

Leveraging Multi scale Backbone with Multilevel supervision for Thermal Image Super Resolution

Sabari Nathan*
Couger Inc
Shibuya, Tokyo, Japan
sabari@couger.co.jp

Priya Kansal*
Couger Inc
Shibuya, Tokyo, Japan
priya@couger.co.jp

Abstract

This paper proposes an attention-based multi-level model with a multi-scale backbone for thermal image super-resolution. The model leverages the multi-scale backbone as well. The thermal image dataset is provided by PBVS 2020 in their thermal image super-resolution challenge. This dataset contains the images with three different resolution scales (low, medium, high) [1]. However, only the medium and high-resolution images are used to train the proposed architecture to generate the super-resolution images in $x2$, $x4$ scales. The proposed architecture is based on the Res2net blocks as the backbone of the network. Along with this, the coordinate convolution layer and dual attention are also used in the architecture. Further, multi-level supervision is implemented to supervise the output image resolution similarity with the real image at each block during training. To test the robustness of the proposed model, we evaluated our model on the Thermal-6 dataset [20]. The results show that our model is efficient to achieve state-of-the-art results on the PBVS dataset. Further the results on the Thermal-6 dataset show that the model has a decent generalization capacity.

1. Introduction

The task of generating a higher resolution image from the lower resolution input images is known as image super resolution [17]. Due to the development of automation in every field, the super-resolution of images is of vital importance. For example, digital cameras, with their advancement, uses super-resolution for image enhancement tasks. The utmost importance of image super-resolution task is when images are captured using low pixel devices and send to some server using the internet connectivity for further analysis. In this case, the image resolution is kept low due to the problem of limited hardware resources at the client

location or sending pictures to the server with a high FPS. These images are further passed to a super-image resolution network on the server for further analysis to get the desired results. However, in recent years, imaging techniques have also been developed a lot. Nowadays, practitioners can capture almost all the visible spectrum regions of an image including the thermal spectrum. Thermal images are infrared radiation emitted by all objects with different temperatures and temperatures above absolute zero. [20] [5]. Unlike the RGB images, the images captured in the thermal spectral band are not affected by the lighting and other environmental conditions, hence these images have wide applications such as medicine, military, object detection, recognition, and tracking [5]. However, capturing these thermal images with a high resolution is quite expensive because of the expensive equipments [20]. Hence, the requirement of high-resolution thermal images at an affordable cost is the need of time. Researchers are working on the thermal image super-resolution as an alternative to this problem. However, image super-resolution is always a challenging problem. Recently, the remarkable performance of neural networks inspired researchers to use these networks for thermal image super-resolution. The approach proposed in this paper is also a deep convolutional neural network-based approach, which exploits the coordinate convolutional layer, multi-scale Res2net connections, and attention modules. The proposed network is novel in the following way:

- Because of the multi-level supervision, this single model can handle the super-resolution task at different scales ($x2$, $x4$).
- This model has more receptivity of spatial and channel information at almost the same computational cost as it exploits the Res2net block as the backbone of the model.
- This model is more robust as the spatial dimensions are expanded in Cartesian space at the start of the network using a coordinate convolutional layer and can retain

*These authors contributed equally to this work

all the spatial and channel information with the help of dual attention at the end.

Rest of the paper is organized in 5 sections. Section 2 consists of the brief review of existing studies. Detailed architecture is discussed in section 3. Section 4 and section 5 deals with details of the experimental set up and the results of ablation studies respectively. Lastly, section 6 deals with the conclusion and future scope of work.

2. Related work

Due to the wide applications, image super-resolution is widely studied in the last few decades. However, with the recent development in deep convolutional neural networks and their impressive performance, researchers in this field have also get attracted to the use of convolutional neural networks for image super-resolution tasks. For example, [3] constructs a three-layer deep convolutional neural network for image super-resolution in which the features of the LR input image are extracted and up-sampled in the last layer. The results of this model outperformed most of the previous non-deep learning-based methods. Similarly, [11][26] develop a model based on the residual learning which is much deeper than the previous methods. There exist a lot of experimentation to improve the performance of the task of super-resolution such as [14] experimented to improve the speed during training by removing the batch normalization. [4][22][27] propose approaches to reduce the complexity and the cost of the image super-resolution task. Further, the HR images generated using generative adversarial networks have also given some impressive results[13]. However, all these approaches have discussed the super-resolution of the images in the RGB spectrum. There are only a few studies that develop the approaches to generate high-resolution(HR) images from the low-resolution (LR) thermal images. Recently [21] propose a deep CNN network with residual blocks exploiting dense connection. Similarly, [17] develop a model based on the Cycle GAN architecture for re-scaling the thermal images from LR to HR. The present work is also a deep convolutional network that is based on some attentions module and exploits the multi-level supervision to train the network.

3. Proposed Architecture

This section provides the details of our proposed architecture for the task of generating x2 and x4 resolution images which are acquired at two different resolutions. The proposed approach is based on the neural network which leverages the multi-scale backbone and multi-level supervision to map the information between LR images and HR images and between MR images and HR images of a different domain. Fig.1 represents the details of the proposed architecture. As shown in the figure, the proposed architecture

is a simple six-block stacked network. Each block consists of one convolutional layer followed by two Res2net blocks [6]. The output of the convolutional layer is up-sampled by the factor 2 for x2 resolution and by factor 4 for x4 resolution using sub-pixel up-scaling layer [22]. Each up-scaled output is then fused for generating the final high-resolution image as well as supervised to improve the pixel-wise resolution as inspired in [10, 9, 16]. However, to increase the dimensional feature information, we first mapped the input image to the Cartesian space using the coordinate convolutional layer [15] as inspired by [10, 9]. A detailed discussion of each part of the network is given in the following sections:

3.1. Dual Attention Module

The output of all up-sampled layers is concatenated. Afterward, two-dimensional attention, both spatial and channel, is employed on this concatenated feature map. Since convolution operations extract informative features by blending cross-channel and spatial features, the attention module emphasizes meaningful features for both the spatial and channel dimensions.

Inspired by [24, 9], for creating the channel attention, we utilized both average pooling and max pooling features. To create the channel attention, we first aggregated the spatial information by creating two pooled feature maps using average pooling and max pooling, thereafter two single-layer perceptrons are used to create the channel attention maps. The output feature maps are then merged and a sigmoid activation is applied to get the final channel attention. Finally, this channel attention is multiplied with the input feature map. Eq.(1) shows the mathematical operation of channel attention.

$$C_A = xX f_a[w_1(w_0 \frac{\sum_{i=1}^n x_i}{n}) + w_1(w_0(max(x_i)))] \quad (1)$$

To apply, spatial attention, the pooling operations are done along the channel axis. Then these two max-pooled and average pooled operations are concatenated and then a convolutional operation is applied with 7x7 filter. Similar to channel attention, the spatial attention is then multiplied to the input feature map. Eq.(2) shows the mathematical operation involve in spatial attention.

$$S_A = xX f_{conv}[\frac{\sum_{i=1}^n x_i}{n} ||max(x_i)] \quad (2)$$

We have used the sequential way of applying attention. The channel attention is applied to the input features generated by the Res2net block. The output is then forwarded to spatial attention. Hence the combined attention is the spatial attention on the channel attention as shown in Eq.(3).

$$attention = S_A(C_A(x)) \quad (3)$$

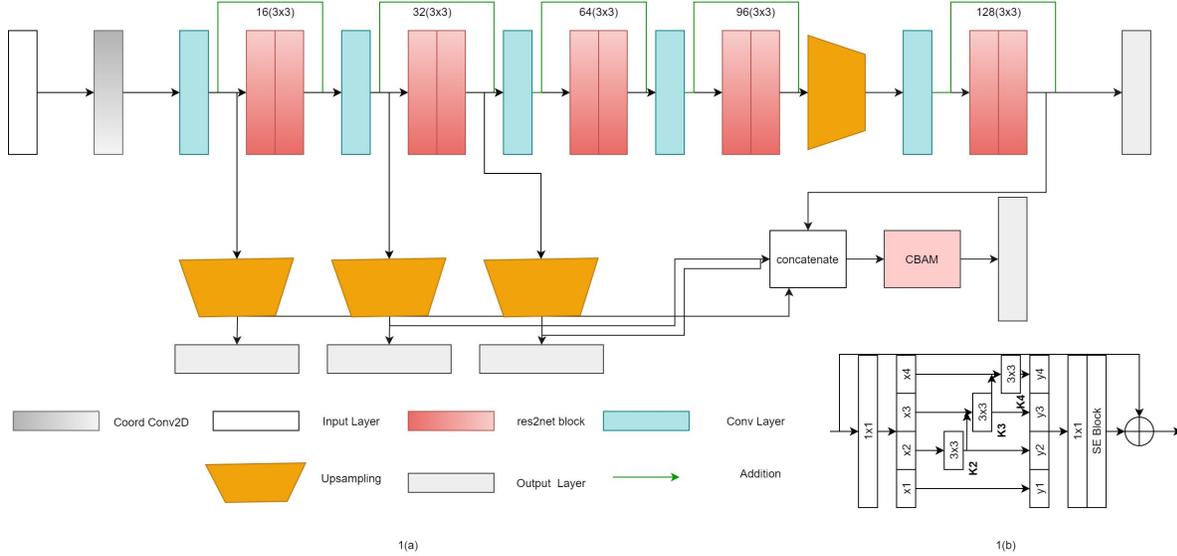


Figure 1: (a) Proposed Model: A Super-Resolution multi-level supervision model with multi-scale Backbone; (b) multi-scale Backbone: A Res2net block

where, x represents the input feature map, f_a and f_{conv} represent the sigmoid activation function and 7×7 convolutional operation respectively. w_0 and w_1 are the shared single layer perceptrons.

3.2. Res2net block based Backbone

As the backbone of the proposed network, Res2net block[6] is used. The Res2net block is a single Res2net block that has a stronger feature extraction ability without any additional computing complexities. As described in [6] and shown in Figure 1(b), firstly, 1×1 convolution is applied to the input feature map. The output is then further equally sliced into n sub-feature map in a way that each feature map has the same spatial dimensions but c/n in the channel, where c is the number of channels in the input feature map. Thereafter, each sub-feature map, except then 1st is forwarded to the 3×3 convolution. The output of the convolution is then added to the subsequent sub-feature map. All the outputs are then concatenated and passed again to 1×1 convolution operation. The Res2net is then added to the input feature map again. The multi-scale structure of Res2net block gives the different number and different combinations of the feature maps and thus a large number of the receptive field. Due to this, the feature information at a very granular level is retained for the next layer.

3.3. Multi-level Supervision

Inspired by [9, 16, 10], to improve the image resolution at different scales using one model, we used multi-level supervision. This helped the model to learn the features according to the original high-resolution image. As the re-

ceptive field increases across successive layers, predictions computed at different layers embed spatial information at different levels. The network is able to update the weights more efficiently, and propagate the gradient in the intermediary level to learn the features at each intermediary scale. The output of the first three Res2net blocks is directly up-sampled and supervised. However, the up-sampled output of the third Res2net block is further passed to two more sets of Res2net blocks, up-sampled and then to one more set of Res2net block. The output of the last block is directly supervised as well as supervised after concatenation along with the previous three up-sampled outputs and the attention. This multi-level supervision guided the network to generate the HR images progressively from the different resolutions respectively. This enabled a single architecture to work on both $\times 2$ and $\times 4$ resolution tasks. The detailed ablation study shows that this multi-level supervision is improving the results significantly.

4. Experimental Setup

4.1. Dataset

PBVS'2020 Dataset For generating super-resolution images, PBVS'2020 [1] provides thermal image captured using three different thermal cameras with three different resolutions (low 160×120 , mid 320×240 and high 640×480). As of the part of PBVS'2021 challenge, [2], only the mid resolution and high-resolution images are used. The high-resolution images are down-sampled by $\times 4$ with added noise for training. For training $\times 2$ model, the mid-resolution images are used as input, and the corresponding same scene

high-resolution images are used as ground truth. Table 1 shows the input-output detail of the dataset and the scales for resolution. A total of 951 images for training and 50 images for validation with each resolution (high and mid) are shared in the development phase whereas 20 images with each resolution are shared for the test phase[17]. A sample image from each resolution is shown in Figure2.

Input		Output	
Scale	Camera	Scale	Camera
High (HR) x1/4	FC-632O	High(HR)	FC-632O
Mid (MR)	Axis Q290	High(HR)	FC-632O

Table 1: Dataset Details

Thermal6 Dataset For testing the robustness of the proposed architecture and the model trained, we tested the results on the Thermal6 dataset also. Thermal6 dataset is acquired using a Tau2 camera with a resolution of 640 x 512. A total of 101 images are there in the dataset which includes the indoor and outdoor environment in day and night both [20].

4.2. Training

The network is trained for five outputs which include the four side layers and one fused output layer with low-resolution input images. All the outputs are supervised using the loss proposed in Eq.7. Total two models are trained to get the high-resolution images which are double (x2) and four times(x4) in scale when compared to input images. To train both x2, x4 resolution models, the high-resolution(HR) images captured from the FC-632O FLIR camera are used as ground truth images. The input for the x2 model is mid-resolution(MR) images captured from Axis Q2901-E camera, whereas for the x4 model, the high-resolution(HR) images are down-sampled to x/4 with added noise as described by the organizers [2] [Refer Table 1]. Adam optimizer is used to update the weights while training. The learning rate is initialized with 0.001 and reduced after 15 epochs to 10 percent if validation loss does not improve. The batch size is set to 4. The total epochs are set to 500. However, training is stopped early when the network got saturated. The dataset is trained using NVIDIA 1080 GTX GPU. The model is evaluated using Peak-Signal-to Noise Ratio (PSNR) and Structural Similarity Index(SSIM) loss.

4.3. Loss Function

To supervise the model outputs, a combination of three different loss functions are used: mean squared error (MSE), SSIM Loss, and Sobel edge loss (SOBEL Loss). MSE is used for maintaining the consistency between input and output; it is defined as:

$$MSE = \frac{1}{N} \sum_{p=1}^P (f(x) - x)^2 \quad (4)$$

where $f(x)$ is the pixel value of generated HR image and x is the pixel value of HR real image.

Pixel wise Structural Similarity Loss is defined as:

$$SSIM \text{ Loss} = \frac{1}{N} \sum_{p=1}^P (1 - SSIM(p)) \quad (5)$$

where $SSIM(p)$ structural similarity index[23] for pixel p .

Sobel loss is the mean squared error of the Sobel edge information of the real image and the generated image. A Sobel filter to detect the edges is applied to the generated and real image and then this information is used to calculate the mean squared error which is equal to the Sobel loss. More information can be found in [25][12]. A mathematical representation is given in equation 3:

$$SOBEL \text{ Loss} = \frac{1}{N} \sum_{i=1}^N (S(f(x)) - S(x))^2 \quad (6)$$

where $S(f(x))$ is the sobel edge information of the generated HR image and $S(x)$ is the sobel edge information of the real image.

Total loss is the sum of the all three losses.

$$\text{Total Loss} = \text{MSE} + \text{SSIM Loss} + \text{SOBEL Loss} \quad (7)$$

5. Experimental Results

5.1. Results

Table 2 shows the results of our proposed model on the validation images of the PBVS’2020 dataset and Thermal6 dataset.

Scale	PBVS(Val)		Thermal6	
	PSNR	SSIM	PSNR	SSIM
x2 Scale	34.1769	0.9116	39.8235	0.9569
x4 Scale	29.813	0.7833	37.4714	0.9308

Table 2: Results of the Proposed Model on PBVS’2020 and Thermal 6

Scale	Bicubic Model	TISR[21]	MLSM[10]	Ours
x2 scale	39.59	41.24	40.89	39.82
x4 scale	34.98	37.85	37.60	37.47

Table 3: Results on Thermal 6 dataset compared to the state-of-the-art

Moreover, the proposed model results on the Thermal-6 dataset is compared with the existing state-of-the-art methods. Table 3 shows the results of the Thermal 6 dataset.

The model which was trained on the PBVS'2020 dataset is used for calculating the evaluation metrics on the Thermal6 dataset. The results are at par with the previous approaches, which show that the model is having a good generalization capacity.



Figure 2: PBVS dataset: Generated output and the Real Output Images in three different scale

Figure 2 and Figure 3 depicts the real image and generated images of PBVS dataset and Thermal6 dataset.

5.2. Ablation Study

To prove the efficiency of our proposed architecture, a wide range of ablation studies have been performed. Table 4, 5, 6 and 8 shows the quantitative results calculated on the experimentation of using and not using the multi-level supervision (MS), the coordinate convolution layer, combined loss, dual attention module, and Res2net block while training respectively.

Multi-level Supervision	x2 Scale		x4 Scale	
	PSNR	SSIM	PSNR	SSIM
Used	34.048	0.911	29.813	0.7833
Not Used	32.573	0.8107	28.329	0.7401

Table 4: Experimentation on Multi-level Supervision

The use of coordinate convolutional layer, attention, the combined loss, and multi-supervision is having proven result in [10]. The results on the proposed model voted for the previous studies and clearly show that the proposed architecture is also having superior performance with a coordinate convolutional layer, multi-level supervision, and attention module when compared to the scenarios where we are not using these.

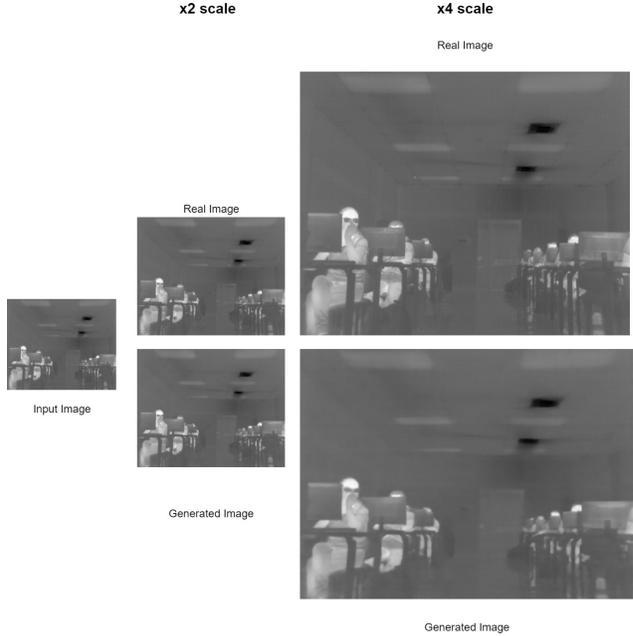


Figure 3: Thermal dataset: Real Image and Generated output image in three different scale

dinate convolutional layer, multi-level supervision, and attention module when compared to the scenarios where we are not using these.

Coordinate Convolutional	x2 Scale		x4 Scale	
	PSNR	SSIM	PSNR	SSIM
Used	32.24	0.8187	27.144	0.7607
Not Used	31.529	0.8157	26.541	0.7135

Table 5: Ablation study of the Coordinate Convolutional layer

Loss	x2		x4	
	PSNR	SSIM	PSNR	SSIM
SSIM+MSE	25.3192	0.6319	26.87	0.6319
MSE	28.9741	0.8144	28.97	0.8144
SSIM+MSE+SOBEL	34.1769	0.911	29.813	0.7833

Table 6: Performance of the proposed architecture for various loss functions

Dual Attention	x2 Scale		x4 Scale	
	PSNR	SSIM	PSNR	SSIM
Used	31.616	0.82532	25.704	0.7086
Not Used	30.332	0.8035	26.312	0.7284

Table 7: Experimentation on the use of Dual Attention

Further, we also experimented by replacing the state-of-the-art backbone blocks. Table 6 shows the results of the proposed architecture with using residual block[7], residual-dense block[28], dense block [8] and Res2net block[6]. According to Table 6, the proposed model is having the best results with Res2net block.

Backbone Block	x2 Scale		x4 Scale	
	PSNR	SSIM	PSNR	SSIM
Res2net	34.048	0.911	29.813	0.7833
Residual	31.4	0.9326	31.44	0.9267
Residual Dense	30.26	0.8521	26.336	0.7515
Dense	29.085	0.8697	24.336	0.74

Table 8: Performance of the proposed architecture for various state-of-the-art blocks as back-bone

Table 9 shows the results on the test data of PBVS 2021 test dataset.

Method	x4 Scale		x2 Scale	
	PSNR	SSIM	PSNR	SSIM
PBVS 2020 Winner[18]	27.72	0.8758	20.02	0.7452
MLSM[10]	27.31	0.8498	20.36	0.7595
PBVS 2021 Winner[19]	30.70	0.929	20.09	0.751
Ours	29.13	0.8469	20.03	0.7484

Table 9: Results on PBVS Test dataset

6. Conclusion

This present paper proposes an attention-based multi-level supervised network with a multi-scale backbone to create high-resolution images with x2 and x4 scale. The network uses residual learning and multi-scale supervision to retain the spatial information throughout training, which helps to improve the robustness of this model. The multi-level supervision also enabled the model to learn the resolution hierarchy throughout the network. The quantitative results and in-depth ablation study results show that the proposed network is not only efficient enough to achieve the results on the PBVS dataset for x2 and x4 scale but also able to generalize the performance on other datasets like Thermal6. In the future, we will exploit the same architec-

ture for images in the RGB spectrum and for other image restoration tasks like image dehazing, image relighting, etc in RGB as well as a thermal spectrum.

7. Acknowledgement

The present work is supported by Couger Inc. Tokyo.

References

- [1] IEEE PBVS'2020. [://vcip-okstate.org/pbvs/20/challenge.html](https://vcip-okstate.org/pbvs/20/challenge.html). 1, 3
- [2] IEEE PBVS'2021. [://pbvs-workshop.github.io/challenge.html](https://pbvs-workshop.github.io/challenge.html). 3, 4
- [3] Chao Dong, Chen Change Loy, Kaiming He, and Xiaoou Tang. Image super-resolution using deep convolutional networks. *IEEE transactions on pattern analysis and machine intelligence*, 38(2):295–307, 2015. 2
- [4] Chao Dong, Chen Change Loy, and Xiaoou Tang. Accelerating the super-resolution convolutional neural network. In *European conference on computer vision*, pages 391–407. Springer, 2016. 2
- [5] Rikke Gade and Thomas B Moeslund. Thermal cameras and applications: a survey. *Machine vision and applications*, 25(1):245–262, 2014. 1
- [6] Shanghua Gao, Ming-Ming Cheng, Kai Zhao, Xin-Yu Zhang, Ming-Hsuan Yang, and Philip HS Torr. Res2net: A new multi-scale backbone architecture. 2, 3, 6
- [7] Kaiming He, Xiangyu Zhang, Shaoqing Ren, and Jian Sun. Deep residual learning for image recognition. *2016 IEEE Conference on CVPR*, Jun 2016. 6
- [8] G. Huang, Z. Liu, L. Van Der Maaten, and K. Q. Weinberger. Densely connected convolutional networks. In *2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 2261–2269, 2017. 6
- [9] Priya Kansal and Sabari Nathan. Eyenet: Attention based convolutional encoder-decoder network for eye region segmentation, 2019. 2, 3
- [10] Priya Kansal and Sabari Nathan. 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 (CVPR) Workshops*, June 2020. 2, 3, 4, 5, 6
- [11] Jiwon Kim, Jung Kwon Lee, and Kyoung Mu Lee. Accurate image super-resolution using very deep convolutional networks. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 1646–1654, 2016. 2
- [12] Josef Kittler. On the accuracy of the sobel edge detector. *Image and Vision Computing*, 1(1):37–42, 1983. 4
- [13] Christian Ledig, Lucas Theis, Ferenc Huszár, Jose Caballero, Andrew Cunningham, Alejandro Acosta, Andrew Aitken, Alykhan Tejani, Johannes Totz, Zehan Wang, et al. Photo-realistic single image super-resolution using a generative adversarial network. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 4681–4690, 2017. 2

- [14] Bee Lim, Sanghyun Son, Heewon Kim, Seungjun Nah, and Kyoung Mu Lee. 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, 2017. 2
- [15] Rosanne Liu, Joel Lehman, Piero Molino, Felipe Petroski Such, Eric Frank, Alex Sergeev, and Jason Yosinski. An intriguing failing of convolutional neural networks and the coordconv solution, 2018. 2
- [16] Sabari Nathan and Priya Kansal. Skeletonnet: Shape pixel to skeleton pixel. In *Proceedings of the IEEE Conference on CVPRw*, 2019. 2, 3
- [17] Rafael E Rivadeneira, Angel D Sappa, and Boris X Vintimilla. Thermal image super-resolution: a novel architecture and dataset. 1, 2, 4
- [18] R. E. Rivadeneira, A. D. Sappa, B. X. Vintimilla, L. Guo, J. Hou, A. Mehri, P. B. Ardakani, H. Patel, V. Chudasama, K. Prajapati, K. P. Upla, R. Ramachandra, K. Raja, C. Busch, F. Almasri, O. Debeir, S. Nathan, P. Kansal, N. Gutierrez, B. Mojra, and W. J. Beksi. Thermal image super-resolution challenge - pbvs 2020. In *2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops (CVPRW)*, pages 432–439, 2020. 6
- [19] R. E. Rivadeneira, A. D. Sappa, B. X. Vintimilla, S. Nathan, P. Kansal, N. Gutierrez, B. Mojra, and W. J. Beksi. Thermal image super-resolution challenge - pbvs 2021. In *2021 IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops (CVPRW)*, 2021. 6
- [20] Rafael E Rivadeneira, Patricia L Suárez, Angel D Sappa, and Boris X Vintimilla. Thermal image superresolution through deep convolutional neural network. In *International Conference on Image Analysis and Recognition*, pages 417–426. Springer, 2019. 1, 4
- [21] Rafael E Rivadeneira, Patricia L Suárez, Angel D Sappa, and Boris X Vintimilla. Thermal image superresolution through deep convolutional neural network. In *International Conference on Image Analysis and Recognition*, pages 417–426. Springer, 2019. 2, 4
- [22] Wenzhe Shi, Jose Caballero, Ferenc Huszár, Johannes Totz, Andrew P Aitken, Rob Bishop, Daniel Rueckert, and Zehan Wang. Real-time single image and video super-resolution using an efficient sub-pixel convolutional neural network. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 1874–1883, 2016. 2
- [23] Zhou Wang, Alan C Bovik, Hamid R Sheikh, and Eero P Simoncelli. Image quality assessment: from error visibility to structural similarity. *IEEE transactions on image processing*, 13(4):600–612, 2004. 4
- [24] Sanghyun Woo, Jongchan Park, Joon-Young Lee, and In So Kweon. Cbam: Convolutional block attention module. In *Proceedings of the European Conference on Computer Vision (ECCV)*, pages 3–19, 2018. 2
- [25] Shanxin Yuan, Radu Timofte, Gregory Slabaugh, Ales Leonardis, Bolun Zheng, Xin Ye, Xiang Tian, Yaowu Chen, Xi Cheng, Zhenyong Fu, et al. Aim 2019 challenge on image demoiring: Methods and results. *arXiv preprint arXiv:1911.03461*, 2019. 4
- [26] Kai Zhang, Wangmeng Zuo, Shuhang Gu, and Lei Zhang. Learning deep cnn denoiser prior for image restoration. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 3929–3938, 2017. 2
- [27] Lei Zhang, Peng Wang, Chunhua Shen, Lingqiao Liu, Wei Wei, Yanning Zhang, and Anton Van Den Hengel. Adaptive importance learning for improving lightweight image super-resolution network. *International Journal of Computer Vision*, 128(2):479–499, 2020. 2
- [28] Y. Zhang, Y. Tian, Y. Kong, B. Zhong, and Y. Fu. Residual dense network for image super-resolution. In *2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 2472–2481, 2018. 6