Thermal Image Super-resolution: A Novel Architecture and Dataset

Rafael E. Rivadeneira[1] [a], Angel D. Sappa[1,2] [b] and Boris X. Vintimilla[1] [c]

[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
[2]*Computer Vision Center, Edifici O, Campus UAB, 08193 Bellaterra, Barcelona, Spain*

Keywords: Thermal Images, Far Infrared, Dataset, Super-resolution.

Abstract: This paper proposes a novel CycleGAN architecture for thermal image super-resolution, together with a large dataset consisting of thermal images at different resolutions. The dataset has been acquired using three thermal cameras at different resolutions, which acquire images from the same scenario at the same time. The thermal cameras are mounted in a rig trying to minimize the baseline distance to make easier the registration problem. The proposed architecture is based on ResNet6 as a Generator and PatchGAN as a Discriminator. The novelty on the proposed unsupervised super-resolution training (CycleGAN) is possible due to the existence of afore-mentioned thermal images—images of the same scenario with different resolutions. The proposed approach is evaluated in the dataset and compared with classical bicubic interpolation. The dataset and the network are available.

1 INTRODUCTION

Image Super-resolution (SR) is an ill-posed problem that refers to the estimation of high-resolution (HR) image/video from a low-resolution (LR) one of the same scene, usually with the use of digital image processing and Machine Learning (ML) techniques; SR has important applications in a wide range of domains, such as surveillance and security (e.g., (Zhang et al., 2010), (Rasti et al., 2016), (Shamsolmoali et al., 2019)), medical imaging (e.g., (Mudunuri and Biswas, 2015), (Robinson et al., 2017),(Huang et al., 2019)), object detection in scene (e.g., (Girshick et al., 2015)), among others.

In recent years, the development of deep learning techniques, have witnessed remarkable progress achieving the performance on various benchmarks of SR, where most of the state-of-the-art are focused on the visible domain. Thermal images, which are in the far-infrared (FIR) spectral band of the electromagnetic spectrum, have become an important tool in several fields (e.g., (Qi and Diakides, 2003), (Herrmann et al., 2018)); unfortunately, due to physical limitations of the technology and the high cost of thermal cameras, these images tend to have a poor resolu-

tion. This poor resolution could be improved by using learning-based super-resolution methods, like those used in the visible spectral domain.

Learning-based super-resolution methods generally work by down-sampling and adding both noise and blur to the given image. These noisy and blurred poor quality images, together with the given images that are considered as the Ground Truths (GT), are used in the learning process. The approaches mention before have been mostly used to tackle the super-resolution problem, however, there are few contributions where the learning process is based on the usage of pair of images (low and high-resolution images) obtained from different cameras. In the current work, a novel learning-based super-resolution approach is proposed to improve the resolution of the given thermal images. Additionally, a large thermal image dataset is acquired[1], containing images with three different resolutions (low, mid, and high) obtained with three different thermal cameras. The manuscript is organized as follows. Section 2 presents works related to the topics tackled in the current work. The proposed dataset is detailed in Section 3. Results are provided in Section 4. Finally, conclusions are given in Section 5.

[a] https://orcid.org/0000-0002-5327-2048
[b] https://orcid.org/0000-0003-2468-0031
[c] https://orcid.org/0000-0001-8904-0209

---

[1]The dataset is available at http://www.cidis.espol.edu.ec/es/dataset

2 RELATED WORK

This section presents a summary of the most recent and relevant contributions to the topics tackled in the current work. Section 2.1 summarizes the state-of-art on image super-resolution, mainly approaches proposed for images from the visible spectrum. Then, Section 2.2 describes recent thermal image datasets freely available in the literature.

2.1 Image Super-resolution

The single image super-resolution (SISR) has been extensively studied in the literature for decades, recently using deep learning techniques better results, with respect to conventional methods, are obtained. The convolutional neural networks (CNNs) have shown a great capability to improve the quality of SR results. (Dong et al., 2015) firstly propose a SRCNN to learn an end-to-end mapping, between the interpolated LR images and their HR counterparts, archiving the state-of-the-art performance. For better performance, FSRCNN (Dong et al., 2016) extracts feature maps at the low-resolution image and up-sample the image at the last layer. Inspired in SRCNN, depth networks start to appear stacking more convolutional layers with residual learning (e.g., (Kim et al., 2016), (Zhang et al., 2017)). In order to speed up the training process the authors in (Lim et al., 2017) propose EDSR, which removes the batch-normalization layer and takes advantage of residual learning (He et al., 2016). Most of the aforementioned CNNs aim at minimizing the mean-square error (MSE) between SR and GT images, tending to suppress the high-frequency details in images. In other words, a supervised training process, using a pair of images, is followed. The main limitation of such approaches lies in the need of having pixel-wise registered SR and GT images to compute the MSE. As mentioned above, in most cases the SR image is obtained from an image down-sampled from the GT.

Different unsupervised training processes have been recently presented in the literature 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 are able to learn from unpaired data sets (Zhu et al., 2017). CycleGAN learns to map images from one domain (source domain) onto another domain (target domain); this functionality makes CycleGAN model appropriate for image SR estimation when there is not a pixel-wise registration.

2.2 Datasets

There is a large variety of datasets available for visible spectrum image super-resolution, recently (Timofte et al., 2017) has released a high-quality (2K resolution) dataset *DIV2K* for visible image restoration, which is split up into 800 images for training, 100 for testing and 100 for validation. Most of the approaches in the literature use common benchmark datasets for evaluating their performance, all of them in visible spectrum domain (e.g., Set5 (Bevilacqua et al., 2012), Set14 (Zeyde et al., 2010), BSD300 (Martin et al., 2001), BSD500 (Arbel et al., 2011), Urban100 (Huang et al., 2015), Manga109 (Matsui et al., 2017), among others). These datasets provide HR images under different categories (e.g., animal, building, food, landscape, people, flora, fauna, car, amongst others) with differ resolutions and amount of images. Some of them even include LR and HR image pairs.

During last years a few thermal datasets have been published; in (Davis and Keck, 2005) a dataset consisting of 284 thermal images, with a resolution of 360×240, is presented. This dataset is acquired with a Raytheon 300D, in a University campus at a walkway and street intersection. It has been generated by capturing images over several daytimes and weather conditions. In (Olmeda et al., 2013) a thermal image dataset, consisting of 15224 images with a resolution 164×129 has been proposed. In this dataset, the images have been acquired with an Indigo Omega imager mounted on vehicle driven in outdoors urban scenarios. In (Hwang et al., 2015) a FLIR-A35 is used to acquire more than 41500 thermal images with a resolution of 320×256. A HR dataset was presented in (Wu et al., 2014); the dataset contains seven different scenes, most of them collected with a FLIR SC8000, with a full resolution of 1024×1024. The dataset consists of 63782 frames with thousands of recording objects; as far as we know, it is the dataset with the largest amount of HR thermal images available in the literature.

Most of the thermal image datasets mentioned above are usually designed for object detection and tracking; some others for applications on the biometric domain or medical applications; and just a few of them are intended for super-resolution tasks. As can be appreciated, most of the datasets contain low-resolution images of the scene. Trying to overcome the limitations mentioned above we have recently presented a small thermal image dataset, which was intended for training super-resolution learning-based approaches (Rivadeneira et al., 2019); it contains a total of 101 images taken with a single HR TAU2 ca-

Figure 1: Example of thermal images acquired by each camera. (*left*) LR image with 160x120 native resolution from Axis Domo P1290. (*middle*) MR image with 320x240 native resolution from Axis Q2901-E. (*right*) HR image with 640x480 resolution from FC-6320 FLIR (native resolution 640x512).
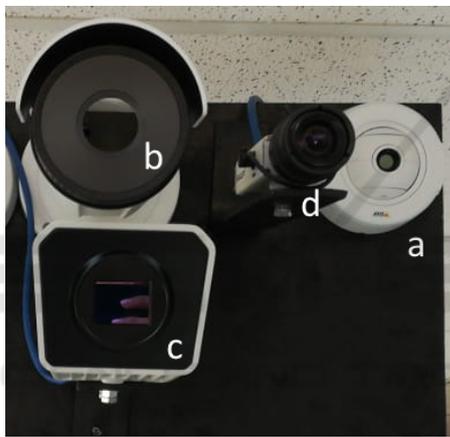

Figure 2: Panel with the cameras: a) Axis Domo P1290 (LR); b) Axis Q2901-E (MR); c) FC-6320 FLIR (HR); d) Basler visible spectrum camera, which is not used in the current work.

mera, having images with a native resolution of 640×512 pixels. Unfortunately, the amount of images in this dataset is not large enough to reach good results when heavy SR learning-based approaches are considered. It should be mentioned that all thermal datasets mentioned above contain images obtained from just one single thermal camera.

3 PROPOSED APPROACH

This section presents the approach proposed in the current work for thermal image SR. Section 3.1 describes the dataset collected in outdoor scenarios with three different thermal cameras; then Section 3.2 presents the architecture proposed for the unsupervised super-resolution.

Figure 3: Mosaic with three different resolution thermal images from each camera for a visual comparison: (*left*) crop from a LR image; (*middle*) crop from a MR image; (*right*) crop from a HR image.

3.1 Data Collection

A challenging dataset has been created using three different thermal cameras. The cameras were physically mounted in a panel as shown in Fig. 2; this structure has been placed in a vehicle to capture different outdoor scenarios. Images were acquired at the same time using a multi-thread develop script. Each camera has a different resolution (low, mid, high); Fig. 3 shows a mosaic build up with the three different resolutions. The cameras have been mounted trying to minimize the baseline distance between the optical axis so that the acquired images are almost registered. The technical information of the cameras is shown in Table 1 and a set of images obtained from these cameras is depicted in Fig. 1.

In spite of the effort during the camera setup, the obtained images capture slightly different regions from the scene (see Fig. 1); these differences come from the camera baselines as well as from the differences on the camera intrinsic parameters. Having in mind these limitations a set of 10 images per each resolution (LR, MR, HR) is selected and registered with

Table 1: Thermal camera specifications (*the HR images have been crop to 640×480).

Image Description	Brand Camera	FOV	Focal Length	Native Resolution
Low (LR)	Axis Domo P1290	35.4	4mm	160×120
Mid (MR)	Axis Q2901-E	35	9mm	320×240
High (HR)	FC-632O FLIR	32	19mm	640×512*

Figure 4: Image registration results. (*top*) From left to right: LR image; MR image; and image resulting from the registration of MR image with the results of SR$_{LR}$ image. (*bottom*) From left to right: MR image; HR image; and image resulting from the registration of HR image with the results of SR$_{MR}$ image.
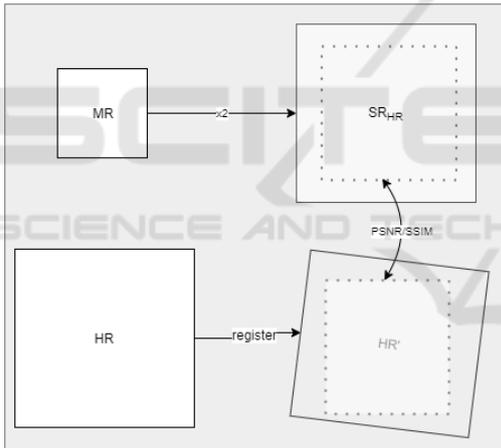

Figure 5: Evaluation diagram between the obtained SR from MR with registered HR (the same process is applied to evaluate SR results from LR, in other words SR$_{LR}$ with MR).

the corresponding SR result. In other words, from a given scene the registration between (SR$_{LR}$, MR) and (SR$_{MR}$, HR) is performed (Fig. 4 presents an illustration of this process). This registration process is performed to evaluate SR results using pixel-wise metrics; actually, to avoid non-overlapped regions (see Fig. 4), just a centered region containing 80% of the image is considered (see illustration in Fig. 5).

The pixel-wise metrics mentioned above to evaluate the results are: *i*) Peak Signal-to-Noise Ratio (PSNR), which is commonly used to measure the reconstruction quality of lossy transformations; and

ii) Structural Similarity Index Metric (SSIM) (Wang et al., 2004), which is based on the independent comparisons of luminance, contrast, and structure. Even the domain of these images are thermal, they are represented like grayscale images, so these metrics can be also used. Image quality assessments (IQA), focused on the perception of human viewers, are avoided in the current work due to they are expensive and time-consuming; furthermore, they are not necessarily consistent in the case of thermal images.

In the current work the evaluations will be performed just for ×2 scale. The average evaluation value will be computed as follows:

$$R_{\times 2} = \frac{1}{N} \sum_1^N eval\left(GT_N, SR_N^{\times 2}\right) \qquad (1)$$

where *eval* corresponds to PSNR or SSIM measures, and *N* is the number of validation images (in the current work 10 images have been considered). As mentioned above that *eval* is computed just on a region of the image to avoid non-overlapped areas (see Fig. 4), which appear due to the bias of the cameras and differences on the intrinsic camera parameters.

3.2 Proposed Architecture

This section presents details of the unsupervised learning approach proposed for estimating SR images. As mentioned above, the current work tackles the SR problem by using images from different cameras, which have been acquired at different resolutions. The proposed approach is based on the usage of a Cycle Generative Adversarial Network (Cycle-GAN) (Zhu et al., 2017), which is able to map information from one domain (low-resolution image) to another domain (high-resolution image). Figure 6 presents an illustration of a CycleGAN architecture used in the current work. It consists of two generators (G$_{L-H}$ and G$_{H-L}$) and two discriminators (D$_H$ and D$_L$). In the generators ResNet with 6 residual blocks (ResNet-6), to avoid degradation of the optimization during the training process, are considered. Each residual block is (Conv -> InstaNorm -> Conv -> InstaNorm -> Relu), with skip connections. Regarding the discriminators, a PatchGAN based architecture is considered. Each block in the discriminator is (Conv -> Conv -> InstaNorm -> Conv -> InstaNorm -> Conv -> LeakyReLU), the shape of the
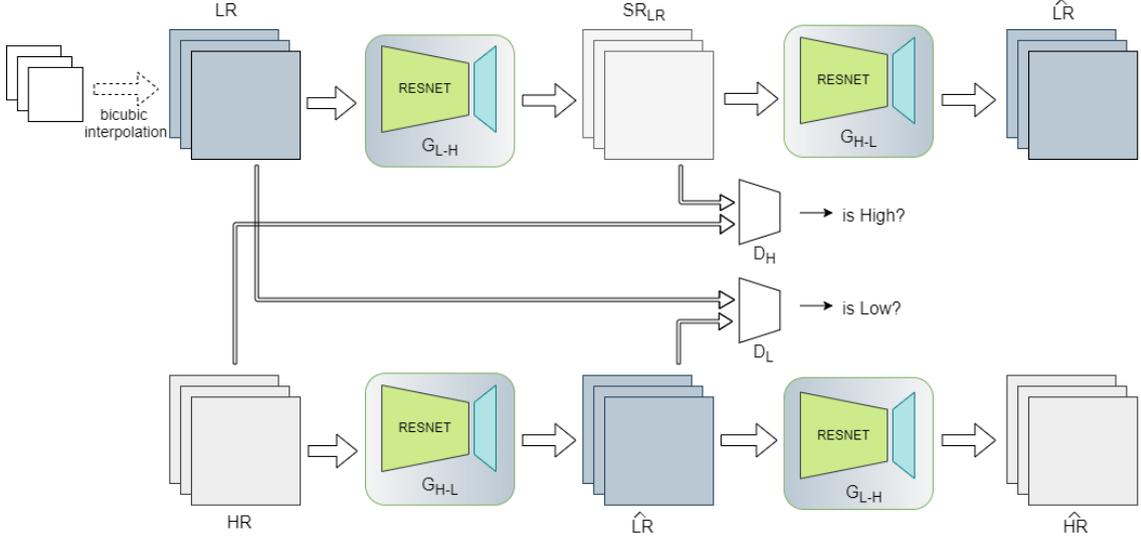
Figure 6: Proposed CycleGAN architecture using 6 residual blocks ResNet; G_{L-H} and G_{H-L} represent generators from lower to higher and from higher to lower resolution respectively. D_H and D_L represent the discriminator for each resolution.

output after last layer is (batch_size, 30, 30, 1), each 30×30 of the output classifies a 70×70 portion of the input images. The discriminator receives two inputs, the target image (classified as a real image) and the generated image (which should be discriminated as real of fake image by the discriminator).

The proposed architecture uses a combination of different loss functions: adversarial loss $\mathcal{L}_{Adversarial}$, cycle loss $\mathcal{L}_{Cycle}$, identity loss $\mathcal{L}_{Identity}$, and structural similarity loss $\mathcal{L}_{SSIM}$, which are detailed next.

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

$$\mathcal{L}_{Adversarial} = -\sum_i log D(G_{L-H}(I_L), I_H), \quad (2)$$

where D is the discriminator, $G_{L-H}(I_L)$ is the generated image, I_L and I_H are the low and high image respectively.

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

$$\mathcal{L}_{Cycled} = \frac{1}{N} \sum_i ||G_{H-L}(G_{L-H}(I_L)) - I_L||, \quad (3)$$

where G_{L-H} and G_{H-L} 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:

$$\mathcal{L}_{Identity} = \frac{1}{N} \sum_i ||G_{L-H}(I_L) - I_H||, \quad (4)$$

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:

$$\mathcal{L}_{SSIM} = \frac{1}{NM} \sum_{p=1}^{P} 1 - SSIM(p), \quad (5)$$

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

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

where λ_i for $i = 1, 2, 3, 4$ are weights for each loss, which have been empirically set.

4 EXPERIMENTAL RESULTS

This section presents the results of the proposed unsupervised learning SR approach. Section 4.1 presents information of the dataset acquired with the platform presented in Section 3.1; then Section 4.2 depicts the parameters used for training the CycleGAN architecture. Finally, Section 4.3 shows the quantitative and qualitative results obtained with the proposed approach.

4.1 Dataset

By using the thermal cameras mounted in the panel shown in Fig. 2 a set of 1021 images per camera was

Figure 7: Examples of thermal images used for training. (*top*) LR images from Axis Domo P1290. (*middle*) MR images from Axis Q2901-E. (*bottom*) HR images from FC-6320 FLIR.

generated. The images correspond to outdoor scenarios containing different objects (e.g., buildings, cars, people, vegetation); the images have been acquired at different daytime, including morning, afternoon and night. Figure 7 shows some examples of the type of images contained in the acquired dataset. The dataset is split up into 951 images for training and 50 for testing; the remaining 20 images are used for validation. All images are saved in one channel in jpg format without compression. The set of images for validation has been registered; half of this set has been used for registering MR images with the corresponding SR_{LR} and the other half for registering SR images with the corresponding SR_{MR}. As detailed in Section 3.1 this registration process is performed for evaluating results at a pixel-wise.

4.2 Training

The proposed CycleGAN model has been implemented with Tensorflow 2.0 using Keras library. The experiments were performed on a workstation with NVIDIA Geforce GTX and 128GB of RAM. The aforementioned set of training images (951 images) has been used without data augmentation; a set of 50 images has been used for testing. The given images have been normalized in a (-1,1) range. The network was trained for 50 epochs, from lower to higher resolution, LR images were firstly up-sampled by bicubic interpolation, due to the fact that input and output images in CycleGAN have to be in the same resolution. No dropout techniques were used since the model does not overfit within 50 epochs. The CycleGAN architecture has been trained using Stochastic AdamOptimizer since it prevents overfit-

ting and leads to convergence faster. The following hyper-parameters were used: learning rate 0.0002 for both the generator and the discriminator networks; epsilon=1e-05; exponential decay rate for the 1st moment momentum 0.5 for discriminator and 0.4 for the generator. The CycleGAN architecture has been trained twice, one for SR_{LR} and once for SR_{MR}

4.3 Results

Quantitative results obtained with the proposed CycleGAN architecture, for both validation, from LR to MR and from MR to HR, are shown in Table 2 and Table 3 respectively, together with the results obtained with (Rivadeneira et al., 2019) and with the bicubic interpolation, which is used as a baseline. As can be appreciated, the proposed architecture achieves a better performance than both, the previous publication and the bicubic interpolation.

Qualitative comparisons, for both evaluations, are depicted in Fig. 8 and Fig. 9. The results have shown that using this architecture to go from a lower resolution to a higher resolution is available, even though the network is trained with images from different cameras where there is not a perfect registration.

Table 2: Results on LR set in a ×2 scale factor, compared with its MR registered validation set.

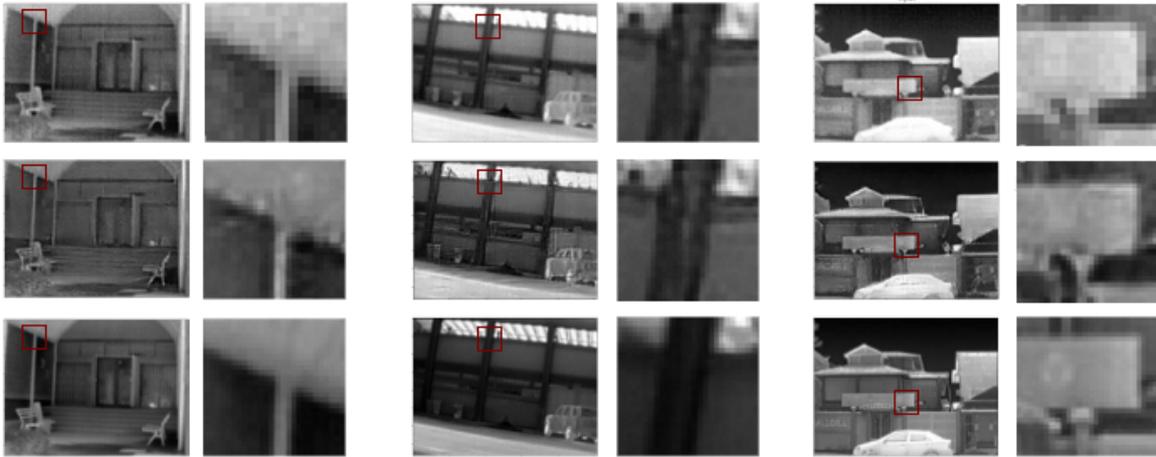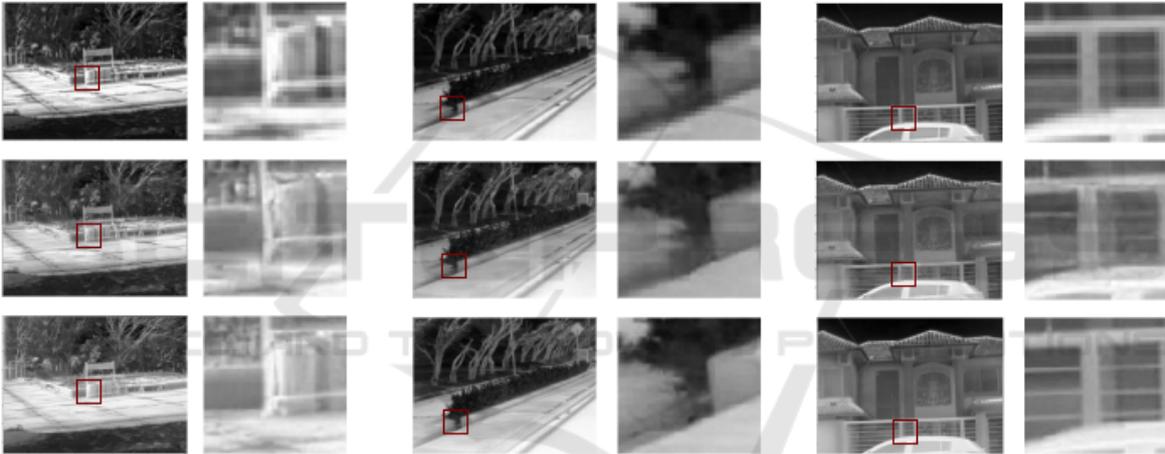| Method	PSNR	SSIM
Bicubic Interpolation	16.46	0.6695
(Rivadeneira et al., 2019)	17.01	0.6704
Current work	21.50	0.7218

Figure 8: SR results on real-world LR images with a ×2 scale factor—these illustrations correspond to the 80% centered area cropped from the images. ($top - row$) Bicubic interpolation image, ($middle - row$) Super-resolution results (SR_{LR}), ($bottom - row$) Ground truth MR image.

Figure 9: SR results on real-world MR images with a ×2 scale factor—these illustrations correspond to the 80% centered area cropped from the images. ($top - row$) Bicubic interpolation image, ($middle - row$) Super-resolution results (SR_{MR}), ($bottom - row$) Ground truth HR image.

Table 3: Results on MR set in a ×2 scale factor, compared with its HR registered validation set.

Method	PSNR	SSIM
Bicubic Interpolation	20.17	0.7553
(Rivadeneira et al., 2019)	20.24	0.7492
Current work	22.42	0.7989

5 CONCLUSIONS

This paper presents a novel CycleGAN based approach for thermal image super-resolution. ResNets with 6 residual blocks are used as the generators, while a PatchGAN based architectures are used as discriminators. The CycleGAN architecture is trained in an unsupervised way using non-paired sets of im-

ages. The network is trained to perform SR at a ×2 scale. It has been trained in two scenarios, to generate mid-resolution images from low-resolution and to generate high-resolution from mid-resolution. Comparisons with a previous approach and with the bicubic interpolation, which is used as a baseline, show the validity of the proposed approach. In order to train the proposed architecture a novel dataset containing images from three different thermal cameras, at different resolutions, has been generated. The proposed CycleGAN based approach allows to tackle the SR problem in a more challenging and realistic scenario than classical super-resolution approaches where the given image is down-sampled to train the network in a supervised way. Future work will be focused on evaluating other loss functions trying to improve the ob-

tained results as well as testing different architectures in the Generator and Discriminator.

ACKNOWLEDGEMENTS

REFERENCES

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

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

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.

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.

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, 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.

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.

He, K., Zhang, X., Ren, S., and Sun, J. (2016). Deep residual learning for image recognition. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 770–778.

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, 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.

Huang, Y., Shao, L., and Frangi, A. F. (2019). Simultaneous super-resolution and cross-modality synthesis in magnetic resonance imaging. In *Deep Learning and Convolutional Neural Networks for Medical Imaging and Clinical Informatics*, pages 437–457. Springer.

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.

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.

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.

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, 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, 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, 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, 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.

Qi, H. and Diakides, N. A. (2003). Thermal infrared imaging in early breast cancer detection-a survey of

recent research. In *Proceedings of the 25th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (IEEE Cat. No. 03CH37439)*, volume 2, pages 1109–1112. IEEE.

Rasti, P., Uiboupin, T., Escalera, S., and Anbarjafari, G. (2016). Convolutional neural network super resolution for face recognition in surveillance monitoring. In *International conference on articulated motion and deformable objects*, pages 175–184. Springer.

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.

Robinson, M. D., Chiu, S. J., Toth, C. A., Izatt, J. A., Lo, J. Y., and Farsiu, S. (2017). New applications of super-resolution in medical imaging. In *Super-Resolution Imaging*, pages 401–430. CRC Press.

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.

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

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.