GAN-based Vision Transformer for High-Quality Thermal Image Enhancement

Mohamed Amine Marnissi
Université de Sfax, Ecole Nationale d'Ingénieurs de Sfax,
Sfax 3038, Tunisie;
mohamed.amine.marnissi@gmail.com

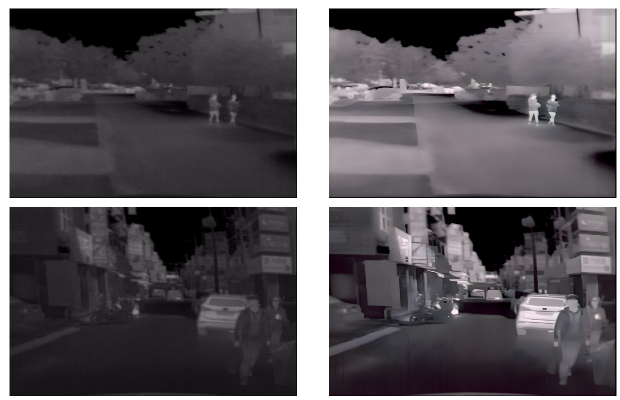
Abir Fathallah
Samovar, CNRS, Télécom SudParis, Institut Polytechnique de Paris,
9 rue Charles Fourier, 91011 Evry Cedex, France
abir.fathallah1803@gmail.com

Abstract

Generative Adversarial Networks (GANs) have shown an outstanding ability to generate high-quality images with visual realism and similarity to real images. This paper presents a new architecture for thermal image enhancement. Precisely, the strengths of architecture-based vision transformers and generative adversarial networks are exploited. The thermal loss feature introduced in our approach is specifically used to produce high-quality images. Thermal image enhancement also relies on fine-tuning based on visible images, resulting in an overall improvement in image quality. A visual quality metric was used to evaluate the performance of the proposed architecture. Significant improvements were found over the original thermal images and other enhancement methods established on a subset of the KAIST dataset. The performance of the proposed enhancement architecture is also verified on the detection results by obtaining better performance with a considerable margin regarding different versions of the YOLO detector.

1. Introduction

Thermal images are often used in a variety of applications, including building inspection [22], medical imaging [3], and military surveillance [28]. However, these images may suffer from a lower signal-to-noise ratio, making denoising more challenging. Additionally, thermal images may have limited dynamic range, which can make it difficult to perform contrast enhancement without losing important information. Therefore, there is a need for developing new methods that specifically address the unique characteristics of thermal images. In order to deal with these limitations, many methods designed to improve the quality of visible images are applied, such as image denoising, contrast enhancement, and super-resolution. Nevertheless, these methods are not always suitable for thermal images owing to their differentiation from visible images, since they are captured using infrared radiation.



Original image TE-VGAN

Figure 1. Thermal images enhancement TE-VGAN.

Recently, the investigation of the enhancement process has been intensified, owing to the development of new deep learning architectures, including Generative Adversarial Networks (GANs) [10, 18, 30] and Vision Transformers (ViT) [20, 31], a type of neural network designed to process images by attending to different regions of the input image. Although the advancement in the enhancement image process, only a few research are addressed improving the quality of thermal image [1, 14, 19, 30]. For example, authors in [19] introduced a GAN-based approach for enhancing the contrast and eliminating the noise with a post-process step for edge enhancement.

The combination of GANs and ViT models has received significant attention in recent research for its potential to improve the quality of image generation with rich spatial information [9, 11, 34]. Typically, the fusion of these models involves using the generative network to produce images and then applying the self-attention mechanism of the ViT model to refine the images by capturing and leveraging the spatial relationships between different image features. In the same context, we propose a hybrid approach that combines GANs and ViT models to introduce a novel model for improving the quality of thermal images. Our proposed model distinguishes itself from previous techniques by leveraging GANs, with a CNN as the generator component, and ViT models as both global and local discriminators with the aim of improving the performance of object detection algorithms by improving object visibility, reducing false positives, and highlighting important features. To the best of our knowledge, this practice has not been explored in the context of combining GANs and ViT techniques. The overall architecture is referred to as the Thermal Enhancement Vision Generative Adversarial Network (TE-VGAN). Fig. 8 demonstrates the effectiveness of our TE-VGAN on two samples from the KAIST dataset.

2. Related works for thermal image enhancement

As already discussed in section 1, thermal images are characterized by low-contrast, low-resolution, and blurred details, which can limit their usefulness in many video analytic applications including object detection, which is our primary goal in this study. To handle thermal image issues, many traditional methods are used for visible imaging in order to enhance thermal image quality. Among these methods, Histogram Equalization (HE) can be employed. For instance, in [25], a multi-objective HE model was suggested to enhance the contrast while preserving the brightness of thermal images. Furthermore, Contrast Limited Adaptive Histogram Equalization (CLAHE) based on local contrast modification was defined in [21]. The technique in [17] belongs as well to traditional methods for thermal image enhancement. It was first based on the computation of a 2D histogram, which incorporated both global and local gray-level distributions of the original image. Then, by applying histogram specification globally and locally, the results were accordingly enhanced.

2.1. CNN-based methods

Convolutional neural networks (CNNs) have recently shown outstanding performance in a lot of computer vision applications such as image classification and object detection and recognition. Some recently published methods for image enhancement have employed CNN architectures to improve the visual quality of thermal or visible images. SRCNN [4] is one of the first attempts to handle this problem. Its basic idea consists of learning a mapping relationship between low-resolution and high-resolution visible images. The SRCNN problem is that the training step is time-consuming. To optimize it, this architecture was extended in [5] to the Fast SRCNN. The latter was composed of three main parts: patch extraction/ representation, non-linear mapping, and reconstruction, which are three standard convolutional layers with a ReLU function, except the last layer because it is for reconstructing the resulting image. VDSR [12] is also one of the most known deep learning methods for enhancement, which aims at augmenting the spatial resolution of visible images. Compared to the SRCNN, the VDSR is deeper since it is composed of 20 layers inspired by the VGG architecture [26], where each layer consists of 64 filters of size 3×3. The idea behind this is to predict the residual image with such a network and to concatenate it with a low-resolution image in order to obtain the desired output image since residual learning converges much faster than others.

While most of the existing methods for image enhancement in the visible domain focus on increasing the spatial resolution of the original image, only a few studies for thermal image enhancement have been proposed to cover other aspects apart from the resolution problem, such as low contrast and blurred edges. The thermal enhancement network [2] was one of the first CNN-based methods for thermal image enhancement, where a relatively shallow CNN was designed to learn an end-to-end mapping from the original image to the target high-resolution image. The authors in [36], an EFTS method was proposed. It consisted of a model based on residual dense blocks, which could perform SR for thermal images while enhancing the visual information of edges. EFTS would generate the SR thermal image by enforcing the CNN network to bear down on edge visual information.

2.2. GAN-based Methods

GANs are deep learning architectures initially introduced by Goodfellow *et al.* in [7]. GANs are generally composed of two sub-networks: generative sub-network G and discriminative sub-network D. These architectures have shown excellent performance in image generation and restoration. They have also been employed in few studies for the enhancement of visible and thermal images. For instance, in [13], the SR GAN (SRGAN) included a deep Residual Network (ResNet) with skip-connection and defined a perceptual loss. The latter consisted of adversarial and content loss functions using high-level feature maps of the VGG network instead of Mean Squerd Error (MSE) function combined with the discriminative network to distinguish between high-resolution and low-resolution images. Another related work that made use of GANs for SR thermal images was presented, in [8]. It was about a deep learning framework based precisely on Deep Convolutional GANs (DCGANs) for thermal face images. This architecture was less deep than SRGAN since it utilized only two networks instead of three in [13], which made it suitable for smaller datasets. The way how residual blocks were organized proved to be efficient in preserving the image edges. To validate the DCGAN architecture, multiple tests were conducted on thermal face datasets.

According to [27], SR methods in the visible spectrum also included the work published in [32], where a rank-content loss in a GAN was used to improve the visual effects in the SR image. To eliminate the effect of artifacts in [35], an image quality assessment metric was employed in a loss function that would improve the stability of image SR. To further reduce the complexity of GAN models for image SR in [6], a Fourier space supervision loss was utilized to recover lost high-frequency information. By doing so, it was shown that the predicted image quality was ameliorated and that the training efficiency was accelerated. [23] was another related work that attempted to enhance the stability and robustness of the training of the SR model using residual blocks and a self-attention layer. Differently, in [24] a measurement loss function was considered in a GAN network in order to extract accurate features by obtaining more detailed information. Compared to the two previous studies that employed GAN architectures for SR in both visible and thermal domains, the authors in [15] focused on the task of enhancing the contrast in thermal images using conditional GANs.

In this paper, we intend to make use of GAN architecture for thermal image enhancement as our first contribution, with significant improvements and modifications such as ViT architecture. Precisely, we put forward a more complete architecture that simultaneously deals with different limitations of thermal images, including low contrast, noise, and blurred edges. Moreover, unlike previous work in the thermal domain, where only grayscale-converted images are used for training, we use in our proposed architecture impaired images only from the thermal domain to train the network. Through extensive experiments and tests, we demonstrate that our approach outperforms existing image enhancement algorithms in terms of contrast and detail enhancement, which proves the effectiveness of our suggested architecture.

3. Thermal Enhancement Approach

3.1. Enhancing Image Contrast with Vision Transformer Discriminator

The input image generator is treated by an attention map for highlighting regions of interest in a thermal image by assigning different weights to each pixel. Then, TE-VGAN architecture employs the U-Net architecture as a generator with the aim of leveraging multi-scale and texture information and extracting multi-level features. The U-Net generator is followed by global and local discriminators. The global discriminator is designed to determine how a real generated image is. It takes the entire image and its ground truth as inputs and produces a probability score indicating whether the input image is real or fake. The local discriminator, on the other hand, assesses the realism of randomly selected patches. Precisely, it takes random patches from both generated and ground truth images and provides a probability score indicating whether each pair of patches is real or fake. The two discriminator architectures rely on ViT, which is particularly useful for processing large input data by leveraging self-attention mechanisms to identify important features and relationships in the input data as shown in Fig.2. The proposed discriminators analyze the generated data according to the details, textures, overall structure and data content. The global discriminator D_{gl} is optimized following Eq.(1).

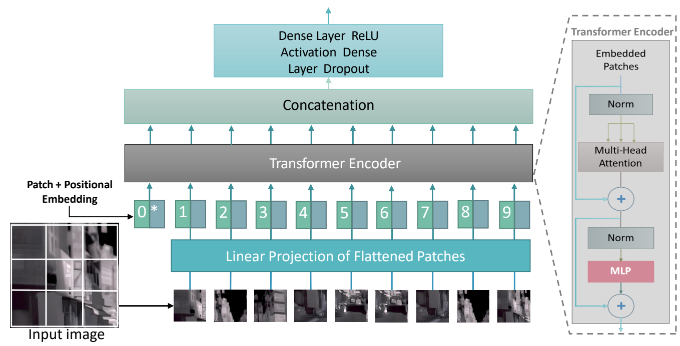


Figure 2. Discriminator-based ViT architecture.

$$\mathcal{L}_{D_{gl}} = \mathbb{E}_{r \sim \mathbb{P}_{real}}[Vit(r, f) - 1)^2] + \mathbb{E}_{f \sim \mathbb{P}_{fake}}[Vit(r, f)^2] \quad (1)$$

where r, represents the real samples, f represents the fake samples and $Vit(r, f)$ designs the output of their discriminator. $\mathbb{P}_{real}$ and $\mathbb{P}_{fake}$ are the distribution of the real and generated data. To optimize the local discriminator D_{lc}, $\mathcal{L}_{D_{lc}}$ is defined by Eq.(2).

$$\mathcal{L}_{D_{lc}} = \mathbb{E}_{x_r \sim \mathbb{Q}_{real}}[Vit(x_r) - 1)^2] + \mathbb{E}_{x_f \sim \mathbb{Q}_{fake}}[Vit(x_f))^2] \quad (2)$$

where $Vit(x_r)$ and $Vit(x_f)$ are the output of the discriminator for the inputs of real and fake patches, respectively, $\mathbb{Q}_{real}$ and $\mathbb{Q}_{fake}$ are the distribution of the real and generated patches.

Considering discriminators include ViTs, the U-Net generator is more efficient at producing high-quality, realistic images closely matching the ground truth. The quality of the generated samples can be significantly improved by optimizing the performance of the generator through adversarial loss function in order to reduce overfitting and simplify

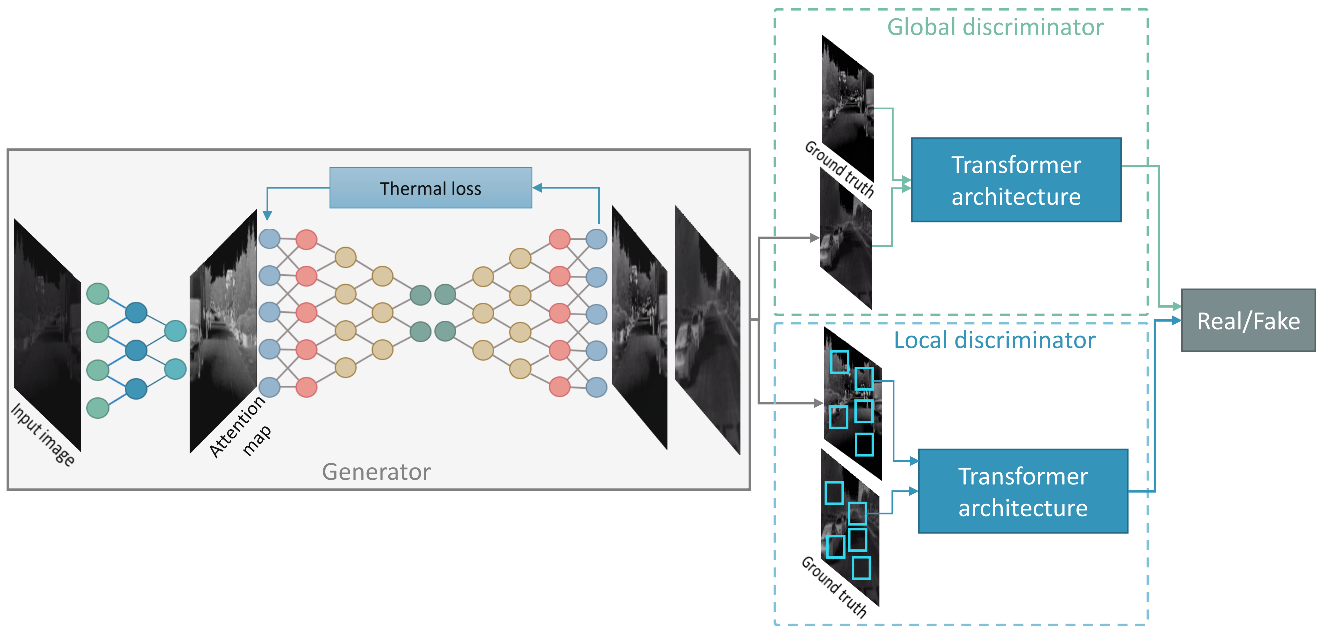


Figure 3. The proposed TE-VGAN architecture composed of generator and two ViT discriminators.

the training process. In particular, the adversarial loss function consists of generator loss and losses of two ViT discriminators. The adversarial loss function $\mathcal{L}_{adv}$ is depicted by Eq.(3).

$$\mathcal{L}_{adv} = \mathcal{L}_G + \mathcal{L}_{Dgl} + \mathcal{L}_{Dlc} \tag{3}$$

where $\mathcal{L}_G$ represents the generator loss.

$$\mathcal{L}_G = \mathbb{E}_{r \sim \mathbb{P}_{fake}}[Vit(f) - 1)^2] \tag{4}$$

3.2. Thermal Loss Optimization

To further enhance the visual quality of the thermal images, we proposed a novel thermal loss function $\mathcal{L}_{thr}$ to optimize our TE-VGAN model. The suggested loss function is designed to capture the intrinsic thermal properties of the images and promote thermal image-specific features during the training process.

$\mathcal{L}_{thr}$ quantifies the dissimilarities between the reconstructed and ground truth thermal images in terms of high-level features to yield visually realistic images, which cannot be discerned by pixel-level measures. To improve the stability of the training process, an instance normalization layer was integrated after each feature map extracted from a pre-trained VGG-16 model, allowing for the preservation of the feature content of the image and enabling the network to dynamically adapt its performance. Moreover, by calculating the L1 error between the input image and the output image, $\mathcal{L}_{thr}$ involves preserving the crucial features of the input image while maintaining the quality of the generated image. This ensures that the generator keeps the same structure and content as the input image. Our thermal loss $\mathcal{L}_{thr}$, defined Eq.(5).

$$\mathcal{L}_{thr} = \parallel r - f \parallel_1 + \frac{1}{W_{i,j}H_{i,j}} \sum_{y=1}^{H_{i,j}} (\phi_{i,j}(r) - \phi_{i,j}(f))^2 \tag{5}$$

where $W_{i,j}$ and $H_{i,j}$ denote the dimensions of the extracted feature maps $\phi_{i,j}$. The indices i and j correspond to the i^{th} max pooling and the j^{th} convolutional layer following the i^{th} max pooling layer, respectively.

Therefore, the learning objective loss used to train TE-VGAN is formulated as follows:

$$\mathcal{L} = \mathcal{L}_{adv} + \lambda_{thr}\mathcal{L}_{thr} \tag{6}$$

where λ_{thr} refers to the weight controlling the share of $\mathcal{L}_{thr}$ in the total objective function $\mathcal{L}$.

3.3. Thermal Image Restoration

Since improving the contrast makes the noise progressively more noticeable, which can negatively impact the quality of thermal images, we alleviate this effect by employing the Real-ESRGAN approach [30]. Real-ESRGAN is a state-of-the-art blind super-resolution method able to restore high-resolution images from low-resolution inputs without prior knowledge of the degradation process. It is based on ESRGAN and leverages synthetic data to simulate complex real-world degradations for model training.

In our present work, the model has been refined using a mixture of thermal and visible images, including images with different types of degradation. The goal of the Real-ESRGAN model involves training the model with visible images and transferring the learned features by fine-tuning the model on thermal images in order to better perceive the challenges of thermal images in terms of lack of texture and style as well as generate higher quality thermal images with finer details. In particular, the Real-ESRGAN generator uses pixel-unshuffle, which reduces the spatial size and increases the channel size of the inputs before feeding them into the main architecture to improve computational efficiency. Furthermore, the U-Net discriminator with skip connections and spectral normalization is used to stabilize the training dynamics and improve the discriminator's ability to handle complex training outputs. The resulting method strikes a good balance between local detail enhancement and noise removal, making it promising for improving thermal image quality.

4. Experimental Results

4.1. Implementation details

The proposed TE-VGAN architecture was trained on low and high-contrast images. We select an unpaired image subset (not the same image) from the KAIST dataset. An example of a training set is provided in Fig.4, composed of low and high-contrast thermal images, with a batch size of 4 images. Practically, we compute the contrast of images as the standard deviation of intensity values. Then, we split the images into low and high-contrast subsets according to an empirically chosen threshold. The Adam optimizer was used with a learning rate of 0.0001 for the first 10 epochs, then gradually decreasing over the next 10 epochs. The model was trained on an NVIDIA Titan X GPU with 12GB of RAM. λ_{thr} defined is set to 0.6.

4.2. Experimental Results

4.2.1 Visual Results of the TE-VGAN Architecture

To assess the effectiveness of our proposed TE-VGAN architecture, we employ the PSNR and SSIM metrics to measure the similarity between the original thermal images and the enhanced ones. Then, we compare the results to the state-of-the-art methods, such as HE and CLAHE [21], as well as the TE-GAN [19]. The evaluation is conducted on the KAIST test set. Table 1 presents the results obtained from the comparative analysis. Our proposed TE-VGAN architecture outperforms other existing methods in terms of the visual quality of the resulting images, as demonstrated in Table 1. To further highlight the efficiency of the TE-VGAN architecture, Fig.5 illustrates the enhanced pedestal form with meticulous details compared to the original form. Specifically, TE-VGAN achieves a better balance between contrast and noise reduction, resulting in enhanced images

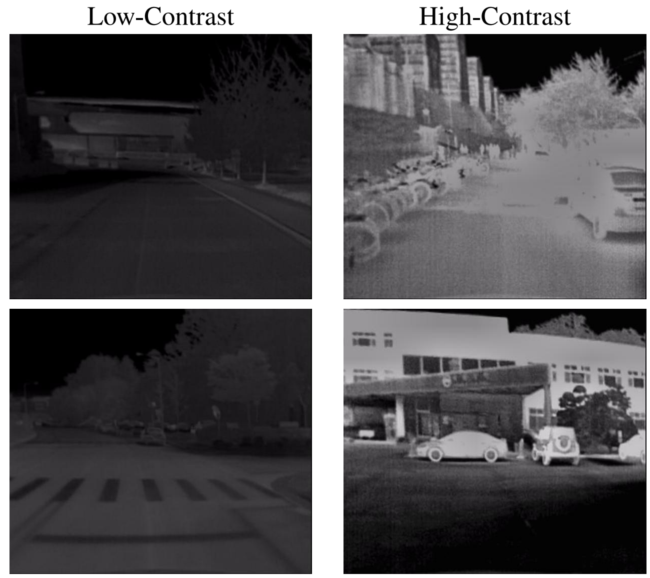


Figure 4. Examples of the training dataset.

Table 1. Comparison of the proposed TE-VGAN to other existing models for thermal image enhancement.

	HE	CLAHE	TE-GAN	TE-VGAN
PSNR	7.81	11.92	13.92	**15.0**
SSIM	0.34	0.37	0.50	**0.69**

Figure 5. Qualitative results of person detection.

with better visual quality. The results obtained from analyzing two sample images from the KAIST dataset are presented in Fig.6. The qualitative analysis shows that while other methods may have improved the contrast slightly, they also increased the visibility of noise.

4.3. Comparison results

To emphasize the significance of each step involved in the TE-VGAN architecture, Fig.7 illustrates the intermediate outcomes of each step. The visual quality of the im-

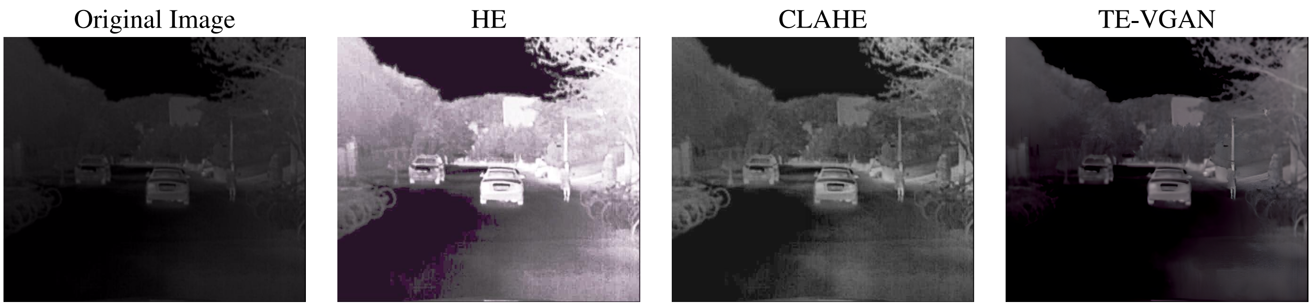


Figure 6. Qualitative results of our TE-VGAN for enhancement compared to other commonly used methods of contrast augmentation: HE, CLAHE.

ages is improved from different aspects by each step, including contrast enhancement, and restoration step, to address the various issues that thermal images commonly suffer from. These findings support the notion that the TE-VGAN method is based on a complementary approach of different steps. To further demonstrate the visual quality of the enhanced images, Fig.8 presents additional qualitative results obtained by applying various super-resolution methods. These outcomes were expected as these methods aim to enhance the resolution of thermal images without addressing the issues of low contrast and noisy details. Only the SRCNN method is an exception since it was trained on visible images, which accounts for the brighter appearance of the enhanced images.

4.3.1 Detection results

Table 2 presents the evaluation of the mAP performance of the YOLOv3 [33], YOLOv6 [16], and YOLOv7 [29] object detectors on images acquired during daytime, nighttime, and both conditions from the KAIST dataset. Furthermore, we compare the aforementioned results with those obtained after utilizing our proposed TE-VGAN architecture. The proposed thermal image enhancement approach resulted in significant improvements in object detection performance across all tested YOLO detectors. When utilizing the enhancement approach, the mAP scores for YOLOv3, YOLOv6, and YOLOv7 detectors improved by 6.8%, 1%, and 5.3%, respectively. Furthermore, as with the YOLOv7 detector, the enhancement's effect on detection performance was more pronounced during nighttime for YOLOv3 and YOLOv6 detectors, with an increase in mAP of 9.1% and 1.5%, respectively, in comparison to the daytime results. These findings indicate that the proposed enhancement approach can effectively improve the visual quality of thermal images, leading to improved object detection performance across different YOLO detectors. Therefore, utilizing the proposed enhancement approach can offer significant advantages in various real-world applications where reliable object detection in thermal images is essential. The object detection results utilizing the YOLOv7 detector with and without TE-VGAN are displayed in Fig.9. It can be shown that a discernible enhancement in detection performance after incorporating the TE-VGAN architecture. Multiple sample images are presented, wherein the enhancement approach corrects false negatives.

Table 2. Detection Performance Comparison of YOLO versions with and without enhancements using mAP(%).

Testing conditions	Model	Without enhancement	With enhancement
Day	YOLOv3 [33]	61.2	**65.5**
Night		66.1	**75.0**
All		62.3	**69.1**
Day	YOLOv6 [16]	67.1	67.7
Night		75.6	**77.1**
All		71.1	**72.2**
Day	YOLOv7 [29]	**62.7**	69.6
Night		76.4	**78.0**
All		66.8	**72.1**

5. Conclusion

In this paper, we introduced a novel hybrid architecture TE-VGAN based GANs and ViT for thermal image enhancement. The proposed method employs GANs with a U-NET-based generator and ViT models as global and local discriminators. The proposed TE-VGAN included a thermal loss to enhance the generated image quality. We leveraged also fine-tuning-based visible images for thermal image restoration. Our experimental evaluations demonstrate that our approach outperforms existing methods, achieving significant improvements in thermal image enhancement, in terms of quantitative and qualitative assessments and object detection tasks.

Figure 7. Details about intermediate results from the proposed TE-VGAN architecture

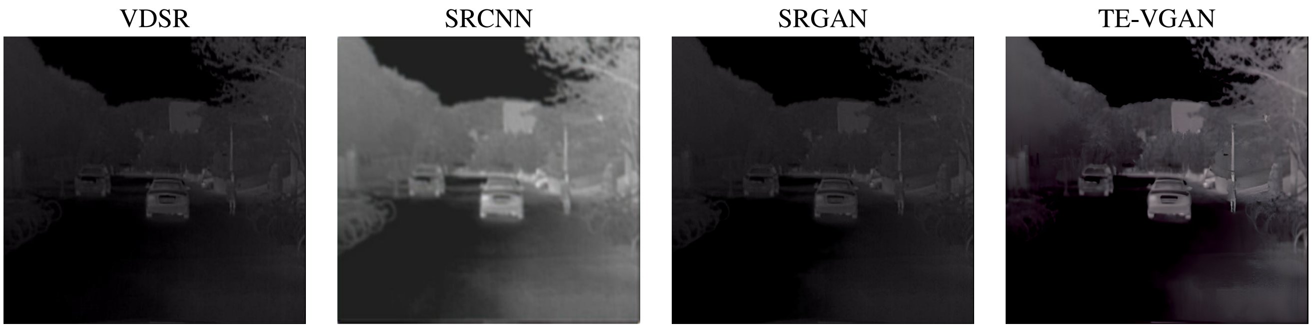


Figure 8. Qualitative results of different super-resolution methods: VDSR, SRCNN, and SRGAN.

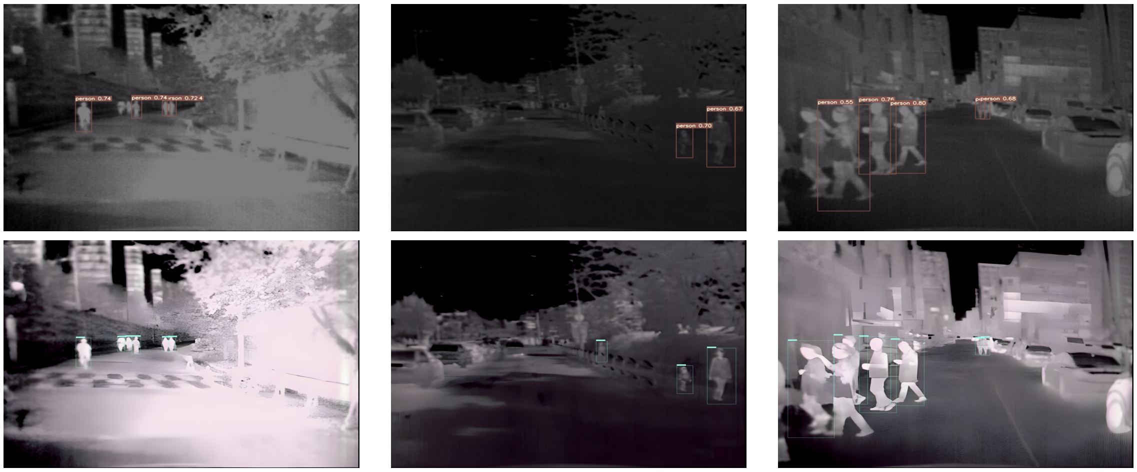


Figure 9. Examples of pedestrian detection using YOLOv7 on KAIST dataset with and without TE-VGAN.

References

[1] Yukyung Choi, Namil Kim, Soonmin Hwang, and In So Kweon. Thermal image enhancement using convolutional neural network. In *2016 IEEE/RSJ international conference on intelligent robots and systems (IROS)*, pages 223–230. IEEE, 2016.

[2] Yukyung Choi, Namil Kim, Soonmin Hwang, and In So Kweon. Thermal image enhancement using convolutional neural network. In *2016 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*. IEEE, 2016.

[3] Soner Civilibal, Kerim Kursat Cevik, and Ahmet Bozkurt. A deep learning approach for automatic detection, segmentation and classification of breast lesions from thermal images. *Expert Systems with Applications*, 212:118774, 2023.

[4] Chao Dong, Chen Change Loy, Kaiming He, and Xiaoou Tang. Image super-resolution using deep convolutional networks. *IEEE transactions on pattern analysis and machine intelligence*, 2015.

[5] Chao Dong, Chen Change Loy, and Xiaoou Tang. Accel-

erating the super-resolution convolutional neural network. *CoRR*, 2016.

[6] Dario Fuoli, Luc Van Gool, and Radu Timofte. Fourier space losses for efficient perceptual image super-resolution. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, 2021.

[7] Ian Goodfellow, Jean Pouget-Abadie, Mehdi Mirza, Bing Xu, David Warde-Farley, Sherjil Ozair, Aaron Courville, and Yoshua Bengio. Generative adversarial nets. In *Advances in neural information processing systems*, 2014.

[8] Axel-Christian Guei and Moulay Akhloufi. Deep learning enhancement of infrared face images using generative adversarial networks. *Applied optics*, 2018.

[9] Ziping He, Kewen Xia, Pedram Ghamisi, Yuhen Hu, Shurui Fan, and Baokai Zu. Hypervitgan: Semisupervised generative adversarial network with transformer for hyperspectral image classification. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, 15:6053–6068, 2022.

[10] Hao Hou, Jun Xu, Yingkun Hou, Xiaotao Hu, Benzheng Wei, and Dinggang Shen. Semi-cycled generative adversarial networks for real-world face super-resolution. *IEEE Transactions on Image Processing*, 2023.

[11] Yunliang Jiang, Lili Yan, Xiongtao Zhang, Yong Liu, and Danfeng Sun. Tcgan: Semantic-aware and structure-preserved gans with individual vision transformer for fast arbitrary one-shot image generation. *arXiv preprint arXiv:2302.08047*, 2023.

[12] Jiwon Kim, Jung Kwon Lee, and Kyoung Mu Lee. Accurate image super-resolution using very deep convolutional networks. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2016.

[13] Christian Ledig, Lucas Theis, Ferenc Huszár, Jose Caballero, Andrew Cunningham, Alejandro Acosta, Andrew Aitken, Alykhan Tejani, Johannes Totz, Zehan 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.

[14] Kyungjae Lee, Junhyeop Lee, Joosung Lee, Sangwon Hwang, and Sangyoun Lee. Brightness-based convolutional neural network for thermal image enhancement. *IEEE Access*, 5:26867–26879, 2017.

[15] Kyungjae Lee, Junhyeop Lee, Joosung Lee, Sangwon Hwang, and Sangyoun Lee. Brightness-based convolutional neural network for thermal image enhancement. *IEEE Access*, 2017.

[16] Chuyi Li, Lulu Li, Yifei Geng, Hongliang Jiang, Meng Cheng, Bo Zhang, Zaidan Ke, Xiaoming Xu, and Xiangxiang Chu. Yolov6 v3. 0: A full-scale reloading. *arXiv preprint arXiv:2301.05586*, 2023.

[17] Chengwei Liu, Xiubao Sui, Xiaodong Kuang, Yuan Liu, Guohua Gu, and Qian Chen. Optimized contrast enhancement for infrared images based on global and local histogram specification. *Remote Sensing*, 2019.

[18] Sameer Malik and Rajiv Soundararajan. Semi-supervised learning for low-light image restoration through quality assisted pseudo-labeling. In *Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision*, pages 4105–4114, 2023.

[19] Mohamed Amine Marnissi, Hajer Fradi, Anis Sahbani, and Najoua Essoukri Ben Amara. Thermal image enhancement using generative adversarial network for pedestrian detection. In *2020 25th International Conference on Pattern Recognition (ICPR)*, pages 6509–6516. IEEE, 2021.

[20] Mohamed Amine Marnissi, Hajer Fradi, Anis Sahbani, and Najoua Essoukri Ben Amara. Thermal image enhancement using generative adversarial network for pedestrian detection. In *2020 25th International Conference on Pattern Recognition (ICPR)*, 2021.

[21] Shelda Mohan and M Ravishankar. Modified contrast limited adaptive histogram equalization based on local contrast enhancement for mammogram images. In *International conference on advances in information technology and mobile communication*, pages 397–403. Springer, 2012.

[22] Maurício M Resende, Enrico B Gambare, Letícia A Silva, Yuri de S Cordeiro, Eneida Almeida, and Renan P Salvador. Infrared thermal imaging to inspect pathologies on façades of historical buildings: A case study on the municipal market of são paulo, brazil. *Case Studies in Construction Materials*, 16:e01122, 2022.

[23] Faezehsadat Shahidi. Breast cancer histopathology image super-resolution using wide-attention gan with improved wasserstein gradient penalty and perceptual loss. *IEEE Access*, 2021.

[24] Ali Shahsavari, Sima Ranjbari, and Toktam Khatibi. Proposing a novel cascade ensemble super resolution generative adversarial network (cesr-gan) method for the reconstruction of super-resolution skin lesion images. *Informatics in Medicine Unlocked*, 2021.

[25] P Shanmugavadivu and K Balasubramanian. Particle swarm optimized multi-objective histogram equalization for image enhancement. *Optics & laser technology*, 2014.

[26] Karen Simonyan and Andrew Zisserman. Very deep convolutional networks for large-scale image recognition. In *International Conference on Learning Representations*, 2015.

[27] Chunwei Tian, Xuanyu Zhang, Jerry Chun-Wen Lin, Wangmeng Zuo, and Yanning Zhang. Generative adversarial networks for image super-resolution: A survey. *arXiv preprint arXiv:2204.13620*, 2022.

[28] Vibashan Vs, Domenick Poster, Suya You, Shuowen Hu, and Vishal M Patel. Meta-uda: Unsupervised domain adaptive thermal object detection using meta-learning. In *Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision*, pages 1412–1423, 2022.

[29] Chien-Yao Wang, Alexey Bochkovskiy, and Hong-Yuan Mark Liao. Yolov7: Trainable bag-of-freebies sets new state-of-the-art for real-time object detectors. *arXiv preprint arXiv:2207.02696*, 2022.

[30] Xintao Wang, Liangbin Xie, Chao Dong, and Ying Shan. Real-esrgan: Training real-world blind super-resolution with pure synthetic data. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pages 1905–1914, 2021.

[31] Xiaogang Xu, Ruixing Wang, Chi-Wing Fu, and Jiaya Jia. Snr-aware low-light image enhancement. In *Proceedings of*

the IEEE/CVF Conference on Computer Vision and Pattern Recognition, pages 17714–17724, 2022.

[32] Wenlong Zhang, Yihao Liu, Chao Dong, and Yu Qiao. Ranksrgan: Generative adversarial networks with ranker for image super-resolution. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, 2019.

[33] Xunxun Zhang and Xu Zhu. Vehicle detection in the aerial infrared images via an improved yolov3 network. In *2019 IEEE 4th International Conference on Signal and Image Processing (ICSIP)*, pages 372–376. IEEE, 2019.

[34] Long Zhao, Zizhao Zhang, Ting Chen, Dimitris Metaxas, and Han Zhang. Improved transformer for high-resolution gans. In M. Ranzato, A. Beygelzimer, Y. Dauphin, P.S. Liang, and J. Wortman Vaughan, editors, *Advances in Neural Information Processing Systems*, volume 34, pages 18367–18380. Curran Associates, Inc., 2021.

[35] Xining Zhu, Lin Zhang, Lijun Zhang, Xiao Liu, Ying Shen, and Shengjie Zhao. Gan-based image super-resolution with a novel quality loss. *Mathematical Problems in Engineering*, 2020.

[36] Yannick Wend Kuni Zoetgnande, Jean-Louis Dillenseger, and Javad Alirezaie. Edge focused super-resolution of thermal images. In *2019 International Joint Conference on Neural Networks (IJCNN)*. IEEE, 2019.