INFRARED IMAGE GENERATION FROM VISIBLE BAND IMAGES (VIS2IR)
TÜBİTAK BİLGEM İLTAREN

PROJECT TEAM

Kerem Er

Damla Selen Gerdan

Emre Kulkul

Selçuk Vural

Umut Yıldız

Selami Utku Yücel

COMPANY MENTOR

Damla Akcaoğlu

Dr. Umut Kayıkcı

ACADEMIC MENTOR

Prof. Haldun Özaktaş

TEACHING ASSISTANT

Ecem Şimşek

Abstract. This project's primary goal is to convert visible images (RGB) into infrared (IR) images, which have essential uses in fields including remote sensing and surveillance. Maintaining physical consistency between the RGB and IR modalities is difficult because of the significant variances between them. This research uses machine learning and image processing techniques to solve the issue and accomplish precise RGB-to-IR image translation. The research intends to use models that can learn cross-domain mappings by investigating highly referenced papers on image-to-image translation. Using the DAGAN model with segmented inputs and corresponding infrared images is the main solution strategy. Segmentation maps are obtained by using MSEG and a custom dataset that has been created from Coaxials, SMOD, Roadscenes, Camel, MSRS and Kaist. A test set has been created by randomly selecting 300 photos from this specified dataset. Also, to ensure physical consistency, the emissivity information from the Heat-assisted detection and ranging paper has been implemented in DAGAN. To ensure robustness and fidelity in translated images, the performance was evaluated by using PSNR, SSIM, MAE, FID and LPIPS scores as metrics. An interface that receives RGB input and outputs the image's infrared version was designed. The designed model will be assessed using test data. The anticipated results include achieving performance comparable to, and potentially surpassing, existing models in the literature based on evaluation metrics.

PROJECT DESCRIPTION

This project addresses a critical challenge faced by our company, TÜBİTAK BİLGEM İLTAREN, in the domain of airborne surveillance systems—namely, the difficulty of deploying infrared (IR) cameras on lightweight aerial platforms due to cost, size, power consumption, and operational limitations. The goal is to develop a software solution that translates visible spectrum (RGB) images into synthetic IR images using deep learning. This enables IR-like vision in situations where using actual IR sensors is infeasible. The outcome of the project is a functioning model and a user-friendly interface that allows engineers to input RGB images and generate realistic IR images in real time.

The motivation behind the project stems from the high cost and impracticality of IR cameras, especially for small-scale or mobile applications such as UAVs. IR imaging plays a vital role in national defense by allowing visibility in low-light or foggy conditions—making it indispensable for border monitoring, military operations, and nighttime surveillance. However, building large, high-quality IR datasets is expensive and slow. Existing solutions like infrared generative adversarial network (IRGAN), multimodal unsupervised image-to-image translation (MUNIT), or unpaired image-to-image translation using cycle-consistent adversarial networks (CycleGAN) offer some capabilities in RGB-to-IR translation but lack physical consistency and domain-specific accuracy, particularly in defense-critical use cases.

To address this gap, our project introduces a hybrid solution that combines semantic understanding and physical realism. The core idea is to use a segmentation-driven approach where an RGB image is first segmented using Multi-domain Semantic Segmentation (MSEG) [1], a high-performance semantic segmentation model. The segmentation map is then used as input to Dual Attention Generative Adversarial Networks (DAGAN) [2], which is trained to synthesize IR images from segmentation data. What differentiates our approach from existing models is the integration of physical properties into the training pipeline—specifically, emissivity values derived from the Heat-assisted detection and ranging (HADAR) [3] model. This allows the generated IR images to better reflect the real-world thermal behavior of objects.

From a system design perspective, our architecture comprises three primary stages: semantic segmentation (via MSEG), emissivity map construction (using material-object mapping), and image synthesis (via DAGAN). Additionally, edge maps from the original RGB image are incorporated to enhance boundary definition. The system was trained on a diverse, custom-compiled dataset consisting of over 3,000 paired RGB-IR images. During testing, the model produced high-quality IR outputs even on out-of-distribution inputs, confirming strong generalization capability.

The final deliverable includes a graphical user interface (GUI) built with Qt Designer in Python. It supports multiple image formats (PNG, JPEG), allows for resolution selection, shows side-by-side comparisons of input and output, and saves results in user-defined directories. It also logs processing statistics and, when ground-truth IR is available, calculates performance metrics such as Peak signal-to-noise ratio (PSNR), Structural similarity index measure (SSIM), and learned perceptual

image patch similarity (LPIPS). The system meets all functional and non-functional requirements and operates smoothly even on mid-range laptops using Windows Subsystem for Linux (WSL).

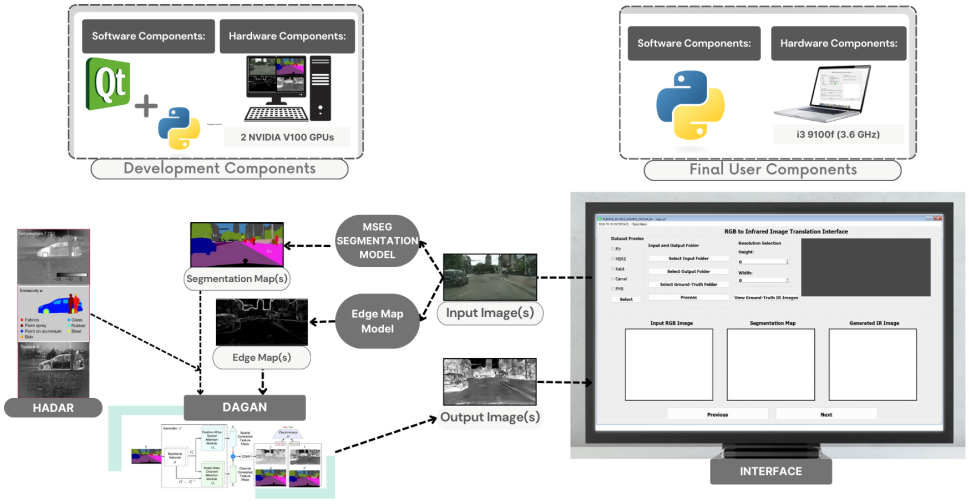

FIGURE 1. Big Picture (Overview of the system pipeline including segmentation, emissivity mapping, edge extraction, and IR generation.)

MILESTONES

There were six milestones achieved in the project.

- **Milestone 1 (Method Selection):** Completed an extensive literature review and empirical evaluation of existing image-to-image translation methods. Finalized DAGAN as the primary model due to its suitability for semantic-to-image synthesis and compatibility with segmentation-based inputs.
- **Milestone 2 (Initial DAGAN implementation):** Trained DAGAN using initial segmentation maps obtained from RGB images and corresponding IR ground truths. This marked the first working version of our end-to-end RGB-to-IR image translation pipeline.
- **Milestone 3 (Dataset collection and segmentation):** 10,000+ paired RGB-IR images processed, with 3,000 paired RGB-IR images the MSEG segmentation model has been chosen.
- **Milestone 4 (Model Finalization):** Integrated emissivity maps from HADAR into the DAGAN pipeline and trained the final model on an expanded and diverse dataset of over 3000 paired images. Achieved high perceptual quality and improved generalization.
- **Milestone 5 (Benchmarking):** Benchmarked the final model against prominent alternatives (Pix2Pix [4], CycleGAN [5], IRGAN [6], IRFormer [7], PID [8], and MUNIT [9]). Our model demonstrated better performance on key metrics FID, and LPIPS.

- **Milestone 6 (User-Interface Development):** Developed a fully functional user interface enabling real-time RGB-to-IR translation. The interface supports image previews, resolution settings, metric display, and IR-ground truth comparisons, making the system accessible for non-expert users.

DESIGN DESCRIPTION

The solution architecture consists of three main stages. First, RGB images are passed through a segmentation model (MSEG) to obtain semantic maps. These maps are then mapped to a reduced taxonomy of 19 classes for compatibility with the image synthesis model. Next, material-specific emissivity values are assigned to each segmented class using a lookup strategy inspired by HADAR, resulting in a spatial emissivity map. Finally, both the segmentation and emissivity maps, along with edge maps derived from the RGB image, are fed into DAGAN, which generates the synthetic IR image.

Emissivity values were derived from the HADAR model's dataset, which provides mean values for common materials such as asphalt, concrete, vegetation, and human skin. Using semantic class-to-material mapping, each segmented region in the image was assigned a physical emissivity value, resulting in a per-pixel emissivity map. This map served as an additional input channel to DAGAN, enriching the model's capacity to generate thermally plausible outputs.

The training process was performed using PyTorch on Google Colab Pro with NVIDIA V100 GPUs. The final model was optimized using a combination of loss functions: adversarial loss for realism, feature matching loss for stability, and perceptual loss to preserve structure. The user interface was developed using Qt Designer in Python and deployed on a Windows machine with WSL for running the Linux-based MSEG model.

A screenshot of the final interface is shown in Figure 2, which includes features such as folder selection, resolution control, image previews, console feedback, and support for optional ground-truth comparison.

RESULTS AND PERFORMANCE EVALUATION

To assess the effectiveness of our RGB-to-IR image translation system, we conducted a thorough evaluation using both quantitative metrics and qualitative visual inspection. The final version of our model—based on DAGAN, MSEG, and emissivity-enhanced inputs—was tested on a curated dataset of 300 unseen RGB-IR image pairs. This dataset was specifically separated to ensure unbiased evaluation of the model's performance under both familiar and out-of-distribution conditions.

Quantitative Evaluation:

We benchmarked our system against six well-known RGB-to-IR image translation models. Pix2Pix, MUNIT, CycleGAN, IRGAN, IRFORMER , and PID. The evaluation consisted of industry-standard metrics. These include PSNR and MAE to measure reconstruction accuracy, SSIM to assess structural similarity, and FID

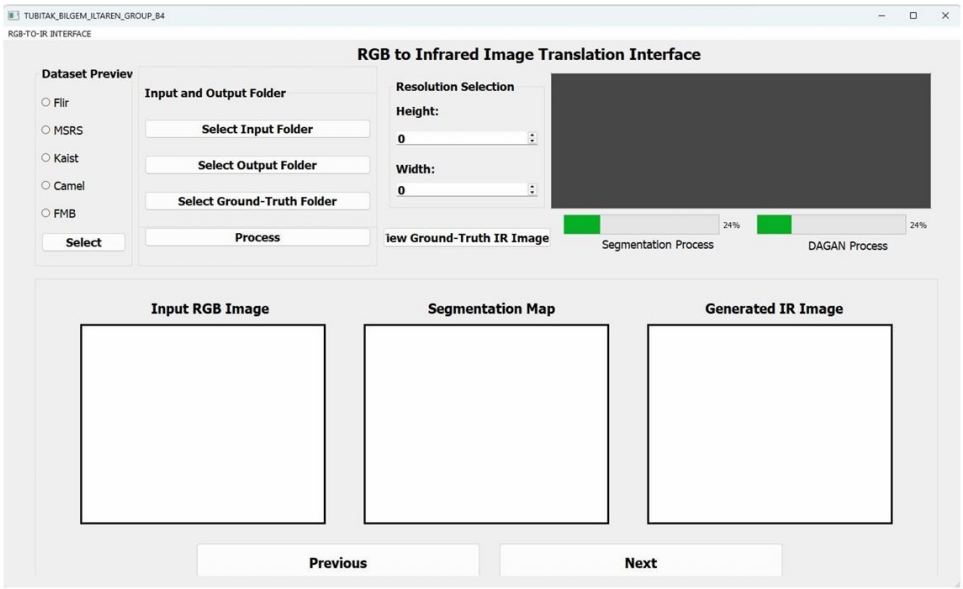

FIGURE 2. Graphical User Interface (Displays RGB input, segmentation map, and generated IR image.)

and LPIPS to evaluate the realism and perceptual closeness of the synthesized images. The following table illustrates the benchmark results.

TABLE 1. Benchmark Results of Our Model vs. Alternatives

Model	FID↓	PSNR↑	SSIM↑	MAE↓	LPIPS↓
DAGAN (Ours)	**66.96**	23.59	0.75	0.049	**0.31**
IRGAN	75.72	**24.28**	**0.76**	**0.045**	0.32
Pix2Pix	106.22	21.51	0.66	0.060	0.45
CycleGAN	172.91	12.70	0.38	0.204	0.53
MUNIT	119.49	9.88	0.26	0.332	0.55
IRFORMER	353.73	16.89	0.65	0.121	0.59
PID	131.45	17.21	0.45	0.105	0.51

As shown in Table 1, our model achieved the best performance in perceptual metrics—FID and LPIPS—which are critical for evaluating visual realism. While IRGAN slightly outperformed in SSIM and MAE, these pixel-wise metrics are known to be less reliable for perceptual tasks. Thus, DAGAN offers a better balance between semantic structure and realistic IR appearance.

Qualitative Evaluation:

To complement the numerical results, we visually examined outputs on both in-distribution and out-of-distribution images. Figure 3 presents a selection of RGB input, its generated IR output, and ground-truth IR image. The model successfully captured fine thermal gradients and responded well to different lighting and object configurations, especially in the case of human detection and material separation.

FIGURE 3. Generated IR Samples (Examples of segmentation map, ground-truth IR image (middle), and generated IR (right) from test set.)

Furthermore, the model was tested on real-world images captured by smartphones around the campus, which were not part of the training distribution. It performed reliably, correctly identifying people, vehicles, and roads, and assigning plausible thermal values based on learned patterns and emissivity data.

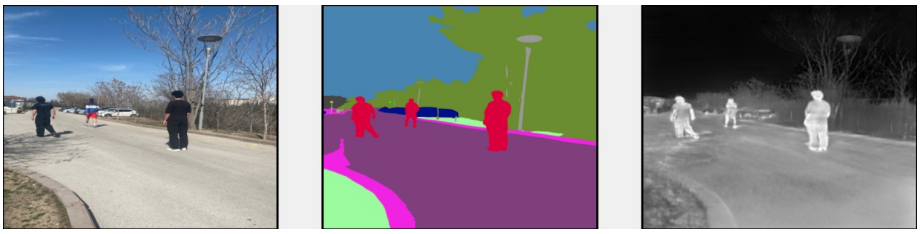

FIGURE 4. Generated IR Samples (Examples of RGB input, segmentation map (middle), and generated IR (right).)

Critical Assessment:

From a performance perspective, the integration of HADAR-based emissivity information has clearly improved the physical realism of the generated infrared images. Unlike many traditional models such as CycleGAN or Pix2Pix, which rely purely on data-driven mappings, our system incorporates physics-informed priors, allowing for semantically and thermally consistent outputs.

Another major strength lies in the semantic understanding introduced via MSEG. Compared to previous approaches where segmentation was either absent or rudimentary, MSEG provides a robust understanding of the scene structure, improving object-level translation accuracy—particularly for complex classes such as humans, vehicles, and infrastructure.

Generalization to out-of-distribution data has also been a notable achievement. The model has shown reliable performance on test images captured in real-world scenarios, even those featuring backgrounds or lighting conditions not present in the training dataset.

Despite these strengths, certain limitations persist. The system's reliance on accurate segmentation remains a key dependency. Errors in semantic maps can propagate into the final output, reducing both structural accuracy and thermal plausibility. In addition, while the emissivity assignment improves realism, it is currently heuristic-based and may not perfectly reflect material diversity in all scenes.

CONCLUSIONS AND FUTURE DIRECTIONS

This project set out to address the practical limitations of IR imaging systems, particularly in resource-constrained platforms such as UAVs, by developing a deep learning-based solution capable of synthesizing IR images from standard RGB inputs. The final system integrates semantic segmentation, physical emissivity modeling, and a generative adversarial network to produce visually and physically consistent IR outputs. In the future, segmentation and IR generation might be unified into a single model to reduce system complexity, heuristic emissivity mapping might be replaced with learned material classification or multispectral estimation, the model might be extended to accept video input to create frame-to-frame IR generation for general surveillance footage, and the IR generation model can be coupled with object detection algorithms for end-to-end surveillance or targeting pipelines.

Overall, the project demonstrates that physically informed, data-driven models can serve as effective alternatives to hardware-constrained IR systems. With continued development, this solution has strong potential for integration into real-world defense and security applications.

REFERENCES

[1] Sakaridis, C., Dai, D., and Van Gool, L., "MSEG: A composite dataset for universal segmentation," *ECCV*, 2022.

[2] Zuo, Y. et al., "DAGAN: Dual Attention Generative Adversarial Networks for Semantic Image Synthesis," *arXiv preprint arXiv:2303.03549*, 2023.

[3] Bao, F. et al., "Heat-assisted detection and ranging," *Nature*, vol. 626, pp. 55–61, 2024.

[4] Isola, P., Zhu, J.-Y., Zhou, T., and Efros, A. A., "Image-to-image translation with conditional adversarial networks," *IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, 2017.

[5] Zhu, J.-Y., Park, T., Isola, P., and Efros, A. A., "Unpaired image-to-image translation using cycle-consistent adversarial networks," *IEEE/CVF International Conference on Computer Vision (ICCV)*, 2017.

[6] Xu, X., et al., "IR-GAN: Infrared image synthesis via conditional generative adversarial network," *IEEE Transactions on Multimedia*, vol. 22, no. 7, pp. 1748–1761, 2020.

[7] Liu, J., et al., "IRFormer: Implicit multispectral transformer for RGB-to-IR translation," *arXiv preprint arXiv:2302.01456*, 2023.

[8] Wang, Y., et al., "Physics-informed diffusion model for infrared image synthesis," *arXiv preprint arXiv:2301.10050*, 2023.

[9] Huang, X., Liu, M.-Y., Belongie, S., and Kautz, J., "Multimodal unsupervised image-to-image translation," *European Conference on Computer Vision (ECCV)*, 2018.

BEHIND THE SCENES