Thermal Image SuperResolution Through Deep Convolutional Neural Network

Rafael E. Rivadeneira[1]([✉]), Patricia L. Suárez[1], Angel D. Sappa[1,2], and Boris X. Vintimilla[1]

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

Abstract. Due to the lack of thermal image datasets, a new dataset has been acquired for proposed a super-resolution approach using a Deep Convolution Neural Network schema. In order to achieve this image enhancement process, a new thermal images dataset is used. Different experiments have been carried out, firstly, the proposed architecture has been trained using only images of the visible spectrum, and later it has been trained with images of the thermal spectrum, the results showed that with the network trained with thermal images, better results are obtained in the process of enhancing the images, maintaining the image details and perspective. The thermal dataset is available at http://www.cidis.espol.edu.ec/es/dataset.

Keywords: Thermal infrared images · Thermal cameras · Image enhancement · Convolutional neural networks

1 Introduction

The electromagnetic spectrum, as shown in Fig. 1, can be split up into several regions, such as the visible spectrum, ultraviolet, X-ray, infrared, radar, radio, among others. The infrared region can be additionally divided into the near (NIR: near-infrared), short (SWIR: short-wavelength infrared), middle (MWIR: mid-wavelength infrared), long (LWIR: long-wavelength infrared) and far (FIR: far-infrared) spectral bands, where the long-wavelength infrared is also known as thermal. All objects emit infrared radiation by themselves, independently of any external energy source, and depending on their temperature they emit a different wavelength in the long wavelength infrared spectrum (i.e., thermal). Thermal cameras capture information in this long wavelength spectral band; they are passive sensors that capture infrared radiation emitted by all objects with a temperature above absolute zero [8], thus it can provide valuable extra
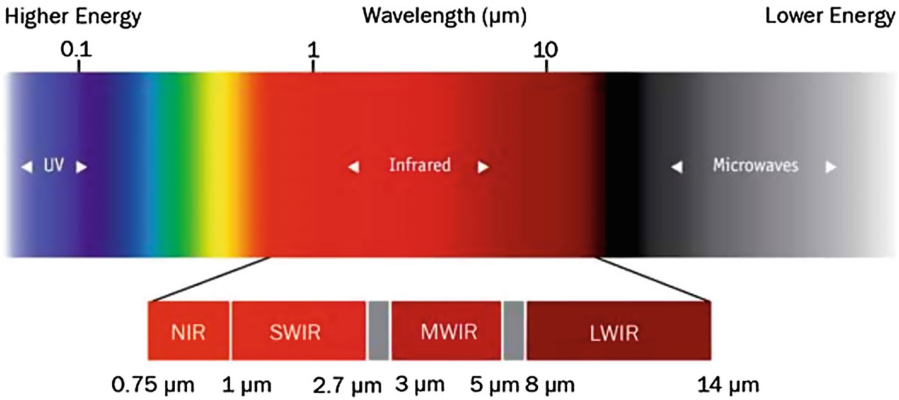
Fig. 1. Electromagnetic spectrum with sub-divided infrared spectrum

information to the visible one (e.g., RGB camera). In particularly, those applications that can be affected by poor lighting conditions, for instance in security and object recognition, where nothing can be captured in total darkness. Contrariwise, thermal cameras are not affected by this lack of illumination. As shown in Fig. 2 thermal images are represented as grayscale images, with dark pixels for cold spots and the whites one for hot spots.

In recent years, infrared imaging field has grown considerably; nowadays, there is a large set of infrared cameras available in the market (e.g, Flir[1], Axis[2], among others) with different technical specifications and costs. Innovative use of infrared imaging technology can therefore play an important role in many applications, such as medicine [16], military [9], objects or materials recognition [3], among others, as well as detection, tracking, and recognition of humans, or even applied for Vegetation Index Estimation [17].

Depending of the thermal camera's specifications, the cost can vary between $ 200.00 and more than $ 20000.00; the latter one has better resolution and higher frame rate. On the contrary, cheap existing thermal cameras have resolution smaller than commercial RGB cameras. This lack of resolution, at a moderate price, is a big limitation when thermal cameras need to be used for general purpose solutions. Hence, a possibility to overcome this limitation could be based on the development of new algorithms that allow to increase image resolution. This possibility has been largely exploited in the visible spectrum domain, where different super-resolution approaches have been proposed from a conventional interpolation (e.g., [7,10,18]). Recently, novel deep learning based approaches have been introduced with large improvements in performance (e.g., [5,11,15,19]). Hence, inspired on those approaches, some contributions have been proposed in the literature to tackle this challenging limitation of thermal imaging; most of these approaches are deep learning based (e.g., [4,13]).

Fig. 2. Thermal image capture with a Tau2 Camera

One of the most relevant approaches for image enhancement has been presented in SRCNN [5]; the approach is based on a convolutional neural network (CNN), where the architecture is trained to learn how to get a high-resolution image from an image with a lower resolution. The authors explored the performance by using different color space representations. They conclude that the best option is obtained by using the Y-channel from the YCbCr color space. The main limitation of their contribution is related with the training time. The approach, named "Accelerating the Super-Resolution Convolutional Neural Network" [6], from the same authors of the previous work, proposes accelerating and compacting their SRCNN structure for faster and better super resolution (SR). The authors introduce a deconvolution layer at the end of the network and adopt smaller filter size but more mapping layers. Yamanaka et al. [19] propose a CNN based approach referred to as "Deep CNN with Residual Net, Skip Connection and Network in Network" (DCSCN) for visible spectrum image super-resolution. According to the authors, this approach has a computation complexity of at least 10 time smaller that state of the art (e.g. VDSR [11], RED [15] and DRCN [12]). Like in the SRCNN in DCSCN the given images are converted to the YCbCr color space and only the Y-channel is considered. All these approaches have been proposed for images enhancement from the visible spectrum.

A CNN based approach for enhancing thermal images has been introduced by Choi et al. in [4], inspired by the proposal in [5]. The authors in [4] compare the accuracy of a network trained in different image spectrum to find the best representation of thermal enhancement. They conclude that a grayscale trained network provided better enhancement than the MWIR-based network
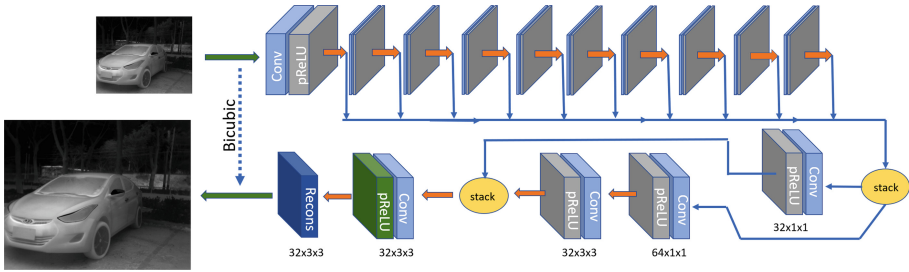
Fig. 3. Proposed convolutional neural network architecture

for thermal image enhancement. On the other hand, Lee et al. [13] also propose a convolutional neural network based on image enhancement for thermal images. The authors evaluate four RGB-based domains, namely, gray, lightness, intensity and V (from HSV color space) with a residual-learning technique. The approach improves the performance of enhancement and speed of convergence. The authors conclude that the V representation is the best one for enhancing thermal images. In [14] the authors proposed a parallelized 1×1 CNNs, named Network in Network to perform image enhancement with a low computational cost; also in [14], uses this technique for image reconstruction.

In most of the previous approaches thermal images have not been considered during the training stage, although intended for thermal image enhancement. They propose to train their CNN based approaches using images from the visible spectrum at different color space representations. On the contrary to all of them, in the current work thermal images are considered for training the proposed CNN architecture. The current work has two main contributions, the first one is the thermal image dataset acquisition used for training and validation. The second is proposed a CNN model designed for thermal spectrum images. The second one is to propose a CNN model designed for thermal spectrum images. Through this paper, terms "thermal images enhancement" and "images super resolution" will be indistinctly used.

The rest of the paper is organized as follows. Section 2 details the collected dataset and describe the approach proposed to enhance thermal images. Experimental results are presented in Sect. 3; and finally, Sect. 4 summarize main contributions of current work.

2   Proposed Approach

In the current work a deep CNN architecture with a residual net and dense connections are proposed. The network uses a thermal dataset to perform a super-resolution to maintain image details.

The architecture, presented in Fig. 3, has a part of the architecture dedicated to obtain the high level characteristics of the image, and another part, to perform the reconstitution of the image. All layers have dropouts and use parametric
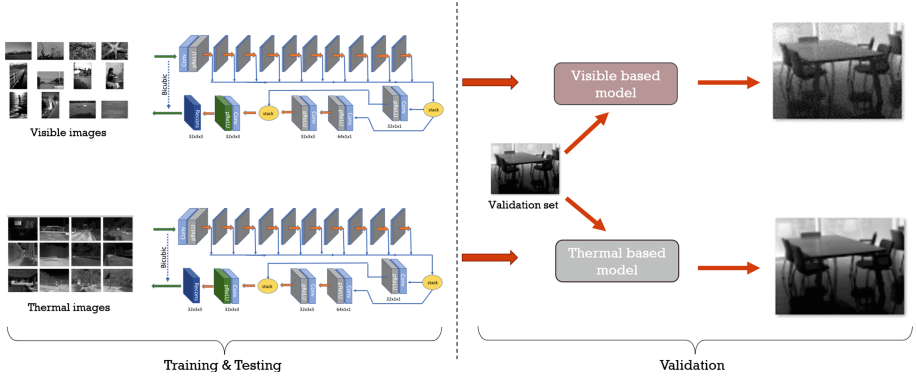
Fig. 4. Training process design of two datasets, using the proposal architecture, to generate two models for validation process.

ReLU as activator (preventing from learning a large negative bias term and getting better performance). Additionally, based on the work of [19], the image generated by bicubic interpolation has been used to enhance the output of the network.

This architecture is used for obtaining thermal image SR. On the contrary to the state-of-the-art approaches, where CNN based architectures are trained with visible spectrum images and used with thermal images. In this work the network is trained with thermal images in order to obtain better results. The latter hypothesis is validated by training the networks twice, one with visible images and one with thermal images. This training process results in two models (see Fig. 4), which are finally validated with thermal images. More details are given below.

3 Experiments Results

In this section, the dataset acquisition and preparation for training and testing are explained. Then the network setup information is provided, and finally the comparison of the two models are depicted.

3.1 Datasets for Training and Testing

As mention above the current work the architecture presented is trained twice, one with visible and other with thermal images. In this section the two datasets used for these training processes are detailed.

Due to the fact that there are not enough thermal image datasets, and the few ones available are in low resolution, a new dataset of 101 thermal images was generated (Fig. 5). This dataset was acquired using a TAU2[3] thermal camera

---

[3] https://www.flir.com/products/tau-2/.

with a 13 mm lens (45° HFOV) in a resolution of 640×512, with a depth of 8 bits and save it in PNG format. These images were acquired in indoors and outdoors environments, in the morning, day and night; they contain objects and people. Controller GUI software of TAU2 camera with the default value was used. In order to increase the variety of images, this dataset was enlarged with $98 + 40$ thermal images from a public dataset[4], acquired with a FLIR T640 using a 41 mm lens with 640×480 resolution. After merging all the images in these three datasets a total of 231 thermal images is obtained for training and testing. All these images were mixed, then 215 were randomly selected for training, 18 randomly selected for testing and the remainder 6 for validation (named as Thermal6). On the other hand, for training the visible model, the BSDS300 [1] is used for training, SET14 [20] for testing and SET5 [2] for validation. Note, as shown in Fig. 4, that the thermal images validation set is used to evaluate both models.

In order to increase the number of training images, a *data augmentation* process is performed, rotating and flipping from top to bottom, from left to right all images. The quality and resolution of the images is maintained getting a total of 1720 and 2400 images for thermal and visible respectively.

3.2   Training

The proposed architecture, has been training using a dense network, also, uses the image generated by bicubic interpolation to improve image details, also the layers for feature extractor uses dropout and ReLU operations, also a learning rate of 0.002 is applied to the model, and uses MSE as a loss function to measure the difference between the ground truth and the output. The model uses Adam Optimizer, which is an adaptive learning rate method, which means, it computes individual learning rates for different parameters. Its name is derived from adaptive moment estimation, and the reason it's called that is because Adam uses estimations of first and second moments of gradient to adapt the learning rate for each weight of the neural network. Each epoch train with a batch of 100000 patches for a total of 63 epochs. Mean Squared Error (MSE) between the ground truth and output is used as a basic loss value.

As presented above, in order to evaluate the proposed approach, the same architecture was trained with the two different datasets, the 1720 thermal images were split up into 48×48 patches with 25 pixels overlapping of adjacent patches, having a total batch of 185760. The 2400 visible images also were split up into 48×48 patches with 25 pixels overlapping, having a batch of 108000 (note that although there are more visible than thermal images the number of thermal patches is larger since thermal images have larger resolution.

The patches obtained above are used as ground truth, while the input patches are obtained by resizing them to half their original resolution. In the current work there is not noise added to the input.

---

[4] https://sites.google.com/view/multispectral/dde.

Fig. 5. Acquired thermal image dataset, with 640×512 resolution, using a Tau2 thermal camera.

The training is performed in Windows Server 2012, with a dual 2.50 GHz CPU E5-2640, using one GPU K20m of 4 GB. Each training consumes approximately 5 GB of RAM and takes approximately 25 h. This architecture is implemented using Tensorflow and Python.

3.3 Results

A fair comparison between the two models trained using the same infrastructure with the same number of batches per epochs and hyper parameters were used.

As show in Fig. 4, two models have been trained with the different dataset, each trained network was validated with a set of six thermal images (Thermal6) and five RGB images (SET5), obtaining a Visible Based Model and a Thermal

Based Model. Table 1 shows that with Thermal6, the thermal trained model shown a PSNR average value higher than the PSNR average value obtained with the Visible trained model. Also, it shows that SET5 got better PSNR values on visible model than thermal model. A qualitative comparison can be appreciated in Fig. 6, where the SR images obtained with the two models, as well as the images with the bicubic interpolation, are depicted. Additionally in this figure the ground truth is presented (values in brackets correspond to the average PSNR presented in Table 1).
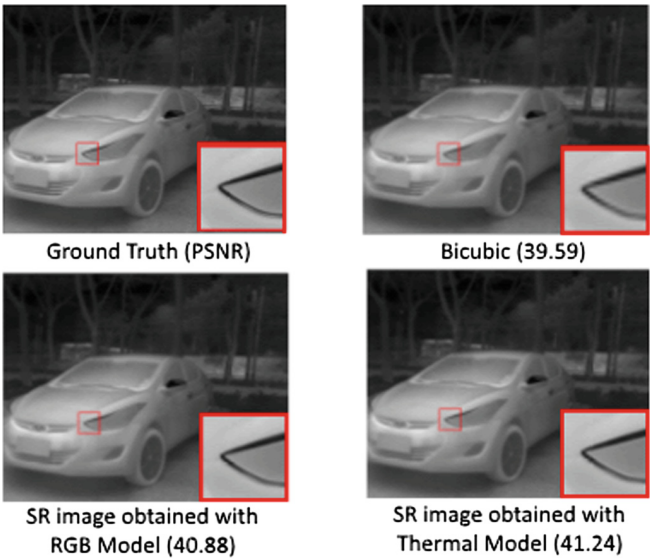

Fig. 6. Enhanced images (twice the original resolution) obtained with different approaches.

Table 1. Average result of PSNR with proposed architecture

Dataset	Scale	Bicubic based model	Visible based model	Thermal
Thermal6	×2	39.59	40.88	**41.24**
	×3	37.68	39.14	**39.62**
	×4	34.98	37.17	**37.85**
SET5	×2	33.64	**37.69**	37.46
	×3	30.37	**34.01**	33.74
	×4	28.41	**31.69**	31.25

4   Conclusions

In the current work, the usage of the proposal network has been considered to obtain thermal image SR. Two models have been obtained by training the same network with two different datasets in order to seek for the best options when thermal images are considered. The experimental results indicate that the network model trained with thermal image dataset is better than using visible image dataset. As an additional contribution a thermal image dataset has been acquired, which is publicly available. As a future work, new CNN architecture will be designed specifically intended for thermal images. Additionally, training the model using a dataset obtained from the combination of different domains (e.g., Y-channel, V-Brightness, Gray and Thermal) will be considered.

References

1. Arbelaez, P., Maire, M., Fowlkes, C., Malik, J.: Contour detection and hierarchical image segmentation. IEEE Trans. Pattern Anal. Mach. Intell. **33**(5), 898–916 (2011)
2. Bevilacqua, M., Roumy, A., Guillemot, C., Alberi-Morel, M.L.: Low-complexity single-image super-resolution based on nonnegative neighbor embedding. BMVA press (2012)
3. Cho, Y., Bianchi-Berthouze, N., Marquardt, N., Julier, S.J.: Deep thermal imaging: proximate material type recognition in the wild through deep learning of spatial surface temperature patterns. In: Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems, p. 2. ACM (2018)
4. Choi, Y., Kim, N., Hwang, S., Kweon, I.S.: Thermal image enhancement using convolutional neural network. In: 2016 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS), pp. 223–230. IEEE (2016)
5. Dong, C., Loy, C.C., He, K., Tang, X.: Image super-resolution using deep convolutional networks. IEEE Trans. Pattern Anal. Mach. Intell. **38**(2), 295–307 (2016)
6. Dong, C., Loy, C.C., Tang, X.: Accelerating the super-resolution convolutional neural network. In: Leibe, B., Matas, J., Sebe, N., Welling, M. (eds.) ECCV 2016. LNCS, vol. 9906, pp. 391–407. Springer, Cham (2016). https://doi.org/10.1007/978-3-319-46475-6_25
7. Duchon, C.E.: Lanczos filtering in one and two dimensions. J. Appl. Meteorol. **18**(8), 1016–1022 (1979)
8. Gade, R., Moeslund, T.B.: Thermal cameras and applications: a survey. Mach. Vis. Appl. **81**, 89–96 (2014)

9. Goldberg, A.C., Fischer, T., Derzko, Z.I.: Application of dual-band infrared focal plane arrays to tactical and strategic military problems. In: Infrared Technology and Applications XXVIII, vol. 4820, pp. 500–515. International Society for Optics and Photonics (2003)
10. Keys, R.: Cubic convolution interpolation for digital image processing. IEEE Trans. Acoust. Speech Signal Process. **29**(6), 1153–1160 (1981)
11. Kim, J., Kwon Lee, J., Mu Lee, K.: Accurate image super-resolution using very deep convolutional networks. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 1646–1654 (2016)
12. Kim, J., Kwon Lee, J., Mu Lee, K.: Deeply-recursive convolutional network for image super-resolution. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 1637–1645 (2016)
13. Lee, K., Lee, J., Lee, J., Hwang, S., Lee, S.: Brightness-based convolutional neural network for thermal image enhancement. IEEE Access **5**, 26867–26879 (2017)
14. Lin, M. Chen, Q., Yan, S.: Network in network. In: International Conference on Learning Representations (ICLR) (2014)
15. Mao, X., Shen, C., Yang, Y.B.: Image restoration using very deep convolutional encoder-decoder networks with symmetric skip connections. In: Advances in Neural Information Processing Systems, pp. 2802–2810 (2016)
16. Ring, E.F.J., Ammer, K.: Infrared thermal imaging in medicine. Physiol. Meas. **33**(3), R33 (2012)
17. Suárez, P.L., Sappa, A.D., Vintimilla, B.X.: Vegetation index estimation from monospectral images. In: Campilho, A., Karray, F., ter Haar Romeny, B. (eds.) ICIAR 2018. LNCS, vol. 10882, pp. 353–362. Springer, Cham (2018). https://doi.org/10.1007/978-3-319-93000-8_40
18. Watson, D.F., Philip, G.M.: Neighborhood-based interpolation. Geobyte **2**(2), 12–16 (1987)
19. Yamanaka, J., Kuwashima, S., Kurita, T.: Fast and accurate image super resolution by deep CNN with skip connection and network in network. CoRR, abs/1707.05425 (2017)
20. Zeyde, R., Elad, M., Protter, M.: On single image scale-up using sparse-representations. In: Boissonnat, J.D., et al. (eds.) Curves and Surfaces 2010. LNCS, vol. 6920, pp. 711–730. Springer, Heidelberg (2012). https://doi.org/10.1007/978-3-642-27413-8_47