Scientific Visualization, 2019, volume 11, number 4, pages 43 - 52, DOI: 10.26583/sv.11.4.04

Segmentation and visualization of obstacles for the enhanced vision system using generative adversarial networks

V.V. Kniaz[1,A,B], M.I. Kozyrev[2,A,C], A.N. Bordodymov[3,A],
A.V. Papazian[4,A], A.V. Yakhanov[5,A]

[A] State Res. Institute of Aviation Systems (GosNIIAS)
[B] Moscow Institute of Physics and Technology (MIPT)
[C] Bauman Moscow State Technical University (BMSTU)

[1] ORCID: 0000-0003-2912-9986, vl.kniaz@gosniias.ru
[2] ORCID: 0000-0001-9901-5664, j18r1l@gmail.com
[3] ORCID: 0000-0001-8159-2375, bordodymov@gmail.com
[4] ORCID: 0000-0003-0119-011X, ares.papazian@yandex.ru
[5] ORCID: 0000-0003-4284-6197, yakhanovalexander@gmail.com

Abstract

Long range infrared cameras may provide increasing crew situational awareness in limited vision and night conditions.

Similar cameras are installed in modern civil aircraft's as part of an improved vision system. Correct thermal image interpretation by the crew requires certain experience, due to the fact that view of the scene very different from the visible range and may change within time of day and season. This paper discusses the deep generative-adversary neural network to automatically convert thermal images to semantically similar color images of the visible range.

Keywords: visualization, deep convolutional neural networks, pilot primary display, visual analytics.

1. Introduction

Increasing crew situational awareness is a guarantee of flight safety. Today's modern civilian aircraft have advanced vision systems. Such system includes an infrared camera that captures images of the cockpit view in the front hemisphere, and a processing unit, that receives the video signal and displays it on the pilot’s multifunctional display. Thermal infrared sensor provides the display of visible objects and terrain in limited vision and night conditions.

The disadvantages of an improved vision system with infrared sensor include difficulties in interpretation thermal image. Since thermal radiation of objects may change within the weather, their appearance on the frame of the improved vision system can vary significantly from the time of day to time of year. For example, the runway may be light against a dark background on sunny days and dark against light during rain. To make it easier for the pilot to detect the visual landmarks seems advisable to pre-process the frame of the improved vision system in order to convert the infrared image into the visible range.

In this paper, we consider a method to convert monochrome thermal images into color images of the visible range. The method uses a modified version of the generative-adversarial network ColorMatchGAN. The network architecture is presented. For training and testing the network, a training sample was collected using the DJI Mavic PRO quadcopter (UAV), equipped with visible and far-infrared cameras. The method of semi-automatic combination of visible and infrared frames is presented. The modified ColorMatchGAN is trained on the collected sample. Testing was carried out on an independent sample of 400 frames.

2. Related work

Computer vision-based situational awareness systems have been widely adopted over the past decade [2, 11]. The most widely used systems based on far-infrared sensors (8-14 µm), which provide an overview of the cockpit in the direction of movement of the aircraft [2, 11]. Such systems are commonly called improved vision systems. The main quality criteria for improved vision systems are the detection range of the runway and obstacles on its surface or in the air. Various algorithms for improving image quality are proposed to increase the detection range of objects [7].

Despite the significant increase in situational awareness provided by modern improved vision systems, interpretation of thermal images can cause significant difficulties for the crew. It is advisable to pre-process the thermal image, which predicts the colors of the object composition and background to facilitate the interpretation of the observed scene. Over the past five years, neural network image processing methods based on generative-adversarial neural networks have been actively developed [3, 4]. The basic idea of a generative-adversarial approach is to train two competing networks: generator G and discriminator D.

The generator's goal is to learn the given distribution of images $B \subset R\ W \times H \times C$ and learn how to reproduce it based on the noise vector z or the input image A. The purpose of the discriminator is the binary classification of the input image into classes: «real» and «model». «Real» images $B \in B$ belongs to the output images. The «model» images $\hat{B}$ are the result of the operation of the generator network G. The adversarial loss function imposes a fine on the network generator if the network discriminator correctly classifies images $\hat{B}$ with the «model» class. Thus, the network generator is trying to build the most plausible images of $\hat{B}$ to confuse the network discriminator.

In recent years, a number of works have been proposed to transform the spectral range of images based on generative-adversarial neural networks [1, 5, 10]. In this paper, we consider a modification of ColorMatchGAN architecture [5]to predict color images from thermal images.

3. THERMAL-to-COLOR image translation method

The purpose of this method is to transform the input image $A \in R\ W \times H$ from far-infrared into the color image $B \in R\ W \times H \times 3$ of the visible spectrum. Required transformation $G: A \rightarrow \hat{B}$ is implemented using a modified generator network based on ColorMatchGAN[5] architecture. This section discusses conditional generative adversarial neural networks that underlie the developed method. A modified network architecture and training sample preparation technique are presented.

Network architecture. Generative-adversarial networks use [3] the adversarial loss function to reduce the likelihood of over-fitting the network. Generative-adversarial networks create an image B for a given random noise vector z, $G : z \rightarrow \hat{B}$ [3, 4]. Conditional generative-competitive networks receive additional information A in addition to the vector z, $G: \{A, z\} \rightarrow \hat{B}$. Usually, A is an image that is transformed by a generative model G. The discriminative model is trained to distinguish between "real" images from the target domain B from "fake" $\hat{B}$ created by the generator. Both models are trained simultaneously. The discriminative model creates an adversarial loss that causes the generator to produce "fake" B images that cannot be distinguished from "real" B. ColorMatchGAN [5] network architecture includes generator U-Net [8] and discriminator PatchGAN [4]. ColorMatchGAN network architecture is presented on Figure 1.

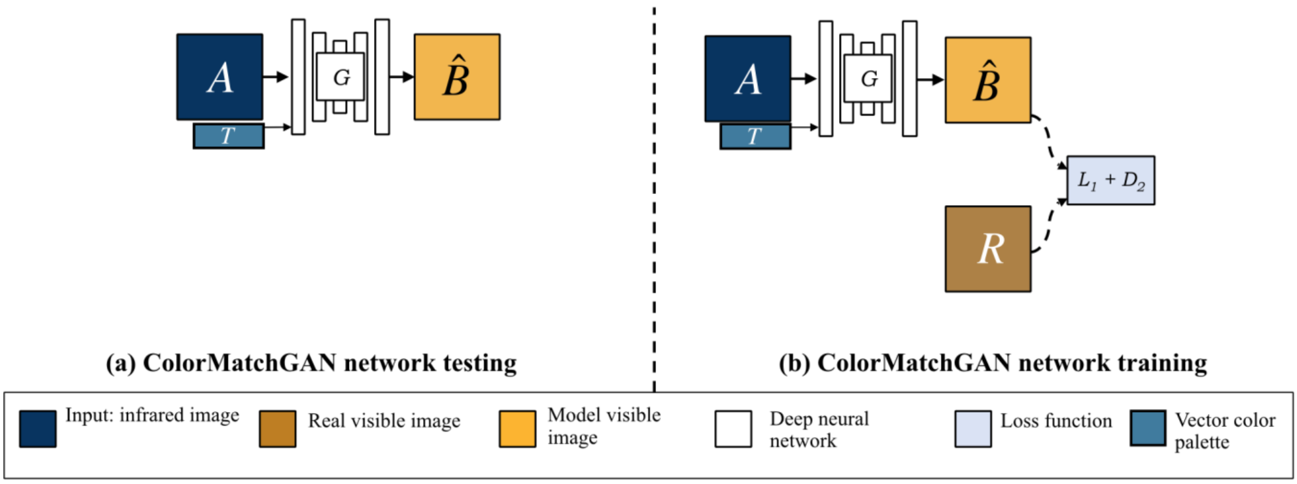


Figure 1: ColorMatchGAN network architecture.

The vector T is based on a histogram of a “real” image converted to the LAB color space, where L expresses lightness and AB express color tone. From a one-dimensional matrix $Z = \ln(\mathrm{flat}(H_{ab}^{T}) + 1)$, where H_{ab} is a two-dimensional histogram of AB from LAB, T matrix was formed, where every element is a copy of Z. Matrix A, being a single channel input image, is concatenated with matrix T and fed forward to ColorMatchGAN neural network.

4. Dataset generation

The LEART training set was used to train the modified network architecture [6]. This sample was collected using a DJI Mavic PRO UAV, equipped with an integrated visible camera, and an additional far-infrared camera (8-14 µm) MH-SM576-6 with a resolution of 640 × 480 pixels. A general view of the UAV is shown in Figure 2.

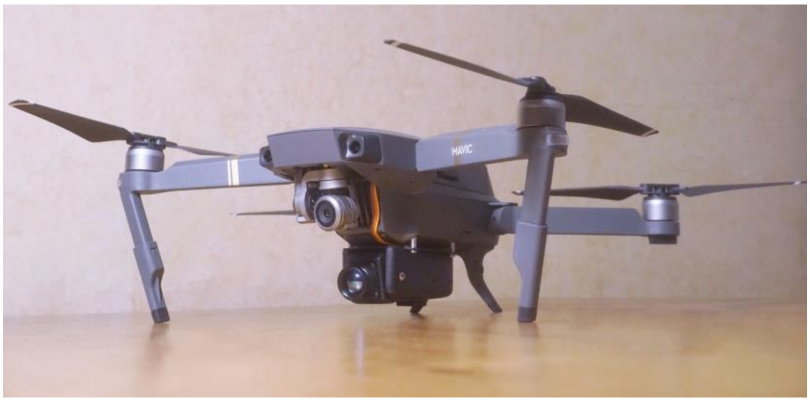


Figure 2: View of the Mavic PRO UAV with cameras of visible and infrared range.

Since the camera of the visible range is mounted on a gyro-stabilized suspension, and the thermal imaging camera is rigidly connected to the body, there is a dynamic discrepancy between color and thermal imaging images. A technique has been developed for the semi-automatic combinations of images of two ranges to eliminate the geometric discrepancy. The technique of combining images of two ranges is based on the use of a homography matrix.

$$H = \begin{bmatrix} h_{00} h_{01} h_{02} \\ h_{10} h_{11} h_{12} \\ h_{20} h_{21} h_{22} \end{bmatrix}$$

Let (x_v , y_v) – be the point in the image of the visible range and (x_t , y_t) – be the point in the thermal image in the same physical place. Then the homography H connects them as follows

$$H = \begin{bmatrix} x_v \\ y_v \\ 1 \end{bmatrix} = H \begin{bmatrix} x_t \\ y_t \\ 1 \end{bmatrix} = \begin{bmatrix} h_{00} h_{01} h_{02} \\ h_{10} h_{11} h_{12} \\ h_{20} h_{21} h_{22} \end{bmatrix} \begin{bmatrix} x_t \\ y_t \\ 1 \end{bmatrix}$$

If the parameters of the homography matrix are known, then you can find the transition from a given point in the image in the visible range (x_v , y_v) to the corresponding point (x_t , y_t) in the infrared image. To calculate the homography matrix, it is necessary to know at least four corresponding points in two images.

Obviously, the process of automatically arranging pairs of points on all frames of a video sequence is a laborious process. It is proposed to use the tracking of points between frames using cross-correlation to automate the task. Four corresponding points are placed on the first frame of the video sequence and are tracked until they are visible in the camera's field of view. Coordinates (x_v^i, y_v^i),(x_t^i, y_t^i) each point, on each frame i, are placed into an array. After that, for each element of the resulting array, the frame of the visible range is converted to the infrared frame.

The proposed technique was implemented as a script in Python. To track the corresponding points, we used the Blender 3D modeling package API. Examples from the training set are shown in Figure 3.

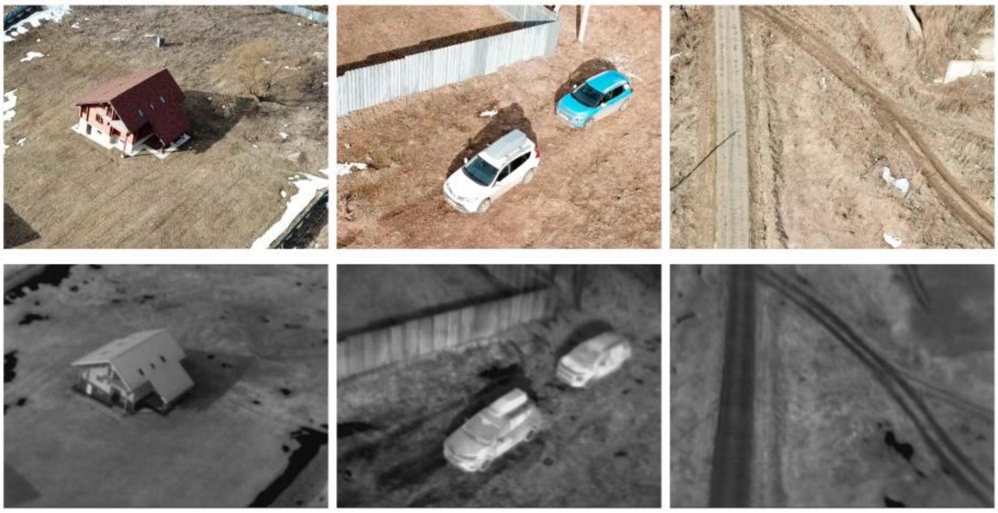


Figure 3: Examples from the LAERT training sample.

Convert images to LAB color space. To train the network, we used the LAB color space, in which lightness is measured along the L axis (in the range from 0 to 100%), displaying the spectral reflection coefficient, the red-green hue is measured along the a axis, and the yellow-blue hue along the b axis (in the range from -120 to +120). To convert an RGB image to LAB, you must first convert the image to the XYZ color space.

$$\begin{bmatrix} X \\ Y \\ Z \end{bmatrix} = [M] \begin{bmatrix} R \\ G \\ B \end{bmatrix} \text{ где } [M] = \begin{bmatrix} S_r X_r S_g X_g S_b X_b \\ S_r Y_r\, S_g Y_g\, S_b Y_b \\ S_r Z_r S_g Z_g S_b Z_b \end{bmatrix}, X_r = \frac{x_r}{y_r}, Y_r = 1, Z_r = \frac{1 - x_r - y_r}{y_r} \begin{bmatrix} S_r \\ S_g \\ S_b \end{bmatrix} = \begin{bmatrix} X_r X_g X_b \\ Y_r\, Y_g\, Y_b \\ Z_r Z_g Z_b \end{bmatrix}^{-1} \begin{bmatrix} X_W \\ Y_W \\ Z_W \end{bmatrix}$$

After translating the image to XYZ, translate it to LAB color space

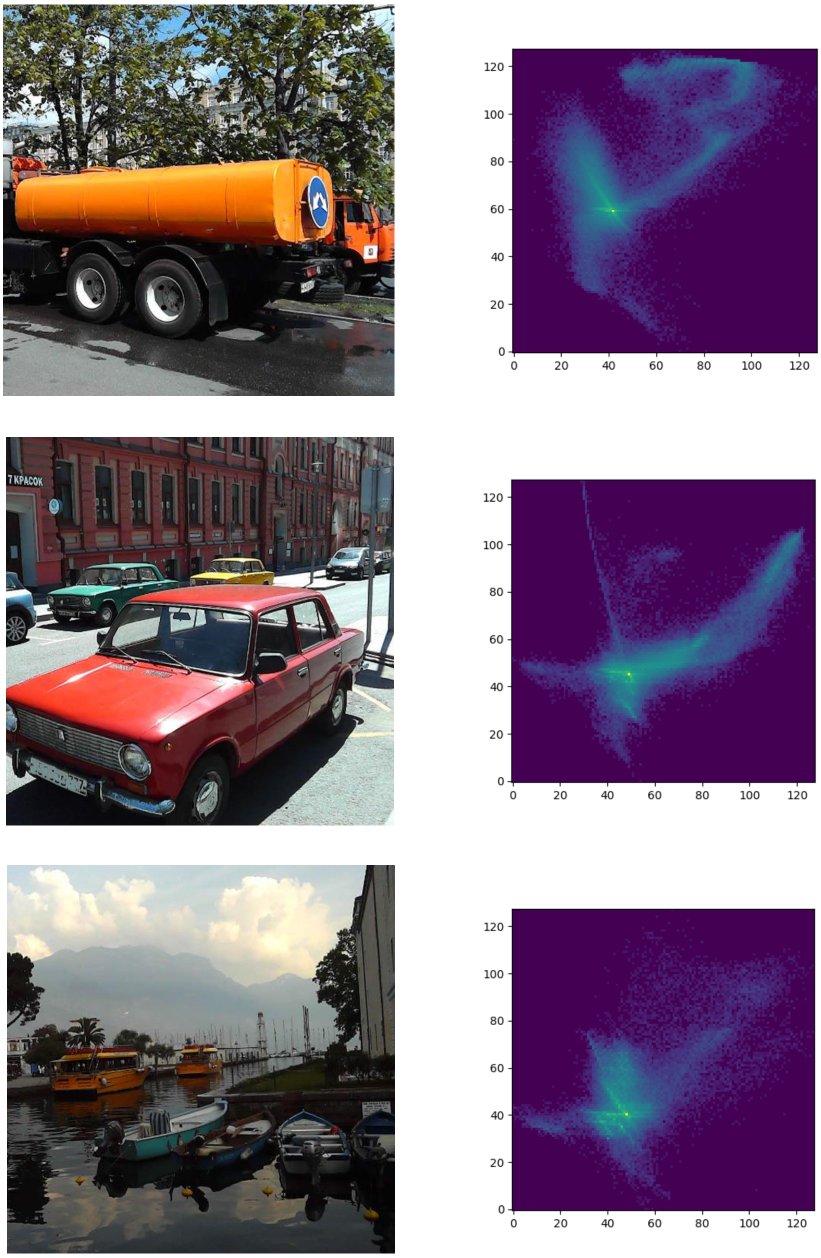
$$L^* = \begin{cases} 116 \left(\frac{Y}{Y_n}\right)^{\frac{1}{3}} - 16 & if \frac{Y}{Y_n} > 0.008856 \\ 903.3 \left(\frac{Y}{Y_n}\right) & if \frac{Y}{Y_n} \leqslant 0.008856 \end{cases}$$

$$a^* = 500 * \left(f\left(\frac{X}{X_n}\right) - f\left(\frac{Y}{Y_n}\right) \right)$$

$$b^* = 200 * \left(f\left(\frac{Y}{Y_n}\right) - f\left(\frac{Z}{Z_n}\right) \right)$$

где $f(t) = \begin{cases} t^{\frac{1}{3}} & if\ \ t > 0.008856 \\ 7.787 * t + \frac{16}{116} & if\ \ t \leqslant 0.008856 \end{cases}$

Since the task of converting a monochrome infrared image to color is incorrect, an additional vector of information of the color palette is required to ensure the stability of color prediction. A two-dimensional histogram of the color frequencies in the Lab color space is constructed to calculate this vector. It is known that, on average, colors often converge to gray in the picture, to increase the branch of saturated colors, the logarithm of the histogram occurs. Examples of constructed two-dimensional histograms in the Lab color space and the original images are shown in Figure 4.



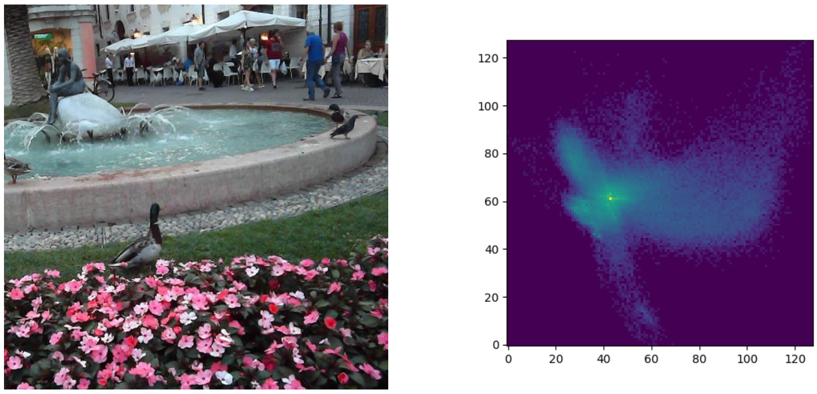
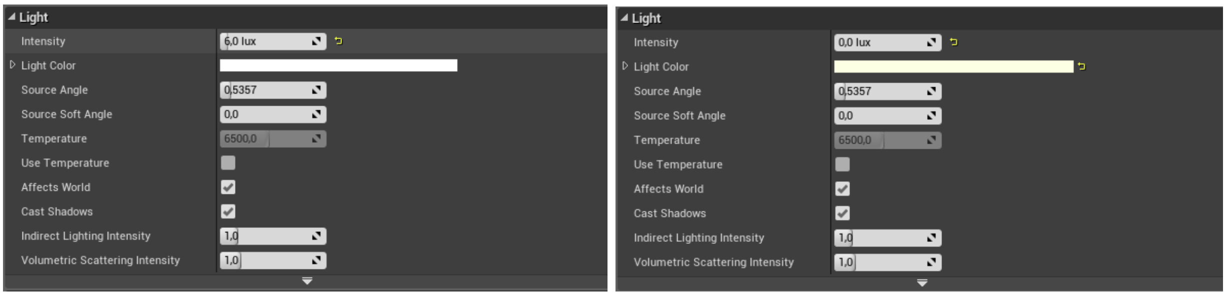


Figure 4: Histograms in Lab space (right), converted to input vector T, constructed from visible range images (left)

Dataset Generation Using Unreal Engine 4 Game Engine

The disadvantage of the LAERT training sample is a small variety of weather conditions and objects. Augmentation was performed using Unreal Engine 4 software to expand the training sample. Based on the survey data from the UAV, the construction of a large-scale orthophotoplane and three-dimensional models of objects using Agisoft Photoscan software prepared textures of the visible and far-infrared range.

Rough three-dimensional models recovered with Agisoft have been edited with Blender software. The resulting scene was imported into the Unreal Engine 4, and the lighting was adjusted (an example of the settings is shown in figure 5). A camera movement scenario has been created that simulates movement on the surface of an ellipse of a given radius. 5000 pairs of images in the visible and infrared ranges from arbitrary angles were formed using the script. Figure 6 shows an example of visible and infrared images taken from the same angle.



Light	
Intensity	6,0 lux
Light Color	▭
Source Angle	0,5357
Source Soft Angle	0,0
Temperature	6500,0
Use Temperature	□
Affects World	☑
Cast Shadows	☑
Indirect Lighting Intensity	1,0
Volumetric Scattering Intensity	1,0

Light	
Intensity	0,0 lux
Light Color	▭
Source Angle	0,5357
Source Soft Angle	0,0
Temperature	6500,0
Use Temperature	□
Affects World	☑
Cast Shadows	☑
Indirect Lighting Intensity	1,0
Volumetric Scattering Intensity	1,0

Figure 5: Scene lighting adjusting in UE4 for visible (left) and infrared (right) ranges

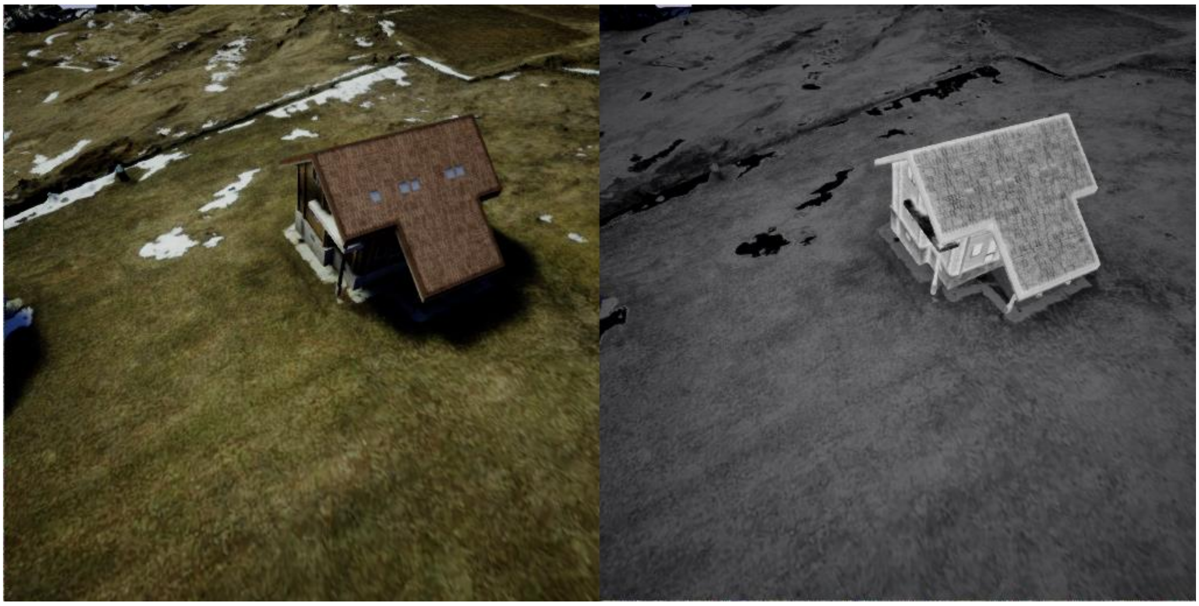


Figure 6: Visible and infrared images taken from the same angle

5. Experiments

ColorMatchGAN was trained on the independent test sub-sample of the LAERT training dataset, using the PyTorch library. In training, NVIDIA 1080 Ti graphics processor was used. The training process took 76 hours for generator G and discriminator D. To optimize the network, we used the Adam gradient descent algorithm with an initial learning rate of 0.0002 and moment parameters β 1 = 0.5, β 2 = 0.999, similarly to [4].

The results of the experimental testing of the network are shown in Figures 7 and 8. A qualitative comparison of the results shows that the ColorMatchGAN network provides an increase in the quality of predicted color images. Quantitative testing using the LPIPS metric [9] shows that the distance between true color images and the ColorMatchGAN prediction is less than the similar distance for images predicted by the pix2pix neural network by 20%.

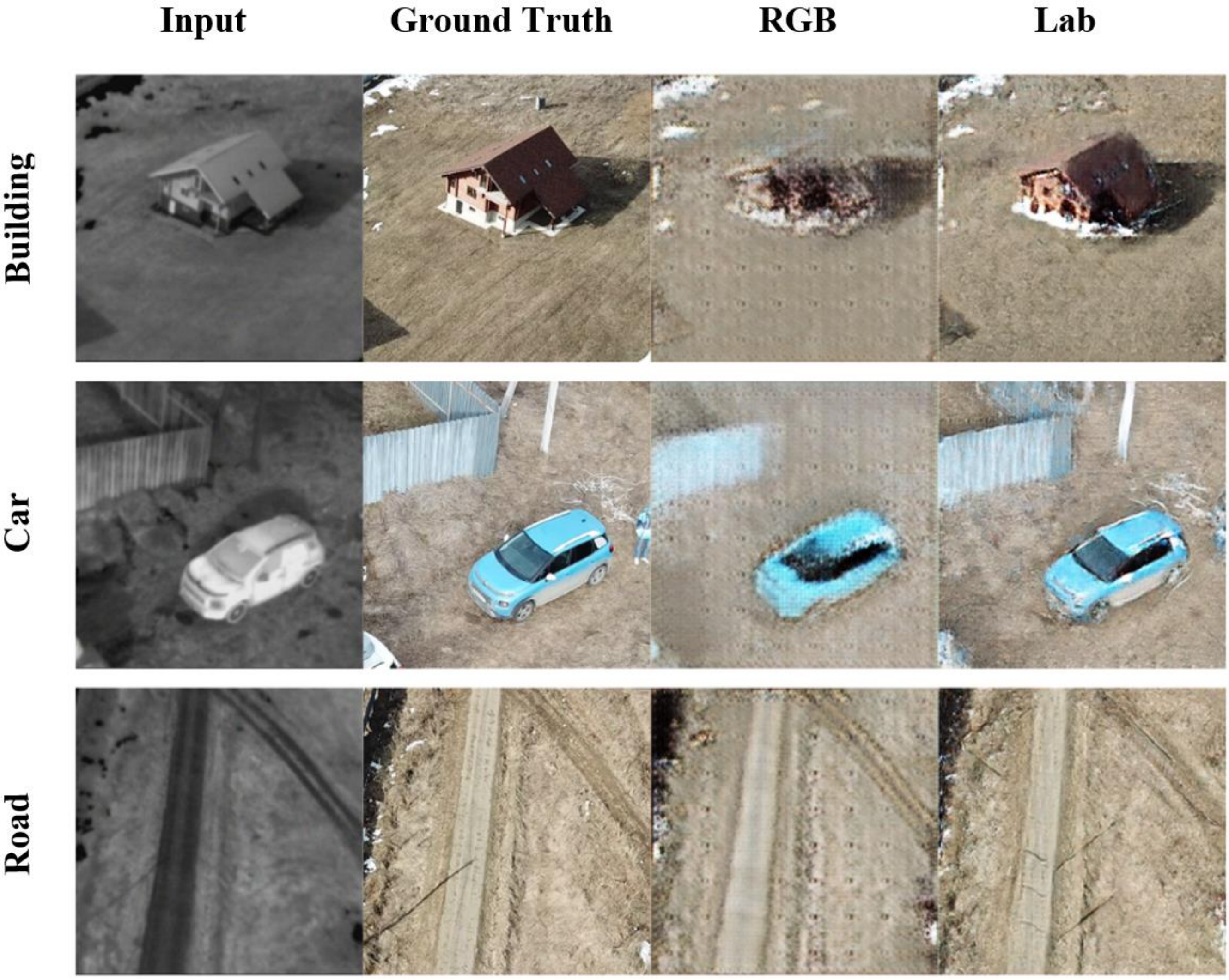


Figure 7: Experimental results for network testing on the LAERT dataset.

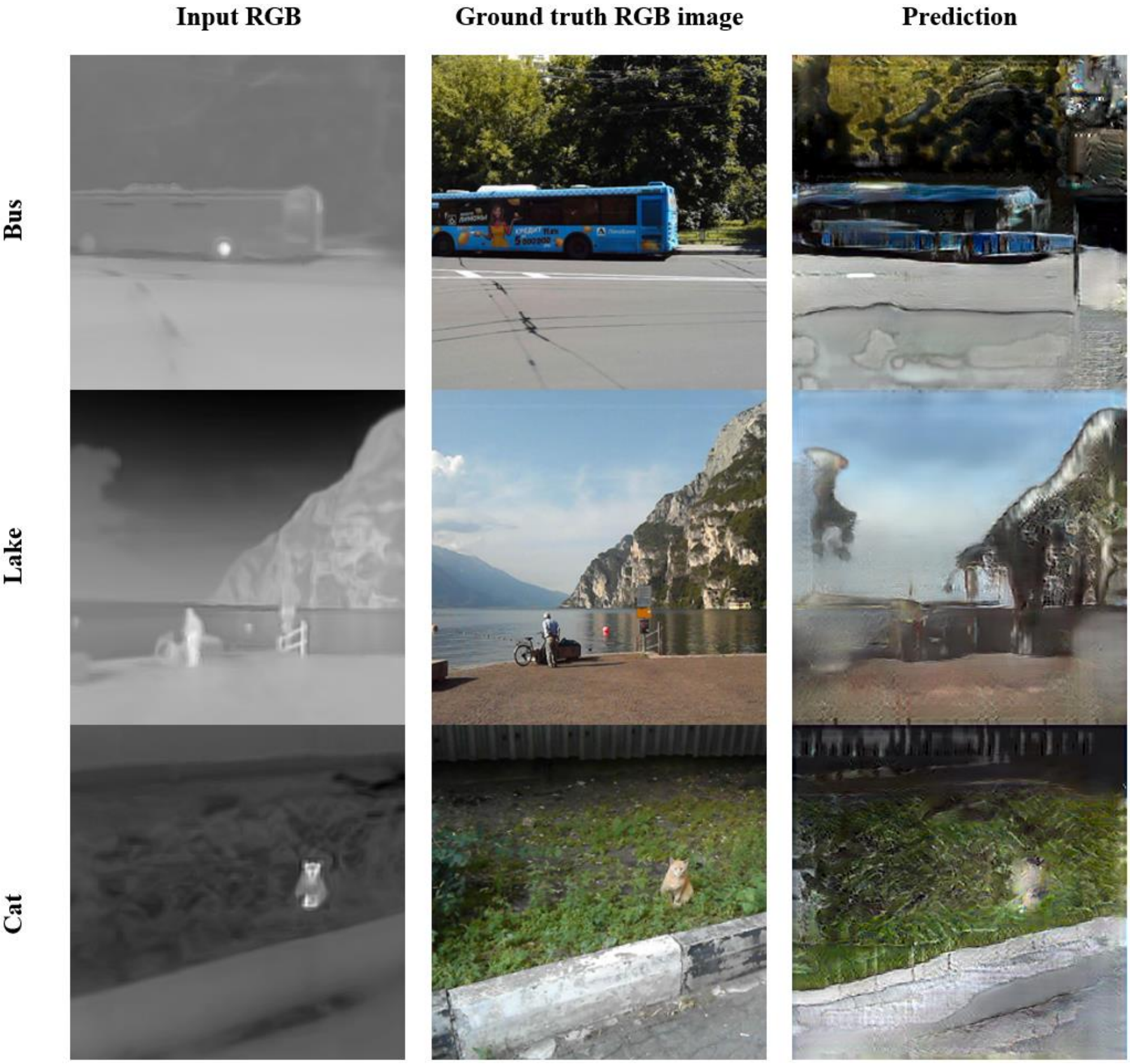


Figure 8: Results of experimental network testing on a ThermalWorld VOC dataset [10].

6. Conclusion

The method of converting images of the far-infrared range into color images of the visible range is considered. The proposed method is based on generative adversarial neural networks. Developed and implemented as a Python script for the PyTorch library is a modification of the ColorMatchGAN network architecture. The proposed modification consists in the transition to the color space Lab to increase the uniform convergence of the learning process. Processing of multispectral training sample LAERT for synchronization and geometrical combination of frames of visible and infrared range is made. A training sample of 4000 frames and an independent test sample of 400 frames were formed.

ACKNOWLEDGEMENTS

The reported study was funded by the Russian Science Foundation (RSF) according to the research project № 19-11-110082.

References

[1] Berg Amanda, Ahlberg Jorgen, Felsberg Michael. Generating Visible Spectrum Images From Thermal Infrared // The IEEE Conference on Computer Vision and Pattern Recognition (CVPR) Workshops. –– 2018. –– June.

[2] Arthur Jarvis J., Norman R. Michael, Kramer Lynda J. et al. Enhanced vision flight deck technology for commercial aircraft low visibility surface operations. –– 2013. –– Access mode: https://doi.org/10.1117/12.2016386 .

[3] Generative adversarial nets / Ian Goodfellow, Jean Pouget-Abadie, Mehdi Mirza et al. // Advances in neural information processing systems. –– 2014. –– P. 2672–2680.

[4] Image-to-Image Translation with Conditional Adversarial Networks / Phillip Isola, Jun-Yan Zhu,Tinghui Zhou, Alexei A Efros // 2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR). –– IEEE, 2017. –– P. 5967– 5976.

[5] Kniaz V. V., Bordodymov A. N. LONG WAVE INFRARED IMAGE COLORIZATION FOR PERSON RE-IDENTIFICATION // ISPRS - International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences . –– 2019. –– Vol. XLII-2/W12. –– P. 111–116. –– Access mode: https://www.int-arch-photogramm-remote-sens-spatial-inf-sci.net/XLII-2-W12/111/2019/ .

[6] Knyaz Vladimir. Multimodal data fusion for object recognition . –– Vol. 110590. –– 2019. –– P. 110590P. –– Access mode: https://doi.org/10.1117/12.2526067 .

[7] Petro Ana Belén, Sbert Catalina, Morel Jean-Michel. Multiscale retinex // Image Processing On Line. –– 2014. –– P. 71–88.

[8] Ronneberger Olaf, Fischer Philipp, Brox Thomas. U-net: Convolutional networks for biomedical image segmentation // International Conference on Medical image computing and computer-assisted intervention / Springer. –– 2015. –– P. 234–241.

[9] The Unreasonable Effectiveness of Deep Features as a Perceptual Metric / Richard Zhang, Phillip Isola, Alexei A Efros et al. // The IEEE Conference on Computer Vision and Pattern Recognition (CVPR). –– 2018. –– Jun.

[10] ThermalGAN: Multimodal Color-to-Thermal Image Translation for Person Re-Identification in Multispectral Dataset / Vladimir V. Kniaz, Vladimir A. Knyaz, Jiří Hladůvka et al. // Computer Vision – ECCV 2018 Workshops. –– Springer International Publishing, 2018.

[11] Vygolov Oleg, Zheltov Sergey. Enhanced, synthetic and combined vision technologies for civil aviation // Computer Vision in Control Systems-2. –– Springer, 2015. –– P. 201–230.