A METHOD FOR SYNTHESIZING THERMAL IMAGES USING GAN MULTI-LAYERED APPROACH

V.A.Mizginov[1,*], V.V.Kniaz[1,2], N.A.Fomin[1]

[1] State Res. Institute of Aviation Systems (GosNIIAS), 125319, 7, Victorenko str., Moscow, Russia
(vl.mizginov, vl.kniaz, nfomin73)@gosniias.ru
[2] Moscow Institute of Physics and Technology (MIPT), Russia

Commission II, WG II/5

KEY WORDS: infrared image, image synthesis, generative adversarial networks, object recognition

ABSTRACT:

The active development of neural network technologies and optoelectronic systems has led to the introduction of computer vision technologies in various fields of science and technology. Deep learning made it possible to solve complex problems that a person had not been able to solve before. The use of multi-spectral optical systems has significantly expanded the field of application of video systems. Tasks such as image recognition, object re-identification, video surveillance require high accuracy, speed and reliability. These qualities are provided by algorithms based on deep convolutional neural networks. However, they require to have large databases of multi-spectral images of various objects to achieve state-of-the-art results. While large and various databases of color images of different objects are widely available in public domain, then similar databases of thermal images are either not available, or they represent a small number of types of objects. The quality of three-dimensional modeling for the thermal imaging spectral range remains at an insufficient level for solving a number of important tasks, which require high precision and reliability. The realistic synthesis of thermal images is especially important due to the complexity and high cost of obtaining real data. This paper is focused on the development of a method for synthesizing thermal imaging images based on generative adversarial neural networks. We developed an algorithm for a multi-spectral image-to-image translation. We have changed to the original GAN architecture and converted the loss function. We presented a new learning approach. For this, we prepared a special training dataset including about 2000 image tensors. The evaluation of the results obtained showed that the proposed method can be used to expand the available databases of thermal images.

1. INTRODUCTION

Nowadays, the use of computer vision is widespread. In different tasks (image recognition, video surveillance, military applications), intelligent systems replace and surpass the results of humans. The ability to use such systems in any visibility and lighting conditions has long attracted various users. Modern computer vision systems are equipped with sensors of various spectra (visible, infrared, radio range). This allows you to significantly expand the scope of their application. Today, methods based on deep convolutional neural networks occupy a leading position in solving problems of human re-identification, image recognition and object detection. Such algorithms are equally well trained on multi-spectral and fusion data. However they require having large databases of multi-spectral images of various objects to achieve state-of-the-art results.

The realistic synthesis of thermal images is especially important due to the complexity and high cost of obtaining real data. The quality of three-dimensional modeling for the thermal imaging spectral range remains at an insufficient level for solving a number of important tasks, which require high precision and reliability. However, modern deep learning algorithms also show excellent results in solving various problems of image-to-image translation. Their use (both with 3D modeling methods and independently) can significantly improve the quality of synthesized thermal images. But only a small amount of multi-spectral data is in the public domain. They contain a small number (hundreds or thousands) of images of some object classes. The goal arises of realistic synthesis of thermal images. The realistic synthesis of thermal images is especially important due to the complexity and high cost of obtaining real data. The quality of three-dimensional modeling for the thermal imaging spectral range remains at an insufficient level for solving a number of important tasks, which require high precision and reliability. However, modern deep learning algorithms also show excellent results in solving various problems of image-to-image translation. Their use (both with 3D modeling methods and independently) can significantly improve the quality of synthesized thermal images.

In this paper, a method for synthesizing thermal imaging images based on generative adversarial neural networks was proposed. We use the paper of Karras (Karras et al., 2018) as a starting point for our research to develop a multi-spectral image-to-image translation. We present the next technical contributions: we have made changes to the base algorithm and presented a new neural network architecture; we developed a new training approach; We made a new dataset to train and test our neural network based on `ThermalWorld` (Kniaz, 2018) . The developed method allows for generating highly realistic thermal images of various objects. Our dataset consists of real RGB- images, thermal images, and semantic segmentation images for several classes: person, car, truck, bus, building, road, trees. We perform an evaluation of our framework on the test split of the dataset. The results of the evaluation showed that our framework generated infrared images that are similar to the ground truth model in both thermal emissivity and geometrical shape. The proposed approach can be used to supplement the existing real infrared image datasets.

*Corresponding author

2. RELATED WORK

2.1 Generative adversarial networks

The emergence of a new type of convolutional neural networks known as generative adversarial neural networks has greatly expanded the understanding of image transformation and generation. For the first time such algorithms were presented in 2014 (Goodfellow et al., 2014). Such a framework can yield specific training algorithms for many kinds of models and optimization algorithm. In this article, the authors explore the special case when the generative model generates samples by passing random noise through a multi-layer perceptron, and the discriminative model is also a multi-layer perceptron. It is referred to in this special case as adversarial nets. The main idea behind GAN is an antagonistic gaming approach (Fig. 1). It consists of two deep convolutional

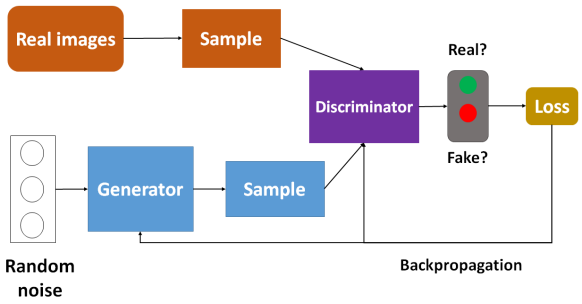

Figure 1. General idea of GAN.

neural networks: a Generator network tries to synthesize an image that visually indistinguishable from a given sample of images in the target domain. A discriminator network tries to distinguish the “fake” images generated by the Generator network from the real image in the target domain. Generator and Discriminator networks are trained simultaneously. The generative model can be thought of as analogous to a team of counterfeiters, trying to produce fake currency and use it without detection, while the discriminative model is analogous to the police, trying to detect the counterfeit currency. Competition in this game drives both teams to improve their methods until the counterfeits are indistinguishable from the genuine articles.

One of the first tasks solved with the help of generative adversarial networks was various image-to-image translation (Zhu et al., 2017),(Isola et al., 2017). In this papers, conditional generative adversarial networks was presented. These networks not only learn the mapping from the input image to the output image but also learn a loss function to train this mapping. This makes it possible to apply the same generic approach to problems that traditionally would require very different loss formulations. This approach is effective at synthesizing photos from label maps, reconstructing objects from edge maps, and colorizing images, among other tasks With the development of GAN architectures, it began to be used for various practical purposes: image denoising (Chen et al., 2018), image super-resolution (Dong et al., 2017), monochrome images colorization (Nazeri, Ng, 2018), (Zbontar, Lecun, 2016), splice detection (Kniaz et al., 2019).

2.2 Multi-spectral image-to-image translation

Recently deep convolutional neural networks began to be applied for image-to-image translation from one spectral range to another. Transformation of the spectral range of an image has been intensively studied in such fields as transfer learning (Paul et al., 2018), cross-domain recognition (Fondje et al., 2020),(Wang et al., 2019),(Kniaz, Bordodymov, 2019). In (Limmer, Lensch, 2016) a methods was proposed for translation of a near-infrared image to a color image.In our previous work, the deep generative-adversarial neural network to automatically convert thermal images to semantically similar color images of the visible range was presented (Kniaz V.V., 2019). The approach was similar to image colorization and style transfer(Ulyanov et al., 2016) problems that were actively studied in recent years. Conversion near-infrared images to visible images without using paired pixel-wise aligned training dataset or rely on a colorful reference image was shown in (Liu et al., 2018). But transformation of LWIR images is more challenging due to the low correlation between the brightness of the color image and a thermal image. If the translation of the infrared image into color is reduced to the problem of colorization, then the inverse transformation is multi-modal. In other words, several synthesized images, the existence of which is physically possible in reality, can correspond to the original input color image. Infrared to visible range image translations were averaged over the entire object, losing the characteristic thermal regions. In our previous works (Kniaz et al., 2016),(Kniaz et al., 2017) we presented the method for transformation of visible range images to infrared images. In (Kniaz, Mizginov, 2018) we proposed a new training method, which extends the traditional GAN training pipeline from the antagonistic game of two players to the game of three players. The third player represents an ”expert” that provides the true negative samples to the discriminator network. In (Kniaz et al., 2018) two-step approach color-to-thermal image translation for person re-identification was developed. One of the last work(Mizginov, Danilov, 2019) presented a method for generation synthetic thermal images using GAN, semantic segmentation and 3D modeling. It showed satisfactory results in translation images of cars.

3. METHOD

As known, generative adversarial networks have achieved significant success in the synthesis of various images from both other images and random noise. High-quality, diverse, and photorealistic images can now be generated by unconditional GANs. However, the synthesis of thermal images from RGB images has a number of difficulties arise: due to the lack of relations between the brightness of objects in thermal images and in color images, unconditional GANs are often unable to accurately predict thermal contrasts in generated images. Moreover, in different conditions, objects can look different in thermal images, while in color images there is no difference between them. Therefore, changing one parameter in the hidden GAN space can cause unwanted changes to other attributes.

The new thermal images generation method is based on our previous research(Kniaz et al., 2018).We have developed two approaches for training convolutional neural networks.In the first of them, we use more input information. The input tensor contains a thermal segmentation image R, a semantic segmentation image with the edge of objects and temperature vector T. Also we used modified neural network StyleGAN in combination with ResNet-18 (He et al., 2016). Our proposed pipeline of the first approach is given in Figure 2. The main idea of the second approach is to train each layer of the generator to synthesize only a certain class of objects in the image. Accordingly, the same number of discriminators evaluate the data synthesized by the generator. The

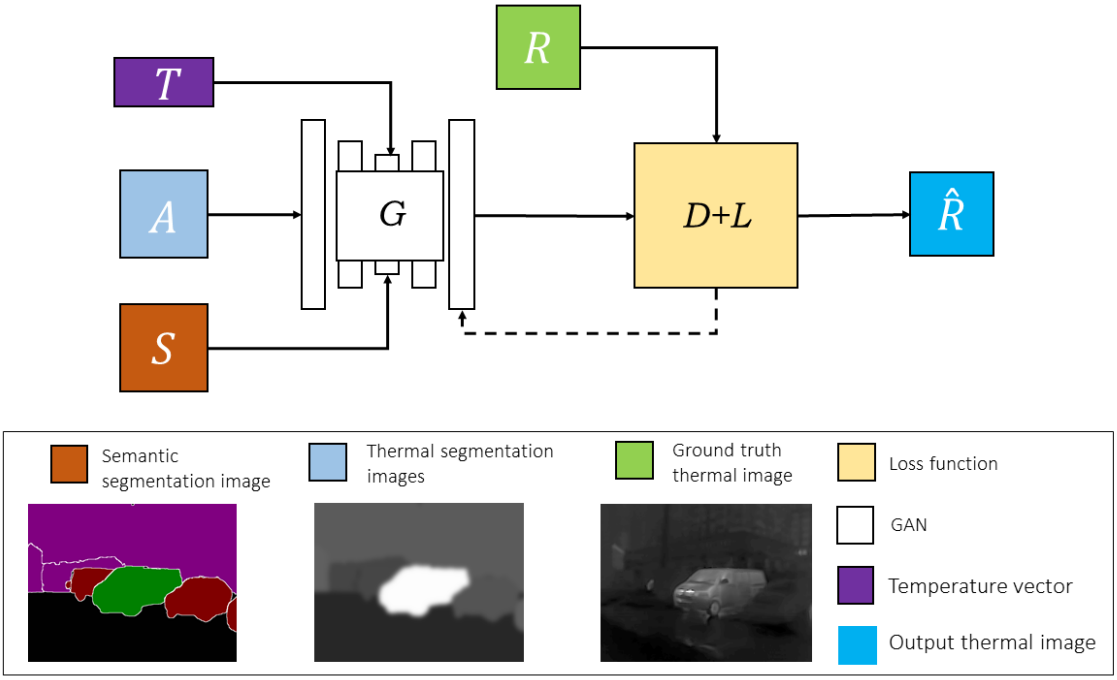

Figure 2. The pipeline of the first approach.

output image is formed by layering the output of each generator. We assume that will allow changing the training parameters for each class of objects without affecting the rest. We use the (Karras et al., 2020) model as the basic GAN architecture. We add a RGB color image A to input tensor from the first approach. It avoids the typical problem of thermal image generation: the brightness is averaged over the entire image and the thermal contrasts of individual objects are lost. Our proposed pipeline of the second approach is given in Figure 3.

3.1 Training Dataset

Our training dataset based on `ThermalWorld` dataset. We made several changes: (1) we transformed all images from the SegmentationClass and SegmentationObject folders to semantic segmentation images with white edges around the each object; (2) we calculated the average temperatures over the area of objects using images from ThermalImages and SegmentationObject folders and saved results as python numpy array. In the first approach, we use a multidimensional tensor, which consists of a semantic segmentation image, a thermal segmentation image, and a temperature vector as input data. This approach allows you to quickly synthesize thermal images from images semantic segmentation images. It does not require a real color images.

For the second approach, we used input data tensor, that consists of a ground true color image,a semantic segmentation images and thermal segmentation image.Moreover semantic segmentation images are divided into several images, each of which displays strictly one of the classes that is present in color and thermal images. An example of one input tensor for the second approach is shown in Figure 4. Such input image tensor makes it possible to feed segmented objects of the target class to the input of each layer of the generator network. This approach allowed the GAN to more accurately predict thermal contrasts of small objects and to avoid brightness averaging over the whole synthesized image. Using a color image in neural network training allows thermal images to be generated directly from real data. All input data are resized by a random number in the range from 380 to 480.For class segmentation S and thermal segmentation R, we subtract the average over the area of each object.We scale and cropp the color segmentation S and the thermal segmentation R to 512x512 pixels.Our dataset includes 6640 semantic segmentation images, RGB color images, thermal segmentation images and ground truth thermal images. The datasets contains such object classes as car, minibus, bus, truck, human and building.

3.2 Network architecture

We propose to use one of the last our development (Kniaz et al., 2020). In contrast with a simple GAN model that reconstructs on image $\hat{R}$ from a latent vector p, a conditional GAN receives an input image A as an additional bit of information. To this end, we propose a new conditional generator architecture inspired by the (Karras et al., 2018, Karras et al., 2019). We hypothesize, that an additional `ResNet-18` encoder could learn a mapping from an input image to the latent code P_T of the style-based decoder. To further increase the convergence of our conditional style-based generator, we add skip connections between intermediate feature maps f_i to noise inputs of the generator. Specifically, we train eight additional pointwise convolutional layers with kernel size 1×1 that translate multichannel feature maps $f_i \in \mathbb{R}^{W_i \times H_i \times C_i}$ to a single channel noise inputs $n_i \in \mathbb{R}^{W_j \times M_j}$. We use a nearest neighborhood interpolation to match the size of feature maps. We modified the architecture in such a way that each generator layer is trained for one specific class of objects. Therefore, we added the same number of discriminators that evaluate the performance of each layer of the generator. We use a residual discriminator similar to (Karras et al., 2019) to provide an adversarial loss function $\mathcal{L}_{adv}$. Two loss functions govern the training process of our conditional generator G: $\mathcal{L}^G_{L_1}$ and $\mathcal{L}^G_{adv}$. The $\mathcal{L}^G_{L_1}$ loss function penalizes the generator for difference in color between the synthesized image $\hat{R}$ and the real image A.

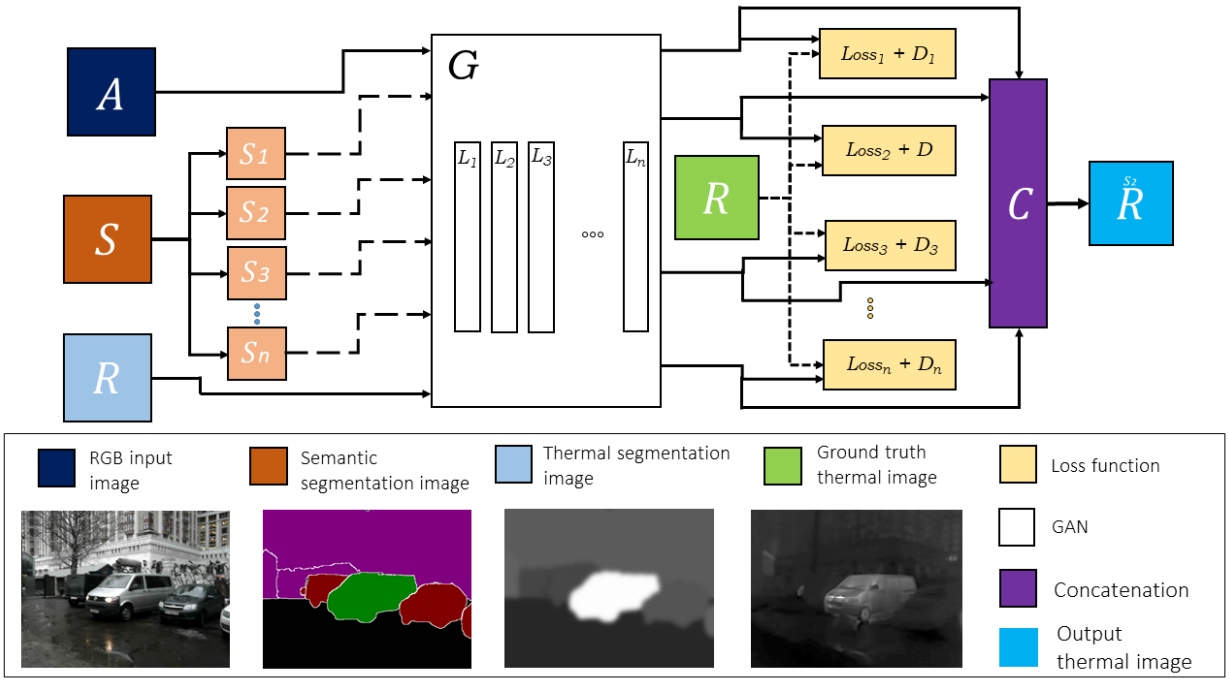

Figure 3. The pipeline of the second approach.

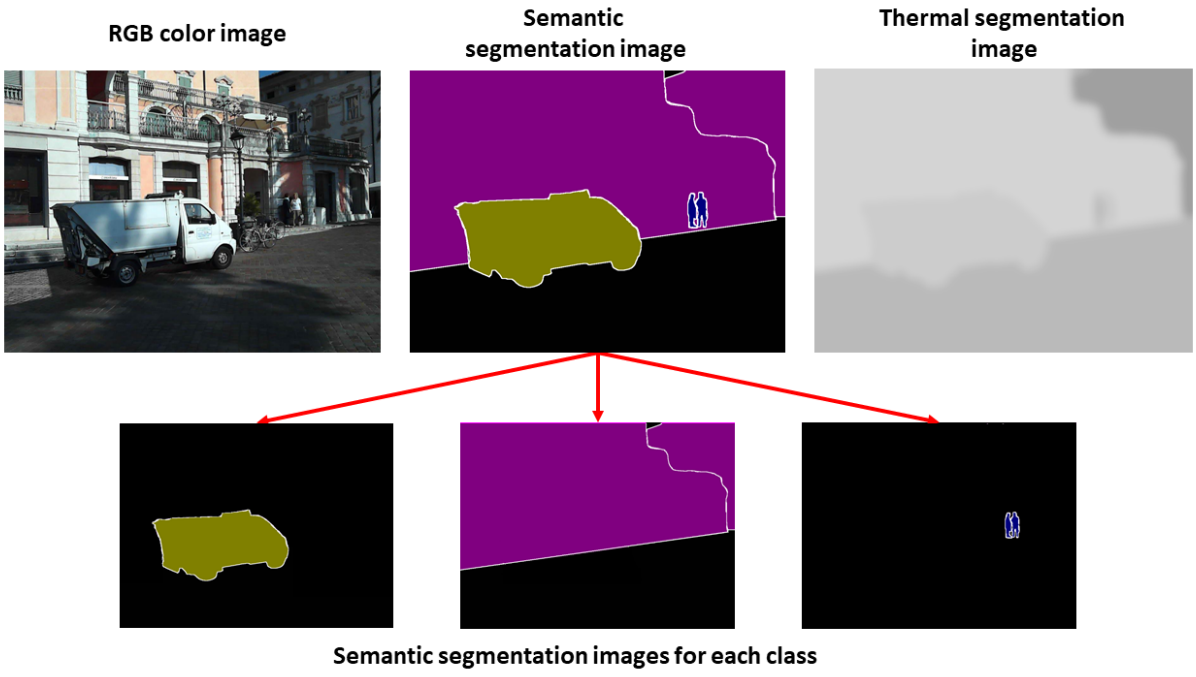

Figure 4. Example of input image tensor for the second approach.

$$\mathcal{L}_1^G(\hat{R}, A) = \mathbb{E}_{A,\hat{R},S}[||R - G(S)||_1]; \quad (1)$$

(WGAN-GP) similar to (Arjovsky et al., 2017)

$$\begin{aligned}\mathcal{L}_{adv}^G(\hat{R}, A) =& \mathbb{E}_{\hat{R}\sim p_G}[D(\hat{R})] - \mathbb{E}_{A\sim p_R}\Big[D(A)\Big] \\ &+\lambda \cdot \mathbb{E}_{\hat{R}\sim p_{\hat{R}}}\Big[(\Big|\Big|\nabla_{\hat{R}} D(\hat{R})\Big|\Big|_2 - 1)^2\Big|\Big|_1\Big],\end{aligned} \quad (2)$$

The adversarial loss is provided by a modified Wasserstein loss where p_G is the distribution of generated samples, P_R is the dis-

tribution of real samples A, and $P_{\hat{R}}$ is the distribution of randomly generated samples. We obtain the final energy to optimize using the weighted sum of two losses

$$\mathcal{L}^G(\hat{R}, A, \nabla_{\hat{R}} D(\hat{R})) = \lambda_{L1} \cdot \mathcal{L}_1^G(\hat{R}, A) + \lambda_{adv} \cdot \mathcal{L}_{adv}^G(\hat{R}, A, \nabla_{\hat{R}} D(\hat{R})); \quad (3)$$

where λ_{L_1}, λ_{adv} are hyperparameters.

4. EXPERIMENTS

4.1 Network Training

We train the generator G synthesized images that were similar in appearance to real images. Each layer of the generator is trained in such a way that it is responsible for synthesizing images of objects of only one class. Also, several discriminators (the same as the classes of objects in the sample) were trained in parallel. The output image is obtained by combining the output of each generator layer. The algorithm was trained on NVIDIA RTX 2080ti GPUs using the PyTorch library and lasted 140 hours.

4.2 GAN Evaluation

We used the independent test dataset to evaluate the GAN. We compare our results with similar methods: $pix2pix$, our last method $ThermalGAN$ and with our method without dividing training into generator layers. The $Qualitative$ evaluation is shown on Fig 5

To $Quantify$, we use the Frechet Inception Distance (Seitzer, 2020). This is a measure of similarity between two datasets of images. It was shown to correlate well with human judgement of visual quality and is most often used to evaluate the quality of samples of Generative Adversarial Networks. FID is calculated by computing the Fréchet distance between two Gaussians fitted to feature representations of the Inception network. The results of evaluation. The results are presented in the table 1.

Method	**Results**
$Pix2pix$	117.082
$ThermalGAN$	93.704
$LayerGAN\ reduced$	75.447
$LayerGAN\ origina$	**34.890**

Table 1. The results of evaluation using FID metric.

The results of assessing the quality of the presented methods using FID showed a significant superiority of the proposed method over existing analogues.

5. CONCLUSION

In this paper, we presented a new method for generation synthetic thermal images using a GAN. We showed that generative adversarial networks leveraging style based architecture became more and more effective for multispectral image-to-image translation. To evaluate the proposed method the Frechet Inception Distance metric was used. The evaluation of the generated infrared images proved that they are similar to the ground truth model in both thermal emission and geometrical shape. This approach surpasses modern baseline methods for image-to-image multispectral translation and our past development. We suppose that the new method will allow for synthesizing realistic thermal images and can be used to supplement the existing training datasets with real infrared images.

REFERENCES

Arjovsky, M., Chintala, S., Bottou, L., 2017. Wasserstein generative adversarial networks. D. Precup, Y. W. Teh (eds), *Proceedings of the 34th International Conference on Machine Learning*, Proceedings of Machine Learning Research, 70, PMLR, International Convention Centre, Sydney, Australia, 214–223.

Chen, J., Chen, J., Chao, H., Yang, M., 2018. Image blind denoising with generative adversarial network based noise modeling. *2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 3155–3164.

Dong, H., Supratak, A., Mai, L., Liu, F., Oehmichen, A., Yu, S., Guo, Y., 2017. TensorLayer: A Versatile Library for Efficient Deep Learning Development. *ACM Multimedia*. http://tensorlayer.org.

Fondje, C. N., Hu, S., Short, N. J., Riggan, B. S., 2020. Cross-domain identification for thermal-to-visible face recognition.

Goodfellow, I. J., Pouget-Abadie, J., Mirza, M., Xu, B., Warde-Farley, D., Ozair, S., Courville, A., Bengio, Y., 2014. Generative adversarial networks.

He, K., Zhang, X., Ren, S., Sun, J., 2016. Deep residual learning for image recognition. *2016 IEEE Conference on Computer Vision and Pattern Recognition, CVPR 2016, Las Vegas, NV, USA, June 27-30, 2016*, 770–778.

Isola, P., Zhu, J.-Y., Zhou, T., Efros, A. A., 2017. Image-to-Image Translation with Conditional Adversarial Networks. *CVPR*.

Karras, T., Laine, S., Aila, T., 2018. A Style-Based Generator Architecture for Generative Adversarial Networks. *CoRR*, abs/1812.04948. http://arxiv.org/abs/1812.04948.

Karras, T., Laine, S., Aittala, M., Hellsten, J., Lehtinen, J., Aila, T., 2019. Analyzing and improving the image quality of stylegan.

Karras, T., Laine, S., Aittala, M., Hellsten, J., Lehtinen, J., Aila, T., 2020. Analyzing and improving the image quality of StyleGAN. *Proc. CVPR*.

Kniaz, V., Gorbatsevich, V., Mizginov, V., 2016. Generation of synthetic infrared images and their visual quality estimation using deep convolutional neural networks. *Scientific Visualization*, 8(4), 67–79.

Kniaz, V. V., 2018. Optical flow-based filtering for effective presentation of the enhanced vision on a HUD. P. Schelkens, T. Ebrahimi, G. Cristóbal (eds), *Optics, Photonics, and Digital Technologies for Imaging Applications V*, 10679, International Society for Optics and Photonics, SPIE, 162 – 171.

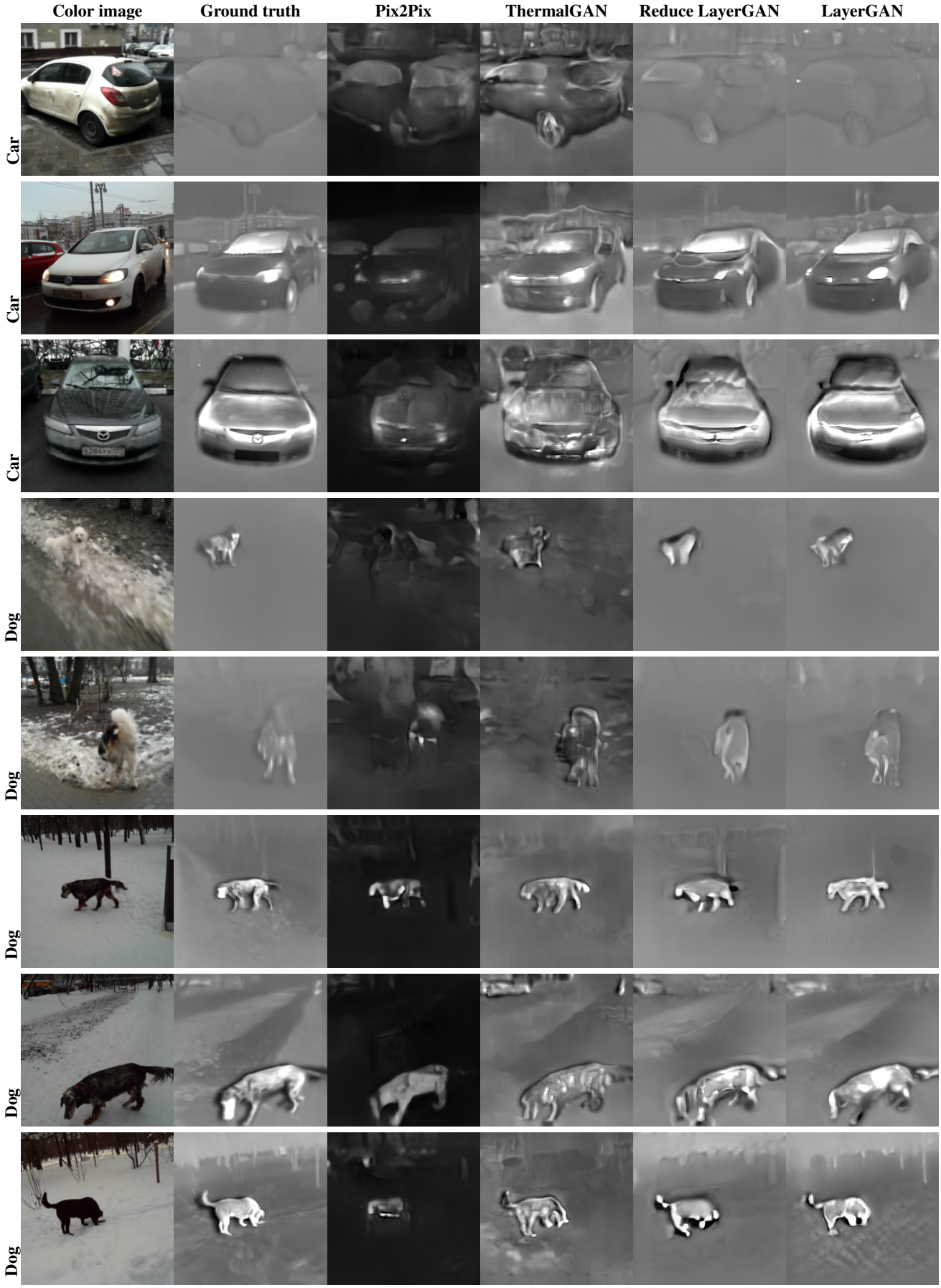
Color image
Ground truth
Pix2Pix
ThermalGAN
Reduce LayerGAN
LayerGAN
Car
Car
Car
Dog
Dog
Dog
Dog
Dog

Figure 5. The Qualitative evaluation different methods.

This contribution has been peer-reviewed.
https://doi.org/10.5194/isprs-archives-XLIV-2-W1-2021-155-2021 | © Author(s) 2021. CC BY 4.0 License.

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

Kniaz, V. V., Gorbatsevich, V. S., Mizginov, V. A., 2017. THERMALNET: A DEEP CONVOLUTIONAL NETWORK FOR SYNTHETIC THERMAL IMAGE GENERATION. *ISPRS - International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, XLII-2/W4, 41–45.

Kniaz, V. V., Knyaz, V. A., Hladůvka, J., Kropatsch, W. G., Mizginov, V. A., 2018. ThermalGAN: Multimodal Color-to-Thermal Image Translation for Person Re-Identification in Multispectral Dataset. *Computer Vision – ECCV 2018 Workshops*, Springer International Publishing.

Kniaz, V. V., Knyaz, V. A., Mizginov, V., Kozyrev, M., Moshkantsev, P., 2020. Structurefromgan: Single image 3d model reconstruction and photorealistic texturing. A. Bartoli, A. Fusiello (eds), *Computer Vision – ECCV 2020 Workshops*, Springer International Publishing, Cham, 595–611.

Kniaz, V. V., Knyaz, V., Remondino, F., 2019. The point where reality meets fantasy: Mixed adversarial generators for image splice detection. H. Wallach, H. Larochelle, A. Beygelzimer, F. d'Alché-Buc, E. Fox, R. Garnett (eds), *Advances in Neural Information Processing Systems*, 32, Curran Associates, Inc.

Kniaz, V. V., Mizginov, V. A., 2018. THERMAL TEXTURE GENERATION AND 3D MODEL RECONSTRUCTION USING SFM AND GAN. *ISPRS - International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, XLII-2, 519–524. https://www.int-arch-photogramm-remote-sens-spatial-inf-sci.net/XLII-2/519/2018/.

Kniaz V.V., Kozyrev M.I., B. A., 2019. Thermal-to-color image translation for visualization on the pilot’s display. *29th International Conference on Computer Graphics, Image Processing and Computer Vision, Visualization Systems and the Virtual Environment GraphiCon’2019*.

Limmer, M., Lensch, H. P. A., 2016. Infrared Colorization Using Deep Convolutional Neural Networks. *CoRR abs/1501.02565*, cs.CV.

Liu, S., John, V., Blasch, E., Liu, Z., Huang, Y., 2018. IR2VI: Enhanced Night Environmental Perception by Unsupervised Thermal Image Translation. *CoRR*, abs/1806.09565. http://arxiv.org/abs/1806.09565.

Mizginov, V. A., Danilov, S. Y., 2019. SYNTHETIC THERMAL BACKGROUND AND OBJECT TEXTURE GENERATION USING GEOMETRIC INFORMATION AND GAN. *ISPRS - International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, XLII-2/W12, 149–154. https://www.int-arch-photogramm-remote-sens-spatial-inf-sci.net/XLII-2-W12/149/2019/.

Nazeri, K., Ng, E., 2018. Image Colorization with Generative Adversarial Networks. *CoRR*, abs/1803.05400. http://arxiv.org/abs/1803.05400.

Paul, A., Vogt, K., Rottensteiner, F., Ostermann, J., Heipke, C., 2018. A COMPARISON OF TWO STRATEGIES FOR AVOIDING NEGATIVE TRANSFER IN DOMAIN ADAPTATION BASED ON LOGISTIC REGRESSION. *The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, XLII-2, 845–852. https://www.int-arch-photogramm-remote-sens-spatial-inf-sci.net/XLII-2/845/2018/.

Seitzer, M., 2020. pytorch-fid: FID Score for PyTorch. `https://github.com/mseitzer/pytorch-fid`. Version 0.1.1.

Ulyanov, D., Lebedev, V., Vedaldi, A., Lempitsky, V. S., 2016. Texture Networks - Feed-forward Synthesis of Textures and Stylized Images. *CoRR abs/1501.02565*, 1603, arXiv:1603.03417.

Wang, G., Zhang, T., Cheng, J., Liu, S., Yang, Y., Hou, Z., 2019. Rgb-infrared cross-modality person re-identification via joint pixel and feature alignment. *The IEEE International Conference on Computer Vision (ICCV)*.

Zbontar, J., Lecun, Y., 2016. Stereo Matching by Training a Convolutional Neural Network to Compare Image Patches. *Journal of Machine Learning Research*.

Zhu, J.-Y., Park, T., Isola, P., Efros, A. A., 2017. Unpaired image-to-image translation using cycle-consistent adversarial networks. *Computer Vision (ICCV), 2017 IEEE International Conference on*.