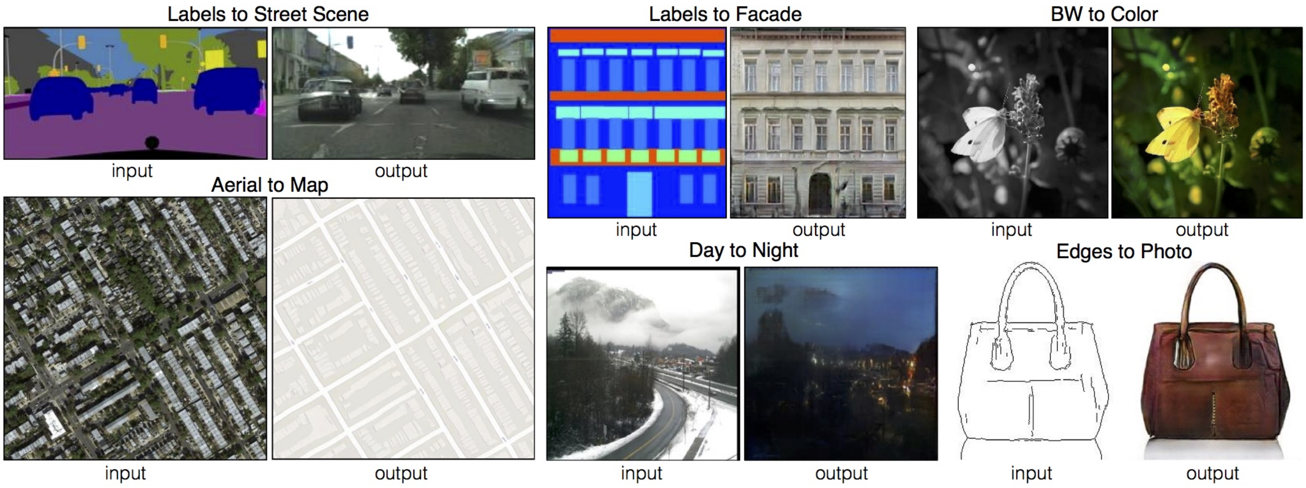
Image-to-Image Translation with Conditional Adversarial Nets

Phillip Isola Jun-Yan Zhu Tinghui Zhou Alexei A. Efros

Univerity of California, Berkeley

In CVPR 2017

[Paper] [GitHub]



Example results on several image-to-image translation problems. In each case we use the same architecture and objective, simply training on different data.

Abstract

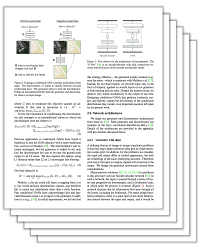
We investigate conditional adversarial networks as a general-purpose solution to image-to-image translation problems. These networks not onl learn the mapping from input image to output image, but also learn a loss function to train this mapping. This makes it possible to apply the sam generic approach to problems that traditionally would require very different loss formulations. We demonstrate that this approach is effective at synthesizing photos from label maps, reconstructing objects from edge maps, and colorizing images, among other tasks. As a community, we n longer hand-engineer our mapping functions, and this work suggests we can achieve reasonable results without hand-engineering our loss functions either.

Try our code

[Torch] [PyTorch]

Ports of our code:
[Tensorflow] (implementation by Christopher Hesse)
[Tensorflow] (implementation by Yen-Chen Lin)
[Chainer] (implementation by pfnet-research)
[Keras] (implementation by Thibault de Boissiere)

Paper

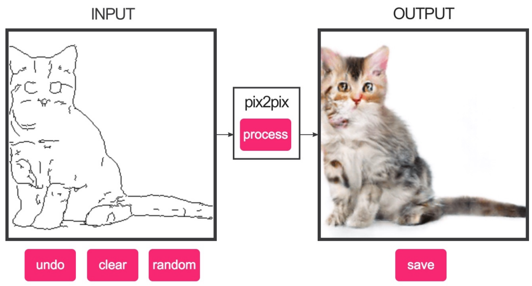


[Download]
7MB

[bibtex]

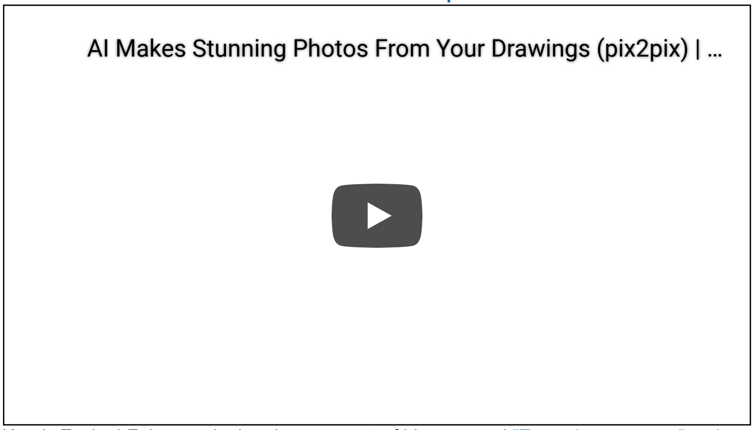
Interactive Demo

(made by Christopher Hesse)



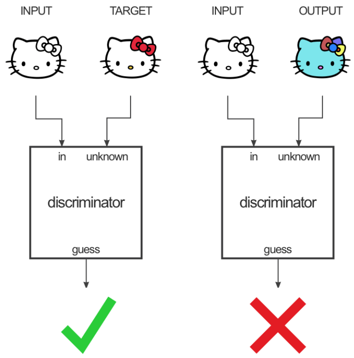
Expository articles and videos

Two-minute Papers



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

Affinelayer blog post



Great explanation by Christopher Hesse, also documenting his tensorflow port of our code.

Experiments

Here we show comprehensive results from each experiment in our paper. Please see the paper for details on these experiments.

Effect of the objective
Cityscapes
Facades

Effect of the generator architecture
Cityscapes

Effect of the discriminator patch scale
Cityscapes
Facades

Additional results
Map to aerial
Aerial to map
Semantic segmentation
Day to night
Edges to handbags
Edges to shoes
Sketches to handbags
Sketches to shoes

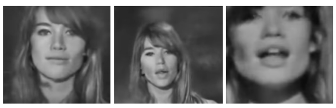
Community contributions: #pix2pix

People have used our code for many creative applications, often posted on twitter with the hashtag #pix2pix. Check them out here! Below we highlight ju a few of the many:

#edges2cats

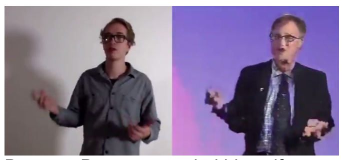
Christopher Hesse trained our model on converting edge maps to photos of cats, and included this in his interactive demo. Apparently, this is what the Internet wanted most, and #edges2cats briefly went viral. The above cats were designed by Vitaly Vidmirov (@vvid).

Alternative Face



Mario Klingemann used our code to translate the appearance of French singer Francoise Hardy onto Kellyanne Conway's infamous "alternative facts" interview. Interesting articles about it can be read here and here.

Person-to-Person



Brannon Dorsey recorded himself mimicking frames from a video of Ray Kurzweil giving a talk. He then used this data to train a Dorsey→Kurzweil translator, allowing him to become a kind of puppeter in control of Kurzweil's appearance.

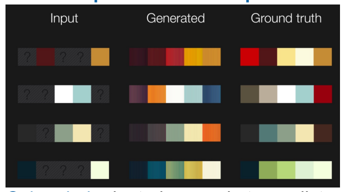
Interactive Anime

Bertrand Gondouin trained our method to translate sketches→Pokemon, resulting in an interactive drawing tool.

Background masking

Kaihu Chen performed a number of interesting experiments using our method, including getting it to mask out the background of a portrait as shown above.

Color palette completion



Colormind adapted our code to predict a complete 5-color palette given a subset of the palette as input. This application stretches the definition of what counts as "image-to-image translation" in an exciting way: if you can visualize your input/output data as images, then image-to-image methods are applicable! (not that this is necessarily the best choice of representation, just one to think about.)

Recent Related Work

Generative adversarial networks have been vigorously explored in the last two years, and many conditional variants have been proposed. Please see th discussion of related work in our paper. Below we point out three papers that especially influenced this work: the original GAN paper from Goodfellow et al., the DCGAN framework, from which our code is derived, and the iGAN paper, from our lab, that first explored the idea of using GANs for mapping us strokes to images.

Ian J. Goodfellow, Jean Pouget-Abadie, Mehdi Mirza, Bing Xu, David Warde-Farley, Sherjil Ozair, Aaron Courville, and Yoshua Bengio. **Generative Adversarial Networks**. NIPS, 2014. [PDF]

Alec Radford, Luke Metz, Soumith Chintala. **Unsupervised Representation Learning with Deep Convolutional Generative Adversarial Networks**. ICLR, 2016. [PDF][Code]

Jun-Yan Zhu, Philipp Krahenbuhl, Eli Shechtman, Alexei A. Efros. **Generative Visual Manipulation on the Natural Image Manifold**. ECCV, 2016. [PD [Webpage][Code]

Also, please check out our follow-up work on image-to-image translation *without* paired training examples:

Jun-Yan Zhu*, Taesung Park*, Phillip Isola, Alexei A. Efros. **Unpaired Image-to-Image Translation using Cycle-Consistent Adversarial Networks**. arXiv, 2017. [PDF][Webpage][Code]

Acknowledgements

We thank Richard Zhang, Deepak Pathak, and Shubham Tulsiani for helpful discussions. Thanks to Saining Xie for help with the HED edge detector. Thanks to the online community for exploring many applications of our work and pointing out typos and errors in the paper and code. This work was supported in part by NSF SMA-1514512, NGA NURI, IARPA via Air Force Research Laboratory, Intel Corp, Berkeley Deep Drive, and hardware donatio by Nvidia. Disclaimer: The views and conclusions contained herein are those of the authors and should not be interpreted as necessarily representing th official policies or endorsements, either expressed or implied, of IARPA, AFRL or the U.S. Government.