A method for creating realistic synthetic images using a generative deep learning model for classifying anomalies in panoramas

Arkhipov, P.O.[1[1]], Philippskih, S. L.[1[2]] and Tsukanov M.V. [1[3]]

Oryol branch of the Federal Research Center "Computer Science and Control" of the Russian Academy of Sciences, arpaul@mail.ru
[2]Oryol branch of the Federal Research Center "Computer Science and Control" of the Russian Academy of Sciences, philippsl@mail.ru
[3]Oryol branch of the Federal Research Center "Computer Science and Control" of the Russian Academy of Sciences, tsukanov.m.v@yandex.ru

Abstract. The article describes a method for creating realistic synthetic examples using a generative deep learning model. When generating training examples, the features of the climate, topography, architecture and landscape of a particular area are taken into account. The generative model is represented by a variational autoencoder, consisting of an encoder, a latent space, and a decoder. To train the autoencoder, a new dataset CSCDroneCL_v1 was created, consisting of manually labeled panoramic images and training examples selected from the open dataset VisDrone2022. Based on the SE-SOTA-ConvNet template, a discriminative neural network model was designed and trained to classify anomalies obtained from multi-temporal panoramas. The architectural compatibility of the generative and discriminative models is ensured by using a single design pattern. The unified architecture of neural network models made it possible to apply the transfer learning method. The use of pre-trained parameters in the variational autoencoder model makes it possible to compensate for the small size of the CSCDroneCL_v1 dataset. The generated synthetic panorama training examples have been added to the VisDrone2022 dataset. Using the new dataset, a discriminative model was trained, resulting in an increase in anomaly classification accuracy of 21.2% for the selected identified class.

Keywords: panoramic image, multiclass classification, transfer learning, variational autoencoder, generative deep learning.

1 Introduction

Machine learning is based on the idea that a computer should automatically create a function to transform input data into output data without a predetermined algorithm. The input of the machine learning model receives examples related to the problem being solved. The model analyzes them and independently finds the statistical

structure in the data. This structure allows you to develop rules for the automatic solution of the problem [1].

The main condition for the successful construction of a machine learning model is that the sample must adequately describe the entire space of the original problem to be solved [1]. The initial dataset must be sufficiently complete, maximally cleared of unnecessary noise and unbiased. If these conditions are met, machine learning methods give the best results when solving problems that are difficult to formalize.

2 The problem of classifying anomalies in multi-temporal panoramas

Earlier, an information technology for correcting brightness and color was developed as part of the task of detecting and classifying anomalies in panoramas of different times when creating panoramic images [2]. An anomaly is understood as a certain area of the image, the characteristics of which differ from the predicted ones formed during image processing. In other words, when comparing two panoramas taken at different times, objects can change their geometric or color characteristics. Such objects of interest need to be identified and classified. This information technology includes many successive stages:

- shooting of the studied area with the help of an unmanned aerial vehicle (UAV);
- correction of brightness and color of received frames;
- stitching frames that have undergone brightness and color correction into panoramas;
- comparison of multi-temporal panoramas and detection of anomalies;
- classification of detected anomalies using a neural network model.

At the current stage, the problem of classification of detected anomalies remains not fully solved [34]. Therefore, in order to solve the problem of multiclass classification of anomalies found in panoramic images, it is necessary to pass to the model many examples of images of real objects that have already been classified by people. For this, publicly available datasets are used. Images in such datasets most often do not take into account the features of climate, topography, architecture and landscape (Fig. 1). For example, it is difficult to find images of cars or houses covered in snow [5, 6, 6].

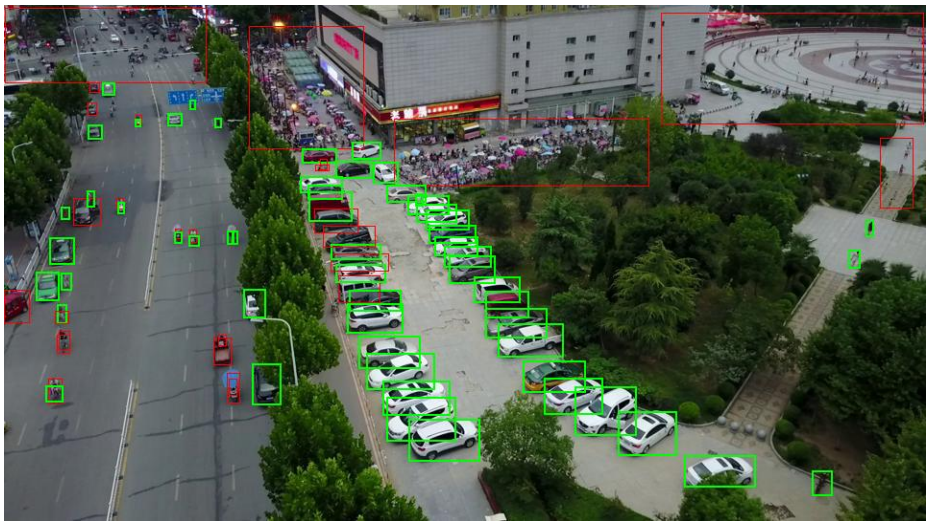

Fig. 1. Images of objects for classification obtained from the VisDrone2022 dataset.

This problem can be solved by creating an extensive dataset that takes into account the peculiarities of the climate and topography of a particular territory, over which photo and video shooting was carried out from the UAV. Preparation of a representative data set will require a lot of effort associated with surveying the area and labeling the resulting images (Fig. 2). It is possible to significantly reduce the amount of labor costs by generating images based on a small number of existing panoramas [7].

Fig. 2. Fragment of a panorama obtained with the help of a UAV and marked manually.

The use of generative deep learning models in the task of classifying anomalies from multi-temporal panoramas greatly simplifies filling in gaps in training samples. It is necessary to manually collect a relatively small number of training examples (base points) over the entire range of feature space, and then the generative model will act as a means of interpolating the missing data between the base points.

3 Generating data with generative deep learning models

Deep learning models are divided into two large classes - discriminative and generative. The discriminant model tries to match the input to the labeled output. A generative model describes how a dataset is generated in terms of a probabilistic model [7]. By sampling from this model, new data can be generated.

Variational autoencoders (VAE) are used to generate new data based on existing images. A variational autoencoder is a deep neural network consisting of two parts: an encoder and a decoder. An encoder is a neural network that compresses high-dimensional input data into a lower-dimensional representation vector. In the case of working with two-dimensional images, the representation vector is a compression of the original image into a latent space of a smaller dimension. A decoder is a neural

network that decompresses a given representation vector back into the original data [8].

Based on the training dataset, the encoder generates a low-dimensional latent space in which each source image is mapped to a multivariate normal distribution around a point in the latent space [9].

In order to generate random data, we need a point μ in the latent space and a standard deviation σ. This data is fed to the input of the decoder and an image is formed from them. It then randomly selects a value of ε from the standard normal distribution.

$$z = \mu + \sigma * \varepsilon \quad (1)$$

where z is a point from the normal distribution;

μ is the mathematical expectation of the normal distribution;

σ is the standard deviation of the normal distribution;

ε is a random value taken from the standard normal distribution.

Each selected point z is input to the decoder. At the output of the decoder, a finished image is formed, which is then used to train the neural network model to classify anomalies [9].

4 Dataset description

The VisDrone2022 dataset [10] was chosen to experiment with machine learning models. To train a discriminative model based on the SE-SOTA-ConvNet template [11, 12], all objects from each frame of the dataset were cut out and six classes wer e formed for training the neural network model (buses, cars, ignored regions, motorcycles, pedestrians and trucks). The entire data set was divided into 3 parts: train (training subset), valid (validation subset) and test (test subset).

For more accurate estimates, the number of images in the <valid> and <test> subsets should be the same (Table 1) and have the same class breakdown as <train> [1].

Table 1. Number of images in train, valid, and test subsets

Subset	Number of images
train	290 207
valid	36 273
test	36 282

To check the quality of the classification of the discriminative model in winter conditions, 811 images of cars obtained by shooting from a UAV and marked manually are used.

To train the generative model, a new dataset was formed, consisting of 811 car images taken in winter and 811 car images from the VisDrone2022 dataset. The new data set has been named CSC DroneCL_v1.

Using the trained generative neural network model, another 10,000 car images were created to train the discriminative model.

5 Model training outcomes

The main idea of the method is the joint use of two neural network models (discriminative and generative) built on the basis of one SE-SOTA-ConvNet design pattern. This makes it possible to significantly facilitate the process of creating and configuring models, generating and validating data. To create a discriminative neural network, you need to specify 3 parameters, to create a generative network – 4 parameters. The data generated by VAE is used to retrain the discriminative network. The quality of the discriminative network is checked using a separate series of real images taken with the help of UAVs.

Based on the SE-SOTA-ConvNet template, the SE-SOTA-ConvNN [12] neural network model was designed and a new variational autoencoder VAENN was developed. A number of experiments were carried out and the best training parameters were chosen. The calculations were carried out on two graphics processing units (GPU) Nvidia A100 in a virtual cloud environment of the hybrid high-performance computing complex (HHPCC) of the Center for Collective Use "High-Performance Computing and Big Data" FRC CSC RAS [13].

To reduce image noise and better optimize when computing on the GPU, the size of the input layer of the neural network was chosen: 32 by 32 pixels (the closest power of two to the median value of 26 pixels in the VisDrone2022 dataset) [10, 14].

The common Adam algorithm, a stochastic gradient descent method based on adaptive moment estimation, was used to train neural networks [15]. The categorical cross entropy was used as a loss function [1]. Before the main training neural network models achieved numerical stability [11].

SE-SOTA-ConvNN neural network parameters: number of dense groups – 4, network width – 64 neurons, compression ratio – 0.8. There are 12 convolutional layers in the network. There are 801,980 parameters in total, of which 797,980 are trainable.

The overall classification accuracy of SE-SOTA-ConvNN on a test subset of the VisDrone2022 dataset for all classes is 92.2%; classification accuracy for the class "Cars" - 96.1%; The accuracy of classifying images of the “Cars” class on the CSC DroneCL_v1 dataset was 71.2%.

The SE-SOTA-ConvNet template was used as the basis for constructing the variational autoencoder VAENN. The encoder consists of four dense groups with a width of 64 neurons and a compression ratio of 0.8. The decoder is a mirror image of the encoder neural network, in which all convolutional layers are replaced by transposed convolutional layers. With the VAENN autoencoder, input and output

images should be as similar as possible. For this reason, a symmetric autoencoder is used.

The encoder and decoder of the variational encoder repeat the architecture of the discriminative network. This made it possible to apply the transfer learning method (the already trained weights of the SE-SOTA-ConvNN neural network model were transferred to the VAENN model) when creating a generative model [1, 14]. The use of pre-trained parameters in the VAENN model made it possible to compensate for the small size of the CSC DroneCL_v1 dataset. Thus, there are 24 convolutional layers in the autoencoder. The total number of network parameters is 2,243,960, of which 2,235,960 are trainable.

After generating 10,000 training examples of the “Cars” class by the variational autoencoder, the SE-SOTA-ConvNN neural network model was retrained. The accuracy of the classification of images of the “Cars” class in the CSC DroneCL_v1 dataset increased to 92.4% after additional training of the model (Table 2).

Table 2. Accuracy of image classification before and after retraining of a discriminative neural network

Dataset	Accuracy before retraining of the network (class "Cars"), %	Accuracy after retraining of the network (class "Cars"), %
VisDrone2022	96.1	96.1
CSC DroneCL_v1	71.2	92.4

6 Conclusions

The developed variational autoencoder - VAENN allows generating high-quality synthetic training examples based on a small number of images taken as a result of real UAV flights. The accuracy of anomaly classification using the SE-SOTA-ConvNN neural network model increased by 21.2%.

However, this method has a number of features:

- due to the use of transfer learning, the size of each variational autoencoder is several times larger than the size of the classifier neural network;
- experiments have shown that it is more convenient to create a separate autoencoder for similar classes of images. This allows you to select the dimension of the latent space to improve the quality of the generated images and remove unnecessary noise. In addition, the small number of classes in the latent space makes it easier to work with the variational autoencoder. The final generative neural network model consists of an ensemble of simple autoencoders. The process of creating and training a large number of neural networks is automated by using a single SE-SOTA-ConvNet design pattern;

– the quality of synthetic examples is assessed for the entire set of generated images. There is no metric that can evaluate the quality of a single image created by a variational autoencoder.

Further refinement of the method involves increasing the degree of automation of the process of generating synthetic training data.

References

1. Chollet, F.: Deep Learning with Python. 2nd edn. Manning Publications Co., Shelter Island, NY (2021).
2. Arkhipov, P.O., Tsukanov, M.V.: Incompatimic model of anomaly detection on different panoramas. Highly Available Systems 17(2), 5–10 (2021).
3. Arkhipov, P.O., Philippskih, S.L., Tsukanov, M.V.: Development of a new model of step convolutional neural network for classification of anomalies on panoramas. Highly Available Systems 17(1), 50–56 (2023).
4. Mokayed, H. et al. Nordic Vehicle Dataset (NVD): Performance of vehicle detectors using newly captured NVD from UAV in different snowy weather conditions. 27 Apr 2023. arXiv:2304.14466v1 [cs.CV].
5. Ye, T. et al. Towards Real-time High-Definition Image Snow Removal: Efficient Pyramid Network with Asymmetrical Encoder-decoder Architecture. 12 Jul 2022. arXiv:2207.05605v1 [cs.CV].
6. Cheng, B., Li, J. at el. Snow Mask Guided Adaptive Residual Network for Image Snow Removal. 11 Jul 2022. arXiv:2207.04754v1 [cs.CV].
7. Foster, D.: Generative Deep Learning. O'Reilly Media, Inc., Sebastopol, CA (2019).
8. Kingma, D., Welling, M.: Auto-Encoding Variational Baye. Machine Learning Group Universiteit van Amsterdam, 10 Dec 2022. arXiv:1312.6114v11 [stat.ML].
9. Hou, X. et al.: Deep Feature Consistent Variational Autoencoder. University of Nottingham, 2 October 2016. arXiv:1610.00291v1 [cs.CV].
10. Zhu, P., Wen, L., Du, D. et al.: Detection and Tracking Meet Drones Challenge. Cornell University, 4 Oct 2021. arXiv: 2001.06303v3 [cs.CV].
11. Ferlitsch, A.: Deep Learning Patterns and Practices. Manning Publications Co., Shelter Island, NY (2021).
12. Philippskikh, S.L.: Classification of images extracted from panoramas using a neural network with a squeeze-excitation module. Intelligent Data Processing: Theory and Applications: Book of abstract of the 14th International Conference, pp. 204-209. Russian Academy of Sciences, Moscow (2022).
13. TsKP "INFORMATIKA" Homepage, https://www.frccsc.ru/ckp, last accessed 2023/05/04.
14. Arkhipov, P. O., Philippskih, S. L.: Building an ensemble of convolutional neural networks for classifying panoramic images. Pattern Recognition and Image Analysis 32(3), 511–514 (2022).
15. Kingma, D., Adam, J.B.: A Method for Stochastic Optimization. Cornell University, 22 Dec 2014. arXiv:1412.6980 [cs.LG].