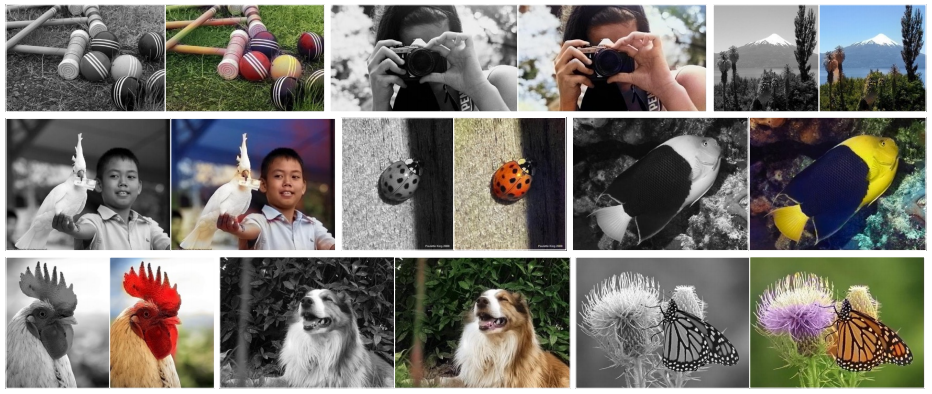
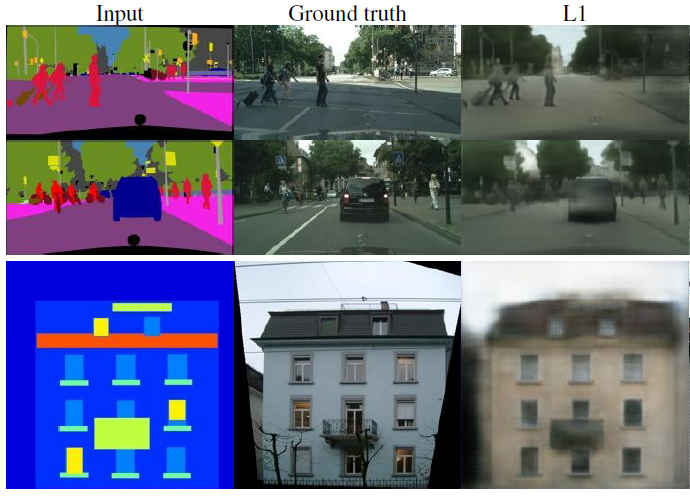
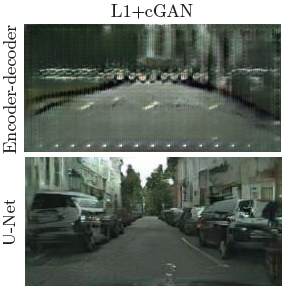
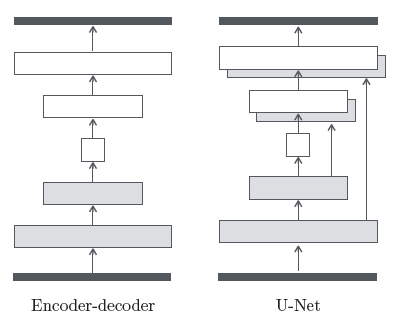
Image to Image Translation with Conditional Adversarial Networks

Jo MinKi

1. Introduction & Related works  
   This work was done by Berkeley’s CV lab, and previous work of these guys was about auto coloring. They did that work with using pair of grayscale image and full colored image. Therefore, the author thought that they could do image to image translation with those kind of paired dataset. The point and weakness of this work is that this model require a ‘paired data’. This feature leads us to the CycleGAN and DiscoGAN.  
   <https://phillipi.github.io/pix2pix/>
2. Related works.  
   1. Colorful Image Colorization (2016 ECCV)  
      
3. Image to Image translation with Neural Network   
     
     
     
     
     
     
     
     
     
     
     
     
     
     
     
     
     
     
     
     
     
     
   1. Naïve approach (L1 loss)  
      The Naïve approach (and also the same method with the previous work) is using CNN with L1 loss. You can see that it works somehow but not looks good. Reconstructed image seems like made by overlapping multiple objects. You can see this when you use every pixel’s value. This model use following loss.  
      This value come from the entire pixel set, so it does not predict the exact pixel value. Thus, image made by this way is not photo-realistic.
   2. GAN  
        
      The way for VAE, finding the data distribution by probabilistic approach is more elegant and sophisticate, but not working very well for the image processing. However, compare to the VAE, GAN focused to train discriminator network so it makes more discriminate decision which leads to the more photo-realistic result. That’s the main reason why GAN works better than VAE for the image generation task, such like image completion. Therefore, Pix2Pix used the GAN. (previous work used just CNN regression model)  
        
      To be more specific, the function that Pix2Pix learn is closer to the ‘Transfer’ function which map the Label image to the Synthetic Image, rather than the ‘Generation’ function (Mapping the random noise to the Target image). Following is the objective function of the Pix2Pix
   3. U-net, PatchGan  
        
        
        
        
        
        
        
        
        
        
        
        
        
        
        
        
        
        
        
        
        
        
        
        
        
        
        
        
      They used the U-net structure for the auto-encoder, which has skip connection between the Encoder and decoder. Because the U-Net can preserve more information from the layer before it compressed, it works better than original auto encoder. I don’t know why it does but there’s some research about that so we can check that.   
        
      I have no idea what the PatchGAN is, but this paper says that this model uses a small unit patch and the loss go through this patches at the backpropagation. Which means, instead of the normal GAN that just use entire image at the discriminator, PatchGAN can get more specific feedback about each patches so it can catch more detail part of the image.

*×*

*×*

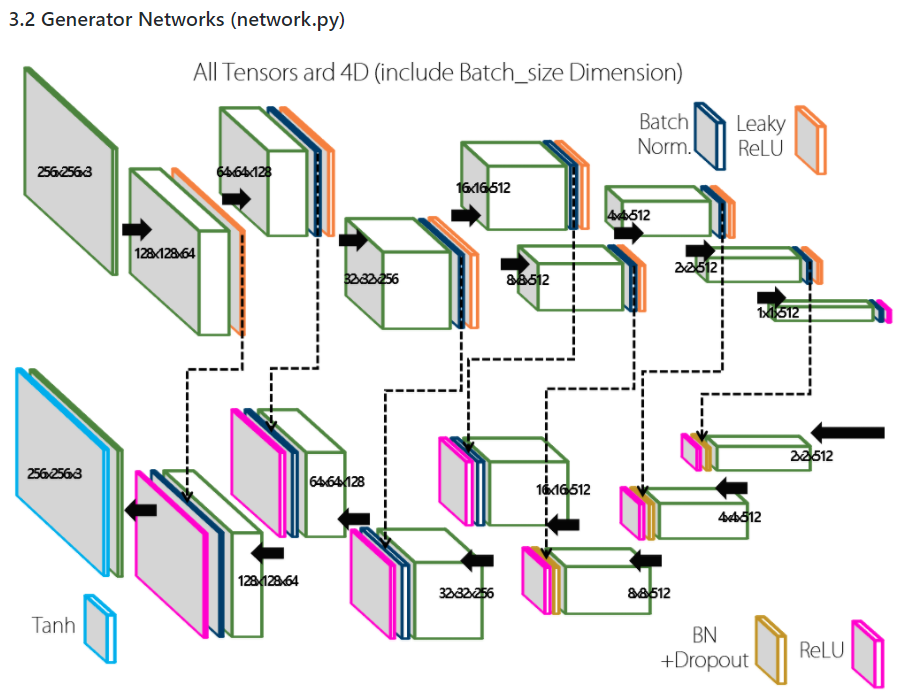
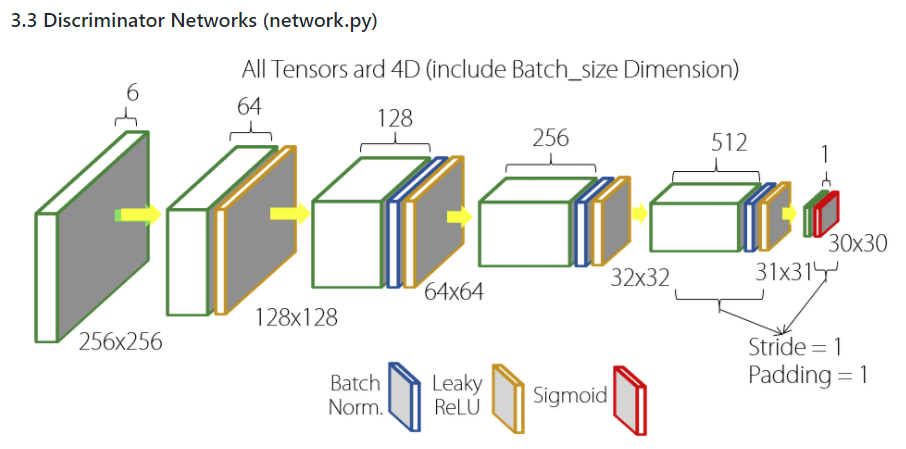
*×*

*×*



Figure 1 From left to right, 1x1 16x16 70x70 256x256

When the patch is small, model can generate more detail result, but the computation cost also gets higher. This paper says that 70x70 is enough.

* 1. Architecture  
       
        
       
     

1. Experiments.   
     
   I don’t know what to say.

Question.

1. Do we know the ? If we don’t, where are we heading for?  
   1. No. we are not knowing the exact function that we should head for. In practically, we are just adding a skip connection on the network so we get the ­­ as a result. So don’t take the explanations on the 3.1 too much seriously.