DatasetGAN: Efficient Labeled Data Factory with Minimal Human Effort [1]

Pershin Maxim, Petrovich Sergey, Nuriev Ainur, Baranovskaya Daria

Higher School of Economics, 2021

We introduce DatasetGAN: an automatic procedure to generate massive datasets of high-quality semantically segmented images requiring minimal human effort. Current deep networks are extremely data-hungry, benefiting from training on large-scale datasets, which are time consuming to annotate. Our method relies on the power of recent GANs to generate realistic images. We show how the GAN latent code can be decoded to produce a semantic segmentation of the image. Training the decoder only needs a few labeled examples to generalize to the rest of the latent space, resulting in an infinite annotated dataset generator! These generated datasets can then be used for training any computer vision architecture just as real datasets are. As only a few images need to be manually segmented, it becomes possible to annotate images in extreme detail and generate datasets with rich object and part segmentations. To showcase the power of our approach, we generated datasets for 7 image segmentation tasks which include pixel-level labels for 34 human face parts, and 32 car parts. Our approach outperforms all semi-supervised baselines significantly and is on par with fully supervised methods, which in some cases require as much as 100x more annotated data as our method.

Introduction

DatasetGAN: Efficient Labeled Data Factory with Minimal Human Effort

Yuxuan Zhang[1,5*] Huan Ling[1,2,3,*] Jun Gao[1,2,3] Kangxue Yin [1]
Jean-Francois Lafleche[1] Adela Barriuso[4] Antonio Torralba[4] Sanja Fidler[1,2,3]
NVIDIA[1] University of Toronto[2] Vector Institute[3] MIT[4] University of Waterloo[5]

y2536zha@uwaterloo.ca, {huling, jung, kangxuey, jlafleche}@nvidia.com
adela.barriuso@gmail.com, torralba@mit.edu, sfidler@nvidia.com

90v2 [cs.CV] 20 Apr 2021

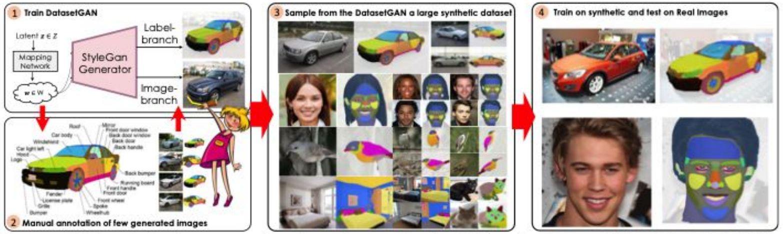


Figure 1: DATASETGAN synthesizes image-annotation pairs, and can produce large high-quality datasets with detailed pixel-wise labels. Figure illustrates the 4 steps. *(1 & 2).* Leverage StyleGAN and annotate only a handful of synthesized images. Train a highly effective branch to generate labels. *(3).* Generate a huge synthetic dataset of annotated images authomatically. *(4).* Train your favorite approach with the synthetic dataset and test on real images.

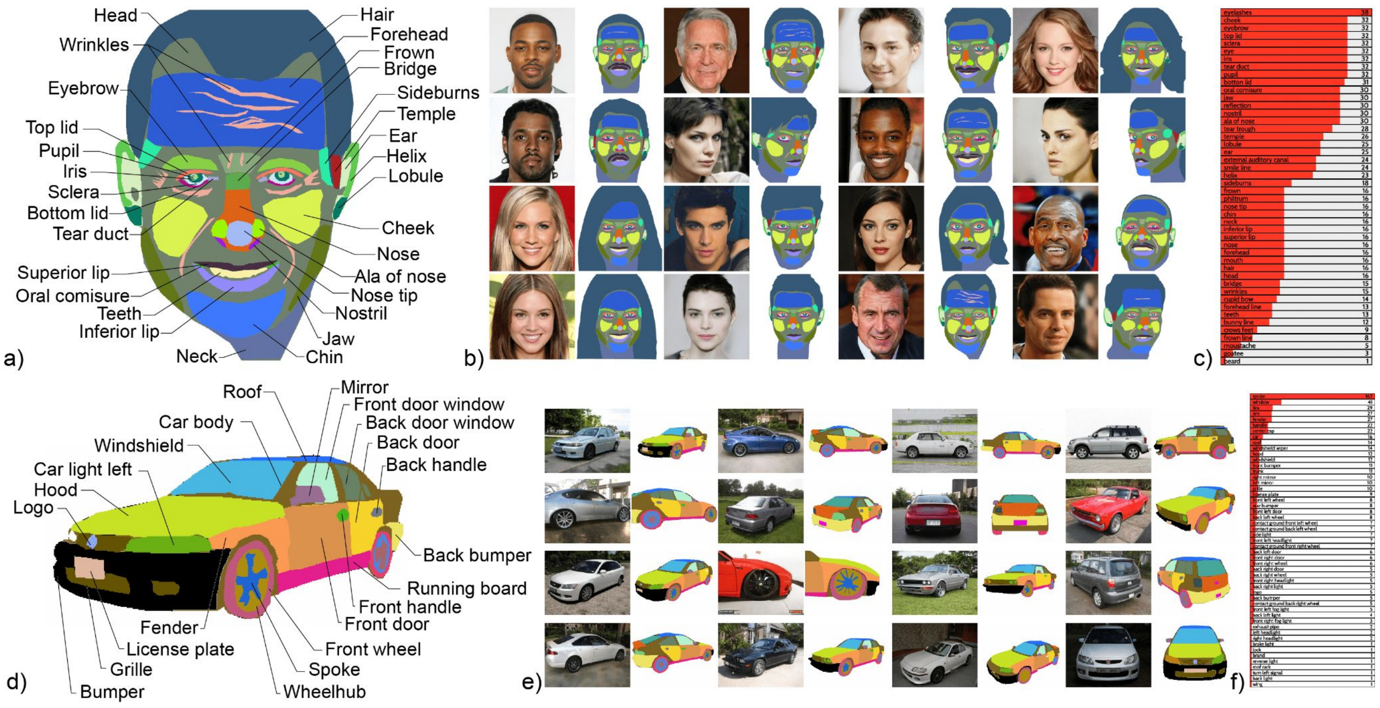
Abstract

We introduce DatasetGAN*: an automatic procedure to generate massive datasets of high-quality semantically seg-*

expensive). Labeling a complex scene with 50 objects can take anywhere between 30 to 90 minutes – clearly a bottleneck in achieving the scale of a dataset that we might desire. In this paper, we aim to synthesize large high quality labeled

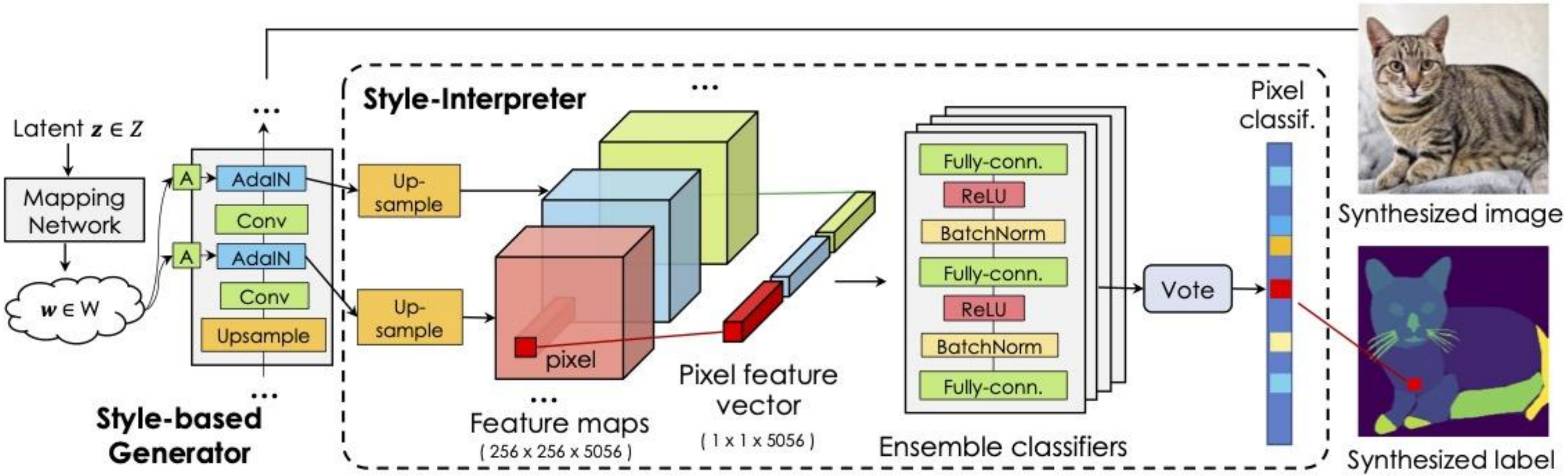
Semantic Segmentation

Input data



Head
Wrinkles
Eyebrow
Top lid
Pupil
Iris
Sclera
Bottom lid
Tear duct
Superior lip
Oral comisure
Teeth
Inferior lip
Neck
Hair
Forehead
Frown
Bridge
Sideburns
Temple
Ear
Helix
Lobule
Cheek
Nose
Ala of nose
Nose tip
Nostril
Jaw
Chin
a)
b)
c)
Roof
Car body
Windshield
Car light left
Hood
Logo
Mirror
Front door window
Back door window
Back door
Back handle
Back bumper
Running board
Front handle
Front door
Front wheel
Spoke
Wheelhub
Fender
License plate
Grille
Bumper
d)
e)
f)

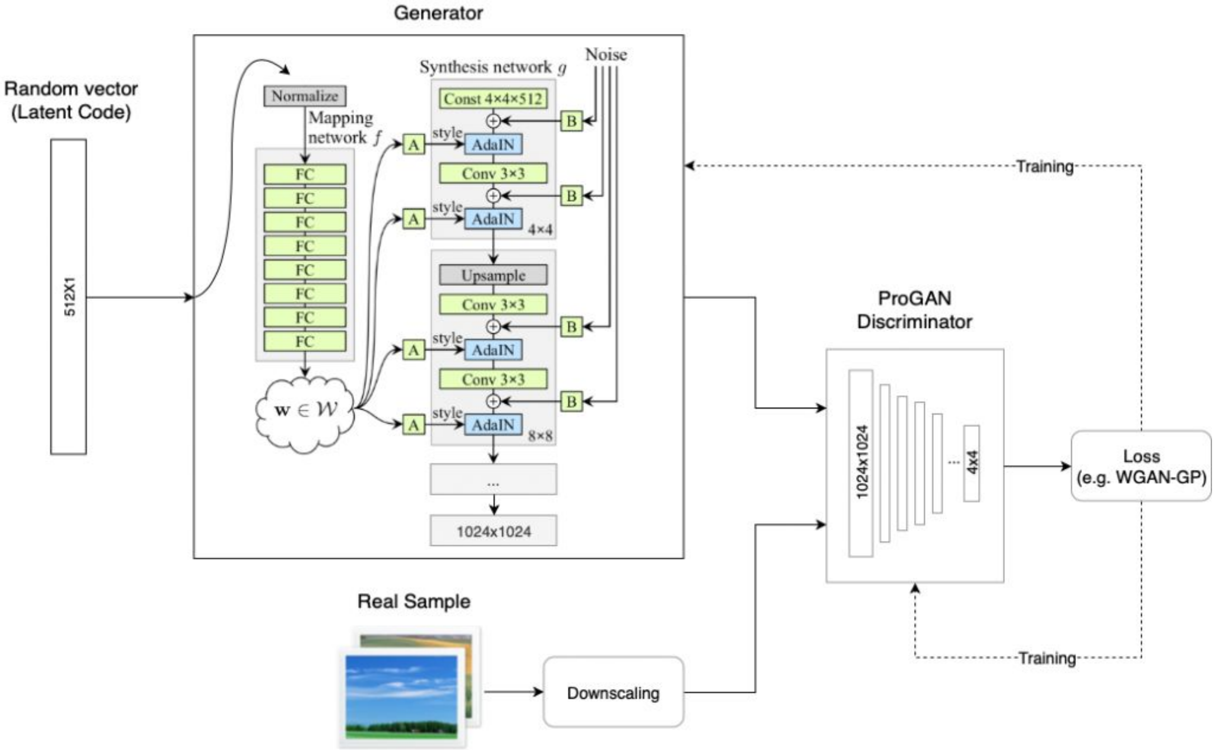
DatasetGAN



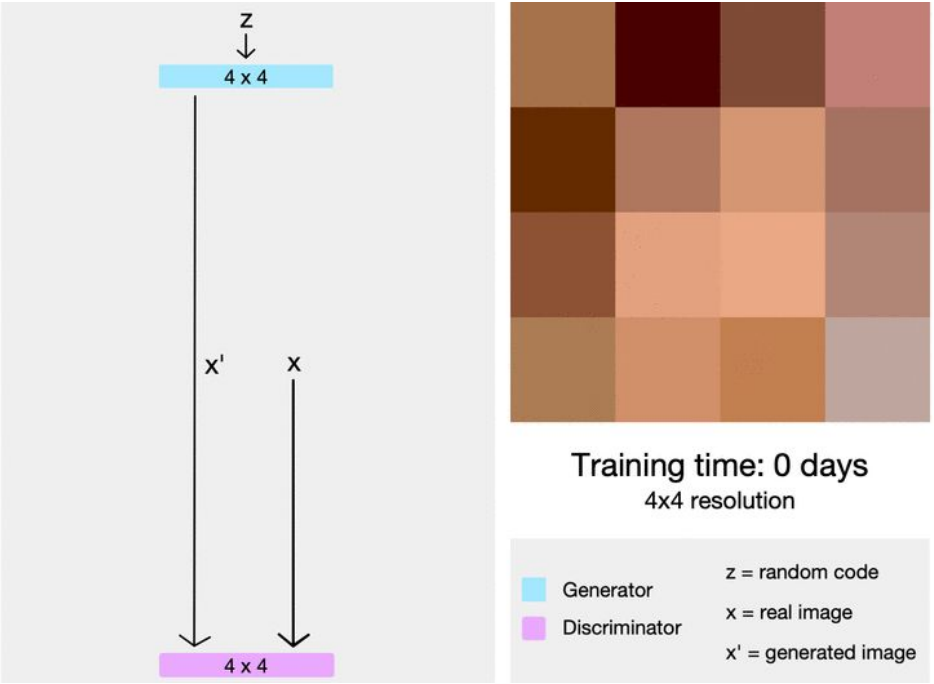
Latent $z \in Z$
Mapping Network
$w \in W$
A
AdaIN
Conv
A
AdaIN
Conv
Upsample
...
...
Style-based Generator
Style-Interpreter
...
Up-sample
Up-sample
pixel
...
Feature maps
(256 x 256 x 5056)
Pixel feature vector
(1 x 1 x 5056)
Fully-conn.
ReLU
BatchNorm
Fully-conn.
ReLU
BatchNorm
Fully-conn.
Ensemble classifiers
Vote
Pixel classif.
Synthesized image
Synthesized label

StyleGAN[2]

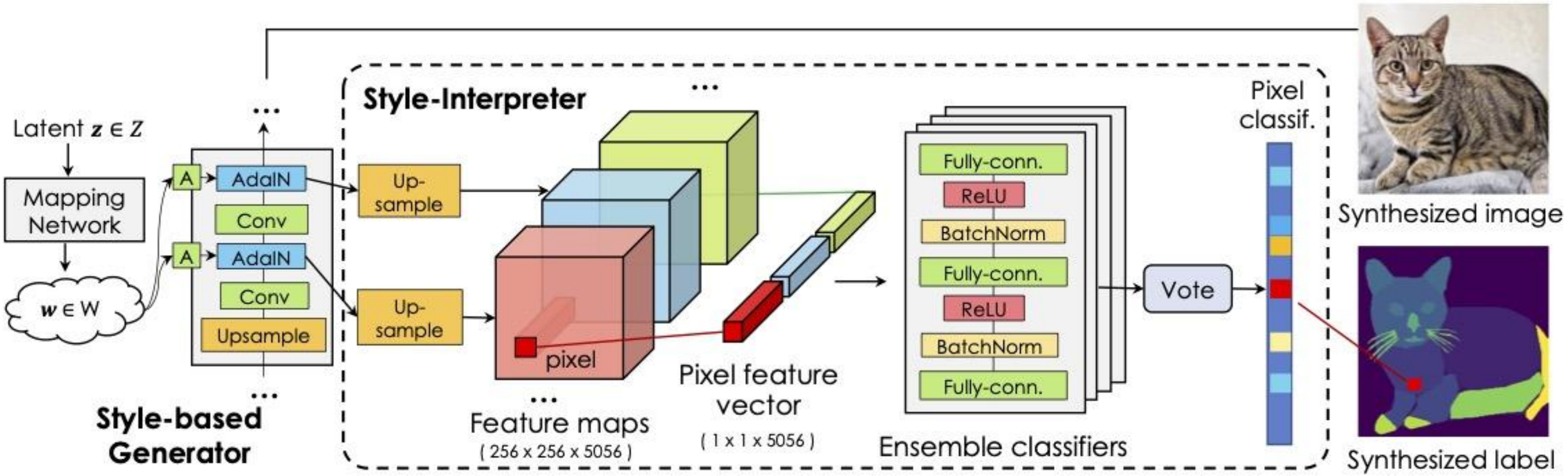
$$AdaIN(x,y) = \sigma(y)(\frac{x-\mu(x)}{\sigma(x)}) + \mu(y)$$



StyleGAN[2]


z
4 x 4
x'
x
4 x 4
Training time: 0 days
4x4 resolution
Generator
Discriminator
z = random code
x = real image
x' = generated image

Training aspects



Measuring uncertainty

$$D_{JS}(p||q) = \frac{1}{2}D_{KL}(p||\frac{p+q}{2}) + \frac{1}{2}D_{KL}(q||\frac{p+q}{2})$$

measure for each pixel between multiple heads

sum over all pixels

filter out top 10% most uncertain images

Experiments

Testing Dataset	ADE-Car-12	ADE-Car-5	Car-20	CelebA-Mask-8 (Face)	Face-34	Bird-11	Cat-16	Bedroom-19
Num of Training Images	16	16	16	16	16	30	30	40
Num of Classes	12	5	20	8	34	11	16	19
Transfer-Learning	24.85	44.92	33.91 ± 0.57	62.83	45.77 ± 1.51	21.33 ± 1.32	21.58 ± 0.61	22.52 ± 1.57
Transfer-Learning (*)	29.71	47.22	✗	64.41	✗	✗	✗	✗
Semi-Supervised [41]	28.68	45.07	44.51 ± 0.94	63.36	48.17 ± 0.66	25.04 ± 0.29	24.85 ± 0.35	30.15 ± 0.52
Semi-Supervised [41] (*)	34.82	48.76	✗	65.53	✗	✗	✗	✗
Ours	**45.64**	**57.77**	$\mathbf{62.33 \pm 0.55}$	**70.01**	$\mathbf{53.46 \pm 1.21}$	$\mathbf{36.76 \pm 2.11}$	$\mathbf{31.26 \pm 0.71}$	$\mathbf{36.83 \pm 0.54}$

✗ means that the method does not apply to this setting due to missing labeled data in the domain.

* means in-domain experiment

Ablation study

Generated Dataset Size	3K	5K	10K	20K
mIOU	43.34	44.37	44.60	**45.04**

Table 3: **Ablation study of synthesized dataset size.** Here, Style-Interpreter is trained on 16 human-labeled images. Results are reported on ADE-Car-12 test set. Performance is slowly saturating.

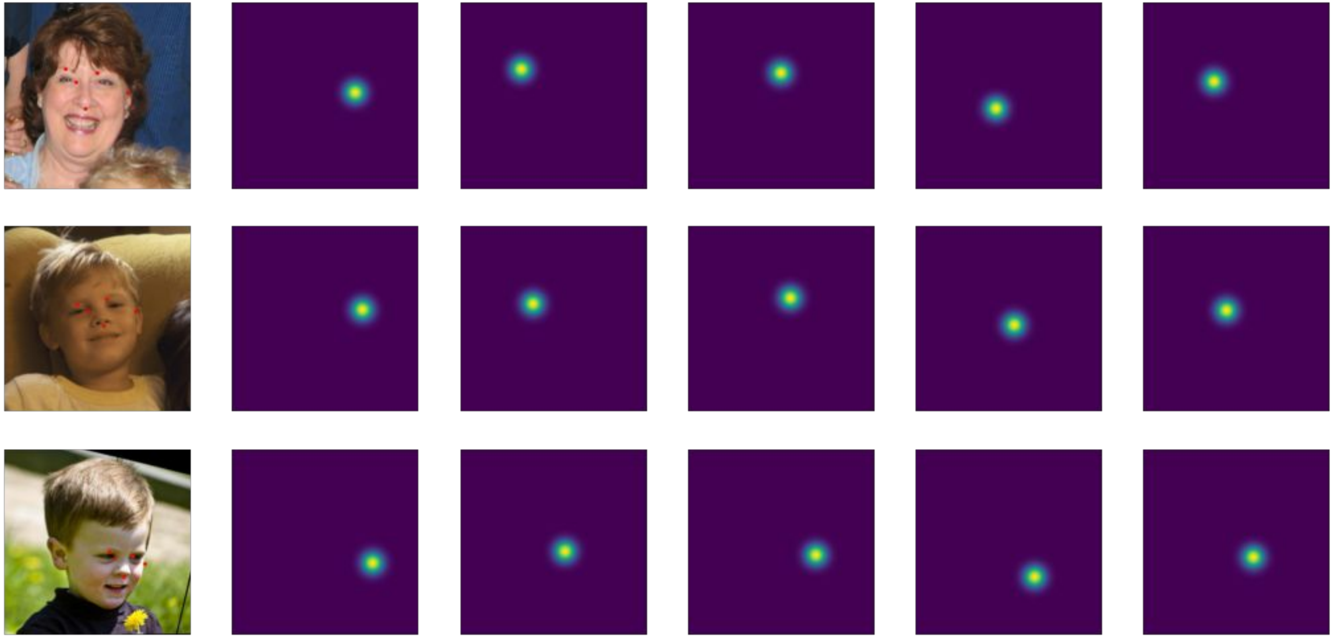
Filtering Ratio	0%	5%	10%	20%
mIOU	44.60	44.89	**45.64**	45.18

Table 4: **Ablation study of the filtering ratio.** We filter out the most uncertain synthesized Image-Annotation pairs. Result shown are reported on ADE-Car-12 test set, using the generated dataset of size 10k. We use 10% in other experiments.

Number of Annotated Images	1	7	13	19
Random	/	40.06 ± 1.32	42.44	44.41
Active Learning	/	40.88	43.49	46.82
Manual	33.92	41.19	43.61	46.74

Table 6: **Data selection.** We compare different strategies for selecting StyleGAN images to be annotated manually. mIoU is reported on ADE-Car-12 test set. We compute mean & var over 5 random runs with 1 & 7 training examples.

Keypoint detection



3D Reconstruction

https://arxiv.org/pdf/2010.09125.pdf

with 🤍 to perfectionists

Conclusion

- — Just plugin for StyleGAN
- — Requires pretrained StyleGAN for each domain

- \+ 10-30 samples enough to train model
- \+ Still useful for exotic domains

Is it actually working?

Note: Training time for 16 images is around one hour. 160G RAM is required to run 16 images training.

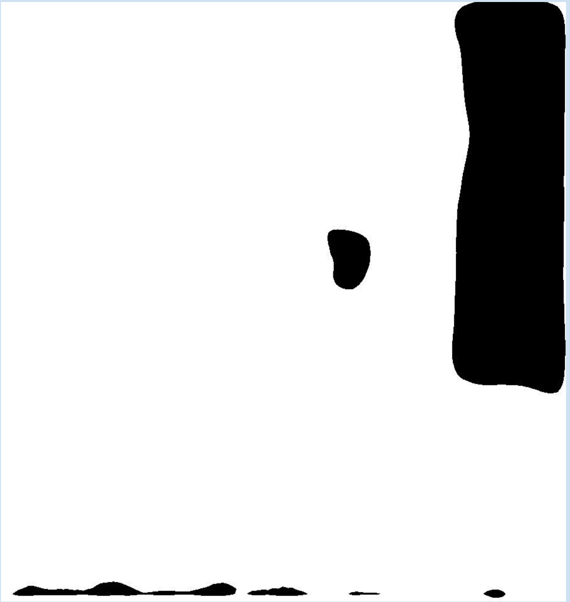
Training time for StyleGAN on 1 GPU takes 14 days 22 hours

Latent vs output

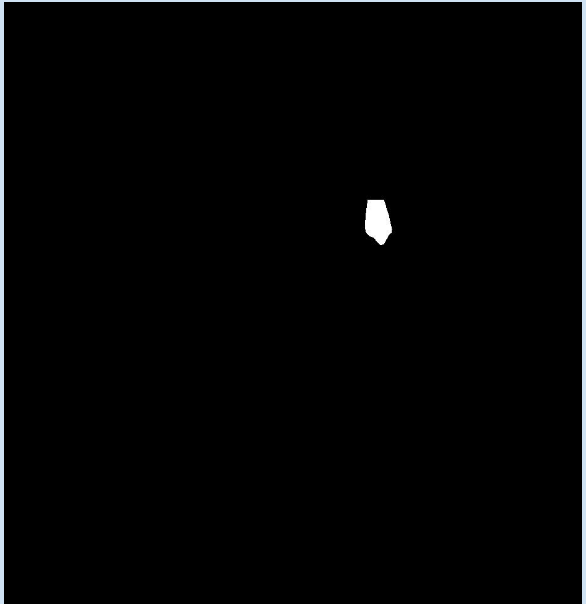
Fine-tuning of pretrained FCN_ResNet50 on 30 StyleGan outputs to predict masks of cat noses

Input

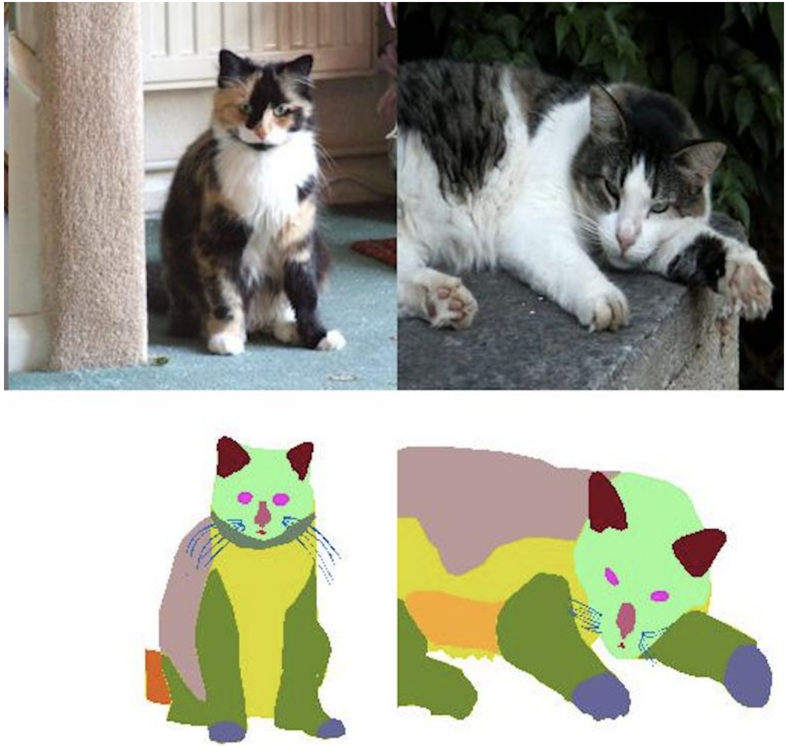
fcn_resnet50



datasetGAN



Opensource code resume



Pros:
- Runnable
- All code and models weights can be downloaded

Cons:
- Code written for GPU, difficult to run on CPU
- Requires pretrained StyleGAN for your task
- Hardcoded randomness (latent z for data generation)
- No documentation, all understanding is only from debugging

```
RuntimeError: Error(s) in loading state_dict for DeepLabV3:
    Missing key(s) in state_dict: "backbone.conv1.weight", "backbone.bn1.weight", "backbc
    Unexpected key(s) in state_dict: "module.layers.0.weight", "module.layers.0.bias", "m
```

Review

Сильные стороны статьи:

1. Актуальность решаемой проблемы
2. Новизна проведенного исследования и предложенного метода
3. Качество предлагаемого алгоритма - модель действительно показывает лучшее качество, чем стандартные transfer learning подходы

Слабые стороны статьи:

1. Статья написана непонятно. Многие основные моменты архитектуры описаны недостаточно подробно
2. Спорная применимость

Оценка по критериям НИПСа:

1. Оценка: 5 из 10
2. Уверенность: 4 из 5

CVPR 2021 Oral

CVPR
VIRTUAL JUNE 19-25

Yuxuan Zhang *[1,4] Huan Ling *[1,2,3] Jun Gao [1,2,3] Kangxue Yin [1]

Jean-Francois Lafleche[1] Adela Barriuso[5] Antonio Torralba[5] Sanja Fidler[1,2,3]

[1]NVIDIA [2]University of Toronto [3]Vector Institute [4]University of Waterloo [5]MIT

Yuxuan Zhang

Princeton University
Verified email at princeton.edu
Machine learning Computer Vision Computer Graphics

FOLLOW

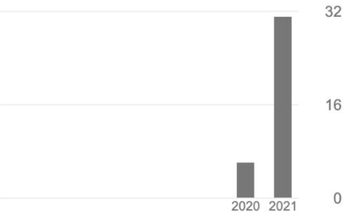
GET MY OWN PROFILE

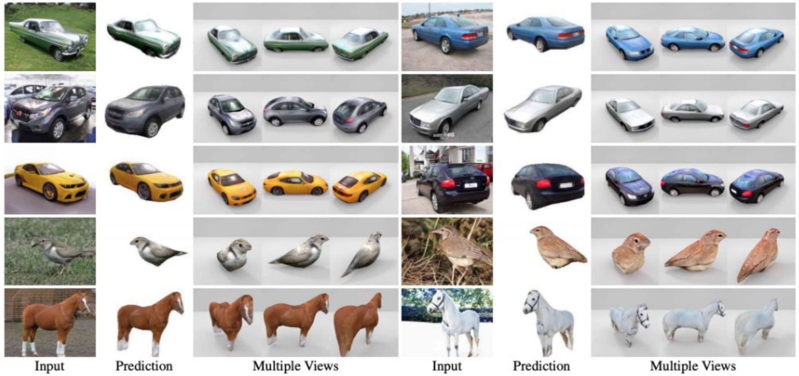
TITLE	CITED BY	YEAR
Deep neural network fingerprinting by conferrable adversarial examples N Lukas, Y Zhang, F Kerschbaum arXiv preprint arXiv:1912.00888	19	2019
Image gans meet differentiable rendering for inverse graphics and interpretable 3d neural rendering Y Zhang, W Chen, H Ling, J Gao, Y Zhang, A Torralba, S Fidler arXiv preprint arXiv:2010.09125	12	2020
Datasetgan: Efficient labeled data factory with minimal human effort Y Zhang, H Ling, J Gao, K Yin, JF Lafleche, A Barriuso, A Torralba, ... Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern …	7	2021

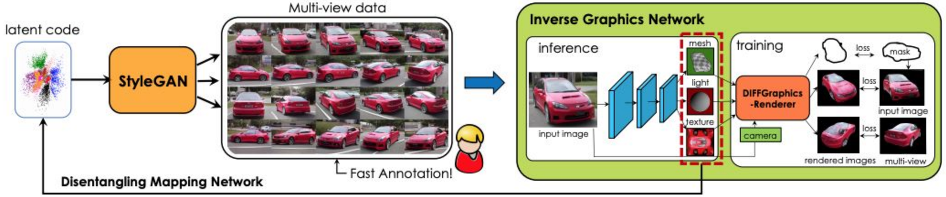
ICLR 2021

Cited by

	All	Since 2016
Citations	38	38
h-index	3	3
i10-index	2	2







Huan Ling

University of Toronto
Verified email at cs.toronto.edu - Homepage
computer vision

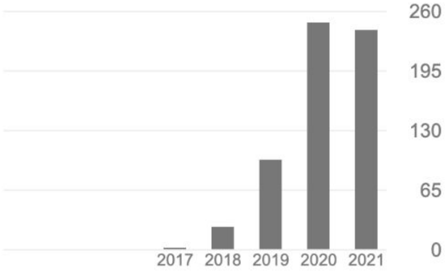
FOLLOW

GET MY OWN PROFILE

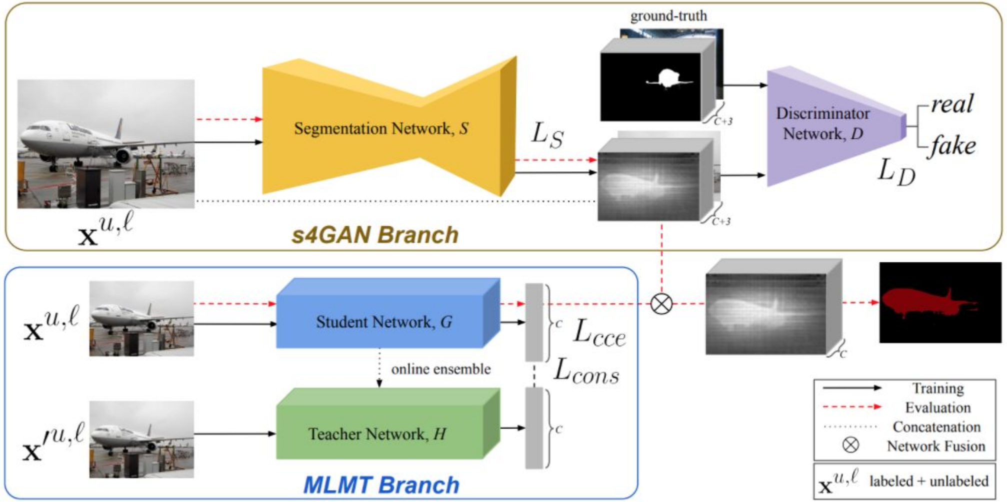
TITLE	CITED BY	YEAR
Efficient interactive annotation of segmentation datasets with polygon-rnn++ D Acuna, H Ling, A Kar, S Fidler Proceedings of the IEEE conference on Computer Vision and Pattern …	228	2018
Learning to predict 3d objects with an interpolation-based differentiable renderer W Chen, H Ling, J Gao, E Smith, J Lehtinen, A Jacobson, S Fidler Advances in Neural Information Processing Systems 32, 9609-9619	124	2019
Fast interactive object annotation with curve-gcn H Ling, J Gao, A Kar, W Chen, S Fidler Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern …	107	2019
Teaching machines to describe images via natural language feedback H Ling, S Fidler arXiv preprint arXiv:1706.00130	49	2017

Cited by

	All	Since 2016
Citations	619	617
h-index	7	7
i10-index	7	7



Baseline



Semi-Supervised Semantic Segmentation
with High- and Low-level Consistency

Citations

[Image gans meet differentiable rendering for inverse graphics and interpretable 3d neural rendering](#) [PDF] arxiv.org
Y Zhang, W Chen, H Ling, J Gao, Y Zhang… - arXiv preprint arXiv …, 2020 - arxiv.org
Differentiable rendering has paved the way to training neural networks to perform" inverse graphics" tasks such as predicting 3D geometry from monocular photographs. To train high performing models, most of the current approaches rely on multi-view imagery which are not …
Cited by 12 Related articles All 5 versions

Semantic segmentation with generative models: Semi-supervised learning and strong out-of-domain generalization [PDF] thecvf.com
D Li, J Yang, K Kreis, A Torralba… - Proceedings of the …, 2021 - openaccess.thecvf.com
Training deep networks with limited labeled data while achieving a strong generalization ability is key in the quest to reduce human annotation efforts. This is the goal of semi-supervised learning, which exploits more widely available unlabeled data to complement …
Cited by 7 Related articles All 4 versions

Segmentation in Style: Unsupervised Semantic Image Segmentation with Stylegan and CLIP [PDF] arxiv.org
D Pakhomov, S Hira, N Wagle, KE Green… - arXiv preprint arXiv …, 2021 - arxiv.org
We introduce a method that allows to automatically segment images into semantically meaningful regions without human supervision. Derived regions are consistent across different images and coincide with human-defined semantic classes on some datasets. In …
Cited by 1 All 3 versions

Learning to See by Looking at Noise [PDF] arxiv.org
M Baradad, J Wulff, T Wang, P Isola… - arXiv preprint arXiv …, 2021 - arxiv.org
Current vision systems are trained on huge datasets, and these datasets come with costs: curation is expensive, they inherit human biases, and there are concerns over privacy and usage rights. To counter these costs, interest has surged in learning from cheaper data …
Cited by 1 Related articles All 4 versions

Adapting to Unseen Vendor Domains for MRI Lesion Segmentation [PDF] arxiv.org
B Mac, AR Moody, A Khademi - arXiv preprint arXiv:2108.06434, 2021 - arxiv.org
One of the key limitations in machine learning models is poor performance on data that is out of the domain of the training distribution. This is especially true for image analysis in magnetic resonance (MR) imaging, as variations in hardware and software create non …
All 3 versions

Generative Models as a Data Source for Multiview Representation Learning [PDF] arxiv.org
A Jahanian, X Puig, Y Tian, P Isola - arXiv preprint arXiv:2106.05258, 2021 - arxiv.org
Generative models are now capable of producing highly realistic images that look nearly indistinguishable from the data on which they are trained. This raises the question: if we have good enough generative models, do we still need datasets? We investigate this …
Related articles All 2 versions

[HTML] Semantic Segmentation for Real-World Applications [HTML] unizar.es
Í Alonso Ruiz - zaguan.unizar.es
En visión por computador, la comprensión de escenas tiene como objetivo extraer información útil de una escena a partir de datos de sensores. Por ejemplo, puede clasificar toda la imagen en una categoría particular o identificar elementos importantes dentro de …

Semantic Segmentation with Generative Models: Semi-Supervised Learning and Strong Out-of-Domain Generalization

Daiqing Li[1*] Junlin Yang[1,3] Karsten Kreis[1] Antonio Torralba[4] Sanja Fidler[1,2,5]
[1] NVIDIA [2] University of Toronto [3] Yale University [4] MIT [5] Vector Institute

Abstract

Training deep networks with limited labeled data while achieving a strong generalization ability is key in the quest to reduce human annotation efforts. This is the goal of semi-supervised learning, which exploits more widely available unlabeled data to complement small labeled data sets. In this paper, we propose a novel framework for discriminative pixel-level tasks using a generative model of both images and labels. Concretely, we learn a generative ad-

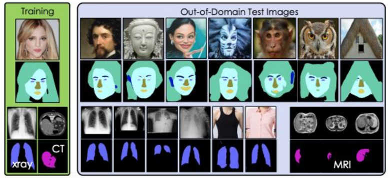


Figure 1: **Out-of-domain Generalization.** Our model trained on real faces generalizes to paintings, sculptures, cartoons and even outputs plau-

Segmentation in Style: Unsupervised Semantic Image Segmentation with Stylegan and CLIP

Daniil Pakhomov, Johns Hopkins University, dpakhom1@jhu.edu
Sanchit Hira, Johns Hopkins University, shira2@jhu.edu
Narayani Wagle, Johns Hopkins University, nwagle@jhu.edu
Kemar E. Green, Johns Hopkins University, kgreen66@jhumi.edu
Nassir Navab, Johns Hopkins University, nassir.navab@jhu.edu

Abstract

We introduce a method that allows to automatically segment images into semantically meaningful regions without human supervision. Derived regions are consistent across different images and coincide with human-defined semantic classes on some datasets. In cases where semantic regions might be hard for human to define and consistently label, our method is still able to find meaningful and consistent semantic classes. In our work, we use pretrained StyleGAN2 [1] generative model: clustering in the feature space of the generative model allows to discover semantic classes. Once classes are discovered, a synthetic dataset with generated images and corresponding segmentation masks can be created. After that a segmentation model is trained on the synthetic dataset and is able to generalize to real images. Additionally, by using CLIP [2] we are able to use prompts defined in a natural language to discover some desired semantic classes. We test our method on publicly available datasets and show state-of-the-art results. The source code for the experiments reported in the paper has been made public [1]

Reference

[1] Yuxuan Zhang, Huan Ling, Jun Gao, Kangxue Yin, Jean-Francois Lafleche, Adela Barriuso, Antonio Torralba, Sanja Fidler “DatasetGAN: Efficient Labeled Data Factory with Minimal Human Effort” https://arxiv.org/abs/2104.06490

[2] Tero Karras, Samuli Laine, Timo Aila “A Style-Based Generator Architecture for Generative Adversarial Networks” https://arxiv.org/abs/1812.04948

[3] Zhang, Yuxuan, et al. "Image gans meet differentiable rendering for inverse graphics and interpretable 3d neural rendering." arXiv preprint arXiv:2010.09125 (2020).

[4] David Acuna, Huan Ling, Amlan Kar, Sanja Fidler “Efficient interactive annotation of segmentation datasets with polygon-rnn++” CVPR 2018 Open Access Repository

[5] Huan Ling, Jun Gao, Amlan Kar, Wenzheng Chen, Sanja Fidler “Fast interactive object annotation with curve-gcn” CVPR 2019 Open Access Repository

[6] Sudhanshu Mittal, Maxim Tatarchenko, Thomas Brox “Semi-Supervised Semantic Segmentation with High- and Low-level Consistency” https://arxiv.org/abs/1908.05724

[7] Daiqing Li, Junlin Yang, Karsten Kreis, Antonio Torralba, Sanja Fidler “Semantic Segmentation With Generative Models: Semi-Supervised Learning and Strong Out-of-Domain Generalization” CVPR 2021 Open Access Repository

[8] Daniil Pakhomov, Sanchit Hira, Narayani Wagle, Kemar E. Green, Nassir Navab “Segmentation in Style: Unsupervised Semantic Image Segmentation with Stylegan and CLIP” https://arxiv.org/abs/2107.12518