This CVPR workshop paper is the Open Access version, provided by the Computer Vision Foundation.
Except for this watermark, it is identical to the accepted version;
the final published version of the proceedings is available on IEEE Xplore.

Thermal Image Super-Resolution Challenge Results - PBVS 2023

Rafael E. Rivadeneira, Angel D. Sappa, Boris X. Vintimilla,
Chenyang Wang, Junjun Jiang, Xianming Liu, Zhiwei Zhong,
Dai Bin, Li Ruodi, Li Shengye

Abstract

This paper presents the results of two tracks from the fourth Thermal Image Super-Resolution (TISR) challenge, held at the Perception Beyond the Visible Spectrum (PBVS) 2023 workshop. Track-1 uses the same thermal image dataset as previous challenges, with 951 training images and 50 validation images at each resolution. In this track, two evaluations were conducted: the first consists of generating a SR image from a HR thermal noisy image downsampled by four, and the second consists of generating a SR image from a mid-resolution image and compare it with its semi-registered HR image (acquired with another camera). The results of Track-1 outperformed those from last year's challenge. On the other hand, Track-2 uses a new acquired dataset consisting of 160 registered visible and thermal images of the same scenario for training and 30 validation images. This year, more than 150 teams participated in the challenge tracks, demonstrating the community's ongoing interest in this topic.

1. Introduction

Image super-resolution (SR) techniques aim to generate a high-resolution (HR) image from a given low-resolution (LR) image. Nowadays, most of the approaches in the literature are deep learning-based solutions, where a downsampled HR image is used as input, augmented with noise and blur, and then used to train the network. While most of these methods have been applied to the visible spectrum, the demand for thermal images in various applications requires techniques that operate in the thermal image domain.

A standard benchmark for evaluating contributions was introduced at the PBVS 2020 workshop through a Thermal Image Super-Resolution (TISR) challenge. The success of these challenges led the way for the fourth TISR challenge in the framework of the PBVS 2023 workshop, which features two tracks. While Track-1 remains unaltered, Track-2 is based on a recently acquired cross-spectral sensor dataset. The dataset includes registered visible and thermal image pairs captured during daylight conditions of a given scene.

Figure 1. A montage of two thermal images captured from the same point of view but with different cameras at different resolutions. The $(left)$ image is a crop from the MR image, while the $(right)$ image is a crop from the HR image [11].

As mentioned above, TISR 2023 challenge[1] has two tracks. Track-1 has two evaluation approaches, like in the previous year challenge [13]. Evaluation 1 consists in generating a $\times 4$ SR image from a noise and downsampled image from the HR camera. Evaluation 2 consists in generating a $\times 2$ SR image from the MR camera (Axis Q2901-E) to be compared with its corresponding semi-registered images obtained from a HR camera (FLIR FC-632O); this second evaluation must tackle two problems, generating SR images acquired with a different camera as well as mapping images from different domains. On the other hand, Track-2 consists in obtaining super-resolution images at $\times 8$ by using a HR visible image of the same scenario as a guidance for the LR thermal image.

The obtained results indicate the growing interest of the

Rafael E. Rivadeneira* (rrivaden@espol.edu.ec), Boris X. Vintimilla (boris.vintimilla@espol.edu.ec)* and Angel D. Sappa*+ are the TISR Challenge - PBVS 2023 organizers, while the other authors participated in the challenge.
*Escuela Superior Politécnica del Litoral, ESPOL, Guayaquil, Ecuador.
+Computer Vision Center, Campus UAB, 08193 Bellaterra, Barcelona, Spain.
Appendix A contains the authors' teams and affiliations.

[1] https://pbvs-workshop.github.io/challenge.html

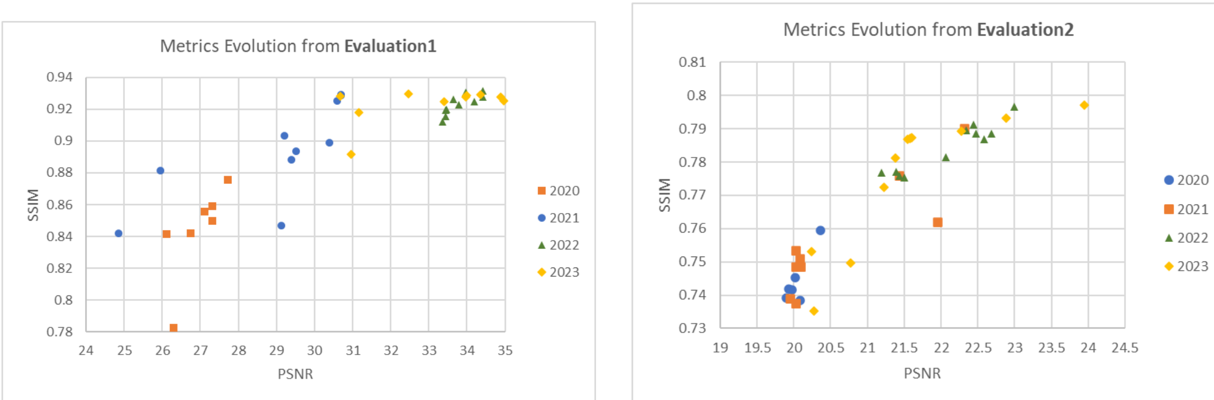


Figure 2. Metrics evolution through all challenges (Track-1).

community in the thermal image SR problem, with metric values increasing each year, as illustrated in Figure 2 for Track-1. The manuscript is organized as follows. Section 2 provides an introduction to the challenge objectives and datasets, while Section 2.3 presents a summary of the results obtained by the different teams. Then, Section 3 offers a brief description of the top approaches. The conclusion is presented in Section 4, followed by an appendix containing information on the teams.

2. TISR 2023 Challenge

The objective of the TISR 2023 challenge is to introduce different approaches for solving the thermal image SR problem and also compare these solutions using the previous year's benchmark. Additionally, this year, a new cross-spectral dataset is introduced to tackle the guided thermal images SR problem and foster further research in this field.

2.1. Thermal Image Datasets

The current challenge utilizes two datasets, one for each track. Track-1 dataset, introduced in [11], is used as in previous TISR challenges (i.e., [12], [14], [13]). This dataset contains 1021 thermal images taken with three different thermal cameras in various lighting conditions, resolution and scenes. Cameras were positioned on a rig to reduce the baseline distance between the optical axis, resulting in nearly registered images. A mosaic made up of images from the MR and HR cameras is presented in Fig. 1.

On the other hand, Track-2 uses the new acquired dataset, which contains visible and thermal registered pairs of images of the same scene taken in daylight conditions. These images have clear edges and are of good quality, making them suitable for training. The dataset includes 200 pairs of images captured using Balser and TAU2 cameras with different resolutions. The images were registered using the Elastix [5] algorithm, obtaining pairs of thermal and visible images with a resolution of 640×480 pixels. The dataset is split up into training, validation, and testing image pair sets, with 160, 30, and 10 images, respectively. No noise has been added to the downsampled images. The HR visible spectrum images serve as a guidance for the LR thermal image to generate a super-resolved HR thermal image. Examples of this dataset can be seen in Fig. 3.

2.2. Evaluation Methodology

The evaluation methodology for Track-1 is the same that the one used in the PBVS 2022 [13] challenge. All team contributions are assessed based on the obtained mean values of peak signal-to-noise ratio (PSNR) and structural similarity (SSIM) measures. Two types of evaluations are conducted, as mentioned abve. In the initial process, a set of 10 LR images obtained from a HR camera are evaluated. Gaussian noise ($\sigma = 10\%$) is introduced, and the HR image is then downsampled by a factor of $\times 4$. Figure 4 (a) provides an illustration of this first evaluation process.

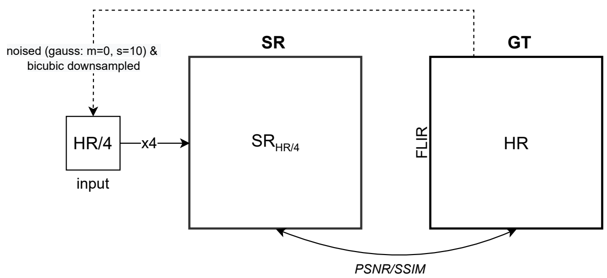
Another set of 10 SR images obtained by a $\times 2$ scale factor from the given MR images is evaluated in the second process. These 10 SR images are evaluated with respect to the corresponding HR GT images (acquired from a different camera, with the same resolution as the computed SR). Feature point-based registration is used to align the images. The evaluation on PSNR and SSIM is performed on 80% of the central cropped region of the image. Figure 4 (b) illustrates this second evaluation process.

The evaluation for Track-2 also uses the mean PSNR and SSIM metrics obtained on a set of 10 LR images (downsampled by a factor of $\times 8$), where no noise has been added. Figure 5 provides an illustration of this track evaluation process.

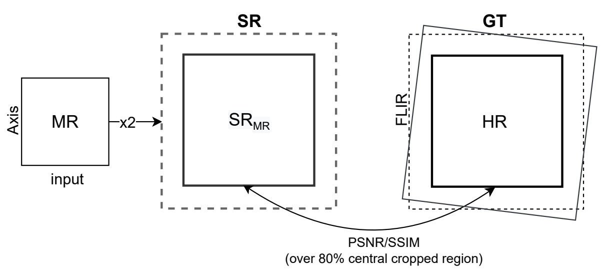
2.3. Challenge Results

The top three results from each participating team for each track are described below. For Track-1, 28 teams reached the final testing phase from the 87 initially registered teams.

Figure 3. Illustrations of the cross-spectral dataset, thermal and visible registered images used in Track-2.



(a) First evaluation process on a set of LR images downsampled by a factor of $\times 4$ with added noise.



(b) Second evaluation process on a set of MR to HR images ($\times 2$).

Figure 4. Evaluations processes for Track-1.

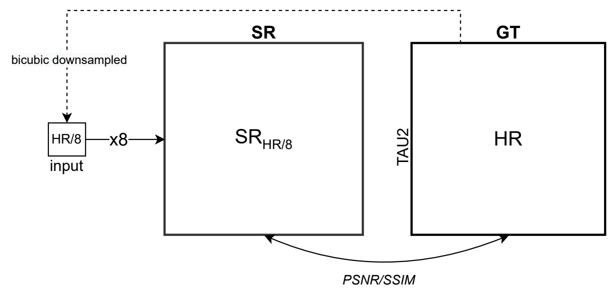


Figure 5. Illustration of the evaluation process for Track-2 on a set of LR images downsampled by a factor of $\times 8$ with no added noise.

Team Approach Track-1	Evaluation 1		Evaluation 2	
	$\times 4$		$\times 2$ (MR to HR)	
	PSNR	SSIM	PSNR	SSIM
AC	34.35	0.9294	**23.95**	**0.7970**
AIR	34.88	0.9279	21.38	0.7812
ANT INS	33.97	0.9276	22.89	0.7932
APTX4869	31.17	0.9178	22.28	0.7893
STAIA	**34.96**	0.9253	20.24	0.7532
TSR	32.46	**0.9296**	20.12	0.7493
PBVS 2022	34.42	**0.9316**	23.00	0.7966

Table 1. Track-1 tops average results for each evaluation of the 2023 TISR challenge (see Section 2.2 for more details). Bold and underline values correspond to the best- and second-best results, respectively for each evaluation. PBVS 2022 shows the highest results obtained for each metric in the previous edition of the challenge.

Table 1 displays the average results (PSNR and SSIM) on the testing images for each team in the two evaluations. As it can be seen, this year's results slightly improved upon the previous year's best results, except for the SSIM of Evaluation 1, where none of the teams achieved a better result than in last year's competition.

On the other hand, in Track-2, 23 teams reached the final testing phase, from the 74 initially registered teams. Table 2 displays the average results (PSNR and SSIM) on the testing images for each team. The CodaLab Competition [9] webpage (Track-1[2]; Track-2[3]) provides further quantitative

[2]https://codalab.lisn.upsaclay.fr/competitions/9649

[3]https://codalab.lisn.upsaclay.fr/competitions/

Team Approach Track-2	×8	
	PSNR	SSIM
AIR	28.74	0.8568
ANT INS	29.41	0.8727
GUIDEDSR	**31.04**	**0.9294**
TU-PAC	28.78	0.8582
SwinIR	27.68	0.8098

Table 2. Track-2 top average results of the 2023 TISR challenge (see Section 2.2 for more details). Bold and underline values correspond to the best- and second-best results. SwinIR results corresponds to a non-guided approach used as a baseline for the comparisons; this non-guided approach reached the best results in the TISR challenge of last year [13].

results that can be analyzed to gain a better understanding of the competition's overall performance.

3. Proposed Approaches and Teams

This section presents a brief description of the approaches proposed by the teams that achieve the highest scores in each metric of the evaluations for each tracks. Illustrations of the architectures for the best results are provided. Teams are alphabetically listed.

3.1. Track-1: AC

In Track-1, for Evaluation 1, the AC team adopts SwinIR [6] as the backbone. For Evaluation 2, the main challenge is the unaligned MR-HR pair. One solution is to align HR with MR. However, MR and HR belong to different resolution spaces, and direct alignment between them is inaccurate. Thus, as illustrated in Fig. 6, the AC team feeds the 2D coordinate map of MR and HR into a super-resolution model M_1 [19] to generate a coarse super-resolved result $I^{\mathrm{SR}}_{\mathrm{In}}$. Then, it uses pretrained PWCNet [16] to estimate the optical flow from HR to $I^{\mathrm{SR}}_{\mathrm{In}}$, generating aligned HR $I^{\mathrm{HR}}_{\mathrm{Warp}}$. Finally, the AC-Team feeds MR into a super-resolution model M_2 and supervises it with $I^{\mathrm{HR}}_{\mathrm{Warp}}$ to reconstruct the final super-resolved result I^{SR}.

It should be noted that the parameters of PWCNet are fixed. During the inference phase, only M_2 is used to recover high-quality images. The AC team adopts the popular Pytorch and trains the model on two NVIDIA V100 GPUs (128G RAM memory) for 60 epochs (approximately four hours). The model is optimized by ADAM with a learning rate of 5e-4. The quantitative results show that the AC-Team achieves 34.35 PSNR & 0.9294 SSIM in Evaluation 1 and best results in Evaluation 2 with **23.95** PSNR & **0.7970** SSIM for Track-1.

The code is available at https://github.com/wcy-cs/PBVS2023_TISR_Track1

3.2. Track-1: AIR

Recent advances in low-level computer vision research have shown that Transformer-based networks achieve impressive performance [4]. To address the limitations of conventional Transformer-based networks that use a limited spatial range of input information, the Hybrid Attention Transformer (HAT) [2] network has been proposed. The HAT network combines channel attention and self-attention schemes to activate more input pixels for reconstruction, resulting in state-of-the-art performance. Building upon the HAT network, this team proposes a **HAT-variant** that places additional emphasis on spatial information. First, the proposed approach replaces the existing *Residual Channel Attention Block (RCAB)* with a *Residual Channel-Spatial Attention Block (RCSAB)* to preserve spatial information more effectively. Specifically, the average pooling layer in the *RCAB* is removed to maintain the spatial resolution of the input feature maps. Furthermore, the use of *RCSAB* does not significantly increase the computational cost compared to *RCAB*, while enabling the creation of more detailed attention maps. Second, in order to enhance full-resolution training and increase the batch size, this team reduced the number of *Residual Hybrid Attention Groups (RHAGs)* from 6 to 5.

All reported implementations were based on the PyTorch framework, while the proposed approaches were conducted using a 16-Core CPU, 2 × V100 GPUs, and 64 GB of RAM for approximately four days. This team used early stopping and an initial learning rate of 0.0001 with the RAdam optimizer [7]. Additionally, low-quality (LQ) images were created by applying JPEG compression (quality=95) and Gaussian noise ($\sigma = \sqrt{10}$). The proposed network was trained using full-resolution high-quality (HQ) images from the challenge training and validation dataset, with a batch size of 4. All images were augmented with random horizontal and vertical flips. The pixel-wise *MSE* loss function was computed for a pair of reconstructed images obtained from **HAT-variant** and high-quality (HQ) images. The study results were quantified using PSNR and SSIM metrics. The quantitative results show that the AIR team achieves 34.88 PSNR and 0.9279 SSIM on Evaluation 1, and 21.38 PSNR and 0.7812 SSIM on Evaluation 2 for Track-1.

3.3. Track-1 & 2: ANT INS

Inspired by Channel Split Convolutional Neural Network(ChaSNet) [10] and Swin Transformer [6], ANT INS team designed a Transformer and Convolution Parallel-based Super Resolution Network (TCP-SRNet). The Channel Split Convolutions are designed to extract local features from images, such as edges, corners, and textures, which are combined in higher layers to form more complex representations

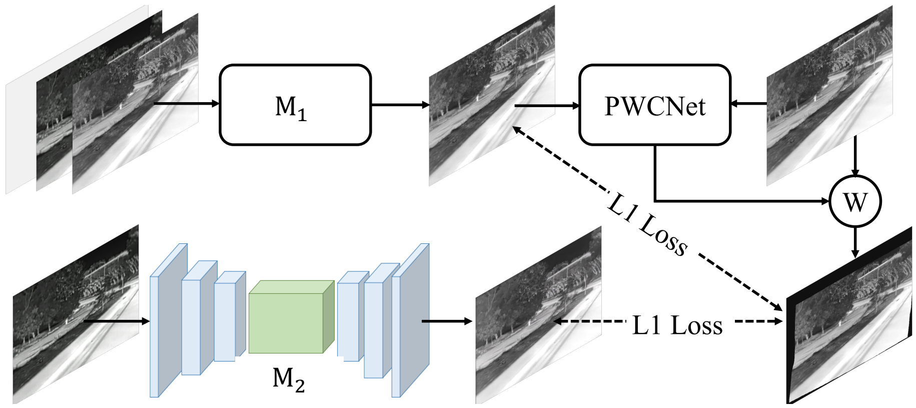


Figure 6. Architecture proposed by AC team, Track-1.

of objects. The Swin Transformers can capture the relationships between different parts of the image, such as the spatial arrangement of objects. TCP-SRNet combines both features to improve the performance of image super-resolution.

For Track-1, in Evaluation 1, TCS-SRNet were trained with an upscaling factor of $\times 4$ with L1 loss, and the inputs are down-sampled with a factor $\times 4$ of the HR images. In Evaluation 2, the inputs are semi-matched, where TCS-SRNet were trained with an upscaling factor of $\times 2$ with L1 loss. Secondly, TCS-SRNet were trained with an upscaling factor of $\times 2$ with L1 loss, Least Squared GAN (LSGAN) loss [8] and SSIM loss. The average of outputs of these two models are the final result for these models.

For Track-2, the process enhances image resolution by splitting LR and HR images into two feature maps, processing them through Transformer and Channel Split Blocks, and using pixel-shuffle operators to increase spatial resolution. The HR visible image is also passed through the convolution layer with a kernel size of 3×3 and 2N channels. The blocks used are Swin-based and Channel Split Convolutional Blocks, similar to those in ChaSNet.

The quantitative results show that the ANT INS team achieves 33.97 PSNR and 0.9276 SSIM in Evaluation 1, and 22.89 PSNR & 0.7932 SSIM in Evaluation 2 for Track-1. Moreover, in Track-2, the team achieves 29.41 PSNR & 0.8727 SSIM.

3.4. Track-1: ATPX4869

The architecture of the network is based on SwinIR [6], with window size as 8, and depth of each Swin Transformer layer as all 6. The number of attention heads of each layer is set to 6, and patch embedding dimension is 180. SwinIR is mainly based on attention operations to extract features of the input image and reconstruct with them. The basic block for SwinIR is Residual Swin Transformer Block, which combines the capability of long-term dependency modeling of self-attention modules and the shift-invariance of convolution blocks. This architecture makes SwinIR a strong baseline for the thermal super-resolution tasks. SwinIR is used with the same settings for both $\times 2$ and $\times 4$ tasks (only upscale is changed).

The parameter settings for $\times 4$ task (noisy LR to HR) is as follows: training loss is weighted sum of L1 loss and MS-SSIM loss:

$$loss_{eval1}(\hat{y}, y) = \mathrm{L1Loss}(\hat{y}, y) + \lambda(1 - \mathrm{MSSSIM}(\hat{y}, y))$$

where λ is set to 0.5 in the experiments. Pretrained model of SwinIR from the BasicSR [18] was used. Data augmentation (flip, rotation) is used and cropped to 128×128 (HR patch size), AdamW optimizer with learning rate 2e-4 is applied for training. The network is trained for 100k iteration with batch size 12, with learning rate halved in 60k, 80k and 95k iterations. In addition, test time augmentation is applied by taking average of outputs from h-flipped, v-flipped and original LR images.

For $\times 2$ task (MR to HR), one of the main obstacles for evaluation 2 task is the misalignment of MR and HR images. Firstly, the ECC Maximization method [3] is used to align the MR to downsampled HR images. The network is trained on the aligned dataset for 50k iterations with batch size 8. AdamW with learning rate 1e-4 and multistep scheduler which halves learning rate in 10k and 20k are applied in this experiment. Loss function is as follows:

$$loss_{eval2}(\hat{y}, y) = \mathrm{TruncL1Loss}(\hat{y}, y) + w(1 - \mathrm{SSIM}(\hat{y}, y))$$

where TruncL1Loss is a truncated absolute mean error, which only penalize the error under a certain threshold:

$$\mathrm{TruncL1Loss}(\hat{y} - y) = \begin{cases} thr, & ||\hat{y} - y|| \geq thr \\ ||\hat{y} - y||, & ||\hat{y} - y|| < thr \end{cases}$$

This loss is chosen to tackle with meaningless large errors caused by possible misaligned pixels. thr is set to 0.05, w is set to 0.5. Test time augmentation same as $\times 4$ task.

Python 3.7 and Pytorch 1.10.0 with CUDA 11.3 is used in the experiments. All trainings are done on 1 Nvidia GeForce RTX 3090 GPU card (24G) of Ubuntu 18.04 server. The quantitative results show that the APTX4869 Team achieves 31.17 PSNR & 0.9178 SSIM on Evaluation 1 and 22.28 PSNR & 0.7893 SSIM on Evaluation 2 for Track-1.

The code is available at https://github.com/jzsherlock4869/TISR_APTX4869

3.5. Track-2: GUIDED SR

This team proposes a model that super-resolves a LR thermal image with a HR RGB image as guidance. Considering that the RGB and thermal images are captured by different imaging pipelines, the team proposes a two-stream network to enhance the LR thermal image. The network architecture is presented in Fig. 7. Specifically, the team first feeds the LR and RGB images into a shallow feature extraction layer to extract shallow features respectively. Then, the extracted features are concatenated and fed into feature fusion layers to fuse multi-modal information. The feature fusion layers are comprised of cascaded NAF Blocks [1]. Finally, the team adopts an HR image reconstruction layer to reconstruct the super-resolution result.

The training process has two steps. In the first step, the network is guided by using L1 loss. In the second step, the network, pre-trained in the first step, is fine-tuned by adopting MSE loss. The team conducted experiments on two NVIDIA 3090 GPUs for two days using the Pytorch framework. The batch-size and patch-size were 8 and 32×32, respectively. The quantitative results show that the GuidedSR team achieves the best results in both metrics, which are **31.04** PSNR & **0.9036** SSIM for Track-2.

Source code can be found in https://github.com/zhwzhong/Guided-SR.

3.6. Track-1: STAIA

Considering the impressive results achieved by Transformer-based deep neural networks in image super-resolution, the HAT model is used as shown in Fig. 8 following HAT [2]; the model parameters are initialized using the ImageNet pretrained model. The batch size is set to 4/gpu, and the high-resolution image block size is 256×256. The original HR image is added with Gaussian noise and downsampled to a LR image as input, and random flip and rotation are used as data augmentation strategies. The model is trained for about 60000 iterations. It was found that using cubic downsampling in OpenCV can achieve higher PSNR (SRx4:34.96) on the test set, and using bicubic downsampling in Pytorch can achieve higher SSIM (SRx4:0.9284). This team directly uses the model structure in HAT, for detailed structure, please refer to HAT [2].

This team uses 4 Nvidia Tesla V100 GPUs and 20-core CPUs to train the neural network model, using the Python programming language and the Pytorch deep learning framework, and the training time is 1 day and 3 hours. The quantitative results show that the team achieves **34.96** PSNR & 0.9253 SSIM on Evaluation 1, and 20.24 PSNR & 0.7532 SSIM on Evaluation 2 for Track-1.

The code is available at https://github.com/daicver/HAT_TISR

3.7. Track-1: TSR

This team recover the HR thermal image from the given LR thermal image by first upsampling the LR image by Bicubic interpolation and then feeding the upsampled results into the network. As shown in Fig. 9, the network consists of three parts, i.e., shallow feature extraction, feature enhancement and HR image reconstruction. Finally, the output of the network is added to the upsampled LR image, generating the super-resolved result. In their model, they employ the NAFBlock [1] as the basic block due to its powerful representation ability.

To improve performance, the model is first pre-trained using ×2 LR-HR pairs and then fine-tuned using ×4 LR-HR pairs. The team chose L1 loss as their loss function. The team conducted experiments using the Pytorch framework on two NVIDIA 3090 GPUs for three days. The batch size and patch size were set to 8 and 32×32, respectively. The quantitative results show that the TSR team achieves 32.46 PSNR & **0.9292** SSIM on Evaluation 1 and 20.12 PSNR & 0.7493 SSIM on Evaluation 2 for Track-1.

Source code can be found in https://github.com/zhwzhong/TSR.

3.8. Track-2: TU-PAC

Guided super-resolution with misaligned images poses two key challenges: effectively using the guide image to upsample the thermal image and avoiding misaligned features for guidance. To address these issues, this team proposes an Attention-based Pixel Adaptive Convolution (APAC) layer based on PACT [15] to upsample thermal images using misaligned guide images. The network is composed of Encoder, Guide, and Decoder branches. Each APAC block refines the guide features using Channel and Spatial Attention to suppress the impact of misaligned features. The refined guide features and thermal features are fed to a PAC^T layer to generate upsampled thermal features for the next layer. Finally, the upsampled features are passed through two convolution layers, and a skip connection adds the bilinear upsampled input to the model output to generate a sharp thermal image.

In addition to APAC, several modifications are proposed to improve the baseline efficacy of the model: SSIM loss is preferred over pixel wise similarity measures such as L1 and MSE to help the model overcome the pixel level misalignment issue. Training time augmentations such as random rotation, horizontal, and vertical flips were used to prevent

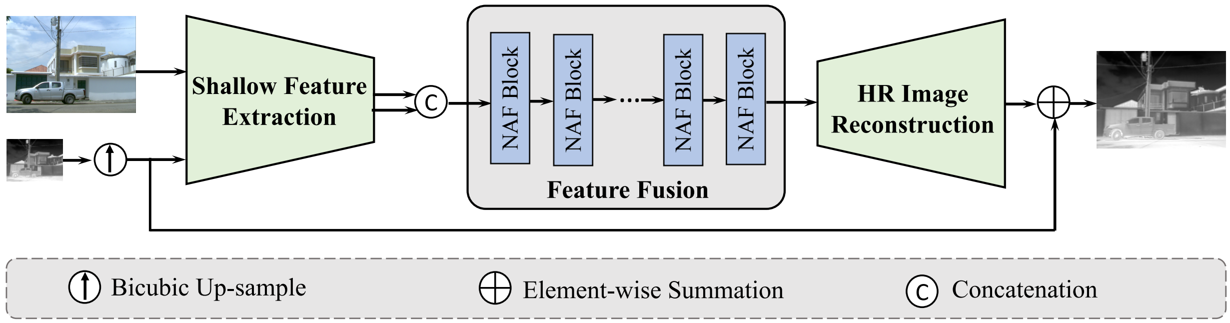


Figure 7. Architecture proposed by GUIDEDSR team, Track-2.

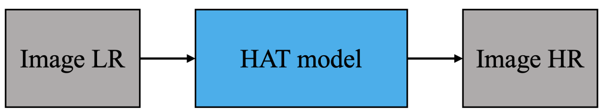


Figure 8. Architecture proposed by STAIA team, Track-1.

overfitting on the small dataset. Test time augmentations introduced in [17] were used to generate final outputs. Additionally, a new test time augmentation is introduced: The guide image is Gaussian blurred by a 3×3 kernel to reduce the effect of noise in the guide image. This augmentation improves both PSNR and SSIM. The model is very efficient with only 1M parameters, trained on a single V100 GPU using PyTorch. The quantitative results show that the TU-PAC team achieves 28.78 PSNR & 0.8582 SSIM for Track-2.

4. Conclusion

This paper provides a summary of the techniques proposed by each team that reached to the final validation phase and submitted their contributions for the Thermal Image Super-Resolution Challenge - PBVS 2023, where two tracks were considered. A new cross-spectral dataset is acquired for Track-2. All the presented approaches are based on deep learning algorithms, using various CNN architectures. This event marks the fourth edition of the challenge focused on thermal image SR, with more participants than in previous editions (87 teams registered for Track-1 and 74 teams registered for Track-2). As a general conclusion, the results for Track-1 indicate that the limit of performance seems to have been reached. On the other hand, regarding the results from Track-2, it can be said that guided approaches improve the baseline results obtained with a non-guided SwinIR, which was used as a reference when considering large super-resolution scales (×8). This dataset will be used as a benchmark in future editions of the challenge. The challenge provides a valuable platform for researchers to collaborate and exchange ideas, leading to significant advancements in the field of thermal image SR.

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 by MCIN/AEI/10.13039/501100011033 and by "ERDF A way of making Europe"; the "CERCA Programme / Generalitat de Catalunya"; and the ESPOL project CIDIS-12-2022. The first author has been supported by Ecuador government under a SENESCYT scholarship contract.

Appendix A. Teams Information

The organization team acknowledge the participants and utilize edited versions of top-performing team submissions to provide additional method explanations.

TISR 2023 organization team:

Members: Rafael E. Rivadeneira[1] (rrivaden@espol.edu.ec), Boris X. Vintimilla[1] and Angel D. Sappa[1,2]

Affiliations:

[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, Campus UAB, 08193 Bellaterra, Barcelona, Spain.

Top Participant Teams:

AC

Members: Chenyang Wang (wangchy02@hit.edu.cn), Junjun Jiang and Xianming Liu.

Affiliation: Harbin Institute of Technology, China.

AIR

Members: Dogun Kim (dgkimk@hanwha.com), Jin Kim and Dongyeon Kang.

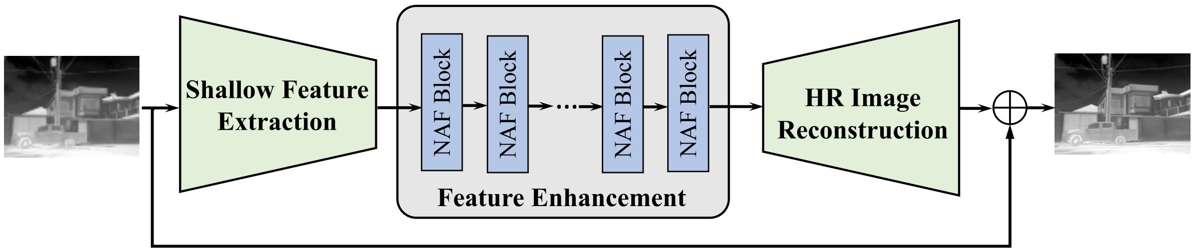


Figure 9. Architecture proposed by TSR team, Track-1.

Affiliation: Hanwha Systems, Seoul, Republic of Korea.

ANT INS

Members: Bo Yan (lengyu.yb@antgroup.com), Zhuang Li, Yadong Li and Hongbin Wang.

Affiliation: Ant Group, China.

APTX4869

Members: Zhuang Jia (jiazhuang@xiaomi.com).

Affiliation: Xiaomi, China

GUIDED SR

Members: Zhiwei Zhong (zhwzhong.hit@gmail.com), Xianming Liu and Junjun Jiang.

Affiliation: Harbin Institute of Technology, China.

STAIA

Members: Dai Bin (daibin2@sensetime.com), Li Ruodi and Li Shengye.

Affiliation: SenseTime, China.

TSR

Members: Zhiwei Zhong (zhwzhong.hit@gmail.com), Xianming Liu and Junjun Jiang.

Affiliation: Harbin Institute of Technology, China.

TU-PAC

Members: Sai Kumar Reddy Manne (manne.sa@northeastern.edu), Michael Wan and Sarah Ostadabbas.

Affiliation: Northeastern University, United States.

References

[1] Liangyu Chen, Xiaojie Chu, Xiangyu Zhang, and Jian Sun. Simple baselines for image restoration. In *Computer Vision–ECCV 2022: 17th European Conference, Tel Aviv, Israel, October 23–27, 2022, Proceedings, Part VII*, pages 17–33. Springer, 2022. 6

[2] X Chen, X Wang, J Zhou, and C Dong. Activating more pixels in image super-resolution transformer. arxiv 2022. *arXiv preprint arXiv:2205.04437*, 2022. 4, 6

[3] Georgios D Evangelidis and Emmanouil Z Psarakis. Parametric image alignment using enhanced correlation coefficient maximization. *IEEE transactions on pattern analysis and machine intelligence*, 30(10):1858–1865, 2008. 5

[4] Salman Khan, Muzammal Naseer, Munawar Hayat, Syed Waqas Zamir, Fahad Shahbaz Khan, and Mubarak Shah. Transformers in vision: A survey. *ACM computing surveys (CSUR)*, 54(10s):1–41, 2022. 4

[5] Stefan Klein, Marius Staring, Keelin Murphy, Max A Viergever, and Josien PW Pluim. Elastix: a toolbox for intensity-based medical image registration. *IEEE transactions on medical imaging*, 29(1):196–205, 2009. 2

[6] Jingyun Liang, Jiezhang Cao, Guolei Sun, Kai Zhang, Luc Van Gool, and Radu Timofte. Swinir: Image restoration using swin transformer. In *Proceedings of the IEEE/CVF international conference on computer vision*, pages 1833–1844, 2021. 4, 5

[7] Liyuan Liu, Haoming Jiang, Pengcheng He, Weizhu Chen, Xiaodong Liu, Jianfeng Gao, and Jiawei Han. On the variance of the adaptive learning rate and beyond. *arXiv preprint arXiv:1908.03265*, 2019. 4

[8] Xudong Mao, Qing Li, Haoran Xie, Raymond YK Lau, Zhen Wang, and Stephen Paul Smolley. Least squares generative adversarial networks. In *Proceedings of the IEEE international conference on computer vision*, pages 2794–2802, 2017. 5

[9] Adrien Pavao, Isabelle Guyon, Anne-Catherine Letournel, Xavier Baró, Hugo Escalante, Sergio Escalera, Tyler Thomas, and Zhen Xu. Codalab competitions: An open source platform to organize scientific challenges. *Technical report*, 2022. 3

[10] Kalpesh Prajapati, Vishal Chudasama, Heena Patel, Anjali Sarvaiya, Kishor P Upla, Kiran Raja, Raghavendra Ramachandra, and Christoph Busch. 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, 2021. 4

[11] Rafael E Rivadeneira, Angel D Sappa, and Boris X Vintimilla. Thermal image super-resolution: A novel architecture and dataset. In *Proc. of the International Joint Conference on Computer Vision, Imaging and Computer Graphics Theory and Applications*, pages 111–119, 2020. 1, 2

[12] Rafael E Rivadeneira, Angel D Sappa, Boris X Vintimilla, Lin Guo, Jiankun Hou, Armin Mehri, Parichehr Behjati Ardakani, Heena Patel, Vishal Chudasama, Kalpesh Prajapati, Kishor P Upla, Raghavendra Ramachandra, Kiran Raja, Christoph Busch, Feras Almasri, Olivier Debeir, Sabari Nathan, Priya Kansal, Nolan Gutierrez, Bardia Mojra, and William J Beksi. Thermal image super-resolution challenge-PBVS 2020. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops*, pages 96–97, 2020. 2

[13] Rafael E Rivadeneira, Angel D Sappa, Boris X Vintimilla, Jin Kim, Dogun Kim, Zhihao Li, Yingchun Jian, Bo Yan, Leilei Cao, Fengliang Qi, et al. Thermal image super-resolution

challenge results-pbvs 2022. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 418–426, 2022. 1, 2, 4

[14] Rafael E Rivadeneira, Angel D Sappa, Boris X Vintimilla, Sabari Nathan, Priya Kansal, Armin Mehri, Parichehr Behjati Ardakani, Anurag Dalal, Aparna Akula, Darshika Sharma, et al. Thermal image super-resolution challenge-pbvs 2021. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 4359–4367, 2021. 2

[15] Hang Su, Varun Jampani, Deqing Sun, Orazio Gallo, Erik Learned-Miller, and Jan Kautz. Pixel-adaptive convolutional neural networks. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 11166–11175, 2019. 6

[16] Deqing Sun, Xiaodong Yang, Ming-Yu Liu, and Jan Kautz. Pwc-net: Cnns for optical flow using pyramid, warping, and cost volume. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 8934–8943, 2018. 4

[17] Radu Timofte, Rasmus Rothe, and Luc Van Gool. Seven ways to improve example-based single image super resolution. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 1865–1873, 2016. 7

[18] Xintao Wang, Liangbin Xie, Ke Yu, Kelvin C.K. Chan, Chen Change Loy, and Chao Dong. BasicSR: Open source image and video restoration toolbox. https://github.com/XPixelGroup/BasicSR, 2022. 5

[19] Zhilu Zhang, Haolin Wang, Ming Liu, Ruohao Wang, Wangmeng Zuo, and Jiawei Zhang. Learning raw-to-srgb mappings with inaccurately aligned supervision. In *ICCV*, pages 4348–4358, 2021. 4