Fusion of Infrared-Visible Images in UE-IoT for Fault Point Detection Based on GAN

BIN LIAO, YOU DU[ID], AND XIANGYUN YIN

School of Electrical and Electronic Engineering, North China Electric Power University, Beijing 102206, China

Corresponding author: You Du (dy815@ncepu.edu.cn)

ABSTRACT In the development of modern intelligent Ubiquitous Electric Internet of Things (UE-IoT), infrared thermal imaging always plays an important role in automated early-warning detection of developing failures of critical assets such as transformers, disconnects and capacity banks in electrical power substation non-intrusively. However, the low resolution and contrast of infrared images hinder the subsequent analysis and recognition of fault points. In contrast, visible images present abundant texture details of the equipment without thermal information. In order to assist the detection of fault points, this paper proposes a Generative adversarial networks (GAN) based infrared and visible image fusion method to produce a composite image with enhanced edges and better quality. The edge loss function is added to represent the perceptual edges. In the discriminator, the proposed method improves the texture similarity between fusion image and visible image by minimizing the Wasserstein distance in VGG (Visual Geometry Group network) feature space. The experimental results show that the fault regions become more salient and the details are enhanced. In this way, it can facilitate the detection of fault points both reliably and accurately.

INDEX TERMS Image fusion, VGG, generative adversarial networks, ubiquitous electric Internet of Things.

I. INTRODUCTION

Since the security and reliability of power infrastructure is growing more important than ever before, preventing the breakdown of power devices due to different faults becomes a main concern. Considering that the complexity between the symptoms and faults, it is not accurate to perform fault diagnosis based on single information source. Worse, it may lead to false or missed judgements [1]. Therefore, the comprehensive analysis and evaluation of multiple information sources is needed. In recent years, the proliferation of Ubiquitous Electric Internet of Things (UE-IoT) in electric power system has led to the deployment of various sensor devices, which transfigure traditional devices or environment becoming more intelligent. Consequently, improved fault detection and remote monitoring can be implemented.

Intelligent inspection robots and unmanned aerial vehicles have been widely used as terminal nodes in the inspection of substations and overhead lines in the Internet of Things [2]. These robots and unmanned aerials often carry infrared and visible light dual sensors to collect relevant equipment images and then sent them to cloud platforms for analysis and processing [3]. Infrared sensor can measure the thermal radi-

ation of the object non-intrusively in real time, wherefore is often used for fault detection [4]. Lopez-Perez and Antonino-Daviu [5] applied infrared images to the fault detection of induction motors. Zhou *et al.* [6] designed a fault diagnosis and remote detection system to realize the analysis and processing of the infrared image of the insulator. Qin *et al.* [7] used infrared thermal imaging technology to monitor the process of electrical performance testing and designed an infrared monitoring system for thermal batteries. Cui *et al.* [8] designed a system to monitor the electrical equipment based on infrared images in UE-IoT. Although infrared images are widely used for fault detection in UE-IoT, relying on the single infrared image obtained by infrared sensor is not accurate enough. The infrared image has fewer details, low contrast and resolution, which leads to inaccurate fault detection. In contrast, the visible image obtained by the optical sensor retains abundant texture information of power equipment. Therefore, by performing edge computing at the terminal and using image fusion algorithms, the useful information from different sensors can be integrated together, which can not only facilitate fault identification and detection [9], but also compress data and save communication traffic [10].

For decades, numerous infrared and visible image fusion methods have been proposed. These methods can be mainly divided into five categories: subspace [11], multi-scale

transform [12]–[15], sparse representation [16], [17], hybrid models [18], [19], and the convolution neutral network (CNN) [20]–[22] based methods. Subspace based fusion methods convert the whole images into uncorrelated components with non-parametric techniques such as PCA or ICA, then the PCA or ICA coefficients matrices are weighted by the Piella metric. Multi-scale transform based methods are the most widely used methods in image fusion, which decompose the source images into several sub-bands at different scales, and then certain fusion rules are applied to fuse the sub-bands. Finally the fusion image is reconstructed by performing inverse transform [23]. Multi-scale transform, including pyramid transform [12], Wavelet transform [13], non-subsampled Contourlet transform (NSCT) [14], and non-subsampled Shearlet transform (NSST) are all introduced in image fusion. Although multi-scale transform based fusion method has ever been the mainstream, they fail to capture the texture details efficiently compared with edge features. The reason is that they adopt the fixed basis functions. Therefore, ringing artifacts are always visible in the fusion images. Unlike the multi-scale transform based fusion methods, sparse representation based methods divide the source images into overlapped images patches, then sparsely represent the image patches as linear combination of atoms from the learned over-complete dictionary. After that, the representation coefficients are fused according to designed fusion rules. Finally, the fusion image is generated through reconstruction algorithm. Sparse representation based methods can express the structured features more efficiently, benefitting from more flexible choice of the basis functions. ASR [16] and CSR [17] are typical examples of this category. Hybrid model-based methods synthesize the advantages of different schemes to improve the fusion quality. Ding *et al.* [18] introduced a fusion method based on the Shearlet transform and sparse PCA features.

In addition to the above methods, the infrared-visible image fusion method based on CNN is also receiving great attention in recent years. Liu *et al.* [20] proposed a new method for infrared and visible image fusion based on convolutional neural network (CNN). Li and Wu [21] studied an auto-encoder-decoder network, in which dense blocks were used in the encoder to extract the features and these features were fused in the decoder to obtain the fusion image. Yu Zhang *et al.* [22] proposed a general image fusion framework based on CNN, which adopted two convolutional layers to extract the salient image features from multiple input images and then fuse these features via feature fusion rules. Finally two convolutional layers were used to combine and reconstruct these features to obtain the fusion image. The CNN based fusion methods can extract more meaningful features and learn the parameters of filters for the image fusion task.

Above mentioned fusion methods for infrared and visible image perform the same representation framework for all input images, this is not appropriate for specific fault detection task. In fact, the thermal information in the infrared image should be preserved in the fusion image, because it reflects the important thermal radiation of devices. On the other hand, the visible image shows the texture structures of devices and should also be reinforced in the fusion image. In addition, most fusion methods in the literature choose fusion rules artificially, which seems a little rigid and lack of adaptability.

Considering the application of image fusion for the fault detection in the UE-IoT, this paper proposes an image fusion method of infrared and visible images based on improved Wasserstein Generative Adversarial Network (WGAN-gp). The GAN include two parts: the generator and the discriminator, contending with each other in a mini-max game. Later, WGAN [24] was proposed to speed the performance and maintain the stability of the training process. GAN shows great potential to generate realistic images and has been introduced in various vision tasks, such as image registration [25], image generation [26], super-resolution [27] and image denoising [28] etc. In our design, the generator is supposed to preserve the thermal information and the visible gradients as much as possible, while the discriminator takes the texture features in the visible image as the constraints. To enforce the GAN convergence faster, gradient penalty is added during the discriminator update.

The main contributions of this paper are as follows:

(1) Our proposed image fusion method is based on GANs, which is a fully end-to-end model and the fusion image can be generated without explicitly designing fusion rules.

(2) The VGG network is employed in the update process of the generator and the discriminator, which help to improve the texture similarity between fusion image and visible image.

(3) Considering that the temperature of the equipment is an important basis for fault detection, we design a specialized loss function for infrared and visible image fusion in fault detection, in which the thermal information is given more weight than detail information.

The rest of this paper is organized as follows. The background and related works are presented in Section II. Section III describes the proposed fusion method. The experimental results and discussion are given in Section IV. Conclusions are presented in Section V.

II. RELATED WORK
A. THE UBIQUITOUS ELECTRIC INTERNET OF THINGS
Ubiquitous Electric Internet of Things (UE-IoT) is a smart service system, which implements modern information technology and advanced communication technologies. It not only realizes the interconnection of all things and human-computer interaction in all links of the power system, but also realizes the identification, perception, interconnection and control of grid infrastructure, people and their environment. The UE-IoT architecture is the same as other forms of IoT systems, which can be divided into four layers: perception layer, network layer, platform layer, and application layer [29]. The network architecture is shown in Fig. 1.
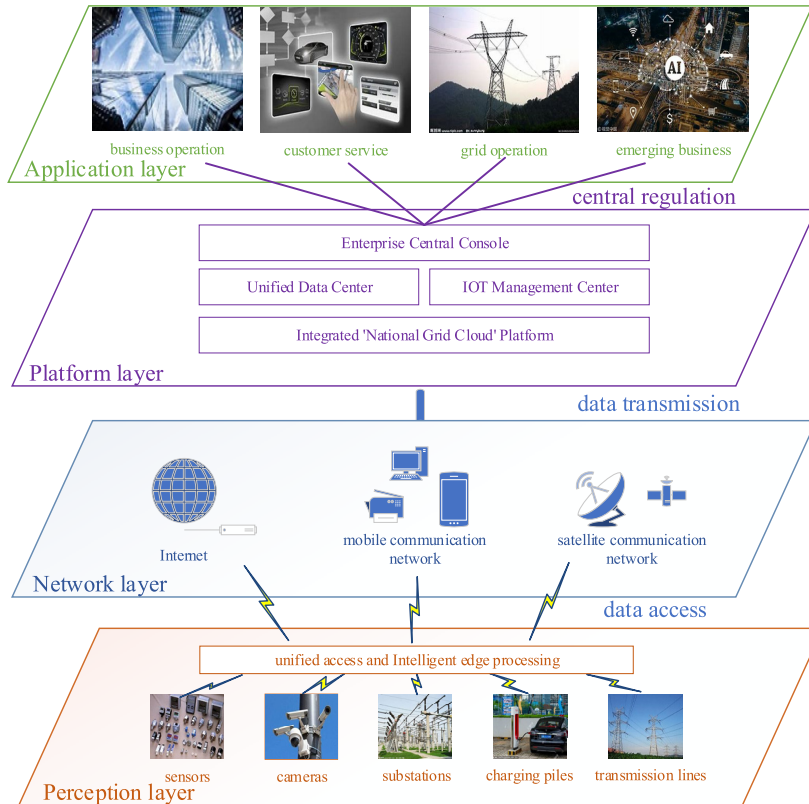
FIGURE 1. Architecture of the ubiquitous electric internet of things.

1) PERCEPTION LAYER

The perception layer is the bottom layer of the ubiquitous electric Internet of Things, and it is also the hub that connects field devices and the network layer. Its main function is to realize information collection and signal processing [30]. Due to the large number of field device terminals to be connected to the perception layer, the terminal operating system is not uniform, and the data provided by the terminal needs to be integrated and processed by edge computing before reaching the platform layer.

2) NETWORK LAYER

The network layer is the second layer in UE-IoT, which consist of access network and transmission network, where the former undertakes the task of data access while the latter realizes the function of information transmission. As a link, the network layer connects the perception layer and the application layer, responsible for transmitting the information obtained by the perception layer to the application layer safely and reliably, and then performs information processing according to different application requirements.

3) PLATFORM LAYER

The platform layer is located above the perception layer and the network layer, and below the application layer. It is the core of the UE-IoT, which is expected to integrate the

huge amount of information resources in the network into a large interconnected network through computing power, and solve problems such as data storage, data mining, and privacy protection.

4) APPLICATION LAYER

The application layer is the top layer of the UE-IoT, and its role is to analyze the data transmitted from the network layer and apply it to some practical scenarios to provide users with a series of services.

B. GENERATIVE ADVERSARIAL NETWORKS

Generative adversarial networks (GANs) are algorithmic architectures proposed by Ian Goodfellow [31],which involve a generator and a discriminator. The generator is trained to produce plausible image with similar characteristics as the input samples, while the discriminator learns to identify the fake images from the real images and penalizes the generator for producing the fake samples. Such adversarial learning process between generator and discriminator continues until the generated data distribution cannot be distinguished from the real one.

Although the original GAN has achieved impressive results in generating realistic images, it is challenging to train a GAN model, confronting with the problems of training instability and non-convergence. In order to solve above

problems, many improved versions of GANs have been proposed. To make the adversarial network training process controllable, the conditional Generative Adversarial Networks (cGAN) [32] introduced conditional variables in training process. In order to solve the gradient vanishing problemin original GANs, the Least Square GANs (LSGAN) [33] adopted the least squares loss function for discriminator, thus performed morestable than original GANs. It was argued [24] that the JS divergence used in original GANs may result in gradient vanishing and wrong optimization direction, so Wasserstein divergence was adopted instead of JS divergence to measure the distance between two distributions, which greatly improved the stability of GANs. However, in practical applications, Wasserstein GAN (WGAN) still suffers from unstable training and slow convergence, and the improvement is not obvious in experiments compared to traditional GANs. As an advanced version of WGAN, namely the WGAN-gp, Ishaan *et al.* [34] pointed out that the reason why such problems appear was that WGAN used weight clipping directly when enforcing Lipschitz constraints. Thus, in WGAN-gp, gradient penalty was applied to meet Lipschitz continuity constraints, which made training process more stable and converge faster than WGAN.

III. PROPOSED METHOD

A. MOTIVATION

Although the existing fault detection methods of power equipment in UE-IoT have achieved remarkable success, there still exist some problems. Most fault detection methods only consider infrared images of power equipment. However, the low resolution of infrared images and blurred target edges make it difficult to locate the fault accurately. On the other hand, the fast development of information technologies puts forward higher requirements for the accuracy and speed of fault detection technology for UE-IoT power equipment.

To integrate the thermal radiation distribution from infrared image and the texture details from visible image into one fusion image, we propose an infrared and visible image fusion method based on improved Wasserstein generative adversarial networks. In order to preserve more edge information in the visible image, we use the first three convolutional layers of VGG19 networks [35] as feature extractors, and calculate the edge loss between extracted features of the generated image and the visible image.

B. METHOD

We assume the fusion of infrared and visible images as the adversarial learning problem between the generator and the discriminator in the GAN. Our work can be divided into two parts, namely training phase and testing phase. As shown in Fig. 2. In the training phase, firstly, a pair of infrared and visible image is stacked and fed into the generator. The generator extracts features from the input images and integrates them, and then generates the fusion image under the guidance of the loss function. Next, features of the fusion image and the

visible image extracted by the trained VGG19 are fed into the discriminator. The discriminator will minimize the difference between the fusion image and the visible image in the feature spaces by the Wasserstein distance. This adversarial learning process between the generator and the discriminator continues until the maximum number of iterations of the adversarial network is reached. In the testing phase, we only input the infrared image and visible image pair into the trained generator, the output of the generator is the final fusion image.

C. ARCHITECTURE

1) GENERATOR ARCHITECTURE

As shown in Fig. 3 (a), our generator architecture is composed of five convolution layers. First of all, the stacked infrared image and visible image are fed into the first layer, where 7×7 filters are applied to extract low level features. Then, these features will be input to the following layer, in which 5×5 filters are used. In the third and fourth layer, we apply 3×3 filters to further extract the high level features. In order to keep the output image with the same size as the input image, the size of filters in the last layer is set to 1×1. Considering that the down sampling and padding operation will lead to the loss of image details and the artifacts around image boundary [36], convolution layer without any down sampling and padding is introduced. Besides, the activation function in each convolutional layer adopts leaky ReLU. Furthermore, the batch norm layer is used in the first four layers except the last one. All parameter settings of the generator are shown in TABLE 1.

2) DISCRIMINATOR ARCHITECTURE

The discriminator consists of four convolutional layers and a fully connected layer, as shown in Fig. 3 (b). The filter size of the first two layers is 5×5, whereas 3×3 filters are used in the second two layers. After each convolutional layer, we adopt the layer normalization rather than the batch normalization, considering that we only penalize the discriminator with respect to each input independently, not the entire batch. In addition, we use leaky ReLU activation function for first four layers, and the stride of filters is set to 2. It should be noted that the discriminator in traditional GANs is simply a classifier, which evaluates the probability of the input image to be real image. But the aim of discriminator in WGANs is to measure the distance between the input image distribution and the real image distribution, which can be regard as a linear regression problem. Thus we use linear layer as the last layer and remove the sigmoid layer. Parameter settings of the discriminator are listed in TABLE 1.

D. LOSS FUNCTION

1) GENERATOR LOSS

Since the generator is designed to generate a fusion image which preserves as much thermal information in the infrared image and detail information in the visible image as possible, we update the generator under the guidance of this
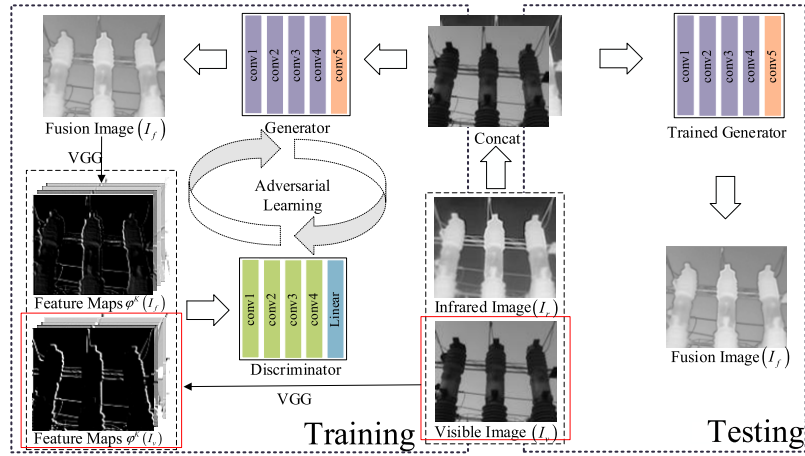
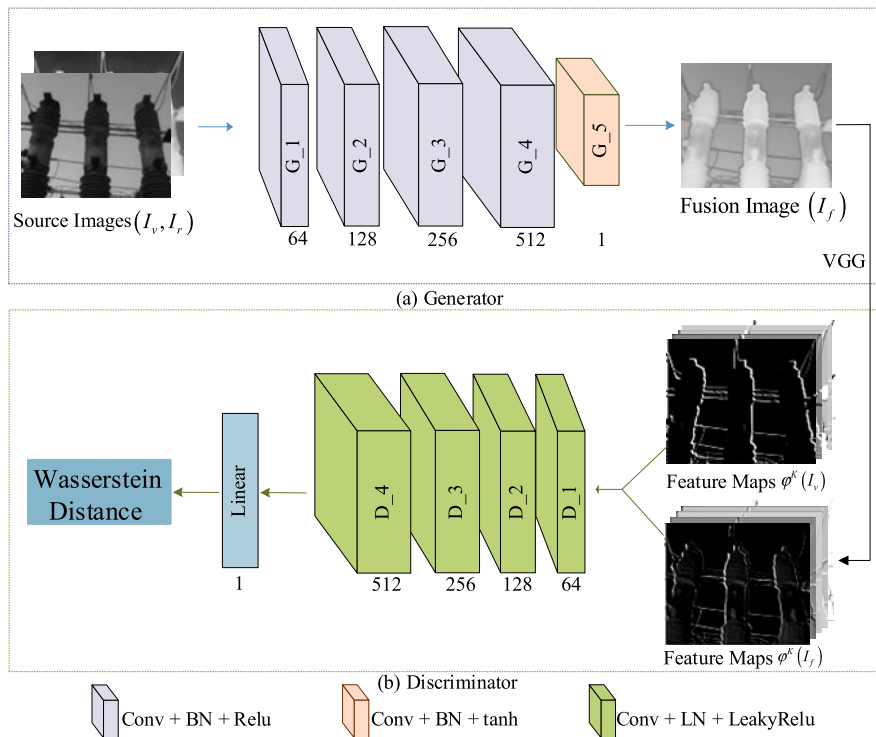**FIGURE 2.** Framework of the proposed method.

FIGURE 3. Architecture of the generator and discriminator.

information. The loss function of generator is designed as:

$$G_{loss} = L_{adversal} + \theta L_{edge} + \gamma L_{thermal} \tag{1}$$

As shown above, the generator loss consists of three parts: adversarial loss $L_{adversal}$, edge information loss L_{edge} and thermal information loss $L_{thermal}$, where θ and γ are weighting factors. Considering that the temperature variation is an important detection indicator for fault detection, we assume that the thermal information in the infrared image is more important than the details in the visible image for our purpose. Thus, we prefer to preserve the thermal information in the fusion image during the update of the generator, and we set θ to 1 and γ to 10.

The $L_{thermal}$ is the thermal information loss between the fusion image and the infrared image. Since it is well known that the thermal information of the infrared image is represented by its pixel intensity [37], we define the $L_{thermal}$ as follows:

$$L_{thermal} = \left\| I_f - I_r \right\|_F^2 \tag{2}$$

where the $\|\cdot\|_F$ denotes the matrix Frobenius norm, and the I_f and I_r represents the fusion image and infrared image, respectively.

To capture the texture and edge information in the visible image, style loss proposed in image style transfer task is employed in [38]. Through extracting the feature maps by the

TABLE 1. Parameter settings of generator and discriminator.

		Filter settings				Normalizer	Activation
		I_c	O_c	K_s	S_s		
Generator	G_1	2	64	7	1	BN	ReLU
	G_2	64	128	5	1	BN	ReLU
	G_3	128	256	3	1	BN	ReLU
	G_4	256	512	3	1	BN	ReLU
	G_5	512	1	1	1	-	LeakyReLU
Discriminator	D_1	64	64	5	2	LN	LeakyReLU
	D_2	64	128	5	2	LN	LeakyReLU
	D_3	128	256	5	2	LN	LeakyReLU
	D_4	256	512	5	2	LN	LeakyReLU
	Linear	512	1	-	-	-	-

BN and LN indicate the BatchNorm layer and LayerNorm layer, respectively. I_c, O_c, K_s and S_s represent the input channel, output channel, kernel size and stride of convolution layers, respectively.

VGG19 network, it evaluates the difference between feature maps from different layers.

$$L_{edge} = \frac{1}{K} \sum_{k=1}^{K} \frac{1}{C^k H^k W^k} \left\| G\left(\varphi^k\left(I_f\right)\right) - G\left(\varphi^k\left(I_v\right)\right) \right\|_F^2 \tag{3}$$

where the $G(\cdot)$ is gram matrix of feature map, and the $\varphi^k(\cdot)$ denotes the feature map in k-th layer of the trained VGG19. The size of the feature map in k-th layer is $C^k \times H^k \times W^k$. For the convenience of calculation, the feature maps are reshaped into $C^k \times H^k W^k$.

The $L_{adversal}$ represents the adversarial loss between generator and discriminator, which is defined as:

$$L_{adversal} = -\frac{1}{N} \sum_{n=1}^{N} D_{\theta_D}\left(\varphi^K\left(I_f^n\right)\right) \tag{4}$$

where the I_f^n denotes the n-th fusion image in training set, the $\varphi^K(\cdot)$ represents the feature map in the last layer extracted by the VGG19 and $D_{\theta_D}(\cdot)$ denotes the output of discriminator.

2) DISCRIMINATOR LOSS

To solve the problems of unstable and slow convergence during the GAN training process, the WGAN-gp [34] added gradient penalty term during update of the discriminator.

Inspired by that, we also introduce the gradient penalty in our work. Thus we fine-tune the loss function based on the original loss function in [34]:

$$D_{loss} = \frac{1}{N} \sum_{n=1}^{N} \left(D_{\theta_D}\left(\varphi^K\left(I_f^n\right)\right) - D_{\theta_D}\left(\varphi^K\left(I_v^n\right)\right) \right. \\ \left. + \lambda L_{penalty} \right) \tag{5}$$

where the first two items are included in the original loss function of discriminator. The $L_{penalty}$ is the gradient penalty term of discriminator, λ is a penalty coefficient. The $L_{penalty}$ is defined as:

$$L_{penalty} = \left(\nabla_\alpha D_{\theta_D}\left(\varphi^K\left(I_\alpha\right)\right) - 1 \right)^2 \tag{6}$$

where the ∇_α represents the gradient operation, α is a random number in (0, 1), and $I_\alpha = \alpha \cdot I_f^n + (1 - \alpha) \cdot I_v^n$. The I_α can be regarded as a random sample between visible image and the fusion image.

IV. EXPERIMENTS

To verify the advantages of the proposed fusion model, we compare our method with other five state-of-the-art methods. The compared methods include adaptive sparse represent (ASR) [16], cross bilateral filter (CBF) [39], latent low rank representation (LATLRR) [40], ResNet and zero-phase component analysis (ResZca) [41], visual saliency map and weighted least square optimization (WLS) [19]. All these methods can be implemented based on their public codes, and the parameters are set the same as in original papers. Experiments are performed on a workstation with 3.2GHz Intel Xeon CPU W-2104, GPU GeForce RTX 2080, and 8 GB memory.

A. TRAINING DETAILS

We select 49 pairs of infrared and visible images from TNO database[1] as our training data. However, it is far from enough to train a good model, so we crop each image into a set of 64×64 image patches by overlapping scheme (i.e., the stride is 16). Thus, we can obtain 45027 pairs of infrared and visible image patches. Since there is no padding operation in the convolutional layer, we pad these patches to the size of 78×78 before feeding them into the generator. In this way, the generated fusion image patch is the size of 78×78. Both generator and discriminator use the Adam [42] solver with the learning rate of 0.002. Penalty coefficient λ is set to 15 and the batch size is set to 1, the max iterations of the GAN network is set to 20000. The procedure and parameter settings are summarized in **Algorithm1**.

B. FUSION METRICS

In recent years, numerous subjective metrics have been proposed to evaluate the quality of the fusion image on the basis of human visual system. Subjective metrics are usually carefully designed by measuring the details or distortion

[1]https : //figshare.com/articles/TNO_Image_Fusion_Dataset/1008029.

Algorithm 1 Training Procedure of the Proposed Method

Parameters: $N_{iters} = 20000, N_{critic} = 5, N$: the Number of Patches $\lambda = 15, \theta = 1, \gamma = 10$

---

for $i = 1$ *to* N_{iters} **do**

   for $j = 1$ *to* N_{crttic} **do**

      for $n = 1$ *to* N **do**

         $I_f^n = G_{\theta_G}\left(I_v^n, I_r^n\right)$

         $L_{thermal}^n = \left\|I_f^n - I_r^n\right\|_F^2$

         $L_{edge}^n =$

         $\frac{1}{K}\sum_{k=1}^{K}\frac{1}{C^k H^k W^k}\left\|G\left(\varphi^k\left(I_f^n\right)\right) - G\left(\varphi^k\left(I_v^n\right)\right)\right\|_F^2$

         $I_\alpha = \alpha \cdot I_f^n + (1-\alpha) \cdot I_v^n$

         $L_{penalty} = \left(\nabla_\alpha D_{\theta_D}\left(\varphi^K\left(I_\alpha\right)\right) - 1\right)^2$

         $D_{loss} = D_{loss} + D_{\theta_D}\left(\varphi^K\left(I_f^n\right)\right) -$

         $D_{\theta_D}\left(\varphi^K\left(I_v^n\right)\right) + \lambda L_{penalty}$

         $G_{loss} =$

         $G_{loss} - D_{\theta_D}\left(\varphi^K\left(I_f^n\right)\right) + \theta L_{edge}^n + \gamma L_{thermal}^n$

      end

      $D_{loss} = D_{loss}/N$

      $G_{loss} = G_{loss}/N$

      Update discriminator by AdamOptimizer:

      $\nabla_{\theta_D}(D_{loss})$

   end

     Update generator by AdamOptimizer:

   $\nabla_{\theta_G}(G_{loss})$

end

---

in the fusion image. Here, we compare the performance of fusion methods under three evaluation indexes, including structural similarity (SSIM) [43], mutual information (MI) [44], and human perception inspired index Q_c [45]. Two sources images are referenced as A and B, and the fusion image is denoted as F. The image variable X can be A or B.

1) STRUCTURE SIMILARITY

SSIM is a common model that measures the structure similarity between the source images and the fusion image. It models distortion as a combination of three different factors: brightness, contrast and structure. The SSIM is defined as:

$$SSIM_{X,F} = \frac{2\mu_X\mu_F}{\mu_X^2 + \mu_F^2} \cdot \frac{2\sigma_X\sigma_F}{\sigma_X^2 + \sigma_F^2} \cdot \frac{\sigma_{XF}}{\sigma_X\sigma_F} \quad (7)$$

$$SSIM = SSIM_{A,F} + SSIM_{B,F} \quad (8)$$

where the μ and σ represent the mean and standard variance of the image patch. The σ_{XF} indicates the standard covariance correlation between the source and fusion image. Sliding windows are applied to calculate the SSIM of the whole image. The closer the value of SSIM is to 1, the better the quality of the image.

2) MUTUAL INFORMATION

Mutual information (MI) is a metric that measures how much information is transmitted from the source images to the fusion image. The MI is defined as:

$$Q_{MI} = MI_{AF} + MI_{BF} \quad (9)$$

where the MI_{AF} and MI_{BF} are the mutual information between the fusion image and the source image A and B, they are defined as follows:

$$MI_{XF} = \sum_{i=0}^{L}\sum_{j=0}^{L} P_{XF}(i,j)\log_2\frac{P_{XF}(i,j)}{P_X(i)P_F(j)} \quad (10)$$

where the P_{XF} represents the joint gray histograms between the source and the fusion image. The P_X and P_F denote the normalized gray histogram of the source image and the fusion image. The greater the Q_{MI} value indicates the better fusion image quality.

3) HUMAN PERCEPTION INSPIRED INDEX

Q_C is a human perception inspired index, it measures the visual effect of the fusion image by comparing the contrast characteristics of the fusion image and the source image. The Q_C is defined as follows:

$$Q_C = \sum_{i,j}\lambda_A(i,j)Q_{AF}(i,j) + \lambda_B(i,j)Q_{BF}(i,j) \quad (11)$$

where Q_{AF} and Q_{BF} represent the degree of contrast information in source image A and source image B retained in the fusion image, respectively. The λ_A and λ_B represent saliency maps of Q_{AF} and Q_{BF}.

C. OBJECTIVE COMPARISONS

In order to compare the fusion effects of various methods in different environments, we select four pairs of infrared and visible images of power equipment as experimental images, including *transformer*, *insulator*, *arrester* and *breaker* respectively. The sizes of these four pairs of images are 296×196, 496×500, 256×256, and 1080×900. The fusion results are shown in Fig. 4. For the "transformer", most methods have integrated the information well in the fusion image, except that the CBF-based method presents unexpected blocking artifacts around the pillar of the insulating sleeve and the ASR-based method shows obvious speckles near the fuel tank edge of the transformer. As to "insulator", the fusion image produced by LATLRR appears some artifacts around the transmission tower, and in the WLS based fusion image the edges of the transmission tower is more blurred than other fusion results. The ResZca and our proposed method perform well in first two pairs of image, but the ResZca based fusion image appears obvious pseudo-Gibbs effects in the neighborhood of the voltage stabilization ring in the third row and the wires in the fourth row, while our proposed method has avoid such artifacts. The texture and edge information in visible images are well preserved. Furthermore, the target in our fusion image is more highlight
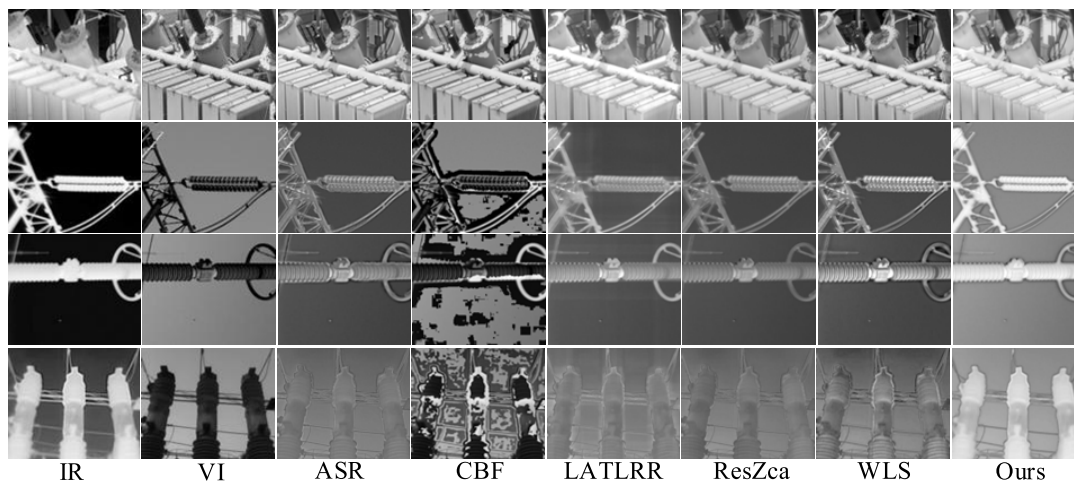
FIGURE 4. Visual result of various methods. From top to bottom: *transformer, insulator, arrester and breaker*. From left to right: infrared image, visible image, fusion images of ASR, CBF, LATLRR, ResZca, WLS and our method.

TABLE 2. Quantitative comparison of various methods.

Images	Metrics	ASR	CBF	LATLRR	ResZca	WLS	Ours
	SSIM	0.6295	0.5944	0.6058	0.6454	0.6348	**0.6485**
transformer	Q_{MI}	0.4525	0.4831	**0.4920**	0.4677	0.4346	0.4502
	Q_C	0.9051	**0.9535**	0.8696	0.8855	0.8692	0.8801
	SSIM	0.3653	0.3908	0.3928	0.3680	0.3640	**0.4402**
insulator	Q_{MI}	0.6543	0.7213	0.5274	0.6866	0.6646	**0.7424**
	Q_C	0.6487	0.6831	0.5322	0.5531	0.5786	**0.7645**
	SSIM	0.4949	0.3662	0.5104	0.5092	0.4834	**0.5176**
arrester	Q_{MI}	0.7472	0.7989	0.6445	0.8024	0.6953	**0.8252**
	Q_C	0.5987	0.5686	0.4957	0.5729	0.5459	**0.7237**
	SSIM	0.7222	0.4968	0.7223	**0.7273**	0.6887	0.6947
breaker	Q_{MI}	0.3873	0..7890	0.2967	0.3370	0.3452	**0.7892**
	Q_C	0.6258	0.7171	0.5931	0.6159	0.6564	**0.7359**

TABLE 3. Running time comparison of various methods (second).

Images	ASR	CBF	LATLRR	ResZca	WLS	Ours
transformer	95.0717	3.4792	22.2991	2.9469	2.8243	**1.0473**
insulator	332.1666	10.8443	124.8803	5.4195	2.5235	**0.9663**
arrester	91.0631	2.6315	22.0793	1.5167	0.5769	**0.4906**
breaker	1333.9	53.9573	895.3056	17.5001	9.1956	**3.1767**

compared with all other methods, which indicate that our method can keep more thermal information in the infrared image.

D. QUANTITATIVE COMPARISONS

The quantitative comparison result of three metrics is shown in Table 2. For the SSIM, the proposed fusion method gets the largest values in first three fusion images, which indicate that more texture and details are preserved in the fusion image. For the MI and Q_C, the values of our last three fusion images are significantly higher than other methods, especially in the last fusion image, which indicate our method not only retains more information in the original image, but also more friendly for human perception system. Furthermore, considering fault detection for power equipment in the UE-IoT, real-time is also a very important criterion. The time costs of different methods are listed in Table 3. As we can see, our method takes much less time than other algorithms, which means that our method is more suitable to meet the real-time requirements of fault detection in UE-IoT than other algorithms.

V. CONCLUSION

In this paper, we propose an infrared and visible image fusion method based on improved Wasserstein generative adversarial network for fault points detection in the UE-IoT. We adopt the trained VGG19 model to extract the features of the generated fusion image and the visible image, and minimize the style differences between these features to ensure that the fusion image can obtain enough edge details. Besides, considering the actual requirements of fault detection, we emphasize the importance of the thermal information during training process. Our method is an end to end model, which can bypass designing fusion rules manually. To verify the performance of proposed method, we compare it with the other five methods. The experimental results show that the fusion image obtained by our method can preserve more details from source images, and has better visual fidelity.

REFERENCES

[1] H. Rui, Z. Yunhao, T. Shiming, Y. Yang, and Y. Wenhai, "Fault point detection of IoT using multi-spectral image fusion based on deep learning," *J. Vis. Commun. Image Represent.*, vol. 64, Oct. 2019, Art. no. 102600.

[2] Z. Zhou, C. Zhang, C. Xu, F. Xiong, Y. Zhang, and T. Umer, "Energy-efficient industrial Internet of UAVs for power line inspection in smart grid," *IEEE Trans. Ind. Informat.*, vol. 14, no. 6, pp. 2705–2714, Jun. 2018.

[3] Z. Zhou, J. Feng, B. Gu, B. Ai, S. Mumtaz, J. Rodriguez, and M. Guizani, "When mobile crowd sensing meets UAV: Energy-efficient task assignment and route planning," *IEEE Trans. Commun.*, vol. 66, no. 11, pp. 5526–5538, Nov. 2018.

[4] Q. Lu and K. Ding, "Research on fault detection system of power equipment based on UV and infrared image," *IOP Conf. Ser., Mater. Sci. Eng.*, vol. 231, Sep. 2017, Art. no. 012062.

[5] D. Lopez-Perez and J. Antonino-Daviu, "Application of infrared thermography to failure detection in industrial induction motors: Case stories," *IEEE Trans. Ind. Appl.*, vol. 53, no. 3, pp. 1901–1908, May/Jun. 2017.

[6] Z. Shen-Pei, L. Xi, Q. Bing-Chen, and H. Hui, "Research on insulator fault diagnosis and remote monitoring system based on infrared images," *Procedia Comput. Sci.*, vol. 109, pp. 1194–1199, Jun. 2017.

[7] Q. Qin, Z.-G. Liu, and L. Jiang, "Thermal battery infrared monitoring system design based on virtual instrument technology," *Proc. SPIE*, vol. 8193, Sep. 2011, Art. no. 81934K.

[8] J. Cui, Y. Zhang, L. Niu, J. Zhan, and J. Huang, "Infrared monitoring studies of winding temperature rise in electrical equipment," *Appl. Mech. Mater.*, vol. 475, pp. 116–119, Dec. 2014.

[9] Y. Li, F. Yu, Q. Cai, K. Yuan, R. Wan, X. Li, M. Qian, P. Liu, J. Guo, J. Yu, T. Zheng, H. Yan, P. Hou, Y. Feng, S. Wang, and L. Ding, "Image fusion of fault detection in power system based on deep learning," *Cluster Comput.*, vol. 22, no. S4, pp. 9435–9443, Jul. 2019.

[10] Z. Zhou, H. Liao, B. Gu, K. M. S. Huq, S. Mumtaz, and J. Rodriguez, "Robust mobile crowd sensing: When deep learning meets edge computing," *IEEE Netw.*, vol. 32, no. 4, pp. 54–60, Jul. 2018.

[11] D. P. Bavirisetti, G. Xiao, and G. Liu, "Multi-sensor image fusion based on fourth order partial differential equations," in *Proc. 20th Int. Conf. Inf. Fusion (Fusion)*, Jul. 2017, pp. 701–709.

[12] D. M. Bulanon, T. F. Burks, and V. Alchanatis, "Image fusion of visible and thermal images for fruit detection," *Biosyst. Eng.*, vol. 103, no. 1, pp. 12–22, May 2009.

[13] G. Pajares and J. M. de la Cruz, "A wavelet-based image fusion tutorial," *Pattern Recognit.*, vol. 37, no. 9, pp. 1855–1872, Sep. 2004.

[14] Q. Zhang and X. Maldague, "An infrared-visible image fusion scheme based on NSCT and compressed sensing," *Proc. SPIE*, vol. 9474, May 2015, Art. no. 94740Y.

[15] H. Yan and Z. Li, "Infrared and visual image fusion based on multi-scale feature decomposition," *Optik*, vol. 203, Feb. 2020, Art. no. 163900.

[16] Y. Liu and Z. Wang, "Simultaneous image fusion and denoising with adaptive sparse representation," *IET Image Process.*, vol. 9, no. 5, pp. 347–357, May 2015.

[17] Y. Liu, X. Chen, R. K. Ward, and Z. J. Wang, "Image fusion with convolutional sparse representation," *IEEE Signal Process. Lett.*, vol. 23, no. 12, pp. 1882–1886, Dec. 2016.

[18] W. Ding, D. Bi, L. He, and Z. Fan, "Infrared and visible image fusion method based on sparse features," *Infr. Phys. Technol.*, vol. 92, pp. 372–380, Aug. 2018.

[19] J. Ma, Z. Zhou, B. Wang, and H. Zong, "Infrared and visible image fusion based on visual saliency map and weighted least square optimization," *Infr. Phys. Technol.*, vol. 82, pp. 8–17, May 2017.

[20] Y. Liu, X. Chen, J. Cheng, H. Peng, and Z. Wang, "Infrared and visible image fusion with convolutional neural networks," *Int. J. Wavelets, Multiresolution Inf. Process.*, vol. 16, no. 3, May 2018, Art. no. 1850018.

[21] H. Li and X.-J. Wu, "DenseFuse: A fusion approach to infrared and visible images," *IEEE Trans. Image Process.*, vol. 28, no. 5, pp. 2614–2623, May 2019.

[22] Y. Zhang, Y. Liu, P. Sun, H. Yan, X. Zhao, and L. Zhang, "IFCNN: A general image fusion framework based on convolutional neural network," *Inf. Fusion*, vol. 54, pp. 99–118, Feb. 2020.

[23] J. Ma, Y. Ma, and C. Li, "Infrared and visible image fusion methods and applications: A survey," *Inf. Fusion*, vol. 45, pp. 153–178, Jan. 2019.

[24] M. Arjovsky, S. Chintala, and L. Bottou, "Wasserstein GAN," 2017, *arXiv:1701.07875*. Accessed: Jan. 2017. [Online]. Available: https://ui.adsabs.harvard.edu/abs/2017arXiv170107875A

[25] D. Mahapatra, Z. Ge, S. Sedai, and R. Chakravorty, "Joint registration and segmentation of xray images using generative adversarial networks," in *Proc. Int. Workshop Mach. Learn. Med. Imag.*, in Lecture Notes in Computer Science, Y. Shi, H. I. Suk, and M. Liu, Eds., 2018, pp. 73–80.

[26] C. Zhang, Y. Feng, B. Qiang, and J. Shang, "Wasserstein generative recurrent adversarial networks for image generating," in *Proc. 24th Int. Conf. Pattern Recognit. (ICPR)*, Aug. 2018, pp. 242–247.

[27] J. Zhang, P. Shamsolmoali, P. Zhang, D. Feng, and J. Yang, "Multispectral image fusion using super-resolution conditional generative adversarial networks," *J. Appl. Remote Sens.*, vol. 13, no. 2, p. 1, Oct. 2018.

[28] S. Chen, D. Shi, M. Sadiq, and M. Zhu, "Image denoising via generative adversarial networks with detail loss," in *Proc. 2nd Int. Conf. Inf. Sci. Syst. (ICISS)*, 2019, pp. 261–265.

[29] Y. Yu, X.-Y. Zhang, Z. Zhao, L. Yu, and X.-J. Li, "The function and application prospect of energy storage in the ubiquitous power Internet of Things," in *Advances in Engineering Research*, D. J. Jakobczak, M. Bhardwaj, and C. S. Chin, Eds. Paris, France: Atlantis Press, 2019, pp. 72–77.

[30] D. P. F. Möller and H. Vakilzadian, "Ubiquitous networks: Power line communication and Internet of Things in smart home environments," in *Proc. IEEE Int. Conf. Electro/Inf. Technol.*, Jun. 2014, pp. 596–601.

[31] I. Goodfellow, J. Pouget-Abadie, M. Mirza, B. Xu, D. Warde-Farley, S. Ozair, A. Courville, and Y. Bengio, "Generative adversarial nets," in *Proc. NIPS*, 2014, pp. 2672–2680.

[32] M. Mirza and S. Osindero, "Conditional generative adversarial nets," 2014, *arXiv:1411.1784*. Accessed: Nov. 2014. [Online]. Available: http://arxiv.org/abs/1411.1784

[33] X. Mao, Q. Li, H. Xie, R. Y. K. Lau, Z. Wang, and S. P. Smolley, "Least squares generative adversarial networks," in *Proc. IEEE Int. Conf. Comput. Vis. (ICCV)*, Oct. 2017, pp. 2813–2821.

[34] G. Ishaan, F. Ahmed, M. Arjovsky, V. Dumoulin, and A. C. Courville, "Improved training of Wasserstein GANs," in *Proc. NIPS*, 2017, pp. 5767–5777.

[35] K. Simonyan and A. Zisserman, "Very deep convolutional networks for large-scale image recognition," 2014, *arXiv:1409.1556*. Accessed: Sep. 2014. [Online]. Available: http://arxiv.org/abs/1409.1556

[36] F. Yu and V. Koltun, "Multi-scale context aggregation by dilated convolutions," 2015, *arXiv:1511.07122*. Accessed: Nov. 2015. [Online]. Available: http://arxiv.org/abs/1511.07122

[37] S. Li, X. Kang, J. Hu, and B. Yang, "Image matting for fusion of multi-focus images in dynamic scenes," *Inf. Fusion*, vol. 14, no. 2, pp. 147–162, Apr. 2013.

[38] C. Wang and M. He, "Image style transfer with multi-target loss for IoT applications," in *Proc. 15th Int. Symp. Pervas. Syst., Algorithms Netw. (I-SPAN)*, Oct. 2018, pp. 296–299.

[39] B. K. S. Kumar, "Image fusion based on pixel significance using cross bilateral filter," *Signal, Image Video Process.*, vol. 9, no. 5, pp. 1193–1204, Jul. 2015.

[40] H. Li and X.-J. Wu, "Infrared and visible image fusion using latent low-rank representation," 2018, *arXiv:1804.08992*. Accessed: Apr. 2018. [Online]. Available: http://arxiv.org/abs/1804.08992

[41] H. Li, X.-J. Wu, and T. S. Durrani, "Infrared and visible image fusion with ResNet and zero-phase component analysis," *Infr. Phys. Technol.*, vol. 102, Nov. 2019, Art. no. 103039.

[42] D. P. Kingma and J. Ba, "Adam: A method for stochastic optimization," 2014, *arXiv:1412.6980*. Accessed: Dec. 2014. [Online]. Available: http://arxiv.org/abs/1412.6980

[43] Z. Wang and A. C. Bovik, "A universal image quality index," *IEEE Signal Process. Lett.*, vol. 9, no. 3, pp. 81–84, Mar. 2002.

[44] G. Qu, D. Zhang, and P. Yan, "Information measure for performance of image fusion," *Electron. Lett.*, vol. 38, no. 7, pp. 313–315, 2002.

[45] Y. Chen and R. S. Blum, "A new automated quality assessment algorithm for image fusion," *Image Vis. Comput.*, vol. 27, no. 10, pp. 1421–1432, Sep. 2009.

YOU DU received the B.S. degree in intelligent science and technology from the School of Automation Engineering, Northeast Electric Power University, in 2018. He is currently pursuing the M.S. degree with North China Electric Power University. His research interests include image fusion and deep learning.

BIN LIAO received the Ph.D. degree in computer applied technology from the Graduate University of Chinese Academy of Sciences, in 2003. He is currently an Associate Professor with the School of Electrical and Electronic Engineering, North China Electric Power University. His current research interests include information fusion and deep learning.

XIANGYUN YIN received the B.S. degree in electronic information engineering from the College of Electronic Information Engineering, North China University of Technology, in 2018. He is currently pursuing the M.S. degree with North China Electric Power University. His current research interests include image fusion and super resolution image reconstruction.

● ● ●