Infrared Image Generation From RGB Images Using CycleGAN

Selma Güzel
Computer Engineering
Yıldız Technical University
İstanbul, Turkey
selma.guzel@std.yildiz.edu.tr

Sırma Yavuz
Computer Engineering
Yıldız Technical University
İstanbul, Turkey
smyavuz@yildiz.edu.tr

2022 International Conference on INnovations in Intelligent SysTems and Applications (INISTA) | 978-1-6654-9810-4/22/$31.00 ©2022 IEEE | DOI: 10.1109/INISTA55318.2022.9894231

***Abstract*—Thermal imaging is more robust than optical imaging against the illumination related issues. Therefore, it is preferred or utilized with RGB data in some of the essential problems such as surveillance, environmental monitoring and so on. Deep learning has been used in various fields including the problems in the scope of thermal imaging and proved its ability to solve lots of problems. However, due to the requirement of large datasets in deep learning and the lack of public thermal data because of the constrains of thermal imaging, deep learning can not be used much in thermal imaging. In this study, we mainly investigate whether using CycleGAN on paired images rather than impaired ones increases the success rate and the effect of our electromagnetic spectrum based normalization approach. Evaluations on public data sets show that our approach has potential to increase the success rate of CycleGAN, but further study is required.**

***Keywords*—IR image generation, thermal imaging, CycleGAN, image generation using CycleGAN.**

I. Introduction

Thermal imaging is an essential task for various problems due to some of its characteristics. Different from optical images, thermal infrared (TIR) images are not sensitive to illumination changes. Hence, their quality is not affected by the problematic conditions according to optical imaging such as darkness, fogginess, overexposure, and underexposure etc. So, they are more useful in certain areas for example, military, surveillance, defense, medicine, electricity, and environmental monitoring.

But, color-to-thermal image translation is a hard task because of some reasons. Firstly, the wavelengths of RGB and IR images are different and it is hard to diagnose the correlation between them. RGB images have three channels containing color information of visible light, while IR images have one channel containing information of invisible light. [11] Secondly, the task of thermal imaging is expensive due to the required equipments and specific constraints. As a result, there are not enough public TIR data which can be used in various studies.

The problems which thermal imaging tries to handle can be solved by deep learning algorithms more accurately considering its known success in nearly every problems including image recognition and object detection. But it is also known that deep learning necessitates a lot of data.

Therefore, generating synthetic thermal infrared datasets from optical images is an important topic [4]. Generation of synthetic datasets using 3D models of various scenes is a time-consuming method that requires long computation time and is not very realistic [5]. On the other hand, Generative Adversarial Networks (GANs) are very efficient to generate images even if the modalities of the source and the target domain are different.

So, in this study, we aim to sythesize IR images from RGB images using CycleGAN on two public datasets to increase the public IR data which can be used in various fields to solve different problems. We investigate if using CycleGAN on paired images after our novel electro magnetic spectrum based normalization increases the success rate significantly. It is not expected to generate an object in an IR image if the object is not seen in RGB data totally. However, despite the wave lenghts of RGB and IR are different, we assume they might be relational due to the fact that they are the parts of a whole electromagnetic spectrum. Therefore, if this probable relation is found by GANs, then RGB images can be used to get their corresponding IR images in an acceptable degree. In addition, even only a part of an object is in the interception of IR and RGB images, the whole object in the IR image can be generated. We tried to find those probabilities in this study.

II. Literature Review

One of the areas in which thermal images are used is person re-identification. However, the heats of objects are changeable according to the environmental conditions. Therefore, in [4], instead of single image, a set of possible person appearances in a thermal image are generated. They used a modified BicycleGAN framework to synthesize multiple color segmentation images for a single color image. In their ThermalGAN framework, in order to find the corresponding original image in the thermal data set, instead of a random noise sample as in BicycleGAN, conditional GANs are conditioned by a temperature vector including the desired background and object temperatures in addition to a single color image. They calculated the average temperature thanks of their manual annotations on the data set they constructed. They called the resulting image as “thermal segmentation”. After predicting temperatures, they predicted the relative local temperature contrasts conditioned by a color image and a thermal segmentation. The sum of a thermal segmentation and temperature contrasts provided the

thermal image. It is reported that their evaluation results are superior than the state-of-the-art [4].

The authors of [4] increased their success on thermal image generation with the other models presented in [6]. In this study, two models are developed for image-to-image translation. In the first model, they used a thermal segmentation image, a semantic segmentation image with the edge of objects and temperature vector as input tensor. In addition, they used modified neural network StyleGAN with ResNet-18 to map an input image to the latent code of the style-based decoder. Moreover, they added skip connections between intermediate feature maps and noise inputs of the generator in order to speed the convergence of the generator. They used modified Wasserstein loss as adversarial loss and L1 loss to increase the similarity in color between the synthesized image and the real image. In the second model, in order to generate class based objects without affecting the rest in the images, the generator is trained layer based and the same number of discriminators assessed the generated data correspondingly. The output of all layers of the generator are combined to form the output image. The basic GAN architecture is used as the model. In this case, an RGB color image is added to the input tensor instead of the feature vector as in the first model to solve the problem of losing separate objects' contrasts due to averaging the brightness over the entire image. Furthermore, semantic segmentation images were divided into several images such that each image shows one class available in color and thermal images to predict thermal contrasts of small objects more precisely and to avoid brightness averaging problem stated above. The evaluation of the methods on the modified version of the ThermalWorld dataset using Frechet Inception Distance (FID) shows that the synthesized thermal images resemble to the ground truth model in both thermal emission and geometrical shape. It is shown that the proposed methods are better than pix2pix and the previous method of the authors.

Another study [11] in person re-identification focuses on disentangling modality-invariant and modality-specific features in images. They accept structure based features such as pose, gender etc. as model-invariant and illumination based methods such as color, texture etc. as model-specific features. Because the goal is generating IR images from RGB images, they combine the RGB image's modality-invariant features with the IR image's modality-specific features. By doing this, it is easier to achieve instance-level alignment between RGB-infrared unpaired-images and generate cross-modality paired-images. There are two modules in their method: a generation module to generate cross-modality paired-images and a feature alignment module to learn both set-level and instance-level aligned features. The generation module consists of three encoders and two decoders. The encoders extract modality-invariant and modality-specific features from RGB images. Then, the RGB decoders take a modality-invariant feature from an RGB image and a modality-specific feature from an IR image as input. By decoding from the across-feature, cross-modality paired-images are generated. In the feature alignment module, there exist an encoder which can translate images from distinct modalities into a common feature space to reduce modality-gap between sets and an encoder which improves features to minimize the difference between instance-level modality-gap. Lastly, generation and feature alignment modules are jointly trained using the re-id loss. As a result, both modality-aligned and identity-discriminative features are learned. It is reported that this is the first work to generate cross-modality paired-images for the RGB-IR Re-ID task. Their experiments performed on SYSU-MM01 and RegDB data sets are measured their success in terms of Cumulative Matching Characteristic (CMC) and mean average precision (mAP) and the results show the proposed method's achievement against state-of-the-art methods.

In [9], due to little texture and color information in TIR images, a sparse U-net architecture based modified pix2pix generative model selecting only low and high level information for symmetric connections is proposed to produce synthetic TIR images from optical RGB data. This does not only preserve enough features but also improve performance because of decrease in the number of network's parameters. In order to improve appearance and structural similarity, intensity and gradient losses are added within the objective function. The experiments on public datasets were evaluated using SSIM and PSNR metrics and it is presented that the method is more successful than the other pix2pix based methods. However, it is also stated that some of the synthetic TIR images at night are different from the real infrared images and they encourage further study to improve the results.

Another study in which a generic image-to-image translation method based on cGANs is used to generate RGB-IR images and vice versa is [8]. In this work, firstly, the authors try to determine if the trained model can be used without any change using images with similar contrasts but different contents. Secondly, inherently problematic images are evaluated. A dual-modality CVC-14 dataset of co-registered visible and thermal images are used for these purposes. Their main goal is improving resolution of the synthesized images, therefore they concentrated to the capability of the generation methods for training of machine learning models. There are three important approaches in this study which we want to emphasize: First, they encourage the image synthesis both from thermal-visible and vice-versa to diagnose the differences between modalities and we can complete that this is also be used to modality fusion to identify all of the seen and unseen features of the objects. Secondly, they remind that despite insufficient results while using different source and target domains in datasets, transfer learning can be used at least as a meaningful starting point. And lastly, they underline the requirement of one-to-many mapping between different modalities because of the fact that an object can have different temperature even in the same environment in different times.

In [5], for thermal image synthesis, 3D modeling is also used in addition to the pix2pix as GAN model. In this study, background and objects' textures are synthesized using GAN whereas semantic and geometric information about objects are generated using 3D modeling. They think that segmentation of objects' thermal zones which are sometimes unchanged in various weather and shooting conditions simplifies predicting the objects' positions and improves visual quality of the generated images. In order to synthesize both thermal zones of objects and depth maps, realistic three-dimensional models were constructed. They measured the success of the results using LPIPS metric and they say that their method produces realistic synthetic images. However, the pre-processing step

seems very time-consuming to us.

III. Problem and Method

CycleGAN is a type of generative adverserial networks (GANs) architecture in which an image in one domain can be converted to the corresponding image in another domain based on reconstruction loss without need of paired data sets [12]. In this paper, thermal image generation from RGB image using CycleGAN on paired after electro magnetic spectrum based normalization is proposed.

We propose two main approaches in this study. The first one is using CycleGAN with paired images instaed of unpaired images. Although reconstruction loss in CycleGAN has capability to generate an RGB image to a thermal correspondence without paired image data sets, according to the literature the results are not sufficient enough. Paired images based architectures similar to pix2pix gives better results in general. Therefore, we propose to use CycleGAN model with paired images to increase the success rate of CycleGAN. We expect that because CycleGAN has ability to generate image even using unpaired data sets to some degree, if we use the model on paired images, it can surpass other only paired images based models.

The second novel approach is taking the positions of RGB and IR wavelengths in the electromagnetic spectrum into account to normalize both RGB and IR images before training. Despite infrared images are constructed according to the heats of the objects different from RGB, both of them are parts of the same whole: electro magnetic spectrum. Hence, if we normalize both the thermal and the RGB images according to their locations on electro magnetic spectrum, the images can be related more logically and conversion might make more sense. Figure 1 shows the electro magnetic spectrum emphasizing visible and infrared spectrums [3].

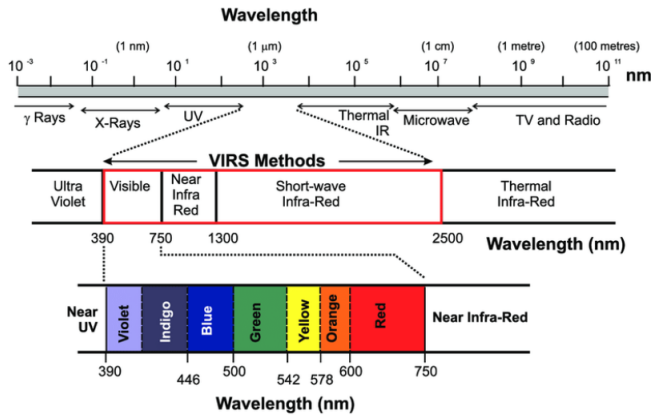

Figure 1: Electromagnetic Spectrum.

Equation (1) shows the normalization of RGB image whose intensity range is [0-255] to range [380-750] for RGB and to range [1000-14000] for IR. Then it is normalized to [-1,1].Equation (2) shows the normalization from range [-1,1] to range [0-255] inversely according to Equation (1).

$$\begin{aligned} x &= ((max - min)/255) * x + min, \\ x &= (x - min)/(max - min), \\ x &= 2 * x - 1 \end{aligned} \tag{1}$$

$$\begin{aligned} x &= (x + 1)/2, \\ x &= x * (max - min) + min, \\ x &= ((x - min) * 255)/(max - min). \end{aligned} \tag{2}$$

IV. Data

We use two data sets for training and testing. The data sets include both indoor and outdoor images.The datasets and their main properties are listed below:

- OSU [2]:We used 6848 training, 864 validation and 864 test rgb-thermal image pairs.
- AAU VAP PAIRS [7]: We used 5568 training, 704 validation and 704 test rgb-thermal image pairs.

V. Implementation Results and Evaluations

We used and modified one of the implementations [1] of the CycleGAN for our studies. The codes are implemented on Google Colab Pro platform with Python version 3.7.12 and Pytorch version 1.10.0+cu111.

First, we tried different architectures and hyper parameters which are number of generator layers, electro-magnetic spectrum based normalization(EMSN),using Leaky_ReLU or ReLU, number of Resnet blocks in Generator. After human assessment, we ended up with four models:

- Unpaired CycleGAN without EMSN
- Unpaired CycleGAN with EMSN
- Paired CycleGAN without EMSN
- CycleGAN with EMSN

We compared their training results. During training, both training and validation images are selected randomly. Figure 2-5 show the validation results of the models after 9 epochs.

It can be seen from Figure 4 and 5 that some of the image pairs could not been generated accurately using unpaired models. For example, in Figure 4, on first three rows, the generated images on the second column are not correspondences of the real images on the first column. It implies that cycle consistency loss can not handle the reconstruction of corresponding image enough. When it comes to EMSN, it didn't improve the results much as sen in Figure 5.

When we used paired versions all of the image pairs were generated to an acceptable degree as expected (See Figure 4 and 5). However, some corruptions occurred and the images are a bit blurry and noisy. For example,in Figure 2, there is a black mark on the generated image on the first row and the second column similar to the human on the image on the first row and first column. However, even if there is not an human on the first row third column, the black mark is still there as seen in first row fourth column.

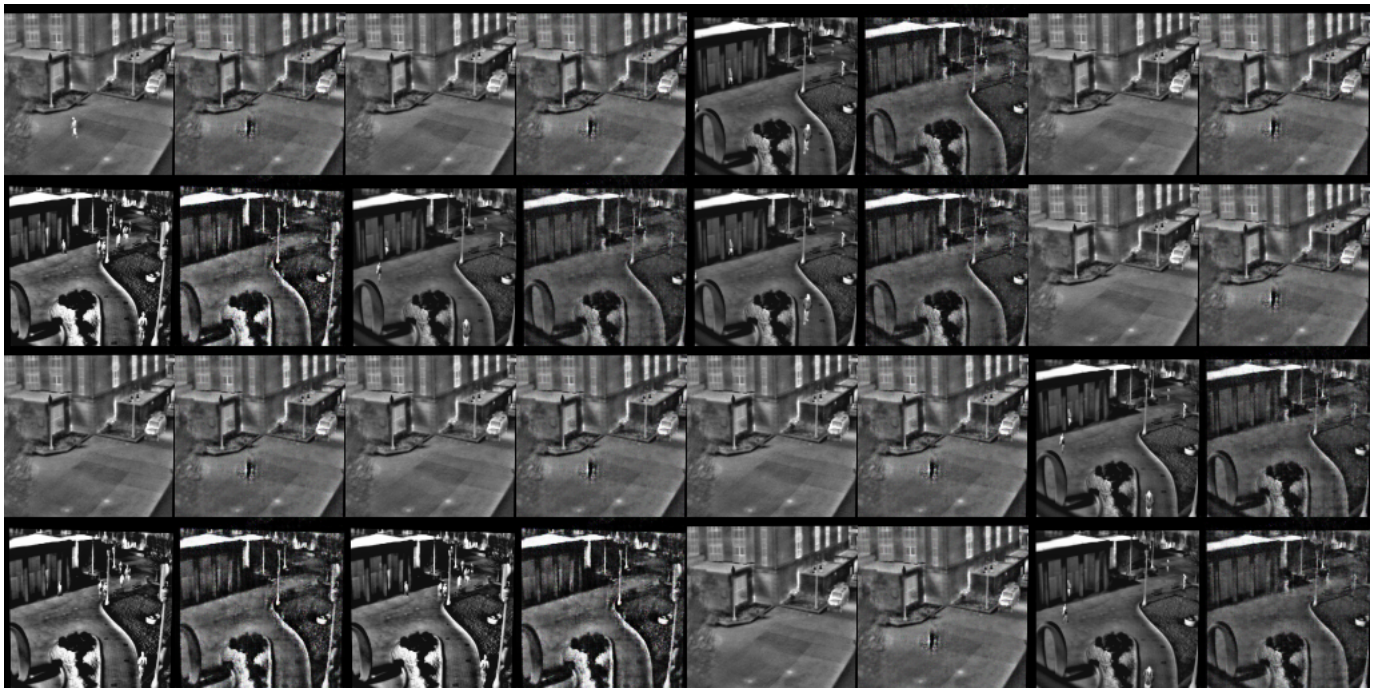

Figure 2: The results of the Paired CycleGAN without EMSN model. From left to right, columns are numerated as 1 to 8. The odd-numbered columns show the real images in the training data set while the even-numbered columns show the generated images by the model which correspond to the real images such that in each even-numbered column, from top to down, every raw corresponds to the same raw in the previous odd-numbered column (first column:second column, third column:fourth column, etc.).

The results of the Paired CycleGAN with EMSN didn't improve the results in some of the images. Furthermore, some of the generated images are more noisy and corrupted.

Which epoch number is the best changes due to randomness for all models. Therefore we evaluated the outputs and used an acceptable model for testing for the next steps. In addition, due to some of the wrong correspondences from unpaired models, we decided to go on only with paired models. In some of the generated images, the humans can not be generated despite most of the small objects can be generated. Furthermore,some false generations decrease the correctness very much. Those problems need further study.

We used five metrics which are used in GAN comparison [10] to evaluate our results and compare it with one of the implementation of the standard unpaired CycleGAN model (CycleGAN) in the literature [1]. For MSE and low scores are preferable while for PSNR and UQI, high scores show superiority. We get the mean of all the images for each metric. Table 1 shows the results.

Table I: Evaluation of Test Results

Dataset	Model	MSE	RMSE	PSNR	UQI
OSU	Paired without EMSN	473,936	17,983	356,064	0,941
OSU	Paired with EMSN	**214,0970**	**12,280**	**358,892**	**0,971**
OSU	CycleGAN	923,878	23,118	355,594	0,909
AAUVAPPAIRS	Paired without EMSN	621,896	23,983	351,964	0,931
AAUVAPPAIRS	Paired with EMSN	663,763	24,106	352,242	0,901
AAUVAPPAIRS	CycleGAN	**513,256**	**21,578**	**352,926**	**0,940**

For OSU, Paired with EMSN is better from all of the models. On the other hand, for AAU VAP PAIRS, CycleGAN is better despite using unpaired images. It can be caused by false positives seen in some of the images when Paired models are used. Hence, we can not derive a generalized result according to these metrics.

VI. Conclusion

In this study, we investigated the ability of the CycleGAN model for image generation if we use the paired RGB and IR images. In addition, we evaluated the effect of our proposed electromagnetic spectrum based normalization. The results are not so accurate but promising.

Electromagnetic spectrum based normalization approach can be improved considering the wavelengths, the camera metrics, etc. Moreoever, not only RGB images but also some other easy to use inputs considering changing and fixed characteristics of images can also be utilized for improving the results.

References

[1] [Online]. Available: https://github.com/udacity/deep-learning-v2-pytorch/tree/master/cycle-gaN

[2] J. W. Davis and V. Sharma, "Background-subtraction using contour-based fusion of thermal and visible imagery," *Computer Vision and Image Understanding*, vol. 106, no. 2, pp. 162–182, 2007, special issue on Advances in Vision Algorithms and Systems beyond the Visible Spectrum. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S1077314206001834

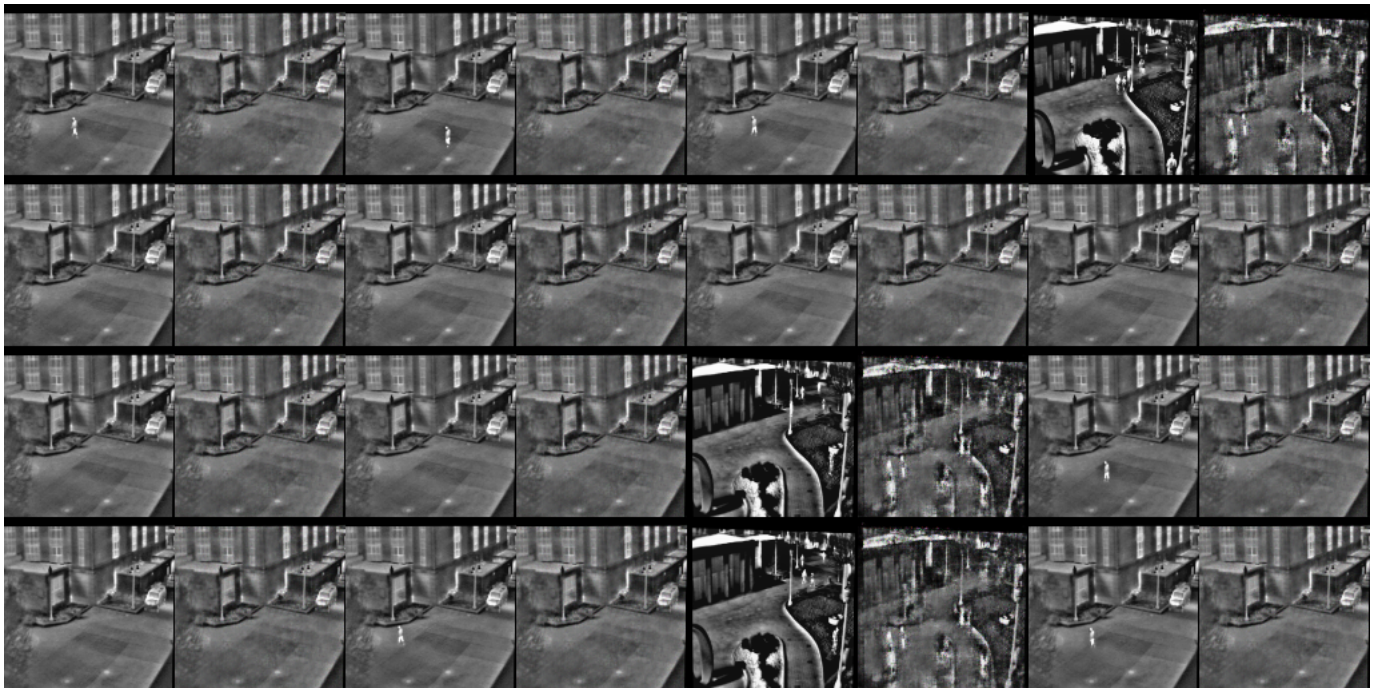

Figure 3: The results of the Paired CycleGAN with EMSN model. From left to right, columns are numerated as 1 to 8. The odd-numbered columns show the real images in the training data set while the even-numbered columns show the generated images by the model which correspond to the real images such that in each even-numbered column, from top to down, every raw corresponds to the same raw in the previous odd-numbered column (first column:second column, third column:fourth column, etc.).

[3] A. Kerr, H. Rafuse, G. Sparkes, J. Hinchey, and H. Sandeman, "Visible/infrared spectroscopy (virs) as a research tool in economic geology; background and pilot studies from new foundland and labrador," pp. 145–166, 03 2011.

[4] V. Kniaz, V. Knyaz, J. Hladuvka, W. Kropatsch, and V. Mizginov, "Thermalgan: Multimodal color-to-thermal image translation for person re-identification in multispectral dataset: Munich, germany, september 8-14, 2018, proceedings, part vi," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, pp. 606–624, 01 2019.

[5] V. Mizginov and S. Danilov, "Synthetic thermal background and object texture generation using geometric information and gan," *ISPRS - International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, vol. XLII-2/W12, pp. 149–154, 05 2019.

[6] V. Mizginov, V. Kniaz, and N. Fomin, "A method for synthesizing thermal images using gan multi-layered approach," *The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, vol. XLIV-2/W1-2021, pp. 155–162, 04 2021.

[7] C. Palmero, A. Clapés, C. Holmberg Bahnsen, A. Møgelmose, T. Moeslund, and S. Escalera, "Multi-modal rgb–depth–thermal human body segmentation," *International Journal of Computer Vision*, vol. 118, 06 2016.

[8] D. Panchard, F. Marelli, E. De Moura Presa, P. Wellig, and M. Liebling, "Perspectives and limitations of visible-thermal image pair synthesis via generative adversarial networks," in *Society of Photo-Optical Instrumentation Engineers (SPIE) Conference Series*, ser. Society of Photo-Optical Instrumentation Engineers (SPIE) Conference Series, vol. 11865, Sep. 2021, p. 1186509.

[9] X. Qian, M. Zhang, and F. Zhang, "Sparse gans for thermal infrared image generation from optical image," *IEEE Access*, vol. 8, pp. 1–1, 01 2020.

[10] P. Raval, 2021. [Online]. Available: https://towardsdatascience.com/measuring-similarity-in-two-images-using-python-b72233eb53c6

[11] G.-A. Wang, T. Zhang, Y. Yang, J. Cheng, J. Chang, X. Liang, and Z.-G. Hou, "Cross-modality paired-images generation for rgb-infrared person re-identification," *Proceedings of the AAAI Conference on Artificial Intelligence*, vol. 34, pp. 12 144–12 151, 04 2020.

[12] J.-Y. Zhu, T. Park, P. Isola, and A. A. Efros, "Unpaired image-to-image translation using cycle-consistent adversarial networks," 2017. [Online]. Available: https://arxiv.org/abs/1703.10593

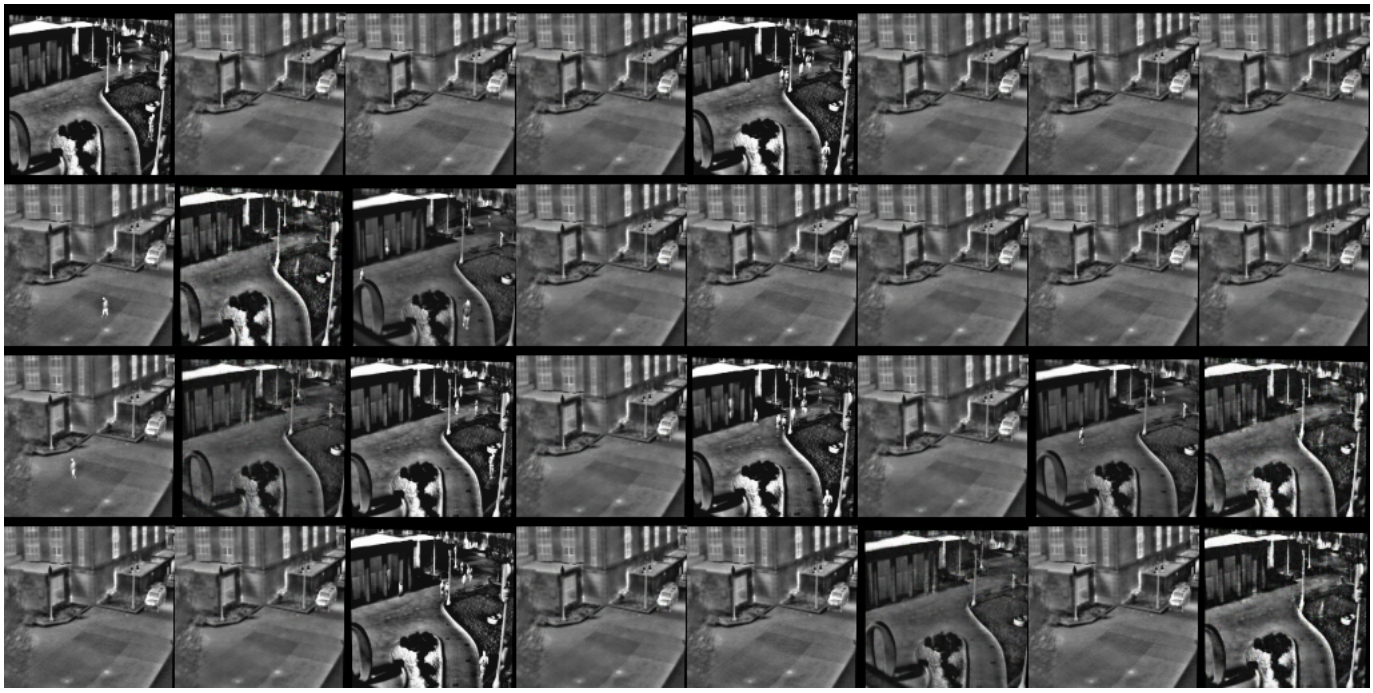

Figure 4: The results of the Unpaired CycleGAN without EMSN model. From left to right, columns are numerated as 1 to 8. The odd-numbered columns show the real images in the training data set while the even-numbered columns show the generated images by the model which correspond to the real images such that in each even-numbered column, from top to down, every raw corresponds to the same raw in the previous odd-numbered column (first column:second column, third column:fourth column, etc.).

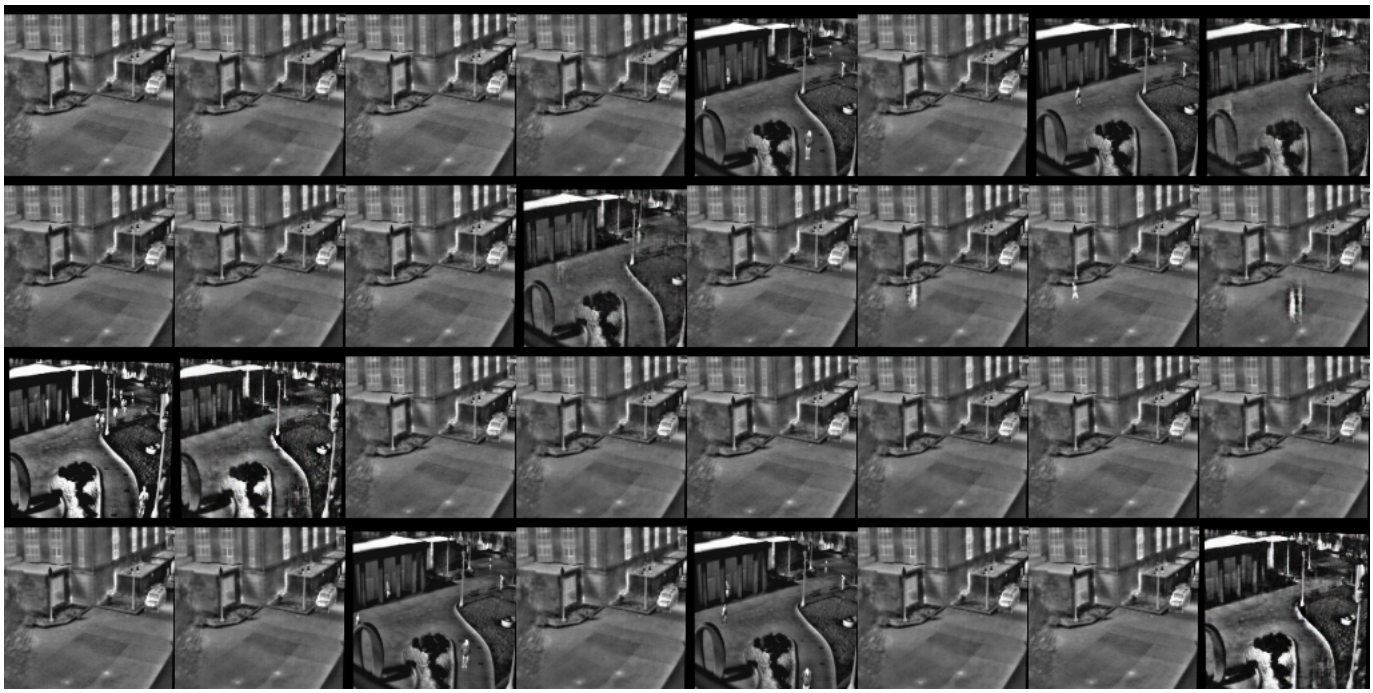

Figure 5: The results of the Unpaired CycleGAN with EMSN model. From left to right, columns are numerated as 1 to 8. The odd-numbered columns show the real images in the training data set while the even-numbered columns show the generated images by the model which correspond to the real images such that in each even-numbered column, from top to down, every raw corresponds to the same raw in the previous odd-numbered column (first column:second column, third column:fourth column, etc.).