Super-Resolution of Thermal Images Using GAN Network

Shashikant Deepak
Dept. Of Electrical Engineering
NIT Rourkela, INDIA
shashikantdeepak@gmail.com

Sanuj Sahoo
Dept. of Electrical Engineering
NIT Rourkela, INDIA
sanutlc1996@gmail.com

Dipti Patra
Dept. Of Electrical Engineering
NIT Rourkela, INDIA
dpatra@nitrkl.ac.in

***Abstract*—Super-resolution (SR) reconstruction of thermal images has been one of the most active research areas specifically for industrial applications. However, most of the conventional RGB SR models available in the literature are not necessarily applicable to thermal images due to their difference in characteristics when compared to normal camera images. The recent advancement in the field of deep learning-based SR has helped achieve unbelievable results. Despite the advancement in models like deep convolution neural networks (CNN) and Generative adversarial networks, there remain multiple problems unsolved that will help improve the spatial resolution of thermal images. Not only the developed model should be computationally efficient but also easily implementable in industrial applications. Motivated to overcome the said limitations, in this work a generative adversarial network (GAN) based single images super-resolution architecture is proposed for thermal camera images. The developed model not only generates at par results with the other model but also is easy to implement and computationally efficient. The modified architecture has an identical layout inspired by SRGAN. In order to make the model faster to train while having less training parameters, the number of residual blocks was reduced to 5. The batch normalization layers were excluded from the residual blocks of both the Generator and Discriminator networks to remove the redundancy. Before each convolution layer, reflective padding is utilized at the edges to preserve the size of the feature maps. The comparative results revealed that the proposed network trained on thermal images produced high-quality images with enhanced details, while still maintaining image features and perspective throughout. The experimental results show that the proposed model has achieved a reduction in computation time compared to the State-of-the-Art method. The suggested strategy has outperformed the SOTA methods with the improvement of approximately 2dB in PSNR along with 0.9825 of SSIM.**

***Index Terms*—Super Resolution; Thermal Images; Generative Adversarial Networks.**

I. Introduction

Thermal imaging is a technique for capturing images in far-infrared bands by using special-purpose thermal cameras. Such images are highly desirable in many industrial application where the objects to be analyzed is invisible to naked eyes and also to normal cameras [1]. Long-Wave IR has a wavelength range of 8 to 15 micro-meters and provides quantifiable thermal data or heat maps of the recorded image representing temperature data. The evolution of thermographic imaging technology allows for the vision of objects beyond the visible range, allowing it to be used in various applications like military, medical, agricultural, and various industries, etc. This system measures temperature using the heat of objects and makes a thermogram, similar to as shown in Fig. I.

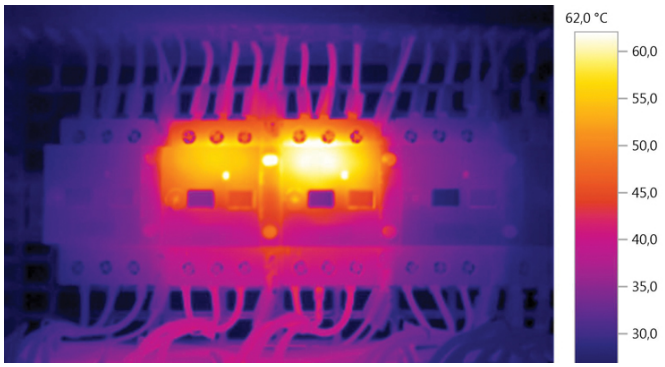

Fig. 1. Thermogram

However, compared to visible range RGB cameras, thermal cameras are constrained by technological limitations and have a poor spatial resolution. Also, thermal cameras are more expensive than visible cameras with similar resolution. The cost of thermal sensors rises dramatically as the resolution of the sensor rises. As a result, thermal cameras are utilized in low-resolution (LR) and low-contrast environments. Such LR images limit the image analysis process and prove disadvantageous for many applications where only high-resolution (HR) images can only serve the purpose. Therefore, there is an ultimate need of a low-cost solution to generate HR images using post-processing techniques like super-resolution (SR) of thermal images. SR is a method that improves spatial resolution by integrating complementary information from one or many LR images to produce HR images. As the SR reconstructed image has the excellent perceptual quality and contains high details and information, it can be employed in a variety of applications such as surveillance, medical diagnosis, industrial maintenance, etc.

Many SR methods are proposed in literature based on different strategies. As mentioned in paper [2], a comprehensive overview of various single image SR, which tries to recover a HR image from a LR image has been presented. But most

978-1-6654-2337-3/21/$31.00 © 2021 IEEE

of them are not necessarily applicable on thermal images. For thermal image datasets, the approach must be more efficient than for visual image datasets due to the difference in their characteristics. Traditional approaches function properly and have demonstrated some patches in natural pictures, which tend to redundantly reoccur multiple times inside the image [3].

In order to learn multi-scale information and create a deep network model for single image SR, the authors in [4] developed a convolution-based inception model. Xudong Zhang [5] proposed a novel SR method in which Compressive sensing theory and deep learning are combined. The approach is divided into two parts: the first is used to create an HR image that contains high-frequency information, and the second part is used to remove the fixed pattern of noise. In 2020, Vishal Chudsama et al. proposed a computationally efficient SR reconstruction method called TherlSuRNet [6], in which progressive upscaling strategy with asymmetrical residual learning in the network was used. This method consists of low-high frequency feature extraction along with upscaling blocks. Here, the author uses asymmetrical strategy and residual learning methods for a different upscaling factor like ×2, ×4, ×8 are used. Sometimes only a thermal image or a RGB image is not much informative. Taking this into consideration, Feras Almasri et al. [7] proposed a multi-modal sensor fusion-based SR approach. Although the generated images are better but some artifacts were noticed in the generated images. Rivadeneira et al. [8] proposed a SISR algorithm to construct an HR thermal infrared image from subpixel shifted aliased LR frames, wherein a stochastic regularized SR approach is applied for x4 scaling factor. These methodologies have yielded better performance, but they lack in exploring the correlation between spatial-spectral features and non-linear mapping. Moreover, the above-mentioned learning-based restoration algorithms ignore multi-scale information. Motivated to overcome the said drawbacks, this paper proposes a simple and computationally efficient model using the generative adversarial network (GAN) for single image SR of thermal images. The proposed model architecture is inspired on the research of [9], [10]. The proposed method shows improvement over the state-of-the-art SR models in both qualitative and quantitative assessments. The rest of the manuscripts is described as follows: A detailed discussion on the proposed architecture and the designed blocks in the model is given in Section II. Section III gives a brief overview of the modalities of the architecture and the specification of the training and testing experiments followed by the result subsection. Final remarks are given in Section VI.

II. Proposed model

This work aims to convert an LR thermal image into an HR thermal image while distinguishing between the HR and SR images. The key idea is to establish a model that can achieve the best perceptual quality of the thermal image and be easy to implement and computationally efficient. The industrial application requires the model to be fast and produce at par results with the most complicated models. Therefore, we design a simple yet powerful model with 5 residual blocks in the generator model in the proposed architecture. To begin, LR images thermal images are created by downscaling HR images by a factor of four using Gaussian pyramids. The image has been downscaled from its original size of (320×240) to (80×60).

The baseline model comprises of Generative adversarial network (GAN), which provides a framework consisting of image generator-discriminator sub-networks. The generator reproduces the image with high perceptual quality images. The primary purpose of this research is to extract features and map them on the original image size. The suggested model has five residual blocks and has a similar layout to the SRGAN model. The proposed model includes a generator and discriminator network. Fig. 2 pictorially gives details of the proposed architecture. The following sections provide a detailed overview of each network module.

A. Generator network

The generator network as depicted in Fig. 2(a) is inspired by the work of [11] and modeled using five residual blocks. Two convolution layers make up each residual block with a kernel size of 33 and 64 channels, followed by an *ElU* activation function. The breakup of the residual blocks used in the architecture is shown in Fig. 3 below.

The batch normalization layers were eliminated from the leftover blocks, as illustrated in Fig.2, as these layers are redundant as indicated in [11]. As presented in the Fig. 2, reflective padding is employed to retain and to supplement the feature maps. Then two subpixel convolution layers are used for upsampling the resolution of output images. Each convolution layer in the generator network used a stride of 2, eventually reducing the size of feature maps. Finally, the pixelshuffle layer upscales the final output of the generator.

B. Discriminator network

To distinguish the generated image from that of the ground truth image, the discriminator sub-network is used. Fig. 2(b) illustrates the network design of the discriminator. Batch normalization layers were also removed in the discriminator too, similar to the generator. Eight convolution layers precede the ELU activation function in the discriminator network. The kernel size of these convolution layers is 33, and the channel count rises by a factor of two in each level, from 64 to 512. The channel count is doubled because we're using convolution of sride 2 between each layer. After the 512 feature maps, average pooling layers are employed in this network followed by two dense layers. Finally, a probability output is drawn out using the sigmoid activation function to distinguish between the ground truth and reconstructed images.

III. Experiment and analysis

A. Collection of data

The dataset used to train the model comprises of 255 images of Transformers with various short circuit rounds selected

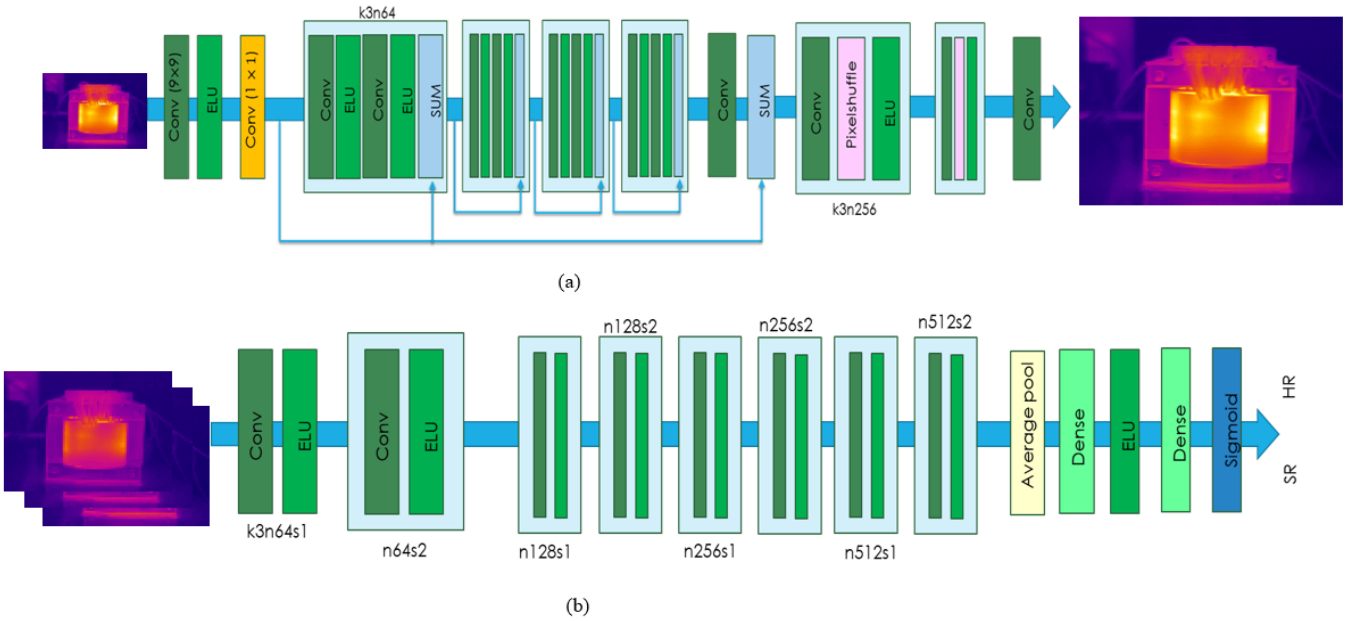

Fig. 2. Architecture of Proposed Model, (a) Generator network, (b) Discriminator network

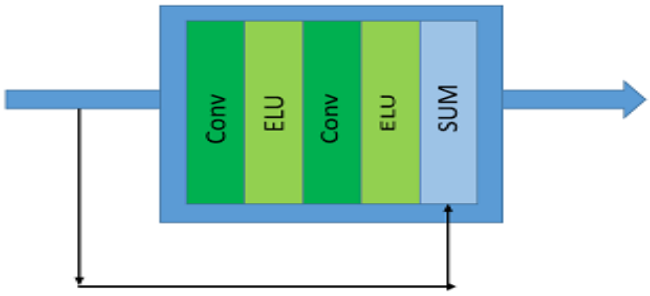

Fig. 3. Residual Block

from 600 to 80 short circuit rounds. This dataset contains the thermal images of electrical machines i.e. transformers. All the faults that occurred in the transformer are the internal faults [12]. The thermal images were acquired under various short-circuiting conditions, i.e., 8 different cases of common core winding short circuit failures were considered. The dataset used to train the model comprises of 255 images of Transformers with various short circuit rounds selected from 600 to 80 short circuit rounds. A Dali-tech T4/T8 infrared thermal image camera with a detector resolution of 384*288 and a measuring range of $-20°$ to + 650°c is used to acquire thermal images at the workbench in an Electrical Machines Laboratory at a temperature of 23°c.

This is a dataset of thermal images (IRT) used in the condition monitoring of electrical equipment, specifically transformers. All artificially generated flaws are internal flaws that are not caused by external components or initial setup component failure. Thermal images of an induction motor are also included in the dataset. All artifact-generated flaws in this dataset are created by internal issues rather than external pieces or failures in the initial setup of electrical components. There are 369 images in all. The induction motor is considered in eight different instances of short circuit failures in the stator windings, stuck rotor problem, and cooling fan failure.

B. Implementation

A normal distribution initialization with zero mean and a standard deviation of 0.02 is used to initialize the generator and discriminator networks. We use the Adam optimizer for both the network with a learning rate of 0.01 and a batch size of 12. The model was trained for 2000 epochs, with the learning rate halved for every 50 epochs. The models are all trained in a Pytorch environment and run on a RTX 2080TI.

IV. Evaluation Metrices

The peak signal to noise ratio (PSNR) and structural similarity (SSIM) metrics given by Eqn. 1 and Eqn. 2 are used to verify efficacy of the proposed method.

$$PSNR = 10\log_{10}\left(\frac{MAX_f^2}{MSE}\right) \tag{1}$$

$$SSIM = \frac{(2\mu_x\mu_y + c_1)(2\sigma_{xy} + c_2)}{(\mu_x^2 + \mu_y^2 + c_1)(\sigma_x^2 + \sigma_y^2 + c_2)} \tag{2}$$

MAX is the maximum pixel value in f^{th} band while f refers to the reconstructed image thermal image.

V. Experiment Result

In order to test the performance of the trained model, the network is tested on the test dataset split from the thermal image dataset and the results for some of the images in Fig. 4 and 5. The pictures were taken under various short

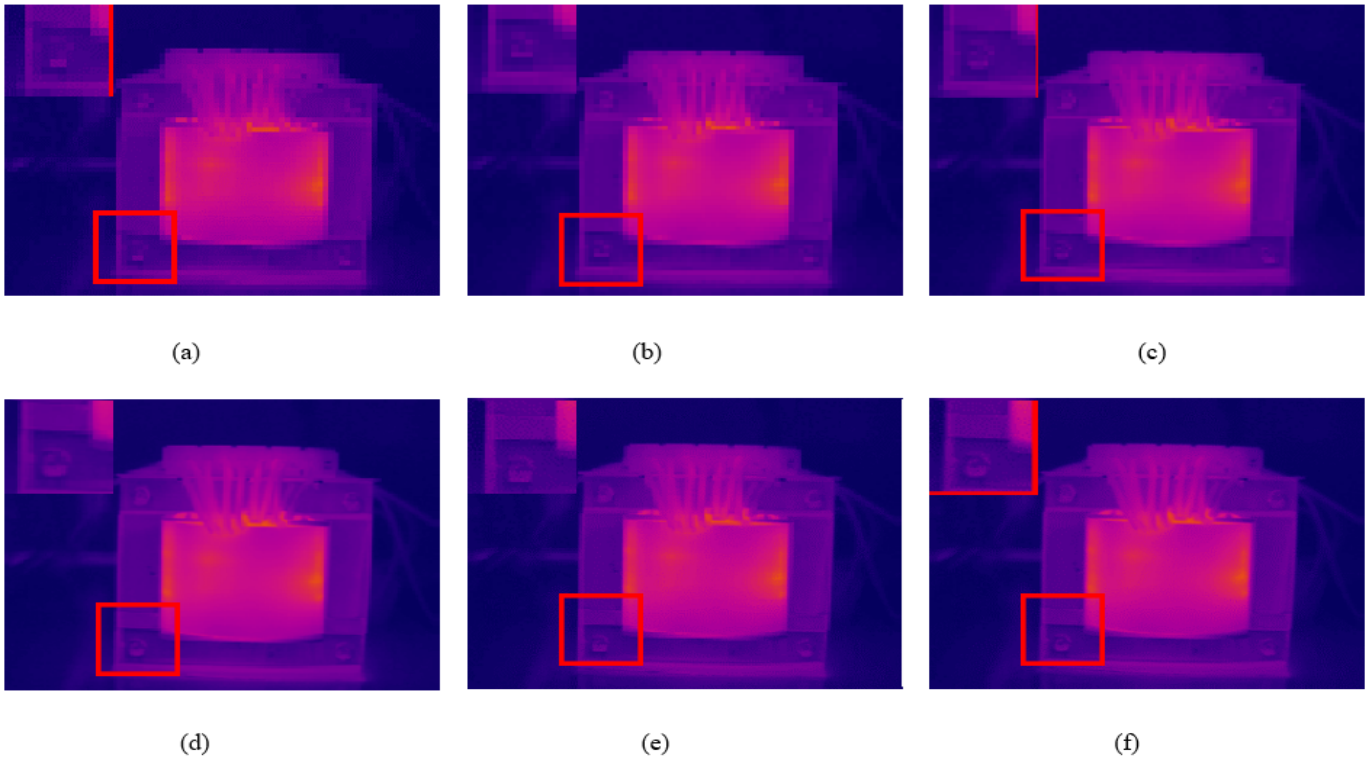

Fig. 4. SR reconstruction result from left to right a 400-round short-circuit situation image: (a) Low resolution image, (b) Bicubic interpolation, (c) SRResNet, (d) SRGAN, (e) Proposed, (f) Ground truth image

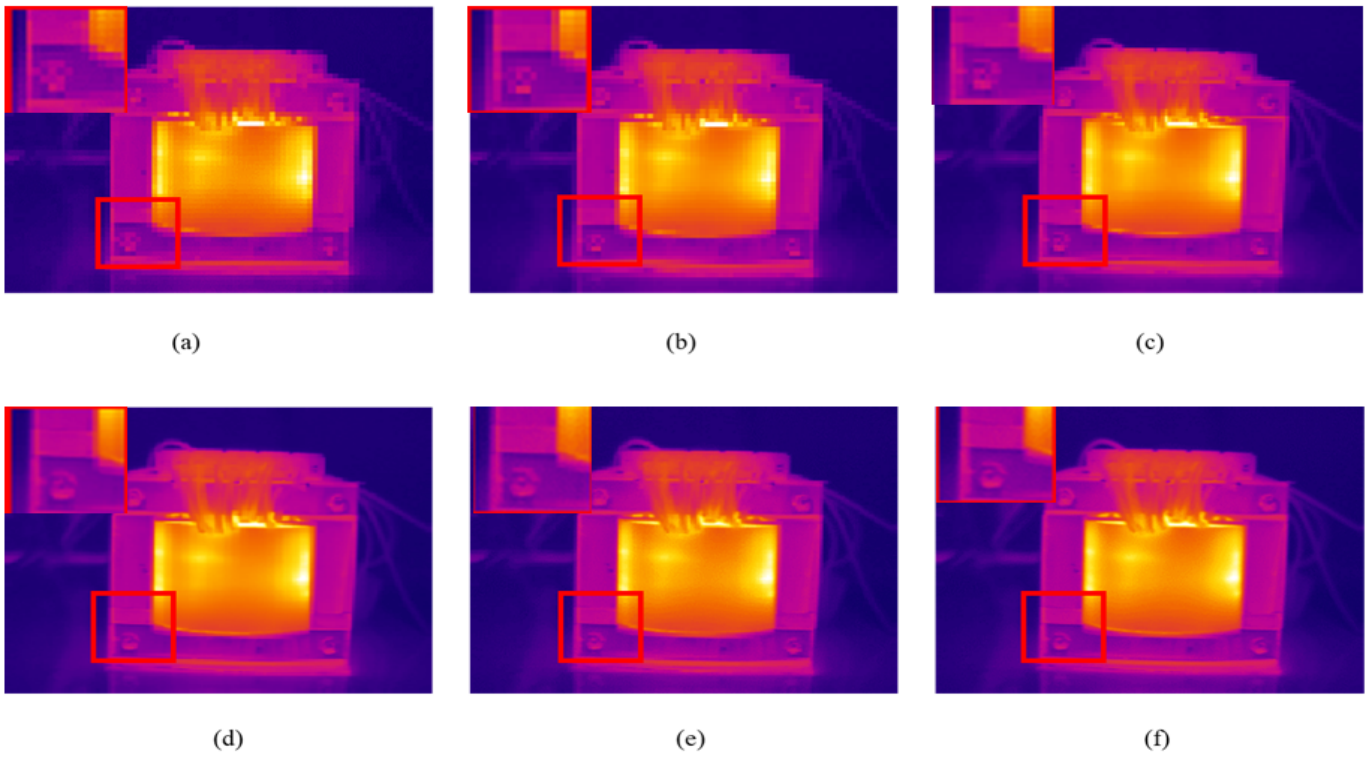

Fig. 5. SR reconstruction result from left to right for a 600-round short-circuit situation image : (a) Low resolution image, (b) Bicubic interpolation, (c) SRResNet, (d) SRGAN, (e) Proposed, (f) Ground truth image

circuit conditions, as shown in Fig. 4 and 5. The results of various SOTA methods are compared with the findings of the proposed model. The illustration depicts the transformer's LR Fig. 4(a) and 5(f) and ground truth images. While Fig. 4 and 5 (b) represents the bicubic interpolation result, (c) represents the SRResNet result, (d) represents the SRGAN model result, and (e) represents the results of the proposed model architecture. We can see that the study activity gives outcomes that are the most accurate representations of the real-world images. The winding in the transformer core is visible and matches the ground truth images. Fig. 4 shows the image of a transformer in a 400-round short-circuit situation, whereas Fig. 5 shows images in a 600-round short-circuit condition. In the ground truth images, the lower half of the transformer is visible; however, bicubic interpolation and SRResNet lose that information. SRGAN generates images that are comparable to ground truth images, but there are some artifacts in the model output. The proposed method reconstructs images which is highly similar to the HR images. Table I shows the results of a fair comparison using the settings

TABLE I
STATISTICAL PARAMETER COMPARISON OF PROPOSED MODEL WITH OTHER SOTA MODELS (400 ROUND SC)
NOTE:THE VALUES REPORTED ARE THE AVERAGE VALUES OVER THE WHOLE TEST DATASET

Methods	PSNR(dB)	SSIM
Bicubic interpolation	35.2932	0.9433
SRResNet	36.8239	0.9581
SRGAN	38.3261	0.9693
Proposed	40.5427	0.9763

specified by the creators of SOTA methods such as Bicubic interpolation, SRResNet, and SRGAN. In terms of PSNR and SSIM, this comparison is conducted for the 400 round of short circuit transformer images. When compared to other statistical measures in table I, the proposed model outperforms others in terms of SSIM and PSNR. Additionally, Table II shows the fair comparison for 600 round of short circuiting transformer images. The visual quality comparison as shown

TABLE II
STATISTICAL PARAMETER COMPARISON OF PROPOSED MODEL WITH OTHER SOTA MODELS (600 ROUND SC)
NOTE:THE VALUES REPORTED ARE THE AVERAGE VALUES OVER THE WHOLE TEST DATASET

Methods	PSNR(dB)	SSIM
Bicubic interpolation	39.6521	0.9573
SRResNet	40.5724	0.9677
SRGAN	41.2967	0.9791
Proposed	43.9239	0.9825

in Fig. 4 and 5 shows that the SR image from the trained generator of the proposed model has high perceptual quality and is significantly similar to the corresponding ground truth image. This conclusion is supported through the tremendous out performance in statistical metrics such as PSNR and SSIM. Therefore, the suggested model outperformed earlier state-of-the-art methods by improving statistical and visual parameters significantly.

VI. CONCLUSION

In this research work a learning-based GAN model is proposed for super resolution of thermography images. The proposed model comprises of a generator and a discriminator networks. In the proposed model architecture, the core of the proposed model performs feature extraction and mapping on the input images. The key motive is to use a computationally efficient, fast and easy to implement model for thermal image super resolution. Moreover, in comparison to other SOTA approaches, the suggested method has shown considerable improvement in statistical indicators such as PSNR and SSIM metrics. Furthermore, when compared visually, the proposed methodology has demonstrated improved perceptual quality.

REFERENCES

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

[2] R. E. Rivadeneira, P. L. Suárez, A. D. Sappa, and B. X. Vintimilla, "Thermal image superresolution through deep convolutional neural network," in *International Conference on Image Analysis and Recognition*. Springer, 2019, pp. 417–426.

[3] X. Chen, G. Zhai, J. Wang, C. Hu, and Y. Chen, "Color guided thermal image super resolution," in *2016 Visual Communications and Image Processing (VCIP)*. IEEE, 2016, pp. 1–4.

[4] Y. Choi, N. Kim, S. Hwang, and I. S. Kweon, "Thermal image enhancement using convolutional neural network," in *RSJ International Conference on Intelligent Robots and Systems (IROS)*. IEEE, 2016, pp. 223–230.

[5] X. Zhang, C. Li, Q. Meng, S. Liu, Y. Zhang, and J. Wang, "Infrared image super resolution by combining compressive sensing and deep learning," *Sensors*, vol. 18, no. 8, p. 2587, 2018.

[6] V. Chudasama, H. Patel, K. Prajapati, K. P. Upla, R. Ramachandra, K. Raja, and C. Busch, "Therisurnet-a computationally efficient thermal image super-resolution network," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops*, 2020, pp. 86–87.

[7] F. Almasri and O. Debeir, "Multimodal sensor fusion in single thermal image super-resolution," in *Asian Conference on Computer Vision*. Springer, 2018, pp. 418–433.

[8] R. E. Rivadeneira, A. D. Sappa, and B. X. Vintimilla, "Thermal image super-resolution: A novel architecture and dataset." in *VISIGRAPP (4: VISAPP)*, 2020, pp. 111–119.

[9] C. Ledig, L. Theis, F. Huszár, J. Caballero, A. Cunningham, A. Acosta, A. Aitken, A. Tejani, J. Totz, Z. Wang *et al.*, "Photo-realistic single image super-resolution using a generative adversarial network," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2017, pp. 4681–4690.

[10] C. Dong, C. C. Loy, K. He, and X. Tang, "Image super-resolution using deep convolutional networks," *IEEE transactions on pattern analysis and machine intelligence*, vol. 38, no. 2, pp. 295–307, 2015.

[11] B. Lim, S. Son, H. Kim, S. Nah, and K. Mu Lee, "Enhanced deep residual networks for single image super-resolution," in *Proceedings of the IEEE conference on computer vision and pattern recognition workshops*, 2017, pp. 136–144.

[12] M. Najafi, S. Baleghi, Yasserand Mehdi, and Miriman, "Thermal images dataset transformer phase dry type," 2020.