

Article

MWIRGAN: Unsupervised Visible-to-MWIR Image Translation with Generative Adversarial Network

Mohammad Shahab Uddin ¹, Chiman Kwan ² and Jiang Li ^{1,*}¹ Department of Electrical and Computer Engineering, Old Dominion University, Norfolk, VA 23625, USA² Applied Research LLC, Rockville, MD 20850, USA

* Correspondence: jli@odu.edu

Abstract: Unsupervised image-to-image translation techniques have been used in many applications, including visible-to-Long-Wave Infrared (visible-to-LWIR) image translation, but very few papers have explored visible-to-Mid-Wave Infrared (visible-to-MWIR) image translation. In this paper, we investigated unsupervised visible-to-MWIR image translation using generative adversarial networks (GANs). We proposed a new model named MWIRGAN for visible-to-MWIR image translation in a fully unsupervised manner. We utilized a perceptual loss to leverage shape identification and location changes of the objects in the translation. The experimental results showed that MWIRGAN was capable of visible-to-MWIR image translation while preserving the object's shape with proper enhancement in the translated images and outperformed several competing state-of-the-art models. In addition, we customized the proposed model to convert game-engine-generated (a commercial software) images to MWIR images. The quantitative results showed that our proposed method could effectively generate MWIR images from game-engine-generated images, greatly benefiting MWIR data augmentation.

Keywords: deep learning; mid-wave infrared (MWIR) videos; image translation; game engine; generative adversarial network



Citation: Uddin, M.S.; Kwan, C.; Li, J. MWIRGAN: Unsupervised Visible-to-MWIR Image Translation with Generative Adversarial Network. *Electronics* **2023**, *12*, 1039. <https://doi.org/10.3390/electronics12041039>

Academic Editor: Yu-Chen Hu

Received: 30 December 2022

Revised: 26 January 2023

Accepted: 16 February 2023

Published: 20 February 2023



Copyright: © 2023 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

1. Introduction

Unsupervised image-to-image translation involves translating an image from one domain to another without labels or image pairs. Infrared (IR) images are very useful in the research areas of defense, but very little research has been conducted in this domain. In addition, there is a scarcity of data challenge in this domain which has a negative effect on the research related to the IR domain. Therefore, there is a need to explore visible/IR image translation to benefit these applications.

There are some research papers in the literature on visible-to-LWIR (Long-Wave Infrared) image conversion [1,2], but much fewer efforts were focused on visible-to-MWIR (Mid-Wave Infrared) image translation. LWIR is less effective for long-range target identification, and it is more prone to be influenced by humidity than MWIR. In addition, MWIR has better thermal contrast than LWIR, and MWIR images are preferred for long-distance applications, including coastal surveillance. MWIR is very effective for the detection of airplanes or missiles. In this paper, we focus on visible-to-MWIR image translation with the generative adversarial network (GAN) to advance research in this domain. Our main contributions are summarized as follows:

- We proposed a GAN-based model named MWIRGAN to perform visible-to-MWIR image translation, which utilizes perceptual loss to guide the translation. The learning is achieved in a fully unsupervised manner.
- We proposed a novel technique to generate MWIR images from game-engine-generated images, which can greatly benefit data augmentation for MWIR image applications.

2. Related Work

2.1. Image-to-Image Translation

GAN models are highly effective for image-to-image translation and have been used for generating synthetic images in various applications [3–8]. Image-to-image translation has many practical applications, such as domain adaption [9], colorization [10], and super-resolution [11]. Pix2Pix GAN [3,12] showed that it is possible to perform an image-to-image translation between two complex domains, but the main drawback is that it needs the paired images from both domains for training. Later on, several GAN-based models were proposed to perform image-to-image translation with unpaired images, such as CycleGAN [4] and DCLGAN [13]. In the early years, most of the image-to-image translation work involved two visible image domains, and very little research was on visible-to-IR image conversion. In recent years, researchers have worked on NIR (near-infrared)-to-visible image translation [14–18], MWIR to grayscale translation [19], and LWIR-to-RGB conversion [12,20] with some success in a supervised manner.

Unsupervised image conversion has also been explored in the recent literature. The authors in [1] modified CycleGAN architecture in an unsupervised manner for image translation between RGB and LWIR. They utilized the faster RCNN network to generate bounding boxes of detected objects to calculate an object-specific loss. The main drawback of their method is that they still needed labeled images to train the faster RCNN network. In addition, the accuracy of the detector influenced the performance of the model. In our proposed model, it neither needs any detector in the model nor paired images for training. Moreover, we focused on long-range targets at 1000–2000 m distance in MWIR images, while their work was on short-range targets at 0–50 m in LWIR images. Our proposed model was built upon the CycleGAN architecture, and we made several modifications to overcome the drawbacks of CycleGAN during visible-to-MWIR image translation.

2.2. Visible-to-IR Image Translation

There are few researchers who have given attention to visible-to-IR image translation. For example, ThermalGAN [2] was proposed to transform color images into LWIR images using a temperature vector, where object temperature and background information were contained in a temperature vector. A thermal segmentation generator and a relative contrast generator were used when generating LWIR images. Pix2Pix GAN [3,12] was proposed for RGB-to-TIR image generation. In [12], the authors showed that Pix2Pix GAN achieved better performances than CycleGAN, but Pix2Pix GAN required paired dataset for training. In [21], the authors also used Pix2Pix GAN to generate infrared images. The main drawback of this method is that it was time-consuming, as some steps were performed manually. Pix2Pix GAN was also modified to generate infrared textures from visible images [22], and, subsequently, the infrared textures were combined with structure from motion (SfM)-constructed 3D models to generate the infrared images. The authors in [23] utilized conditional GAN to construct NIR spectral bands from RGB images with a paired dataset for training. Cycle GAN [4,24] was also used for visible-to-IR image translation.

AttentionGAN was also proposed in [25] for visible-to-MWIR video conversion for long distances where they used attention maps, generated by a modified resnet-18, to focus on the regions of interest. However, the attention maps were generated from two different domains, and domain mismatch might have existed, negatively affecting the performance of the model. In our proposed model, we utilized the learned perceptual image patch similarity (LPIPS) metric [26] from the same domain to calculate the perceptual loss. Therefore, the domain mismatch is not an issue in our model. Our model can perform visible-to-MWIR translation without the deformation of the objects in the images while keeping fine details.

3. Method

3.1. Architecture

We used a model based on the architecture of CycleGAN [4]. The proposed model can perform unsupervised visible-to-MWIR image conversion using an unpaired dataset, as shown in Figure 1, which consists of two generators (G_X and G_Y) and two discriminators (D_X and D_Y). We used the ResNet [27] architecture for the generators proposed in CycleGAN with some modifications mentioned in [28]. Each generator has 9 residual blocks along with several convolution layers. The architecture of the generator is [c7s1-64, d128, d256, R256, R256, R256, R256, R256, R256, R256, R256, u128, u64, c7s1-3], where c7s1-k represents a 7×7 Convolution-BatchNorm-ReLU layer having k filters with stride 1, dk represents a 3×3 Convolution-BatchNorm-ReLU layer having k filters with stride 2, Rk indicates a residual block containing two 3×3 convolutional layers with stride 1 and k filters in both layers, and uk is a 3×3 fractional-strided-Convolution-BatchNorm-ReLU layer having k filters with stride $1/2$.

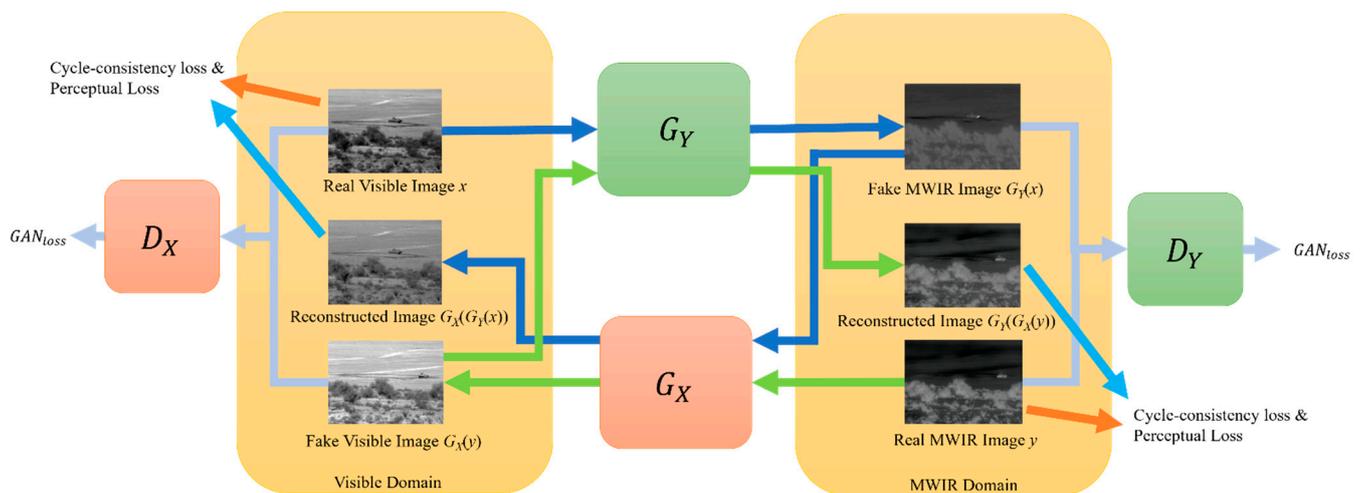


Figure 1. Framework of the proposed model for converting visible images to MWIR images. Our proposed model has two generators and two discriminators. We calculate the cycle-consistency loss and the perceptual loss using the real images (x , y) and the reconstructed images ($G_X(G_Y(x))$, $G_Y(G_X(y))$) along with the GAN loss during training our model. Arrows with different colors indicate signal flows through the generators and discriminators.

We used the architecture of the patch-based structure [29] for our discriminators, which achieved good performances in the previous work [3,8,28,29]. The architecture of our discriminators is [C64, C128, C256, C512, C512], where Ck represents a 4×4 Convolution-BatchNorm-LeakyReLU layer having k filters with stride 2. We used an adaptive average pooling layer and a convolution layer after the last layer to obtain one-dimensional output from the discriminators. A patch-based discriminator works with the image patches to determine whether the patches are real or fake. These patch-based discriminators have several advantages. They have fewer parameters than a global discriminator and can work on any image size in a fully convolutional manner [3,8]. In addition, a patch-based discriminator contributes to the fast convergence of the generator [8].

3.2. Loss Function

The overall loss function consists of three components: GAN loss, perceptual loss, and cycle consistency loss. In our model, there are two generators, G_Y and G_X . Given the two domains, visible and MWIR, letting G_Y perform visible to MWIR mapping and G_X MWIR to visible mapping, x is an image in the visible domain and y in the MWIR domain, $G_Y(x)$ is then a generated MWIR image, and $G_X(y)$ is a generated visible image. During training, the discriminator D_Y discriminates y from $G_Y(x)$ and D_X discriminates x from $G_X(y)$.

The GAN loss component is defined as [5]

$$L_{GAN} = L_{GAN}(G_Y, D_Y, X, Y) + L_{GAN}(G_X, D_X, Y, X) \quad (1)$$

GAN loss alone cannot generate good-quality images [5]. Since there are no paired images in the two domains, we used the cycle-consistency loss [4,24] to measure the differences between the original images (x, y) and the reconstructed images ($G_X(G_Y(x)), G_Y(G_X(y))$), encouraging the model to learn mappings that are both plausible and consistent with the input data:

$$L_{cyc} = ||G_X(G_Y(x)) - x||_1 + ||G_Y(G_X(y)) - y||_1 \quad (2)$$

In our experiments, we noticed that GAN loss and cycle-consistency loss in CycleGAN cannot perform well during the translation between the two different modalities, though they worked great when two domains from the same modality were involved, which was also mentioned by the authors in [1] when they tried to perform a translation between RGB and LWIR. We also noticed that CycleGAN cannot translate the details of the objects from visible to MWIR, and the shapes of objects were often distorted. To tackle this challenge, we proposed to use the perceptual loss, LPIPS [26], by following the idea in [30]. Typically, researchers use the LPIPS loss to evaluate the perceptual similarity between two images. To calculate LPIPS, we need to collect the activations of two image patches from a pre-defined network, and LPIPS is computed based on the activations. A larger LPIPS stands for more differences that exist between the two images. Previous research shows that LPIPS matches human perception better than other metrics [26]. In our experiments, we used activations from the pre-trained AlexNet [31] model to calculate LPIPS. We calculated LPIPS between an image and its recovered version through the two translations to model geometric changes between the two images. Perceptual loss helps preserve the shape of the objects between visible images and MWIR images:

$$L_{perc} = \text{LPIPS}(G_X(G_Y(x)), x) + \text{LPIPS}(G_Y(G_X(y)), y) \quad (3)$$

In summary, the overall loss function of our model is

$$MWIRGAN_{loss} = L_{GAN} + \alpha L_{cyc} + \beta L_{perc} \quad (4)$$

where α and β are two hyperparameters.

4. Experiment

4.1. Dataset

We used the ATR dataset [32], provided by the Defense Systems Information Analysis Center (DSIAC), to validate our method for visible-to-MWIR image translation. In this dataset, there are visible and MWIR videos of different vehicles taken at long distances. We extracted frames from the visible and MWIR videos to train our model. During training, 3600 frames from each domain were used to train our model. We extracted 1875 frames from the visible domain to test the trained model. The frame size of MWIR videos is 640×512 , and the frame size of optical videos is 640×480 . Figure 2 shows two training samples from both domains.

This dataset is complex and challenging [25]. First, due to the very long distances, the sizes of the targets in the images are very small. Second, the orientation of the targets in the image changes drastically, and illuminations are different among the videos. Third, the cameras are not fixed and move during video capturing, making the video not stable. Finally, there are many differences between visible videos and MWIR videos including (1) the wavelength range of MWIR imagers is between 3 and 5 microns, but the range is between 0.4 and 0.8 microns for optical imagers; (2) MWIR cameras use heat radiation from objects for imaging and do not need any external illumination, while optical cameras need external illumination; and (3) MWIR images have many advantages including tolerance of

smoke, dust, heat turbulence, and fog. Overall, MWIR is highly effective during the night for surveillance purposes, while optical cameras are preferred in the daytime.

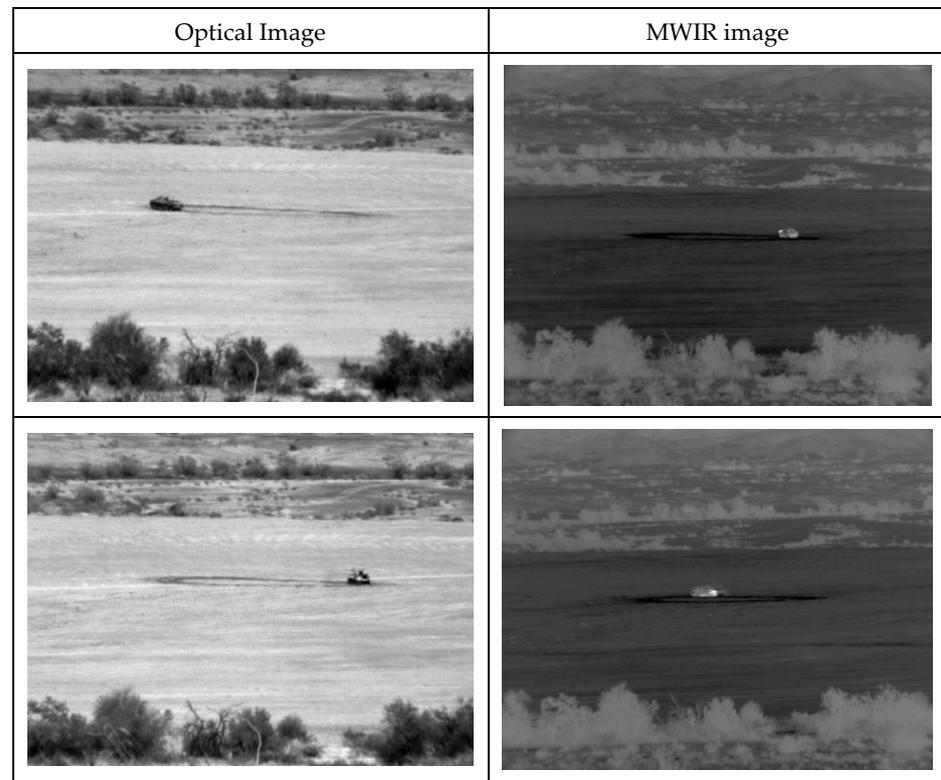


Figure 2. Unpaired training samples from optical and MWIR image domains.

4.2. Training Parameters

We randomly cropped 256×256 patches from the training images of both domains for training and used full images for testing. We noticed that using cropped patches for training achieved better results than using full images for training. We set batch sizes as 1 and 50 for the number of image buffers as suggested in [4]. We chose the hyperparameters as $\alpha = 6$ and $\beta = 1$ in Equation (4) after trial and error. We used the Adam optimizer [33] during training for optimization and trained our model for 30 epochs with a learning rate of 0.0001 which decayed to 0 after 8 epochs of training. The Pytorch framework was utilized to develop our model, and the model was trained and tested on an NVIDIA GPU with 12G memory.

5. Experiment Results

5.1. Qualitative Comparison

Figure 3 shows the MWIR images generated by the proposed model, CycleGAN [4] and AttentionGAN [25]. We can see that the MWIR images generated by AttentionGAN contain visible stripe artifacts. Overall, the quality of the images is not as good when compared to the real MWIR images. Moreover, we noticed several drawbacks of the CycleGAN model. MWIR images generated by CycleGAN lost the vehicle objects in the translated images, and the fine details of the objects were not preserved during this translation. On the other hand, our method performed better than the two competing models. The translated MWIR images preserved fine details of the objects, and the engine parts of the vehicles were appropriately highlighted due to high temperatures. Our proposed model resolved all the issues related to the MWIR images generated by AttentionGAN and CycleGAN.

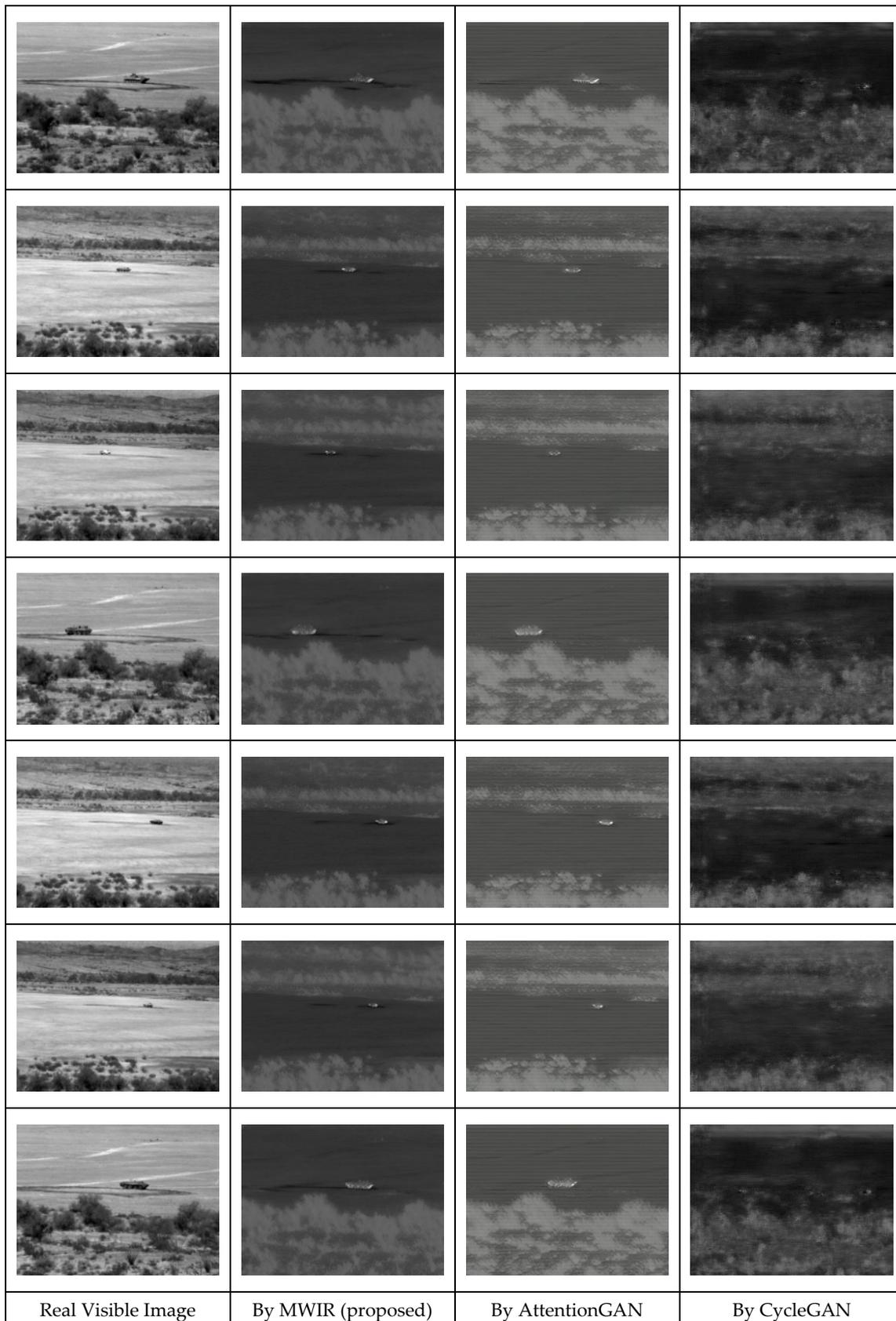


Figure 3. MWIR images generated from visible images by different models. The proposed model generated best-quality MWIR images. MWIR images generated by the AttentionGAN have visible stripe artifacts. The vehicle objects are not visible in the MWIR images generated by the CycleGAN.

5.2. Quantitative Comparison

We utilized Fréchet Inception Distance (FID) [34,35] and Kernel Inception Distance (KID) [36] to quantitatively evaluate model performances in image translation. The FID and KID scores have been widely used to assess the quality of the generated images with respect to real images. For both FID and KID, smaller values denote a higher quality of synthetic images with respect to real images. Table 1 shows the FID and KID scores for the MWIR images generated by different models. It is observed that our proposed model (FID: 1.72, KID: 22.82) achieved the best results as compared to the CycleGAN model (FID: 3.28, KID: 49.69) and the AttentionGAN model (FID: 2.46, KID: 36.59) for visible-to-MWIR image translation, which is consistent with the qualitative observations in Figure 3. The FID and KID scores of the proposed model are the lowest among all the models considered.

Table 1. Quantitative comparison among different methods. Bold numbers indicate best-performing methods.

Model	FID Score	KID Score
CycleGAN	3.28	49.69
AttentionGAN	2.46	36.59
MWIRGAN (proposed)	1.72	22.82

6. Ablation Study

We conducted an ablation study to investigate the contribution of each of the loss function components in the proposed model. Figure 4 shows the results of the ablation study. First, we removed the perceptual loss from the overall loss function and kept the cycle-consistency loss and GAN loss. With this configuration, we noticed that the model experienced mode collapsed due to the absence of perceptual loss and could not translate objects in the visible domain to the MWIR domain. For visible images captured at the 1000 m distance, the model could not highlight vehicles properly (such as the engine part), and for the 2000 m distance images, the model totally omitted vehicles that were missed in the translated MWIR images. Second, we removed the cycle-consistency loss from the total loss and performed image translation with the modified model. This configuration achieved the worst result as it could not translate the visible image into the MWIR domain, keeping all objects intact. Table 2 shows the quantitative results of the ablation study. We can see that both loss function components are important, and the proposed model did not function properly without either of them.

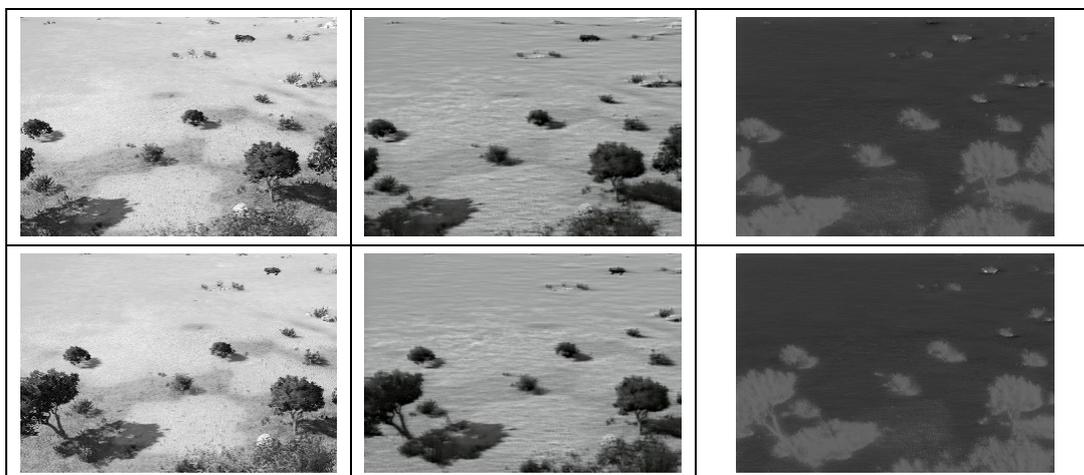


Figure 4. Cont.

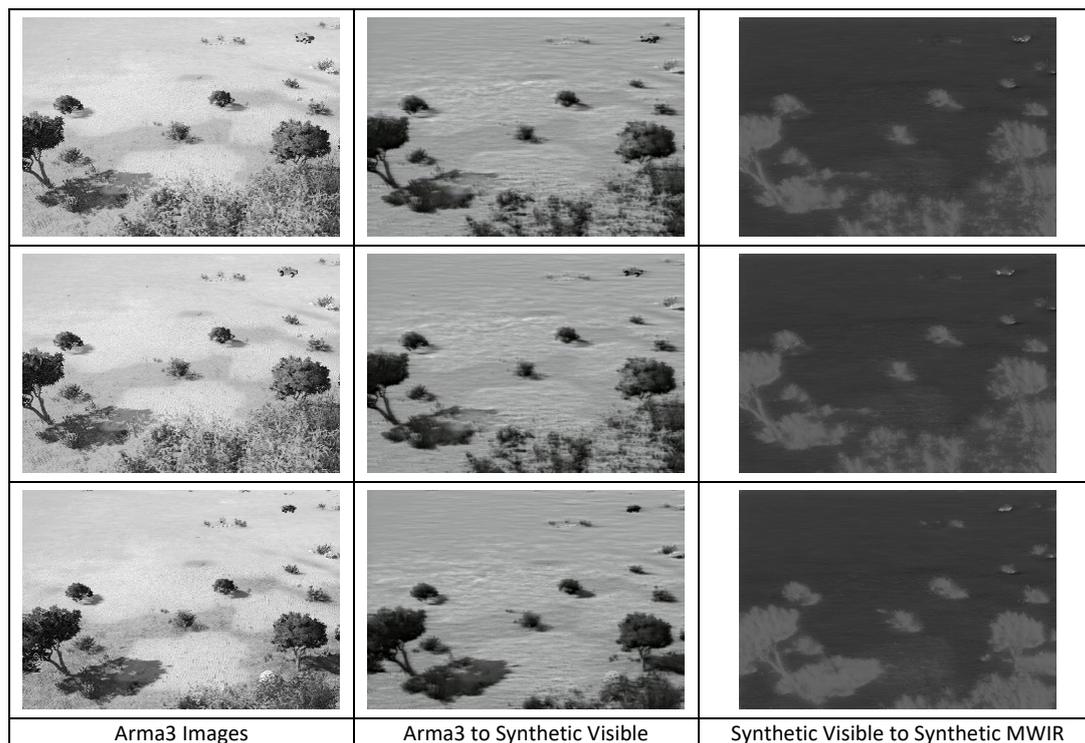


Figure 4. Ablation study results. The results show that both perceptual loss and cycle-consistency loss are essential for obtaining good-quality MWIR images.

Table 2. Qualitative comparison of our ablation study. Bold numbers indicate best-performing methods.

Model	FID Score	KID Score
MWIRGAN (proposed)	1.72	22.82
MWIRGAN w/o Perceptual loss	2.23	32.68
MWIRGAN w/o cycle-consistency loss	4.08	64.73

7. Converting Game-Engine-Generated Images to MWIR Images

As a case study, we applied the proposed model to convert synthetic visible images from Arma 3 [37], generated by the Real Virtuality 4 engine [37], to MWIR images. This was a two-step process. First, we trained the proposed model to convert the game-engine-generated synthetic visible images to real visible images. We utilized 3600 unpaired game-engine-generated images and 3600 real visible images to train the model. Second, we applied our original trained model to convert the converted real visible images to MWIR images. Figure 5 shows the diagram of the two-step method for converting game-engine-generated images to MWIR images. Figure 6 shows the results of the proposed model. It is observed that our proposed technique could generate MWIR images from Arma-3-generated images. In the generated MWIR images, vehicles were correctly highlighted with engine parts being brighter than other objects. The quantitative results of the generated MWIR images from the game engine images are listed in Table 3. We also compared the generated MWIR images with those generated from real visible images. According to the quantitative results, the quality of synthetic MWIR images from Arma 3 (using our model) was better (FID: 2.72, KID: 41.84) than the MWIR image from the real visible image (FID: 3.28, KID: 49.69) generated by CycleGAN. As expected, the quality of the synthetic MWIR images from the Arma 3 images (FID: 2.72, KID: 41.84) was worse than that of the synthetic MWIR images from the real visible images (FID: 1.72, KID: 22.82).

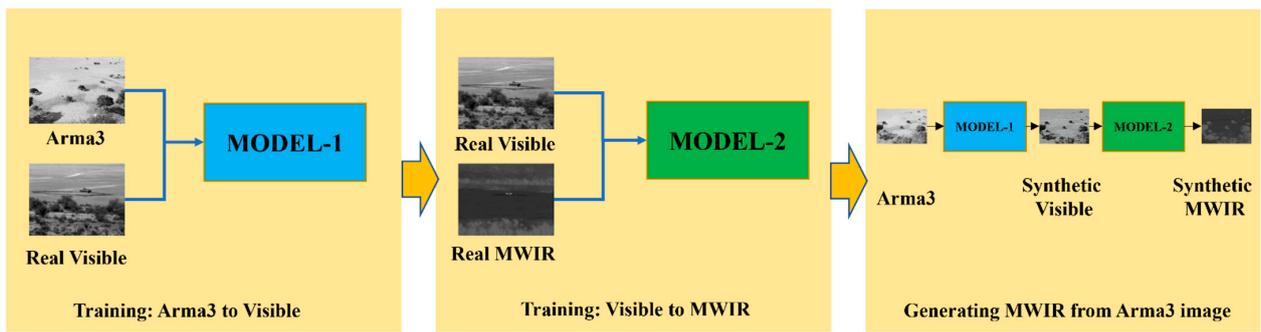


Figure 5. Illustration of the proposed two-step technique for generating MWIR images from game-engine-generated images.

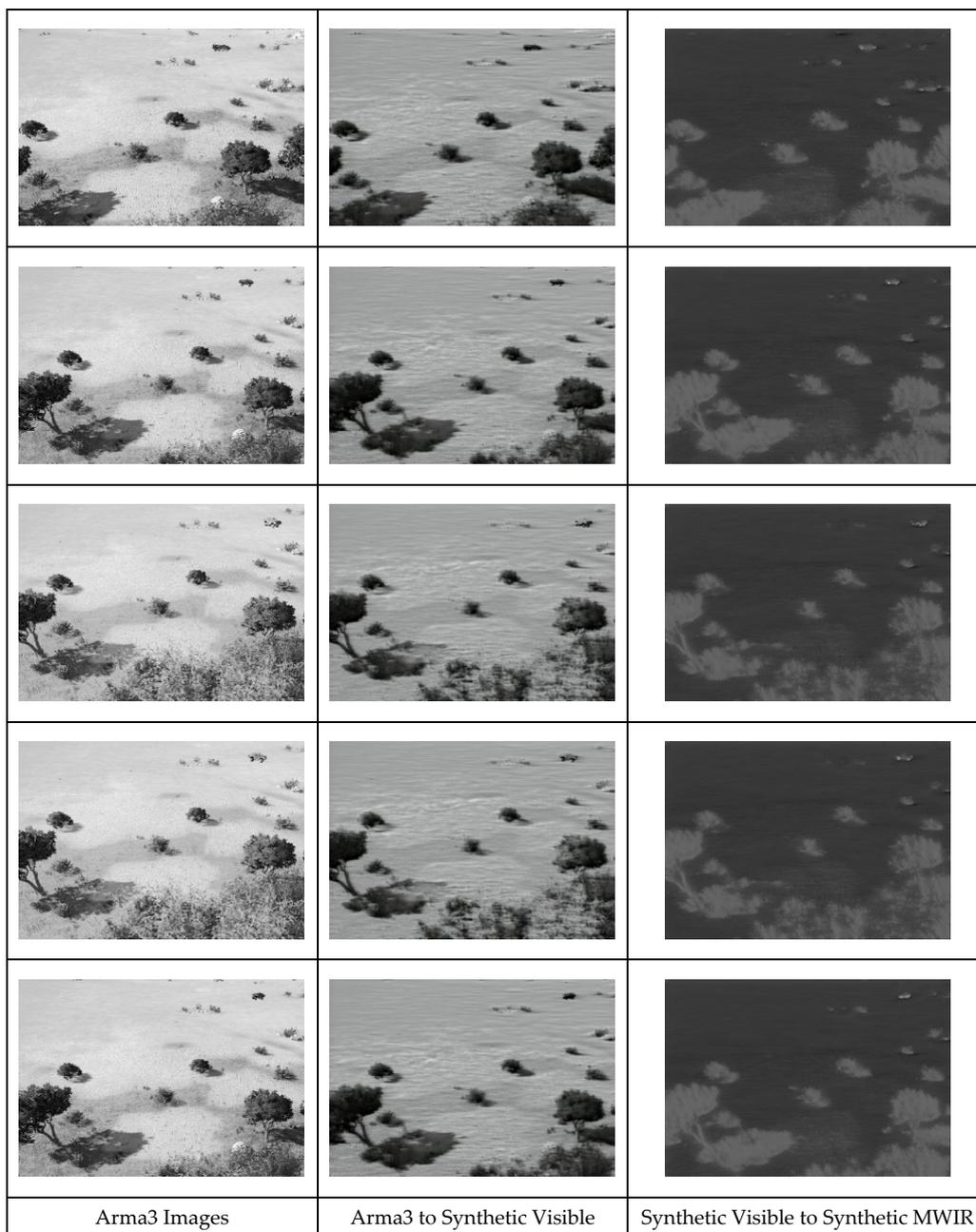


Figure 6. MWIR images generated from game-engine-generated synthetic images using our technique.

Table 3. Quantitative evaluation of the translation from game-engine-generated images to MWIR images.

Model	Task	FID Score	KID Score
Cycle GAN (real visible to MWIR)	MWIR from real visible image	3.28	49.69
MWIRGAN (real visible to MWIR)	MWIR from real visible image	1.72	22.82
MWIRGAN (arma3 to MWIR)	MWIR from real game-engine-generated image	2.72	41.84

8. Conclusions

In this paper, we proposed the MWIRGAN model for visible-to-MWIR image translation. We generated MWIR images from visible images in an unsupervised manner without using any paired datasets. Our proposed method was able to translate visible images into MWIR images successfully by keeping fine details and overall structures in the image. The proposed model outperformed several competing state-of-the-art methods. In addition, we customized the proposed model to convert game-engine-generated images to MWIR images in a two-step process. This work will mitigate the data scarcity challenge in the MWIR domain and will advance the research in this area.

Author Contributions: Conceptualization, C.K. and J.L.; methodology, M.S.U., C.K., and J.L.; software, M.S.U.; formal analysis, M.S.U., C.K., and J.L.; validation, M.S.U., C.K., and J.L.; investigation: M.S.U., C.K., and J.L.; resources, C.K. and J.L.; data curation, C.K. and M.S.U.; writing—original draft preparation, M.S.U., C.K., and J.L.; writing—review and editing, M.S.U. and J.L.; supervision, C.K. and J.L.; project administration, C.K. and J.L. All authors have read and agreed to the published version of the manuscript.

Funding: This research received no external funding.

Data Availability Statement: The ATR dataset is available on <https://dsiac.org/databases/atr-algorithm-development-image-database/>, accessed on 1 January 2020.

Conflicts of Interest: The authors declare no conflict of interest.

References

- Abbott, R.; Robertson, N.M.; Martinez del Rincon, J.; Connor, B. Unsupervised Object Detection via LWIR/RGB Translation. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops, Seattle, WA, USA, 14–16 June 2020; pp. 90–91.
- 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), Munich, Germany, 8–14 September 2018.
- 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.
- 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.
- Goodfellow, I.; Pouget-Abadie, J.; Mirza, M.; Xu, B.; Warde-Farley, D.; Ozair, S.; Courville, A.; Bengio, Y. Generative adversarial nets. In *Advances in Neural Information Processing Systems*; Ghahramani, Z., Welling, M., Cortes, C., Lawrence, N.D., Weinberger, K.Q., Eds.; Curran Associates, Inc.: Red Hook, NY, USA, 2014; Volume 27, pp. 2672–2680.
- Liu, M.-Y.; Tuzel, O. Coupled generative adversarial networks. *arXiv* **2016**, arXiv:1606.07536.
- Brock, A.; Donahue, J.; Simonyan, K. Large scale GAN training for high fidelity natural image synthesis. *arXiv* **2018**, arXiv:1809.11096.
- Kim, T.; Cha, M.; Kim, H.; Lee, J.K.; Kim, J. Learning to discover cross-domain relations with generative adversarial networks. In Proceedings of the International Conference on Machine Learning, Sydney, Australia, 6–11 August 2017.
- Hoffman, J.; Tzeng, E.; Park, T.; Zhu, J.Y.; Isola, P.; Saenko, K.; Efros, A.; Darrell, T. Cycada: Cycle-consistent adversarial domain adaptation. In *International Conference on Machine Learning*; PMLR: Grenoble, France, 2018; pp. 1989–1998.
- Zhang, R.; Isola, P.; Efros, A.A. October. Colorful image colorization. In *European Conference on Computer Vision*; Springer: Cham, Switzerland, 2016; pp. 649–666.
- Wu, B.; Duan, H.; Liu, Z.; Sun, G. SRPGAN: Perceptual generative adversarial network for single image super resolution. *arXiv* **2017**, arXiv:1712.05927.
- 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. [[CrossRef](#)] [[PubMed](#)]

13. Han, J.; Shoeiby, M.; Petersson, L.; Armin, M.A. Dual contrastive learning for unsupervised image-to-image translation. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, Online, 19–25 June 2021; pp. 746–755.
14. Sun, T.; Jung, C.; Fu, Q.; Han, Q. NIR to RGB domain translation using asymmetric cycle generative adversarial networks. *IEEE Access* **2019**, *7*, 112459–112469. [[CrossRef](#)]
15. Perera, P.; Abavisani, M.; Patel, V.M. In2i: Unsupervised multi-image-to-image translation using generative adversarial networks. In Proceedings of the 24th International Conference on Pattern Recognition (ICPR), Beijing, China, 20–24 August 2018.
16. Uddin, M.S.; Li, J. Generative adversarial networks for visible to infrared video conversion. In *Recent Advances in Image Restoration with Applications to Real World Problems*; Kwan, C., Ed.; IntechOpen: London, UK, 2020; pp. 285–289. [[CrossRef](#)]
17. Mehri, A.; Sappa, A.D. Colorizing near infrared images through a cyclic adversarial approach of unpaired samples. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops (CVPRW), Long Beach, CA, USA, 16–17 June 2019.
18. Suarez, P.; Sappa, A.; Vintimilla, B. Infrared image colorization based on a triplet DCGAN architecture. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops (CVPRW), Honolulu, HI, USA, 21–26 July 2017; pp. 212–217.
19. Liu, S.; John, V.; Blasch, E.; Liu, Z.; Huang, Y. IR2VI: Enhanced night environmental perception by unsupervised thermal image translation. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops (CVPRW), Salt Lake City, UT, USA, 18–22 June 2018.
20. Ahlberg, J.; Felsberg, M. Generating visible spectrum images from thermal infrared. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops (CVPRW), Salt Lake City, UT, USA, 18–22 June 2018; pp. 1224–122409.
21. Mizginov, V.A.; Danilov, S.Y. "Synthetic Thermal Background and Object Texture Generation Using Geometric Information and GAN. *Int. Arch. Photogramm. Remote Sens. Spat. Inf. Sci.* **2019**, *XLII-2/W12*, 149–154. [[CrossRef](#)]
22. Kniaz, V.V.; Mizginov, V.A. Thermal Texture Generation and 3D Model Reconstruction Using SFM and GAN. *Int. Arch. Photogramm. Remote Sens. Spat. Inf. Sci.* **2018**, *XLII-2*, 519–524. [[CrossRef](#)]
23. Yuan, X.; Tian, J.; Reiertz, P. Generating artificial near infrared spectral band from rgb image using conditional generative adversarial network. *ISPRS Ann. Photogramm. Remote Sens. Spat. Inf. Sci.* **2020**, *V-3-2020*, 279–285. [[CrossRef](#)]
24. Yun, K.; Yu, K.; Osborne, J.; Eldin, S.; Nguyen, L.; Huyen, A.; Lu, T. Improved visible to IR image transformation using synthetic data augmentation with cycle-consistent adversarial networks. In *Pattern Recognition and Tracking XXX*; International Society for Optics and Photonics: Bellingham, WA, USA, 2019; Volume 10995, p. 1099502.
25. Uddin, M.S.; Hoque, R.; Islam, K.A.; Kwan, C.; Gribben, D.; Li, J. Converting optical videos to infrared videos using attention gan and its impact on target detection and classification performance. *Remote Sens.* **2021**, *13*, 3257. [[CrossRef](#)]
26. Zhang, R.; Isola, P.; Efros, A.A.; Shechtman, E.; Wang, O. The unreasonable effectiveness of deep features as a perceptual metric. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Salt Lake City, UT, USA, 18–23 June 2018; pp. 586–595.
27. 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.
28. Tang, H.; Xu, D.; Sebe, N.; Yan, Y. Attention-guided generative adversarial networks for unsupervised image-to-image translation. In Proceedings of the 2019 International Joint Conference on Neural Networks (IJCNN), Budapest, Hungary, 14–19 July 2019; IEEE: Piscataway, NJ, USA, 2019; pp. 1–8.
29. Radford, A.; Metz, L.; Chintala, S. Unsupervised Representation Learning with Deep Convolutional Generative Adversarial Networks. *arXiv* **2015**, arXiv:1511.06434.
30. He, J.; Uddin, M.S.; Canan, M.; Sousa-Poza, A.; Kovacic, S.; Li, J. Ship Deck Segmentation Using an Attention Based Generative Adversarial Network. In Proceedings of the 2022 IEEE 13th Annual Ubiquitous Computing, Electronics & Mobile Communication Conference (UEMCON), New York, NY, USA, 26–29 October 2022; IEEE: Piscataway, NJ, USA, 2022; pp. 367–372.
31. Krizhevsky, A.; Sutskever, I.; Hinton, G.E. Imagenet classification with deep convolutional neural networks. *Commun. ACM* **2017**, *60*, 84–90. [[CrossRef](#)]
32. ATR Dataset. Available online: <https://www.dsiac.org/resources/available-databases/atr-algorithm-development-image-database/> (accessed on 1 January 2020).
33. Kingma, D.; Ba, J. Adam: A method for stochastic optimization. In Proceedings of the ICLR, Banff, AB, Canada, 14–16 April 2014.
34. Heusel, M.; Ramsauer, H.; Unterthiner, T.; Nessler, B.; Hochreiter, S. Gans trained by a two time-scale update rule converge to a local nash equilibrium. In *Advances in Neural Information Processing Systems*; MIT Press: Cambridge, MA, USA, 2017; pp. 6626–6637.
35. Fréchet, M. Sur la distance de deux lois de probabilité. *Comptes Rendus Hebd. Seances L Acad. Des Sci.* **1957**, *244*, 689–692.
36. Bińkowski, M.; Sutherland, D.J.; Arbel, M.; Gretton, A. Demystifying mmd gans. *arXiv* **2018**, arXiv:1801.01401.
37. Arma3. Available online: <https://arma3.com/features/engine> (accessed on 1 January 2020).

Disclaimer/Publisher's Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.