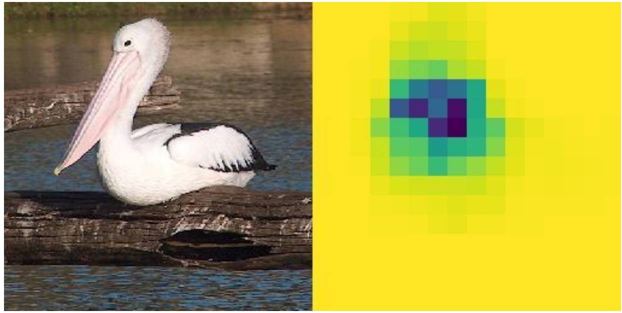
Image-to-Image Translation with Conditional Adversarial Networks

Pengkai Zhu, Weiwei Tao, Hari Saran

Task

https://www.researchgate.net/figure

http://theusbport.com/google-creatives-explain-style-transfer-turn-photos-art/20027



Isola et al., 2016

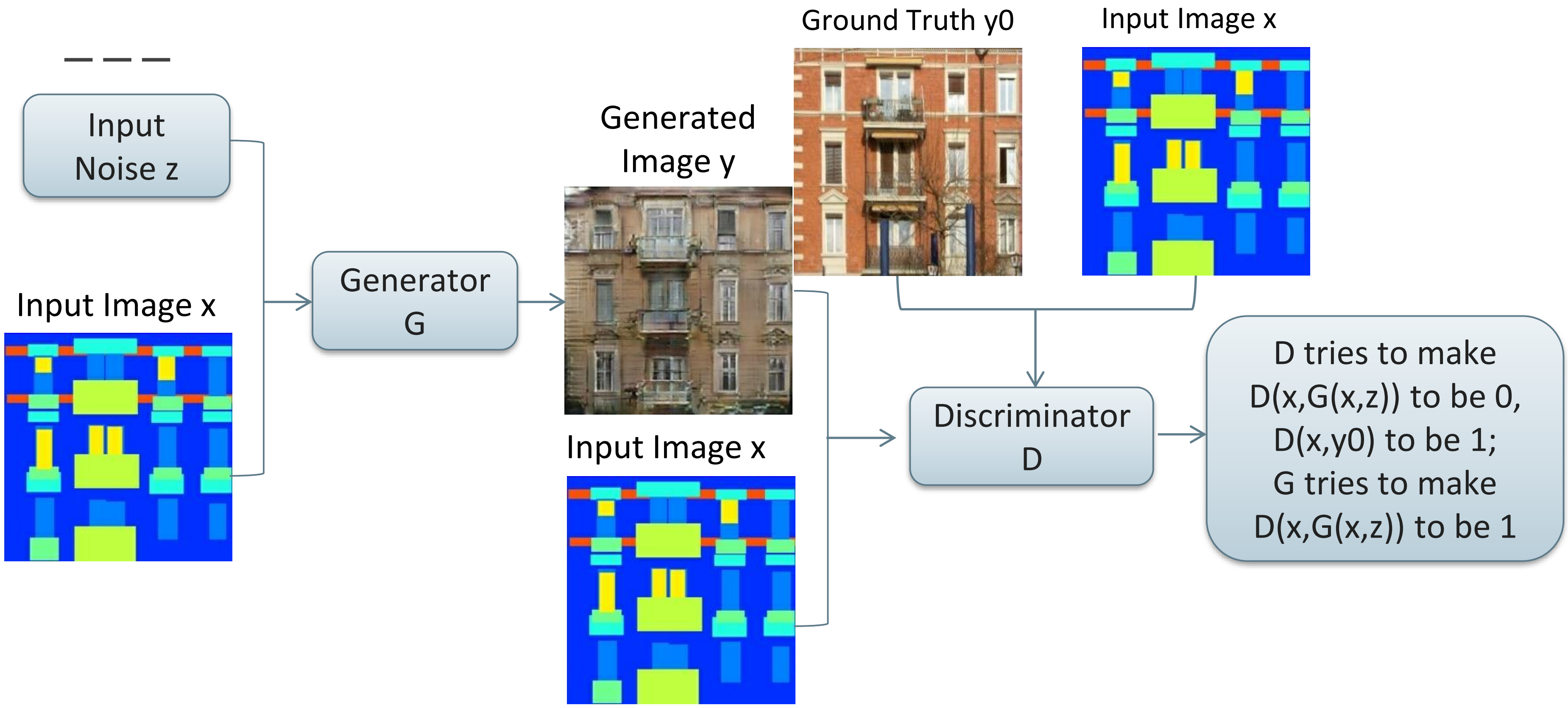
http://techpp.com/2010/02/18/convert-photos-to-sketches-online-for-free/

Related work

- CNN is used to generate colorful images (Zhang et al., 2016).
- GAN (Goodfellow et al., 2014) works well for generating small images.
- Conditional GAN is developed to generate high quality natural images (Denton et al.,2015).
- Deep convolutional GANs (DCGANs) are more stable to be trained (Radford et al., 2016).
- The cGAN objective function is mixed with the traditional L2 loss to generate images, which is proved to give better performance (Pathak et al., 2016).

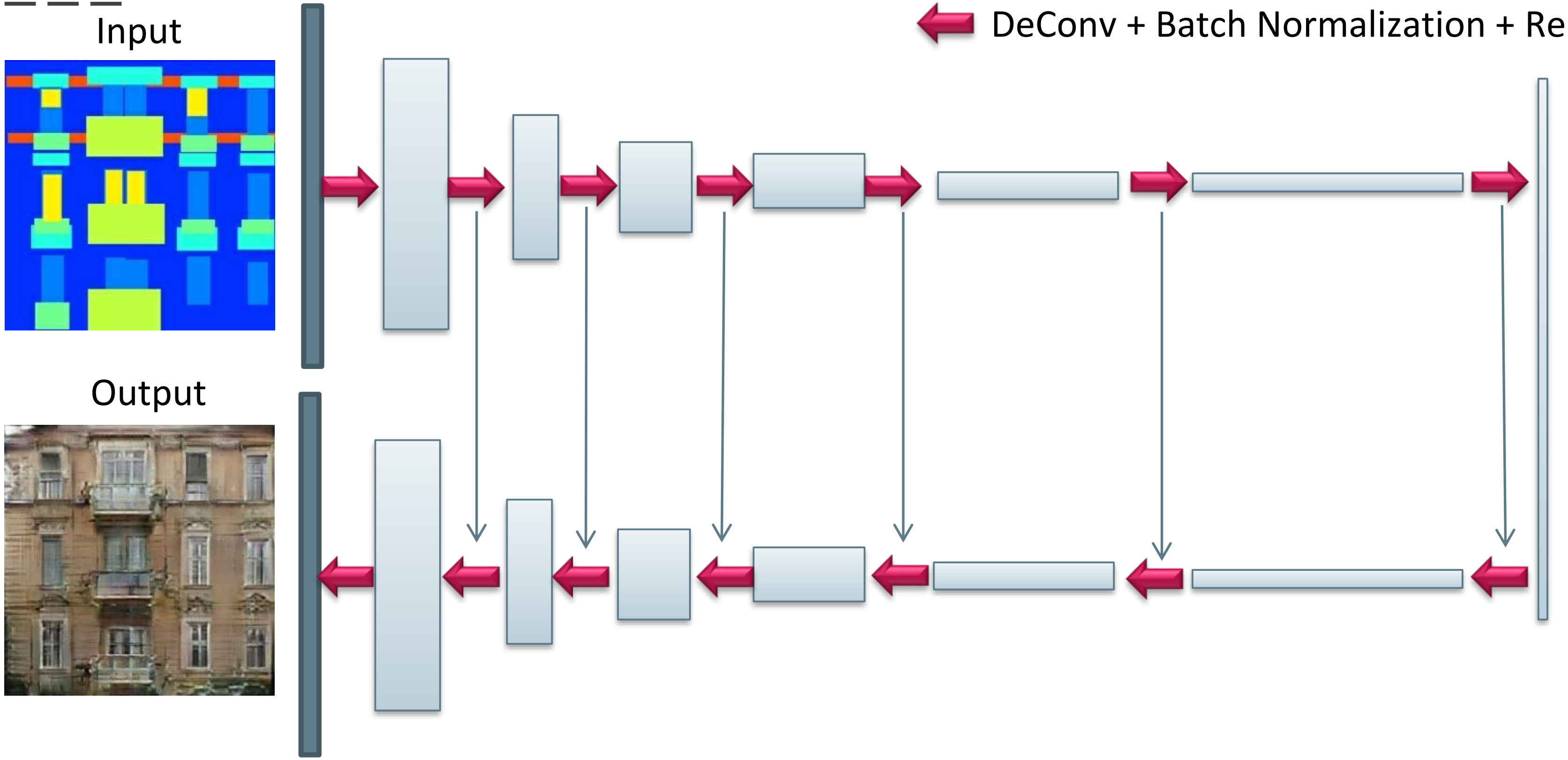
Call for a method which can deal with all kinds of image processing tasks! (Isola et al., 2016)

cGAN framework



Encoder-decoder with skips

Skip
Conv + Batch Normalization + ReLU
DeConv + Batch Normalization + ReLU



Implementation details

Objective function:

$$G^* = \arg\min_G \max_D \mathcal{L}_{cGAN}(G,D) + \lambda \mathcal{L}_1(G)$$

Generator Encoders	conv1	conv2	conv3	conv4	conv5	conv6	conv7	conv8
Input size Input channel	256×256 3	128×128 64	64×64 128	32×32 256	16×16 512	8×8 512	4×4 512	2×2 512
Generator Decoders	unconv1	unconv2	unconv3	unconv4	unconv5	unconv6	unconv7	unconv8
Input size Input channel	1×1 512	2×2 512	4×4 512	8×8 512	16×16 512	32×32 256	64×64 128	128×128 64

Discriminator	conv1	conv2	conv3	conv4	conv5
Input size Input channel	256×256 6	128×128 64	64×64 128	32×32 256	31×31 512

Datasets

Dataset	Training Samples	Validation Samples	Type
Facades	400	100	Structural label ⟷ Photo
Google Maps	1096	1098	Map ⟷ Aerial Photo
Edge to Shoes	16K	200	Edges → Shoes Photo
ImageNet	10K	2500	Photo → Probability Map

Evaluation metric

It is still an open problem

- Plausibility to people:

  Examples illustration

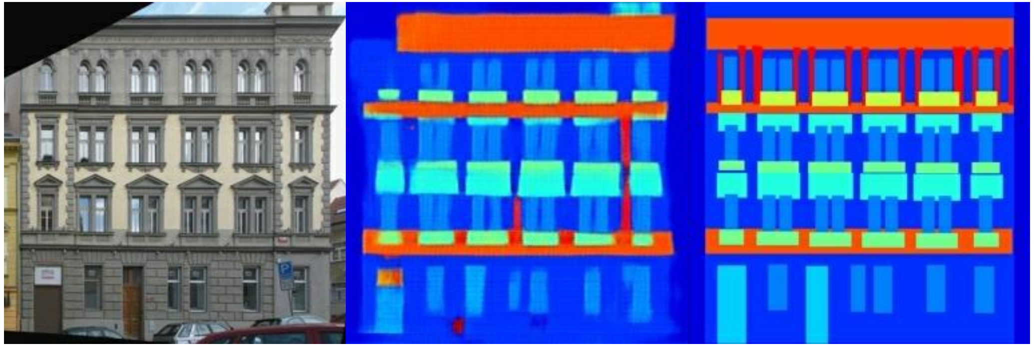
- Quantitative metric:

  Pixel by pixel mean absolute error (MAE) between synthetic image and ground truth

Results

Label to Architecture

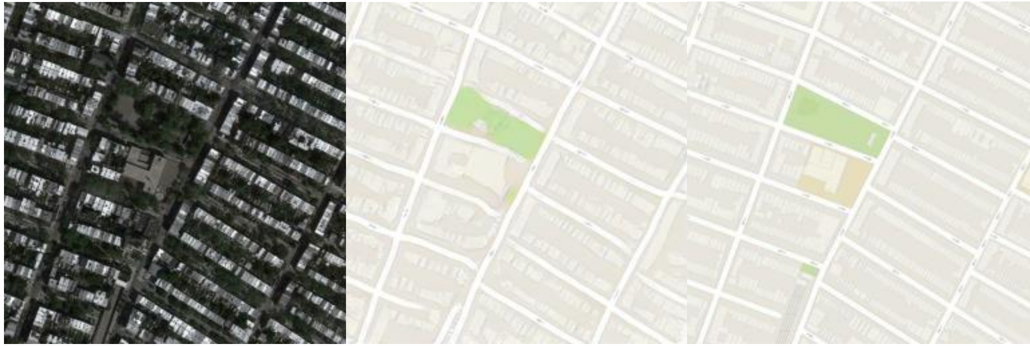
Input Output Ground Truth



Architecture to Label

Results

Aerial Photo to Map



Input Output Ground Truth

Map to Aerial Photo

Results

Edge to Shoes

Input Output Ground Truth

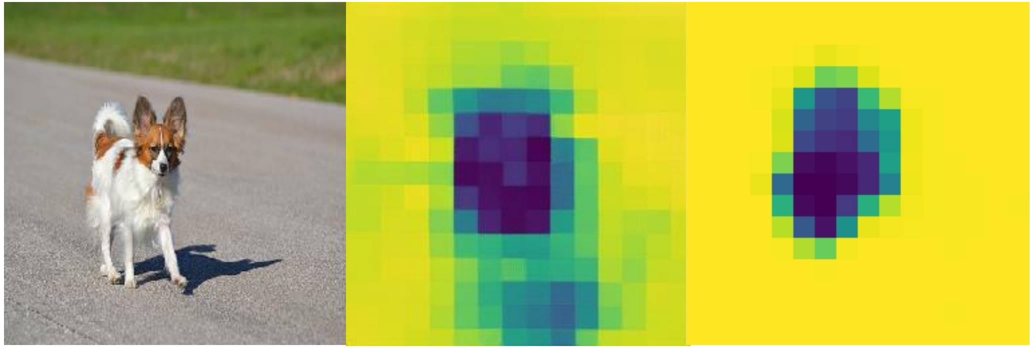
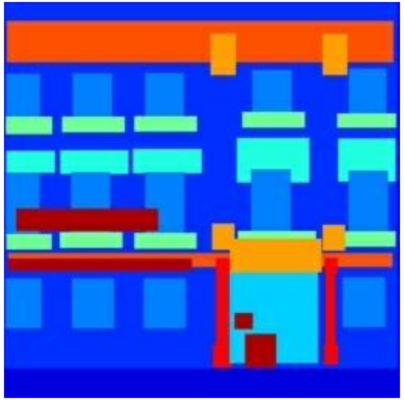


Photo to Probability Map

Loss function analysis

$$G^* = \arg\min_G \max_D \mathcal{L}_{cGAN}(G, D) + \lambda \mathcal{L}_1(G)$$

Input



Ground Truth GAN + L1 GAN GAN + L2

Mean absolute error

Task	MAE
Label to Architecture (cGAN+L1)	0.1757 ± 0.0420
Label to Architecture (cGAN+L2)	0.1778 ± 0.0405
Label to Architecture (cGAN only)	0.1930 ± 0.0525
Architecture to Label	0.1006 ± 0.0278
Aerial Photo to Map	0.0262 ± 0.0086
Map to Aerial Photo	0.1306 ± 0.0250
Edge to Shoes	0.0765 ± 0.0250
Photo to Probability Map	0.2229 ± 0.0831

Conclusion

- cGAN is a promising approach for image translation
- General framework for multiple tasks
- Blurry, artifacts issue
- Pretrained model for generator