RWN: ROBUST WATERMARKING NETWORK FOR IMAGE CROPPING LOCALIZATION

Qichao Ying⋆, Xiaoxiao Hu⋆, Xiangyu Zhang⋆, Zhenxing Qian⋆, Sheng Li and Xinpeng Zhang⋆

* Fudan University

ABSTRACT

Image cropping can be maliciously used to manipulate the layout of an image and alter the underlying meaning. Previous image cropping detection schemes only predict whether an image has been cropped, ignoring which part of the image is cropped. This paper presents a novel robust watermarking network for image cropping localization. We train an anti-cropping processor (ACP) that embeds a watermark into a target image. The visually indistinguishable protected image is then posted on the social network instead of the original image. At the recipient’s side, ACP extracts the watermark from the attacked image, and we conduct feature matching on the original and extracted watermark to locate the position of the cropping. We further extend our scheme to detect tampering attacks on the attacked image, and a simple yet efficient method (JPEG-Mixup) is proposed that noticeably improves the generalization of JPEG robustness. We demonstrate that our scheme is the first to provide high-accuracy and robust image cropping localization.

Index Terms— image cropping localization, image tamper detection, robustness, image forensics, watermarking

1. INTRODUCTION

Image manipulation leads to severe security threats, where misleading photojournalism, copyright violation, or fabricating stories can be a means for some writers or politicians to potentially influence public opinion. Researchers have developed a number of schemes to detect various kinds of digital attacks, e.g., tampering [1, 2], DeepFake [3] and cropping [4, 5, 6]. Among them, cropping is an extremely cheap and effective way to manipulate the layout of an image and alter the underlying meaning.

Existing cropping detection algorithms [4, 5] mainly focus on predicting whether an image is cropped. They are represented by detecting the exposing evidence of asymmetrical image cropping, e.g., the shift of the image center [4], and the inconsistency of JPEG blocking artifacts [5, 7]. For example, Fanfani et al. [4] exploits the camera principal point insensitive to image processing operations. Yerushalmi et al. [5] detects whether there are vanishing points and lines on structured image content. In [7], the block artifact grids (BAGs) are extracted blindly with a new extraction algorithm, and then abnormal BAGs can be detected with a marking procedure as a trail of cropping detection. However, different cropping behaviors are with varied intentions, and a simple binary classification cannot distinguish benign cropping behaviors from those malicious attacks, e.g., removal of visible watermarks or critical objects. Van et al. [6] proposes an image cropping localization scheme to study how image cropping causes chromatic and vignetting alteration. But the tiny traces only exist in high-definition images and can be easily destroyed by image processing. Experiments show that the scheme cannot be applied to highly-compressed or low-quality images.

Watermarking [8, 9, 10] aims at hiding information imperceptibly into the host data for covert purposes. The technology focuses on robustness against possible digital attacks, e.g., image compression and noise addition. Watermarking has been widely used in copyright protection and content authentication of images in multimedia. Recently, many novel watermarking schemes for image protection are proposed [11, 12, 13]. The uprising of deep networks has given birth to a series of novel watermarking methods with enhanced robustness [14, 15, 16]. Shin et al. [14] is the first in including a differentiable approximation to JPEG in the watermarking model. Later, Zhu et al. [15] proposes a powerful and comprehensive network that is robust against a variety kinds of attacks. Subsequent watermarking works [16] design more sophisticated and accurate algorithms to simulate image post-processing attacks, especially the JPEG compression. For example, in our previous work [13], we use robust watermarking to not only detect image tamper but also conduct image self-recovery.

This paper explores the potential of robust watermarking on image cropping localization. We propose a robust watermarking network (RWN) for image cropping localization by protecting the original image through watermarking. If the protected image is cropped, RWN is expected to locate the cropped region at the recipient’s side. We train an anti-cropping processor (ACP) based on an invertible neural network [17] to embed a watermark into an original image. The watermark is shared with the recipient. ACP produces a visually indistinguishable protected image, which is then posted on the social network instead of the unprotected version. On receiving the attacked version of the protected image, we use ACP again to extract the watermark and conduct an efficient and classical feature matching algorithms (SURF [18]) between the original and extracted watermark to determine where the cropping was positioned in the original image plane. Considering that image tampering is also very common and might destroy the embedded watermark, we further use the tamper detector to predict the tamper mask, and the features within the predicted tamper mask are discarded to prevent mismatching. Besides, we explore a simple yet efficient method (JPEG-Mixup) to improve the generalization of JPEG robustness. We test our scheme by introducing man-made hybrid attacks. The results demonstrate that our scheme can accurately locate the cropped region. We also show the effectiveness of tamper detection by comparison with some state-of-the-art image forgery detection schemes [1, 19, 2].

The highlights of this paper are three-folded. First, we present the first high-accuracy robust image cropping localization scheme. We innovatively use normalizing flows to build an invertible function for image forensic problems. Second, with the embedded watermark, RWN can also conduct high-accuracy tamper detection, which is comparable with the state-of-the-art works. Third, we explore a simple yet efficient method (JPEG-Mixup) for improved JPEG simulation. Experiments show that JPEG-Mixup can remarkably improve the generalization of RWN on robustness against JPEG compression.

This work is supported by National Natural Science Foundation of China under Grant U20B2051, 62072114, U20A20178, U1936214. Qichao Ying and Xiaoxiao Hu contribute equally. Corresponding author: Zhenxing Qian. Email: {qcying20, xhxu21, xyzhang20, zxqian, lisheng, zhangxinpeng}@fudan.edu.cn. We sincerely thank the anonymous reviewers for their constructive and insightful suggestions on improving this paper.
2. METHOD

2.1. Pipeline Design

The proposed RWN consists of five stages, namely, watermark embedding, image redistribution, tamper detection, watermark extraction and cropping localization. We embed a watermark W into an original image I. The protected image I_M is generated and we upload it onto the social cloud instead of the unprotected original image. The attacker generates the attacked image I_A by freely adding three kinds of attacks (benign attacks, cropping, tampering) on I_M.

On the recipient’s side, the tamper detector T predicts the tamper mask on the attacked image to see which parts of the image are tampered, and we also extract the watermark as W from the received image. Afterwards, we rectify the extracted watermark by W = W · (1 − M) to discard the tampered contents. Finally, with the original watermark W as reference, we use a feature matching algorithm to locate the position of the cropping in the original image plane. An image is determined as cropped if there is valid matching result.

Modeling. We regard the embedding and extraction as the inverse problem, even if the protected image is cropped and attacked. The formulation of the inverse problem is:

\[
(I_M, R) = F(I, W)
\]

\[
\hat{I}, \hat{W} = P^{-1}(A(I_M), \hat{R}),
\]

subject to \(E(\hat{I}) = I_G, E(\hat{W}) = W_G\) and \(E(G) = \hat{G}\). Here, R and \(\hat{R}\) are the additional output and input of the network to keep the channel numbers as four. \(A\) denotes the function of image post-processing that attacks the hidden watermark. We let \(R = 0.5 \cdot \hat{I}\) where \(\hat{I}\) is a matrix full of one. The ground-truth cropped watermark \(W_G\) and the ground-truth cropped original image \(I_G\) can be generated by sharing the cropping mask used by \(A\). Besides the invertibility, we regulate that \(E(I_M) = \hat{I}\) for the imperceptibility of the embedding.

Fig. 1 shows the pipeline of our scheme. We employ an Anti-Cropping Processor (ACP) \(P\) that jointly learns the paired function of both watermark embedding and extraction. We implement \(A\) using a differentiable attack layer that simulates the attacker’s behavior. The rest of the components include a tamper detector \(T\), a discriminator \(D\) and a feature matcher \(M\), which is the SURF algorithm.

2.2. Network Implementation

Considering the efficiency of the invertible U-Net proposed in [20], we build our ACP on top of this architecture. The network consists of six invertible blocks each of which contains a Haar wavelet transformation and a double-side affine coupling. The network ends with a conditional split layer. The number of the input and output channels is four. The sizes of \(R\) and \(\hat{R}\) are of the same as \(W\). \(R\) and \(\hat{R}\) are not required to be the same.

In the attacking layer \(A\), we first convert the marked images into the corresponding 8-bit RGB-formatted image. We adopt the Straight-Through Estimator [21] on calculating the gradients. Then, we build differentiable methods to simulate the hybrid cropping attacks, which include benign attacks, cropping and tampering attack. The benign attacks are represented by JPEG compression, scaling, etc., which do not alter the semantic meaning of the image. We take the implementation from [15] except that we build our own JPEG simulator. For cropping, we randomly crop a portion of the protected image I_M. For tampering attack, we first randomly select random areas using a binary matrix M inside I_M, and generate the tampered image by

\[
I_A = I_M \cdot M + I_M \cdot (1 - M)\hat{R},
\]

where \(I_M\) refers to the source of the tamper.

Although there are already many schemes which include a carefully-designed JPEG simulator, e.g., JPEG-SS [16], JPEG-Mask [15], MBRS [22], the real-world JPEG robustness of these schemes is still limited. We believe it mainly attribute to that the networks are over-fitted to a fixed compression mode. In this paper, we propose to apply the Mix-up strategy [23] over these implementations so that the generated results are more flexible. For \(i \in [1, k]\), we randomly sample the quality factor \(QF_i\) that follows \(QF_i \in [10, 100]\), together with a previous differentiable implementation of JPEG simulator \(S_i\), where \(S_i \in \{JPEG-SS, JPEG-Mask, MBRS\}\). We generate the corresponding pseudo-JPEG images \(I_{jpg}^i, I_{jpg}^2, ..., I_{jpg}^k\). The output of JPEG-Mixup is produced by mixing the pseudo-JPEG images with arbitrary contributing rates \(\epsilon_i\).

\[
I_{jpg} = \sum_{i \in [1,k]} \epsilon_i \cdot S_i(I_M, QF_i).
\]

where \(\sum \epsilon_i = 1\). The proposed mix-up based JPEG simulation (JPEG-Mixup) is fairly simple yet effective. Compared to previous schemes, the interpolation technique can generate more diversified pseudo-JPEG images for RWN. In the experiments, we show the performance gain in JPEG robustness using JPEG-Mixup in comparison with the other simulators.

Finally, we build the tamper detector T upon U-Net [24], a traditional image segmentation network, and implement the discriminator D using Patch-GAN [25].

2.3. Objective Loss Function and Training Details

Loss Function. For ACP, the first part of the loss is reconstruction loss \(L_{rec}\) for \(E(W) = W_G, E(I) = I_G\) and \(E(I_M) = I\). \(E(\cdot)\) is the expectation operator.

\[
L_{rec} = \mathcal{F}(I, I_M) + \mathcal{F}(W_G, \hat{W}) + \mathcal{F}(I_G, \hat{I})
\]

where \(\mathcal{F}\) is the \(\ell_1\) distance function. The second part is the nullification loss for \(E(R) = \hat{R}\).

\[
L_{van} = \mathcal{F}(R, \hat{R}).
\]
The third part is the adversarial loss $L_{adv}$ that aim to fool the adversarial networks to make wrong predictions on whether an image is an original image $P$ or a generated image $\hat{P}$. We accept the least squared adversarial loss proposed in [26], which is defined as:

$$L_{adv}(\hat{P}) = \|1 - D_{\theta}(\hat{P})\|^2_2.$$  

(7)

In sum, the total loss for ACP is

$$L_T = L_{adv} + \alpha \cdot L_{div} + \beta \cdot L_{un},$$  

(8)

where $\alpha$ and $\beta$ are hyper-parameters. For the tamper detector, we minimize the binary cross entropy (BCE) loss between the predicted tamper mask $\hat{M}$ and the ground-truth mask $M$.

$$L_T = -[M \log \hat{M} + (1 - M) \log (1 - \hat{M})].$$  

(9)

### 3. EXPERIMENTAL RESULTS

We resize the images to the size of $256 \times 256$. The hyper-parameters are set as $\alpha = 1, \beta = 8$. We set $k = 3$, which is the generated pseudo-JPEG images used in JPEG-Mixup. The batch size is set as 16, and we use Batch Normalization (BN), Adam optimizer [27] and the learning rate is $1 \times 10^{-4}$. We train the scheme on the COCO training/test set [28] with automatically generated attacks. The dataset contains 117266 training samples and 40670 testing samples. The scheme is tested with human-participated attacks, where we have invited some volunteers to crop the provided marked images and perform further image post-processing attack on them. The cropping rate is roughly $\delta \in [0.5, 1)$. During testing, we use real JPEG coder to compress the protected images.

#### 3.1. Real-World Performance on Cropping Localization

**Quality of the Protected Images.** In Fig. 2, we randomly sample different pairs of images as the original and watermark. From the figures, generally, little detail of the watermark can be found. Though the magnified difference is visible, it does not matter if the difference is visible, only that the marked image is perceptually close to the original image. We have conducted more embedding experiments over 1000 images from the test set, and the average PSNR between the protected images and the original images is 36.23dB, and the average SSIM [29] is 0.983.

#### 3.2. Robustness

We test the robustness of our scheme by conducting image post-processing attack on the marked image. In the last two groups, tampering attack is introduced. In Table 1, we can observe that our scheme provides high-accuracy cropping localization despite larger cropping rate and attack.

#### 3.3. Tamper Detection

Table 1. Average performance of cropping localization under different cropping rate and attack.

| Rate | Index | NoAttack | JPEG | Scaling | MedianBlur | AWGN |
|------|-------|----------|------|---------|------------|------|
| 90%  | IoU   | 0.919    | 0.895| 0.803  | 0.838      | 0.784|
|      | SSIM  | 0.949    | 0.926| 0.957  | 0.953      | 0.832|
| 70%  | IoU   | 0.858    | 0.813| 0.878  | 0.915      | 0.763|
|      | SSIM  | 0.942    | 0.878| 0.940  | 0.940      | 0.804|
| 50%  | IoU   | 0.821    | 0.706| 0.603  | 0.540      | 0.513|
|      | SSIM  | 0.914    | 0.7223| 0.942  | 0.840      | 0.788|

Table 2. F1 score comparison for tamper detection among our scheme and the state-of-the-art methods.

| Method     | NoAttack | JPEG | Blurring | Scaling | AWGN |
|------------|----------|------|----------|---------|------|
| Proposed   | 0.773    | 0.736| 0.695    | 0.745   | 0.573|
| Mantra-Net [1] | 0.566    | 0.480| 0.557    | 0.540   | 0.347|
| MVSS-Net [2] | 0.545    | 0.364| 0.399    | 0.485   | 0.323|
| CAT-Net [19] | 0.467    | 0.433| 0.419    | 0.428   | 0.365|

Accuracy. Fig. 2 further shows the results of watermark extraction, feature matching and cropping localization. We see that the cropping mask is accurately predicted. As a result, even without the prior knowledge of the original image, we know the relative position of the attacked image in the original image plane. We have conducted more experiments over 1000 images under different cropping rate and different kinds of benign attacks. Here for space limit, we take $QF=70$ as an example. The average performances are reported in Table 1. Higher IoU (Intersection over Union) and SSIM [29] indicates more accurate localization result. The average IoUs are above 0.8 where the protected images do not undergo further attacks except cropping. The scheme is proven to be agnostic to the cropping rate in that the performance does not degrade significantly with larger cropping rate.

Robustness. We test the robustness of our scheme by conducting image post-processing attack on the marked image. In the last two groups, tampering attack is introduced. In Table 1, we can observe that our scheme provides high-accuracy cropping localization despite

![Fig. 2. Results of watermark embedding, extraction and cropping localization. The attacks involved in each group are different, namely, (a) cropping alone, (b) cropping & scaling and (c) cropping & JPEG & tamper.](image-url)
the presence of the attacks along with cropping. The results promote
the real-world application of RWN. Thanks to the robustness of fea-
ture matching of SURF, the experiments show that in most cases our
scheme do not require a precise watermark extraction.

**Comparison.** Van et al. [6] is fragile where the targeted image can-
not be compressed. If the original images are in JPEG formats, the
required traces are much likely destroyed already. The F1 score of
[6] are 0.575 on natural images where the resolution and image qual-
ity are randomized. In contrast, RWN does not have any restriction
on the original image. After watermarking, we can robustly locate
the cropped area with a much higher F1 score, i.e., 0.867.

### 3.2. Accuracy of Tamper Detection

**Baseline Introduction.** Great efforts have been made on combating
daily image forgeries. Kown et al. [19] proposes to model quantized
DCT coefficient distribution to trace compression artifacts. Mantra-
Net [1] uses fully convolutional networks for feature extraction and
further uses long short-term memory (LSTM) cells for pixel-wise
anomaly detection. In MVSS-Net [2], a system with multi-view fea-
ture learning and multi-scale supervision is developed to jointly ex-
plot the noise view and the boundary artifact to learn manipulation
detection features.

**Comparison.** Fig. 3 shows the tamper detection results on two
test images. Note that for fair comparison, we add the same attacks
on the original images for [1, 19, 2]. We observe that although the
images are tampered by a variety of random attacks, we succeed in
localizing the tampered areas. In Table 2, we provide the average
results over 1000 images. The F1 score of RWN is 0.773 on uncom-
pressed images and 0.736 on JPEG attacked images ($QF = 80$).
The performance of Mantra-Net on JPEG images is much worse
than that on plain-text images. We believe the reason is that less
statistical trace is preserved in the compressed version. In contrast,
the embedded watermark signal serves as the alternative trace for
tamper detection, which is designed to resist benign attacks. The
performance on JPEG images does not drop too much. Therefore,
robust watermarking can successfully aid tamper detection by hid-
ing crafted traces similar to [12, 11].

### 3.3. Ablation Study

We discuss the influences of the network design and the training
strategies in RWN. In each ablation test, we fine-tune the network
till the accuracy of cropping localization is close to the baseline. We
also use fixed attacks during testing.

**Influence of INN architecture.** In Test 1, we train two individual
networks to implement the hiding network and revealing network.
The normalizing-flow-based ACP is replaced in order to measure
the effectiveness brought by invertible problem formulation.

**Influence of JPEG Simulator.** In Test 2, we implement the JPEG
attack with that proposed in [22, 15, 16]. For fair comparison, we
train the implementations together with the baseline under the same
losses, batch size, QF and cropping rate.

**Influence of Discriminator.** In Test 3, we do not use the discrimi-
ator $D$ to monitor the quality of the marked images.

Fig. 4 shows the detailed comparison results. First, the baseline
results outperform the encoder-decoder network design. Second,
while MBRS [22] can provide decent robustness, the extraction
performance is even better using our JPEG simulator. Specifically,
the average SSIM between $W$ and $W_{G}$ using MBRS [22] is 0.797
compared to 0.878 reported in Table 1. The Mix-Up strategy pre-
vents the networks from being over-fitted to any single JPEG sim-
ulator, which helps the scheme significantly improve its real-world
robustness. Third, without the discriminator, the extraction result is
also worse than the baseline.

4. CONCLUSION

This paper presents a novel robust watermarking network for image
cropping rate localization. We use an invertible pipeline for water-
mark embedding and extraction, where the watermark is shared with
the recipient. On receiving the attacked version of the protected im-
age, we use ACP again to extract the watermark and conduct SURF
feature matching algorithms to determine where the cropping rate is
positioned in the original image plane. We also extend our scheme
to detect tampering attacks on the attacked image, and JPEG-Mixup
is proposed that noticeably improves the generalization of JPEG ro-
 bustness. Experiments verify that RWN is effective in real-world
applications and comparable to state-of-the-art methods.
5. REFERENCES

[1] Yue Wu, Wael AbdAlmageed, and Premkumar Natarajan, “Mantra-net: Manipulation tracing network for detection and localization of image forgeries with anomalous features,” in Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2019, pp. 9543–9552.

[2] Chengbo Dong, Xinru Chen, Ruohan Hu, Juann Cao, and Xirong Li, “Mvss-net: Multi-view multi-scale supervised networks for image manipulation detection,” arXiv preprint arXiv:2112.08935, 2021.

[3] Lingzhi Li, Jianmin Bao, Ting Zhang, Hao Yang, Dong Chen, Fang Wen, and Baining Guo, “Face x-ray for more general face forgery detection,” in Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2020, pp. 5001–5010.

[4] Marco Fanfani, Massimo Iuliani, Fabio Bellavia, Carlo Colombo, and Alessandro Piva, “A vision-based fully automated approach to robust image cropping detection,” Signal Processing: Image Communication, vol. 80, pp. 115629, 2020.

[5] Ido Yerushalmy and Hagit Hel-Or, “Digital image forgery detection based on lens and sensor aberration,” International journal of computer vision, vol. 92, no. 1, pp. 71–91, 2011.

[6] Basile Van Hoock and Carl Vondrick, “Dissecting image crops,” arXiv preprint arXiv:2011.11831, 2020.

[7] Weihai Li, Yuan Yuan, and Nenghai Yu, “Passive detection of doctored jpeg image via block artifact grid extraction,” Signal Processing, vol. 89, no. 9, pp. 1821–1829, 2009.

[8] Qichao Ying, Jingzhi Lin, Zhenxing Qian, Haisheng Xu, and Xinpeng Zhang, “Robust digital watermarking for color images in combined dft and dt-cwt domains,” Mathematical Biosciences and Engineering, vol. 16, no. 5, pp. 4788–4801, 2019.

[9] Ying Huang, Baoning Niu, Hu Guan, and Shuwu Zhang, “Enhancing image watermarking with adaptive embedding parameter and psnr guarantee,” IEEE Transactions on Multimedia, vol. 21, no. 10, pp. 2447–2460, 2019.

[10] Ning Liu, Palak Amin, and KP Subbalakshmi, “Security and robustness enhancement for image data hiding,” IEEE Transactions on Multimedia, vol. 9, no. 3, pp. 466–474, 2007.

[11] David Khachaturov, Ilia Shumailov, Yiren Zhao, Nicolas Papernot, and Ross Anderson, “Markpainting: Adversarial machine learning meets inpainting,” arXiv preprint arXiv:2106.00660, 2021.

[12] Minghao Yin, Yongbing Zhang, Xi Li, and Shiqi Wang, “When deep fool meets deep prior: Adversarial attack on super-resolution network,” in Proceedings of the 26th ACM international conference on Multimedia, 2018, pp. 1930–1938.

[13] Qichao Ying, Zhenxing Qian, Hang Zhou, Haisheng Xu, Xinpeng Zhang, and Siyi Li, “From image to image: Immunized image generation,” in Proceedings of the 29th ACM international conference on Multimedia, 2021, pp. 1–9.

[14] Richard Shin and Dawn Song, “Jpeg-resistant adversarial images,” in NIPS 2017 Workshop on Machine Learning and Computer Security, 2017, vol. 1.

[15] Jiren Zhu, Russell Kaplan, Justin Johnson, and Li Fei-Fei, “Hidden: Hiding data with deep networks,” in Proceedings of the European conference on computer vision (ECCV), 2018, pp. 657–672.

[16] Kunlin Liu, Dongdong Chen, Jing Liao, Weiming Zhang, Hang Zhou, Jie Zhang, Wenbo Zhou, and Nenghai Yu, “Jpeg robust invertible grayscale,” IEEE Transactions on Visualization and Computer Graphics, 2021.

[17] Diederik P Kingma and Prafulla Dharwal, “Glow: Generative flow with invertible 1x1 convolutions,” arXiv preprint arXiv:1807.03039, 2018.

[18] Herbert Bay, Tinne Tuytelaars, and Luc Van Gool, “Surf: Speeded up robust features,” in European conference on computer vision. Springer, 2006, pp. 404–417.

[19] Myung-Joon Kwon, In-Jae Yu, Seung-Hun Nam, and Heung-Kyu Lee, “Cat-net: Compression artifact tracing network for detection and localization of image splicing,” in Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision, 2021, pp. 375–384.

[20] Mingqing Xiao, Shuxin Zheng, Chang Liu, Yaolong Wang, Di He, Guolin Ke, Jiang Bian, Zhouchen Lin, and Tie-Yan Liu, “Invertible image rescaling,” in European Conference on Computer Vision. Springer, 2020, pp. 126–144.

[21] Yoshua Bengio, Nicholas Léonard, and Aaron Courville, “Estimating or propagating gradients through stochastic neurons for conditional computation,” arXiv preprint arXiv:1308.3432, 2013.

[22] Zhaoyang Jia, Han Fang, and Weiming Zhang, “Mbrs: Enhancing robustness of dnn-based watermarking by mini-batch of real and simulated jpeg compression,” arXiv preprint arXiv:2108.08211, 2021.

[23] Hongyi Zhang, Moustapha Cisse, Yann N Dauphin, and David Lopez-Paz, “mixup: Beyond empirical risk minimization,” arXiv preprint arXiv:1710.09412, 2017.

[24] Olaf Ronneberger, Philipp Fischer, and Thomas 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.

[25] Phillip Isola, Jun-Yan Zhu, Tinghui Zhou, and Alexei A Efros, “Image-to-image translation with conditional adversarial networks,” in Proceedings of the IEEE conference on computer vision and pattern recognition, 2017, pp. 1125–1134.

[26] Xiaodong Mao, Qing Li, Haoran Xie, Raymond YK Lau, Zhen Wang, and Stephen Paul Smolley, “Least squares generative adversarial networks,” in Proceedings of the IEEE international conference on computer vision, 2017, pp. 2794–2802.

[27] Diederik P Kingma and Jimmy Ba, “Adam: A method for stochastic optimization,” arXiv preprint arXiv:1412.6980, 2014.

[28] Tsung-Yi Lin, Michael Maire, Serge Belongie, James Hays, Pietro Perona, Deva Ramanan, Piotr Dollár, and C Lawrence Zitnick, “Microsoft coco: Common objects in context,” in European conference on computer vision. Springer, 2014, pp. 740–755.

[29] Zhou Wang, Alan C Bovik, Hamid R Sheikh, and Eero P Simoncelli, “Image quality assessment: from error visibility to structural similarity,” IEEE transactions on image processing, vol. 13, no. 4, pp. 600–612, 2004.