Thermal Image Super-Resolution Challenge Results - PBVS 2025

Rafael E. Rivadeneira, Angel D. Sappa and Riad Hammoud,
Jiyong Rao, Hang Zhong, Yu Wang, and Shengjie Zhao,
Zhiwei Zhong, Yung-Hui Li and Shiqi Wang,
Qiangqiang Shen, Hanzhang Wang and Xuanqi Zhang,

Abstract

This paper presents the results of the Sixth Thermal Image Super-Resolution Challenge held within the Perception Beyond the Visible Spectrum (PBVS) workshop at CVPR 2025. The challenge maintains the same cross-spectral benchmark dataset as the previous year, consisting of 1000 thermal images each paired with corresponding high-resolution RGB images. Track 1 focuses on single thermal image super-resolution, enhancing low-resolution infrared images by a factor of $\times 8$, whereas Track 2 addresses guided thermal image super-resolution, performing super-resolution at scale factors of $\times 8$ and $\times 16$ by leveraging high-resolution RGB images as auxiliary inputs. The 2025 edition attracted increased participation, with 128 teams competing in Track 1 and 86 teams in Track 2. The paper describes methodologies employed by the top participating teams, emphasizing innovations in transformer-based and hybrid architectures, and provides a detailed comparative analysis of the results between the 2024 and 2025 challenges. This analysis reveals significant progress in thermal image reconstruction accuracy, showcasing notable advances achieved by the leading methodologies.

1. Introduction

Thermal Image Super-Resolution (TISR) has gained considerable importance in recent years, primarily driven by its critical applications in fields such as security [7], surveillance [26], autonomous driving [5], and industrial inspection [1], where high-quality thermal imaging significantly enhances decision-making capabilities. Since its inception at the Perception Beyond the Visible Spectrum (PBVS) workshop at CVPR 2020 [21], the Thermal Image Super-Resolution Challenge has become a pivotal benchmark, annually showcasing the latest advancements in this specialized field. The ongoing relevance and impact of the TISR challenges are underscored by the growing number of participants and the continuous improvement in the quality of submitted results.

Reflecting on its evolution, the TISR challenge series has played a key role in stimulating methodological innovations and benchmarking progress. After the successful fourth TISR challenge at PBVS 2023 [20] and the fifth edition held in 2024 [24], the 2025 edition marks the sixth consecutive year of the challenge, demonstrating sustained community interest and growth in participant numbers and quality of submissions. Consistent with previous editions, the same benchmark dataset comprising 1000 thermal images paired with registered high-resolution RGB images is utilized. Both challenge tracks remain open for ongoing benchmarking and comparative analysis through their respective CodaLab competitions (Track 1[1] and Track 2[2]), allowing continuous evaluation of advancements.

The Sixth edition of the TISR challenge (PBVS 2025) attracted increased participation with 128 registered teams for Track-1 and 86 for Track-2. Track 1 continues to focus on single thermal image super-resolution, requiring participants to reconstruct enhanced thermal images from low-resolution input by an $\times 8$ factor. Track 2 addresses guided thermal image super-resolution, using high-resolution RGB images to achieve scaling factors of $\times 8$ and $\times 16$. Both tracks leverage the same dataset of 1000 registered thermal-RGB image pairs, ensuring consistency and facilitating a clear comparative evaluation of progress.

This paper details the methodologies and outcomes of top-performing teams, with an emphasis on novel architec-

Rafael E. Rivadeneira[1] (rrivaden@espol.edu.ec), Angel D. Sappa[1,2] and Riad Hammoud[3] are the TISR Challenge - PBVS 2025 organizers. The remaining authors are team members who achieved top results in the challenge.
[1]Escuela Superior Politécnica del Litoral, ESPOL, Guayaquil, Ecuador.
[2]Computer Vision Center, Campus UAB, 08193 Bellaterra, Barcelona, Spain.
[3]Plus.ai, California, USA.
Appendix A contains the authors' teams and affiliations.

[1]https://codalab.lisn.upsaclay.fr/competitions/21247
[2]https://codalab.lisn.upsaclay.fr/competitions/21248

Figure 1. Illustrations from the cross-spectral dataset showing aligned thermal and visible images [25].

tures and training techniques introduced since the previous year. A comparison between 2024 and 2025 highlights some advancements, particularly in transformer-based and hybrid convolutional-transformer models, underscoring the continued progress and heightened competitiveness within the field of thermal image super-resolution. The results demonstrate an increasing community interest in the thermal image SR problem, as evidenced by the annual improvement in metric values, illustrated in Figure 2.

The structure of this manuscript is organized as follows. Section 2 describes the objectives of the Sixth Thermal Image Super-Resolution Challenge and provides details about the benchmark dataset employed. Section 2.3 presents a summary of this year's participation and a comparative analysis of results obtained between 2024 and 2025. Section 3 outlines the methodologies of the top-performing teams, emphasizing recent technical advancements. Finally, Section 4 offers concluding remarks, while appendix shows teams information and their affiliations.

2. TISR 2025 Challenge

Continuing the tradition from previous editions [20–23, 25], the 2025 Thermal Image Super-Resolution challenge serves as an essential benchmark event for evaluating advances in thermal image enhancement methods. As in earlier editions, two tracks are maintained: single-image super-resolution (Track 1) and guided thermal image super-resolution (Track 2), utilizing the CIDIS dataset [25].

2.1. Thermal Image Datasets

The dataset used for the TISR 2025 challenge is the same Cross-spectral Image Dataset for Image Super-resolution (CIDIS) [25] introduced in 2024, depicted in Figure1. CIDIS consists of 1000 newly acquired pairs of simultaneously captured thermal and RGB images under diverse conditions, using a Basler camera for visible spectrum and a FLIR TAU2 camera for thermal spectrum, standardized to a resolution of 640×480 pixels. This new dataset was introduced due to the previous datasets used from 2020 to 2023 reaching their maximum achievable results, as concluded in the 2023 challenge paper [20]. Accurate cross-spectral registration was ensured using advanced techniques such as LightGlue [15], Elastix [12], and other state-of-the-art alignment algorithms.

The dataset is structured into three subsets: 700 image pairs allocated for training, 200 for validation, and the remaining 100 image pairs reserved exclusively for testing. Two distinct tracks leverage CIDIS: Track 1 challenges participants to achieve single thermal image super-resolution at an upscale factor of $\times 8$, while Track 2 involves guided super-resolution, utilizing corresponding high-resolution RGB images to enhance thermal image resolution at factors of $\times 8$ and $\times 16$. This structure facilitates direct comparisons and robust evaluation of methodological improvements achieved since the previous year's challenge.

2.2. Evaluation Methodology

The evaluation methodology for the TISR 2025 challenge adheres to the established protocols used in previous PBVS workshops [20, 24], ensuring consistency and comparability across years. Both challenge tracks utilize the standardized test subset of 100 thermal-RGB image pairs from the CIDIS dataset [25]. For Track 1 (Single Thermal Image Super-Resolution), participants perform super-resolution on 20 low-resolution (LR) thermal images, generated by downsampling original high-resolution captures by a factor of $\times 8$ without introducing additional noise. The evaluation protocol for this track is illustrated in Figure 3.

Track 2 (Guided Thermal Image Super-Resolution) employs the remaining 80 image pairs from the test set, split evenly for two distinct evaluations: 40 images downsampled by a factor of $\times 8$ (Evaluation 1) and 40 images downsampled by a factor of $\times 16$ (Evaluation 2). High-resolution RGB images are provided as auxiliary inputs to guide the super-

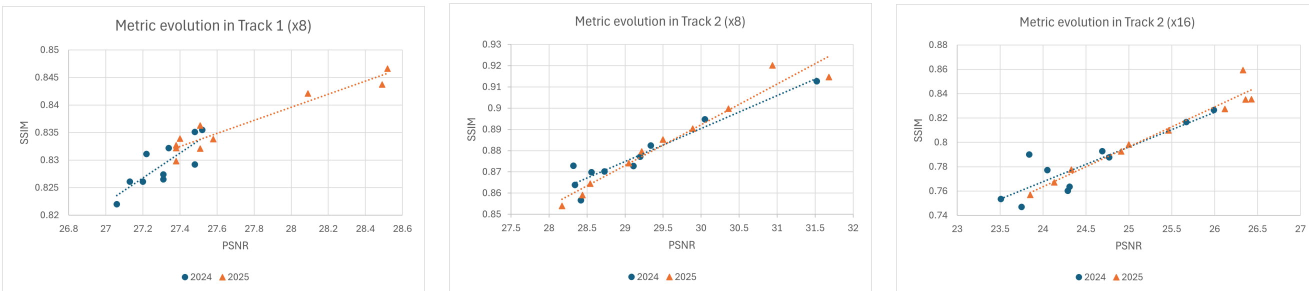

Figure 2. Top 10 metrics evolution through all tracks in 2024 & 2025 challenges.

resolution task, following the same downsampling approach without adding noise. This dual-scale evaluation enables assessing the robustness of guided methods across varying degrees of difficulty. Figure 4 illustrate this track evaluation.

Performance is quantitatively evaluated primarily using two standard image quality metrics—Peak Signal-to-Noise Ratio (PSNR) and Structural Similarity Index Measure (SSIM) [27]—which determine the winners of the challenge. Additionally, metrics such as RMSE, ALEX, VGG, and SQUEEZE are provided for reference purposes only, offering supplementary insights into image quality, as detailed on the official CodaLab competition results page. This comprehensive evaluation protocol facilitates a clear comparative analysis of methodological advancements over results from the 2024 challenge.

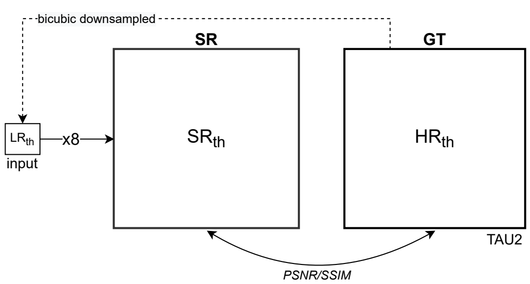

Figure 3. Illustration of the evaluations process for Track-1 on a set of LR images downsampled by a factor of ×8 with no added noise [24].

2.3. Challenge Results

The top-performing teams' results for each track of the 2025 TISR challenge are summarized below. In Track-1, out of 128 teams that initially registered, 34 advanced to the final evaluation phase. Table 1 summarizes the average PSNR and SSIM metrics obtained by the three best-performing teams on the testing set. The top-ranked team, **TongJi-SR**, achieved a PSNR of 28.52 and an SSIM of 0.8466, closely followed by the **InVilab Uantwerp** and **AiMF SR** teams. A qualitative visualization of the top result is depicted in Figure 5.

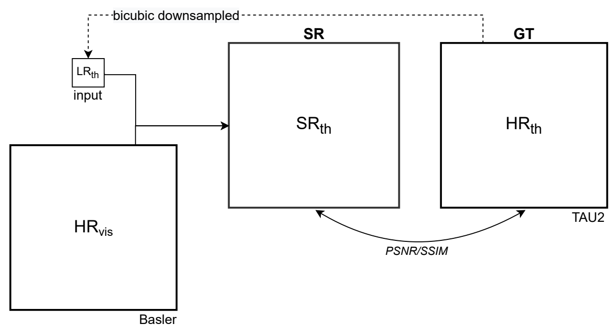

Figure 4. Illustration of the evaluations process for Track-2 on a set of LR images downsampled by a factor of ×8 and ×16 with no added noise and using the corresponding HR visible images as guidance [24].

Team [# param.] (Track-1: SINGLE)	×8	
	PSNR	**SSIM**
AiMF-SR [70.02M]	28.09	0.8421
InVilab [21.6M]	<u>28.49</u>	<u>0.8437</u>
TongJi-SR [27.73M]	**28.52**	**0.8466**

Table 1. Top three results (PSNR and SSIM) of Track-1 (Single TISR, ×8). Bold and underline indicate first- and second-best results, respectively. See Section 2.2 for more details.

Team [# param.] (Track-2: GUIDED)	**Eval 1** (×8)		**Eval 2** (×16)	
	PSNR	**SSIM**	**PSNR**	**SSIM**
CityU HH [132.00M]	**31.68**	<u>0.9147</u>	- - -	- - -
SwinPaste [3.3M]	<u>30.94</u>	**0.9201**	26.33	**0.8593**
UMKC MCC [20.00M]	30.36	0.8997	<u>26.36</u>	0.8351
QQ [64M]	- - -	- - -	**26.43**	<u>0.8354</u>

Table 2. Top three results (PSNR and SSIM) of Track-2 (Guided TISR ×8 and ×168). Bold and underline values correspond to the best- and second-best results, respectively for each evaluation. See Section 2.2 for more details.

For Track 2, out of the 86 initially registered teams, 15 progressed to the final testing stage. Table 1 detail the top three performing teams for the two evaluation scales (×8 and

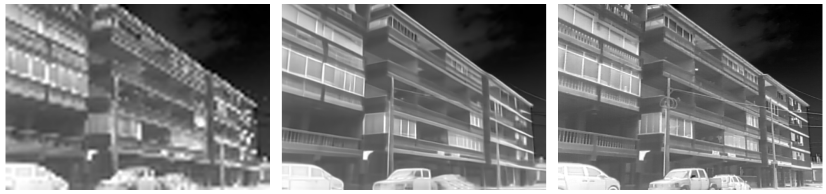

Figure 5. Results on Track-1 testing set. Images from left to right: LR, super-resolution result (TongJi-SR Team), and GT.

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 (SwinPaste Team), and GT.

$\times 16$). Team **CityU HH** achieved the highest PSNR at scale factor $\times 8$ (31.68), while the best SSIM was achieved by the **SwinPaste** team (0.9201). For the $\times 16$ super-resolution evaluation, team **QQ** obtained the best PSNR of 26.43, while the highest SSIM of 0.8593 was achieved by the **SwinPaste** team. Figure 6 provides qualitative examples of the top-performing team's outputs in both evaluations.

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 Figure 6 show qualitative result of Top1 result in both evaluation. Detailed quantitative outcomes are accessible on the CodaLab Competition [18] webpage for both Tracks (Track 1[3]; Track 2[4]).

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 by track/alphabetical order.

3.1. Track 1: AIMF-SR

The AiMF-SR team adopted a transformer-based architecture for the single-image thermal super-resolution task. Specifically, the team utilized the HMANet [4] model, employing the HMA-Medium variant, which integrates Residual Hybrid Transformer Blocks (RHTBs) for enhanced global feature representation. The architecture comprises three main modules: shallow feature extraction, deep feature transformation via RHTBs, and image reconstruction, as depicted in Fig. 7.

Initially, the model was fine-tuned from an ImageNet pre-trained HMA model (with a scale factor ×4). Subsequent fine-tuning involved training on DF2K, Urban100 datasets, and additional thermal images downscaled via bicubic interpolation (×4). The final training phase utilized pairs of LR-HR thermal images at an ×8 scale factor. A MultiStep scheduler with the Adam optimizer was employed, setting the initial learning rate to 1×10^{-5} for 250K iterations, with scheduled decays. Additional fine-tuning for the ×8 SR pairs was conducted over 100K iterations at a learning rate of 5×10^{-5} without a warm-up phase. The loss function was based on a combination of L1 Loss, FFT Loss and SSIM loss with 1.0, 0.05, 0.05 weights, respectively, to ensure stable training and robust super-resolution outputs.

Experiments were conducted on a single NVIDIA A100 GPU with 80 GB of VRAM, using the PyTorch framework. Training lasted approximately three days, with a batch size of 16 and a patch size of 64×64. The final model comprises approximately 70M parameters. Quantitatively, the AiMF-SR team achieved competitive results, obtaining a PSNR of 28.09 and SSIM of 0.8421 on Track-1.

Source code is available at: https://github.com/seanko29/PBVS25_TISR25_Track1

[3] https://codalab.lisn.upsaclay.fr/competitions/21247

[4] https://codalab.lisn.upsaclay.fr/competitions/21248

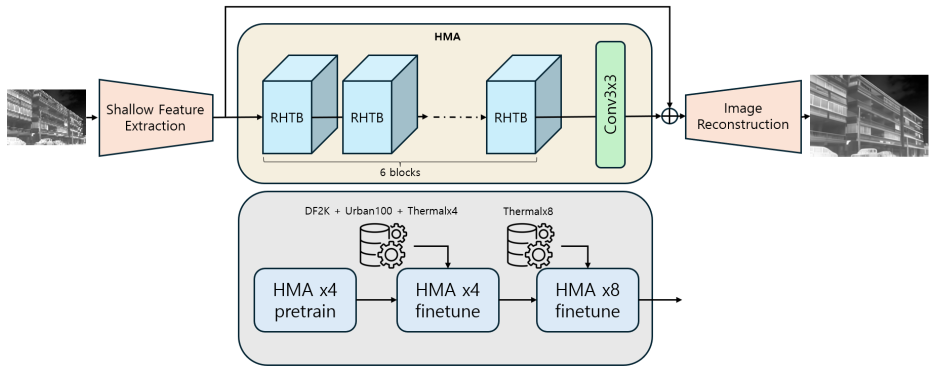

Figure 7. Architecture proposed by the AiMF-SR team on Track 1.

3.2. Track 1: INVILAB

The InViLab Uantwerp team proposed an advanced transformer-based architecture specifically tailored for thermal image super-resolution. Their solution incorporates SRFormer [29] and Hybrid Attention Transformer (HAT) [3], enhanced by their novel Thermal Gradient Spatial Attention (TGSA) module, as shown in Fig. 8. The TGSA module explicitly targets thermally significant regions—such as edges and hotspots—by computing local thermal gradients using Sobel filters, generating spatial attention maps that guide transformer self-attention mechanisms.

The team trained their network with a combined Charbonnier and L1 loss function optimized for thermal imagery. Extensive data augmentation was applied, including synthetic infrared images via RGB-to-IR translation [2], geometric transformations, temperature scaling, and Gaussian blur. To further enhance the robustness and generalization capabilities, they utilized self-ensembling techniques, aggregating predictions from rotated inputs (0°, 90°, 180°, and 270°) and applying median filtering.

Training was conducted using four NVIDIA A100 GPUs, implemented in PyTorch along with BasicSR and xFormers libraries. The total number of parameters in the proposed model is approximately 21.6M. The quantitative results indicate that the InViLab Uantwerp team achieved excellent performance in Track 1, obtaining a PSNR of 28.49 and SSIM of 0.8437.

Source code is publicly available at: https://github.com/thomasdekerf/TISR_track1

3.3. Track 1: TONGJI-SR

The TongJi-SR team proposed a single-image super-resolution approach based on the DRCT architecture [10], which leverages the Swin Transformer [17] for effective global spatial modeling. Their network, illustrated in Figure 9, comprises three main components: shallow feature extraction, deep feature extraction, and image reconstruction.

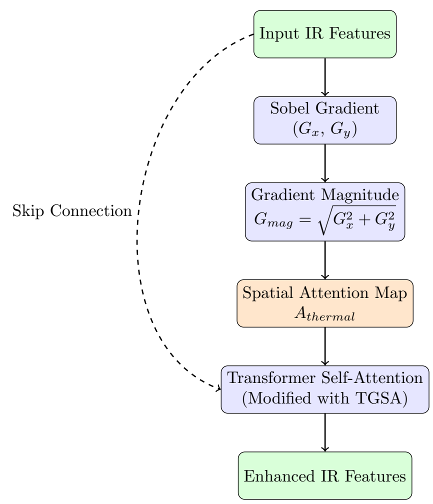

Figure 8. Architecture proposed by the InViLab Uantwerp team on Track 1.

Specifically, the shallow and deep feature modules capture rich hierarchical representations, while the reconstruction module synthesizes the final high-resolution thermal image.

To enhance the model's representational capacity, the TongJi-SR team combined Mean Squared Error (MSE) loss and Structural Similarity Index (SSIM) loss with weights of 1 and 0.02, respectively. Additionally, multi-scale supervision [13] and combination-based data augmentation [6] strategies were employed. In this augmentation method, four randomly selected training images were quarter-cropped and recombined to increase the dataset's diversity. Use the 5-fold cross validation to select the best testing-performance model.

Training was conducted using PyTorch on NVIDIA 4090 GPUs. The model has approximately 27.73M parameters and was initialized with ImageNet pre-trained weights. It was trained with an Adam optimizer at a learning rate of 1×10^{-4}, for a total of 130K iterations (80K pre-training and 50K fine-tuning iterations). A 5-fold cross-validation

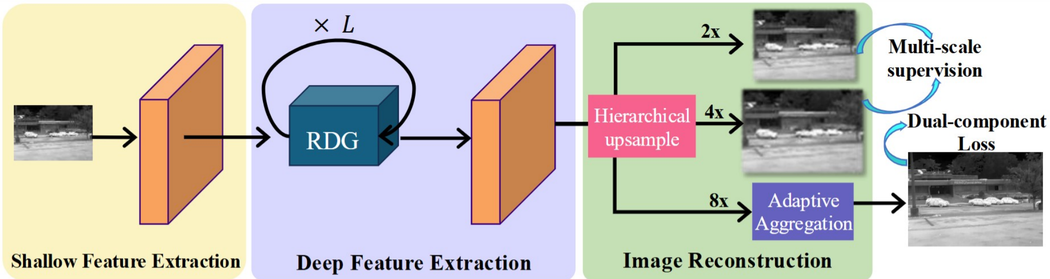

Figure 9. Architecture proposed by the TongJi-SR team on Track 1.

strategy was implemented for improved robustness. The TongJi-SR team achieved the top performance in Track 1, obtaining a PSNR of **28.52** and SSIM of **0.8466**.

Source code is publicly available at: https://github.com/Raojiyong/PBVS_TSR

3.4. Track 2: CityU HH

The CityU HH team proposed a guided thermal super-resolution approach that leverages both low-resolution (LR) thermal images and corresponding high-resolution (HR) RGB images. The network architecture, illustrated in Figure 10, consists of three main components: feature embedding, deep feature extraction, and image reconstruction.

Initially, the LR thermal image and HR RGB image are concatenated and processed through a patch embedding module. The embedded features are then refined by multiple cascaded Gated CNN Blocks inspired by recent advances in Mamba-based architectures [28] and deep residual learning [9]. These gated convolutional blocks effectively capture long-range dependencies and spatial contextual information. Finally, the extracted deep features undergo a patch unembedding operation to reconstruct the super-resolved thermal image, complemented by a global residual connection from the input thermal image to stabilize training.

The training process utilized an L1 loss function between the predicted high-resolution thermal images and the ground truth, ensuring robust reconstruction performance.

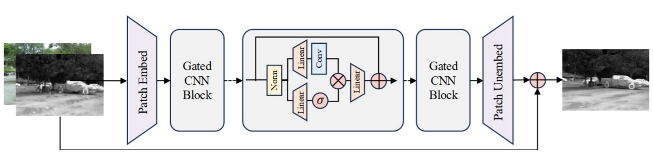

Figure 10. Architecture proposed by the CityU HH team on Track 2.

The CityU HH team trained their model on an NVIDIA A100 GPU using PyTorch for two days. The model comprises approximately 132M parameters. The training batch size was set to 16 with an image patch size of 128×128 pixels. In Track 2, the CityU HH team achieved outstanding quantitative performance, obtaining the highest PSNR of **31.68** and a competitive SSIM of 0.9147 at the ×8 evaluation.

Source code is publicly available at: https://github.com/zhwzhong/Cityu-HH.git

3.5. Track 2: QQ

The QQ team proposed a guided thermal super-resolution architecture designed to effectively integrate low-resolution thermal images with corresponding high-resolution RGB images. As illustrated in Figure 11, it adopts a U-Net shaped framework enhanced by combining Transformer and Mamba blocks to exploit complementary features. Each basic block within their architecture splits input features into two parallel branches: one branch processes features using Swin Transformer blocks [14], while the other leverages Mamba blocks [8, 16] for efficient global feature modeling. The outputs from both branches are then concatenated and aggregated, achieving rich contextual feature representations.

The QQ team trained their model by directly minimizing the L1 loss between predicted super-resolved thermal images and the ground truth, which significantly improved visual fidelity and reconstruction accuracy.

Training was implemented using PyTorch on a single NVIDIA RTX 3090 GPU with 24 GB RAM over approximately two days. The model has around 64M parameters. The QQ team achieved the highest PSNR of **26.43** and an SSIM of 0.8354 in the ×16 evaluation of Track 2.

Source code is available at: https://github.com/hitszqq/track2

3.6. Track 2: SWINPASTE

The SwinPaste team proposed a guided thermal super-resolution model inspired by the SwinFuSR architecture [24], emphasizing cross-domain fusion of RGB and thermal modalities. Their method introduced a novel pre-training strategy, termed the "paste"technique, where multiple thermal images were blended onto a transparent canvas to generate augmented training samples, enhancing the model's generalization capabilities.

As illustrated in Figure 12, the proposed architecture comprises three key modules: shallow feature extraction,

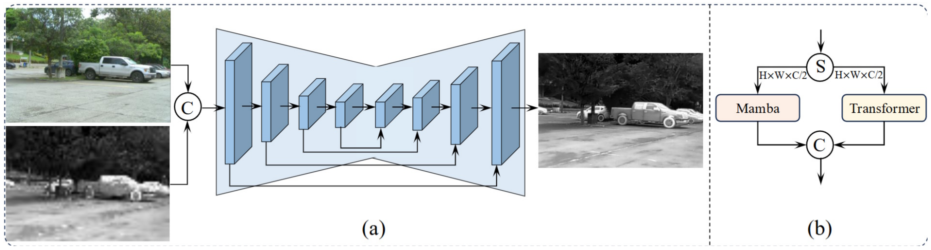

Figure 11. Architecture proposed by the QQ team on Track 2.

deep feature fusion, and image reconstruction. The shallow feature module extracts low-level thermal and RGB features using convolutional layers and Swin Transformer blocks [14]. The deep feature module integrates thermal and RGB features through an attention-guided cross-domain fusion mechanism. Finally, the reconstruction module utilizes upsampling layers and multi-scale supervision to recover high-frequency details in the thermal images.

The model was initially pre-trained on more than 10K augmented image pairs created via the "paste"strategy, followed by fine-tuning on the CIDIS training data provided by the competition. The team optimized their model using the L1 loss with an Adam optimizer.

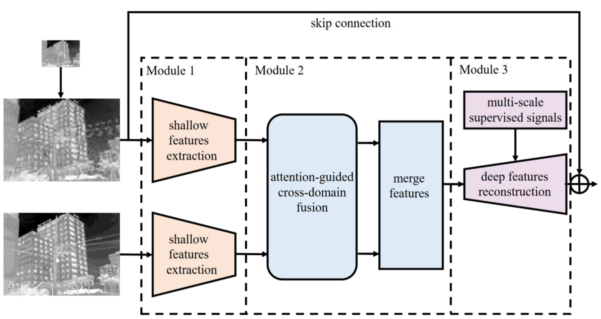

Figure 12. Architecture proposed by the SwinPaste team on Track 2.

Training was performed on two NVIDIA A100 GPUs (40GB RAM each) using PyTorch over approximately six days (four days pre-training, two days fine-tuning). The total number of parameters of their model is about 3.3M. The SwinPaste team achieved competitive results, obtaining a PSNR of 30.94 and the best SSIM of **0.9201** in the $\times 8$ evaluation, and a PSNR of 26.33 with the highest SSIM of **0.8593** in the $\times 16$ evaluation.

Source code is publicly available at: https://github.com/zoniazhong/SwinPaste

3.7. Track 2: UMKC MCC

The UMKC MCC team proposed a guided thermal super-resolution method inspired by MSFFCT [19] and TSFNet [11], employing a multi-scale, cross-modal fusion architecture, as shown in Figure 13.

Their network uses bicubic interpolation for alignment, deformable convolutions for robust feature extraction, and a dual-branch structure (pixel unshuffling and DCT-based frequency analysis) combined via residual channel attention blocks. Training leveraged a combination of L1, SSIM, perceptual, and DCT-based losses, alongside extensive data augmentation.

Training was conducted using two NVIDIA RTX A6000 GPU, lasting approximately three days. The model has around 24M parameters, making it relatively lightweight compared to other teams. Despite this smaller size, the UMKC MCC team obtained highly competitive results, achieving a PSNR of 30.36 and SSIM of 0.8997 at the $\times 8$ evaluation, and a PSNR of 26.36 with an SSIM of 0.8351 at the $\times 16$ evaluation in Track 2.

Source code is publicly available at: https://drive.google.com/drive/folders/16uOfBJopZP_NqKGlva0fkfLnpXN0tZd1?usp=sharing.

4. Conclusion

The sixth Thermal Image Super-Resolution challenge at CVPR 2025 demonstrated significant advancements over the 2024 edition, attracting a record number of participants and revealing clear improvements in both quantitative metrics (PSNR and SSIM) and methodological innovations. Hybrid architectures combining transformers, convolutional networks, and advanced cross-modal fusion modules emerged prominently, with specialized attention techniques and novel data augmentation strategies enhancing thermal image reconstruction quality. Overall, results confirmed the community's continued progress, yet highlighted ongoing opportunities for developing more computationally efficient and practically deployable models for future research.

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

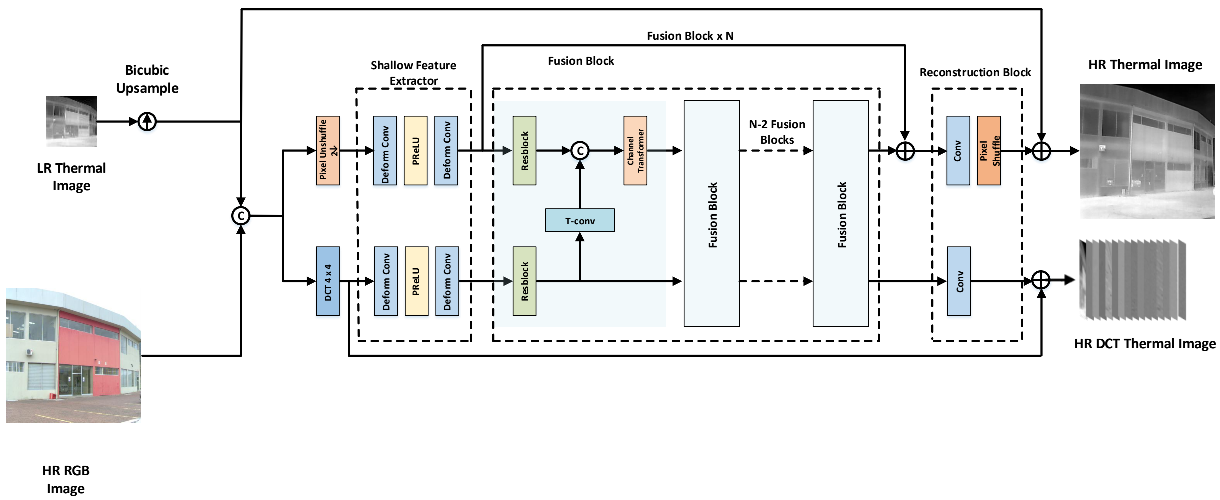

Figure 13. Architecture proposed by the UMKC MCC team on Track 2.

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 2025 organization team:

Members: Rafael E. Rivadeneira[1] (rrivaden@espol.edu.ec), Angel D. Sappa[1,2] and Riad Hammoud[3]

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.
[3]Plus.ai, California, USA

Top Participant Teams:

AiMF-SR

Members: Donggeun Ko (sean.ko@aimfuture.ai), Youngsang Kwak and Jaehwa Kwak.

Affiliation: AiM Future Inc., Seoul, South Korea.

CityU HH

Members: Zhiwei Zhong[1] (zhwzhong.cs@gmail.com), Yung-Hui Li[2] and Shiqi Wang[1].

Affiliation:
[1] Department of Computer Science, City University of Hong Kong, Hong Kong.
[2] AI Research Center, Hon Hai Research Institute, Taiwan.

InVilab Uantwerp

Members: Thomas De Kerf (thomas.dekerf@uantwerpen.be) and Rhys Evans.

Affiliation: InViLab Research Group, Faculty of applied engineering, University of Antwerp, Bélgica.

QQ

Members: Qiangqiang Shen[1] (aurorac9213@gmail.com), Hanzhang Wang[1] and Xuanqi Zhang[2].

Affiliation:
[1]Harbin Institute of Technology, Shenzhen.
[2]University of British Columbia, Canada.

SwinPaste

Members: Hang Zhong (zoniazhong@163.com), Yu Wang, and Shengjie Zhao.

Affiliation: School of Computer Science and Technology, Tongji University, China.

TongJi-SR

Members: Jiyong Rao (raojyon@gmail.com), Yu Wang, and Shengjie Zhao.

Affiliation: School of Computer Science and Technology, Tongji University, China.

UMKC MCC

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

Affiliation: [1]University of Missouri, USA.

References

[1] Hussein N AL-Jubori, AL-Darraji Izzat, and Houssem Jerbi. Defect detection using thermography camera techniques: A review. *Al-Khwarizmi Engineering Journal*, 20(4):70–88, 2024. 1

[2] Sayed Pedram Haeri Boroujeni and Abolfazl Razi. Ic-gan: An improved conditional generative adversarial network for rgb-to-ir image translation with applications to forest fire monitoring. *Expert Systems with Applications*, 238:121962, 2024. 5

[3] 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. 5

[4] Shu-Chuan Chu, Zhi-Chao Dou, Jeng-Shyang Pan, Shaowei Weng, and Junbao Li. Hmanet: Hybrid multi-axis aggregation network for image super-resolution. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 6257–6266, 2024. 4

[5] Moataz Eltahan and Khaled Elsayed. Enhancing autonomous driving by exploiting thermal object detection through feature fusion. *International Journal of Intelligent Transportation Systems Research*, 22(1):146–158, 2024. 1

[6] Zheng Ge, Songtao Liu, Feng Wang, Zeming Li, and Jian Sun. Yolox: Exceeding yolo series in 2021. *arXiv preprint arXiv:2107.08430*, 2021. 5

[7] Xinyue Hao, Shaojuan Luo, Meiyun Chen, Chunhua He, Tao Wang, and Heng Wu. Infrared small target detection with super-resolution and yolo. *Optics & Laser Technology*, 177:111221, 2024. 1

[8] Ali Hatamizadeh and Jan Kautz. Mambavision: A hybrid mamba-transformer vision backbone. *arXiv preprint arXiv:2407.08083*, 2024. 6

[9] Kaiming He, Xiangyu Zhang, Shaoqing Ren, and Jian Sun. Deep residual learning for image recognition. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 770–778, 2016. 6

[10] Chih-Chung Hsu, Chia-Ming Lee, and Yi-Shiuan Chou. Drct: Saving image super-resolution away from information bottleneck. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 6133–6142, 2024. 5

[11] 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*, 2023. 7

[12] 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

[13] Juncheng Li, Faming Fang, Kangfu Mei, and Guixu Zhang. Multi-scale residual network for image super-resolution. In *Proceedings of the European conference on computer vision (ECCV)*, pages 517–532, 2018. 5

[14] 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. 6, 7

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

[16] Yue Liu, Yunjie Tian, Yuzhong Zhao, Hongtian Yu, Lingxi Xie, Yaowei Wang, Qixiang Ye, Jianbin Jiao, and Yunfan Liu. Vmamba: Visual state space model. *Advances in neural information processing systems*, 37:103031–103063, 2024. 6

[17] Ze Liu, Yutong Lin, Yue Cao, Han Hu, Yixuan Wei, Zheng Zhang, Stephen Lin, and Baining Guo. Swin transformer: Hierarchical vision transformer using shifted windows. In *Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV)*, 2021. 5

[18] 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. 4

[19] Raghunath Sai Puttagunta, Birendra Kathariya, Zhu Li, and George York. Multi-scale feature fusion using channel transformers for guided thermal image super resolution. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR) Workshops*, pages 3086–3095, June 2024. 7

[20] 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

[21] 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

[22] 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

[23] 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

[24] Rafael E Rivadeneira, Angel D Sappa, Chenyang Wang, Junjun Jiang, Zhiwei Zhong, Peilin Chen, and Shiqi Wang. Thermal image super-resolution challenge results-pbvs 2024. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pages 3113–3122, 2024. 1, 2, 3, 6

[25] 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. 2

[26] Yu Wang, Wenjun Duan, Junqing Li, Dongdong Shen, and Peiyong Duan. Thermal-adaptation-behavior-based thermal sensation evaluation model with surveillance cameras. *Sensors*, 24(4):1219, 2024. 1

[27] 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. 3

[28] Weihao Yu and Xinchao Wang. Mambaout: Do we really need mamba for vision? In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2025. 6

[29] Yupeng Zhou, Zhen Li, Chun-Le Guo, Li Liu, Ming-Ming Cheng, and Qibin Hou. Srformerv2: Taking a closer look at permuted self-attention for image super-resolution, 2024. 5