Thermal Image Super-Resolution Challenge Results - PBVS 2024

Rafael E. Rivadeneira and Angel D. Sappa,
Chenyang Wang and Junjun Jiang,
Zhiwei Zhong, Peilin Chen and Shiqi Wang

Abstract

This paper outlines the advancements and results of the Fifth Thermal Image Super-Resolution challenge, hosted at the Perception Beyond the Visible Spectrum CVPR 2024 workshop. The challenge employed a novel benchmark cross-spectral dataset consisting of 1000 thermal images, each paired with its corresponding registered RGB image. The challenge featured two tracks: Track-1 focused on Single Thermal Image Super-Resolution with an $\times 8$ upscale factor, while Track-2 extended its evaluation to include both $\times 8$ and $\times 16$ scaling factors, utilizing high-resolution RGB images to guide the super-resolution process for low-resolution thermal images. The participation of over 175 teams highlights the research community's strong engagement and dedication to enhancing image resolution techniques across both single and cross-spectral methodologies. This year's challenge sets new benchmarks and provides valuable insights into future directions for research in thermal image super-resolution.

1. Introduction

The field of image super-resolution (SR), particularly focusing on enhancing the resolution of thermal images, has seen notable advancements in recent years. The primary approach involves using deep learning techniques to convert low-resolution (LR) images into high-resolution (HR) counterparts. These methods typically involve training on downsampled HR images that have been artificially augmented with noise and blur to improve the network's ability to enhance image quality. Despite the prevalence of such methods for visible spectrum images, there is a growing need for specialized SR techniques tailored for the thermal spectrum due to its wide range of applications.

Rafael E. Rivadeneira* (rrivaden@espol.edu.ec) and Angel D. Sappa*+ are the TISR Challenge - PBVS 2024 organizers. The remaining authors are team members who achieved top results 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.

Figure 1. A montage of visible and thermal images, captured from the same point of view but with different cameras sensor [35].

Reflecting on the progression of this field, the Thermal Image Super-Resolution (TISR) challenge, first introduced at the Perception Beyond the Visible Spectrum (PBVS) CVPR 2020 workshop [31], has become a pivotal benchmark for evaluating advancements in thermal image SR. The success of these annual challenges, culminating in the fourth TISR challenge at PBVS 2023 [30], has prepared the way for ongoing innovation and benchmarking within this specialized area. Each challenge has progressively built upon the learning's and datasets of its predecessors, contributing to a rich foundation for current and future research. The dataset used in previous challenge is available and also the CodaLab (Track-1[1], Track-2[2]) is open for benchmark comparisons.

The 2024 TISR challenge maintains the structure of previous years challenges, with two different tracks: Track-1 focuses on single-image super-resolution (SISR) with an $\times 8$ upscale factor. Track-2, on the other hand, includes both $\times 8$ and $\times 16$ scaling factors and introduces the concept of using RGB images as a guide in the super-resolution of LR thermal input images. This dual-track approach not only sustains the traditional focus on SISR but also expands the challenge

[1] https://codalab.lisn.upsaclay.fr/competitions/9649
[2] https://codalab.lisn.upsaclay.fr/competitions/9666

Figure 2. Illustrations of the cross-spectral dataset, thermal and visible registered images [35].

to explore the potential of cross-spectral methodologies in enhancing SR techniques. This year's challenge, make use of a new registered cross-spectral dataset using visible and thermal cameras sensor, a mosaic is presented in Fig. 1.

The structure of the manuscript is outlined as follow. Section 2 introduces the aims of the challenge and the used datasets. This is followed by Section 2.3, which provides an overview of the outcomes achieved by the teams in participating in the two tracks. Subsequently, Section 3 gives a concise overview of the leading methods proposed in the competition. The manuscript concludes with Section 4, and supplementary details about the participating teams are included in the appendix.

2. TISR 2024 Challenge

Similar to past challenges (i.e., [31], [34], [32], [30]), the TISR 2024 challenge aims to showcase a variety of methods for addressing the thermal image super-resolution issue, serving as a benchmark in the field. Furthermore, this year introduces a novel cross-spectral dataset designed to address the challenge of guided thermal image SR, encouraging additional exploration and research within this domain.

2.1. Thermal Image Datasets

This year's challenge utilizes an innovative dataset known as CIDIS (Cross-spectral Image Dataset for Image Super-resolution) [35], featuring 1000 pairs of visible and thermal images. These images were captured using a Basler camera for the visible spectrum and a FLIR TAU2 camera for the thermal spectrum, introducing new challenges and opportunities for Super-resolution technology. The dataset provides a comprehensive collection of 640×480 resolution images for both visible and thermal spectrums, as depicted in Fig. 2. It includes registered pairs of visible and thermal images taken in daylight conditions, which were precisely aligned using leading registration techniques such as Elastix [24], Imregister [25], LightGlue [20], and Nemar [1] models.

The dataset comprises 1000 registered pairs of images, split into 700 images for training, 200 for validation, and 100 for testing. The testing set is further divided into 20 images for Track-1 testing, and 40 images each for Track-2 Evaluation 1 and Track-2 Evaluation 2. These high-quality images feature distinct edges, making them ideal for SR model development and training. As mentioned above, the dataset's images were captured using Basler and FLIR TAU2 cameras, both offering different resolutions, but for this dataset, all images were standardized to a resolution of 640x480 pixels. Visible spectrum images serve as a guidance for enhancing a low-resolution thermal images, aiming to generate super-resolved HR thermal images. Examples from this dataset are showcased in Fig. 2.

2.2. Evaluation Methodology

The evaluation methodology for both tracks follows the same protocol as that used in the PBVS 2023 challenge [30]. Contributions from all teams are assessed based on the average values of two key metrics: peak signal-to-noise ratio (PSNR) and structural similarity index measure (SSIM) [39]. As previously mentioned, the challenge features two tracks, each with its distinct evaluation process utilizing the 100 image set aside for testing as follows. In the first track, a set of 20 LR images, derived from high-resolution camera captures, are considered. These images are not subjected to additional noise and are downsampled by a factor of $\times 8$. Figure 3 illustrates the evaluation process for Track-1.

For Track-2, the remaining set of 80 LR images is split into two groups, with one group being downsampled by a factor of $\times 8$ and the other by $\times 16$, to facilitate Evaluation 1 and Evaluation 2, respectively. Similar to Track-1, no noise is added to these downsampled images. Figure 4 depicts the evaluation process for Track-2.

This methodology ensures a standardized and fair comparison of the participating teams' approaches, allowing

for a comprehensive analysis of their effectiveness in enhancing the resolution of thermal images. By leveraging a combination of PSNR and SSIM metrics, the evaluation process aims to quantify both the fidelity and perceptual quality of the super-resolved images, offering insights into the advancements achieved in thermal image super-resolution technology through this challenge.

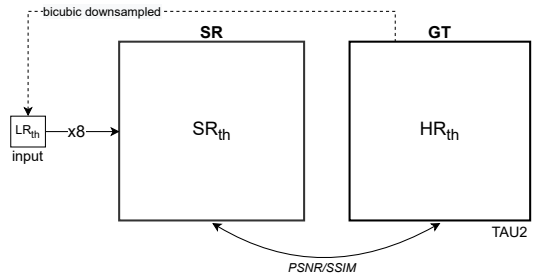

Figure 3. Illustration of the evaluations process for Track-1 on a set of LR images downsampled by a factor of $\times 8$ with no added noise.

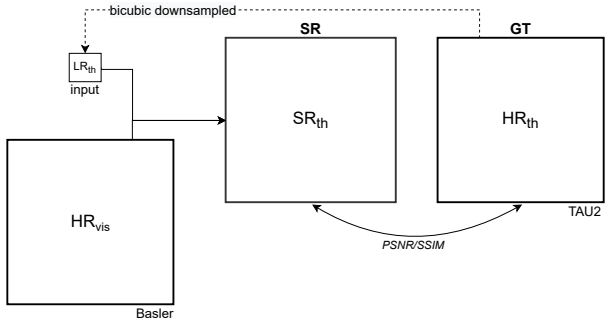

Figure 4. Illustration of the evaluations process for Track-2 on a set of LR images downsampled by a factor of $\times 8$ and $\times 16$ with no added noise and using the corresponding HR visible images as guidance.

2.3. Challenge Results

The top three results from each participating team for each track are detailed in the following. In Track-1, 19 teams progressed to the final testing phase out of the 113 teams that initially registered. Table 1 presents the average results (PSNR and SSIM) for the testing images across each team in the two evaluations. Figure 5 show qualitative results of Top1 result.

For Track-2, out of the initial 76 teams that registered, 16 advanced to the final testing stage. Table 2 showcases the average performance metrics (PSNR and SSIM) for the testing images across these teams, while Fig. 6 show qualitative result of Top1 result in both evaluation. Detailed quantitative outcomes, enhancing the understanding of the overall competition's performance, are accessible on the CodaLab Competition [27] webpage for both Tracks (Track-1[3]; Track-2[4]).

Team [# param.] **(Track-1: SINGLE)**	$\times 8$	
	PSNR	**SSIM**
AC-TSR $[132.98M]$	**27.52**	**0.8355**
CTYUN-AI $[20.08M]$	<u>27.48</u>	<u>0.8351</u>
HBNU $[21.20M]$	27.34	0.8322
HSC-SCA $[289.10M]$	<u>27.48</u>	0.8292

Table 1. Track-1 top average results for Single Image SR of the 2024 TISR challenge (see Section 2.2 for more details). Bold and underline values correspond to the best- and second-best results.

Team [# param.] **(Track-2: GUIDED)**	**Eval 1** $(\times 8)$		**Eval 2** $(\times 16)$	
	PSNR	**SSIM**	**PSNR**	**SSIM**
AIR $[3.04M]$	- - -	- - -	24.77	0.7878
GUIDEDSR $[600M]$	**31.52**	**0.9127**	**25.99**	**0.8266**
UMKC MCC $[12.17M]$	<u>30.05</u>	<u>0.8947</u>	<u>25.67</u>	<u>0.8167</u>
VISION IC $[3.30M]$	29.34	0.8824	24.69	0.7928

Table 2. Track-2 tops average results for Guided Thermal Image SR in each evaluation of the 2024 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.

3. Proposed Approaches and Teams

This section presents a overview of the methodologies employed by the teams that secured the top positions in each metric of the evaluations across both tracks. Visual representations of the architectures yielding the best outcomes are included. The teams are organized in alphabetical order.

3.1. Track-1: AC-TSR

The AC-TSR architecture first upscales the LR image to match the size of the high-resolution image, and then feeds it into the super-resolution network. As shown in Fig. 7, the network comprises three parts: feature extraction, feature enhancement, and image reconstruction. Specifically, the LR thermal image is fed into the feature extraction implemented by convolutional layers to extract shallow features.

Then, the extracted features are fed into feature enhancement, which comprises N cascaded NAFBlocks [4], to enhance the image features. The enhanced feature is then fed into the image reconstruction layer, followed by a skip connection from the LR thermal image, generating the final super-resolved result I_{SR}. L1 loss is chosen as the loss function to constrain the network. To enhance the representability of the network, the model is first trained with a $\times 4$ LR-HR

[3] https://codalab.lisn.upsaclay.fr/competitions/17013

[4] https://codalab.lisn.upsaclay.fr/competitions/17014

Figure 5. Results on Track-1 testing set. Images from left to right: LR, super-resolution result (AC-TSR Team), and GT.

Figure 6. Results on Track-2 testing set: top row shows Evaluation 1 ($\times$8), bottom row shows Evaluation 2 ($\times$16). Images from left to right: HR Visible, LR, super-resolution result (GUIDED SR Team), and GT.

pair and then fine-tuned with a $\times$8 LR-HR pair. In addition, data augmentation is employed to increase the quantity and diversity of the data, and self-ensemble is also used in the EDSR [18] method.

The AC-TSR team conducts experiments on two NVIDIA 3090 GPUs for two days using the PyTorch [26] framework. The batch size and patch size are 8 and 32$\times$32, respectively. The model size is 132.98M parameters. The quantitative results show that the AC-TSR team achieves the best results in both metrics: **27.52** PSNR and **0.8355** SSIM for Track-1.

Source code can be found in https://github.com/wcy-cs/AC-TSR.

3.2. Track-1: CTYUN-AI

The team employed the HAT-L [6] model to enhance PSNR and SSIM metrics, leveraging a model architecture that processes LR images through a Shallow Feature Extraction layer before concatenating the shallow features for deep feature extraction via Residual Hybrid Attention Groups (RHAG) [6] and a convolution layer, ultimately upsampling to reconstruct high-resolution images, as shown in Fig. 8.

A dual-component loss function combining Mean Squared Error Loss (MSELoss) and SSIM Loss with respective weights of 1 and 0.02 is adopted to balance the scale of the loss functions effectively. The model, comprising 20.8M parameters and initiated with ImageNet pre-trained weights, is trained with an Adam optimizer at a learning rate of 0.0001 for 2000 steps, incorporating a warmup phase at a reduced learning rate for the latter half of the training. To ensure robustness, a 5-fold cross-validation strategy is implemented, training models on partitions of the data and averaging predictions from models with the highest validation accuracy to determine the final test outcomes.

The CTYUN-AI team conducts experiments on two NVIDIA 3090 GPUs, each with 24GB of VRAM, over a total of 20 hours. The model size is 20.8M parameters. The quantitative results show that this team achieves the second best results in both metrics: 27.48 PSNR and 0.8351 SSIM for Track-1.

Source code can be found in https://github.com/upczww/TISR.

3.3. Track-1: HBNU

Given the outstanding results obtained by Transformer-based deep neural networks in image super resolution, the HBNU team chooses to utilize Hybrid Attention Transformer (HAT) [5]. This team first applies Nearest Neighbor upsampling to the LR image, followed by the application of Cutblur [40], an image augmentation technique specifically designed for image super-resolution. After downsampling the image using a desubpixel layer [37], they then directly employ the model architecture of HAT [5]. For a detailed understanding of the model structure, please refer to HAT [5]. In the training stage, inspired by HAT [5]'s same-task pre-training strategy, this team performs pre-training using the DF2K (DIV2K [11]+Flicker2K [8]) to avoid the massive time required for pre-training on ImageNet [7]. Figure 9 shows the proposed approach.

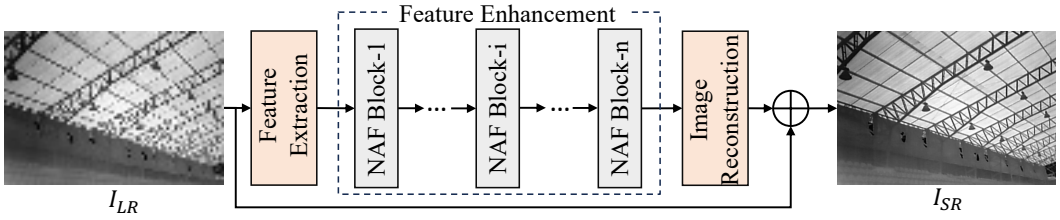

Figure 7. Architecture proposed by the AC-TSR team.

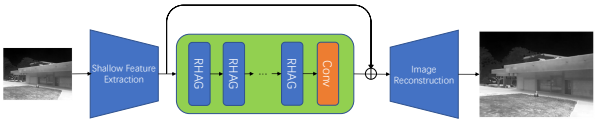

Figure 8. Architecture proposed by CTYUN-AI team on Track-1.

This team extracts sub-images in advance to increase I/O speed by cropping the images. The crop sizes are 480×480 for HR images and 60×60 for LR images for pre-training, and 432×432 for HR images and 54×54 for LR images for fine-tuning. The batch size is 32 for pre-training and 8 for fine-tuning, with the patch size set to 48×48 in both cases. To mitigate overfitting on the small dataset, training-time augmentations like random rotation, horizontal, and vertical flips are employed. Additionally, test-time augmentations, outlined in [28], are utilized to produce the final outputs. The L1 loss is selected as loss function. Adam optimizer is used for model training, with a learning rate set to 2e-4 for pre-training and halved for fine-tuning.

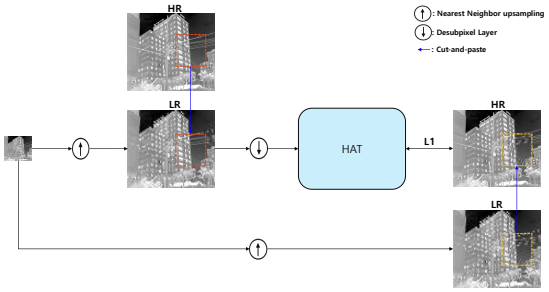

Figure 9. Architecture proposed by HBNU team on Track-1.

This team utilizes 4 NVIDIA GeForce RTX 3090 GPUs for pre-training and 2 NVIDIA RTX A6000 GPUs for fine-tuning, all supported by a 20-core CPU. Python programming language and the PyTorch deep learning framework is used. The model size is 21.2M parameters. The quantitative results show that this team achieves 27.34 PSNR & 0.8322 SSIM for Track-1.

Source code can be found in https://github.com/huiwoni/thermal_Image_SR

3.4. Track-1: HSC-SCA

The HSC-SCA team proposes a pre-trained SwinIR model as shown in Fig. 10, enhanced through the DIV2K dataset, to upscale LR thermal images to $\times 8$ HR images, utilizing DIV2K and Urban100 datasets for re-training to preserve pre-trained weights and focus on structural details in complex urban settings [10, 16, 36]. Grayscale images are specifically chosen to emphasize shape enhancement, with a tanh layer for output normalization and a cosine annealing approach for learning rate adjustment [22]. AdamW is used for optimization, with the model's loss comprising pixel-wise, perceptual, and adversarial components. Initial training involves DIV2K, Urban100, and a thermal dataset, followed by fine-tuning with adjusted datasets and loss structures, selecting three models for a global ensemble based on the lowest validation loss.

The initial training phase mixed Charbonnier and SSIM losses with a VGG19-based perceptual loss and a pix2pix discriminator for adversarial loss, maintaining a loss ratio of $1 : 10^{-3} : 10^{-3}$, and shifted to MSE for pixel-wise loss with adjusted loss ratios in the fine-tuning phase [2, 12]. Two ensemble methods are employed: geometric self-ensemble for manipulated input averaging and a global ensemble for averaging results across different models [19]. The learning rate begins at $2 \cdot 10^{-4}$, decreasing to 10^{-4} in fine-tuning. Re-training includes a sliding window approach for extracting and augmenting LR-HR image pairs (48×48 and 384×384, respectively), supplemented by random transformations in the tuning phase for enhanced augmentation [29].

The HSC-SCA team conducts experiments on AMD Ryzen 9 7950X 16-Core Processor (4.50 GHz) / 64 GB RAM / 4090 1 way (24 GB VRAM), using pytorch, with a training time of about 4 days. The model size is 289.1M. The quantitative results show that this team achieves the second best results in PSNR metric 27.48 PSNR and 0.8292 in SSIM metric for Track-1.

Source code can be found in https://github.com/jsyoonDL/PBVS2024_TISR_Track1_x8.

3.5. Track-2: AIR

Recent advances in Image Super-resolution research have shown that Transformer-based networks deliver impressive

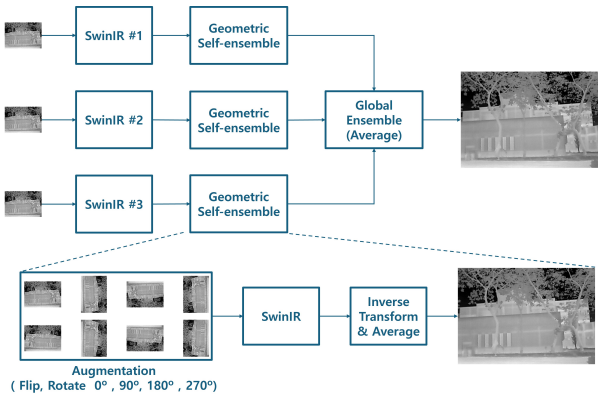

Figure 10. Architecture proposed by HSC-SCA team on Track-1.

results [14]. Although they excel in capturing global information, they are less adept at constructing high-frequency details compared to CNNs [9]. To overcome this limitation, the authors of CRAFT [15] developed a model that effectively captures and integrates high-frequency information with global insights. The AIR team has adapted this model for Guided Thermal Image Super-resolution, useful even for RGB images, and introduces the Guided CRAFT model. This approach starts with extracting features from thermal images using RCRFG from CRAFT. Then, it proposes a Guided RCRFG that enhances the performance of Guided Thermal Image Super-resolution by utilizing these features along with RGB images. The architecture is detailed in Fig. 11.

The proposed network is trained using high-quality (HQ) images with a resolution of 320×240 from the provided train and validation dataset, employing a batch size of 4. All images are augmented by random horizontal, vertical flips and translates. The pixel-wise *L1* loss function is calculated for a pair of reconstructed images obtained from Guided CRAFT and HQ images.

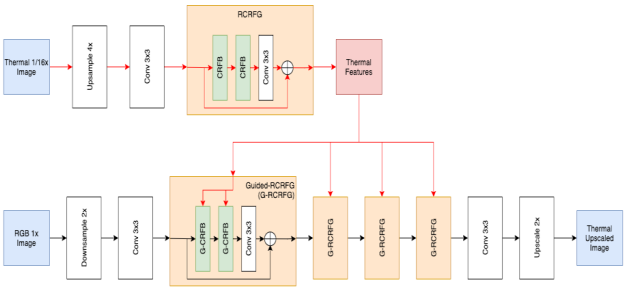

Figure 11. Architecture proposed by AIR team on Track-2.

All reported implementations are based on PyTorch framework, while the proposed approach is conducted with 16-Core CPU, 1 × A100 GPU, 64Gib RAM for approximately five days. This team uses early stopping and an initial learning rate of 0.0002 with the RAdam optimizer [21]. The model size is 3.04M parameters. The quantitative results show that the AIR team achieves 24.77 PSNR and 0.7878 SSIM in evaluation 2 for Track-2.

Source code can be found in https://github.com/DoGunKIM93/guided-craft.

3.6. Track-2: GUIDED SR

The GUIDED SR team proposes a hybrid framework, and it is based on the following two observations: 1) the widely used two-stream structure can better capture complementary information between different modalities [33, 41, 42]; 2) the mix of experts (MOE) strategy can further improve the effectiveness of the algorithm by splitting a task into several parts and using specialized experts to handle each one. Specifically, as depicted in Fig. 12, the proposed framework takes a paired RGB and thermal image (RGB-T) as input and produces a high-resolution thermal image as output. It contains five expert models, and the architecture details of the expert model are listed in Fig. 13 (more details for this architecture can be found in NAFNet [4]). Although these models have the same network design, they have different functions. Experts 1-4 process RGB-T pairs that are rotated at various angles, striving to rebuild the images from multiple perspectives. Expert 5 serves as a fusion module that integrates the output of the other experts.

To make these expert models cooperate with each other, each expert model is first trained individually and then their trained parameters are integrated for fine-tuning. The $\mathcal{L}_1$ loss is used as the default loss function. The experiments are conducted on four NVIDIA A40 GPUs, each with 48GB of RAM, and it takes about five days to train our framework. The batch size and patch size are set to 8 and 32 × 32, respectively. The model size is 600M parameters. The quantitative results show that this team achieves the best results in both evaluations across both metrics: **31.52** PSNR and **0.9127** SSIM in evaluation 1, and **25.99** PSNR and **0.8266** SSIM in evaluation 2, respectively, for Track-2.

Source code can be found in https://github.com/zhwzhong/GuidedSR-2024.

3.7. Track-2: UMKC MCC

The UMKC MCC team presents a multi-scale architecture inspired by [13], as shown in Fig. 14. It uses bicubic interpolation to upsample a LR thermal image (×8 or ×16 for GTISR tasks) to align with a high-resolution RGB image. The upsampled thermal image is then concatenated with the RGB image. The concatenated image is downsampled by 2× and 4× in two parallel pixel unshuffling streams. The stream outputs are then processed by shallow feature extractors composed of two deformable convolutions with a PReLU activation function in between.

The extracted features are inputted into a single fusion

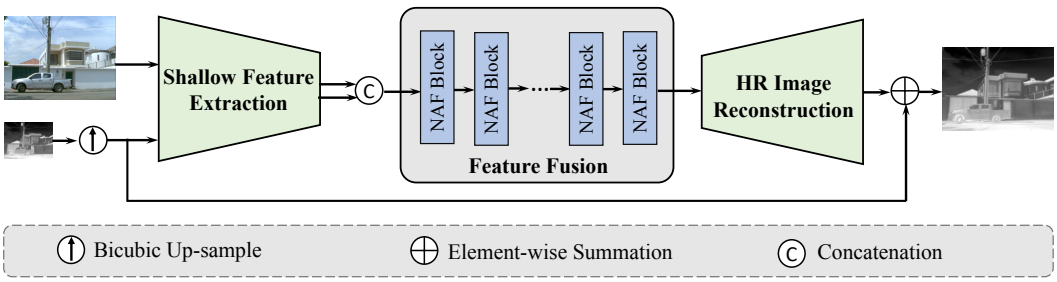

Figure 12. Architecture proposed by GUIDED SR team, Track-2.

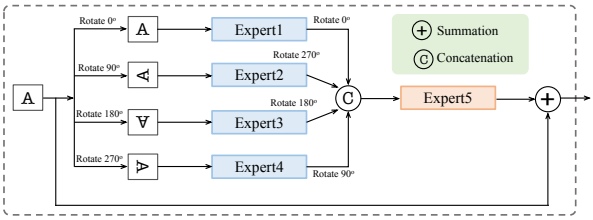

Figure 13. Detailed architecture proposed by GUIDED SR team.

block. The first component of the fusion block is a residual block with enhanced channel attention [38]. The $4\times$ downsampled features are upsampled to match the size of the $2\times$ downsampled features. Those features are concatenated and then passed through a channel transformer [3]. These fusion blocks repeat N times. The $2\times$ downsampled features also undergo a reconstruction block to obtain a high-resolution thermal image.

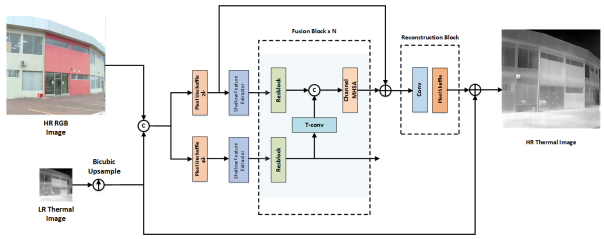

Figure 14. Architecture proposed by UMKC MCC team, Track-2.

During training, each image is randomly cropped to either 32×32 (for the $\times 8$ GTISR task) or 16×16 (for the $\times 16$ GTISR task). Each patch is then augmented through mixup and flipping. It took 2 days to train the model with a batch size of 8 on 2 NVIDIA RTX A6000 GPUs. The model has 12.17 M parameters and optimal performance is achieved with fusion blocks of size 48. The model size is 12.17M parameters. The quantitative results show that this team achieves the second best results in both evaluations across both metrics: 30.05 PSNR and 0.8947 SSIM in evaluation 1, and 25.67 PSNR and 0.8167 SSIM in evaluation 2, respectively, for Track-2.

The code is available at https://drive.google.com/file/d/1MnbKL4OpD-yjkY6fUWxbGR18F1pYZ5Hr

3.8. Track-2: VISION IC

The VISION IC team proposes a novel method named SwinFuSR for RGB guided thermal image super-resolution (Fig. 15), inspired by SwinFusion [23], as a solution for the PBVS 2024 TISR track-2 challenge. The architecture is composed of three modules: (1) shallow features extraction using convolutional layers followed by N Swin Transformer blocks [17], (2) deep features extraction using L attention-guided cross-domain fusion blocks for the fusion of IR and RGB features followed by concatenation and convolution to merge the branches and (3) deep features reconstruction using P Swin Transformer layers to refine the merged features and three convolution layers to return to image space.

In the first two modules, the architecture is divided into two branches, similarly to SwinFusion: one dedicated to the RGB image and the other to the IR image. A bicubic interpolation is performed on the IR image so that its dimensions match those of its paired RGB image. A skip connection from the IR interpolated image and to reconstructed image is introduced for faster convergence and better performance. This strategy allows us to obtain good convergence properties with an L_1 loss, then refine the optimization with an L_2. The used patch size is 128×128 and batch size is 16. The two input patches are augmented with random horizontal and vertical flip and random rotations and then normalized between 0 and 1.

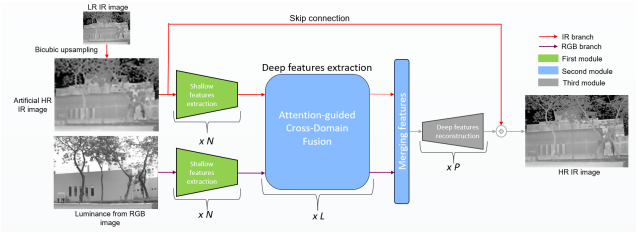

Figure 15. Architecture proposed by VISION IC team, Track-2.

The number of heads, the window size and the embedding dimensions are 6, 9 and 60 respectively. Hyperparameters are set with $N = 2$, $L = 3$ and $P = 3$. For training, a PyTorch framework is used. The learning rate is set to 4×10^{-4} until 3300 epochs and then it is reduced to 1×10^{-4} for the remainder. The Adam optimizer is used. The run lasted 72 hours (4300 epochs) with two Tesla V100 GPUs with 32.0 GB of RAM each. The model size is 3.30M parameters. The quantitative results show that this team achieves the following results: 29.34 PSNR and 0.8824 SSIM in evaluation 1, and 24.69 PSNR and 0.7928 SSIM in evaluation 2, respectively, for Track-2.

Source code can be found in https://github.com/VisionICLab/SwinFuSR.

4. Conclusion

This paper highlights the innovative solutions proposed by participants teams of the Thermal Image Super-Resolution Challenge at PBVS 2024, incorporating both traditional and novel tracks with a focus on cross-spectral datasets. The challenge highlights the use of transformer-based models, attention mechanisms, and hybrid architectures for enhanced detail and texture recovery, alongside strategic data augmentation and advanced loss functions to optimize model performance. This year's challenge has seen an unprecedented level of participation with over 175 teams across the tracks, reflecting a growing interest in thermal image super-resolution. The outcomes from Track-1 demonstrate robust performance, setting new benchmarks for the field. Meanwhile, results from Track-2 reveal that guided super-resolution techniques notably enhance image quality, establishing a foundation for future research. The introduced dataset will serve as a critical benchmark for upcoming challenges, promoting collaboration and advancing thermal image super-resolution technology.

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"; and by the ESPOL project CIDIS-12-2022. The second author acknowledges the support of the Generalitat de Catalunya CERCA Program to CVC's general activities, and the Departament de Recerca i Universitats from Generalitat de Catalunya with reference 2021SGR01499.

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 2024 organization team:

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

Affiliations:

[1]Escuela Superior Politécnica Del Litoral, ESPOL, 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-TSR

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

Affiliation: School of Computer Science and Technology, Harbin Institute of Technology, Harbin 150001, China.

AIR

Members: Jin Kim (kim.jin@hanwha.com), Dongyeon Kang, Dogun Kim.

Affiliation: Hanwha Systems, Seoul, Republic of Korea.

CTYUN-AI

Members: Weiwei Zhou (zhouweiwei@chinatelecom.cn), Chengkun Ling and Jiada Lu.

Affiliation: China Telecom Cloud, Guangzhou, China.

GUIDED SR

Members: Zhiwei Zhong (zhwzhong.cs@gmail.com), Peilin Chen and Shiqi Wang.

Affiliation: Department of Computer Science, City University of Hong Kong, Hong Kong SAR, China.

HBNU

Members: Huiwon Gwon (huiwon/20191515@edu.hanbat.ac.kr), Hyejeong Jo and Sunhee Jo.

Affiliation: Hanbat National University, Daejeon, Republic of Korea.

HSC-SCA

Members: Jiseok Yoon (jsyoon2118@hanwha.com), Wonseok Jang and Haseok Song.

Affiliation: Hanwha Systems / Seongnam-si, Gyeonggi-do, Republic of Korea.

UMKC MCC

Members: Raghunath Sai Puttagunta (rpyc8@umsystem.edu), Zhu Li and George York.

Affiliation: SenseTime, China.

VISION IC

Members: Cyprien Arnold (cyp.arnold@gmail.com), and Lama Seoud.

Affiliation: Polytechnique Montréal, Montréal, Canada.

References

[1] Moab Arar, Yiftach Ginger, Dov Danon, Amit H Bermano, and Daniel Cohen-Or. Unsupervised multi-modal image registration via geometry preserving image-to-image translation. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pages 13410–13419, 2020. 2

[2] Jonathan T. Barron. A general and adaptive robust loss function. In *2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 4326–4334, 2019. 5

[3] Yuanhao Cai, Jing Lin, Zudi Lin, Haoqian Wang, Yulun Zhang, Hanspeter Pfister, Radu Timofte, and Luc Van Gool. Mst++: Multi-stage spectral-wise transformer for efficient spectral reconstruction, 2022. 7

[4] 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. 3, 6

[5] Xiangyu Chen, Xintao Wang, Wenlong Zhang, Xiangtao Kong, Yu Qiao, Jiantao Zhou, and Chao. Dong. Hat: Hybrid attention transformer for image restoration. *arXiv preprint arXiv:2309.05239*, 2023. 4

[6] Xiangyu Chen, Xintao Wang, Jiantao Zhou, Yu Qiao, and Chao Dong. Activating more pixels in image super-resolution transformer. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 22367–22377, 2023. 4

[7] Jia Deng, Wei Dong, Richard Socher, Li-Jia Li, Kai Li, and Li Fei-Fei. Imagenet: A large-scale hierarchical image database. In *2009 IEEE Conference on Computer Vision and Pattern Recognition*, pages 248–255, 2009. 4

[8] R. Timofte et al. Ntire 2017 challenge on single image super-resolution: Methods and results. In *2017 IEEE Conference on Computer Vision and Pattern Recognition Workshops (CVPRW)*, pages 1110–1121, 2017. 4

[9] Guangwei Gao, Zixiang Xu, Juncheng Li, Jian Yang, Tieyong Zeng, and Guo-Jun Qi. Ctcnet: A cnn-transformer cooperation network for face image super-resolution. *IEEE Transactions on Image Processing*, 32:1978–1991, 2023. 6

[10] Jia-Bin Huang, Abhishek Singh, and Narendra Ahuja. Single image super-resolution from transformed self-exemplars. In *2015 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 5197–5206, 2015. 5

[11] Andrey Ignatov, Radu Timofte, et al. Pirm challenge on perceptual image enhancement on smartphones: report. In *European Conference on Computer Vision (ECCV) Workshops*, January 2019. 4

[12] Phillip Isola, Jun-Yan Zhu, Tinghui Zhou, and Alexei A. Efros. Image-to-image translation with conditional adversarial networks, 2018. 5

[13] Birendra Kathariya, Zhu Li, and Geert Van der Auwera. Joint pixel and frequency feature learning and fusion via channel-wise transformer for high-efficiency learned in-loop filter in vvc. *IEEE Transactions on Circuits and Systems for Video Technology*, pages 1–1, 2023. 6

[14] Dawa Chyophel Lepcha, Bhawna Goyal, Ayush Dogra, and Vishal Goyal. Image super-resolution: A comprehensive review, recent trends, challenges and applications. *Information Fusion*, 91:230–260, 2023. 6

[15] Ao Li, Le Zhang, Yun Liu, and Ce Zhu. Feature modulation transformer: Cross-refinement of global representation via high-frequency prior for image super-resolution. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pages 12514–12524, 2023. 6

[16] Jingyun Liang, Jiezhang Cao, Guolei Sun, Kai Zhang, Luc Van Gool, and Radu Timofte. Swinir: Image restoration using swin transformer, 2021. 5

[17] Jingyun Liang, Jiezhang Cao, Guolei Sun, Kai Zhang, Luc Van Gool, and Radu Timofte. SwinIR: Image restoration using swin transformer. In *2021 IEEE/CVF International Conference on Computer Vision Workshops (ICCVW)*, pages 1833–1844. IEEE. 7

[18] 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/CVF Conference on Computer Vision and Pattern Recognition Workshops*, pages 1132–1940, 2017. 4

[19] Bee Lim, Sanghyun Son, Heewon Kim, Seungjun Nah, and Kyoung Mu Lee. Enhanced deep residual networks for single image super-resolution, 2017. 5

[20] Philipp Lindenberger, Paul-Edouard Sarlin, and Marc Pollefeys. Lightglue: Local feature matching at light speed. *arXiv preprint arXiv:2306.13643*, 2023. 2

[21] 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. 6

[22] Ilya Loshchilov and Frank Hutter. Sgdr: Stochastic gradient descent with warm restarts, 2017. 5

[23] Jiayi Ma, Linfeng Tang, Fan Fan, Jun Huang, Xiaoguang Mei, and Yong Ma. SwinFusion: Cross-domain long-range learning for general image fusion via swin transformer. 9(7):1200–1217. 7

[24] Kasper Marstal, Floris Berendsen, Marius Staring, and Stefan Klein. Simpleelastix: A user-friendly, multi-lingual library for medical image registration. In *Proceedings of the IEEE conference on computer vision and pattern recognition workshops*, pages 134–142, 2016. 2

[25] D Muthukumaran and M Sivakumar. Medical image registration: a matlab based approach. *Int. J. Sci. Res. Comput. Sci. Eng. Inf. Technol*, 2(1):29–34, 2017. 2

[26] Adam Paszke, Sam Gross, Soumith Chintala, Gregory Chanan, Edward Yang, Zachary DeVito, Zeming Lin, Alban Desmaison, Luca Antiga, and Adam Lerer. Automatic differentiation in pytorch. In *Proceedings of the Conference on Neural Information Processing Systems Workshop*, pages 4–9, 2017. 4

[27] 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

[28] Rasmus Rothe Radu Timofte 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*, 2016. 5

[29] Edgar Riba, Dmytro Mishkin, Daniel Ponsa, Ethan Rublee, and Gary Bradski. Kornia: an open source differentiable computer vision library for pytorch. In *Winter Conference on Applications of Computer Vision*, 2020. 5

[30] Rafael E Rivadeneira, Angel D Sappa, Boris X Vintimilla, Dai Bin, Li Ruodi, Li Shengye, Zhiwei Zhong, Xianming Liu, Junjun Jiang, and Chenyang Wang. Thermal image super-resolution challenge results-pbvs 2023. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 470–478, 2023. 1, 2

[31] 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. 1, 2

[32] 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. 2

[33] Rafael E. Rivadeneira, Angel D. Sappa, Boris X. Vintimilla, Jin Kim, Dogun Kim, Zhihao Li, Yingchun Jian, Bo Yan, Leilei Cao, Fengliang Qi, Hongbin Wang, Rongyuan Wu, Lingchen Sun, Yongqiang Zhao, Lin Li, Kai Wang, Yicheng Wang, Xuanming Zhang, Huiyuan Wei, Chonghua Lv, Qigong Sun, Xiaolin Tian, Zhuang Jia, Jiakui Hu, Chenyang Wang, Zhiwei Zhong, Xianming Liu, and Junjun Jiang. Thermal image super-resolution challenge results - pbvs 2022. In *2022 IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops (CVPRW)*, pages 417–425, 2022. 6

[34] 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

[35] Rafael E Rivadeneira, Henry O. Velesaca, and Angel D. Sappa. Cross-spectral image registration: a comparative study and a new benchmark dataset. In *Proceedings of International Conference on Innovations in Computational Intelligence and Computer Vision (ICICV)*, 2024. 1, 2

[36] Radu Timofte, Shuhang Gu, Jiqing Wu, and Luc Van Gool. Ntire 2018 challenge on single image super-resolution: Methods and results. In *2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops (CVPRW)*, pages 965–96511, 2018. 5

[37] Thang Vu, Cao Van Nguyen, Trung X. Pham, Tung M. Luu, and Chang D. Yoo. Fast and efficient image quality enhancement via desubpixel convolutional neural networks. *The European Conference on Computer Vision (ECCV) Workshops*, 2018. 4

[38] Qilong Wang, Banggu Wu, Pengfei Zhu, Peihua Li, Wangmeng Zuo, and Qinghua Hu. Eca-net: Efficient channel attention for deep convolutional neural networks, 2020. 7

[39] 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. 2

[40] Jaejun Yoo, Namhyuk Ahn, and Kyung-Ah. Sohn. Rethinking data augmentation for image super-resolution: A comprehensive analysis and a new strategy. *arXiv preprint arXiv:2004.00448*, 2020. 4

[41] Zhiwei Zhong, Xianming Liu, Junjun Jiang, Debin Zhao, and Xiangyang Ji. Deep attentional guided image filtering. *IEEE Transactions on Neural Networks and Learning Systems*, 2023. 6

[42] Zhiwei Zhong, Xianming Liu, Junjun Jiang, Debin Zhao, and Xiangyang Ji. Guided depth map super-resolution: A survey. *ACM Comput. Surv.*, 55(14s), jul 2023. 6