SSIM index is used to predict the perceived quality of digital images and videos. A similar approach was applied in various computer vision areas.

The author's main contributions to the image demoiréing problem contain the use of the novel architecture for this task, innovative two-part loss function, and the untypical use of the *cross-sampling* training procedure.

Key words: image demoiréing, computer vision, attention U-Net, cross-sampling.

1. Introduction

Image demoiréing is a relatively new issue in the field of Computer Vision (CV). Essentially, it is a specific case of picture restoration. The problem of moiré fringes appears when an opaque ruled pattern with transparent gaps is overlaid on another similar pattern. For this kind of interference pattern to emerge, the two patterns must not be completely identical, but one should be slightly geometrically transformed with respect to the other, or should have a slightly dissimilar pitch. Moiré patterns appear e.g. in digital photography and television. In the first example, it occurs when a pattern on an object being photographed interferes with the shape of the light sensors to generate undesirable artifacts. The term originates from the French noun *moiré*, a type of textile, traditionally made of silk (now also cotton or synthetic material) with a rippled appearance. Examples of images from *ultra-high-definition demoiréing dataset* (UHDM) [27] with moiré fringes are depicted in Fig. 1.

Nowadays, taking photos of electronic displays is a common way of transferring image data and it is widely practiced in industries and everyday life. Because of signal interference between the pixel matrix of the display screen and the *Bayer filters* (color filter arrays (CFA) for arranging RGB color filters on a square grid of photosensors)

Fig. 1. Moiré examples from UHDM database [27]. Images with moiré patterns are presented on the left. On the right images without disruption are shown.

in a camera sensor objectionable moiré patterns significantly disrupt captured screen images. In this work, we tackle the problem of deep learning-based image demoiréing capable of improving the quality of such images.

Recently, deep neural network-based solutions achieve great results in many CV-related problems. Various image restoration and image deblurring issues were resolved with the help of these machine learning algorithms. The presented approach is based

on a type of *U-Net* convolutional neural network (CNN) [18]. The network consists of two main paths (a contracting path and an expansive path) which gives it the u-shaped architecture. Each contracting path is followed by blocks of activation functions and max pooling operations. As a result, the spatial information is reduced while the feature information is increased. The second path is intended to reverse this state. It is a particular example of the symmetrical *autoencoder* architecture.

The author's main contributions to the image demoiréing problem are the proposal of an unused before and highly efficient architecture, the presentation of a loss function implementation which is innovative for this area, and the untypical use of the *cross-sampling* training procedure.

1.1. Related Works

Leading architectures dedicated to image deblurring and image restoration are built on straight deep convolutional neural networks (DCNNs) [2, 13], generative adversarial networks (GANs) [11, 23], Transformer-based blocks [21, 25, 28] and U-Net-shaped hierarchical structures [22].

In the context of real-world (nonsynthetic) moiré removal, the number of studies is meaningfully lower. The first real-world committed dataset was proposed in [26] (the dataset is available from [20]) where authors used a multiresolution fully convolutional neural network. In other papers, a dynamic feature encoding module [8] and a novel multiscale bandpass convolutional neural networks were also suggested [29]. The algorithms strictly dedicated to high-resolution images analysis are also available – the multi-stage framework FHDe2Net [4] was applied to 1080p resolution image demoiréing on the FHDMi database and the ESDNet [27] architecture was Performed well trained and performed well on UHDM dataset. The first mentioned framework consists of two branches. The global to local cascades branch removes moiré patterns from the picture while the other part of the structure conserves high-resolution details. The second solution is based on a semantic-aligned scale-aware module to address the scale variation of moiré patterns. Most of the proposed deep learning algorithms have high computational complexity. We focused on developing a lightweight and effective model on lower resolution images ($512 \cdot 512$) to balance computational costs.

To solve the considered problem we propose a special variant of the U-Net architecture [18]. It is a convolutional network originally designed for biomedical image segmentation. The network is fast and it was successfully adopted in different areas of CV, like super-resolution [9, 12], depth estimation [1, 6] and image denoising [5]. In our research we used the *Attention U-Net* [14] – the U-Net-type architecture with *attention gates* applied. These extra operations automatically learn to focus on target structures without additional supervision. The concept comes from Natural Language Processing (NLP) for image captioning [3]. In this paper, we present our training procedure methods and our results achieved on the test datasets.

2. Experimental setup

2.1. Datasets

As there are few papers addressing the image demoiréing problem, there are also not too many datasets for this task. Most of them are generated by moiré promoted software. However, for some time real-world data is also available.

2.1.1. UHDM

The UHDM dataset [27] is a collection of ultra-high-definition images. It contains 5,000 pairs of images in resolution $4032 \cdot 3024$ and $4624 \cdot 3274$. The dataset was collected using various mobile phones which affects resolution and quality diversity. To produce realistic moiré images, authors shoot clean pictures displayed on the screen with a phone camera , and the phone was fixed on a smartphone gimbal, which allows them to conveniently and flexibly adjust the camera view through its control button. UHDM is randomly split into 4,500 pairs for the training procedure and 500 pairs for validation. In our research we used the same distribution of subsets.

2.1.2. TIP2018

Authors of the frequently cited publication [26] created a benchmark of $135\,000$ image pairs available from [20]. Collected images have a wide variety of moiré effects. Each pair contains an image contaminated with a moiré pattern and its corresponding uncontaminated reference image. Image references are enhanced with a black border which is explained by the observation that dark colors are least affected by the moiré effect. Displayed images were captured using a mobile phone. 90% of images are used as the training set and 10% are used for validation and testing. During the validation, we used these data to tune the parameters of a classifier when the test is executed to assess the final performance [17]. We kept these proportions.

2.2. Proposed methods

2.2.1. Attention U-Net

In our research, we proposed widely respected *Attention U-Net* architecture [14]. The algorithm uses self-attention gating modules that can be utilized in CNN-based standard image analysis models for dense label predictions.

Mentioned gates are located in the standard U-Net architecture to highlight salient features that are passed through the skip connections. Attention gates filter the neuron activation during backward and forward passes to down-weighted background regions and up-weighted spatial regions which are more relevant for a given problem.

This kind of neural network architecture has been never successfully used before in image demoiréing or image restoration tasks.

2.2.2. Loss function

In this paper, three loss functions are considered. First, we applied the *mean squared error* (MSE) function which measures the average squared difference between the estimated values and the actual value. Later, we experimented with combinations of MSE loss and *the structural similarity index measure* (SSIM) loss [24]. We defined the second one as the SSIM value according to (1). The SSIM measure is used to assess the perceived quality of digital images and videos. The SSIM value remains between -1 and 1. A value closer to 1 indicates better image quality. A similar approach was used in depth estimation models [1] where authors empirically found and set 0.1 as a reasonable weight for MSE part in the loss formula. We decided to start from the same value which might be understood as a significant approximation. Achieved results were compared also with the ten times bigger MSE-part multiplier. To determine the most accurate value the remarkably wider range should be considered. Because of technical limitations, we decided to investigate just these two cases. In [15] the author compares similarity and distance measures. The author showed and explained the theoretical similarities between SSIM and MSE functions which depend on the same parameters. Nevertheless, the experimental simulations showed a great difference between these two metrics. It can be observed that MSE is quite insensitive to all types of distortions while SSIM responds considerably to even minor blurriness and noise changes. By combining these two methods we can control more output image parameters.

The equations crucial for our computations are presented below.

$$\mathrm{SSIM}(x, y) = \frac{(2\mu_x\mu_y + C_1) + (2\sigma_{xy} + C_2)}{(\mu_x^2 + \mu_y^2 + C_1)(\sigma_x^2 + \sigma_y^2 + C_2)}, \tag{1}$$

where:

  μ_x – the pixel sample mean of x,
  μ_y – the pixel sample mean of y,
  σ_x^2 – the variance of x,
  σ_y^2 – the variance of y,
  σ_{xy} – the variance of x and y,
  C_1, C_2 – two variables to stabilize the division with weak denominator.

$$\mathrm{MSE}(x, y) = \frac{1}{n}\sum_n (x_n - y_n)^2, \tag{2}$$

where:

  n – number of pixels,
  x_n – ground truth pixel,
  y_n – predicted pixel.

$$L_{\text{SSIM}}(x, y) = 1 - \text{SSIM}(x, y) \,. \tag{3}$$

During experiments we applied and analysed three loss functions described by (4-6).

$$L_1(x, y) = \text{MSE}(x, y) \,, \tag{4}$$

$$L_2(x, y) = 0.1 \cdot \text{MSE}(x, y) + L_{\text{SSIM}}(x, y) \,, \tag{5}$$

$$L_3(x, y) = \text{MSE}(x, y) + L_{\text{SSIM}}(x, y) \,. \tag{6}$$

2.2.3. Training procedures

To train our implemented solutions we used ADAM optimizer [10]. It is an algorithm for first-order gradient-based optimization of stochastic objective functions. The method is efficient and requires less memory than fundamental stochastic gradient descent. The training lasted 100 epochs and every single training epoch was followed by validation. During the validation SSIM and *peak signal-to-noise ratio* (PSNR) metrics were monitored. The equation for PSNR is shown below.

$$\text{PSNR}(x, y) = 20 \cdot \log_{10}(I_{\text{max}}) - 10 \cdot 10 \log_{10}(\text{MSE}(x, y)) \tag{7}$$

where:

I_{max} – the maximum possible pixel value of the image. When the pixels are represented using 8 bits per sample, this is 255.

To reduce computational costs we decided to use a *cross-sampling* technique. It is a method used for balancing uneven or extensive datasets. During the training, we used all elements from the UHDM training set and only 15 500 elements from TIP2018 training set. After every epoch, the elements from the second-mentioned set were randomly drawn. For comparison, we also trained one particular model on the whole training dataset.

As a result of time-consuming prevention, the resolution of the whole collected database (UHDM and TIP2018) was reduced to a dimension of $512 \cdot 512$ pixels. Tests and validations were made on the UHDM and the TIP2018 datasets separately. During the examination of our methods, the testing resolution also decreased to $512 \cdot 512$. The final comparison with other algorithms was made on the images with original resolution.

We implemented our proposed network using PyTorch [16] and it was trained on the NVIDIA TITAN V100 GPU with 32 GB memory.

3. Experimental results

3.1. Loss functions comparison

During epochs validating we monitored two main metrics: SSIM and PSNR. These two measuring tools are widely used in image quality assessment. To minimize computational costs we decided to use the *cross-sampling* method in experiments with loss functions.

In Figs. 2a and b we can see how the validation metrics change in the training time for the UHDM validation dataset.

We can observe that by optimizing SSIM value – PSNR value increases slower than without using SSIM-related loss function. It is explained by strong PSNR-MSE relation which is presented in [15]. The two-part loss function minimizes MSE less efficiently which is a necessary compromise if we want to maximize the SSIM measure. It is a newer measurement tool that is designed based on three main factors, i.e., luminance, contrast, and structure to better suit the operation of the human visual system [19].

Validation measures in function of epochs for the TIP2018 validation set are depicted in Figs. 3a and b.

A similar tendency of the SSIM-PSNR dependence can be observed but the proportions are slightly different. It is observed that with decreasing MSE part in the MSE – SSIM combinational function – PSNR increase.

All results for test datasets are located in the Table 1. According these results, the most proficient loss function is $0.1 \cdot \text{MSE} + \text{SSIM}$ which achieves the best results with SSIM measure for both datasets. On the other hand, the PSNR was not so undervalued as using MSE+SSIM minimizing. Choosing the most accurate loss function should depend on task specification but we present the pros and cons of these calculations.
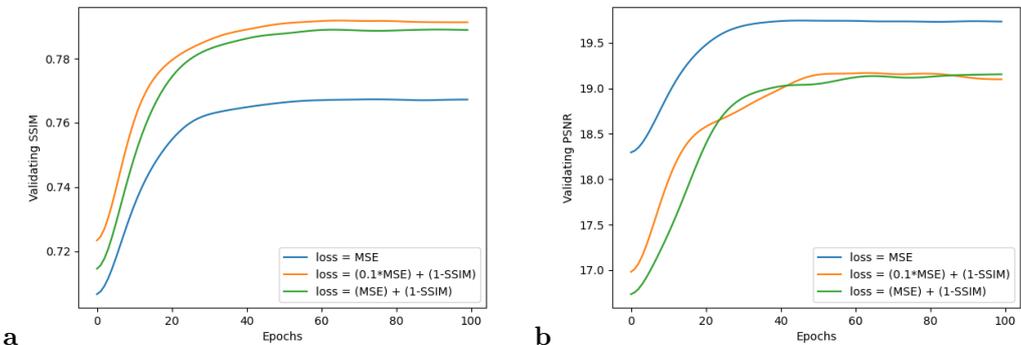
Fig. 2. Validating with two measures in function of epochs for the UHDM validation dataset: (**a**) with SSIM; (**b**) with PSNR. Graph color represents the loss function applied.

a   b
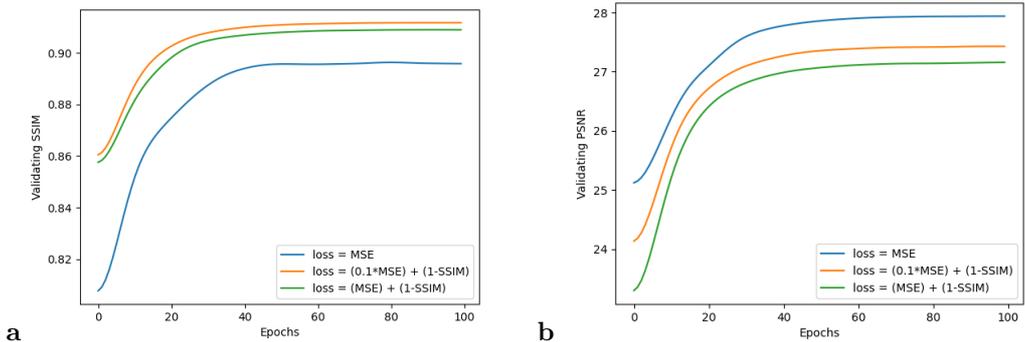
Fig. 3. Validating with two measures in function of epochs for the TIP2018 validation dataset: (**a**) with
SSIM; (**b**) with PSNR. Graph color represents the loss function applied.

The presented SSIM-related loss function includes point-wise differences but also opti-
mizes the process of distortion removal by looking at regions around each point. For the
optimization of the algorithm to be better suited to the operation of the human visual
system, we recommend the proposed cost function.

3.2. Influence of the cross-sampling implementation

We examined the influence of the *cross-sampling* method which reduced the training time
almost eight times. Because of a lack of resources was decided to train the model on
full TIP2018 and UHDM datasets just once. We made decision to use $0.1 \cdot \text{MSE} + \text{SSIM}$
loss function to perform it. Later we compared the achieved measures with the cross-
sampling method.

Tab. 1. Final metrics for different datasets and loss functions.

DATASET	METRIC	LOSS FUNCTION	VALUE
UHDM	SSIM	MSE	0.77
		$0.1 \cdot \text{MSE} + L_{\text{SSIM}}$	0.80
		$\text{MSE} + L_{\text{SSIM}}$	0.79
	PSNR	MSE	19.80
		$0.1 \cdot \text{MSE} + L_{\text{SSIM}}$	19.31
		$\text{MSE} + L_{\text{SSIM}}$	19.23
TIP2018	SSIM	MSE	0.90
		$0.1 \cdot \text{MSE} + L_{\text{SSIM}}$	0.91
		$\text{MSE} + L_{\text{SSIM}}$	0.91
	PSNR	MSE	27.94
		$0.1 \cdot \text{MSE} + L_{\text{SSIM}}$	27.46
		$\text{MSE} + L_{\text{SSIM}}$	27.18
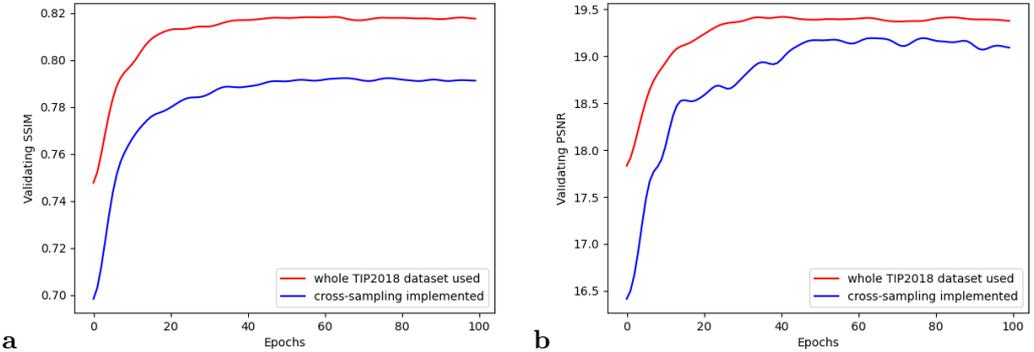
Fig. 4. Validating with two measures in function of epochs for the UHDM validation dataset: (**a**) with SSIM; (**b**) with PSNR. Graph color represents the loss function applied.
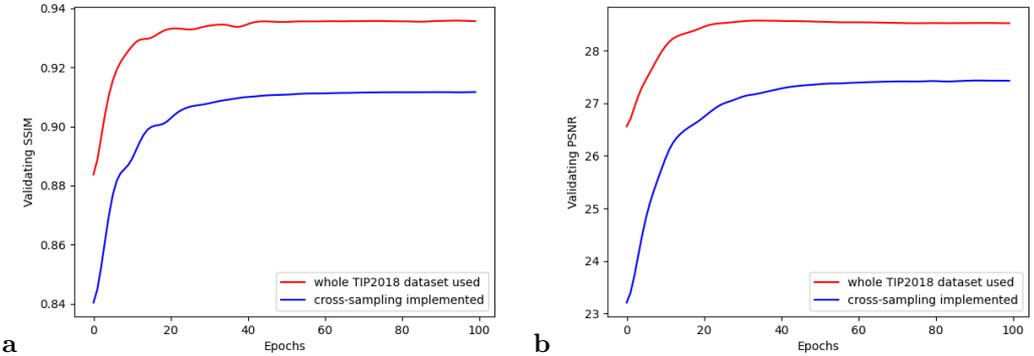
Fig. 5. Validating with two measures in function of epochs for the TIP2018 validation dataset: (**a**) with SSIM; (**b**) with PSNR. Graph color represents the loss function applied.

Charts with measures in the function of epoch iterations for the UHDM dataset are presented in Figs. 4a and b.

It can be noticed that differentials and rate of increase of the SSIM measure are similar for both methods. The correlation between PSNR measures is shaped differently and obtained efficiency gap is narrowed.

The corresponding charts for the TIP2018 dataset are shown in Figs. 5a and b.

Accurate results for both discussed training methods are presented in the Tab. 2.

We provided two versions of our model: *cross-sampled Attention U-net for image demoiréing (cs-AUid)* and *Attention U-net for image demoiréing (AUid)*. The first solution is much less time-consuming and easier to train while the second one achieves better results.

Tab. 2. Comparison of the *cross-sampling* method and full-dataset training metrics.

DATASET	METRIC	CROSS-SAMPLING	VALUE
UHDM	SSIM	YES	0.80
		NO	0.82
	PSNR	YES	19.31
		NO	19.48
TIP2018	SSIM	YES	0.91
		NO	0.94
	PSNR	YES	27.46
		NO	28.58

3.3. Comparison with other algorithms

In Tab. 3 we presented our results in comparison with the best available and documented solutions. Based on the following outcome, we can conclude that our method outperforms most of the other techniques in SSIM metric. We can suspect that it is an advantage of the relatively innovative loss function and efficient *Attention U-Net* algorithm. The results of *cs-AUid* model might be considered as a satisfying replacement that can easily be trained even with second-rate computational resources. It is less time-consuming and based on effective training procedures.

Notwithstanding, we need to remember that results presented in the previous chapter were obtained for a resolution $512 \cdot 512$. For scientific cases, we tested our trained model also on the benchmark images with a much higher number of image pixels.

In Figs. 6 and 7 the examples of the model outcome are shown. The images were made with our *cs-AUid* approach. We can observe that moiré is barely visible.

4. Conclusions

In this work, we proposed a convolutional neural network based *Attention U-Net* for image demoiréing. We presented two training procedures and we analyzed three loss functions. Our innovation lies in efficient datasets management and proper architecture choice. Our solution might be interpreted as an efficient alternative for more complex and time-consuming models.

Tab. 3. Comparison with other *state-of-the-art* algorithms. *cs-AUid* corresponds to cross-sampling method while *AUid* means large dataset used in the training. Both models were trained with $0.1 \cdot \mathrm{MSE} + L_{\mathrm{SSIM}}$ loss function. Compared solutions: TIP2018 [20], MopNet [7], MBCNN [29], FHDe2Net [4], ESDNet-L [27].

DATASET	MEASURE	INPUT	TIP2018	MopNet	MBCNN	FHDe2Net	ESDNet-L	cs-AUid	AUid
UHDM	SSIM	0.51	0.76	0.76	0.79	0.75	0.80	0.76	0.78
	PSNR	17.12	19.91	19.49	21.41	20.34	22.42	18.20	18.25
TIP2018	SSIM	0.74	0.87	0.89	0.89	0.90	0.92	0.88	0.90
	PSNR	20.30	26.77	27.75	30.03	27.78	30.11	26.03	26.82
Number of parameters (millions)			1.4	58.6	15.2	13.6	10.6	6.4	6.4

Fig. 6. Examples of images from UHDM dataset. From left: moiréd image, original image, demoiréd image (cs-AUid output).

Fig. 7. Examples of images from TIP2018 dataset. Image references are enhanced with a black border which is explained by the observation that dark colors are least affected by the moiré effect. From left: moiréd image, original image, demoiréd image (cs-AUid output).

The next part of the research should be adapting models for higher resolutions. Presented models were trained and validated on the $512 \cdot 512$ pixel maps. Basing on the results collected in Tab. 3 we can notice that with the increase in image resolution the quality of the images indicated with the respective quality measures significantly decreased. We also could study the loss function in the wider range of SSIM-MSE relation in the measure L_2 according to (5) to optimize the model parameters. In the next steps, the PSNR also should be maximized. In [15] its author presented a new similarity measure denoted there as CMSC which can be examined as a loss function in our further research.

References

[1] I. Alhashim and P. Wonka. High quality monocular depth estimation via transfer learning. *arXiv*, 2018. doi:10.48550/arXiv.1812.11941.

[2] Y. Zhang an Y. Tian, Y. Kong, B. Zhong, and Y. Fu. Residual dense network for image restoration. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 43(7):2480–2495, 2020. doi:10.1109/tpami.2020.2968521.

[3] P. Anderson, X. He, C. Buehler, et al. Bottom-up and top-down attention for image captioning and visual question answering. In *Proc. 2018 IEEE/CVF Conf. Computer Vision and Pattern Recognition (CVPR)*, pages 6077–6086, Salt Lake City, Utah, 18-23 Jun 2018. IEEE. doi:10.1109/CVPR.2018.00636.

[4] X. Cheng, Z. Fu, and J. Yang. Multi-scale dynamic feature encoding network for image demoiréing. In *Proc. 2019 IEEE/CVF Int. Conf. Computer Vision Workshops (ICCVW)*, pages 3486–3493, Seoul, Korea, 27 Oct – 2 Nov 2019. IEEE. doi:10.1109/ICCVW.2019.00432.

[5] J. Gurrola-Ramos, O. Dalmau, and T. E. Alarcon. A residual dense U-Net neural network for image denoising. *IEEE Access*, 9:31742–31754, 2021. doi:10.1109/ACCESS.2021.3061062.

[6] C. Wang K. Batmanghelich D. Tao H. Fu, M. Gong. Deep ordinal regression network for monocular depth estimation. *2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2018. doi:10.1109/cvpr.2018.00214.

[7] B. He, C. Wang, B. Shi, and L. Y. Duan. Mop moiré patterns using MopNet. In *Proc. 2019 IEEE/CVF Int. Conf. Computer Vision (ICCV)*, pages 2424–2432, Seoul, Korea, 27 Oct – 2 Nov 2019. IEEE. doi:10.1109/ICCV.2019.00251.

[8] B. He, C. Wang, B. Shi, and L.-Y. Duan. FHDe^{2}Net: Full high definition demoireing network. In A. Vedaldi, H. Bischof, T. Brox, and J.-M. Frahm, editors, *Computer Vision – Proc. ECCV 2020*, volume 12367 of *Lecture Notes in Computer Science*, pages 713–729, Glasgow, UK, 23–28 Aug 2020. Springer International Publishing. doi:10.1007/978-3-030-58542-6_43.

[9] X. Hu, M. A. Naiel, A. Wong, et al. RUNet: A robust UNet architecture for image super-resolution. In *Proc. 2019 IEEE/CVF Conf. Computer Vision and Pattern Recognition Workshops (CVPRW)*, pages 505–507, Long Beach, California, 16-20 Jun 2019. IEEE. doi:10.1109/CVPRW.2019.00073.

[10] D. P. Kingma and J. Ba. Adam: A method for stochastic optimization. In Y. Bengio and Y. LeCun, editors, *Proc. 3rd Int. Conf. Learning Representations, ICLR 2015*, San Diego, CA, 7-9 May 2015. Accessible in arXiv. doi:10.48550/arXiv.1412.6980.

[11] O. Kupyn, T. Martyniuk, J. Wu, and Z. Wang. DeblurGAN-v2: Deblurring (orders-of-magnitude)

faster and better. In *Proc. 2019 IEEE/CVF Int. Conf. Computer Vision (ICCV)*, pages 8877–8886, Seoul, Korea, 27 Oct – 2 Nov 2019. IEEE. doi:10.1109/ICCV.2019.00897.

[12] Z. Lu and Y. Chen. Single image super resolution based on a modified U-net with mixed gradient loss. *Signal, Image and Video Processing*, 16(5):1143–1151, 2022. doi:10.1007/s11760-021-02063-5.

[13] S. Nah, T. H. Kim, and K. M. Lee. Deep multi-scale convolutional neural network for dynamic scene deblurring. In *Proc. 2017 IEEE Conf. Computer Vision and Pattern Recognition (CVPR)*, pages 257–265, Honolulu, Hawaii, 21-26 Jul 2017. IEEE. doi:10.1109/CVPR.2017.35.

[14] O. Oktay, J. Schlemper, L. L. Folgoc, et al. Attention U-Net: Learning where to look for the pancreas. *arXiv*, 2018. doi:10.48550/arXiv.1804.03999.

[15] G. Palubinskas. Image similarity/distance measures: what is really behind MSE and SSIM? *International Journal of Image and Data Fusion*, 8(1):32–53, 2016. doi:10.1080/19479832.2016.1273259.

[16] A. Paszke, S. Gross, F. Massa, et al. PyTorch: An imperative style, high-performance deep learning library. In *Advances in Neural Information Processing Systems 32 – Proc. 33rd Conf. Neural Information Processing Systems (NeurIPS 2019)*, volume 11, pages 8024–8035, Vancouver, Canada, 8-14 Dec 2019. Accessible in arXiv. doi:10.48550/arXiv.1912.01703.

[17] B. D. Ripley. *Pattern Recognition and Neural Networks*. Cambridge University Press, 1996. doi:10.1017/CBO9780511812651.

[18] O. Ronneberger, P. Fischer, and T. Brox. U-Net: Convolutional networks for biomedical image segmentation. In N. Navab, J. Hornegger, W. M. Wells, et al., editors, *Medical Image Computing and Computer-Assisted Intervention – Proc. MICCAI 2015*, volume 9351 of *Lecture Notes in Computer Science*, pages 234–241, Munich, Germany, 5-9 Oct 2015. Springer International Publishing. doi:10.1007/978-3-319-24574-4_28.

[19] De R. I. M. Setiadi. PSNR vs SSIM: imperceptibility quality assessment for image steganography. *Multimedia Tools and Applications*, 80(6):8423–8444, 2021. doi:10.1007/s11042-020-10035-z.

[20] Y. Sun, Y. Yu, and W. Wang. Moiré photo restoration using multiresolution convolutional neural networks, 8 May 2018. https://paperswithcode.com/paper/moire-photo-restoration-using-multiresolution. [Accessed: May, 2022].

[21] F.-J Tsai, Y.-T. Peng, Y.-Y. Lin, et al. Stripformer: Strip transformer for fast image deblurring. In S. Avidan, G. Brostow, M. Cissé, et al., editors, *Computer Vision – Proc. ECCV 2022*, volume 13679, Part XIX of *Lecture Notes in Computer Science*, pages 146–162, Tel Aviv, Israel, 23-27 Oct 2022. Springer Nature Switzerland. doi:10.1007/978-3-031-19800-7_9.

[22] Z. Tu, H. Talebi, H. Zhang, et al. MAXIM: Multi-Axis MLP for image processing. In *Proc. 2022 IEEE/CVF Conf. Computer Vision and Pattern Recognition (CVPR)*, pages 5759–5770, New Orleans, Louisiana, 18-24 Jun 2022. IEEE. doi:10.1109/CVPR52688.2022.00568.

[23] X. Wang, Y. Li, H. Zhang, and Y. Shan. Towards real-world blind face restoration with generative facial prior. In *Proc. 2021 IEEE/CVF Conf. Computer Vision and Pattern Recognition (CVPR)*, pages 9164–9174, Virtual conference, 20-25 Jun 2021. IEEE. doi:10.1109/CVPR46437.2021.00905.

[24] Z. Wang, A. C. Bovik, H. R. Sheikh, and E. P. Simoncelli. Image quality assessment: from error visibility to structural similarity. *IEEE Trans. Image Processing*, 13(4):600–612, Apr 2004. doi:10.1109/TIP.2003.819861.

[25] Z. Wang, X. Cun, J. Bao, et al. Uformer: A general U-shaped transformer for image restoration. In *Proc. 2022 IEEE/CVF Conf. Computer Vision and Pattern Recognition (CVPR)*, pages 17662–17672, New Orleans, Louisiana, 18-24 Jun 2022. IEEE. doi:10.1109/CVPR52688.2022.01716.

[26] W. Wang Y. Sun, Y. Yu. Moiré photo restoration using multiresolution convolutional neural networks. *IEEE Trans. Image Processing*, 27(8):4160–4172, 2018. doi:10.1109/tip.2018.2834737.

[27] X. Yu, P. Dai, W. Li, L. Ma, et al. Towards efficient and scale-robust ultra-high-definition image demoiréing. In S. Avidan, G. Brostow, M. Cissé, et al., editors, *Computer Vision – Proc. ECCV 2022*, volume 13678, Part XVIII of *Lecture Notes in Computer Science*, pages 646–662, Tel Aviv, Israel, 23-27 Oct 2022. Springer Nature Switzerland. doi:10.1007/978-3-031-19797-0_37.

[28] S. W. Zamir, A. Arora, S. Khan, et al. Restormer: Efficient transformer for high-resolution image restoration. In *Proc. 2022 IEEE/CVF Conf. Computer Vision and Pattern Recognition (CVPR)*, pages 5718–5729, New Orleans, Louisiana, 18-24 Jun 2022. IEEE. doi:10.1109/CVPR52688.2022.00564.

[29] B. Zheng, S. Yuan, G. Slabaugh, and A. Leonardis. Image demoireing with learnable bandpass filters. In *Proc. 2020 IEEE/CVF Conf. Computer Vision and Pattern Recognition (CVPR)*, pages 3633–3642, Virtual conference, 14-19 Jun 2020. IEEE. doi:10.1109/CVPR42600.2020.00369.