Image-to-Image Translation with Conditional Adversarial Networks

Wenjie Niu

August 10, 2018
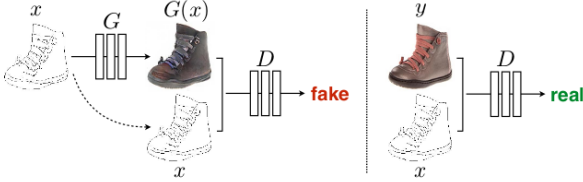

Figure 1. Training a conditional GAN to map edges→ photo. The discriminator, D, learns to classify between fake (synthesized by the generator) and real {edge, photo} tuples. The generator, G, learns to fool the discriminator. Unlike an unconditional GAN, both the generator and discriminator observe the input edge map. [3]

Abstract

Conditional adversarial networks not only 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 same generic approach to problems that traditionally would require very different loss formulations. Indeed, since the release of the pix2pix software associated with this paper, a large number of internet users (many of them artists) have posted their own experiments with our system, further demonstrating its wide applicability and ease of adoption without the need for parameter tweaking.

1. Introduction

Many problems in image processing, computer graphics, and computer vision can be posed as translating an input image into a corresponding output image. Just as a concept may be expressed in either English or French, a scene may be rendered as an RGB image, a gradient field, an edge map, a semantic label map, etc.

In analogy to automatic language translation, they define automatic image-to-image translation as the task of translating one possible representation of a scene into another, given sufficient training data (see Figure 2).

The community has already taken significant steps in this direction, with convolutional neural nets (CNNs) becoming the common workhorse behind a wide variety of image prediction problems. CNNs learn to minimize a loss function an objective that scores the quality of results and although the learning process is automatic, a lot of manual effort still goes into designing effective losses. In other words, we still have to tell the CNN what we wish it to minimize.

GANs have been vigorously studied in the last two years and many of the techniques it is explored in this paper have been previously proposed. Nonetheless, earlier papers have focused on specific applications, and it has remained unclear how effective image-conditional GANs can be as a general-purpose solution for image-to-image translation. Their primary contribution is to demonstrate that on a wide variety of problems, conditional GANs produce reasonable results. Their second contribution is to present a simple framework sufficient to achieve good results, and to analyze the effects of several important architectural choices.

2. Method

GANs are generative models that learn a mapping from random noise vector z to output image y, $G : z \rightarrow y$ [1]. In contrast, conditional GANs learn a mapping from observed image x and random noise vector z, to y, $G : \{x, z\} \rightarrow y$. The generator G is trained to produce outputs that cannot be distinguished from "real" images by an adversarially trained discriminator, D, which is trained to do as well as possible at detecting the generators "fakes". This training procedure is diagrammed in Figure 1.

2.1. Network architectures

They adapt our generator and discriminator architectures from those in [4]. Both generator and discriminator use modules of the form convolution-BatchNorm-ReLu [2]. Details of the architecture are provided in the supplemental materials online, with key features discussed below.

References

[1] I. J. Goodfellow, J. Pouget-Abadie, M. Mirza, B. Xu, D. Warde-Farley, S. Ozair, A. Courville, and Y. Bengio. Generative adversarial nets. In *NIPS*, 2014. 1
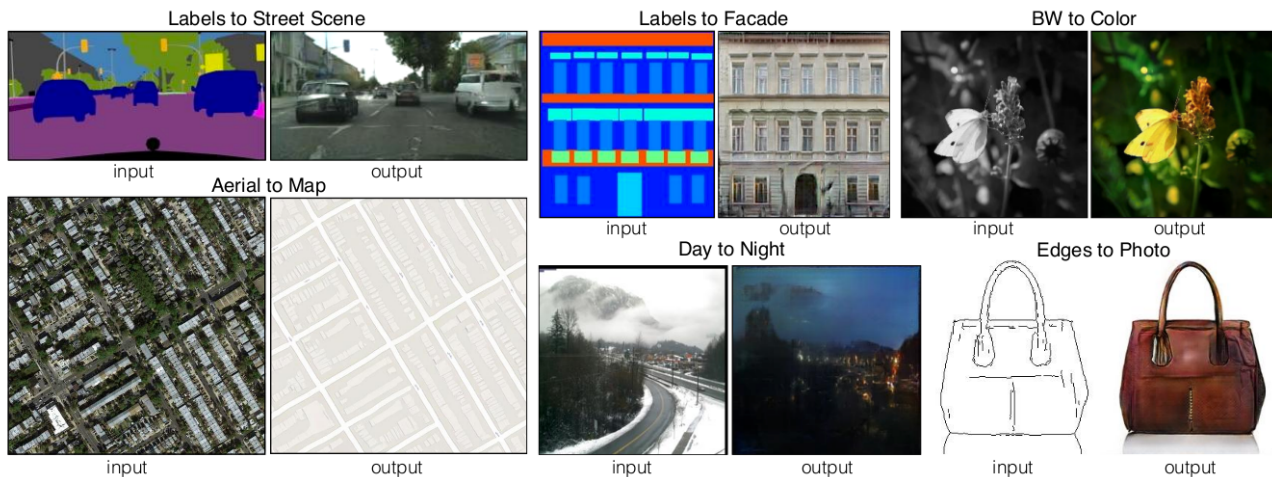
Figure 2. Many problems in image processing, graphics, and vision involve translating an input image into a corresponding output image. These problems are often treated with application-specific algorithms, even though the setting is always the same: map pixels to pixels. Conditional adversarial nets are a general-purpose solution that appears to work well on a wide variety of these problems. Here we show results of the method on several. In each case we use the same architecture and objective, and simply train on different data. [3]

[2] S. Ioffe and C. Szegedy. Batch normalization: accelerating deep network training by reducing internal covariate shift. In *ICML*, 2015. 1

[3] P. Isola, J. Y. Zhu, T. Zhou, and A. A. Efros. Image-to-image translation with conditional adversarial networks. In *CVPR*, 2017. 1, 2

[4] A. Radford, L. Metz, and S. Chintala. Unsupervised representation learning with deep convolutional generative adversarial networks. In *ICLR*, 2016. 1