A Multi-Task Learning Framework for Image Restoration Using a Novel Generative Adversarial Network

Rim Walha[1,2][a], Fadoua Drira[1][b] and Rania Bedhief[1]

[1]*REGIM-Lab, ENIS, University of Sfax, Tunisia*

[2]*Higher Institute of Computer Science and Multimedia of Sfax, University of Sfax, Tunisia*

Keywords: Multi-Task Learning, Deep Learning, Real-World Degradations, Image Restoration.

Abstract: In the last years, deep learning has gained growing popularity in image restoration, becoming the efficient mainstream for the subsequent higher level computer vision processing tasks. In particular, image restoration is a challenging task due to the high variations of degradations faced in the real-world scenarios. In this study, we introduce an efficient multi-task generative adversarial learning based framework as a practical solution suitable for various types of image degradations. We apply recent advancements in deep learning to design, build and train such a framework that can deal with several image restoration tasks treated simultaneously. More precisely, the main specificities of the proposed architecture are: (1) the introduction of a novel generator based on an encoder with separate decoders, (2) the utilization of low-level multi-scale features within the encoder component of our architecture, (3) the incorporation of the multi-scale transformer technique in each decoder in order to learn and share the low-level features representations among different tasks. Our experimental study demonstrates the efficiency and the robustness of the proposed framework for two specific image restoration tasks including image deblurring and image denoising. Moreover, it achieves high performance results that exceed those of state-of-the-art methods when evaluated on the same datasets.

1 INTRODUCTION

Nowadays, the popularity of computer vision applications reveals a pressing need for high quality images to guarantee efficient based systems. Nevertheless, on the ascendant massive use of mobile internet and the ubiquitous presence of cameras on various devices, image quality could inevitably be corrupted by several degradations during the acquisition and transmission processes. The presence of these undesirable artifacts including blur, noise, low resolution could adversely affect the overall performance. An adequate solution to this problem is image restoration; a process that aims to recover an image from its degraded version. For simplicity, we limit this study to two main image restoration tasks including denoising and deblurring. Indeed, the principal challenge in image denoising is to recover a clean signal A from the noisy observation B corrupted by an additive noise N, namely: $B = A + N$. Sophisticated filters have been proposed in the literature, most of them could

be classified into six categories: wavelet-based, linear, non-linear, adaptive, total variation, and partial differential equation based filters. Classical denoising methods are mainly based on modifying transform coefficients (Guo et al., 2019) or averaging neighborhood pixels (Walha et al., 2014). Major difficulties in noise removal consist in feature/edge/texture preservation while smoothing away noise in flat regions without additional processing artefacts (Drira et al., 2012; Walha et al., 2015; Walha et al., 2018). These difficulties concern also the image deblurring. The latter is the process of removing blurs. It could be blind or non-blind according to the usage of blur kernel information. Non-Blind process refers to handle an image by a given known blur kernel, while blind process aims to restore sharpness in an image without prior knowledge about the blur kernel (Guemri et al., 2017). In general, a blurred image Y can be modeled as: $Y = K * X + N$, where K is the blur kernel, X is the sharp image and N is the additive noise.

Earlier studies formulated the image restoration as an inverse problem, emphasizing the definition of a model for corrupted images while taking into account priors of clean images. This model is then exploited to

---

[a] https://orcid.org/0000-0002-0483-6329

[b] https://orcid.org/0000-0001-6706-4218

minimize an objective function, aiming to reconstruct a clear image from its degraded version (Walha et al., 2013). Due to the success of deep learning to achieve good outstanding performance in various computer vision applications (Harizi et al., 2022a; Harizi et al., 2022b), recent image restoration studies have focused on proposing solutions based on these architectures. The main advantage here is that no explicit modeling of image prior is required. Well-known state-of-the-art deep learning based solutions could be classified into two groups: Single Task Learning (STL) and Multi-Task Learning (MTL) based solutions. On one hand, the first group treats each degradation independently therefore various networks are conceived for different degradation types. On the other hand, the second group focuses on the proposition of single networks that can deal with a combined set of degradations. It aims to optimize the performance across multiple task predictors through some transfer knowledge between them (Caruana, 1997).

In this work, our main concern is the proposition of a new deep auto-encoder based multi-task generative adversarial learning framework for image restoration. The latter is inspired by the recent success of deep architectures including Convolution Auto-Encoder (CAE), Transformers and Generative Adversarial Networks (GAN). More specifically, we propose a multi-task end-to-end framework based on a single encoder which learns multi-scale features representations to be shared between different tasks. The framework encompasses separate decoders containing multi-scale transformer blocks useful for further analyzing local image structure and fine details across multi-scales in order to achieve an effective restoration. Each decoder focuses on its specific restoration task. The proposed framework could be obviously extended to cope with other restoration tasks.

The rest of this paper is organized as follows: Section 2 outlines related works on MTL and STL in deep networks-based image restoration. Section 3 details the proposed multi-task learning framework. Section 4 presents our experimental study. The study is closed with conclusions and emerging aspects for future research in Section 5.

2 RELATED WORK

To address image restoration, deep neural networks employ STL methods, emphasizing specialized architectures for individual degradation types. Numerous recent publications explore this area; (Koh et al., 2021; Wang et al., 2022; Zamir et al., 2022; Li et al., 2018) to name a few. These methods enable a net-

work to restore various types of degradation using the same architecture. In (Wang et al., 2022), UFormer, a transformer-based architecture, was introduced for image restoration. It relies on a learnable multi-scale restoration modulator incorporated into the decoder. Despite its effectiveness, single-stage methods like UFormer often exhibit high network complexity. Later, Cheng et al. overcome the recourse to complicated architecture for image denoising via subspace learning (Cheng et al., 2021). A simple baseline that adopts the single-stage UNet as architecture was suggested in (Chen et al., 2022). A set of different networks could collaborate to tackle complex image restoration tasks. Such process evolves multi-stage methods that decompose the overall task into smaller easier sub-tasks, each stage is based on a lightweight sub-network. In (Zamir et al., 2021), the authors propose MPRNet as a multi-stage progressive image restoration architecture, composed of two encoder-decoder sub-networks and one original resolution sub-network. Another multi-stage restoration method proposed in (Chen et al., 2021), is called HiNet. Feature fusion and attention-guided map are introduced across stages. Also, a multi-axis MLP based architecture called MAXIM was proposed in (Tu et al., 2022).

Other recent studies proceed via MTL methods in deep neural networks to deal with combined degradations. In fact, MTL is a subfield of machine learning useful in domain-related tasks (Crawshaw, 2020). It is a mechanism of learning multiple tasks simultaneously using a shared model compared to STL. Generally, MTL improves the generalization capability and accuracy performance mainly for correlated or related tasks. Giving this context, the network could benefit from domain-specific knowledge encapsulated in the training samples of the different tasks. Good representations could be thus learned with less amount of data and reduced overfitting. For instance, Liu et al. proposed a two-step training based framework to restore images with unknown degradation factors (Liu et al., 2019). These steps include MTL and fine tuning. Martyniuk (Martyniuk, 2019) presented an end-to-end pipeline that contains a generic encoder and separate decoders. The author introduced a new architecture for the generator inspired by the feature pyramid networks to deal with deblurring, dehazing and rain-drops removal tasks.

In conclusion, even though we noticed a limited number of propositions dealing with MTL, the latter is an active research area with promising issues. It could be very useful mainly for real-time applications. Indeed, the combinaison of mutiple tasks into the same learning model reduces the computational
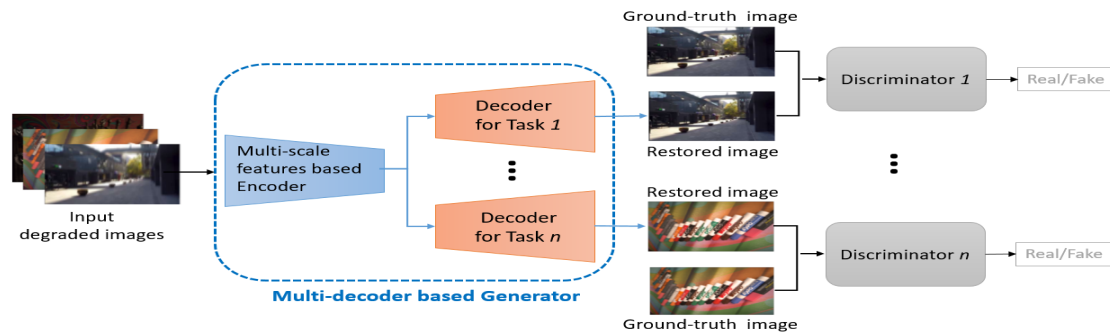
Figure 1: Overview of the proposed multi-task generative adversarial network for image restoration.

complexity. This justifies the motivation of our study.

3 PROPOSED MULTI-TASK GENERATIVE ADVERSARIAL LEARNING NETWORK FOR IMAGE RESTORATION

In this section, we describe the proposed image restoration framework that relies on a multi-task GAN. An overview of the proposed framework is depicted in Fig. 1. As illustrated in this figure, our generator comprises a single multi-scale features based encoder and separate decoders. Separate discriminators are used to distinguish fake images from real images. More details about each part of our proposition are given in the following subsections.

3.1 Preliminaries

In this section, we explain the preliminary works on which is based our framework. Indeed, Generative Adversarial Network is the core of our proposed framework. It represents a deep neural network architecture designed by Goodfellow et al. (Goodfellow et al., 2014). Figure 2 illustrates its basic framework. It is comprised of two neural networks – a generative model and a discriminative model – trained with an adversarial loss function to generate data that resembles a distribution. The first neural network, the generator is used to generate new samples as close as possible to given samples. The second neural network, the discriminator, is to discriminate between two different classes of data from the generator and to determine the real or the fake.

Despite their promising results, the main limitation of GANs is their unstable learning generally caused by the gradient vanishing and the mode collapse. To improve the learning stability of GAN-based models, several variant have been proposed.

Coupling GAN with Auto-encoders (AE) as a secondary network is one among interesting proposed solutions. Indeed, each network has its own learning process. Furthermore, AE could represent data samples with lower dimensionality. The AE model encompasses three components: the encoder, the decoder, and the loss function to compare the output to the target image. Our investigation concerns GAN and AE based models with a context of MTL.

Figure 2: Generative adversarial network basic framework.

3.2 Encoder Proposed Structure

In order to capture relevant features that describes fine details of an image, we propose a multi-scale features based encoder as illustrated in Fig. 3. More precisely, in order to deal with degradations at different scales, the proposed encoder starts with extracting features from the input degraded image x by using the Feature Pyramid Network (FPN) (Lin et al., 2017) with Mobilnet-V2 backbone (Sandler et al., 2018). This network outputs k multi-scale feature maps, each of them is a $2D$-dimensional vector corresponding to the features extracted at different scales of the image. These maps are referred to as annotation vectors:

$$FPN(x) = \{M0, M1, ..., Mk\} \tag{1}$$

These multi-scale feature maps (except $M0$) are further analyzed, as shown in Fig. 3, and are passed through Convolution Maxpooling Blocks (CMB) to extract low-level features. Each block comprises a stack of eight units, each of them is formed by a 2D-convolution layer followed by a 2D-maxpooling layer and a ReLu activation layer. Especially, the number of filters used in the convolution layers ranges from
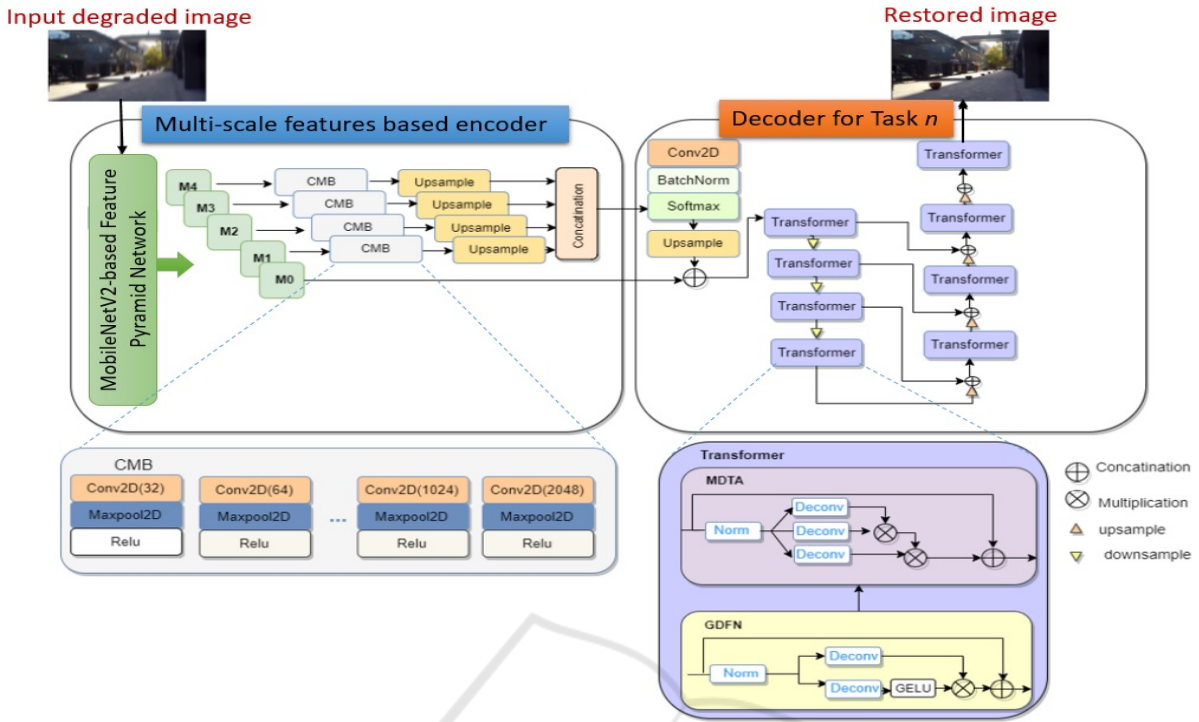
Figure 3: Illustration of the encoder-decoder structure within the proposed multi-decoder based generator.

32 to 2048. The maxpooling is performed on the feature maps obtained from the convolution layer to progressively reduce the spatial size of the representation that minimizes the number of parameters and computations in the proposed network. The output of each CMB unit is transmitted to the next unit and the final output multi-scale low-level features are up-sampled and concatenated into one tensor which contains relevant information on different scale levels.

3.3 Decoders Proposed Structure

The proposed multi-task GAN-based image restoration framework involves multi-decoder based generator. More precisely, separate decoders are used to deal with different image restoration tasks. The same architecture is used for each decoder. The separate decoders share the low-level multi-scale features generated by the proposed encoder. As illustrated in Fig. 3, these features are further analyzed in the decoder part using a 2D-convolution layer followed by batch normalization and softmax layers, then upsampled and concatenated with the $M0$ feature map. This extra skip-connection with $M0$ is applied for preserving the features representation between the encoder and the decoder parts, which helps the decoder to recover information that might have been lost during feed-forward convolutions. This leads to achieve better sharpness and structural similarity in the generated image.

Motivated by the outstanding success of transformers in recent computer vision works, we propose to integrate in the decoders part the multi-scale transformer blocks. This enhances our framework's ability to handle various degradations simultaneously, yielding promising results. As shown in Fig. 3, extra skip connections between the transformer blocks are used. Especially, inspired from (Zamir et al., 2022), each transformer block consists of multi-deconvolution head transposed attention (MDTA) and gated-deconvolution feed-forward network (GDFN). The transformer block is designed as a multi-scale block to adjust features in multiple layers of the decoder part. In fact, the MDTA is composed by depth-wise convolutions to emphasize the local context in order to produce the global attention map with layer normalized tensor. We apply 1×1 convolutions to aggregate pixel-wise cross-channel context followed by 3×3 depth-wise convolutions to encode channel-wise spatial context. GDFN consists of depth-wise convolutions, which helps to encode information from the neighboring pixel positions and it is useful for learning local image structure for effective restoration. GDFN contains also the non-linearity GELU and the layer normalization. It controls the information flow through the respective hierarchical scale, thereby allowing each scale to focus on the fine details complimentary to the other scale. GDFN focuses on

3.4 Loss Function

Likewise with other multi-scale restoration networks (Martyniuk, 2019), we use a mixing loss L that combines a content loss L_{cont} with an adversarial loss L_{adv}. Hence, the overall loss function L can be formulated as follows:

$$L = L_{cont} + \beta \cdot L_{adv} \qquad (2)$$

where β is a weight constant. The adversarial loss corresponds to the discriminators loss function, whereas the content loss corresponds to the generator loss function. In this work, we use the Wasserstein distance GAN Gradient Penalty, WGAN-GP (Gulrajani et al., 2017), as the discriminators loss function. In fact, WGAN-GP is shown to be efficient for improving the stability of training and also to be robust to the choice of the generator structure. The game between the generator G and the discriminator $D_{t \in \{1..n\}}$, called as a critic, relies on the WGAN objective function L_{WGAN} which is constructed using the Kantorovich-Rubinstein duality and defined as follows:

$$L_{WGAN} = E_{x \sim P_r} \left[D_t(x) \right] - E_{\hat{x} \sim P_g} \left[D_t(\hat{x}) \right] \qquad (3)$$

where P_r presents the data distribution over real sample x and P_g constitutes the generator's distribution, defined by $\hat{x} = G(z)$; the input z corresponds to a sample from a noise distribution. The discriminator tries to maximize the L_{WGAN} function during the training phase by maximizing the difference between its results on real samples and its results on fake samples.

In the new form of Wasserstein metric, D_t is demanded to be K-Lipschitz continuous. The idea is that there exists a real constant $K \geq 0$, called a Lipschitz constant, and the critic value approximates $K \cdot W(P_r, P_\upsilon)$, where $W(P_r, P_\upsilon)$ is the Wasserstein distance that measures the distance between the distributions P_r and P_υ. Here, each discriminator D_t approximates the distance between real and fake samples.

The WGAN concept requires that the discriminator relies within the space of 1-Lipschitz functions. In order to enforce the Lipschitz constraint and to maintain a stable learning process with gradient descents, Gulrajani et al. (Gulrajani et al., 2017) suggest to add to L_{WGAN} a gradient penalty term which is defined as:

$$\eta E_{\hat{x} \sim P_{\hat{x}}} \left[\left(\| \nabla_{\hat{x}} D_t(\hat{x}) \|_2 - 1 \right)^2 \right] \qquad (4)$$

Thereby, the adversarial loss function consists of two parts which are the WGAN loss function L_{WGAN} and the gradient penalty. This can be formulated by:

$$L_{adv} = L_{WGAN} + \eta E_{\hat{x} \sim P_{\hat{x}}} \left[\left(\| \nabla_{\hat{x}} D_t(\hat{x}) \|_2 - 1 \right)^2 \right] \quad (5)$$

For the content loss L_{cont}, the classical choice can be the Mean Absolute Error (MAE) loss or the Mean Squared Error (MSE) loss on raw pixels. Using those functions leads to blurry artifacts on generated images. In this work, the content loss has two loss components: the L_1 loss for preserving colors and the perceptual loss function L_X as follows:

$$L_{cont} = L_X + 0.5 \cdot L_1 \qquad (6)$$

The perceptual loss function L_X (Eq.7) presents an L2-loss (Johnson et al., 2016), but relies on the difference between the generated and target images according to the feature maps within the generator part.

$$L_X = \frac{1}{W_{i,j} H_{i,j}} \sum_{x=1}^{W_{i,j}} \sum_{y=1}^{H_{i,j}} \left(\Phi_{i,j} \left(I^S \right)_{x,y} - \Phi_{i,j}(G(I^B))_{x,y} \right)^2$$
$$(7)$$

where I^B and I^S are respectively a degraded image and a ground-truth image, $H_{i,j}$ and $W_{i,j}$ denotes the dimensions of the feature maps, and $\Phi_{i,j}$ represents the feature map generated by the j-th convolution layer.

4 EXPERIMENTAL EVALUATION

4.1 Datasets and Settings

In order to evaluate the proposed image restoration framework, two well-known datasets are used, including the GoPro dataset (Nah et al., 2017) and the SSID dataset (Abdelhamed et al., 2018). Especially, the SIDD, a Smartphone Image Denoising dataset, contains 30000 noisy images taken from ten real-world scenes under various lighting conditions and using five smartphone cameras. This dataset aims to address the problems of smartphones images denoising, where the small sensor and aperture size cause noticeable noise even in pictures taken at base ISO. Further processing is applied to provide ground-truth images along with the noisy images. The GoPro dataset, widely used for image deblurring, consists of 3214 pairs of blurry and sharp images captured in the wild at 1280×720 resolution. It is divided into 2103 training pairs and 1111 test pairs. The images are derived from high-speed camera videos at 240 frames per second, with blurry images obtained by averaging successive frames.

In this work, we used PyTorch for our implementation. The proposed multi-task framework was trained on randomly cropped image patches of size 256×256 for 100 iterations per task. Horizontal and vertical flips and rotations are adopted for data augmentation. The training was performed on "Google Collaboratory Pro" using a GPU. The parameter β in

Eq.(4), that controls the relative importance of the loss terms, is set to 0.01. The FPN within the encoder generated 5 feature maps ($k \in 0..4$).

4.2 Comparison with State-of-the-Art Methods

The proposed network is evaluated on natural scene images and compared with state-of-the-art deblurring networks and denoising networks. Performance Evaluation can be broadly categorized into quantitative and qualitative evaluations. For the quantitative evaluation, we use the Peak Signal-to-Noise Ratio (PSNR) and the Structural SIMilarity index (SSIM).

Table 1 (respectively Table 2) illustrates the values of PSNR and SSIM generated by different methods performed on the GoPro dataset (respectively the SIDD dataset) for the deblurring task (respectively the denoising task). The proposed multi-task GAN-based framework achieves a new state-of-the-art PSNR value of 36.17 dB on the GoPro dataset, as shown in Table 1. Results, illustrated in Figure 4, demonstrate superior deblurring performance with sharper images and improved preservation of edges and local details compared to state-of-the-art deblurring networks.

Our framework, capable of handling various degradations, is tested on the SIDD dataset for denoising. Results in Table 2 show its superior performance compared to other state-of-the-art denoising networks, as measured by PSNR. The proposed framework demonstrates the second-best SSIM index performance and effectively removes real noise while preserving fine details in denoised images, as shown in Figure 5 with magnified regions from the SIDD dataset. Our effective deblurring and denoising results stem from utilizing low-level multi-scale features in our encoder, enabling detailed analysis of corrupted images. Moreover, the proposed decoder structure relies on multiple instances of transformer block which boost the GAN overall performance in image restoration.

5 CONCLUSION AND OPEN ISSUES

In this paper, we proposed an effective multi-task GAN-based image restoration framework that addresses various degradations. Key features include: (1) a unique generator with separate decoders, (2) utilization of low-level multi-scale features in the encoder, and (3) integration of multi-scale Transformer

Table 1: PSNR and SSIM values from various methods on the GoPro dataset for image deblurring, with the top two results emphasized in bold and underlined.

Restoration method	PSNR	SSIM
MIMO-UNet (Cho et al., 2021)	32.68	0.959
HiNet (Chen et al., 2021)	32.71	0.959
MAXIM (Tu et al., 2022)	32.86	<u>0.961</u>
Restormer (Zamir et al., 2022)	32.92	<u>0.961</u>
MSCNN (Nah et al., 2017)	29.20	0.916
MPRNet (Zamir et al., 2021)	32.66	0.959
SRNet (Tao et al., 2018)	30.10	0.932
NAFNet (Chen et al., 2022)	32.88	<u>0.961</u>
MTLGAN (Martyniuk, 2019)	27.30	0.810
DeblurGAN-v2 (Kupyn et al., 2019)	29.55	0.934
MFC-Net (Xia et al., 2022)	31.04	0.916
SVRNN (Ren et al., 2022)	30.46	0.936
UFormer (Wang et al., 2022)	32.97	**0.967**
Our proposition	**36.17**	<u>0.961</u>

Table 2: PSNR and SSIM values from various methods on the SIDD dataset for image denoising. Best first and second results are highlighted in bold and underlined, respectively.

Restoration method	PSNR	SSIM
MPRNet (Zamir et al., 2021)	39.71	0.958
MIRNet (Zamir et al., 2020)	39.72	0.959
NBNet (Cheng et al., 2021)	39.75	0.959
UFormer (Wang et al., 2022)	39.89	0.960
MAXIM (Tu et al., 2022)	39.96	0.960
HiNet (Chen et al., 2021)	39.99	0.958
Restormer (Zamir et al., 2022)	40.02	0.960
NAFNet (Chen et al., 2022)	40.30	**0.962**
Our proposition	**42.41**	<u>0.961</u>

technique in each decoder to learn and share low-level feature representations across tasks. Extensive experimental results on two datasets demonstrate the framework's superior quantitative results compared to state-of-the-art methods for various degradations. Additionally, qualitative evaluations on degraded images confirm the framework's ability to visually reconstruct plausible deblurred and denoised images efficiently. As a perspective of this work, we plan to explore our multi-task framework for various applications, with a focus on evaluating its effectiveness in tasks like image super-resolution, underwater image restoration, and image dehazing.

(a)

(b)

(c)

(d)

Figure 4: Examples of deblurring results on GoPro dataset. Left to right: input blurry images, output images obtained by the proposed framework, ground truth images.
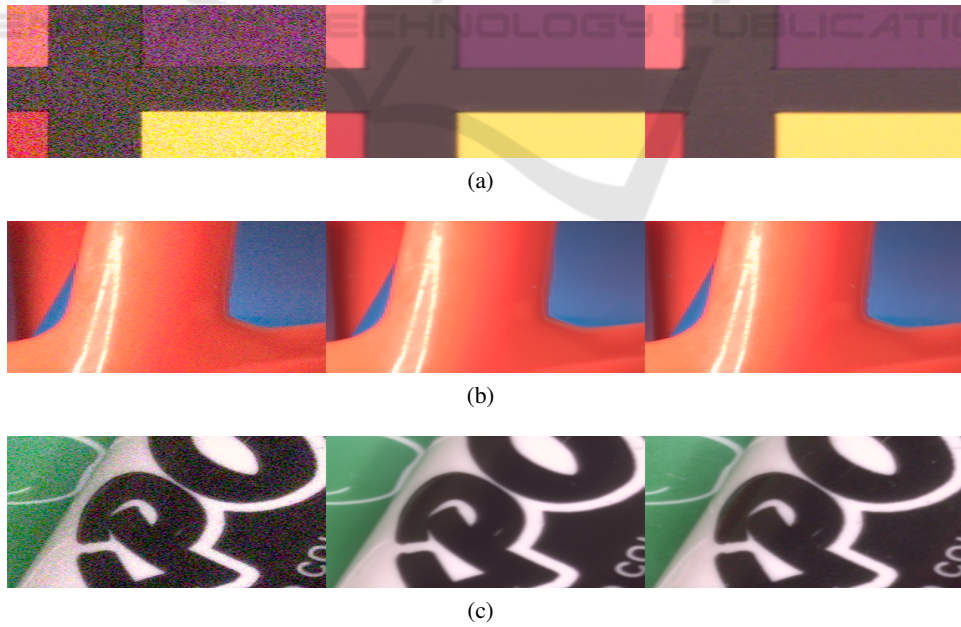

(a)

(b)

(c)

Figure 5: Examples of denoising results on SIDD dataset. Left to right: input noisy images, images reconstructed by the proposed framework, ground truth images.

REFERENCES

Abdelhamed, A., Lin, S., and Brown, M. S. (2018). A high-quality denoising dataset for smartphone cameras. In *CVPR*, pages 1692–1700.

Caruana, R. (1997). Multitask learning. *Mach. Learn.*, 28(1):41–75.

Chen, L., Chu, X., Zhang, X., and Sun, J. (2022). Simple baselines for image restoration. In *ECCV, Part VII*, volume 13667, pages 17–33.

Chen, L., Lu, X., Zhang, J., Chu, X., and Chen, C. (2021). Hinet: Half instance normalization network for image restoration. In *CVPR*, pages 182–192.

Cheng, S., Wang, Y., Huang, H., Liu, D., Fan, H., and Liu, S. (2021). Nbnet: Noise basis learning for image denoising with subspace projection. In *CVPR*, pages 4896–4906.

Cho, S., Ji, S., Hong, J., Jung, S., and Ko, S. (2021). Rethinking coarse-to-fine approach in single image deblurring. In *ICCV 2021*, pages 4621–4630.

Crawshaw, M. (2020). Multi-task learning with deep neural networks: A survey. *CoRR*, abs/2009.09796.

Drira, F., Lebourgeois, F., and Emptoz, H. (2012). A new pde-based approach for singularity-preserving regularization: application to degraded characters restoration. *IJDAR*, 15(3):183–212.

Goodfellow, I. J., Pouget-Abadie, J., Mirza, M., Xu, B., Warde-Farley, D., Ozair, S., Courville, A. C., and Bengio, Y. (2014). Generative adversarial nets. In *NIPS 2014*, pages 2672–2680.

Guemri, K., Drira, F., Walha, R., Alimi, A. M., and Lebourgeois, F. (2017). Edge based blind single image deblurring with sparse priors. In *VISAPP 2017*, pages 174–181.

Gulrajani, I., Ahmed, F., Arjovsky, M., Dumoulin, V., and Courville, A. C. (2017). Improved training of wasserstein gans. In *NIPS 2017*, pages 5767–5777.

Guo, S., Yan, Z., Zhang, K., Zuo, W., and Zhang, L. (2019). Toward convolutional blind denoising of real photographs. In *CVPR 2019*, pages 1712–1722.

Harizi, R., Walha, R., and Drira, F. (2022a). Deep-learning based end-to-end system for text reading in the wild. *Multim. Tools Appl.*, 81(17):24691–24719.

Harizi, R., Walha, R., Drira, F., and Zaied, M. (2022b). Convolutional neural network with joint stepwise character/word modeling based system for scene text recognition. *Multim. Tools Appl.*, 81(3):3091–3106.

Johnson, J., Alahi, A., and Fei-Fei, L. (2016). Perceptual losses for real-time style transfer and super-resolution. In *ECCV, Part II*, volume 9906, pages 694–711. Springer.

Koh, J., Lee, J., and Yoon, S. (2021). Single-image deblurring with neural networks: A comparative survey. *Comput. Vis. Image Underst.*, 203:103134.

Kupyn, O., Martyniuk, T., Wu, J., and Wang, Z. (2019). Deblurgan-v2: Deblurring (orders-of-magnitude) faster and better. In *ICCV 2019*, pages 8877–8886.

Li, X., Wu, J., Lin, Z., Liu, H., and Zha, H. (2018). Recurrent squeeze-and-excitation context aggregation net for single image deraining. In *ECCV, Part VII*, volume 11211, pages 262–277.

Lin, T., Dollár, P., Girshick, R. B., He, K., Hariharan, B., and Belongie, S. J. (2017). Feature pyramid networks for object detection. In *CVPR 2017*, pages 936–944.

Liu, X., Suganuma, M., Luo, X., and Okatani, T. (2019). Restoring images with unknown degradation factors by recurrent use of a multi-branch network. *arXiv: CVPR*.

Martyniuk, T. (2019). *Multi-task learning for image restoration*. PhD thesis, Faculty of Applied Sciences, Ukrain.

Nah, S., Kim, T. H., and Lee, K. M. (2017). Deep multi-scale convolutional neural network for dynamic scene deblurring. In *CVPR*, pages 257–265.

Ren, W., Zhang, J., Pan, J., Liu, S., Ren, J. S., Du, J., Cao, X., and Yang, M. (2022). Deblurring dynamic scenes via spatially varying recurrent neural networks. *IEEE Trans. Pattern Anal. Mach. Intell.*, 44(8):3974–3987.

Sandler, M., Howard, A. G., Zhu, M., Zhmoginov, A., and Chen, L. (2018). Mobilenetv2: Inverted residuals and linear bottlenecks. In *CVPR*, pages 4510–4520.

Tao, X., Gao, H., Shen, X., Wang, J., and Jia, J. (2018). Scale-recurrent network for deep image deblurring. In *CVPR*, pages 8174–8182.

Tu, Z., Talebi, H., Zhang, H., Yang, F., Milanfar, P., Bovik, A., and Li, Y. (2022). Maxim: Multi-axis mlp for image processing. In *CVPR*, pages 5759–5770.

Walha, R., Drira, F., Alimi, A. M., Lebourgeois, F., and Garcia, C. (2014). A sparse coding based approach for the resolution enhancement and restoration of printed and handwritten textual images. In *ICFHR 2014*, pages 696–701.

Walha, R., Drira, F., Lebourgeois, F., Garcia, C., and Alimi, A. M. (2013). Single textual image super-resolution using multiple learned dictionaries based sparse coding. In *ICIAP 2013, Part II*, volume 8157, pages 439–448.

Walha, R., Drira, F., Lebourgeois, F., Garcia, C., and Alimi, A. M. (2015). Joint denoising and magnification of noisy low-resolution textual images. In *ICDAR 2015*, pages 871–875.

Walha, R., Drira, F., Lebourgeois, F., Garcia, C., and Alimi, A. M. (2018). Handling noise in textual image resolution enhancement using online and offline learned dictionaries. *Int. J. Document Anal. Recognit.*, 21(1-2):137–157.

Wang, Z., Cun, X., Bao, J., Zhou, W., Liu, J., and Li, H. (2022). Uformer: A general u-shaped transformer for image restoration. In *CVPR*, pages 17662–17672.

Xia, H., Wu, B., Tan, Y., Tang, X., and Song, S. (2022). Mfc-net: Multi-scale fusion coding network for image deblurring. *Appl. Intell.*, 52(11):13232–13249.

Zamir, S. W., Arora, A., Khan, S., Hayat, M., Khan, F. S., and Yang, M. (2022). Restormer: Efficient transformer for high-resolution image restoration. In *CVPR*, pages 5718–5729.

Zamir, S. W., Arora, A., Khan, S. H., Hayat, M., Khan, F. S., Yang, M., and Shao, L. (2020). Learning enriched features for real image restoration and enhancement. In *ECCV, Part XXV*, volume 12370, pages 492–511.

Zamir, S. W., Arora, A., Khan, S. H., Hayat, M., Khan, F. S., Yang, M., and Shao, L. (2021). Multi-stage progressive image restoration. In *CVPR*, pages 14821–14831.