Thermal image super-resolution: a novel unsupervised approach

Rafael E. Rivadeneira[1 [0000−0002−5327−2048]], Angel D. Sappa[1,2 [0000−0003−2468−0031]], and Boris X. Vintimilla[1 [0000−0001−8904−0209]]

[1] Escuela Superior Politécnica del Litoral, ESPOL, Facultad de Ingeniería en Electricidad y Computación, CIDIS,
Campus Gustavo Galindo Km. 30.5 Vía Perimetral, P.O. Box 09-01-5863,
Guayaquil, Ecuador
{rrivaden, asappa, boris.vintimilla}@espol.edu.ec
[2] Computer Vision Center, Edifici O, Campus UAB, 08193
Bellaterra, Barcelona, Spain

Abstract. This paper proposes the use of a CycleGAN architecture for thermal image super-resolution under a transfer domain strategy, where middle-resolution images from one camera are transferred to a higher resolution domain of another camera. The proposed approach is trained with a large dataset acquired using three thermal cameras at different resolutions. An unsupervised learning process is followed to train the architecture. Additional loss function is proposed trying to improve results from the state of the art approaches. Following the first thermal image super-resolution challenge (PBVS-CVPR2020) evaluations are performed. A comparison with previous works is presented showing the proposed approach reaches the best results.

Keywords: Thermal image Super-Resolution, Thermal images, Datasets, Challenge, unpair thermal images.

1 Introduction

Single Image Super-Resolution (SISR) is a challenging ill-posed problem that refers to the task of restoring high-resolution (HR) images from a low-resolution (LR) image of the same scene, usually with the use of digital image processing or Machine Learning (ML) techniques. Super-Resolution is wide used in several applications, such as medical imaging (e.g., [Mudunuri and Biswas, 2015]), object detection (e.g., [Girshick et al., 2015]), security (e.g., [Zhang et al., 2010], [Shamsolmoali et al., 2019]), among others. In recent years deep learning techniques have been applied to SISR problem achieving remarkable results with respect to state-of-the-art approaches. Most of these techniques are focused on the visible domain—i.e., RGB images. Long-wavelength infrared (LWIR) images, also referred to in the literature as thermal images, have become very important in several challenging fields (e.g., dark environments for military security, medicine for mama cancer detection [Herrmann et al., 2018] or for car driving

assistance [Ding et al., 2019] are just three examples where these images can be used). Thermal cameras capture the information of LWIR spectra, which is the radiation emitted by the object's surface when their temperature is above zero [Gade and Moeslund, 2014].

Unfortunately, most of the thermal cameras in the market have poor resolution, due to the technology limitation and the high price of that technology. Thermal cameras with a high-resolution are considerably expensive with respect to low-resolution ones. Due to this limitation and a large number of applications based on their use, single thermal image super-resolution (SThISR) has become an attractive research topic in the computer vision community.

In the visible spectrum exists thousands of images captured with HD cameras, which are very useful for training networks used in the SISR problem. On the contrary to the visible spectrum domain, thermal images tend to have a poor resolution and there are a few HD datasets. Due to the lack of thermal images, a novel dataset was recently proposed in [Rivadeneira et al., 2020a] containing images with three different resolutions (low, mid, and high) obtained with three different thermal cameras. This dataset has been used in the first thermal image super-resolution challenge on PBVS-CVPR2020 conference, where several teams have participated and a baseline has been obtained. The current work is focused on two topics; firstly, a novel CycleGAN architecture is proposed, which makes use of a novel loss function (SOBEL cycle loss) to achieve better results than the ones obtained in PBVS-CVPR2020 challenge [Rivadeneira et al., 2020b]; secondly, the dataset presented in PBVS-CVPR2020 is enlarged with new images that help to generalize the training phase to ensure that the architecture is enough to SR any thermal image characteristics.

The manuscript is organized as follows. Section 2 presents works related to the topics tackled in the current work. The used datasets and the proposed architecture are detailed in Section 3. Results are provided in Section 4. Finally, conclusions are given in Section 5.

2 Related Work

Single Image Super-Resolution (SISR) is a classic problem in the computer vision community, most often for images from the visible spectrum. In this section, common thermal image datasets used as benchmarks by the community, together with the state of the art SISR approaches in the thermal image domain, are reviewed.

2.1 Benchmark Datasets

Visible spectrum HD images, for training SR networks and evaluating their performance, is not a problem due it large variety of datasets available in the literature (e.g., [Timofte et al., 2017], [Bevilacqua et al., 2012], [Zeyde et al., 2010], [Martin et al., 2001], [Arbel et al., 2011], [Huang et al., 2015], [Matsui et al., 2017],

among others). These HR images have been acquired in different scenarios covering a large set of objects' category (e.g., building, people, food, cars, among others) at different resolutions. On the contrary to the visible spectrum, in the thermal image domain, there are just a few datasets available in the literature, most of them in low resolution (e.g., [Davis and Keck, 2005], [Olmeda et al., 2013], [Hwang et al., 2015], among others); actually, thermal image datasets available in the literature have been designed for other specific applications (e.g., biometric domain, medical, security) but used to tackle the thermal image super-resolution problem. Up to our knowledge, [Wu et al., 2014] is the largest HR thermal image dataset available in the literature; this dataset consists of full-resolution 1024×1024 images, collected with a FLIR SC8000, containing 63782 frames; the main drawback of this dataset is that all the images come from the same scenario.

Trying to overcome the lack of datasets intended for thermal images SR task, in [Rivadeneira et al., 2019] a novel dataset is presented. It consists of 101 images acquired with a HR TAU2 FLIR camera, with a native resolution of 640×512 pixels of different scenarios (e.g. indoor, outdoor, day, night). A very large dataset (FLIR-ADAS) has been released by FLIR[3], it provides an annotated thermal images set for training and validation object detection neural networks. This dataset was acquired also with a TAU2 thermal camera but mounted on a vehicle. Provided images are with a resolution of 640×512. It contains a total of 14452 thermal images sampled from short videos taken on streets and highways. This dataset was intended for driving assistance applications, although can be used for the super-resolution problem.

Most of these datasets are not large enough to reach good results when heavy SR learning-based approaches are considered; furthermore, the datasets mentioned above contain images obtained from just one thermal camera. Having in mind all these limitations, recently [Rivadeneira et al., 2020a] presents a novel dataset that consists of a set of 1021 thermal images acquired with three different thermal cameras, which acquire images at different resolutions. The dataset contains images of outdoor scenarios with different daylight conditions (e.g., morning, afternoon, and night) and objects (e.g., buildings, cars, vegetation), mounting the cameras in a rig trying to minimize the baseline distance between the optical axis to get an almost registered image. This dataset has been used as a benchmark in the first thermal image super-resolution challenge organized on the workshop *Perception Beyond the Visible Spectrum* of CVPR2020 conference [Rivadeneira et al., 2020b].

2.2 Super-Resolution

The image SR is a classical issue, and still a challenging problem in the computer vision community and can be categorized as single-image SR (SISR) and multi-image SR (MISR), where SISR task is more challenging than MISR due to the lack of features that can be obtained in just one image rather than multiple

[3] FREE FLIR Thermal Dataset for Algorithm Training https://www.flir.in/oem/adas/adas-dataset-form/

images of the same scene. SISR has been studied in the literature for years and can be roughly classified as interpolation-based SR (conventional and traditional methods) and deep learning-based SR.

SRCNN [Dong et al., 2015] for the first time introduced deep learning in the SR field, showing the capability to improve the quality of SR results in comparison to traditional methods. Inspired in SRCNN, [Kim et al., 2016] proposes a VDSR network showing significant improvement. The authors propose to use more convolutional layers, increasing the depth of the network from 3 to 20 layers, and adopt global residual learning to predict the difference between generated image from the ground-truth (GT) image instead pixel-wise. FSRCNN [Dong et al., 2016] gets better computational performance by extracting the feature maps on a low-resolution image and just in the last layer upsampled it reducing the computational cost. Inspired by these works, different approaches to the image SR problem have been published using deeper networks using more convolutional layers with residual learning (e.g., [Kim et al., 2016], [Zhang et al., 2017]). Recently, several SR approaches using CNN (e.g., SRFeat-M [Park et al., 2018], EDSR [Lim et al., 2017], RCAN [Zhang et al., 2018]) have been proposed obtaining state-of-the-art performance for visible LR images. The CNNs mentioned above aim to minimize the difference between SR and GT images by using a supervised training process with a pair of images having a pixel-wise registration. In general, the strategy followed by these approaches is to down-sample the given HR image, add random noise or blur it, and then use it as the input LR image.

To overcome the limitation of having a pixel-wise registration between SR and GT images, unsupervised approaches have been proposed. For instance, [Shi et al., 2018] proposes a single image super-resolution approach, referred to as SRGAN, which achieves impressive state-of-the-art performance. This approach is inspired by the seminal Generative Adversarial Network (GAN) presented in [Goodfellow et al., 2014]. In recent literature, different unsupervised training processes have been presented for applications such as transferring style [Chang et al., 2018], image colorization [Mehri and Sappa, 2019], image enhancement [Chen et al., 2018], feature estimation [Suarez et al., 2019], among others. All these approaches are based on two-way GANs (CycleGAN) networks that can learn from unpaired data sets [Zhu et al., 2017]. CycleGAN can be used to learn how to map images from one domain (source domain) into another domain (target domain). This functionality makes CycleGAN models appropriate for image SR estimation when there is not a pixel-wise registration.

Most of the SR approaches mentioned above are focused on images from the visible spectrum. Based on SRCNN, [Choi et al., 2016] propose the first approach named Thermal Enhancement Network (TEN). Due to the lack of thermal image dataset, TEN uses RGB images for training. In [Rivadeneira et al., 2019], a dataset of 101 HR thermal images have been considered and, in conclusion, the authors state that better results are obtained if the network is trained using images from the same spectral band. Recently, [Mandanici et al., 2019] uses a concept of multi-image SR (MISR) for thermal imaging SR. As mentioned

above, in [Rivadeneira et al., 2020a] a novel dataset using three different camera resolutions has been proposed; this dataset is used to train a CycleGAN architecture that makes a transfer domain from a LR image (from one camera) to a HR image (of another camera), without pairing the images. Using this dataset as a reference, in [Rivadeneira et al., 2020b] two kinds of evaluations are proposed, the first evaluation consists of down-sampling a HR thermal images by $\times 2$, $\times 3$ and $\times 4$ and comparing their SR results with the corresponding GT images. The second evaluation consists in obtaining the $\times 2$ SR from a given MR thermal image and comparing it with its corresponding semi registered HR image. Results from this work are considered as baseline measures for future works in the thermal images super-resolution—MLVC-Lab [Chudasama et al., 2020] and Couger AI [Kansal and Nathan, 2020] corresponds to the approaches with the best results according to the mentioned evaluations.

3 Proposed Approach

In Section 3.1, details of the proposed architecture are given together with information about the proposed loss function. Additionally, in Section 3.2, the datasets used for training and validation are described. Finally, the process followed to evaluate the performance of the proposed approach is introduced.

3.1 Architecture

The proposed approach is based on the usage of Cycle Generative Adversarial Network (CycleGAN) [Zhu et al., 2017], widely used for map feature maps from one domain to another domain. In the current work, this framework is used to tackle the SR problem by mapping information from the mid-resolution (MR) to the high-resolution (HR) domain. As shown in Fig. 1, the proposed approach consists of two generators (MR to HR and HR to MR), with their corresponding discriminators (DISC MR and DISC HR) that validate the generated images. As generators, a ResNet with 6 residual blocks (ResNet-6) is considered. It uses optimization to avoid degradation in the training phase. The residual blocks have convolutional layers, with instant normalization and ReLu, and skip connections. As discriminators a patchGAN architecture is considered; the generated image and a non paired GT image are used to validate if the output is real or not.

Following the architecture presented in [Rivadeneira et al., 2020a], a combination of different loss functions is used: *i*) adversarial loss $\mathcal{L}_{Adversarial}$, *ii*) cycle loss $\mathcal{L}_{Cycle}$, *iii*) identity loss $\mathcal{L}_{Identity}$, and *iv*) structural similarity loss $\mathcal{L}_{SSIM}$; additionally, another loss term, Sobel loss $\mathcal{L}_{Sobel}$, is proposed. Sobel loss consist in apply Sobel filter edge detector [Kittler, 1983] to the input image and the cycled generated image, and get the mean square difference between both images, helping to evaluate the contour consistency between the two images. Details on each of these loss terms are given below—Fig. 1 illustrates these terms.

The **adversarial loss** is designed to minimize the cross-entropy to improve the texture loss:

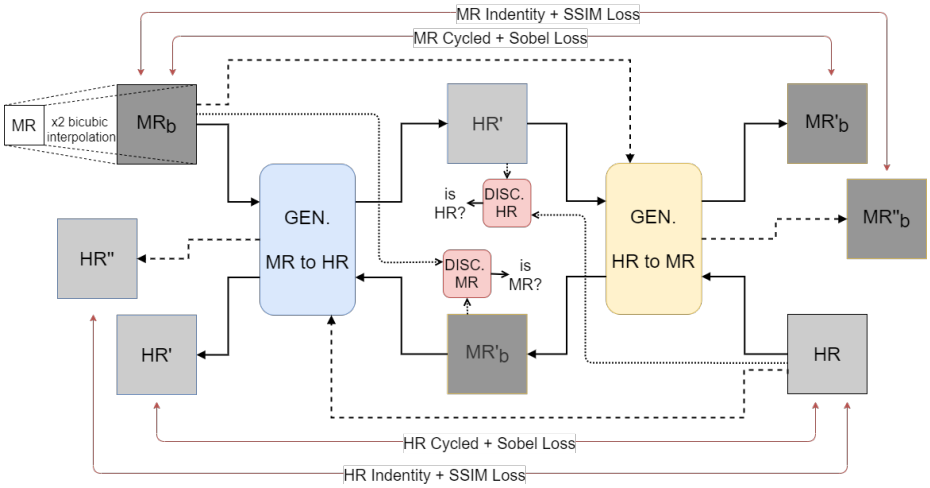

Fig. 1. CycleGAN architecture with 6 blocks ResNet for MR to HR generator and for HR to MR; with cycled + Sobel Loss and Identity + SSIM loss, and it respective discriminators,

$$\mathcal{L}_{Adversarial} = -\sum_{i} logD(G_{M2H}(I_M), I_H), \tag{1}$$

where D is the discriminator, $G_{M2H}(I_M)$ is the generated image, I_M and I_H are the low and high-resolution images respectively.

The **cycled loss** ($\mathcal{L}_{Cycled}$) is used to determinate the consistency between input and cycled output; it is defined as:

$$\mathcal{L}_{Cycled} = \frac{1}{N}\sum_{i} ||G_{H2M}(G_{M2H}(I_M)) - I_M||, \tag{2}$$

where G_{M2H} and G_{H2M} are the generators that go from one domain to the other domain.

The **identity loss** ($\mathcal{L}_{Identity}$) is used for maintaining the consistency between input and output; it is defined as:

$$\mathcal{L}_{Identity} = \frac{1}{N}\sum_{i} ||G_{H2M}(I_M) - I_M||, \tag{3}$$

where G is the generated image and I is the input image.

The **structural similarity loss** ($\mathcal{L}_{SSIM}$) for a pixel P is defined as:

$$\mathcal{L}_{SSIM} = \frac{1}{NM}\sum_{p=1}^{P} 1 - SSIM(p), \tag{4}$$

where $SSIM(p)$ is the Structural Similarity Index (see [Wang et al., 2004] for more details) centered in pixel p of the patch (P). The **Sobel loss** ($\mathcal{L}_{Sobel}$) is

used to determinate the edge consistency between input and cycled output; it is defined as:

$$\mathcal{L}_{Sobel} = \frac{1}{N}\sum_{i} ||Sobel(G_{H2M}(G_{M2H}(I_M))) - Sobel(I_M)||, \quad (5)$$

where G_{M2H} and G_{H2M} are the generators that go from one domain to the other domain and Sobel gets the contour of the images.

The **total loss function** ($\mathcal{L}_{total}$) used in this work is the weighted sum of the individual loss function terms:

$$\mathcal{L}_{total} = \lambda_1\mathcal{L}_{Adversarial} + \lambda_2\mathcal{L}_{Cycled} + \lambda_3\mathcal{L}_{Identity} + \lambda_4\mathcal{L}_{SSIM} + \lambda_5\mathcal{L}_{Sobel}, \quad (6)$$

where λ_i are weights empirically set for each loss function.

3.2 Datasets

The proposed approach is trained by using two datasets. The novel dataset from [Rivadeneira et al., 2020a] and the FLIR-ADAS mentioned in Section 2.1. For the first dataset, only mid-resolution (MR) and high-resolution (HR) images, acquired with two different cameras at different resolutions (mid and high resolution) are considered; each resolution set has 951 images and 50 images are left for testing. Figure 2 shows some illustrations of this dataset, just images from the mid-resolution and high-resolution are depicted.

For the second dataset, which contains 8862 training images, just one out of nine images have been selected, resulting in a sub-set of 985 images. This subsampling process has been applied in order to have more different scenarios since these images correspond to a video sequence, consecutive images are quite similar. Figure 3 shows some illustrations from this second dataset.

Both datasets have HR images with a native resolution of 640x512; these images have been cropped to 640x480 pixels, centered, to be exactly x2 of MR images; both datasets have 8bits and are saved in jpg format, so both have similar scenarios but acquired in different places and conditions.

Fig. 2. Examples of thermal images. (*top*) MR images from Axis Q2901-E. (*bottom*) HR images from FC-6320 FLIR [Rivadeneira et al., 2020a].

The main idea is to train the network with a shuffle mix of images between these two datasets, having the same proportion of images, sizes, and the same condition's scenarios. As mention in section 3.3, the validation is done with the same set of images used in the PBVS-CVPR2020 challenge [Rivadeneira et al., 2020b], to compare the results with the most recent results in the state-of-the-art literature.

Fig. 3. Examples of the *Free FLIR Thermal Dataset for Algorithm Training* (FLIR-ADAS).

3.3 Evaluation

The quantitative evaluations of the proposed method are performed as proposed in [Rivadeneira et al., 2020a] for MR to HR case; this evaluation has been adopted in the PBVS-CVPR2020 Challenge [Rivadeneira et al., 2020b], referred to as *evaluation2*, which consists in getting the average results of PSNR and SSIM measures on the generated SR of mid-resolution images and compared with the semi registered high-resolution image obtained from the other camera. This process is illustrated in Fig. 4. Just a centered region containing 80% of the image is considered in order to use these measures. For a fair comparison, the images from the validation set are the same ones used in the previous works mentioned above. Results from this *evaluation2* are compared with those presented in [Rivadeneira et al., 2020a].

4 Experimental Results

This section presents the results obtained with the unsupervised SThISR architecture proposed in this work. Section 4.1 describes the settings used for training the proposed approach, while Section 4.2 presents the quantitative results.

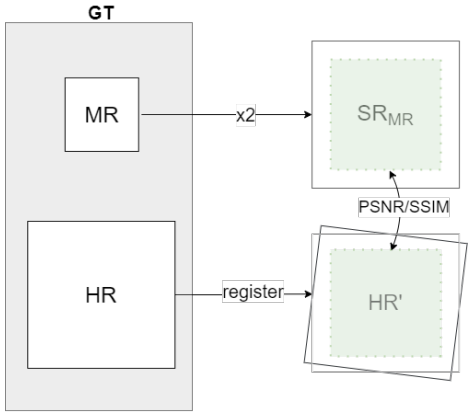

Fig. 4. PBVS-CVPR2020 Challenge evaluation2 approach [Rivadeneira et al., 2020b] .

4.1 Settings

The proposed approach is trained on a NVIDIA Geforce GTX mounted in a workstation with 128GB of RAM, using Python programming language, Tensorflow 2.0, and Keras library. Only the two datasets mentioned in 2.1 are considered, no data-augmentation process has been applied to the given input data.

Images are up-sampled by bicubic interpolation, due to the CycleGAN transfer domain (from mid to high resolution) needs images at the same resolution. Images are normalized in a [-1,1] range. The network was trained for 100 epochs without dropout since the model does not present overfit. The generator is a ResNet with 6 residual blocks (ResNet-6) using Stochastic AdamOptimizer to prevent over fittings and lead to faster convergent and avoiding degradation in the training phase. The discriminators use a patchGAN architecture, and it validates if the generated image together with the GT images is real or not. For each epoch in the training phase, input mages were shuffle random mix. The hyper-parameters used were 0.0002 for learning rate for both the generator and the discriminator networks; epsilon=1e-05; exponential decay rate for the 1^{st} moment momentum 0.5 for the discriminator and 0.4 for the generator. For λ_i values weights in losses, the set of values that obtain the best results were: for $\mathcal{L}_{Cycled}$=10, $\mathcal{L}_{Identity}$=5, $\mathcal{L}_{SSIM}$=5 and $\mathcal{L}_{Sobel}$=10. The proposed architecture has been trained twice, one with just the first dataset and once with both datasets.

4.2 Results

The quantitative results obtained for each training, together with previous works and other approaches from the PBVS-CVPR2020 Challenge are shown in Table 1 using PSNR and SSIM measures comparison. The best results are highlighted

in bold and the second-best result is underline. As can be appreciated, the proposed SThISR approach achieves better results than other works. Between both current work results, using just one dataset gets seven-tenths better PSNR results rather than using both datasets. SSIM measure gets higher result using the two datasets buts just by one-thousandth. These results show that using just the first dataset the proposed approach archive better results, meaning that this dataset is varied enough to train a network and that it is possible to do a Single Thermal Image Super-Resolution between two different domains using images acquired with different camera resolutions and without registration.

Approachs'	**PSNR**	**SSIM**
Bicubic Interpolation	20,24	0,7515
[Rivadeneira et al., 2020a]	22,42	0,7989
MLVC-Lab^{+}	20,02	0,7452
COUGER AI^{+}	20,36	0,7595
Current Work^{1}	**22,98**	<u>0,8032</u>
Current Work^{2}	<u>22,27</u>	**0,8045**

Table 1. Quantitative average results of evaluation detail in Section 3.3. $^{+}$Winner approaches at the PBVS-CVPR2020 Challenge. Work^{1} using just first dataset; Work^{2} using both datasets. Bold and underline values correspond to the first and second best results respectively.

Regarding the quality of the obtained results, Fig. 5 shows the worst and best super-resolution results from the validation set. The worst result gets 20.11/0.6464 PSNR and SSIM measures respectively; it should be mentioned that although it is the worst result from the whole validation set, it is considerably better than the results obtained with a bicubic interpolation: 17.36/0.6193 PSNR and SSIM respectively. In the case of the best result, a 26.06/0.8651 PSNR and SSIM measures respectively are obtained, in this case, the bicubic interpolation reaches 19.41/0.8021 PSNR and SSIM respectively. In conclusion, it could be stated that the most challenging scenarios are those with objects at different depth and complex textures, on the contrary, it can be appreciated that scenes with planar surfaces are more simple to obtain their corresponding super-resolution representation.

5 Conclusions

This paper presents an extended version of the work presented at VISAPP 2020. Two datasets are considered during the training stage, and adjusting different hyper-parameters values on loss function in CycleGAN and adding a Sobel loss. The proposed approach has shown an improvement on previous work and

Fig. 5. Examples quality results. (*top*) from left to right, MR image, HR image, worst results. (*bottom*) best results.

achieved better results on state-of-the-art values comparing to the results from the first challenge on SR thermal images in terms of PSNR and SSIM measures. It should be mentioned that the proposed SThISR architecture is trained using an unpaired set of images. The first dataset has large variability, showing that its not necessary the use of other datasets.

ACKNOWLEDGEMENTS

This work has been partially supported by the Spanish Government under Project TIN2017-89723-P; and the "CERCA Programme / Generalitat de Catalunya". The first author has been supported by Ecuador government under a SENESCYT scholarship contract.

References

[Arbel et al., 2011] Arbel, P., Maire, M., Fowlkes, C., and Malik, J. (2011). Contour detection and hierarchical image segmentation. *IEEE Trans. Pattern Anal. Mach. Intell*, 33(5):898–916.

[Bevilacqua et al., 2012] Bevilacqua, M., Roumy, A., Guillemot, C., and Alberi-Morel, M. L. (2012). Low-complexity single-image super-resolution based on nonnegative neighbor embedding.

[Chang et al., 2018] Chang, H., Lu, J., Yu, F., and Finkelstein, A. (2018). Pairedcyclegan: Asymmetric style transfer for applying and removing makeup. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 40–48.

[Chen et al., 2018] Chen, Y.-S., Wang, Y.-C., Kao, M.-H., and Chuang, Y.-Y. (2018). Deep photo enhancer: Unpaired learning for image enhancement from photographs

with gans. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 6306–6314.

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

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

[Davis and Keck, 2005] Davis, J. W. and Keck, M. A. (2005). A two-stage template approach to person detection in thermal imagery. In *2005 Seventh IEEE Workshops on Applications of Computer Vision (WACV/MOTION'05)-Volume 1*, volume 1, pages 364–369. IEEE.

[Ding et al., 2019] Ding, M., Zhang, X., Chen, W.-H., Wei, L., and Cao, Y.-F. (2019). Thermal infrared pedestrian tracking via fusion of features in driving assistance system of intelligent vehicles. *Proceedings of the Institution of Mechanical Engineers, Part G: Journal of Aerospace Engineering*, 233(16):6089–6103.

[Dong et al., 2015] Dong, C., Loy, C. C., He, K., and Tang, X. (2015). Image super-resolution using deep convolutional networks. *IEEE transactions on pattern analysis and machine intelligence*, 38(2):295–307.

[Dong et al., 2016] Dong, C., Loy, C. C., and Tang, X. (2016). Accelerating the super-resolution convolutional neural network. In *European conference on computer vision*, pages 391–407. Springer.

[Gade and Moeslund, 2014] Gade, R. and Moeslund, T. B. (2014). Thermal cameras and applications: a survey. *Machine vision and applications*, 25(1):245–262.

[Girshick et al., 2015] Girshick, R., Donahue, J., Darrell, T., and Malik, J. (2015). Region-based convolutional networks for accurate object detection and segmentation. *IEEE transactions on pattern analysis and machine intelligence*, 38(1):142–158.

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

[Herrmann et al., 2018] Herrmann, C., Ruf, M., and Beyerer, J. (2018). Cnn-based thermal infrared person detection by domain adaptation. In *Autonomous Systems: Sensors, Vehicles, Security, and the Internet of Everything*, volume 10643, page 1064308. International Society for Optics and Photonics.

[Huang et al., 2015] Huang, J.-B., Singh, A., and Ahuja, N. (2015). Single image super-resolution from transformed self-exemplars. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 5197–5206.

[Hwang et al., 2015] Hwang, S., Park, J., Kim, N., Choi, Y., and So Kweon, I. (2015). Multispectral pedestrian detection: Benchmark dataset and baseline. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 1037–1045.

[Kansal and Nathan, 2020] Kansal, P. and Nathan, S. (2020). A multi-level supervision model: A novel approach for thermal image super resolution. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops*, pages 94–95.

[Kim et al., 2016] Kim, J., Kwon Lee, J., and Mu Lee, K. (2016). Accurate image super-resolution using very deep convolutional networks. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 1646–1654.

[Kittler, 1983] Kittler, J. (1983). On the accuracy of the sobel edge detector. *Image and Vision Computing*, 1(1):37–42.

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

[Mandanici et al., 2019] Mandanici, E., Tavasci, L., Corsini, F., and Gandolfi, S. (2019). A multi-image super-resolution algorithm applied to thermal imagery. *Applied Geomatics*, 11(3):215–228.

[Martin et al., 2001] Martin, D., Fowlkes, C., Tal, D., Malik, J., et al. (2001). A database of human segmented natural images and its application to evaluating segmentation algorithms and measuring ecological statistics. Iccv Vancouver:.

[Matsui et al., 2017] Matsui, Y., Ito, K., Aramaki, Y., Fujimoto, A., Ogawa, T., Yamasaki, T., and Aizawa, K. (2017). Sketch-based manga retrieval using manga109 dataset. *Multimedia Tools and Applications*, 76(20):21811–21838.

[Mehri and Sappa, 2019] Mehri, A. and Sappa, A. D. (2019). Colorizing near infrared images through a cyclic adversarial approach of unpaired samples. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops*, pages 0–0.

[Mudunuri and Biswas, 2015] Mudunuri, S. P. and Biswas, S. (2015). Low resolution face recognition across variations in pose and illumination. *IEEE transactions on pattern analysis and machine intelligence*, 38(5):1034–1040.

[Olmeda et al., 2013] Olmeda, D., Premebida, C., Nunes, U., Armingol, J. M., and de la Escalera, A. (2013). Pedestrian detection in far infrared images. *Integrated Computer-Aided Engineering*, 20(4):347–360.

[Park et al., 2018] Park, S.-J., Son, H., Cho, S., Hong, K.-S., and Lee, S. (2018). Srfeat: Single image super-resolution with feature discrimination. In *Proceedings of the European Conference on Computer Vision (ECCV)*, pages 439–455.

[Rivadeneira et al., 2020a] Rivadeneira, R. E., Sappa, A. D., and Vintimilla, B. X. (2020a). Thermal image super-resolution: A novel architecture and dataset. In *VISIGRAPP (4: VISAPP)*, pages 111–119.

[Rivadeneira et al., 2020b] Rivadeneira, R. E., Sappa, A. D., Vintimilla, B. X., Guo, L., Hou, J., Mehri, A., Behjati Ardakani, P., Patel, H., Chudasama, V., Prajapati, K., et al. (2020b). Thermal image super-resolution challenge-pbvs 2020. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops*, pages 96–97.

[Rivadeneira et al., 2019] Rivadeneira, R. E., Suárez, P. L., Sappa, A. D., and Vintimilla, B. X. (2019). Thermal image superresolution through deep convolutional neural network. In *International Conference on Image Analysis and Recognition*, pages 417–426. Springer.

[Shamsolmoali et al., 2019] Shamsolmoali, P., Zareapoor, M., Jain, D. K., Jain, V. K., and Yang, J. (2019). Deep convolution network for surveillance records super-resolution. *Multimedia Tools and Applications*, 78(17):23815–23829.

[Shi et al., 2018] Shi, W., Ledig, C., Wang, Z., Theis, L., and Huszar, F. (2018). Super resolution using a generative adversarial network. US Patent App. 15/706,428.

[Suarez et al., 2019] Suarez, P. L., Sappa, A. D., Vintimilla, B. X., and Hammoud, R. I. (2019). Image vegetation index through a cycle generative adversarial network. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops*, pages 0–0.

[Timofte et al., 2017] Timofte, R., Agustsson, E., Van Gool, L., Yang, M.-H., and Zhang, L. (2017). Ntire 2017 challenge on single image super-resolution: Methods and results. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops*, pages 114–125.

[Wang et al., 2004] Wang, Z., Bovik, A. C., Sheikh, H. R., Simoncelli, E. P., et al. (2004). Image quality assessment: from error visibility to structural similarity. *IEEE transactions on image processing*, 13(4):600–612.

[Wu et al., 2014] Wu, Z., Fuller, N., Theriault, D., and Betke, M. (2014). A thermal infrared video benchmark for visual analysis. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops*, pages 201–208.

[Zeyde et al., 2010] Zeyde, R., Elad, M., and Protter, M. (2010). On single image scale-up using sparse-representations. In *International conference on curves and surfaces*, pages 711–730. Springer.

[Zhang et al., 2017] Zhang, K., Zuo, W., Gu, S., and Zhang, L. (2017). Learning deep cnn denoiser prior for image restoration. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 3929–3938.

[Zhang et al., 2010] Zhang, L., Zhang, H., Shen, H., and Li, P. (2010). A super-resolution reconstruction algorithm for surveillance images. *Signal Processing*, 90(3):848–859.

[Zhang et al., 2018] Zhang, Y., Li, K., Li, K., Wang, L., Zhong, B., and Fu, Y. (2018). Image super-resolution using very deep residual channel attention networks. In *Proceedings of the European Conference on Computer Vision (ECCV)*, pages 286–301.

[Zhu et al., 2017] Zhu, J.-Y., Park, T., Isola, P., and Efros, A. A. (2017). Unpaired image-to-image translation using cycle-consistent adversarial networks. In *Proceedings of the IEEE international conference on computer vision*, pages 2223–2232.