Generative Adversarial Nets

Priyanka Mehta
Sudhanshu Srivastava

Outline

What is a GAN?

How does GAN work?

Newer Architectures

Applications of GAN

Future possible applications

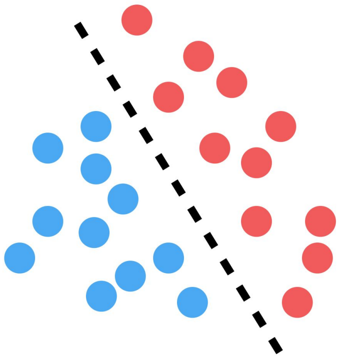
Generative Adversarial Networks

What: a network that uses two kinds of models working against each other to solve a problem

Why: It can solve complex problems while requiring much less input data, with reduced computational complexity

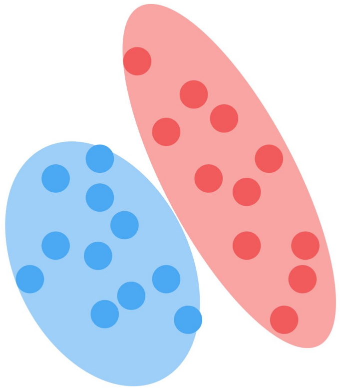
Background: Discriminative vs Generative Models

Discriminative



- Categorize samples
- Widely used
- Backpropogation & dropout
- Piecewise linear units with "well-behaved gradients"

Generative



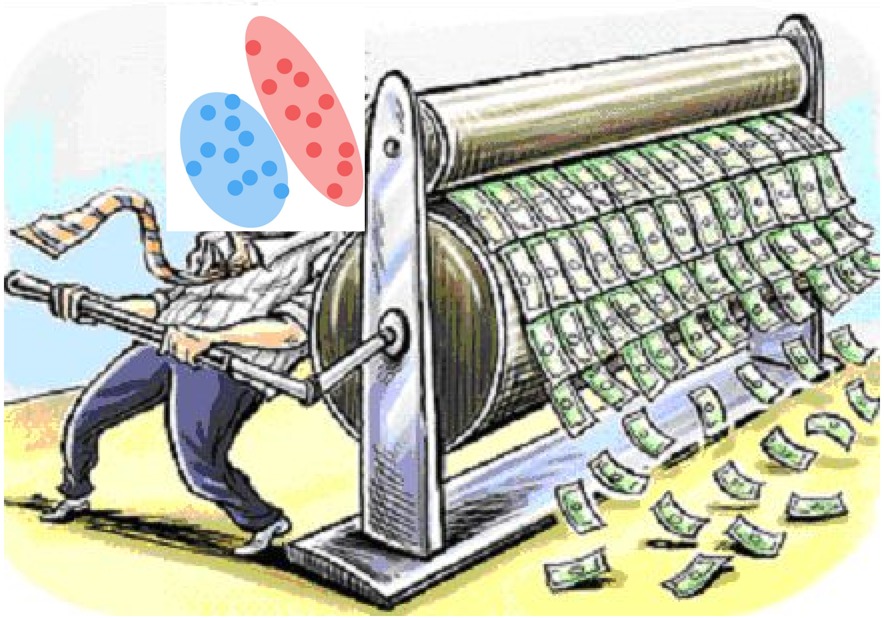
- Model the distribution that produces the samples
- Less widely used
- "Intractable probabilistic computations"
- Hard to use piecewise linear units

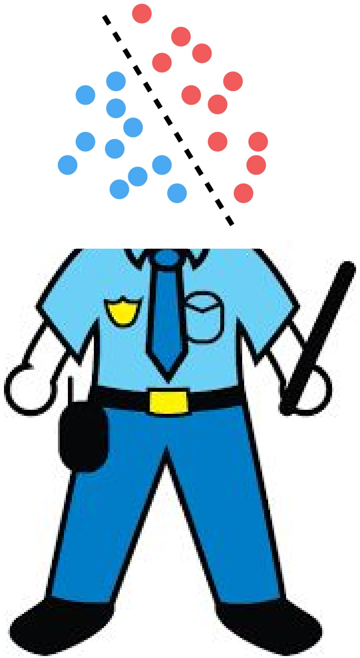
Image source: http://joelouismarino.github.io/blog_posts/blog_VAE.html

Adversarial Nets

Generative model (G) competes with discriminative model (D):
G tries to generate data indistinguishable from training data
D tries to distinguish them

Adversarial Nets





Adversarial Nets Implementation details

- Two Networks, Generator(G), Discriminator(D).
- G takes random noise Z, generates output G(Z).
- Aim: G(Z) must look like samples from the distribution we are interested in.

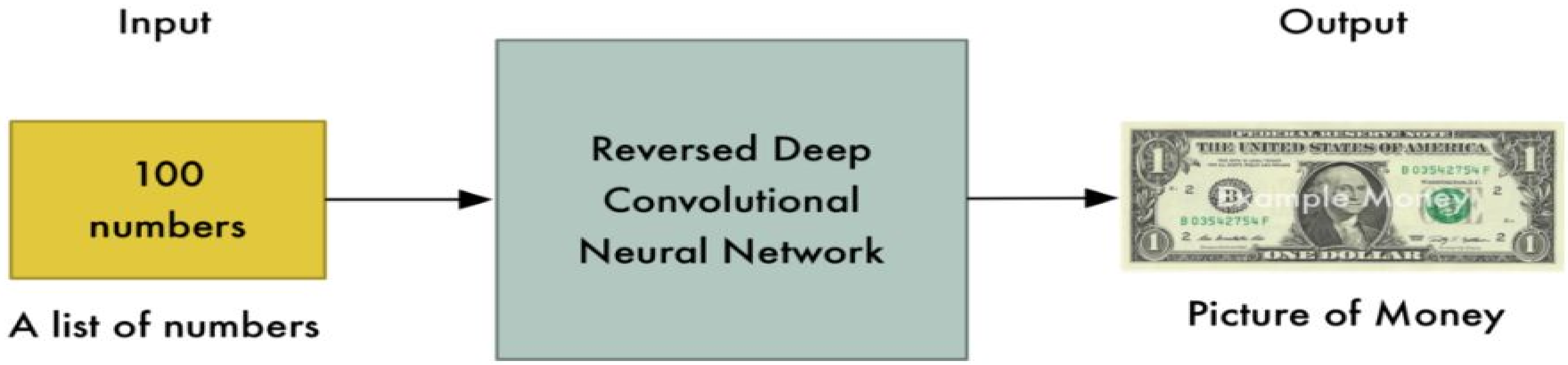


Image courtesy: **Adam Geitgey, *Machine Learning is Fun***

GAN Implementation details

- For an input X, D produces the probability that X is a sample from the distribution.

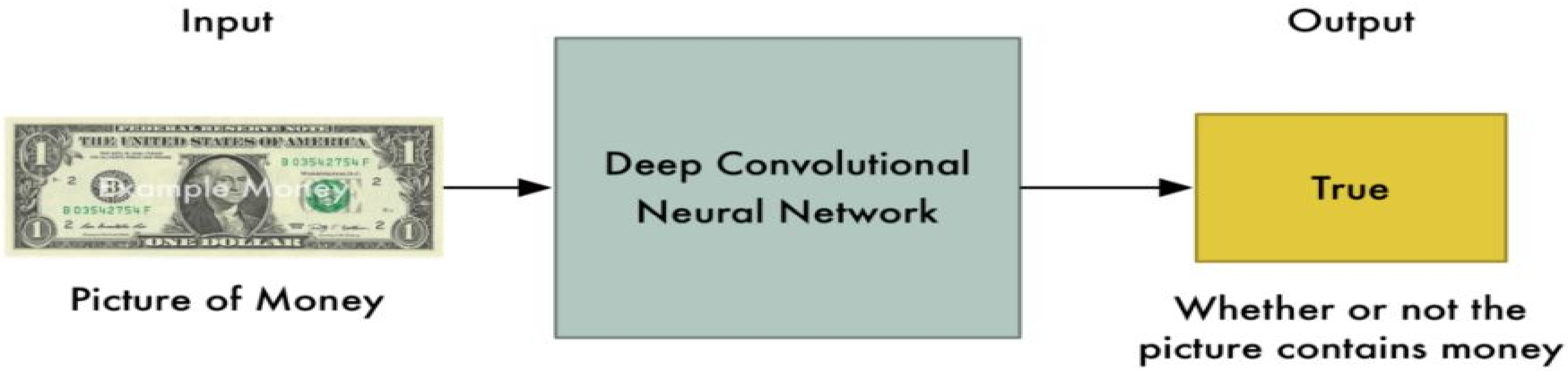


Image courtesy: Adam Geitgey, *Machine Learning is Fun*

GAN implementation details

- $D(X)$ should be closer to 1 if X is a sample from the distribution, and closer to 0 if X is generated by G.
- In each iteration, take a genuine X, and a generated sample, $G(Z)$.
- Aim: Maximize: $\log(D(X)) + \log(1-D(G(Z)))$
- Achieved by Gradient Ascent- calculate gradient of the above function, ascend along the Gradient (update weights of D according to the gradient).

GAN Implementation Details

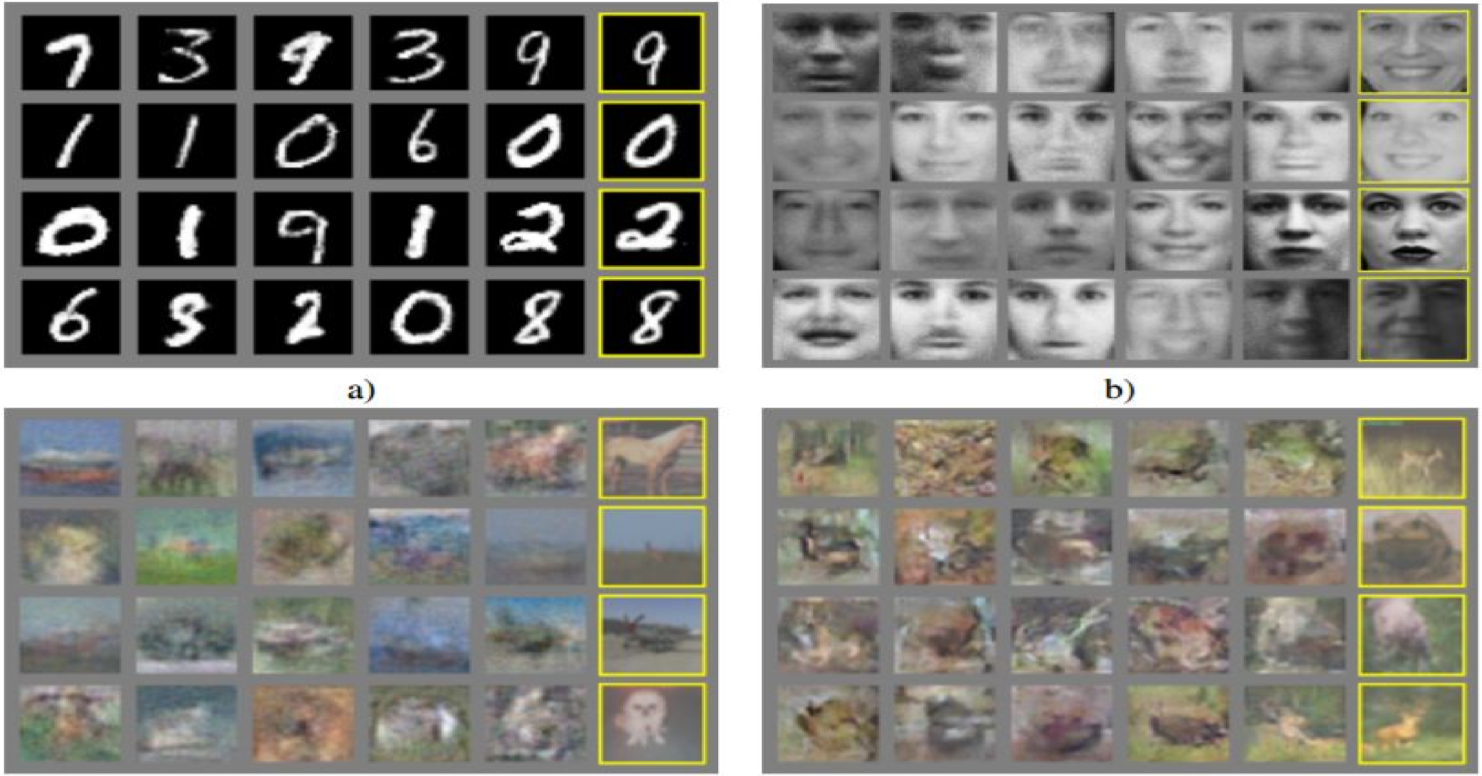
- G should be able to fool D, so D(G(Z)) should be close to 1.
- Aim: Minimize log(1-D(G(Z))).
- Achieved by Gradient Descent.

In Game Theory, such a scenario is called a **Min-Max Game**.

Summary of the Math

With enough time and sufficiently complicated D and G, the distribution generated by G converges to the distribution of interest.

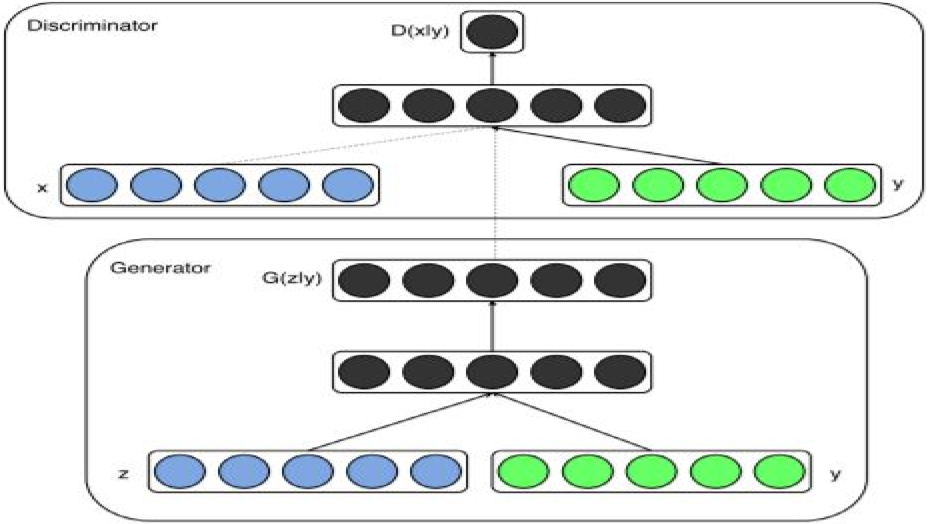
GAN paper results



a) b) c) d)

Application 1: Conditional GAN (Mirza et al., 2014)

Natural Question- I don't want a replica of the entire MNIST dataset, I just want the number 9.



X and Z are the same as previous discussion, Y is the subset of interest.

CGAN (contd.)

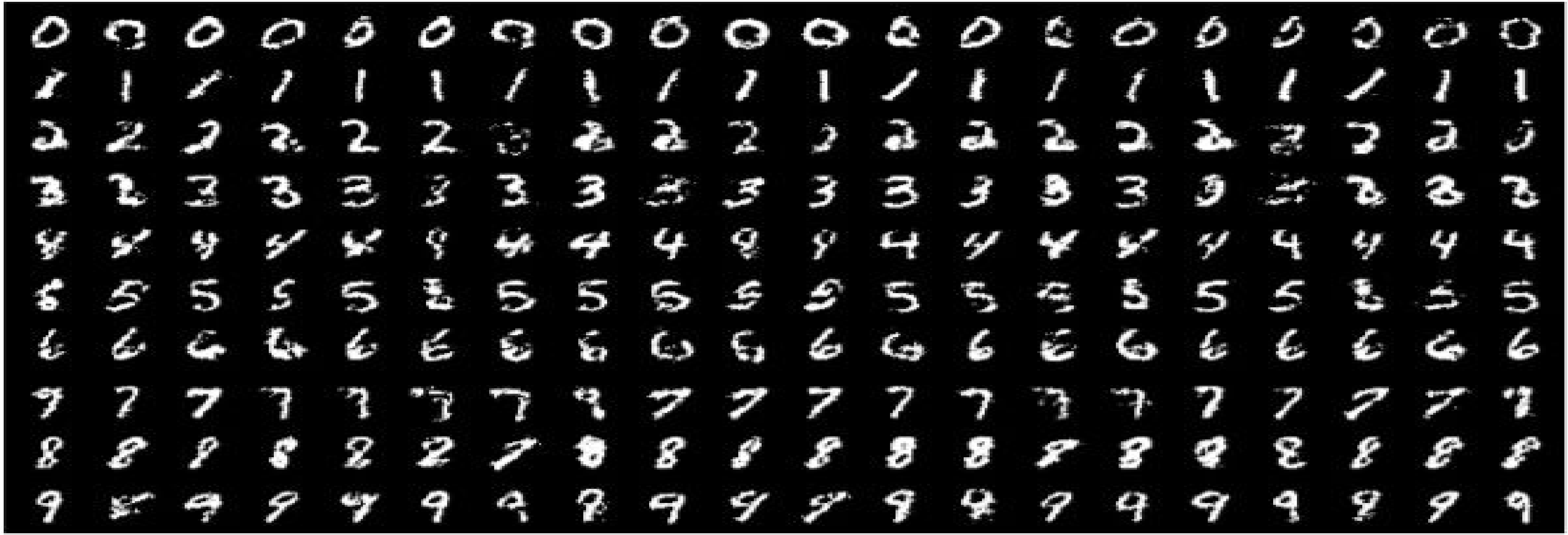


Figure 2: Generated MNIST digits, each row conditioned on one label

Application 2: Image from Captions (Reed et al.,2016)

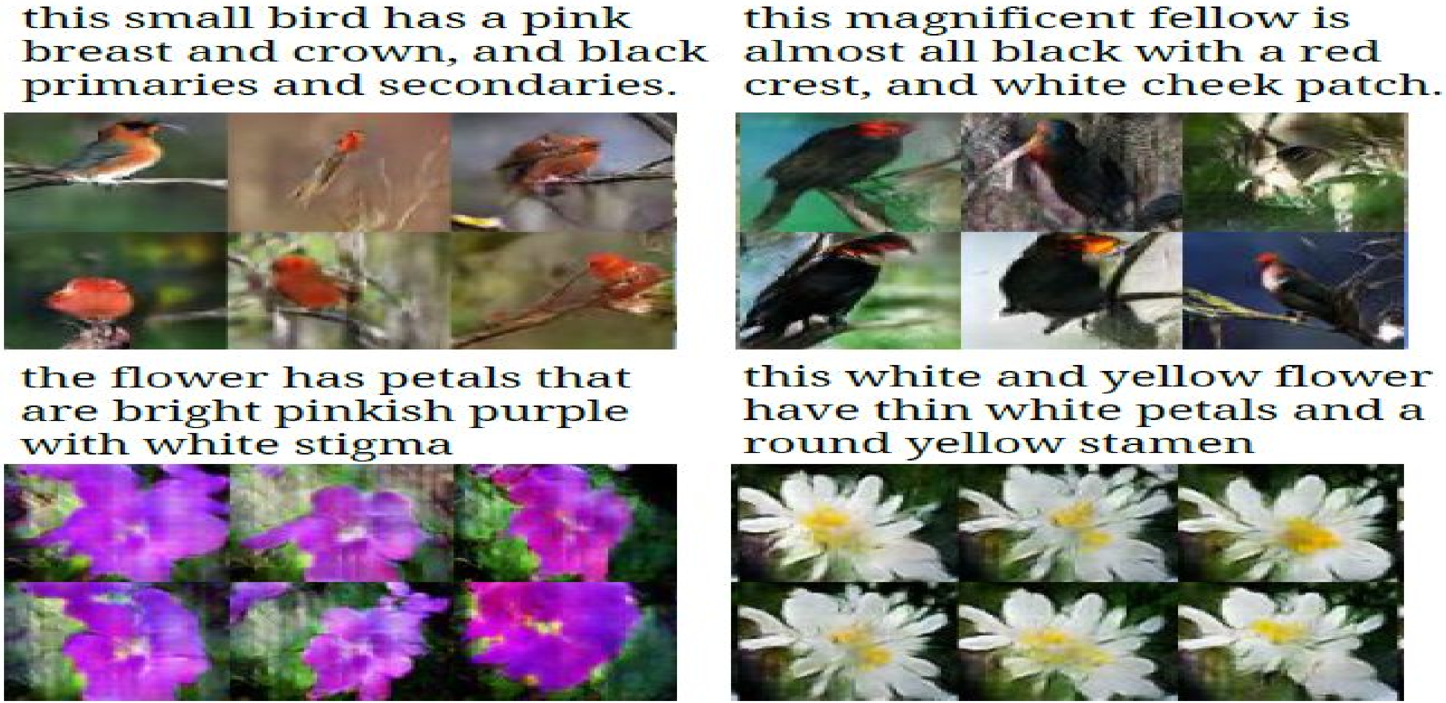
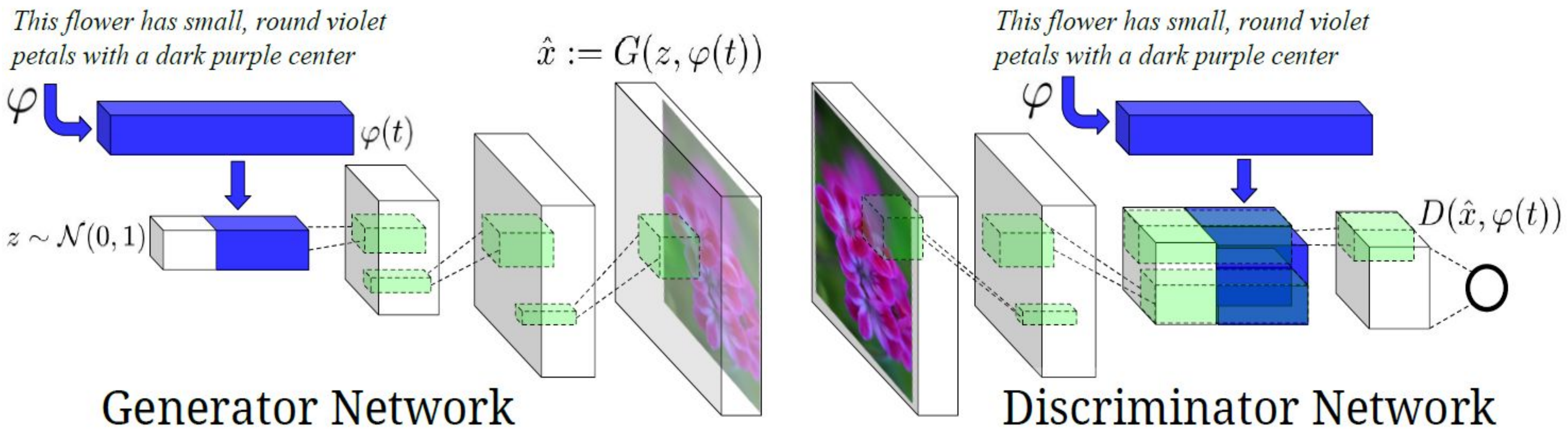


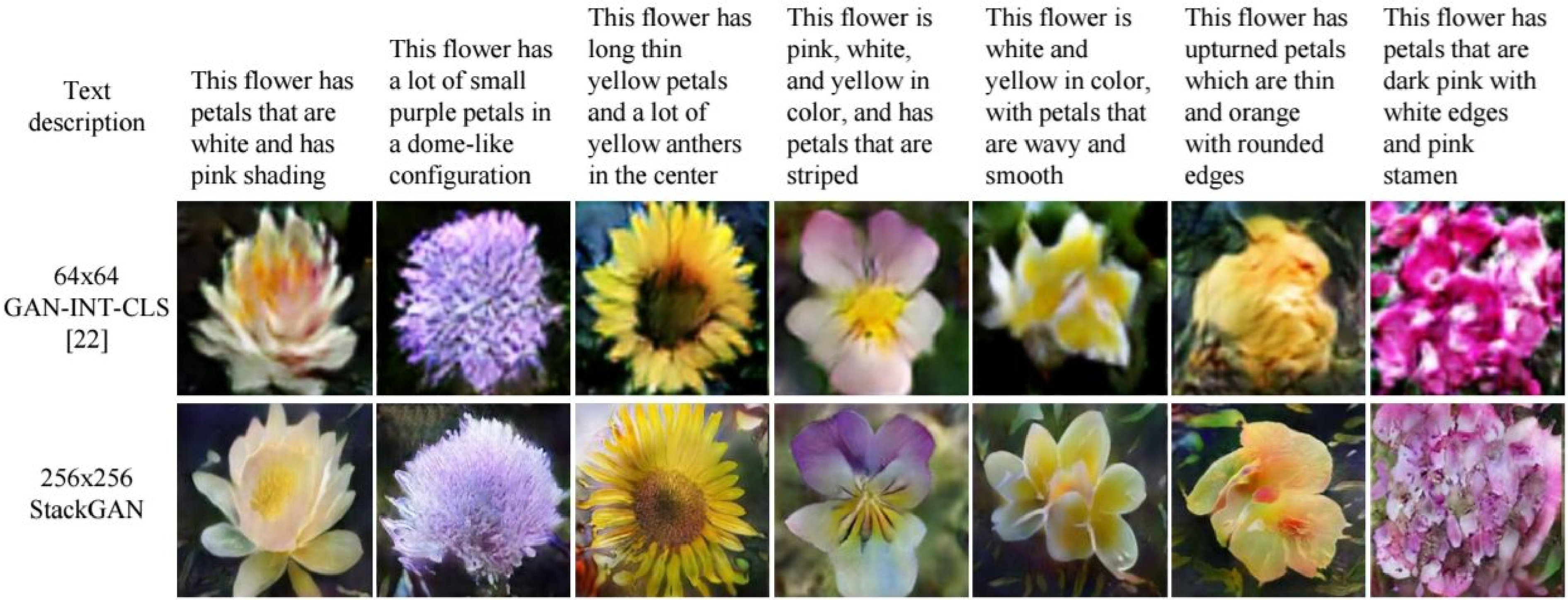
Image from Captions (contd.)



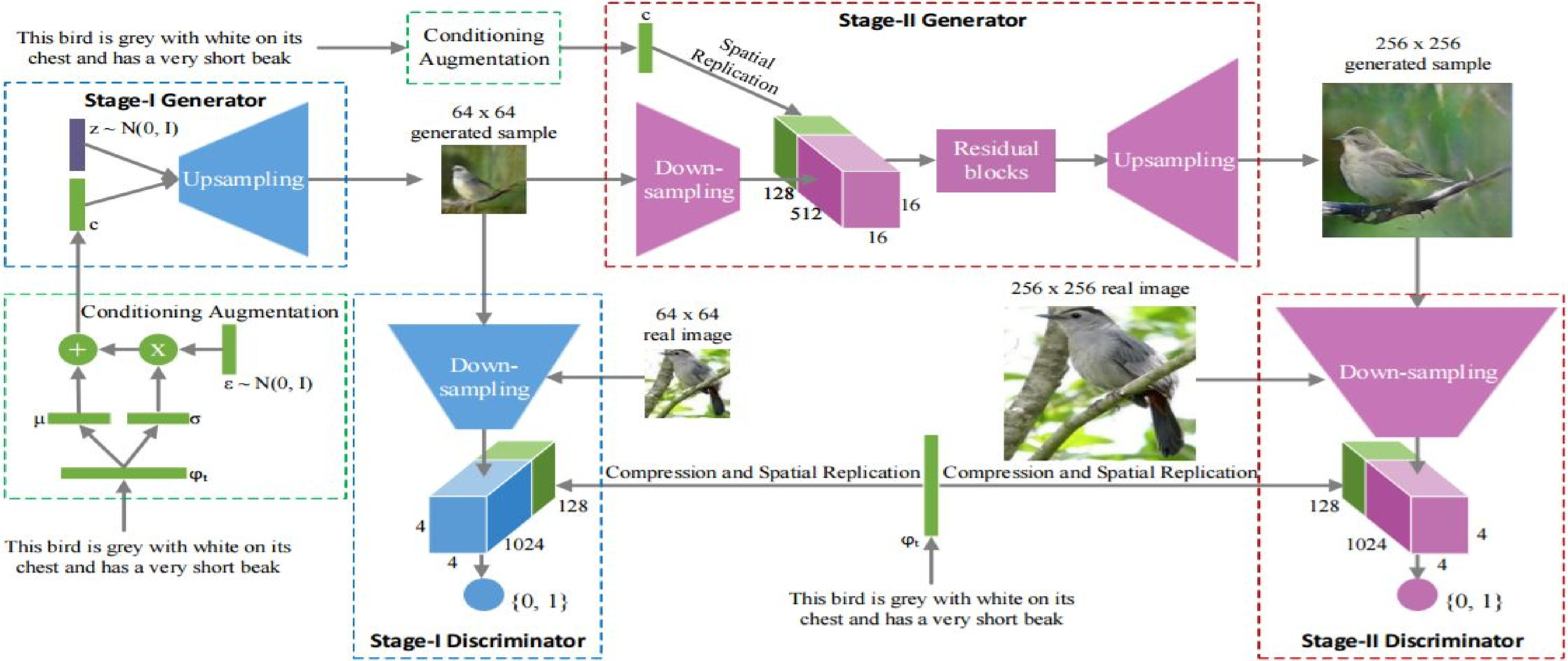
Phi is an autoencoder, which gives a vector of size 128. Z is a vector of size 100.

Image courtesy Reed et al.

StackGAN:Hyper-Realistic Images from Captions(Zhang et al, 2016)



StackGAN(contd.)



Application 3: “Style” Detection

Can we use a GAN to capture categorically and continuously varying details about a visual image?

Like the way digits are written, or the way hair is styled!

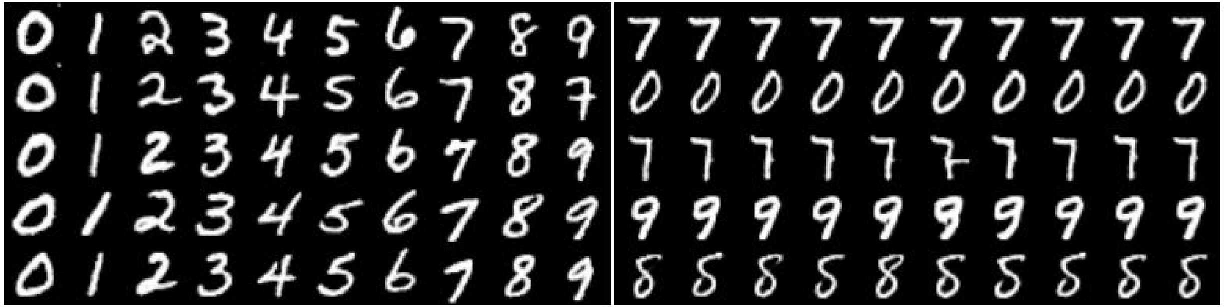
InfoGAN: Interpretable Representation Learning by Information Maximizing Generative Adversarial Nets

Chen et al., 2016

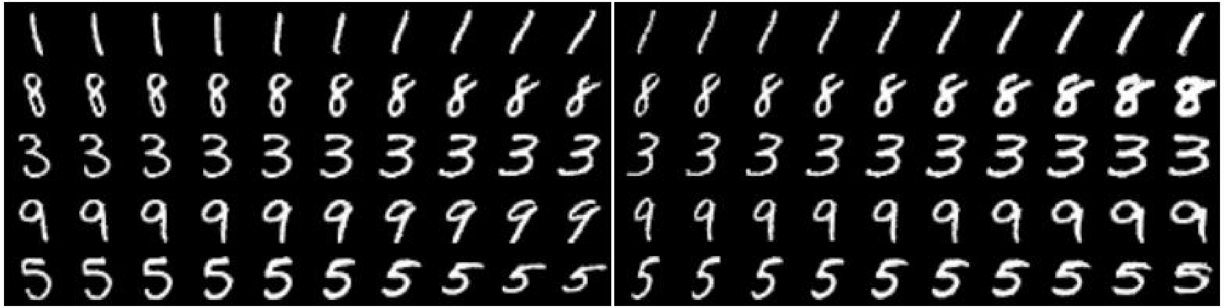
Slightly modified original GAN--instead of just noise for the generative model, adds in “latent code” that encodes salient features

(ie if we know the data is written numbers, we can make sure the “random noise” contains a variable for “digit”, a variable for “stroke angle”, and a variable for “stroke thickness”)

Goal is to learn the latent variables governing the samples in a dataset



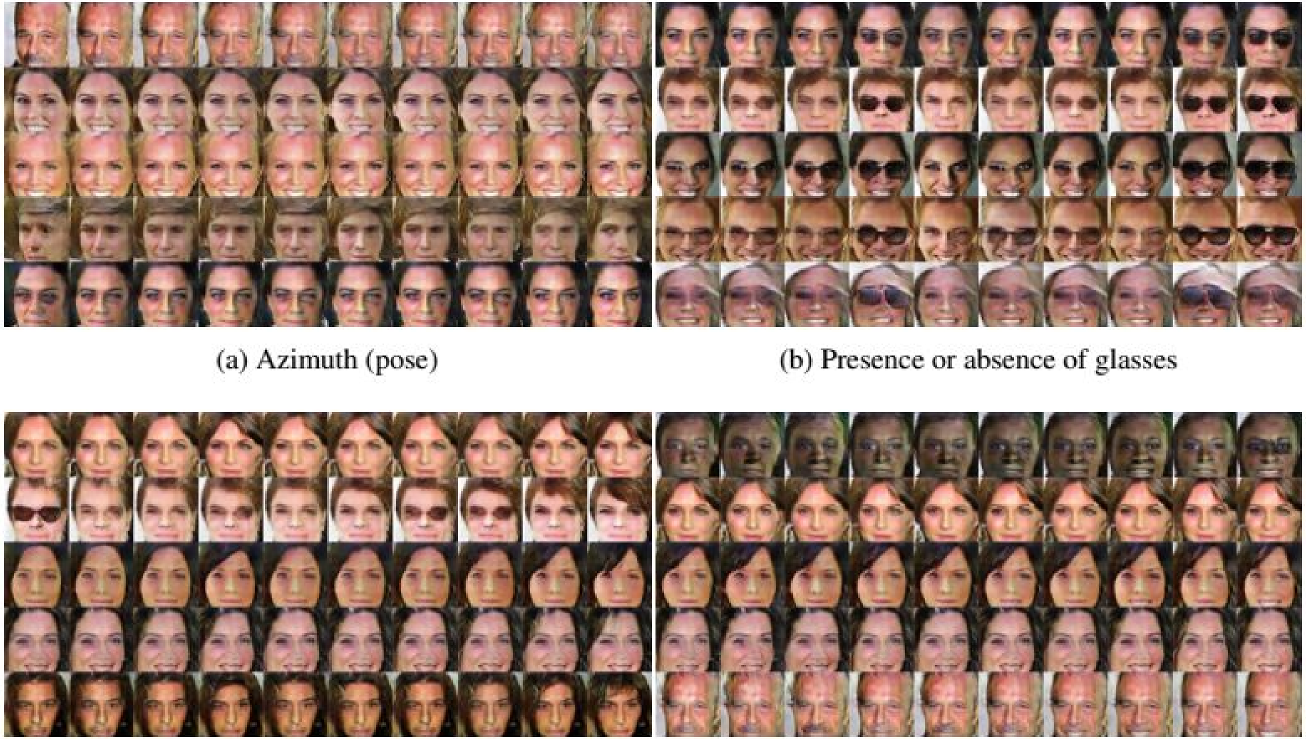
(a) Varying c_1 on InfoGAN (Digit type)   (b) Varying c_1 on regular GAN (No clear meaning)



(c) Varying c_2 from -2 to 2 on InfoGAN (Rotation)   (d) Varying c_3 from -2 to 2 on InfoGAN (Width)

Figure 2: **Manipulating latent codes on MNIST:** *In all figures of latent code manipulation, we will use the convention that in each one latent code varies from left to right while the other latent codes and noise are fixed. The different rows correspond to different random samples of fixed latent codes and noise. For instance, in (a), one column contains five samples from the same category in c_1, and a row shows the generated images for* 10 *possible categories in c_1 with other noise fixed.* In (a), each category in c_1 largely corresponds to one digit type; in (b), varying c_1 on a GAN trained without information regularization results in non-interpretable variations; in (c), a small value of c_2 denotes left leaning digit whereas a high value corresponds to right leaning digit; in (d), c_3 smoothly controls the width. We reorder (a) for visualization purpose, as the categorical code is inherently unordered.

Chen et al., 2016



(a) Azimuth (pose) (b) Presence or absence of glasses

(c) Hair style (d) Emotion

Figure 6: **Manipulating latent codes on CelebA:** (a) shows that a categorical code can capture the azimuth of face by discretizing this variation of continuous nature; in (b) a subset of the categorical code is devoted to signal the presence of glasses; (c) shows variation in hair style, roughly ordered from less hair to more hair; (d) shows change in emotion, roughly ordered from stern to happy.

Chen et al., 2016

Discussion Question

Applications for this application? What could this generative “style detection” be useful for?

Application: Medical Imaging!

Problems for medical image neural nets:

-old methods: go pixel-by-pixel -> information loss

-newer methods: semantic segmentation, but it needs LOTS of data to train -> medical imaging does not have that luxury

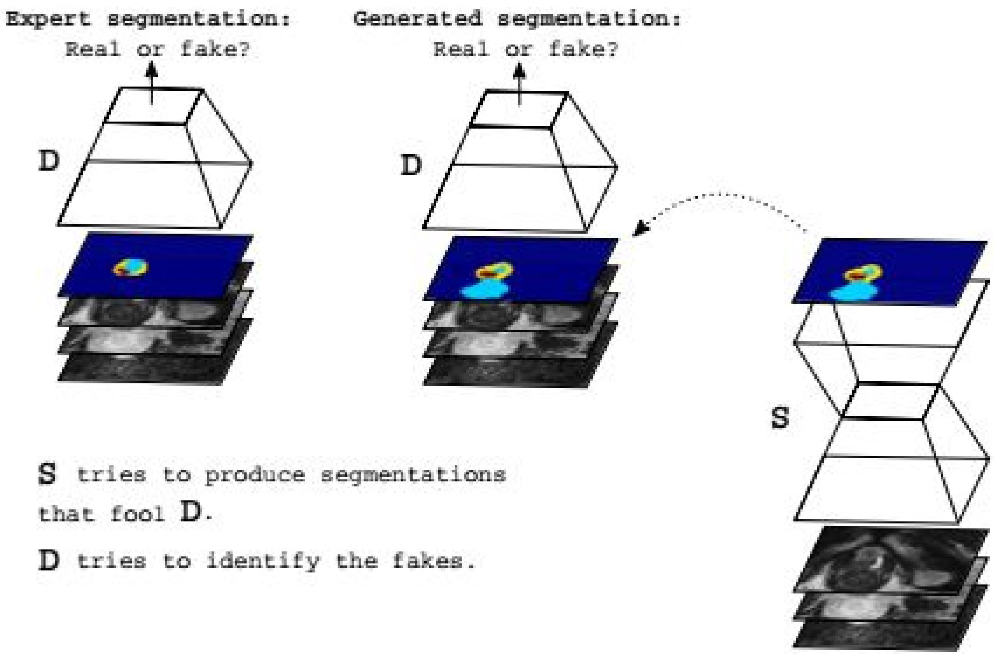
Adversarial Networks for the Detection of Aggressive Prostate Cancer Kohl et al,., 2017

Solution: Adversarial nets! (Just like the initial paper)

This paper -> $$\min_G \max_D \left(\mathbb{E}_{\boldsymbol{x}\sim p_{\mathrm{data}}(\boldsymbol{x})} [\log D(\boldsymbol{x})] + \mathbb{E}_{\boldsymbol{z}\sim p_{\boldsymbol{z}}(\boldsymbol{z})} [\log (1 - D(G(\boldsymbol{z})))] \right)$$

Original paper -> $$\min_G \max_D V(D,G) = \mathbb{E}_{\boldsymbol{x}\sim p_{\mathrm{data}}(\boldsymbol{x})}[\log D(\boldsymbol{x})] + \mathbb{E}_{\boldsymbol{z}\sim p_{\boldsymbol{z}}(\boldsymbol{z})}[\log(1 - D(G(\boldsymbol{z})))].$$

Adversarial Networks for the Detection of Aggressive Prostate Cancer Kohl et al,., 2017



Adversarial Networks for the Detection of Aggressive Prostate Cancer Kohl et al,., 2017

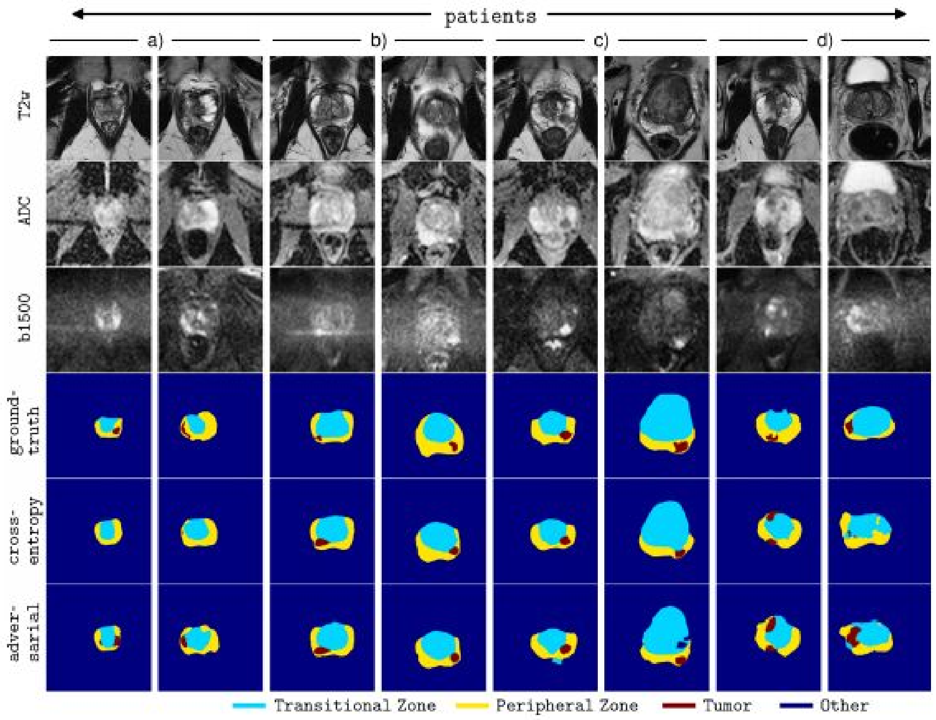
MRI dataset: 152 patients with suspicious results

Trained on 188 tumor-diagnosed images and 475 tumor-free-diagnosed images

Table 1: Experimental results of the four-fold cross-validation for GS $\geq$ 7 Tumor.

training scheme loss	cross-entropy $\mathcal{L}_{mce}$	adversarial $\mathcal{L}_S$ & $\mathcal{L}_D$	hybrid $\mathcal{L}_{mce}/2 + \mathcal{L}_S$ & $\mathcal{L}_D$
tumor DSC	0.35 ± 0.29	$\mathbf{0.41 \pm 0.28}$	0.39 ± 0.29
tumor sensitivity	0.37 ± 0.33	$\mathbf{0.55 \pm 0.36}$	0.49 ± 0.35
tumor specificity	0.98 ± 0.14	0.98 ± 0.14	0.98 ± 0.14

Adversarial Networks for the Detection of Aggressive Prostate Cancer Kohl et al,., 2017



Discussion Question

Medical use of adversarial networks: is it meaningless unless it can be as good as human doctors/almost no chance of dangerous error? Or is it useful the way it is now?

Other cool existing applications

- **Generating Video with Scene Dynamics:** http://web.mit.edu/vondrick/tinyvideo/
- **Draw a face from text description**- could help in generating a sketch of a criminal.(Attribute2Image: Conditional Image Generation from Visual Attributes , Yan et al.)
- **Remove snow/rain from an image**- could help Self-Driving Cars by clearing their field of view when the weather is bad. (Image De-raining Using a Conditional Generative Adversarial Network, Zhang et al.)
- **Effects of aging on a given face -** Face Aging With Conditional Generative Adversarial Networks, Antipov et al.

Other cool existing applications

- Music Generation: C-RNN-GAN, Mogren et al
- Interactive Image Generation : https://www.youtube.com/watch?v=9c4z6YsBGQ0&feature=youtu.be

Other potential applications

- Hyper-realistic Virtual Conversational Agent using inverse Lip-reading. Can be made to look like any person you want. Will also be useful in the animation industry.
- Translating other languages to ASL.
- Once the training is complete, the Discriminator can be used to detect fake/photoshopped images.
- Demand-based art.
- ‘Virtually’ trying on clothes on e-commerce websites.

Discussion

How does everyone feel about GANs?

Can you think of other uses for them?

Sources

Chen, X., Duan, Y., Houthooft, R., Schulman, J., Sutskever, I., & Abbeel, P. (2016). Infogan: Interpretable representation learning by information maximizing generative adversarial nets. In *Advances in Neural Information Processing Systems* (pp. 2172-2180).

Kohl, S., Bonekamp, D., Schlemmer, H. P., Yaqubi, K., Hohenfellner, M., Hadaschik, B., ... & Maier-Hein, K. (2017). Adversarial Networks for the Detection of Aggressive Prostate Cancer. *arXiv preprint arXiv:1702.08014*.