MobileAR-GAN: MobileNet based Efficient Attentive Recurrent Generative Adversarial Network for Infrared to Visual Transformations

Nand Kumar Yadav, *Student Member, IEEE,* Satish Kumar Singh, *Senior Member, IEEE,* and Shiv Ram Dubey, *Senior Member, IEEE*

***Abstract*—Deep learning has recently shown outstanding performance for different applications, including image-to-image translation by Generative Adversarial Networks (GANs). However, GAN models are very complex as build with multiple deep networks and requires huge computational resources for the training as well as inference. Hence, the real-time deployment of GAN models is not feasible at present. In this paper, we propose MobileNet based Efficient Attentive Recurrent Generative Adversarial Network (MobileAR-GAN) for resource-constrained Infrared to Visual translation. The proposed model utilizes the light-weight MobileNet and enhances its capacity using the Attention and Recurrent modules, leading to an efficient yet effective model. We consider the Infrared to Visible Image Translation task to validate the efficiency and performance of the proposed model. The proposed MobileAR-GAN outperforms most of the existing GAN models in terms of both the efficiency as well as the quality of the generated images. We also test the MobileAR-GAN model over the resource-limited Jetson TX2 board with a very compelling results. The proposed model shows promising results over state-of-the-art methods. Compared to light-weight models such as Pix2pix and GAN-Compression methods, an average improvement gain of 19.19% and 17.05% is observed by the proposed model in terms of SSIM metric. It is observed that the proposed model can be deployed on edge devices to transform the images taken at night time using infrared camera to the corresponding visible images with satisfactory performance.**

***Index Terms*—GAN, Attention GAN, Transformation, MobileNet, Attention networks, Deep Learning, Computer vision, CNN.**

I. Introduction

The Near Infrared (NIR) camera is used widely for night vision display. The NIR to Visual RGB domain transformation is a widely accepted task for surveillance-based systems and night vision-equipped devices. The NIR is cost-effective technology and uses near-infrared rays with low-cost LEDs [1]. In contrast, the NIR to visible domain transformation takes place over night-day scenes transformation, and the limited computational resources draw the boundary for translation by using fewer parameters. It is a computer vision problem related to image generation and synthesis. Computer vision allows to deal with several tasks associated with image-based applications like restoration [2], image synthesis [3], transfer learning [4], activity detection [5], pose estimation [6], and many more. We use GAN to generate realistic fair samples in visible domain from NIR domain. GAN refers to the combination of two CNN networks with adversarial approaches. These networks oppose each other and learn in an adversarial fashion. Computer vision problems like image colorization, segmentation [7], and segmented map to real scene image generation [8] can be solved efficiently using GANs. Deep learning methods [9], [10], [11] categorize into supervised learning and unsupervised learning. Unsupervised methods become advantageous over supervised methods in terms of fewer extensive manual work. GANs with unsupervised learning approaches gained much attention recently. The advantage of using the GAN-based method is, it's unsupervised or semi-supervised nature. GAN shows better transformations compared to other machine learning and deep learning-based methods. Various tasks like image colorization, image super resolutions, and image style transfer used GAN for effective results. Various image-to-image transformation methods have already been proposed for inter-domain transformation [12], [10], [13], [14]. However, the existing GAN models are quite extensive in terms of the computational resources. Hence, it is required to develop the light-weight GAN models for real-time applications. The proposed method is applicable in various industry-based methods in measurement and control, such as surveillance-based security systems, cross-border defense systems, etc. as depicted in Fig. 1. Moreover, due to the lightweight nature of the proposed model, it can be integrated with self-driving cars to tackle the low light visibility problem as well as robotics devices for obstacle avoidance and object grasping in low light conditions.

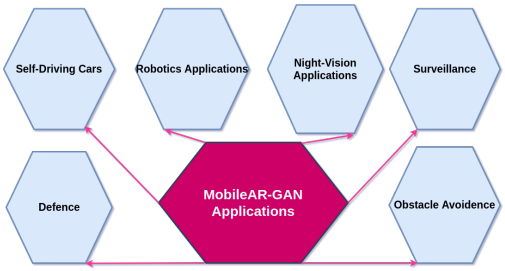

Fig. 1: Different possible application scenarios of the proposed light-weight MobileAR-GAN model in real-life.

Motivated from the need of light-weight GAN models, we

N.K. Yadav, S.K. Singh and S.R. Dubey are with the Computer Vision and Biometrics Laboratory at Indian Institute of Information Technology, Allahabad, Prayagraj-211015, U.P., India (email: nandkmyadav@gmail.com, sk.singh@iiita.ac.in, srdubey@iiita.ac.in).

The final paper is availabe at: https://ieeexplore.ieee.org/document/9754574.

make following contributions through this paper:

- We introduce a MobileNet based Efficient Attentive Recurrent Generative Adversarial Network (MobileAR-GAN) for image-to-image translation over limited computational devices.
- Our proposed MobileAR-GAN boosts the feature selection of GAN by combining the attention network in Mobile-net based inverted residual block with enhanced attention to problem-specific features.
- The proposed Mobile-GAN's learning space narrowed down to optimal learning due to the attention guidance and the recurrent network's objective.
- The proposed MobileAR-GAN model is tested using Nvidia-Jetson board with very satisfactory performance in real time for NIR to Visible transformation.

The remaining paper is arranged as: a literature survey is presented in section II; the proposed MobileAR-GAN architecture is described in section III; network architecture shown in section III-B; the experimental setting and result analysis in section IV; Impact of various losses and architectural modules in section V; and in the end, the concluding remarks highlighted in section VI.

II. RELATED WORK

Generative Adversarial Networks (GAN) [15] have emerged as the latest unsupervised method in deep learning, which is competent enough for generating new samples within the limits of training data distribution. Various promising GAN methods have been proposed for image-to-image translations [10], [16], [17], which can handle the new sample generation with pre-specified features. ConditionalGAN [18] proposed generating new samples with some pre-specified features by embedding a conditional vector for generated image. For multi-domain image-to-image translation, Coupled Generative Adversarial Network (CoGAN) [11] was proposed, which learned the joint distribution of two datasets by using the marginal distribution of two different datasets. Based on ConditionalGAN, Pix2pix [9] proposed for conditional image transformation, conditional embedding helps the generator network to generate samples with some pre-specified properties in image generation. Pix2pix required paired images for conditional embedding in the image translation task. To reduce the extensive manual work for paired data collection, CycleGAN [10] proposed an unpaired image-to-image translation technique with cycle-consistency loss (cyclic loss). Simultaneously, DualGAN [13] proposed for the unpaired image translation with reconstruction loss very similar to CycleGAN. In recent literature, it is observed that the utilization of attention mechanism with GAN can reduce the training time leading to faster convergence by exploiting the important regions with higher priority, such as Self-attention GAN [19], Spatial Attention [20], and Light-weight Attention [21]. The attention-based method boosts the CNN performance and learns long-range dependencies by using relatively low computational cost than traditional CNN and attentive to a specific region. Recently proposed GAN methods used attention-based networks for image-to-image translation techniques and found

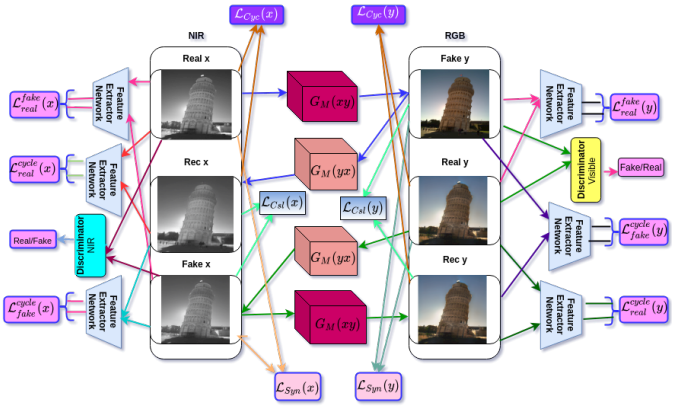

Fig. 2: Proposed MobileAR-GAN architecture.

improved results with realistic generated samples. Mejjati et al. [16] presented an attentive Generative Adversarial Network (AGGAN) for image-to-image translation, which draws the attention maps by utilizing generators output and secures the background details by employing inverse attention map. The AGGAN reported more promising image translation results than earlier non-attentive models. Tang et al. also presented the new attention-based image translation method known as AttentionGAN [17], which utilizes ResNet-based architecture without additional network integration to calculate the attention mask. Convergence results of these methods show that attention-based networks converge faster within fewer training epochs than traditional networks. PCSGAN [22] proposed by Babu et al. by using image-to-image translation for NIR-to-Visual transformation. Infrared Colorization Using Deep Convolutional Neural Networks is proposed by Matthias Limmer et al. [23]. NIR Image Colorization using SPADE Generator and Grayscale Approximated Self-Reconstruction is done in [24]. Hickman et al. [25] proposed color fusion for NIR images using the RGBN fusion scheme, which used color Weight Map (CWM) to enhance features' visibility by controlling color distortion within the scene. However, these GAN models suffer from the high computational requirement and limited performance in challenging scenario such as visible image synthesis from NIR images.

III. PROPOSED MOBILEAR-GAN MODEL

This section presents a novel MobileNet based Efficient Attentive Recurrent Generative Adversarial Network (MobileAR-GAN) for visual image synthesis from the NIR scene images the proposed method shown in Fig. 2. The proposed MobileAR-GAN generator architecture is described in Fig. 3. A paired dataset $P_{k=1}^{n} = (X_k, Y_k)_{k=1}^{n}$ used for the training purposes, where X_k and Y_k denotes NIR and corresponding RGB scene images in paired manner for the input source and the target domains. We utilize the cyclic-synthesized loss [26] within the CycleGAN framework. For each generator network Recurrent Mobile network integrates with attention mechanisms. Thus, the proposed MobileAR-GAN translates the images from source domain (X) to target domain (Y) and target domain (Y) to source domain (X) in cyclic manner. We use two Attention Guided Generator

Networks, $G_{M_{XY}}$ and $G_{M_{YX}}$, i.e., $G_{M_{XY}}$ to perform image translation from domain X to domain Y ($X \rightarrow Y$) and $G_{M_{YX}}$ to generate the image in domain X from domain Y ($Y \rightarrow X$). The generator networks consist of an inbuilt attention mechanism by using attention gates. We use auto-encoder based architecture as the backbone of each generator network. The decoder part uses the attention gates, which prioritizes the synthesization in important and specific regions to improve the quality of the generated images. The proposed MobileAR-GAN uses other losses, which help to optimize the learning space by superimposing different curvatures in the learning space.

A. Proposed Mobile-Net Generator

To reduce the computational parameters, we recursively use an Inverted residual block with a combination of U-net-based architecture. In which a Recurrent Mobile Network block uses a U-net backbone. In the encoding part of the network, we combine the recurrent mobile network with the max-pool layer and down-sample it up to bottleneck. While the decoder uses attention gates additionally to perform the region-specific representational learning, the applicability of attention gates helps in fast convergence and better representational learning. At the same time, the Mobile block helps the model to reduce the computational parameters in the Recurrent Mobile network. Each recurrent mobile network uses only a 1×1 Conv2d layer with a recurrent Mobile block. Each recurrent mobile block has two times recursion. While for effective computing with less overhead, the advantage of auto-mixed precision-based networks is also utilized to optimize data size during training and testing. However, using attention increases the network parameters slightly, but increases the computational efficiency compared to GAN-Compression [27] as highlighted in Tables V and VI.

B. Network Modules

Following are the details of the proposed network:

Recurrent Mobile Network: Firstly 1×1 convolution operation is performed and then output is passed to Recurrent Mobile block where $t = 2$ times recursion performed as shown in Table III. Depthwise separable convolution operation used in Mobile-Net to reduce the trainable parameters for Generator architecture. In depthwise separable convolution operation channel wise spatial operation is performed. Basically, first pointwise 1×1 convolution operation is performed and then depthwise 3×3 convolution takes place. This approach is termed as Inverted Residual Block (IR) which is based on the MobileNet concept used in Mobile Block MB. We propose to use the MB block in an recurrent fashion leading to Recurrent Mobile Block. Consider x and x_1 are the input and output of Recurrent Mobile Block and t as the number of recursion call, then x_1^t can be given recursively as,

$$x_1^t = \begin{cases} MB(x), & \text{if } t = 1, \\ MB(x + x_1^{t-1}), & \text{if } t > 1 \end{cases} \quad (1)$$

Note that traditional Residual block uses wide $(3 \times 3) \rightarrow$ narrow $(1 \times 1) \rightarrow$ wide (3×3) approach while Mobile Block uses Inverted Residual block [28] consist of narrow $(1 \times 1) \rightarrow$ wide $(3 \times 3) \rightarrow$ narrow (1×1) approach.

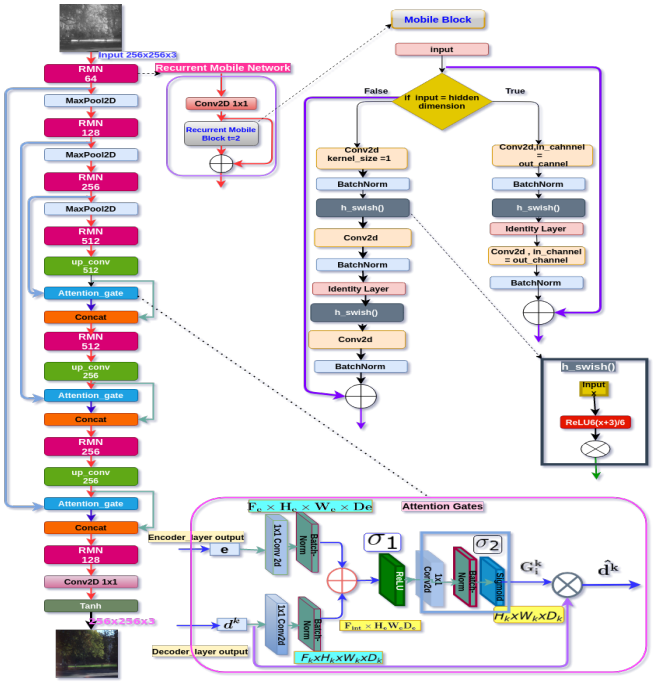

Fig. 3: Proposed MobileAR-GAN Generator architecture.

Attention Gates: For more elaborating attentive representation learning we use attention gates in the decoder part of proposed generator network. Attention gates [29] help to focus on more important feature learning while transformation takes place in decoder network. The detailed attention block is illustrated in Table II, where W_e denotes the encoder layer input and W_d denotes the decoder layer input for Attention Gates. The c^{th} channel of attention gate output in k^{th} layer of decoder is defined as $\hat{d}_{i,c}^k = d_{i,c}^k \times G_i^k$ where $d_{i,c}^k$ is the c^th channel of the input to attention gate and G_i^k is the learnt attention scores. The attention score is computed as $G_i^k = \sigma_2(\varphi^T(\sigma_1(W_d^T d_i^k + W_e^T e_i + b_e)) + b_\varphi)$ where σ_1 is notation used for ReLU activation function, σ_2 represents the Sigmoid activation function, φ represents 1×1 convolution operation followed by batch normalization, d_i^k denotes the k^{th} layer decoder's output after upscaling, e_i is the output of corresponding encoder layer, W_d and W_e are the weights of the conv layers corresponding to decoder and encoder branch, respectively, b_e and b_φ are the bias terms.

Automatic Mixed Precision: Some operations associated with the linear and convolutional layers become faster for float16 data type (half precision), which drastically reduces the computational cost of most operations performed using these layers in deep learning. We implement automatic mixed-precision training [30] using TORCH.CUDA.AMP package available with the torch library. These operations help in faster training and testing using limited computational resources. Additional required packages are provided in Supplementary.

Generator Architecture: The detailed network architecture is represented in Fig. 3. Generator Architecture is illustrated in Table I with sub-architectures Recurrent Mobile Block and

TABLE I: **Generator Network Architecture**

R1 (Recurrent Mobile Network), in_channel =3, out_channel =64, t=2
MaxPool2d, kernel_size = 2, stride = 2
R2 (Recurrent Mobile Network), in_channel =64, out_channel =128, t=2
MaxPool2d, kernel_size = 2, stride = 2
R3 (Recurrent Mobile Network), in_channel =128, out_channel =256, t=2
MaxPool2d, kernel_size = 2, stride = 2
R4 (Recurrent Mobile Network), in_channel =256, out_channel =512, t=2
UP(1) Upconv, in_channels =512, out_channels=256
Attention_Gates, input = (UP(1), R3)
Concat input = (UP(1), R3)
R5 (Recurrent Mobile Network), in_channel =512, out_channel =256, t=2
UP(2) Upconv, in_channels =256, out_channels=128
Attention_Gates, input = (UP(2), R2)
Concat input = (UP(2), R2)
R6 (Recurrent Mobile Network), in_channel =256, out_channel =128, t=2
UP(3) Upconv, in_channels =128, out_channels=64
Attention_Gates, input = (UP(3), R1)
Concat input = (UP(3), R1)
R7 (Recurrent Mobile Network), in_channel =128, out_channel =64, t=2
Conv2D, in_channels =64, out_channels =3, kernel_size = 1, stride = 1
Tanh

TABLE II: **Attention gates**

W_e block, input = in			
Operation	**kernel_size**	**stride**	**channels in, out**
Conv2d + Batch_Norm	1	1	inp, inp/2
W_d block, input = in			
Operation	**kernel_size**	**stride**	**channels in, out**
Conv2d + Batch_Norm	1	1	inp, inp/2
Q = ReLU(output(W_d) + output(W_e))			
φ **block**, input = inp/2			
Operation	**kernel_size**	**stride**	**channels in, out**
Conv2d + Batch_Norm	1	1	inp, inp/inp
Sigmoid			
Out = φ(Q) * input((W_d))			

TABLE III: **Recurrent Mobile Network & Upconv Block**

Recurrent Mobile Network	**Upconv Block (UP)**
Input (in_ch)	Input (in_ch), Output (in_ch/2)
C = Conv2d 1x1 (in_ch)	Upsample, Scale_factor = 2
Recurrent Mobile Block (MB), t=2	Conv2d Kernel =3, stride =1, pad =1
C + MB	BatchNorm+ReLU6

TABLE IV: **Mobile Block**

INPUT (in_ch = h_dim*2)
Conv2d (in_ch, h_dim, 1, 1, 0, bias=False)
Batch_Norm (h_dim), h_swish
Conv2d (h_dim, h_dim, 3, 1, 1, groups= h_dim, bias=False)
Batch_Norm (h_dim) , Identity, h_swish
Conv2d(hidden_dim, in_ch, 1, 1, 0, bias=False), Batch_Norm(in_ch)
Input+Output(Mobile Block)

Mobile Block illustrated in Table III and IV, respectively. Attention Gates and Up conv blocks illustrated in Table II and Table III respectively. We use CycleGAN architecture as baseline for our methods which consists two generator $G_{M_{XY}}$, $G_{M_{YX}}$ and two discriminator networks D_X and D_Y.

Discriminator Architecture: We use PatchGAN Discriminator as proposed in Pix2pix [9]. The detailed architecture of the same is illustrated in Supplementary under Table 2.

C. Objective Function

Adversarial Loss (Al): Adversarial loss evaluates between generator network and discriminator network during adversarial training of both the networks. The Generator network generates artificially synthesized image samples. The discriminator network labeled the artificially synthesized image as fake/real by evaluating the data distribution of synthesized images and their correlation with corresponding actual image data distribution. In the training, the samples x,y obtained through domains X and Y respectively. Adversarial loss for $X \rightarrow Y$ translation described as below,

$$\mathcal{L}_{GAN}(G_{M_{XY}}, D_Y) = Min_{G_{M_{XY}}} Max_{D_Y} = \mathbb{E}_{y \sim p_{data(y)}}[\log D_Y(y)] + \mathbb{E}_{x \sim p_{data(x)}}[\log(1 - D_Y(G_{M_{XY}}(x)))]$$

Same as above, the adversarial loss for opposite domain $Y \rightarrow X$ transformation defined as ($\mathcal{L}_{GAN}(G_{M_{YX}}, D_X)$).

Cycle-consistency Loss (Cl): For the Cyclic training of the images, Cycle-consistency loss [10] used to reduce the domain gap between the reconstructed and input images, computed through the L1 distance measurement among the real image and the artificially synthesized cyclic image (reconstructed image) for forward and opposite of it, backward cycle-consistency loss calculated. For $x \in X$ and $y \in Y$, cycle-consistency loss during forward cyclic transformation described as $\mathcal{L}_{Cyc_F}$ and for the backward transformation cycle-consistency loss described as $\mathcal{L}_{Cyc_B}$ illustrated below.

$$\mathcal{L}_{Cyc_F} = ||x - G_{M_{YX}}(G_{M_{XY}}(x))||_1$$
$$\mathcal{L}_{Cyc_B} = ||y - G_{M_{XY}}(G_{M_{YX}}(y))||_1$$

Synthesized Loss (Sl): Synthesized loss is calculated as L_1 distance between the artificially synthesized images and corresponding ground truth images. The synthesized losses for the domains X and Y are defined as,

$$\mathcal{L}_{Sl_X} = ||x - G_{M_{YX}}(y)||_1$$
$$\mathcal{L}_{Sl_Y} = ||y - G_{M_{XY}}(x)||_1$$

Cycle-Synthesized Loss (Csl): The cycle-synthesized loss [26] helps to reduce the gap between pixels of the artificially synthesized images and the reconstructed images obtained through cyclic training in the opposite domain. Cycle-synthesized loss is computed as an L_1 distance measure between the images. The cycle-synthesized loss computed as,

$$\mathcal{L}_{Csl_1} = ||G_{M_{YX}}(G_{M_{XY}}(x)) - G_{M_{YX}}(y)||_1$$
$$\mathcal{L}_{Csl_2} = ||G_{M_{XY}}(G_{YX}(y)) - G_{M_{XY}}(x)||_1$$

where $G_{M_{YX}}(y)$ and $G_{M_{XY}}(x)$ are the synthesized images and $G_{M_{YX}}(G_{M_{XY}}(x))$ and $G_{M_{XY}}(G_{M_{YX}}(y))$ are the cyclic reconstructed images.

Feature Reconstruction Loss (FR): To reduce the feature gap between artificially synthesized images and the target images, we compute the loss among similar feature representations for each domain. Detailed FR loss is described

in Supplementary. Below are feature reconstruction losses computed among different domains:

$$\mathcal{L}_{real}^{fake}(X) = l_{feat}^{\Psi,l}(x, G_{M_{XY}}(x))$$
$$\mathcal{L}_{real}^{fake}(Y) = l_{feat}^{\Psi,l}(y, G_{M_{YX}}(y))$$
$$\mathcal{L}_{real}^{cycle}(X) = l_{feat}^{\Psi,l}(x, G_{M_{YX}}(G_{M_{XY}}(x)))$$
$$\mathcal{L}_{real}^{cycle}(Y) = l_{feat}^{\Psi,l}(y, G_{M_{XY}}(G_{M_{YX}}(y)))$$
$$\mathcal{L}_{fake}^{cycle}(X) = l_{feat}^{\Psi,l}(G_{M_{YX}}(y), G_{M_{YX}}(G_{M_{XY}}(x)))$$
$$\mathcal{L}_{fake}^{cycle}(Y) = l_{feat}^{\Psi,l}(G_{M_{XY}}(x), G_{M_{XY}}(G_{M_{YX}}(y)))$$

Final Objective Function: The final objective function for the proposed MobileAR-GAN is illustrated as below:

$$\mathcal{L}(G_{M_{XY}}, G_{M_{YX}}, D_X, D_Y) =$$
$$\mathcal{L}_{GAN} + \mathcal{L}_{Cyc} + \mathcal{L}_{Csl} + \mathcal{L}_{Sl} + \mathcal{L}_{FR} \textbf{ where}$$
$$\mathcal{L}_{GAN} = \lambda_g(\mathcal{L}_{GAN}(G_{M_{XY}}, D_Y) + \mathcal{L}_{GAN}(G_{M_{YX}}, D_X))$$
$$\mathcal{L}_{Cyc} = \lambda_{Cyc}(\mathcal{L}_{Cyc_F} + \mathcal{L}_{Cyc_B})$$
$$\mathcal{L}_{Csl} = \lambda_{Csl}(\mathcal{L}_{Csl_1} + \mathcal{L}_{Csl_2})$$
$$\mathcal{L}_{Sl} = \lambda_{Sl}(\mathcal{L}_{Sl_X} + \mathcal{L}_{Sl_Y})$$
$$\mathcal{L}_{FR} = \lambda_{feat}(\mathcal{L}_{real}^{fake}(X) + \mathcal{L}_{real}^{fake}(Y) + \mathcal{L}_{real}^{cycle}(X) + \mathcal{L}_{real}^{cycle}(Y) + \mathcal{L}_{fake}^{cycle}(X) + \mathcal{L}_{fake}^{cycle}(Y)).$$

IV. Experimental Results and Observations

In this section, first we describe the experimental settings and then quantitative results and analysis followed by qualitative results comparison. Finally, we justify the efficiency of the proposed model in terms of the computational performance and inference time.

A. Experimental Settings

Evaluation Metrics: For the quantitative analysis of the results with previous methods, we use the standard evaluation measures, such as Structural Similarity Index (SSIM), Color_loss, Learned Perceptual Image Patch Similarity (LPIPS), Peak-to-Signal-Noise Ratio (PSNR), and Visual Information Fidelity (Vif). More information about these measures can be found in the Supplementary under section named as Evaluation metrics.

Datasets Used: We test the proposed model over two benchmark datasets: RGB-NIR Scene Dataset and Outdoor Multispectral Images with Vegetation (OMSIV) dataset for scene synthesis. The RGB-NIR Scene and OMSIV datasets contain the NIR and corresponding RGB images. RGB-NIR Scene Dataset [1] contains 477 images of 9 categories in NIR and RGB domains, out of which we use 333 image pairs for training and 144 image pairs for testing. We refer this dataset as NIR-RGB scene dataset as we use it for NIR to RGB domain translation. The Outdoor Multispectral Images with Vegetation (OMSIV)[2] dataset consists the NIR and multispectral images. For the OMSIV dataset we use 400 images for training and 100 images for testing.

[1]https://ivrlwww.epfl.ch/supplementary˙material/cvpr11/index.html
[2]https://github.com/xavysp/ssmid-dataset

Parameter Settings: For both datasets, training and testing images are resized to 256×256. Cyclic architecture is used with pool size of 50 as initially proposed in CycleGAN [10]. We use diffGrad optimizer [31] with learning rate 0.0002 and momentum values as $\beta_1 = 0.5$, $\beta_2 = 0.999$. For comparison with baseline methods, Pix2pix is used with its original implementation. The same is also applied for CycleGAN and DualGAN. Also, for a fair comparison, CycleGAN is trained in a paired manner. For comparative analysis of different methods with the proposed MobileAR-GAN, each non-attentive method is used to train for 200 epochs. Attention-based approaches like AGGAN, AttentionGAN, and MobileAR-GAN are required to train for comparatively fewer epochs, i.e., 100 showing a higher convergence rate compared to non-attention based methods. We train the proposed MobileAR-GAN in a Cyclic manner, but test it for NIR to Visible transformation based on the application. The proposed MobileAR-GAN and GAN-Compression(Mobile-ResNet)) models are tested for NIR to Visible transformation, for fair comparison we tested both models without knowledge distillation [32]. The loss weight hyper-parameters are considered from the source papers and empirical observations and set in the final objective function as 1, 10, 1, 15, and 1 for Adversarial loss (λ_g), Cycle-Consistency loss (λ_{Cyc}), Cycle-Synthesized loss (λ_{Csl}), Synthesized loss (λ_{Sl}), and Feature reconstruction loss (λ_{FR}), respectively. All the testing operations as mentioned in Tables V and VI are performed on same machine having Tesla V100 GPU with Xeon-processor 2.40 GHz.

B. Quantitative Result Analysis

The proposed method is compared with recent state-of-the-art non-attention-based method such as Pix2Pix [9], CycleGAN [10], DualGAN [13], PCSGAN [22], ThermalGAN [33], and GAN-Compression [27] with Mobile-ResNet model [27] as well as attention-based methods such as AGGAN [16] and AttentionGAN [17]. MobileAR-GAN shows more realistic and natural-looking images compared to the state-of-the-art attention and non-attention-based GAN models. The near infrared to visual color synthesis quantitative results using the proposed MobileAR-GAN and state-of-the-art methods are reported in Table V for the NIR-RGB Scene dataset and Table VI for the OMSIV dataset. It can be seen MobileAR-GAN performs better than the state-of-art methods in terms of the SSIM, LPIPS, Color_Loss and Vif for both NIR-RGB Scene and OMSIV datasets. Following are the observations:

- The proposed method reported gain in SSIM score over NIR-RGB Scene dataset is $\{19.15\%, 15.02\%, 59.45\%, 0.12\%, 4.97\%, 7.93\%, 23.74\%, 15.19\%\}$ compared to non-attentive approaches such as Pix2Pix, CycleGAN, DualGAN, PCSGAN and attention-based approaches such as AGGAN, Attention-GAN, as well as recent state-of-the-art non-attentive approaches like Thermal-GAN, GAN-Compression(Mobile-ResNet), respectively (Refer Table V).
- The gain reported in PSNR score over NIR-RGB Scene dataset is $\{0.39\%, 0.56\%, 1.00\%, 0.14\%, 1.03\%, 0.92\%, 0.42\%, 0.74\%\}$ higher than non-attentive approaches

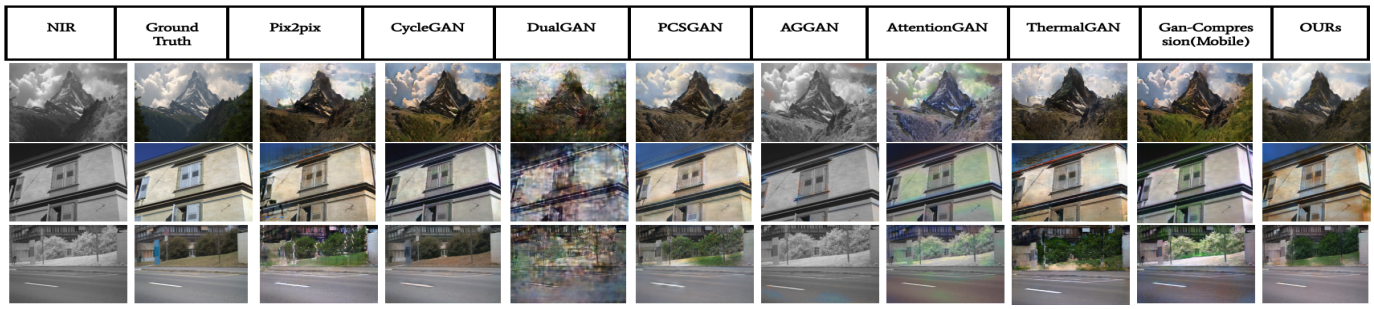

Fig. 4: Qualitative comparison by observational analysis of NIR-Scene to Visible Scene transformation, using NIR-RGB Scene dataset over different methods. As shown above the proposed MobileAR-GAN generates better quality real-looking and fair images.

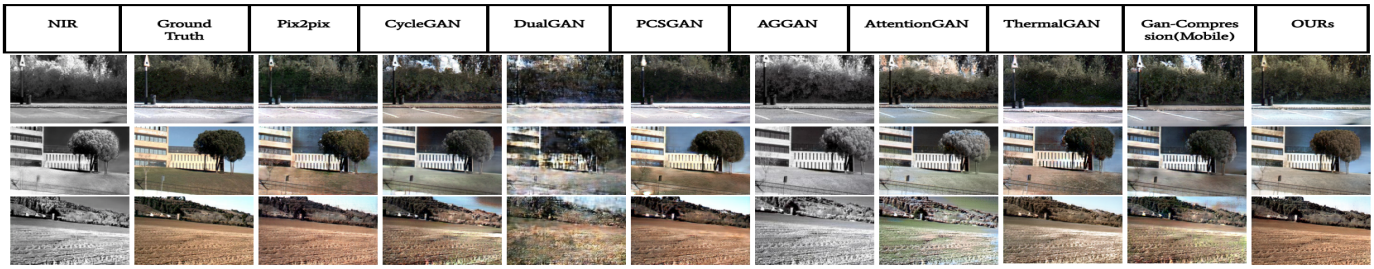

Fig. 5: Qualitative comparison by observational analysis of NIR-Scene to Visible Scene transformation, using OMSIV dataset over different methods. As shown above the proposed MobileAR-GAN generates better quality real-looking and fair images.

TABLE V: We show the empirical test results of various state-of-the-art methods, including Pix2pix, CycleGAN, DualGAN, PCSGAN, AGGAN, Attention-GAN, ThermalGAN, and GAN-compression(Mobile-ResNet) models along with our proposed MobileAR-GAN to show a comparative quantitative analysis over the NIR-RGB Scene dataset.

Method	**Pix2pix**	**CycleGAN**	**DualGAN**	**PCSGAN**	**AGGAN**	**AttentionGAN**	**ThermalGAN**	**Gan-compr-ession(Mob)**	**MobileAR-GAN**
SSIM	0.5796	0.6004	0.4331	0.6898	0.6579	0.6398	0.5581	0.5995	**0.6906**
PSNR	28.17	28.12	28.00	28.24	27.99	28.02	28.21	28.07	**28.28**
LPIPS	0.179	0.177	0.263	**0.130**	0.162	0.187	0.202	0.174	0.133
Vif	0.7919	0.7905	0.7902	0.7931	0.7826	0.7890	0.7925	0.7901	**0.7932**
Color_loss	35.56	42.13	45.31	34.46	45.50	44.69	145.23	42.63	**32.58**
Generator-Param	54.41 M	11.38 M	84.61 M	11.38M	11.45 M	11.82 M	66.99 M	**2.0** M	2.48 M
Computational_Per.	18.15 GMac	56.86 GMac	25.77 GMac	56.86 GMac	45.32 GMac	71.48 GMac	57.80 GMac	18.39 GMac	**18.00 GMac**
Testing_Time	75 Secs	89 Secs	55 secs	54 Secs	40 secs	293 Secs	84 Secs	**18 Secs**	19 Secs

TABLE VI: We show the empirical test results of various state-of-the-art methods, including Pix2pix, CycleGAN, DualGAN, PCSGAN, AGGAN, Attention-GAN, ThermalGAN, and GAN-compression(Mobile-ResNet) models along with our proposed MobileAR-GAN to show a comparative quantitative analysis over the OMSIV Scene dataset.

Method	**Pix2pix**	**CycleGAN**	**DualGAN**	**PCSGAN**	**AGGAN**	**AttentionGAN**	**ThermalGAN**	**Gan-Compr-ession(Mob)**	**MobileAR-GAN**
SSIM	0.5874	0.5888	0.3914	0.6446	0.2201	0.5847	0.5532	0.5890	**0.7004**
PSNR	28.60	28.20	28.14	28.62	27.89	28.11	28.58	28.26	**28.74**
LPIPS	0.140	0.161	0.238	0.127	0.393	0.177	0.170	0.161	**0.108**
Vif	0.8049	0.7928	0.7930	**0.8072**	0.7846	0.7908	0.8043	0.7933	0.8052
Color_loss	27.90	43.71	46.70	26.31	80.19	49.71	112.19	41.38	**24.40**
Generator-Param	54.41 M	11.38 M	84.61 M	11.38M	11.45 M	11.82 M	66.99M	**2.0M**	2.48 M
Computational_Per.	18.15 GMac	56.86 GMac	25.77 GMac	56.86 GMac	45.32 GMac	71.48 GMac	57.80 GMac	18.39 GMac	**18.0 GMac**
Testing_Time	53 Secs	69 Secs	38 Secs	40 Secs	33 secs	227 Secs	61 Secs	**13 Secs**	14 Secs

such as Pix2Pix, CycleGAN, DualGAN, PCSGAN and attention-based approaches such as AGGAN, Attention-GAN, as well as recent non-attention based Thermal-GAN, GAN-Compression(Mobile-ResNet), respectively, as reported in Table V.

- The gain in SSIM score over OMSIV dataset (as reported in Table VI) is $\{19.23\%, 18.95\%, 78.94\%, 8.65\%, 218.22\%, 19.78\%, 26.60\%, 18.91\%\}$ higher than non-attention-based methods such as Pix2Pix, CycleGAN, DualGAN, PCSGAN and attention-based methods such as AGGAN, AttentionGAN, as well as recent non-attention based ThermalGAN, GAN-Compression(Mobile-ResNet), respectively.
- The gain in PSNR score over OMSIV dataset (see Table VI) is $\{0.48\%, 1.91\%, 2.13\%$, $0.41\%, 3.04\%, 2.42\%, 0.55\%, 1.69\%\}$ higher than non-attention-based methods such as Pix2Pix, CycleGAN, DualGAN, PCSGAN and attention-based methods such as AGGAN, Attention-GAN, as well as recent non-attention based Thermal-GAN, GAN-Compression(Mobile-ResNet), respectively.

On the other hand, the proposed MobileAR-GAN shows lower score for LPIPS and Color_loss for both NIR-RGB Scene and OMSIV multispectral vegetation datasets.

- The proposed MobileAR-GAN shows reduction for

LPIPS as reported in Table V, over NIR-RGB Scene dataset by {25.69%, 24.85%, 49.42%, −2.30% 17.90%, 28.87%, 34.15%, 23.56%} than Pix2Pix, CycleGAN, DualGAN, PCSGAN, AGGAN, AttentionGAN, as well as recent non-attention based ThermalGAN, GAN-Compression(Mobile-ResNet), respectively.

- The proposed MobileAR-GAN shows reduction for LPIPS as reported in Table VI, over OMSIV dataset by {22.85%, 32.91%, 54.62%, 14.96%, 72.51%, 38.98%, 36.47%, 32.91%} than Pix2Pix, CycleGAN, DualGAN, PCSGAN, AGGAN, AttentionGAN, as well as recent non-attention based ThermalGAN, GAN-Compression(Mobile-ResNet), respectively.
- The proposed MobileAR-GAN shows reduction for Color_loss as reported in Table V, over NIR-RGB Scene dataset by {8.38%, 22.66%, 28.09%, 5.49%, 28.39%, 27.09%, 77.56%, 23.57%} than Pix2Pix, CycleGAN, DualGAN, PCSGAN, AGGAN, AttentionGAN, as well as recent non-attention based ThermalGAN, GAN-Compression(Mobile-ResNet), respectively.
- The proposed MobileAR-GAN shows reduction for Color_loss as reported in Table VI, over OMSIV dataset by {12.54%, 44.17%, 47.75%, 7.25%, 69.57%, 50.91%, 78.38%, 41.03%} than Pix2Pix, CycleGAN, DualGAN, PCSGAN, AGGAN, AttentionGAN, as well as recent non-attention based ThermalGAN, GAN-Compression(Mobile-ResNet), respectively.

C. Computational Analysis

For the Network analysis we perform Computational-parameter analysis over the state-of-the-art non-attention-based methods such as Pix2Pix [9], CycleGAN [10], DualGAN [13], PCSGAN [22], ThermalGAN [33], GAN-Compression(Mobile-ResNet) [27], as well as attention-based method models such as AGGAN [16] and AttentionGAN [17]. It can be seen in the Table V and VI, MobileAR-GAN takes less computational parameters and testing time compared to Pix2Pix, CycleGAN, DualGAN, PCSGAN, AGGAN, AttentionGAN and ThermalGAN. While at the same time, we observe either superior or very much comparable efficiency by the proposed model as compared to GAN-Compression(Mobile-ResNet) in terms of Generator-Parameters and Testing time.

- The proposed MobileAR-GAN shows reduction for generator parameters by {95.44%, 78.20%, 43.91%, 78.20%, 78.34%, 79.01%, 96.29%} than Pix2Pix, CycleGAN, DualGAN, PCSGAN, AGGAN, AttentionGAN and ThermalGAN, respectively, as reported in Table V and VI.
- The proposed MobileAR-GAN shows reduction for testing time over NIR-RGB Scene dataset as reported in Table V by {74.66%, 78.65%, 65.45%, 64.81%, 52.50%, 93.51%, 77.38%, −5.55%} than Pix2Pix, CycleGAN, DualGAN, PCSGAN, AGGAN, AttentionGAN, ThermalGAN, and GAN-Compression(Mobile-ResNet), respectively.
- The proposed MobileAR-GAN shows reduction for testing time over OMSIV Multi-spectral dataset as reported in Table VI by {73.58%, 79.71%, 63.15%, 65.00%, 57.57%, 93.83%, 77.04%, −7.69%} than Pix2Pix, CycleGAN, DualGAN, PCSGAN, AGGAN, AttentionGAN, ThermalGAN, and GAN-Compression(Mobile-ResNet), respectively.
- The proposed MobileAR-GAN shows reduction for computational performance in terms of GMac by {0.82%, 68.34%, 30.15%, 68.34%, 60.28%, 74.81%, 68.85%, 2.12%} than Pix2Pix, CycleGAN, DualGAN, PCSGAN, AGGAN, AttentionGAN , ThermalGAN, and GAN-Compression(Mobile-ResNet), respectively, as reported in Table V and VI.

D. Qualitative Result Analysis

The empirical test results analysis between the generated images and ground truth images using the proposed MobileAR-GAN and other compared sate-of-the-art GAN models (i.e., Pix2Pix, CycleGAN, DualGAN, PCSGAN, AGGAN, AttentionGAN, ThermalGAN and GAN-Compression(Mobile-ResNet)) is shown in Fig. 4 and 5 for the sample images from NIR-RGB Scene and OMSIV datasets, respectively. It is evident in Fig. 4, the proposed MobileAR-GAN results better than existing state-of-the-art methods on the NIR-RGB Scene dataset. A similar observation is also made in Fig. 5 that MobileAR-GAN results are visually much better over OMSIV dataset.

Moreover, the MobileAR-GAN model produces better and visually appealing results corresponding to the non-attention-based approaches like Pix2Pix, CycleGAN, DualGAN, PCSGAN, and ThermalGAN as the proposed model able to focus on the essential visual characteristics in a better way with the help of attention module. On the other hand, the existing attention-based models such as AGGAN and AttentionGAN fail to produce the qualitative images with color consistency. However, the proposed method is able to tackle this issue with the help of inverted residual block with attention modules and performs better even in fewer epochs of training as compared to the existing methods.

V. Impact of Different Losses and Architectural Modules.

Impact of Different Losses: In order to justify the requirement and impact of the used losses in the proposed approach, we execute an ablation study by evaluating the different combinations of loss functions. The qualitative comparison of various losses is shown in Supplementary for the samples from NIR-RGB Scene and OMSIV datasets, respectively. For NIR-RGB Scene dataset, the proposed MobileAR-GAN reports better results in terms of the SSIM, Vif, PSNR, LPIPS and Color_loss in % when adversarial loss, cycle loss, synthesized loss, feature reconstruction loss and Cycle synthesized loss are combined. We gain an increment of {304.56%, 1.47%, 0.48%} for SSIM, PSNR, and Vif, respectively, and reduction of {57.09%, 56.57%} for LPIPS and Color_loss, respectively, by the proposed MobileAR-GAN model when compared to only adversarial loss as reported in Table VII. MobileAR-GAN shows an increment of {17.58%, 0.92%, 0.69%} in SSIM,

TABLE VII: The quantitative analysis of different loss combinations over the proposed MobileAR-GAN model for the NIR-RGB Scene dataset and OMSIV Scene dataset.

	NIR-RGB Scene dataset					OMSIV Scene dataset				
Method	**SSIM**	**PSNR**	**Vif**	**LPIPS**	**Color_loss**	**SSIM**	**PSNR**	**Vif**	**LPIPS**	**Color_loss**
Al	0.1707	27.87	0.7894	0.310	75.03	0.0910	27.90	0.7794	0.399	89.58
Al+Cl	0.5873	28.02	0.7877	0.201	45.35	0.5659	28.17	0.7925	0.179	43.94
Al+Cl+Sl	0.6301	28.18	0.7905	0.167	34.21	0.6354	28.51	0.8006	0.141	27.74
Al+Cl+Sl+Csl	0.6378	28.16	0.7917	0.165	34.73	0.6506	28.50	0.8015	0.130	26.59
Al+Cl+Sl+Csl+FR	**0.6906**	**28.28**	**0.7932**	**0.133**	**32.58**	**0.7004**	**28.74**	**0.8052**	**0.108**	**24.40**

TABLE VIII: The quantitative analysis of different modules combinations over the proposed MobileAR-GAN model for NIR-RGB Scene dataset and OMSIV Scene dataset.

	NIR_RGB Scene dataset					OMSIV Scene dataset				
Method	**SSIM**	**PSNR**	**Vif**	**LPIPS**	**Color_loss**	**SSIM**	**PSNR**	**Vif**	**LPIPS**	**Color_loss**
U-Net (1x1 Conv only)	0.4800	28.07	0.7924	0.203	38.87	0.4260	28.20	0.7950	0.192	39.03
U-Net + Recurrent Mobile	0.5852	28.20	0.7907	0.168	35.05	0.6492	28.68	0.8064	0.111	25.02
(U-Net + Recurrent Mobile + Attention) Proposed	**0.6906**	**28.28**	**0.7932**	**0.133**	**32.58**	**0.7004**	**28.74**	**0.8052**	**0.108**	**24.40**

PSNR and Vif scores, respectively, and reduction of $\{33.83\%, 28.15\%\}$ in LPIPS and Color_loss scores, respectively, as compared with the combination of adversarial loss and cycle loss, as depicted in Table VII. The proposed MobileAR-GAN gains $\{9.60\%, 0.35\%, 0.34\%\}$ in terms of the SSIM, PSNR and Vif, respectively, with a reduction of $\{20.35\%, 4.76\%\}$ in terms of the LPIPS and Color_loss, respectively, as compared to the combination of adversarial loss, cycle loss and synthesized loss (Refer Table VII). However, MobileAR-GAN achieves $\{8.27\%, 0.42\%, 0.18\%\}$ improvements using the SSIM, PSNR and Vif, respectively while leads to reduction of $\{19.39\%, 6.19\%\}$ in the LPIPS and Color_loss scores, respectively as compared to the combination of adversarial loss, cycle loss, synthesized loss and cycle-synthesized loss is considered, as illustrated in Table VII. The similar trend of importance of different losses used is also observed over OMSIV dataset as illustrated in Table VII. The findings from this experiment justify the inclusion of relevant losses in the proposed model.

Impact of different Architectural Modules: In order to justify the combination of different modules and their effect on the performance of the proposed MobileAR-GAN model, we perform a comparison with only U-Net and only U-Net+ Recurrent Mobile module based models in Table VIII. We found that the performance of the proposed model is better when all the modules such as U-Net, Recurrent Mobile and Attention are utilized. Basically, without attention the model is not able to exploit the gradients from essential regions. Thus, the performance improves when attention is integrated into the network. We report the saliency map in Fig. 2 in Supplementary which also justify the need of different modules in the proposed model.

VI. Conclusion

The proposed MobileAR-GAN method shows outstanding performance in terms of visual quality in the generated visible images from NIR images. At the same time, it is very efficient to be used with computational-limited devices. The utilization of Attention and Recurrent modules with MobileNet based generator network leads to higher quality images with small number of parameters. It is observed that the proposed model outperforms the state-of-the-art GAN models for NIR to visible translation. An ablation study on loss functions justify the utilization of relevant objective functions for the proposed model. For the fair comparison we compared all the methods without knowledge distillations. The proposed MobileAR-GAN shows better computational efficiency compared to state-of-the-art Pix2pix and GAN-Compression methods with improvement of 0.82% and 2.12% respectively, in terms of the no. of operations. In order to demonstrate the efficiency of the proposed model, we have successfully deployed it on edge computing NVIDIA Jetson device and observed very promising processing speed in real time. Hence, the proposed MobileAR-GAN model has a huge potential for real time measurement and processing of visual data for image-to-image translation applications.

References

[1] M. A. T. Monsalve, G. Osorio, N. L. Montes, S. Lopez, S. Cubero, and J. Blasco, "Characterization of a multispectral imaging system based on narrow bandwidth power leds," *IEEE Transactions on Instrumentation and Measurement*, vol. 70, pp. 1–11, 2021.

[2] X. Mao, C. Shen, and Y.-B. Yang, "Image restoration using very deep convolutional encoder-decoder networks with symmetric skip connections," in *Advances in neural information processing systems*, 2016, pp. 2802–2810.

[3] L. Wang, V. Sindagi, and V. Patel, "High-quality facial photo-sketch synthesis using multi-adversarial networks," in *2018 13th IEEE international conference on automatic face & gesture recognition (FG 2018)*. IEEE, 2018, pp. 83–90.

[4] Z. Zhou, Y. Xiang, H. Xu, Z. Yi, D. Shi, and Z. Wang, "A novel transfer learning-based intelligent nonintrusive load-monitoring with limited measurements," *IEEE Transactions on Instrumentation and Measurement*, vol. 70, pp. 1–8, 2021.

[5] W. Huang, L. Zhang, W. Gao, F. Min, and J. He, "Shallow convolutional neural networks for human activity recognition using wearable sensors," *IEEE Transactions on Instrumentation and Measurement*, vol. 70, pp. 1–11, 2021.

[6] V. Crescitelli, A. Kosuge, and T. Oshima, "Poison: Human pose estimation in insufficient lighting conditions using sensor fusion," *IEEE Transactions on Instrumentation and Measurement*, vol. 70, pp. 1–8, 2021.

[7] N. Souly, C. Spampinato, and M. Shah, "Semi supervised semantic segmentation using generative adversarial network," in *Proceedings of the IEEE International Conference on Computer Vision*, 2017, pp. 5688–5696.

[8] R. Abdal, Y. Qin, and P. Wonka, "Image2stylegan: How to embed images into the stylegan latent space?" in *Proceedings of the IEEE International Conference on Computer Vision*, 2019, pp. 4432–4441.

[9] P. Isola, J.-Y. Zhu, T. Zhou, and A. 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.

[10] J.-Y. Zhu, T. Park, P. Isola, and A. A. Efros, "Unpaired image-to-image translation using cycle-consistent adversarial networks," in *Proceedings of the IEEE international conference on computer vision*, 2017, pp. 2223–2232.
[11] M.-Y. Liu and O. Tuzel, "Coupled generative adversarial networks," in *Advances in neural information processing systems*, 2016, pp. 469–477.
[12] P. Isola, J.-Y. Zhu, T. Zhou, and A. 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.
[13] Z. Yi, H. Zhang, P. Tan, and M. Gong, "Dualgan: Unsupervised dual learning for image-to-image translation," in *Proceedings of the IEEE international conference on computer vision*, 2017, pp. 2849–2857.
[14] B. Liao and Y. Chen, "An image quality assessment algorithm based on dual-scale edge structure similarity," 10 2007, pp. 56–56.
[15] I. Goodfellow, J. Pouget-Abadie, M. Mirza, B. Xu, D. Warde-Farley, S. Ozair, A. Courville, and Y. Bengio, "Generative adversarial nets," in *Advances in neural information processing systems*, 2014, pp. 2672–2680.
[16] Y. A. Mejjati, C. Richardt, J. Tompkin, D. Cosker, and K. I. Kim, "Unsupervised attention-guided image-to-image translation," in *Advances in Neural Information Processing Systems*, 2018, pp. 3693–3703.
[17] H. Tang, D. Xu, N. Sebe, and Y. Yan, "Attention-guided generative adversarial networks for unsupervised image-to-image translation," in *2019 International Joint Conference on Neural Networks (IJCNN)*. IEEE, 2019, pp. 1–8.
[18] M. Mirza and S. Osindero, "Conditional generative adversarial nets," *ArXiv*, vol. abs/1411.1784, 2014.
[19] H. Zhang, I. Goodfellow, D. Metaxas, and A. Odena, "Self-attention generative adversarial networks," in *International Conference on Machine Learning*, 2019, pp. 7354–7363.
[20] A. R. Lejbølle, K. Nasrollahi, B. Krogh, and T. B. Moeslund, "Person re-identification using spatial and layer-wise attention," *IEEE Transactions on Information Forensics and Security*, vol. 15, pp. 1216–1231, 2020.
[21] Y. Cui, Y. An, W. Sun, H. Hu, and X. Song, "Lightweight attention module for deep learning on classification and segmentation of 3-d point clouds," *IEEE Transactions on Instrumentation and Measurement*, vol. 70, pp. 1–12, 2021.
[22] "Pcsgan: Perceptual cyclic-synthesized generative adversarial networks for thermal and nir to visible image transformation," *Neurocomputing*, vol. 413, pp. 41–50, 2020.
[23] M. Limmer and H. P. Lensch, "Infrared colorization using deep convolutional neural networks," in *2016 15th IEEE International Conference on Machine Learning and Applications (ICMLA)*. IEEE, 2016, pp. 61–68.
[24] T. Sun and C. Jung, "Nir image colorization using spade generator and grayscale approximated self-reconstruction," in *2020 IEEE International Conference on Visual Communications and Image Processing (VCIP)*, 2020, pp. 463–466.
[25] D. L. Hickman, "Colour fusion of rgb and nir imagery for surveillance applications," in *Electro-Optical and Infrared Systems: Technology and Applications XVII*, vol. 11537. International Society for Optics and Photonics, 2020, p. 115370H.
[26] K. B. Kancharagunta and S. R. Dubey, "Csgan: Cyclic-synthesized generative adversarial networks for image-to-image transformation," *arXiv preprint arXiv:1901.03554*, 2019.
[27] M. Li, J. Lin, Y. Ding, Z. Liu, J.-Y. Zhu, and S. Han, "Gan compression: Efficient architectures for interactive conditional gans," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2020, pp. 5284–5294.
[28] M. Sandler, A. Howard, M. Zhu, A. Zhmoginov, and L.-C. Chen, "Mobilenetv2: Inverted residuals and linear bottlenecks," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2018, pp. 4510–4520.
[29] O. Oktay, J. Schlemper, L. L. Folgoc, M. Lee, M. Heinrich, K. Misawa, K. Mori, S. McDonagh, N. Y. Hammerla, B. Kainz *et al.*, "Attention u-net: Learning where to look for the pancreas," *arXiv preprint arXiv:1804.03999*, 2018.
[30] P. Micikevicius, S. Narang, J. Alben, G. Diamos, E. Elsen, D. Garcia, B. Ginsburg, M. Houston, O. Kuchaiev, G. Venkatesh *et al.*, "Mixed precision training," in *International Conference on Learning Representations*, 2018.
[31] S. R. Dubey, S. Chakraborty, S. K. Roy, S. Mukherjee, S. K. Singh, and B. B. Chaudhuri, "diffgrad: An optimization method for convolutional neural networks," *IEEE transactions on neural networks and learning systems*, vol. 31, no. 11, pp. 4500–4511, 2019.
[32] G. Hinton, O. Vinyals, J. Dean *et al.*, "Distilling the knowledge in a neural network."
[33] V. V. Kniaz, V. A. Knyaz, J. Hladuvka, W. G. Kropatsch, and V. Mizginov, "Thermalgan: Multimodal color-to-thermal image translation for person re-identification in multispectral dataset," in *Proceedings of the European Conference on Computer Vision (ECCV) Workshops*, September 2018.

Nand Kumar Yadav is with the Indian Institute of Information Technology (IIIT), Allahabad as Research Scholar. He received the M.Tech. degree IIIT Allahabad in 2018. His research interest includes Computer Vision, Deep Learning, Biometrics, Convolutional Neural Networks, Image Feature Description, Image Retrieval, Image-to-Image Transformation.

Satish Kumar Singh is with the Indian Institute of Information Technology Allahabad India, as an Associate Professor and heading the Computer Vision and Biometrics Lab (CVBL). His areas of interest include Image Processing, Computer Vision, Biometrics, Deep Learning, and Pattern Recognition. He is the senior member of IEEE. Presently Dr. Singh is Section Chair IEEE Uttar Pradesh Section (2021) and a member of IEEE India Council (2021). He also served as the Vice-Chair, Operations, Outreach and Strategic Planning of IEEE India Council (2020) & Vice-Chair IEEE Uttar Pradesh Section (2019 & 2020). Dr. Singh is also serving as the Chair of IEEE Signal Processing Society Chapter of Uttar Pradesh Section. He has executed the projects funded by DRDO, Govt. of India. He has published several papers in IEEE Transactions and Reputed Journals.

Shiv Ram Dubey is with the Indian Institute of Information Technology (IIIT), Allahabad as Assistant Professor. Earlier, he was with the IIIT Sri City as Assistant Professor. He received the Ph.D. degree IIIT Allahabad in 2016. Before that he was a Project Officer at Indian Institute of Technology, Madras. He was a recipient of several awards including Best PhD Award, Early Career Research Award from SERB, Govt. of India and NVIDIA GPU Grant Award Twice from NVIDIA. His research interest includes Computer Vision, Deep Learning, Biometrics, Convolutional Neural Networks, Image Feature Description, Image Retrieval, Image-to-Image Transformation, Face Detection and Recognition, Facial Expression Recognition, Texture and Hyperspectral Image Analysis.