IEEE Access
Multidisciplinary | Rapid Review | Open Access Journal

Received 7 August 2022, accepted 23 August 2022, date of publication 1 September 2022, date of current version 12 September 2022.

Digital Object Identifier 10.1109/ACCESS.2022.3203736

RESEARCH ARTICLE

EQAdap: Equipollent Domain Adaptation Approach to Image Deblurring

IBSA JALATA[1], NAGA VENKATA SAI RAVITEJA CHAPPA[1], THANH-DAT TRUONG[1], (Graduate Student Member, IEEE), PIERCE HELTON[1], CHASE RAINWATER[2], AND KHOA LUU[1], (Member, IEEE)

[1]Department of Computer Science and Computer Engineering, University of Arkansas, Fayetteville, AR 72701, USA
[2]Department of Industrial Engineering, University of Arkansas, Fayetteville, AR 72701, USA

Corresponding author: Naga Venkata Sai Raviteja Chappa (nchappa@uark.edu)

This work was supported in part by the NSF Data Science, in part by the Data Analytics that are Robust and Trusted (DART), in part by the Arkansas Biosciences Institute (ABI) Grant Program, and in part by the NSF West Virginia Arkansas- Collaborative Research and Education in Smart Health (WVAR-CRESH) Grant.

ABSTRACT In this paper, we present an end-to-end unsupervised domain adaptation approach to image deblurring. This work focuses on learning and generalizing the complex latent space of the source domain and transferring the extracted information to the unlabeled target domain. While fully supervised image deblurring methods have achieved high accuracy on large-scale vision datasets, they are unable to well generalize well on a new test environment or a new domain. Therefore, in this work, we introduce a novel Bijective Maximum Likelihood loss for the unsupervised domain adaptation approach to image deblurring. We evaluate our proposed method on GoPro, RealBlur_J, RealBlur_R, and HIDE datasets. Through intensive experiments, we demonstrate our state-of-the-art performance on the standard benchmarks.

INDEX TERMS Deep neural networks (DNNs), instance level affinity-domain adaptation (ILA-DA), unsupervised domain adaptation (UDA).

I. INTRODUCTION

Image blurring is a challenging problem in computer vision. Image blurring happens when the object being recorded changes during the recording of a single exposure, due to rapid movement or long exposure time. For the blurred image, the underlying scene dynamics are unraveled and the the sharp version of the blurred image can be recovered by the inverse problem called deblurring. Though easy motion patterns, e.g. object moving at moderate speed, defocused camera, camera shake, are extensively studied and formulated in previous methods, more complicated motion dynamics, i.e. medium to high blur, have been difficult to address properly.

Recently, image deblurring has been experiencing a revival because of deep learning methods, particularly Convolutional Neural Networks (CNNs) [1], [2], [3], [4]. CNN methods address the challenges observed in methods that use a hand-crafted technique with empirical observations. The method learns general prior by capturing image features from large-scale data that give a performance gain over the other hand-crafted methods. There are many CNN-based methods with variant models that achieve better performance. The methods and functional units commonly used in deblurring include generative models, encoder and decoder approaches, dilated convolutions, recursive residual learning, attention methods, and dense connections.

The procedure that aims to attenuate the challenges discussed above is referred to as Unsupervised Domain Adaptation. This method involves training a deep learning model on the labeled source dataset and adapting to the unlabeled target dataset to make sure the performance is maintained on the new domain.

Contributions of This Work: In this work, we introduce a novel Equipollent Domain Adaptation (EQAdap) approach to Image Deblurring. The contributions can summarized as follows.

The associate editor coordinating the review of this manuscript and approving it for publication was Jiachen Yang.

Firstly, we present a novel metric of domain adaptation to image deblurring that utilizes both data from the source domain and unlabeled target domain in an unsupervised manner. Particularly, along with the supervised training on the source domain, the new metric also includes a new unsupervised loss that allows training on the target domain without annotation. Secondly, the intensive experiments on three benchmarks, i.e. GoPro → RealBlur_J, GoPro → RealBlur_R and GoPro → HIDE have shown the performance of our approaches. Also, we introduce a new experimental setting that shows our state-of-the-art (SOTA) performance compared to prior SOTA approaches.

II. RELATED WORKS

In this paper, we are focused on image blurring and image deblurring, which are briefly introduced as follows, respectively.

A. IMAGE BLURRING

The blurring is mathematically formulated as,

$$B = K * I + N \tag{1}$$

where B is a blurry image, I is a sharp image, and N is additive noise. K is a known (non-blind) and unknown (blind) blur kernel. In equation 1, $*$ represents the convolution operator [5], [6]. Many of the deblurring problems fall under the categories of non-blind and blind deblurring. Non-blind deblurring (NBD) methods attempt to restore the original image, given the blur estimate. Most of the methods depend on traditional approaches such as Wiener filter [7] and Richardson-Lucy deconvolution [8] which are known to cause ringing artifacts and to obtain sharp image (I) estimates. Some methods of non-blind deblurring use a Maximum a posteriori (MAP) estimation, which employs an augmented optimization objective that incorporates a prior distribution. Although image global priors [9], [10] are commonly used in NBD, local priors that are patch-based [11] have been effective. The existing image prior in MAP is assumed to be combined with one specific data term for deblurring which is based on the l2 norm that models image noise with a Gaussian distribution [9]. However, in the presence of outliers and serious noise in the input image, The Laplacian model [12] shows effectiveness and produces good results in a reasonable amount of time compared to the Gaussian model. Moreover, the gradient of natural images is well represented by hyper-Laplacian methods. Prior methods have tried to address the problem of outliers. Cho *et al.* [13] discussed the severe ringing artifacts caused by outliers in input images. In the method, they used Expectation-Maximization to develop a deconvolution method [14] to address the non-linear property of the image formation due to saturated pixels. They use a forward model that is a modified version of the Richardson-Lucy algorithm. In recent time, CNNs have been widely used to deal with image noise and saturation: [15] captured the characteristics of degradation by utilizing both traditional and CNN based methods. However, the methods were found to be ineffective since their networks need to be fine tuned for every kernel. CNNs have also been used to learn image priors and perform outlier-robust image restoration. The work in [16] uses a CNN for estimating blur kernels from local patches and predicts the probabilistic distribution of motion blur field using a Markov random field -> model, but the scope of their network is limited to a single specific blur kernel. Some of the recent works on NBD employ machine learning frameworks such as Gaussian conditional random fields [29] or shrinkage fields [6], whereas the most recent work in [17] uses CNN based regularization. However, none of these methods can handle noisy blur kernels.

Blind deblurring methods try to restore the original image from blurred images without the presence of a blur kernel. Previously, most blind deblurring methods [17], [18] were developed based on non-blind deblurring methods to restore sharp images [4], [15]. Pan *et al.* [18] proposed a method by removing the outliers in the intermediate latent images and extracting reliable edges for kernel estimation. Dong *et al.* [19] approached outliers differently than the prior methods [4], [11]. The method avoids the heuristic outliers detection step and focuses on measuring the goodness-of-fit so that the outliers have a minimum effect in the blur kernel estimation process.

Prior methods have also used convolutional neural networks for blind deblurring. The approach is usually data driven and takes advantage of the large learning capacity of neural networks on the given datasets. Tao *et al.* [3] used an encoder-decoder approach by incorporating it with a scale recurrent network to restore sharp images. The recurrent network captures a significant cue from blurred image and the number of trainable parameters is reduced significantly. Kupyn *et al.* [20] used an end to end learning method, GAN to generate a high quality image. Kupyn *et al.* [21] later introduced the Feature Pyramid Network as a backbone for the generator of DeblurGAN-v2, and it is based on a relativistic conditional GAN with a double-scale discriminator. The methods achieved a good performance and higher efficiency that it is predecessor, and the method was applied in video deblurring and domain specific deblurring methods.

In recent years, DNNs have been widely employed for image deblurring. Early works substituted some modules in the conventional optimization-based framework with DNNs [5], [22]. Chakrabarti [22] used DNNs to predict the complex Fourier coefficients of the blur kernel. Sun *et al.* [16] explicitly estimated the blur kernel at the patch level. Gong *et al.* [23] utilized DNNs to estimate the motion flow from blurry images. The clean images were obtained via non-blind deconvolution. Nah *et al.* [2] adopted a kernel free method to generate a large-scale dynamic scene deblurring dataset by averaging the consecutive frames in high-speed videos. Furthermore, they proposed a multi-scale architecture to progressively restore the latent sharp image. Since then, various networks were proposed in an end-to-end manner and have redefined the state-of-the-art results. That includes: deep hierarchical multi-patch network, selective sharing scheme, incremental temporal training, efficient

pixel adaptive and feature attentive design. However, those methods are sub-optimal since the same generic model is applied to every test image and fails to explore the specific internal information.

1) DOMAIN ADAPTATION

Recently, unsupervised domain adaptation (UDA) has become a prominent research focus in the field of computer vision. It has four primary approaches: adversarial learning [24], [25], [26], [27], [28], [29], [30], self-training [31], entropy minimization [32], [33], [34] and domain discrepancy minimization [35], [36], [37].

Sharma *et al.* [38] proposed an instance affinity based criterion during the process of transfer called Instance Level Affinity-Domain Adaptation(ILA-DA). They initially proposed a reliable and efficient method to extract similar and dissimilar samples across the source and target followed by utilization of multi-sample contrastive loss to drive the alignment of the domain. Wang and Jiang [39] proposed coupled generative adversarial networks (CoGAN) for the problem of zero-shot domain adaptation (ZSDA) and introduced a couple of classfiers to control the training process. Wang and Jiang [40] introduced a new solution to the ZSDA problem; their proposed network structure extends the coupled generative adversarial networks (CoGAN) into a conditional model. Na *et al.* [41] proposed a UDA method that handles large discrepancies present in the domain. They introduced a fixed ratio-based mixup to augment multiple intermediate domains between the source and target domain. They train the source- and target-dominant models from the augmented domains which have complementary characteristics. Hoffman *et al.* [26] proposed a novel method which adapts representations at both the pixel- and feature-level, enforces cycle-consistency and leverages a task loss which does not require aligned pairs.

B. BIJECTIVE DEEP NETWORK

Statistical Machine Learning algorithms learn the structure of the dataset by placing the data into a parametric distribution $p(x;\theta)$. For a given data that is represented with distribution we can create new data from the prior distribution. Unfortunately, it takes a longer time to process using statistical methods. Among the generative models, flow based models learn the data distribution $p(x)$ by applying the log-likelihood. In general, flow based models try to learn a continuous, differentiable non-linear transformation into a simpler distribution. In RealNVP [42] and NICE [43], coupling layers were introduced by stacking a sequence of invertible bijective transformation functions. The bijective function computes the jacobian determinant in trivial way without losing the ability to learn complex non-linear transformations. Germain *et al.* [44] introduced a simpler way to calculate the jacobian determinant. The method presents autoregressive autoencoders that can estimate a relatable distribution.

Kingma and Dhariwal [45] presents a 1×1 convolution replacing a fixed permutation that prior methods use. This helps during the process of optimization since learning a permutation matrix is not continuous that is amenable to gradient ascent. Hoogeboom *et al.* [46] proposed an $n \times n$ convolution that is more flexible since it operates on both spatial and channel dimension. Moreover, in their method, the authors presented an emerging and invertible periodic convolution, which chained specific invertible autoregressive convolutions and used a Fourier transform to transfer data to the frequency domain.

III. THE PROPOSED METHOD

In this section, we present a novel deblurring method consisting of flow-based invertible modules with domain adaptation from a labeled source dataset to an unlabeled target dataset. Flow-based invertible frameworks are known for transforming distributions from an input to a latent space using a bijective function. In our work, a flow based network is trained on clean images from the source dataset. The network generalizes the complex distribution of the source dataset. We formulated the training in a way that the MPRNET [15] network reconstructs clean images from the source dataset and target dataset. For the source dataset, the loss is measured using the difference between the ground truth and the reconstructed image. For the target dataset, the reconstructed image is fed into the flow based invertible network to calculate the loss.

A. PROBLEM FORMULATION

Given the blurry input image $B_s \in \mathbb{R}^{HXWX3}$ from the source dataset, and a blurry input image $B_t \in \mathbb{R}^{HXWX3}$ from the target dataset, our proposed method predicts the desired deblurred images $I \in \mathbb{R}^{HXWX3}$. Let $\mathcal{F}$ be a non-linear function that employs the mapping from $B \subseteq \mathbb{R}^2$ to $I \subseteq \mathbb{R}^2$, i.e. $\mathcal{F} : B \rightarrow I$ where $B = B_s \cup B_t$.

We parameterize the non-linear function $\mathcal{F}$ by the proposed method with parameters θ_F. Generally, given a pair of blurred and sharp images with N training samples, i.e. $\mathcal{D}_s = (B_{si}, Y_i)_{i=1}^{N}$, and blurred images without sharp images with M training samples, i.e. $\mathcal{D}_t = (B_{ti})_{i=1}^{M}$, the framework could learn and generate the deblurred image. Specifically, the learning objective can be formulated as below:

$$\theta_F^* = \arg\min_{\theta_F} \Big[\mathbb{E}_{(B_s,Y)\in\mathcal{D}} \mathcal{L}_s(\mathcal{F}(B_s;\theta_F), Y) + \mathbb{E}_{(B_t)\in\mathcal{D}} \mathcal{L}_t(\mathcal{F}(B_t;\theta_F)) \Big] \quad (2)$$

where Y is the ground truth, $\mathcal{F}(B;\theta_F)$ is the predicted sharp image and $\mathbb{E}$ is the loss function between the generated image and the ground truth image. $\mathcal{L}_s$ and $\mathcal{L}_t$ are the objective losses defined on the source domain and target domain, respectively.

B. DEBLURRING NETWORK WITH SUPERVISED APPROACH

In the proposed framework, the encoder-decoder network that captures the contextual information of the source dataset is used for the supervised approach. The network reconstructs a deblurred image from a blurred source and the target dataset.

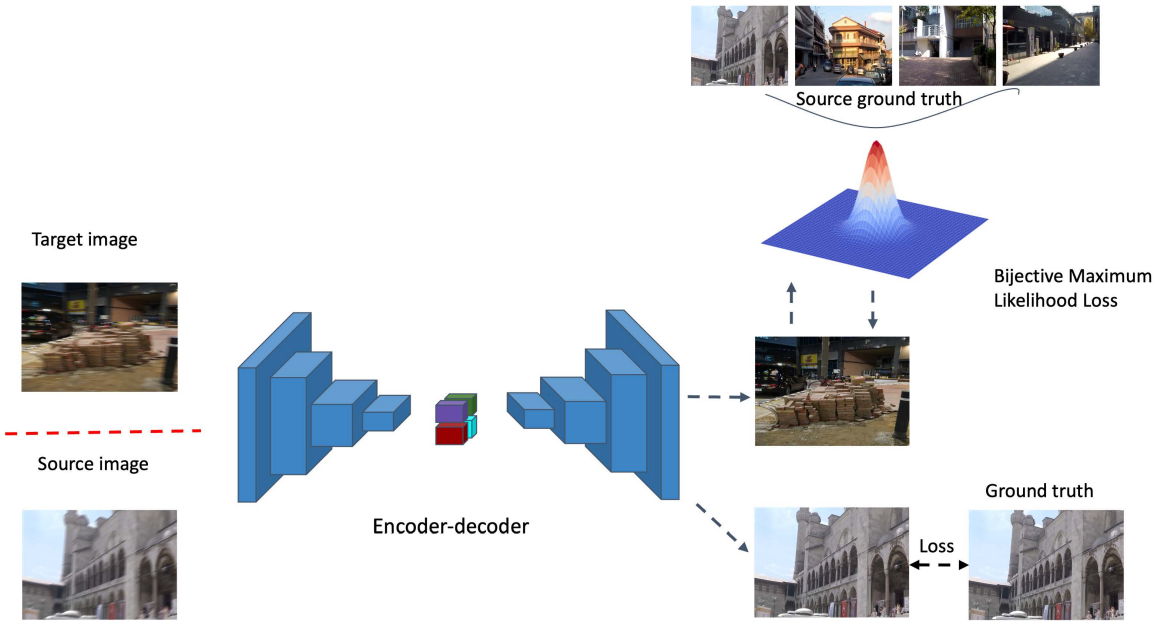


FIGURE 1. Proposed framework. The RGB image from source and target dataset forwarded to the encoder-decoder network sequentially. The network reconstruct deblurred images. The supervised loss is employed on the source training samples. The bijective maximum likelihood loss is computed on target training samples.

Meanwhile, since the ground-truth images of the source dataset are available, the sub-networks learn in a supervised manner from the source dataset. The encoder-decoder framework for the supervised network is chosen because it can generate semantically robust features.

The encoder and decoder of the network are built based on standard U-Net. It gradually maps the input to lower representations and slowly applies reverse mapping to reconstruct a new image at the same size as the original one. To extract features at each scale, we comprise channel attention blocks (CABs) which helps in enhancing the discriminative ability of the network. A supervised attention map (SAM) is plugged in every two stages to facilitate gradual learning of the module. Moreover, with the help of SAM we generate the attention maps that repress the less informative features and allow only the useful ones to pass to the next stage.

As shown in Figure 1, the supervised part of the network takes advantage of the labeled source dataset. During training, the network reconstructs sharp images from the target domain and source domain alternatively. Due to the high squared penalty that produces a blurry and over-smoothed visual effect, we do not use the standard mean squared error (MSE) loss function found in deblurring topics. We formulate the loss function for the supervised approach as follows:

$$\mathcal{L}_s = \mathcal{L}_{char}(B_s, Y) + \lambda \times \mathcal{L}_{edge}(B_s, Y) \tag{3}$$

where Y is the ground truth, $\mathcal{L}_{char}$ is the Charbonnier loss [47], and $\mathcal{L}_{edge}$ is the edge loss [48]. To balance the the weight of the two losses, the weight parameter λ is set to 0.05.

C. DEBLURRING NETWORK WITH UNSUPERVISED APPROACH

The flow-based network trained on the source domain complements the supervised deblurring network in a way that extracts deeper and expressive features and provides a broader interpretation of both the source and target domain datasets. As it is shown in Figure 1, the supervised network reconstructs the corresponding images from source and target datasets. For the source domain instances with a ground truth label, we employ a supervised approach to training using the loss function in equation 3. For the target domain, since the ground truth is unavailable we calculate the loss from the Bijective deep network. Initially, the Bijective deep network is trained on the ground truth of the source dataset where the model generalizes the distribution of the sharp image of the source domain.

Given a probability mass function of the distribution of an image from target domain denoted as $p_t(I)$, a reconstructed image from target domain denoted as $q_t(I)$, and the real distribution learned from ground-truth of the source domain $\mathcal{D}_s$, the efficiency of the function on the target dataset can be formulated as:

$$Eff = \int \mathcal{L}_t(p_t(I), q_s(I))p_t(I)dI \tag{4}$$

where the efficiency of the function on the target dataset is denoted as Eff. $\mathcal{L}_t(p_t(I), q_s(I))$ defines the distance between two distributions $p_t(I)$ and $q_s(I)$. For the target domain, the ground-truth is unavailable, so in equation 4 we replace $q_t(I)$ by $q_s(I)$. Although the distributions $q_t(I)$ and $q_s(I)$ may vary in image space, they have similar distributions in terms of

representing high resolution images. Thus, we adopted the prior knowledge acquired from the source domain where the labeled target data is not required for the computation. Among several candidates to estimate the divergence between the distribution $p_t(I)$ and $q_s(I)$, we choose $\mathcal{L}_t$ as the Kullback–Leibler divergence [49], and we can prove that the upper bound of *Eff* is as follows:

$$Eff \leq \underbrace{\mathbb{E}_I[-\log(q_s(I))]}_{\mathcal{L}_{llk}} \tag{5}$$

where $\mathcal{L}_{llk}$ is our Maximum Likelihood Loss. In the next section, we will further describe the learning process of $q_s(I)$ on the clear images of the source domain.

D. LEARNING BIJECTIVE MAPPING ON SOURCE DATASET

In this section, we present the learning process of the bijective network $\mathcal{G}$ on the set of clear images of the source domain. Let $\mathcal{G} : \mathbb{R}^{H\times W\times 3} \rightarrow \mathbb{R}^{H\times W\times 3}$ be the bijective network that maps the clear image Y into the latent space, i.e. $Z = \mathcal{G}(Y, \theta_G)$ (θ_G is the set of parameters of the deep network $\mathcal{G}$). By the change of variable theorem, the distribution $q_Y(Y)$ of clear images can be formed as follows:

$$q_Y(Y) = q_Z(\mathcal{G}(Y, \theta_G)) \det \left| \frac{\partial \mathcal{G}(Y)}{\partial Y} \right| \tag{6}$$

where q_Z is the prior distribution of the latent space, and $\det \left| \frac{\partial \mathcal{G}(Y,\theta_G)}{\partial Y} \right|$ is the Jacobian determinant of $\mathcal{G}(Y, \theta_G)$ with respect to Y. Then, the bijective network $\mathcal{G}$ is learned by minimizing the negative log-likelihood as follows:

$$\theta_G^* = \arg\min_{\theta_G} \mathbb{E}_Y - \left[\log q_Z(\mathcal{G}(Y, \theta_G)) + \log \det \left| \frac{\partial \mathcal{G}(Y)}{\partial Y} \right| \right] \tag{7}$$

Generally, there may have been various choices for the prior distribution $q_Z(\cdot)$. However, the ideal prior distribution should be easy in sampling and simple in the density estimation. Therefore, the Normal distribution is chosen as the prior distribution $q_Z(\cdot)$.

Additionally, to enhance the ability of the bijective network $\mathcal{G}$ so that $\mathcal{G}$ can model the complex structures of the image, we decompose $\mathcal{G}$ into multiple sub-functions, i.e. $\mathcal{G} = \mathcal{G}_1 \circ \mathcal{G}_2 \circ \ldots \circ \mathcal{G}_K$ (K is the number of sub-functions and $\circ$ denotes the compositional function). Each subfunction $\mathcal{G}_i$ is designed as a non-linear function. Several deep neural architectures can be adopted for $\mathcal{G}_i$ [42], [50].

IV. EXPERIMENTS

In this section, we firstly overview the datasets and implementation details in our experiments. Particularly, GoPro is used as a source dataset, while RealBlur_J, RealBlur_R, HIDE, and RED datasets are used as target datasets. Then, we discuss the quantitative and qualitative comparisons briefly described in the experimental subsection. We also present the empirical performance of the proposed method in the ablation study.

TABLE 1. Experimental results of our approach on two benchmarks: GoPro → RealBlur_R and GoPro → RealBlur_J.

Method	RealBlur-R		RealBlur-J	
	PSNR	SSIM	PSNR	SSIM
Xu *et al* [54]	34.46	0.937	27.14	0.830
DeblurGAN [20]	33.79	0.903	27.97	0.834
SRN *et al* [3]	35.66	0.947	28.56	0.867
DeblurGAN-v2 *et al* [21]	35.26	0.944	28.70	0.866
DMPHN *et al* [55]	35.70	0.948	28.42	0.860
Zhang *et al* [4]	35.48	0.947	27.80	0.847
Pan *et al* [56]	34.01	0.916	27.22	0.790
Nah*et al* [2]	32.51	0.841	27.87	0.827
Hu *et al* [57]	33.67	0.916	26.41	0.803
Baseline(MPRNET) [48]	35.99	0.952	28.70	0.873
Our method	**37.31**	**0.972**	**30.76**	**0.922**

TABLE 2. Experimental results of our approach on the benchmark of GoPro → HIDE.

Method	HIDE	
	PSNR	SSIM
DeblurGAN [20]	24.51	0.871
SRN *et al* [3]	28.36	0.915
DeblurGAN-v2 *et al* [21]	26.61	0.875
DMPHN *et al* [55]	29.09	0.924
Gao *et al* [58]	29.11	0.913
Nah *et al* [51]	25.73	0.874
Suin *et al* [59]	29.98	0.930
Baseline(MPRNET) [48]	30.96	0.939
Our method	**32.14**	**0.953**

A. DATASET

In this work, we perform extensive experiments on several different datasets.

1) GoPro DATASET

is captured from a high-speed camera and contains pairs of blurry images and corresponding ground truth sharp images. 2,103 images with the size of 1,280 × 720 are provided for training, and 1,111 test images with the same size are provided by [51]. To avoid the problem of overfitting, various data augmentation techniques are performed. With respect to geometric transformations, patches are randomly flipped (rotated by 90 degrees) horizontally and vertically. RGB channels are randomly permuted, with respect to color.

2) HIDE DATASET

The blurred images are synthesized by averaging 11 sequential frames from a video recorded with 240 fps camera and the middle frame from the video is taken as the sharp image [52]. The dataset is split into 6,397 training and 2,025 testing images.

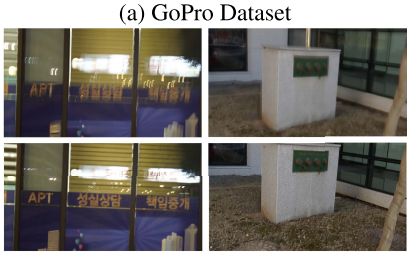
3) REAL BLUR DATASET

This dataset [53] consists of 182 scenes from *RealBlur-R* and *RealBlur-J* as the training set, while the other 50 scenes are used as our test set. Each training set includes 3,758 image pairs containing 182 reference pairs, while the test set

TABLE 3. Ablation study on different domain adaptation settings.

Domain Adaptation	Datasets	PSNR		
		RealBlur_J	RealBlur_R	HIDE
✗	RealBlur_R(Source) → RealBlur_R(Target)	27.31	-	28.42
✗	RealBlur_J(Source) → RealBlur_J(Target)	-	30.18	27.84
✓	GoPro(Source) → RealBlur_J(Target)	30.76	37.31	32.14

(a) GoPro Dataset



(b) RealBlur Dataset

(c) HIDE Dataset

FIGURE 2. Examples of deblurring datasets including blurry images (top row) and corresponding clear images (bottom row).

consists of 980 image pairs without any reference pairs. Here, we include reference pairs in the training set so the network can map to the sharp images.

B. TRAINING DETAILS

In these experiments, we adopt an encoder and decoder network for supervised learning as part of the framework from the standard UNet [60]. On the network, we add channel attention blocks (CABs) [61] to extract and represent the low-frequency information that is challenging for other standard CNN based networks. For the decoder part of the network, we used bilinear upsampling then we add convolution to reconstruct the deblurred image. For the bijective network, we incorporate the multi-scale architecture structure. Each scale of the network is built with ActNorm, Invertible 1×1 Convolution, and Affine Coupling Layer [45]. The model is implemented in PyTorch [62] and trained on NVIDIA Quadro P8000 GPUs with 48GB of VRAM.

As shown in Fig 2, we have trained model with patch sizes of 64 and 128 and different learning rates. We crop part of an image randomly from the set of input datasets and fed it to the network. We couldn't train our model using a larger patch size due to the high computational demand. We achieved the best results using a patch size of 64 and a learning rate of 1e-4. The different PSNR & SSIM values for the different learning rates is demonstrated in Table 4 and Table 5.

TABLE 4. PSNR and SSIM values for the network trained on GoPro [51] as the source domain and RealBlur_J as the target domain and tested on RealBlur_J, RealBlur_R [53], and HIDE [52] testing dataset. The input images are cropped with 64 × 64 size with different learning rates.

LR	RealBlur_J		RealBlur_R		HIDE	
	PSNR	SSIM	PSNR	SSIM	PSNR	SSIM
1e-4	31.02	0.93	36.52	0.96	31.78	0.83
2.5e-4	30.76	0.92	37.31	0.97	32.14	0.95
5e-4	27.65	0.77	34.26	0.83	28.24	0.82

TABLE 5. PSNR and SSIM values for the network trained on GoPro [2] as the source domain and RealBlur_J as the target domain and tested on RealBlur_J, RealBlur_R [53], and HIDE testing dataset [52]. The input images are cropped with 128 × 128 size with different learning rates (LR).

LR	RealBlur_J		RealBlur_R		HIDE	
	PSNR	SSIM	PSNR	SSIM	PSNR	SSIM
1e-4	30.38	0.91	35.14	0.89	29.21	0.85
2.5e-4	29.67	0.87	35.42	0.92	29.84	0.88
5e-4	26.42	0.75	33.28	0.84	27.81	0.84

C. EXPERIMENTAL RESULTS

1) QUANTITATIVE COMPARISONS

We compare our proposed method with several state-of-art methods, such as Hu *et al.* [57], Nah *et al.* [51], DeblurGAN [20] and MPRNET [48]. Compared to these methods, we achieved a higher PSNR and SSIM value. It is worth noticing that our network is trained on GoPro as a source dataset. For the first experimental setting, we use the model trained on GoPro to test on the testing dataset of RealBlur_J, RealBlur_R, HIDE, and REDS. In the second experimental setting, we train on both source and target domain and test on the dataset listed above.

Table 1 and Table 2 shows that our method achieved significantly better scores on publicly available datasets, such as RealBlur-R, RealBlur-J and HIDE datasets. For example compared to the baseline MPRNET [15], we obtain a performance gain of 1.32 dB.

2) QUALITATIVE COMPARISONS

Fig. 3 depicts some deblurring results from the testsets of RealBlur_J, RealBlur_R and HIDE. The qualitative results of prior works suffer from incomplete deblurring and poor generalization while our proposed method restores more

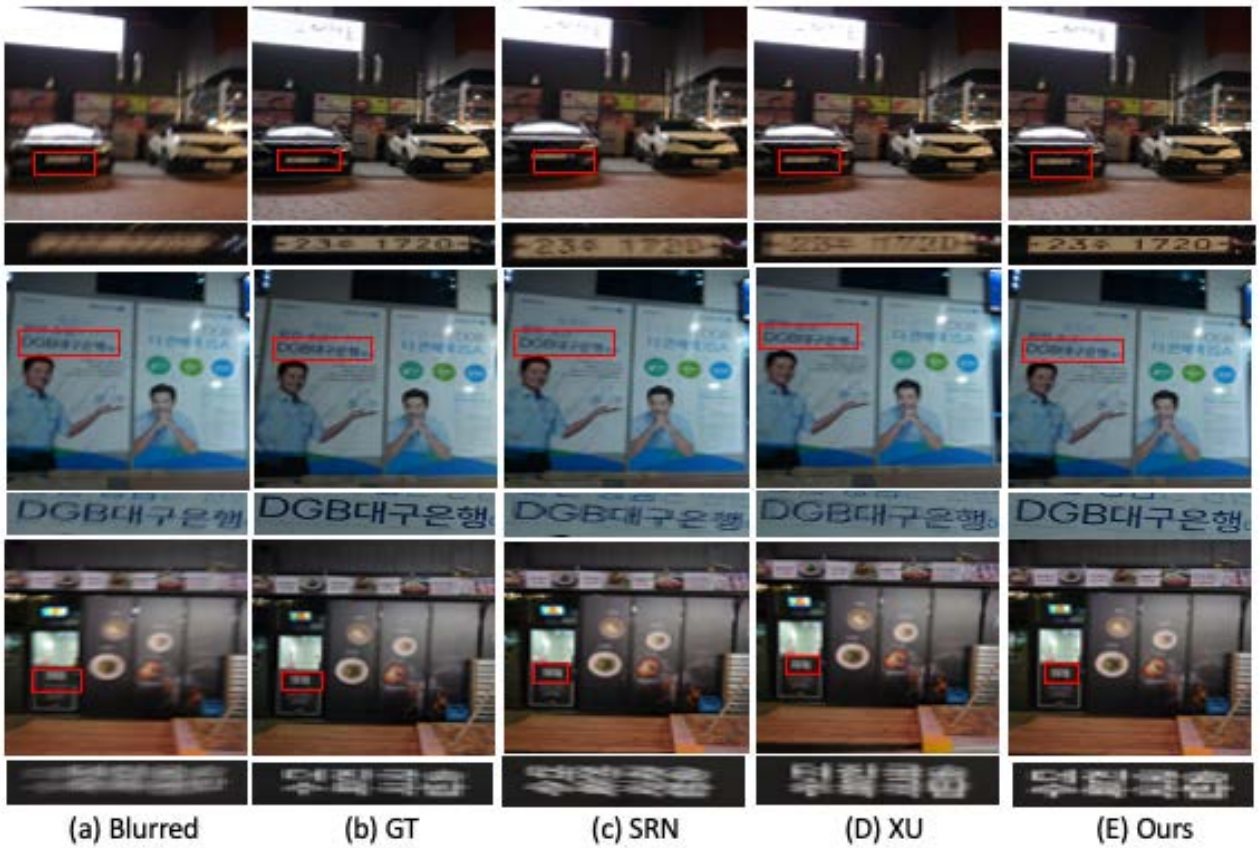


FIGURE 3. Qualitative comparison for image deblurring on RealBlur_J dataset. Our proposed method shows better restored detailed compared to the state-of-the-art methods.

perceptually-faithful and sharper images. We believe the improvement of the images happened because the model effectively learned the necessary cues from the distribution of the source dataset using our proposed domain adaptation framework.

D. ABLATION STUDY

In this section, we present the ablation experiments to analyze the contribution of domain adaptation. We performed the evaluation on RealBlur_J, RealBlur_R [53] and HIDE [52] datasets where the model was trained on images with patch size of 64 × 64 for 40,000 iterations. The results are shown in Table 3.

1) WITHOUT DOMAIN ADAPTATION

It is evident from Table 3 that the model did not perform well without domain adaptation. This demonstrates the significant increase in performance given by the model when domain adaptation was used.

2) CHOICES OF SOURCE AND TARGET DATASETS

In our tests without domain adaptation, we applied the same source and target dataset that were used while training and made sure that we did not test on the same dataset used for training.Consequently, we did not test for RealBlur_R and RealBlur_J when we trained the model.

We demonstrate the efficiency of the proposed method by removing domain adaptation i.e., both the source and target are the same datasets as shown in Table 3. This table shows a considerable change(↓) in PSNR from 30.76 to 27.31 when trained on RealBlur_R dataset and 37.31 to 30.18 when trained on RealBlur_J dataset. When tested on HIDE, the PSNR went from 32.14 to 27.84 & 28.42 for the models trained on RealBlur_J and RealBlur_R, respectively.

V. CONCLUSION

In this work, we have proposed a novel domain adaptation approach to image deblurring. The bijective network learned on the source dataset is introduced to model the complex and diverse structure of the clear images. In addition to the supervised learning on the source dataset, we further propose a new maximum likelihood to learn the image deblurring model on the target dataset in an unsupervised manner. Through extensive experiments on GoPro → RealBlur_J, GoPro → RealBlur_R, and GoPro → HIDE, our proposed method outperforms the prior methods, establishing new state-of-the-art benchmarks for image deblurring.

REFERENCES

[1] S. Su, M. Delbracio, J. Wang, G. Sapiro, W. Heidrich, and O. Wang, "Deep video deblurring for hand-held cameras," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jul. 2017, pp. 1279–1288.

[2] S. Nah, T. H. Kim, and K. M. Lee, "Deep multi-scale convolutional neural network for dynamic scene deblurring," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jul. 2017, pp. 3883–3891.

[3] X. Tao, H. Gao, X. Shen, J. Wang, and J. Jia, "Scale-recurrent network for deep image deblurring," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit.*, Jun. 2018, pp. 8174–8182.

[4] J. Zhang, J. Pan, J. Ren, Y. Song, L. Bao, R. W. H. Lau, and M.-H. Yang, "Dynamic scene deblurring using spatially variant recurrent neural networks," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit.*, Jun. 2018, pp. 2521–2529.

[5] A. Gupta, N. Joshi, C. L. Zitnick, M. Cohen, and B. Curless, "Single image deblurring using motion density functions," in *Proc. Eur. Conf. Comput. Vis.* Cham, Switzerland: Springer, 2010, pp. 171–184.

[6] O. Whyte, J. Sivic, A. Zisserman, and J. Ponce, "Non-uniform deblurring for shaken images," *Int. J. Comput. Vis.*, vol. 98, no. 2, pp. 168–186, 2012.

[7] N. Wiener, *Extrapolation, Interpolation, and Smoothing of Stationary Time Series: With Engineering Applications*, vol. 8. Cambridge, MA, USA: MIT Press, 1964.

[8] W. H. Richardson, "Bayesian-based iterative method of image restoration," *J. Opt. Soc. Amer.*, vol. 62, no. 1, pp. 55–59, Jan. 1972.

[9] D. Krishnan and R. Fergus, "Fast image deconvolution using hyper-Laplacian priors," in *Proc. Adv. Neural Inf. Process. Syst.*, vol. 22, 2009, pp. 1033–1041.

[10] A. Levin, R. Fergus, F. Durand, and W. T. Freeman, "Image and depth from a conventional camera with a coded aperture," *ACM Trans. Graph.*, vol. 26, no. 3, p. 70, 2007.

[11] D. Zoran and Y. Weiss, "From learning models of natural image patches to whole image restoration," in *Proc. Int. Conf. Comput. Vis.*, Nov. 2011, pp. 479–486.

[12] L. Bar, N. Kiryati, and N. Sochen, "Image deblurring in the presence of impulsive noise," *Int. J. Comput. Vis.*, vol. 70, no. 3, pp. 279–298, 2006.

[13] S. Cho, J. Wang, and S. Lee, "Handling outliers in non-blind image deconvolution," in *Proc. Int. Conf. Comput. Vis.*, Nov. 2011, pp. 495–502.

[14] O. Whyte, J. Sivic, and A. Zisserman, "Deblurring shaken and partially saturated images," *Int. J. Comput. Vis.*, vol. 110, no. 2, pp. 185–201, 2014.

[15] L. Xu, J. S. Ren, C. Liu, and J. Jia, "Deep convolutional neural network for image deconvolution," in *Proc. Adv. Neural Inf. Process. Syst.*, vol. 27, 2014, pp. 1790–1798.

[16] J. Sun, W. Cao, Z. Xu, and J. Ponce, "Learning a convolutional neural network for non-uniform motion blur removal," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2015, pp. 769–777.

[17] L. Chen, F. Fang, J. Zhang, J. Liu, and G. Zhang, "OID: Outlier identifying and discarding in blind image deblurring," in *Proc. 16th Eur. Conf. Comput. Vis. (ECCV)*, Glasgow, U.K., Aug. 2020, pp. 598–613.

[18] J. Pan, Z. Lin, Z. Su, and M.-H. Yang, "Robust kernel estimation with outliers handling for image deblurring," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2016, pp. 2800–2808.

[19] J. Dong, J. Pan, Z. Su, and M.-H. Yang, "Blind image deblurring with outlier handling," in *Proc. IEEE Int. Conf. Comput. Vis. (ICCV)*, Oct. 2017, pp. 2478–2486.

[20] O. Kupyn, V. Budzan, M. Mykhailych, D. Mishkin, and J. Matas, "DeblurGAN: Blind motion deblurring using conditional adversarial networks," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit.*, Jun. 2018, pp. 8183–8192.

[21] O. Kupyn, T. Martyniuk, J. Wu, and Z. Wang, "DeblurGAN-v2: Deblurring (orders-of-magnitude) faster and better," in *Proc. IEEE/CVF Int. Conf. Comput. Vis. (ICCV)*, Oct. 2019, pp. 8878–8887.

[22] A. Chakrabarti, "A neural approach to blind motion deblurring," in *Proc. Eur. Conf. Comput. Vis.* Cham, Switzerland: Springer, 2016, pp. 221–235.

[23] D. Gong, J. Yang, L. Liu, Y. Zhang, I. Reid, C. Shen, A. Van Den Hengel, and Q. Shi, "From motion blur to motion flow: A deep learning solution for removing heterogeneous motion blur," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jul. 2017, pp. 2319–2328.

[24] Y. Chen, W. Li, and L. V. Gool, "ROAD: Reality oriented adaptation for semantic segmentation of urban scenes," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit.*, Jun. 2018.

[25] Y.-H. Chen, W.-Y. Chen, Y.-T. Chen, B.-C. Tsai, Y.-C.-F. Wang, and M. Sun, "No more discrimination: Cross city adaptation of road scene segmenters," in *Proc. IEEE Int. Conf. Comput. Vis. (ICCV)*, Oct. 2017, pp. 1992–2001.

[26] J. Hoffman, E. Tzeng, T. Park, J.-Y. Zhu, P. Isola, K. Saenko, A. Efros, and T. Darrell, "CyCADA: Cycle-consistent adversarial domain adaptation," in *Proc. ICML*, 2018, pp. 1989–1998.

[27] J. Hoffman, D. Wang, F. Yu, and T. Darrell, "FCNs in the wild: Pixel-level adversarial and constraint-based adaptation," 2016, *arXiv:1612.02649*.

[28] W. Hong, Z. Wang, M. Yang, and J. Yuan, "Conditional generative adversarial network for structured domain adaptation," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit.*, Jun. 2018.

[29] J. Wang and J. Jiang, "Learning across tasks for zero-shot domain adaptation from a single source domain," *IEEE Trans. Pattern Anal. Mach. Intell.*, early access, Jun. 14, 2021, doi: 10.1109/TPAMI.2021.3088859.

[30] J. Wang, M.-M. Cheng, and J. Jiang, "Domain shift preservation for zero-shot domain adaptation," *IEEE Trans. Image Process.*, vol. 30, pp. 5505–5517, 2021.

[31] Y. Zou, Z. Yu, B. V. Kumar, and J. Wang, "Unsupervised domain adaptation for semantic segmentation via class-balanced self-training," in *Proc. ECCV*, 2018.

[32] Z. Murez, S. Kolouri, D. Kriegman, R. Ramamoorthi, and K. Kim, "Image to image translation for domain adaptation," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit.*, Jun. 2018, pp. 4500–4509.

[33] F. Pan, I. Shin, F. Rameau, S. Lee, and I. S. Kweon, "Unsupervised intra-domain adaptation for semantic segmentation through self-supervision," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2020.

[34] J.-Y. Zhu, T. Park, P. Isola, and A. A. Efros, "Unpaired image-to-image translation using cycle-consistent adversarial networks," in *Proc. IEEE Int. Conf. Comput. Vis. (ICCV)*, Oct. 2017.

[35] Y. Ganin and V. Lempitsky, "Unsupervised domain adaptation by back-propagation," in *Proc. ICML*, 2015, pp. 1180–1189.

[36] M. Long, Y. Cao, J. Wang, and M. I. Jordan, "Learning transferable features with deep adaptation networks," in *Proc. ICML*, 2015, pp. 97–105.

[37] E. Tzeng, J. Hoffman, K. Saenko, and T. Darrell, "Adversarial discriminative domain adaptation," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jul. 2017.

[38] A. Sharma, T. Kalluri, and M. Chandraker, "Instance level affinity-based transfer for unsupervised domain adaptation," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2021, pp. 5361–5371.

[39] J. Wang and J. Jiang, "Adversarial learning for zero-shot domain adaptation," in *Proc. Eur. Conf. Comput. Vis.* Cham, Switzerland: Springer, 2020, pp. 329–344.

[40] J. Wang and J. Jiang, "Conditional coupled generative adversarial networks for zero-shot domain adaptation," in *Proc. IEEE/CVF Int. Conf. Comput. Vis. (ICCV)*, Oct. 2019, pp. 3375–3384.

[41] J. Na, H. Jung, H. J. Chang, and W. Hwang, "FixBi: Bridging domain spaces for unsupervised domain adaptation," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2021, pp. 1094–1103.

[42] L. Dinh, J. Sohl-Dickstein, and S. Bengio, "Density estimation using real NVP," 2017, *arXiv:1605.08803*.

[43] L. Dinh, D. Krueger, and Y. Bengio, "NICE: Non-linear independent components estimation," 2014, *arXiv:1410.8516*.

[44] M. Germain, K. Gregor, I. Murray, and H. Larochelle, "MADE: Masked autoencoder for distribution estimation," in *Proc. Int. Conf. Mach. Learn.* 2015, pp. 881–889.

[45] D. P. Kingma and P. Dhariwal, "Glow: Generative flow with invertible 1×1 convolutions," 2018, *arXiv:1807.03039*.

[46] E. Hoogeboom, R. V. D. Berg, and M. Welling, "Emerging convolutions for generative normalizing flows," in *Proc. Int. Conf. Mach. Learn.* 2019, pp. 2771–2780.

[47] P. Charbonnier, L. Blanc-Féraud, G. Aubert, and M. Barlaud, "Two deterministic half-quadratic regularization algorithms for computed imaging," in *Proc. IEEE Int. Conf. Image Process. (ICIP)*, vol. 2, Sep. 1994, pp. 168–172.

[48] S. W. Zamir, A. Arora, S. Khan, M. Hayat, F. S. Khan, M.-H. Yang, and L. Shao, "Multi-stage progressive image restoration," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit.*, Jun. 2021, pp. 14821–14831.

[49] T.-D. Truong, C. N. Duong, N. Le, S. L. Phung, C. Rainwater, and K. Luu, "BiMaL: Bijective maximum likelihood approach to domain adaptation in semantic scene segmentation," in *Proc. IEEE/CVF Int. Conf. Comput. Vis. (ICCV)*, Oct. 2021.

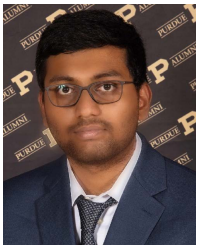
[50] D. P. Kingma and P. Dhariwal, "Glow: Generative flow with invertible 1×1 convolutions," in *Proc. NIPS*, S. Bengio, H. Wallach, H. Larochelle, K. Grauman, N. Cesa-Bianchi, and R. Garnett, Eds., 2018.

[51] S. Nah, T. H. Kim, and K. M. Lee, "Deep multi-scale convolutional neural network for dynamic scene deblurring," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Jul. 2018, pp. 3883–3891.

[52] Z. Shen, W. Wang, X. Lu, J. Shen, H. Ling, T. Xu, and L. Shao, "Human-aware motion deblurring," in *Proc. IEEE/CVF Int. Conf. Comput. Vis. (ICCV)*, Oct. 2019.

[53] J. Rim, H. Lee, J. Won, and S. Cho, "Real-world blur dataset for learning and benchmarking deblurring algorithms," in *Proc. Eur. Conf. Comput. Vis. (ECCV)*, 2020.
[54] L. Xu, S. Zheng, and J. Jia, "Unnatural L_0 sparse representation for natural image deblurring," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Jun. 2013, pp. 1107–1114.
[55] H. Zhang, Y. Dai, H. Li, and P. Koniusz, "Deep stacked hierarchical multi-patch network for image deblurring," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2019, pp. 5978–5986.
[56] J. Pan, D. Sun, H. Pfister, and M.-H. Yang, "Blind image deblurring using dark channel prior," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2016, pp. 1628–1636.
[57] Z. Hu, S. Cho, J. Wang, and M.-H. Yang, "Deblurring low-light images with light streaks," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Jun. 2014, pp. 3382–3389.
[58] H. Gao, X. Tao, X. Shen, and J. Jia, "Dynamic scene deblurring with parameter selective sharing and nested skip connections," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2019, pp. 3848–3856.
[59] M. Suin, K. Purohit, and A. N. Rajagopalan, "Spatially-attentive patch-hierarchical network for adaptive motion deblurring," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2020, pp. 3606–3615.
[60] O. Ronneberger, P. Fischer, and T. Brox, "U-Net: Convolutional networks for biomedical image segmentation," in *Proc. Int. Conf. Med. Image Comput. Comput.-Assist. Intervent.* Cham, Switzerland: Springer, 2015, pp. 234–241.
[61] Y. Zhang, K. Li, K. Li, L. Wang, B. Zhong, and Y. Fu, "Image super-resolution using very deep residual channel attention networks," in *Proc. Eur. Conf. Comput. Vis. (ECCV)*, 2018, pp. 286–301.
[62] J. Fu, J. Liu, H. Tian, Y. Li, Y. Bao, Z. Fang, and H. Lu, "Dual attention network for scene segmentation," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2019, pp. 3146–3154.

IBSA JALATA received the B.Sc. degree from Addis Ababa University, Ethiopia, and the M.Sc. degree from Chonbuk National University, South Korea. He is currently pursuing the Ph.D. degree with the University of Arkansas, Arkansas. He had been working as a Research Assistance with the Media and Communication Laboratory, Chonbuk National University. He has been working as a Lecturer with Adama Science and Technology University, Ethiopia. His research interests include activity recognition, action localization, computer vision, face detection, and recognition, and biomedical image processing.



NAGA VENKATA SAI RAVITEJA CHAPPA received the B.Tech. degree in electronics and communication engineering from MVGR College of Engineering, JNTU, Kakinada, India, in 2018, and the master's degree in computer engineering (electrical and computer engineering) from Purdue University, Indianapolis. He is currently pursuing the Ph.D. degree with the Department of Computer Science and Computer Engineering, University of Arkansas. His research interests include computer vision applications, ML on edge-devices, wearable and ubiquitous computing, and human activity recognition based on sensors.

THANH-DAT TRUONG (Graduate Student Member, IEEE) received the B.Sc. degree in computer science from Honors Program, Faculty of Information Technology, University of Science, VNU, in 2019. He is currently pursuing the Ph.D. degree with the Department of Computer Science and Computer Engineering, University of Arkansas. He was a Research Intern with the Coordinated Laboratory Science, University of Illinois Urbana-Champaign. He worked as a Research Assistant with the Artificial Intelligence Laboratory, University of Science, VNU, when he was an undergraduate student. His papers appear at top tier conferences such as IEEE/CVF Computer Vision and Pattern Recognition (CVPR), IEEE/CVF International Conference on Computer Vision (ICCV), and Canadian Conference on Computer and Robot Vision. His research interests include face recognition, face detection, domain adaptation, deep generative model, and adversarial learning. He is also a Reviewer of top-tier journals and conferences including IEEE TRANSACTION ON PATTERN ANALYSIS AND MACHINE INTELLIGENCE (TPAMI), *Journal of Computers, Environment and Urban Systems*, IEEE ACCESS, IEEE/CVF Computer Vision and Pattern Recognition, Winter Conference on Applications of Computer Vision (WACV), and International Conference on Pattern Recognition (ICPR).

PIERCE HELTON is currently pursuing the bachelor's degree with the Department of Computer Science and Computer Engineering, University of Arkansas. He is expected to graduate in fall of 2022. He is also working with the Computer Vision and Image Understanding Laboratory, University of Arkansas. He has plans to work in industry and potentially pursue the M.S. degree in computer science after earning his B.S. His research interests include machine learning, deep learning, and computer vision.

CHASE RAINWATER received the Ph.D. degree in industrial and systems engineering from the University of Florida. He is currently a Professor with the Department of Industrial Engineering (INEG), University of Arkansas. His primary research interests include large-scale optimization, homeland security, and supply chain logistics. In addition, he is active in the areas of reliability, cybersecurity, and healthcare planning. He is a member of IIE, INFORMS, and MORS. His teaching honors include being named INEG Outstanding Teacher, in 2012 and 2014.

KHOA LUU (Member, IEEE) was a Research Project Director with the Cylab Biometrics Center, Carnegie Mellon University (CMU), USA. He is currently an Assistant Professor and the Director of the Computer Vision and Image Understanding (CVIU) Laboratory, Department of Computer Science and Computer Engineering, University of Arkansas, Fayetteville, USA. He has received four patents. He has coauthored more than 120 papers in conferences and journals. His research interests include face recognition, biometrics, video and image understanding, autonomous car driving, robot vision, computer vision, machine learning, and quantum machine learning.

Dr. Luu is a PC Member of AAAI and ICPRAI, in 2020 and 2022, respectively. He received two best paper awards. He was a Vice-Chair of Montreal Chapter IEEE SMCS, Canada, from September 2009 to March 2011. He is a Co-organizer and the Chair of CVPR Precognition Workshop, in 2019, 2020, 2021, and 2022; MICCAI Workshop, in 2019, 2020; and ICCV Workshop, in 2021. He is serving as an Associate Editor of IEEE ACCESS journal. He is currently a Reviewer for several top-tier conferences and journals, such as CVPR, ICCV, ECCV, NeurIPS, ICLR, *FG*, BTAS, IEEE TRANSACTIONS ON PATTERN ANALYSIS AND MACHINE INTELLIGENCE, IEEE TRANSACTIONS ON IMAGE PROCESSING, *Journal of Pattern Recognition*, *Journal of Image and Vision Computing*, *Journal of Signal Processing*, *Journal of Intelligence Review*, IEEE ACCESS, and IEEE TRANSACTIONS.