Alleviating Semantics Distortion in Unsupervised Low-Level Image-to-Image Translation via Structure Consistency Constraint

Youngjae Kim

2024.03.11

Medical Imaging & Intelligent Reality Lab (MI2RL)
Convergence Medicine/Radiology,
University of Ulsan College of Medicine
Asan Medical Center
South Korea

About the paper

Alleviating Semantics Distortion in Unsupervised Low-Level Image-to-Image Translation via Structure Consistency Constraint

Jiaxian Guo[1] Jiachen Li[2] Huan Fu[1]
Mingming Gong[3] Kun Zhang[4,6] Dacheng Tao[1,5]
[1] The University of Sydney [2] Shanghai Jiao Tong University [3] The University of Melbourne
[4] Carnegie Mellon University [5] JD Explore Academy
[6] Mohamed bin Zayed University of Artificial Intelligence
jguo5934@uni.sydney.edu.au lijc0804@sjtu.edu.cn hufu6371@uni.sydney.edu.au
mingming.gong@unimelb.edu.au kunz1@cmu.edu dacheng.tao@gmail.com

Jiaxian Guo

FOLLOW

GET MY OWN PROFILE

Ph.D. Student, Computer Science Department, The University of Sydney
Verified email at uni.sydney.edu.au - Homepage

Generative Model Reinforcement Learning Domain Adaptation Deep Learning
Natural language processing

TITLE	CITED BY	YEAR

Cited by

	All	Since 2019
Citations	1334	1268
h-index	10	10
i10-index	10	10

Introduction

- Image-to-image translation, (or domain mapping)

  - aims to translate an image in the source domain X properly to the target domain Y

- However, since paired data are often unavailable or expensive to obtain,

  => Unsupervised I2I translation has attracted intense attention in recent years

- Finding $\mathbf{G}_{XY}$ such that the translated images and target domain images have similar distributions

$$P_{GXY\,(X)} \approx P_{Y}$$

- Due to an infinite number of functions that can satisfy the adversarial loss,

  GAN alone could learn a function far away from the true one.

  => various constraints are placed.

  (ex. CycleGAN, DistanceGAN, GcGAN, DRIT++, MUNIT ..)

Introduction

- However, in most unpaired datasets, not only style but also the underlying semantic distributions differ across source and target datasets

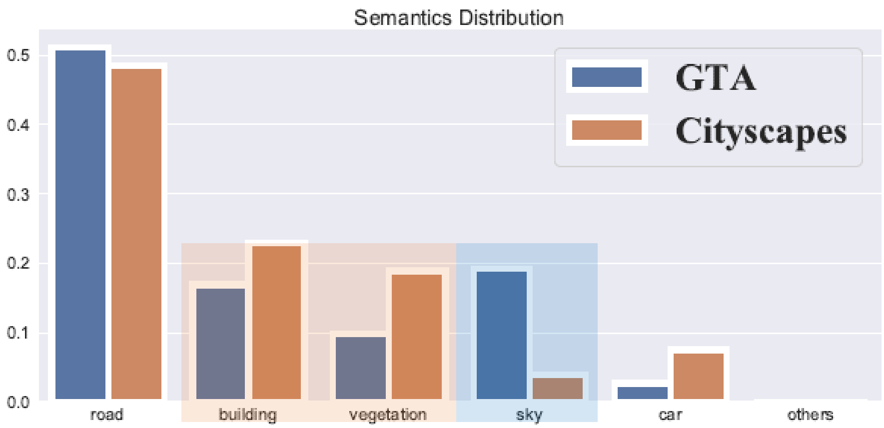
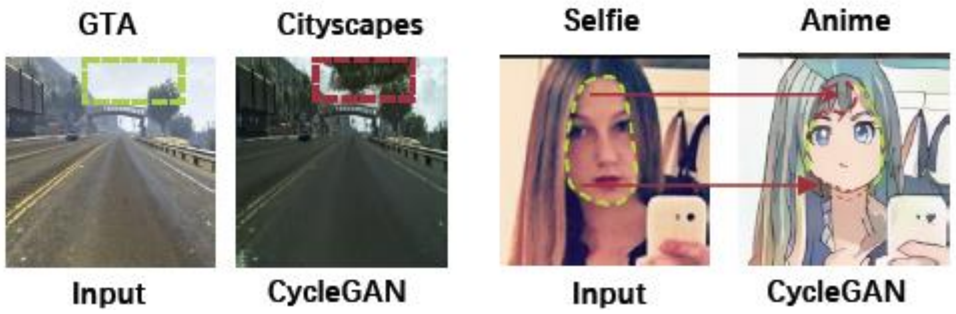


Figure 1. Class distributions in GTA and Cityscapes. We can see that the ratio of the sky in GTA is significantly higher than it in Cityscapes, and thus the distribution matching based method has to translate the sky to vegetation/building to align the distributions.



- resulting in a semantic mismatch between input and translated images : **semantics distortion problem**.

- In low-level I2I, the difference between domains arises from the low-level information e.g., resolution, illumination, color rather than geometry variation, while the structure (e.g.the shapes of objects) in images is most invariant across the source and target domains, i.e., **the semantics of an image is highly related to its structure** (shape of objects).

Methodology

- As we know, geometric structures in an images are often outlined by colors.
- Hope to preserve the geometry structure during translation
- = expect the color translation to be consistent between the input and output images.
- Ex) green leaf (summer) -> yellow leaf (autumn) :
- easy to identify it as leaf. (would be way harder to identify, if leaf color is changed to random color) - motivation

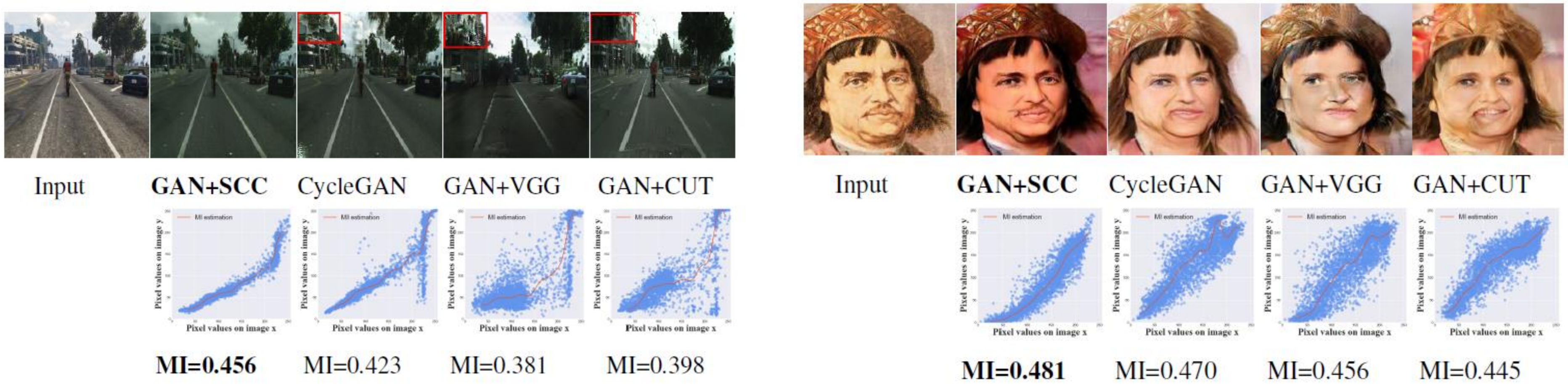


Figure 3. Unsupervised image translation examples on GTA $\rightarrow$ Cityscapes. Portrait $\rightarrow$ Photo. The top row is the translated results by each method. The bottom row is the scatter plot of the pixel values in the input image x and its corresponding pixel value in the translated image $\hat{y}$, which shows the non-linear dependency of pixel values in two images. Obviously, the stronger the dependency between pixel values in the input image (X-axis) and the translated images (Y-axis), the better the geometry structure of the input image is maintained. MI stands for the mutual information estimated by our rSMI method. Specifically, the VGG refers to the Contextual loss [39] of VGG features.

Methodology

- To alleviate semantics distortion problem in low-level I2I translation…

- **promote the structure consistency** of the source and translated images because the image structure is highly related to its semantics in this task.

  -> Our work is the first to explore such constraints for unsupervised image-to-image translation

$x_i \in \mathcal{X}$

$\hat{y}_i = G_{XY}(x_i)$

V^{x_i}

$V^{\hat{y}_i}$

Random variables for pixels in x_i and $\hat{y}_i$

$\{v_j^{x_i}\}_{j=1}^{M} \xrightarrow{\text{Sampled from}} P_{V^{x_i}}$

$\{v_j^{\hat{y}_i}\}_{j=1}^{M} \xrightarrow{\text{Sampled from}} P_{V^{\hat{y}_i}}$

* M is the number of pixels of the image

Joint distribution

$$MI(V^{x_i}, V^{\hat{y}_i}) = \mathbb{E}_{(v^{x_i}, v^{\hat{y}_i}) \sim P_{(V^{x_i}, V^{\hat{y}_i})}} \left(\log \frac{P_{(V^{x_i}, V^{\hat{y}_i})}}{P_{V^{x_i}} \otimes P_{V^{\hat{y}_i}}} \right) \quad (1)$$

Product of marginal distributions

→ Mutual information can be calculated

(= how dependent one random variable is to the other)

(= how different the pixel values of input image is from the generated image)

Because V^{x_i} and $V^{\hat{y}_i}$ are low-dimensional, a straightforward way to estimate (1) is to estimate the distributions P based on the histogram of the images.

Methodology

- How to use the MI in this task with efficient backpropagation?
- Proposing extension version of SMI (squared loss MI)

Research

Open Access

Mutual information estimation reveals global associations between stimuli and biological processes

Taiji Suzuki[1], Masashi Sugiyama[2], Takafumi Kanamori[3] and Jun Sese*[4]

Address: [1]Department of Mathematical Informatics, The University of Tokyo, 7-3-1 Hongo, Bunkyo-ku, Tokyo 113-8656, Japan, [2]Department of Computer Science, Tokyo Institute of Technology, 2-12-1 O-okayama, Meguro-ku, Tokyo 152-8552, Japan, [3]Department of Computer Science and Mathematical Informatics, Nagoya University, Furocho, Chikusaku, Nagoya 464-8603, Japan and [4]Department of Information Science, Ochanomizu University, 2-1-1 Ohtsuka, Bunkyo-ku, Tokyo 112-8610, Japan

Email: Taiji Suzuki - s-taiji@stat.t.u-tokyo.ac.jp; Masashi Sugiyama -
Jun Sese* - sesejun@is.ocha.ac.jp

* Corresponding author

from The Seventh Asia Pacific Bioinformatics Conference (APBC 2009)
Beijing, China. 13–16 January 2009

Published: 30 January 2009

BMC Bioinformatics 2009, **10**(Suppl 1):S52 doi:10.1186/1471-2105-10-S1-

This article is available from: http://www.biomedcentral.com/1471-2105/1

Density-Difference Estimation

Masashi Sugiyama
Tokyo Institute of Technology, Japan.
sugi@cs.titech.ac.jp
http://sugiyama-www.cs.titech.ac.jp/~sugi

Takafumi Kanamori
Nagoya University, Japan.
kanamori@is.nagoya-u.ac.jp

Taiji Suzuki
The University of Tokyo, Japan.
s-taiji@stat.t.u-tokyo.ac.jp

Marthinus Christoffel du Plessis
Tokyo Institute of Technology, Japan.
christo@sg.cs.titech.ac.jp

Song Liu
Tokyo Institute of Technology, Japan.
song@sg.cs.titech.ac.jp

Ichiro Takeuchi
Nagoya Institute of Technology, Japan.
takeuchi.ichiro@nitech.ac.jp

$P_{V^{x_i}} \otimes P_{V^{\hat{y}_i}}$ as S_i

$P_{(V^{x_i}, V^{\hat{y}_i})}$ as Q_i

$$\begin{aligned} SMI(V^{x_i}, V^{\hat{y}_i}) &= D_{PE}(P_{V^{x_i}} \otimes P_{V^{\hat{y}_i}} || P_{(V^{x_i}, V^{\hat{y}_i})}) \\ &= D_{PE}(S_i || Q_i) \\ &= \mathbb{E}_{Q_i}[(\frac{S_i}{Q_i} - 1)^2]. \end{aligned} \tag{2}$$

* Representing SMI using Pearson divergence.

$$D_{\text{Pearson}}(P||Q) = \sum_x \frac{(P(x)-Q(x))^2}{P(x)}$$

$\frac{S_i}{Q_i}$ is unbounded, so SMI value can be infinity.
→ Cause instability in the backpropagation

Methodology

- How to use the MI in this task with efficient backpropagation?

$P_{V^{x_i}} \otimes P_{V^{\hat{y}_i}}$ as S_i

$P_{(V^{x_i}, V^{\hat{y}_i})}$ as Q_i

$$\begin{aligned} SMI(V^{x_i}, V^{\hat{y}_i}) &= D_{PE}(P_{V^{x_i}} \otimes P_{V^{\hat{y}_i}} || P_{(V^{x_i}, V^{\hat{y}_i})}) \\ &= D_{PE}(S_i || Q_i) \\ &= \mathbb{E}_{Q_i}[(\frac{S_i}{Q_i} - 1)^2]. \end{aligned} \quad (2)$$

* Representing SMI using Pearson divergence.

$$D_{\text{Pearson}}(P||Q) = \sum_x \frac{(P(x)-Q(x))^2}{P(x)}$$

$\frac{S_i}{Q_i}$ is unbounded, so SMI value can be infinity.
→ Cause numeric instability in the backpropagation

(relative Pearson Divergence)

$$D_{rPE}(S_i \,||\, Q_i) = D_{PE}(S_i \,||\, \beta S_i + (1-\beta)Q_i). \quad (3)$$

$\beta \in (0, 1)$

$$Q_i \longrightarrow \beta S_i + (1-\beta)Q_i$$

= Keeping the density ratio bounded to $[0, \frac{1}{\beta}]$

(Stable backpropagation)

$$\begin{aligned} rSMI(V^{x_i}, V^{\hat{y}_i}) &= D_{rPE}(P_{V^{x_i}} \otimes P_{V^{\hat{y}_i}} || P_{(V^{x_i}, V^{\hat{y}_i})}) \\ &= \mathbb{E}_{\beta S_i + (1-\beta)Q_i}[(\frac{S_i}{\beta S_i + (1-\beta)Q_i} - 1)^2] \end{aligned} \quad (4)$$

Methodology

- How to use the MI in this task with efficient backpropagation?

$\phi \in \mathbb{R}^m$ is the kernel function

$\alpha \in \mathbb{R}^m$ Parameter vector to solve

* m is the number of kernels

$$rSMI(V^{x_i}, V^{\hat{y}_i}) = D_{rPE}(P_{V^{x_i}} \otimes P_{V^{\hat{y}_i}} || P_{(V^{x_i}, V^{\hat{y}_i})})$$
$$= \mathbb{E}_{\beta S_i + (1-\beta) Q_i}[(\frac{S_i}{\beta S_i + (1-\beta) Q_i} - 1)^2] \quad (4)$$

(5)-(6)
(to estimate the rSMI, linear combination of kernel functions was used)

$$\widehat{rSMI}(V^{x_i}, V^{\hat{y}_i}) = 2\hat{\alpha}^T \hat{h} - \hat{\alpha}^T \hat{H} \hat{\alpha} - 1. \quad (7)$$

- Resource friendly → efficient backpropagation

$$\mathcal{L}_{SCC} = \frac{1}{N} \sum_{i=1}^{N} \widehat{rSMI}(V^{x_i}, V^{G_{XY}(x_i)}), \quad (8)$$

$$\min_{G_{XY}} \max_{D_Y} \mathcal{L}_{GAN+SCC}(G_{XY}, D_Y) = \mathcal{L}_{GAN}(G_{XY}, D_Y) - \lambda_{SCC} \mathcal{L}_{SCC}(G_{XY}), \quad (9)$$

- Can be used in various image to image translation frameworks, e.g., CycleGAN and CUT

Experiments / Results

- Digits Translation
- Segmentation in Cityscapes
- Maps
- Simulation to Real (GTA to Real, Real to Anime)

Experiments / Results

(a) GAN (b) GAN + SCC (c) CycleGAN (d) CycleGAN + SCC

Figure 5. Qualitative comparisons on SVHN→MNIST. From Figure (a) and (b), we can see that the GAN method has no collapse solution by combining with our SCC. Also, the semantics distortion problem in CycleGAN is alleviated after incorporating with SCC.

Table 1. Classification accuracy for digits experiments.

	Translated Images as Test set			Translated Images as Training set		
Method	S → M	M → M-M	M-M → M	S → M	M → M-M	M-M → M
GAN alone	21.3±9.5	54.6±40.5	80.3±3.5	28.6±10.8	45.7±31.2	95.5±0.4
+ SCC	37.3±1.2	96.3±0.2	90.9±0.5	47.9±2.3	86.2±1.9	96.0±0.1
CycleGAN	26.1±8.1	95.3±0.4	84.7±2.5	31.6±5.6	83.8±3.0	95.9±0.4
+ SCC	38.0±0.5	**96.7±0.1**	91.5±0.3	47.4±2.0	87.7±2.1	**96.1±0.2**
GcGAN-*rot*	32.5±2.0	95.0±0.6	85.9±0.8	40.9±6.5	84.6±2.8	96.0±0.1
+ SCC	36.5±1.3	96.4±0.3	91.8 ±1.0	47.5±1.2	89.5±0.6	96.1±0.1
GcGAN -*vf*	33.3±4.2	95.2±0.4	84.5±1.5	31.6±5.6	83.8±3.0	95.9±0.4
+ SCC	37.0±0.8	96.6±0.3	91.8±0.8	49.5±4.9	87.8±2.3	96.0±0.1
Cyc + rot + SCC	39.0±0.5	96.5±0.3	91.8±1.0	50.5±1.8	**89.8±0.5**	**96.1±0.1**
Cyc + vf + SCC	**44.6±6.8**	**96.7±0.3**	**92.0±0.8**	**51.3±5.4**	89.0±0.8	**96.1±0.1**

- S : SVHN
- M : MNIST
- M-M : MNIST-M

- Most of methods had promising improvements in both accuracy and stability.

Experiments / Results

Table 2. Quantitative scores on GTA → Citycapes,Citycapes parsing → image and Photo → Map. The scores with * are reproduced on a single GPU using the codes provided by the authors. More qualitative results are given at the Appendix A.7.2.

Methods	GTA → Citycapes			Citycapes parsing → image			Photo → Map		
	pixel acc ↑	class acc ↑	mean IoU ↑	pixel acc↑	class acc↑	mean IoU↑	RMSE ↓	acc%(δ_1) ↑	acc%(δ_2) ↑
CoGAN	\	\	\	0.40	0.10	0.06	\	\	\
BiGAN/ALI	\	\	\	0.19	0.06	0.02	\	\	\
SimGAN	\	\	\	0.20	0.10	0.04	\	\	\
DistanceGAN	\	\	\	0.53	0.19	0.11	\	\	\
GAN + VGG	0.216	0.098	0.041	0.551	0.199	0.133	34.38	28.1	48.8
DRIT++	0.423	0.138	0.071	\	\	\	32.12	29.8	52.1
GAN *	0.382	0.137	0.068	0.437	0.161	0.098	33.22	19.3	42.0
+ SCC	0.487	0.148	0.089	0.642	0.215	0.155	28.91	38.6	61.8
GcGAN-rot *	0.405	0.139	0.068	0.551	0.197	0.129	27.98	42.8	64.6
+ SCC	0.445	0.162	0.080	0.651	0.228	0.162	**26.55**	**44.7**	**66.5**
CycleGAN *	0.232	0.127	0.043	0.52	0.17	0.11	26.81	43.1	65.6
+ SCC	0.386	0.161	0.076	0.571	0.192	0.134	26.61	44.7	66.2
CUT *	0.546	0.165	0.095	0.695	0.259	0.178	28.48	40.1	61.2
+ SCC	**0.572**	**0.185**	**0.11**	**0.699**	**0.263**	**0.182**	27.34	39.2	60.5

Input Label **GAN+SCC** GAN+VGG CUT GcGAN DRIT++ CycleGAN **CycleGAN+SCC**

Experiments / Results

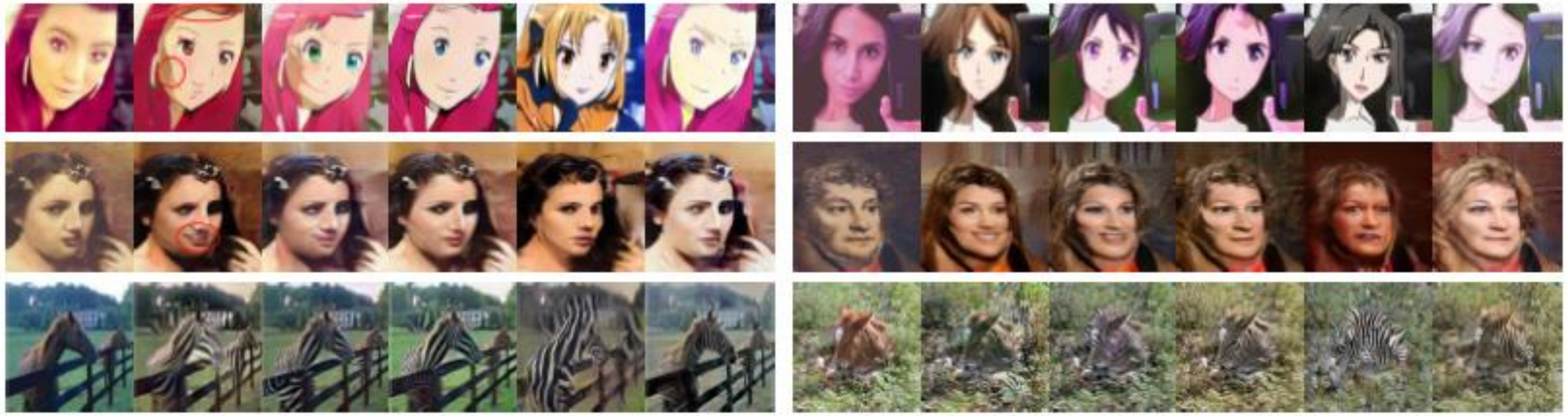


Figure 7. Qualitative results on Selfie → Anime, Portrait → Photo, Horse → Zebra datasets. More qualitative results are given in A.7.3. We can see that the no matter personal identification or horse shape is better preserved by the translation model empowered by our SCC.

Figure 8. Sensitivity analysis examples on Selfie → Anime and GTA → Cityscapes. Obviously, the semantics distortion problem in CycleGAN is alleviated after incorporating with our SCC.

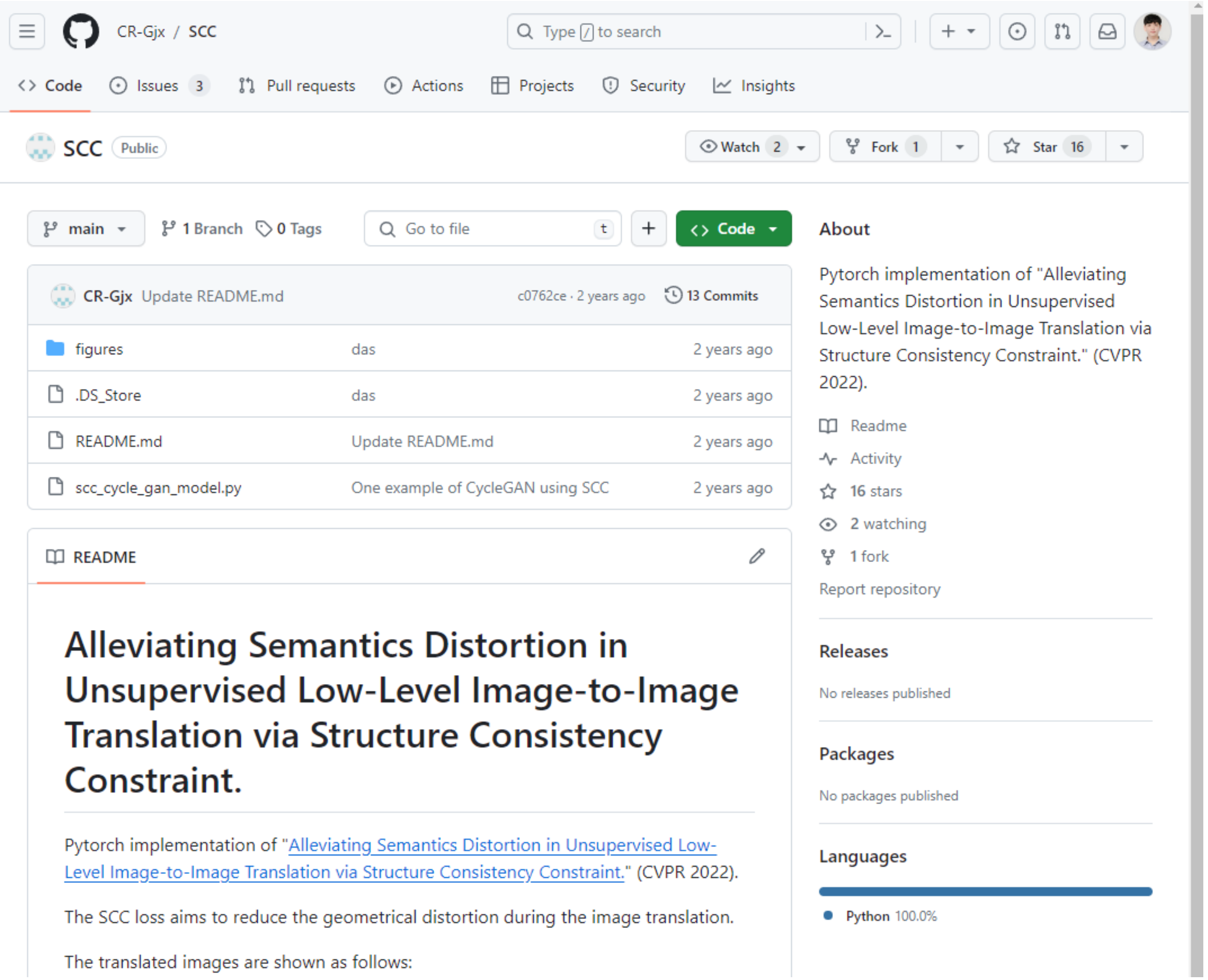
Table 4. The segmentation scores for different λ_{SCC} of the model CycleGAN + SCC in the datasets GTA2cityscapes.

λ_{SCC}	0	1	3	5	7	9
pixel acc ↑	0.232	0.292	0.322	0.360	0.382	0.386
class acc ↑	0.127	0.136	0.143	0.160	0.160	0.161
mean IoU ↑	0.0432	0.055	0.059	0.070	0.075	0.076

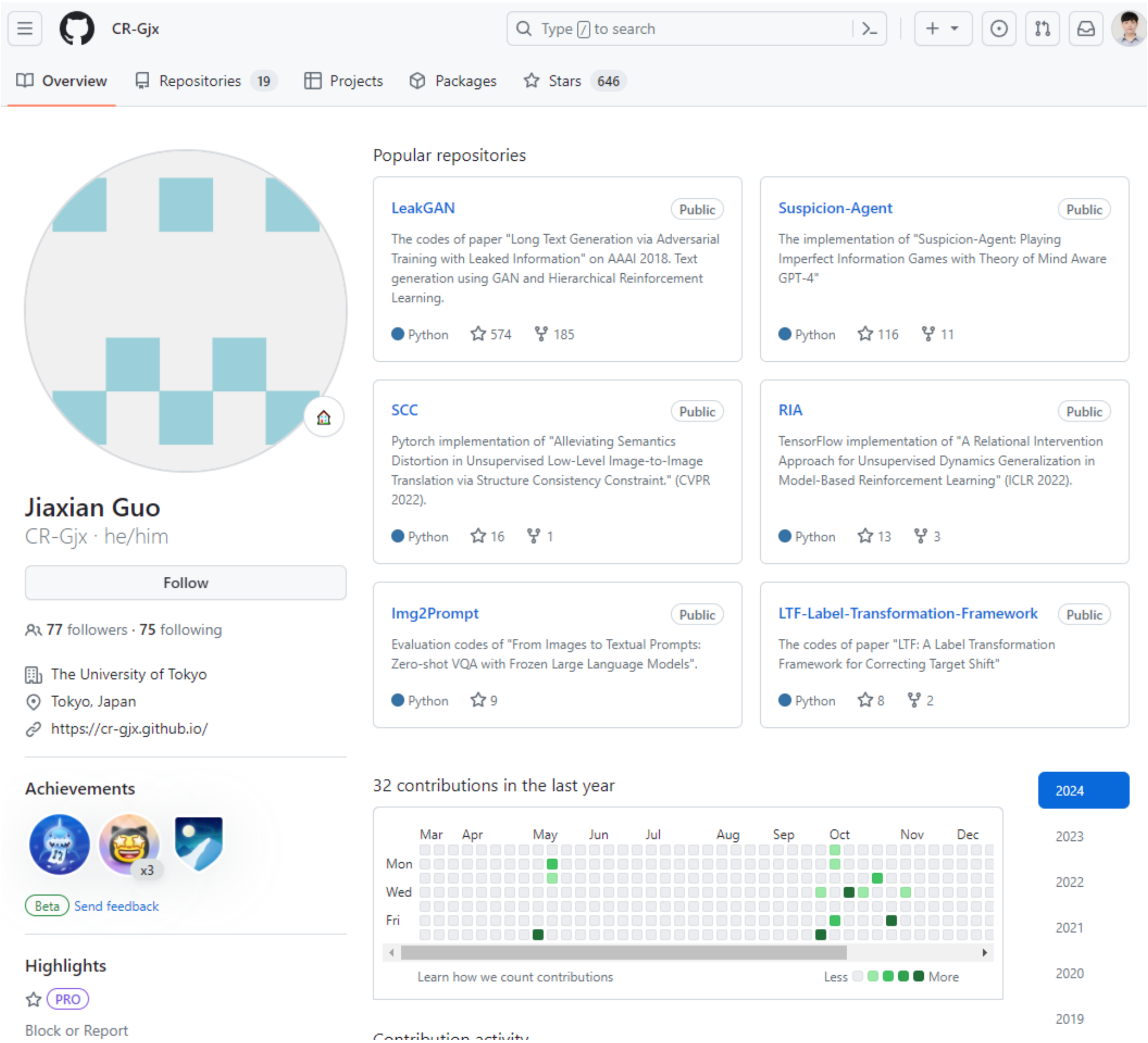
conclusion

- Proposed structure consistency constraint(SCC) to improve structure consistency in pixel wise level for unsupervised image to image translation
- Evaluation was done in wide range of applications
- Demonstrates that SCC can achieve high-quality translation by keeping the geometry of the original domain.

github


CR-Gjx / SCC
Type / to search
Code
Issues 3
Pull requests
Actions
Projects
Security
Insights
SCC Public
Watch 2
Fork 1
Star 16
main
1 Branch
0 Tags
Go to file
Code
CR-Gjx Update README.md
c0762ce · 2 years ago
13 Commits
figures das 2 years ago
.DS_Store das 2 years ago
README.md Update README.md 2 years ago
scc_cycle_gan_model.py One example of CycleGAN using SCC 2 years ago
README
Alleviating Semantics Distortion in Unsupervised Low-Level Image-to-Image Translation via Structure Consistency Constraint.
Pytorch implementation of "Alleviating Semantics Distortion in Unsupervised Low-Level Image-to-Image Translation via Structure Consistency Constraint." (CVPR 2022).
The SCC loss aims to reduce the geometrical distortion during the image translation.
The translated images are shown as follows:
About
Pytorch implementation of "Alleviating Semantics Distortion in Unsupervised Low-Level Image-to-Image Translation via Structure Consistency Constraint." (CVPR 2022).
Readme
Activity
16 stars
2 watching
1 fork
Report repository
Releases
No releases published
Packages
No packages published
Languages
Python 100.0%

github



CR-Gjx
Type / to search
Overview
Repositories 19
Projects
Packages
Stars 646
Popular repositories
LeakGAN
Public
The codes of paper "Long Text Generation via Adversarial Training with Leaked Information" on AAAI 2018. Text generation using GAN and Hierarchical Reinforcement Learning.
Python 574 185
Suspicion-Agent
Public
The implementation of "Suspicion-Agent: Playing Imperfect Information Games with Theory of Mind Aware GPT-4"
Python 116 11
SCC
Public
Pytorch implementation of "Alleviating Semantics Distortion in Unsupervised Low-Level Image-to-Image Translation via Structure Consistency Constraint." (CVPR 2022).
Python 16 1
RIA
Public
TensorFlow implementation of "A Relational Intervention Approach for Unsupervised Dynamics Generalization in Model-Based Reinforcement Learning" (ICLR 2022).
Python 13 3
Img2Prompt
Public
Evaluation codes of "From Images to Textual Prompts: Zero-shot VQA with Frozen Large Language Models".
Python 9
LTF-Label-Transformation-Framework
Public
The codes of paper "LTF: A Label Transformation Framework for Correcting Target Shift"
Python 8 2
Jiaxian Guo
CR-Gjx · he/him
Follow
77 followers · 75 following
The University of Tokyo
Tokyo, Japan
https://cr-gjx.github.io/
Achievements
x3
Beta Send feedback
Highlights
PRO
Block or Report
32 contributions in the last year
Mar Apr May Jun Jul Aug Sep Oct Nov Dec
Mon
Wed
Fri
Learn how we count contributions
Less More
2024
2023
2022
2021
2020
2019

Thank you