April 10, 2017

Introduction to

Generative Adversarial Networks

- Simon Kohl, KIT & DKFZ

Karlsruhe Machine Learning, Statistics and AI Meetup

Is human creativity untouchable?

By:
HikingArtist.com

- What is creativity?
- The ability to plausibly rearrange/abstract higher-order concepts (shapes, ideas, ...)?
- Is everything humans can think up a sample from a learned distribution over previous experience?
- Can we emulate creativity with models that learn plausible distributions?

Is human creativity untouchable?

By:
HikingArtist.com

- What is creativity?
- The ability to plausibly rearrange/abstract higher-order concepts (shapes, ideas, ...)?
- Is everything humans can think up a sample from a learned distribution over previous experience?
- Can we emulate creativity with models that learn plausible distributions?

Two samples from the distribution over possible Christiano Ronaldos:

Generative Models

Goal: unsupervised learning of a distribution $\rightarrow$ sample from it

Deep Approaches [1]:

1. **model-based**: explicitly learning a distribution
   Variational Autoencoder, Boltzmann Machines, ..
2. **model-free**: implicitly learning a distribution
   Generative Stochastic Networks (Markov Chain)

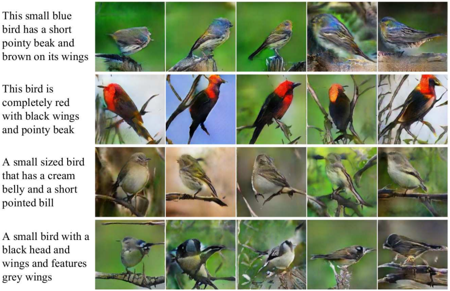
$\rightarrow$ most prominent *disadvantage*:

failure to represent distributions over high-dimensional spaces at computationally acceptable cost.

New Scheme [2]:

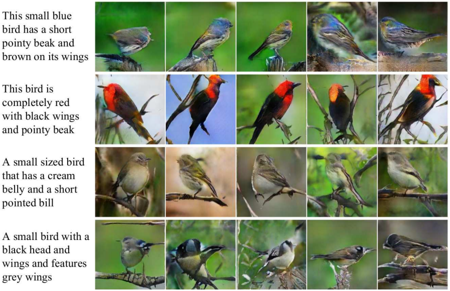
Implicitly learning a distribution by back-propagating through a learned higher-order discriminator-function $\rightarrow$ *Generative Adversarial Networks (GANs)*

GAN examples



Zhang *et al.*, 2016 [3]

GAN examples

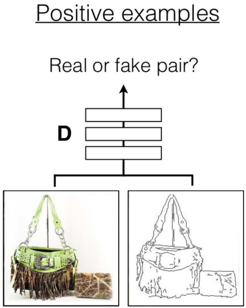


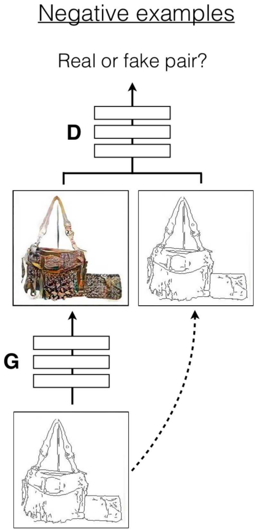
Zhang *et al.*, 2016 [3]

Isola *et al.*, 2016 [4]

GAN Framework

conditional GAN





G tries to synthesize fake images that fool **D**

D tries to identify the fakes

adapted from [4]

GAN Framework

- **Goal**: $p_g \rightarrow p_{target}$
- GAN: **two deep nets**, that compete/collaborate:

  1. the generator
     - **i)** $G = G(\theta_G; \boldsymbol{z}), \quad \boldsymbol{z} \sim p_{\boldsymbol{z}},$ (unconditional)
     - **ii)** $G = G(\theta_G; \boldsymbol{z}, \boldsymbol{x}), \boldsymbol{z} \sim p_{\boldsymbol{z}}$ & $\boldsymbol{x} \sim p_{\boldsymbol{source}}$ (conditional)
  2. the discriminator $D = D(\theta_D; \boldsymbol{y}), \boldsymbol{y} \sim p_{\boldsymbol{target}}$ *or* $\boldsymbol{y} \sim p_{\boldsymbol{g}}$

- Training is a MiniMax Game with value-function:

$$\min_{\boldsymbol{G}} \max_{\boldsymbol{D}} \left(\mathbb{E}_{\boldsymbol{y} \sim p_{\text{target}}} [\log D(\boldsymbol{y})] + \mathbb{E}_{\boldsymbol{z} \sim p_{\boldsymbol{z}}} [\log (1 - D(G(\boldsymbol{z})))] \right) \tag{1}$$

  1. train D: minimize

$$\mathcal{L}_D(\boldsymbol{\theta}_D, \boldsymbol{\theta}_G) = -\frac{1}{N} \sum_i^N \left(\log D(\underbrace{\boldsymbol{y}_i}_{real}) + \log (1 - D(\underbrace{G(\boldsymbol{x}_i, \boldsymbol{z}_i)}_{fake})) \right) \tag{2}$$

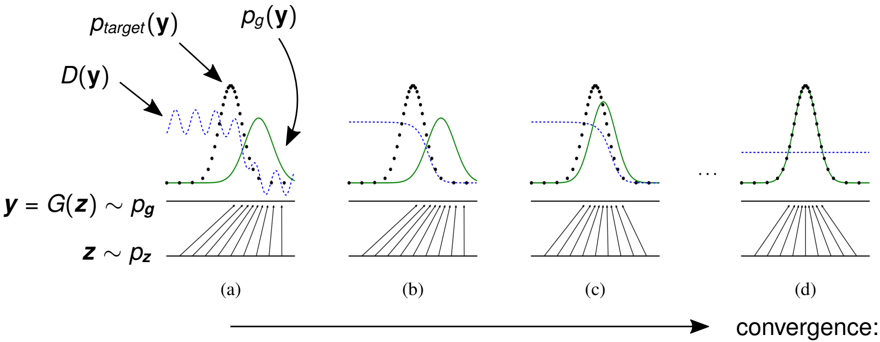
  2. train G: minimize

$$\mathcal{L}_G = -\mathcal{L}_D \tag{3}$$

GAN Framework

- **Properties**:
  1. globally optimal Discriminator (for given G):

$$D^*(\mathbf{y}) = \frac{p_{target}(\mathbf{y})}{p_{target}(\mathbf{y}) + p_g(\mathbf{y})} \tag{4}$$



convergence:
$D(\mathbf{y}) = 1/2$

adapted from [2]

GAN Framework

- **Properties**:

  1. globally optimal Discriminator: ...
  2. stability:

     i) $\mathcal{L}_G = \underbrace{-log\, D(G(\boldsymbol{x},\boldsymbol{z}))}_{stronger\ gradients\ [2]}\ or\ \underbrace{\|f(\boldsymbol{y}) - f(G(\boldsymbol{x},\boldsymbol{z}))\|_2^2}_{featuremap-matching\ [5]}$

     ii) keep D optimal, i.e one may train it more often than G

  3. choice of divergence to minimize:

     i) Eq. 1 equivalent to minimizing Jensen-Shannon divergence

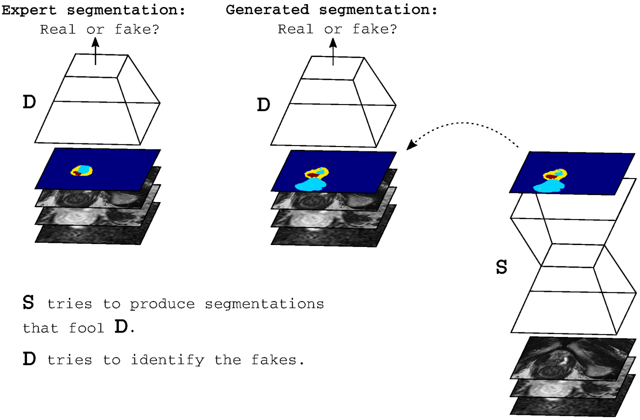
     ii) Alternative: Use Earth-Mover distance (= Wasserstein-1) [6]

$$\min_{G}\max_{D}\left(\mathbb{E}_{\boldsymbol{y}\sim p_{\text{target}}}[D(\boldsymbol{y})] - \mathbb{E}_{\boldsymbol{z}\sim p_{\boldsymbol{z}}}[D(G(\boldsymbol{z}))]\right) \tag{5}$$

     iii) Very recent further improvement:
     penalize gradient norms of D vs. weight-clipping of D [7]

Adversarial Networks ..

.. for the Detection of Aggressive Prostate Cancer [8]

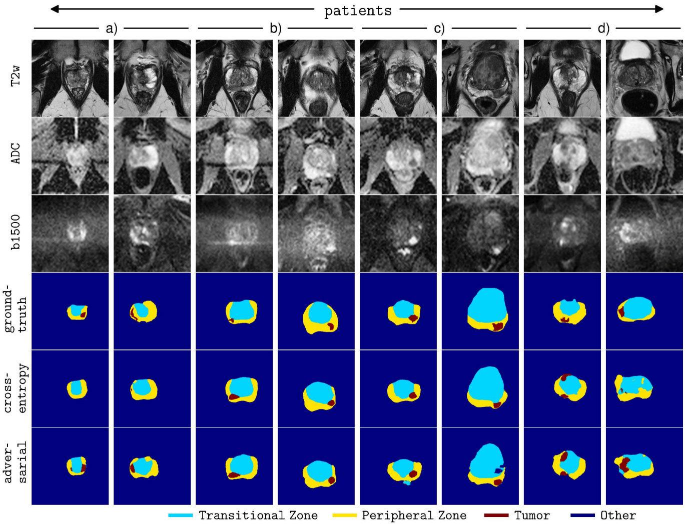


S tries to produce segmentations that fool **D**.

D tries to identify the fakes.

Adversarial Networks ..

.. for the Detection of Aggressive Prostate Cancer [8]



Conclusion

- GANs are very versatile:

  i) text-to-image, domain-adaptation, image super-resolution, replace Monte-Carlo simulations (physics), generate music, ...

  ii) ‘non-generative’ tasks:
  segmentation (special case of image-to-image)

- current research: make training and modeling more robust

- interesting concept from a learning-theory point of view:
  teacher/student- or adversarial set-up

- Are we witnessing another ‘insult to humanity’?

References I

[1] Ian Goodfellow.
Nips 2016 Tutorial: Generative Adversarial Networks.
arXiv preprint arXiv:1701.00160, 2016.

[2] Ian Goodfellow *et al.*
Generative Adversarial Nets.
In *NIPS*, pages 2672–2680, 2014.

[3] Han Zhang *et al.*
StackGAN: Text to Photo-realistic Image Synthesis with Stacked Generative Adversarial Networks.
arXiv preprint arXiv:1612.03242, 2016.

[4] Phillip Isola *et al.*
Image-to-Image Translation with Conditional Adversarial Networks.
arXiv preprint arXiv:1611.07004, 2016.

[5] Tim Salimans *et al.*
Improved Techniques for Training GANs.
In *Advances in Neural Information Processing Systems*, pages 2226–2234, 2016.

[6] Martin Arjovsky *et al.*
Wasserstein GAN.
arXiv preprint arXiv:1701.07875, 2017.

References II

[7] Ishaan Gulrajani, Faruk Ahmed, Martin Arjovsky, Vincent Dumoulin, and Aaron Courville.
Improved training of wasserstein gans.
arXiv preprint arXiv:1704.00028, 2017.

[8] Simon Kohl, David Bonekamp, Heinz-Peter Schlemmer, Kaneschka Yaqubi, Markus Hohenfellner, Boris Hadaschik, Jan-Philipp Radtke, and Klaus Maier-Hein.
Adversarial networks for the detection of aggressive prostate cancer.
arXiv preprint arXiv:1702.08014, 2017.

Paper Results

'Adversarial Networks for the Detection of Aggressive Prostate Cancer' [8]

Table: Experimental results of the four-fold cross-validation for GS ≥ 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	**0.41 ± 0.28**	0.39 ± 0.29
tumor sensitivity	0.37 ± 0.33	**0.55 ± 0.36**	0.49 ± 0.35
tumor specificity	0.98 ± 0.14	0.98 ± 0.14	0.98 ± 0.14