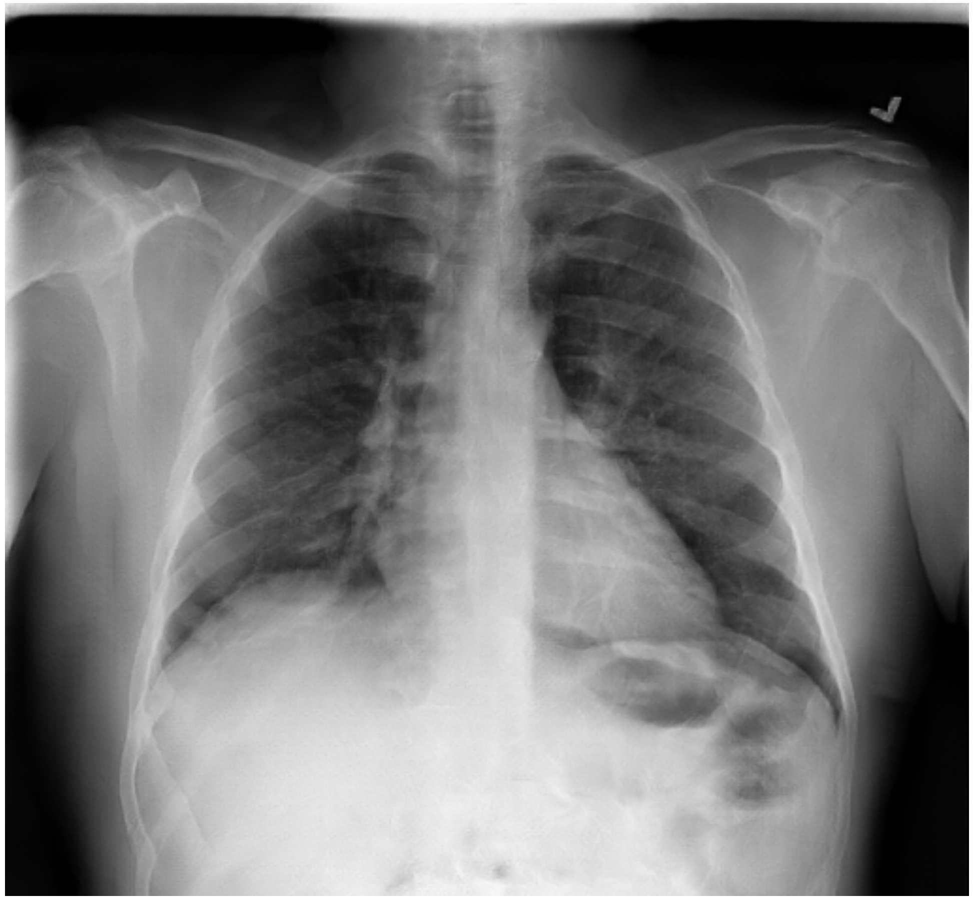
Synthesizing fMRI with Generative Adversarial Networks: Applications, Promises and Pitfalls

SANMI KOYEJO & BLISS CHAPMAN
CS @ ILLINOIS

Synthetic

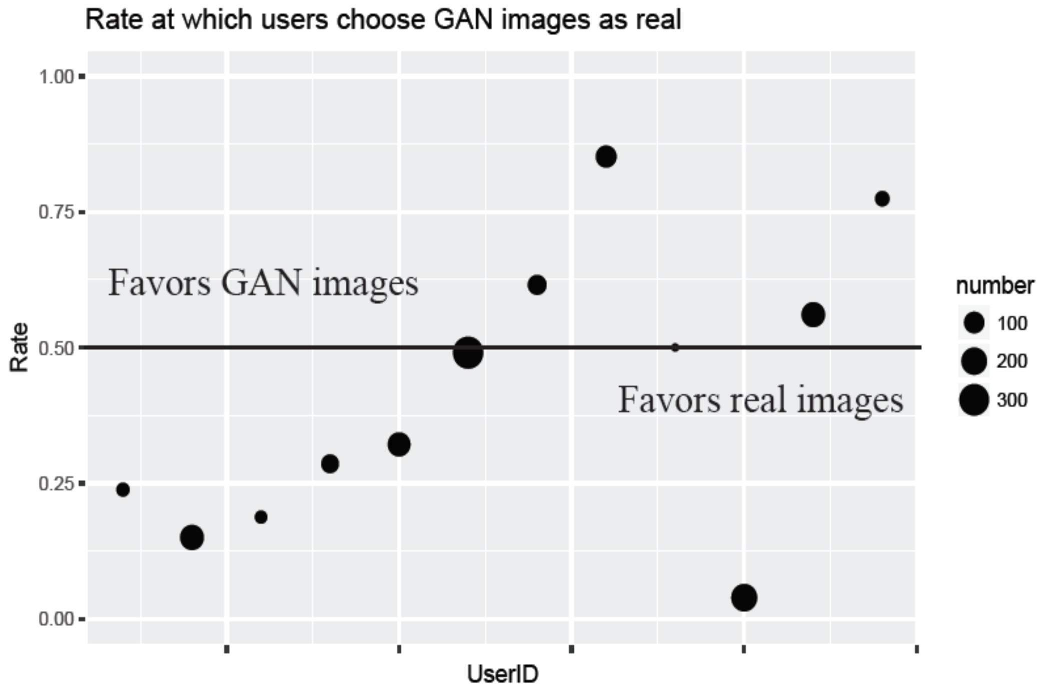
Real

Karras, T., et al. “Progressive growing of gans for improved quality, stability, and variation.” ICLR 2018.

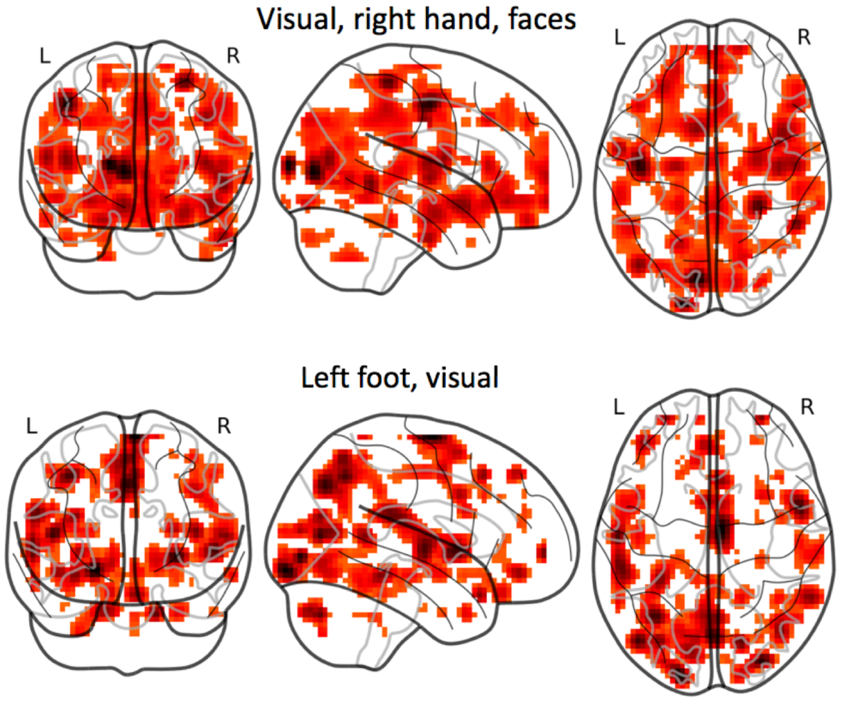


Synthetic

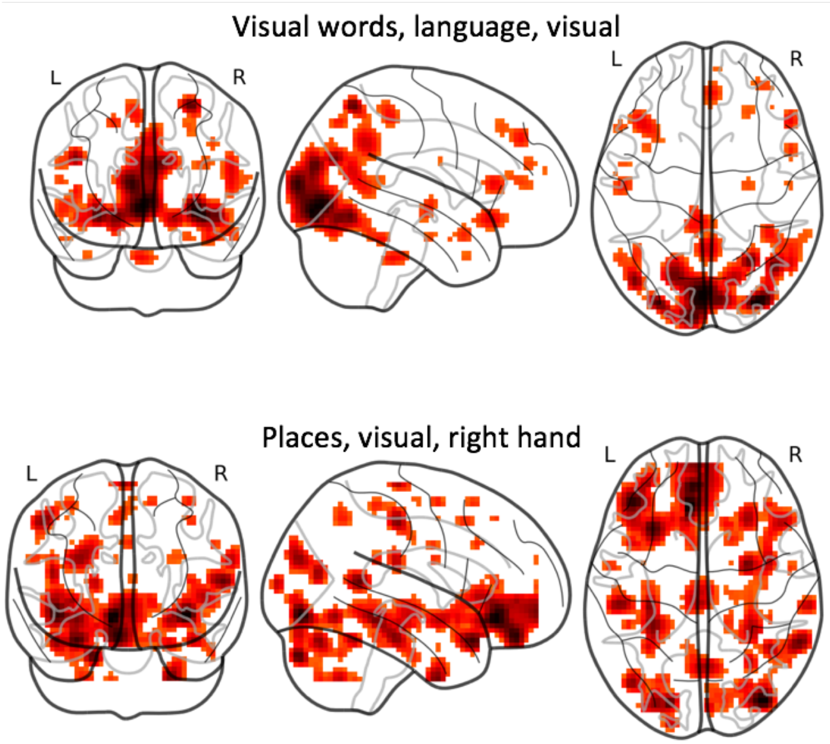
Real



Experienced radiologists were asked to choose which of a real lung x-ray and a GAN generated image were real. Subjects favored real images slightly (on average GAN images were identified as real 39% of the time) but subject behavior varied widely. Size of blob identifies number of pairs viewed; note one subject preferred GAN images over 80% of the time, another could identify real images nearly exactly.



Real

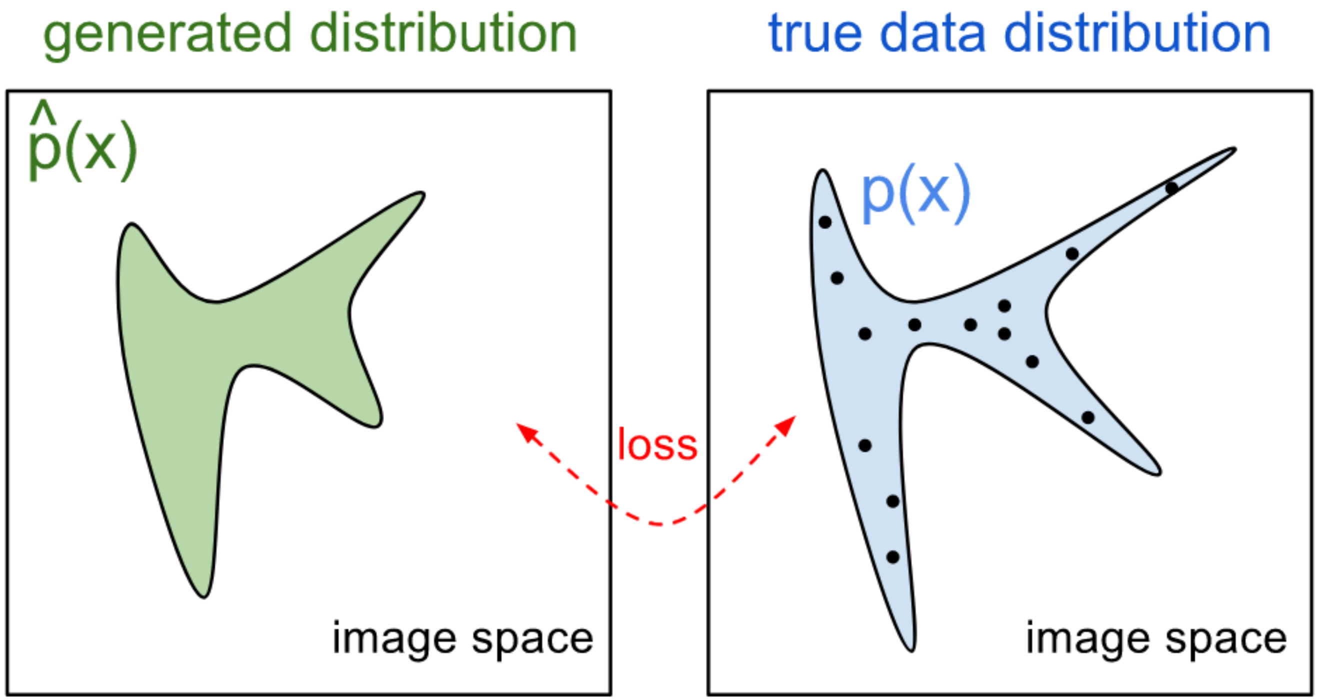


Synthetic

Outline

- ❑ Generative Adversarial Networks (GANs)
- ❑ Synthesizing fMRI
- ❑ Applications
- ❑ Pitfalls
- ❑ Tutorial

Generative Models



Source: https://blog.openai.com/generative-models/

Low Dimensions

1) Select a parametric family e.g. Gaussians

2) Find the parameters of the distribution e.g. using maximum likelihood

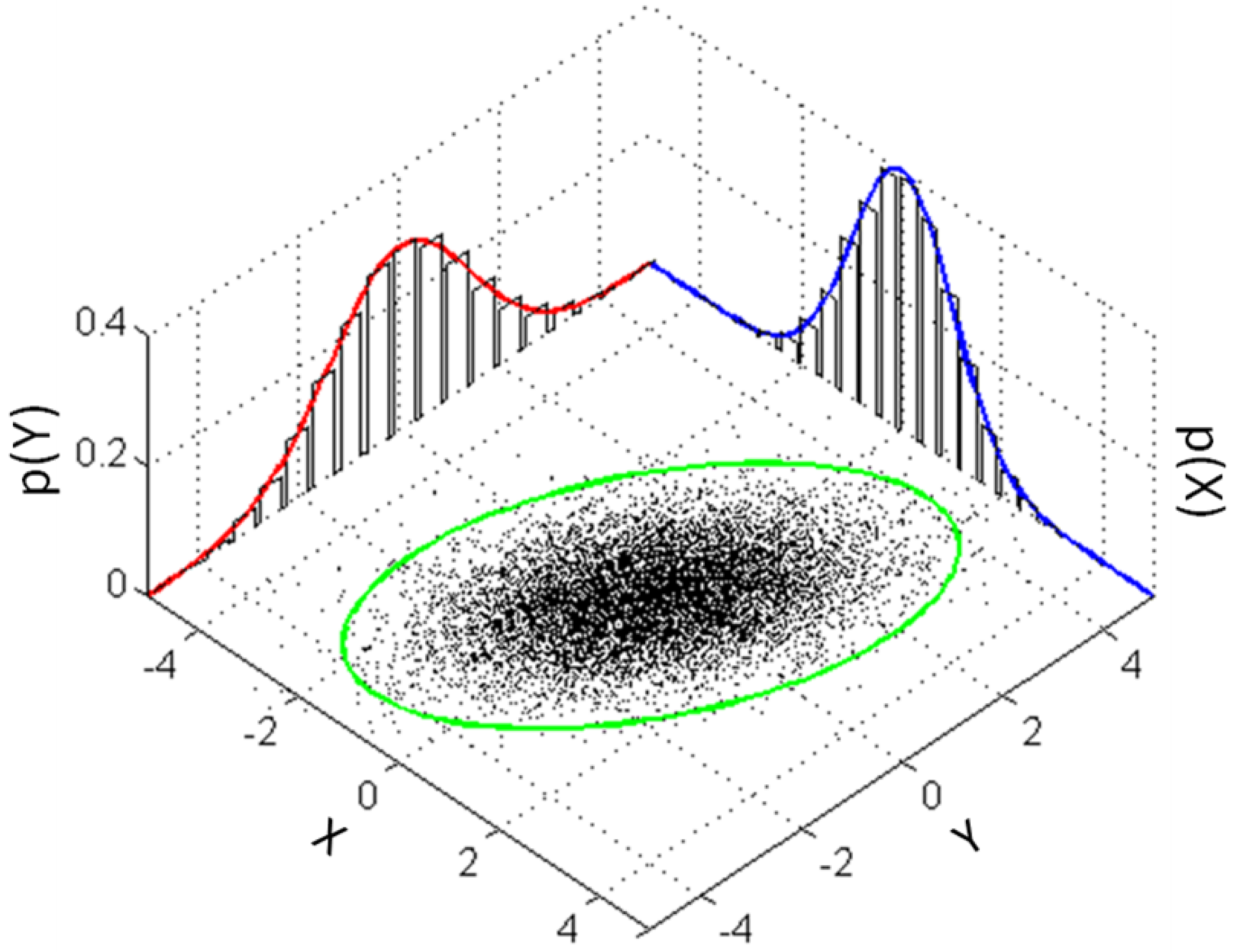
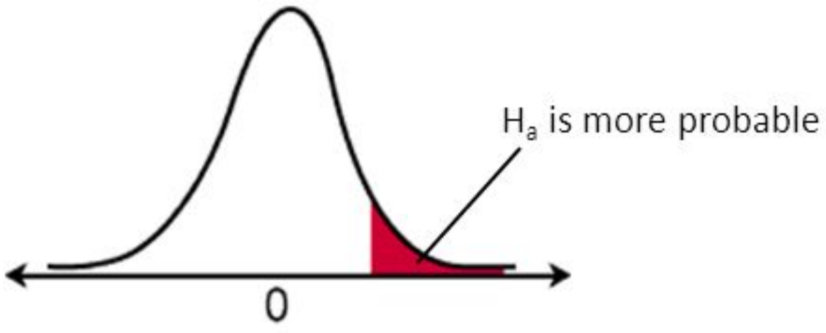
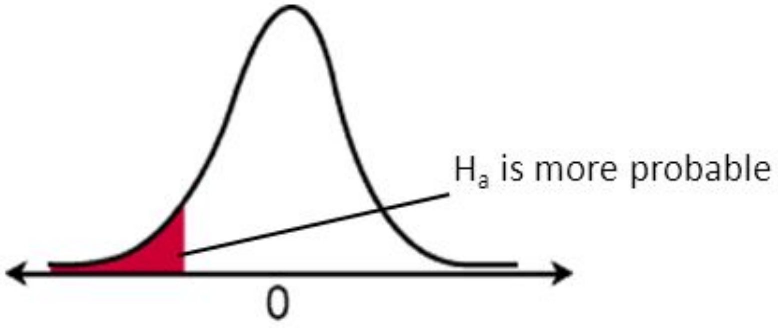


Image Source: Wikipedia



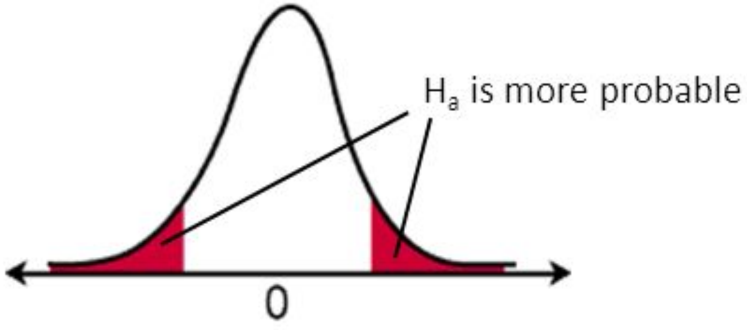
Right-tail test

H_a: $\mu >$ value



Left-tail test

H_a: $\mu <$ value



Two-tail test

H_a: $\mu \neq$ value

Image source: http://slideplayer.com/slide/9325599/

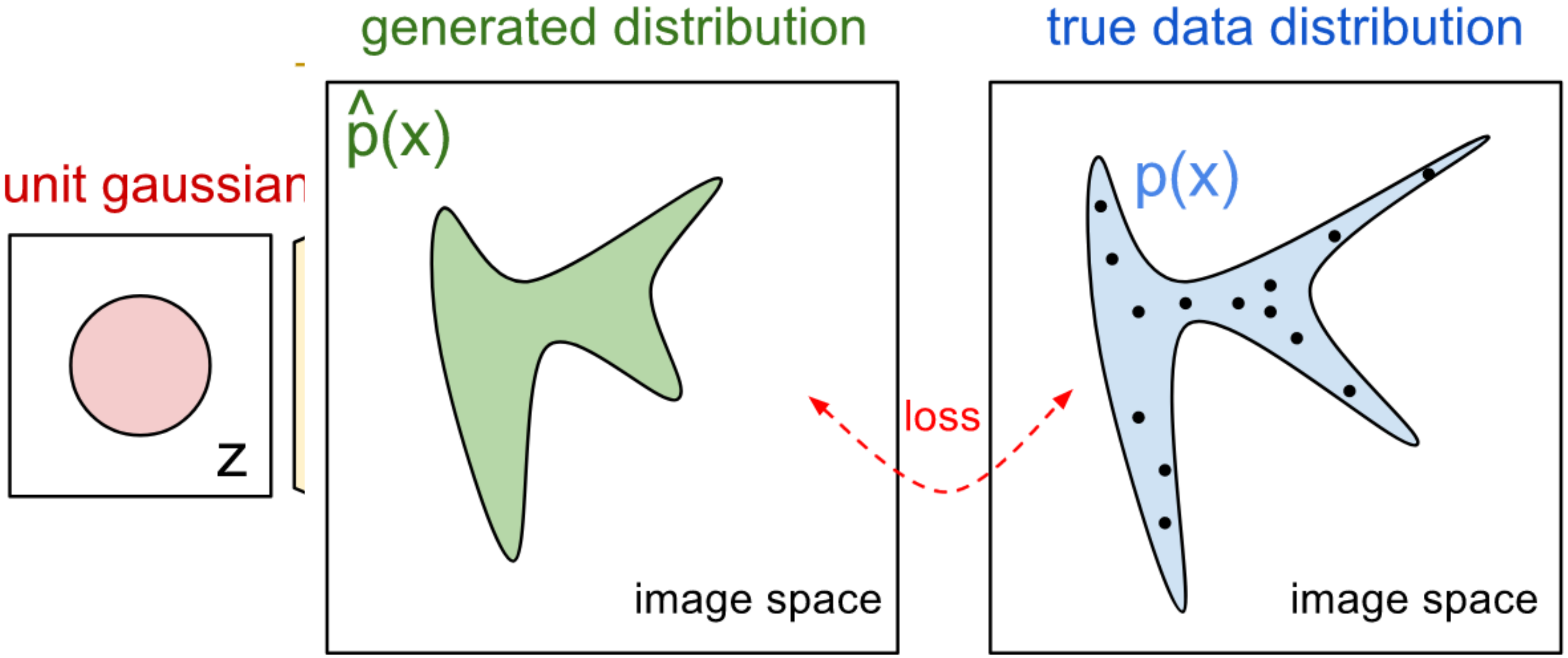
Generative Models in High Dimensions

None of the standard distributions / fitting methods work for high dimensional complex data.

Modern strategies:

- Variational Autoencoders (VAE): Kingma, Diederik P., and Max Welling. "Auto-encoding variational Bayes." *NIPS 2014*.
- Generative Adversarial Networks (GAN): Goodfellow, Ian, et al. "Generative adversarial nets." *NIPS* 2014.
- and many others…

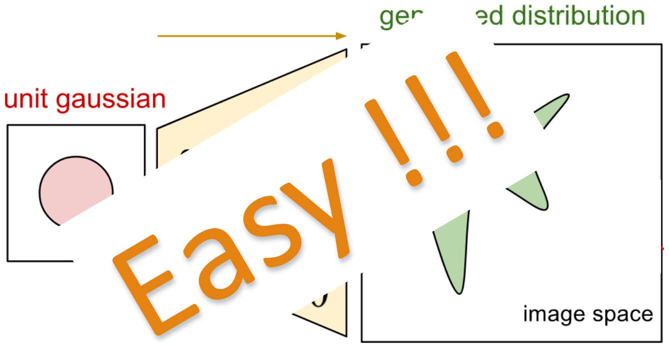
Implicit Generative Models



Source: https://blog.openai.com/generative-models/

Using the Implicit Generative Model

Synthesis
(Sampling)



$$z \sim \mathcal{N}(0,1); \; x = f_\theta(z)$$

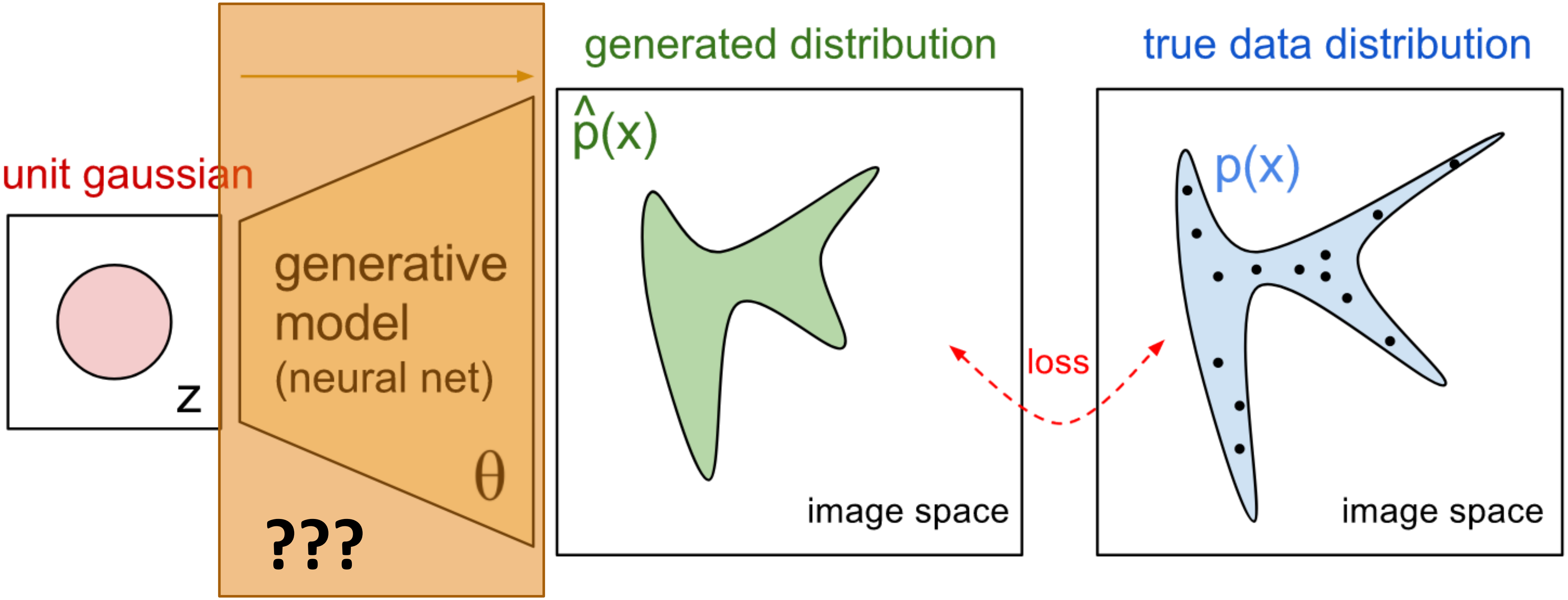
Scoring / fitting
(likelihood)

Hard!!!

$$\hat{p}(x) = \int_z \mathbf{1}_{[x=f_\theta(z)]} \mathcal{N}(0,1) dz$$

Source: https://blog.openai.com/generative-models/

Estimating the Model (i.e. neural net)

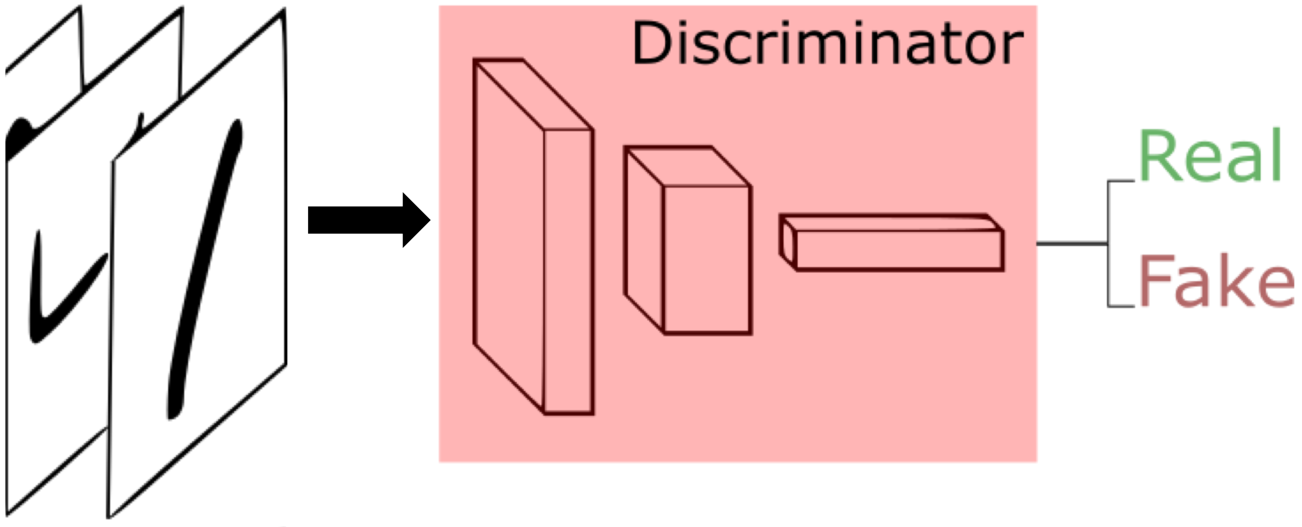


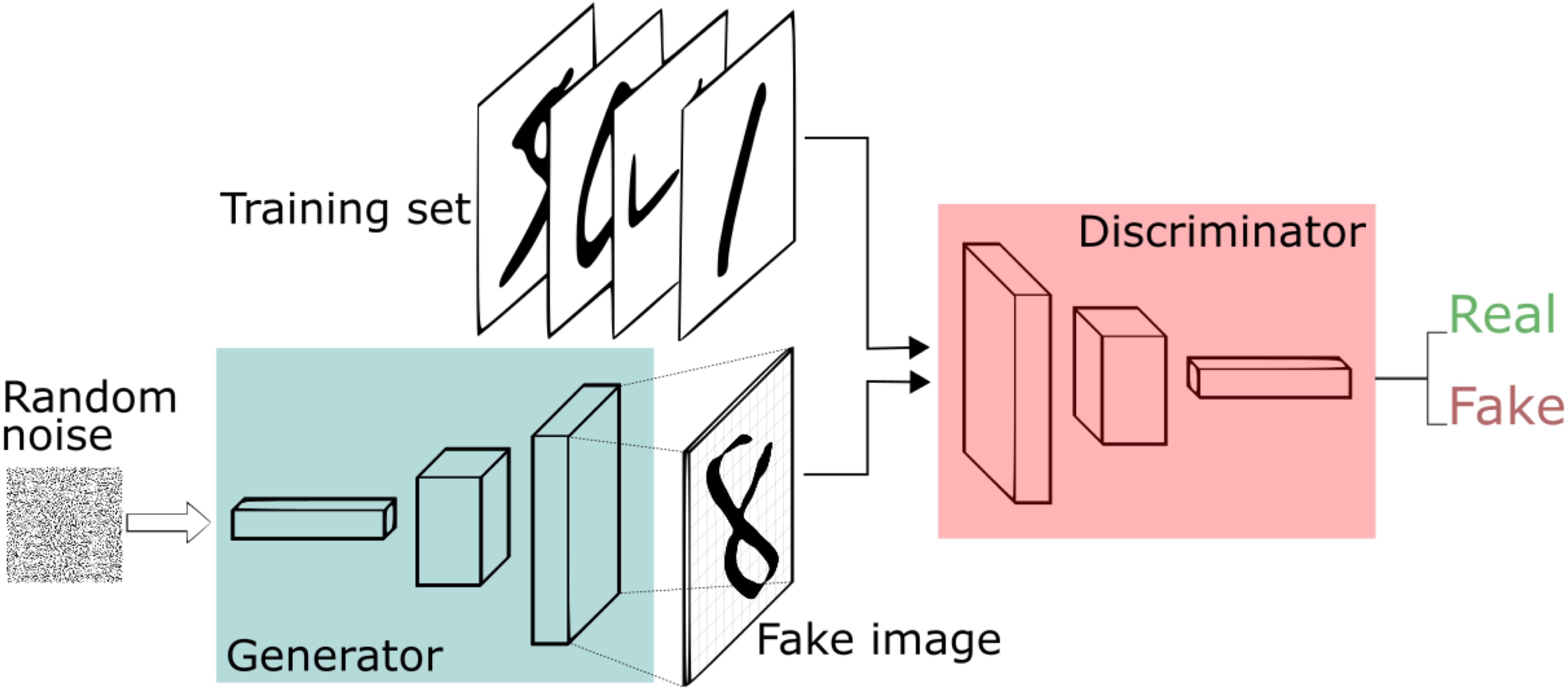
Learning a Loss Function

Instead of a standard loss function, the GAN uses a classifier to determine if the generated data “looks” real or fake

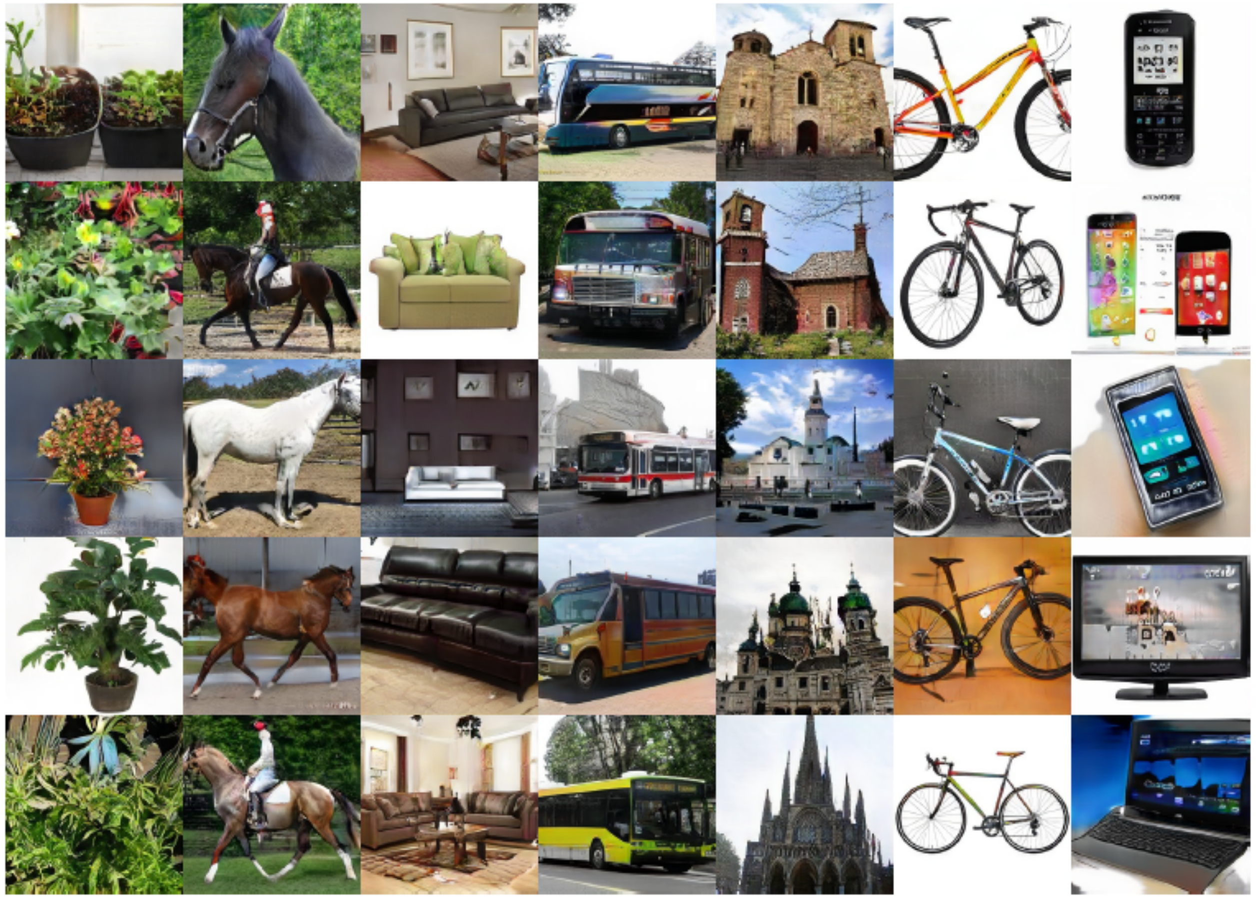
Interpreted as a two-player game:

- Generator tries to fool the classifier
- Discriminator tries to avoid getting fooled




Training set
Discriminator
Real
Fake
Random
noise
Generator
Fake image

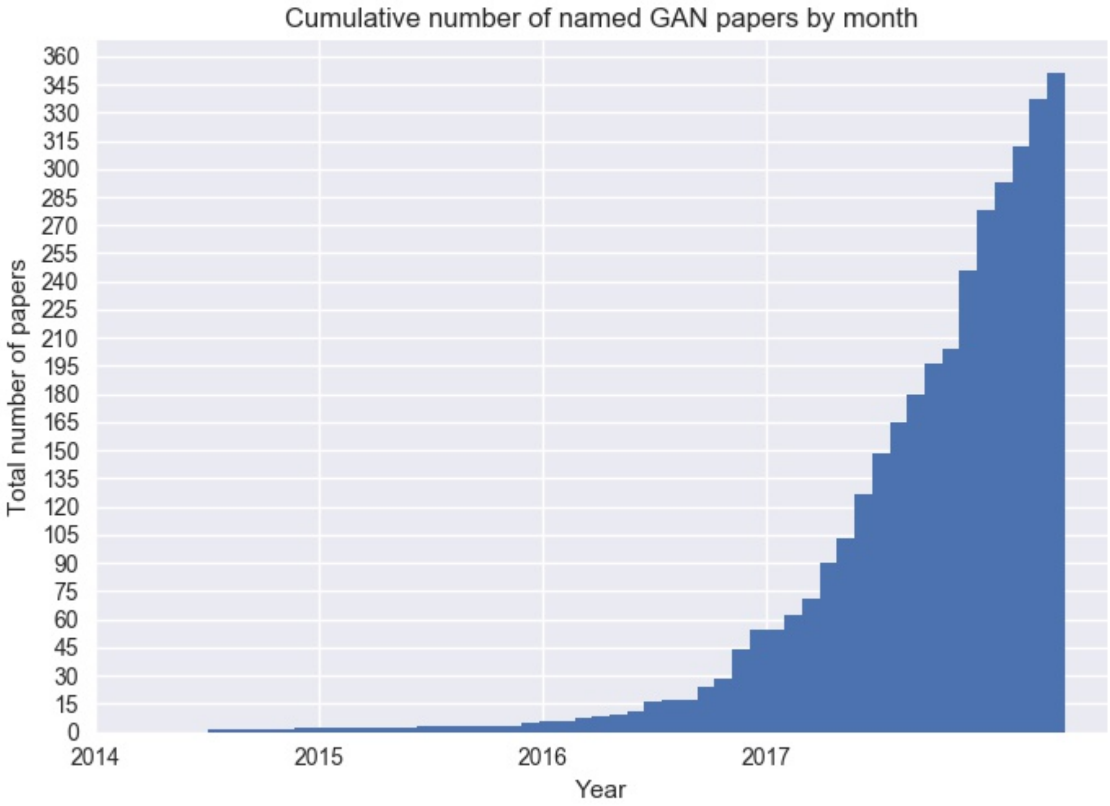
GANs in 2018



POTTEDPLANT HORSE SOFA BUS CHURCHOUTDOOR BICYCLE TVMONITOR

Karras, T., et al. “Progressive growing of gans for improved quality, stability, and variation.” ICLR 2018.

Lots of Innovation



- 3D-ED-GAN—Shape Inpainting using 3D Generative Adversarial Network and Recurrent Convolutional Networks
- 3D-GAN—Learning a Probabilistic Latent Space of Object Shapes via 3D Generative-Adversarial Modeling (github)
- 3D-IWGAN—Improved Adversarial Systems for 3D Object Generation and Reconstruction (github)
- 3D-RecGAN—3D Object Reconstruction from a Single Depth View with Adversarial Learning (github)
- ABC-GAN—ABC-GAN: Adaptive Blur and Control for improved training stability of Generative Adversarial Networks(github)
- ABC-GAN—GANs for LIFE: Generative Adversarial Networks for Likelihood Free Inference
- AC-GAN—Conditional Image Synthesis With Auxiliary Classifier GANs

GANs for fMRI

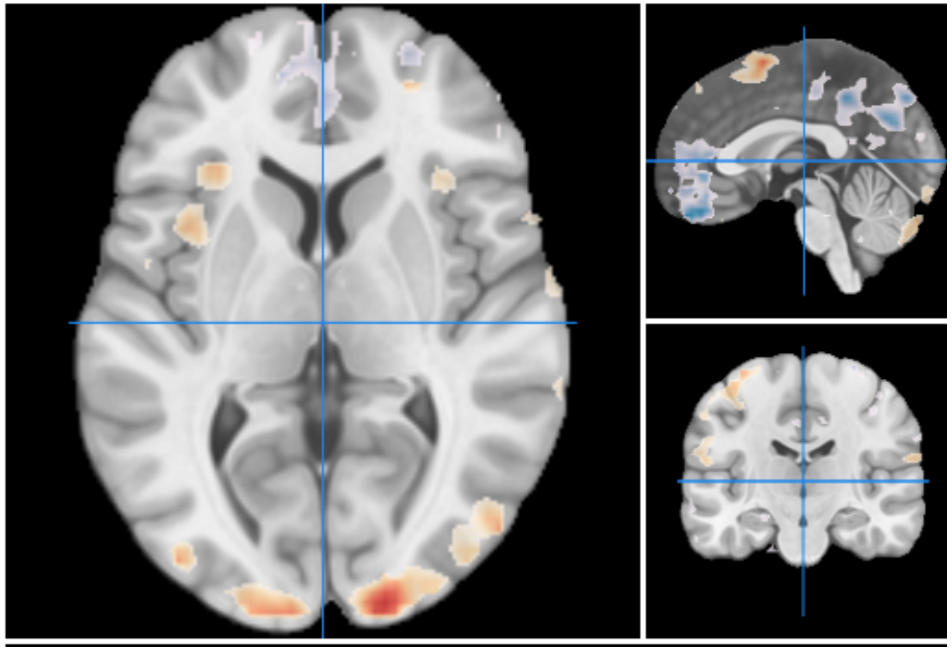
Joint work with Sanmi Koyejo, Peiye Zhuang, Ran Li, Alex Schwing

Dataset - BrainPedia

NEUROVAULT

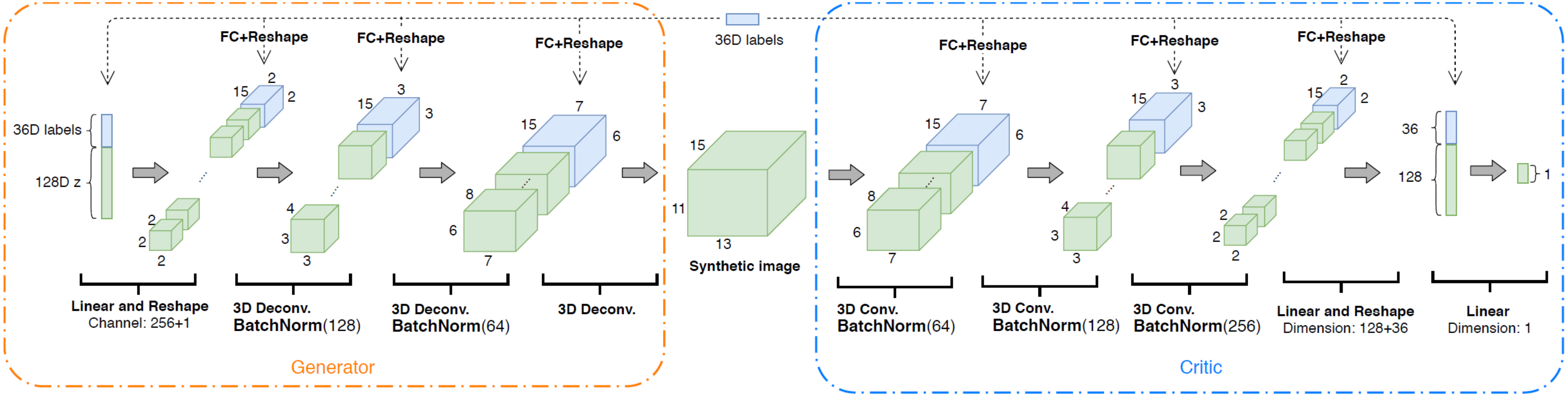
6573 z-score task contrasts · 45 total tasks · 19 cognitive labels

Add Date	Oct. 26, 2016, 7:31 p.m.
Authors	None
Contributors	
Description	BrainPedia is a collection of SPMs obtained from about 30 protocoles from OpenfMRI, the Human Connectome Project and Neurospin research center that map a wide set of cognitive functions.
DOI	None
Field Strength	None
id	1952
Journal	None

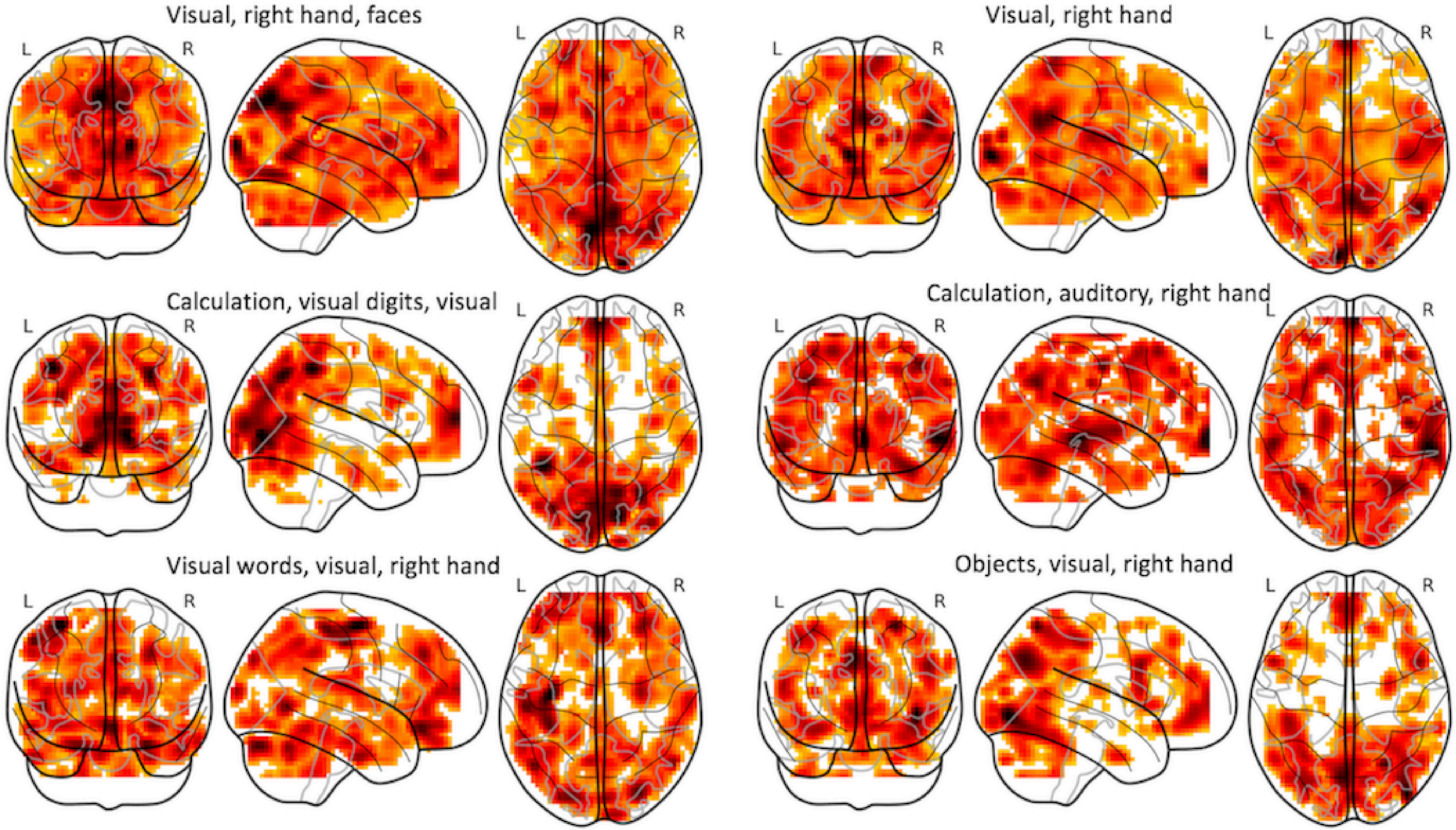


Contributed by bthirion

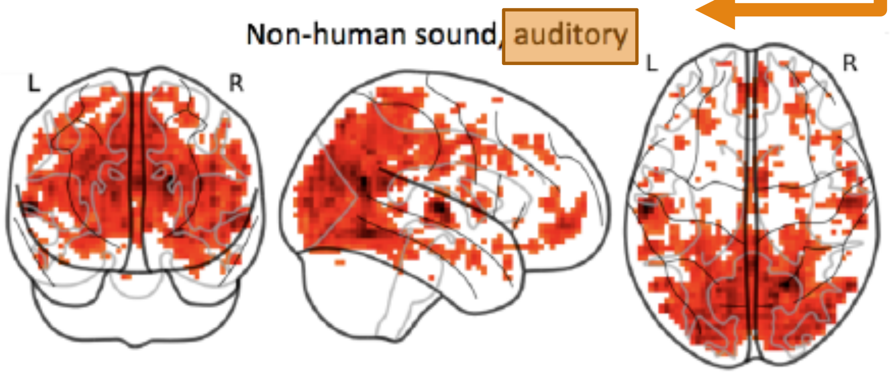
3D Conditional GAN



$$z \sim \mathcal{N}(0,1);\ x|k = f_\theta(z,k) \qquad D : \hat{x}, k \mapsto \{\text{real},\ \text{fake}\}$$


Visual, right hand, faces
L
R
L
R
Visual, right hand
L
R
L
R
Calculation, visual digits, visual
L
R
L
R
Calculation, auditory, right hand
L
R
L
R
Visual words, visual, right hand
L
R
L
R
Objects, visual, right hand
L
R
L
R

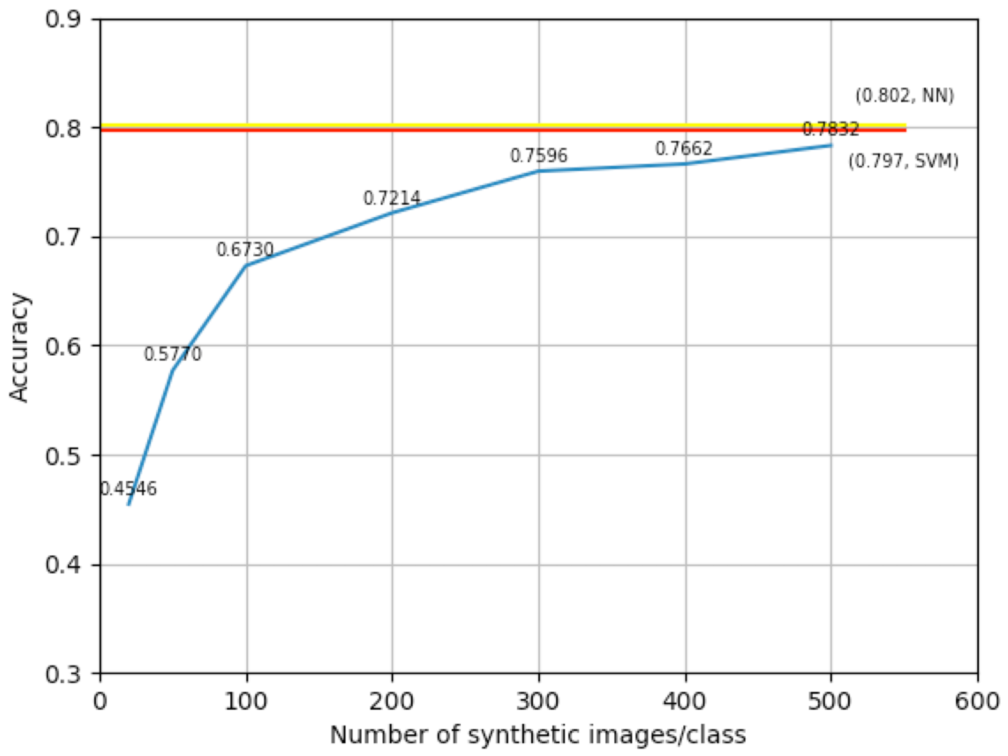
Evaluation: Neurosynth Decoder



Analysis	Correlation
auditory cortex	0.207
auditory	0.198
heschi	0.195
heschl gyrus	0.194
sounds	0.190
fractional anisotropy	0.189
fa	0.188
pitch	0.187
anisotropy fa	0.186
anisotropy	0.185

Source: http://neurosynth.org/decode/

Evaluation: Classifier



1) Use subset of data to train a generative model.

2) Use synthesized data to train a classifier.

3) Evaluate classifier on test set of real data.

Application: Classifier Data Augmentation

Downsampling	Input	Classifier	Accuracy	Macro F1	Precision	Recall
4.0×	Real	SVM	0.797	0.797	0.813	0.797
	Real	NN	0.802	0.802	0.817	0.802
	Real+Synth.	SVM	0.806	0.803	0.823	0.807
	Real+Synth.	NN	**0.819**	**0.817**	**0.830**	**0.819**
2.0×	Real	SVM	0.855	0.857	0.867	0.857
	Real	NN	0.863	0.863	0.872	0.863
	Real+Synth.	SVM	0.860	0.863	0.860	0.857
	Real+Synth.	NN	**0.891**	**0.894**	**0.906**	**0.891**

Comparison to Traditional Models

Training data	Accuracy	F1	Precision	Recall
Synth. data from GMM (20 images/class)	0.203	0.309	0.309	0.202
Synth. data from GMM (500 images/class)	0.720	0.725	0.765	0.720
Real+Synth. (from GMM)	0.793	0.798	0.824	0.793
Synth. data from ICW-GAN (20 images/class)	**0.458**	**0.433**	**0.537**	**0.458**
Synth. data from ICW-GAN (500 images/class)	**0.783**	**0.776**	**0.805**	**0.783**
Real+Synth. (from ICW-GAN)	**0.819**	**0.817**	**0.830**	**0.819**

Cognitive Neuroscience Applications

Data augmentation for decoding

Two-sample testing: Determine sample size / power for new experiments

One sample testing: Determine if an individual differs from the population

Realistic data source for fitting complex biologically plausible models

…

Summary of Pitfalls

- GANs are susceptible to mode collapse
- High dimensional generative models are difficult to evaluate
- GANs are hard to visualize and explain
- Likelihood function is intractable, thus new methods are required for standard probabilistic tasks e.g. hypothesis testing
- Model training is compute-intensive
- Synthesis is not always realistic
- …

Karras, T., et al. "Progressive growing of gans for improved quality, stability, and variation." ICLR 2018.

Conclusion

Synthetic fMRI is an exciting and promising technology for accelerating brain data analysis.

Lots of work to do:

- Higher resolution / higher quality synthesis
- Improved evaluation
- Modeling of individual differences
- Outlier detection
- Data imputation

- Potential clinical applications
- Extension to time series
- …

1. Clone the repository:

```
git clone https://github.com/BlissChapman/OHBMDeepBrainSynthesisTutorial
```

2. Install dependencies:

```
cd OHBMDeepBrainSynthesisTutorial
python3 -m pip install -r requirements.txt --user
```

3. Start a Jupyter notebook server:

```
jupyter notebook
```

4. Select Tutorial.ipynb and enjoy!

Thank You!

Questions: bliss.chapman@gmail.com