Depth-Conditioned Thermal-like Image Generation

Patricia L. Suárez[1]
[1]*ESPOL Polytechnic University*
Guayaquil, Ecuador
plsuarez@espol.edu.ec

Angel Sappa[1,2]
[2]*Computer Vision Center*
Barcelona, Spain
sappa@ieee.org

2024 14th International Conference on Pattern Recognition Systems (ICPRS) | 979-8-3503-7565-7/24/$31.00 ©2024 IEEE | DOI: 10.1109/ICPRS62101.2024.10677832

***Abstract*—This paper proposes a novel approach to generate thermal-like representations from RGB images by using the corresponding depth map as an additional constraint. The given RGB images are converted to the HSV color space and the brightness channel is used as input together with the spatial information provided by the depth map of the given scene. This depth map is used as prior information by the generative network. By training a generative model with paired input images and their corresponding depth maps, the model learns the mapping from the RGB images to thermal-like representations. Experimental results demonstrate that the method outperforms state-of-the-art approaches, producing superior-quality thermal images with improved shape and sharpness, attributed to using depth maps as complement information.**

***Index Terms*—thermal-like representations, conditioned generative models, depth maps**

I. Introduction

Thermal imaging plays a crucial role across diverse sectors, including security, healthcare, and environmental monitoring [1]. However, conventional thermal imaging techniques are often constrained by factors such as equipment costs and weather conditions. In response to these challenges, the synthesis of thermal images from standard RGB images can be considered a real alternative, offering the potential to increase visual data and facilitate applications where traditional thermal images are not possible to acquire by high costs or lack of data sources. The evolution of thermal imaging from its military origins to widespread applications in industrial inspection and medical diagnostics has spurred the development of synthetic thermal image generation. Overcoming challenges such as cost and accessibility in acquiring real thermal imagery, this field merges computer vision, machine learning, and physics-based modeling to simulate thermal radiation and produce images for training machine learning models and testing algorithms.

Key challenges in this domain include precise material properties estimation, incorporating atmospheric effects, and validating synthetic images against real-world data. Despite these obstacles, synthetic thermal image generation presents exciting opportunities for applications such as training autonomous vehicles in challenging weather conditions, biomedical diagnosing, and enhancing security systems through improved thermal surveillance capabilities like security screening (e.g., [2], [3], [4]).

In recent years, research efforts have focused on leveraging advanced computational techniques to synthesize thermal-like images from RGB inputs. These efforts have led to the development of various methodologies, including generative models based on deep learning architectures such as Generative Adversarial Networks (GANs) and Variational Autoencoders (VAEs). By learning the underlying patterns and characteristics of thermal imagery from large datasets of paired RGB and thermal images, these models can generate convincing thermal-like representations from RGB inputs (e.g., [5], [6]). Lately, conditional synthetic thermal image generation has emerged as a transformative field, offering innovative solutions to challenges in acquiring real thermal imagery [7]. Taking advantage of numerical modeling of heat transfer and advanced technologies such as generative adversarial networks, researchers are pioneering novel approaches to create synthetic thermal images for various applications.

The current work presents a novel methodology for generating thermal images conditioned on the depth information of the given scene. These depth maps provide valuable spatial information about a scene, enriching the generation process with improved shape and sharpness. The model also converts the RGB input images to the HSV color space, from which just the brightness channel (H) is used to facilitate the convergence of the model to the thermal domain enhanced with the information extracted from the depth maps. Our approach seamlessly integrates depth cues into the generation process, resulting in thermal-like images that closely approximate authentic thermal images.

The proposed methodology involves training a paired GAN generative model, which learns to generate thermal representations from RGB images while considering the provided depth information as a conditioning to the generative process. This integration allows the model to generate thermal images with improved shape and sharpness, very similar to real thermal images. The manuscript is organized as follows. Section II provides a comprehensive review of related work in the fields of thermal image synthesis using deep learning with or without prior information. Section III presents the approach proposed for conditional generative thermal-like image generation using depth maps as prior. Experimental results and comparisons with different approaches are given in Section IV. Finally, conclusions are presented in Section V.

II. Related Work

In this section, the existing literature on thermal image synthesis with deep learning-based techniques such as con-

ditional generative GAN networks using prior information or any other type of convolutional networks is reviewed. By examining the different strategies and techniques employed in these approaches, we seek to gain insight into advances and challenges in the field and lay the foundation for our proposed methodology.

In the field of thermal imaging, high-quality image synthesis using convolutional networks such as conditional generative adversarial networks or deep learning-based architectures using complementary information as priors has emerged as a promising approach. Researchers have explored various methodologies to leverage conditional GANs along with prior information to generate realistic and visually appealing thermal images. By conditioning GANs to specific antecedents, such as semantic information, distances or physical characteristics of the scene, in different research work authors manage to improve synthetic thermal images so that they are not only visually pleasing but also semantically better defined. In [8], a method for detecting pedestrians in thermal images is proposed, addressing the limitations of infrared cameras, such as low contrast and blurred details. The proposed architecture, TE-GAN, is a thermal enhancement model based on generative adversarial networks. This architecture consists of two important modules: contrast enhancement and denoising, followed by a post-processing step for edge restoration to improve the overall image quality. Following a similar strategy to synthesize thermal faces, Zhang et al. [9] propose a GAN-based multi-stream feature-level fusion technique. This involves using a generator sub-network, which is built using an encoder-decoder network with dense residual blocks. Additionally, a multi-scale discriminator sub-network is employed. Another approach to facial recognition is the one proposed by [10], where a method for generating thermal images from visible spectrum has been implemented to perform facial emotion recognition. This technique leverages facial thermal imaging as an efficient modality for recognizing emotions. The process involves using CycleGAN, a type of GAN, to translate images from the visible spectrum to thermal images, enhancing the capabilities of facial emotion recognition systems.

On the contrary to previous approaches, in [11] the authors propose a generic framework to generate thermal images of any scenario from the corresponding RGB image. The authors propose an unpaired cyclic adversarial generative model, to obtain the synthetic representation of thermal images from visible images. The model allows simulation of the temperature information of the objects in the scene.

The domain translation problem has been also studied in a generic framework by the remote sensing community, where the development of models capable of translating co-aligned images between different modalities (e.g., RGB-IR, Synthetic Aperture Radar (SAR) - Electro-Optical (EO), SAR-IR, SAR-RGB) has been tackled. This challenge has motivated the research community to set-up competitions in different forums to evaluate the performance of different contribution. One example of these competitions is the one held annually within the framework of the PBVS-CVPR workshop [12]. In summary, domain translation has recently become an active research topic of interest in various communities. According to the state-of-the-art, all proposed techniques focus on generating a representation in different domain from the given one by using some generative approach. However, none of the approaches reviewed in the literature make use of additional information that could facilitate or improve the results. In the current work, thermal image generation uses depth information to enhance the obtained results.

III. Proposed Approach

This section presents the approach proposed for generating thermal-like images by means of a CycleGAN architecture. The proposed approach takes advantage of the inherent characteristics of the HSV color space, specifically focusing on the brightness channel, to facilitate a more homogeneous translation to the thermal domain. Also, thermal imaging primarily captures heat intensity rather than the visible spectrum colors. In the HSV model, the 'Value' component aligns directly with brightness or intensity, which acts as a proxy for thermal radiation in synthetic imaging. This allows the model to prioritize intensity variations, much like thermal cameras, which focus more on heat emitted by objects rather than their color. Furthermore, depth information is used as an additional input to improve the quality and sharpness of the generated thermal images. It should be noted that the input is just an RGB image, which is converted to the HSV color space, depth information is estimated by means of the approach proposed in [13]. By combining intensity information, the V channel, with spatial cues, our approach offers a comprehensive solution to produce high-fidelity pseudothermal images with greater structural coherence, contextual relevance, and representation of temperatures as close as possible to real thermal images.

By incorporating depth-aware constraints, our model learns to preserve depth-related characteristics across image translations, ensuring that the thermal features align with the spatial arrangement of objects within the scene. This results in pseudo-thermal images that not only exhibit enhanced visual fidelity but also convey a deeper understanding of the scene's composition and spatial relationships. Depth maps provide information about the 3D structure of the scene (e.g., surface orientation). This knowledge can be used to infer thermal gradients, occlusions, and object boundaries in the thermal images, allowing for more accurate synthesis, especially in cases where visible spectrum data alone might not provide enough information.

Building upon the Cycled GAN framework introduced by Zhu et al. [14], and inspired by [11], our architecture facilitates domain transfer between paired image domains, enabling the synthesis of thermal-like images. To ensure accurate translation of pixel intensities to the far infrared spectrum, we integrate the contrastive loss function presented in [15], focusing on learning the relationship between input embeddings from nearby regions, thereby enhancing image quality by predicting missing information based on environmental context.

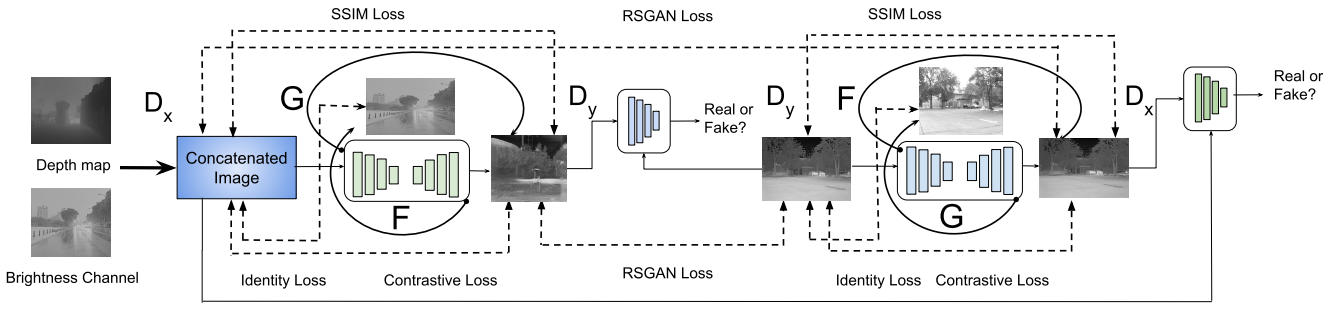

Fig. 1. Proposed Cycle GAN architecture.

Overall, the choice of color space is an important consideration in image processing, and the experiments with the proposed approach demonstrate the significant impact that this choice can have on the quality and accuracy of the resulting output. In the case of obtaining synthetic thermal images, the use of the brightness channel, from HSV representation, was found to be highly effective and should be considered a key factor in the design of similar models in the future.

In the current work the Least square GAN loss is incorporated, in place of the traditional GAN loss suggested by [16]. This leads to better training of the GAN, and consequently, more accurate synthetic thermal images. The use of the least square GAN loss in our architecture offers several advantages over the standard GAN loss providing more stability and helping synthesize better similarity of textures and temperatures of the images. The definition of the least square loss for both the generator and discriminator components can be expressed as follows:

$$L_D^{\text{LS-GAN}} = \tfrac{1}{2}\mathbb{E}_{y\sim\mathbb{P}}[(D(y)-1)^2] + \tfrac{1}{2}\mathbb{E}_{x\sim\mathbb{Q}}[D(x)^2] \quad (1)$$

$$L_G^{\text{LS-GAN}} = \tfrac{1}{2}\mathbb{E}_{x\sim\mathbb{Q}}[(D(x)-1)^2], \quad (2)$$

where x represents a real thermal from the real data distribution, x represents a generated (fake) thermal from the generator, $D(y)$ represents the discriminator's output (probability) for a real thermal y, and $D(x)$ represents the discriminator's output (probability) for a generated synthesize thermal x. To enhance the synthesis of thermal images, instance normalization is employed, which adjusts the features of each synthetic image individually. Applying this normalization process effectively reduces style differences between the generated and real-synthetic thermal samples, leading to improved overall quality and realism in the synthesized depth maps.

The contrastive loss has been implemented in our architecture to enable the model to be trained by learning the similarities between the latent space generated by the network. As outlined in [17], contrastive learning methods require only a definition of the similarity distribution to sample a positive input $\mathbf{x}^+ \sim p^+(\cdot \mid \mathbf{x})$, and data distribution for a negative input $\mathbf{x}^- \sim p^-(\cdot \mid \mathbf{x})$, for a given input image $\mathbf{x}$. The authors in [17] argue that the shape of the tensor $V_l \in \mathbb{R}^{S_l \times D_l}$ is determined by the architecture of the network, where S_l denotes the number of spatial locations of the tensor. The notation $v_l^s \in \mathbb{R}^{D_l}$ is used to refer to the D_l-dimensional feature vector at the s^{th} spatial location. Additionally, $\bar{v}_l^s \in \mathbb{R}^{(S_l-1)\times D_l}$ is defined as the collection of feature vectors at all other spatial locations except s. Also, $\hat{v}_l^s$ is the predicted feature vector at spatial location s and layer l, derived from $\hat{Y}$. The primary goal of training the model is to minimize the difference between the predicted output $\hat{Y}$ and the true output Y. By including both, the loss function explicitly penalizes discrepancies between these two sets of representations.

Hence, the proposed contrastive loss function is defined as follows:

$$\mathcal{L}_{\text{contrasive}}(\hat{Y}, Y) = \sum_{l=1}^{L}\sum_{s=1}^{S_l} \ell_{\text{contr}}\ (\hat{v}_l^s, v_l^s, \bar{v}_l^s)\,. \quad (3)$$

The training process focuses on learning discriminative features between images, which can be beneficial in capturing the most representative differences in temperature patterns. Therefore, it helps to explicitly model features related to these important features in a synthetic thermal image. In the context of paired images, contrast loss can still be beneficial if there are subtle differences between domains that are not fully captured by consistent cyclic loss. It helps the model learn these differences explicitly, potentially leading to more accurate translations.

To constrain the pixel intensity levels within the bounds of the target domain during data transformation, the model incorporates the identity loss function. This implies that the generative network preserves essential attributes, such as thermal intensity levels and object shapes while ensuring the stability of the formation model. Specifically, the generative network aims to maintain $G(x) \approx x$ and $F(y) \approx y$ and ensures minimal deviation in pixel intensity levels from the original domain during the transformation process. The identity loss function is defined as follows:

$$\begin{aligned}\mathcal{L}_{\text{identity}}\,(G,F) = {} & E_{x\sim p_{\text{data}}(x)}[\|G(x)-x\|] \\ & + E_{y\sim p_{\text{data}}(y)}[\|F(y)-y\|].\end{aligned} \quad (4)$$

.

Furthermore, another index used as a reference is the structural similarity index, proposed in [19]. This index assesses images by considering the sensitivity of the human visual perception system to alterations in local structure. The underlying concept of this loss function is to aid the learning model in generating visually enhanced images. The structural similarity loss is defined as:

$$\mathcal{L}_{\text{SSIM}}(x, y) = 1 - SSIM(x, y), \tag{5}$$

where $SSIM(x, y)$ is the Structural Similarity Index (see [19], x represents the output of the neural network that we are trying to optimize, and y is the reference or ground truth image. It represents the target image that the model aims to reproduce or approximate as closely as possible

The final loss function (L_{final}) used in our model combines the earlier mentioned loss components and it is formulated as follows:

$$\begin{aligned} L_{\text{final}} &= \mathcal{L}_{\text{LSGAN}}(G, D, X, Y) + \lambda_X \mathcal{L}_{\text{Lcontrative}}\,(Gs, F, X) \\ &+ \lambda_Y \mathcal{L}_{\text{Lcontrastive}}(F, G, Y) + \gamma \mathcal{L}_{\text{Identity}(G,F)} + \beta \mathcal{L}_{\text{SSIM(x,y)}}, \end{aligned} \tag{6}$$

where λ_X and λ_Y represent the weights attributed to the contrastive loss function for the domains X and Y, respectively. These values are empirically determined based on experimental outcomes. The contrastive loss component $\mathcal{L}_{\text{contrastive}}$ evaluates the similarity of the latent spaces generated by the generator networks G and F within the embedding network for corresponding input images; γ and β are the weights of the $identity$ and $SSIM$ loss functions respectively, they are defined according to the results of the experiments.

IV. Experimental Results

A. Datasets

The M3FD data set [20] has been utilized to train the proposed model. The data set was captured using a binocular optical and infrared sensor and consists of 4,500 image pairs of outdoor scenes. For training, 3,000 image pairs were used while 890 pairs were used for validation and the remaining images for testing of the trained model. The images were preprocessed to generate realistic synthetic far-infrared images by transferring them to the HSV color space and selecting the brightness channel for training. To test the model's robustness, the FLIR ADAS V2 dataset [21] with 300 pairs has been used. Also, an additional proprietary data set called Thermal Stereo [22], which includes 200 pairs of registered visible-thermal images, was used. The model trained with the M3FD data set was evaluated and compared against the results obtained from other experiments.

B. Results and Comparisons

The proposed approach is evaluated by comparing it with the results from the state-of-the-art models for unpaired image translation (e.g., [14] and [18]). These models are well-known for their capability to generate synthetic images from the visible spectrum to another domain. In this study, the concept is adapted by introducing modifications to the loss functions and pre-processing techniques to generate synthetic thermal images.

This section provides an overview of the quantitative and qualitative results obtained through the proposed methodology. The data set used for training is also detailed and the preprocessing techniques used on the images are described. Furthermore, it performs a comparative analysis using similarity metrics and evaluates the PSNR present in the generated synthetic images.

Table I presents the average results obtained from the model in [14], the approaches presented in [18], both CUT and FastCUT, and the approach proposed in the current work. The evaluation process uses samples from the M3FD, FLIR ADAS V2 datasets as well as our dataset consisting of outdoor scenes named Thermal Stereo. The SSIM obtained with each dataset, together with their corresponding PSNR values are depicted. Visual representations of the synthetic thermal images generated from these validation sets are depicted in Fig. 2, Fig. 3, and Fig. 4 respectively. In conclusion, the proposed approach builds upon the previous state-of-the-art model, introducing modifications to generate synthetic thermal images. The evaluation results demonstrate the effectiveness of the approach in producing high-quality synthetic thermal images, as evidenced by the quantitative metrics and visual comparisons.

V. Conclusions

This paper presents a novel approach for synthesizing thermal-like images using depth maps as additional information. The brightness channel from the HSV representation is used as input by the generative model. It generates a thermal-like image representation that closely simulates authentic thermal images while incorporating spatial information provided by depth cues. Through training a generative model on paired RGB images and their corresponding depth maps, we enable the model to learn the mapping from RGB inputs to thermal-like representations. Experimental results demonstrate that our model improves state-of-the-art techniques, obtaining synthetic thermal images with better representation of temperatures and improved quality.

Acknowledgements

This material is based upon work supported by the Air Force Office of Scientific Research under award number FA9550-22-1-0261; and partially supported by the Grant PID2021-128945NB-I00 funded by MCIN/AEI/10.13039/501100011033 and by "ERDF A way of making Europe"; and by the ESPOL project CIDIS-12-2022. The second author acknowledges the support of the Generalitat de Catalunya CERCA Program to CVC's general activities, and the Departament de Recerca i Universitats from Generalitat de Catalunya with reference 2021SGR01499.

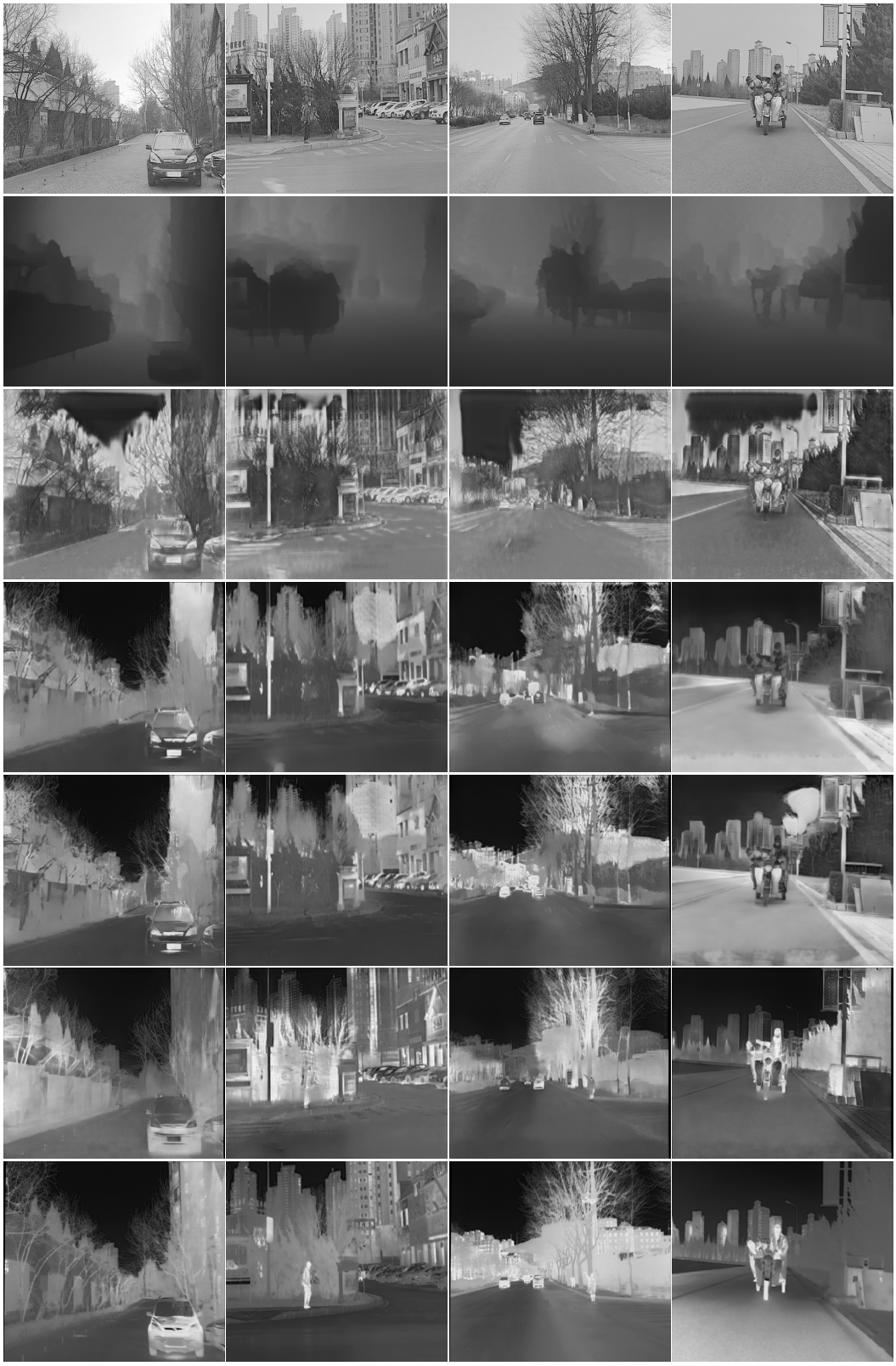

Fig. 2. Experimental results from M3FD dataset: (*1st. row*) Brightness channel from the HSV representation; (*2nd. row*) Depth map obtained from [13]; (*3rd. row*) Thermal representation from [14]; (*4th. row*) Thermal representation from [18] (FastCUT); (*5th. row*) Thermal representation from [18] (FastCUT); (*6th. row*) Thermal representation from the proposed approach; (*7th. row*) Ground truth images.

Fig. 3. Experimental results from Thermal Stereo dataset: (*1st. row*) Brightness channel from the HSV representation; (*2nd. row*) Depth map obtained from [13]; (*3rd. row*) Thermal representation from [14]; (*4th. row*) Thermal representation from [18] (CUT); (*5th. row*) Thermal representation from [18] (FastCUT); (*6th. row*) Thermal representation from the proposed approach; (*7th. row*) Ground truth images.

Fig. 4. Experimental results from FLIR ADAS V2 dataset: (*1st. row*) Brightness channel from the HSV representation; (*2nd. row*) Depth map obtained from [13]; (*3rd. row*) Thermal representation from [14]; (*4th. row*) Thermal representation from [18] (CUT); (*5th. row*) Thermal representation from [18] (FastCUT); (*6th. row*) Thermal representation from the proposed approach; (*7th. row*) Ground truth images.

TABLE I
AVERAGE RESULTS FROM THE VALIDATION SETS (M3FD-THERMAL, STEREO, AND FLIR ADAS V2). BEST RESULTS IN **BOLD**.

Approaches	M3FD		Thermal Stereo		FLIR ADAS V2	
	PSNR	**SSIM**	**PSNR**	**SSIM**	**PSNR**	**SSIM**
Zhu et al. [14]	12.612	0.492	10.658	0.404	11.163	0.447
CUT [18]	13.178	0.6644	12.152	0.526	11.214	0.509
FastCUT [18]	13.241	0.683	12.581	0.631	12.117	0.605
Proposed Approach	**14.799**	**0.781**	**17.635**	**0.751**	**13.517**	**0.701**

REFERENCES

[1] M. Teutsch, A. D. Sappa, and R. I. Hammoud, *Computer vision in the infrared spectrum: challenges and approaches*. Springer, 2022.

[2] D. Perpetuini, C. Filippini, D. Cardone, and A. Merla, "An overview of thermal infrared imaging-based screenings during pandemic emergencies," *International Journal of Environmental Research and Public Health*, vol. 18, no. 6, p. 3286, 2021.

[3] P. Wang, H. Sun, X. Bai, S. Guo, and D. Jin, "Traffic thermal infrared texture generation based on siamese semantic CycleGAN," *Infrared Physics & Technology*, vol. 116, p. 103748, 2021.

[4] Z. Wang, J. Zhan, Y. Li, Z. Zhong, and Z. Cao, "A new scheme of vehicle detection for severe weather based on multi-sensor fusion," *Measurement*, vol. 191, p. 110737, 2022.

[5] X. Li, J. Li, Y. Li, A. Ozcan, and M. Jarrahi, "High-throughput terahertz imaging: progress and challenges," *Light: Science & Applications*, vol. 12, no. 1, p. 233, 2023.

[6] R. Blythman, A. Elrasad, E. O'Connell, P. Kielty, M. O'Byrne, M. Moustafa, C. Ryan, and J. Lemley, "Synthetic thermal image generation for human-machine interaction in vehicles," in *2020 Twelfth International Conference on Quality of Multimedia Experience (QoMEX)*. IEEE, 2020, pp. 1–6.

[7] V. Mizginov and S. Y. Danilov, "Synthetic thermal background and object texture generation using geometric information and GAN," *The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, vol. 42, pp. 149–154, 2019.

[8] M. A. Marnissi, H. Fradi, A. Sahbani, and N. E. Ben Amara, "Thermal image enhancement using generative adversarial network for pedestrian detection," in *25th International Conference on Pattern Recognition (ICPR)*, 2021, pp. 6509–6516.

[9] L. Zhang, A. Gonzalez-Garcia, J. Van De Weijer, M. Danelljan, and F. S. Khan, "Synthetic data generation for end-to-end thermal infrared tracking," *Transactions on Image Processing*, vol. 28, no. 4, pp. 1837–1850, 2018.

[10] G. Pons, A. El Ali, and P. Cesar, "ET-CycleGAN: Generating thermal images from images in the visible spectrum for facial emotion recognition," in *Companion publication of the international conference on multimodal interaction*, 2020, pp. 87–91.

[11] P. L. Suárez and A. D. Sappa, "Toward a thermal image-like representation." in *VISIGRAPP (4: VISAPP)*, 2023, pp. 133–140.

[12] S. Low, O. Nina, A. D. Sappa, E. Blasch, and N. Inkawhich, "Multi-modal aerial view image challenge: Translation from synthetic aperture radar to electro-optical domain results - PBVS 2023," in *2023 IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops (CVPRW)*, 2023, pp. 515–523.

[13] P. L. Suárez, D. Carpio, and A. Sappa, "A deep learning based approach for synthesizing realistic depth maps," in *International Conference on Image Analysis and Processing*. Springer, 2023, pp. 369–380.

[14] J.-Y. Zhu, T. Park, P. Isola, and A. A. Efros, "Unpaired image-to-image translation using cycle-consistent adversarial networks," in *Proceedings of the IEEE international conference on computer vision*, 2017, pp. 2223–2232.

[15] R. Liu, Y. Ge, C. L. Choi, X. Wang, and H. Li, "Divco: Diverse conditional image synthesis via contrastive generative adversarial network," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, June 2021, pp. 16 377–16 386.

[16] I. Goodfellow, J. Pouget-Abadie, M. Mirza, B. Xu, D. Warde-Farley, S. Ozair, A. Courville, and Y. Bengio, "Generative adversarial networks," *Communications of the ACM*, vol. 63, no. 11, pp. 139–144, 2020.

[17] A. Andonian, T. Park, B. Russell, P. Isola, J.-Y. Zhu, and R. Zhang, "Contrastive feature loss for image prediction," in *Proceedings of the IEEE/CVF international conference on computer vision*, 2021, pp. 1934–1943.

[18] T. Park, A. A. Efros, R. Zhang, and J.-Y. Zhu, "Contrastive learning for conditional image synthesis," in *ECCV*, 2020.

[19] Z. Wang, A. C. Bovik, H. R. Sheikh, and E. P. Simoncelli, "Image quality assessment: from error visibility to structural similarity," *IEEE Transactions on Image Processing*, vol. 13, no. 4, pp. 600–612, 2004.

[20] J. Liu, X. Fan, Z. Huang, G. Wu, R. Liu, W. Zhong, and Z. Luo, "Target-aware dual adversarial learning and a multi-scenario multi-modality benchmark to fuse infrared and visible for object detection," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2022, pp. 5802–5811.

[21] D. FLIR, Thermal. (2023) Free teledyne flir thermal dataset for algorithm training. Accessed on November 07, 2023. [Online]. Available: https://www.flir.com/news-center/camera-cores–components/flir-releases-first-european-thermal-imaging-dataset-for-automotive-driver-assistance-development/

[22] R. Rivadeneira, H. Velesaca, and A. D. Sappa, "Cross-spectral image registration: a comparative study and a new benchmark dataset," in *Proceedings of International Conference on Innovations in Computational Intelligence and Computer Vision (ICICV)*, 2024.