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.

DIRT-SR38: A Real Paired Cross-Sensor Thermal Sequence Dataset for Super-Resolution in Clinical Dynamic Infrared Thermography

Warre Clarys Simon Verspeek Thomas De Kerf
Jan Verstockt Steve Vanlanduit Filip Thiessen Gunther Steenackers
University of Antwerp, Belgium
{warre.clarys,simon.verspeek,thomas.dekerf}@uantwerpen.be
{jan.verstockt,steve.vanlanduit,gunther.steenackers}@uantwerpen.be
filip.thiessen@clinic12b.be

Abstract

*Thermal super-resolution (SR) is commonly trained and evaluated using synthetic degradations such as bicubic downsampling, which do not reflect real infrared sensor characteristics such as optics, noise, aliasing, and in-camera processing. We introduce **DIRT-SR38**, the first publicly available real paired cross-sensor thermal sequence dataset acquired under a standardized clinical dynamic infrared thermography (DIRT) cooling–rewarming protocol. The dataset contains synchronized low-resolution (LR) and high-resolution (HR) thermal recordings from 38 healthy female participants, with 810 aligned LR–HR frame pairs per subject (30,780 pairs in total). We benchmark representative SR methods and show a large performance gap between real-paired training and conventional synthetic bicubic training, confirming that synthetic degradations do not adequately model real thermal imaging pipelines. We further show that synthetic pretraining followed by fine-tuning on limited real paired data can recover strong performance, highlighting the value of real paired datasets even at modest scale.*

1. Introduction

Thermal imaging enables contactless observation of surface temperature and is widely used in biomedical monitoring, surveillance, and industrial inspection. However, compared to visible-spectrum cameras, thermal sensors typically offer substantially lower spatial resolution due to focal-plane array constraints, optics, and hardware trade-offs. Image super-resolution (SR) provides a practical way to enhance spatial detail without modifying acquisition hardware, but progress in thermal SR remains limited by the lack of realistic training and evaluation data.

Most learning-based SR pipelines are developed and benchmarked using synthetic degradations, commonly generated by bicubic or Gaussian downsampling of high-resolution (HR) images. While convenient, such models are poor proxies for real thermal imaging pipelines, where degradations arise from sensor-specific modulation transfer characteristics, aliasing, NETD-driven noise, internal non-uniformity correction (NUC), and proprietary in-camera processing. As a result, models trained on synthetic downsampling often generalize poorly to real infrared data, a discrepancy that has been highlighted in real-world SR benchmarks [2, 31], leaving a persistent gap between reported benchmark performance and practical deployment.

The value of physically measured low-resolution (LR) and HR pairs has recently been demonstrated in adjacent spectral domains. In hyperspectral imaging, Ulrichsen *et al.* [29] showed that SR models trained on authentic cross-resolution acquisitions substantially outperform those trained with synthetic downsampling. Within thermal imaging, real paired datasets and benchmarking efforts have begun to emerge, including Real-infraredSR [35] and multi-camera initiatives such as the PBVS Thermal Image Super-Resolution Challenge [23, 25]. However, existing thermal SR datasets primarily contain static images or generic scenes, leaving clinically motivated, temporally resolved thermography largely unaddressed.

Dynamic infrared thermography (DIRT) measures the thermal recovery process following a controlled stimulus (typically cooling), and is increasingly used to study perfusion-related rewarming patterns in biomedical applications. In reconstructive surgery, DIRT has been investigated for perforator visualization and selection in deep inferior epigastric perforator (DIEP) flap breast reconstruction, where accurate spatial depiction of rewarming dynamics is clinically relevant [7, 14, 18, 33]. Recent work further explores automated and AI-based analysis of DIRT sequences for perforator mapping [3, 5]. Despite this growing interest,

to date no publicly available dataset provides *real paired* cross-sensor thermal *sequences* acquired under a standardized clinical DIRT cooling–rewarming protocol.

In this work, we introduce **DIRT-SR38**, a real paired cross-sensor thermal sequence dataset acquired under a standardized abdominal DIRT cooling–rewarming protocol using synchronized dual-camera recordings. The protocol reflects clinical perforator mapping for DIEP flap breast reconstruction and includes recordings from 38 healthy female participants. For each subject, we provide a steady-state baseline sequence and a dynamic rewarming sequence, yielding 810 aligned LR and HR frame pairs per participant (30,780 paired samples in total).

We benchmark representative state-of-the-art SR methods on the real paired data and quantify the domain gap between real-paired training and conventional synthetic bicubic training. Our results show consistent degradation under synthetic training, highlighting the importance of physically measured LR–HR pairs in realistic clinical settings. We further show that synthetic pretraining followed by limited real-data fine-tuning can recover strong performance, indicating that real paired data remain highly valuable even when only modest amounts are available.

The main contributions of this work are:

- **DIRT-SR38:** the first publicly available real paired cross-sensor thermal sequence dataset acquired under a standardized clinical DIRT protocol, comprising 38 subjects and 30,780 aligned LR–HR frame pairs.
- A reproducible acquisition and pairing pipeline for clinical thermal SR, including synchronized dual-sensor recording, cross-sensor geometric registration, and LR temporal stabilization preprocessing.
- A benchmark of representative SR methods showing a substantial performance gap between real-paired training and synthetic bicubic training on thermal data.
- An analysis of synthetic pretraining followed by real fine-tuning, showing that strong performance can be achieved with substantially fewer real subjects, highlighting the value of real paired datasets even in limited-data settings.

The **DIRT-SR38** dataset, including the official train/validation/test split used in this work, is publicly available on Zenodo at https://doi.org/10.5281/zenodo.19241293 [8].

2. Related Work

2.1. Single-Image Super-Resolution

Deep learning has become the dominant paradigm for single-image SR, evolving from early convolutional networks [9, 15] to adversarial approaches [16, 30] and transformer-based architectures [4, 17, 19]. Beyond distortion-oriented reconstruction, recent work addresses blind and real-world SR, where the degradation process is unknown [2, 31, 34].

Cai *et al.* [2] introduced a real-world SR benchmark highlighting the gap between synthetic bicubic training and real camera degradations. Zhang *et al.* [34] proposed practical degradation modeling to approximate real imaging pipelines, while Real-ESRGAN [31] further refined synthetic degradation simulation for blind SR training.

2.2. Real Paired Spectral and Thermal SR

Real paired datasets provide direct supervision from physically measured LR and HR observations. In hyperspectral imaging, Ulrichsen *et al.* [29] demonstrated that training on authentic cross-resolution acquisitions leads to improved reconstruction fidelity compared to synthetic downsampling.

In the infrared domain, Zhou *et al.* [35] introduced RealinfraredSR, a dataset acquired using a thermal imager, and analyzed discrepancies between synthetic and measured degradations. To address this mismatch, they collect real LR–HR image pairs by varying the focal length of an infrared zoom lens. Similarly, Rivadeneira *et al.* [23, 24] proposed a multi-camera acquisition setup and later organized the PBVS Thermal Image Super-Resolution challenge [25]. The PBVS dataset contains over one thousand aligned thermal scenes acquired at different native resolutions and serves as a benchmark for real thermal SR evaluation.

Existing real paired thermal SR datasets are primarily static image-based and target general scenes. To date, publicly available datasets do not focus on standardized clinical DIRT sequences acquired under controlled cooling–rewarming protocols.

2.3. Dynamic Infrared Thermography

DIRT analyzes tissue rewarming after a controlled thermal challenge and has been investigated for perforator mapping and perfusion assessment in reconstructive surgery [7, 14, 18, 33]. Standardization of cooling methodology and acquisition setup is discussed in clinical literature [6, 28], as protocol variations may influence thermal recovery dynamics.

Recent studies explore automated and AI-based analysis of DIRT data for perforator detection and vascular assessment [3, 5]. However, these works operate on single-sensor acquisitions, and no public dataset currently provides real paired cross-sensor thermal sequences under a standardized DIRT protocol.

3. Dataset

3.1. Overview

DIRT-SR38 consists of synchronized LR and HR thermal image sequences acquired from 38 participants under

a standardized abdominal DIRT cooling–rewarming protocol. Each subject includes a 30-second steady-state baseline sequence (90 frames) and a 4-minute dynamic rewarming sequence (720 frames), resulting in 810 aligned LR–HR frame pairs per subject and 30,780 paired samples in total.

All recordings were originally acquired in radiometric mode. However, the released dataset contains intensity-normalized PNG images rather than absolute temperature values. Normalization is performed per sequence and applied consistently to aligned LR–HR pairs to remove global offsets and sensor drift effects while preserving relative spatial temperature patterns.

Both 8-bit and 16-bit normalized PNG versions are provided. In these benchmark formats, pixel values represent relative thermal intensity within each sequence and should not be interpreted as calibrated degrees Celsius. LR inputs are provided at 108×108 resolution and HR targets at 216×216, corresponding to a $\times 2$ scaling factor.

The dataset and official split are available on Zenodo [8].

3.2. Participants

The dataset includes recordings from 38 healthy adult female volunteers ($\geq$21 years of age; age range 21–59 years, mean 33.7 ± 9.2 years; BMI range 19.1–32.4 kg/m^2, mean 23.2 ± 2.6 kg/m^2).

Only female participants were included, as the dataset is acquired under a clinically standardized DIRT protocol for DIEP flap breast reconstruction planning, which is exclusively performed in women.

This study was conducted in accordance with institutional ethical guidelines for research involving human participants. Written informed consent was obtained from all volunteers prior to data acquisition.

The dataset is fully anonymized prior to release. Each participant is represented by a random subject identifier (P01–P38), and no metadata enabling re-identification is provided.

3.3. DIRT Protocol

Data are collected under a standardized cooling–rewarming DIRT protocol used for perforator mapping in DIEP flap breast reconstruction planning [6]. All acquisitions were performed in a temperature-controlled room maintained at 21°C to ensure consistent thermal recovery dynamics across subjects.

Each measurement consists of:

- a 30-second steady-state baseline recording,
- a four-minute rewarming sequence after a controlled abdominal cooling stimulus.

The abdominal region-of-interest (ROI) corresponds to the donor site used in DIEP flap breast reconstruction procedures (Fig. 1). This region is clinically relevant, as the perforator vessels that supply the DIEP flap originate in this abdominal area. In DIRT-based perforator mapping, spatial rewarming patterns within this region provide information on vascular supply and are therefore of primary clinical interest. All recordings are centered on this standardized abdominal ROI.

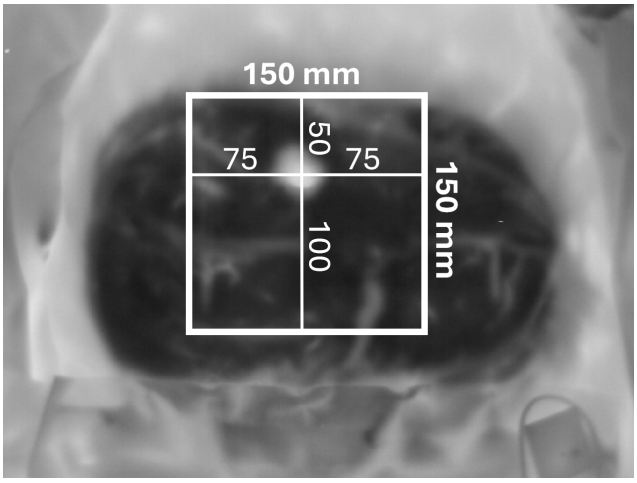


Figure 1. Clinically relevant abdominal ROI used in all DIRT-SR38 acquisitions. The square covers a 150×150 mm area centered on the umbilicus, corresponding to the standard donor-site region for DIEP flap breast reconstruction. All LR–HR pairs are aligned and cropped to this ROI.

3.4. Sensors

The HR modality is captured using a FLIR A655sc (Teledyne FLIR, USA), an uncooled microbolometer operating in the 7.5–14 μm spectral range. The detector provides 640×480 pixels with a 17 μm pixel pitch and a thermal sensitivity (NETD) below 30 mK. A 25° lens is used, resulting in a horizontal field-of-view (HFOV) of approximately 25°.

The LR modality is acquired using a Mosaic Core SEEK Thermal module (Seek Thermal Inc., USA), also operating in the 7.5–14 μm spectral band. The sensor provides 320×240 pixels with a NETD of approximately 70 mK and a HFOV of 24°. Compared to the HR camera, the LR sensor differs in spatial sampling density, optical characteristics, modulation transfer function (MTF), and internal non-uniformity correction (NUC) behavior, resulting in distinct noise and aliasing properties. The key specifications of both sensors are summarized in Table 1.

Both sensors were intrinsically calibrated prior to data collection using a planar thermal calibration target. Lens distortion was corrected and all released frames are provided in the undistorted image domain.

All recordings are acquired in radiometric mode and stored in calibrated degrees Celsius. Emissivity is set to 0.98, consistent with human skin. Ambient temperature parameters are manually configured prior to acquisition to en-

Table 1. Thermal camera specifications.

	FLIR A655sc	SEEK Mosaic Core
Resolution	640×480	320×240
Pixel pitch (µm)	17	12
Spectrum (µm)	7.5–14	7.5–14
NETD (mK)	< 30	~ 70
HFOV (°)	25	24
Frame rate (Hz)	50	27
Radiometric	Yes	Yes

sure stable absolute temperature estimation.

3.5. Temporal Synchronization

The two modalities are recorded asynchronously (LR: 27 Hz; HR: 50 Hz). Frame-level timestamps are stored for both streams. For each released LR frame (subsampled to 3 Hz), the temporally closest HR frame is selected via post-hoc timestamp-based nearest-neighbor matching. Given the 20 ms sampling interval of the HR sensor, the maximum temporal misalignment is bounded by 10 ms. At the final 3 Hz release rate, this corresponds to a relative deviation below 3%, which is negligible with respect to the slow thermal dynamics observed in DIRT.

3.6. Geometric Configuration

Both thermal cameras are rigidly mounted on a custom 3D-printed dual-sensor mount at a working distance of 1000 mm from the subject, with an inter-sensor baseline of 80 mm between optical centers (Fig. 2). To ensure a consistent acquisition distance across all measurements, an auxiliary IFM O3D303 time-of-flight sensor (IFM Electronic GmbH, Germany) is integrated into the mount and used exclusively for distance verification and positioning prior to recording; no depth data are stored or included in the released dataset. The cameras are arranged in a symmetric convergent configuration such that their optical axes intersect at the acquisition distance of 1000 mm, with the inter-axis convergence angle mechanically constrained to 4.58° by the 3D-printed mount.

At this working distance, the slight difference in horizontal field-of-view (25° for HR; 24° for LR) produces near-complete spatial overlap within the 150×150 mm abdominal ROI while keeping parallax effects geometrically consistent and limited in magnitude.

This controlled acquisition geometry ensures realistic cross-sensor observations while preserving sufficient spatial sampling density in the HR modality to serve as a reference domain.

3.7. Geometric Alignment

To obtain spatially aligned LR–HR image pairs, we register the LR sensor to the HR reference using a global planar homography. Because all acquisitions use a fixed camera rig, a controlled working distance of 1000 mm, and standardized supine subject positioning, the abdominal ROI lies approximately on the same imaging plane across subjects. A one-time calibration is therefore performed by manually selecting $N = 12$ cross-sensor correspondences. The homography matrix $\mathbf{H}$ is estimated using RANSAC-based least squares with a reprojection threshold of 3 px [11, 12]:

$$\mathbf{x}_{\mathrm{LR}} \sim \mathbf{H}\,\mathbf{x}_{\mathrm{HR}}, \tag{1}$$

where $\mathbf{x}$ denotes homogeneous image coordinates. The resulting calibration achieves a reprojection root-mean-square error (RMSE) of 1.12 px on the selected correspondences.

Since the camera positions remain fixed throughout data collection, the estimated homography is reused for all subjects and sequences. For each measurement, we define a 150×150 mm abdominal ROI relative to the umbilicus location. In the HR camera, this region corresponds to a 216×216 pixel crop. The LR content is projected onto this HR ROI grid using perspective warping.

To compensate for small residual misalignments, we apply an Enhanced Correlation Coefficient (ECC) [10] refinement step with a translation-only motion model. ECC is performed in the gradient domain (Sobel magnitude) and optimized using a maximum of 200 iterations with a convergence tolerance of 10^{-6}.

Across the dataset, ECC refinement was evaluated over 2052 sampled LR–HR frame pairs across subjects. The

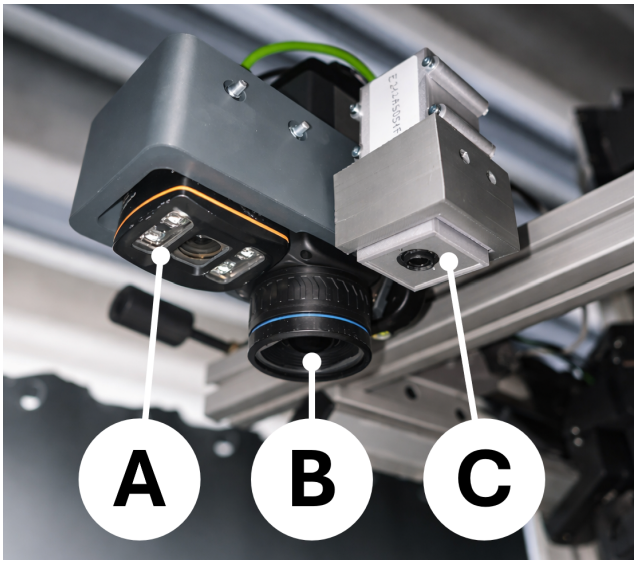


Figure 2. Dual-sensor acquisition mount used for **DIRT-SR38**. (A) IFM O3D303 ToF sensor used for distance verification only (not included in the dataset), (B) FLIR A655sc HR thermal camera, (C) SEEK Mosaic Core LR thermal camera.

refinement requires subpixel corrections on average (mean $|d_x| = 0.36$ px, mean $|d_y| = 0.60$ px), with 95th percentiles of 1.44 px and 1.60 px, respectively. The mean ECC correlation coefficient after convergence is 0.94, indicating stable cross-sensor registration.

3.8. Thermal Preprocessing

Thermal sensors may exhibit temporal drift and fixed-pattern effects, and commonly rely on NUC procedures, which can introduce global intensity discontinuities in practice [13, 20–22].

Preprocessing is applied exclusively to the LR modality; HR images remain unmodified and serve as the reference domain. To improve temporal consistency of the LR modality, we apply two preprocessing steps after geometric alignment: (i) robust NUC jump correction and (ii) monotonic warm-up stabilization.

3.8.1. NUC Jump Detection and Correction

Let I_t denote the aligned LR ROI at frame t, and let m_t be its median intensity. The median statistic is used due to its robustness against localized thermal extrema and spatial outliers. We compute the temporal difference

$$\Delta_t = m_t - m_{t-1}. \tag{2}$$

Abrupt global intensity offsets are consistent with gain/offset-based non-uniformity correction and drift phenomena reported for infrared focal-plane arrays [13, 21, 22]. To detect abrupt global intensity shifts, we evaluate a robust z-score of Δ_t using the median and median absolute deviation (MAD):

$$z_t = \frac{\Delta_t - \mathrm{median}(\Delta)}{1.4826 \cdot \mathrm{MAD}(\Delta)} [27]. \tag{3}$$

A frame is marked as a NUC event when $|z_t| \geq 8$. Detected jumps partition the sequence into temporal segments. For each segment boundary, we estimate the median intensity level immediately before and after the jump and apply a segment-wise additive offset to enforce median continuity across segments. Since the correction is spatially uniform within each frame, relative spatial temperature gradients remain unaffected.

3.8.2. Monotonic Warm-up Stabilization

During early acquisition, thermal sensors may exhibit stabilization effects that violate the expected monotonic rewarming behavior under the DIRT protocol. To enforce physically plausible dynamics, we apply isotonic regression to the sequence of median intensities $\{m_t\}$. Specifically, we solve

$$\min_{\hat{m}_t} \sum_t (m_t - \hat{m}_t)^2 \quad \text{s.t.} \quad \hat{m}_{t+1} \geq \hat{m}_t, \tag{4}$$

using the pool-adjacent-violators algorithm (PAVA) [1, 26]. The resulting correction $\hat{m}_t - m_t$ is applied as a frame-wise additive offset.

Together, these preprocessing steps reduce sensor-induced temporal artifacts while preserving relative temperature patterns relevant for dynamic infrared thermography analysis.

3.9. Low-Resolution Resampling

After geometric alignment, the HR ROI is defined at 216×216 pixels. Over the same physical region, the corresponding LR ROI spans 112×112 pixels, reflecting the native cross-sensor resolution difference and slight field-of-view mismatch.

To enable standardized $\times 2$ SR benchmarking, LR inputs are resampled to 108×108 pixels using area-based interpolation while HR targets remain at 216×216. The small adjustment in sampling density enforces an integer scaling factor while adding limited interpolation blur relative to the intrinsic sensor degradation.

3.10. Intended Use and Scope

DIRT-SR38 is intended for research on thermal image SR, cross-sensor degradation modeling, and DIRT analysis under controlled acquisition conditions; it is not intended for clinical diagnosis, medical decision making, or direct patient care.

4. Experiments

4.1. Experimental Setup

We evaluate $\times 2$ cross-sensor thermal SR on **DIRT-SR38**. The dataset contains recordings from 38 participants, each providing a steady-state baseline sequence and a dynamic rewarming sequence acquired under the standardized DIRT protocol.

To ensure subject independence between training and evaluation, we adopt a subject-wise split. Specifically, 30 subjects are used for training (24,300 paired frames), 4 for validation (3,240 pairs), and 4 for testing (3,240 pairs). This protocol prevents information leakage between subjects and reflects realistic deployment scenarios in which models must generalize to previously unseen individuals.

Each sample consists of a LR input image of size 108×108 and a corresponding HR target of size 216×216, corresponding to a $\times 2$ scaling factor.

All models are trained and evaluated on normalized 8-bit thermal images derived from the original radiometric recordings. The input therefore consists of single-channel thermal intensity images.

Performance is evaluated using Peak Signal-to-Noise Ratio (PSNR) and Structural Similarity Index (SSIM). Metrics are computed per frame over the full 216×216 ROI and

averaged across the complete test set.

4.2. Benchmark Methods

To obtain an architecturally diverse benchmark, we select methods spanning classical interpolation, transformer-based reconstruction, diffusion-based generative SR, and adversarial real-world SR. This diversity enables a structured comparison between distortion-oriented reconstruction models and perceptual or generative approaches under physically measured cross-sensor thermal degradations.

All benchmark models are implemented using their official architectures and training configurations as described in the original publications. No architecture modifications or task-specific hyperparameter tuning were performed, ensuring a fair and reproducible comparison.

Bicubic interpolation serves as a classical non-learning baseline and provides a lower-bound reference under purely interpolation-based upscaling.

SwinIR [17] is a shifted-window transformer for image restoration and represents strong distortion-oriented reconstruction optimized for PSNR/SSIM performance.

HAT [4] extends transformer-based SR with hybrid spatial and channel attention mechanisms, serving as a high-capacity state-of-the-art reconstruction model.

SinSR [32] formulates SR as a single-step diffusion process, allowing us to evaluate whether generative prior modeling benefits cross-sensor thermal SR.

ESRGAN [30] introduces adversarial perceptual training for texture enhancement, while **Real-ESRGAN** [31] extends this framework with degradation-aware modeling for real-world image restoration. Although originally designed for RGB imaging pipelines, it provides a useful baseline for evaluating real-world cross-sensor SR.

4.3. Training Regimes

We investigate three training regimes in order to analyze the difference between synthetic and real cross-sensor degradations.

4.3.1. Real-Paired Training

Models are trained directly on the physically measured LR–HR pairs provided by **DIRT-SR38**. This regime captures the true cross-sensor degradation characteristics present in the dataset, including differences in spatial sampling, optical transfer characteristics, and sensor noise.

4.3.2. Synthetic Training

In the synthetic regime, LR inputs are generated from the HR targets using bicubic downsampling with scale factor $\times 2$. This mirrors common single-image SR practice and serves as a reference setting for synthetic degradation modeling.

Table 2. Quantitative $\times 2$ SR results on DIRT-SR38. Training on measured real LR–HR pairs consistently outperforms training on synthetic bicubic degradations, confirming a substantial real-vs-synthetic domain gap. SinSR performs best overall, followed closely by HAT.

Method		PSNR ↑	SSIM ↑
Bicubic	Real	22.47	0.6853
	Synthetic	23.40	0.7709
SwinIR	Real	34.80	0.9724
	Synthetic	23.10	0.7709
HAT	Real	<u>35.25</u>	<u>0.9759</u>
	Synthetic	<u>26.03</u>	**0.9710**
SinSR	Real	**35.63**	**0.9772**
	Synthetic	**26.10**	<u>0.9616</u>
ESRGAN	Real	29.98	0.9641
	Synthetic	22.14	0.7140
Real-ESRGAN	Real	31.39	0.9657
	Synthetic	23.18	0.7424

4.3.3. Synthetic Pretraining with Real Fine-Tuning

Collecting real paired LR–HR observations is expensive and often results in relatively small training datasets. A common strategy in image restoration is therefore to first train models on synthetic degradations and subsequently adapt them to real measurements through fine-tuning.

Following this paradigm, we investigate a two-stage training strategy. Models are first pretrained using synthetic LR images generated from the HR targets via bicubic downsampling ($\times 2$). After convergence, the pretrained models are fine-tuned on the real paired cross-sensor observations provided by **DIRT-SR38**.

To analyze the data efficiency of this approach, we vary the number of subjects used during the real-data fine-tuning stage while keeping the validation and test sets fixed. Specifically, fine-tuning is performed using subsets of $\{2, 4, 6, 8, 10\}$ training subjects, corresponding to $\{1620, 3240, 4860, 6480, 8100\}$ real LR–HR frame pairs. For reference, the full training set contains 30 subjects (24,300 paired frames), meaning that the fine-tuning stage uses between 6.7% and 33.3% of the available real training data.

This training regime allows us to study how quickly pretrained models adapt to real cross-sensor degradations and whether synthetic pretraining improves reconstruction performance when only limited real paired data are available.

4.4. Results

4.4.1. Real-Paired vs. Synthetic Training

Table 2 summarizes SR performance on the **DIRT-SR38** test set for both real-paired and synthetic training regimes.

All learning-based approaches significantly outperform bicubic interpolation when trained on real paired

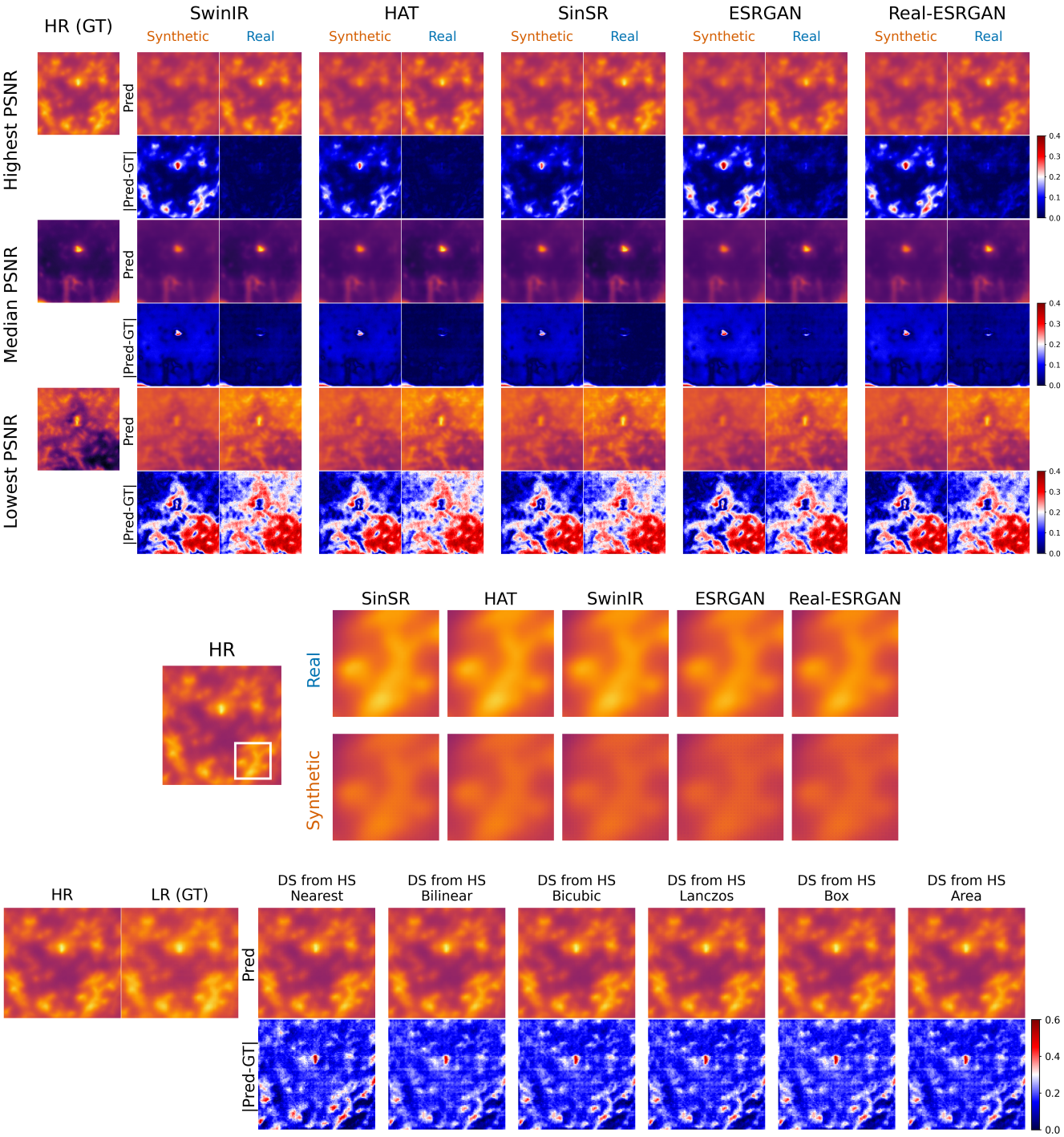


Figure 3. Qualitative analysis on **DIRT-SR38**. **Top:** three representative test subjects with highest, median, and lowest PSNR (top to bottom). For each subject, the leftmost image is the HR reference, followed by reconstructions from SwinIR, HAT, SinSR, ESRGAN, and Real-ESRGAN, each shown for models trained on synthetic bicubic LR data and on real paired LR–HR data. The row below each reconstruction shows the absolute error map $|\mathrm{Pred} - \mathrm{GT}|$ on a shared color scale. Real-paired training generally yields lower errors and reconstructions closer to the HR target. **Middle:** zoomed view of the boxed HR region, comparing real-trained (top row) and synthetic-trained (bottom row) models; real-paired training better preserves local thermal structure and contrast. **Bottom:** measured LR image versus synthetic LR images produced by standard HR downsampling operators, together with absolute error maps relative to measured LR. Despite their visual similarity, all interpolation-based degradations show structured residual errors, indicating that common synthetic downsampling does not match the real LR sensor formation process.

data. Among the evaluated methods, the diffusion-based SinSR model achieves the highest reconstruction accuracy (35.63 dB PSNR and 0.9772 SSIM), followed closely by the transformer-based HAT model (35.25 dB / 0.9759). SwinIR also performs strongly, confirming that distortion-oriented transformer architectures are effective for thermal SR.

GAN-based approaches such as ESRGAN and Real-ESRGAN obtain lower PSNR than the reconstruction-focused models. This is consistent with their perceptual training objectives, which emphasize visual realism rather than strict pixel-wise fidelity. In thermal imaging, where relative spatial temperature patterns are important, distortion-oriented reconstruction objectives appear better aligned with the evaluation metrics used in this work.

A substantial performance gap is observed when models are trained on synthetic bicubic downsampling instead of real sensor observations. For example, SinSR drops from 35.63 dB to 26.10 dB when trained on synthetic data, while HAT decreases from 35.25 dB to 26.03 dB. Across the evaluated architectures, the PSNR reduction ranges from approximately 7 to more than 11 dB. This consistent degradation indicates that bicubic downsampling does not adequately reproduce the degradation characteristics of the LR thermal sensor.

The qualitative comparisons in Fig. 3 support these quantitative trends. Across representative easy, median, and difficult test cases, models trained on real paired data generally produce reconstructions that more closely match the HR reference and yield lower absolute errors than the same architectures trained on synthetic bicubic degradations. The bottom panel further shows that, although interpolation-based synthetic LR images may appear visually plausible, they still differ systematically from the measured LR observations. These discrepancies help explain the large performance gap between synthetic and real-paired training.

Overall, the results highlight the importance of physically measured LR–HR pairs for developing and evaluating SR methods for thermal imaging systems.

4.4.2. Synthetic Pretraining with Real Fine-Tuning

We evaluate whether synthetic pretraining improves data efficiency when adapting to real cross-sensor thermal degradations. Due to computational constraints, this experiment is conducted only with HAT, which provides a strong reconstruction baseline while being substantially more efficient to train than SinSR.

HAT is first pretrained on synthetic bicubic $\times 2$ downsampling and subsequently fine-tuned on real LR–HR pairs from **DIRT-SR38**. During fine-tuning, we vary the number of subjects used for real supervision while keeping the validation and test sets fixed.

Table 3 shows that performance improves consistently as more real paired subjects are used for fine-tuning. With only two subjects, the model already reaches 32.94 dB PSNR, indicating that synthetic pretraining provides a useful initialization for real-data adaptation. Performance increases to 35.66 dB with eight subjects and saturates thereafter, with only a marginal gain at ten subjects.

Table 3. HAT performance with synthetic pretraining and fine-tuning on increasing amounts of real paired **DIRT-SR38** data. While synthetic pretraining alone performs poorly, adding even a few real subjects strongly improves PSNR. Fine-tuning on 8–10 subjects matches or slightly exceeds training on the full real-only set, indicating improved data efficiency.

Training Regime	Real LR–HR Pairs	PSNR ↑
Synthetic	0	26.03
Real (30 subjects)	24,300	35.25
Synthetic + Real (2 subjects)	1,620	32.94
Synthetic + Real (4 subjects)	3,240	33.76
Synthetic + Real (6 subjects)	4,860	34.33
Synthetic + Real (8 subjects)	6,480	35.66
Synthetic + Real (10 subjects)	8,100	35.66

Notably, performance after fine-tuning with 8–10 subjects matches and slightly exceeds the result obtained by training HAT directly on the full real training set (Table 2), despite using substantially fewer real subjects. This suggests that synthetic pretraining can reduce the amount of real paired data needed for strong performance. Since this experiment is limited to HAT, we do not claim that the same behavior necessarily transfers to other architectures.

5. Conclusion

We presented **DIRT-SR38**, the first publicly available real paired cross-sensor thermal sequence dataset acquired under a standardized clinical DIRT cooling–rewarming protocol. The dataset contains 30,780 aligned LR–HR frame pairs from 38 participants and supports realistic benchmarking of thermal image super-resolution under clinically relevant acquisition conditions.

Our experiments show a pronounced domain gap between synthetic bicubic training and real-paired cross-sensor training: across all evaluated SR models, physically measured LR–HR pairs yield substantially better performance than synthetically downsampled data. Using HAT, we further show that synthetic pretraining followed by real fine-tuning can recover strong performance with only a limited amount of real paired data.

Acknowledgments

This work was supported by the Research Foundation Flanders (FWO) (1SB9825N; W. Clarys, 12A6126N; T. De Kerf, T001723N) and by a Tier-1 HPC starting grant from the Flemish Supercomputer Center (VSC).

References

[1] R. E. Barlow and H. D. Brunk. The isotonic regression problem and its dual. *Journal of the American Statistical Association*, 67(337):140–147, 1972. 5

[2] Jianrui Cai, Hui Zeng, Hongwei Yong, Zisheng Cao, and Lei Zhang. Toward real-world single image super-resolution: A new benchmark and a new model. In *Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV)*, pages 3086–3095, 2019. 1, 2

[3] Edgar Cardenas De La Hoz, Jan Verstockt, Simon Verspeek, Warre Clarys, Filip E. F. Thiessen, Thierry Tondu, Wiebren A. A. Tjalma, Gunther Steenackers, and Steve Vanlanduit. Automated thermographic detection of blood vessels for diep flap reconstructive surgery. *International Journal of Computer Assisted Radiology and Surgery*, 19(9):1733–1741, 2024. 1, 2

[4] 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 (CVPR)*, pages 22367–22377, 2023. 2, 6

[5] Warre Clarys, Rhys Evans, Simon Verspeek, Jan Verstockt, Hai Zhang, Veronique Verhoeven, Wiebren A. A. Tjalma, Filip Thiessen, and Gunther Steenackers. Optimising neural networks for perforator detection in diep flap breast reconstruction using dynamic infrared thermography. *Quantitative InfraRed Thermography Journal*, pages 1–11, 2025. 1, 2

[6] Warre Clarys, Simon Verspeek, Jan Verstockt, Stefan Hummelink, Gunther Steenackers, and Filip Thiessen. Comparative study of cooling techniques for perforator detection in diep flap reconstruction using dynamic infrared thermography. *J. Plast. Reconstr. Aesthet. Surg.*, 109:144–151, 2025. 2, 3

[7] Warre Clarys, Valentine Hotome, Stefan Hummelink, Simon Verspeek, Veronique Verhoeven, Wiebren A. A. Tjalma, Gunther Steenackers, and Filip Thiessen. Dynamic infrared thermography in free flap surgery: A systematic review. *JPRAS Open*, 48:281–304, 2026. 1, 2

[8] Warre Clarys, Simon Verspeek, Thomas De Kerf, Jan Verstockt, Steve Vanlanduit, Filip Thiessen, and Gunther Steenackers. Dirt-sr38: Real paired cross-sensor thermal sequence dataset, 2026. 2, 3

[9] Chao Dong, Chen Change Loy, Kaiming He, and Xiaoou Tang. Image super-resolution using deep convolutional networks. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 38(2):295–307, 2016. 2

[10] Georgios D. Evangelidis and Emmanouil Z. Psarakis. Parametric image alignment using enhanced correlation coefficient maximization. *IEEE Trans. Pattern Anal. Mach. Intell.*, 30(10):1858–1865, 2008. 4

[11] Martin A. Fischler and Robert C. Bolles. Random sample consensus: A paradigm for model fitting with applications to image analysis and automated cartography. *Communications of the ACM*, 24(6):381–395, 1981. 4

[12] Richard Hartley and Andrew Zisserman. *Multiple View Geometry in Computer Vision*. Cambridge University Press, 2 edition, 2003. 4

[13] Majeed M. Hayat, Sergio N. Torres, Ernest Armstrong, Stephen C. Cain, and Brian Yasuda. Statistical algorithm for nonuniformity correction in focal-plane arrays. *Applied Optics*, 38(5):772–780, 1999. 5

[14] Hannah Eliza John, Vachara Niumsawatt, Warren M. Rozen, and Iain S. Whitaker. Clinical applications of dynamic infrared thermography in plastic surgery: A systematic review. *Gland Surgery*, 5(2):122–132, 2016. 1, 2

[15] Jiwon Kim, Jung Kwon Lee, and Kyoung Mu Lee. Accurate image super-resolution using very deep convolutional networks. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 1646–1654, 2016. 2

[16] Christian Ledig, Lucas Theis, Ferenc Huszár, Jose Caballero, Andrew Cunningham, Alejandro Acosta, Andrew P. Aitken, Alykhan Tejani, Johannes Totz, Zehan Wang, and Wenzhe Shi. Photo-realistic single image super-resolution using a generative adversarial network. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 4681–4690, 2017. 2

[17] 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 (ICCV) Workshops*, pages 1833–1844, 2021. 2, 6

[18] Olivia M. Lin, David J. Hunter-Smith, and Warren M. Rozen. Thermal challenges in dynamic infrared thermography used for perforator mapping. *Journal of Reconstructive Microsurgery*, 40(4):268–275, 2024. 1, 2

[19] Zhisheng Lu, Juncheng Li, Hong Liu, Chaoyan Huang, Linlin Zhang, and Tieyong Zeng. Transformer for single image super-resolution. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR) Workshops*, pages 457–466, 2022. 2

[20] Robert Olbrycht, Bogusław Wiecek, and Gilbert De Mey. Thermal drift compensation method for microbolometer thermal cameras. *Applied Optics*, 51(11):1788–1794, 2012. 5

[21] Tomasz Orżanowski. Nonuniformity correction algorithm with efficient pixel offset estimation for infrared focal plane arrays. *SpringerPlus*, 5:1831, 2016. 5

[22] D. L. Perry and E. L. Dereniak. Linear theory of nonuniformity correction in infrared staring sensors. *Optical Engineering*, 32(8):1854–1859, 1993. 5

[23] Rafael E. Rivadeneira, Angel D. Sappa, and Boris X. Vintimilla. Thermal image super-resolution: A novel architecture and dataset. In *Proceedings of the 15th International Conference on Computer Vision Theory and Applications (VISAPP)*, pages 111–119, 2020. 1, 2

[24] 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/CVF Conference on*

Computer Vision and Pattern Recognition Workshops, 2020. 2

[25] 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 *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR) Workshops*, pages 418–426, 2022. 1, 2

[26] Tim Robertson, F. T. Wright, and Richard L. Dykstra. *Order Restricted Statistical Inference*. John Wiley & Sons, 1988. 5

[27] Peter J. Rousseeuw and Christophe Croux. Alternatives to the median absolute deviation. *Journal of the American Statistical Association*, 88(424):1273–1283, 1993. 5

[28] Filip E. F. Thiessen, Thierry Tondu, Nicolas Vermeersch, Ben Cloostermans, Ralv Lundahl, Bart Ribbens, Lawek Berzenji, Veronique Verhoeven, Guy Hubens, Gunther Steenackers, and Wiebren A. A. Tjalma. Dynamic infrared thermography (dirt) in deep inferior epigastric perforator (diep) flap breast reconstruction: Standardization of the measurement set-up. *Gland Surgery*, 8(6):799–805, 2019. 2

[29] Alexander Ulrichsen, Thomas De Kerf, David Dunphy, Paul Murray, Steve Vanlanduit, and Stephen Marshall. A true hyperspectral image super-resolution dataset. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR) Workshops*, pages 4421–4430, 2025. 1, 2

[30] Xintao Wang, Ke Yu, Shixiang Wu, Jinjin Gu, Yihao Liu, Chao Dong, Yu Qiao, and Chen Change Loy. Esrgan: Enhanced super-resolution generative adversarial networks. In *Computer Vision – ECCV 2018 Workshops*, pages 63–79. Springer, 2018. 2, 6

[31] Xintao Wang, Liangbin Xie, Chao Dong, and Ying Shan. Real-esrgan: Training real-world blind super-resolution with pure synthetic data. In *Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV) Workshops*, pages 1905–1914, 2021. 1, 2, 6

[32] Yufei Wang, Wenhan Yang, Xinyuan Chen, Yaohui Wang, Lanqing Guo, Lap-Pui Chau, Ziwei Liu, Yu Qiao, Alex C. Kot, and Bihan Wen. Sinsr: Diffusion-based image super-resolution in a single step. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 25796–25805, 2024. 6

[33] Sverre Weum, James B. Mercer, and Louis de Weerd. Evaluation of dynamic infrared thermography as an alternative to ct angiography for perforator mapping in breast reconstruction: a clinical study. *BMC Medical Imaging*, 16:43, 2016. 1, 2

[34] Kai Zhang, Jingyun Liang, Luc Van Gool, and Radu Timofte. Designing a practical degradation model for deep blind image super-resolution. In *Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV)*, pages 4791–4800, 2021. 2

[35] Yicheng Zhou, Yuan Liu, Liyin Yuan, Qian Chen, Guohua Gu, and Xiubao Sui. Real-infraredsr: Real-world infrared image super-resolution via thermal imager. *Optics Express*, 31(22):36171–36187, 2023. 1, 2