

Лаборатория компьютерной
графики и мультимедиа
ВМК МГУ имени М.В. Ломоносова

**«Синтетические данные и порождающие модели в
распознавании дорожных знаков и других задачах
компьютерного зрения»**

Антон Конушин

2020 год

Распознавание дорожных знаков



Рассматривается задача выделения и классификации дорожных знаков на изображении

Системы помощи водителю и управления беспилотным автомобилем



Обновление карт





Актуальные проблемы

Современные методы распознавания объектов используют методы машинного обучения. Качество распознавания зависит от размера и разнообразия обучающей выборки.

1. Выборку данных большого размера сложно собрать и разметить.
2. Есть редко встречающиеся или новые классы объектов.



Поэтому современные алгоритмы распознают только часть существующих дорожных знаков. Данная работа фокусируется на проблеме распознавания редких дорожных знаков.

Синтез обучающих данных



Одним из возможных решений проблемы отсутствия эталонных данных является синтез данных. Синтез данных успешно используется для обучения некоторых алгоритмов компьютерного зрения.

1. Методы компьютерной графики



вычисление оптического потока



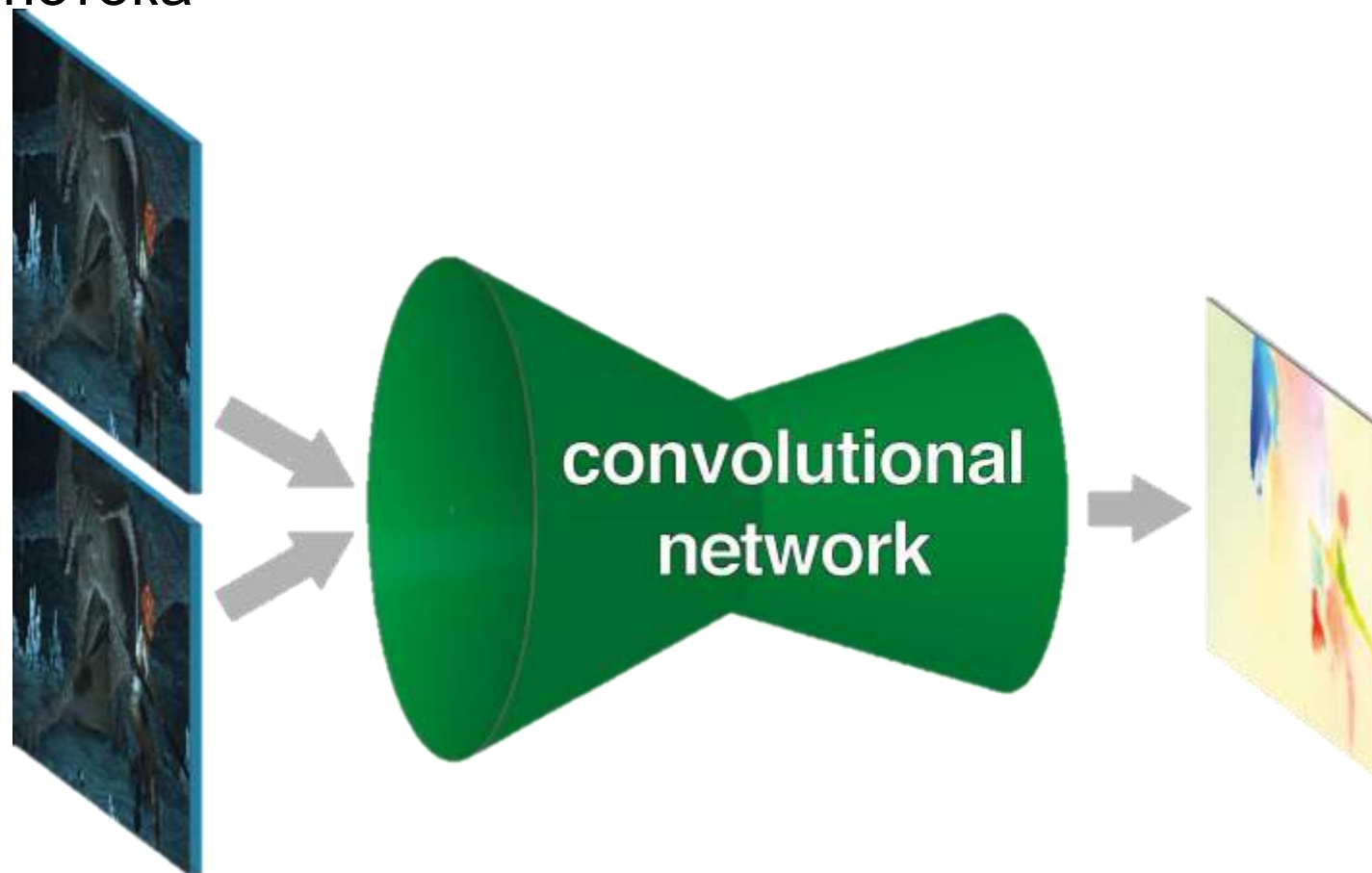
выделение автомобилей

2. Нейросетевые методы генерации изображений — перспективная и мало исследованная область синтеза обучающих выборок

FlowNet



Прямолинейное применение свёрточных нейросетей к вычислению оптического потока



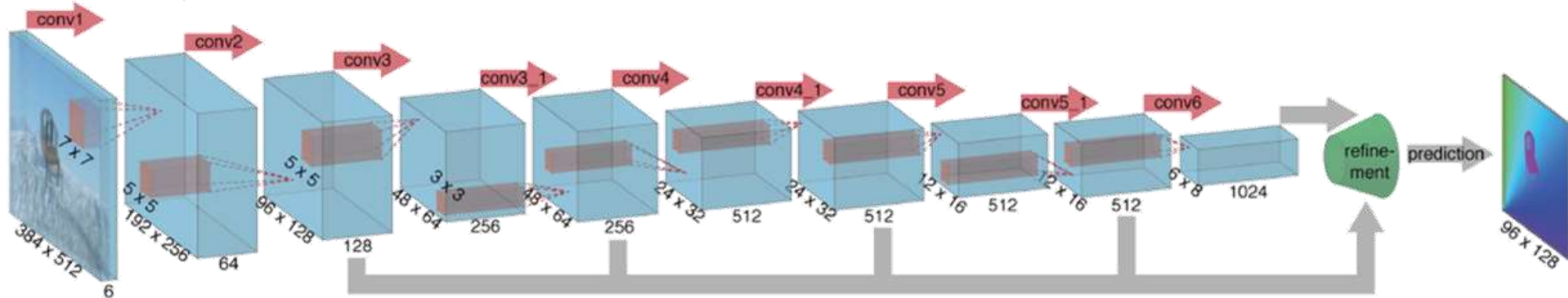
Source: <https://arxiv.org/abs/1504.06852>

A. Dosovitskiy et. al. FlowNet: Learning Optical Flow with Convolutional Networks. 2015

FlowNet



FlowNetSimple



- Объединяем 2 кадра в би канальное изображение
- Применяем нейросеть
- Можно сделать сложнее, объединив признаки с 2х картинок с помощью специального слоя сравнения патчей

MPI Sintel



Open source 3D movie “Sintel”



1064 training and 564 test frames



Various rendering options



Source: http://ps.is.tue.mpg.de/research_projects/mpi-sintel-flow
D. Butler et. al. A naturalistic open source movie for optical flow evaluation. ECCV 2012

Flying chairs



	Frame pairs	Frames with ground truth	Ground truth density per frame
Middlebury	72	8	100%
KITTI	194	194	~50%
Sintel	1041	1041	100%
Flying Chairs	22872	22872	100%

- Не хватает данных для обучения
- Поэтому синтетический набор из летающих стульев

Source: arxiv.org/abs/1504.06852



Evaluation ground truth Evaluation area

	Method	Setting	Code	FI-bg	FI-fg	FI-all	Density	Runtime	Environment	Compare
1	PSPO			4.35 %	15.21 %	6.15 %	100.00 %	5 min	1 core @ 2.5 Ghz (Matlab + C/C++)	<input type="checkbox"/>
2	ISF			5.40 %	10.29 %	6.22 %	100.00 %	10 min	1 core @ 3 Ghz (C/C++)	<input type="checkbox"/>
A. Behl, O. Jafari, S. Mustikovela, H. Alhaija, C. Rother and A. Geiger: Bounding Boxes, Segmentations and Object Coordinates: How Important is Recognition for 3D Scene Flow Estimation in Autonomous Driving Scenarios? . International Conference on Computer Vision (ICCV) 2017.										
3	PRSM		code	5.33 %	13.40 %	6.68 %	100.00 %	300 s	1 core @ 2.5 Ghz (C/C++)	<input type="checkbox"/>
C. Vogel, K. Schindler and S. Roth: 3D Scene Flow Estimation with a Piecewise Rigid Scene Model . ijcv 2015.										
4	OSF+TC			5.76 %	13.31 %	7.02 %	100.00 %	50 min	1 core @ 2.5 Ghz (C/C++)	<input type="checkbox"/>
M. Neoral and J. Šochman: Object Scene Flow with Temporal Consistency . 22nd Computer Vision Winter Workshop (CVWW) 2017.										
5	SSF			5.63 %	14.71 %	7.14 %	100.00 %	5 min	1 core @ 2.5 Ghz (Matlab + C/C++)	<input type="checkbox"/>
Z. Ren, D. Sun, J. Kautz and E. Sudderth: Cascaded Scene Flow Prediction using Semantic Segmentation . International Conference on 3D Vision (3DV) 2017.										
6	SOSF			5.42 %	17.24 %	7.39 %	100.00 %	55 min	1 core @ 2.5 Ghz (Matlab + C/C++)	<input type="checkbox"/>
7	OSF		code	5.62 %	18.92 %	7.83 %	100.00 %	50 min	1 core @ 2.5 Ghz (C/C++)	<input type="checkbox"/>
M. Menze and A. Geiger: Object Scene Flow for Autonomous Vehicles . Conference on Computer Vision and Pattern Recognition (CVPR) 2015.										
8	PWC-Net			9.66 %	9.31 %	9.60 %	100.00 %	0.03 s	NVIDIA Pascal Titan X	<input type="checkbox"/>
9	MirrorFlow			8.93 %	17.07 %	10.29 %	100.00 %	11 min	4 core @ 2.2 Ghz (C/C++)	<input type="checkbox"/>
J. Hur and S. Roth: MirrorFlow: Exploiting Symmetries in Joint Optical Flow and Occlusion Estimation . ICCV 2017.										
10	FlowNet2			10.75 %	8.75 %	10.41 %	100.00 %	0.12 s	GPU Nvidia GeForce GTX 1080	<input type="checkbox"/>
11	SDF			8.61 %	23.01 %	11.01 %	100.00 %	TBA	1 core @ 2.5 Ghz (C/C++)	<input type="checkbox"/>
M. Bai*, W. Luo*, K. Kundu and R. Urtasun: Exploiting Semantic Information and Deep Matching for Optical Flow . ECCV 2016.										
12	UnFlow			10.15 %	15.93 %	11.11 %	100.00 %	0.12 s	GPU @ 1.5 Ghz (Python + C/C++)	<input type="checkbox"/>
S. Meister, J. Hur and S. Roth: UnFlow: Unsupervised Learning of Optical Flow with a Bidirectional Census Loss . AAAI 2018.										
13	FSF+MS			8.48 %	25.43 %	11.30 %	100.00 %	2.7 s	4 cores @ 3.5 Ghz (C/C++)	<input type="checkbox"/>
T. Tani, S. Sinha and Y. Sato: Fast Multi-frame Stereo Scene Flow with Motion Segmentation . IEEE Conference on Computer Vision and Pattern Recognition (CVPR) 2017.										
14	CNNF+PMBP			10.08 %	18.56 %	11.49 %	100.00 %	45 min	1 cores @ 3.5 Ghz (C/C++)	<input type="checkbox"/>
15	MR-Flow		code	10.13 %	22.51 %	12.19 %	100.00 %	8 min	1 core @ 2.5 Ghz (Python + C/C++)	<input type="checkbox"/>
J. Wulff, L. Sevilla-Lara and M. Black: Optical Flow in Mostly Rigid Scenes . IEEE Conf. on Computer Vision and Pattern Recognition (CVPR) 2017.										

Figure 8. Screenshot of the KITTI 2015 public table. PWC-Net has the lowest percentage of error (FI-all) among all optical flow methods, only inferior to scene flow methods that use additional stereo input information.



	GTSRB	GTSDDB	STS	BTSD	LISA	RTSD
Количество кадров	-	900	4000	25630	6610	179138
Количество знаков	43	43	7	108	47	205
Физических знаков	1728	1213	-	4565	-	15630
Изображений знаков	51839	1213	3488	13444	7855	104358

статистики публичных баз дорожных знаков

Существующие публичные выборки не подходят для обучения и оценки системы распознавания редких дорожных знаков (детектор и классификатор), поэтому была собрана база российских дорожных знаков (RTSD), подходящая для данной задачи.

База RTSD



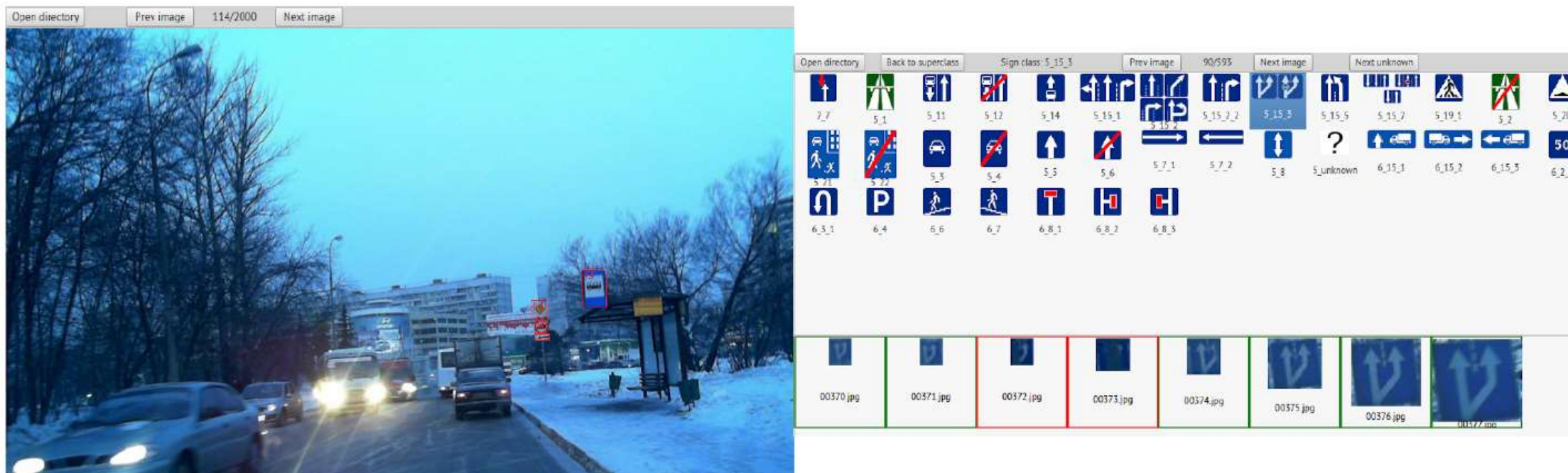
кадры из RTSD

получены с HD видеорегистраторов, установленных на автомобилях

Разметка данных



1. Ручная разметка с помощью 3 фрилансеров.
Трудоемкость — 500 человеко-часов.



2. Обучение детектора, анализ ошибок и доразметка выборки.
Добавлено 9 тысяч новых знаков.

Разбиение на тренировочную и тестовую части

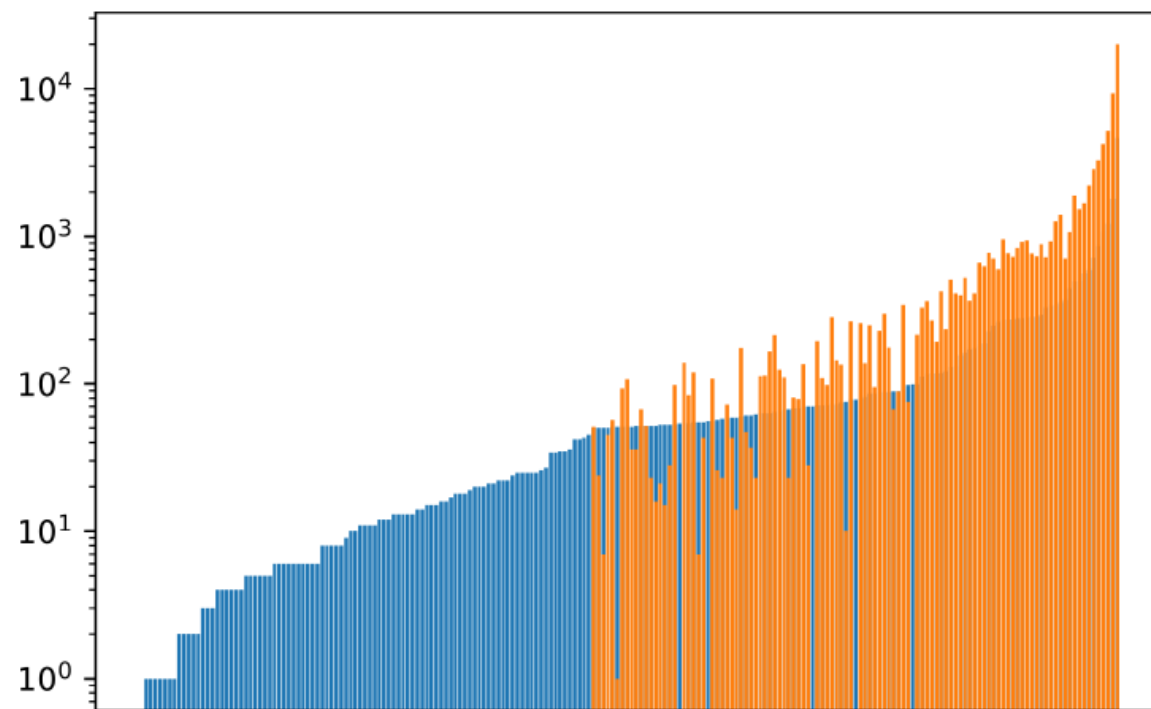


Выборка делится в соотношении примерно 4 к 1.

В тестовую часть входят все 205 классов знаков.

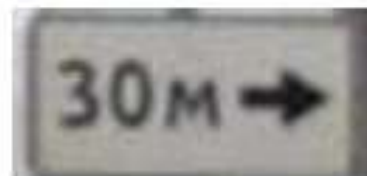
Изображения одного экземпляра знака входят целиком либо в тестовую, либо в тренировочную часть.

В тренировочную часть вошло 106 классов. 99 классов — редкие знаки.



количество дорожных знаков
в **тренировочной** и **тестовой** части

Примеры редких дорожных знаков



Исследование применимости синтетических методов



Базовый метод

Генерация синтетических изображений дорожных знаков по иконке с помощью набора эвристических преобразований:



контраст



поворот



движение



размытие



фон

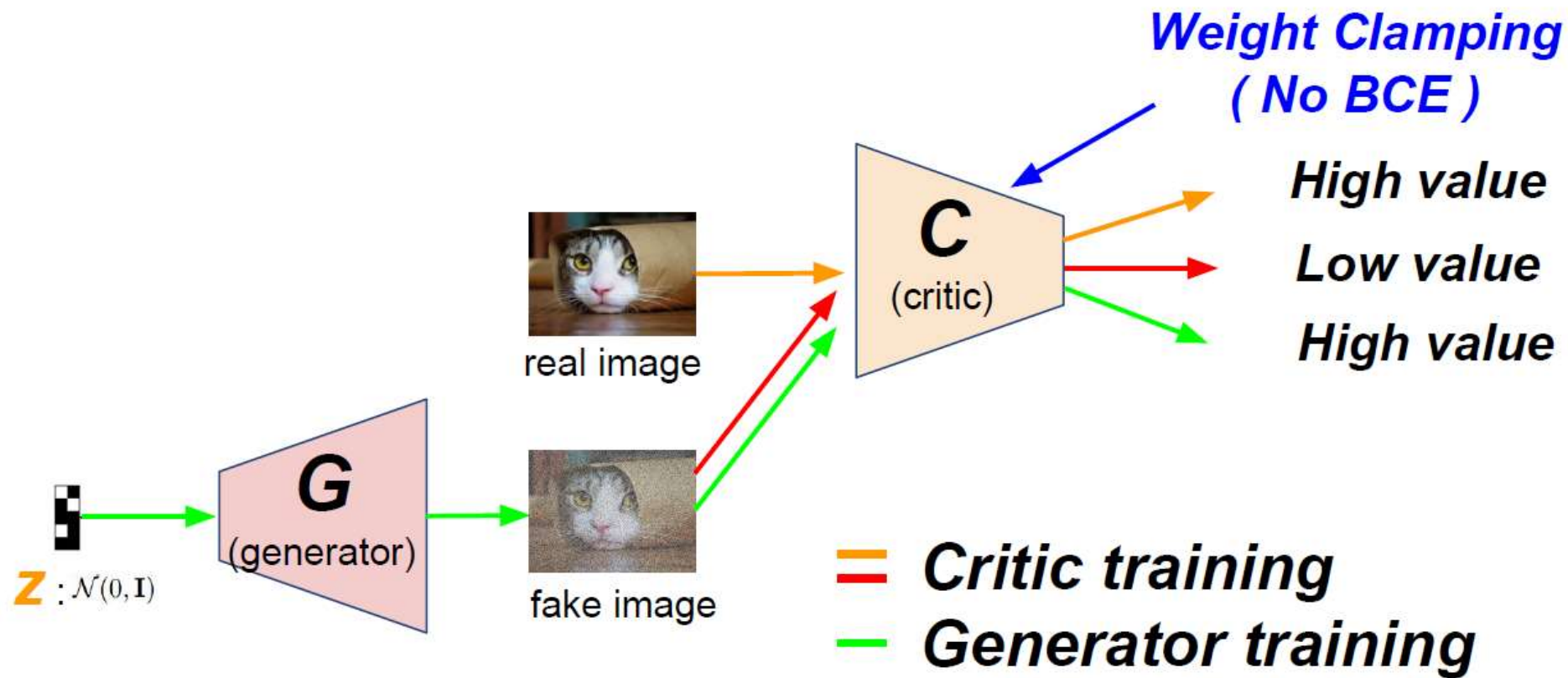
Получение более реалистичных синтетических изображений с помощью данного метода — очень трудоемкая задача.

Chigorin, Konushin. A system for large-scale traffic sign recognition and mapping. CMRT 2013

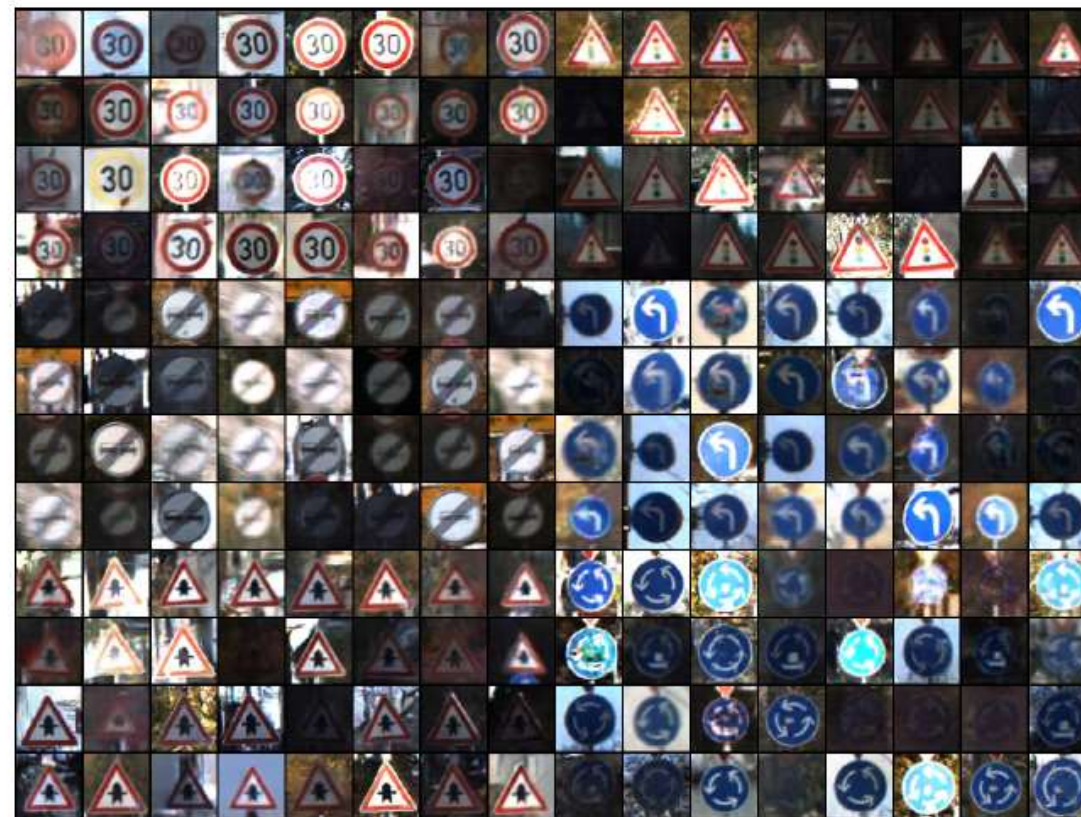
Wasserstein GAN



- Martin et al, Wasserstein GAN, 2017



Примеры сгенерированных знаков

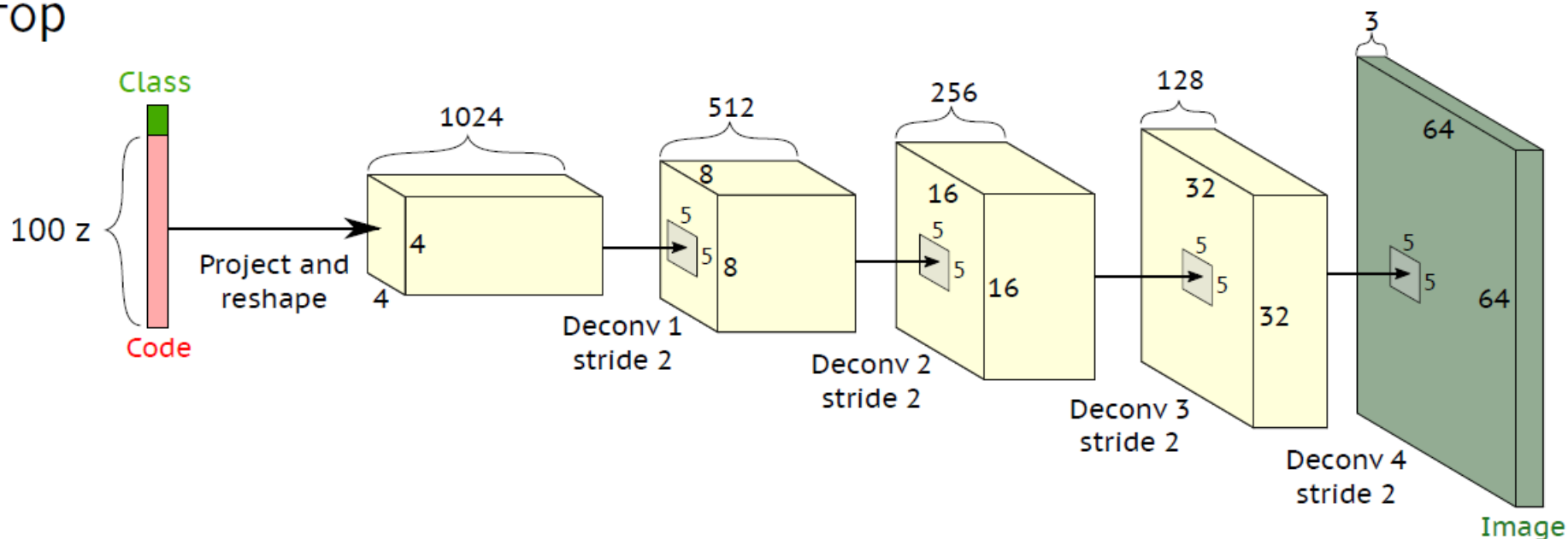


В. Шахуро, А. Конушин. Синтез обучающих выборок для классификации знаков с помощью нейросетей. Компьютерная оптика 2018; 42(1)

Условные генеративные сети



Генератор



Критик

Сверточная нейросеть с двумя выходами: мера реалистичности и класс изображения

Odena et al. Conditional Image Synthesis With Auxiliary Classifier GANs. ICML 2017

Обучение



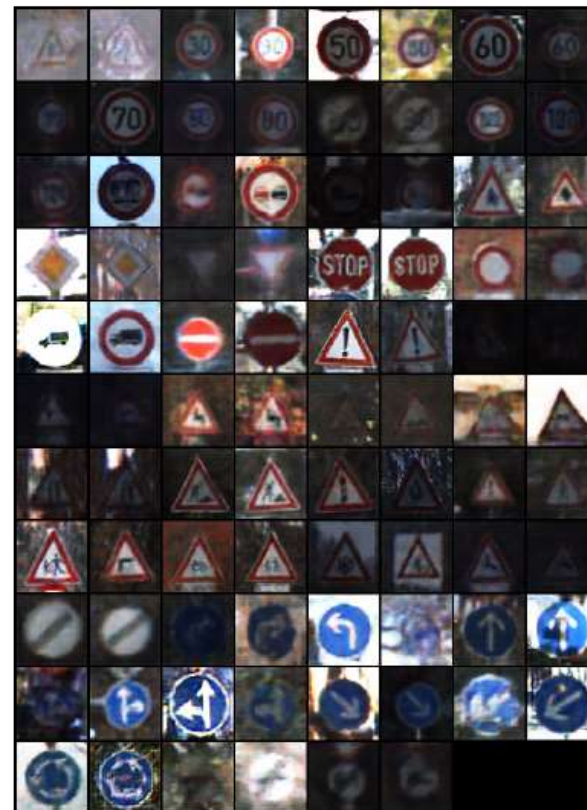
Функция потерь:

$$L = L_C + W(p_r, p_g) + L_R$$

Чтобы критик использовал метку класса, используется перекрестная энтропия:

$$L_C = E[\log p(c|x_{real})] + E[\log p(c|x_{fake})]$$

При обучении на несбалансированной выборке каждому обучающему примеру с меткой y добавим вес $\frac{1}{k_p(y)}$, где k — количество объектов данного класса



сгенерированные
изображения

П. Хрушков, В. Шахуро, А. Конушин. Синтез изображений дорожных знаков с помощью условных порождающих противоборствующих нейросетей. ГрафиКон 2018

Экспериментальная оценка

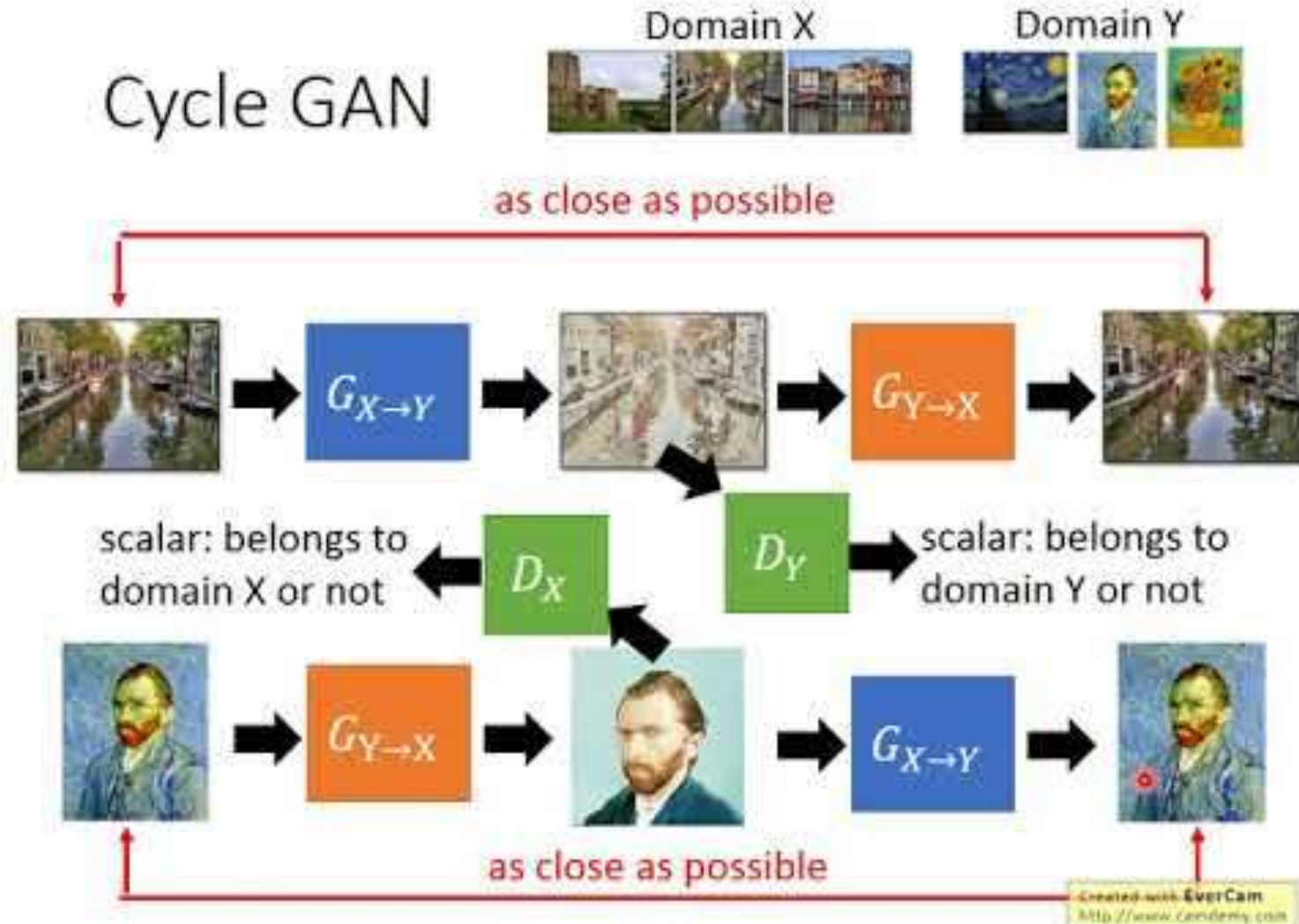


	Тренировочная выборка			
	39 тыс. без размн.	215 тыс. без размн.	39 тыс. с размн.	215 тыс. с размн.
Реальные данные	96.6	—	98.4	—
WGAN синтетика	95.3	96.1	97.6	98.1
Условная WGAN	79.2	83.7	81.3	81.5
Синтетика по иконке	46.5	53.7	67.8	69.7

Экспериментальная оценка проводилась на базе немецких знаков GTSRB: 43 класса знаков, ~ 50 тыс. изображений для обучения и тестирования

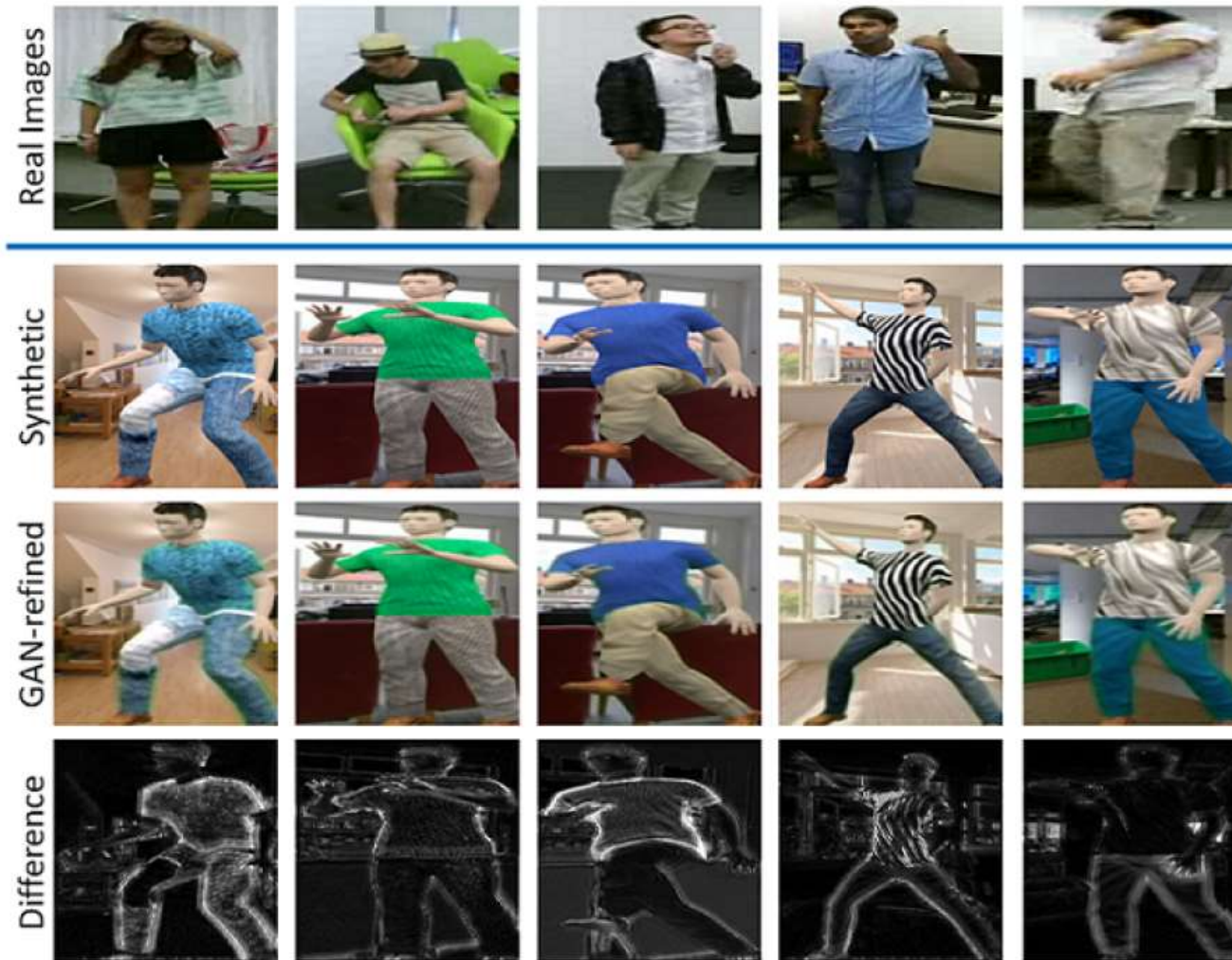
В качестве классификатора использовалась сверточная нейросеть

CycleGAN



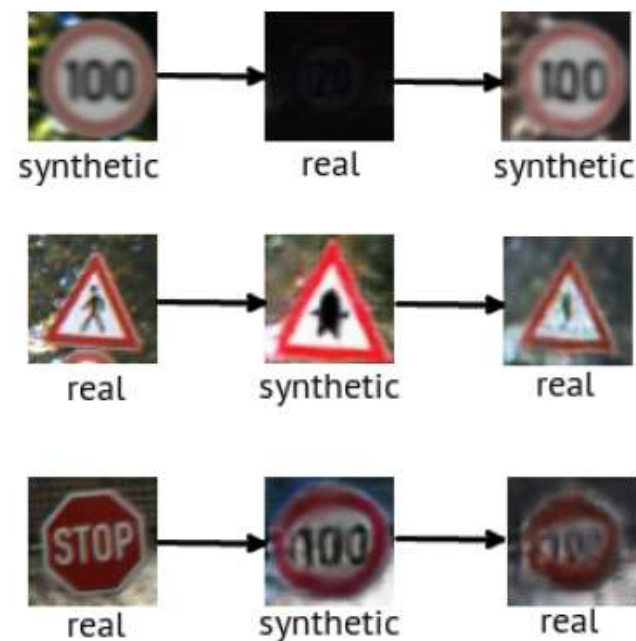
Zhu, Jun-Yan, et al. "Unpaired image-to-image translation using cycle-consistent adversarial networks." *arXiv preprint*(2017).

Повышение визуального реализма синтетики



Jian Liu, Ajmal Mian. Learning Human Pose Models from Synthesized Data for Robust RGB-D Action Recognition. <https://arxiv.org/abs/1707.00823>

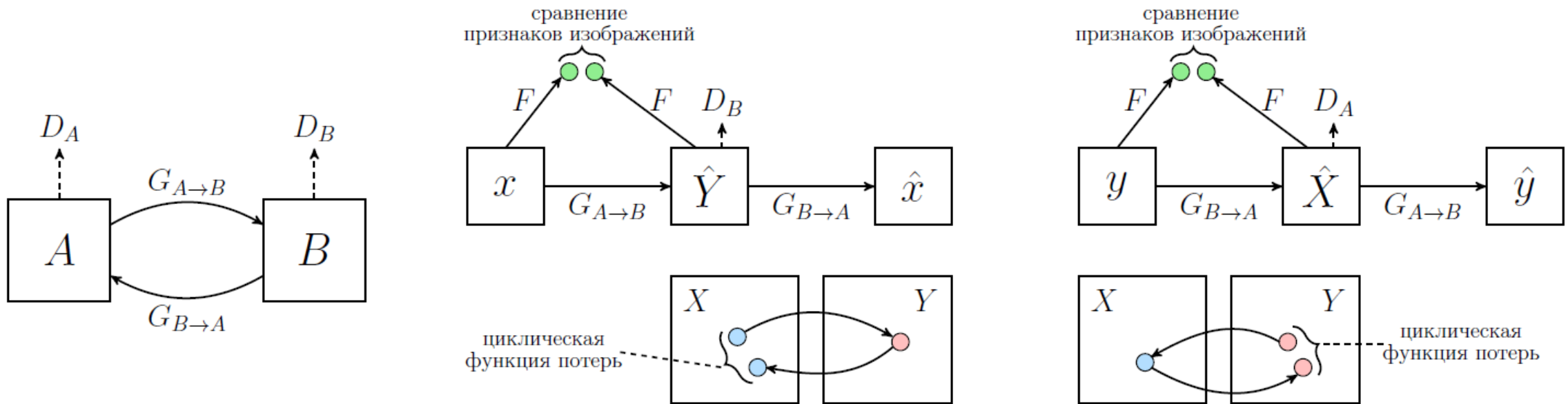
Повышение визуального реализма синтетики



примеры простых синтетических изображений дорожных знаков

J.-Y. Zhu, T. Park et al. Unpaired Image-to-Image Translation using Cycle-Consistent Adversarial Networks. ICCV 2017

Схема предложенного алгоритма



Для сохранения класса объекта в функцию потерь добавляется сравнение нейросетевых признаков \bullet исходного и преобразованного изображения

Экспериментальная оценка



Тренировочная выборка	Точность классификации
GTSRB	98.2
GTSRB с заменой на простую синтетику «ограничение скорости»	87.7
GTSRB с заменой на GAN-синтетику «ограничение скорости»	95.8
GTSRB с заменой на простую синтетику «синие круги»	93.8
GTSRB с заменой на GAN-синтетику «синие круги»	94.4
GTSRB с заменой на простую синтетику «красные треугольники»	94.6
GTSRB с заменой на GAN-синтетику «красные треугольники»	95.8

Метод классификации редких дорожных знаков



Б. Фаизов, В. Шахуро, В. Санжаров, А. Конушин. Классификация редких дорожных знаков. Компьютерная оптика 2020; 44(2)

Синтез обучающих данных для детекторов



Базовые методы: случайное размещение знаков



synt



cgi



cgi-gan

Улучшенный метод: вставка синтетических знаков вместо реальных



исходный кадр



inpainting



вставка знака

Результаты экспериментов



Тренир. выборка	все	част.	редк.
Real	0.8908	0.8920	0.8602
Synt	0.1390	0.1385	0.1483
CGI	0.1070	0.1063	0.1323
InpaintSynt	0.5523	0.5626	0.5526
InpaintGAN	0.6212	0.6236	0.5426
Real + Synt	0.8848	0.8862	0.8554
Real + CGI	0.8856	0.8872	0.8572
Real + GAN	0.8853	0.8869	0.8521
Real + InpaintSynt	0.8861	0.8871	0.8663
Real + InpaintGAN	0.8898	0.8913	0.8660

Использование синтетических данных повышает качество выделения редких знаков при незначительном уменьшении качества выделения частых знаков. Метрика качества — площадь под кривой точность-полнота.

Результаты экспериментов (детектор и классификатор)



Детектор и классификатор WideResNet учатся на одной выборке

Тренир. выборка	все	част.	редк.
Real	0.7544	0.8246	0.0909
Synt	0.0997	0.0999	0.0909
CGI	0.0881	0.0881	0.0853
InpaintSynt	0.1589	0.1596	0.1379
InpaintGAN	0.1985	0.1880	0.1991
Real + Synt	0.8022	0.8438	0.3396
Real + CGI	0.8384	0.8515	0.4851
Real + GAN	0.8445	0.8568	0.4983
Real + InpaintSynt	0.7641	0.8293	0.3400
Real + InpaintGAN	0.8499	0.8586	0.5136

Использование синтетических данных значительно повышает качество распознавания как частых, так и редких знаков.

Результаты экспериментов (детектор и классификатор)

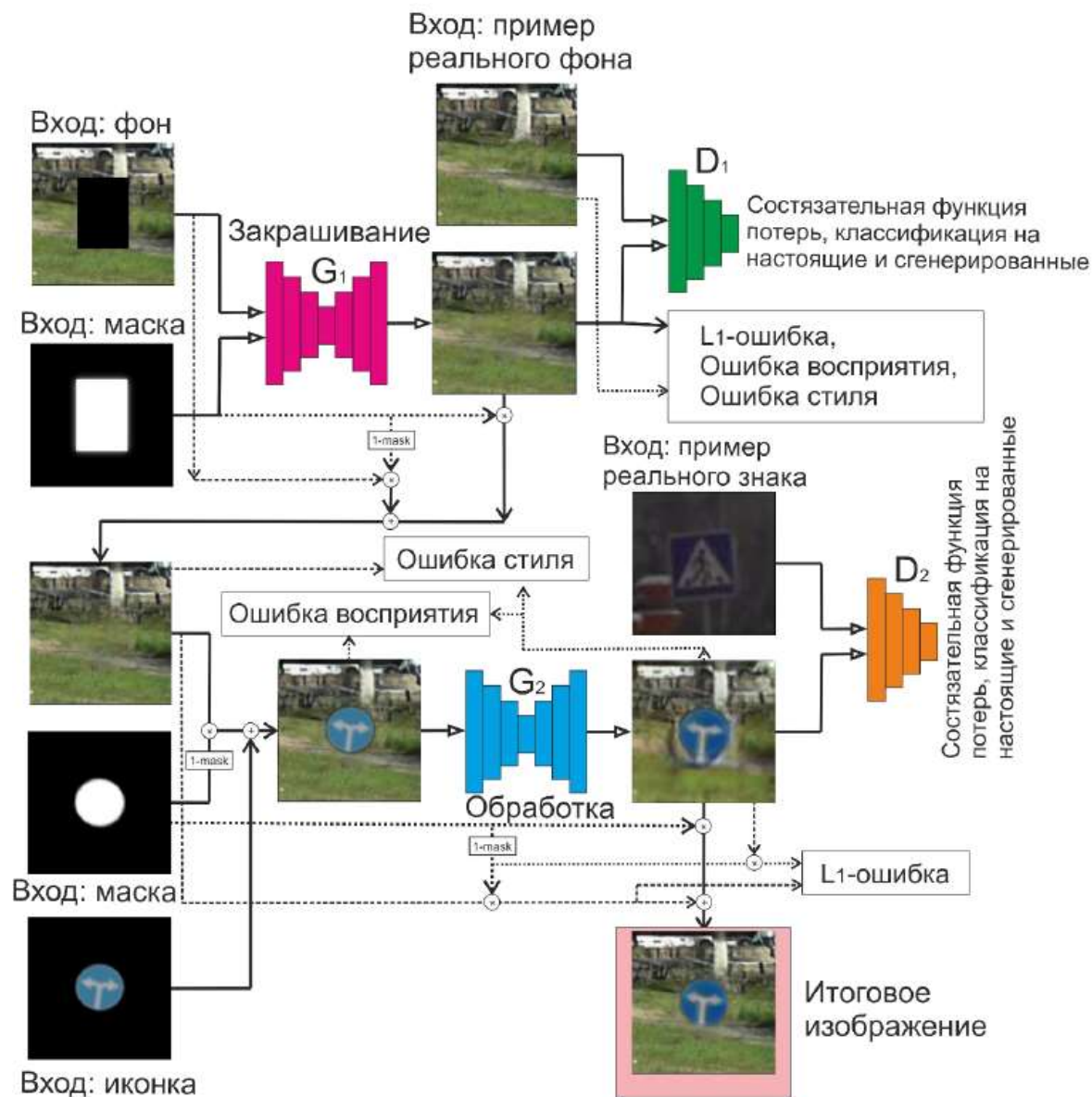


Детектор и классификатор редких знаков с использованием индекса из синт. данных, улучшенных с помощью GAN

Тренир. выборка детектора	все	част.	редк.
Real	0.8606	0.8673	0.5896
Real+GAN	0.8514	0.8589	0.5857
Real+InpaintSynt	0.8440	0.8528	0.5942
Real+InpaintGAN	0.8611	0.8662	0.6013

Использование синтетических данных и предложенного метода классификации (задача 5) значительно повышает качество распознавания знаков.

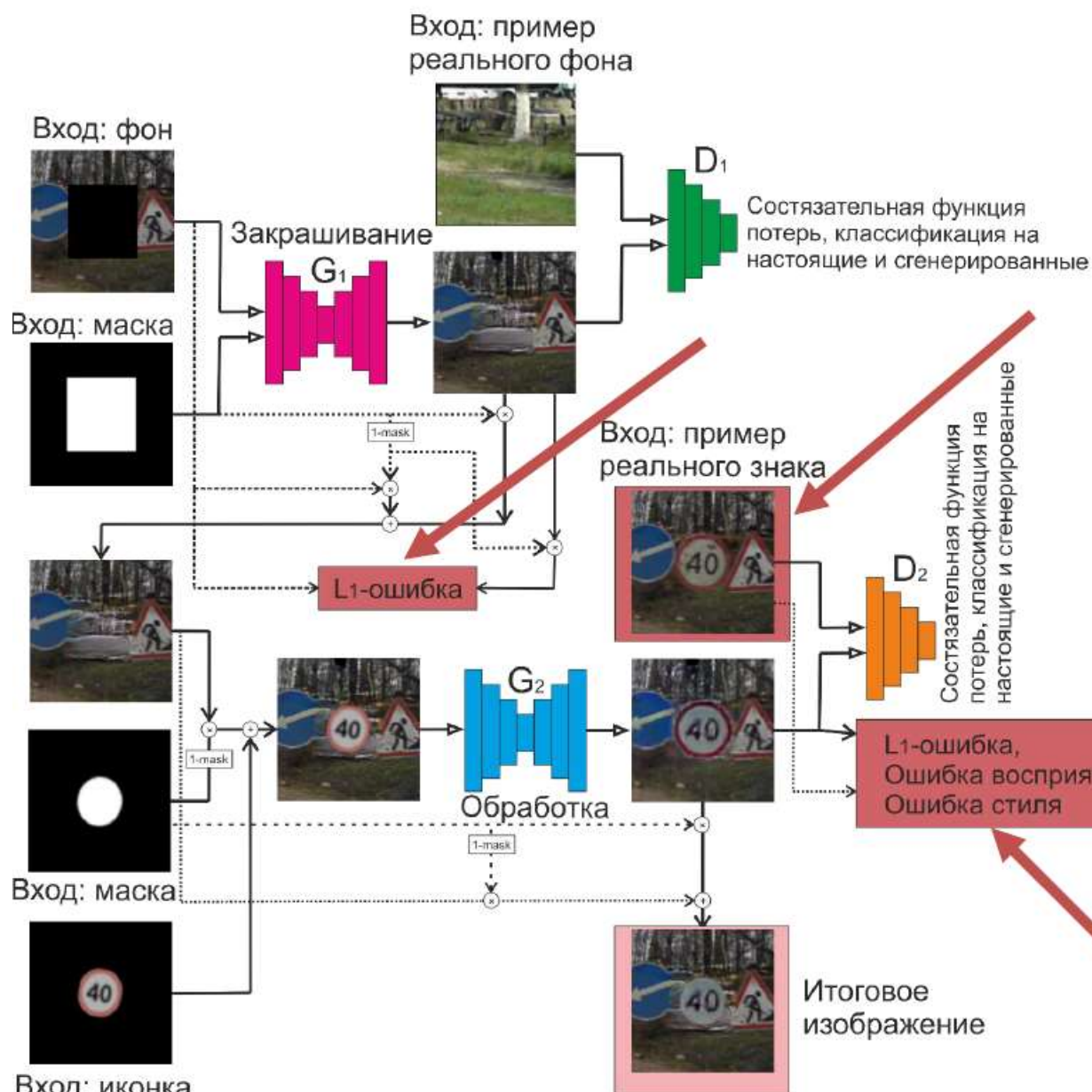
Исследование более сложных методов (Вариант 1)



Обозначения:

- G_1 – закрашивание части изображения.
- D_1 – классификация реальной картинки и закрашенной
- G_2 – пост-обработка встроенной иконки
- D_2 – классификация реальных знаков и встроенных

Исследование более сложных методов (Вариант 2)



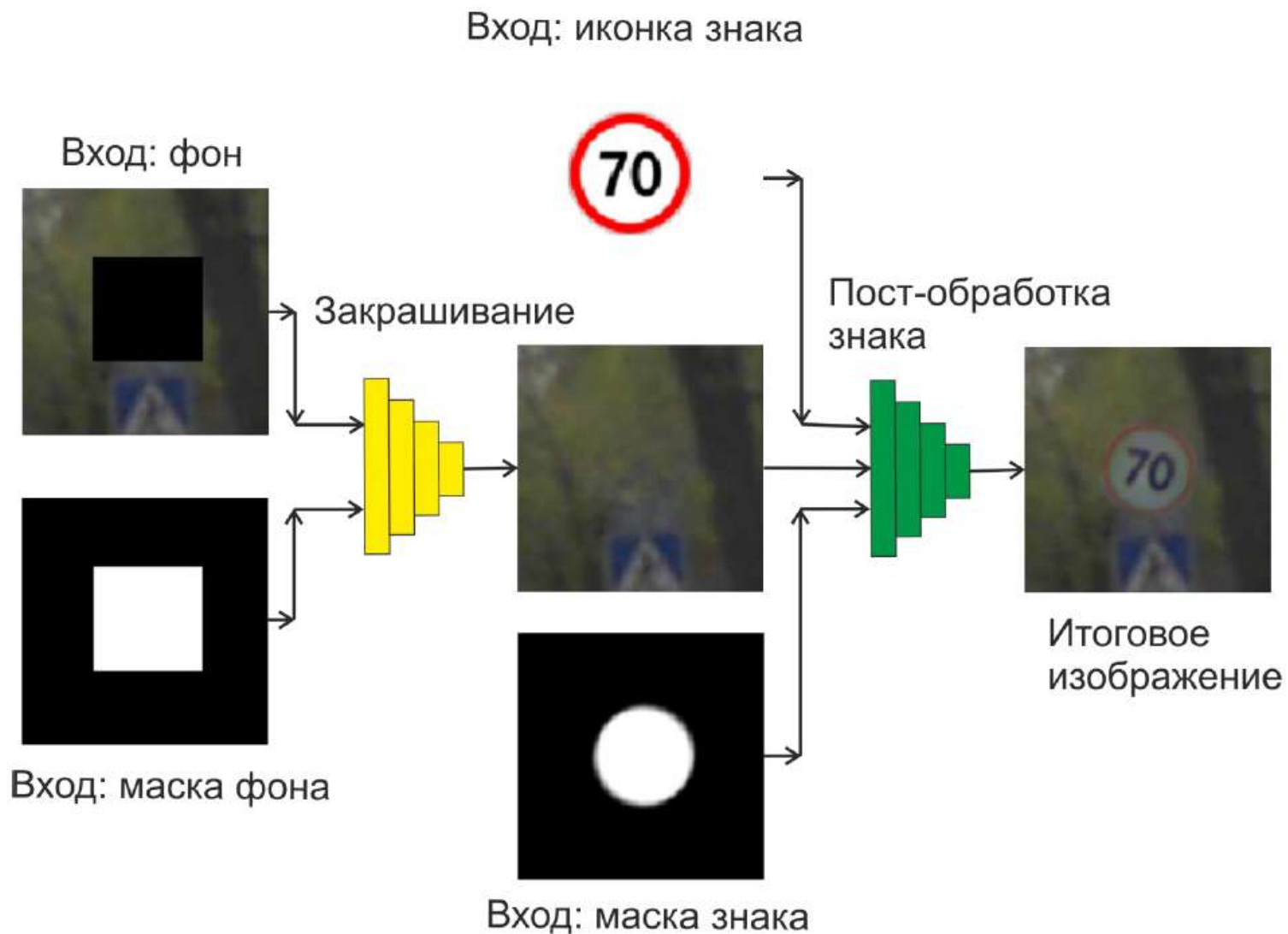
Добавляется дополнительный поток данных при обучении

В нем заранее известен только правильно встроенный знак, а в первом потоке данных известна правильная закрашенная область

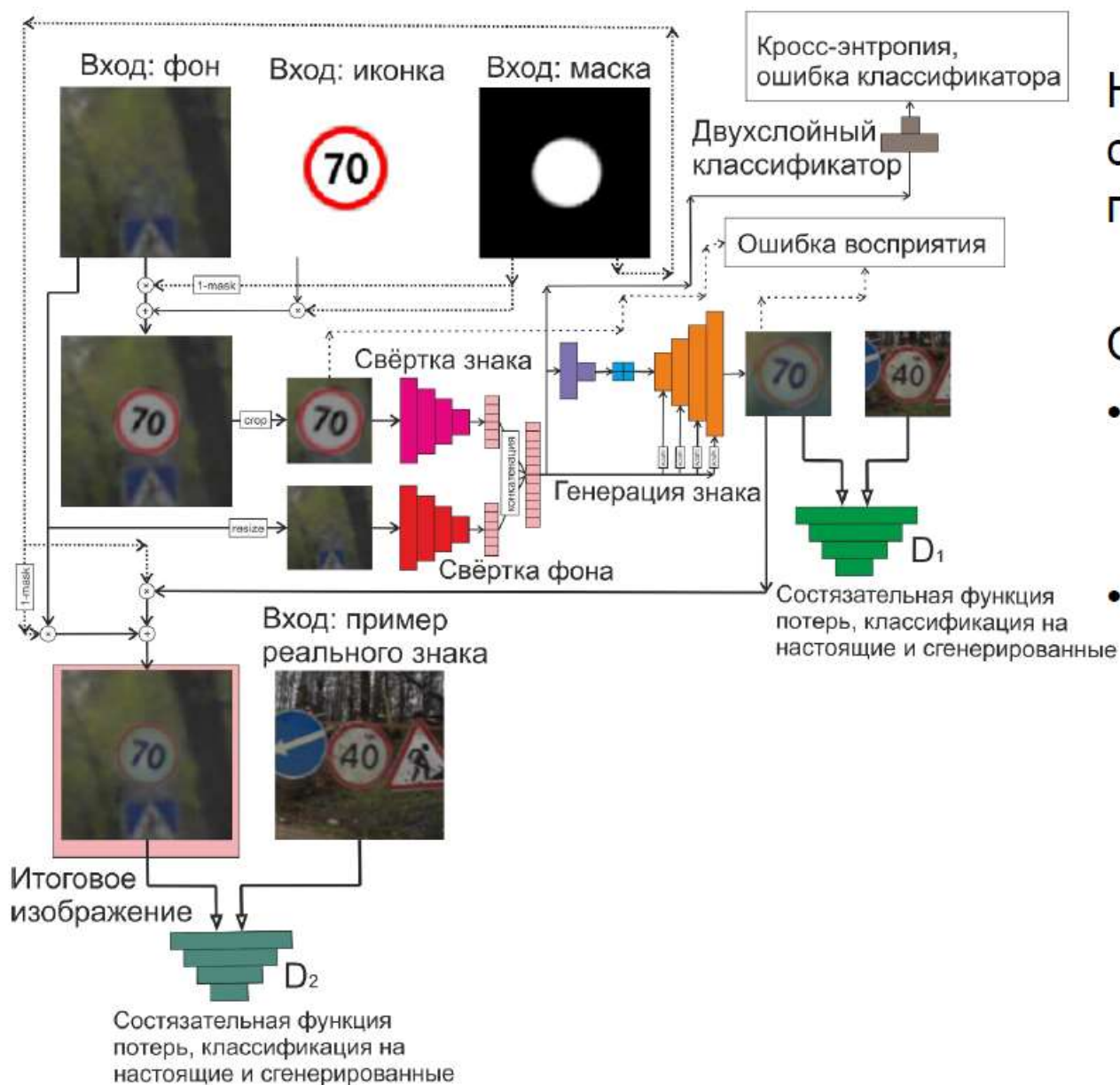
Исследование более сложных методов (Вариант 3)



Потоки данных в третьем подходе:



Третий подход



На входе - изображения с уже закрасненным предыдущим знаком

Обозначения:

- D_1 – классификация реальных знаков и синтетических
- D_2 – классификация знака с учетом фона

Примеры работы



№	icon	real	synt	cgi	cgi_gan	Первый метод	Второй метод	Третий метод
1								
2								
3								
4								
5								
6								

Примеры встраивания



Примеры встраивания



Примеры работы



Экспериментальная оценка



Для оценки качества обучалась система распознавания дорожных знаков:

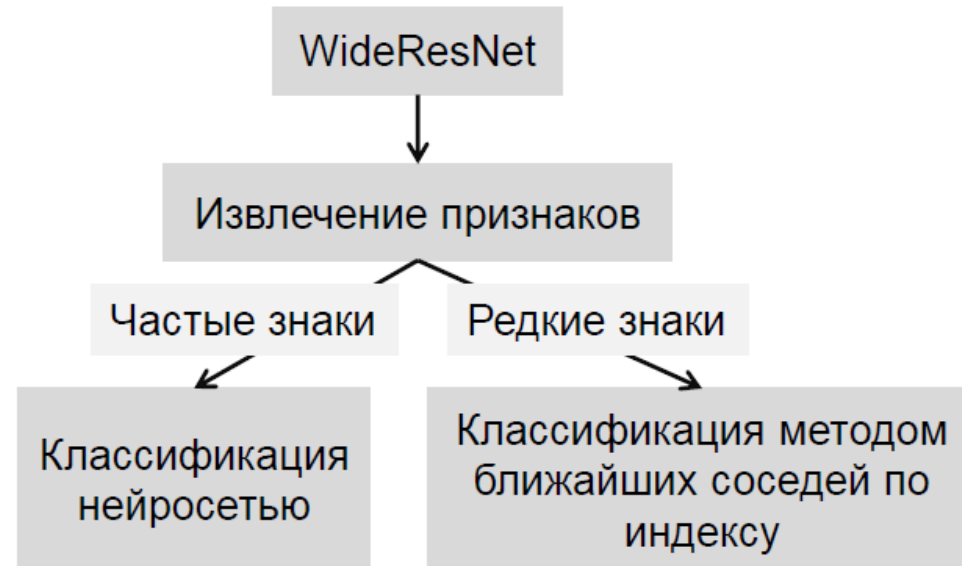
- **Детектор:** модифицированный для поиска дорожных знаков Fast-RCNN, в котором используются Inception-подобные блоки
- **Классификатор:** обычная и улучшенная ¹⁾ для задачи классификации редких дорожных знаков нейронная сеть WideResNet

Выборка RTSD для задачи детекции:

	Обучение	Тестирование
Картинок	47639	11389
Знаков	80277	25232

Выборка RTSD для задачи классификации:

	Обучение	Тестирование
Всего	79896	25613
Редких	0	1622
Частых	79896	23991



¹⁾ Фаизов, Б.В., Шахуро В.И., Санжаров В.В., Конушин А.С. Классификация редких дорожных знаков // Компьютерная оптика. – 2020. – Т. 44, № 2. – С. 236-243.

Экспериментальная оценка классификатора



	Обычный классификатор			Улучшенный с индексом из синтетики		
	все	редкие	частые	все	редкие	частые
rtsd	88,87	0,00	94,88			
rtsd + cgi_gan	92,75	53,82	95,39	93,52	70,16	95,09
rtsd + первый тип	91,67	68,74	93,22	93,84	74,97	95,11
rtsd + второй тип	92,03	68,19	93,64	93,98	75,46	95,23
rtsd + третий тип	92,82	69,67	94,39	94,11	76,33	95,31

только cgi_gan	49,8	43,90	50,19	60,55	54,69	60,94
только первый тип	67,50	69,05	67,39	72,57	72,13	72,60
только второй тип	73,61	65,60	74,15	72,35	71,89	72,38
только третий тип	69,94	69,42	69,97	73,03	73,98	72,96

Мерой выбрана точность классификации

- лучший результат

- предложенные методы

Экспериментальная оценка детектора



	Без классификатора			С классификатором		
	все	редкие	частые	все	редкие	частые
rtsd	89,09	85,86	89,25	86,01	58,56	86,61
rtsd + cgi	88,56	85,72	88,72	83,84	48,51	85,15
rtsd + inpaint	88,61	86,63	88,71	76,41	34,00	82,93
rtsd + первый тип	88,98	86,59	89,09	85,81	59,98	86,40
rtsd + второй тип	88,98	86,29	89,13	86,11	60,13	86,62
rtsd + третий тип	89,01	86,78	89,13	85,39	64,20	86,13

only cgi	10,70	13,23	10,63	8,81	8,53	8,81
only inpaint	55,23	55,26	56,26	15,89	13,79	15,96
только первый тип	38,22	38,97	38,17	19,85	18,80	19,91
только второй тип	37,20	42,97	37,13	25,88	24,42	25,94
только третий тип	62,12	54,26	62,36	39,99	32,51	40,35

Мерой выбрана площадь под кривой точности-полноты детектора (AUC)



- лучший результат



- предложенные методы

Резюме



- Пока доступная CGI не позволяет добиться фотореализма
- Обучение чисто на синтетических данных существенно проигрывает обучению на реальных данных
- Интересно фокусироваться на «простейших» реальных объектах или на решении таких задач, в которых мы рассматриваем производные признаки, не столь зависящих от реализма
- Добавление синтетики позволяет повысить качество распознавания в тех случаях, когда реальных изображений недостаточно