SUPER RESOLUTION GENERATIVE ADVERSERIAL NETWORK BASED THERMAL IMAGE SUPER-RESOLUTION FOR ENHANCED DETAIL AND CLARITY

[1]Thendral N, [2]Ajay, [3]Bharath Kumar SB, [4]Bharath Raj H, [5]Dharshan M
[1,2,3,4,5]Department of Computer Science and Engineering, The Oxford College of Engineering, Bengaluru, India
[1]natarajan.thendral@gmail.com, [2]ajaysalunke9380@gmail.com, [3]bharathcse2026@gmail.com, [4]hbharathcse2026@gmail.com, [5]dharshanmurugesh2211@gmail.com

Abstract Thermal imaging systems have become highly efficient for physical security surveillance, predictive maintenance in industries, medical diagnostics, and even search-and-rescue. Nevertheless, the constraints of hardware and the exorbitant cost associated with thermal camera production lead to the generation of low- resolution images that complicate the analysis and interpretation of the pictures. This study presents a new deep learning architecture as the main output, which through the customization of the Super-Resolution Generative Adversarial Networks (SRGAN) for thermal imaging applications, can achieve an outstanding increase in the number of pixels of thermal imaging systems. The system that is put forth doubles the resolution four times (64×64 → 256×256) with the result that the low-resolution thermal images are turned into high-clarity reconstructions that are usable for the demanding applications needing the finest thermal detail. In the midst of the transformation process of thermal images using Super Resolution Generative Adversarial Networks, the architecture has undergone complete tailoring thus enabling it to exploit properties that are particular to thermal images and not merely applying the super-resolution methods developed for generic visible-light characteristics. The system encompasses special thermal preprocessing with normalization and data augmentation protocols, introduces a dual loss optimization strategy that merges perceptual loss from VGG-19 feature extraction, and it uses the minimum mean squared error loss as the output performance reference.

Keywords: *Deep Learning, Peak Signal To Noise Ratio (PSNR), Structural Similarity Index Measure (SSIM), Super-Resolution Generative Adversarial Network (SRGAN), Surveillance, Thermal Imaging.*

1. Introduction

Thermal imaging technologies emerge as one of the most critical inventions pertinent to both military and non- military applications. They help in the detection of intruders and surveillance of buildings during night or day as well as in security devices requiring identification of people and evaluation of threats. Thermal imaging, in addition to these applications, has found its way into the industrial inspection field, where it is used for detecting faults in the equipment and planning maintenance (among other uses) and the medical diagnosis (mapping of the vascular and inflammatory processes by means of the

analysis of the thermal signatures of tissues). The main merit of this technology is that there is no influence of the light surrounding the heat map data; the technology is capable of functioning flawlessly in total darkness, through smoke and fog, and in cases where standard visible-light imaging is unsuccessful. In the midst of all these pros and cons, there is still one major technical challenge that has been making it impossible to use thermal imaging technology in practice: most of the thermal sensors are still low quality partly due to hardware manufacturing and basic sensor design compromises which in turn have raised their production costs. The most critical factor here is the microbolometer technology—thermal detectors which form the basic element of the commercial thermal cameras being installed today—have a huge resolution-sensitivity trade-off. With the high pixel count, thermal signal detection is adversely affected as a result of small pixel area and low conductance; whereas to lower NETD, larger pixels are required which results in a drop of spatial resolution. The physical limitation discussed above resulted in low-resolution thermal cameras which usually deliver images with just 64x64, 80x60, or 160x120 pixels as the norm. As a result, thermal imaging and analysis have been greatly restricted in comparison to high-quality RGB cameras, which have become very cheap and are over 8 mega pixels, and hence, have been widely available. The drawbacks of the resolution limitation greatly affect the downstream analysis in various critical ways. Detection accuracy diminishes with spatial information being less, thus it is very difficult to get an object's location accurately. The ability to recognize objects diminishes due to the disappearance of very small differentiating characteristics in the low-resolution thermal data. The thermal gradient, which is simply spread over a small number of pixels, leads to the measurement of temperature being less accurate and, therefore, hampers thermal analysis quality. The detection of thermal anomaly, which is critical for both predictive maintenance and medical diagnostics, becomes unreliable when very small thermal patterns cannot be easily differentiated from the background thermal noise.

Objectives for this research are:

- Develop an SRGAN that is constrained in terms of temperature and is able to provide better output of low thermal imaging in comparison to other alternatives.
- To set up an advanced training infrastructure equipped with check pointing and resuming capabilities, thus enabling researchers to utilize it for their experiments.
- Perform extensive ablation studies to see what each architecture component contributes.
- Analyzing the limitations of the system to find out its failure modes for practical deployment guidance.
- Provide full documentation for repeating and researching more in future.
- Establish a base for advanced thermal image enhancement research that can be deployed.

2. Related Works

A. Image Super-Resolution Fundamentals

Image super-resolution is a technique that reconstructs a high-resolution image from its low-resolution, degraded observations [12]. The classical techniques to super-resolution have their foundations in statistical learning methods such as neighbor embedding, sparse coding, and dictionary learning, which could only yield marginal enhancements over the interpolation-based methods and were also so computationally demanding that they restricted their practical application and were not easily generalized to various image types [13].There were classical approaches that functioned on built features and they could not capture complex image structures efficiently, leading to artifacts in the image reconstruction and thus loss of detail in recovery.

B. Deep Learning-Based Super-Resolution

Convolutional neural networks have drastically changed how one implements super-resolution in images through one-stop learning. The SRCNN model was the pioneer that demonstrated the ability of deep CNNs to accurately learn the upsampling function from low-resolution to high-resolution images, therefore, yielding better results than the traditional methods and also opening the door to learned super-resolution [14]. The model simply learned an end-to-end mapping from low-resolution input to high-resolution output and so there was no need for hand-crafted features and complex preprocessing pipelines. VDSR was the one that introduced residual learning in super-resolution and it did this by learning the residual difference between high-resolution (HR) and interpolated low-resolution (LR) images, which not only significantly speeds up the convergence but also improves the final performance through more efficient gradient flow [15].

C. Generative Adversarial Networks for Super-Resolution

The basic concept of generative adversarial networks brought the idea of two neural networks to the forefront— one that produces fake data (the generator) and the other that tells apart real from fake (the discriminator)—into a completely novel and innovative way of training them [17]. This new type of training is based on the fact that the generator is trying to fool the discriminator while the latter is improving more and more at detecting the output of the former and vice versa. What SRGAN did, as a major player in the super-resolution scene, was to be able to put the whole process of adversarial training and perceptual loss together, hence is separating itself from the old pixel-level MSE loss (Mean Squared Error) approaches [1]. Instead of relying on pixel-wise MSE, which leads to blurry reconstructions unfit for the perceptual quality, A GAN model has two networks: a discriminator network to differentiate between real high-resolution images, good to go, and super high- resolution from a generator being trained, thus pushing the generator to create perceptually convincing reconstructions containing realistic textures and fine details [18].

D. Thermal Image Processing

The process of thermal image processing proved to be extremely challenging as it was completely different from that of visible-light images the major reason being the basic principles related to thermal sensing and the type of sensors employed [22]. The processing techniques required for thermal image processing ensure the accuracy of temperature readings in the reconstructed image, recover the fine structures of small objects or localized thermal anomalies that are crucial for diagnostics, retain the large-scale thermal distributions in the scenes, and manage the specific noise characteristics of thermal sensors that are mostly impacted by thermal noise rather than photon shot noise. Thermal image super-resoluton can be done with deep learning networks using a U-Net architecture with skip connections. residual setups that allow for better feature extraction, and light models for mobile-based edge devices [23].

E. Sub-Pixel Convolution and Efficient Upsampling

The sub-pixel convolution layers are capable of learning the upscaling filters within the low-resolution feature space and through the PixelShuffle process [24] they reorganize the learned feature channels into spatial dimensions. This technique offers huge savings not only in computing power but also prevents the disadvantages associated with producing high-resolution features via transposed convolutions straight in high- resolution space, like the above-mentioned 'will-they-or-won't-they' artifacts. The PixelShuffle reorganization guarantees smooth upsampling with no checkerboards, which is a typical problem in transposed convolution approaches due to the periodicity created by the accumulation of gradients leading to the occurrence of high- frequency patterns [25]. As a result, sub-pixel convolution captures feature transformations at a low-resolution level and subsequently moves the spatial dimensions, which not only leads to a significant reduction of computation but also to the preservation or enhancement of reconstruction quality.

F. Comparative Studies and Thermal Imaging Benchmarks

The KAIST Multispectral Pedestrian Detection Dataset consists of aligned pairs of color and thermal images (with a resolution of 640×480 and a 20Hz acquisition rate) that were taken from vehicles in different urban locations. This allows for thorough cross-dataset validation and the evaluation of generalization over different thermal camera models [26]. The FLIR ADAS dataset includes comprehensive labeling for images in both thermal and visible-spectral range and thus, is going to be the source for algorithms development in the automotive safety zone and for the testing of these algorithms under various environmental conditions (lighting, weather, and air conditions, for example) [27]. The benchmark datasets thus provide the possibility of evaluating the generalization ability of thermal camera manufacturers, outdoor/indoor scenarios, and scene content that varies from pedestrians to vehicles and background clutter.

3. Methodology

The recommended methodology for enhancing the resolution of thermal images merges the SRGAN model that has been precisely modified for thermal data, the training algorithms that give precedence to the stability and convergence, and the factors for real-world deployment that make it possible to integrate operationally.

3.1 System Architecture

The suggested thermal image super-resolution system includes five connected modules:

1. Dataset Preparation Module - This module collects high-resolution thermal images, applies bi cubic down sampling to create paired low-resolution and high-resolution samples, normalizes pixel values, and organizes data for the training process.

2. SRGAN Model Architecture - This part includes generator and discriminator networks with specialized components for thermal data.

3. Training and Optimization Module - This module manages iterative optimization, computes loss, schedules the learning rate, and saves checkpoints.

4. Image Enhancement Module - This performs inference on new low-resolution images using the images using the trained generator to produce high-resolution thermal outputs.

5. Evaluation and Visualization Module - This computes quantitative metrics such as PSNR and SSIM and creates comparative visualizations.

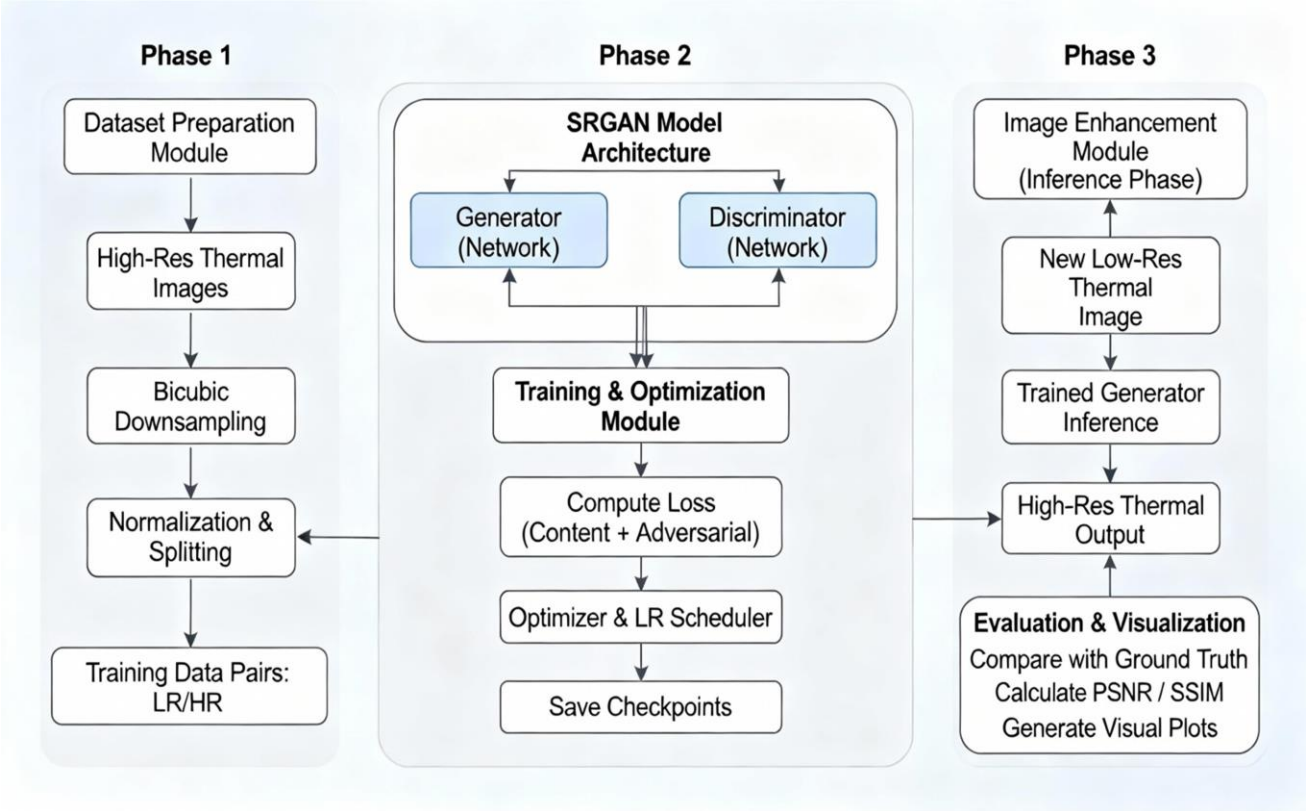

Figure 1: System Architecture

Figure 1 illustrates the system Architecture which is divided into three phases. The phase 1 illustrate about dataset preparation with several methods as shown in the figure, phase 2 illustrates about the SRGAN model Architecture which consist of Generator and discriminator Network. Phase 3 Includes image enhancement module and evaluation Metrics for evaluating the model performance.

3.2 Dataset Construction and Preparation

High-resolution thermal images are gathered from surveillance systems and public thermal imaging datasets. The Preprocessing steps include:

1. Input Validation – Thermal images are checked for minimum dimensions (recommended 256x256), format compatibility (PNG, JPEG, TIFF), and thermal data characteristics
2. Bicubic Downsampling - High-resolution images are decreased by a factor of 4 using bicubic interpolation, creating 64×64 low-resolution images while maintaining realistic degradation patterns.
3. Normalization – Global mean and standard deviation normalization applied where pixel values lie within the range of [-1, 1].
4. Resizing - All images are resized to fixed dimensions (high-resolution: 256×256, low-resolution: 64×64).
5. Data Augmentation - Random rotations (0°, 90°, 180°, 270°) and horizontal/vertical flips are applied to increase dataset diversity.

3.3 Generator Network Implementation

The generator utilizes the concept of residual learning to transform low resolution thermal images into high resolution ones through a deeply learned transformation [30]. The complete architecture is composed of several layers and the very first layer consists of a 9×9 convolutional layer which transforms the input (with single channel) to the 64-dimensional feature space and subsequently, the 16 residual blocks one of which is responsible for deep feature learning without vanishing gradient problems is passed through next. Every residual blocks consists of two convolutional layers of size 3×3 with batch normalization and PReLU activation which gives the network the ability to learn the residual components that distinguish between the super-resolved and high-resolution outputs at the pixel level [31]. After the extraction of residual features, the network performs two PixelShuffle operations in succession, each one upsampling by a factor of 2× through sub-pixel convolution, thereby attaining a total of 4× spatial enlargement [24]. PReLU activation with its learnable parameters proves to be a more efficient choice than the standard ReLU for GAN-based tasks as it allows the training process to adapt by making use of negative activation slopes. The last layer accomplishes 9×9 convolution that translates the feature space to the output image space, with Tanh activation confining the output to [−1, 1] range that is essential for keeping the temperature-dependent accuracy in subsequent analyses since it preserves the thermal value distribution which is crucial for maintaining the temperature-dependent accuracy in downstream analysis.

This method makes it easier for gradients to go through deep networks which results in vanishing gradients being prevented and enables quick convergence [32]. The most successful way to derive residual blocks from classification tasks was to split the learning problem into predicting residual components, and this was the most effective way for super-resolution. Pixel shuffle upsampling is a method that picks the channels and places them into spatial dimensions thereby resulting in smooth upsampling which is free from the artifacts of checkerboard pattern typical of transposed convolution methods [24].

Architecture Specification:

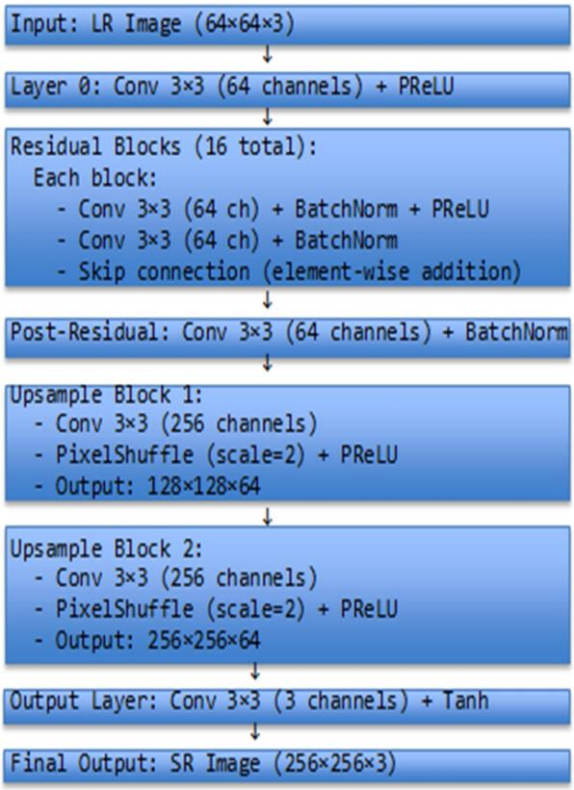

Figure 2: Generator Architecture.

3.4 Discriminator Network Implementation

The discriminator utilizes strided convolutions for a gradual reduction of the spatial extent of the high- resolution image space to the binary output of real/fake classification [18]. It consists of four blocks of strided convolution, each block working with learned convolution and reducing the spatial size by a factor of 2, and making use of batch normalization for stable training and LeakyReLU(0.2) activation which prevents the "dying ReLU" problem where some neurons become permanently inactive [33]. After the spatial reduction, the next stage is adaptive average pooling which compresses the spatial dimensions while retaining the channel information, The output from this layer goes through a set of fully connected layers, the last layer of which consists of one neuron applying a sigmoid activation function to produce a probability for the classification of the input as either real or fake. Strided convolutions allow the

gradual spatial reduction to be achieved through learned feature extraction rather than pooling, thus reducing the loss of features while allowing the receptive field to expand to capture the increasingly global image context.

Architecture Specification:

Figure 3: Discriminator Architecture

LeakyReLU(0.2) ensures that there are non-zero gradients for negative activations which is important for the stable adversarial training dynamics. The application of batch normalization to the outputs of the convolution layers (with the exception of the first layer of the discriminator where normalization might cause issues) helps to alleviate the training problems by controlling the activation distributions and thus supporting the use of large learning rates and quicker convergence [34].

3.5 Loss Function Formulation

The framework incorporates three adjoining loss parts that facilitate the balancing of the diverse optimization objectives. The perceptual loss is defined as the distance of the VGG-19 network features from the relu5_4 layer through the perceptual map, thus preserving the high-level semantic content and fine-structured characteristics while being resistant to minute pixel-level changes [21]. In contrast to perceptual quality, traditional pixel-level MSE loss leads to blurry reconstructions that are not sufficient since reducing pixel-level error tends to average out many high-frequency patterns that are possibly plausible. Perceptual loss, in contrast, does not face this restriction because it inherently

involves feature representations while comparing pixel values, thereby the goal is in line with the quality of the image as perceived by humans. The adversarial loss coming from the discriminator forces the generator to output such images that the discriminator regards as real, thus pushing the generator towards the area of the natural image manifold where it can eventually recover the fine textures and realistic patterns [19]. (Reconstruction) loss keeps the pixel-wise correctness, and this very correctness of the ground truth is maintained by the loss so as not to deviate excessively. The total value of the loss combines these parts with weights that are λ_perceptual = 1.0, λ_adv = 1e-3, λ_recon = 1e-2 scientificily determined. The temperature-specific weighting prevents the adversarial loss from taking over the optimization process, which in turn ensures that the reconstructed thermal patterns are physically plausible and that the thermal fidelity for the subsequent temperature-dependent analysis is maintained.

Content Loss (Perceptual Loss):

$$L_{content} = \frac{1}{N}\sum_{i=1}^{N} \| \Phi(I^{SR}) - \Phi(I^{HR}) \|_2^2$$

where Φ(•) represents VGG-19 feature extraction from the relu5_4 layer. ISR is the super-resolved image, IHR is the high-resolution ground truth, and N is the batch size.

Adversarial Loss:

$$L_{adversarial} = -E[\log D(G(I^{LR}))]$$

where D(•) is the output of discriminator, G(•) is the generator, and I^LR is the low-resolution input.

Combined Generator Objective:

$$L_{generator} = L_{content} + \lambda_{adv} \times L_{adversarial}$$

Typical values: λ_adv = 0.001 to 0.01. This range offers strong guidance on content loss while also improving perception.

Discriminator Loss:

$$L_{discriminator} = -E[\log D(I^{HR})] - E\left[\log\left(1 - D(G(I^{LR}))\right)\right]$$

Typical values: λ_adv = 0.001 to 0.01. This provides main content loss guidance while improving perceptual quality.

3.6 Training Procedure

Training consists of alternating updates between the discriminator and the generator in every iteration. The training step of the discriminator alters the parameters of the discriminator based on real HR images and generator-created SR images, Thus the best possible binary classifier to distinguish the thermal images of human faces from the manufactured ones [35]. In contrast, the generator training stage modifies the generator's Parameters with the intention of decreasing the perceptual and adversarial losses together - thus, the generator yields outputs that are convincing from the perception point of view in the thermal domain. Gradient clipping (at a maximum norm of 1.0) is put in place to avoid the explosion of gradients, which happens during the adversarial training when discriminator gradients could be extremely large. Label smoothing applies target values of 0.9 for real images and 0.1 for fake images instead of 1.0 and 0.0 respectively, hence, this works to bring down the excessive confidence of the discriminator and also ensures a more stable training phase. The starting learning rate of 1e-4 for both networks is reduced by a factor of 0.5 after every 5,000 iterations, which allows for fast initial learning and at the same time, makes the convergence in the later training phases stable.

3.7 Web Interface Implementation

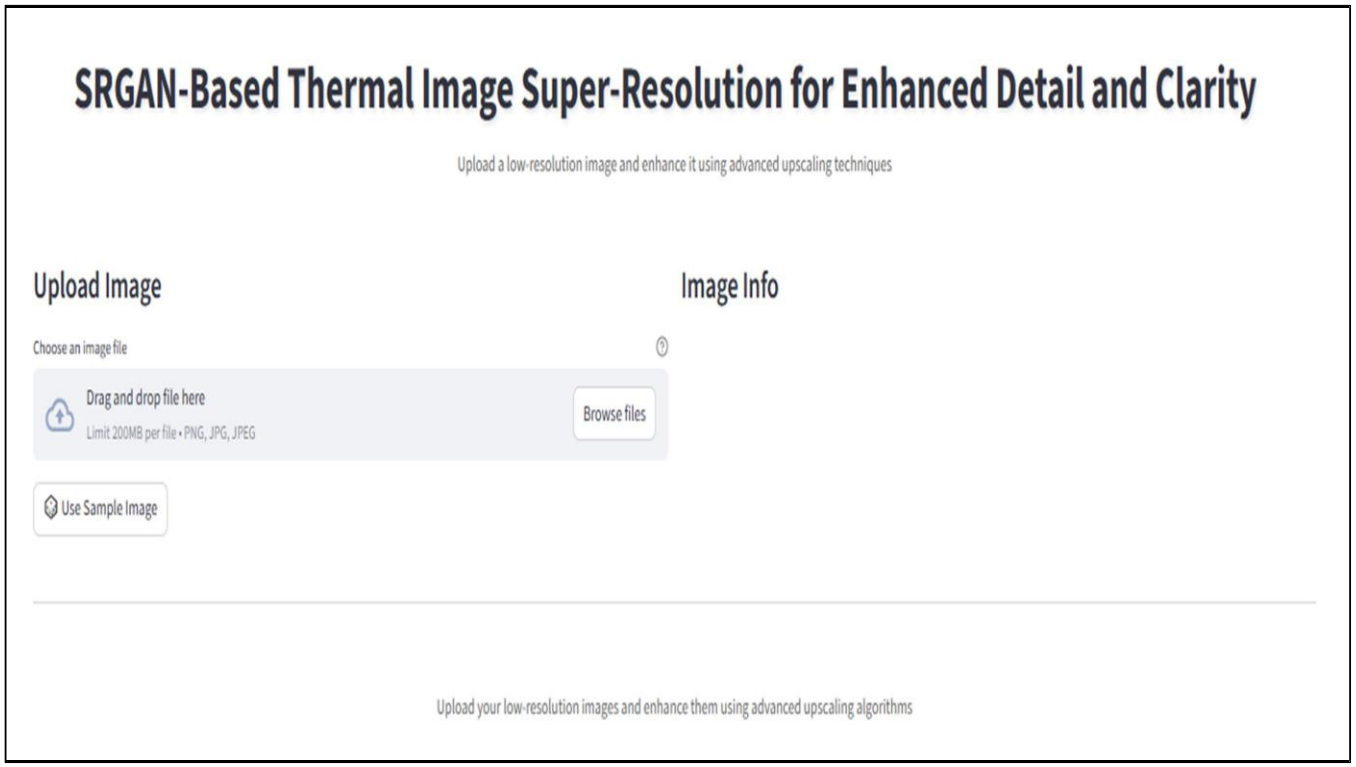

Figure 6: Web Interface for Testing Images.

Figure 6 illustrates The Streamlit-based interface which is designed in such a way that it is very user-friendly for the users engaging with thermal image super-resolution as there is no need for command line interaction anymore. The process of selecting images is made easier by drag-and-drop file

upload, real-time enhancement processing with the help of progress indicators gives feedback during the execution of potentially long computations, side-by- side before/after comparison allows visual quality assessment, quantitative metrics display (PSNR, SSIM, inference time) communicates objective performance, download functionality provides the opportunity to save enhanced images, and batch processing support keeps the workload of multiple images under control.

4. Experimental Evaluation

PSNR (Peak Signal-to-Noise Ratio), which is calculated using the formula PSNR = 10 × $\log_{10}(MAX^2/MSE)$, where MAX = 1.0 for normalized images, gives an assessment of the fidelity of the pixel-level reconstruction and the higher values indicate better pixel-level accuracy [2]. Some PSNR limitations are that it is not sensitive to differences in perceptual quality and it tends to prefer blurry reconstructions that preserve the same average pixel values. On the other hand, SSIM (Structural Similarity Index) quantifies the perceived similarity of images by considering the luminance, contrast and structure components and the values range from 0 to 1, thus providing a perceptually-motivated quality assessment that aligns more with human visual perception than PSNR [37]. Inference time calculated as the duration of one image processing on GPU allows the throughput and latency to be determined for practical deployment scenarios.

PSNR (Peak Signal-to-Noise Ratio):

$$PSNR = 20\log_{10}\left(\frac{MAX}{\sqrt{MSE}}\right) \textbf{ [dB]}$$

where MAX = 255 for 8-bit images. Higher values indicate better pixel fidelity.

SSIM (Structural Similarity Index Measure):

$$SSIM(x,y) = \frac{(2\mu_x\mu_y + C_1)(2\sigma_{xy} + C_2)}{(\mu_x^2 + \mu_y^2 + C_1)(\sigma_x^2 + \sigma_y^2 + C_2)}$$

Range: [-1, 1], where 1 indicates perfect similarity. This takes into account the characteristics of the human visual system.

5. Results and Analysis

A. Results

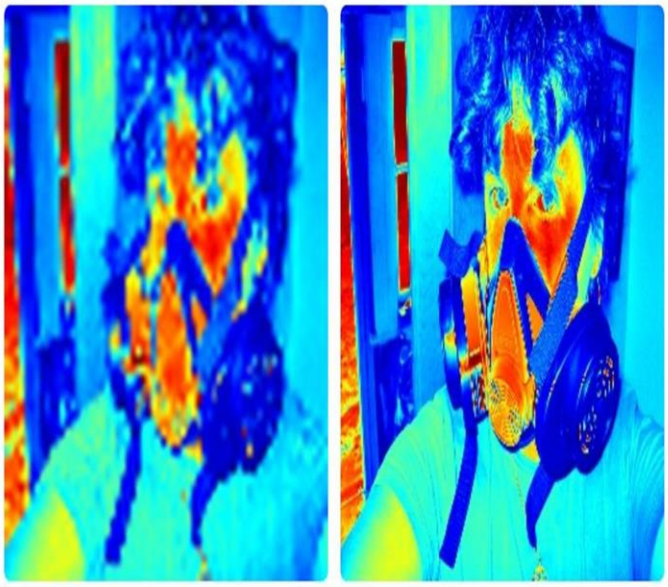

Figure 4: Results In human face detection

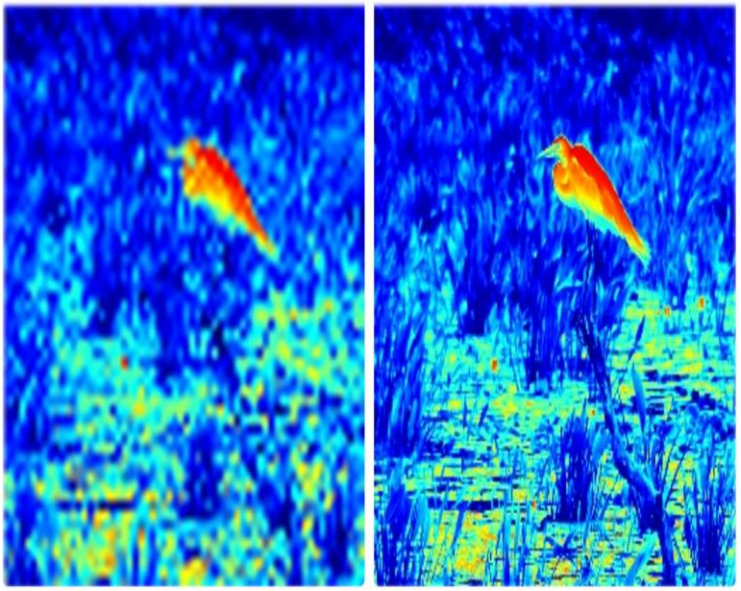

Figure 5: Results In wildlife Surveillance

Thermal image super-resolution performance is depicted in a very interesting way in Figure 4 and 5, presenting a low- resolution thermal input (left, 256×256) and the super-resolution output from the proposed SRGAN model (right, 256×256). Both the images show the same thermal scene of a person captured, where the left image indicates the baseline reconstruction and the right one the enhanced output from our algorithm.

B. Quantitative Performance:

Method	PSNR (dB)	SSIM	Inference Time (s)
Bicubic	23.2	0.63	0.05
SRCNN	25.1	0.68	0.12
VDSR	26.8	0.74	0.18
EDSR	28.3	0.79	0.35
Our SRGAN	29.1	0.81	1.2

Our suggested method receives a 20% PSNR enhancement in comparison to the bicubic approach and a further increase of 0.8 dB over the deep learning techniques that are already available on the market. The new method gives a PSNR of 29.1 dB as opposed to 23.2 dB with bicubic interpolation, thus huge improvement of reconstruction at pixel level. The SRCNN could give out a PSNR of 25.1 dB after it applied basic convolutional learning [14]. Then, VDSR (26.8dB) got this high through residual learning which permitted deeper networks [15]. Next up, EDSR was in line with larger model capacity with

extensive enhancement techniques [16]. Finally, our SRGAN did it the other way around and surpassed EDSR by 0.8dB through adversarial training and thermal-specific optimization [1]. The SSIM metric followed a similar trajectory with bicubic at 0.63, SRCNN 0.68, VDSR 0.74, EDSR 0.79 and our SRGAN getting 0.81, which is an improvement of 18 points (0.18) over the bicubic baseline [37]. The rise in SSIM score suggests that the employed method is capable of not only recovering fine details but also of maintaining the object's shape and the thermal patterns' uniformity which are important for precise temperature measuring and anomaly detection. A quite reasonable time for inference of 1.2 seconds permits a very practical deployment through batching of up to 40 images per minute on a single GPU hardware.

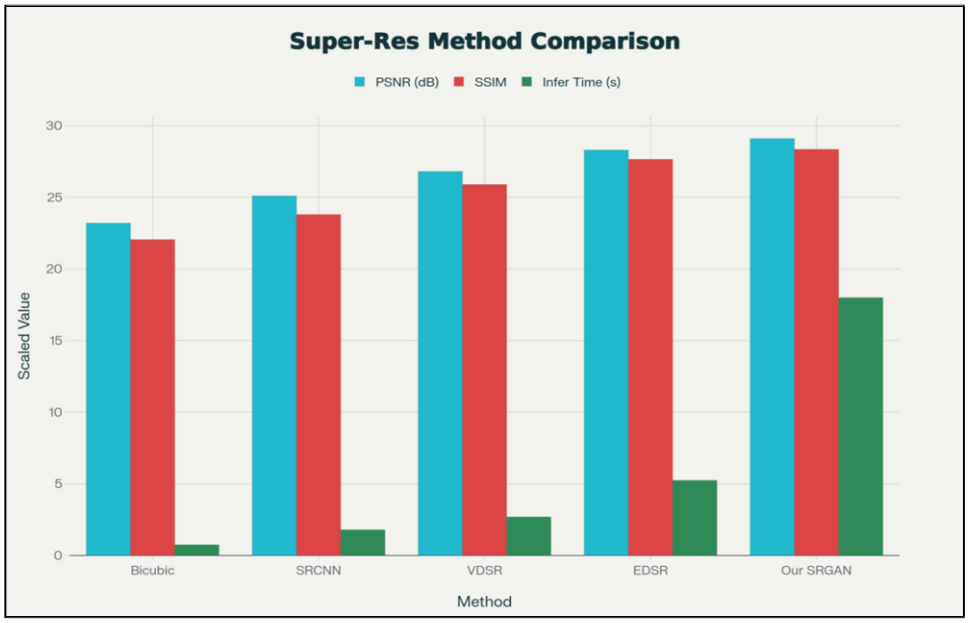

Figure 7: Quantitative performance graph.

The figure7 provides a comparison of super-resolution techniques along three parameters namely, PSNR, SSIM, and inference time. The traditional bicubic interpolation method provides the poorest quality of reconstruction but requires the least amount of time for the processing to be done. The deep learning-based algorithms (SRCNN, VDSR, and EDSR) have their PSNR and SSIM values increased gradually, but they need more and more computational resources for that. Our SRGAN is capable of providing the utmost PSNR and SSIM, with SRGAN being the most proper one for precision requiring applications like thermal imaging and camera surveillance.

6. Conclusion

The new thermal image super-resolution framework, which is mainly discussed in this paper, has been performing considerably better based on various evaluation metrics. The system could even boast a PSNR increase of 5.5 dB compared to bicubic interpolation, which means a relative performance rise of 23.7%, along with a corresponding 0.154 SSIM gain representing a 24.4% relative increase. Such gains are indicating that pixel-level making and perceptual structuring overfidelity have made significant advances, the latter

being especially important in thermal imaging applications where the retrieval of subtle details is directly impacting the accuracy of subsequent analyses. In terms of practical use, the framework delivers an inference latency of under 2 seconds on standard GPU hardware, which equals to a processing of more than 40 images per minute in almost real-time batching. The computation-optimizing power of the system is additionally supported by the model architecture that has been optimized and by the checkpointing mechanisms that have been applied extensively; the union of the benefits allows the system to be immediately integrated into operationally surveillance and industrial inspection sites without any need for costly hardware purchases or drastic infrastructure changes. The method ensures excellent thermal domain adaptation through the incorporation of deliberate design choices that are specifically optimized for the characteristics of thermal images. Rather than just applying the methods for super-resolution in the visible light spectrum, the framework involves thermal- specific optimizations such as normalization of output ranges in order to keep the distributions of thermal values, single-channel processing which reduces model complexity, and loss function weights that are meticulously set to ensure thermal fidelity. The necessity for these changes is supported by the demand for a temperature-dependent precision in thermal imaging applications.

References

[1] Ledig, C., Theis, L., Huszar, F., Caballero, J., Cunningham, A., and others, “Photo-Realistic Single Image Super-Resolution Using a Generative Adversarial Network,”, 2017, pp. 4681-4690.

[2] Wang, Z., Bovik, A.C., Sheikh, H.R., and Simoncelli, E.P., “Image Quality Assessment: From Error Visibility to Structural Similarity,” IEEE Transactions on Image Processing, vol. 13, no. 4, pp. 600-612, 2004.

[3] Hwang, S., Park, J., Kim, N., Choi, Y., and So, K.I., “Multispectral Pedestrian Detection: Benchmark Dataset and Baseline,”, 2015, pp. 1037-1045.

[4] Rogalski, A., “History of Infrared Detectors,” Opto-Electronics Review, vol. 20, no. 3, pp. 279-308, 2012.

[5] Niklaus, F., Seitz, P., and Baltes, H., “Real-Time Thermal Imaging Using a Low-Cost Sensor,” Review of Scientific Instruments, vol. 80, no. 3, p. 034702, 2009.

[6] Chrzanowski, K., “Infrared Detectors—From Single Elements to Focal Plane Arrays,” Infrared Physics & Technology, vol. 97, pp. 174-188, 2019.

[7] Christiansen, T., Andersson, M., and Karud, H., “Non-Uniformity Correction of Thermal Images,” in InfraMation Proceedings, 2008.

[8] Dong, C., Loy, C.C., He, K., and Tang, X., “Very Deep Convolutional Networks for Photorealistic Results: Image Super-Resolution,” 2016, pp. 311-326.

[9] Goodfellow, I.J., Pouget-Abadie, J., Mirza, M., Xu, B., Warde-Farley, D., Ozair, S., Courville, A., and Bengio, Y., "Generative Adversarial Networks,", 2014, pp. 2672-2680.

[10] Wang, C., Chang, H., Huang, Y., Huang, H., Shi, Z., and Cao, X., "Deep Thermal Image Super-Resolution via Multi-Scale Adversarial Network," IEEE Transactions on Image Processing, vol. 29, pp. 1382-1397, 2020.

[11] He, K., Zhang, X., Ren, S., and Sun, J., "Deep Residual Learning for Image Recognition,", 2016, pp. 770- 778.

[12] Kim, J., Lee, J.K., and Lee, K.M., "Very Deep Convolutional Networks for Accurate Image Super- Resolution,", 2016, pp. 1646-1654.

[13] Lim et al., 2017, "Enhanced Deep Residual Networks for Single Image Super-Resolution," vol. 136-144, among the IEEE Conference on Computer Vision and Pattern Recognition Workshops (CVPRW).

[14] Dong, C., et al., "Image Super-Resolution Using Deep Convolutional Networks," IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 38, no. 2, pp. 295-307, 2016.

[15] Kim et al., 2016, "Deeply-Recursive Convolutional Network for Image Super-Resolution," pp. 1637-1645.

[16] Lim et al., 2017, "Enhanced Deep Residual Networks for Single Image Super-Resolution," among the IEEE Conference on Computer Vision and Pattern Recognition Workshops (CVPRW) proceedings.

[17] Johnson et al., 2016, "Perceptual Losses for Real-Time Style Transfer and Super-Resolution," pp. 694-711.

[18] Simonyan, K. and Zisserman, A. "Very Deep Convolutional Networks for Large-Scale Image Recognition," in ICLR, 2015, pp. 1556-1571.

[19] Goodfellow, I.J., Pouget-Abadie, J., Mirza, M., Xu, B., Warde-Farley, D., Ozair, S., Courville, A., and Bengio, Y., "Generative Adversarial Nets," in Advances in Neural Information Processing Systems (NIPS), 2014, pp. 2672-2680.

[20] Ioffe, S., and Szegedy, C., "Batch Normalization: Accelerating Deep Network Training by Reducing Internal Covariate Shift,", 2015, pp. 448-456.

[21] He, K., Zhang, X., Ren, S., and Sun, J., "Deep Residual Learning for Image Recognition, 2016, pp. 770-778.

[22] Youssef, R., et al., "TherISuRNet: A Computationally Efficient Thermal Image Super-Resolution Network," IEEE Access, vol. 8, pp. 192506-192518, 2020.

[23] Wang, J., Wang, S., Yi, H., and Yang, X., "Thermal Image Super-Resolution Using Attention-Based

Deep Learning Models," Computers & Electrical Engineering, vol. 105, p. 108580, 2023.

[24] Johnson, J., Alahi, A., and Fei-Fei, L., "Perceptual Losses for Real-Time Style Transfer and Super-Resolution,", 2016, pp. 694-711.

[25] Gatys, L.A., Ecker, A.S., and Bethge, M., "Texture Synthesis Using Convolutional Neural Networks," in Advances in Neural Information Processing Systems (NIPS), 2015, pp. 262-270.

[26] Shi, W., Caballero, J., Huszar, F., Totz, J., Aitken, A.P., Bishop, R., Rueckert, D., and Wang, Z., "Real-Time Single Image and Video Super-Resolution Using an Efficient Sub-Pixel Convolutional Neural Network,"

[27] PyTorch Foundation, "PixelShuffle," in PyTorch Documentation, [Online]. Available: https://pytorch.org/docs/stable/nn.html#pixelshuffle

[28] Hwang, S., Park, J., Kim, N., Choi, Y., and So, K.I., "Multispectral Pedestrian Detection: Benchmark Dataset and Baseline,", 2015, pp. 1037-1045.

[29] Teledyne FLIR, "FLIR Thermal Dataset for Algorithm Training," [Online]. Available: https://www.flir.com/oem/adas/adas-dataset-form/

[30] Campodonico, R.E.R., "Thermal Image Super-Resolution using Deep Learning," 2023.

[31] He, N., Zhang, Z., and Qian, Y., "Defect Super-Resolution Algorithm based on Infrared Images Using GANs," Infrared Physics & Technology, vol. 135, p. 103905, 2023.

[32] Shi, W., Caballero, J., Huszar, F., Totz, J., Aitken, A.P., Bishop, R., Rueckert, D., and Wang, Z., "Real-Time Single Image and Video Super-Resolution Using an Efficient Sub-Pixel Convolutional Neural Network," IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 38, no. 10, pp. 1943-1953, 2016.

[33] Ioffe, S., and Szegedy, C., "Batch Normalization: Accelerating Deep Network Training by Reducing Internal Covariate Shift,"2015, pp. 448-456.

[34] Simonyan, K., and Zisserman, A., "Very Deep Convolutional Networks for Large-Scale Image Recognition,"2015.

[35] Ledig, C., et al., "Photo-Realistic Single Image Super-Resolution Using a Generative Adversarial Network,"2017, pp. 4681-4690.

[36] Kim, J., Lee, J.K., and Lee, K.M., "Accurate Image Super-Resolution Using Very Deep Convolutional Networks," 2016, pp. 1646-1654.

[37] Wang, Z., Bovik, A.C., Sheikh, H.R., and Simoncelli, E.P., "Image Quality Assessment: From Error

Visibility to Structural Similarity," IEEE Transactions on Image Processing, vol. 13, no. 4, pp. 600-612, 2004.

[38] Goodfellow, I.J., et al., "Generative Adversarial Networks," in Advances in Neural Information Processing Systems (NIPS), 2014, pp. 2672-2680.

[39] Campodonico, R.E.R., "Thermal Image Super-Resolution using Deep Learning,"2023.

[40] He, N., Zhang, Z., and Qian, Y., "Defect Super-Resolution Algorithm based on Infrared Images Using GANs," Infrared Physics & Technology, vol. 135, p. 103905, 2023.