International Journal of Information Technology & Computer Engineering

ISSN 2347–3657

Volume 13, Issue 2, 2025

Image Translation and Reconstruction Using a Single Dual Mode Lightweight Encoder

Manik Rao patil, G. Rakesh, Koyyada Pooja, Parshaveni Nagaraju

[1]Assistant Professor, Department Of IT, Guru Nanak Institutions Technical Campus (Autonomous), India.

[2,3,4]B.Tech Students, Department Of IT, Guru Nanak Institutions Technical Campus (Autonomous), India

ABSTRACT

In this work, we propose a novel Dual-Mode Web-Based Image Processor designed to address the challenges of image translation across different modalities. Traditional computer vision models often rely on a single sensor modality, such as RGB or thermal images, but fail to fully exploit the complementary strengths of both. Our architecture leverages a single lightweight encoder that efficiently encodes both grayscale and thermal images into compact latent vectors. This encoding enables cross-modal image translation, including grayscale image colorization and thermal image reconstruction, facilitating flexibility in handling multiple downstream tasks. Our approach reduces the computational burden by utilizing a compact encoder and optimizing for both data compression and robust image translation across varied lighting conditions. The model employs four distinct generators and two discriminators in an adversarial framework, incorporating reconstruction error terms to ensure consistency and contrast preservation. Experimental results demonstrate competitive quality in translation and reconstruction across various lighting scenarios, with comprehensive evaluations across multiple metrics. Additionally, ablation studies validate the effectiveness of the proposed loss terms, confirming their role in improving model performance.

1-INTRODUCTION

In recent years, computer vision has seen significant advancements, primarily due to the use of deep learning techniques that leverage large-scale datasets and powerful models. However, most existing models in the field of image translation rely on single sensor modalities, such as RGB or thermal images, and often fail to harness the complementary strengths of both. This limitation restricts their applicability in real-world scenarios where both types of images are available and can offer unique insights. To address this challenge, we propose a Dual-Mode Web-Based Image Processor that combines grayscale and thermal image inputs into a unified processing framework. By utilizing a single, lightweight encoder, our system efficiently encodes both types of images into compact latent vectors, facilitating the cross-modal translation tasks of grayscale image colorization and thermal image reconstruction. This approach enables flexibility in handling different image modalities and downstream tasks, while also reducing the computational burden typically associated with large, task-specific models. Moreover, our architecture optimizes for data compression, making it suitable for environments with limited resources, such as edge computing and real-time applications. The model uses an adversarial framework, including four distinct generators and two discriminators, to ensure high-quality translation and reconstruction of images under varying lighting conditions. Our method is evaluated through extensive experiments, demonstrating competitive performance in multiple metrics, and is further validated through ablation studies that highlight the importance of the proposed loss functions.

EXISTING SYSTEM

For single-directional image translation using large, task-specific models, several existing systems have been developed, each tailored to specific image translation tasks. These systems typically focus on translating images between specific modalities or tasks, such as converting between RGB and grayscale, day-to-night images, or semantic segmentation to photo-realistic images. The models are often large and optimized for specific tasks, which makes them highly effective in their narrow scope but also less flexible for other applications.

CycleGAN is a widely used model for unpaired image-to-image translation, where the goal is to translate an image in one domain (e.g., grayscale) into another domain (e.g., RGB). This model is task-specific for the translation between two domains, but it operates in a single direction and requires the cycle consistency loss to ensure the translation is reversible.

Pix2Pix is another image translation model that is optimized for paired image datasets. It is based on conditional GANs and requires a dataset where the input (e.g., grayscale or edge map) and target (e.g., color image) are paired. It performs single-directional translation for specific tasks, such as converting sketches to photos or satellite images to maps.

PROPOSED SYSTEM

The Dual-Mode Web-Based Image Processor is a novel approach that leverages both grayscale and thermal images for various computer vision tasks, including cross-modal image translation, grayscale image colorization, and thermal image reconstruction. Unlike traditional systems that rely on large, task-specific models or are restricted to single-directional image translation, this system

offers the flexibility of dual-mode operation with significant advantages in terms of efficiency, adaptability, and performance on resource-constrained platforms.

The system features a lightweight dual-mode encoder that is capable of encoding both grayscale and thermal images into compact latent vectors. This encoder uses a shared set of parameters to efficiently process images from different modalities, ensuring consistency and minimal computational overhead.

The encoder compresses both input modalities into low-dimensional latent vectors, enabling data compression for easier storage and transmission, which is essential for edge computing or embedded systems. The latent vectors of grayscale images are passed to a colorization generator, which generates corresponding RGB (colorized) images. This process helps restore the color information that is lost in grayscale images, creating high-quality color representations of grayscale inputs.

2-LITERATURE SURVEY

Title: Random reconstructed unpaired image-to-image translation
Author: X. Zhang, C. Fan, Z. Xiao, L. Zhao, H. Chen, and X. Chang
Year: 2023
Description: The goal of unpaired image-to-image translation is to learn a mapping from a source domain to a target domain without using any labeled examples of paired images. This problem can be solved by learning the conditional distribution of source images in the target domain. A major limitation of existing unpaired image-to-image translation algorithms is that they generate untruthful images which are over-colored and lack details, while the translation of realistic images must be rich in details. To address this limitation, we propose a random reconstructed unpaired image-to-image translation (RRUIT) framework by generative adversarial network, which uses random reconstruction to preserve the high-level features in the source and adopts an adversarial strategy to learn the distribution in the target. We update the proposed objective function with two loss functions. The auxiliary loss guides the generator to create a coarse image, while the coarse-to-fine block next to the generator block produces an image that obeys the distribution of the target domain. The coarse-to-fine block contains two sub-modules based on the densely connected atrous spatial pyramid pooling which enriches the details of generated images. We conduct extensive experiments on photorealistic stylization and artistic stylization. The experimental results confirm the superiority of the proposed RRUIT.

Title: EBAT: Enhanced bidirectional and autoregressive transformers for removing hairs in hairy dermoscopic images
Author: Y. Lee and W. You
Year: 2023.
Description: A great progress in deep learning technologies for skin cancer detection from dermoscopic images has been made for a decade. While its performance is vulnerable to a large amount of hairs densely covering the skin surface, the existing image processing methods frequently fail to remove hairs in hairy skin images. In this paper, we propose, as a deep learning approach to removing hairs, a generative image in painting network where bidirectional autoregressive transformers (BATs) are employed to learn image features and are systematically integrated with convolutional neural networks (CNNs) in multiple spatial scales in order to reconstruct missing regions. Each patch split from a masked image is unfolded and processed through BATs, and re-folded to constitute diverse shapes of feature maps through kernel-based unfolding-folding operations. By introducing the multi-scale features extracted by collaborative learning of transformers and CNNs to the texture generator network, our method can effectively reconstruct minute details of local regions as well as global structure which might not be easily inferred from neighbor pixels in hairy skin images. Quantitative and qualitative evaluations show not only that our multi-scale dual-modality strategy is much robust to reconstruct hair-shaped missing regions compared to the existing transformer-based image in painting method called BAT-Fill, but also that our framework outperforms the state-of-the-art image in painting models in removing hairs from hairy dermoscopic images.

Title: IPLF: A novel image pair learning fusion network for infrared and visible image
Author: D. Zhu, W. Zhan, Y. Jiang, X. Xu, and R. Guo
Year: 2023.
Description: Infrared and visible image fusion focuses on the integration of complementary information, whose fused results with abundant texture details and salient targets can cater for human visual observation well. However, the loss of some important details is still a challenge in the image fusion process. In this paper, a novel method based on cross-modality reinforcement module and multi-attention fusion strategy is proposed to boost the end-to-end CNN backbone. Specifically, a cross-modality architecture is applied to compensate the spectral differences from heterogeneous images; and the multi-scale strip pooling is utilized as a further feature representation tool to model long-rang independencies precisely; then, a detail injection block is devised to fulfill the enhancement

requirements of texture-contrast and target intensity; sequentially, a multi-attention fusion module is proposed to integrate features progressively. Extensive comparative experiments are conducted on several datasets to demonstrate the superiority of the proposed method in both quantitative metrics and visual perception, such as the average of visual information fidelity metric based on all the experiment samples reach to 0.964.

Title: Solo GAN: Multi-domain multimodal unpaired image-to-image translation via a single generative adversarial network
Author: S. Huang, C. He, and R. Cheng
Year: 2022
Description: Despite significant advances in image-to-image (I2I) translation with generative adversarial networks (GANs), it remains challenging to effectively translate an image to a set of diverse images in multiple target domains using a single pair of generator and discriminator. Existing I2I translation methods adopt multiple domain-specific content encoders for different domains, where each domain-specific content encoder is trained with images from the same domain only. Nevertheless, we argue that the content (domain-invariance) features should be learned from images among all of the domains. Consequently, each domain-specific content encoder of existing schemes fails to extract the domain-invariant features efficiently. To address this issue, we present a flexible and general SoloGAN model for efficient multimodal I2I translation among multiple domains with unpaired data. In contrast to existing methods, the SoloGAN algorithm uses a single projection discriminator with an additional auxiliary classifier and shares the encoder and generator for all domains. Consequently, the SoloGAN can be trained effectively with images from all domains such that the domain-invariance content representation can be efficiently extracted. Qualitative and quantitative results over a wide range of datasets against several counterparts and variants of the SoloGAN demonstrate the merits of the method, especially for challenging I2I translation datasets, i.e., datasets involving extreme shape variations or need to keep the complex backgrounds unchanged after translations. Furthermore, we demonstrate the contribution of each component in SoloGAN by ablation studies.

Title: Thermal infrared image colorization for nighttime driving scenes with top-down guided attention
Author: S. Mei, R. Jiang, M. Ma, and C. Song
Year: 2022
Description: Benefitting from insensitivity to light and high penetration of foggy environments, infrared cameras are widely used for sensing in nighttime traffic scenes. However, the low contrast and lack of chromaticity of thermal infrared (TIR) images hinder the human interpretation and portability of high-level computer vision algorithms. Colorization to translate a nighttime TIR image into a daytime color (NTIR2DC) image may be a promising way to facilitate nighttime scene perception. Despite recent impressive advances in image translation, semantic encoding entanglement and geometric distortion in the NTIR2DC task remain under-addressed. Hence, we propose a top-down attention And gradient alignment based GAN, referred to as PearlGAN. A top-down guided attention module and an elaborate attentional loss are first designed to reduce the semantic encoding ambiguity during translation. Then, a structured gradient alignment loss is introduced to encourage edge consistency between the translated and input images. In addition, pixel-level annotation is carried out on a subset of FLIR and KAIST datasets to evaluate the semantic preservation performance of multiple translation methods. Furthermore, a new metric is devised to evaluate the geometric consistency in the translation process. Extensive experiments demonstrate the superiority of the proposed PearlGAN over other image translation methods for the NTIR2DC task.

3-PROJECT DESCRIPTION

The Dual-Mode Web-Based Image Processor is an innovative solution designed to tackle the challenges of image translation across different modalities, such as grayscale and thermal images. Traditional computer vision models typically rely on a single sensor modality, either RGB or thermal images, but often fail to take full advantage of the complementary strengths these modalities can offer. This system addresses that gap by utilizing a single lightweight encoder that efficiently processes both grayscale and thermal images, encoding them into compact latent vectors. These latent vectors facilitate cross-modal image translation, enabling two primary tasks: grayscale image colorization and thermal image reconstruction. This flexibility allows the system to handle multiple image processing tasks, making it versatile and adaptable to various applications.
The architecture is designed to be computationally efficient, reducing the burden of traditional models by leveraging a compact encoder and optimizing for both data compression and effective image translation across varying lighting conditions. The system uses an adversarial framework with four distinct generators and two discriminators to generate high-quality, realistic images. It also incorporates reconstruction error terms, ensuring that the translated and reconstructed images preserve essential details like contrast and consistency. The

model is evaluated using various metrics, demonstrating competitive performance in image translation and reconstruction tasks, even in challenging lighting environments. Finally, ablation studies validate the effectiveness of the loss terms used in the system, further confirming their impact on improving the overall performance of the model. This makes the Dual-Mode Web-Based Image Processor a robust and efficient tool for a wide range of image processing applications.

TECHNIQUE USED OR ALGORITHM USED
EXISTING TECHNIQUE: -

- **Single-directional image translation using large**

The algorithm for single-directional image translation using large, task-specific models operates by first accepting input images from a specific modality, such as grayscale or thermal images, and passing them through a preprocessing pipeline to normalize the images for consistent input. The model utilizes a large, task-specific architecture, typically based on Generative Adversarial Networks (GANs), where a generator network is trained to translate images from one modality to another. For example, the system may translate grayscale images to RGB images or thermal images to visible spectrum images. The generator leverages a series of convolutional layers to extract high-level features from the input image and then generates the corresponding output image in the target modality. A discriminator network is used in an adversarial setup to assess whether the generated image closely resembles a real image from the target domain. The model is trained to minimize an adversarial loss, encouraging the generator to create realistic images, while simultaneously minimizing a reconstruction loss to preserve pixel-level accuracy between the generated and target images. In the case of grayscale-to-colorization tasks, the generator learns to restore the lost color information, while in thermal-to-visible reconstruction, it learns to infer visible details from the thermal input. The training procedure is usually supervised, requiring paired datasets where each input image has a corresponding target image. After the model is trained, it can perform single-directional image translation from grayscale to color or thermal to visible images by processing new inputs through the trained generator. The overall architecture is highly task-specific, meaning it performs best for the particular modality translation for which it was trained. Due to the size and complexity of the model, these systems are typically computationally expensive, but they achieve high-quality translations within their narrow domain. Evaluation is conducted using image quality metrics such as Peak Signal-to-Noise Ratio (PSNR), Structural Similarity Index (SSIM), and Fréchet Inception Distance (FID) to assess the fidelity and realism of the generated output

PROPOSED TECHNIQUE USED OR ALGORITHM USED:

- **Dual-Mode Web-Based Image Processor:**

The proposed system operates by processing grayscale and thermal images to perform various tasks such as grayscale image colorization, thermal image reconstruction, and cross-modal image translation. The system is composed of a dual-mode encoder, adversarial training components, and generators designed for each task

The latent vector representation is then passed to one of the task-specific generators, depending on the required output. If the input is a grayscale image, the system utilizes a colorization generator to generate a colorized version of the grayscale image. This generator works by transforming the latent vector back into an RGB image, thus restoring the color information. In the case of a thermal image, the system employs a reconstruction generator, which uses the latent vector to generate a visible spectrum image that mimics the details captured by RGB cameras. Additionally, the system supports cross-modal translation, allowing for tasks such as translating a thermal image into a grayscale image or vice versa.

Training the model involves adversarial learning using a Generative Adversarial Network (GAN) framework. The model includes a discriminator, which evaluates the quality of the generated images by distinguishing between real and fake images. The generator tries to minimize the adversarial loss, creating more realistic images that the discriminator cannot distinguish from real images. Along with the adversarial loss, the system also employs reconstruction loss and consistency loss to ensure that the generated images maintain visual fidelity and structural consistency with the original images

Finally, the performance of the system is evaluated using various quality assessment metrics, such as PSNR (Peak Signal-to-Noise Ratio), SSIM (Structural Similarity Index), and FID (Fréchet Inception Distance). These metrics help assess the visual quality and realism of the generated images. Additionally, ablation studies are performed to evaluate the contribution of each loss term (adversarial, reconstruction, and consistency loss) to the overall system performance. The results from these evaluations demonstrate that the system performs efficiently across different image translation tasks while maintaining high-quality outputs.

International Journal of Information Technology & Computer Engineering

ISSN 2347–3657

Volume 13, Issue 2, 2025

4-REQUIREMENTS ENGINEERING

We can see from the results that on each database, the error rates are very low due to the discriminatory power of features and the regression capabilities of classifiers. Comparing the highest accuracies (corresponding to the lowest error rates) to those of previous works, our results are very competitive.

Hardware Requirements

The hardware requirements may serve as the basis for a contract for the implementation of the system and should therefore be a complete and consistent specification of the whole system. They are used by software engineers as the starting point for the system design. It should what the system do and not how it should be implemented.

- PROCESSOR : DUAL CORE 2 DUOS.
- RAM : 4GB DD RAM
- HARD DISK : 250 GB

Software Requirements

The software requirements document is the specification of the system. It should include both a definition and a specification of requirements. It is a set of what the system should do rather than how it should do it. The software requirements provide a basis for creating the software requirements specification. It is useful in estimating cost, planning team activities, performing tasks and tracking the teams and tracking the team's progress throughout the development activity.

- Operating System : Windows 7/8/10
- Platform : Anaconda
- Programming Language : Python
- Front End : VS code

Functional Requirements

A functional requirement defines a function of a software-system or its component. A function is described as a set of inputs, the behavior, Firstly, the system is the first that achieves the standard notion of semantic security for data confidentiality in attribute-based deduplication systems by resorting to the hybrid cloud architecture.

Non-Functional Requirements

The major non-functional Requirements of the system are as follows

Usability

The system is designed with completely automated process hence there is no or less user intervention.

Reliability

The system is more reliable because of the qualities that are inherited from the chosen platform python. The code built by using python is more reliable.

Performance

This system is developing in the high level languages and using the advanced back-end technologies it will give response to the end user on client system with in very less time.

Supportability

The system is designed to be the cross platform supportable. The system is supported on a wide range of hardware and any software platform, which is built into the system.

Implementation

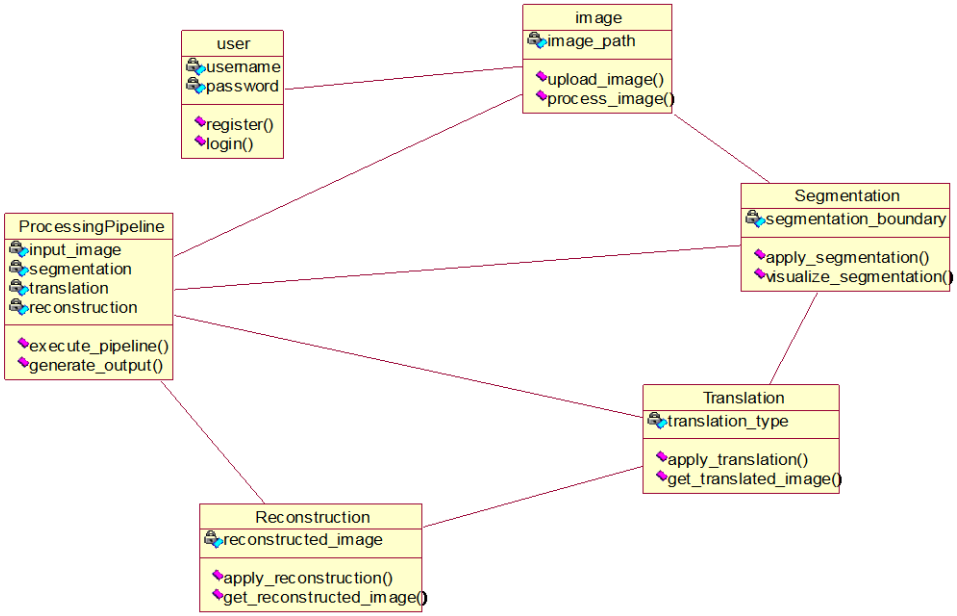
The system is implemented in web environment using Jupyter notebook software. The server is used as the intellignce server and windows 10 professional is used as the platform. Interface the user interface is based on Jupyter notebook provides server system.

5-DESIGN ENGINEERING

Design Engineering deals with the various UML [Unified Modeling language] diagrams for the implementation of project. Design is a meaningful engineering representation of a thing that is to be built. Software design is a process through which the requirements are translated into representation of the software. Design is the place where quality is rendered in software engineering.

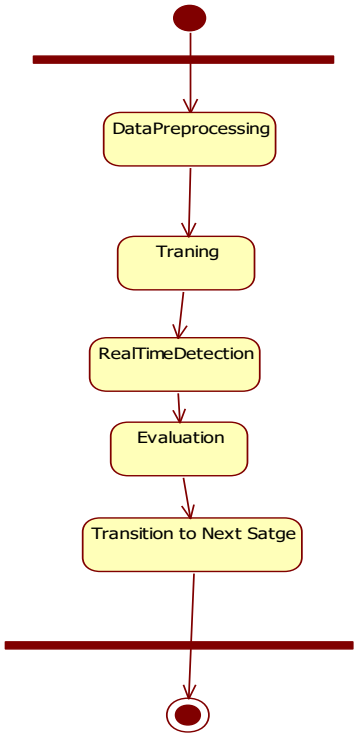
Class DIAGRAM

International Journal of Information Technology & Computer Engineering
ISSN 2347–3657
Volume 13, Issue 2, 2025



In this class diagram represents how the classes with attributes and methods are linked together to perform the verification with security. From the above diagram shown the various classes involved in our project.

ACTIVITY DIAGRAM



Activity diagrams are graphical representations of workflows of stepwise activities and actions with support for choice, iteration and concurrency. In the Unified Modeling Language, activity diagrams can be used to describe the business and operational step-by-step workflows of components in a system. An activity diagram shows the overall flow of control.

International Journal of Information Technology & Computer Engineering

ISSN 2347–3657

Volume 13, Issue 2, 2025

SYSTEM ARCHITECTURE

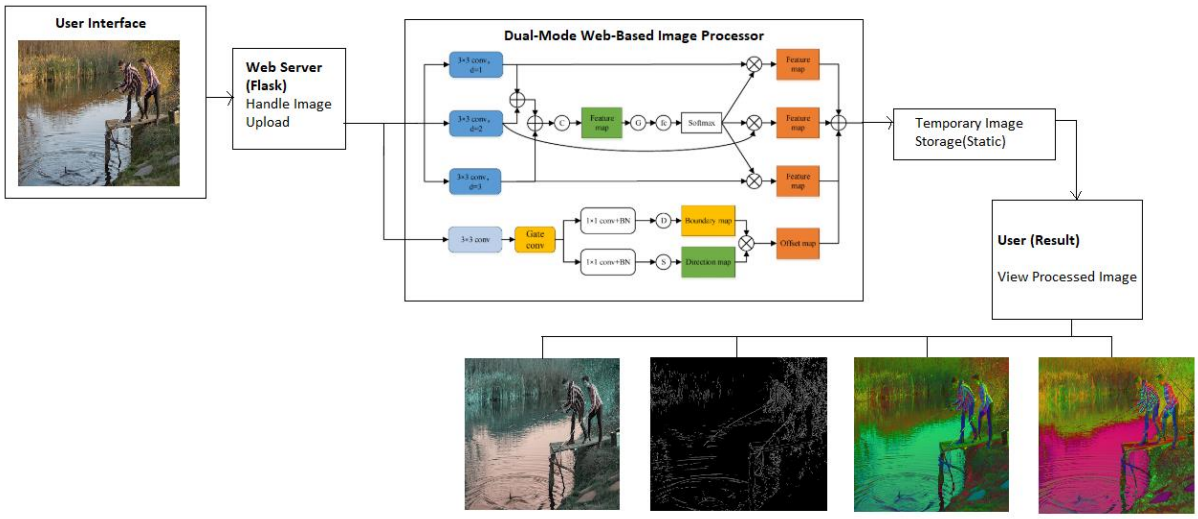


Fig 1: System Architecture

6-CONCLUSION

In this work, we presented a Dual-Mode Web-Based Image Processor that effectively addresses the challenges of image translation across different modalities, specifically grayscale and thermal images. By leveraging a single lightweight encoder, the system efficiently encodes both image types into compact latent vectors, enabling versatile tasks such as grayscale image colorization and thermal image reconstruction. The proposed approach offers significant computational benefits, reducing the typical burden of large, task-specific models while ensuring high-quality image translation even under varied lighting conditions. The use of adversarial learning and the incorporation of reconstruction error terms further enhance the model's robustness, ensuring consistency and contrast preservation. The experimental results demonstrated competitive performance in image translation and reconstruction, confirming the effectiveness of the proposed architecture. Moreover, ablation studies validated the role of the loss terms in improving the overall performance, establishing the proposed model as a powerful tool for cross-modal image translation.

REFERENCES

[1] B. D. Lee and M. H. Sunwoo, ‘‘HDR image reconstruction using segmented image learning,’’ IEEE Access, vol. 9, pp. 142729–142742, 2021.

[2] Y. Lee and W. You, ‘‘EBAT: Enhanced bidirectional and autoregressive transformers for removing hairs in hairy dermoscopic images,’’ IEEE Access, vol. 11, pp. 14225–14235, 2023.

[3] L. Zhai, Y. Wang, S. Cui, and Y. Zhou, ‘‘A comprehensive review of deep learning-based real-world image restoration,’’ IEEE Access, vol. 11,pp. 21049–21067, 2023.

[4] G. Batchuluun, J. K. Kang, D. T. Nguyen, T. D. Pham, M. Arsalan, and K. R. Park, ‘‘Deep learning-based thermal image reconstruction and object detection,’’ IEEE Access, vol. 9, pp. 5951–5971, 2021.

[5] J. Smith, P. Loncomilla, and J. Ruiz-Del-Solar, ‘‘Human pose estimation using thermal images,’’ IEEE Access, vol. 11, pp. 35352–35370, 2023.

[6] G. Batchuluun, Y. W. Lee, D. T. Nguyen, T. D. Pham, and K. R. Park, ‘‘Thermal image reconstruction using deep learning,’’ IEEE Access, vol. 8, pp. 126839–126858, 2020.

[7] M. R. U. Saputra, N. Trigoni, P. P. De Gusmao, C. X. Lu, Y. Almalioglu, S. Rosa, C. Chen, J. Wahlstrom, W. Wang, and A. Markham, ‘‘DeepTIO: A deep thermal-inertial odometry with visual hallucination,’’ IEEE Robot. Autom. Lett., vol. 5, no. 2, pp. 1672–1679, Apr. 2020.

[8] P. Saha and S. Mukhopadhyay, ‘‘Multispectral information fusion with reinforcement learning for object tracking in IoT edge devices,’’ IEEE Sensors J., vol. 20, no. 8, pp. 4333–4344, Apr. 2020.

[9] A. Kaur, D. Mishra, S. Jain, and M. Sarkar, ‘‘Content driven onchip compression and time efficient reconstruction for image sensor applications,’’ IEEE Sensors J., vol. 18, no. 22, pp. 9169–9179, Nov. 2018.

[10] Y. L. Coelho, F. D. A. Santos, A. Frizera-Neto, and T. F. Bastos-Filho, ‘‘A lightweight framework for human activity recognition on wearable devices,’’ IEEE Sensors J., vol. 21, no. 21, pp. 24471–24481, Nov. 2021.
[11] D. Hong, J. Chanussot, N. Yokoya, J. Kang, and X. X. Zhu, ‘‘Learningshared cross-modality representation using multispectral-LiDAR and hyperspectral data,’’ IEEE Geosci. Remote Sens. Lett., vol. 17, no. 8, pp. 1470–1474, Aug. 2020.
[12] R. J. Piechocki, X. Wang, and M. J. Bocus, ‘‘Multimodal sensor fusion in the latent representation space,’’ Sci. Rep., vol. 13, no. 1, p. 2005,Feb. 2023.
[13] C. Choi, J. H. Choi, J. Li, and S. Malla, ‘‘Shared cross-modal trajectory prediction for autonomous driving,’’ in Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit., Jun. 2021, pp. 244–253.
[14] D. Zhu, W. Zhan, Y. Jiang, X. Xu, and R. Guo, ‘‘IPLF: A novel image pair learning fusion network for infrared and visible image,’’ IEEE Sensors J., vol. 22, no. 9, pp. 8808–8817, May 2022.
[15] X. Xiang, N. Lv, Z. Yu, M. Zhai, and A. El Saddik, ‘‘Cross-modality person re-identification based on dual-path multi-branch network,’’ IEEE Sensors J., vol. 19, no. 23, pp. 11706–11713, Dec. 2019.
[16] S. Huang, C. He, and R. Cheng, ‘‘SoloGAN: Multi-domain multimodal unpaired image-to-image translation via a single generative adversarial network,’’ IEEE Trans. Artif. Intell., vol. 3, no. 5, pp. 722–737, Oct. 2022.
[17] C. Wang, C. Xu, and D. Tao, ‘‘Self-supervised pose adaptation for cross-domain image animation,’’ IEEE Trans. Artif. Intell., vol. 1, no. 1, pp. 34–46, Aug. 2020.
[18] P. Isola, J.-Y. Zhu, T. Zhou, and A. A. Efros,‘‘Image-to-image translation with conditional adversarial networks,’’ in Proc. CVPR, 2017, pp. 1125–1134.
[19] J.-Y. Zhu, T. Park, P. Isola, and A. A. Efros, ‘‘Unpaired image-to-image translation using cycle-consistent adversarial networks,’’ in Proc. ICCV, Oct. 2017, pp. 2223–2232.
[20] Z. Yi, H. Zhang, P. Tan, and M. Gong, ‘‘DualGAN: Unsupervised dual learning for image-to-image translation,’’ in Proc. IEEE Int. Conf. Comput. Vis. (ICCV), Oct. 2017, pp. 2868–2876.
[21] T. Kim, M. Cha, H. Kim, J. K. Lee, and J. Kim, ‘‘Learning to discover cross-domain relations with generative adversarial networks,’’ in Proc. 34th Int. Conf. Mach. Learn., vol. 4, 2017, pp. 2941–2949.
[22] X. Zhang, C. Fan, Z. Xiao, L. Zhao, H. Chen, and X. Chang, ‘‘Random reconstructed unpaired image-to-image translation,’’ IEEE Trans. Ind. Informat., vol. 19, no. 3, pp. 3144–3154, Mar. 2023.
[23] X. Huang, M. Y. Liu, S. Belongie, and J. Kautz, ‘‘Multimodal unsupervised image-to-image translation,’’ in Proc. Eur. Conf. Comput. Vis. (Lecture Notes in Computer Science), vol. 11207, Apr. 2018, pp. 179–196
[24] Y. Dalva, S. F. Altindis, and A. Dundar, ‘‘VecGAN: Image-to-image translation with interpretable latent directions,’’ in Proc. Eur. Conf. Comput. Vis. Springer, 2022, pp. 153–169.
[25] A. Nyberg, A. Eldesokey, D. Bergström, and D. Gustafsson, ‘‘Unpaired thermal to visible spectrum transfer using adversarial training,’’ in Proc. Eur. Conf. Comput. Vis. (Lecture Notes in Computer Science), vol. 11134, 2019, pp. 657–669.
[26] H. Patel and K. P. Upla, ‘‘An unsupervised approach for thermal to visible image translation using autoencoder and generative adversarial network,’’ Mach. Vis. Appl., vol. 32, no. 4, pp. 1–18, Jul. 2021.
[27] T. Wang, T. Zhang, L. Liu, A. Wiliem, and B. Lovell, ‘‘CannyGAN: Edgepreserving image translation with disentangled features,’’ in Proc. IEEE Int. Conf. Image Process. (ICIP), Sep. 2019, pp. 514–518.
[28] T. Wang, T. Zhang, and B. C. Lovell, ‘‘EBIT: Weakly-supervised image translation with edge and boundary enhancement,’’ Pattern Recognit. Lett., vol. 138, pp. 534–539, Oct. 2020.
[29] F. Luo, Y. Li, G. Zeng, P. Peng, G. Wang, and Y. Li, ‘‘Thermal infrared image colorization for nighttime driving scenes with top-down guided attention,’’ IEEE Trans. Intell. Transp. Syst., vol. 23, no. 9, pp. 15808–15823, Sep. 2022.
[30] V. V. Kniaz, V. A. Knyaz, J. Hladuvka, W. G. Kropatsch, and V. Mizginov, ‘‘ThermalGAN: Multimodal color-to-thermal image translation for person re-identification in multispectral dataset,’’ in Proc. Eur. Conf. Comput. Vis. (ECCV) Workshops, vol. 11134, 2018, pp. 606–624, 2018.