A Unified Framework for RGB-Infrared Transfer

Subjects: Computer Science, Artificial Intelligence

Contributor: Qiyang Sun , Xia Wang , Changda Yan , Xin Zhang

Infrared(IR) images (both 0.7-3 μm and 8-15 μm) offer radiation intensity texture information that visible images lack, making them particularly helpful in daytime, nighttime, and complex scenes. Many researchers are studying how to translate RGB images into infrared images for deep learning-based visual tasks such as object tracking, crowd counting, panoramic segmentation, and image fusion in urban scenarios. The utilization of the RGB-IR dataset in the aforementioned tasks holds the potential to provide comprehensive multi-band fusion data for urban scenes, thereby facilitating precise modeling across different scenarios. In addressing the challenge of accurately generating high-radiance textures for the targets in the infrared spectrum, the proposed approach aims to ensure alignment between the generated infrared images and the radiation feature of ground-truth IR images.

infrared image image-to-image translation multi-modal controls vector quantization transformer

1. Introduction

Complex illumination scenarios have adversely affected the accuracy of visible light data in recent years. Unfortunately, these factors are beyond our control and significantly reduce the usefulness of images. This situation poses a significant challenge in processing and training, limiting the range of applications for these data. Infrared (IR) images (both 0.7–3 μm and 8–15 μm) offer radiation intensity texture information that visible images lack, making them particularly helpful in daytime, nighttime, and complex scenes. In low-light conditions, infrared images captured through thermal radiation (8–15 μm) provide enriched semantic information. Objects exhibiting high thermal temperatures can reveal discernible features within intricate scenes. Therefore, based on deep learning, cross-modal image translation has become a hot topic in remote sensing research in recent years. Many researchers are studying how to translate RGB images into infrared images for deep learning-based visual tasks such as object tracking, crowd counting, panoramic segmentation, and image fusion in urban scenarios. The utilization of the RGB-IR dataset in the aforementioned tasks holds the potential to provide comprehensive multi-band fusion data for urban scenes, thereby facilitating precise modeling across different scenarios.

A large-scale neural network algorithm based on RGB features can be trained using large monomodal public datasets, such as ImageNet [1], PASCAL VOC [2], and MS COCO [3]. However, compared to RGB datasets, infrared image public datasets often suffer from limitations such as limited scene diversity, a lack of diverse target categories, low data volume, and low resolution. Therefore, researchers have developed a large number of deep-learning-based style transfer algorithms to achieve the conversion from RGB images to infrared images through end-to-end translation, such as CNN [4][5][6], GAN [7][8][9], attention networks [10][11][12], etc., to learn and fit the mapping relationship between RGB and IR images. These RGB-IR algorithms approach the task by solving a pixel-level conditional generation problem. IR images convert the radiation intensity field into grayscale images, leading to a mapping relationship between IR and RGB images that is not based on spectral physical characteristics. As a result, there is no strict pixel-level correspondence [13]. The research conducted in [14][15] indicated that, while a conditioned generative model can successfully generate customized IR images, these models primarily focus on studying the texture or content transformation from RGB to IR, without considering the diverse types of migration mapping relationship between different visual fields. The mono-modality transformation predominantly relies on simplistic semantic matching and transferring strategies, leading to unrealistic expression of radiation information. Due to the global feature extraction and generation mechanisms of the transfer model, vehicles and pedestrians exhibit significant disparities between the generated infrared textures and the ground-truth, and they may even be overlooked in some results. Consequently, this limits their flexibility and versatility in various scenarios and tasks. In practical applications, it is crucial for the model to accurately translate complex and diverse scenes, data, and task requirements. Therefore, designing a unified visibility-infrared migration framework suitable for multi-scene and multi-task purposes holds significant practical value.

To this end, researchers propose a novel multi-modal translation approach. This method not only enhances the overall naturalness of human-computer interaction but also consolidates the information from multiple data sources to generate more comprehensive results.

2. Translation from Image to Image

The translation from image to image was first discussed in [16], to learn the mapping function between source and target domains. This style of transfer work mainly deals with two significant challenges: First, the imaging principles of IR and RGB sensors differ, and the radiation field where IR is located varies significantly from the color space. Consequently, traditional methods find it difficult to determine the mapping relationship between RGB and IR. Second, the mainstream infrared migration methods are based on end-to-end generative adversarial networks. Among them, cycle consistency is used to handle unpaired data [17][18], while the enhanced attribute space is proposed to provide diversity [19]. Most algorithms for translating infrared images introduced architectures based on Cycle-GAN, such as Drit++ et al. [20][21][22][23]. In addition, some other algorithms also provide appropriate structural solutions for this task.For example, FastCUT [24] adopts one-sided translation without using cycle consistency to improve diversity [25][26], and U-GAT-T [27] focuses explicitly on geometric transformations of content in translation. Kuang et al. [28] improved the pix2pix method and proposed TIC-CGAN, the first GAN application to translate thermal IR (8–15 μm) images in traffic scenes. The generator in ThermalGAN [29] utilized a U-Net-based architecture, and the authors used a unique dataset named ThermalWorld to enhance training. In DRIT [21], the authors introduced the use of multiple generators. Each generator focused on learning attributes of different scenes, and a classifier based on ResNet [30] was used to determine which generator's output was most suitable for a given input image.

Wang et al. [31] proposed an attention-based hierarchical infrared image coloring network (AHTIC-Net) to enhance the realistic and rich texture information of small objects in translated images. It employed a multi-scale structure to extract features of objects with different sizes, thereby improving the model's focus on small objects during training. In recent years, many migration models have leaned towards using universal style transfer (UST) methods. Representative UST methods include AdaIN [32], WCT [33], and Avatar-Net [34]. These methods have been continuously expanded upon [35][36][37]. However, they are limited in terms of disentanglement and reconstruction of image content during the stylization process. In addition, the research on extracting image content structure and texture style features has become increasingly mature. Gatys et al. [38] found that the layers in CNN can extract content structure and style texture, they proposed an optimization-based iterative generation method for stylized images. Li and Justin [39][40] used an end-to-end model to achieve real-time style transfer with a specific style. To enable more efficient applications, Stylebank et al. [41][42][43] combined multiple types into one model and achieved excellent stylization results. Chen et al. [44] proposed an internal–external style transfer algorithm (IEST) that includes two contrastive losses, which can generate more natural stylized effects. However, the existing encoder–transfer–decoder style transfer methods cannot handle random dependencies, which may result in the loss of detailed information.

Recently, the effectiveness of vector quantization (VQ) technology as an intermediate representation for generative models has been demonstrated [45][46]. Therefore, in the RGB-IR study, the researchers explore the suitability of using vectorization as an encoder in RGB-IR tasks, where the latent representation obtained through vector quantization serves as the intermediate vector for RGB-IR tasks.

References

1. Russakovsky, O.; Deng, J.; Su, H.; Krause, J.; Satheesh, S.; Ma, S.; Huang, Z.; Karpathy, A.; Khosla, A.; Bernstein, M.; et al. Imagenet large scale visual recognition challenge. Int. J. Comput. Vis. 2015, 115, 211–252.
2. Everingham, M.; Eslami, S.A.; Van Gool, L.; Williams, C.K.; Winn, J.; Zisserman, A. The pascal visual object classes challenge: A retrospective. Int. J. Comput. Vis. 2015, 111, 98–136.
3. Lin, T.Y.; Maire, M.; Belongie, S.; Hays, J.; Perona, P.; Ramanan, D.; Dollár, P.; Zitnick, C.L. Microsoft coco: Common objects in context. In Proceedings, Part V 13, Proceedings of the Computer Vision—ECCV 2014: 13th European Conference, Zurich, Switzerland, 6–12 September 2014; Springer: Berlin/Heidelberg, Germany, 2014; pp. 740–755.
4. Chang, Y.; Luo, B. Bidirectional convolutional LSTM neural network for remote sensing image super-resolution. Remote Sens. 2019, 11, 2333.
5. Gu, J.; Sun, X.; Zhang, Y.; Fu, K.; Wang, L. Deep residual squeeze and excitation network for remote sensing image super-resolution. Remote Sens. 2019, 11, 1817.

6. Lu, T.; Wang, J.; Zhang, Y.; Wang, Z.; Jiang, J. Satellite image super-resolution via multi-scale residual deep neural network. Remote Sens. 2019, 11, 1588.
7. Haut, J.M.; Fernandez-Beltran, R.; Paoletti, M.E.; Plaza, J.; Plaza, A.; Pla, F. A new deep generative network for unsupervised remote sensing single-image super-resolution. IEEE Trans. Geosci. Remote Sens. 2018, 56, 6792–6810.
8. Lei, S.; Shi, Z.; Zou, Z. Coupled adversarial training for remote sensing image super-resolution. IEEE Trans. Geosci. Remote Sens. 2019, 58, 3633–3643.
9. Xiong, Y.; Guo, S.; Chen, J.; Deng, X.; Sun, L.; Zheng, X.; Xu, W. Improved SRGAN for remote sensing image super-resolution across locations and sensors. Remote Sens. 2020, 12, 1263.
10. Zhang, D.; Shao, J.; Li, X.; Shen, H.T. Remote sensing image super-resolution via mixed high-order attention network. IEEE Trans. Geosci. Remote Sens. 2020, 59, 5183–5196.
11. Salvetti, F.; Mazzia, V.; Khaliq, A.; Chiaberge, M. Multi-image super resolution of remotely sensed images using residual attention deep neural networks. Remote Sens. 2020, 12, 2207.
12. Zhang, S.; Yuan, Q.; Li, J.; Sun, J.; Zhang, X. Scene-adaptive remote sensing image super-resolution using a multiscale attention network. IEEE Trans. Geosci. Remote Sens. 2020, 58, 4764–4779.
13. Yang, S.; Sun, M.; Lou, X.; Yang, H.; Zhou, H. An unpaired thermal infrared image translation method using GMA-CycleGAN. Remote Sens. 2023, 15, 663.
14. Huang, S.; Jin, X.; Jiang, Q.; Liu, L. Deep learning for image colorization: Current and future prospects. Eng. Appl. Artif. Intell. 2022, 114, 105006.
15. Liang, W.; Ding, D.; Wei, G. An improved DualGAN for near-infrared image colorization. Infrared Phys. Technol. 2021, 116, 103764.
16. Isola, P.; Zhu, J.Y.; Zhou, T.; Efros, A.A. Image-to-image translation with conditional adversarial networks. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Honolulu, HI, USA, 21–26 July 2017; pp. 1125–1134.
17. Zhu, J.Y.; Park, T.; Isola, P.; Efros, A.A. Unpaired image-to-image translation using cycle-consistent adversarial networks. In Proceedings of the IEEE International Conference on Computer Vision, Venice, Italy, 22–29 October 2017; pp. 2223–2232.
18. Liu, M.Y.; Breuel, T.; Kautz, J. Unsupervised image-to-image translation networks. In Proceedings of the Advances in Neural Information Processing Systems, Long Beach, CA, USA, 4–9 December 2017; Volume 30.
19. Zhu, J.Y.; Zhang, R.; Pathak, D.; Darrell, T.; Efros, A.A.; Wang, O.; Shechtman, E. Toward multimodal image-to-image translation. In Proceedings of the Advances in Neural Information Processing Systems, Long Beach, CA, USA, 4–9 December 2017; Volume 30.

20. Zhang, L.; Gonzalez-Garcia, A.; Van De Weijer, J.; Danelljan, M.; Khan, F.S. Synthetic data generation for end-to-end thermal infrared tracking. IEEE Trans. Image Process. 2018, 28, 1837–1850.
21. Cui, Z.; Pan, J.; Zhang, S.; Xiao, L.; Yang, J. Intelligence Science and Big Data Engineering. Visual Data Engineering. In Proceedings, Part I, Proceedings of the 9th International Conference, IScIDE 2019, Nanjing, China, 17–20 October 2019; Springer Nature: Berlin/Heidelberg, Germany, 2019; Volume 11935.
22. Lee, H.Y.; Tseng, H.Y.; Huang, J.B.; Singh, M.; Yang, M.H. Diverse image-to-image translation via disentangled representations. In Proceedings of the European Conference on Computer Vision (ECCV), Munich, Germany, 8–14 September 2018; pp. 35–51.
23. Lee, H.Y.; Tseng, H.Y.; Mao, Q.; Huang, J.B.; Lu, Y.D.; Singh, M.; Yang, M.H. Drit++: Diverse image-to-image translation via disentangled representations. Int. J. Comput. Vis. 2020, 128, 2402–2417.
24. Park, T.; Efros, A.A.; Zhang, R.; Zhu, J.Y. Contrastive learning for unpaired image-to-image translation. In Proceedings, Part IX 16, Proceedings of the Computer Vision—ECCV 2020: 16th European Conference, Glasgow, UK, 23–28 August 2020; Springer: Berlin/Heidelberg, Germany, 2020; pp. 319–345.
25. Mao, Q.; Lee, H.Y.; Tseng, H.Y.; Ma, S.; Yang, M.H. Mode seeking generative adversarial networks for diverse image synthesis. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, Long Beach, CA, USA, 15–20 June 2019; pp. 1429–1437.
26. Mao, Q.; Tseng, H.Y.; Lee, H.Y.; Huang, J.B.; Ma, S.; Yang, M.H. Continuous and diverse image-to-image translation via signed attribute vectors. Int. J. Comput. Vis. 2022, 130, 517–549.
27. Lee, H.Y.; Li, Y.H.; Lee, T.H.; Aslam, M.S. Progressively Unsupervised Generative Attentional Networks with Adaptive Layer-Instance Normalization for Image-to-Image Translation. Sensors 2023, 23, 6858.
28. Kuang, X.; Zhu, J.; Sui, X.; Liu, Y.; Liu, C.; Chen, Q.; Gu, G. Thermal infrared colorization via conditional generative adversarial network. Infrared Phys. Technol. 2020, 107, 103338.
29. Kniaz, V.V.; Knyaz, V.A.; Hladuvka, J.; Kropatsch, W.G.; Mizginov, V. Thermalgan: Multimodal color-to-thermal image translation for person re-identification in multispectral dataset. In Proceedings of the European Conference on Computer Vision (ECCV) Workshops, Munich, Germany, 8–14 September 2018.
30. He, K.; Zhang, X.; Ren, S.; Sun, J. Deep residual learning for image recognition. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Las Vegas, NV, USA, 27–30 June 2016; pp. 770–778.

31. Wang, H.; Cheng, C.; Zhang, X.; Sun, H. Towards high-quality thermal infrared image colorization via attention-based hierarchical network. Neurocomputing 2022, 501, 318–327.

32. Huang, X.; Belongie, S. Arbitrary style transfer in real-time with adaptive instance normalization. In Proceedings of the IEEE International Conference on Computer Vision, Venice, Italy, 22–29 October 2017; pp. 1501–1510.

33. Li, Y.; Fang, C.; Yang, J.; Wang, Z.; Lu, X.; Yang, M.H. Universal style transfer via feature transforms. In Proceedings of the Advances in Neural Information Processing Systems, Long Beach, CA, USA, 4–9 December 2017; Volume 30.

34. Sheng, L.; Lin, Z.; Shao, J.; Wang, X. Avatar-net: Multi-scale zero-shot style transfer by feature decoration. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Salt Lake City, UT, USA, 18–23 June 2018; pp. 8242–8250.

35. Gu, S.; Chen, C.; Liao, J.; Yuan, L. Arbitrary style transfer with deep feature reshuffle. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Salt Lake City, UT, USA, 18–23 June 2018; pp. 8222–8231.

36. Jing, Y.; Liu, X.; Ding, Y.; Wang, X.; Ding, E.; Song, M.; Wen, S. Dynamic instance normalization for arbitrary style transfer. In Proceedings of the AAAI Conference on Artificial Intelligence, New York, NY, USA, 7–12 February 2020; Volume 34, pp. 4369–4376.

37. An, J.; Li, T.; Huang, H.; Shen, L.; Wang, X.; Tang, Y.; Ma, J.; Liu, W.; Luo, J. Real-time universal style transfer on high-resolution images via zero-channel pruning. arXiv 2020, arXiv:2006.09029.

38. Gatys, L.A.; Ecker, A.S.; Bethge, M. Image style transfer using convolutional neural networks. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Las Vegas, NV, USA, 27–30 June 2016; pp. 2414–2423.

39. Johnson, J.; Alahi, A.; Fei-Fei, L. Perceptual losses for real-time style transfer and super-resolution. In Proceedings, Part II 14, Proceedings of the Computer Vision—ECCV 2016: 14th European Conference, Amsterdam, The Netherlands, 11–14 October 2016; Springer: Berlin/Heidelberg, Germany, 2016; pp. 694–711.

40. Li, C.; Wand, M. Precomputed real-time texture synthesis with markovian generative adversarial networks. In Proceedings, Part II 14, Proceedings of the Computer Vision—ECCV 2016: 14th European Conference, Amsterdam, The Netherlands, 11–14 October 2016; Springer: Berlin/Heidelberg, Germany, 2016; pp. 702–716.

41. Chen, D.; Yuan, L.; Liao, J.; Yu, N.; Hua, G. Stylebank: An explicit representation for neural image style transfer. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Honolulu, HI, USA, 21–26 July 2017; pp. 1897–1906.

42. Dumoulin, V.; Shlens, J.; Kudlur, M. A learned representation for artistic style. arXiv 2016, arXiv:1610.07629.

43. Lin, M.; Tang, F.; Dong, W.; Li, X.; Xu, C.; Ma, C. Distribution aligned multimodal and multi-domain image stylization. ACM Trans. Multimed. Comput. Commun. Appl. (TOMM) 2021, 17, 1–17.
44. Chen, H.; Wang, Z.; Zhang, H.; Zuo, Z.; Li, A.; Xing, W.; Lu, D. Artistic style transfer with internal-external learning and contrastive learning. In Proceedings of the Advances in Neural Information Processing Systems, Online, 6–14 December 2021; Volume 34, pp. 26561–26573.
45. Esser, P.; Rombach, R.; Ommer, B. Taming transformers for high-resolution image synthesis. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, Nashville, TN, USA, 20–25 June 2021; pp. 12873–12883.
46. Razavi, A.; Van den Oord, A.; Vinyals, O. Generating diverse high-fidelity images with vq-vae-2. In Proceedings of the Advances in Neural Information Processing Systems, Vancouver, BC, Canada, 8–14 December 2019; Volume 32.

Retrieved from https://encyclopedia.pub/entry/history/show/119262