Physics-Informed Generative Adversarial Network for Infrared Image Generation

Murdock Aubry
University of Toronto
Department of Computer Science
murdock@cs.toronto.edu

Rémi Grzeczkowicz
University of Toronto
Department of Computer Science
remigrz@cs.toronto.edu

Abstract

Infrared (IR) thermography, or thermal imaging, captures an image of an object by detecting the infrared radiation emitted from it. Generating IR images from RGB data reduces the need for additional sensors, improving efficiency in terms of time, cost, and flexibility. However, current methods treat IR image generation as a stylistic transformation [11], rather than focusing on accurately reflecting the physical properties of the IR spectrum. This limitation often leads to poor performance in low-light and high-entropy environments. To address this, [11] introduced TeVNet, a physics-informed model that uses a specialized loss function to evaluate the physical accuracy of generated IR images. Building on this, we integrate their physics-based learning module into IR-GAN [10], a conditional generative adversarial network (GAN) for RGB to IR image generation. Additionally, we incorporate pre-processing modules inspired by [2], which enhance the defining features of an input embedding, resulting in stronger output fidelity. We benchmark our model against other leading methods and observe substantial performance improvements under various image conditions. Following an ablation study, we conclude that the physics-based loss terms *significantly* improve the performance of the base IR-GAN model, outperforming all other models in most metrics. Our results emphasize the importance of embedding physical principles into existing architectures to enhance both performance and realism. Our code is available at: https://github.com/jacedoir/irgan.

1 Introduction

Infrared imaging technology has garnered considerable attention for its reliable sensing capabilities in low visibility conditions. It has various civilian (image segmentation for self driving cars [5], forest fire monitoring [1]) and military applications [3]. With the rise of Generative Adversarial Networks [4] and convolutional neural networks (CNNs) [6], multi-spectral image translation has become possible. Consequently, numerous studies have attempted to reliably and accurately convert the abundant RGB images into infrared images. However, most existing image translation methods overlook the underlying physical principles, treating infrared images as mere stylistic variations [11]. This limitation restricts their practical application. In this project, we obviate the limitations of current methods by incorporating both a physics-based loss module and image pre-processing techniques. Utilizing IR-GAN [10], a generative adversarial network for RGB to IR image translation as our foundation, we implement a physics-based loss inspired by [11], as well as extensive image pre-processing methods [2] to significantly improve performance.

1.1 Contributions

To the best of our knowledge, we are the first to integrate a physics-based loss term into a generative adversarial network for *any* image generation task. By embedding fundamental physical principles within a data-driven framework, we bridge the gap between conventional deep learning approaches and physics-informed methodologies, presenting a novel paradigm for improving generative models. Furthermore, our method incorporates pre-processing techniques that are often overlooked in the infrared image generation literature, highlighting their potential to enhance performance. Together, these innovations establish a new benchmark for leveraging physics-aware strategies in generative tasks.

1.2 Limitations

Our ablation studies (Table 2) demonstrate that the physics-based loss term objectively enhances model performance, resulting in significant improvements beyond the training set when compared to the standard IR-GAN architecture.

Additionally, the studies indicate that the preprocessing module limits our performance and requires further investigation.

The preprocessing techniques did not result in any significant improvement in performance across tasks or metrics. Originally designed to produce 16-channel outputs for a forward pass through a transformer block, the GAN architecture in this implementation requires 3-channel inputs. Future work should prioritize the development of preprocessing techniques optimized for 3-channel outputs.

2 Related Work

Thermal GAN [8] pioneered adversarial training for infrared (IR) image generation, synthesizing thermal images from RGB inputs and inspiring further advancements. Decao, Ma, et al. [10] introduced IR GAN, using a conditional GAN with a U-Net-based UConvNeXt architecture to improve texture and edge preservation. They later proposed a dual-encoder architecture for aerial imagery [9], though challenges with high-entropy images persist.

Wadsworth et al. [13] provide a paired IR-RGB dataset and an inverse IR-to-RGB model but offer limited architectural innovation, often missing key elements like vehicles and pedestrians. Sirui Wang et al. [14] propose a ”Lightweight Pyramid Across-Scale” module with minor gains over IR GAN but lack implementation details and focus on low-entropy images.

Yijia Chen et al. [2] incorporate pre-processing techniques and a vision transformer block, leveraging self-attention for architecture design insights. Finally, Mao et al. [11] propose TevNet, a physics-based loss network, which inspires this study by promising to enhance GAN performance.

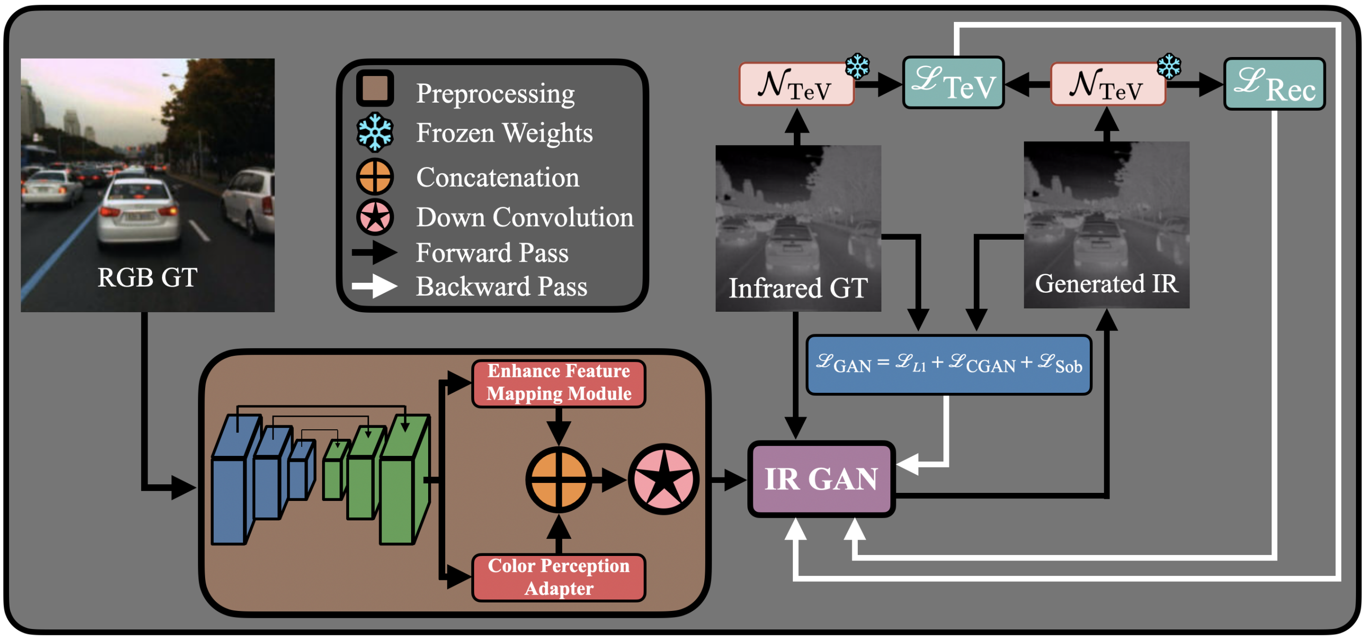
RGB GT
Preprocessing
Frozen Weights
Concatenation
Down Convolution
Forward Pass
Backward Pass
$\mathcal{N}_{\text{TeV}}$
$\mathscr{L}_{\text{TeV}}$
$\mathcal{N}_{\text{TeV}}$
$\mathscr{L}_{\text{Rec}}$
Infrared GT
Generated IR
$\mathscr{L}_{\text{GAN}} = \mathscr{L}_{L1} + \mathscr{L}_{\text{CGAN}} + \mathscr{L}_{\text{Sob}}$
Enhance Feature Mapping Module
Color Perception Adapter
IR GAN

Figure 1: **Network Architecture.** The figure displays the architectural construction of the entire physics-informed IR-GAN with convolutional pre-processing steps. **This is our original diagram.**

3 Methods

The global architecture for the methodology proposed in this work is delineated in Figure 7. Our model consists of three main components:

1. **IR-GAN.** The generative backbone of the model, employing a UConvNeXt-based conditional GAN architecture to enhance feature extraction and achieve high-quality infrared image synthesis.
2. **TeV Decomposition.** Image decomposition model to ensure that the generated sample abides by physical laws.

3. **Pre-processing.** A series of convolutional layers that extract latent infrared features, map them onto the infrared domain, and enhance fine-grained texture representation through multi-scale processing.

3.1 Infrared Generative Adversarial Network

The IR-GAN architecture, illustrated in Figure 2, is a conditional generative adversarial network (CGAN) designed to generate high-quality infrared (IR) images from visible images. This approach improves upon a standard CGAN by incorporating a UConvNeXt backbone, which enhances feature extraction and image generation performance.

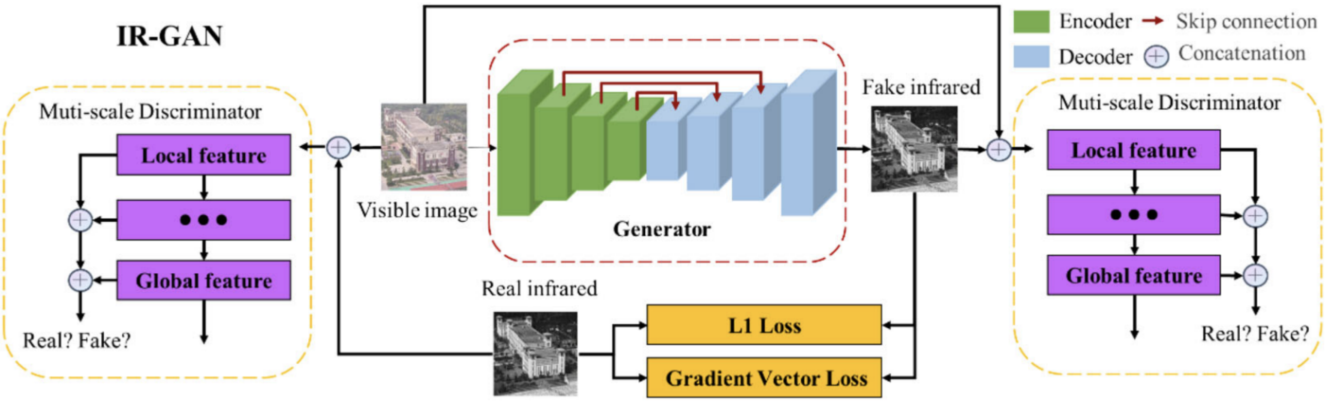

Figure 2: **Infrared GAN Architecture.** The figure illustrates the architectural design of IR-GAN [10], highlighting its core components.

The IR-GAN loss function can be expressed as follows:

$$\mathcal{L}_{\text{GAN}} = \arg\min_{G}\min_{D} \ell_{\text{CGAN}}(G, D) + \lambda_{L1}\ell_{L1}(G) + \lambda_{\text{Sobelov}}(G_{\text{Sobelov}}) \tag{1}$$

where

$$\min_{G}\min_{D} \ell_{\text{CGAN}}(G, D) = \mathbb{E}_{x,y\sim p_{\text{data}}}[\log D(x, y)] + \mathbb{E}_{x,y\sim p_{\text{data}}}[\log(1 - D(x, y))] \tag{2}$$

and

$$\ell_{L1}(G) = E_{x,y}\left[\|y - G(x)\|_1\right] \qquad \text{and} \qquad \ell_{\text{Sobelov}}(G_{\text{Sobelov}}) = E_{x,y}\left[\|y_{\text{Sobelov}} - G(x)_{\text{Sobelov}}\right]. \tag{3}$$

In our implementation, we utilize $\lambda_{L1} = 100.0$ and $\lambda_{\text{Sobelov}} = 15.0$.

3.2 TeV Decomposition

We implement the physics-based module introduced by [11] into IR-GAN [10]. The loss function of the vanilla IR-GAN model can be expressed as Equation 8. We propose implementing an additional architectural component which decomposes infrared images into its temperature $\mathbf{T}$, emissivity $\mathbf{e}$, and thermal texture $\mathbf{V}$. This is achieved by implementing the TeVNet decomposition

$$\tilde{\mathbf{e}}, \tilde{\mathbf{T}}, \tilde{\mathbf{V}} = \mathcal{N}_{\text{TeV}}(\mathcal{S}). \tag{4}$$

The TeVNet module is then trained to recontruct the infrared image given the three channel components:

$$\tilde{\mathcal{S}} = \text{Rec}(\tilde{\mathbf{e}}, \tilde{\mathbf{T}}, \tilde{\mathbf{V}}). \tag{5}$$

We can then construct two distinct loss terms:

- **Reconstruction Loss.** The first loss is the reconstruction loss of the generated image, referred to as $\mathcal{L}_{\text{Rec}}$, and is given by

$$\mathcal{L}_{\text{Rec}} = \|\text{Rec}(\mathcal{N}_{\text{TeV}}(\hat{\mathbf{x}}_0)) - \hat{\mathbf{x}}_0\|_2^2. \tag{6}$$

  If the generator accurately captures and maps the distribution of the infrared domain, the $\mathcal{L}_{\text{Rec}}$ of the generated image,x_0 will be minimal.

- **TeV Decomposition Loss.** This second loss function measures discrepancies in TeV space. Provided the ground truth infrared image, we compute

$$\mathcal{L}_{\text{TeV}} = \|\mathcal{N}(\hat{\mathbf{x}}_0) - \mathcal{N}(\mathbf{x}_0)\|_2^2. \tag{7}$$

  where $\hat{\mathbf{x}}_0$ is the translated image and $\mathbf{x}_0$ is the ground truth image. TeVNet quantifies the physically unreasonable aspects of the translated images from the TeV space perspective.

TeVNet is trained according to Figure 3, taken from [11]. The cumulative loss function which is used to train the GAN is then given by

$$\mathcal{L}_{\text{GAN}} = \arg\min_{G}\min_{D} \ell_{\text{CGAN}}(G, D) + \lambda_{L1}\ell_{L1}(G) + \lambda_{\text{Sobelov}}(G_{\text{Sobelov}}) + \lambda_{Rec}\mathcal{L}_{\text{Rec}} + \lambda_{\text{TeV}}\mathcal{L}_{\text{TeV}} \tag{8}$$

where we set $\lambda_{Rec} = \lambda_{TeV} = 15.0$.

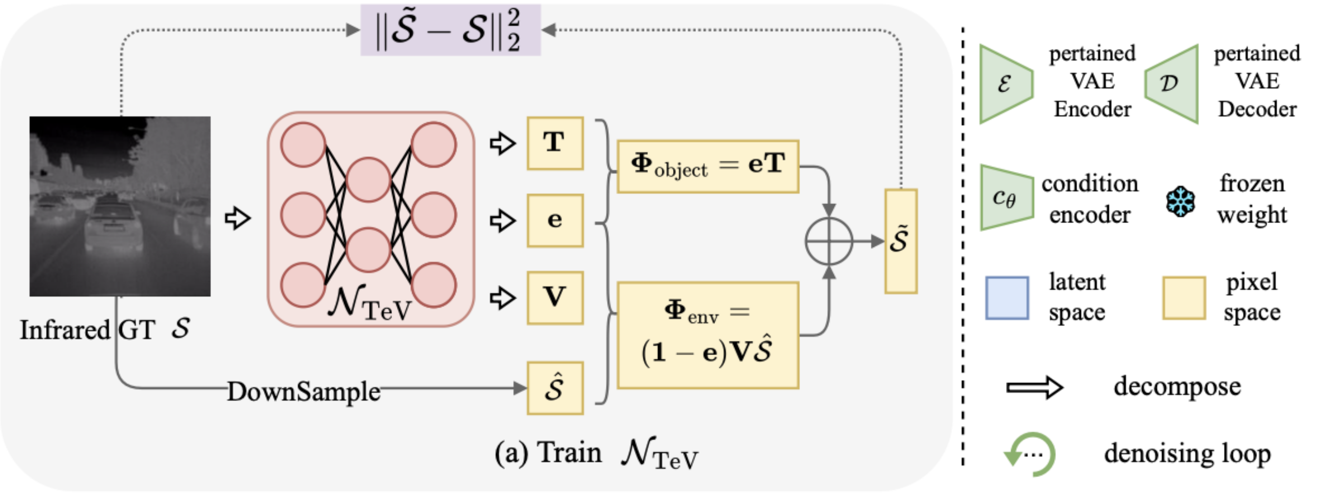
$\|\tilde{\mathcal{S}} - \mathcal{S}\|_2^2$
Infrared GT $\mathcal{S}$
$\mathcal{N}_{\text{TeV}}$
T
e
V
$\hat{\mathcal{S}}$
DownSample
$\mathbf{\Phi}_{\text{object}} = \mathbf{eT}$
$\mathbf{\Phi}_{\text{env}} = (\mathbf{1} - \mathbf{e})\mathbf{V}\hat{\mathcal{S}}$
$\tilde{\mathcal{S}}$
(a) Train $\mathcal{N}_{\text{TeV}}$
$\mathcal{E}$
pertained VAE Encoder
$\mathcal{D}$
pertained VAE Decoder
c_θ
condition encoder
frozen weight
latent space
pixel space
decompose
denoising loop

Figure 3: **TeVNet Training Architecture** This will serve as our foundational architecture. Subsequently, we will implement preprocessing and physical loss to develop our unique architecture. This schematic was taken directly from [11].

3.3 Image Pre-processing

To combat the prominence of issues in low-light and high-entropy settings, we propose implementing the pre-processing techniques of [2]. In particular, utilizing the Color Perception Adapter (CPA) and Enhancing Feature Mapping Module (EFMM), the architecture is displayed in Figure 7.

3.3.1 Color Perception Adapter (CPA)

The Color Perception Adapter (CPA) bridges the visible and infrared image domains by incorporating critical infrared spectral details into the processing pipeline. Conventional cameras capture RGB images but lack sensitivity to the infrared spectrum, which contains valuable non-visible information. CPA addresses this gap with two components:

1. **Infrared Feature Extraction:** Extracts latent infrared features from visible images, identifying imperceptible characteristics.

2. **Feature Adaptation and Mapping:** Maps these features onto the infrared pixel domain to reconstruct infrared aspects of the input.

By combining RGB and infrared features through convolutional processing, CPA generates detailed infrared images, enhancing spectral representation for downstream applications.

3.3.2 Enhanced Feature Mapping Module

The Enhanced Feature Mapping Module (EFMM) is designed to extract fine-grained texture features through a multi-scale processing framework. By employing downsampling at various scales, the EFMM reduces the spatial resolution of input features, enabling the detection of broader patterns and structures.

These multi-scale features are subsequently processed through convolutional layers followed by upsampling operations. This process results in the generation of multiple single-channel Detail Perception Enhancement Modules (DPEMs), each carefully calibrated to match the dimensions of the original input features.

The refined features from these enhancement modules are then integrated with the original features and subjected to additional convolutional processing. The outcome is a set of highly detailed, discriminative feature representations that significantly enhance the model's capacity to capture and utilize fine-grained details.

3.4 Training Data

We train both the TeVNet module and the infrared GAN on both the KAIST [7] and VEDAI [12] openly available datasets. The KAIST dataset comprises videos of driving scenes, categorized into three typical environments: campus, road, and downtown. Each scene type is captured during both daytime and nighttime conditions, providing a diverse range of lighting scenarios. In contrast, the VEDAI dataset focuses on aerial images of vehicles, offering a variety of challenges such as multiple orientations, lighting and shadow variations, specular reflections, and occlusions.

These datasets were chosen to test our model in a variety of settings. The dichotomy between the human-scale images of KAIST and the large-scale images with nuanced details of VEDAI allows us to systematically and holistically study the performance of our model and test the impact of each of the pre-processing and TeV decomposition modules in isolation.

3.5 Training Specifications

TeVNet was trained for 200 epochs on each dataset using one NVIDIA T4 GPU for 12 GPU hours. IRGAN was trained using one NVIDIA RTX 3090. On the KAIST dataset, the model was trained for 80 epochs during 82 GPU hours. For VEDAI, it was trained for 200 epochs during 4 GPU hours.

4 Results

4.1 Sample Outputs

4.1.1 Pre-processing Outputs

This section presents the comparison of pre-processed and generated images for the VEDAI dataset. The figure below highlights the different stages of processing, including the original RGB images, pre-processed versions, the generated infrared (IR) images, and the ground truth IR images. By showcasing these outputs, we can better understand the impact of pre-processing and the quality of the generated IR images relative to the true ground truth.

4.1.2 TeVNet Decomposition Outputs

Figure 5 displays sample outputs from the TeVNet module, which decomposes infrared images into three key components: temperature (T), emissivity (e), and thermal texture (Φ_{env}). The decomposition provides valuable insights into the distinct physical properties of the thermal scene, which are essential for accurate infrared image generation and analysis. In the figure, the top row showcases outputs from the KAIST dataset, while the bottom row presents results from the VEDAI dataset.

By separating these components, TeVNet enhances our understanding of the underlying thermal characteristics and aids in generating refined infrared images. This decomposition can also serve as a foundation for further analysis and optimization of infrared-based techniques in various practical applications.

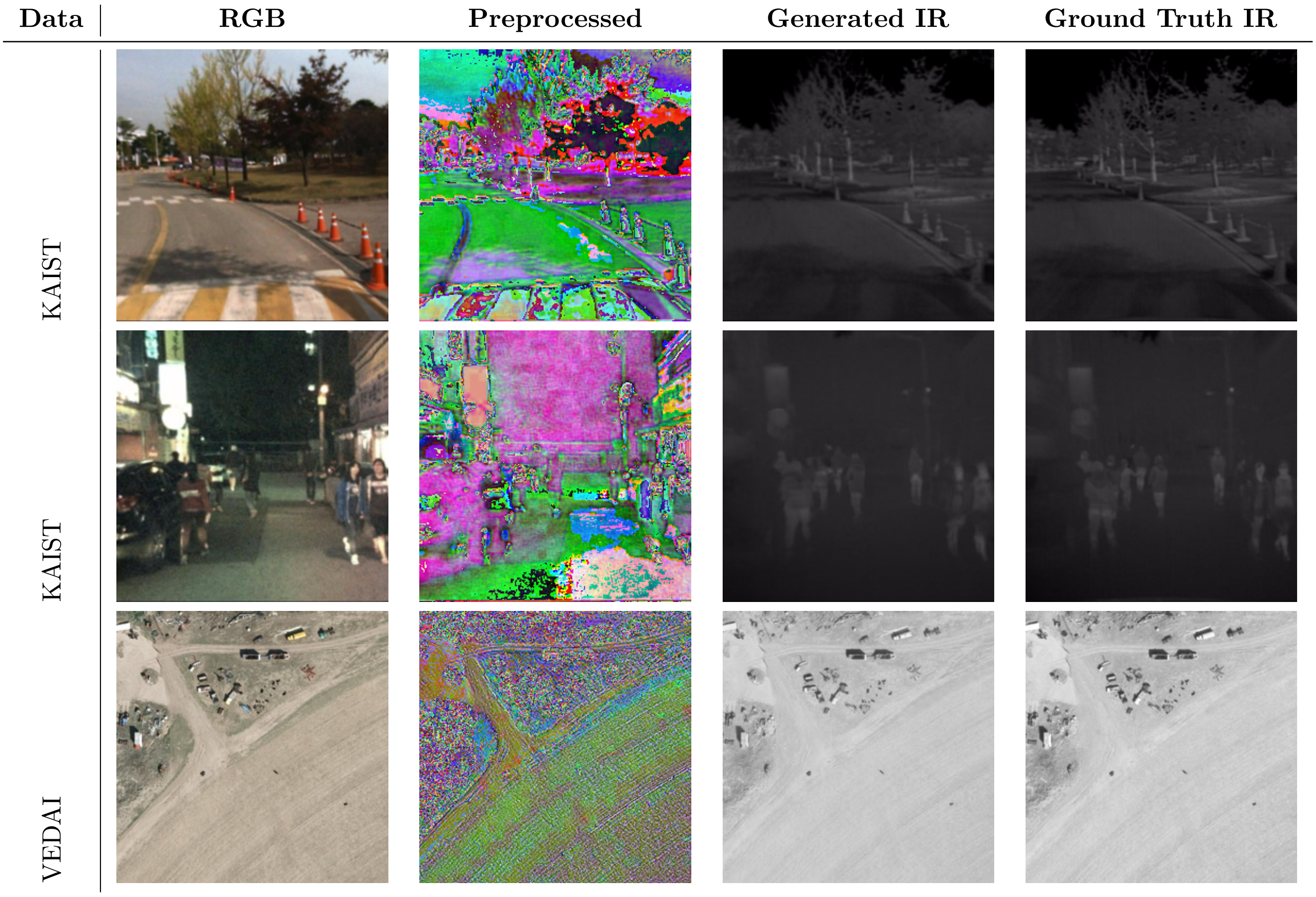

Figure 4: Output Comparisons. The figure displays sample outputs.

4.1.3 Full Model Outputs

A sample of a generated infrared image produced by our models is presented in Figure 6, alongside outputs from other relevant models. Visually, our results are superior to those produced by other GANs. The results using TeV, appear to be as well-aligned as PID. We also observe that without TeV, our results seem less physically accurate; for example, there is less white, indicating less heat, at the back of the car. Notably, this image is sampled from a scene that our model has never encountered before, yet it is able to generate an effective result. Finally, we observe that our additions enhance the apparent quality of the generated images compared to those produced by IRGAN.

4.2 Benchmarks

In this section, we perform explicit benchmarking experiments against other prominent methods, utilizing the Fréchet Inception Distance (FID) [15], Peak Signal-to-noise Ratio (PSNR), Learned Perceptual Image Patch Similarity (LPIPS), the L1 distance and the Structural Similarity Index Measure (SSIM) as metrics. The results are contained in Table 1, and are complemented by qualitative examples that provide anecdotal evidence of the model's capabilities in Figures 5, 4 and 6.

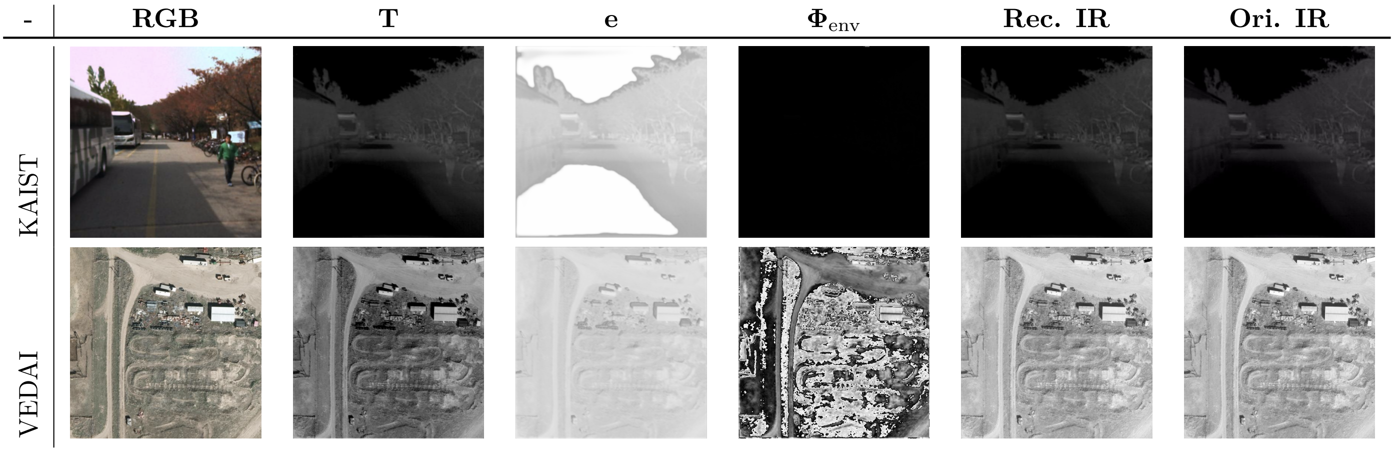

Figure 5: **TeVNet Decomposition.** Sample outputs of the TeVNet module on both the KAIST dataset (top row) and VEDAI dataset (bottom row). The right-most columns display the reconstructed IR image and the original IR image, respectively.

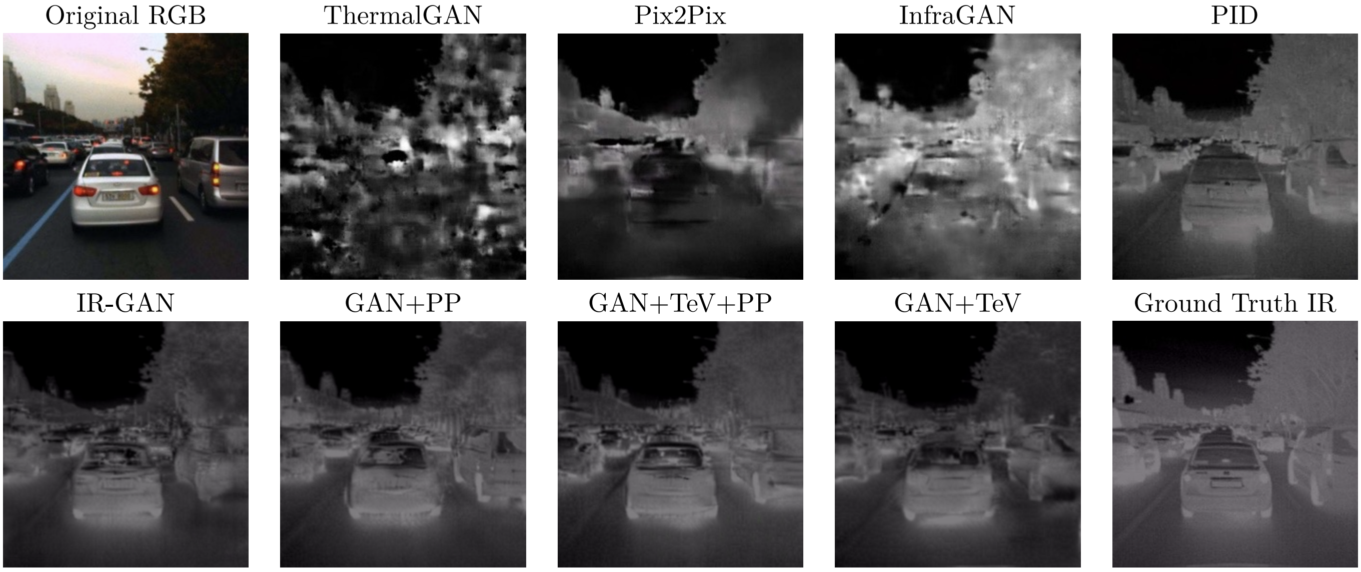

Figure 6: **Full Output Comparisons.** The figure displays sample outputs from various other relevant methods.

Table 1: **Benchmark Results.** Performance of various RBG to IR generative models on both the KAIST and VEDAI datasets using a variety of metrics. Here, "**Ours**" refers to IR-GAN trained with TeVNet decomposition.

Dataset	**Model**	**PSNR↑**	**FID↓**	**LPIPS↓**	**L1↓**	**SSIM↑**
KAIST	ThermalGAN	19.74	277.85	0.242	0.165	0.66
	Pix2Pix	21.25	132.04	0.196	0.137	0.69
	InfraGAN	22.97	222.96	0.157	0.121	0.76
	PID	23.60	51.69	**0.137**	-	**0.7913**
	IR-GAN	23.34	**9.82**	0.223	0.065	0.6768
	Ours	**23.63**	10.21	0.218	**0.063**	0.6863
VEDAI	ThermalGAN	28.54	-	0.019	0.068	**0.8971**
	Pix2Pix	**26.46**	-	-	0.175	0.8754
	InfraGAN	29.26	-	**0.011**	0.055	-
	IR-GAN	30.55	44.97	0.040	**0.022**	0.8866
	Ours	27.15	48.80	0.050	0.039	0.8760

4.3 Ablations

We propose two modifications to the typical CGAN implementation: an additional set of physics-informed loss terms, and extensive image pre-processing. To quantify the effect of each modification, we perform ablation studies, where we benchmark the original IR-GAN architecture against all permutations of such modifications. The results are displayed in Table 2 and are presented for both the KAIST and VEDAI datasets.

Table 2: **Ablation Results.** Various performance metrics for the model proposed in this paper when removing distinct modules. Here, PP refers to pre-processing, and TeV refers to the TeVNet module.

Dataset	**Model**	**PSNR↑**	**FID↓**	**LPIPS↓**	**L1↓**	**SSIM↑**
KAIST	IR-GAN	23.34	9.82	0.223	0.065	0.6768
	GAN+TeV+PP	23.63	10.21	0.218	0.063	0.6863
	GAN+PP	19.32	210.24	0.330	0.095	0.5544
	GAN+TeV	**24.33**	**9.69**	**0.201**	**0.059**	**0.6955**
VEDAI	IR-GAN	30.55	44.97	**0.04**	**0.022**	0.8866
	GAN+TeV+PP	27.15	48.80	0.05	0.039	0.8760
	GAN+PP	26.60	49.91	0.05	0.042	0.8735
	GAN+TeV	**30.59**	**44.19**	**0.04**	**0.022**	**0.9023**

5 Discussion

The KAIST dataset offers small, human-scale images rich with diverse objects such as cars, people, and bikes, while the VEDAI dataset comprises large-scale aerial imagery captured from a bird's-eye view. Both datasets include varied lighting conditions, encompassing both daytime and nighttime scenes, providing robust benchmarks for model evaluation. Our approach consistently outperforms PID and IR-GAN across tasks.

Ablation studies reveal that pre-processing, particularly the down-convolution step, generally degrades performance. This is likely because the down-convolution reduces the input to three channels, discarding valuable spectral information that could enhance the model's understanding of complex features. Unlike [2], which employs transformer blocks with larger input dimensions (16 channels) to preserve feature richness, the down-projection step limits the model's capacity to capture nuanced details. This insight underscores the importance of retaining higher-dimensional feature representations in generative tasks and may be the subject of future work.

Notably, our GAN+TeV model achieves the strongest overall results, as shown in Table 2. By decomposing images into temperature, emissivity, and thermal texture components, TeVNet captures essential thermal properties that are otherwise entangled in conventional image representations. This decomposition not only enriches the model's understanding of the input data but also enables more precise and realistic synthesis of infrared imagery.

6 Conclusion

Our benchmarking experiments demonstrate that the proposed model, which incorporates TeVNet decomposition and image preprocessing, achieves excellent performance on both the KAIST and VEDAI datasets. Additionally, removing the preprocessing module enables the model to reach state-of-the-art performance, as discussed in Section 5. Specifically, our approach outperforms established methods such as IR-GAN, PID, and others in terms of PSNR, FID, and LPIPS metrics on both datasets. It also achieves competitive results on LPIPS and SSIM, highlighting its capability to generate high-fidelity infrared imagery.

Ablation studies further validate the significance of TeVNet in enhancing the model's performance by capturing critical thermal properties through decomposition into temperature, emissivity, and thermal texture components. In contrast, the pre-processing step, particularly down-convolution, adversely impacts performance by potentially discarding valuable spectral information. These findings emphasize the importance of preserving higher-dimensional feature representations for improved generative accuracy.

Overall, the integration of TeVNet decomposition establishes a robust framework for infrared image generation, offering superior perceptual quality and fidelity while setting a new benchmark for generative models in this domain.

References

[1] 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.

[2] Yijia Chen, Pinghua Chen, Xiangxin Zhou, Yingtie Lei, Ziyang Zhou, and Mingxian Li. Implicit multi-spectral transformer: An lightweight and effective visible to infrared image translation model. *arXiv preprint arXiv:2404.07072*, 2024.

[3] Carlo Corsi. Infrared: a key technology for security systems. *Advances in Optical technologies*, 2012(1):838752, 2012.

[4] Ian Goodfellow, Jean Pouget-Abadie, Mehdi Mirza, Bing Xu, David Warde-Farley, Sherjil Ozair, Aaron Courville, and Yoshua Bengio. Generative adversarial nets. *Advances in neural information processing systems*, 27, 2014.

[5] Qishen Ha, Kohei Watanabe, Takumi Karasawa, Yoshitaka Ushiku, and Tatsuya Harada. Mfnet: Towards real-time semantic segmentation for autonomous vehicles with multi-spectral scenes. In *2017 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, pages 5108–5115, 2017.

[6] Changxin Huang, Binbin Liang, Wei Li, and Songchen Han. A convolutional neural network architecture for vehicle logo recognition. In *2017 IEEE International Conference on Unmanned Systems (ICUS)*, pages 282–287. IEEE, 2017.

[7] Soonmin Hwang, Jaesik Park, Namil Kim, Yukyung Choi, and In So Kweon. Multispectral pedestrian detection: Benchmark dataset and baseline. In *2015 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 1037–1045, 2015.

[8] Vladimir Vladimirovich Kniaz, Vladimir Alexandrovich Knyaz, Jiří Hladůvka, Walter G. Kropatsch, and Vladimir Mizginov. Thermalgan: Multimodal color-to-thermal image translation for person re-identification in multispectral dataset. In *ECCV Workshops*, 2018.

[9] Decao Ma, Juan Su, Shaopeng Li, and Yong Xian. Aerialirgan: unpaired aerial visible-to-infrared image translation with dual-encoder structure. *Scientific Reports*, 14(1):22105, 2024.

[10] Decao Ma, Yong Xian, Bing Li, Shaopeng Li, and Daqiao Zhang. Visible-to-infrared image translation based on an improved cgan. *The Visual Computer*, 40:1–10, 04 2023.

[11] Fangyuan Mao, Jilin Mei, Shun Lu, Fuyang Liu, Liang Chen, Fangzhou Zhao, and Yu Hu. Pid: Physics-informed diffusion model for infrared image generation. *arXiv preprint arXiv:2407.09299*, 2024.

[12] Sébastien Razakarivony and Frédéric Jurie. Vehicle detection in aerial imagery : A small target detection benchmark. *J. Vis. Commun. Image Represent.*, 34:187–203, 2016.

[13] Emma Wadsworth, Advait Mahajan, Raksha Prasad, and Rajesh Menon. Deep learning for thermal-rgb image-to-image translation. *Infrared Physics & Technology*, 141:105442, 2024.

[14] Sirui Wang, Guiling Sun, Liang Dong, and Bowen Zheng. Pas-gan: A gan based on the pyramid across-scale module for visible-infrared image transformation. *Infrared Physics Technology*, 139:105314, 2024.

[15] Yu Yu, Weibin Zhang, and Yun Deng. Frechet inception distance (fid) for evaluating gans. *China University of Mining Technology Beijing Graduate School*, 3, 2021.

Appendix

Dataset	$\mathcal{L}_{\text{rec}}$	$\mathcal{L}_{\text{TeV}}$	SSIM↑	PSNR↑	LPIPS↓	FID↓
KAIST	$k_1 = 0$	$k_2 = 0$	0.7892	23.31	0.1380	64.54
	$k_1 = 0$	$k_2 = 50$	0.7892	23.53	0.1366	50.99
	$k_1 = 50$	$k_2 = 0$	0.7901	**23.63**	**0.1363**	**50.17**
	$k_1 = 50$	$k_2 = 50$	0.7893	23.43	0.1370	50.66
	$k_1 = 50$	$k_2 = 5$	**0.7913**	23.60	0.1366	51.69
FLIR	$k_1 = 50$	$k_2 = 50$	**0.4006**	17.26	0.3599	**84.26**
	$k_1 = 50$	$k_2 = 5$	0.3991	**17.35**	**0.3577**	84.28

Figure 7: **TeVNet Ablation.** Ablation study of the TeVNet archiecture taken directly from [11].