GU J Sci, Part C, 10(2): 286-299 (2022)

Gazi University

Journal of Science

PART C: DESIGN AND TECHNOLOGY

http://dergipark.gov.tr/gujsc

Generative Adversarial Network for Generating Synthetic Infrared Image from Visible Image

Utku ULUSOY[1*] Koray YILMAZ[1] Gülay ÖZŞAHİN[2]

[1]*TÜBİTAK BİLGEM Advanced Technologies Research Institute (İLTAREN), Ankara, Turkey*

[2]*TÜBİTAK BİLGEM Information Technologies Institute (BTE), Ankara, Turkey*

Article Info

Research article
Received:18/10/2021
Revision:14/12//2021
Accepted:13/04/2022

Keywords

Generative Adverserial Network
Pix2Pix
Machine Learning
Normalization
Infrared Image
Image to Image Translation

Abstract

One of the most important discoveries in the field of deep learning in recent years is the Generative Adversarial Networks (GAN). It offers great convenience and flexibility for image-to-image conversion processes. This study aims to obtain thermal images from visible band colour images by using Pix2Pix Network, which is a Conditionally Generative Adversarial Network (cGAN). For this purpose, a data set has been prepared by taking facial images at different angles in the visible and infrared bands. By applying image processing methods on this created dataset, pixel-by-pixel matching process was performed. Synthetic thermal face images were obtained thanks to this learning network fed with facial images consisting of visible and long wavelength infrared image (LWIR) pairs. In the generator and discriminator deep networks of the Pix2Pix GAN, Batch Normalization and Instance Normalization methods are applied and their effects on the outputs are examined. The same process has also been tested on the Google Maps dataset and thus its effects on different datasets have been demonstrated. Similarity values between synthetic outputs and real images of both studies has been calculated with several image quality metrics. As a result of offered most suitable method, it creates some details more successfully such as reflections, saturations and artifacts etc. These details in the infrared band images cannot be noticed or predicted by a person. Thanks to this study, hopeful results were obtained to produce infrared synthetic image.

1. INTRODUCTION

Artificial intelligence was defined by John McCarthy as the science or engineering of making intelligence machines and especially intelligent computer programs [1]. Artificial intelligence has been affecting many areas of our lives with recent developments. With the new deep learning methods developed recently, big data can be processed and interpreted. Electronic hardwares or computers with high processing power are required to process big data using artificial intelligence techniques. Developments in graphics processing unit (GPUs) technologies have also accelerated the work in this field. Many tasks such as making sense of large-scale data, making inferences, learning and producing unique new data can be performed by artificial intelligence.

One of the most important deep network discoveries in recent years is Generative Adversarial Networks which were developed by Goodfellow et al. in 2014 [2]. Generative Adversarial Networks have two deep networks called generator and discriminator. It produces its outputs by taking advantage of the competition between these networks. A lot of progress has been made in this field after the development of the first version of the Generative Adversarial Networks capable of generating images in 2014. The Pix2Pix Network, which was developed as a type of Generative Adversarial Network, is one of them. Developed as a Conditionally Generative Adversarial Network type, the Pix2Pix network provides much superiority over traditional models in converting between images [3].

*Corresponding author, e-mail: utku.ulusoy@tubitak.gov.tr
DOI:10.29109/gujsc.1011486

Synthetic image generation is the process of producing fake images similar to real data. Synthetic data is used in many areas to imitate data that is impossible to measure. Thanks to synthetic data, the datasets included in the research are enriched and great flexibility is gained to the researches.

This study shows how different parameters in the generator and discriminator code blocks affect the output of the GAN. A dataset has been created from the images taken with visible band and long wavelength infrared cameras. Pix2Pix network, which is a type of cGAN, is trained by making use different parameters of this dataset. Synthetic thermal images obtained from visible band images, not used in training, have been created by this deep learning network. Thanks to this study, the results of different normalization methods in the Pix2Pix network are shown.

Within this article, information about the materials and methods used in the study is presented in Section 2, the detailed description of the research and the experimental results and graphics of the research are shown in Section 3. Topics open to discussion in this context are included in Section 4. The main achievements of the study, the results and inferences that can guide future studies are given in Section 5.

2. MATERIALS AND METHODS

Performing operations on digital images by using the capabilities of computers is called digital image processing [4]. Image processing is used in many areas such as image classification, image features extraction, pattern recognition, image restoration, and image enhancement.

The term artificial intelligence is used to describe systems or machines that can perform human-specific tasks by imitating human intelligence [5]. It finds its place in many areas of our lives, and appears as artificial intelligence chat engines, suggestion systems, smart assistants etc. [6]. Machine learning can be defined as a subset of artificial intelligence. It produces new outputs with the algorithms that learn using training datasets. Deep learning is also a subset of machine learning. In deep learning, the system can be trained both supervised and unsupervised methods. One of the first deep learning network methods that can generate images is named as Generative Adversarial Networks.

In deep learning, our system is designed with multiple layers and it is a machine learning method that produces results with a dataset given to it. Pix2Pix, a type of GAN, was used in this study. Before the use of Pix2Pix Network, pixel-by-pixel mapping was made with digital image processing algorithms in the dataset created in order to realize the efficient operation of the network.

2.1. Generative Adversarial Networks

Since deep learning methods produce better results than classical image processing methods, their use in the field of digital image processing has become widespread. Thanks to the GAN developed by Goodfellow et al. in 2014, deep networks have become capable of producing images [2]. Generative Adversarial Networks have two different deep networks, unlike classical deep network models. These are called generators and discriminators. GAN is shown in Figure 1. The generator network tries to generate pictures similar to real pictures from randomly generated vectors (usually a noise is used for these vectors). The discriminator network takes these pictures as input values and tries to detect whether they are real or fake. The loss values are calculated with the contention between these two deep networks at the end of each epoch. Gradients are calculated with back-propagation to reduce these losses. Generator starts to better render real images from randomly generated vector with each update.

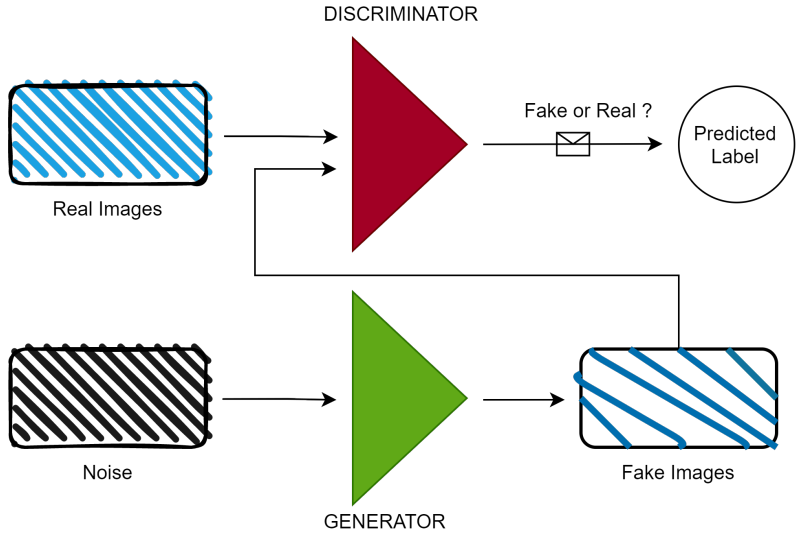


Figure 1. *GAN model*

Despite its popularity, the early work of the GAN had many limitations on the generator network [7]. In recent years, studies on GAN have increased and many GAN types have been created in order to produce deep networks that give better outputs by eliminating these deficiencies. Some of the best known of them are: cGAN, DCGAN, InfoGAN, DiscoGAN, VanillaGAN, Pix2Pix GAN etc. Pix2Pix GAN, which is one of the popular GAN types, was used in this study.

2.2. Pix2Pix GAN Network

Pix2Pix is a Conditional Generative Adversarial Network, developed to solve image-to-image conversion problems [3]. The study, carried out by P. Isola et al. in 2018, is not the first research in the field of Conditional GANs, but it differs from other studies in many ways. Pix2Pix uses a 'U-Net' based architecture in the generator deep network block. The difference between the Encoder-Decoder architecture and U-Net is shown in Figure 2. U-Net is an encoder-decoder architecture including skip connections. It uses skip connections to overcome the bottleneck limitation.

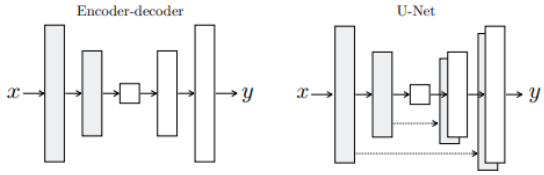


Figure 2. *Encoder-decoder and U-Net network models [3]*

In the discriminator deep network block, small image segments are evaluated using the convolutional 'PatchGAN' classifier. In addition, Pix2Pix, unlike cGAN, uses the L1 norm instead of the L2 norm in its objective function. L1 causes less turbidity. Generator and discriminator deep networks in the Pix2Pix network use convolution, BatchNorm and ReLu modules. The success value that is tried to be obtained in the Pix2Pix network is given in Equation 1.

$$G^* = \arg \min_G \max_D \mathcal{L}_{cGAN}(G,D) + \lambda \mathcal{L}_{L1}(G) \quad (1)$$

The objective of conditional GAN (G*) is shown in the Equation 1, G tries to minimize this objective. Conversely, D tries to maximize it. Mixing the GAN objective with more traditional loss such as L2 distance is a benefical approach. In the pix2pix network, they explore that using L1 distance rather than L2 generates less bluring outputs [3]. Importance of L1 loss can be changed with hyperparameter lambda λ.

In conventional GAN models, the output image y is obtained from the random noise vector z, G: z → y. In conditional GAN models, the output image y is obtained from the observed image x and the random noise

vector as a condition, G: {x, z} → y [8]. The conditional GAN architecture is shown schematically in Figure 3. In the cGAN's generator network G, condition class label c and random noise vector z generate a data G(z). Discriminator network D compares the observed image x, generated data from the generator and condition class label. Then it classifies as a "real" or "fake".

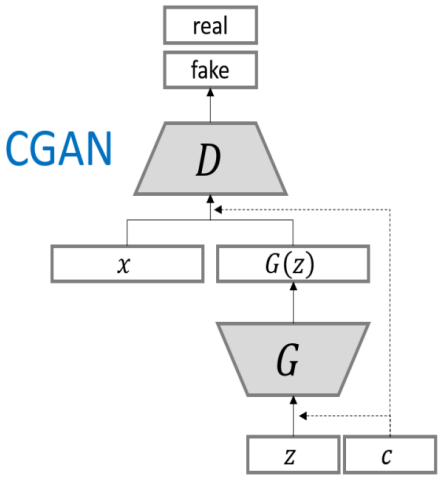


Figure 3. *cGAN model [9]*

In the Pix2Pix network, two images are given as input. The first one is used for fake image production in the generator and the other one is used for evaluating the fake image corresponding to this image in the discriminator. The pixels of these two images must match exactly. Unlike popular traditional GAN models, the Pix2Pix GAN is a supervised and unidirectional model [10]. For Pix2Pix, as in other GAN deep network architectures, the training parameters (learning coefficient, number of iterations, ReLu value, etc.) must be adjusted very precisely, otherwise the output images will be distorted. Examples of images produced with Pix2Pix GAN are given in Figure 4.

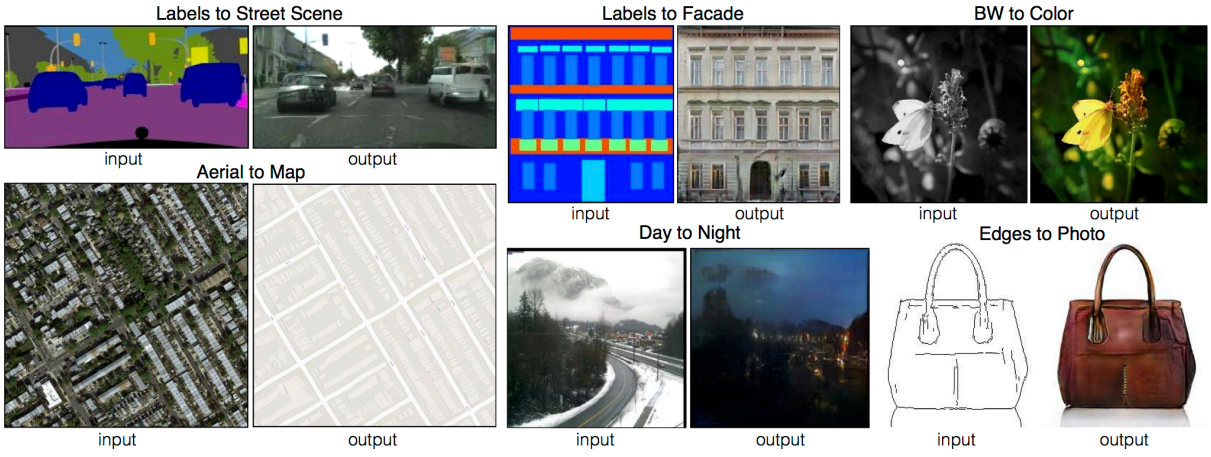


Figure 4. *Example results on several image-to-image translation problems via Pix2Pix network [3]*

3. EXPERIMENTAL RESULTS

This study aims to obtain long wavelength infrared (IR) synthetic images from visible band camera images. The research carried out for this purpose consists of 3 steps.

1- Setting up the Pix2Pix network

2- Creating the dataset

3- Creating infrared synthetic image and evaluating research results

The experimental results were concluded by performing the steps given above in order.

3.1. Setting up the Pix2Pix network

A working Pix2Pix network was built using the Python programming language. Pytorch was used as an artificial intelligence library. The work was carried out on the Visual Studio Code IDE. Before working on the network, Google Maps dataset was used to measure the operability of the network. The dataset consists of satellite images and Google Maps images. There are 1097 map image pairs for training and validation. An example from the dataset is shown in Figure 5.

Figure 5. *Sample image from Google Maps' dataset*

The operation of the Pix2Pix network was tested on this dataset and the operability of the network was tested by observing the results of the study. After this successful setting up, the step of creating a dataset was started.

3.2. Creating the dataset

In our study, a dataset consisting of visible band camera image and LWIR camera image pairs is needed to be used as input in our Pix2Pix network. For this purpose, it was decided to create a dataset consisting of human faces both because it is a heat source and in order to examine the performance in details. It is very difficult to establish pixel-by-pixel matching between color images and grayscale images [11]. Pixel-by-pixel mapping of these images to be used while feeding the Pix2Pix deep network is a very important issue. In order to ensure the realization of this match, the image acquisition conditions have been created to meet these conditions. Image processing methods were used in the Matlab environment in order to increase the matching in the received images from cameras. The experimental setup created to obtain LWIR and visible band images is as in Figure 6.

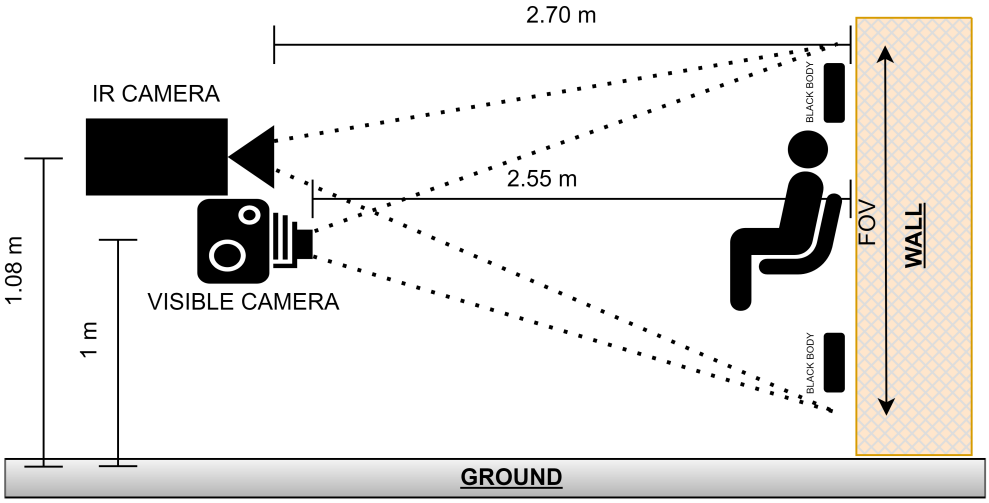


Figure 6. *Cameras and scene locations scheme*

In order to match with minimum error in LWIR camera images and visible band camera images, blackbodies (BB) at 40°C temperatures were used to determine the reference points in the Field of Views (FOVs) of both cameras. The upper right and lower left FOV references of the cameras are determined by these blackbodies. The lenses of the cameras are brought to the closest position to each other in the same direction and inclination as far as the physical conditions allow. The FOV of the cameras was created by paying attention to include the whole scene while ensuring that they overlap at the highest level.

Images of 23 men and women were taken to create the dataset. While taking these images, images were taken in VGA format using 7 different angles, with 640 x 480 dimensions. These angles are shown in Figure 7.

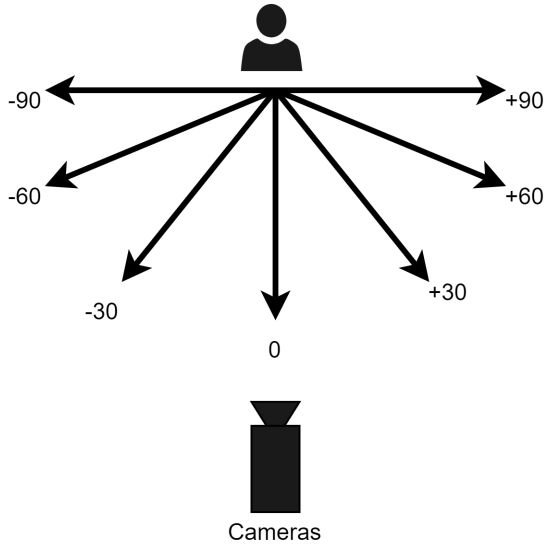


Figure 7. *Photos shooting angles*

In order to see the performance of the Pix2Pix GAN network in details, both glasses and non-glasses images of people with glasses were taken for each angle. Objects containing glass, such as glasses, are objects that have low trasmittance level at infrared wavelengths. Such objects become more prominent in infrared images by creating contrast. Infrared image pairs with and without glasses are shown in Figure 8.

Figure 8. *Glasses effect*

Here, it is aimed to check whether the deep network makes such a detail separation during the image transformation. The acquired images were subjected to image processing methods such as scaling, shifting and rotation in the Matlab environment and consequently pixel-by-pixel matching was performed. Pixel matching was controlled with imshowpair methods in matlab. The matching errors caused by the experimental setup are minimized in this way. The images before and after the image matching performed in the Matlab environment are shown in Figure 9.a and Figure 9.b.

Figure 9. *a) Acquired IR and visible images taken from cameras on top of each other using imshowpair method in matlab environment, b) superimposing images using pixel-by-pixel matching methods, c) display of mapped and cropped images for use in dataset with imshowpair method, d) representation of the visible and ir image pair to be used in the dataset*

Infrared and visible band images were combined in Matlab after cutting at 470 x 470 so 250 infrared and visible image pairs were obtained in 940 x 470 dimensions (Figure 9.c). 43 of these pairs were randomly selected as a validation dataset. These images were used to fine-tune the network parameters during training. Moreover, 7 image pairs of a person who was not included in the training and validation dataset were selected as a test dataset. Images of a person were not included in the training data in order to show the result of the Pix2Pix network creating unmemorized infrared image of an individual. In addition, the results were examined by giving face images with and without glasses in the validation data. The remaining 200 image sets are labeled as training data. An example of this data set is also shown in Figure 9.d. This created dataset is given as an input to the Pix2Pix deep network.

If we want to test the discriminator's ability to generalize data it has never seen before, it is necessary to split our dataset into train and test and also increase the test datas. However, the aim of this study is slightly different. We did not choose this as we aimed to test the performance of the experiments. One of the reasons why we do not prefer it is that our dataset is limited. Relying on our experience, the train and validation

data were kept high and it was aimed to increase the training of the network. Test datas were used to examine performance on unknown images. In addition, different parameters (MSE, SNR etc.) are given below to measure this performance.

3.3. Creating infrared synthetic image and evaluating research results

The Pix2Pix deep network uses convolution, BatchNorm and ReLu modules in both the generator and the discriminator.When there is a change in the input distribution to our network, and hidden layers of network try to learn new distribution. These layers update their weights after each batch. This problematic change in the distribution is named as an Internal Covariate Shift. As a result of Internal Covariate Shift, undesirable results are obtained at the output. The main reason for this is the incompatibility of the data used as training and test data. Datasets with similar functional properties do not always take values in the same range. During the training, the input values of each layer change depending on the change in the previous layer. This causes both a lower learning rate to be used and a very fine tuning of the initial parameters. The decrease in learning rate also slows down training. In order to prevent this, the error rate is reduced by normalizing the input of each layer [12]. For this purpose, in the Pix2Pix network, Batch Normalization is applied in order to prevent these internal covariate shifts that occur while the weights are updated during the back propagation stage, and it is aimed to give the network more stable and better results. Optimization of the neural network has been increased by performing batch normalization. Thus, it provided much better results in DCGAN (Deep Convolutional GAN) [13-14].

In 2016, a study was carried out by Ulyanov et al. to obtain repetitive patterns (texture) and stylized images. In this study, it was seen that adding Batch Normalization after each convolution layer and concatenation layer gave significant better results [15]. It has been observed that applying normalization before each layer causes oscillation and unstable conditions on the results [13]. In 2017, D.Ulyanov et al. discovered that small changes they made on the fast stylization method they had performed before gave much better results [16]. As a result of mentioned study, the Instance Normalization method, also known as Contrast Normalization, has emerged, which gives much better results than Batch Normalization in image production. In Contrast Normalization, unlike Batch Normalization, each image is normalized separately.

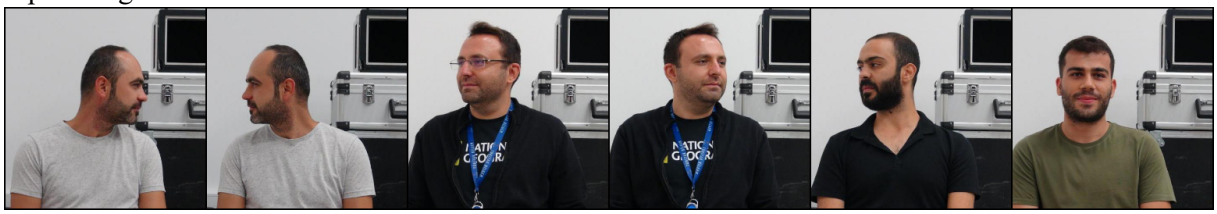
In this study, while trying to obtain infrared images from visible band images, we aimed to determine which normalization pair is the best used in generator and discriminator deep networks. We created four different weight normalization usage tables in our experiments on our own dataset [Table 1].

Table 1. *Experiments Control Table*

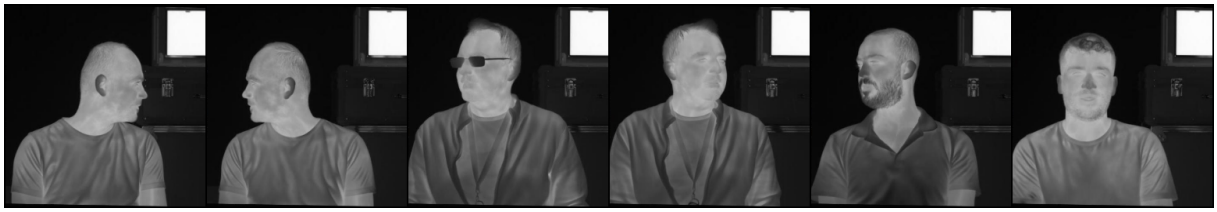
Experiment Name	***Discriminator***	***Generator***
Experiment 1	*Instance Normalization*	*Batch Normalization*
Experiment 2	*Instance Normalization*	*Instance Normalization*
Experiment 3	*Batch Normalization*	*Batch Normalization*
Experiment 4	*Batch Normalization*	*Instance Normalization*

Sample results of four different experiments performed are given in Figure 10. As can be seen from the images, it is very difficult to distinguish the similarities between synthetic infrared images and real images.

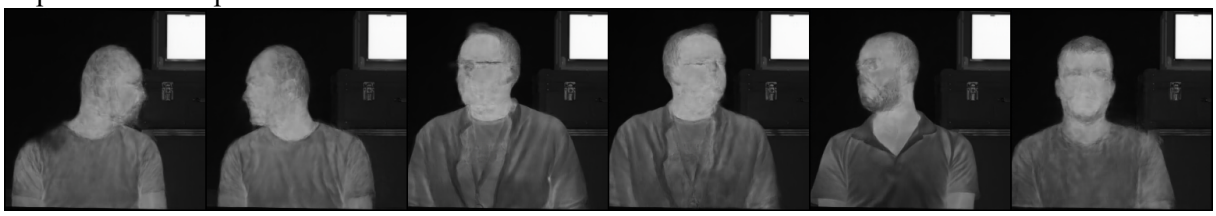
Input Images:



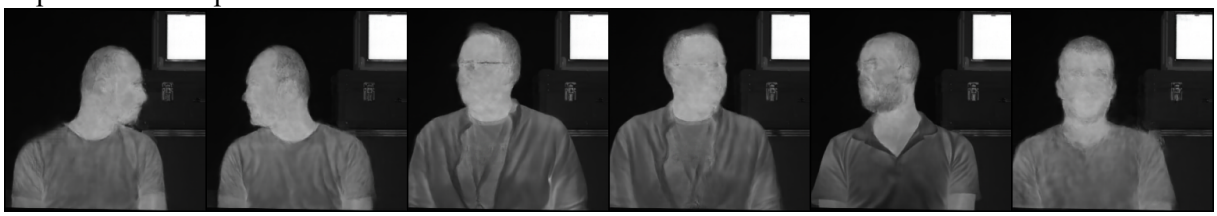
Ground Truth:



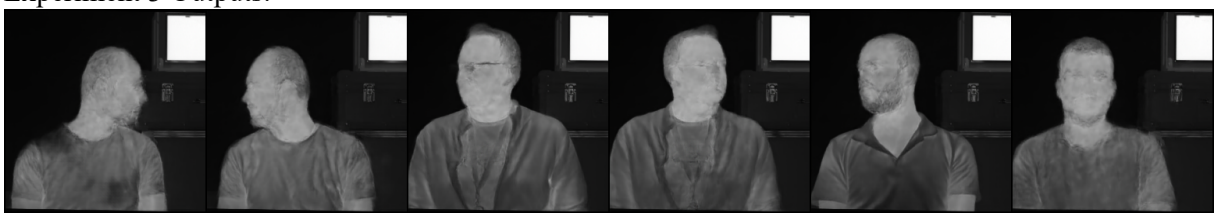
Experiment 1 Outputs:



Experiment 2 Outputs:



Experiment 3 Outputs:



Experiment 4 Outputs:

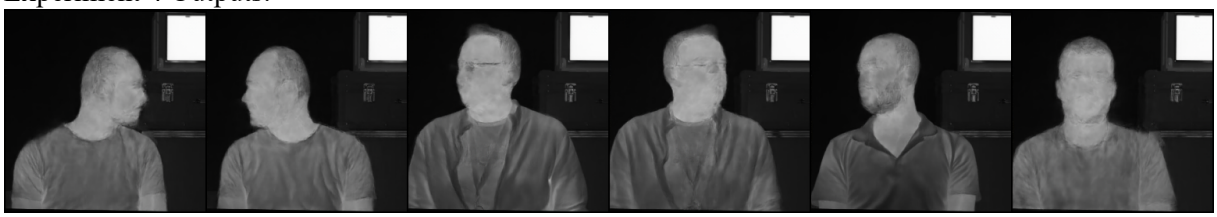


Figure 10. *Sample of input images, ground truth and experiments outputs*

For this reason, Mean Square Error (MSE), Structural Similarity Index Measure (SSIM), Signal to Noise Ratio (SNR) and Peak Signal to Noise Ratio (PSNR) parameters between real images and synthetic images created to evaluate the best result among experiments were calculated. The results of these are shown in the chart below [Figure 11].

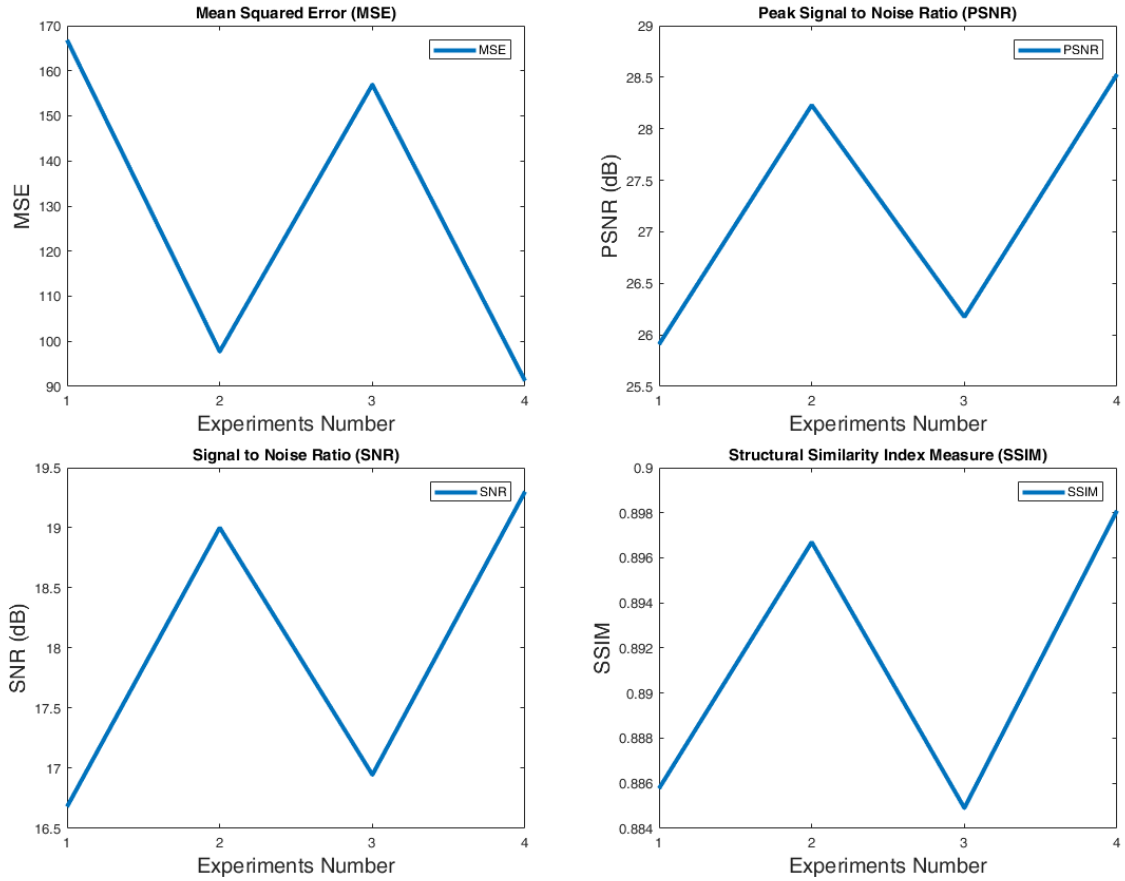


Figure 11. *Own dataset experiments result*

Four different synthetic infrared images were obtained in four different experiments. The obtained output results were compared with the real infrared accuracy images (ground truth images). When the results are compared, it is seen that the best results are obtained when Batch Normalization is used on the discriminator and Instance Normalization is used on the generator as in Experiment 4 [Figure 12]. The second best results are Experiment 2 results. According to Experiment 2, the use of Contrast Normalization in generator and discriminator produces better results.

Figure 12. *Best experiment result with never trained face image (left to right: visible real image, infrared image and fake infrared image)*

We would like to point out that our study is not valid for all datasets. For this purpose, the same experiment was used to obtain a google map from the satellite image on Google Maps dataset. SNR, PSNR, SSIM and MSE values of the output results were calculated according to the actual data and are shown in Figure 13.

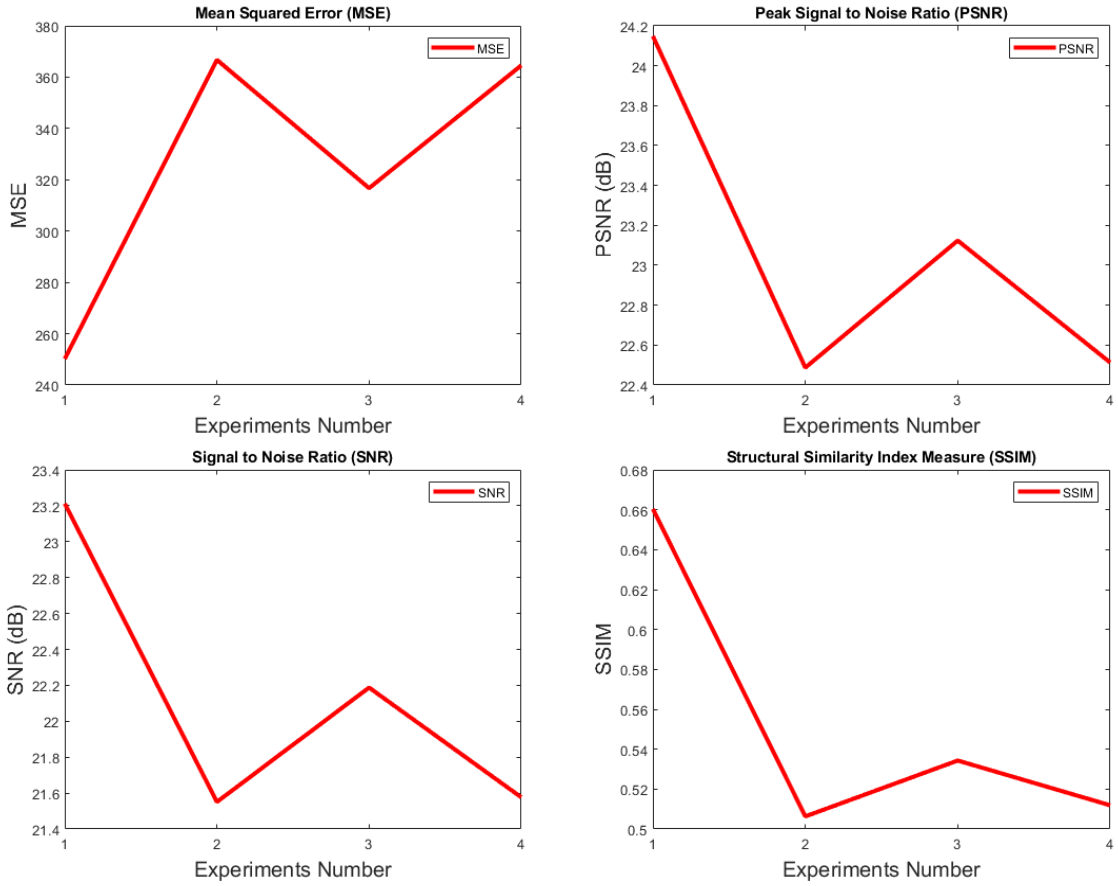


Figure 13. *Google maps dataset experiments result*

In this study carried out on the Google Maps dataset, it was observed that using Instance Normalization on the discriminator and Batch Normalization on the generator, which are the conditions of Experiment 1, gave much better results than other methods. The worst performance was obtained under Experiment 4 conditions. Figure 14 shows the correct and incorrect results obtained from the discriminator network under the conditions of Experiment 4.

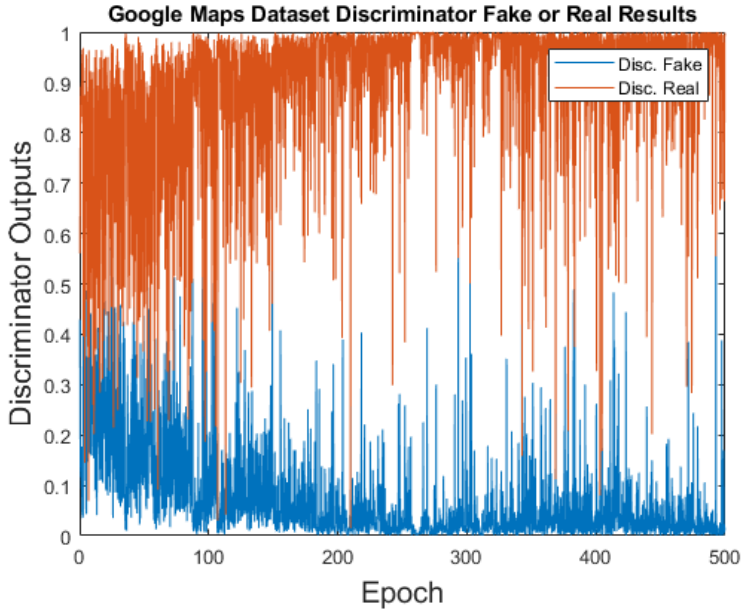


Figure 14. *Google Maps dataset discriminator fake or real loss results for Experiment 4*

In this model, which was trained using RGB 3-channel images, it was observed that the learning process became stable after approximately 200 epochs. The discriminator network has been quite successful in separating the generator's outputs from the real.

As shown in Figure 15, the discriminator outputs of Experiment 4 applied on own dataset, which we determined as the most successful model in generating infrared images from visible band images, are given.

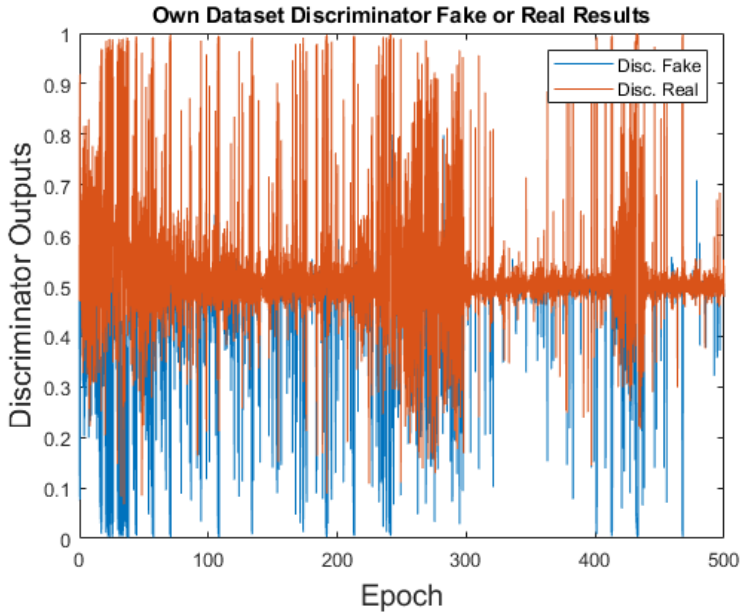


Figure 15. *Own dataset discriminator fake or real loss results for Experiment 4*

The results of the model trained for 500 epochs are as in the Figure 15. It is seen that the deep network has difficulty in making the necessary discrimination due to the fact that the same data is repeated in all 3 channels of the infrared image. In the production of infrared synthetic data, performance of the generator deep network has also increased. It has been observed that as the number of epochs is increased; the steady state increases and the loss values decrease, creating better results for the generator.

4. DISCUSSION

It is predicted that more successful results can be obtained with more sampling of the dataset created from the visible band and infrared images used in the research. In the study carried out, images of people with and without glasses were used and it was observed that the Pix2Pix network could not fully distinguish these details in people. The pre-trained Pix2Pix network fed with glasses-free input data started to create thermal human figures with glasses [Figure 10]. There is a possibility that the Patch Network used on the discriminator in Pix2Pix networks may cause this. The possibility that the number of training data in the network may prevent making this separation should also be considered. Since the samples taken at 30º and 60º angles do not match at the contrast change points where the frequency is high depending on the camera positions, additional studies are needed to make the infrared images created at these angles more successful. In the studies carried out on Google Maps dataset and own dataset, it has seen that the most successful results were in the different experiments. Input images of Google Maps dataset were RGB and outputs of network were also RGB images. On the other hand, input images of own dataset were RGB and outputs of network were Grayscale images. For these reason, the synthetic images produced have different properties. As a result, the effects of normalization methods on these produced images differ.

5. CONCLUSION AND FUTURE WORKS

With the help of this study, a Pix2Pix deep network was created that obtains infrared synthetic images from the images of the visible band camera. Different normalization methods were tried on the Pix2Pix network and their effects on the network were examined. The most suitable method is proposed for infrared image conversion from visible band images. In addition, the effects of changes such as batch size, color toning

and horizontal rotation on the deep network output should not be ignored. Infrared cameras are costly radiometric devices. Obtaining infrared images in researches in this field can be challenging. Thanks to this study, synthetic infrared image generation was performed and an improvement in this field was proposed. It is also possible to convert to the temperature and radiance values in the desired calibration range from the obtained synthetic images. With the help of the trained datasets, conversions can be performed between infrared images in different band gaps (Longwave IR, Midwave IR, Shortwave IR etc.). Thus, reflections, saturations and artifacts that a person cannot notice or predict with the naked eye can be created by artificial intelligence. As it is known, Pix2Pix is a unidirectional architecture. In future studies, it is planned to increase the dataset and present improvements and different methods in the formation of visible band facial images from infrared images [Figure 16].

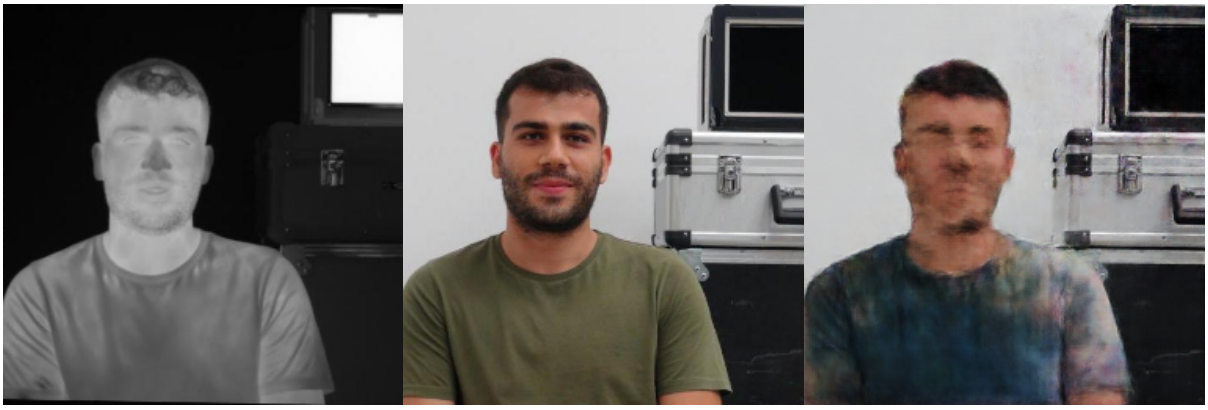


Figure 16. *Infrared image to visible image conversion*

It is considered that an improvement study to be carried out in this area may be beneficial in terms of extracting real face image features from the data obtained from infrared cameras in security, intelligence and military fields.

ACKNOWLEDGEMENTS

We would like to thank TÜBİTAK BİLGEM İLTAREN and Turkish Land Forces. This work is partially carried out at their laboratories and equipments. We are also grateful to Dr. Demet Sevil ARMAĞAN ŞAHINKAYA, İlhami BEKTAŞ, Emrah ODUNCU and Umut KAYIKÇI for their support, without whom this study could not be accomplished.

REFERENCES

[1] McCarthy J., What is Artificial Intelligence?, Stanford University, Available at: http://www-formal.stanford.edu/jmc/whatisai.pdf, (Accessed: 2021-10-05), 2.

[2] Goodfellow I. J., Pouget-Abadie J., Mirza M., Xu B., Warde-Farley D., Ozair S., Courville A. C., and Bengio Y., Generative Adversarial Nets, In Proceedings of NIPS, (2014) 2672–2680.

[3] Isola P., Zhu J., Zhou T., Efros A., Image-to-Image Translation with Conditional Adversarial Networks, 2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), (2017).

[4] Balcı O., Terzi Ş. B., Balaban Ö., Map Generation & Manipulation With Generative Adversarial Network, Journal of Computational Design, 1 (3), (2020) 95-114.

[5] Demirhan A., Kılıç Y. A., Güler İ., Artificial Intelligence Applications in Medicine, Yoğun Bakım Dergisi 9(1):31-41, (2010).

[6] ELMAS Ç., Yapay zeka uygulamaları, Ankara: Seçkin Yayıncılık 2018, pp.479.

[7] Turhan C. G., Bilge H. S., Variational Autoencoded Compositional Pattern Generative Adversarial Network for Handwritten.

[8] M. Mirza, S. Osindero, Conditional Generative Adversarial nets., arXiv preprint arXiv:1411.1784, (2014).

[9] A. Mino, G. Spanakis, LoGAN: Generating Logos with a Generative Adversarial Neural Network Conditioned on color, 17th IEEE International Conference on Machine Learning and Applications, (2018) 966.

[10] Altun S., Talu M.F., Review of Generative Adversarial Networks for Image-to-Image Translation and Image Synthesis, Avrupa Bilim ve Teknoloji Dergisi, Ejosat Özel Sayı 2021 (HORA), (2021) 53-60.

[11] Liu M., Breuel T., Kautz J., Unsupervised Image-to-Image Translation Networks, arXiv preprint arXiv:1703.00848, (2017).

[12] Ioffe S., Szegedy C., Batch Normalization: Accelerating Deep Network Training by Reducing Internal Covariate Shift, ICML'15: Proceedings of the 32nd International Conference on International Conference on Machine Learning, Volume 37 (2015) 448–456.

[13] Radford A., Metz L., Chintala S., Unsupervised Representation Learning with Deep Convolutional Generative Adversarial Networks, arXiv:1511.06434, (2015).

[14] Salimans T., Goodfellow I., Zaremba W., Cheung V., Radford A., Chen X., Improved Techniques for Training GANs, arXiv:1606.03498, (2016).

[15] Ulyanov D., Lebedev V., Vedaldi A., Lempitsky V., Texture Networks: Feed-forward Synthesis of Textures and Stylized Images, ICML'16: Proceedings of the 33rd International Conference on International Conference on Machine Learning, Volume 48 (2016) 1349–1357.

[16] Ulyanov D., Vedaldi A., Lempitsky V., Instance Normalization: The Missing Ingredient for Fast Stylization, arXiv preprint arXiv:1607.08022, (2016).