Image-to-Image Translation with Conditional GAN

Weini Yu, Zhouchangwan Yu, Jason Hu

STANFORD UNIVERSITY

Introduction

- **Image-to-image translation** are tasks that take in input images and generate or manipulate them into a different visual space. Traditionally this task requires hand-crafted machinery.
- In this study we explore image translation using **conditional-general adversarial networks (C-GAN)**, in which we translate images using GAN conditioned on input images and generate desired output image.

Data

- Paired aerial images and corresponding maps scraped from Google Maps.
- Image size 600 x 600 in JPEG
- 1097 training, 1098 validation, and 1098 test examples.

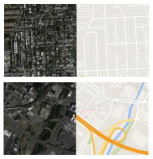


Fig 1. Training data samples

Preprocessing

- Resize to 286 x 286
- Random crop to 256 x 256
- Random horizontal flip
- Zero center and normalize all pixel values to [-1, 1]

Method

Conditional-GAN

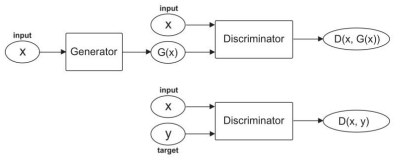


Figure 2. Conditional-GAN model

Conditional adversarial loss:

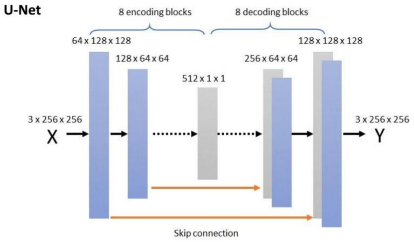
$L_{cGAN}(G, D) = E_{x,y}[logD(x, y)] + E_x[\log(1 - D(x, G(x)))]$

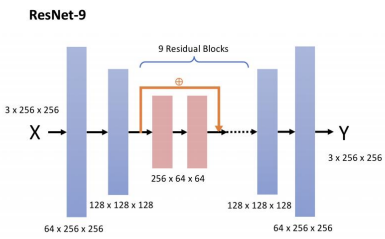
Objective of minmax game:

$(G^*, D^*) = arg \min_G \max_D (L_{cGAN}(G, D) + \lambda L_{L1}(G))$

Method

Generator





Discriminator

PatchGAN

- Classifies whether an image is real or not with receptive field of 70 x 70
- 5 downsampling convolutional-norm-LeakyReLU layers

ImageGAN

- Classifies whether an image is real or not with receptive field of the entire image
- 7 downsampling convolutional-norm-LeakyReLU layers

Results and Analysis

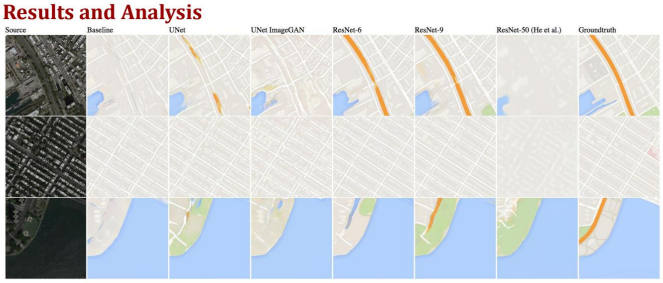


Fig 2. Generated map images of different architecture and hyperparameters

Model	MSE
Baseline	$1.5661e^{-2}$
U-Net	$1.5058e^{-2}$
ResNet-6	$1.3298e^{-2}$
ResNet-9	$\mathbf{1.1969e^{-2}}$
ResNet50 (He et al.)	$1.5848e^{-2}$

Table 1. Mean Squared Error between different models and ground truth

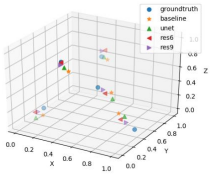


Fig 3. t-SNE plot of 5 random generated images from different models and ground truth

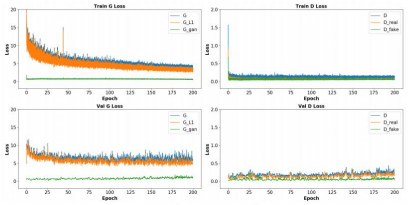


Fig 4. Loss of ResNet-9 PatchGAN

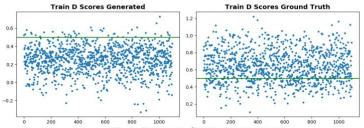


Fig 5. Train set discriminator score

Conclusions

- C-GAN is effective in translating image from one domain to another.
- Residual connection is very effective in image translation tasks because it makes deeper networks easier to train.
- Our residual-based networks outperform U-Net model in aerial-map translation.

Future Work

- Explore residual-based network for discriminator
- Experiment with dynamic training frequency that allows generator to train more often than discriminator in the beginning and gradually slow down