A Generative Model for Guided Thermal Image Super-Resolution

Patricia L. Suárez[1][a] and Angel D. Sappa[1,2][b]
[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 Super-Resolution, HSV Color Space, Luminance-Driven Bicubic Image.

Abstract: This paper presents a novel approach for thermal super-resolution based on a fusion prior, low-resolution thermal image and H brightness channel of the corresponding visible spectrum image. The method combines bicubic interpolation of the $\times 8$ scale target image with the brightness component. To enhance the guidance process, the original RGB image is converted to HSV, and the brightness channel is extracted. Bicubic interpolation is then applied to the low-resolution thermal image, resulting in a Bicubic-Brightness channel blend. This luminance-bicubic fusion is used as an input image to help the training process. With this fused image, the cyclic adversarial generative network obtains high-resolution thermal image results. Experimental evaluations show that the proposed approach significantly improves spatial resolution and pixel intensity levels compared to other state-of-the-art techniques, making it a promising method to obtain high-resolution thermal.

1 INTRODUCTION

Super-resolution is the process of upgrading the resolution and quality of an image to obtain a higher-resolution version, reconstructing missing high-frequency information, and improving image clarity. It is a vital technique in image processing and computer vision, as it addresses the challenge of obtaining sharper, more detailed images from low-resolution sources. This involves the use of advanced algorithms and mathematical methods to infer high-resolution detail based on available low-resolution data. Super-resolution is the process of upgrading the resolution and quality of an image to obtain a higher-resolution version, reconstructing missing high-frequency information, and improving image clarity. It is a vital technique in image processing and machine vision as it addresses the challenge of getting sharper, more detailed images from low-resolution sources. This involves the use of advanced algorithms and mathematical methods to infer high-resolution detail based on available low-resolution data (e.g., (Mehri et al., 2021b), (Mehri et al., 2021a)). Lately, a variant of super-resolution has emerged that takes advantage of the guidance of an additional high-resolution image, known as a "guide image." This guide image has similar content to the low-resolution image but is acquired at a higher resolution. This additional information helps improve the quality and accuracy of the super-resolved output by providing more detailed and accurate information about the scene. Guided super-resolution strategy has been used to improve depth-map super-resolution; where a high-resolution RGB image is used as a guiding image (e.g., (Liu et al., 2017), (Guo et al., 2018)).

Thermal imaging technology has become relevant in various fields, including surveillance, medical imaging, and industrial applications, due to its ability to capture temperature variations and reveal hidden patterns in the infrared spectrum. However, the intrinsic limitations of thermal sensors often result in low-resolution images, which can make accurate analysis and decision-making processes difficult. Therefore, thermal super-resolution has emerged as a very popular technique to improve the quality, visibility, and accuracy of thermal images (e.g., (Rivadeneira et al., 2023), (Mandanici et al., 2019), (Prajapati et al., 2021), (Zhang et al., 2021)). By increasing resolution and reducing noise, thermal image super-resolution improves object clarity and detail in thermal images, benefiting various applications. Improved visibility of objects makes them appear sharper and more defined, aiding tasks such as surveillance and object identification. Additionally, thermal image super-resolution

[a] https://orcid.org/0000-0002-3684-0656
[b] https://orcid.org/0000-0003-2468-0031

Suárez, P. and Sappa, A.
A Generative Model for Guided Thermal Image Super-Resolution.
DOI: 10.5220/0012469400003660

In *Proceedings of the 19th International Joint Conference on Computer Vision, Imaging and Computer Graphics Theory and Applications (VISIGRAPP 2024) - Volume 3: VISAPP*, pages 765-771
ISBN: 978-989-758-679-8; ISSN: 2184-4321

provides more detailed information, increasing the accuracy of thermal image analysis. The guidance strategy mentioned above has also being used in the thermal image super-resolution problem, where a high-resolution visible spectrum image is considereded as a guidance for the super-resolution process (e.g., (Almasri and Debeir, 2019), (Almasri and Debeir, 2018), (Gupta and Mitra, 2021), (Gupta and Mitra, 2020)).

In the present work, the problem of guided super-resolution is addressed using an adversarial generative model, which allows obtaining improved synthetic images (super-resolved) using as input a fused image (HSV color space brightness channel and the low-resolution thermal image).

The performance and quality of the obtained synthesized super-resolved thermal images are comprehensively evaluated through extensive experiments and comparisons with state-of-the-art methods. The manuscript is structured as follows; Section 2, describes related state-of-the-art approaches. Section 3 introduces the proposed approach. Next, Section 4, shows experimental results and comparison with state-of-art approaches. Both quantitative and qualitative results are provided showing the improvements achieved with the proposed approach. Finally, conclusions and future works are given in Section 5.

2 RELATED WORK

Single image super-resolution (SISR) is a challenging task in image processing that aims to reconstruct high-resolution images from low-resolution images. In recent years, there has been a growing interest in using prior information to improve the performance of SISR methods. Prior information can be used to guide the reconstruction process and improve the quality of the output images. In this related work, we review some of the recent advances in SISR using prior information. One of the approaches presented in (Chudasama et al., 2020) proposes a CNN network named TherISuRNet which employs a progressive enhancement method that incorporates asymmetric residual learning, ensuring computational efficiency for a variety of enhancement factors, including $\times 2$ and $\times 3$, and $\times 4$. This architecture is specifically designed to include separate modules for extracting low and high-frequency features, complemented by up-sampling blocks. Another approach that uses a generative network is the one proposed in (Deepak et al., 2021) an architecture for the super-resolution of a single image based on a Generative Adversarial Network (GAN) is presented. This model has been specifically designed to improve the images of thermal cameras. The model achieves ease of implementation and computational efficiency. To speed up the model training process and reduce the number of training parameters, the number of residual blocks has been reduced to just 5, allowing for faster training. Additionally, the batch normalization layers are removed from the generator and discriminant networks to eliminate redundancy in the model architecture. Reflective padding is carefully incorporated before each convolutional layer to ensure that feature map sizes are preserved at the edges.

In (Thuan et al., 2022) a method is proposed to increase the resolution of thermal images using edge features of corresponding high-resolution visible images. This method is based on a Generative Adversarial Network (GAN) that uses residual dense blocks and can perform super-resolution. The dataset used in this method contains raw image data of indoor scenes captured by low-resolution thermal cameras. Another paper presented by (Zhang et al., 2022) introduces a novel network called Heat-Transfer-Inspired Network (HTI-Net) for SR image reconstruction, drawing inspiration from heat transfer principles. Their approach involves redesigning the ResNet network using a second-order mixed-difference equation derived from finite difference theory, allowing for enhanced feature reuse by integrating multiple information sources. Furthermore, a pixel value flow equation (PVFE) in the image domain has been developed, based on the thermal conduction differential equation (TCDE) in the thermal field, to tap into deep potential feature information. In Wang et al. (Wang et al., 2022), the authors introduce a deep-learning-based approach for fusing infrared and visible images, focusing on multimodal super-resolution reconstruction. Their method uses an encoder-decoder architecture to achieve this. Their approach was found to yield various imaging modalities and demonstrated superior performance in both visual effects and objective assessments. Additionally, the authors conduct a systematic review of the applications, methodologies, datasets, and evaluation metrics relevant to infrared (IR) image super-resolution. They categorized IR image super-resolution methods into two groups: traditional methods and deep learning-based methods. Traditional methods were further divided into three subcategories: frequency domain-based, dictionary-based, and other miscellaneous methods.

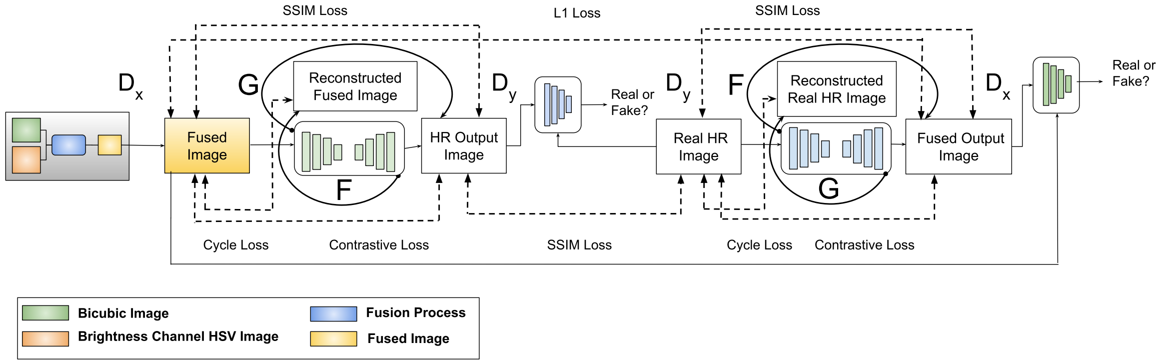

Figure 1: CycleGAN proposed architecture.

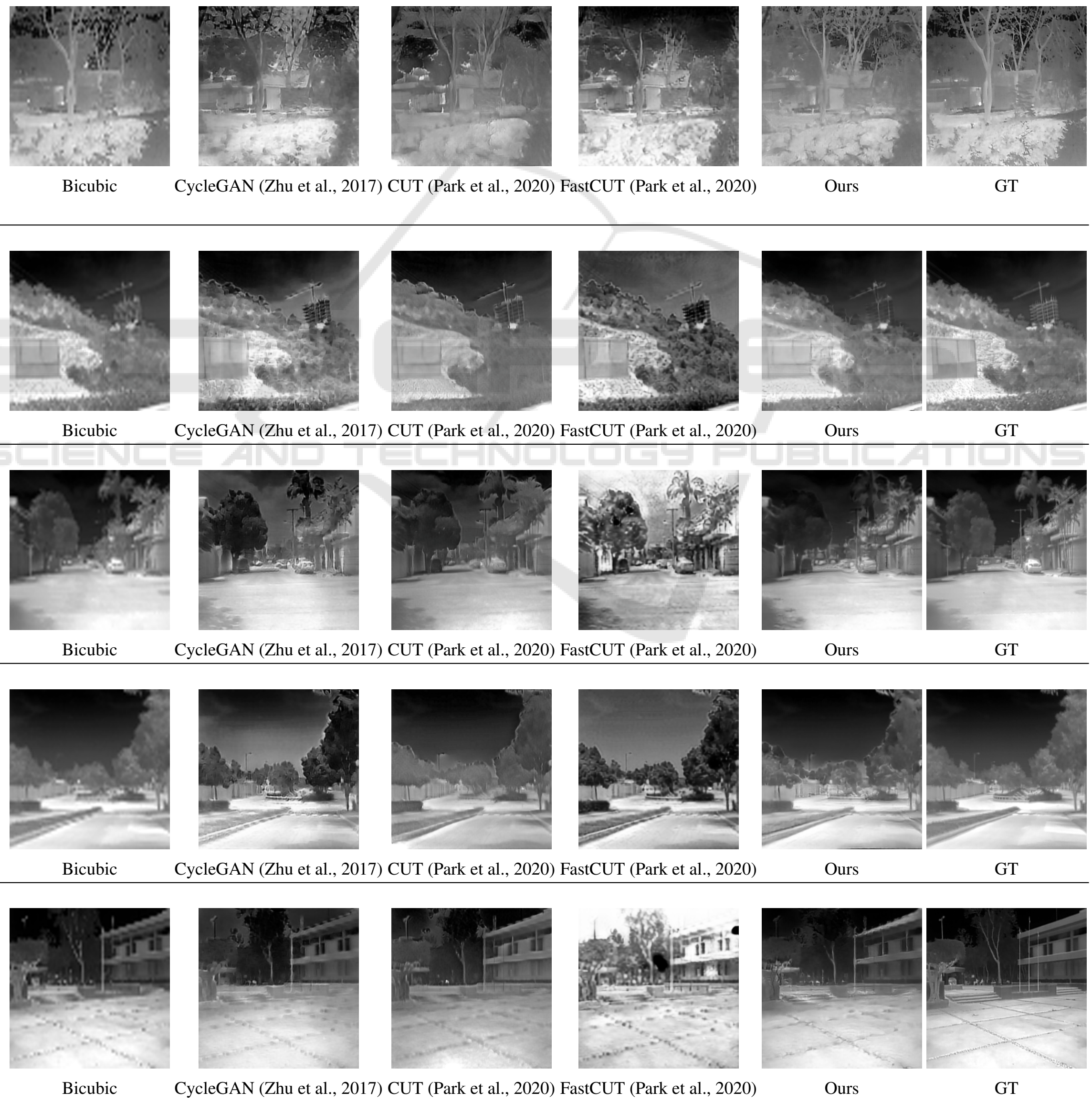

Figure 2: Results from the state-of-the-art and the proposed approaches.

Figure 3: Results from the state-of-the-art and the proposed approaches using Thermal Stereo testing dataset.

3 PROPOSED STRATEGY FOR THERMAL-LIKE SUPER-RESOLUTION

In this section, the applied strategy for achieving an $\times 8$ super-resolution of thermal images, which involves the utilization of a CycleGAN-based architecture (Zhu et al., 2017), trained with unpair images is presented. This sophisticated architecture is engineered to transform low-resolution thermal images into remarkably detailed, high-resolution counterparts.

The preprocessing phase is a crucial step in this process. It involves the fusion of the bicubic image, which is derived from the initial low-resolution thermal image, with the brightness channel of the HSV (Hue, Saturation, Value) color space. It is noteworthy that the RGB image was initially converted to the HSV color space, and then its brightness channel was extracted to obtain the luminance-bicubic fused image used as input in the architecture. The fusion process in the preprocessing phase helps the model generate high-resolution thermal images by integrating the thermal information with the brightness channel of the RGB image converted to HSV color space.

This fusion allows the model to obtain a more comprehensive representation of the image's content to enhance the understanding of the image content, preserving structural details and textures. Additionally, this integration of two different data sources allows the model to better preserve crucial details during the super-resolution process. Therefore, by taking advantage of the brightness channel of an HSV image, which contains information about the overall intensity of the image, the model can accurately represent the essential features and characteristics of the original content. It effectively enhances the model's ability to preserve crucial details during the super-resolution process.

Additionally, a residual layer is implemented as a skip connection at the end of the network (Zhang et al., 2018). This residual layer plays a pivotal role in fine-tuning the super-resolved output, ensuring that the essential features and characteristics of the original image are faithfully preserved. The architecture of the proposed approach is presented in Fig. 1.

For the training process, inspired on (Suárez and Sappa, 2023) multiple loss functions are employed, each serving a unique purpose in guiding the model toward the desired outcome. These loss functions are the L1 loss (Mean Absolute Error) which helps to calculate the absolute pixel-wise difference between the super-resolved and high-resolution images. It encourages the model to minimize these differences, ensuring accurate pixel-level reconstruction. This L1 loss is defined as:

$$L_{\text{L1\_regularized}}(G) = \frac{1}{N}\sum_{i=1}^{N}|G(x_i) - y_i| + \beta \cdot R(G), \quad (1)$$

where x_i and y_i are the pixel values at the same position (i, j) in the two images. N is the total number of pixels in the images. $\sum_{i=1}^{N}$ represents the sum over all the pixels in the images. $|x_i - y_i|$ calculates the absolute difference between the corresponding pixel values in the two images. β is the regularization hyperparameter that controls the importance of regularization in the loss function. $R(G)$ represents the regularization term, which may take the form of a parameter norm L1 applied to the model.

Also, the contrastive loss has been included, this loss function introduces a contrastive comparison between the discriminator's assessment of real and generated images. It is designed to encourage similar data points to be closer in the learned feature space while pushing dissimilar data points farther apart. This loss is defined as:

$$\mathcal{L}_{\text{contrastive}}(\hat{Y}, Y) = \sum_{l=1}^{L}\sum_{s=1}^{S_l} \ell_{\text{contr}}\left(\hat{v}_l^s, v_l^s, \bar{v}_l^s\right), \quad (2)$$

where $v_l \in \mathbb{R}^{S_l \times D_l}$ represents a tensor whose shape depends on the model architecture. The variable S_l denotes the number of spatial locations of the tensor. Consequently, the notation $v_l^s \in \mathbb{R}^{D_l}$ is employed to refer to the D_l-dimensional feature vector at the s-th spatial location. Additionally, $\bar{v}_l^s \in \mathbb{R}^{(S_l-1)\times D_l}$ represents the collection of feature vectors at all other spatial locations except the s-th one.

Additionally, the SSIM loss (Structural Similarity Index) has been used to evaluate the structural similarity between the super-resolved and high-resolution images. It focuses on preserving the structural details and textures in the output. The SSIM Loss is defined as:

$$\text{LSSIM} = 1 - \text{SSIM}(s, r), \quad (3)$$

where, s and r are the high-resolution images obtained from the model and the original high-resolution image, respectively, that are being compared. SSIM() is the SSIM function, which measures the similarity between two images (s, r). Finally, the cyclic consistency loss helps to ensure the cyclic consistency of the model's transformations. This loss function measures the difference between the original high-resolution image and the result of applying the model twice in succession. This cycle consistency loss is defined as:

$$\begin{aligned}\mathcal{L}_{\text{cycle}}(F, G) = {} & \mathbb{E}_{s\sim p_{\text{data}}(s)}[\|s - G(F(s))\|_1] \\ & + \mathbb{E}_{r\sim p_{\text{data}}(r)}[\|r - F(G(r))\|_1],\end{aligned} \quad (4)$$

where, s and r are the high-resolution images obtained from the model and the original high-resolution image, respectively, that are being compared.

These loss functions collectively guide the model during the training process, steering it toward generating high-quality, super-resolved thermal images that faithfully capture the intricate details and characteristics of the original content. The selection of λ values enables us to finely adjust the balance between these various objectives throughout the training process. This resulting loss is represented as:

$$\begin{aligned}\mathcal{L}_{\text{final}} = {} & \lambda_1 \mathcal{L}_{\text{cont}}(G, H, X) + \lambda_2 \mathcal{L}_{\text{cont}}(G, H, Y) \\ & + \lambda_3 \mathcal{L}_{\text{SSIM}(x,y)} + \lambda_4 \mathcal{L}_{\text{cycle}(G,F)} \\ & + \lambda_5 \mathcal{L}_{\text{L1\_regularizeed}(F,G)},\end{aligned} \quad (5)$$

where λ_i are empirically defined.

4 EXPERIMENTAL RESULTS

4.1 Datasets

In this research, our dataset Thermal Stereo, has been used and contains pairs of high-quality visible and thermal images captured under daylight conditions. The dataset consists of 200 image pairs taken with Basler and TAU2 cameras, each having different resolutions. The Elastix algorithm (Klein et al., 2009) has been employed to register these pairs of thermal and visible images, all possessing a resolution of 640×480 pixels.

For the experiments, the dataset has been split up into three subsets: training, validation, and test image pairs, comprising 160, 30, and 10 image pairs, respectively. It is important to note that the high-resolution visible spectrum images are used as references to enhance the low-resolution (LR) thermal images and generate high-resolution thermal images. No additional noise has been introduced to the downscaled images during this process.

4.2 Results

This section presents the experimental results of the proposed approach for a super-resolution model based on RGB images and its corresponding thermal low-resolution image. To compare the results, other state-of-the-art generative models with similar structural characteristics have been selected. All models have been trained with the same data set, to guarantee an evaluation that follows the same parameters and execution environment. Also, to evaluate the performance of these methods, widely used metrics such as SSIM (structural similarity index) and PSNR (Peak Signal-to-Noise Ratio) have been used.

The quantitative results of this comparison are summarized in Table 1, which shows the values obtained from the proposed super-resolution method in comparison with other similar generative approaches. In all cases, improvements can be observed in both metrics. Qualitatively, the images super-resolved with our strategy have greater contour detail than the images produced by the other generative models. Figures 2 and 3 show results for a sample of the test set, which was obtained with each super-resolution model for comparison.

The experimental results demonstrate that our approach generates high-quality, super-resolved thermal images that faithfully capture the details and levels of pixel intensity of the super-resolved thermal image. The approach is compared with various state-of-the-art methods, and it has demonstrated good performance in both quantitative and qualitative evaluations.

Table 1: Average results on super-resolution using our testing set. Best results in **bold**.

Approaches	**NYU Dataset**	
	PSNR	**SSIM**
Bicubic (Han et al., 2021)	27,244	0.792
CycleGAN (Zhu et al., 2017)	27,311	0.794
CUT (Park et al., 2020)	27,417	0.793
FastCUT (Park et al., 2020)	27,501	0.793
Proposed Approach	**27,893**	**0.802**

5 CONCLUSIONS

In conclusion, the proposed thermal super-resolution strategy has proven to be robust in the quality and tonality of the pixels. It has also shown better results in quantitative metrics. In future work, we are going to explore the integration of additional loss functions, such as perceptual loss or adversarial losses, which can further refine the ability of our model to generate high-quality super-resolved synthesized thermal images. In addition, the exploration of novel architectures, including attention mechanisms, or recurrent models could open new ways to improve thermal super-resolution techniques.

ACKNOWLEDGEMENTS

This material is based upon work supported by the Air Force Office of Scientific Research under award number FA9550-22-1-0261, and partially supported by the Grant PID2021-128945NB-I00 funded byMCIN/AEI/10.13039/501100011033 and by "ERDF A way of making Europe"; the "CERCA75046 Programme / Generalitat de Catalunya"; and the ESPOL project CIDIS-12-2022.

REFERENCES

Almasri, F. and Debeir, O. (2018). Rgb guided thermal super-resolution enhancement. In *2018 4th International Conference on Cloud Computing Technologies and Applications (Cloudtech)*, pages 1–5. IEEE.

Almasri, F. and Debeir, O. (2019). Multimodal sensor fusion in single thermal image super-resolution. In *Computer Vision–ACCV 2018 Workshops: 14th Asian Conference on Computer Vision, Perth, Australia, December 2–6, 2018, Revised Selected Papers 14*, pages 418–433. Springer.

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 (CVPR) Workshops*.

Deepak, S., Sahoo, S., and Patra, D. (2021). Super-resolution of thermal images using gan network. In *2021 Advanced Communication Technologies and Signal Processing (ACTS)*, pages 1–5.

Guo, C., Li, C., Guo, J., Cong, R., Fu, H., and Han, P. (2018). Hierarchical features driven residual learning for depth map super-resolution. *IEEE Transactions on Image Processing*, 28(5):2545–2557.

Gupta, H. and Mitra, K. (2020). Pyramidal edge-maps and attention based guided thermal super-resolution. In *Computer Vision–ECCV 2020 Workshops: Glasgow, UK, August 23–28, 2020, Proceedings, Part III 16*, pages 698–715. Springer.

Gupta, H. and Mitra, K. (2021). Toward unaligned guided thermal super-resolution. *IEEE Transactions on Image Processing*, 31:433–445.

Han, J., Shoeiby, M., Petersson, L., and Armin, M. A. (2021). Dual contrastive learning for unsupervised image-to-image translation. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops*.

Klein, S., Staring, M., Murphy, K., Viergever, M. A., and Pluim, J. P. W. (2009). elastix: A toolbox for intensity-based medical image registration. *IEEE Transactions on Medical Imaging*, 29(1):196–205.

Liu, W., Chen, X., Yang, J., and Wu, Q. (2017). Robust color guided depth map restoration. *IEEE Transactions on Image Processing*, 26:315–327.

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

Mehri, A., Ardakani, P. B., and Sappa, A. D. (2021a). Linet: A lightweight network for image super resolution. In *2020 25th International Conference on Pattern Recognition (ICPR)*, pages 7196–7202. IEEE.

Mehri, A., Ardakani, P. B., and Sappa, A. D. (2021b). Mprnet: Multi-path residual network for lightweight image super resolution. In *Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision*, pages 2704–2713.

Park, T., Efros, A. A., Zhang, R., and Zhu, J.-Y. (2020). Contrastive learning for unpaired image-to-image translation. In *European Conference on Computer Vision*.

Prajapati, K., Chudasama, V., Patel, H., Sarvaiya, A., Upla, K. P., Raja, K., Ramachandra, R., and Busch, C. (2021). Channel split convolutional neural network (chasnet) for thermal image super-resolution. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 4368–4377.

Rivadeneira, R. E., Sappa, A. D., Vintimilla, B. X., Bin, D., Ruodi, L., Shengye, L., Zhong, Z., Liu, X., Jiang, J., and Wang, C. (2023). Thermal image super-resolution challenge results. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops*.

Suárez, P. L. and Sappa, A. D. (2023). Toward a thermal image-like representation. In *Proceedings of the International joint Conference on Computer Vision*.

Thuan, N. D., Dong, T. P., Manh, B. Q., Thai, H. A., Trung, T. Q., and Hong, H. S. (2022). Edge-focus thermal image super-resolution using generative adversarial network. In *2022 International Conference on Multimedia Analysis and Pattern Recognition (MAPR)*, pages 1–6.

Wang, B., Zou, Y., Zhang, L., Li, Y., Chen, Q., and Zuo, C. (2022). Multimodal super-resolution reconstruction of infrared and visible images via deep learning. *Optics and Lasers in Engineering*, 156:107078.

Zhang, M., Wu, Q., Guo, J., Li, Y., and Gao, X. (2022). Heat transfer-inspired network for image super-resolution reconstruction. *IEEE Transactions on neural networks and learning systems*.

Zhang, W., Sui, X., Gu, G., Chen, Q., and Cao, H. (2021). Infrared thermal imaging super-resolution via multi-scale spatio-temporal feature fusion network. *IEEE Sensors Journal*, 21(17):19176–19185.

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, 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.