LKAT-GAN: A GAN for Thermal Infrared Image Colorization Based on Large Kernel and AttentionUNet-Transformer

Youwei He, Xin Jin, *Member, IEEE*, Qian Jiang, Zien Cheng, Puming Wang, and Wei Zhou, *Member, IEEE*

***Abstract*—Because thermal infrared (TIR) images are not affected by light and foggy environments, which are widely used in various night traffic scenarios. Especially, TIR images also play an important role in autonomous vehicles. However, low contrast and lack of chromaticity have always been their problems. Image colorization is a vital technique to improve the quality of TIR images, which is beneficial to human interpretation and downstream tasks. Despite thermal infrared image colorization methods have been rapidly improved, the detail blurriness and color distortion in colorized images remain under-addressed. Mostly because these methods cannot effectively extract the ambiguous feature information of TIR images. Hence, we propose a large kernel (LK) U-Net and Attention_U-Net-Transformer (ViT-Based) based generative adversarial network. An LK_U-Net is designed to extract the feature of TIR images. Then, a branch structure composed of Attention_U-Net and ViT-Based can provide the network with semantic information from different perspectives to decode features. In addition, a composite loss function is employed to ensure the network generates a high-quality colorized image. The proposed method is evaluated on KAIST and IRVI datasets. Experimental results demonstrate the superiority of the proposed LKAT-GAN over other methods for thermal infrared image colorization tasks.**

***Index Terms*—Thermal infrared image colorization, Generative adversarial networks, Transformer, Large kernel, Attention mechanism.**

I. Introduction

INFRARED images can be used to distinguish different objects from the background based on their radiation, and it works well not only day and night time but in bad weather and dim environments, such as rainy and foggy weather. Hence, the TIR imaging technique is widely used in the military, surveillance system, vehicle imaging system, nighttime traffic, and other scenarios [1], [2], [3], [4]. However, thermal infrared image lacks color information and the majority of details compared to RGB image in the day, which seriously hinders the development of its downstream tasks, such as image recognition [5], [6], object detection [7], and other fields [8], [9], [10], [11]. What is more, thermal infrared image does not conform to human visual habits, which leads it difficult to interact with human applications. Therefore, thermal infrared image colorization is increasingly important to improve the quality of the thermal infrared image.

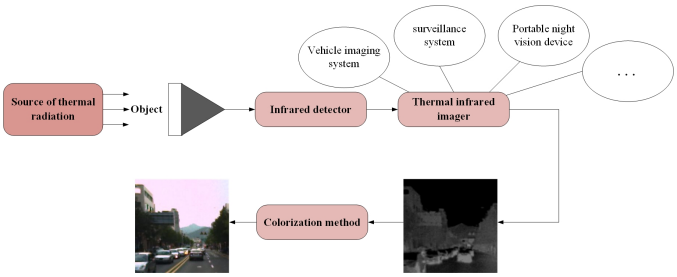

Fig. 1. **Thermal infrared image colorization can effectively solve the limitation of thermal infrared images, which is conducive to various applications of thermal infrared images.**

As a key technique of the thermal infrared image enhancement, image colorization can effectively improve the quality of the thermal infrared image. As shown in Fig. 1. It can make thermal infrared images more in line with human visual habits, and restore their grayscale details. Current grayscale image-based colorization methods have achieved impressive results [12], [13], [14], [15]. On the one hand, only colors are needed to be estimated in common grayscale image-based colorization methods, yet texture and color information both need to be reconstructed for thermal infrared image colorization. On the other hand, the large difference between the thermal infrared image and the visible light color image also makes the design of colorization algorithm more difficult. Thus the main issues are how to make our network effectively extract the blurred feature information of thermal infrared images and how to make the network generate high-quality colorized images with color and grayscale details through an appropriate loss function.

In the past few years, conventional methods have primarily focused on the color recovery of thermal infrared images, resulting in the colorized thermal infrared images are far from real images [13], [16]. Recently, deep learning has made great progress in computer vision tasks. In particular, the

This study is supported by the National Natural Science Foundation of China (Nos. 62101481, 62002313, 62261060, 62166047, 62162067), Major Scientific and Technological Project of Yunnan Province (No. 202202AD080002), Basic Research Project of Yunnan Province (Nos. 202201AU070033, 202201AT070112, 202001BB050076, 202005AC160007), The Fund Project of Yunnan Province Education Department (No.2022j0008), Key Laboratory in Software Engineering of Yunnan Province (No. 2020SE408), The open project of Engineering Research Center of Cyberspace in 2021-2022 (No. KJAQ202112012), and 13th Graduate Research Innovation Project of Yunnan University in China (Grant No. 2021Y405), Research and Application of Object detection based on Artificial Intelligence. (*Corresponding authors: Xin Jin; Wei Zhou.*)

Youwei He, Xin Jin, Qian Jiang, Zien Cheng, Puming Wang, and Wei Zhou are with Engineering Research Center of Cyberspace, Yunnan University, Kunming 650091, Yunnan, China, and also with School of Software, Yunnan University, Kunming 650091, Yunnan, China. (e-mail: xinxin_jin@163.com, jiangqian_1221@163.com, zwei@ynu.edu.cn).

Manuscript received **; revised **.

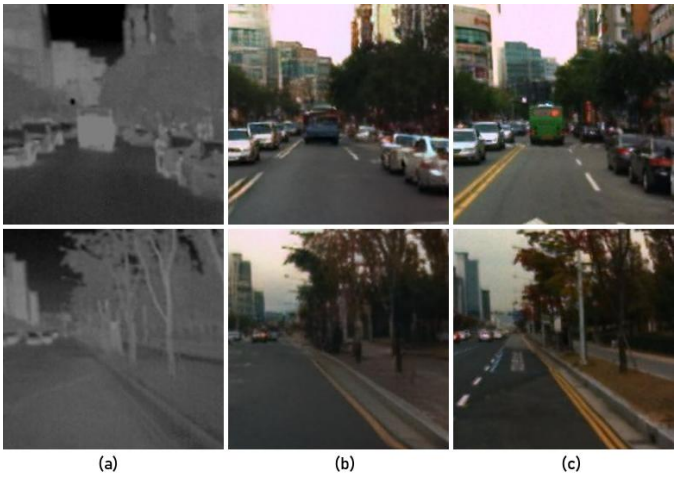

Fig. 2. **(a) TIR images. (b) Colorized results by our method. (c) True RGB images. (Compared with TIR images, the details and color information of colorized results have been significantly improved, such as buildings, roads, vehicles, plants, and other objects. It is easier to distinguish the objects when they are colorized.)**

successful applications of convolutional neural network (CNN) and adversarial neural network (GAN) have made thermal infrared image colorization methods be a great success, but the results are still different from the real RGB image due to the limitations of the thermal infrared image.

In this paper, we propose an LKAT-GAN for thermal infrared image colorization, it can effectively extract the high-level semantic and multi-scale local feature information of the thermal infrared image, and the image quality (texture and color) of the colorized image is higher. Moreover, our approach has better robustness across different datasets. We use a condition GAN-based framework to solve the problems of blurred image details and lack of texture caused by CNN. The generator (G) is composed of three parts. The U-Net structure combined with LK is the first step to extract thermal infrared image features. Second, these features obtained by the first step are further processed through a branch structure, which includes an Attention_U-Net module and a Vision Transformer (ViT) based module [17]. Finally, the colorized thermal infrared image is obtained. Our discriminator (D) is a 70×70 PatchGAN from an image-to-image translation network [18]. Our loss function refers to the work of Kuang [19], which includes content loss, adversarial loss, perceptual loss, and total variation loss. Overall, the main contributions of this paper are summarized as follows:

- A novel GAN-based network is proposed for the task of thermal infrared image colorization, in which the generator consists of LK_U-Net and Attention_U-Net-Vit branch structure.
- Combined with the enhancement of the large kernel receptive field, we design an LK_U-Net that can effectively improve the ability to extract the high-level semantic and multi-scale local feature information of thermal infrared images.
- To obtain the feature information of different perspectives, we design a U-Net with attention and ViT-Based branch module to jointly process the local and global features of images.
- On the task of thermal infrared image colorization, the experimental results demonstrate that our method can achieve state-of-the-art performance.

The remainder structure of this paper is as follows. In section II, we review some related methods of image colorization and large kernel net (LK-Net). In section III, we explain the architecture of the proposed LKAT-GAN. Section IV presents the experimental results on KAIST dataset. Section V makes the conclusions of our work.

II. Related Works

In this section, we briefly review the previous work on visible image colorization, thermal infrared image colorization and LK-Net.

A. Visible Image Colorization

Image colorization refers to the process of converting a single channel thermal infrared image to a three channels colorized image. Previous methods of grayscale image coloration require user involvement, such as Scribble-based method [20], it is based on previous color scribbles and then propagates them to the rest of the grayscale image. Example-based method [21] is a kind of colorization technique that utilize color information obtained from reference images to guide colorization.

Recently, deep learning-based colorization methods [18], [22], [23], [24], [25], [26] outperform conventional methods due to their robustness and generalization. Larsson et al. [23] proposed a fully automatic method based on CNN, a pre-trained VGG-Net is used to augment the original grayscale input. Zhao et al. [24] added image segmentation information for better colorization performance. With the introduction of GAN models [26], many image colorization tasks have adopted the GAN structure [13], [18], [27], [28]. In addition, zhang et al. [29] proposed a novel quantization metrics that can be used to measure chromatic aberration between the colorized image and the ground turth. The GAN structure is composed of two parts: generator and discriminator. The generator is responsible for generating images that are close to the real labels to fool the discriminator, and the discriminator is responsible for identifying which one is fake. Compared with conventional pixel-level losses, the adversarial loss of GAN minimizes various differences between the generated images in the target domain and the real images, thus greatly improves the performance of the results.

B. Thermal Infrared Image Colorization

Due to the lack of details and the low quality of thermal infrared images, the colorization of thermal infrared images that directly using grayscale colorization networks is often unsatisfactory. Moreover, unlike grayscale image colorization, which only generate color information, thermal infrared images require enhance details as well. Therefore, the specific design for the task of thermal infrared image colorization is required to enable it to effectively extract the features of thermal infrared images and obtain better performance.

We consider thermal infrared image colorization methods can be divided into two main categories: CNN-based [30], [31] and GAN-based [19], [27], [32]. Because the data distribution of thermal infrared image is quite different from that of visible image and the number of datasets of thermal infrared image is large, the characteristics of CNN network lead to it cannot be well adapted to the task of thermal infrared image colorization. TIR [31] is the first CNN-based colorization network, it is based on the U-Net architecture to colorize the thermal infrared image. TIR is lightweight enough, but the colorized performance is not satisfactory. Due to the advantages of the GAN-based network structure in generating images [33], [34], [35], most of these thermal infrared image colorization networks are based on GAN architecture. Pix2Pix [18] is a typical GAN-based colorization network, which generates a certain improvement in details and color of the colorized image. However, it still looks unnatural compared to the true image. After that, Kuang et al. [19] improved the model based on Pix2Pix, and the composite loss function jointly constrained generated images, this method has significantly improved the quality of colorized thermal infrared images. Lou et al. [36] proposed a network based on ToDayGAN [37] to translate a nighttime thermal infrared (NTIR) image into a daytime colorized image. They proposed a new attention module, an attentional loss and a structured gradient alignment loss, and the colorized NTIR images have a satisfying visual effect. Liao et al. [32] proposed a new attention mechanism, and combined U-Net++ and U-Net3+ structures to form the MS_U-Net, which further improves the quality of the colorized image. Recently, some networks have been proposed to colorize thermal infrared videos, such as Recycle-GAN [38] and I2VGAN [39]. They pay more attention to the fluency and realism of the videos.

C. Large Kernel Net

Recently, ViTs have shown leading performances on many computer vision tasks. What makes the performance gap between CNNs and ViTs? There are many works to explain it from different views [40], [41], [42], [43], [44]. Ding et al. [45] found that multi-head self-attention (MHSA) plays a key role in ViTs, which has large receptive fields (e.g. $\geq$ 7×7), and each layer of MHSA can collect information from a relatively large scale. However, it is a typical fashion to use small spatial convolutions (e.g. 3×3) to enlarge the receptive fields. Only some old-fashioned networks use large convolutions (Kernel size is greater than 5) as the main part of the network, such as AlexNet [46], Inceptions [47], [48]. What if we replace many small kernels with a few large kernels? Ding et al. [45] introduced large depth-wise convolutions into conventional networks (the size of the convolution kernel is set to range from 3×3 to 31×31), and proved that the large kernel can greatly improve the representation ability and the receptive field of the network. In addition, the problems of a large number of parameters and calculations have been alleviated to a certain extent by depth-wise convolutions. Besides, if the attention structure in Swin-Transformer [49] is replaced by a large convolutions structure, the performance is similar. This shows that large kernels can have relatively large receptive fields like MHSA. The combination of large convolution and small convolution, features can be extracted well.

III. Proposed Method

We design a GAN containing LK_U-Net and Attention_U-Net-Vit structure to extract the features of thermal infrared images and colorize them more precisely. In this section, we first summarize our proposed LKAT-GAN. Then, the details of our network and the loss functions in training are introduced.

A. Overview

The overall architecture is shown in Fig. 3. The proposed method consists of G and D. G is based on two modules and D is the 70×70 PatchGAN. A colorized image will be generated by G with the input of a thermal infrared image, while D guides G to generate a more realistic result. First, we use the LK_U-Net to extract the input features of the image, then a set of feature maps will be got. Second, through a branch structure that includes an Attention_U-Net and a Vit-Based net with different perspectives to process these feature maps. Then, these feature maps from the branch structure are concatenated to a convolution layer, and colorized image is obtained. Finally, the output image and the true image are inputted to the discriminator for discrimination. To get a high-quality image, the total loss is composed of content loss, adversarial loss, perceptual loss and total variant loss. We will present the details of each loss term in the following sections.

B. LK_U-Net

A large kernel means that the network can extract image features in a larger scale, and it will greatly improve the receptive field of the network [45]. Therefore, we use large convolutions that can improve the networks ability to extract the high-level features of thermal infrared images. However, LK-Net [45] is similar to the most network structures in image classification tasks, which pays more attention to the high-level semantic features of images. For thermal infrared image colorization tasks, LK-Net loses many details of the original input image, so that the results are not ideal. In this work, we adapt core structure of LK-Net (LK Blocks and ConvFFN Blocks) to build our model called LK_U-Net for thermal infrared image colorization. With the combination of LK-Net and U-Net network, LK_U-Net has large enough receptive fields to extract the high-level semantic information and process the multi-scale features of the image without damaging some detailed features.

The LK_U-Net module is shown in Fig. 4.:

Down as the beginning layers. We first use the 1×1 convolution layer to capture image details and adjust the number of channels. Then, a 3×3 convolutional layer is arranged for down-sampling.

Stage as the major layers of LK_U-Net, which consists of two components LK Blocks and ConvFFN Blocks. LK Blocks use shortcuts and DW large kernels. Before and after

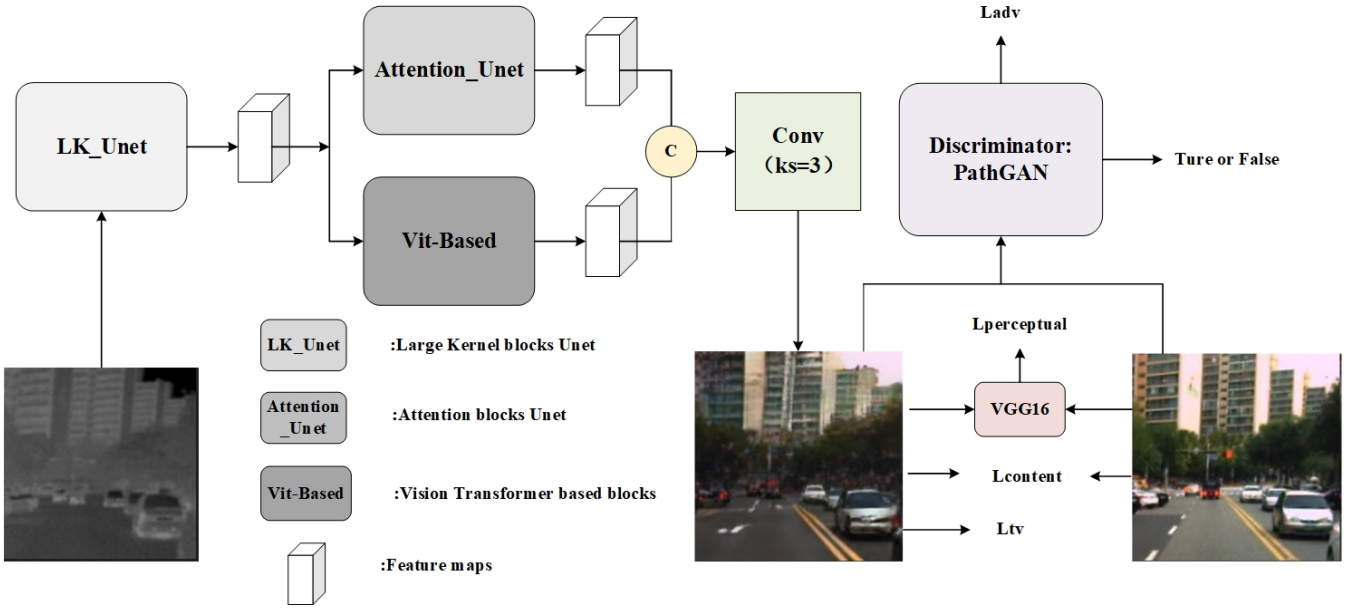

Fig. 3. **The overall framework. Including a generator, a discriminator and the loss function. Different colors represent different modules.**

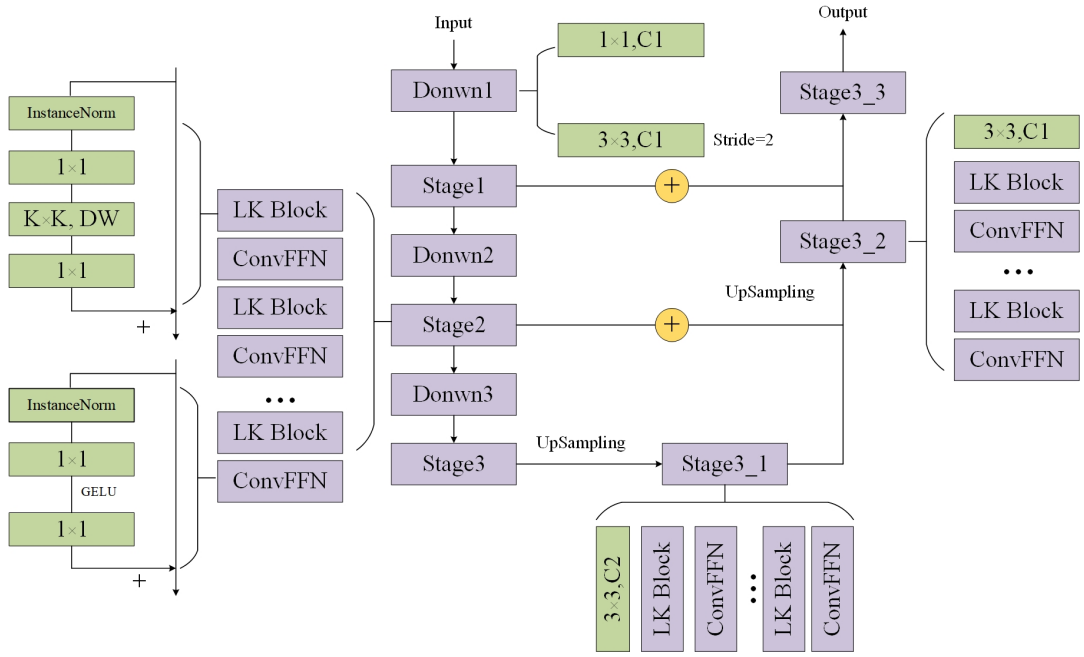

Fig. 4. **The LK_U-Net module. Except for depth-wise (DW) large kernel, the other components include 3×3, 1×1 Conv, and InstanceNorm. Activation function to use GELU.**

DW large kernels, the 1×1 convolution is used as a common practice. The representational ability of the model is not only related to the ability of large convolution layers to provide sufficient receptive field and gather spatial information but is closely related to depth. Therefore, 1×1 layers are arranged to increase the depth. ConvFFN Blocks consist of InstanceNorm [50], two 1×1 layers, GELU [51], and the number of internal channels of the ConvFFN Blocks is 4× as the input. ConvFFN works like the Feed-Forward Network (FFN), which has been widely used in Transformers [49], [52] and MLPS [53], [54]. For the UpSamping, we use the nearest interpolation method to improve the resolution of the image, and then a 3×3 convolutional layer is utilized to follow, which can change the channel dimension and make the sampling process learnable.

In summary, the ambiguous feature of the thermal infrared image can be effectively and precisely extracted without losing too much information by the LK_U-Net structure. It lays a solid foundation for the subsequent module to further process the feature information, as shown in Fig. 5. There are three hyper-parameters in this module: the number of LK and ConvFFN Blocks B, the kernel_size K, and the channel dimension C. These hyper-parameters are de-

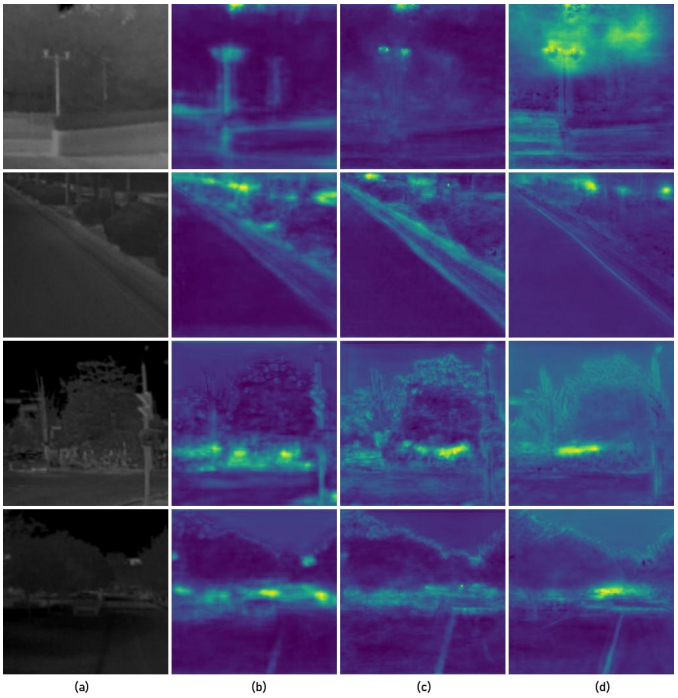

Fig. 5. **Feature maps of the first module of the network. (a) Thermal infrared images. (b) the LK blocks are replaced by 3×3 convolution blocks in LKAT-GAN. (c) the kernel_size in LK blocks replaced from 13×13 to 3×3. (d) LKAT-GAN. (As shown in (d), the feature images extracted by LK_U-Net have clearer edge information and clearer semantic information.)**

fined by $[B_1, B_2, B_3, B_4, B_5, B_6]$, $[K_1, K_2, K_3, K_4, K_5, K_6]$, $[C_1, C_2, C_3, C_4, C_5, C_6]$.

C. Attention_U-Net and ViT-Based branch module

The attention mechanism of deep learning is derived from the bionics of the human visual attention mechanism, which can obtain image mosaic detailed information to suppress other useless information and improve the representation ability of the network [55]. As shown in Fig. 6. Inspired by the excellent performance of attention mechanisms in image processing, we utilize CBAM [56] as a part of our network. CBAM uses a combination of channel and spatial attention mechanisms to enhance the capacity of extracting features. Given a feature map, the CBAM module can serially generate attention feature map information in both channels and space dimensions, then the information of two feature maps are multiplied with the previous input feature map for adaptive feature correction, which will produce the final feature map. Besides, CBAM is also a lightweight module, unlike the self-attention mechanism that consumes a lot of computing power. Based on the above advantages, we embed the CBAM mechanism into our U-Net substructure, so that it can pay attention to some important details of the image and improve the colorized quality of the thermal infrared image.

Since Attention_U-Net is a fully convolutional network, the multi-scale features of images can be extracted well, but the lack of global feature information will lead to the fuzzy details of the colorized image. To extract the global feature information of the image, we employ the encoder and decoder

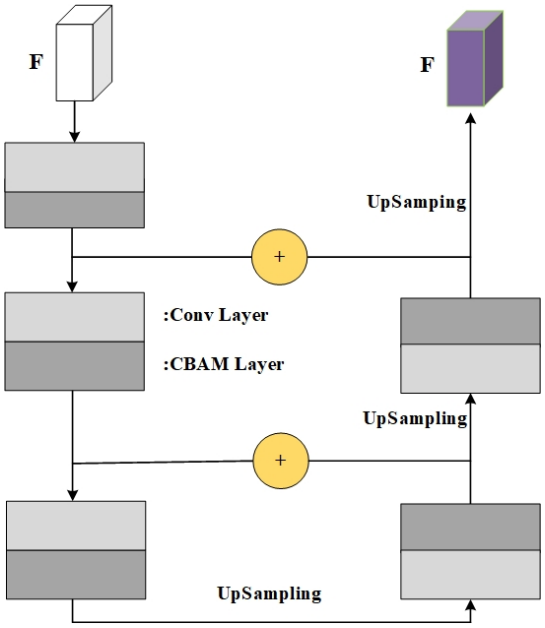

Fig. 6. **The Attention_U-Net module. A CBAM [56] block is added to each convolution layer.**

in the Masked Autoencoders Are Scalable Vision Learners (MAE) [17], as shown in Fig. 7, which is usually composed of MSHA; besides, we remove the mask operation, and the decoder is retained in MAE. More perspectives guide to better results, the branch module can fully combine the global and local feature information of the image, and the quality of the final colorized image can be effectively improved from our practice. Thus, we use the Attention_U-Net and ViT-Based branch structure to further process the feature information generated by the LK_U-Net module, then the final result of the generator is obtained.

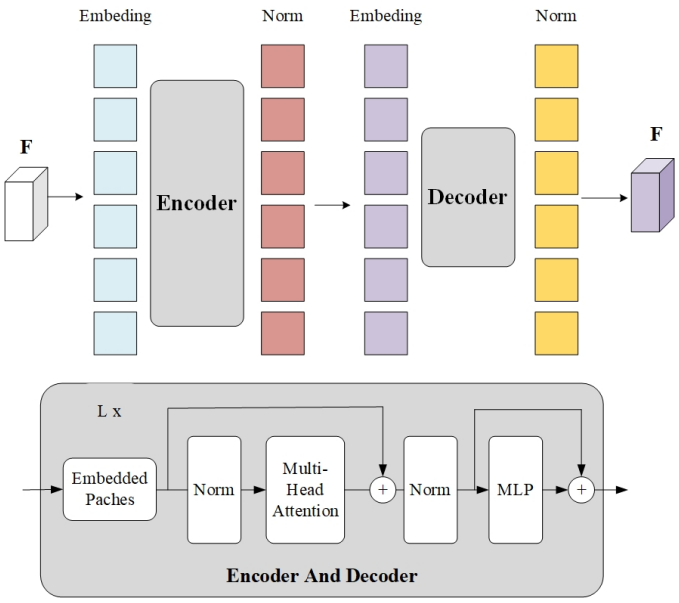

Fig. 7. **The ViT-Based module. The encoder is applied to the small subset of visible patches. Then, the full set of encoded patches is processed by a small decoder.**

To solve the problem that the texture is not clear enough due to the lack of global feature information of the network, we design an Attention_U-Net and Vit-Based branch structure, which can extract both local and global features to jointly process the input. Moreover, the size and dimension of the output are the same, then concatenate as the final output of this structure.

D. Discriminator

We use patchGAN [18] as the discriminator, which models the image as a Markov random field and enhances the high-frequency correctness of local image patches. The input image is mapped to a single scalar output for a regular GAN, which signifies "real" or "fake". However, the patchGAN maps the input into an N×N array of input X, where each X_ij signifies whether the patch ij in the image is real or fake. In this way, the receptive field of the discriminator can be improved so that it can focus on more areas. We follow the experimental results in [18], the discriminator works well when N is set to 70.

E. Loss Function

The successful practice of TIC-CGAN [19] proves that a comprehensive loss function is suitable for the thermal infrared image colorization tasks. Therefore, The loss function of our method refers to TIC-CGAN, which consists of 4 parts: Content loss $L_{content}$, Adversarial loss L_{adv}, Perceptual loss $L_{perceptual}$, and Total Variant loss L_{tv}.

Content loss can ensure that the difference between the generated colorized image and the labeled image in chrominance and luminance is minimized. We choose the MSE loss instead of MAE to implement better robustness. $L_{content}$ is defined as follows:

$$L_{content} = \mathbb{E}_{x,y}[||y - G(x,z)||_1] \tag{1}$$

Adversarial loss is the basic loss function of GAN networks, which can make the generated image more realistic and have more detailed information. Here, we use the Adversarial loss with conditions. L_{adv} is defined as follows:

$$L_{adv} = \mathbb{E}_{x,y}[\log D(x,y)] + \mathbb{E}[\log(1 - D(x, G(x,z)))] \tag{2}$$

Perceptual loss constrains the perceptual difference between the generated colorized image and the labeled image. Based on the experience of existing researches, the VGG-19[57] network pre-trained on the Image-Net dataset is utilized as the perceptual feature extractor. $L_{perceptual}$ is defined as follows:

$$L_{perceptua} = \mathbb{E}_{x,y}[||(\Phi_i(G(x,z) - \Phi_i(y)||_1] \tag{3}$$

where Φ_i represents the i-th layer features extracted by the pre-trained network.

Total Variant loss can effectively enhance the spatial smoothness of the generated colorized thermal infrared image and reduce the noise of the generated image. L_{tv} is defined as follows:

$$L_{tv} = \frac{1}{WH} \sum |\nabla_x G(\widetilde{y})| + |\nabla_y G(\widetilde{x})| \tag{4}$$

Therefore, the total loss function of our network is defined as follows:

$$\begin{aligned} L_{total} = \lambda_{content} L_{content} + \lambda_{adv} L_{adv} \\ + \lambda_{perceptua} L_{perceptua} + \lambda_{tv} L_{tv} \end{aligned} \tag{5}$$

where $\lambda_{content}$, λ_{adv}, $\lambda_{perceptua}$, and λ_{tv} are parameters that control the weight of each loss function.

IV. Experiments

In this section, we first introduce the dataset, evaluation metrics and implementation details. Then the ablation experiments are performed to demonstrate the contribution of each component in LKAT-GAN, thus the best model is selected. At last, we present the quantitative and qualitative experimental results on KAIST and IRVI datasets.

A. Dataset And Implementation Details

KAIST dataset. Since the infrared and visible images need accurate pixel-to-pixel correspondence for thermal infrared image colorization tasks. Thus, we choose KAIST multi-spectral pedestrian dataset [58] and perform comprehensive experiments. KAIST dataset provides 95k, day and night thermal and color image pairs with a resolution of 640 × 512. Considering the poor quality of night images, we only choose the day thermal and color image pairs to train and test our model. 26,387 day thermal and color image pairs are used in the training set, and 23,925 day thermal and color image pairs are used for evaluation. Due to the sufficient size of the dataset, we resize the resolution of all images in this dataset to 256 × 256 by randomly cropping.

IRVI dataset. This dataset is proposed by Li et al. [39] and consists of two scenarios: traffic and monitoring scenarios, it contains 24,351 pairs of day thermal and color images with a resolution of 256 × 256, in which 22,079 pairs of images are used as the training set and 2,272 pairs of images are used as the testing set. Although the overall image quality of this dataset is higher than that of KAIST, the scene in IRVI dataset is relatively single. Thus, it is used as our second dataset to demonstrate the performance of the proposed model.

Implementation details. We implement our model in the PyTorch framework and train for 20 epochs with a batch size of 1 on an NVIDIA 3090 GPU. In the course of training, we use the Adam optimizer [59] with $(\beta_1, \beta_2) = (0.5, 0.999)$ and the learning rate is set to 0.0002. We fix hyper-parameters of LK_U-Net, B=[2, 6, 2, 2, 6, 2], C=[64, 128, 256, 256, 128, 64], K=[13, 13, 13, 13, 13, 13]. For loss function, we follow TIC-CGAN [19], in which $\lambda_{content}$, λ_{adv}, $\lambda_{perceptua}$, and λ_{tv} are set to 1, 0.03, 1, and 1. For the comparison methods, we use the published settings for model training and testing. In this paper, there are two ways to evaluate: subjective and objective evaluation. In objective evaluation, two important evaluation metrics are used: structure similarity index (SSIM) and peak signal-to-noise ratio (PSNR). Among them, SSIM can reflect the image quality from brightness, contrast and structure. And PSNR can measure the image quality from noise intensity. The higher values of SSIM and PSNR are better.

TABLE I
QUANTITATIVE COMPARISONS OF ABLATION STUDIES OF LKAT-GAN ON THE KAIST DATASET.

	Without LK block	With LKS block	Without Attention block	Without ViT block	LKAT -GAN
PSNR	15.39	15.62	15.75	**15.86**	15.79
SSIM	0.5290	0.5341	0.5363	0.5359	**0.5435**

TABLE II
AVERAGE RESULTS OF 23,925 IMAGES ON THE KAIST TEST DATASET. THE BEST RESULTS ARE IN BOLD.

Method	SSIM	PSNR
CycleGAN [60]	0.3046	12.17
Pix2pix [18]	0.4981	15.10
SCGAN [27]	0.5252	15.48
TIC-CGAN [19]	0.5337	15.52
PealGAN [36]	0.4292	14.83
MUGAN [32]	0.5352	15.55
LKAT-GAN	**0.5435**	**15.79**

TABLE III
RESULTS OF IRVI DATASET ON THE TEST DATASET (THE BEST RESULTS ARE IN BOLD).

Method	Traffic		Monitoring	
	SSIM	PSNR	SSIM	PSNR
I2VGAN [39]	0.60	17.02	0.46	17.30
SCGAN [27]	0.65	17.72	0.50	16.55
TIC-CGAN [19]	0.62	16.94	0.52	17.65
PealGAN [36]	0.60	16.34	0.48	13.84
MUGAN [32]	**0.66**	**18.42**	0.46	14.72
LKAT-GAN	0.65	17.43	**0.58**	**18.51**

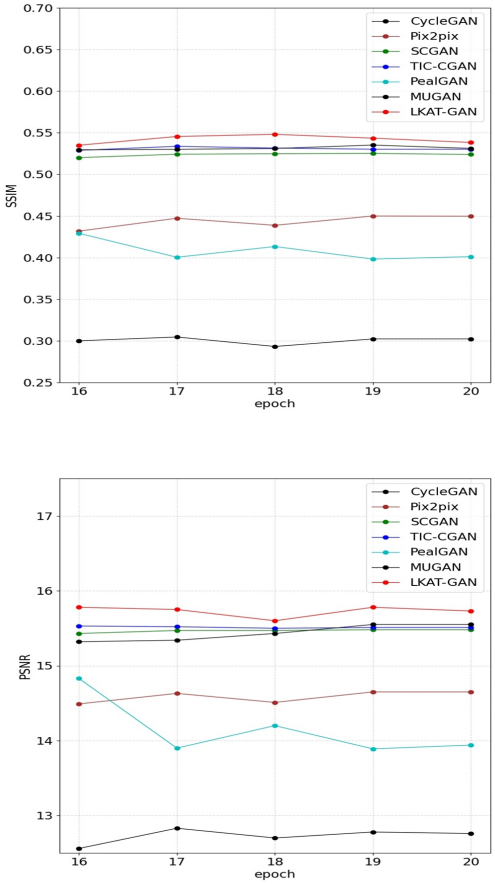

Fig. 8. **Results of different methods on KAIST datasets. The X-axis is the epoch and the Y-axis is the index.**

B. Ablation Study

We conduct the ablation study to evaluate the contribution of each component in LKAT-GAN, and select the model with the best performance. There are four experimental settings to exclude some parts from the original structure. **1) LK block:** the LK blocks are replaced by 3×3 convolution blocks in LKAT-GAN; **2) LKS block:** the structure of LK is kept, but the K×K in LK replaced from 13×13 to 3×3; **3) Attention block:** attention blocks in Attention_U-Net are removed; **4) ViT block:** ViT-Based blocks are removed.

We will verify the effective of each part of the model on the colorized results through the experiments. The objective evaluation is shown in Table I. The SSIM index of LKAT-GAN is the highest among all ablation methods. Although the PSNR index does not reach the maximum value among all ablation methods, the LKAT-GAN has the best performance generally. The network has the worst performance when we do not use the LK block. Obviously, when the first step cannot effectively extract the features, the subsequent structure will also be deeply affected on this basis. While retaining the overall structure of LK_U-Net and setting the kernel_size to 3×3, the performance of the network is also relatively poor, which proves the superiority of large convolutions. Based on using the LK_U-Net with large convolution, all networks perform well with the changed structures (Attention_U-Net and ViT-Based). Of course, the colorized result is the best when the three modules are used at the same time. The subjective results are shown in Fig. 9, we can clearly see that the LKAT-GAN shows the best performance in various details, such as the car

Fig. 9. **The colorized results for the ablation study on the KAIST Dataset. (a) Thermal infrared images. (b) Without LK block. (c) With LKS block. (d) Without attention block. (e) Without ViT block. (g) LKAT-GAN. (h) True RGB images. (As shown in (f), the vehicle texture information and road details on the side of the image are more accurate, and the overall image looks more natural.)**

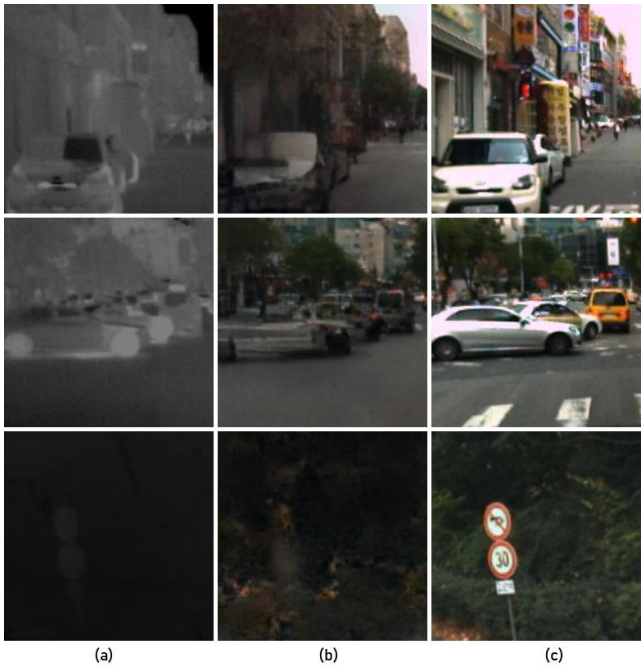

Fig. 10. **Examples of failure cases for our approach on the KAIST Dataset. (a) Thermal infrared images. (b) Error results. (c) True RGB images.**

on the side of the image and the overall quality of the colorized image. It should be noted that the number of model parameters and training time will be greatly increased after adding the VIT-Based structure. Therefore, there is a trade-off between the colorization performance and the model complexity.

C. Experiments on KAIST Dataset

The comparison methods include six thermal infrared image colorization networks, as CycleGAN [60], Pix2pix [18], SCGAN [27], TIC-CGAN [19], PealGAN [36], and MUGAN [32]. Among them, CycleGAN and PealGAN are unsupervised models, and the other methods are supervised models. All of them are trained on the NVIDIA 3090 GPU for 20 epochs. To display the objective evaluation of each model, we use line graphs to show the scores of all methods from 16 to 20 epochs, as shown in Fig. 8, the SSIM and PSNR values of the proposed are the best in each epochs compared to the comparison methods, which proves the superiority of our method. The best scores for these methods are shown in Table II. According to the experimental results, we can find that LKAT-GAN has the best performance both in SSIM and PSNR. In addition, the highest SSIM values well proves that the result generated by our method is closer to the ground truth. In quantitative comparison, LKAT-GAN also has the best performance among these thermal infrared image colorization networks both in details and color. The qualitative comparison of these methods is shown in Fig. 11. The overall quality of the image produced by SCGAN is unsatisfactory, the colors and details are not well reconstructed. Similarly, TICCGAN is deficient in image detail reconstruction. PealGAN has the advantage of addressing unpair datasets as an unsupervised model. Under the premise of paired datasets, the performance of the unsupervised model is inferior to that of the supervised model. For MUGAN, although the results are better than TICCGAN and SCGAN, the detail reconstruction of the colorized image is not satisfactory enough, and there has a small amount of noise in the results at some cases. Obviously, LKAT-GAN has a more realistic color and richer details than the compared methods, especially in some objects, such as cars, loads, and buildings. In general, the LKAT-GAN has state-of-the-art performance

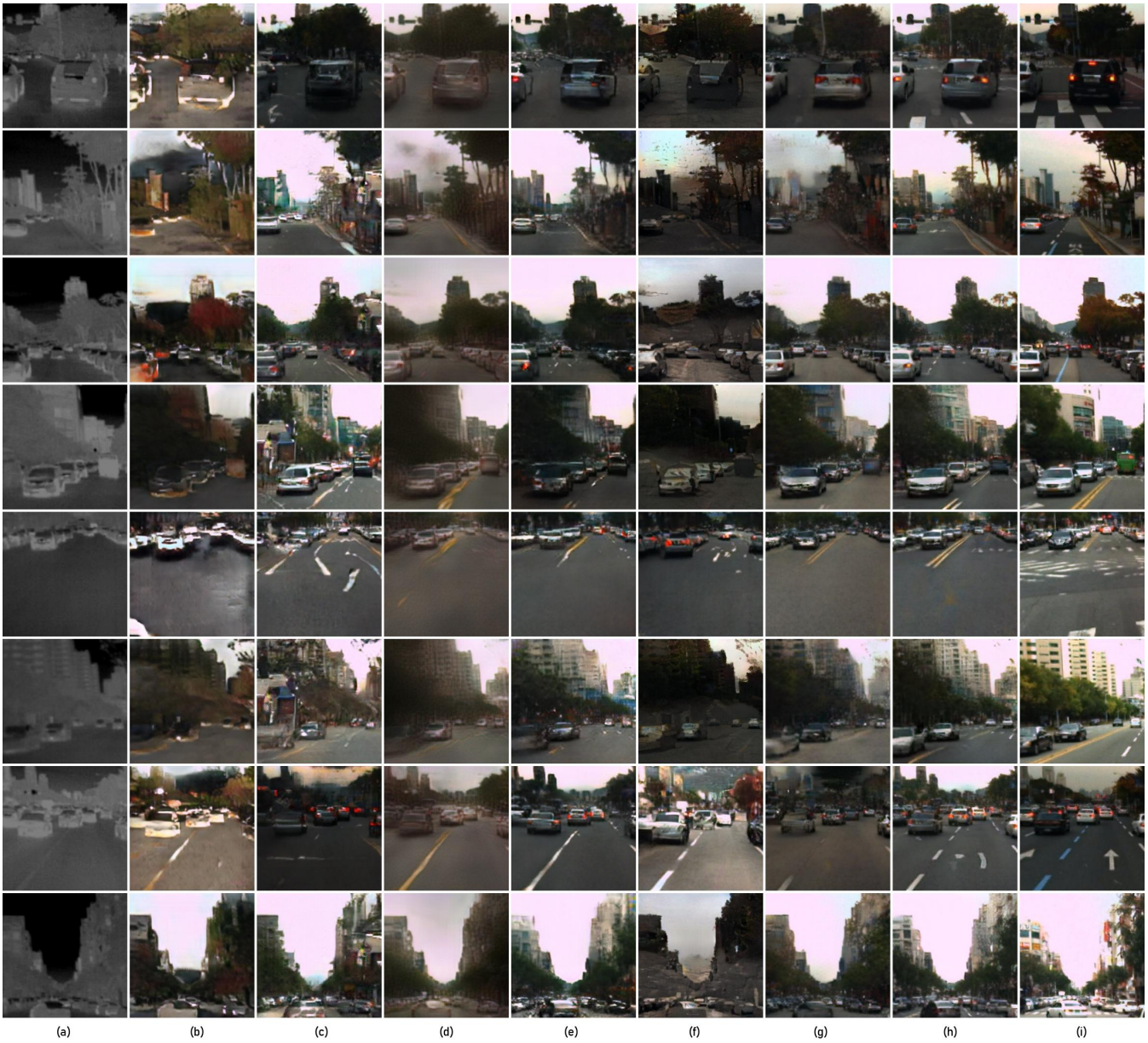

Fig. 11. **Colorized results using three different methods on the KAIST Dataset. (a) Thermal infrared images. (b) CycleGAN [60]. (c) Pix2pix [18]. (d) SCGAN [27]. (e) TIC-CGAN [19]. (f) PealGAN [36]. (g) MUGAN [32]. (h) LKAT-GAN. (i) True RGB images. (As shown in (h), the details and colors of cars, roads, buildings and other objects are well restored. In addition, the results of LKAT-GAN are clearer from the overall effect of the image.)**

compared with the comparative methods in both objective and subjective evaluation on KAIST Dataset.

D. Failure cases

Although our network performs well on most testing images, it also fails in some scenarios. Fig. 10. shows some common types of failures. The first type of failure cases is because the buildings in thermal infrared images have complex shapes and colors, yet there are few similar scenes in the training set that leads to the network can only recover the basic shape and relatively simple color of the buildings. The second type of failure cases will occur when the car is turning. In thermal infrared images, the body of the turning car cannot be distinguished from the surrounding environment, resulting in the failure of colorization for the car. The third type is that the available information provided by thermal infrared images is too little, and the network cannot extract useful features that leads to the failure of colorization. Improving the quality of the dataset (increasing the number of images of the corresponding scenes and improving the quality of images) maybe can effectively solve the above problems.

E. Experiments on IRVI Dataset

We compare I2VGAN [39], SCGAN [27], TICC-GAN [19], PealGAN [36], and MUGAN [32] with our method on IRVI dataset as a supplementary experiment. Due to the limitation of the number of images on the dataset, we train Traffic and Monitoring scenarios together, then test different models

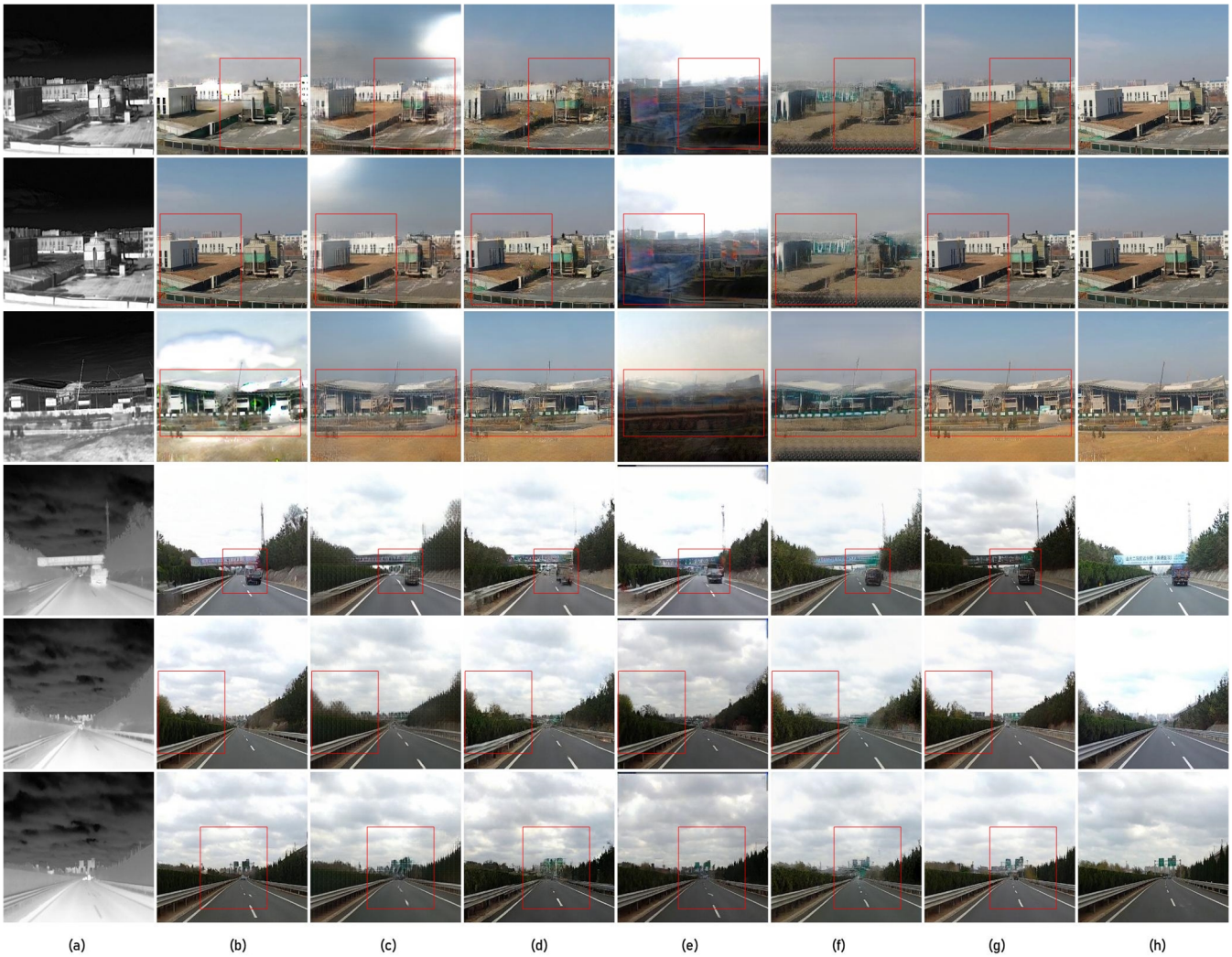

Fig. 12. **The colorized results of IRVI Dataset. (a) Thermal infrared images. (b) I2VGAN [39]. (c) TICC-GAN [19]. (d) SCGAN [27]. (e) PealGAN [36]. (f) MUGAN [32]. (g) LKAT-GAN. (h) True RGB images. The differences in the results are circled in red.** (As shown in (g), the clarity, color, and detail of the results are the best in both the Monitoring and Traffic scenarios.)

on Traffic and Monitoring scenarios. As shown in Table III, our method achieve state-of-the-art performance in Monitoring scenario, and it also performed well in Traffic scenario. Our model is more complex and contains ViT module, and the dataset is not large enough, which leads to the over-fitting of the network. It leads the performance of MUGAN is higher than that of our method in Traffic scenario, but LKAT-GAN can effectively process the global and local feature information of the thermal infrared image, the result looks more natural in detail and color for MUGAN. More importantly, MUGAN performs poorly in the Monitoring scenario, which indicates that our model has better robustness in multiple scenarios. As shown in Fig. 12, the first three lines show the colorized results of the different models in Monitoring scenario and the last three lines show the results of the different models in Traffic scenario. The results of SCGAN show distinct white spots and blurred architectural details. For I2VGAN, the color of the result is not realistic enough and the overall quality is poor. TICCGAN has achieved good results in color restoration, but the details of the lawn and building are not satisfactory. For PealGAN, it performs well in Traffic scenario as an unsupervised model, but performs poorly in Monitoring scenario. MUGAN shows its superiority in Traffic scenario, the edge details and overall color of the results are accurate. However, MUGAN fails in Monitoring scenario. Compared with these methods, the proposed method achieves the best results in both color and details, such as vehicles, trees, and traffic signs both in Monitoring and Traffic scenarios.

V. Conclusion

In this work, we propose a novel method called LKAT-GAN for thermal infrared image colorization. Benefitting from the LK_U-Net and the Attention_U-Net-Transformer branch structure, LKAT-GAN can effectively extract the ambiguous feature information of thermal infrared images and correctly learn the semantic information in the images. Moreover, we propose the LK_U-Net to extract the high-level semantic and multi-scale local features and reduce the loss of input image information. Besides, a branch structure is also arranged to decode the features from different perspectives to improve the

representational capacity of the network. The experimental results on KAIST and IRVI datasets demonstrate the superiority of LKAT-GAN over other methods for the thermal infrared image colorization tasks. Meanwhile, the results of the LKAT-GAN visualization show some phenomena, such as it can sometimes give better color than the ground truth and generate more fine detail by retrieving extra structural information from the IR input, and it also happens that the output has completely wrong contents but looks authentic, due to the nature of neural network based generative model. These characteristics also show the advantages of LKAT-GAN in the application of thermal infrared image colorization. In the future, the colorization of thermal infrared image at night and the actual deployment application of the network are the directions of our future research.

REFERENCES

[1] F. Erden and A. E. Çetin, "Hand gesture based remote control system using infrared sensors and a camera," *IEEE Transactions on Consumer Electronics*, vol. 60, no. 4, pp. 675–680, 2014. I

[2] R. Gade and T. B. Moeslund, "Thermal cameras and applications: a survey," *Machine vision and applications*, vol. 25, no. 1, pp. 245–262, 2014. I

[3] S. Huang, X. Jin, Q. Jiang, and L. Liu, "Deep learning for image colorization: Current and future prospects," *Engineering Applications of Artificial Intelligence*, vol. 114, p. 105006, 2022. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0952197622001920 I

[4] R. G. Wijnhoven and P. H. de With, "Identity verification using computer vision for automatic garage door opening," *IEEE Transactions on Consumer Electronics*, vol. 57, no. 2, pp. 906–914, 2011. I

[5] J. Kang, D. V. Anderson, and M. H. Hayes, "Face recognition for vehicle personalization with near infrared frame differencing," *IEEE Transactions on Consumer Electronics*, vol. 62, no. 3, pp. 316–324, 2016. I

[6] C. Chen, Y. Xu, and X. Yang, "User tailored colorization using automatic scribbles and hierarchical features," *Digital Signal Processing*, vol. 87, pp. 155–165, 2019. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S1051200418302033 I

[7] Y. Tang, M. Zhu, Z. Chen, C. Wu, B. Chen, C. Li, and L. Li, "Seismic performance evaluation of recycled aggregate concrete-filled steel tubular columns with field strain detected via a novel mark-free vision method," *Structures*, vol. 37, pp. 426–441, 2022. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S2352012421011747 I

[8] C. Zou, H. Mo, C. Gao, R. Du, and H. Fu, "Language-based colorization of scene sketches," *ACM Trans. Graph.*, vol. 38, no. 6, December 2019. [Online]. Available: https://doi.org/10.1145/3355089.3356561 I

[9] Y.-G. Shin, K.-A. Choi, S.-T. Kim, and S.-J. Ko, "A novel single ir light based gaze estimation method using virtual glints," *IEEE Transactions on Consumer Electronics*, vol. 61, no. 2, pp. 254–260, 2015. I

[10] J. Ma, L. Tang, F. Fan, J. Huang, X. Mei, and Y. Ma, "Swinfusion: Cross-domain long-range learning for general image fusion via swin transformer," *IEEE/CAA Journal of Automatica Sinica*, vol. 9, no. 7, pp. 1200–1217, 2022. I

[11] M. Dias, J. Monteiro, J. Estima, J. Silva, and B. Martins, "Semantic segmentation and colorization of grayscale aerial imagery with w-net models," *Expert systems*, vol. 37, no. 6, p. e12622, 2020. I

[12] R. K. Gupta, A. Y.-S. Chia, D. Rajan, E. S. Ng, and H. Zhiyong, "Image colorization using similar images," in *Proceedings of the 20th ACM International Conference on Multimedia*, ser. MM '12. New York, NY, USA: Association for Computing Machinery, 2012, p. 369–378. [Online]. Available: https://doi.org/10.1145/2393347.2393402 I

[13] S. Iizuka, E. Simo-Serra, and H. Ishikawa, "Let there be color! joint end-to-end learning of global and local image priors for automatic image colorization with simultaneous classification," *ACM Trans. Graph.*, vol. 35, no. 4, jul 2016. [Online]. Available: https://doi.org/10.1145/2897824.2925974 I, I, II-A

[14] A. Royer, A. Kolesnikov, and C. H. Lampert, "Probabilistic image colorization," *arXiv preprint arXiv:1705.04258*, 2017. I

[15] A. Deshpande, J. Lu, M.-C. Yeh, M. Jin Chong, and D. Forsyth, "Learning diverse image colorization," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, July 2017. I

[16] G. Larsson, M. Maire, and G. Shakhnarovich, "Learning representations for automatic colorization," in *European conference on computer vision*. Springer, 2016, pp. 577–593. I

[17] K. He, X. Chen, S. Xie, Y. Li, P. Dollár, and R. Girshick, "Masked autoencoders are scalable vision learners," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, June 2022, pp. 16 000–16 009. I, III-C

[18] 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 (CVPR)*, July 2017. I, II-A, II-A, II-B, III-D, III-D, II, IV-C, 11

[19] X. Kuang, J. Zhu, X. Sui, Y. Liu, C. Liu, Q. Chen, and G. Gu, "Thermal infrared colorization via conditional generative adversarial network," *Infrared Physics & Technology*, vol. 107, p. 103338, 2020. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S1350449519311387 I, II-B, II-B, III-E, IV-A, II, III, IV-C, 11, IV-E, 12

[20] A. Levin, D. Lischinski, and Y. Weiss, "Colorization using optimization," in *ACM SIGGRAPH 2004 Papers*, ser. SIGGRAPH '04. New York, NY, USA: Association for Computing Machinery, 2004, p. 689–694. [Online]. Available: https://doi.org/10.1145/1186562.1015780 II-A

[21] E. Reinhard, M. Adhikhmin, B. Gooch, and P. Shirley, "Color transfer between images," *IEEE Computer graphics and applications*, vol. 21, no. 5, pp. 34–41, 2001. II-A

[22] Z. Cheng, Q. Yang, and B. Sheng, "Deep colorization," in *Proceedings of the IEEE International Conference on Computer Vision (ICCV)*, December 2015. II-A

[23] G. Larsson, M. Maire, and G. Shakhnarovich, "Learning representations for automatic colorization," in *Computer Vision – ECCV 2016*, B. Leibe, J. Matas, N. Sebe, and M. Welling, Eds. Cham: Springer International Publishing, 2016, pp. 577–593. II-A, II-A

[24] J. Zhao, J. Han, L. Shao, and C. G. Snoek, "Pixelated semantic colorization," *International Journal of Computer Vision*, vol. 128, no. 4, pp. 818–834, 2020. II-A, II-A

[25] J. Zhao, L. Liu, C. G. Snoek, J. Han, and L. Shao, "Pixel-level semantics guided image colorization," *arXiv preprint arXiv:1808.01597*, 2018. II-A

[26] I. Goodfellow, J. Pouget-Abadie, M. Mirza, B. Xu, D. Warde-Farley, S. Ozair, A. Courville, and Y. Bengio, "Generative adversarial networks," *Commun. ACM*, vol. 63, no. 11, p. 139–144, oct 2020. [Online]. Available: https://doi.org/10.1145/3422622 II-A, II-A

[27] Y. Zhao, L.-M. Po, K.-W. Cheung, W.-Y. Yu, and Y. A. U. Rehman, "Scgan: Saliency map-guided colorization with generative adversarial network," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 31, no. 8, pp. 3062–3077, 2021. II-A, II-B, II, III, IV-C, 11, IV-E, 12

[28] J. Q. L. J. L. S.-J. W. P. Y. S. Huang Shanshan, Jin Xin, "A fully-automatic image colorization scheme using improved cyclegan with skip connections," *Multimedia Tools and Applications volume*, vol. 80, p. 26465–26492, July 2021. [Online]. Available: https://doi.org/10.1007/s11042-021-10881-5 II-A

[29] S. Zhao, Z. Zhang, R. Hong, M. Xu, H. Zhang, M. Wang, and S. Yan, "Crnet: Unsupervised color retention network for blind motion deblurring," ser. MM '22. New York, NY, USA: Association for Computing Machinery, 2022. [Online]. Available: https://doi.org/10.1145/3503161.3547962 II-A

[30] L. Tang, Y. Deng, Y. Ma, J. Huang, and J. Ma, "Superfusion: A versatile image registration and fusion network with semantic awareness," *IEEE/CAA Journal of Automatica Sinica*, vol. 9, no. 12, pp. 2121–2137, 2022. II-B

[31] A. Berg, J. Ahlberg, and M. Felsberg, "Generating visible spectrum images from thermal infrared," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR) Workshops*, June 2018. II-B, II-B

[32] H. Liao, Q. Jiang, X. Jin, L. Liu, L. Liu, S.-J. Lee, and W. Zhou, "Mugan: Thermal infrared image colorization using mixed-skipping unet and generative adversarial network," *IEEE Transactions on Intelligent Vehicles*, pp. 1–16, 2022. II-B, II-B, II, III, IV-C, 11, IV-E, 12

[33] R. Gao, X. Hou, J. Qin, J. Chen, L. Liu, F. Zhu, Z. Zhang, and L. Shao, "Zero-vae-gan: Generating unseen features for generalized and transductive zero-shot learning," *IEEE Transactions on Image Processing*, vol. 29, pp. 3665–3680, 2020. II-B

[34] X. Gao, Z. Zhang, T. Mu, X. Zhang, C. Cui, and M. Wang, "Self-attention driven adversarial similarity learning network," *Pattern*

Recognition, vol. 105, p. 107331, 2020. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0031320320301345 II-B
[35] Y. Wei, Z. Zhang, Y. Wang, M. Xu, Y. Yang, S. Yan, and M. Wang, "Deraincyclegan: Rain attentive cyclegan for single image deraining and rainmaking," *IEEE Transactions on Image Processing*, vol. 30, pp. 4788–4801, 2021. II-B
[36] F. Luo, Y. Li, G. Zeng, P. Peng, G. Wang, and Y. Li, "Thermal infrared image colorization for nighttime driving scenes with top-down guided attention," *IEEE Transactions on Intelligent Transportation Systems*, 2022. II-B, II, III, IV-C, 11, IV-E, 12
[37] A. Anoosheh, T. Sattler, R. Timofte, M. Pollefeys, and L. Van Gool, "Night-to-day image translation for retrieval-based localization," in *2019 International Conference on Robotics and Automation (ICRA)*. IEEE, 2019, pp. 5958–5964. II-B
[38] A. Bansal, S. Ma, D. Ramanan, and Y. Sheikh, "Recycle-gan: Unsupervised video retargeting," in *Proceedings of the European Conference on Computer Vision (ECCV)*, September 2018. II-B
[39] S. Li, B. Han, Z. Yu, C. H. Liu, K. Chen, and S. Wang, "I2v-gan: Unpaired infrared-to-visible video translation," in *Proceedings of the 29th ACM International Conference on Multimedia*, 2021, pp. 3061–3069. II-B, IV-A, III, IV-E, 12
[40] J.-B. Cordonnier, A. Loukas, and M. Jaggi, "On the relationship between self-attention and convolutional layers," *arXiv preprint arXiv:1911.03584*, 2019. II-C
[41] Q. Han, Z. Fan, Q. Dai, L. Sun, M.-M. Cheng, J. Liu, and J. Wang, "Demystifying local vision transformer: Sparse connectivity, weight sharing, and dynamic weight," *arXiv preprint arXiv:2106.04263*, 2021. II-C
[42] G. Hinton, "How to represent part-whole hierarchies in a neural network," *arXiv preprint arXiv:2102.12627*, 2021. II-C
[43] F. Wu, A. Fan, A. Baevski, Y. N. Dauphin, and M. Auli, "Pay less attention with lightweight and dynamic convolutions," *arXiv preprint arXiv:1901.10430*, 2019. II-C
[44] X. Zhu, D. Cheng, Z. Zhang, S. Lin, and J. Dai, "An empirical study of spatial attention mechanisms in deep networks," in *Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV)*, October 2019. II-C
[45] X. Ding, X. Zhang, J. Han, and G. Ding, "Scaling up your kernels to 31x31: Revisiting large kernel design in cnns," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, June 2022, pp. 11 963–11 975. II-C, II-C, III-B, III-B
[46] A. Krizhevsky, I. Sutskever, and G. E. Hinton, "Imagenet classification with deep convolutional neural networks," *Communications of the ACM*, vol. 60, no. 6, pp. 84–90, 2017. II-C
[47] C. Szegedy, S. Ioffe, V. Vanhoucke, and A. A. Alemi, "Inception-v4, inception-resnet and the impact of residual connections on learning," in *Thirty-first AAAI conference on artificial intelligence*, 2017. II-C
[48] C. Szegedy, V. Vanhoucke, S. Ioffe, J. Shlens, and Z. Wojna, "Rethinking the inception architecture for computer vision," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, June 2016. II-C
[49] Z. Liu, Y. Lin, Y. Cao, H. Hu, Y. Wei, Z. Zhang, S. Lin, and B. Guo, "Swin transformer: Hierarchical vision transformer using shifted windows," in *Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV)*, October 2021, pp. 10 012–10 022. II-C, III-B
[50] D. Ulyanov, A. Vedaldi, and V. Lempitsky, "Instance normalization: The missing ingredient for fast stylization," *arXiv preprint arXiv:1607.08022*, 2016. III-B
[51] D. Hendrycks and K. Gimpel, "Gaussian error linear units (gelus)," *arXiv preprint arXiv:1606.08415*, 2016. III-B
[52] A. Dosovitskiy, L. Beyer, A. Kolesnikov, D. Weissenborn, X. Zhai, T. Unterthiner, M. Dehghani, M. Minderer, G. Heigold, S. Gelly *et al.*, "An image is worth 16x16 words: Transformers for image recognition at scale," *arXiv preprint arXiv:2010.11929*, 2020. III-B
[53] X. Ding, H. Chen, X. Zhang, J. Han, and G. Ding, "Repmlpnet: Hierarchical vision mlp with re-parameterized locality," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, June 2022, pp. 578–587. III-B
[54] I. O. Tolstikhin, N. Houlsby, A. Kolesnikov, L. Beyer, X. Zhai, T. Unterthiner, J. Yung, A. Steiner, D. Keysers, J. Uszkoreit, M. Lucic, and A. Dosovitskiy, "Mlp-mixer: An all-mlp architecture for vision," in *Advances in Neural Information Processing Systems*, M. Ranzato, A. Beygelzimer, Y. Dauphin, P. Liang, and J. W. Vaughan, Eds., vol. 34. Curran Associates, Inc., 2021, pp. 24 261–24 272. [Online]. Available: https://proceedings.neurips.cc/paper/2021/file/cba0a4ee5ccd02fda0fe3f9a3e7b89fe-Paper.pdf III-B
[55] Z. Niu, G. Zhong, and H. Yu, "A review on the attention mechanism of deep learning," *Neurocomputing*, vol. 452, pp. 48–62, 2021. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S092523122100477X III-C
[56] S. Woo, J. Park, J.-Y. Lee, and I. S. Kweon, "Cbam: Convolutional block attention module," in *Proceedings of the European Conference on Computer Vision (ECCV)*, September 2018. III-C, 6
[57] K. Simonyan and A. Zisserman, "Very deep convolutional networks for large-scale image recognition," *arXiv preprint arXiv:1409.1556*, 2014. III-E
[58] S. Hwang, J. Park, N. Kim, Y. Choi, and I. So Kweon, "Multispectral pedestrian detection: Benchmark dataset and baseline," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, June 2015. IV-A
[59] D. P. Kingma and J. Ba, "Adam: A method for stochastic optimization," *arXiv preprint arXiv:1412.6980*, 2014. IV-A
[60] A. Almahairi, S. Rajeshwar, A. Sordoni, P. Bachman, and A. Courville, "Augmented CycleGAN: Learning many-to-many mappings from unpaired data," in *Proceedings of the 35th International Conference on Machine Learning*, ser. Proceedings of Machine Learning Research, J. Dy and A. Krause, Eds., vol. 80. PMLR, 10–15 Jul 2018, pp. 195–204. [Online]. Available: https://proceedings.mlr.press/v80/almahairi18a.html II, IV-C, 11

Xin Jin received the B.S. degree in electronics and information engineering from Henan Normal University, Xinxiang, China, in 2013, and the Ph.D. degree in communication and information systems from Yunnan University, Kunming, China, in 2018. He was a Post-Doctoral Fellow with the School of Software, Yunnan University from 2018 to 2020. He is an Associate Professor with the School of Software, Yunnan University. His research interests include pulse coupled neural networks and its applications, image processing, information fusion, optimization algorithm, and fuzzy set theory.

Qian Jiang received the B.S. degree in thermal energy and power engineering and the M.S. degree in power engineering and engineering thermo-physics from Central South University (CSU), Changsha, China, in 2012 and 2015, respectively, and the Ph.D. degree in communication and information systems from Yunnan University, Kunming, China, in 2019. She was a Post-Doctoral Fellow with the School of Software, Yunnan University from 2019 to 2021. She is an Associate Professor with the School of Software, Yunnan University. Her research interests include deep neural networks, fuzzy set theory, bio-informatics, image processing, and information fusion.

Puming Wang received the B.E. degree in Communication Engineering from Xidian University, Xian, China and the M.E. in Control Engineering from Wuhan University of Technology, Wuhan, China. He received the Ph.D degree in School of Computer Science and Technology at Huazhong University of Science and Technology, Wuhan, China. He is an associate professor in the School of Software at Yunnan University, Kunming, China. His research interests include big data, artificial intelligence, and network security.