Thermal Image Enhancement using Generative Adversarial Network for Pedestrian Detection

Mohamed Amine Marnissi, Hajer Fradi, Anis Sahbani, Najoua Essoukri Ben Amara

Thermal Image Enhancement using Generative Adversarial Network for Pedestrian Detection

Mohamed Amine Marnissi*†, Hajer Fradi‡, Anis Sahbani§¶ and Najoua Essoukri Ben Amara*
*Université de Sousse, Ecole Nationale d'Ingénieurs de Sousse, LATIS-Laboratory of Advanced Technology and Intelligent Systems, 4023, Sousse, Tunisie;
‡Université de Sousse, Institut Supérieur des Sciences Appliquées, LATIS-Laboratory of Advanced Technology and Intelligent Systems, 4023, Sousse, Tunisie;
†Université de Sfax, Ecole Nationale d'Ingénieurs de Sfax, 3038, Sfax, Tunisie;
§Enova Robotics S.A.
¶Sorbonne Université, CNRS, Ins titule for intelligent Systems and Robotics (ISIR), Paris, France
Emails: medamine.marnissi@eniso.u-sousse.tn, hajer.fradi@issatso.rnu.tn, anis.sahbani@enovarobotics.com and najoua.benamara@eniso.rnu.tn

***Abstract*—Infrared imaging has recently played an important role in a wide range of applications including video surveillance, robotics and night vision. However, infrared cameras often suffer from some limitations, essentially about low-contrast and blurred details. These problems contribute to the loss of observation of target objects in infrared images, which could limit the feasibility of different infrared imaging applications. In this paper, we mainly focus on the problem of pedestrian detection on thermal images. Particularly, we emphasis the need for enhancing the visual quality of images before performing the detection step. To address that, we propose a novel thermal enhancement architecture called TE-GAN based on Generative Adversarial Network, and composed of two modules contrast enhancement and denoising with a post-processing step for edge restoration in order to improve the overall image quality. The effectiveness of the proposed architecture is assessed by means of visual quality metrics and better results are obtained compared to the original thermal images and to the obtained results by other existing enhancement methods. These results have been conducted on a subset of KAIST dataset that we make available to encourage research in this direction[1]. Using the same dataset, the impact of the proposed enhancement architecture has been demonstrated on the detection results by obtaining better performance with a significant margin using YOLOv3 detector.**

I. Introduction

An infrared camera is a device that forms an image using infrared radiations, compared to the commonly used cameras that form an image using visible light [1]. This camera can detect infrared radiations emitted by an object having a high temperature. Instead of the 400-700 nm range of the visible light camera, infrared one operates in wavelengths as long as 700 nm - 1mm [2]. The main advantage using infrared cameras is that the temperature can be easily measured for dangerous objects while keeping the user out of danger since the thermography is a non-contact method. In addition, thermography allows the user to capture fast-moving targets and fast-changing thermal patterns of objects even in bad lighting conditions.

Because of all the aforementioned reasons, the past few decades have witnessed a widespread growth in the use of infrared cameras in many fields, including military and civilian ones especially for automotive applications, medical imaging, robotics and video surveillance [3]. Despite the usefulness of these cameras, there are some limitations that have to be considered, essentially about the compromise between the cost and the image quality. It is important to mention that the fabrication of high-resolution infrared cameras is extremely difficult and the manufacturing cost is more expensive regarding similar quality in visible cameras. Consequently, low-resolution thermal cameras are more commonly used. Low-resolution together with bad acquisition conditions in some cases, present multiple challenges such as low-contrast, noisy information and blurred details. These challenges make infrared imaging applications hard to perform well. It is essentially the case of various video analysis applications such as object detection, tracking and recognition.

Precisely, in this paper we focus on the problem of pedestrians detection and localization from low resolution infrared cameras for surveillance applications. The problem of pedestrian detection has been extensively studied using visible datasets and good results are usually obtained [4]. However, in some situations, for instance, in nighttime or in bad lighting conditions, the performance of the state-of-art detectors dramatically drop. Here comes the importance of using thermal cameras to detect persons because they could better distinguish humans which are warmer than surrounding objects.

Even though pedestrian detection using infrared cameras is more convenient during nighttime, it is still subject to errors if we take into account the highlighted problems mainly when the resolution is not sufficient. For these reasons, we intend in this present work to enhance the visual quality of thermal images in order to improve the detection performance. Basically inspired from EnlightenGAN [5] and DnCNN [6], we propose a new thermal enhancement architecture using Generative Adversarial Network (GAN). The proposed architecture is composed of two modules with a post-processing step: contrast enhancement and denoising modules and an edge restoration step. These different cues are merged together in one single architecture to complement each others and to further improve the overall quality. The proposed architecture covering multiple aspects is considered as the main contribution of this current paper. It has also the advantage to be

[1]https://github.com/AmineMarnissi/TE-GAN

trained on impaired images. In addition, its effectiveness is demonstrated on the detection performance by obtaining better results with a significant margin.

The rest of the paper is organized as follows: in Section II the related work for thermal images enhancement and adversarial learning are presented. Then, our proposed approach for thermal image enhancement is introduced in Section III. Details about the used person detector is given in Section IV. The experimental results are discussed in Section V. Finally, in Section VI we conclude and give some potential future works.

II. Related work

A. Thermal Image Enhancement

In this section, we give an overview of the existing methods for thermal image enhancement from traditional methods to deep learning ones. This overview includes enhancement methods for visible images as well.

1) Traditional Methods: Generally, infrared images are characterized by low-contrast, low-resolution and blurred details. To solve these issues, traditional methods already used for visible imaging could be adopted to enhance thermal images. Among these methods, Histogram Equalization (HE) is a commonly known algorithm that could readily augment the contrast. It has been employed either in its basic form or in an extended one. For instance, in [7], a multi-objective HE model has been proposed to enhance the contrast while preserving the brightness of thermal images. Also, Contrast Limited Adaptive Equalization (CLAHE) based on local contrast modification (LCM) is defined in [8]. It is proposed to highlight fine hidden details and to adjust the level of contrast enhancement. It is important to mention that the presented equalization histogram methods usually generate close results, but their application could accentuate the noise in the image.

2) Learning Methods: Deep Neural Networks (DNNs) have recently shown outstanding performance in many computer vision applications such as image classification, object detection and recognition. Some recently published methods for image enhancement have employed DNN architectures to improve the visual quality of thermal or visible images. One of the first attempts to handle this problem was published in [9] and is referred as SRCNN. It employs Convolutional Neural Network (CNN) for Super-Resolution. Its basic idea consists of learning a mapping relationship between low-resolution (LR) and high-resolution (HR) visible images. VDSR [10] is also one of the most known deep learning methods for enhancement which aims at augmenting the spatial resolution of visible images.

While most of the existing methods for image enhancement in visible domain focus on increasing the spatial resolution of the original image, only few studies for thermal image enhancement that cover other aspects such as low contrast and blurred edges have been conducted. More in details, TEN architecture [11] is 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. Fan *et al.* [12] also designed a CNN architecture but to improve the contrast between the target and the background by highlighting the target and suppressing background clutters. In [13], Lee *et al.* propose to incorporate the brightness domain with a residual-learning technique in order to improve the performance of enhancement and the speed of convergence. Related work includes as well CDN-MRF [14], which introduces a cascaded architecture composed of two consecutive deep networks with different receptive fields that are jointly trained to increase the spatial resolution of thermal image by a large scale factor. Finally, in [15], an edge-focused method is proposed. It consists of a model based on residual dense blocks, that can perform super-resolution for thermal images, while enhancing the visual information of edges.

B. Adversarial Learning

Generative Adversarial Networks (GANs) are deep learning architectures initially introduced by Goodfellow *et al.* in [16]. 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. Also, they have been employed in few studies for image enhancement in visible and thermal domains. For instance, in [17], SRGAN (which refers to Super-Resolution GAN) that includes deep residual network (ResNet) with skip-connection and defines a perceptual loss is proposed. Another related work that made use of GANs for super-resolution infrared image is presented in [18]. Precisely, it employs deep convolutional generative adversarial networks (DCGAN) for infrared face images. Compared to the two previous works that employ GAN architectures for super-resolution in both visible and infrared domains, in [19], the authors rather focused on the problem of enhancing the contrast in infrared images using the conditional generative adversarial networks. Also, in [20], a refined convolutional neural architecture that produces results with higher contrast and sharper details is proposed. But in this approach, visible images are used for training and the network is applied to infrared images.

Following the same strategy, in this current paper, we make use of GAN for thermal image enhancement. But, we propose a more complete architecture that simultaneously deals with different aspects (low contrast, noise, and blurred edges). Moreover, differently from previous works in thermal domain, where only grayscale converted images are used for training, our proposed architecture is properly trained on impaired images from thermal domain.

III. Proposed Approach for Thermal Image Enhancement

In this paper, we propose a new generative adversarial network for thermal image enhancement task. Our proposed architecture is composed of two main modules, the first one is proposed to improve the contrast of the image. Since, by augmenting the contrast, the noise will be gradually more visible, we mitigate this effect by a second module that aims at removing the underlying noise. Both models operate

instantly and simultaneously in an end-to-end architecture. The overall proposed architecture called TE-GAN, which stands for Thermal Enhancement Generative Adversarial Network is shown in Fig.1. The remainder of this section describes each of these architecture components.

A. Contrast Enhancement Module:

1) U-NET Generator : The first generator in our proposed architecture is an attention-guided U-Net generator. U-Net [21] has been widely applied in various applications, including semantic segmentation [22] and image restoration [23]. It aims at extracting multi-level features from different depth layers and at generating high quality images using multi-scale and texture information. Moreover, the attention map has shown its usefulness to improve the visual quality, which is an element-wise difference 1-I, where I represents an illumination channel of the thermal image normalized to [0,1]. It is based on re-sizing each feature map and multiplying with all intermediate feature maps and the output image.

In our proposed TE-GAN architecture, the U-NET generator is composed of 8 convolutional blocks. Each block includes two 3 x 3 convolutional layers, followed by LeakyReLu and a batch normalization layer. At the upsampling part, to minimize the checkerboard effects, the standard deconvolutional layer is changed by one bilinear upsampling layer plus one convolutional layer.

2) Adversarial Loss: GAN architectures are mainly composed of a generator which is a CNN network aiming at producing the target image. This generator is coupled with a discriminator network which learns to distinguish between fake and real data in order to generate better image quality. In our particular case, since our goal is to improve the contrast, using only one global discriminator is not sufficient to adaptively enhance the local regions. Therefore, we resort to a local discriminator that randomly takes cropped patches from ground-truth and generated images, to be able to distinguish real from enhanced images.

Differently from classifying the complete image, the Patch-GAN or Markovian discriminator [24] can be used to classify single patches in a given image as real or fake. These methods have the advantage of running faster since fewer parameters are used. In our architecture, we employ PatchGAN for real and fake discrimination in both local and global discriminators. In addition, for the global discriminator structure, we modify the relativistic discriminator [25] by replacing the sigmoid function with the least square GAN loss (LSGAN) [26]. The relativistic discriminator estimates the probability that real data is more realistic than fake one. The standard function of relativistic discriminator is defined as:

$$D_{Ra}(x_r, x_f) = \sigma(C(x_r) - \mathbb{E}_{x_f \sim \mathbb{P}_{fake}}[C(x_f)]) \quad (1)$$

$$D_{Ra}(x_f, x_r) = \sigma(C(x_f) - \mathbb{E}_{x_r \sim \mathbb{P}_{real}}[C(x_r)]) \quad (2)$$

where C denotes the discriminator network, x_r and x_f are sampled from the real and fake distributions respectively and σ represents the sigmoid function. The loss functions for the global discriminator D and the generator G are defined as:

$$L_D^{Global} = \mathbb{E}_{x_r \sim \mathbb{P}_{real}}[D_{Ra}(x_r, x_f) - 1)^2] + \mathbb{E}_{x_f \sim \mathbb{P}_{fake}}[D_{Ra}(x_f, x_r)^2] \quad (3)$$

$$L_G^{Global} = \mathbb{E}_{x_f \sim \mathbb{P}_{fake}}[D_{Ra}(x_f, x_r) - 1)^2] + \mathbb{E}_{x_r \sim \mathbb{P}_{real}}[D_{Ra}(x_r, x_f)^2] \quad (4)$$

For the local discriminator, 5 patches from the output and real images are randomly cropped each time. The loss functions are defined as:

$$L_D^{Local} = \mathbb{E}_{x_r \sim \mathbb{P}_{real-patches}}[D(x_r) - 1)^2] + \mathbb{E}_{x_f \sim \mathbb{P}_{fake-patches}}[D(x_f) - 0)^2] \quad (5)$$

$$L_G^{Local} = \mathbb{E}_{x_r \sim \mathbb{P}_{fake-patches}}[D(x_f) - 1)^2] \quad (6)$$

3) Perceptual Loss: In [27], Johnson *et al.* proposed a perceptual loss for computing the distance between the output image and its ground truth based on high-level representations extracted from VGG pre-trained model. In our proposed architecture, we use self feature preserving loss [28], which enables the auto-adjustment of the network for preserving the features content of the image. In addition, we add an instance normalization layer [29] after each feature map before feeding them into L_{SFP}^{Global} and L_{SFP}^{Local} in order to stabilize the training. Precisely, the loss function is applied for local and global discriminators L_{SFP}^{Local}, L_{SFP}^{Global}, respectively, and is defined as:

$$L_{SFP}(I^L) = \frac{1}{W_{i,j} H_{i,j}} \sum_{y=1}^{H_{i,j}} (\phi_{i,j}(I^L) - \phi_{i,j}(G(I^L)))^2 \quad (7)$$

where $W_{i,j}$ and $H_{i,j}$ are the dimensions of the extracted feature maps. I^L is the input low-light of thermal image and G(I^L) denotes the output of the generator. $\phi_{i,j}$ is the feature map extracted from a pre-trained VGG-16 model. *i* is the *i*-th max pooling, and j is the *j*-th convolutional layer after *i*-th max pooling layer.

B. Denoising Module

1) Denoising Generator: One of the most common problems in image enhancement task in visible and thermal images is the distribution of noise. Especially, in our case when the contrast is augmented, much noise in the full image could be revealed. For this reason, we propose a second generative adversarial network to complement the first one. This network aims at enhancing the thermal image once again by removing the noise level. Our denoising GAN consists of a generator and a discriminator networks. The generator is a CNN network that utilizes residual learning and batch normalization to speed up and to augment the performance of the training step. The global discriminator employs the PatchGAN structure and the relativistic function defined in Eq. 4.

In our approach, the architecture of the denoising generator network is composed of 7 convolutional layers using 64 filters of size 3 x 3, each one of them is followed by a ReLU activation layer and a batch normalization layer except for the first and last layers. To remove the noise from image in the hidden layer, our model makes use of the residual

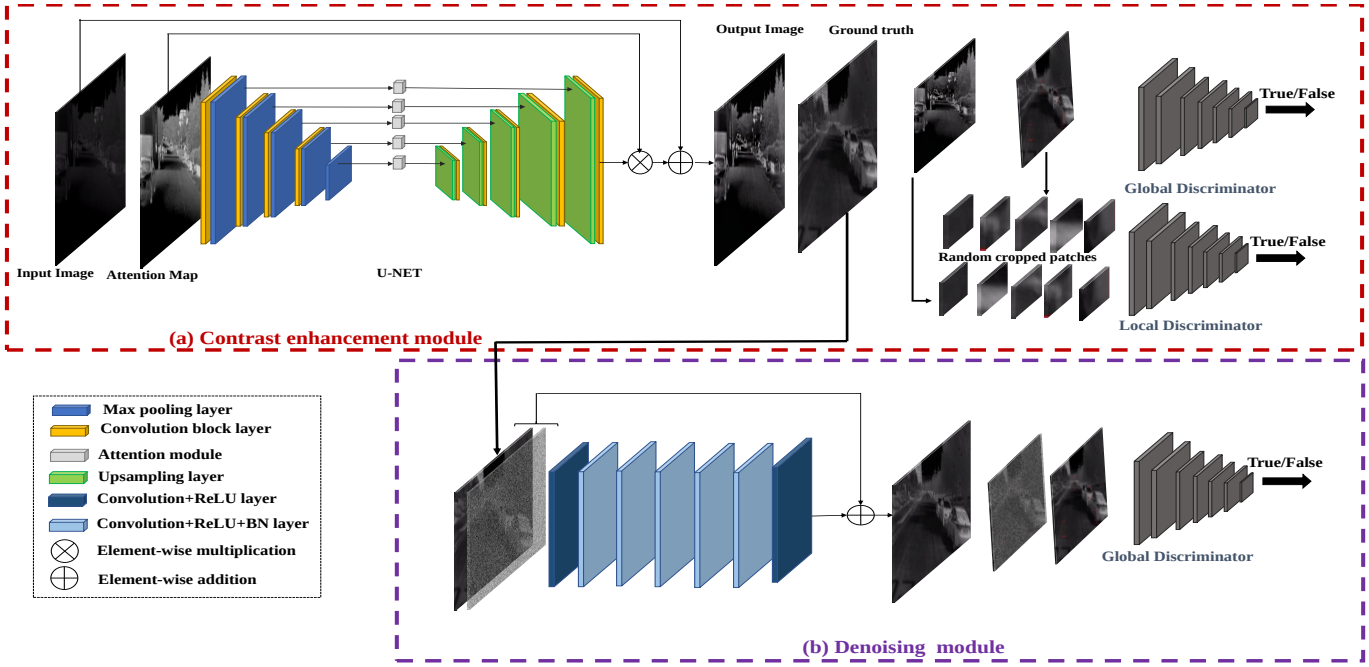

Fig. 1. The proposed TE-GAN architecture composed of two modules: (a) contrast enhancement and (b) denoising.

learning strategy by employing an individual residual unit to predict the residual image. The input of the network is a noisy observation image from the ground-truth label of high-contrast and a random level of Gaussian noise [0,50].

2) Content Loss: The content loss function ensures that the content information found in the ground-truth label image is also captured in the generated image. In the second step in our architecture, for which the goal is to remove the noise from thermal image, we use pixel-wise MSE that stimulates the network to minimize the low-level content errors between the input (noisy image) and the denoised generated image.

$$L_{content} = \frac{1}{WH} \sum_{i=1}^{W} \sum_{j=1}^{H} \parallel y_{i,j} - G(x)_{i,j} \parallel_1 \qquad (8)$$

where W and H denote the height and width of the thermal image respectively and $\parallel \ . \ \parallel_1$ denotes the L_1 norm. $y_{i,j}$ represents the pixel values of ground truth image and $G(x)_{i,j}$ are the pixel values of the generated image.

C. Summary of the proposed TE-GAN architecture

To train TE-GAN architecture by comparing the generated images to the ground truth ones, the previous loss functions defined in Eq.7 and Eq.8 are added to the adversarial loss L_{adv} (which includes global and local losses of the first generator and the global loss of the second generator defined in Eq.4, Eq.6) in an overall loss function L defined as:

$$L = L_{SFP}^{Global} + L_{SFP}^{Local} + L_{adv} + L_{content} \qquad (9)$$

Once both of contrast enhancement and denoising modules are performed, we propose to apply a convolutional edge enhancement filter that improves the edges in thermal images and decreases the visual blur effects. Using Pillow library, we choose to apply "edge_enhance_more" filter of size 3x3, with 9 in the center and -1 for the remaining values.

IV. Pedestrian Detection

Object detection is a common task in the research area of video analysis and its results lay the foundations of a wide range of applications. It consists of precisely identifying and localizing pertinent objects in a single image by classification or by regression. The current popular detectors make use of deep learning networks such as Fast R-CNN [30], Faster R-CNN [31], Single Shot Detector (SSD) [32] and You Only Look Once (YOLO)[33]. Generally, object detection models can be divided into two categories. The first category requires two single shots: the first one consists of generating region proposal networks (RPN) and the second one aims at detecting objects of each proposal such as the case of Fast R-CNN and Faster R-CNN. The second category is based on one shot to detect several objects such as SSD and YOLO detectors.

YOLOv3 [34] is the third version of object detection algorithm in YOLO family. It is one of the most popular and recent real-time object detectors, which improves the accuracy compared to many methods such as ResNet and Feature Pyramid Networks (FPN) structure. The YOLOv3 network makes the prediction at 3 scales to enable multi-scale detection as the case of FPN.

It is important to mention that by applying available pre-trained models of YOLOv3 detector trained on visible datasets, using thermal images, a drop in the detection accuracy can be clearly observed, since the two domains (visible and thermal) exhibit different visual characteristics. This justifies the need for training the detector on a thermal dataset. Once the model

is obtained on thermal domain, the detection performance can be compared to the results on the same testing dataset but after enhancing the visual quality of images by our proposed enhancement architecture presented in Section III. Obviously, a prior employment of the enhancement approach can be explored for other applications such as object tracking and recognition in order to improve the overall performance in thermal domain.

V. Experimental results

A. Dataset and Experiments

The proposed approach is evaluated on KAIST (Korea Advanced Institute of Science & Technology) dataset [35]. It is one of the largest multi-spectral pedestrian dataset composed of aligned visible and Long-Wave Infrared (LWIR) images under adverse illumination conditions, day and night. It approximately consists of 95k frames on urban traffic environment and of dense annotations for 1182 different pedestrians. This dataset is divided into a training set of 50.2k images from Set 00 to Set 05, and a test set of 45.1k images from Set 06 to Set 11. In our work, only thermal images from this dataset are used.

For thermal image enhancement, we noticed the absence of available standard datasets that simultaneously contain original and enhanced thermal images. This problem represents an obstacle to evaluate enhancement techniques that improve the visual quality of thermal images through deep learning architectures. In some previous works [11], [20], visible images are used for training after converting them to grayscale, but this approach is not convincing enough since thermal images are visually different compared to grayscale ones.

To mitigate this problem, in this current study, we build an unpaired thermal subset of KAIST dataset, composed of two parts: low and high contrast 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. Since the contrast in KAIST dataset is not usually sufficient, even for images in which the contrast is above the predefined threshold, we apply CLAHE to better augment it. We make the built subset of KAIST available to boost research in this direction[2]. Our obtained results using the proposed TE-GAN architecture on this subset of KAIST are available as well to encourage comparisons in the future.

As already mentioned, since there is no standard dataset and common evaluation protocol for thermal image enhancement, only results using some traditional methods such as Histogram Equalization (HE) and Contrast Limited Adaptive Histogram Equalization (CLAHE) [8] on our constructed subset of KAIST are reported and compared to our obtained results in terms of Peak Signal-to-Noise Ratio (PSNR) and Structure Similarity Index (SSIM) [19]. To better prove the effectiveness of our proposed enhancement architecture, our results are qualitatively evaluated and compared to SRCNN [9], VDSR [10] and SRGAN [17].

For pedestrian detection, we train YOLOv3 detector following the benchmark protocol that comes with KAIST dataset and we adopt the evaluation method presented in [36]. Precisely, we select every 3 frames from training sets and every 20 frames from testing sets, and we only consider the non-occluded, non-truncated and large (> 50) instances. This results in a training set of 7601 images (4755 day, 2846 night) and a testing set of 2252 images (1455 day, 797 night).

The performance of pedestrian detector in thermal images is evaluated before applying the enhancement step, in terms of mean Average Precision (mAP) of detections at Intersection Over Union (IOU) equal to 0.5 regarding the ground truth boxes. Also, the Log Average Miss Rate (LAMR) over the range of $[10^{-2},10^{0}]$ against the False Positives Per Image (FPPI) is reported. These results are compared to those obtained on enhanced thermal images after applying our proposed TE-GAN architecture.

B. Implementation details

We choose to train TE-GAN architecture over 200 epochs: the first 100 epochs with a learning rate of 0.0004 and the last 100 epochs with a learning rate decayed linearly to 0. For optimization, we use the Adam optimizer with a mini-batch of 4 images. We learned our model on an NVIDIA Titan X GPU with 12GB RAM. Also, YOLOv3 detector is trained on 100 epochs initialized with Darknet-53 pretrained model on COCO dataset, with a mini-batch size of 8 images using the Adam optimizer as well.

C. Results and Analysis

1) Results of the proposed TE-GAN architecture: To evaluate the proposed TE-GAN architecture, PSNR and SSIM metrics are calculated between the original images and the enhanced ones. The results are reported in Table I and compared to HE and CLAHE [8] methods on the testing set of KAIST (only thermal data).

TABLE I
Comparison of the proposed TE-GAN architecture to other existing methods of contrast augmentation in terms of PSNR and SSIM

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

As shown in the table, our proposed architecture TE-GAN gives better visual quality of images after enhancement compared to other commonly used methods of contrast augmentation. The corresponding qualitative results on two sample images from KAIST dataset are also shown in Figure 2. From these results, we can clearly observe that for other methods even though the contrast is slightly improved, the visible noise is more accentuated.

[2]https://github.com/AmineMarnissi/TE-GAN

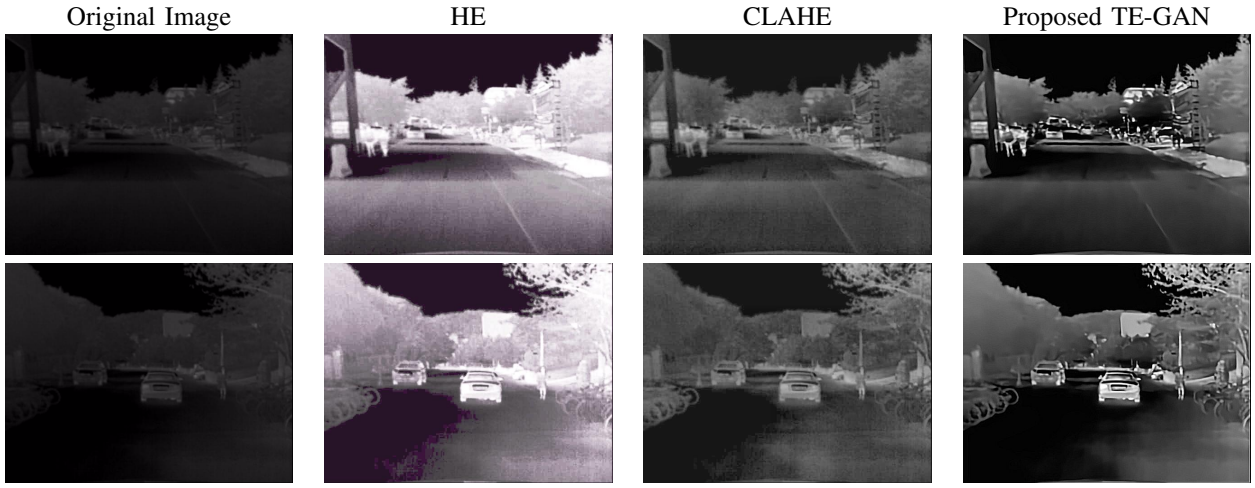

Fig. 2. Qualitative results of our proposed architecture TE-GAN for enhancement compared to other commonly used methods of contrast augmentation: HE and CLAHE on two sample images.

To better highlight the importance and the relevance of each step in TEN-GAN architecture, in Figure 3, the intermediate results of each step are visualized on the same sample images shown in Figure 2. As demonstrated in the figure, each step (contrast enhancement, denoising and edge enhancement) affects the visual quality from different aspect in order to respond to all aforementioned problems from which suffer thermal images. These results justify the complementary aspect of different steps on which the proposed TE-GAN is based.

Always about the visual quality of enhanced images, we show in Figure 4, other qualitative results by employing different super-resolution methods on the same sample images of Figure 2. The obtained results in this figure are expected since such SR methods aim at augmenting the resolution of thermal images and do not address the problems of low contrast and noisy details. Only SRCNN method is excepted since the architecture has been trained on visible images, which justifies that the enhanced images appear brighter.

2) Results of pedestrian detection : In Table II, we evaluate the performance of YOLOv3 detector in terms of mAP and LAMR for images at daytime, nighttime and both from KAIST dataset. These results are compared to those obtained after enhancing the visual quality of the same data by applying our proposed TE-GAN architecture.

TABLE II
COMPARISON OF THE DETECTION PERFORMANCE OF YOLOV3 DETECTOR WITH AND WITHOUT ENHANCEMENT

Testing conditions	Metric	Without enhancement	With enhancement
Day	mAP	0.61	**0.63**
	LAMR	0.41	**0.40**
Night	mAP	0.66	**0.73**
	LAMR	0.26	**0.20**
All	mAP	0.62	**0.65**
	LAMR	0.45	**0.43**

The baseline detector (YOLOv3) trained and tested on thermal images without enhancement achieves a mAP of 62% and LAMR of 45%. These results are improved while considering the enhancement architecture by a margin of 3% in terms of mAP and 2% in terms of LAMR, which proves the effectiveness of enhancing the image visual quality before detection. By comparing day and night results, as expected, the margin of improvement between with and without enhancement is more significant in nighttime since the temperature decreases during the night. These results comply with our main proposal stated at the beginning of the paper. In Figure 5, we show some results of detections with and without enhancement. As depicted in this figure the detection performance is improved. We show different sample images, where some false positives or false negatives are corrected by TE-GAN enhancement architecture.

VI. CONCLUSION

In this paper, we proposed a novel thermal enhancement architecture TE-GAN, composed of contrast enhancement and denoising modules using Generative Adversarial Network. This architecture has the advantage of improving the overall quality of thermal images. By means of tests on KAIST dataset, the effectiveness of the proposed architecture is proven by obtaining better quantitative and qualitative results compared to the original thermal images and to the obtained results by other existing enhancement methods.

Furthermore, the impact of the proposed enhancement architecture has been demonstrated on the detection results by obtaining better performance with a significant margin using YOLOv3 detector. There are several possible extensions of this paper. For instance, given the importance of enhancing the visual quality of thermal datasets for video analysis, this work could be extended to other applications such as person tracking and activity recognition. Also, an extension of the proposed TE-GAN architecture to incorporate a super-resolution module could be investigated as well.

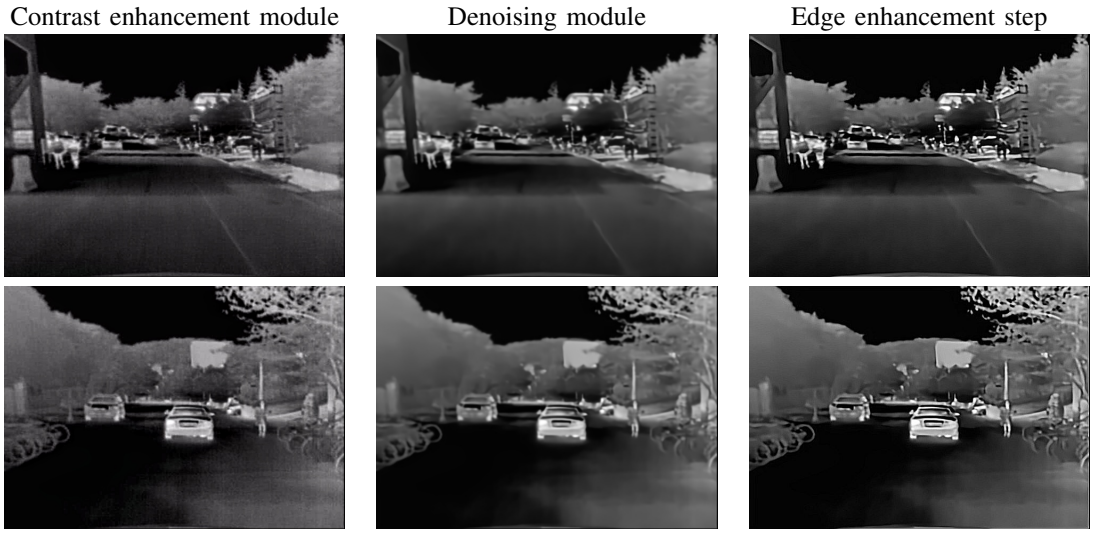

Fig. 3. Details of intermediate results from the proposed TE-GAN architecture, showing the effects of each step on the same sample images of Figure 2.

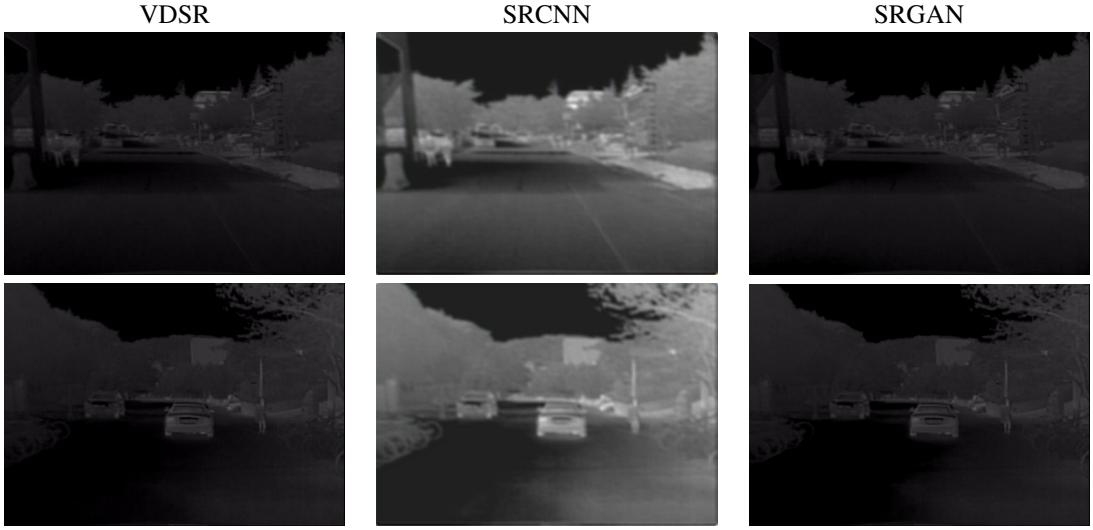

Fig. 4. Qualitative results of different super-resolution methods: VDSR, SRCNN and SRGAN on the same sample images of Figure 2.

REFERENCES

[1] 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.

[2] R. Bonaldi, "Functional finishes for high-performance apparel," in *High-Performance Apparel*. Elsevier, 2018, pp. 129–156.

[3] I. T. Ćirić, Ž. M. Ćojbašić, D. D. Ristić-Durrant, V. D. Nikolić, M. V. Ćirić, M. B. Simonović, and I. R. Pavlović, "Thermal vision based intelligent system for human detection and tracking in mobile robot control system," *Thermal science*, vol. 20, pp. 1553–1559, 2016.

[4] A. Mhalla, T. Chateau, S. Gazzah, and N. Essoukri Ben Amara, "An embedded computer-vision system for multi-object detection in traffic surveillance," *IEEE Transactions on Intelligent Transportation Systems*, vol. 20, no. 11, pp. 4006–4018, 2018.

[5] Y. Jiang, X. Gong, D. Liu, Y. Cheng, C. Fang, X. Shen, J. Yang, P. Zhou, and Z. Wang, "Enlightengan: Deep light enhancement without paired supervision," *arXiv preprint arXiv:1906.06972*, 2019.

[6] K. Zhang, W. Zuo, Y. Chen, D. Meng, and L. Zhang, "Beyond a gaussian denoiser: Residual learning of deep cnn for image denoising," *IEEE Transactions on Image Processing*, vol. 26, no. 7, pp. 3142–3155, 2017.

[7] P. Shanmugavadivu and K. Balasubramanian, "Particle swarm optimized multi-objective histogram equalization for image enhancement," *Optics & laser technology*, vol. 57, pp. 243–251, 2014.

[8] S. 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*. Springer, 2012, pp. 397–403.

[9] 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.

[10] J. Kim, J. Kwon Lee, and K. Mu Lee, "Accurate image super-resolution using very deep convolutional networks," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2016, pp. 1646–1654.

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

[12] Z. Fan, D. Bi, L. Xiong, S. Ma, L. He, and W. Ding, "Dim infrared image enhancement based on convolutional neural network," *Neurocomputing*, vol. 272, pp. 396–404, 2018.

[13] K. Lee, J. Lee, J. Lee, S. Hwang, and S. Lee, "Brightness-based

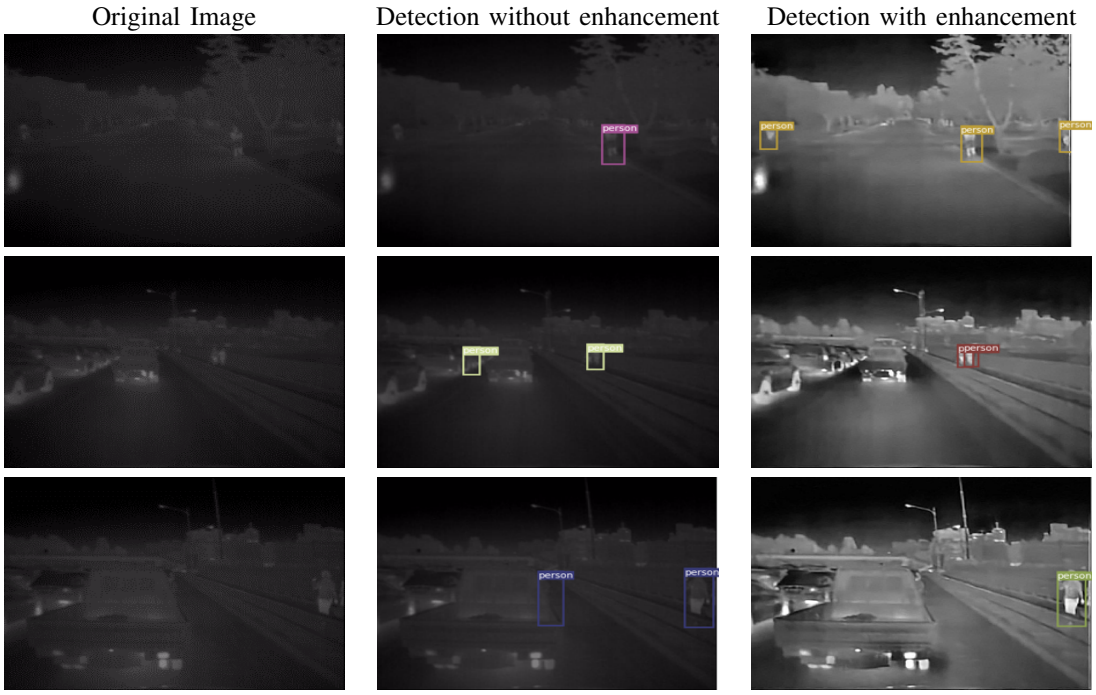

Fig. 5. Some results of pedestrian detection using YOLOv3 on thermal images from KAIST dataset with and without enhancement. Note that the bounding boxes are drawn on random colors.

convolutional neural network for thermal image enhancement," *IEEE Access*, vol. 5, pp. 26 867–26 879, 2017.

[14] Z. He, S. Tang, J. Yang, Y. Cao, M. Y. Yang, and Y. Cao, "Cascaded deep networks with multiple receptive fields for infrared image super-resolution," *IEEE transactions on circuits and systems for video technology*, vol. 29, no. 8, pp. 2310–2322, 2018.

[15] Y. W. K. Zoetgnande, J.-L. Dillenseger, and J. Alirezaie, "Edge focused super-resolution of thermal images," in *2019 International Joint Conference on Neural Networks (IJCNN)*. IEEE, 2019, pp. 1–8.

[16] I. Goodfellow, J. Pouget-Abadie, M. Mirza, B. Xu, D. Warde-Farley, S. Ozair, A. Courville, and Y. Bengio, "Generative adversarial nets," in *Advances in neural information processing systems*, 2014, pp. 2672–2680.

[17] 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.

[18] A.-C. Guei and M. Akhloufi, "Deep learning enhancement of infrared face images using generative adversarial networks," *Applied optics*, vol. 57, no. 18, pp. D98–D107, 2018.

[19] K. Lee, J. Lee, J. Lee, S. Hwang, and S. Lee, "Brightness-based convolutional neural network for thermal image enhancement," *IEEE Access*, vol. 5, pp. 26 867–26 879, 2017.

[20] X. Kuang, X. Sui, Y. Liu, Q. Chen, and G. Gu, "Single infrared image enhancement using a deep convolutional neural network," *Neurocomputing*, vol. 332, pp. 119–128, 2019.

[21] O. Ronneberger, P. Fischer, and T. Brox, "U-net: Convolutional networks for biomedical image segmentation," in *International Conference on Medical image computing and computer-assisted intervention*. Springer, 2015, pp. 234–241.

[22] O. Mechi, M. Mehri, R. Ingold, and N. E. B. Amara, "Text line segmentation in historical document images using an adaptive u-net architecture," in *2019 International Conference on Document Analysis and Recognition (ICDAR)*. IEEE, 2019, pp. 369–374.

[23] D. Liu, B. Wen, X. Liu, Z. Wang, and T. S. Huang, "When image denoising meets high-level vision tasks: A deep learning approach," *arXiv preprint arXiv:1706.04284*, 2017.

[24] P. Isola, J.-Y. Zhu, T. Zhou, and A. A. Efros, "Image-to-image translation with conditional adversarial networks," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2017, pp. 1125–1134.

[25] A. Jolicoeur-Martineau, "The relativistic discriminator: a key element missing from standard gan," *arXiv preprint arXiv:1807.00734*, 2018.

[26] X. Mao, Q. Li, H. Xie, R. Y. Lau, and Z. Wang, "Multi-class generative adversarial networks with the l2 loss function," *arXiv preprint arXiv:1611.04076*, vol. 5, pp. 1057–7149, 2016.

[27] J. Johnson, A. Alahi, and L. Fei-Fei, "Perceptual losses for real-time style transfer and super-resolution," in *European conference on computer vision*. Springer, 2016, pp. 694–711.

[28] B. RichardWebster, S. E. Anthony, and W. J. Scheirer, "Psyphy: A psychophysics driven evaluation framework for visual recognition," *IEEE transactions on pattern analysis and machine intelligence*, vol. 41, no. 9, pp. 2280–2286, 2018.

[29] D. Ulyanov, A. Vedaldi, and V. Lempitsky, "Improved texture networks: Maximizing quality and diversity in feed-forward stylization and texture synthesis," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2017, pp. 6924–6932.

[30] R. Girshick, "Fast r-cnn," in *Proceedings of the IEEE international conference on computer vision*, 2015, pp. 1440–1448.

[31] S. Ren, K. He, R. Girshick, and J. Sun, "Faster r-cnn: Towards real-time object detection with region proposal networks," in *Advances in neural information processing systems*, 2015, pp. 91–99.

[32] W. Liu, D. Anguelov, D. Erhan, C. Szegedy, S. Reed, C.-Y. Fu, and A. C. Berg, "Ssd: Single shot multibox detector," in *European conference on computer vision*. Springer, 2016, pp. 21–37.

[33] J. Redmon, S. Divvala, R. Girshick, and A. Farhadi, "You only look once: Unified, real-time object detection," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2016, pp. 779–788.

[34] J. Redmon and A. Farhadi, "Yolov3: An incremental improvement," *arXiv preprint arXiv:1804.02767*, 2018.

[35] S. Hwang, J. Park, N. Kim, Y. Choi, and I. So Kweon, "Multispectral pedestrian detection: Benchmark dataset and baseline," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2015, pp. 1037–1045.

[36] D. Ghose, S. M. Desai, S. Bhattacharya, D. Chakraborty, M. Fiterau, and T. Rahman, "Pedestrian detection in thermal images using saliency maps," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops*, 2019.