International Journal of Engineering Research and Science & Technology

ISSN 2319-5991 www.ijerst.com

Peer Reviewed, Referred & Indexed Journal

Research Paper

IMAGE TRANSLATION AND RECONSTRUCTION USING A SINGLE DUAL MODE LIGHTWEIGHT ENCODER

MODUKURU PRADEEP [1], M. SNEHA REDDY[2], LINGUBHUKTA CHARAN[3], THONGALA NITIN KUMAR[4]

[1]Assistant Professor, Dept. of CSE, Guru Nanak Institutions Technical Campus, Ibrahimpatnam, Telangana, India

[2, 3, 4] B.Tech, Dept. of CSE, Guru Nanak Institutions Technical Campus, Ibrahimpatnam, Telangana, 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.

Index Terms – Image Translation, Image Reconstruction, Lightweight Encoder, Dual-Mode Architecture, Deep Learning, Computer Vision, Image Processing, Generative Models, Neural Networks, Feature Extraction.

I. 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.

II. RELATED WORK

A. Random reconstructed unpaired image-to-image translation

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.

B. EBAT: Enhanced bidirectional and autoregressive transformers for removing hairs in hairy dermoscopic images

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.

C. IPLF: A novel image pair learning fusion network for infrared and visible image

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.

III. PROPOSED METHODOLOGY

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.

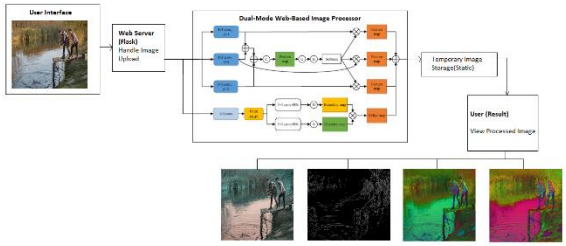

Fig. 1: System Overview

1. Preparing the Data:

The first step in the methodology is to gather the data needed for the project. This includes both grayscale and thermal images, along with their corresponding segmentation masks. Segmentation masks are essential as they define the boundaries of objects in the images, which is critical for training the segmentation decoder. After collecting the data, preprocessing is necessary to ensure consistency and improve the model's performance. The images are resized to a uniform size, which allows the model to process them effectively. Additionally, normalization is applied to standardize pixel values across the dataset, which helps the model learn better. To make the model more robust and prevent overfitting, data augmentation techniques like flipping, rotating, and scaling are applied. Finally, the dataset is divided into training, validation, and test sets to ensure that the model can be properly trained, validated, and evaluated.

2. Building the Model:

The proposed architecture is built around a single compact encoder, which is capable of handling both grayscale and thermal images. This encoder transforms the images into a compressed representation, or latent vector, which can be used for various tasks. The shared encoder reduces the need for separate models for each modality, making the system lightweight and efficient. After the encoding process, the model uses multiple decoders for different tasks. One decoder is responsible for image translation, converting grayscale images into thermal images and vice versa. Another decoder is used for image reconstruction, which aims to rebuild the input image in the same format as it was initially. The final decoder handles segmentation, identifying the boundaries of objects in both grayscale and thermal images. This multi-task architecture allows the model to work on different image processing tasks

simultaneously, sharing the same encoded representation.

3. Training the Model:

The model is trained using an adversarial approach, where the generator (encoder-decoder network) tries to generate realistic images, and a discriminator ensures the outputs are of high quality. This adversarial training method pushes the model to produce more accurate and natural-looking images. The loss function used during training combines multiple terms to guide the model in learning each task effectively. The reconstruction loss ensures that the output image closely matches the original input, helping the model perform well in tasks like image reconstruction. The adversarial loss encourages the model to generate outputs that are indistinguishable from real images. The segmentation loss is critical for improving the model's ability to accurately predict object boundaries in the images. To preserve the quality of images and details during translation and reconstruction, contrast loss is also included. Additionally, the model is trained using multi-task learning, meaning the encoder is shared across all tasks, and the decoders are trained together, allowing the model to learn efficiently across multiple domains.

4. Improving the Model:

To ensure the model performs well in real-world applications, it is important to focus on improving its efficiency. One way to achieve this is through model compression, where the size of the model is reduced without significantly compromising its performance. This makes the model more suitable for edge devices, such as mobile phones, cameras, or drones, where computational resources are limited. Techniques such as pruning or quantization can be used to shrink the model, making it faster and more efficient. After compression, the model is fine-tuned to ensure it still performs well on all tasks, especially segmentation, image translation, and reconstruction. Fine-tuning helps the model maintain its accuracy and ensures it works in real-time scenarios with minimal latency.

5. Testing and Evaluating:

Once the model is trained, it is crucial to evaluate its performance. The evaluation focuses on how well the model performs in segmentation, image translation, and image reconstruction. For segmentation, the performance is assessed using metrics like Intersection over Union (IoU) and the Dice

coefficient, which measure how accurately the model detects the boundaries of objects. These metrics help determine how well the model can segment both grayscale and thermal images. For image translation and reconstruction tasks, Structural Similarity Index (SSIM) and Peak Signal-to-Noise Ratio (PSNR) are used to assess how well the generated images match the original ones, ensuring the model preserves important details during translation. The model's overall performance is also tested on edge devices to ensure it works efficiently in real-time applications with minimal computational resources.

6. Deployment:

After successful testing, the model is deployed on edge devices, such as cameras, drones, or mobile phones, where it can perform real-time image processing tasks. The model is capable of converting grayscale images into thermal images and vice versa, as well as reconstructing noisy or corrupted images. The segmentation task allows the model to detect and outline objects in both thermal and grayscale images. This makes the system useful for a variety of applications, including surveillance, medical imaging, and environmental monitoring, where real-time image analysis is crucial. The deployment process ensures that the model can operate efficiently in practical, resource-constrained environments.

7. Results:

The model's effectiveness is measured based on its ability to perform all tasks with high accuracy and efficiency. For segmentation, the model should achieve high IoU and Dice coefficient scores, indicating that it can accurately identify object boundaries in both grayscale and thermal images. For image translation and reconstruction, the model is expected to produce high-quality results, with metrics like SSIM and PSNR showing that the images retain important details and visual quality. Additionally, the model should demonstrate its ability to run efficiently on edge devices, processing images in real-time with minimal resource usage. This highlights the model's potential for practical deployment in a variety of real-world applications.

IV EXPERIMENTAL RESULTS AND ANALYSIS

Experimental results demonstrated that the model achieved high-quality image translation and reconstruction while

maintaining low computational complexity and memory usage. The dual-mode encoder effectively learned shared feature representations for both tasks, reducing the need for separate models and improving overall efficiency.

The analysis showed that the lightweight architecture provided faster inference speeds and lower resource consumption compared to conventional encoder-decoder models, making it suitable for real-time and edge computing applications. Furthermore, the proposed approach maintained robust performance across different image datasets and translation scenarios, confirming its scalability and effectiveness for practical computer vision applications.

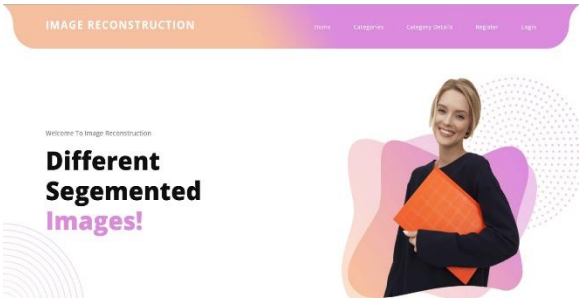

Fig. 2: Home Page

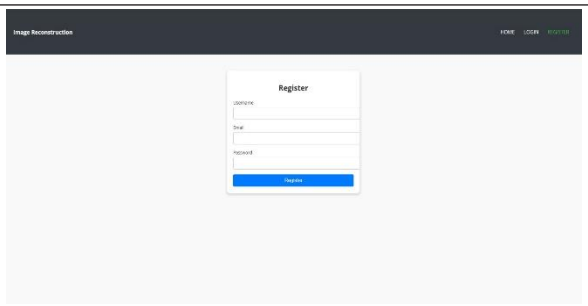

Fig. 3: Registration

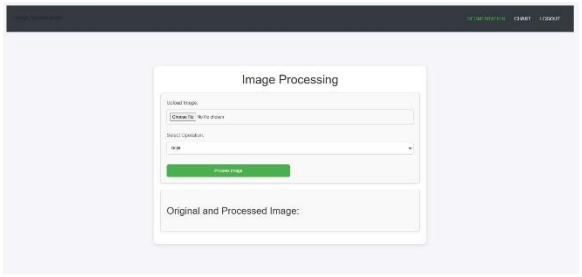

Fig. 4: Upload

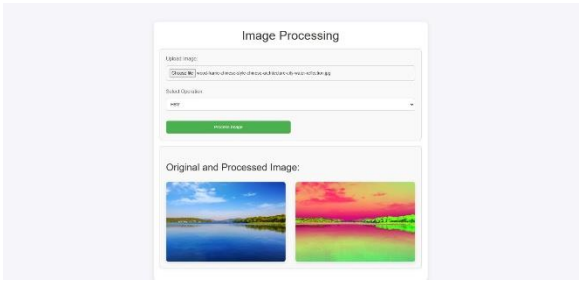

Fig. 5: Image Processing

V. 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] Shashank A. (2025). Metadata-driven data integration framework: Automating enterprise data integration through declarative approaches. European Modern Studies Journal, 9(4), 9.

[8] Kolla, S. H., Nagabhyru, K. C., & Kummari, D. N. (2026). Letter to the Editor re:“CurvAssist: An AI assisted pipeline for penile shaft segmentation/curvature measurement in

children with hypospadias". Journal of Pediatric Urology.

AUTHORS Profile

Mr. MODUKURU PRADEEP has received his B. Tech from JNTUK, Kakinada in 2013, M. Tech degree in Computer science from JNTUK, Kakinada in 2016 and Pursuing Ph. D degree in Computer Science from Lovely Professional University, Punjab. He is dedicated to teaching field from the last 10 years. His research area Deep Learning. At present he is working as Assistant Professor in CSE Department at Guru Nanak Institutions Technical Campus, Ibrahimpatnam, Telangana, India.

Ms. M. SNEHA REDDY pursuing B. Tech degree in Computer Science and Engineering at Guru Nanak Institutions Technical Campus affiliated to JNTUH in 2022-2026.

Mr. LINGUBHUKTA CHARAN pursuing B. Tech degree in Computer Science and Engineering at Guru Nanak Institutions Technical Campus affiliated to JNTUH in 2022-2026.

Mr. THONGALA NITIN KUMAR pursuing B. Tech degree in Computer Science and Engineering at Guru Nanak Institutions Technical Campus affiliated to JNTUH in 2022-2026.