Доменная адаптация с помощью Style Transfer и GAN

Бетев Иван

образование

Примерчик

Манхэттен

Париж

Дак кто (who?) такая Domain Adaptation

Доменная адаптация - это дисциплина машинного обучения, которая имеет дело со сценариями, в которых модель, обученная на исходном распределении, используется в контексте другого (но связанного) целевого распределения

1

Machine Learning

2

Transfer Learning

3

Domain Adaptation

Почему не устраивает fine-tuning?

1) Повторное обучение многопараметрической глубокой сети на больших данных
2) Временные и вычислительные затраты
3) Неразмеченные данные

Синтетика

https://t.me/neural_wise_wolf_bot

Бенчмарки

Benchmarks

Add a Result

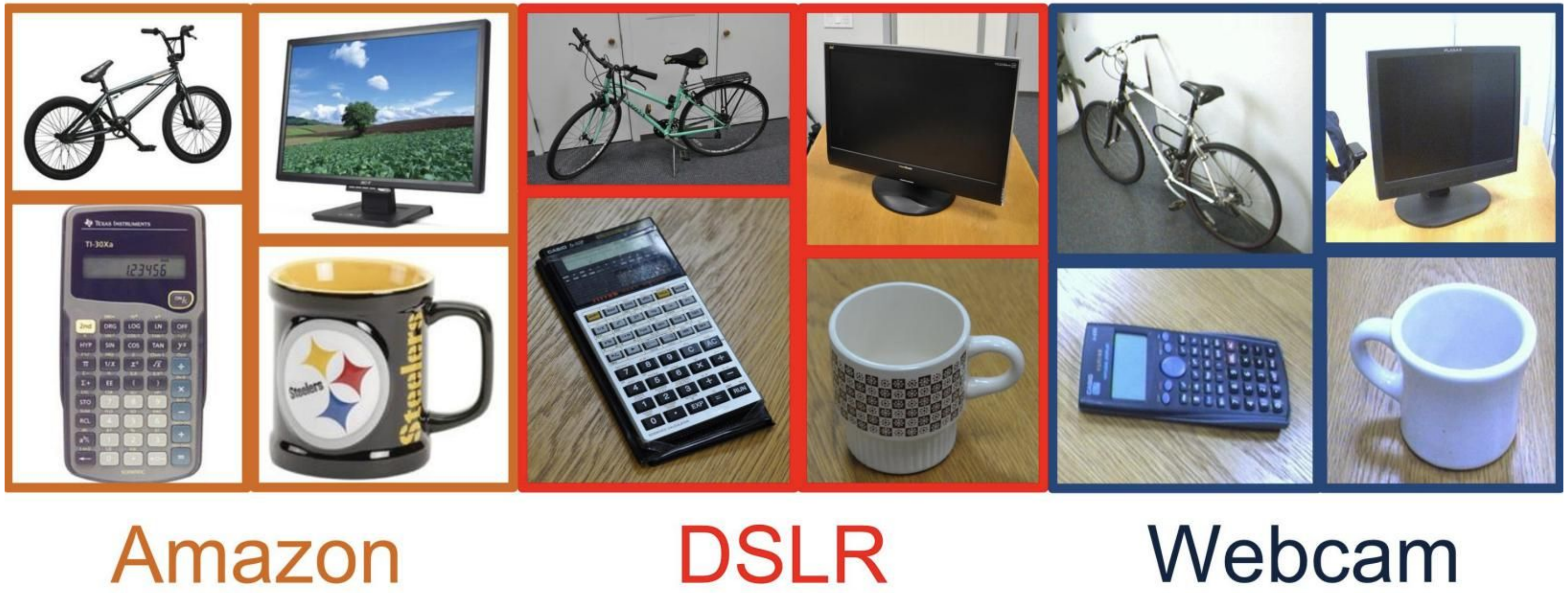
These leaderboards are used to track progress in Domain Adaptation

Trend	Dataset	Best Model	Paper	Code	Compare
	Office-31	CDTrans			See all
	SYNTHIA-to-Cityscapes	ILM-ASSL			See all
	GTA5 to Cityscapes	ILM-ASSL			See all
	VisDA2017	MIC			See all
	Office-Home	MIC			See all
	ImageCLEF-DA	MCC+NWD			See all
	MNIST-to-USPS	DFA-MCD			See all
	USPS-to-MNIST	SHOT			See all
	SVHN-to-MNIST	Mean teacher			See all
	SVNH-to-MNIST	SRDA (RAN)			See all

Show all 47 benchmarks

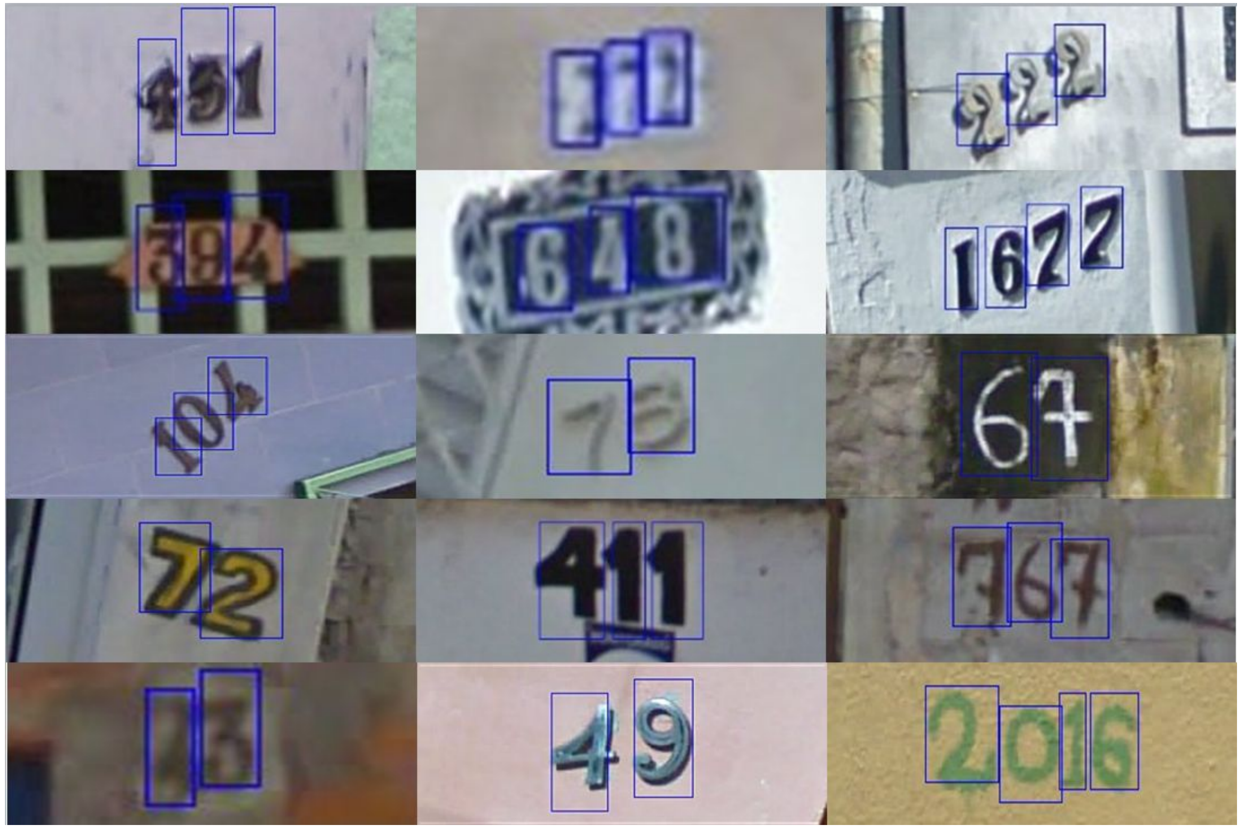
https://paperswithcode.com/task/domain-adaptation/latest

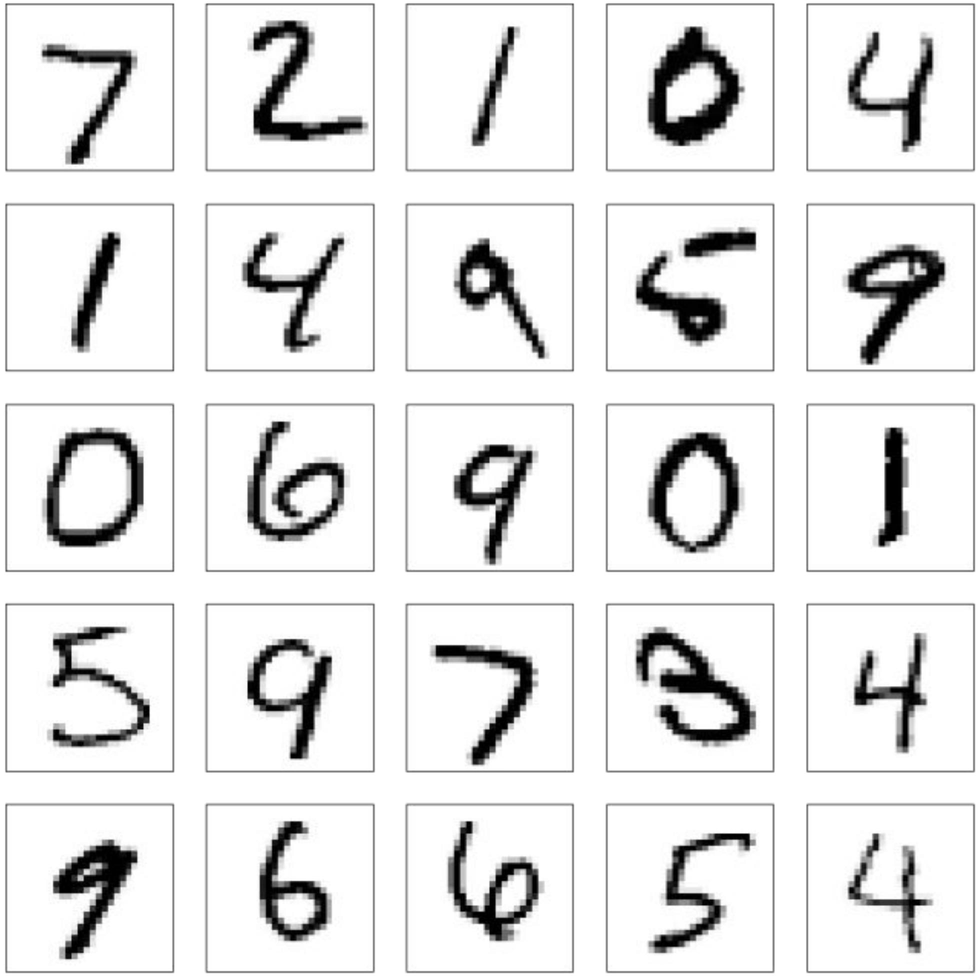
Office-31



https://faculty.cc.gatech.edu/~judy/domainadapt/#datasets_code

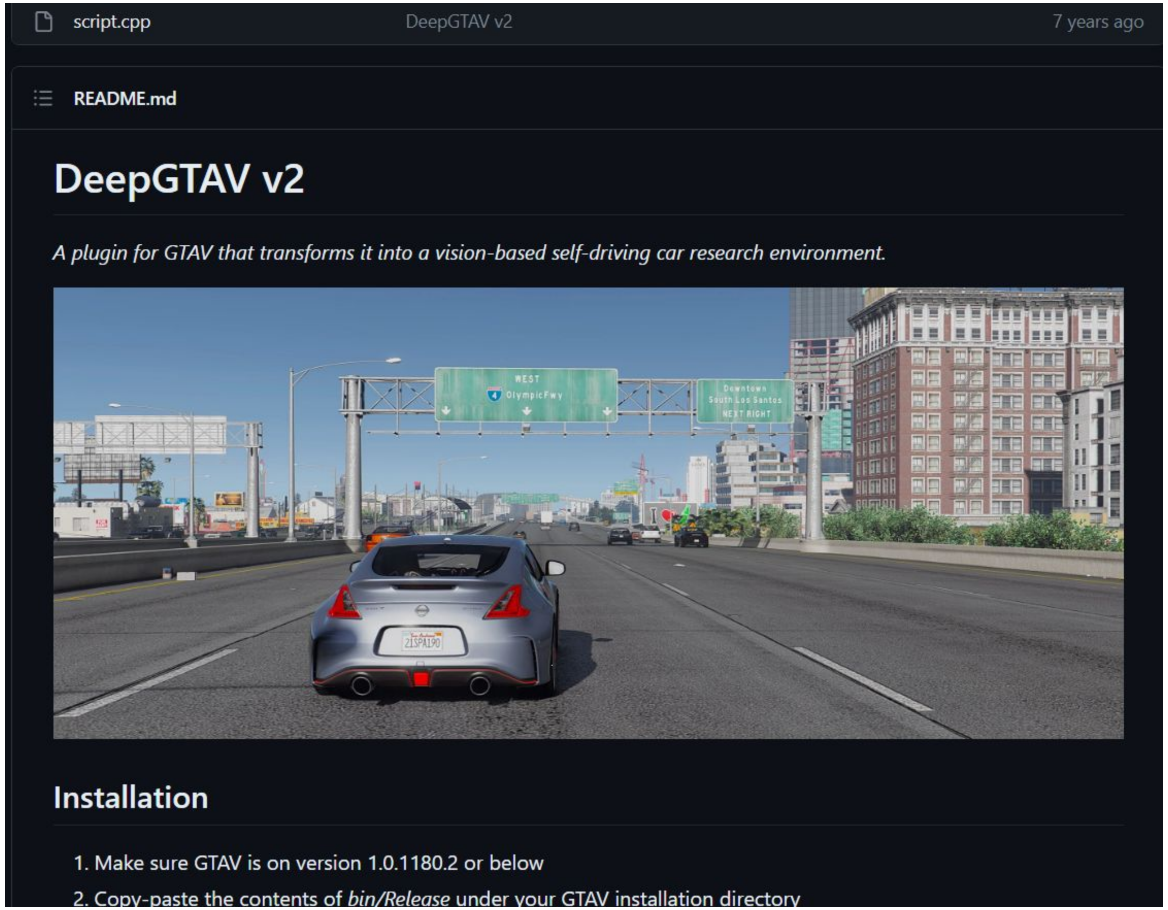
SVNH-to-MNIST





http://ufldl.stanford.edu/housenumbers/

GTA5 to Cityscapes



https://github.com/aitorzip/DeepGTAV

https://arxiv.org/pdf/1608.02192v1.pdf

Виды доменной адаптации

One-step DA Approaches	Brief Description	Subsettings
Discrepancy-based	fine-tuning the deep network with labeled or unlabeled target data to diminish the domain shift	class criterion [118], [86], [79], [98] [53], [45], [75], [139], [130], [29], [118], [28]
		statistic criterion [74], [130], [73] [75], [120], [32], [109], [87], [144]
		architecture criterion [69], [54], [68], [95], [128], [89]
		geometric criterion [16]
Adversarial-based	using domain discriminators to encourage domain confusion through an adversarial objective	generative models [70], [4], [57]
		non-generative models [119], [118], [26], [25], [117] [85]
Reconstruction-based	using the data reconstruction as an auxiliary task to ensure feature invariance	encoder-decoder reconstruction [5], [33], [31], [144]
		adversarial reconstruction [131], [143], [59]

Multi-step Approaches	Brief Description
Hand-crafted	users determine the intermediate domains based on experience [129]
Instance-based	selecting certain parts of data from the auxiliary datasets to compose the intermediate domains [114], [16]
Representation-based	freeze weights of one network and use their intermediate representations as input to the new network [96]

https://arxiv.org/pdf/1802.03601.pdf

Adversarial

- **Adversarial Reconstruction:** the reconstruction error is measured as the difference between the reconstructed and original images within each image domain by a cyclic mapping obtained via a GAN discriminator, such as dual GAN [131], cycle GAN [143] and disco GAN [59].

Цель работы

Real-Time Monocular Depth Estimation using Synthetic Data with Domain Adaptation via Image Style Transfer

Amir Atapour-Abarghouei[1] Toby P. Breckon[1,2]
[1]Department of Computer Science – [2]Department of Engineering
Durham University, UK
{amir.atapour-abarghouei,toby.breckon}@durham.ac.uk

Abstract

Monocular depth estimation using learning-based approaches has become promising in recent years. However, most monocular depth estimators either need to rely on large quantities of ground truth depth data, which is extremely expensive and difficult to obtain, or predict disparity as an intermediary step using a secondary supervisory signal leading to blurring and other artefacts. Training a depth estimation model using pixel-perfect synthetic data can resolve most of these issues but introduces the problem of domain bias. This is the inability to apply a model trained on synthetic data to real-world scenarios. With advances in image style transfer and its connections with domain adaptation (Maximum Mean Discrepancy), we take advantage of style transfer and adversarial training to predict pixel perfect depth from a single real-world color image based on training over a large corpus of synthetic environment data. Experimental results indicate the efficacy of our approach compared to contemporary state-of-the-art techniques.

1. Introduction

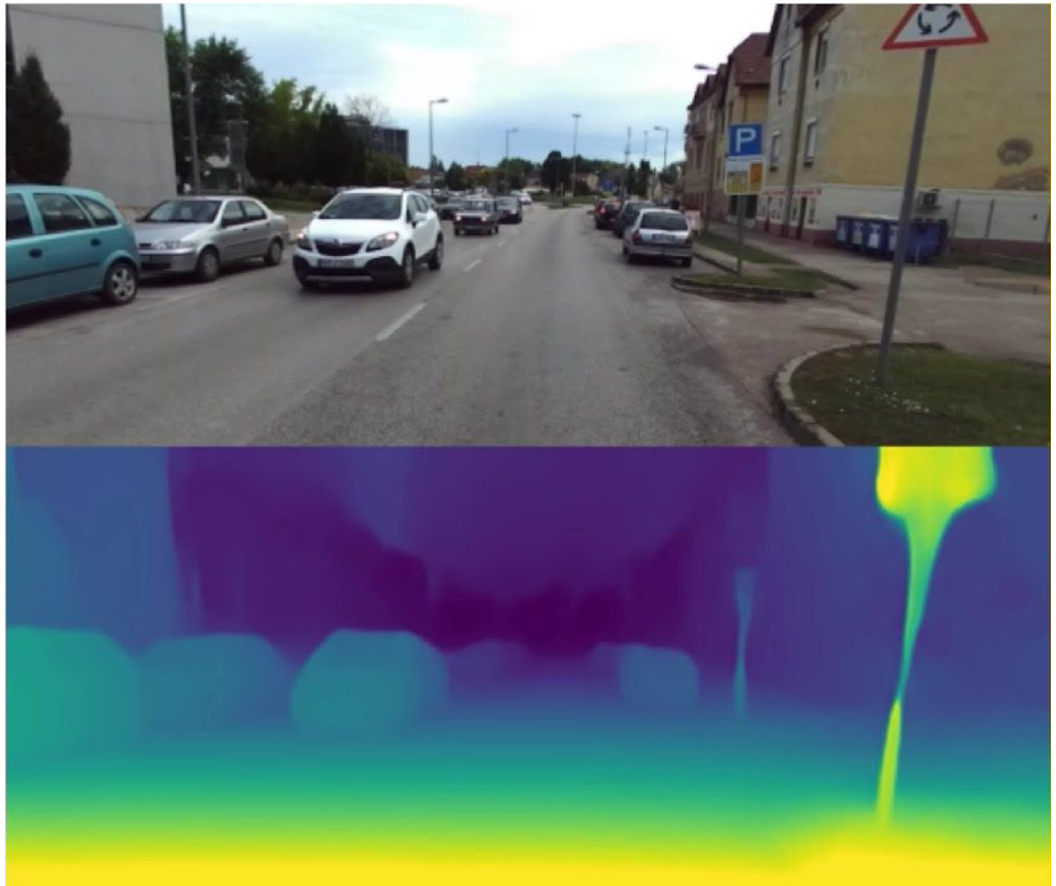
As 3D imagery has become the staple requirement within many computer vision applications, accurate and ef-

Figure 1: Our monocular depth estimation (KITTI [55]).

learning approaches have recently emerged that take advantage of off-line training on ground truth depth data to make monocular depth prediction possible [39, 48, 18, 17, 42, 91]. However, since ground truth depth is extremely difficult and expensive to acquire in the real world, when it is obtained it is often sparse and flawed, constraining the practical use of many of these approaches.

Other monocular approaches, sometimes referred to as *unsupervised*, do not require direct ground truth depth, but instead utilize a secondary supervisory signal during training which indirectly results in producing the desired depth [26, 22, 83, 12]. Training data for these approaches is abundant and easily obtainable but they suffer from undesirable artefacts, such as blurring and incoherent content, due to the nature of their secondary supervision.

However, an often overlooked fact is that the same tech-



https://breckon.org/toby/publications/papers/abarghouei18monocular.pdf

Датасет - Source domen

The KITTI Vision Benchmark Suite

A project of Karlsruhe Institute of Technology and Toyota Technological Institute at Chicago

home setup stereo flow sceneflow depth odometry object tracking road semantics raw data submit results

A. Geiger | P. Lenz | C. Stiller | R. Urtasun — Log in

Road/Lane Detection Evaluation 2013

This benchmark has been created in collaboration with Jannik Fritsch and Tobias Kuehnl from Honda Research Institute Europe GmbH. The road and lane estimation benchmark consists of 289 training and 290 test images. It contains three different categories of road scenes:

- uu - urban unmarked (98/100)
- um - urban marked (95/96)
- umm - urban multiple marked lanes (96/94)
- urban - combination of the three above

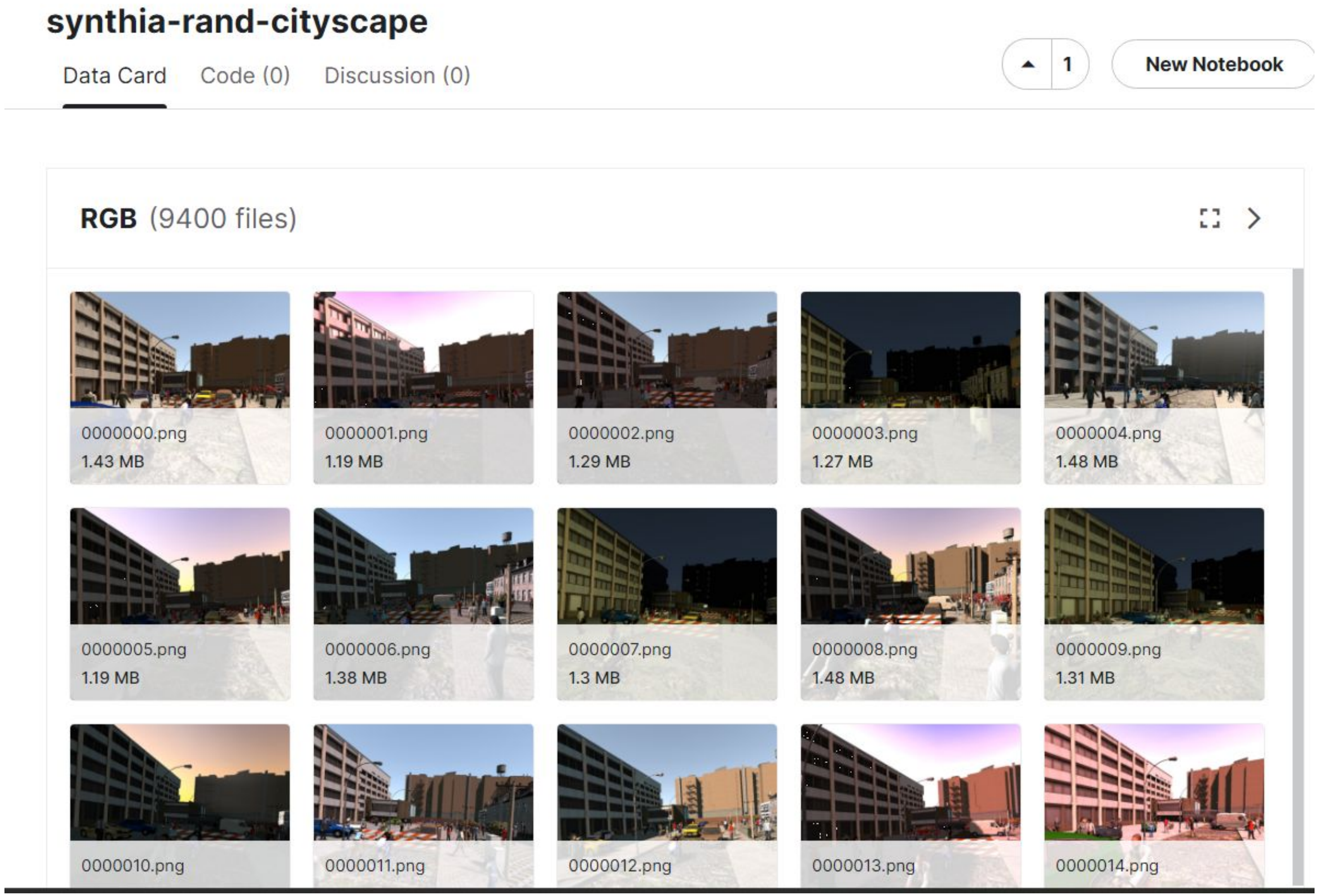
Ground truth has been generated by manual annotation of the images and is available for two different road terrain types: road - the road area, i.e, the composition of all lanes, and lane - the ego-lane, i.e., the lane the vehicle is currently driving on (only available for category "um"). Ground truth is provided for training images only.

- Download base kit with: left color images, calibration and training labels (0.5 GB)
- Download right color image extension (0.5 GB)

https://www.cvlibs.net/datasets/kitti/eval_road.php

Датасет - Target Domain

SYNTHIA, The ***SYNTHetic collection of Imagery and Annotations***



https://www.kaggle.com/datasets/pengweili/synthiarandcityscape

Pipeline

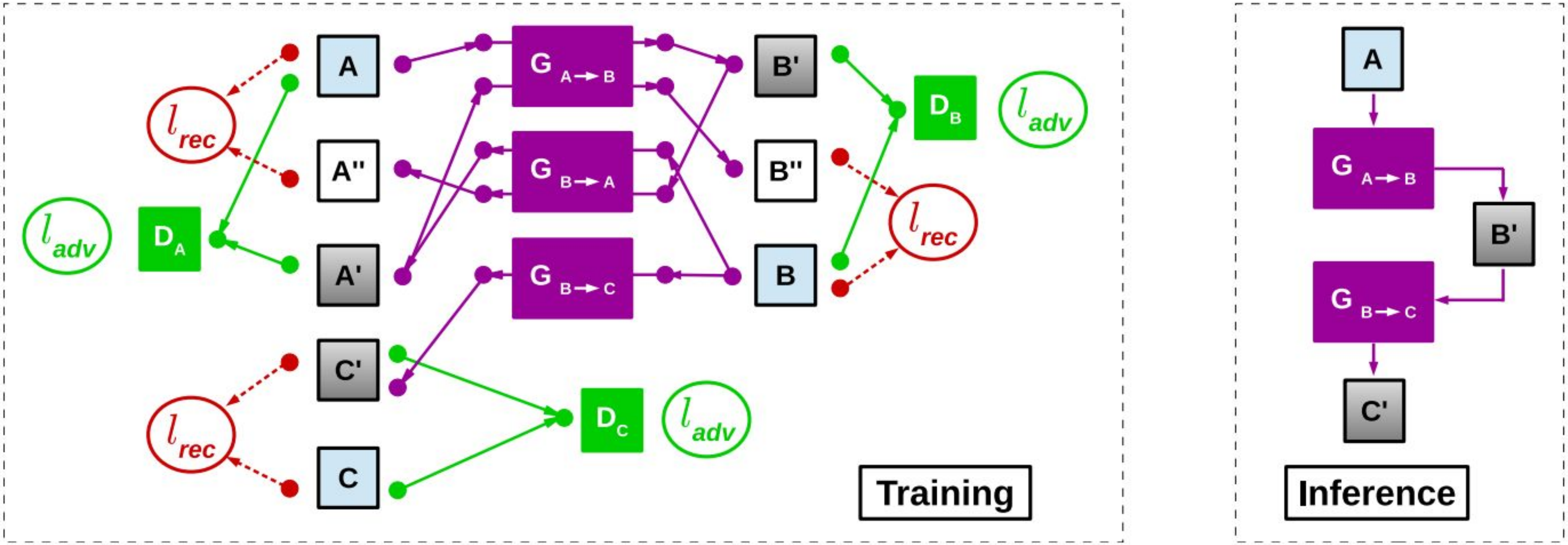


Figure 2: Our approach using [90]. Domain A (real-world RGB) is transformed into B (synthetic RGB) and then to C (pixel-perfect depth). A, B, C denote ground truth, A', B', C' generated images, and A'', B'' cyclically regenerated images.

1 stage - monocular depth estimation model

RGB $\xrightarrow{\text{mapping}}$ depth

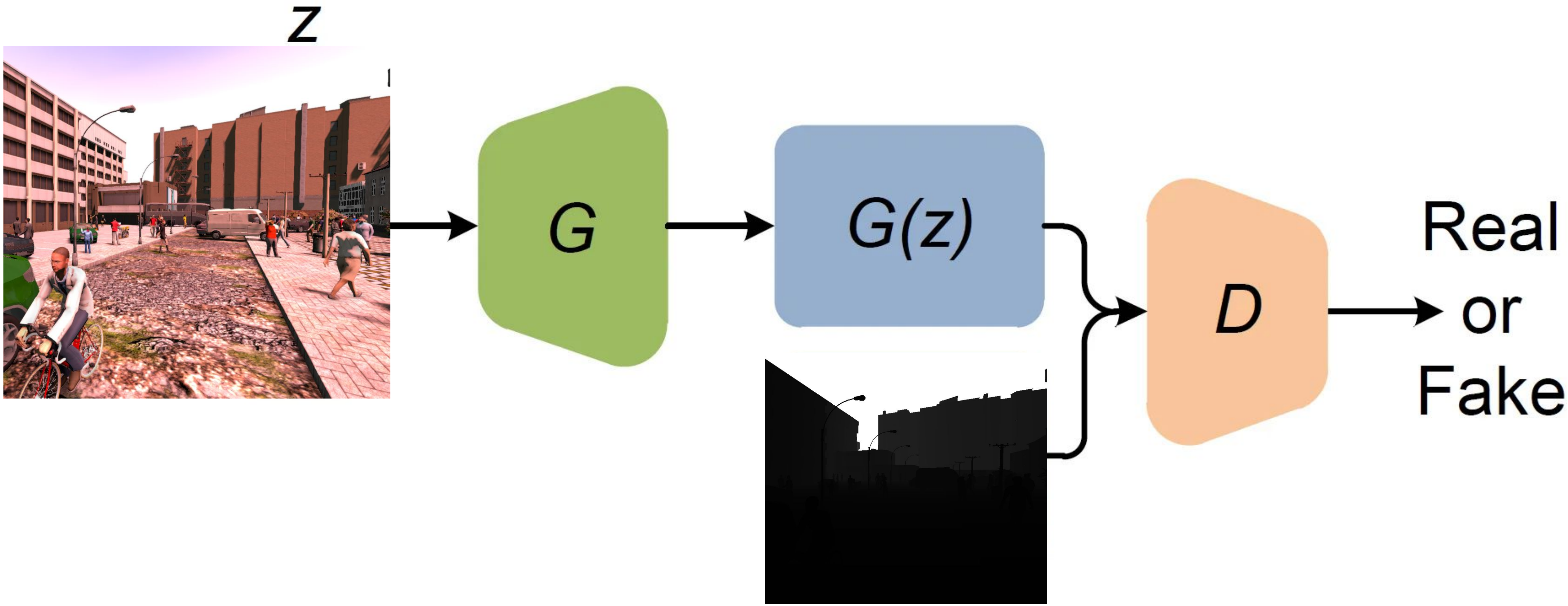
Наивное решение:

- НН, минимизирующая reconstruction loss (Euclidean distance) между значениями пикселей output и ground truth

Проблемы:
Модель, обученная предсказывать глубину через l1 или l2, склонна генерировать значения, являющимися средним между всеми возможными. Результаты получаются очень размытыми

Решение проблемы - GAN!

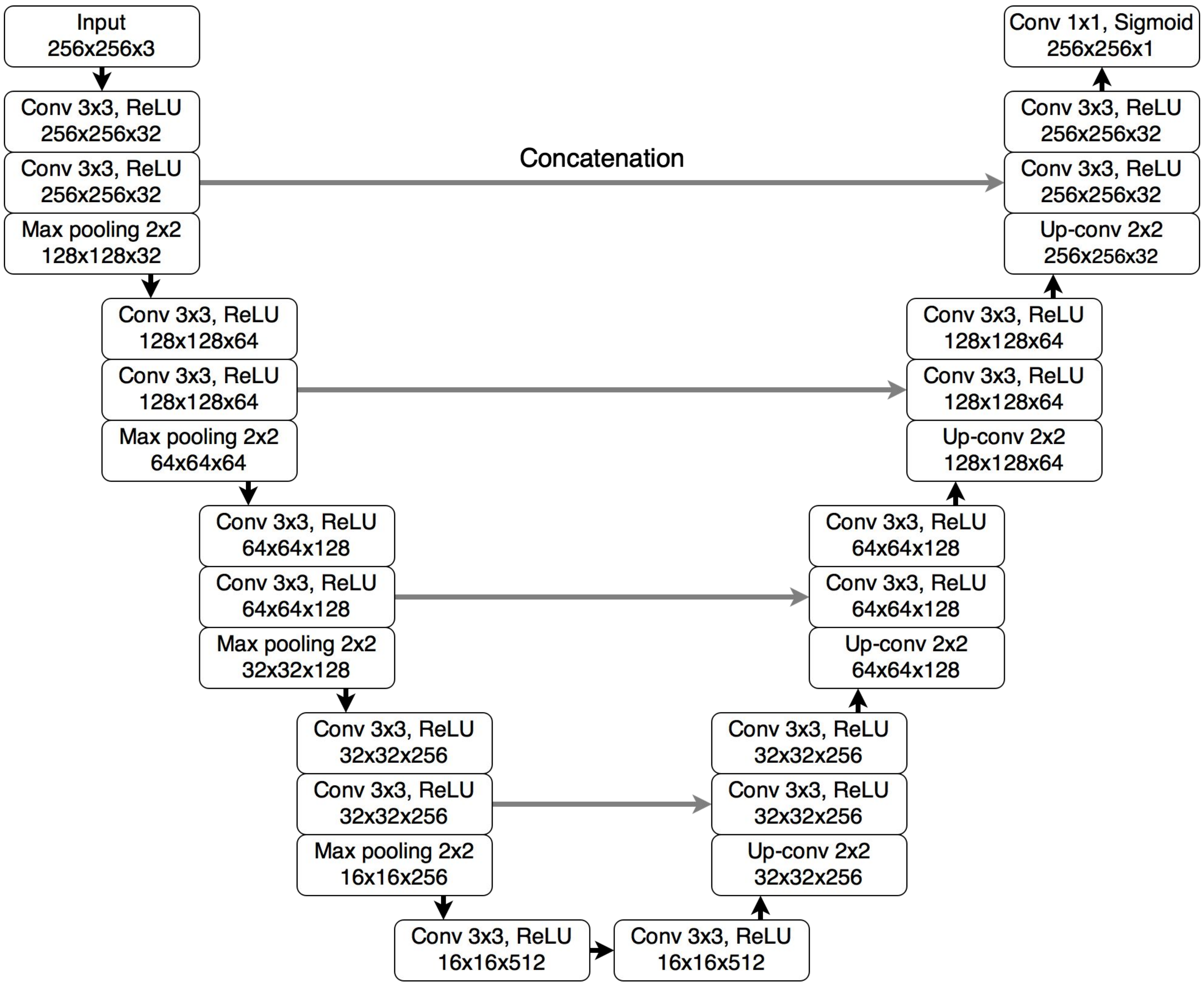
Добавим состязательную часть обучения



Генератор

1) енкодер - декодер
2) skip connections

UNET



Дискриминатор

Conv2d
BatchNorm
LeakyReLU(0.2)

x5 + fc fc

Лосс

$$\mathcal{L}_{rec} = ||G(x) - y||_1$$

$$\mathcal{L}_{adv} = \min_G \max_D \mathop{\mathbb{E}}_{x,y\sim\mathbb{P}_d(x,y)}[logD(x,y)] + \mathop{\mathbb{E}}_{x\sim\mathbb{P}_d(x)}[log(1 - D(x, G(x)))]$$

$$\mathcal{L} = \lambda\mathcal{L}_{rec} + (1-\lambda)\mathcal{L}_{adv}$$

Время поумничать

Algorithm 1 WGAN with gradient penalty. We use default values of $\lambda = 10$, $n_{\text{critic}} = 5$, $\alpha = 0.0001$, $\beta_1 = 0$, $\beta_2 = 0.9$.

Require: The gradient penalty coefficient λ, the number of critic iterations per generator iteration n_{critic}, the batch size m, Adam hyperparameters α, β_1, β_2.
Require: initial critic parameters w_0, initial generator parameters θ_0.

while θ has not converged **do**
    for $t = 1, \ldots, n_{\text{critic}}$ **do**
        for $i = 1, \ldots, m$ **do**
            Sample real data $\boldsymbol{x} \sim \mathbb{P}_r$, latent variable $\boldsymbol{z} \sim p(\boldsymbol{z})$, a random number $\epsilon \sim U[0, 1]$.
            $\tilde{\boldsymbol{x}} \leftarrow G_\theta(\boldsymbol{z})$
            $\hat{\boldsymbol{x}} \leftarrow \epsilon \boldsymbol{x} + (1 - \epsilon)\tilde{\boldsymbol{x}}$
            $L^{(i)} \leftarrow D_w(\tilde{\boldsymbol{x}}) - D_w(\boldsymbol{x}) + \lambda(\|\nabla_{\hat{\boldsymbol{x}}} D_w(\hat{\boldsymbol{x}})\|_2 - 1)^2$
        end for
        $w \leftarrow \text{Adam}(\nabla_w \frac{1}{m} \sum_{i=1}^{m} L^{(i)}, w, \alpha, \beta_1, \beta_2)$
    end for
    Sample a batch of latent variables $\{\boldsymbol{z}^{(i)}\}_{i=1}^{m} \sim p(\boldsymbol{z})$.
    $\theta \leftarrow \text{Adam}(\nabla_\theta \frac{1}{m} \sum_{i=1}^{m} -D_w(G_\theta(\boldsymbol{z})), \theta, \alpha, \beta_1, \beta_2)$
end while

А давайте заменим Adversarial Loss из Vanilla GAN на WGAN-GP

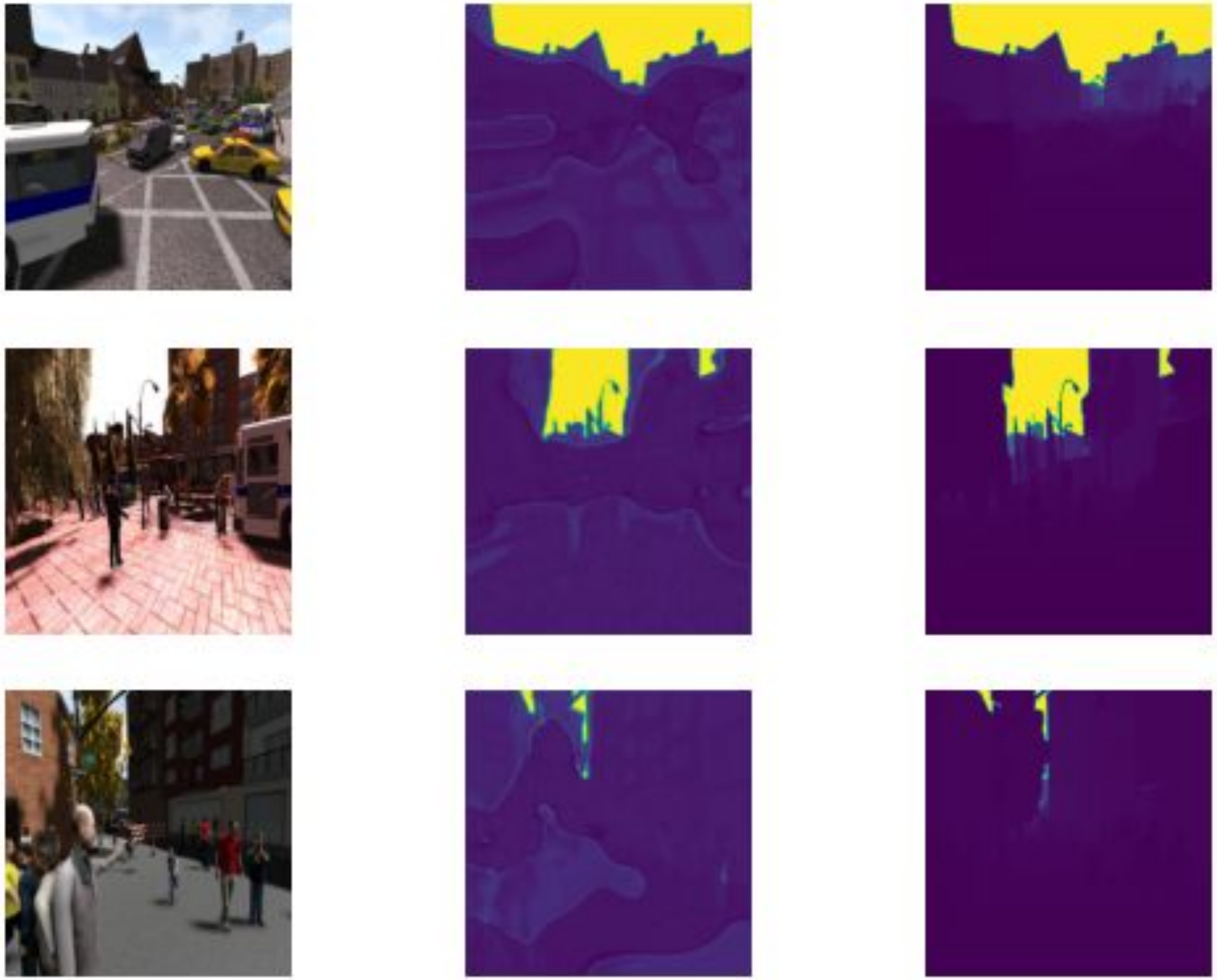
Особенности обучения

- ADAM
- b1 = 0.5
- b2 = 0.999
- lr = 0.0002
- lambda = 0.9 (в статье 0.99)

Аугментации

- RandomGorizontalFlip

Результаты



RMSE

Method	Training Data	Error Metrics (lower, better)				Accuracy Metrics (higher, better)		
		Abs. Rel.	Sq. Rel.	RMSE	RMSE log	$\sigma < 1.25$	$\sigma < 1.25^2$	$\sigma < 1.25^3$
Train Set Mean	K	0.403	0.530	8.709	0.403	0.593	0.776	0.878
Eigen *et al*. Coarse	K	0.214	1.605	6.563	0.292	0.673	0.884	0.957
Eigen *et al*. Fine	K	0.203	1.548	6.307	0.282	0.702	0.890	0.958
Liu *et al*.	K	0.202	1.614	6.523	0.275	0.678	0.895	0.965
Zhou *et al*.	K	0.208	1.768	6.856	0.283	0.678	0.885	0.957
Zhou *et al*.	K+CS	0.198	1.836	6.565	0.275	0.718	0.901	0.960
Godard *et al*.	K	0.148	1.344	5.927	0.247	0.803	0.922	0.964
Godard *et al*.	K+CS	0.124	1.076	5.311	0.219	0.847	0.942	0.973
Our Approach	K+S*	**0.110**	**0.929**	**4.726**	**0.194**	**0.923**	**0.967**	**0.984**

100%

: 9.161529301712905

Stage 2 - Style Transfer aka Domain Adaptation

Хотим выучить mapping функцию $\mathcal{D} : X \to Y$

X - Реальные изображения
Y - Синтетические данные

В терминах идентичности распределении D(X) и Y
После этого спокойно можно применять depth estimation model

В нашем случае используем adversarial и cycle-consistently training
для перевода двух множеств неразмеченных данных из разных доменов

CyCADA

CyCADA: Cycle-Consistent Adversarial Domain Adaptation

Judy Hoffman, Eric Tzeng, Taesung Park, Jun-Yan Zhu
BAIR, UC Berkeley
{jhoffman,etzeng,taesung_park,junyanz}@eecs.berkeley

Phillip Isola
OpenAI*
isola@eecs.berkeley

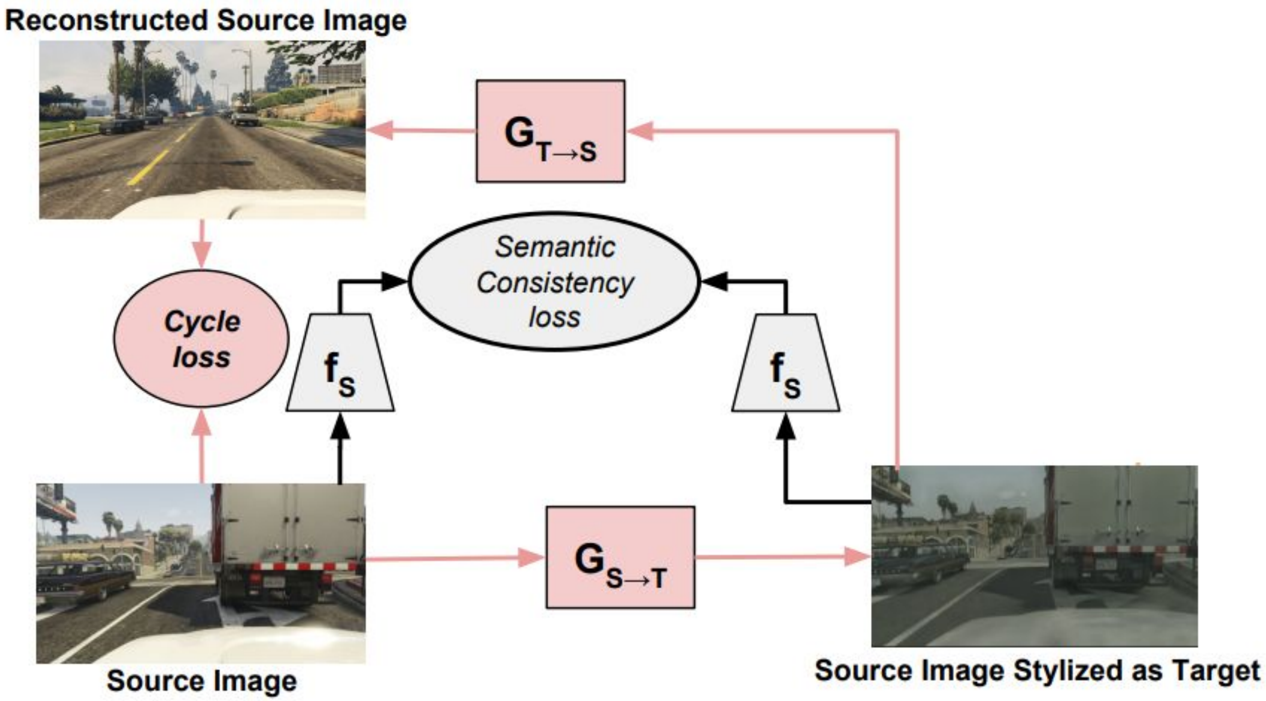
Kate Saenko
CS, Boston University
saenko@bu

Alexei A. Efros, Trevor Darrell
BAIR, UC Berkeley
{efros,trevor}@eecs.berkeley

Abstract

Domain adaptation is critical for success in new, unseen environments. Adversarial adaptation models applied in feature spaces discover domain invariant representations, but are difficult to visualize and sometimes fail to capture pixel-level and low-level domain shifts. Recent work has shown that generative adversarial networks combined with cycle-consistency constraints are surprisingly effective at mapping images between domains, even without the use of aligned image pairs. We propose a novel discriminatively-trained Cycle-Consistent Adversarial Domain Adaptation model. CyCADA adapts representations at both the pixel-level and feature-level, enforces cycle-consistency while leveraging a task loss, and does not require aligned pairs. Our model can be applied in a variety of visual recognition and prediction settings. We show new state-of-the-art results across multiple adaptation tasks, including digit classification and semantic segmentation of road scenes demonstrating transfer from synthetic to real world domains.

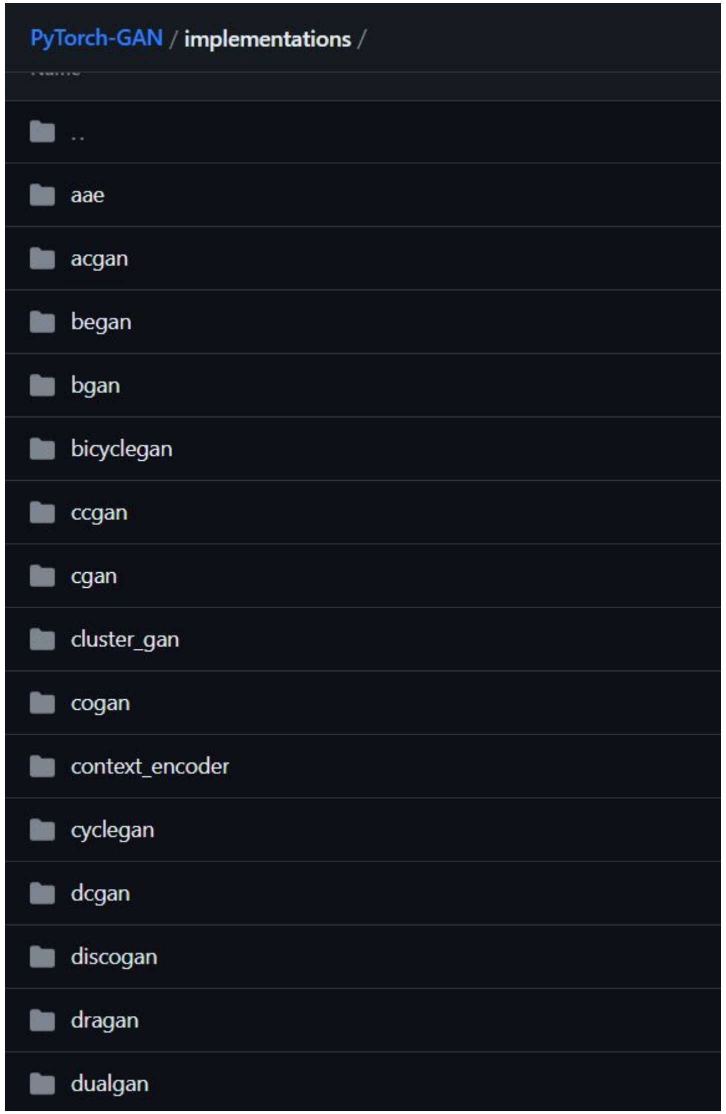
1 Introduction



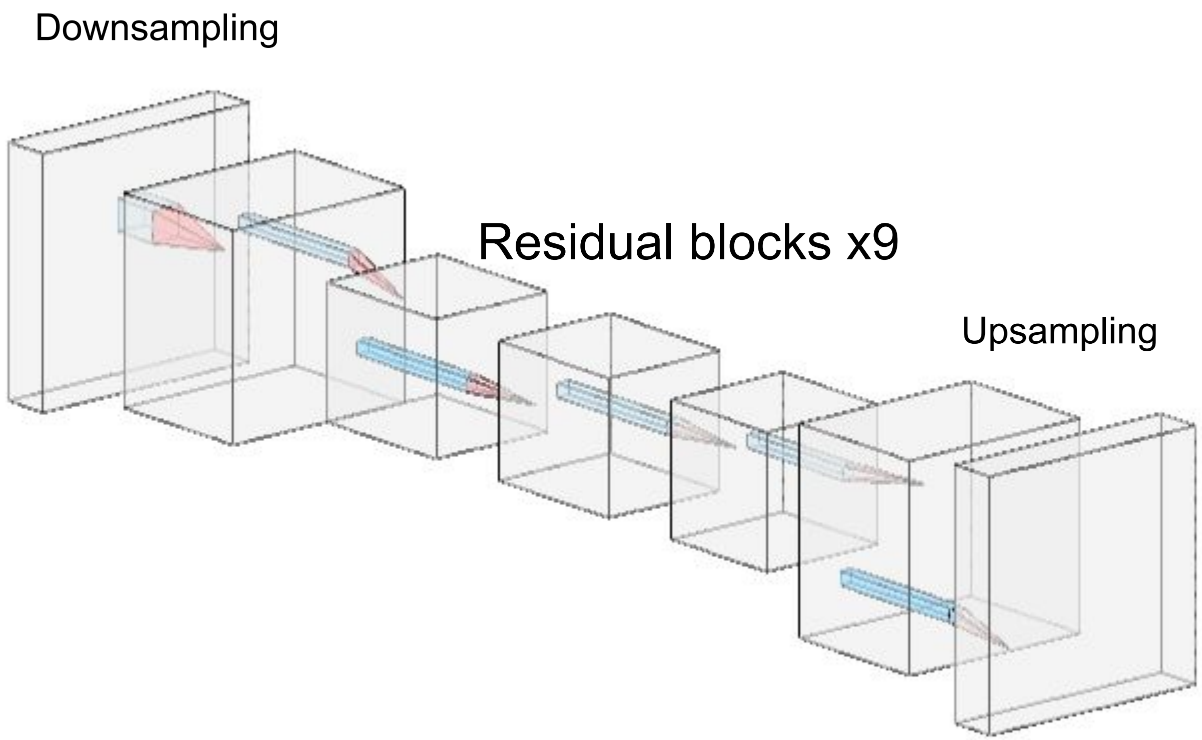
https://arxiv.org/pdf/1711.03213.pdf

Отдадим должное

https://github.com/eriklindernoren/PyTorch-GAN/tree/master



Генераторы



http://alexlenail.me/NN-SVG/AlexNet.html

Дискриминаторы

Conv2d
BatchNorm
LeakyReLU(0.2)

x5 + fc fc

Лосс

$$\mathcal{L}_{adv}(X \rightarrow Y) = \min_{G_{XtoY}} \max_{D_Y} \mathop{\mathbb{E}}_{y \sim \mathbb{P}_d(y)} [log D_Y(y)] + \mathop{\mathbb{E}}_{x \sim \mathbb{P}_d(x)} [log(1 - D_Y(G_{XtoY}(x)))]$$

$$\mathcal{L}_{cyc} = ||G_{YtoX}(G_{XtoY}(x)) - x||_1 + ||G_{XtoY}(G_{YtoX}(y)) - y||_1$$

$$\mathcal{L}_{adv}(Y \rightarrow X) = \min_{G_{YtoX}} \max_{D_X} \mathop{\mathbb{E}}_{x \sim \mathbb{P}_d(x)} [log D_X(x)] + \mathop{\mathbb{E}}_{y \sim \mathbb{P}_d(y)} [log(1 - D_X(G_{YtoX}(y)))]$$

$$\mathcal{L} = \mathcal{L}_{adv}(X \rightarrow Y) + \mathcal{L}_{adv}(Y \rightarrow X) + \lambda \mathcal{L}_{cyc}$$

Снова умничаем

Так как на выходе мы должны получить изображение аналогичного контента, добавим Identity Loss в виде l1 с коэффициентом lambda_id = 5.

loss_identity = 0.5 * (criterion_identity(G_BA(real_A), real_A) + criterion_identity(G_AB(real_B), real_B))

Pipeline обучения

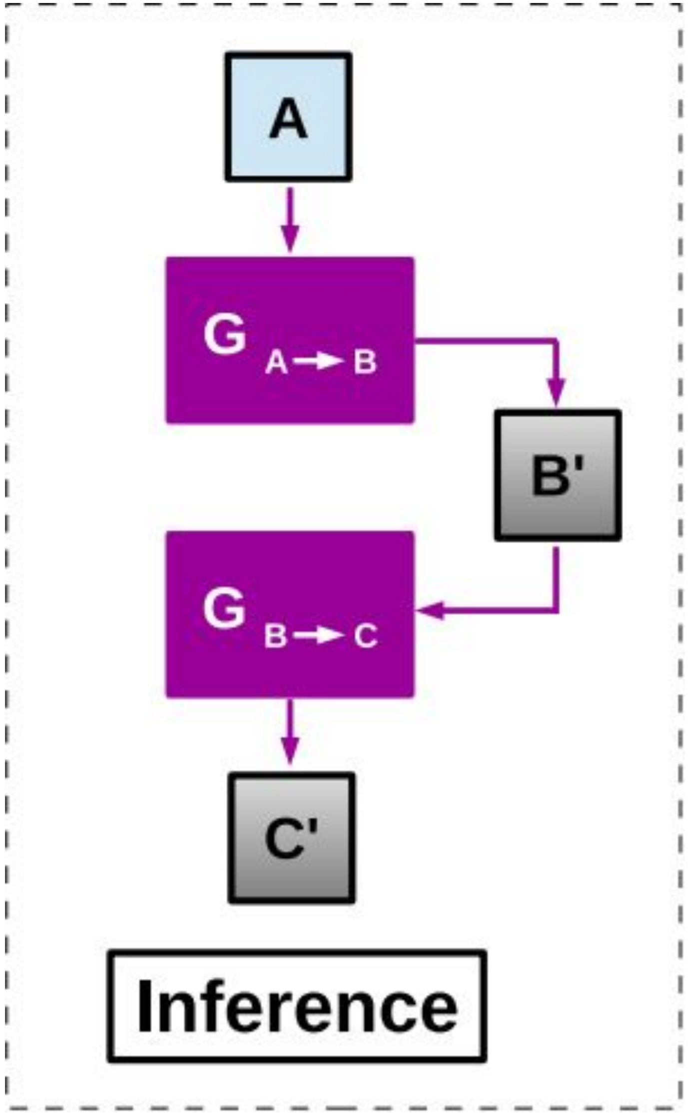
1) прогоняем генераторы на Identity Loss
2) по стандартной форме прогнали генераторы и посчитали ошибку
3) циклично прогоняем и считаем cycle loss
4) учим дискриминаторы стандартно

Особенности обучения

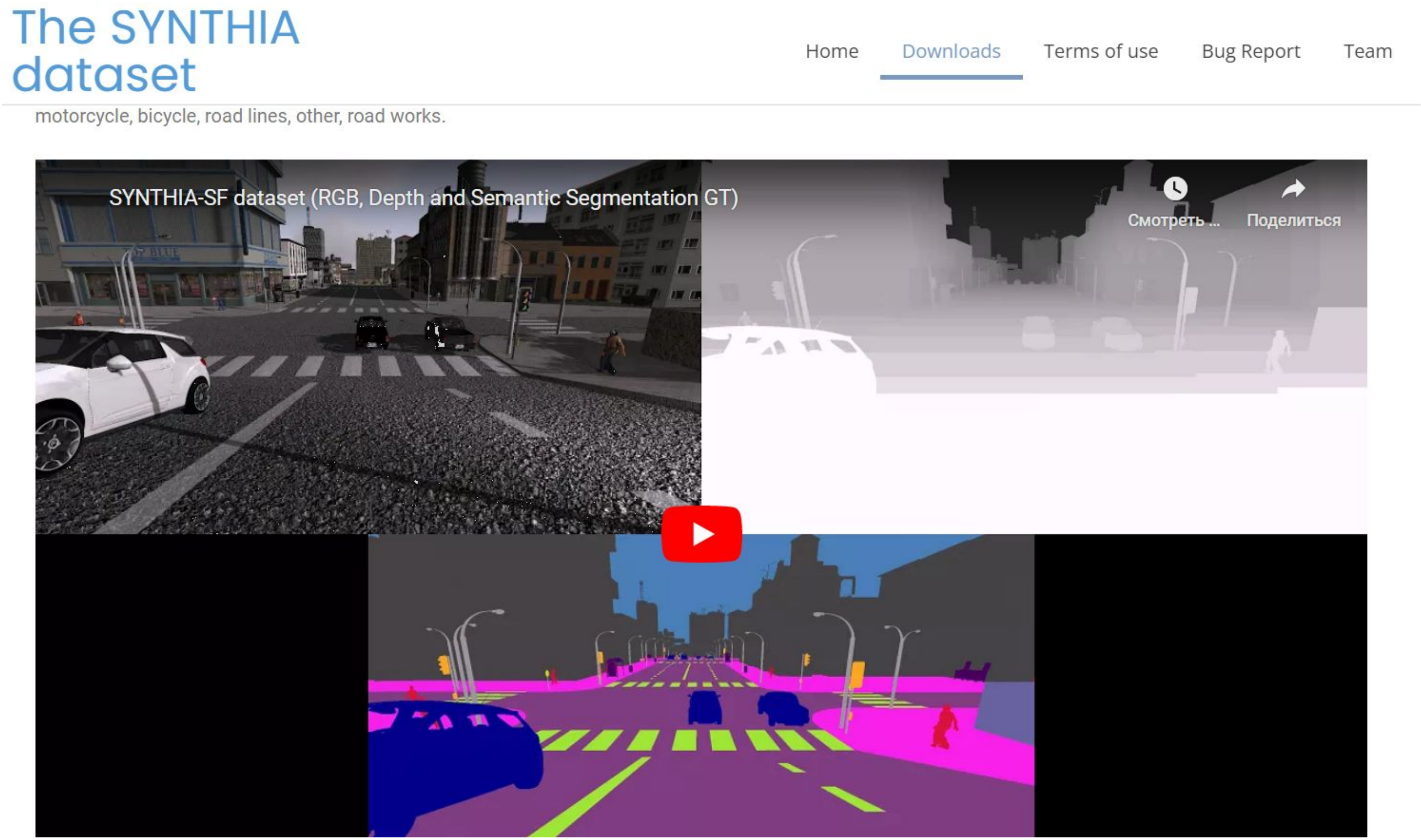
- ADAM
- b1 = 0.5
- b2 = 0.999
- lr = 0.0001
- lambda_cyc = 10

Результаты

Объединяем в одно целое



Датасет (((



https://synthia-dataset.net/downloads/

Что останется за рамками ...

1) обучить на хорошем датасете
2) эксперименты с архитектурами
3) эксперименты с параметрами
4) добавить Wasserstein dist в second stage

GitHub

https://github.com/betyavan/HSE_DL_project

Спасибо!

образование