Image Inpainting

于建民、鄭筱樺、陳英傑、黃子瑋、張馭荃

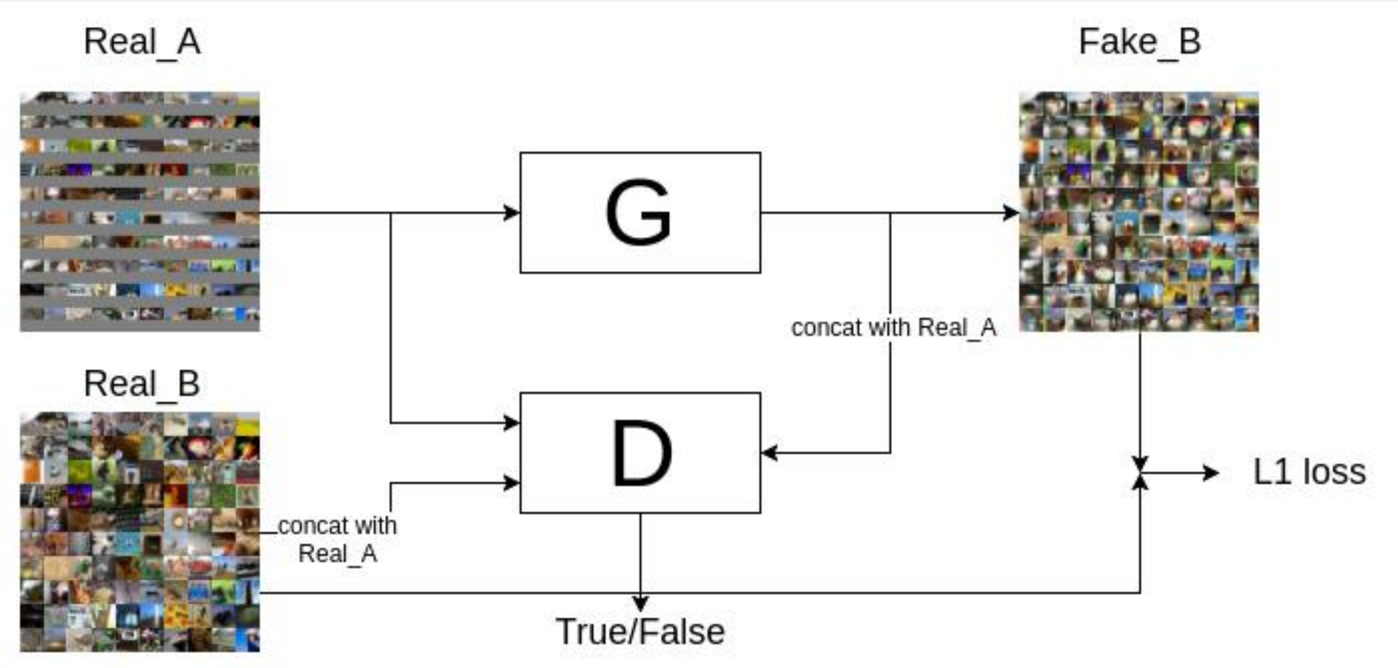
Team Name: IWantGPU

INTRODUCTION

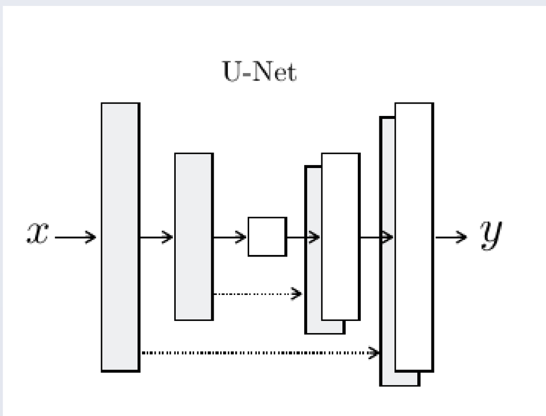
- Image inpainting is the process of reconstructing lost or deteriorated parts of images. We remove the lower half of the images, and aim to recover the lost part or generate reasonable images.

MODELS

- **Pix2Pix**



- Pix2Pix is a conditional gan which conditioned on an input image and generate a corresponding output image. In our case, the input image is an image whose lower-half part is sliced. We want to generate the full image.
- Generator: U-Net based
  - U-Net is an encoder-decoder with skip connections between mirrored layers in the encoder and decoder stacks.
  - Use skip connections to pass information directly across the net.
  - Provide noise in the form of dropout.



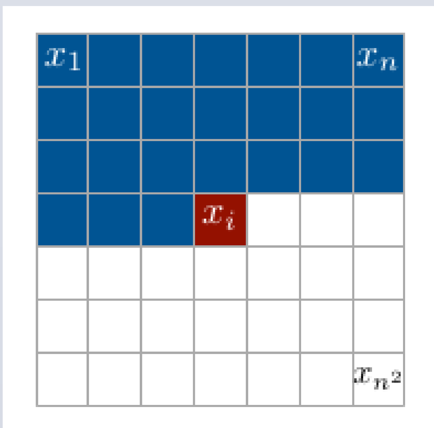
- Discriminator: convolutional PatchGAN classifier
  - Penalize structure at the scale of image patches: classify if each N × N patch in an image is real or fake
  - Since N can be smaller than image size, PatchGAN has fewer parameters. This can run faster and be applied on arbitrarily large images.
- Loss function

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

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

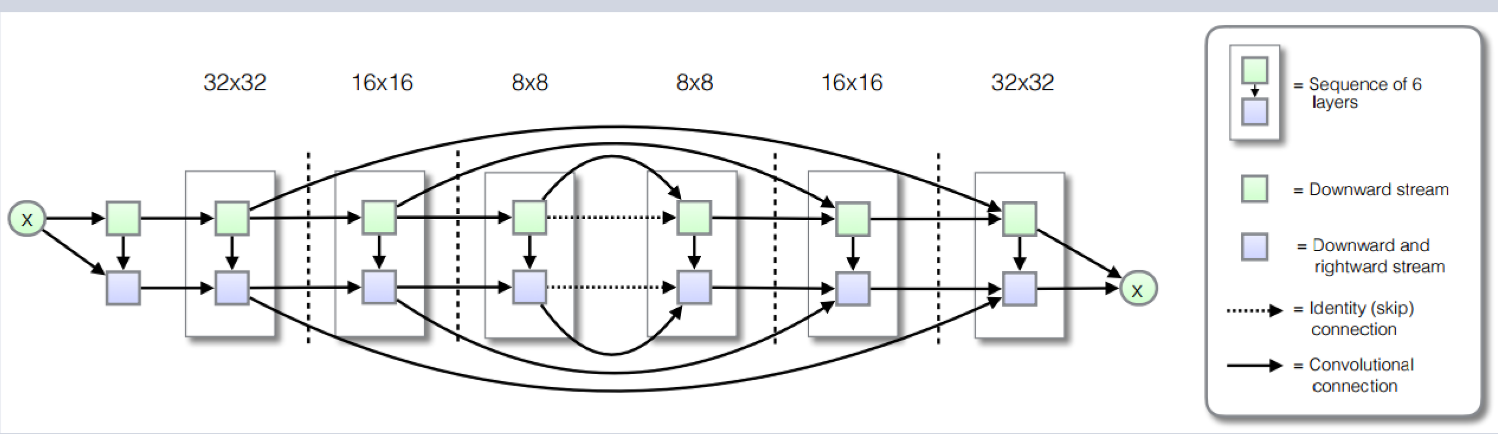
$$G^* = \arg\min_G \max_D \mathcal{L}_{cGAN}(G,D) + \lambda \mathcal{L}_{L1}(G).$$

- **Pixel CNN++**
- **Original Pixel CNN**: a generative model of images with tractable likelihoods.
  - The probabilty density function of an image x is factorized into all its pixels as $p(x) = \Pi_i p(x_i|x_{<i})$
  - Thus, the generation proceeds row by row and pixel by pixel.
  - Pixels as discrete variables: each channel variable takes one of 256 integer values and is modeled with a softmax layer.



Context

  - Using residual connections helps training this deep model with up to 12 layers.
- **Pixel CNN++**: simplify the structure and improve its performance.
  - Use mixture of logistic distribution to model the sub-pixel color intensity, and then round it to the nearest 8-bit representation.
  - Simplify the conditioning of sub-pixels.
  - Add short-cut connections.



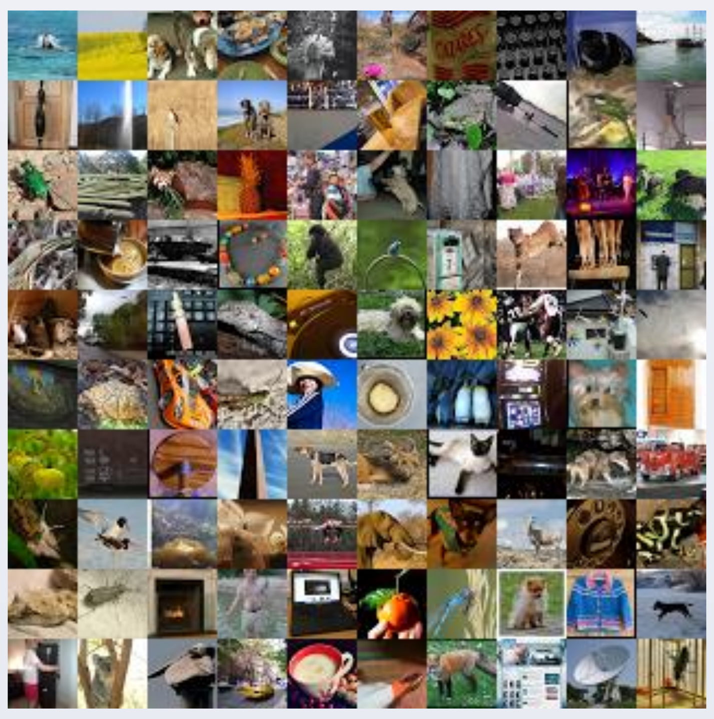
  - Add dropout as regularization to avoid overfitting.

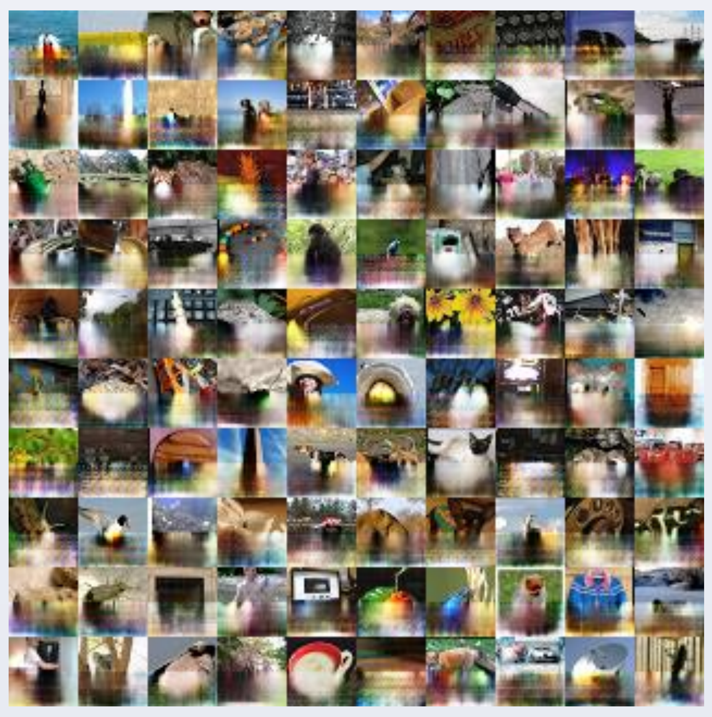
DATASET

- We use the 32x32 images from image-net
  - 1281149 images for training, 50000 images for testing

EXPERIMENTS

- **Pix2pix**
- **Pixel Discriminator + U-Net Generator**


PixelUnet_real


PixelUnet_fake

- **PatchGAN Discriminator + U-Net Generator**


PatchGanUnet_real


PatchGanUnet_fake

- **Pixel CNN++**


Test Data


Our Model

CONCLUSIONS

- We tried a conditional GAN model Pix2Pix to inpaint broken images. However, the generator only sees upper half of the image as condition, which is complex in dimensionality and sometimes insufficient. Therefore, we encounter some mode collapse problem, and the generated images are blurred.
- As for the Pixel CNN++ method, although the bpd loss of our model is higher compared to the pre-trained model, the generated images are perceptually reasonable. In general, it is hard to completely recover the lost part of the image, even using the state-of-the-art pre-trained model. Nevertheless, this generation method can capture the overall object and its shape, generating highly convincing image from broken one.

REFERENCES

- Phillip Isola, Jun-Yan Zhu, Tinghui Zhou, Alexei A. Efros (2016), Image-to-Image Translation with Conditional Adversarial Networks.
- Aaron van den Oord, Nal Kalchbrenner, Koray Kavukcuoglu (2016), Pixel Recurrent Neural Networks.
- Tim Salimans, Andrej Karpathy, Xi Chen, Diederik P. Kingma (2017), PixelCNN++: Improving the PixelCNN with Discretized Logistic Mixture Likelihood and Other Modifications.