Image-to-Image Translation

Pix2Pix

CVPR

2017

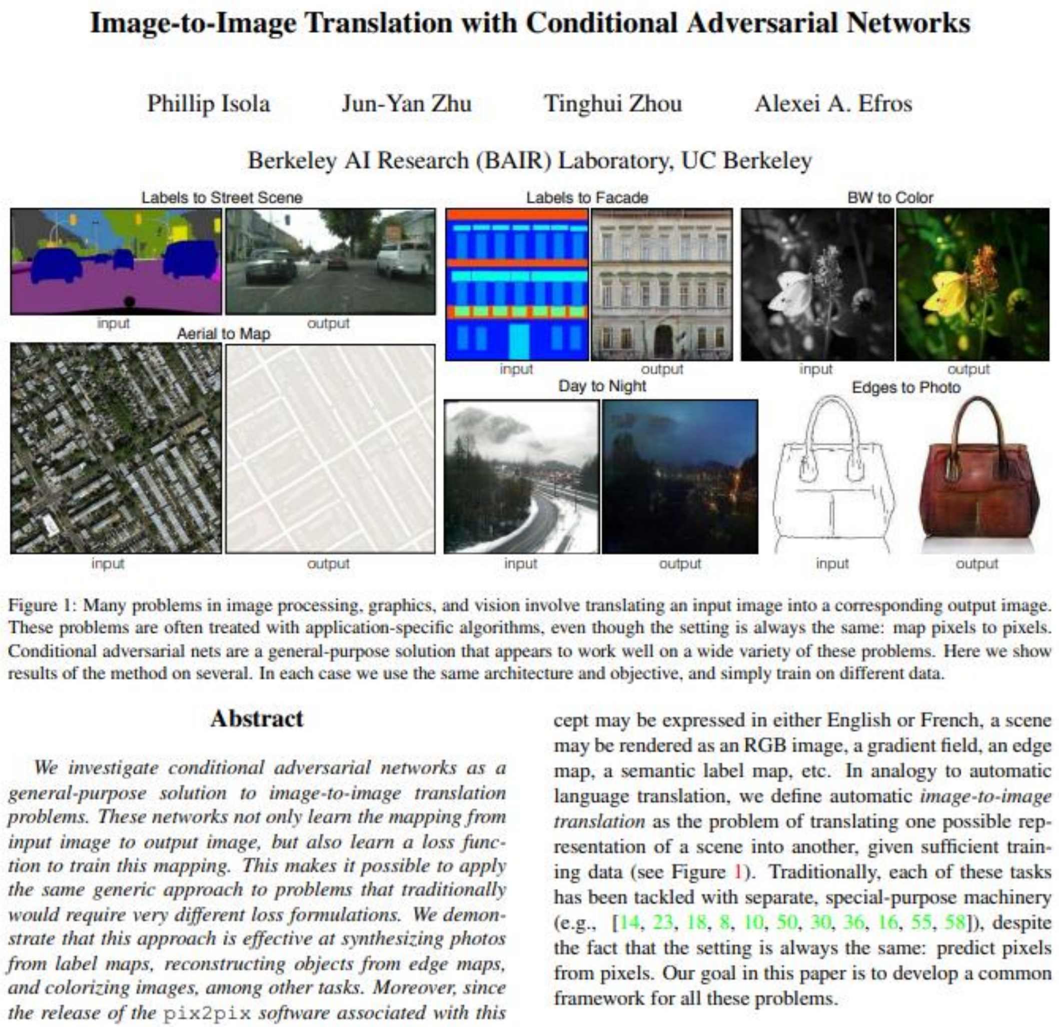

Image-to-Image Translation with Conditional Adversarial Networks

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

Berkeley AI Research (BAIR) Laboratory, UC Berkeley

Figure 1: 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.

Abstract

We investigate conditional adversarial networks as a general-purpose solution to image-to-image translation problems. These 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. We demonstrate that this approach is effective at synthesizing photos from label maps, reconstructing objects from edge maps, and colorizing images, among other tasks. Moreover, since the release of the pix2pix *software associated with this*

cept 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, we define automatic *image-to-image translation* as the problem of translating one possible representation of a scene into another, given sufficient training data (see Figure 1). Traditionally, each of these tasks has been tackled with separate, special-purpose machinery (e.g., [14, 23, 18, 8, 10, 50, 30, 36, 16, 55, 58]), despite the fact that the setting is always the same: predict pixels from pixels. Our goal in this paper is to develop a common framework for all these problems.

Isola, P., Zhu, J. Y., Zhou, T., & Efros, A. A. (2017). Image-to-image translation with conditional adversarial networks. In *Proceedings of the IEEE conference on computer vision and pattern recognition* (pp. 1125-1134).

Image-to-Image Translation

Pix2Pix

Background & Goal

Previous research limitation

- Training CNN-based models to minimize L2 distance tends to produce blurred results

Goal

- To produce realistic images instead of blurred ones
- Develop a common Framework

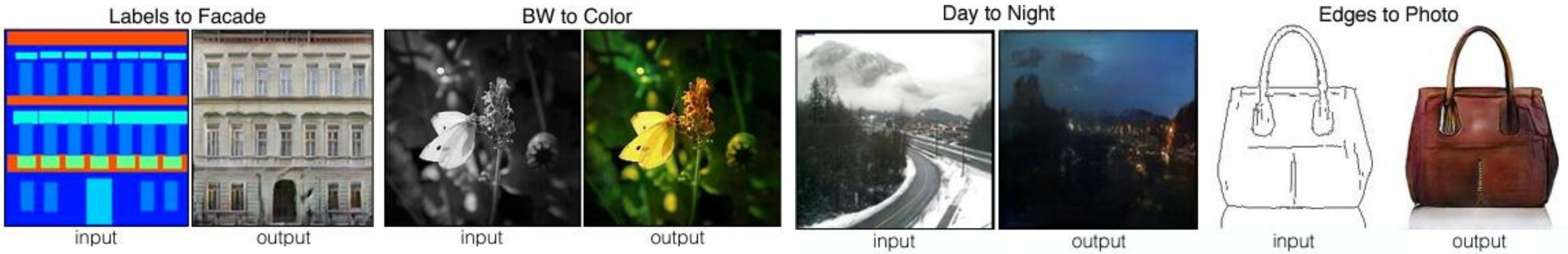

Fig 1. Many problems in image processing involve translating an input image into a corresponding output image

Isola, P., Zhu, J. Y., Zhou, T., & Efros, A. A. (2017). Image-to-image translation with conditional adversarial networks. In *Proceedings of the IEEE conference on computer vision and pattern recognition* (pp. 1125-1134).

Image-to-Image Translation

Pix2Pix

Network Architectures

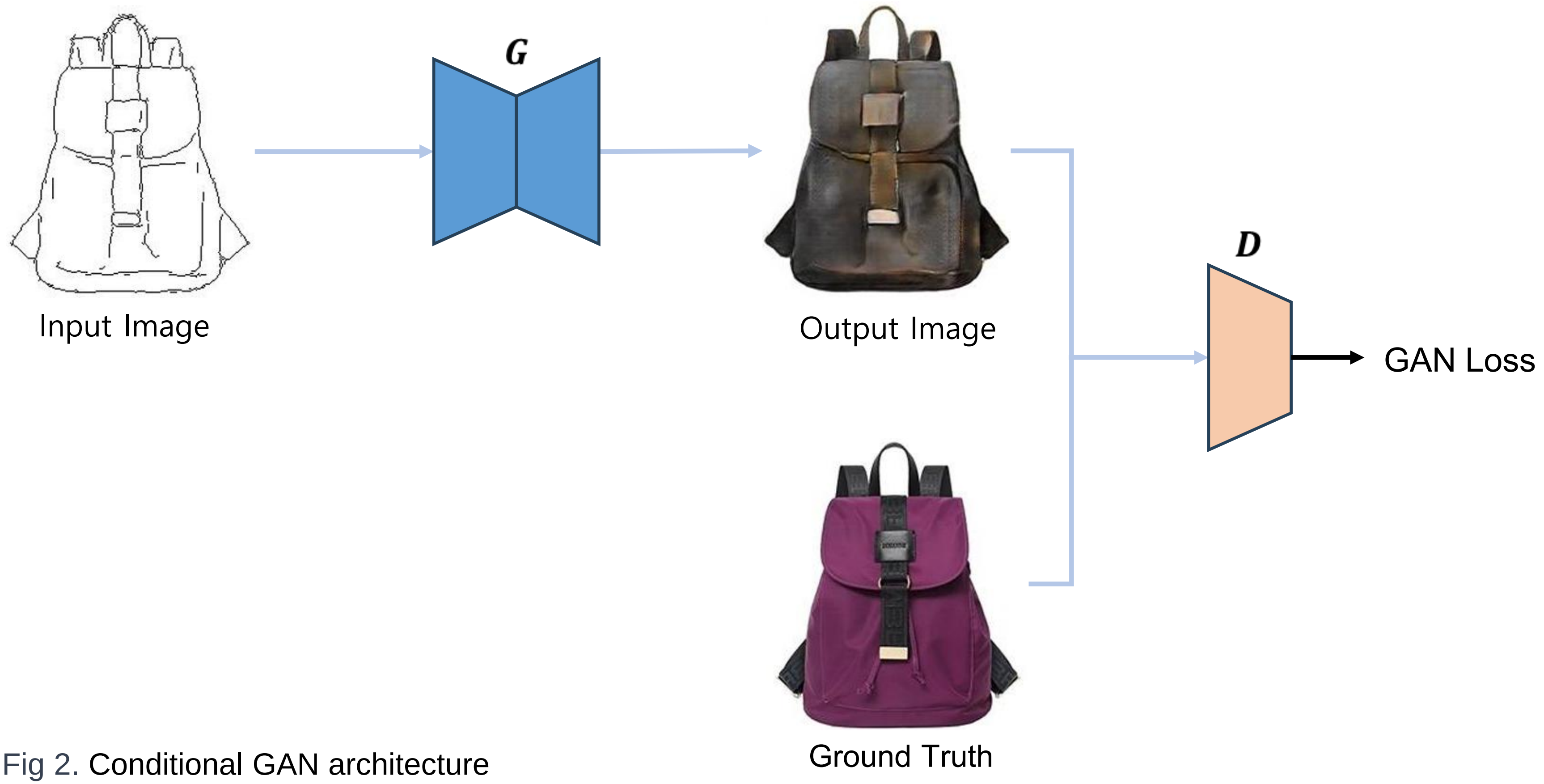

Fig 2. Conditional GAN architecture

Image-to-Image Translation

Pix2Pix

Network Architectures

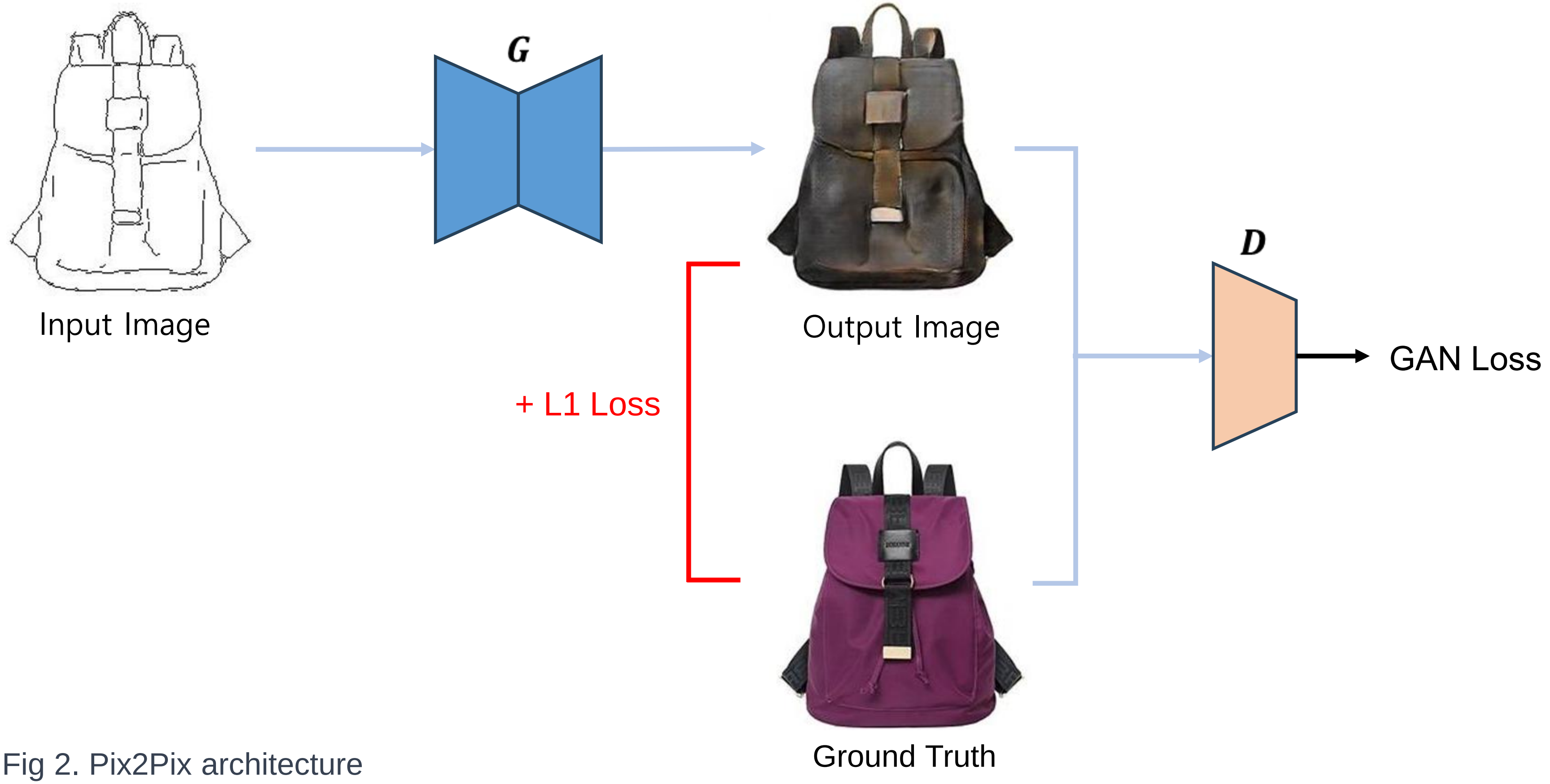

Fig 2. Pix2Pix architecture

Image-to-Image Translation

Pix2Pix

Network Architectures

Generator = U-Net

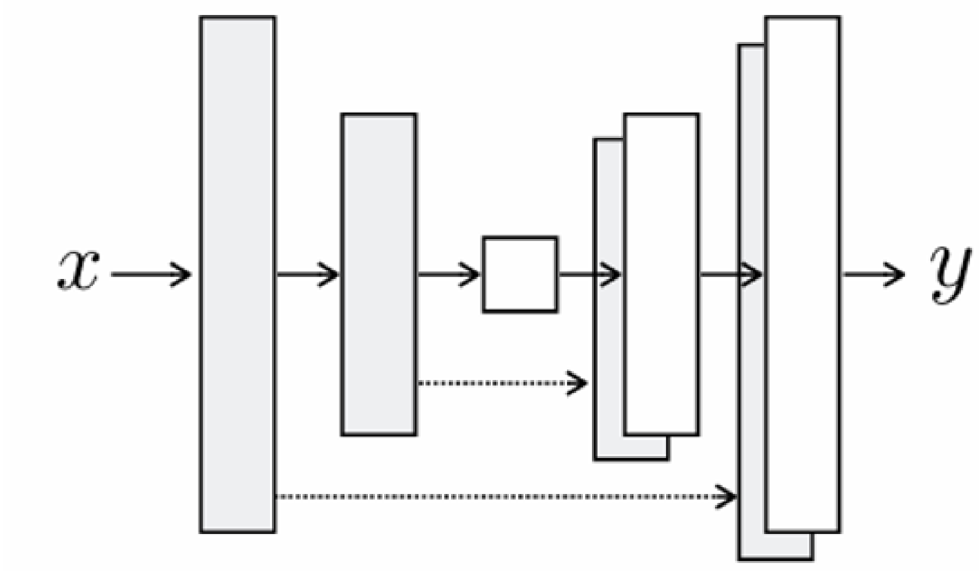

Fig 3. Generator of Pix2Pix

Discriminator = PatchGAN

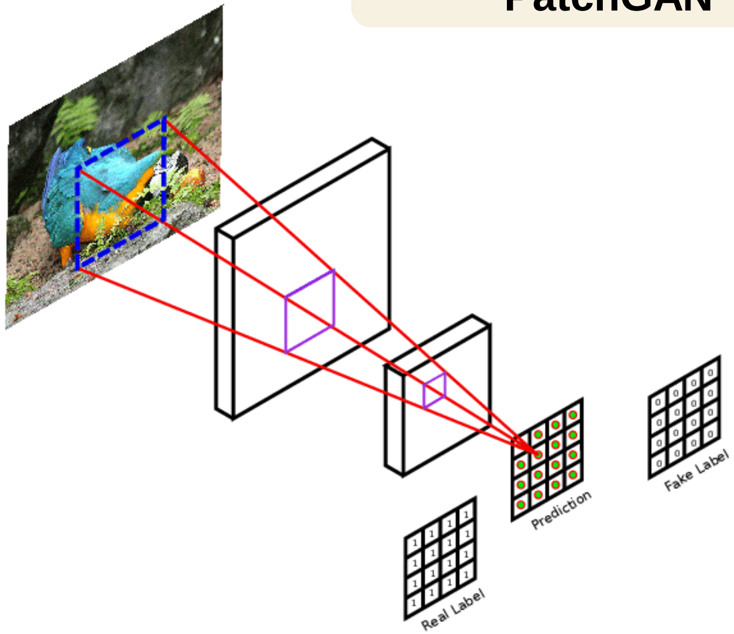

Fig 4. Discriminator of Pix2Pix

Isola, P., Zhu, J. Y., Zhou, T., & Efros, A. A. (2017). Image-to-image translation with conditional adversarial networks. In *Proceedings of the IEEE conference on computer vision and pattern recognition* (pp. 1125-1134).

Image-to-Image Translation

Pix2Pix

Loss Functions

GAN loss

- Conditional GAN learns a mapping from an image x and a random noise vector z to y. $G : \{x, z\} \rightarrow y$

$$\mathcal{L}_{cGAN}(G, D) = \mathbb{E}_{x,y}[\log D(x, y)] + \mathbb{E}_{x,z}[\log(1 - D(x, G(x, z))],$$

Eq 1. Conditional GANs Loss Function

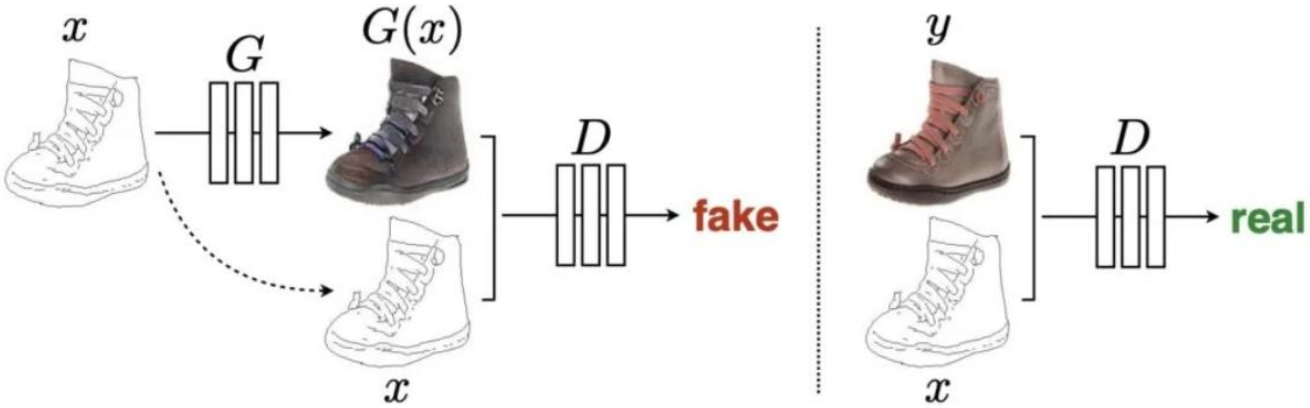

Fig 5. Overview of conditional GAN

Isola, P., Zhu, J. Y., Zhou, T., & Efros, A. A. (2017). Image-to-image translation with conditional adversarial networks. In *Proceedings of the IEEE conference on computer vision and pattern recognition* (pp. 1125-1134).

Image-to-Image Translation

Pix2Pix

Loss functions

L1 loss

- It is beneficial to combine the GAN loss with a more traditional loss (pixel-wise loss)
- Using L1 distance rather than L2 as L1 encourages less blurring

$$\mathcal{L}_{L1}(G) = \mathbb{E}_{x,y,z}[\|y - G(x,z)\|_1].$$

Eq 2. L1 Distance

Ignore random noise

- Condition image itself contains highly informative prior.
- Since the L1 loss drives the output to resemble the ground truth, generator learned to ignore the noise.

$$L_{pix2pix}(G,D) = \arg\min_G \max_D \mathcal{L}_{cGAN}(G,D) + \lambda\mathcal{L}_{L1}(G).$$

Eq 3. Final Loss Function

Image-to-Image Translation

Pix2Pix

Experiment Results

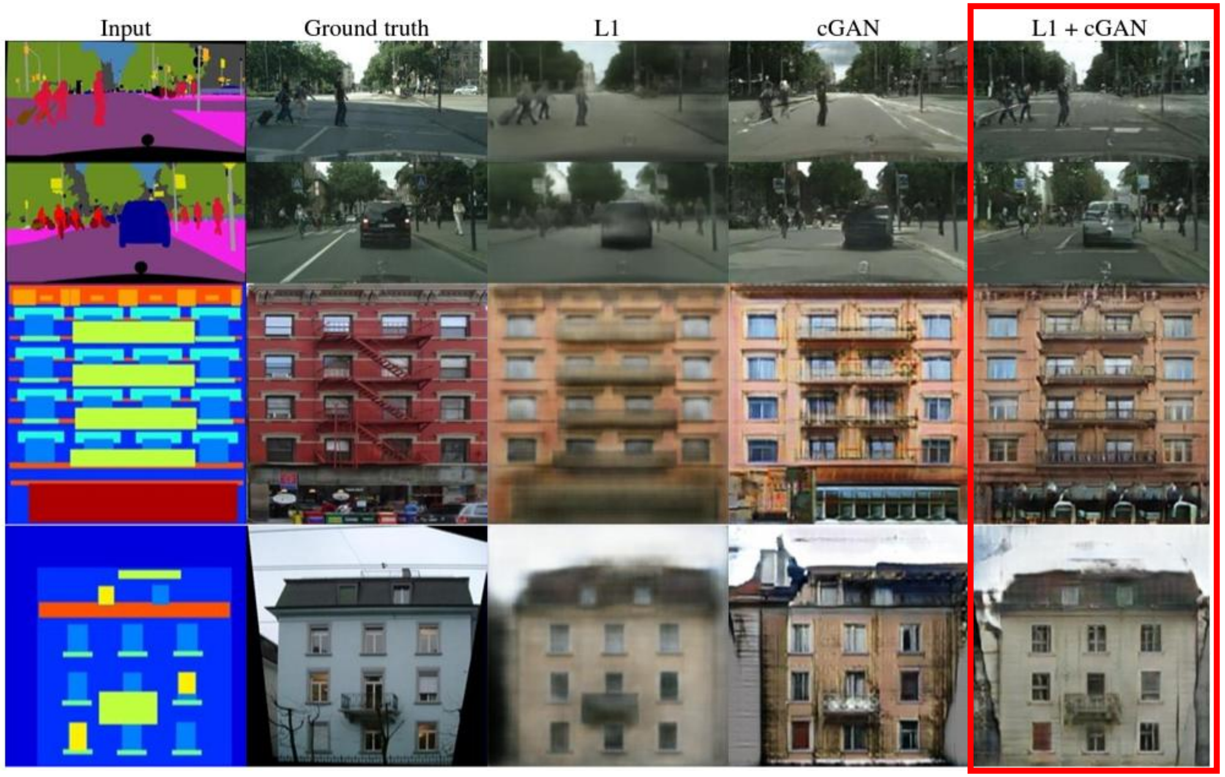

Fig 6. Results by losses

Isola, P., Zhu, J. Y., Zhou, T., & Efros, A. A. (2017). Image-to-image translation with conditional adversarial networks. In *Proceedings of the IEEE conference on computer vision and pattern recognition* (pp. 1125-1134).

Pix2Pix

Implications & Limitations

Implications

- A general framework for supervised image-to-image translation.
- By combining conditional GANs, the model preserves more realistic details than pixel-wise loss.

Limitations

- As a supervised learning method, it requires a dataset with accurately paired input-output images.
- Obtaining such datasets is challenging, making it difficult to apply to various applications.