National Research University Higher School of Economics

as a manuscript

Shakhuro Vladislav

Rare traffic sign recognition using synthetic training data

PhD Dissertation Summary

for the purpose of obtaining academic degree
Doctor of Philosophy in Computer Science

Moscow — 2020

This work was prepared at National Research University Higher School of Economics.

Academic Supervisor: Anton S. Konushin, PhD, Asociate Professor, National Research University Higher School of Economics

Subject of the work

Object recognition on images is one of the basic tasks in computer vision. Such task is usually solved using machine learning. Machine learning methods often require big training samples. Recognition quality of modern machine learning learning algorithms depends heavily on size of training sample. For instance, in [1] is shown that usage of training sample that is 4.4 images in total for training a deep learning model allows to achieve human quality of face recognition on LFW sample [2].

Usage of synthetic training samples allows to solve following problems:

- complexity of obtaining large training samples (that are necessary for training deep neural networks);
- labour intensity of labelling samples;
- the need to recreate training sample in case of domain change;
- obtaining samples of rare object classes;
- obtaining qualitative models using complex synthetic training data.

This dissertation researches methods for generating synthetic training samples for the traffic sign recognition task. Traffic sign is a flat synthetic object with standard appearance. Automated computer generation of synthetic images of traffic signs allows to obtain big training samples quickly. Traffic sign recognition task can be solved in two stages: detection and classification. During the detection stage all traffic signs are selected with bounding boxes, during the second stage all selected traffic signs are classified: class label from a fixed set of labels is assigned to each sign. Traffic sign recognition is an important component in two applications.

First application is self-driving car control. Self-driving cars are cars that are capable of driving safely on public roads without human assistance. Key component of self-driving car control is object detection. Pedestrians, cars, traffic lights and traffic sign are the main objects of interest.

Second application of traffic sign recognition is automated mapping using car dashboard cameras. Such application is important because nowadays creating and maintaining maps up-to-date requires either large financial resources (if a commercial solution is used), either lots of human annotators' time (if crowdsourcing is used for labelling data).

Traffic sign detector used in such applications must fulfill following requirements:

1. Detect wide range of traffic sign casses, including signs that are very rare in training samples. Examples of such sign classes are shown in fig. 1. Existing works usually recognize restricted set of classes which have enough samples in training data. At the same time number of traffic signs is an order of magnitude larger.
2. Be tolerant to different lighting and weather conditions: dark time, snow, rain, etc.

Figure 1 — Examples of rare russian traffic sign classes

3. Satisfy recall (near 100%) and precision (1 false positive per 1 minute of video stream, i.e. 90%) requirements.

Traffic sign recognition system has to be trained on representative sample to fulfill aforesaid requirements. Synthetic training samples can be used instead of costly labelling of large sample.

Synthetic training generation task is actively researched in computer vision. Despite that there are few methods for synthetic traffic sign generation. Several methods [3—5] use traffic sign icons to generate synthetic samples using empirical transformations (rotation, shift, blur, change of contrast and colorness). Experimental evaluation of simple synthetic data generation method shows that adding synthetic image to real images enhances classification quality. However, current methods don't improve quality of traffic sign detector and don't allow to training traffic sign classifier only on synthetic samples with acceptable quality.

A promising research direction in sythetic training data generation nowadays is usage of generative neural networks. A method called generative adversarial networks (GAN) was proposed in [6]. The key idea of method is training a neural network to assess quality of generated image instead of using an analytical function. So analytical function is replaced with model that is trained on real and synthetic data. Such method dramatically improved the quality of synthetic images. Adversarial neural networks nowadays are successfully used for generating realistically looking synthetic images [7], domain translation [8] (for instance, translating satellite images to maps, image stylization). It was shown that usage of synthetic data generated with GAN allows to improve human recognition [9].

The goal of dissertation is to improve the quality of traffic sign recognition using synthetic training data.

Several **objectives** are set to achieve that goal:

1. Collect and label a sample of frames with wide range of traffic sign classes with full annotation of occuring classes of signs. The sample must be suitable for evaluation of rare traffic sign recognition.
2. Research applicability of modern generative neural networks for generating training samples for traffic sign classification.
3. Develop a method for conditional generation of training samples for traffic sign classification.

4. Develop a method for improving visual realism of synthetic images of rare traffic signs.
5. Develop a method for rare traffic sign classification.
6. Develop a method for generation of training samples for rare traffic sign detection.

Main results of the dissertation:

1. A samples of russian traffic signs is collected and labelled (Russian Traffic Sign Dataset, RTSD). This dataset has more frames and traffic sign classes compared to other public datasets. The dataset contains frames captured in different lighting, weather conditions and seasons. The sample contains 205 traffic sign classes, out of which 99 classes are rare (they are contained only in the test split of the dataset).
2. Generative adversarial networks are researched in application to traffic sign generation task. Addition of synthetic training samples generated using neural networks to the real data improves quality of traffic sign classification. A method for training conditional Wasserstein GAN is proposed.
3. A method for improving synthetic images of traffic signs is proposed. Training data obtained using that methods improves accuracy of traffic sign classification.
4. A new method for traffic sign classification is developed. Such method allows to classify either frequent (classes that are present in training and testing samples), or rare (classes that are present only in testing sample) classes of traffic signs. The method is trained on real and synthetic training samples.
5. A method for generating sythetic training samples for traffic sign detection is developed. That proposed methods improves quality of rare traffic sign detection.

Novelty of the results:

1. The task of traffic sign recognition with large number of classes is investigated for the first time with sufficienlty large sample.
2. The work investigates applicability of generative adversarial networks for generating synthetic traffic signs for the first time. Researched methods are evaluated of traffic sign classification task.
3. A new method for traffic sign classification is proposed. Such method preserves quality of frequent sign classification and improves quality of rare traffic sign classification.
4. A new method for generating training samples for traffic sign detector is proposed. The method improves accuracy of rare traffic sign detection.

Publications. Main results of the dissertation are published in 5 periodicals that are indexed in Scopus database.

First-tier publications:

1. Shakhuro, V. Image synthesis with neural networks for traffic sign classification / V. I. Shakhuro, A. S. Konouchine // Computer Optics. — 2018. — vol. 42, № 1. — p. 105—112. (Scopus, Q2).
2. Shakhuro V. Classification of rare traffic signs / B. V. Faizov, V. I. Shakhuro, V. V. Sanzharov, A. S. Konouchine // Computer Optics. — 2020. — vol. 44, № 2. — p. 236—243. (Scopus, Q2).

Second-tier publications:

3. Shakhuro, V. Russian traffic sign images dataset / V. I. Shakhuro, A. S. Konushin // Computer Optics. — 2016. — vol. 40, № 2. — p. 294—300. (Scopus, Q3).

Other publications:

4. Shakhuro, V. Rare Traffic Sign Recognition Using Synthetic Training Data / V. Shakhuro, B. Faizov, A. Konushin // Proceedings of the 3rd International Conference on Video and Image Processing. — Shanghai, China : Association for Computing Machinery, 2019. — p. 23—26. (Scopus).
5. Shakhuro, V. Generation of synthetic traffic sign images using conditional generative adversarial networks / P. Хрушков, V. Shakhuro, A. Konushin // Graphicon. — 2018. — p. 242—246. (Scopus)

Author conducted main theoretical and practical research that is stated in dissertation. In [1,3] research advisor A.S.Konushin possesses problem statement, V.I.Shakhuro possesses all obtained results. In [4] research advisor A.S.Konushin possesses problem statement, V.I.Shakhuro possesses all obtained results. Contribution of B.V.Faizov is technical help with traffic sign classification. In [2] A.S.Konushin possesses problem statement, V.I.Shakhuro — all results. Contribution of B.V.Faizov is technical help with implementation of the proposed rare traffic sign classification method. Contribution of V.V.Sanzharov is technical implementation of ray tracing for generating synthetic traffic signs. In [5] research advisor A.S.Konushin possesses problem statement, V.I.Shakhuro possesses all obtained results. Contribution of P.V.Khrushkov is help with technical implementation of the proposed conditional generative neural network.

Approbation of the research.

The results of the dissertation are reported at the following conferences and workshops:

- Internation Conference on Video and Image Processing (ICVIP) 2019, China, Shanghai, December 22-24, 2019;
- International Conference on Computer Graphics, Image Processing and Machine Vision, GraphiCon 2018, Tomsk, Russia, September 24-27, 2018;
- Computer Vision research seminar at Faculty of Computational Mathematics and Cybernetics, Moscow State University, Russia;

– PhD Computer Science research seminar at Higher School of Economics, Moscow, Russia;
– Microsoft Research PhD Summer School, UK, Cambridge, 2015.

Content of the work

The **introduction** describes relevance of research that is conducted in dissertation, formulates goal and objectives of the work, novelty and practical importance of the work.

The **first chapter** reviews existing methods for generating synthetic training data. Disadvantages of existing methods are described that justify significance of this dissertation.

The **second chapter** describes Russian Traffic Sign Dataset (RTSD) that is collected and labelled in this work. Review of existing public traffic sign dataset in the time of research (2015) show that there is no available dataset that is suitable for training traffic sign recognition system (detector and classifier) with large number of classes.

RTSD dataset consists of frames provided by Geocenter-Consulting company[1]. Frames are obtained using HD or FullHD dashboard cameras. Cameras take photos at rate 5 fps. Frames are captured at different seasons (winter, spring, autumn), time of day (morning, afternoon, evening) and weather (rain, snow, bright sun). Sample frames are shown in fig. 2.

There are several samples for testing traffic sign recognition algorithms. Samples contain groups of classes called «prescription» (blue circles), «prohibition» (red triangles), «restriction» (circles with red border), «main road» (yellow rhombus), «service» (rectangles with blue border), «special regulation» (blue rectangle). General appearance of traffic signs is shown in fig. 3.

Piotr Dollar toolbox [10] is used for implementation of cascaded traffic sign detector using integral channel features. Detector is trained using parameters from [11]: 10 channel for feature computation (pixel color in LUV color space, gradient magnitude, six gradient orientations), cascade of 400 decision stumps of depth 2 trained in 4 iterations with bootstrapping (2000 negative samples per iteration), {50, 100, 200, 400} decision stumps are trained in iterations accordingly. Multiscale traffic sign detection (sign size from 16×16 to 128×128 pixels) is conducted using pyramid that consists of 50 image scales. A classifier is trained for every group of sign classes that uses images of size 56×56 pixels.

Area Under ROC Curve (AUC) metric is used to evaluate quality of detectors. That metric is widely used for assessing quality of traffic sign detection [11; 12].

Traffic sign detectors trained of different RTSD samples showed insufficient quality for applications (near 100% recall and 90% precision). Only

[1] http://geocenter-consulting.ru

Figure 2 — Sample frames from RTSD that demonstrate different seasons, weather and lighting conditions.

Figure 3 — General appearance of traffic signs in RTSD samples

detectors trained for narrow group of traffic sign classes (main road and red triangles) obtained required accuracy.

Convolutional neural network [13] is used for traffic sign classification. It obtained 98% accuracy on German Traffic Sign Recognition Dataset (GTSRB) [14].

RTSD was specially divided into train and test split that is suitable for rare traffic sign recognition task. A special procedure that uses all available sign classes was implemented.

Histogram of image distribution per class in train and test samples is shown in fig. 4. Note that division procedure approximately splits the dataset in 4 to 1 ratio with limit on minimal number of samples in train split and guarantees that all images of one physic sign will go to either train or test part. Such split can be used for training and evaluating frequent and rare traffic sign recognition algorithms.

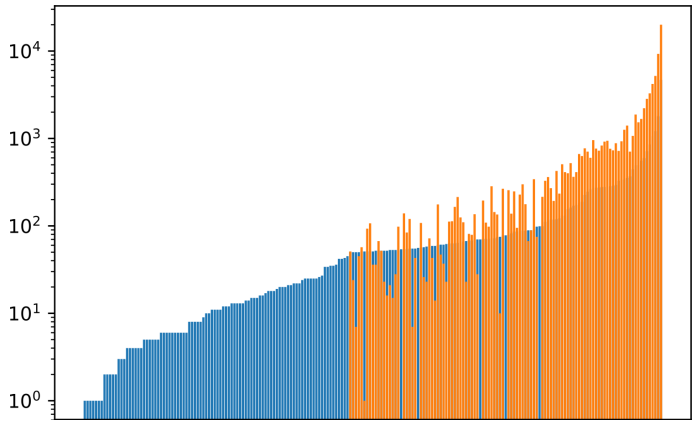

Figure 4 — Number of images per sign class in train (orange columns) and test (blue columns) parts of RTSD. Classes are sorted in ascending order of image number in test part.

In conclusion in this chapter the Russian Traffic Sign Dataset (RTSD) was presented. This dataset exceed other public traffic sign datasets in number of frames, sign classes, physical signs and their images. Aside of that the dataset contains frames with different weather and lighting conditions and seasons. Object detection algorithm based on soft cascade and gradient features and convolutional neural network classifier were evaluated on the dataset. Evaluation showed that current detection and classification methods achieve insufficient quality for applications.

The third chapter is devoted to generation of synthetic training data for traffic sign classification. Three methods for traffic sign generation are

considered: direct image generation from noise, conditional image generation from noise and synthetic traffic sign image processing.

Unconditional traffic sign image generation. A sample of real images of the same size $H \times W \times C$ from the distribution $p_r(x)$ of real images is given. Here H, W, C are height, width and number of color channels of the image. The goal is to train neural network $g_\theta(z)$ that obtains multivariate noise $z \sim p(z)$ (for instance, gaussian) and transforms that noise into images of size $H \times W \times C$ that are similar to real image. Wasserstein metric can be used to metric the realism of images generated using neural network [15]:

$$W(p_r, p_g) = \max_{w \in \mathcal{W}} \mathsf{E}_{x \sim p_r(x)} f_w(x) - \mathsf{E}_{z \sim p(z)} f_w(g_\theta(z)).$$

Two neural networks are trained alternately: critic f_w function that is used for computing Wasserstein metric and generator g_θ function. Weights of the critic are clipped after each train step to guarantee that f_w is Lipschitz function with bounded constant (that is needed for Wasserstein metric).

Generator neural network obtains multivariate noise at input and outputs random image of traffic sign. At the same time generator must be able to generate images of required traffic sign classes. In other words, distribution of output images must be conditional, i.e. depend on class label c. In work [16] generative neural networks are trained to sample from conditional distributions. Generator obtains traffic sign class label at input in addition to noise. Classes is coded using binary vector coded with one-hot method: one element of the vector is 1, others are 0.

Unfortunately such approach is unsuitable for Wasserstein metric since that loss is different from usual loss of generative adversarial networks. In practice, neural network trained with Wasserstein metric learns to ignore class label. N neural networks were trained instead of single network, where N is the number of sign classes. Every neural network is trained on narrow class of traffic signs. As a result, quality of generated samples became higher.

Generator uses DCGAN architecture [7]. It uses transposed convolutions to increase the resolution of 100-component noise vector to 3-channel image of size 64×64 pixels. Neural network classifier uses input resolution 48×48 pixels.

Generation using neural network was compared to simple synthetic traffic generation from icons [17]. Several transformations are applied to icon:

- gaussian blur;
- additive gaussian noise;
- color change in HSV color space;
- motion blur;
- placing transformed icon on background cropped from dashcam frames.

Sample images generated using single class neural network generator are shown in fig. 5. Note that generative neural networks trained with Wasserstein

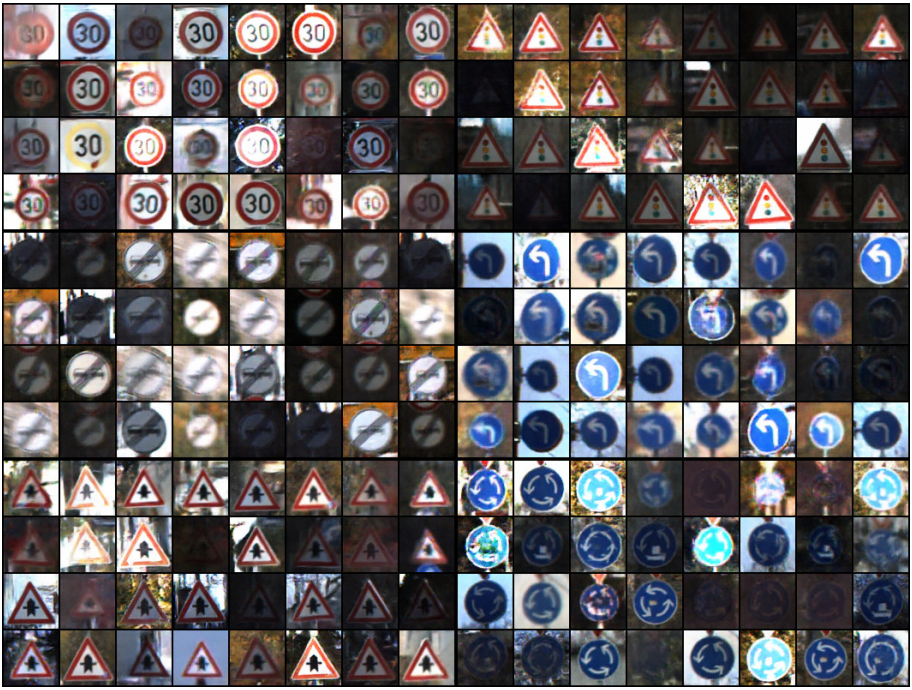

Figure 5 — Sample images generated using per class trained neural network generators.

metric produce photorealistic images that are undistinguishable from real images. At the same time synthetic images generated from sign icon substantially differ from real images. More sophisticated transformation has to be applied to obtain more realistic synthetic images. Sample synthetic images obtained from icons are shown in fig. 6.

Experimental evaluation of synthetic training data showed that:

1. Classifier trained on neural network synthetic data achieves 0.3% lower accuracy than classifier trained on real data. Note that larger synthetic training sample has to be used to achieve comparable quality.
2. Synthetic images can be used in addition to real data. But it's an ineffective method compared to simple image augmentation using rotation, shift and scaling. Such simple augmentation technique allows to achieve higher classification accuracy.
3. Synthetic images obtained using neural network are substantially more realistic than simple synthetic images generated from icon. Accuracy of classifier trained on complex synthetic data is higher (53.7% vs 96.1% without augmentation and 69.7% vs 98.1% with augmentation).

Figure 6 — Sample synthetic images generated from sign icons.

Conditional generation of traffic sign images In this section a new method for training conditional generative networks using Wasserstein metric is proposed. Critic additionally trains to classify generated images. The following cross-entropy term is used for training:

$$L_C = \mathsf{E}_{x \sim p_r(x)}[\log p(c|x)] + \mathsf{E}_{z \sim p(z)}[\log p(c|g_\theta(z))]$$

Here c is real or synthetic image class label.

Critic function f_w must be Lipschitz with constant 1 for training neural networks (see more in [15]). In [18] additional term in loss function is used to restrict critic neural network:

$$L_R = \lambda(||\nabla_{\hat{x}} f_w(\hat{x})|| - 1)^2.$$

Here $\hat{x} = tx + (1-t)g_\theta(z)$, $z \sim p(z)$, $t \sim U[0;1]$ is convex combination of real and synthetic images. Loss function L that is used for training generator is a weighted sum of $W(p_r, p_g)$, L_C and L_R.

L_R term contains convex combination $\hat{x}$ of synthetic and real image. Note that it's meaningless to compute convex combination of two images of different classes. Training procedure is the following:

1. Sample a minibatch of real images.
2. Generate synthetic images with the same labels as images in minibatch. So every real image has corresponding synthetic image with the same class label.

Notice that every generator training iteration consume class labels for generating synthetic images. Class labels should be sampled with the same probability as they occur in training sample. Consider full probability of a real image $p_r(x)$:

$$p_r(x) = p(c_1)p_r(x|c_1) + \cdots + p(c_k)p_r(x|c_k).$$

Even if the generator synthesizes images conditioned on class label well (i.e. qualitatively estimated conditional distributions $p_r(x|c)$), but the class labels are sampled with probabilities $p(c_1), \ldots, p(c_k)$, then the result distribution differs from $p_r(x)$ and loss function penalizes generator for the «incorrect» distribution.

Consider that a label c_i occurs more often than a label c_j. Then c_i probability measure is weighted in $p_r(x)$ more than c_j probability measure. Therefore generator trains to generate images with label c_i with higher quality than images with label c_j. In theory WGAN is able to ideally estimate real distribution, but generator has restricted resources and can't keep too much information. For this reason if the WGAN has an alternative: images of which class can be generated with lower quality, then this class will be c_j, not c_i.

A reweighted distribution of real images can be used to cope with that problem:

$$p'_r(x) = \frac{p_r(x|c_1) + \cdots + p_r(x|c_k)}{k}.$$

Note that

$$\mathsf{E}_{x\sim p'_r(x)}[f(x)] = \mathsf{E}_{x,c\sim p_r(x,c)}\left[\frac{f(x)}{kp(c)}\right].$$

Then to obtain the balanced version of real images distribution every training sample with label c is weighted with $\frac{1}{kp(c)}$ coefficient.

Sample images generated using conditional WGAN and auxiliary classifier are show in fig. 7. Many samples are visually indistinguishable from real images, but some classes mix with each other (for instance, speed limit 20 and «watch out, children», speed limits 100 and 120).

Quantitative evaluation was carried using neural network traffic sign classifier. Classifier accuracy on different training samples is shown in table 1. Conditional WGAN generates more realistic images than simple generation using sign icon, but worse than 43 neural networks trained to generate per-class images.

Postprocessing of synthetic traffic sign images. In this section a method for postprocessing synthetic traffic sign images is proposed. The method transforms synthetic images from CGI collection into more realistic. CGI dataset is obtained using ray tracing rendering and is described in chapter 4. Postprocessing neural network is trained using auxiliary generator network and cyclic loss function from [8]. Two neural networks are used during training for two image domains A (synthetic traffic signs) and B (real traffic signs). First generator network transforms images from domain A to domain B (from synthetic

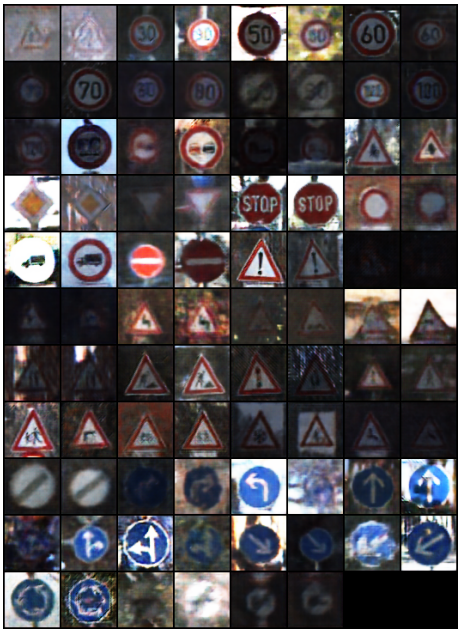

Figure 7 — Sample images generated using conditional Wasserstein GAN.

Training sample	39k imgs w/o augm.	215k imgs w/o augm.	39k imgs w augm.	215k imgs w augm.
Real	96.6	—	98.4	—
WGAN synth.	95.3	96.1	97.6	98.1
Real + WGAN synth.	—	97.7	—	98.4
Conditional WGAN	79.2	83.7	81.3	81.5
Real + conditional WGAN	—	95.2	—	95.5
Synth. from icon	46.5	53.7	67.8	69.7
Real + synth. from icon	—	96.5	—	97.9

Table 1 — Traffic sign classifier results on different training samples.

to real images), second network transforms images from domain B to domain A (from real to synthetic images). Two image datasets from domains A and B are used during training. Neural networks are regularized with cyclic consistency loss: sequential application of two generators should output image that is equal to input image.

Experiments show that trained generator changes sign class during transformation from synthetic to real domain. Additional term called identity

Figure 8 — Sample transformed synthetic traffic sign images with sign class preservation.

loss [19] is added to loss function to train generator to preserve traffic sign class. It penalizes generator for feature vector difference of original and transformed iamages. Feature vectors are neuron activations of penultimate layer of traffic sign classifier. Loss function is defined in the following way:

$$\mathcal{L} = \mathcal{L}_{cyclegan} + \mathsf{E}_{x \sim p_A(x)}[||F(G_{A \to B}(x)) - F(x)||_1]$$

Here F is a neural network used for extracting features of images. Sample images obtained using described generator are shown in fig. 8.

An experiment was conducted to show efficiency of the proposed method. During this experiment some real images of signs were replaced with rendered and processed synthetic images. Several groups of classes (speed limit, blue circles, triangles with red border) were replaced. Neural network classifier was trained on this samples and evaluated on testing sample of real images. Classifier accuracy showed that proposed postprocessing method make synthetic images more realistic. In the chapter 4 it will be shown that proposed method allow to improve accuracy of rare traffic sign classifier.

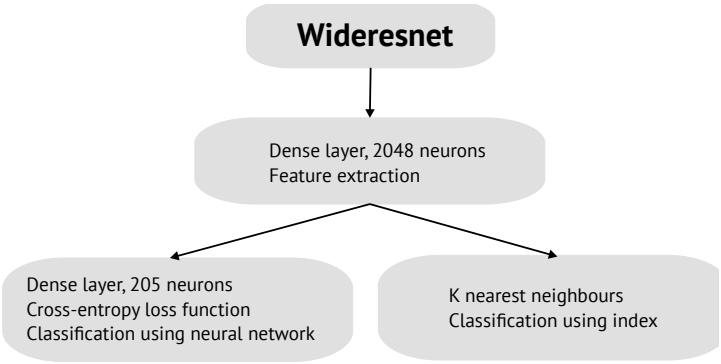

Figure 9 — Scheme of the baseline classifier.

Rare traffic sign classification task is investigated in the **fourth chapter**. Several types of synthetic data were used for training traffic sign classifier: simple synthetic (Synt), which is generated from icon using simple transformations, photorealistic synthetic (CGI), which is rendered from 3D sign models, and photorealistic synthetic with postprocessing (CGI-GAN). Neural network postprocessing for CGI-GAN sample is described in chapter 3.

Baseline classifier. k nearest neighbours classifier is used for baseline. It uses activations from penultimate layer of convolutional neural network Wideresnet [20] trained for traffic sign classification. Scheme of the method is shown in fig. 9.

Improvement of the neural network features. An additional term was added to loss function to improve baseline traffic sign classification method. It aids spatial separation of the features. This term is called contrastive loss function. It stimulates image features of different classes to be far enough from each other in feature space. The term is defined using the following formula [21]:

$$L(x_1,x_2,y) = \frac{1}{2}(1-y)D^2(x_1,x_2) + \frac{1}{2}y(\max(0, m - D(x_1,x_2)))^2, \quad (1)$$

$$D(x_1,x_2) = ||f(x_1) - f(x_2)||_2. \quad (2)$$

Here x_1, x_2 are two images, $f(x)$ is a feature vector obtained from neural network for image x, $D(x_1, x_2)$ is a Euclidean distance between feature vectors, $m > 0$ is a numerical hyperparameter, a threshold which regulates how far feature vectors should be, y is a binary variable which is 0 in case x_1 and x_2 are images of the same class and 1 otherwise. Similar to the baseline classifier exracted feature vectors are scaled and are used from train k nearest neighbours (k-NN) classifier. Scheme of the classifier is shown in fig. 10.

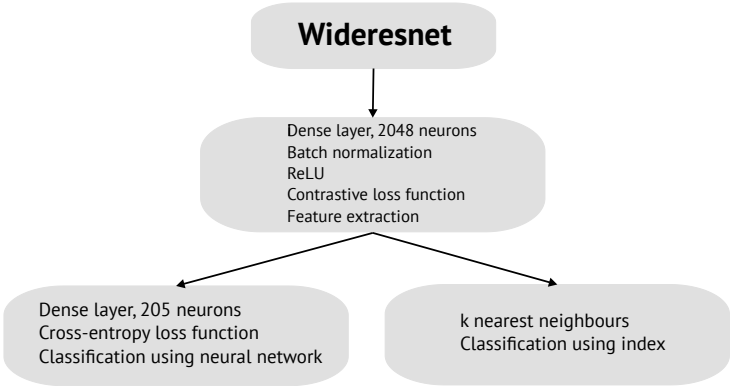

Figure 10 — Scheme of the neural network for extracting improved feature vectors

Proposed classifier Experimental evaluation of the neural network classifier and k-NN classifier showed that first classifier works best on frequent classes and second classifier works best on rare sign classes. To obtain optimal quality either for frequent either for rare sign classes, a hybrid method is proposed. In this method an image is first classified using a binary rare/frequent classifier and then the according method is used for fine-grained classification.

Rare and frequent sign binary classifier is based on idea from [21]. An additional output is added to Wideresnet neural network that outputs trained feature vectors. Modified loss function is used for training:

$$L(x_1,x_2,y) = \frac{1}{2}z(1-y)D^2(x_1,x_2) + \frac{1}{2}zy(\max(0, m - D(x_1,x_2)))^2.$$

This loss function compared to loss (1) has additional variable z that is equal to 0 if both images has rare class and 1 if even one of the images has frequent class. Neural network obtains input minibatches of size 64 during training. Then feature vectors of each image are extracted. Loss function on every new output of the neural network is computed as average of all image pairs in minibatch. Categorical cross-entropy loss function is used to train neural network classification layer. Addition of a new variable z in loss function make it independent of image pairs that both have rare classes.

Classification layer is used only during training step. Learned feature vectors are used to train decision forest with 1000 trees. This forest is used for binary sign classification. Scheme of the method is shown in fig. 11.

Experimental evaluation. Russian Traffic Sign Dataset is used for evaluation. It contains 205 sign classes, out of which 99 classes occur only in test set. Several conclusions can be drawn from experimental results:

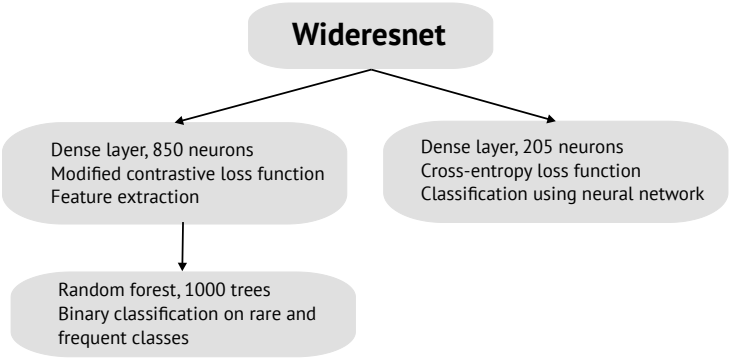

Figure 11 — Scheme of the classifier for rare and frequent sign classification

1. Mixing real and synthetic data improves neural network classifier accuracy.
2. CycleGAN with identity loss improves realism of synthetic data. Classifier that is trained on this data shows highest accuracy in experiments.
3. Classifier trained on purely synthetic training data achieves substantially lower quality than classifier trained on real data.
4. Contrastive loss function improves training of neural feature vectors and aids better separation of sign classes in space.
5. Qualitative binary classifier of frequent and rare sign classes can be trained on purely synthetic data using specially trained neural feature vectors.
6. Proposed hybrid classifier combines neural network classifier for frequent sign classes with k-NN classifier for rare sign classes. It improves quality on rare classes without decrease in quality on frequent classes.

A new traffic sign classification method was proposed in this chapter. It improves quality of rare traffic sign classification and slightly decreases quality on frequent classes. Method uses binary classifier that uses neural network feature vectors.

Synthetic training data generation for traffic sign detector is described in **fifth chapter**.

Random placement of synthetic traffic signs. In three baseline methods synthetic traffic signs are randomly placed in frame. These baseline methods are called Synt (simple synthetic images generated using sign icon), CGI (computer-generated imagery) and CGI-GAN (CGI images improved with CycleGAN). Methods were described in fourth chapter.

Figure 12 — An example of frame with real sign, inpaited real sign and synthetic traffic traffic sign

Replacement of real traffic signs with synthetic traffic signs. In real data traffic signs are located not randomly. Most signs are placed in the upper half of the frame and in one of the two clasters, on the left or on the right side of the road. Modern neural network detectors analyze large area around the object, therefore placement of synthetic traffic sign should be realistic. Random placement isn't realistic.

An advanced method for generating synthetic training data for traffic sign detection is proposed. Synthetic dataset is based on real dataset with labelled traffic signs. All real signs in the dataset are inpainted with specially trained neural network. This neural network has encoder-decoder architecture with skip-connections [22]. It's trained using random backround crops. All crops are resized to 128×128 pixels. Second neural network called discriminator is used for training. It trains to distiguish real background samples from inpainted. Generator and discriminator are trained using Waserstein loss function [15]. A synthetic sign (Synt or CGI-GAN) is placed on top of inpainted real sign. Such methods will be named InpaintSynt and InpaintGAN. Sample original frame, inpainted real sign and a frame with inserted synthetic traffic sign is shown in fig. 12.

Experimental evaluation PVANet detector [23] with Focal loss [24] and two traffic sign classifiers on top of Wideresnet [20] were used for experimental evaluation of the proposed methods. First classifier is a trained Wideresnet with $k = 2$ and $depth = 8$. Classifier obtains images of size 64×64 pixels on input and predicts one of the 205 classes. Second classifier is specially trained for rare traffic sign classification. It is described in chapter 4. Area Under Curve (AUC) metric is used for estimating detection quality. Modified AUC is used to evaluate detection and classification. Detected bounding box is considered true positive if it is correctly classified.

Following conclusing can be drawn from results of experiments:

1. Detector trained only on frequent classes of traffic signs can detect also rare classes of traffic sign. We can assume that detector learns to detect signs of the general appearance. If a rare sign class is similar to a frequent sign class (for instance, a circle with red border), the detector will be able to detect it.

2. Random placement of synthetic traffic signs doesn't improve detection quality. Synthetic signs placed instead of real signs (InpaintSynt, InpaintGAN) compared to baseline methods obtain better quality.
3. Mixing real and synthetic data worsens detection quality. InpaintSynt and InpaintGAN methods are the exceptions. Usage of these methods improves quality of rare traffic sign detection, but worsen a little frequent traffic sign detection quality.

Two conclusions can be drawn from detection and classification experiments:

1. All types of synthetic data improve classification quality of rare and frequent classes of traffic signs. CGI-GAN and InpaintGAN training data show best classification quality.
2. Rare traffic sign classifier trained on CGI-GAN images substantially improves quality of traffic sign recognition even with detector trained only on real data. Best rare traffic recognition quality is obtained using real and improved synthetic data (InpaintGAN).

In this chapter several methods for generating data for traffic sign detection were considered. We considered three random placement methods and an improved method that replaces real traffic signs with synthetic signs. Proposed method was evaluated on traffic sign detection and showed its' efficiency compared to baseline methods. Besides that detector was evaluated together with two traffic sign classifiers: baseline neural network classifier and rare traffic sign classifier from chapter 4. Proposed rare traffic sign classifier showed better results with traffic sign detector than baseline classifier.

In **conclusion** of the dissertain main results of the work are described:

1. Russian Traffic Sign Dataset is collected and labelled. This dataset is suitable for training and evaluating rare traffic sign recognition algorithms.
2. It was shown that generative neural networks can be successfully applied for traffic sign image generation.
3. A method for training conditional Wasserstein GAN is proposed.
4. A method for postprocessing synthetic images using neural networks is developed. It was evaluated on traffic sign classification task.
5. A method for traffic sign classification is proposed. Is improves quality on rare sign classes. The method works in two stages. It first classifies whether a sign is of frequent or rare class and then applies neural network for frequent sign classification and k nearest neighbours for rare sign classification.
6. A method for generating traffic sign detection data is proposed. The method used real training data with sign replaced with synthetic signs. Experimental evaluation of the proposed method showed that it is better than randomly placing synthetic traffic signs on frame.

Further improvement of the proposed methods is possible in the following directions:

- Better postprocessing of the inserted signs that considers context of the inserted sign;
- Better placement of synthetic signs in frame. A special generative neural network can be trained to sample possible places of a sign in image.

Список литературы

1. Deepface: Closing the gap to human-level performance in face verification [Text] / Y. Taigman [et al.] // Proceedings of the IEEE conference on computer vision and pattern recognition. — 2014. — P. 1701—1708.
2. Labeled Faces in the Wild: A Database for Studying Face Recognition in Unconstrained Environments [Text] : tech. rep. / G. B. Huang [et al.] ; University of Massachusetts, Amherst. — 10/2007. — No. 07—49.
3. Classifier training based on synthetically generated samples [Text] / H. Hoessler [et al.] // International Conference on Computer Vision Systems: Proceedings (2007). — 2007.
4. Evaluation of traffic sign recognition methods trained on synthetically generated data [Text] / B. Moiseev [et al.] // International Conference on Advanced Concepts for Intelligent Vision Systems. — Springer. 2013. — P. 576—583.
5. *Chigorin, A.* A system for large-scale automatic traffic sign recognition and mapping [Text] / A. Chigorin, A. Konushin // CMRT13–City Models, Roads and Traffic. — 2013. — Vol. 2013. — P. 13—17.
6. Generative adversarial nets [Text] / I. Goodfellow [et al.] // Advances in neural information processing systems. — 2014. — P. 2672—2680.
7. *Radford, A.* Unsupervised Representation Learning with Deep Convolutional Generative Adversarial Networks [Text] / A. Radford, L. Metz, S. Chintala // CoRR. — 2015. — Vol. abs/1511.06434.
8. Unpaired image-to-image translation using cycle-consistent adversarial networks [Text] / J.-Y. Zhu [et al.] // Proceedings of the IEEE international conference on computer vision. — 2017. — P. 2223—2232.
9. *Zheng, Z.* Unlabeled samples generated by gan improve the person re-identification baseline in vitro [Text] / Z. Zheng, L. Zheng, Y. Yang // Proceedings of the IEEE International Conference on Computer Vision. — 2017. — P. 3754—3762.
10. *Dollár, P.* Piotr's Computer Vision Matlab Toolbox (PMT) [Text] / P. Dollár. — https://github.com/pdollar/toolbox.

11. Traffic sign recognition—How far are we from the solution? [Text] / M. Mathias [et al.] // The 2013 international joint conference on Neural networks (IJCNN). — IEEE. 2013. — P. 1—8.

12. Detection of traffic signs in real-world images: The German Traffic Sign Detection Benchmark [Text] / S. Houben [et al.] // The 2013 international joint conference on neural networks (IJCNN). — IEEE. 2013. — P. 1—8.

13. Multi-column deep neural network for traffic sign classification [Text] / D. CireşAn [et al.] // Neural networks. — 2012. — Vol. 32. — P. 333—338.

14. Man vs. computer: Benchmarking machine learning algorithms for traffic sign recognition [Text] / J. Stallkamp [et al.] // Neural Networks. — 2012. — Vol. 32. — P. 323—332. — Selected Papers from IJCNN 2011.

15. *Arjovsky, M.* Wasserstein Generative Adversarial Networks [Text] / M. Arjovsky, S. Chintala, L. Bottou // Proceedings of the 34th International Conference on Machine Learning - Volume 70. — Sydney, NSW, Australia : JMLR.org, 2017. — P. 214—223. — (ICML'17).

16. *Mirza, M.* Conditional Generative Adversarial Nets [Text] / M. Mirza, S. Osindero. — 2014. — arXiv: 1411.1784 [cs.LG].

17. Evaluation of traffic sign recognition methods trained on synthetically generated data [Text] / B. Moiseev [et al.] // International Conference on Advanced Concepts for Intelligent Vision Systems. — Springer. 2013. — P. 576—583.

18. Improved Training of Wasserstein GANs [Text] / I. Gulrajani [et al.] // Proceedings of the 31st International Conference on Neural Information Processing Systems. — Long Beach, California, USA : Curran Associates Inc., 2017. — P. 5769—5779. — (NIPS'17).

19. Beyond face rotation: Global and local perception gan for photorealistic and identity preserving frontal view synthesis [Text] / R. Huang [et al.] // Proceedings of the IEEE International Conference on Computer Vision. — 2017. — P. 2439—2448.

20. *Zagoruyko, S.* Wide Residual Networks [Text] / S. Zagoruyko, N. Komodakis. — 2016. — arXiv: 1605.07146 [cs.CV].

21. Metric learning for novelty and anomaly detection [Text] / M. Masana [et al.] // arXiv preprint arXiv:1808.05492. — 2018.

22. Faceshop: Deep sketch-based face image editing [Text] / T. Portenier [et al.] // arXiv preprint arXiv:1804.08972. — 2018.

23. PVANET: Deep but Lightweight Neural Networks for Real-time Object Detection [Text] / K.-H. Kim [et al.]. — 2016. — arXiv: 1608.08021 [cs.CV].

24. Focal loss for dense object detection [Text] / T.-Y. Lin [et al.] // Proceedings of the IEEE international conference on computer vision. — 2017. — P. 2980—2988.