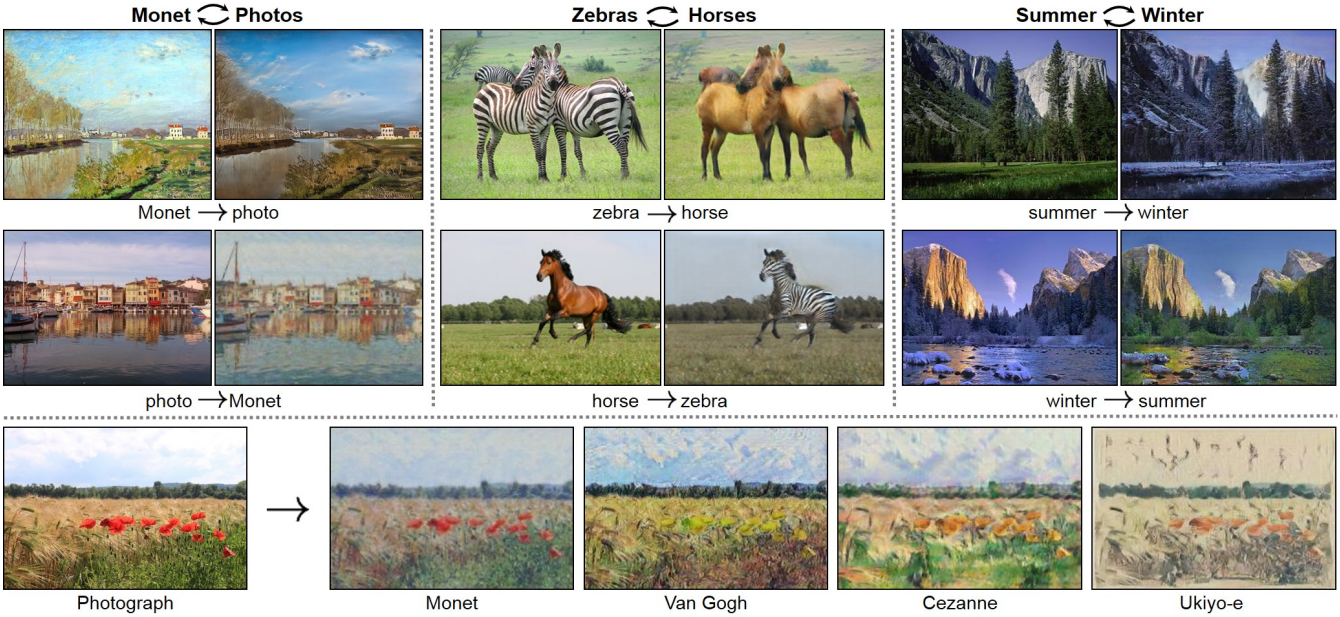
Unpaired Image-to-Image Translation using Cycle-Consistent Adversarial Networks

Jun-Yan Zhu* **Taesung Park*** **Phillip Isola** **Alexei A. Efros**

UC Berkeley

In ICCV 2017

Paper | **PyTorch code** | **Torch code**



Abstract

Image-to-image translation is a class of vision and graphics problems where the goal is to learn the mapping between an input image and an output image using a training set of aligned image pairs. However, for many tasks, paired training data will not be available. We present an approach for learning to translate an image from a source domain X to a target domain Y in the absence of paired examples. Our goal is to learn a mapping G: X → Y such that the distribution of images from G(X) is indistinguishable from the distribution Y using an adversarial loss. Because this mapping is highly under-constrained, we couple it with an inverse mapping F: Y → X and introduce a cycle consistency loss to push F(G(X)) ≈ X (and vice versa). Qualitative results are presented on several tasks where paired training data does not exist, including collection style transfer, object transfiguration, season transfer, photo enhancement, etc. Quantitative comparisons against several prior methods demonstrate the superiority of our approach.

Paper

arxiv 1703.10593, 2017.

Citation

Jun-Yan Zhu*, Taesung Park*, Phillip Isola, and Alexei A. Efros. "Unpaired Image-to-Image Translation using Cycle-Consistent Adversarial Networks", in IEEE International Conference on Computer Vision (ICCV), 2017.
(* indicates equal contributions) **Bibtex**

Code: PyTorch | Torch

If you have questions about our PyTorch code, please check out **model training/test tips** and **frequently asked questions**.

Course

CycleGAN course assignment **code** and **handout** designed by Prof. **Roger Grosse** for **"Intro to Neural Networks and Machine Learning"** at University of Toronto. Please contact the instructor if you would like to adopt this assignment in your course.

Other Implementations

Tensorflow (Harry Yang) | **Tensorflow** (Archit Rathore) | **Tensorflow** (Van Huy) | **Tensorflow** (Xiaowei Hu) | **TensorLayer** (luoxier)
Tensorflow-simple (Zhenliang He) | **Chainer** (Yanghua Jin) | **Minimal PyTorch** (yunjey) | **Mxnet** (Ldpe2G) | **lasagne/keras** (tjwei)

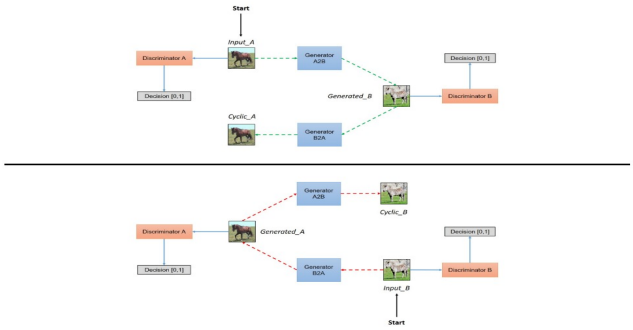
ICCV Spotlight Talk

Expository Articles and Videos

Two minute papers

Karoly Zsolnai-Feher made the above as part of his very cool **"Two minute papers"** series.

Understanding and Implementing CycleGAN



Nice explanation by Hardik Bansal and Archit Rathore, with Tensorflow code documentation.

Creative Applications of CycleGAN

Researchers, developers and artists have tried our code on various image manipulation and artistic creatiion tasks. Here we highlight a few of the many compelling examples. Search **CycleGAN** in Twitter for more applications.

Converting Monet into Thomas Kinkade

What if **Claude Monet** had lived to see the rise of Americana pastoral kitsch in the style of **Thomas Kinkade**? And what if he resorted to it to support himself in his old age? Using CycleGAN, our great **David Fouhey** finally realized the dream of Claude Monet revisiting his cherished work in light of Thomas Kinkade, the self-stylized painter of light.

Resurrecting Ancient Cities

Jack Clark used our code to convert ancient maps of **Babylon**, **Jerusalem** and **London** into modern Google Maps and satellite views.

Animal Transfiguration

Tatsuya Hatanaka trained our method t translate black bears to pandas. See mo examples and download the models at th **website**. **Matt Powell** performe transfiguration between different species of bird

Portrait to Dollface

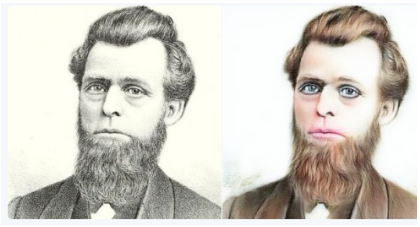
Mario Klingemann used our code to translate portraits into dollface. See how the characters in Game of Thrones look like in the doll world.

Face ↔ Ramen

Takuya Kato performed a magical and hilarious Face ↔ Ramen translation with CycleGAN. Check out more results **here**

Colorizing legacy photographs

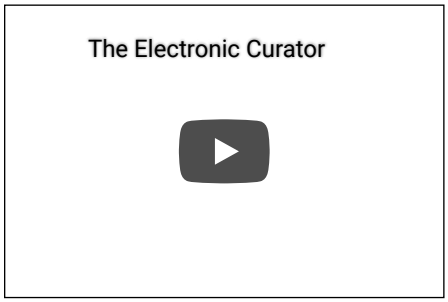
Mario Klingemann trained our method to tu legacy black and white photos into color version



Cats ↔ Dogs

itok_msi produced cats ↔ dogs CycleGAN results with a local+global discriminator and a smaller cycle loss.

The Electronic Curator



Eran Hadas and Eyal Gruss used CycleGAN to convert human faces into vegetable portraits. They built a real-time art demo which allows users to interact with the model with their own faces.

Turning Fortnite into PUBG

Chintan Trivedi used CycleGAN to transla between Fornite and PURB, two popular Batt Royale games with hundreds of millions of user Now you can enjoy the gameplay of one game the visuals of the other. Check out his **blog** f more cool demos.

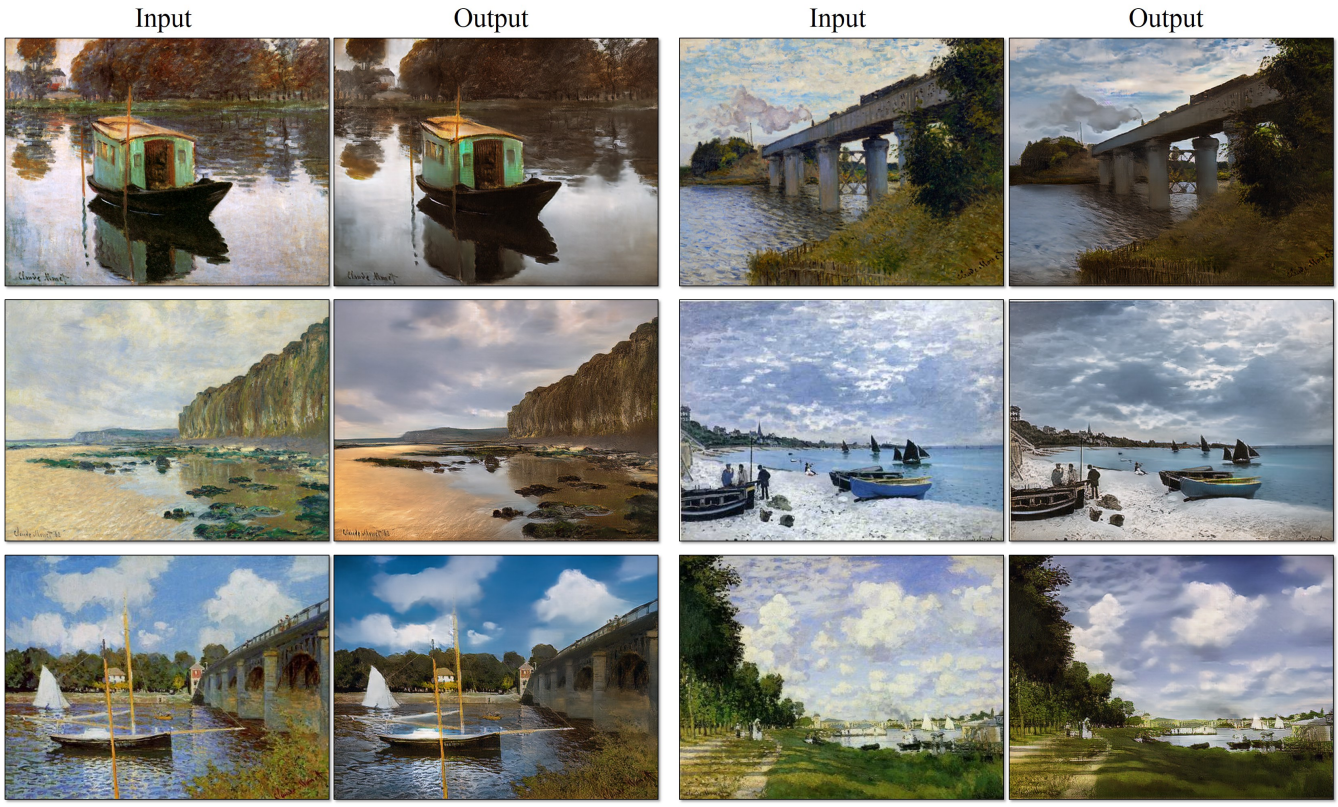
Popular Press

Applications in our Paper

Monet Paintings → Photos

Mapping Monet paintings to landscape photographs from Flickr:

Best results | **Random training set results** | **Random test set results**



Collection Style Transfer

Transferring input images into artistic styles of Monet, Van Gogh, Ukiyo-e, and Cezanne.

Results on the author's personal photos

Random training set results | **Random test set results**

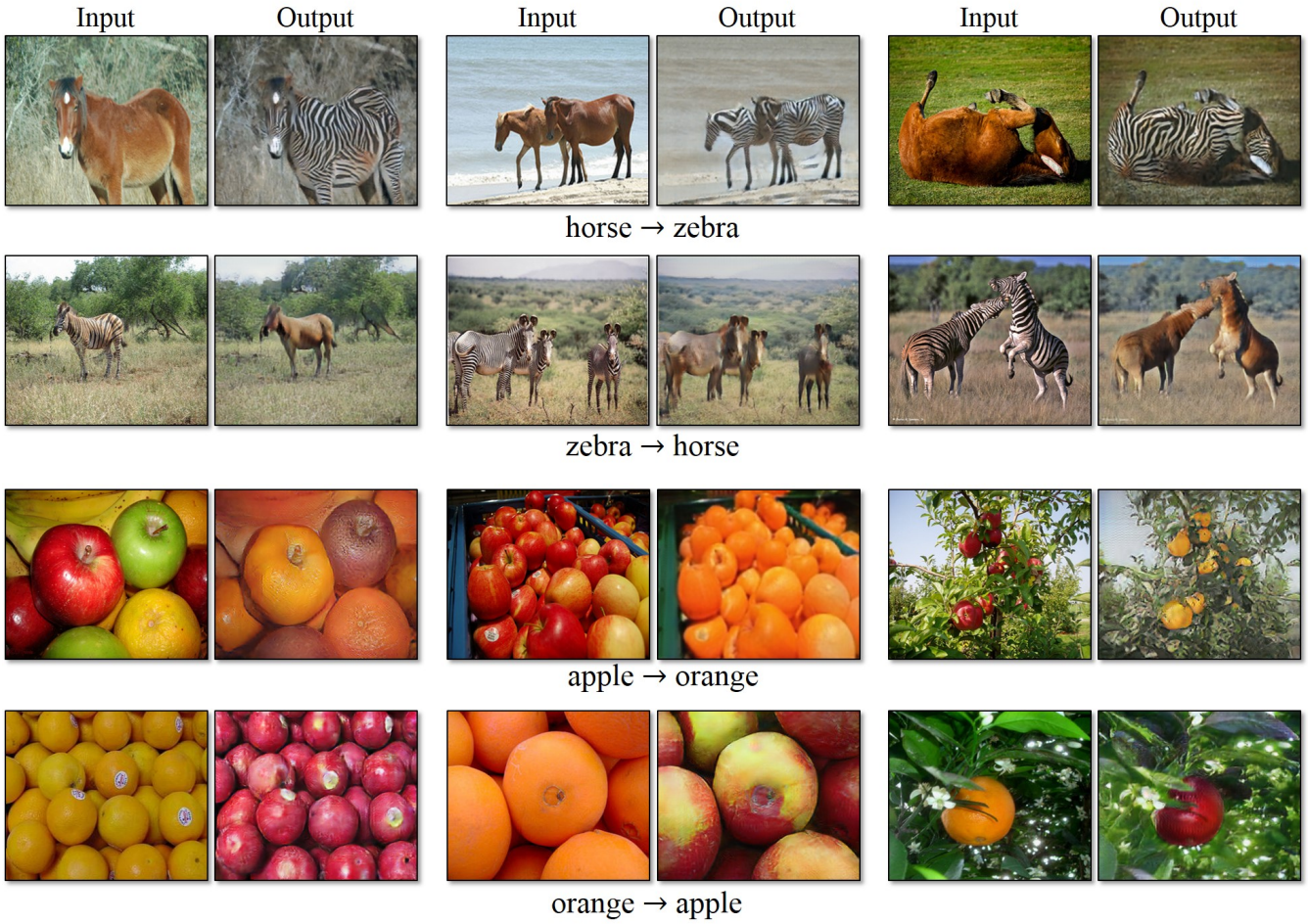
Object Transfiguration

Object transfiguration between horses and zebras:

Best results | **Random training set results** | **Random test set results**

Object transfiguration between apples and oranges:

Best results | **Random training set results** | **Random test set results**



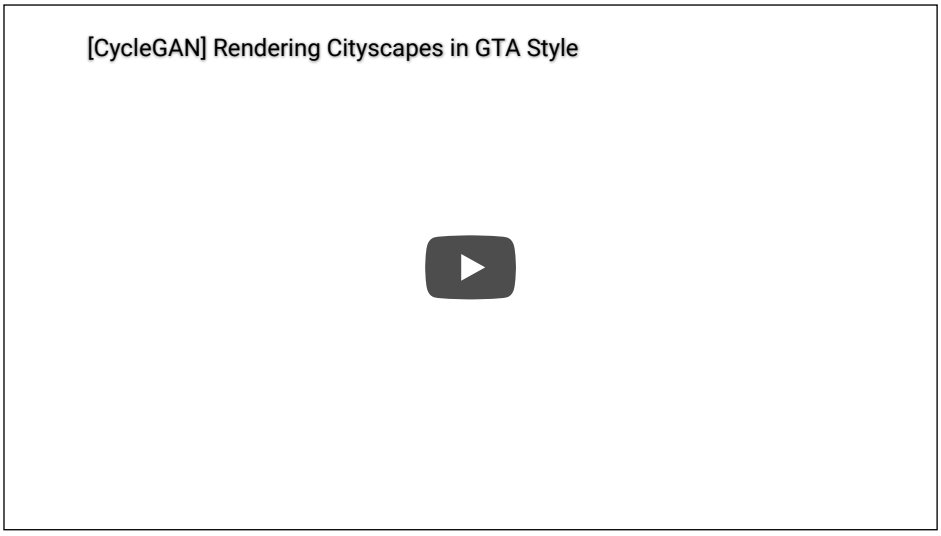
Horse Video to Zebra Video

Turning a horse video into a zebra video (by CycleGAN)

Driving Applications (CG → Real and Day → Night)

Translation between driving scenes in different style. Each frame was rendered independently.

Between **Cityscapes** and **GTA dataset**



Between Day and Night driving using the **Berkeley Deep Drive** dataset (not public yet)

The GTA → Cityscapes results of CycleGAN can be used for domain adaptation for segmentation.
A segmentation model trained on the Cityscapes-style GTA images yields mIoU of 37.0 on the segmentation task on Cityscapes.
More information can be found at **Cycada**.
You can download the **original GTA images (18GB)** and **the translated Cityscapes-style GTA images (16GB)**.

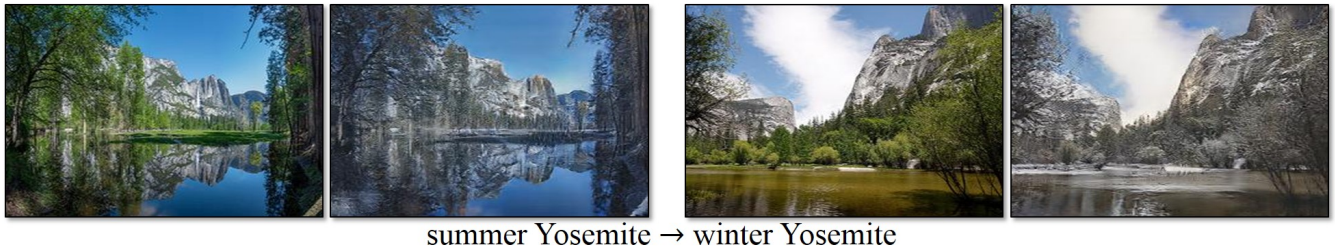
GTA → Cityscapes

Season Transfer

Transferring seasons of Yosemite in the Flickr photos:

Best results | **Random training set results** | **Random test set results**

winter Yosemite → summer Yosemite



summer Yosemite → winter Yosemite

Photo Enhancement

iPhone photos → DSLR photos: generating photos with shallower depth of field.

Best Results | **Random training set results** | **Random test set results**

Input Output Input Output Input Output

Experiments and comparisons

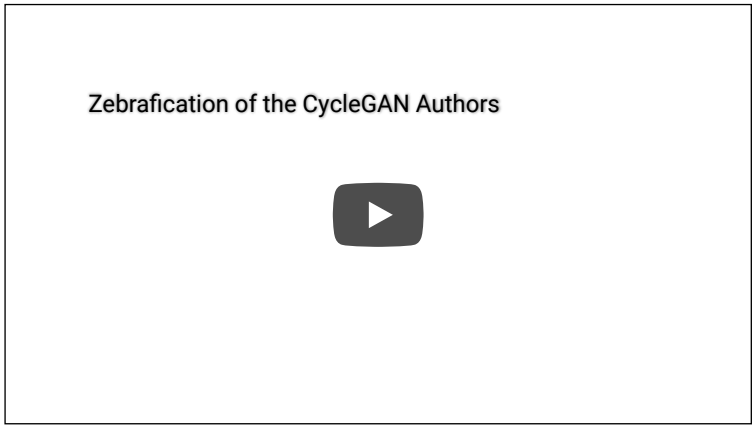
- **Comparison on Cityscapes**: different methods for mapping labels ↔ photos trained on Cityscapes.
- **Comparison on Maps**: different methods for mapping aerialphotos ↔ maps on Google Maps.
- **Facade results**: CycleGAN for mapping labels ↔ facades on **CMP** Facades datasets.
- **Ablation studies**: different variants of our method for mapping labels ↔ photos trained on Cityscapes.
- **Image reconstruction results**: the reconstructed images F(G(x)) and G(F(y)) from various experiments.
- **Style transfer comparison**: we compare our method with neural style transfer [Gatys et al. '15].
- **Identity mapping loss**: the effect of the identity mapping loss on Monet to Photo.

Failure Cases

Our model does not work well when a test image looks unusual compared to training images, as shown in the left figure. See more typical failure cases **[here]**. On translation tasks that involve color and texture changes, like many of those reported above, the method often succeeds. We have also explored tasks that require geometric changes, with little success. For example, on the task of dog ↔ cat transfiguration, the learned translation degenerates into making minimal changes to the input. Handling more varied and extreme transformations, especially geometric changes, is an important problem for future work. We also observe a lingering gap between the results achievable with paired training data and those achieved by our unpaired method. In some cases, this gap may be very hard -- or even impossible, -- to close: for example, our method sometimes permutes the labels for tree and building in the output of the cityscapes photos → labels task. To resolve this ambiguity may require some form of weak semantic

supervision. Integrating weak or semi-supervised data may lead to substantially more powerful translators, still at a fraction of the annotation cost of the fully-supervised systems.

Meet the Authors of CycleGAN



Related Work

- Ian J. Goodfellow, Jean Pouget-Abadie, Mehdi Mirza, Bing Xu, David Warde-Farley, Sherjil Ozair, Aaron Courville, Yoshua Bengio **"Generative Adversarial Networks"**, in NIPS 2014.
- Alec Radford, Luke Metz and Soumith Chintala **"Unsupervised Representation Learning with Deep Convolutional Generative Adversarial Networks"**, in ICLR 2016.
- Jun-Yan Zhu, Philipp Krähenbühl, Eli Shechtman, and Alexei A. Efros. **"Generative Visual Manipulation on the Natural Image Manifold"**, in ECCV 2016.
- Phillip Isola, Jun-Yan Zhu, Tinghui Zhou, and Alexei A. Efros. **"Image-to-Image Translation with Conditional Adversarial Networks"**, in CVPR 2017.

Acknowledgement

We thank Aaron Hertzmann, Shiry Ginosar, Deepak Pathak, Bryan Russell, Eli Shechtman, Richard Zhang, and Tinghui Zhou for many helpful comments. This work was supported in part by NSF SMA-1514512, NSF IIS-1633310, a Google Research Award, Intel Corp, and hardware donations from NVIDIA. JYZ is supported by the Facebook Graduate Fellowship, and TP is supported by the Samsung Scholarship. The photographs used in style transfer were taken by AE, mostly in France.