Photo: Detail from Sofia Crespo's Tribute to Manolo Part 2

COMP547

DEEP UNSUPERVISED LEARNING

Lecture #01 – Introduction

KOÇ UNIVERSITY

Aykut Erdem // Koç University // Spring 2021

Welcome to COMP547

- This courses gives an overview of deep unsupervised learning,
- In particular, we will cover deep generative models and self-supervised learning approaches.
- You will develop fundamental and practical skills at applying deep unsupervised learning to your research.

A little about me...

Koç University
Associate Professor
2020-now

Hacettepe University
Associate Professor
2010-2020

Universitá Ca' Foscari di Venezia
Post-doctoral Researcher
2008-2010

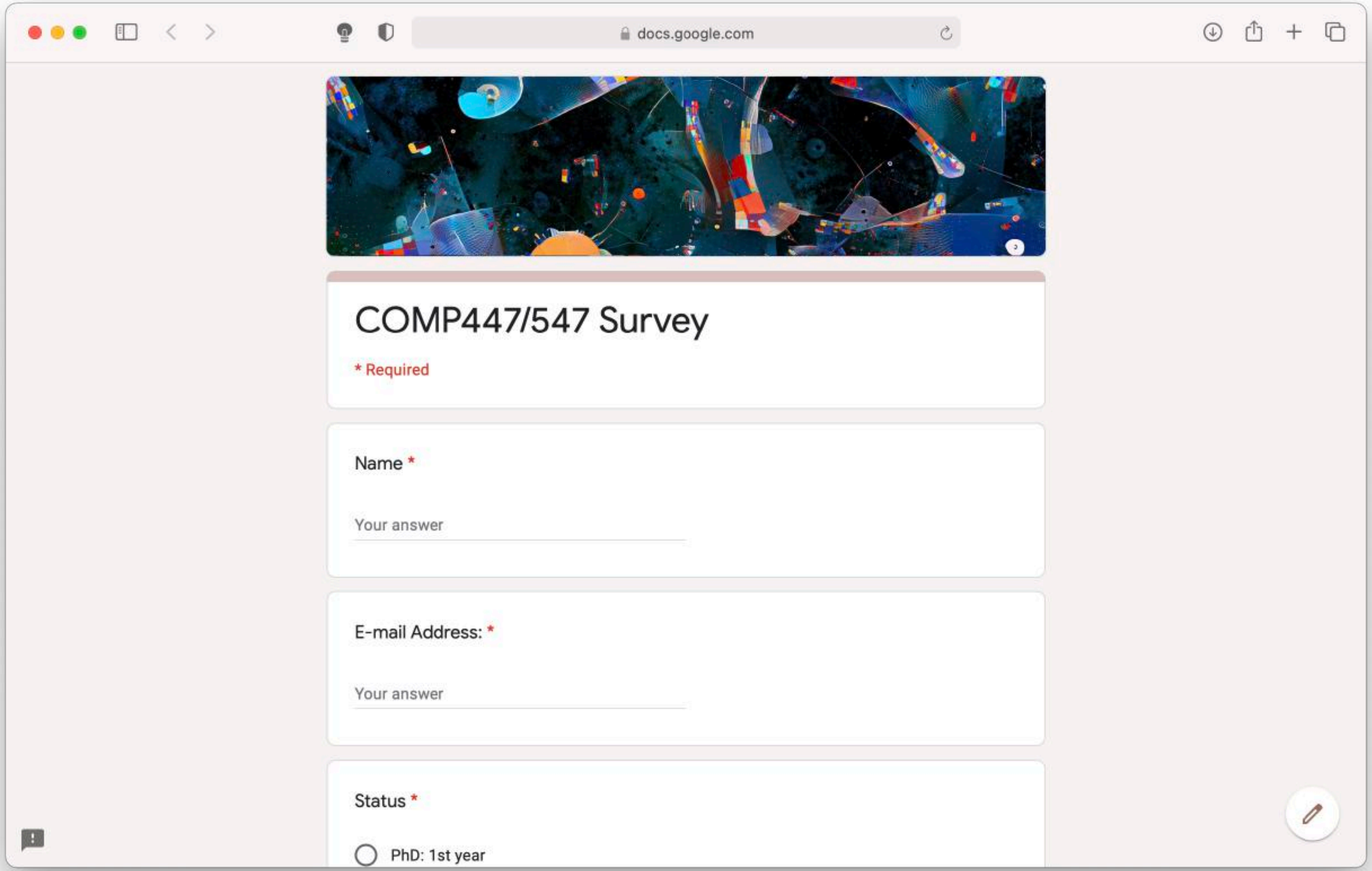
Middle East Technical University
1997-2008
Ph.D., 2008
M.Sc., 2003
B.Sc., 2001

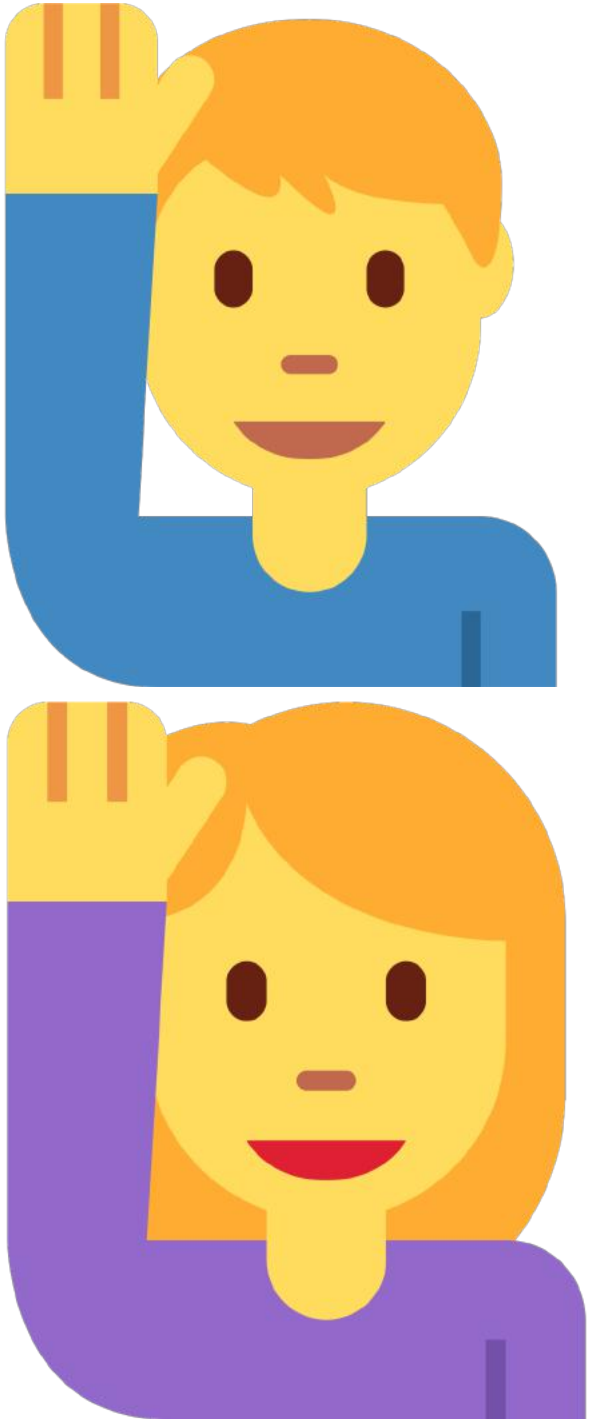
MIT
Fall 2007
Visiting Student

VirginiaTech
Visiting Research Scholar
Summer 2006

- I explore better ways to understand, interpret and manipulate visual data.
- My research interests span a diverse set of topics, ranging from image editing to visual saliency estimation, and to multimodal learning for integrated vision and language.

What about you?





https://forms.gle/2NNYSXBiGGXHskucA

Lecture overview

- course logistics
- course topics
- what is deep unsupervised learning

- **Disclaimer:** Some of the material and slides for this lecture were borrowed from

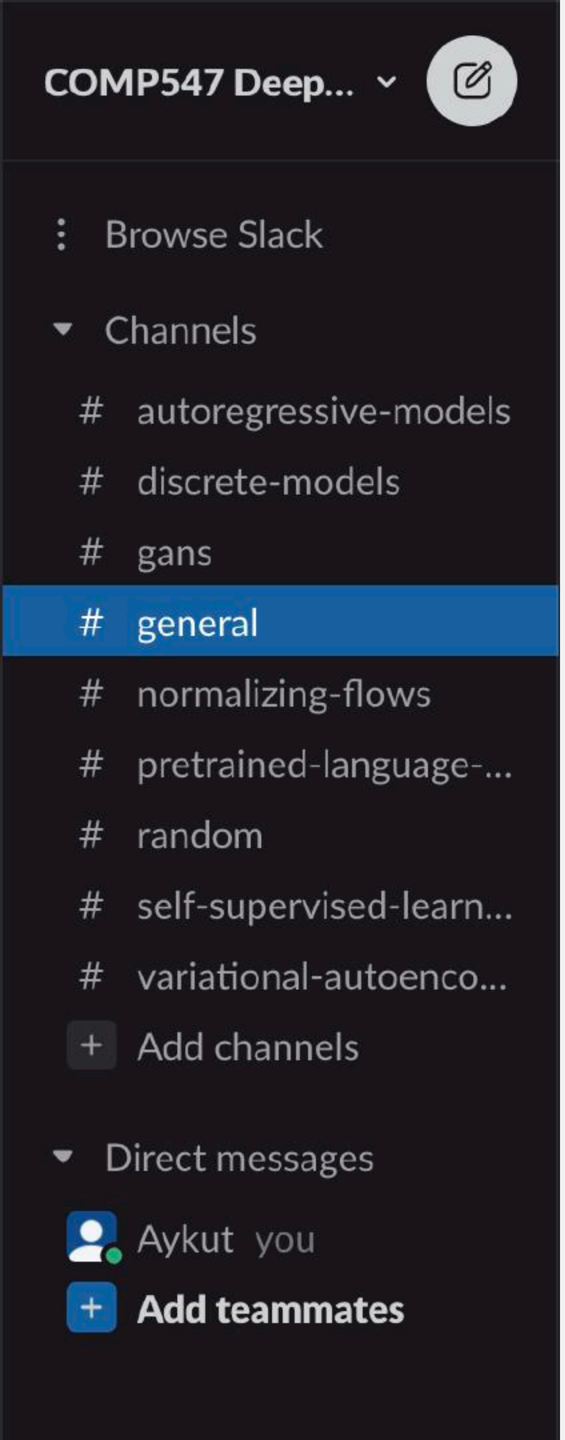
  —Pieter Abbeel, Peter Chen, Jonathan Ho, Aravind Srinivas' Berkeley CS294-158 class

Course Logistics

Course Information

Time/Location	8:30-9:45am Monday and Wednesday (Zoom)
Instructor	Aykut Erdem
Course webpage	https://aykuterdem.github.io/classes/comp547/

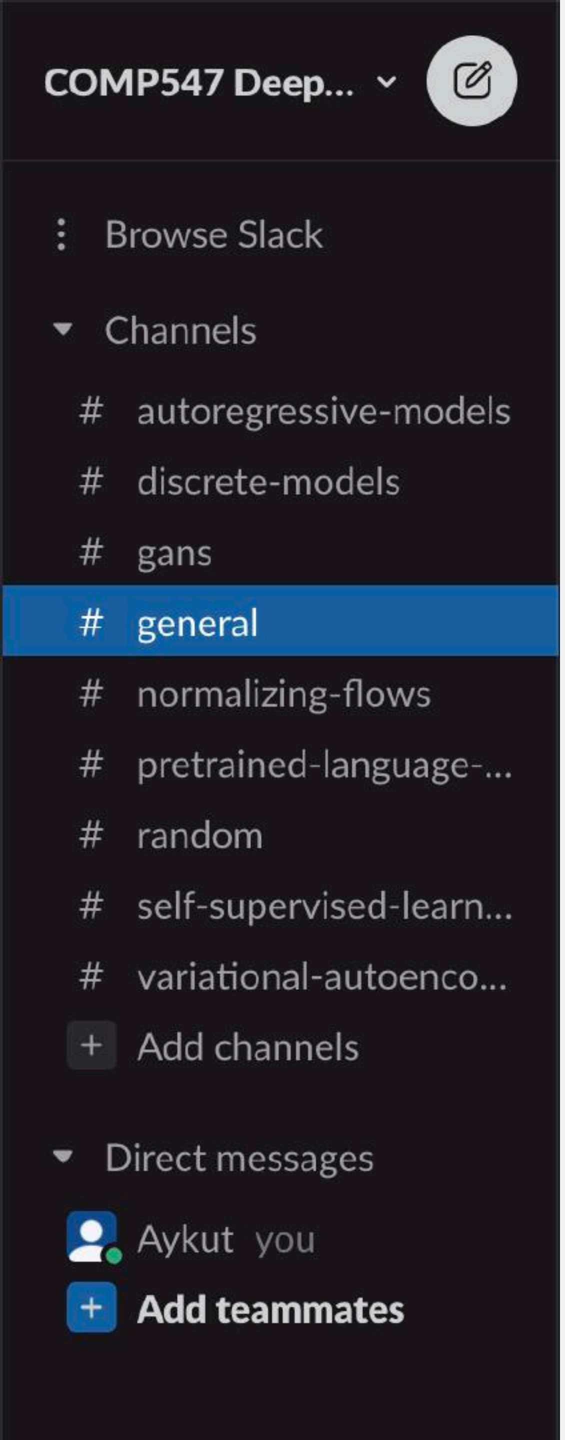
- Blackboard for course related announcements and collecting and grading your submissions.

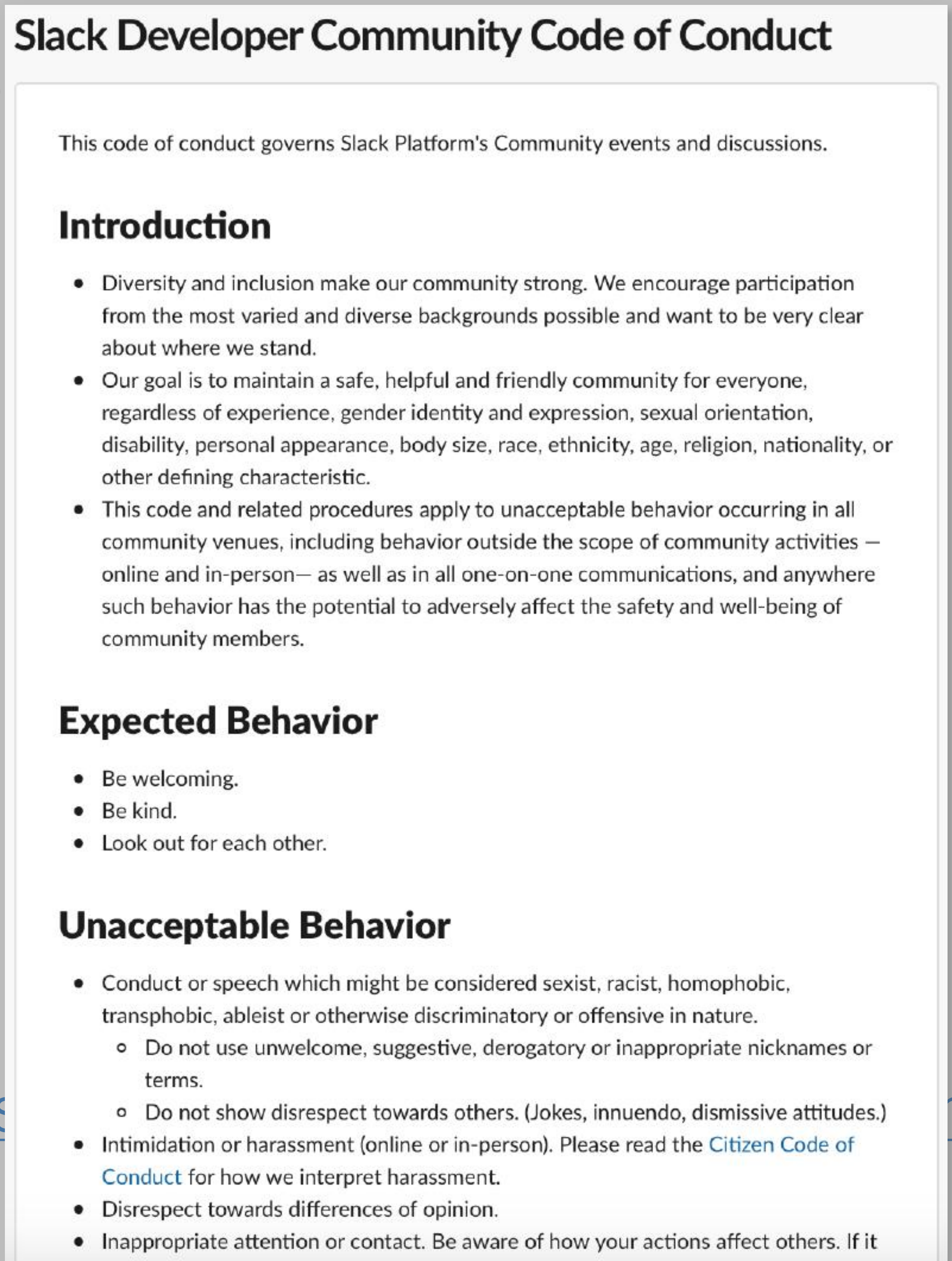


COMP547 Slack Workspace

https://join.slack.com/t/comp547/signup

CO pace





Slack Developer Community Code of Conduct

This code of conduct governs Slack Platform's Community events and discussions.

Introduction

- Diversity and inclusion make our community strong. We encourage participation from the most varied and diverse backgrounds possible and want to be very clear about where we stand.
- Our goal is to maintain a safe, helpful and friendly community for everyone, regardless of experience, gender identity and expression, sexual orientation, disability, personal appearance, body size, race, ethnicity, age, religion, nationality, or other defining characteristic.
- This code and related procedures apply to unacceptable behavior occurring in all community venues, including behavior outside the scope of community activities — online and in-person— as well as in all one-on-one communications, and anywhere such behavior has the potential to adversely affect the safety and well-being of community members.

Expected Behavior

- Be welcoming.
- Be kind.
- Look out for each other.

Unacceptable Behavior

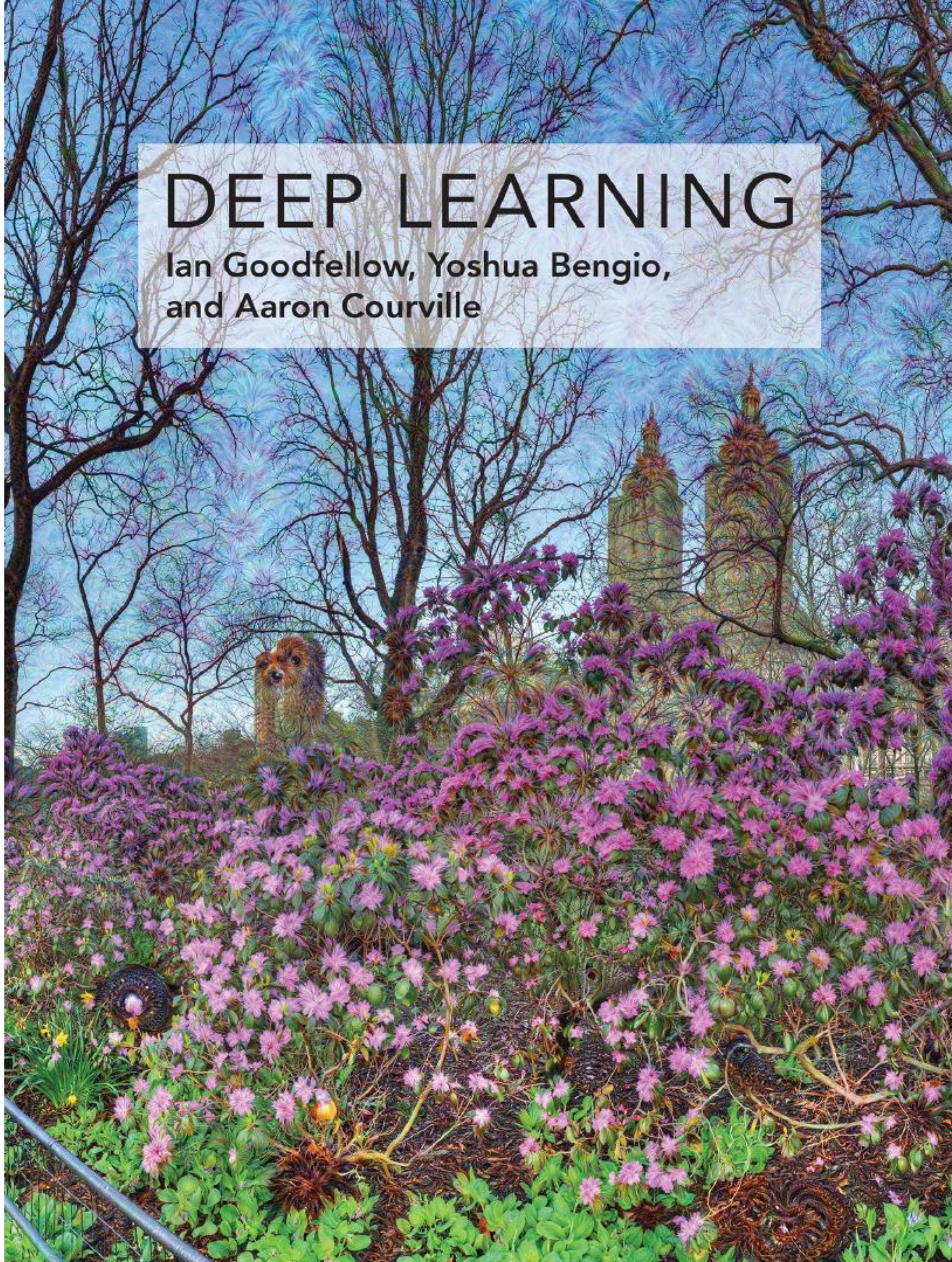
- Conduct or speech which might be considered sexist, racist, homophobic, transphobic, ableist or otherwise discriminatory or offensive in nature.
  - Do not use unwelcome, suggestive, derogatory or inappropriate nicknames or terms.
  - Do not show disrespect towards others. (Jokes, innuendo, dismissive attitudes.)
- Intimidation or harassment (online or in-person). Please read the Citizen Code of Conduct for how we interpret harassment.
- Disrespect towards differences of opinion.
- Inappropriate attention or contact. Be aware of how your actions affect others. If it

https nup

Reference Book

- Goodfellow, Bengio, and Courville, Deep Learning, MIT Press, 2016 (draft available online)

- In addition, we will extensively use online materials (video lectures, blog posts, surveys, papers, etc.)



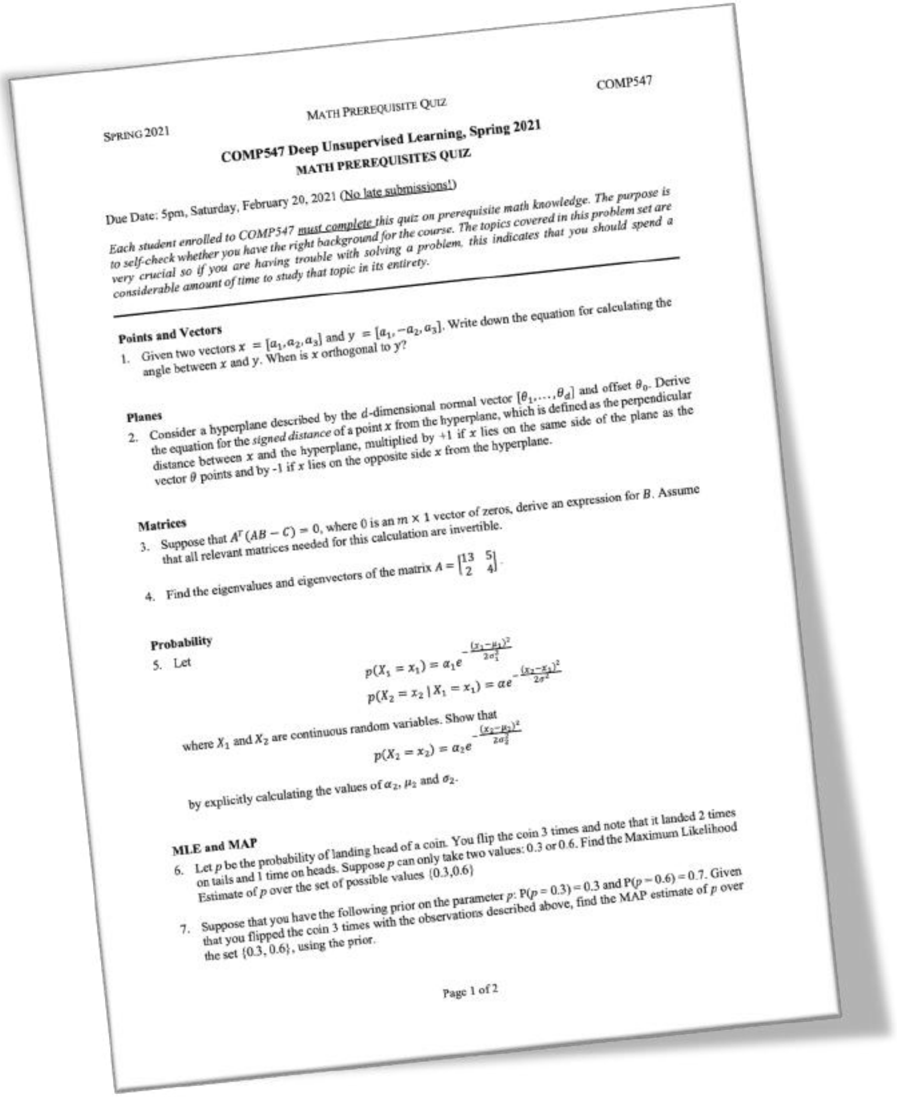
Instruction Style

- Students are responsible for studying and keeping up with the course material outside of class time.
  - Reading particular book chapters, papers or blogs, or
  - Watching some video lectures.

- After the first six lectures, each week we will discuss a paper on the topics of the previous week.

Prerequisites

- Calculus (MATH106, MATH203) and linear algebra (**MATH107**)
  - Derivatives,
  - Tensors, matrix operations
- Probability and statistics (**ENGR200**)
- Machine learning (ENGR421)
- Deep learning (COMP541)
- Programming (Python)



SPRING 2021 | MATH PREREQUISITE QUIZ | COMP547

COMP547 Deep Unsupervised Learning, Spring 2021

MATH PREREQUISITES QUIZ

Due Date: 5pm, Saturday, February 20, 2021 (No late submissions!)

Each student enrolled to COMP547 must complete this quiz on prerequisite math knowledge. The purpose is to self-check whether you have the right background for the course. The topics covered in this problem set are very crucial so if you are having trouble with solving a problem, this indicates that you should spend a considerable amount of time to study that topic in its entirety.

Points and Vectors

1. Given two vectors $x = [a_1, a_2, a_3]$ and $y = [a_1, -a_2, a_3]$. Write down the equation for calculating the angle between x and y. When is x orthogonal to y?

Planes

2. Consider a hyperplane described by the d-dimensional normal vector $[\theta_1, \ldots, \theta_d]$ and offset θ_0. Derive the equation for the *signed distance* of a point x from the hyperplane, which is defined as the perpendicular distance between x and the hyperplane, multiplied by +1 if x lies on the same side of the plane as the vector θ points and by -1 if x lies on the opposite side x from the hyperplane.

Matrices

3. Suppose that $A^T(AB - C) = 0$, where 0 is an $m \times 1$ vector of zeros, derive an expression for B. Assume that all relevant matrices needed for this calculation are invertible.

4. Find the eigenvalues and eigenvectors of the matrix $A = \begin{bmatrix} 13 & 5 \\ 2 & 4 \end{bmatrix}$.

Probability

5. Let

$$p(X_1 = x_1) = \alpha_1 e^{-\frac{(x_1-\mu_1)^2}{2\sigma_1^2}}$$
$$p(X_2 = x_2 \mid X_1 = x_1) = \alpha e^{-\frac{(x_2-x_1)^2}{2\sigma^2}}$$

where X_1 and X_2 are continuous random variables. Show that

$$p(X_2 = x_2) = \alpha_2 e^{-\frac{(x_2-\mu_2)^2}{2\sigma_2^2}}$$

by explicitly calculating the values of α_2, μ_2 and σ_2.

MLE and MAP

6. Let p be the probability of landing head of a coin. You flip the coin 3 times and note that it landed 2 times on tails and 1 time on heads. Suppose p can only take two values: 0.3 or 0.6. Find the Maximum Likelihood Estimate of p over the set of possible values {0.3,0.6}

7. Suppose that you have the following prior on the parameter p: $P(p = 0.3) = 0.3$ and $P(p = 0.6) = 0.7$. Given that you flipped the coin 3 times with the observations described above, find the MAP estimate of p over the set {0.3, 0.6}, using the prior.

Page 1 of 2

Math Prerequisite Quiz

Each student enrolled to COMP547 must complete this quiz!

Topics Covered in ENGR421

- Basics of Statistical Learning
  - Loss function, MLE, MAP, Bayesian estimation, bias-variance tradeoff, overfitting, regularization, cross-validation

- Supervised Learning
  - Nearest Neighbor, Naïve Bayes, Logistic Regression, Support Vector Machines, Kernels, Neural Networks, Decision Trees
  - Ensemble Methods: Bagging, Boosting, Random Forests

- Unsupervised Learning
  - Clustering: K-Means, Gaussian mixture models
  - Dimensionality reduction: PCA, SVD

Topics Covered in COMP541

- Basic linear models for classification and regression
- Stochastic Gradient Descent (Backpropagation) Learning
- AutoGrad
- Multilayer Perceptron (MLP)
- Convolutional Neural Networks
- Recurrent Neural Networks
- Attention
- Transformers
- Visualization
- Optimization
- Generalization
- Generative Models
- Reinforcement Learning

Course Topics

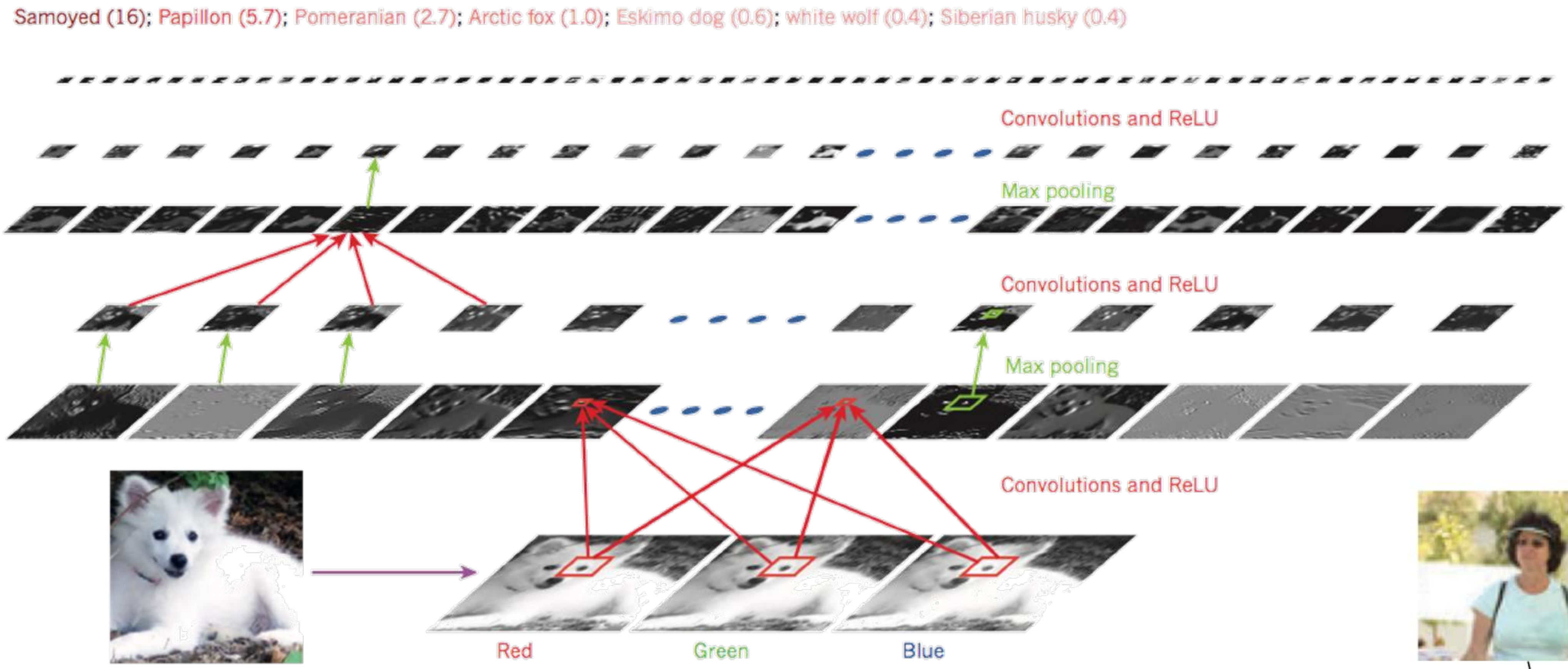
Topics Covered in This Semester

- Neural Building Blocks: CNNs and RNNs
- Neural Building Blocks: Attention and Transformers
- Autoregressive Models
- Normalizing Flow Models
- Variational Autoencoders
- Generative Adversarial Networks
- Discrete Latent Variable Models
- Self-Supervised Learning
- Pretraining Language Models

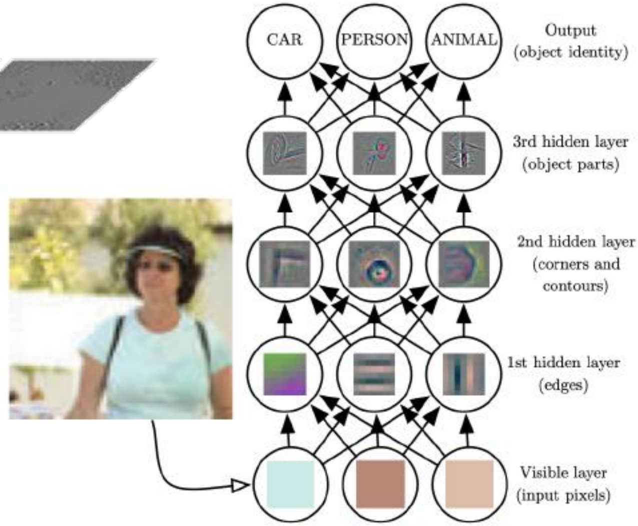
Schedule

Week	Topic	Assignments
Feb 15-17	Introduction to the course, Basics I: Spatial Processing with CNNs	
Feb 22-24	Basics II: Sequential Processing with NNs	
Mar 1-3	Autoregressive Models	Assg 1 out
Mar 8-10	Normalizing Flow Models	
Mar 15-17	Variational Autoencoders	Assg 1 due, Assg 2 out
Mar 22-24	Generative Adversarial Networks	
Mar 29-31	Generative Adversarial Networks (cont'd)	Assg 2 due, Assg 3 out
Apr 5-7	*No classes - Spring Break*	Project proposal due
Apr 12-14	Discrete Latent Variable Models	
Apr 19-21	Strengths and Weaknesses of Current Models	Assg 3 due
Apr 26-28	Self-Supervised Learning	
May 3-5	Project Progress Presentations	Project progress reports due
May 10-12	Pretraining Language Models	
May 17	Final Project Discussions	
May 24-26	Midterm Exam	
May 31, Jun 2	Final Project Presentations	
Jun 7-9	Final Project Presentations	Final project reports due

Week 1 (2): Neural building blocks: CNNs

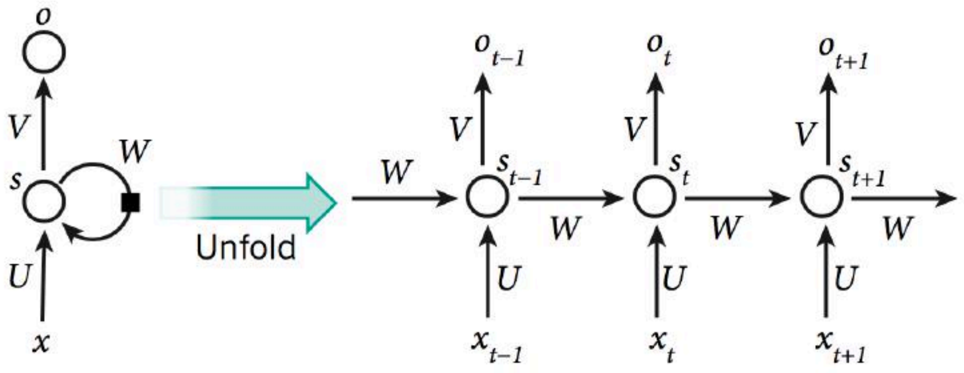


A Convolutional Neural Network (CNN)

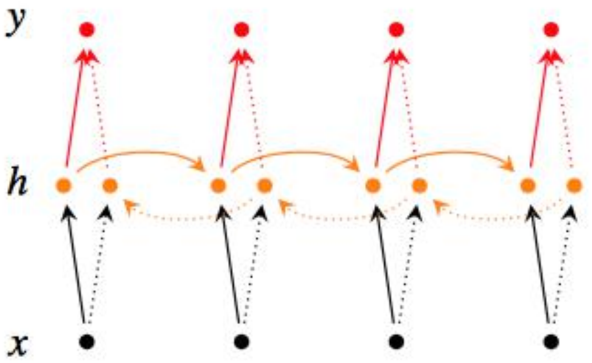


Y. LeCun, Y. Bengio, G. Hinton, **"Deep Learning"**, Nature, Vol. 521, 28 May 2015

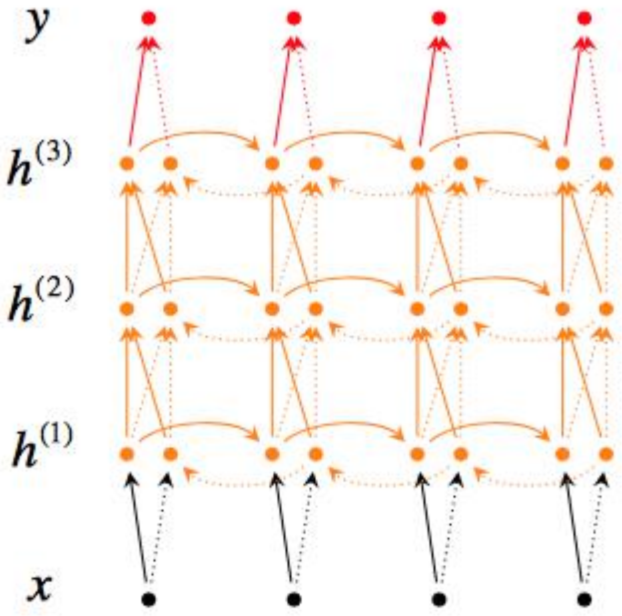
Week 2 (1): Neural building blocks: RNNs



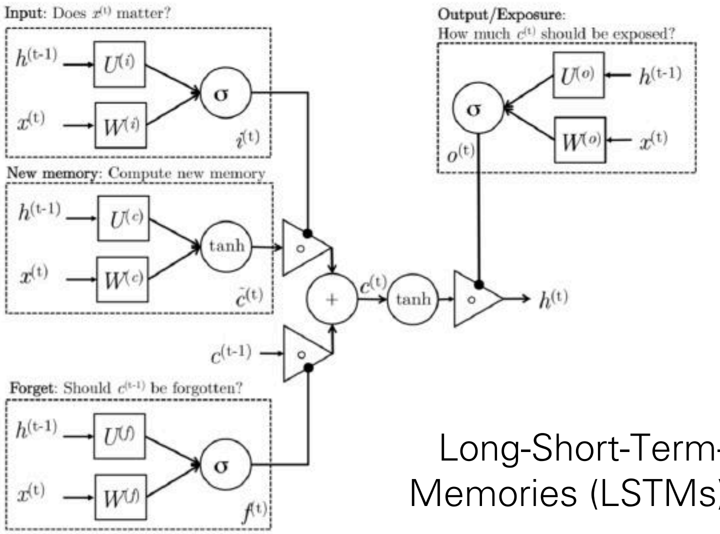
A Recurrent Neural Network (RNN)
(unfolded across time-steps)



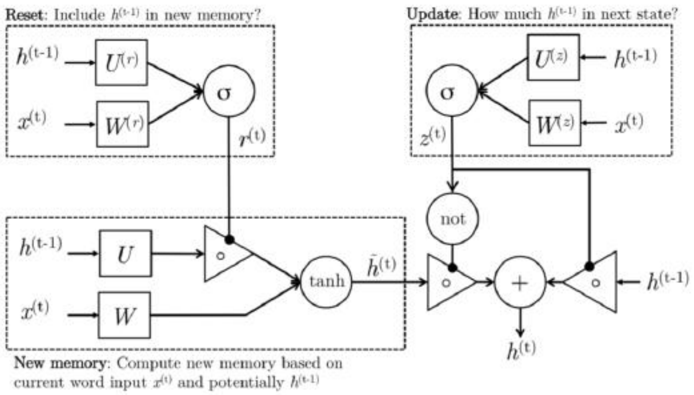
A bi-directional RNN



A deep bi-directional RNN



Long-Short-Term-Memories (LSTMs)

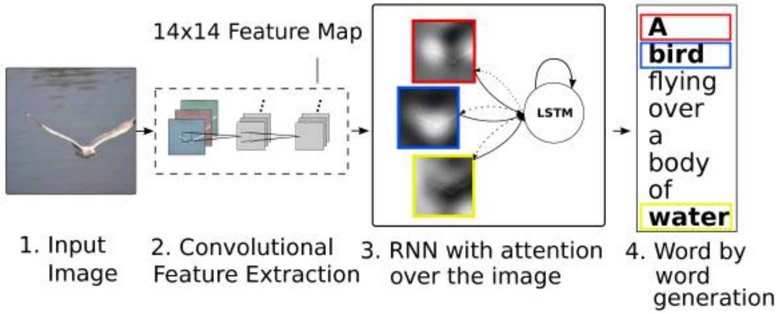


Gated Recurrent Units (GRUs)

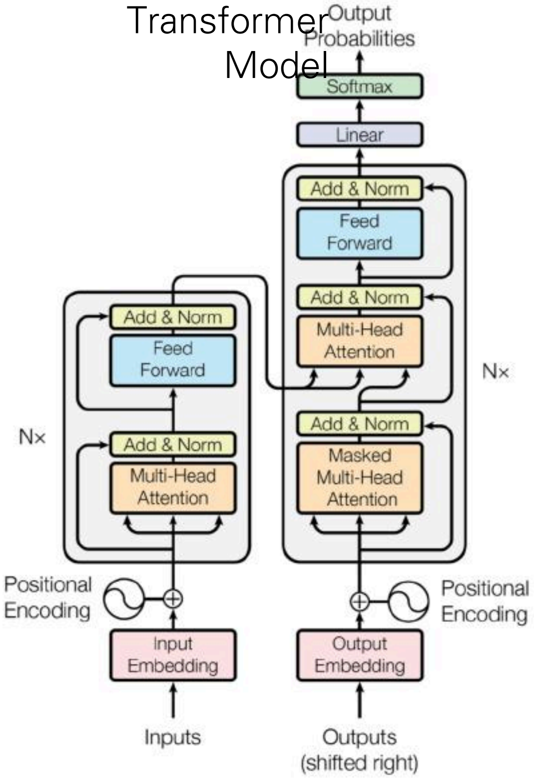
C. Manning and R Socher, **Stanford CS224n** Lecture 8 Notes
Y. LeCun, Y. Bengio, G. Hinton, "**Deep Learning**", Nature, Vol. 521, 28 May 2015

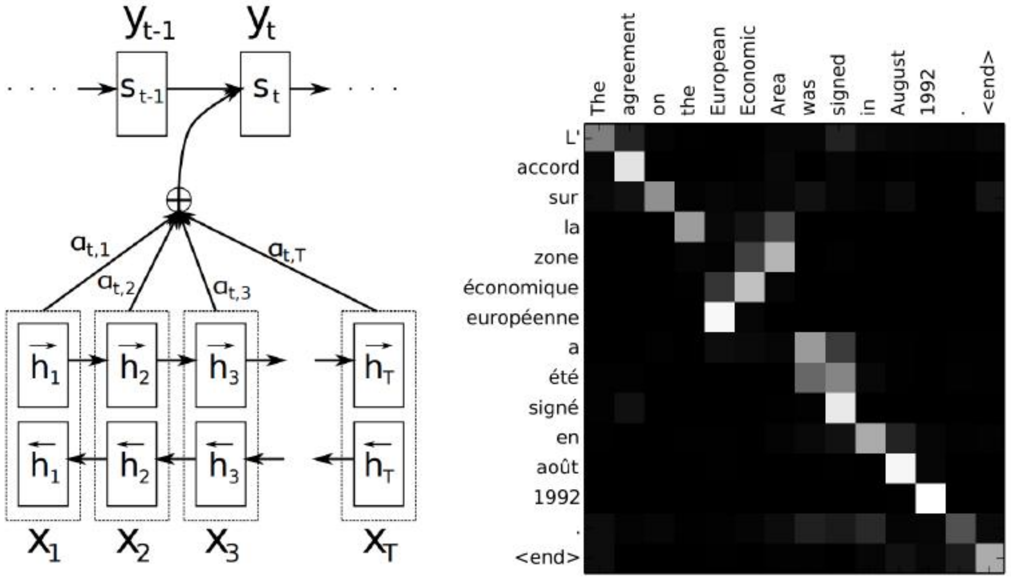
Week 2 (2): Neural building blocks: Attention mechanisms, Transformers

Spatial Attention in Image Captioning



Transformer Model





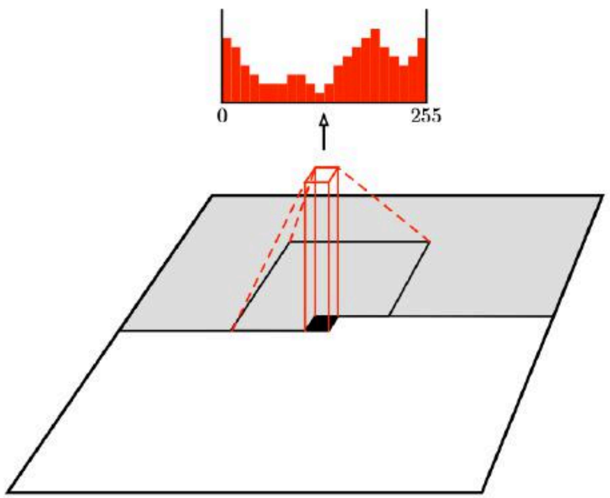
Seq2Seq with Attenion

K. Xu et al., "**Show, Attend and Tell: Neural Image Caption Generation with Visual Attention**", ICML 2015

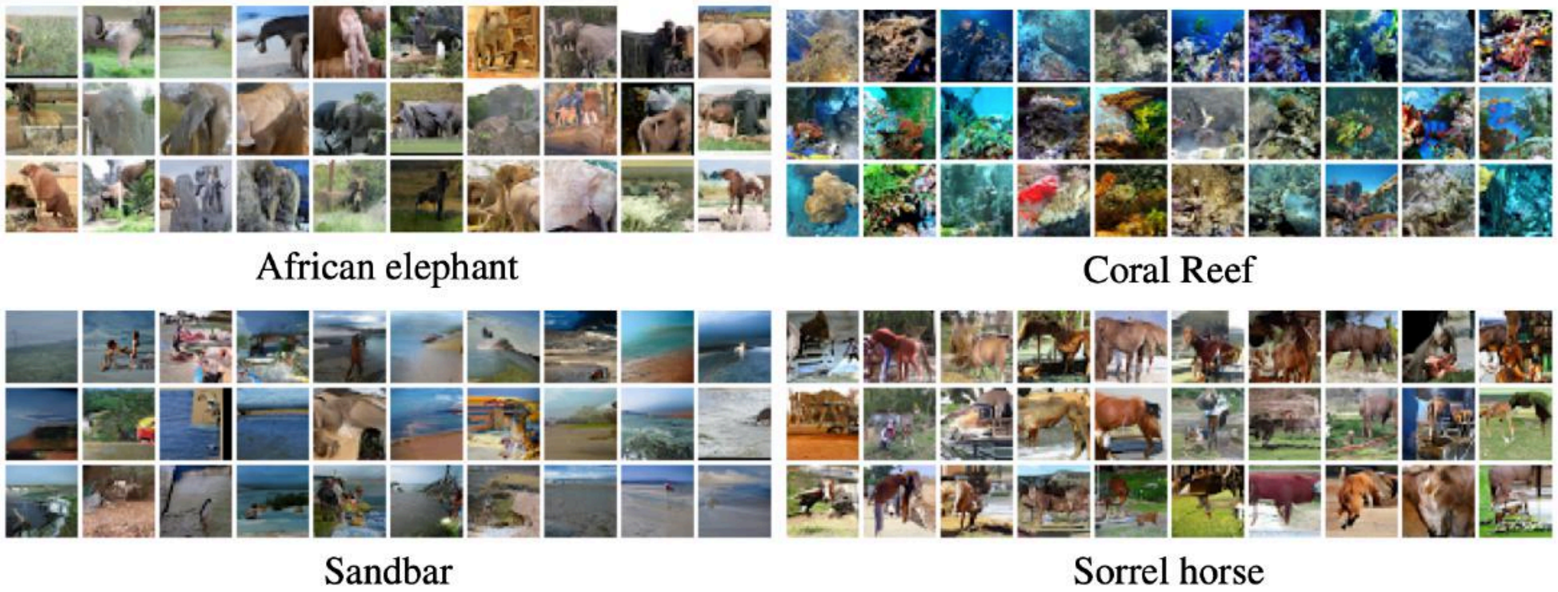
D. Bahdanau, K. Cho and Y. Bengio, "**Neural Machine Translation by Jointly Learning to Align and Translate**", ICLR 2015

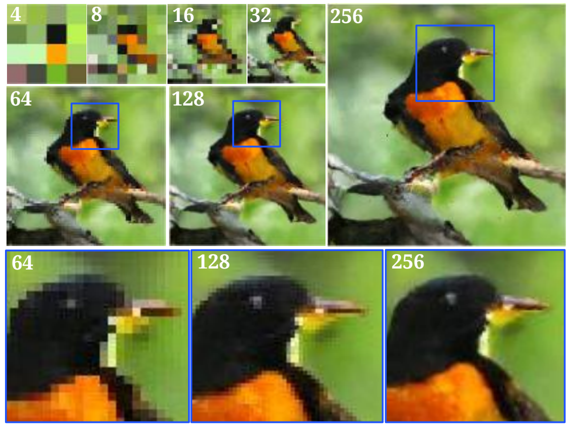
A. Vaswani et al., "**Attention Is All You Need**", NIPS 2016

Week 3: Autoregressive Models



PixelCNN



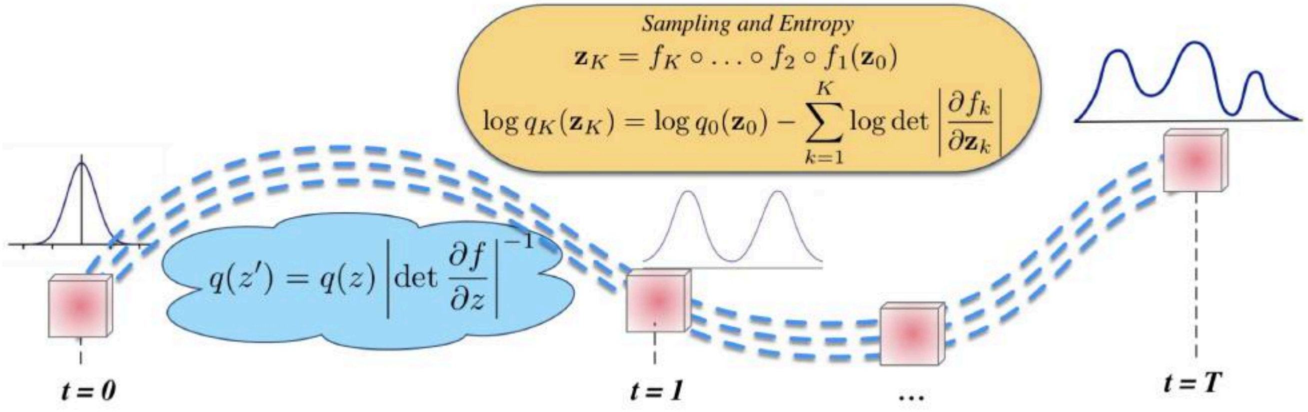


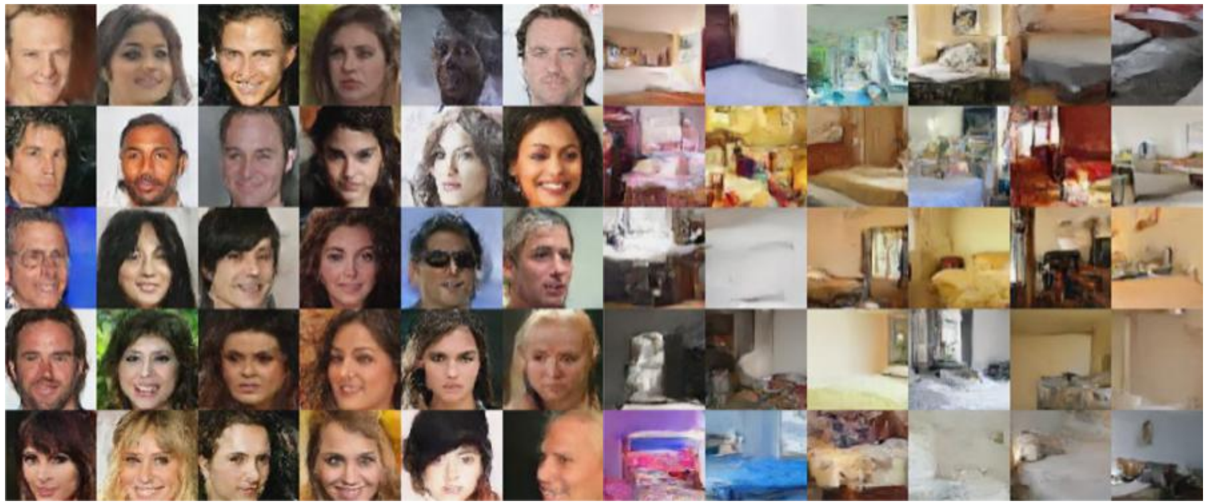
"A yellow bird with a black head, orange eyes and an orange bill."

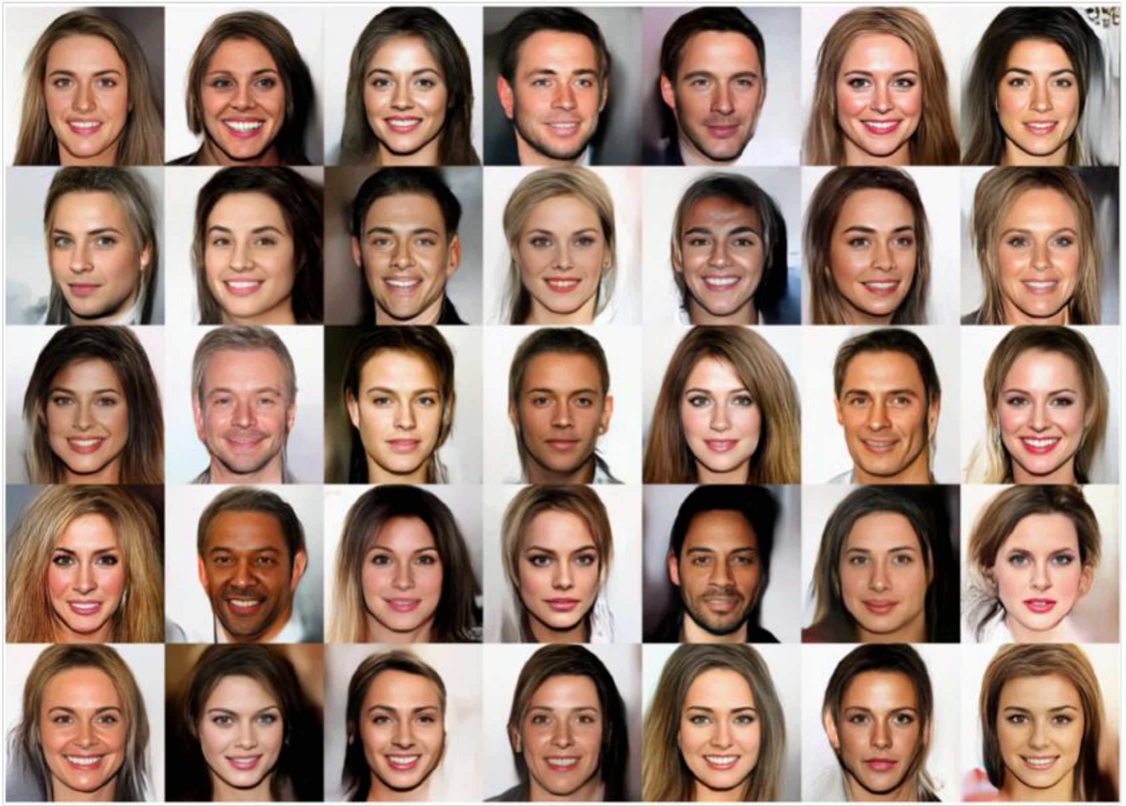
Class conditioned samples generated by PixelCNN

A. van den Oord et al., "**Conditional Image Generation with PixelCNN Decoders**", NeurIPS 2016
S. Reed et al., "**Parallel Multiscale Autoregressive Density Estimation**", ICML 2017

Week 4: Normalizing Flow Models

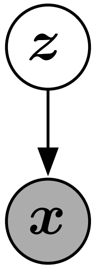




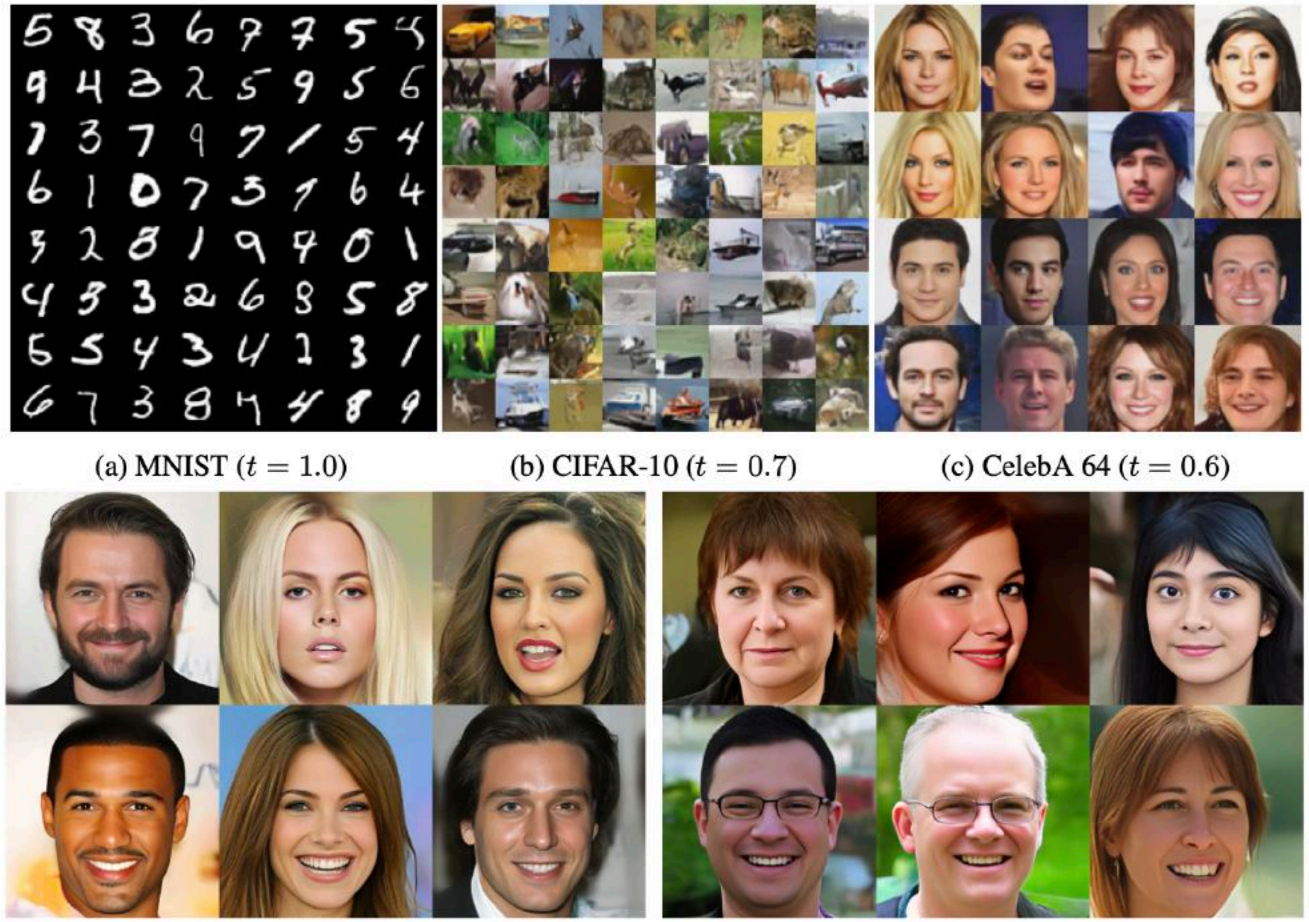


S. Mohamed, D. Rezende, **Deep Generative Models,** UAI 2017 Tutorial
L. Dinh, S. Sohl-Dickstein S. Bengio, "**Density Estimation Using Real NVP**", ICLR 2017
D.P. Kingma, P. Dhariwal, "**Glow: Generative Flow with Invertible 1×1 Convolutions**", NeurIPS 2018

Week 5: Variational Autoencoders



$$\log p(\boldsymbol{x}) \geq \log p(\boldsymbol{x}) - D_{\mathrm{KL}}\left(q(\boldsymbol{z}) \| p(\boldsymbol{z} \mid \boldsymbol{x})\right)$$
$$= \mathbb{E}_{\boldsymbol{z}\sim q} \log p(\boldsymbol{x}, \boldsymbol{z}) + H(q)$$



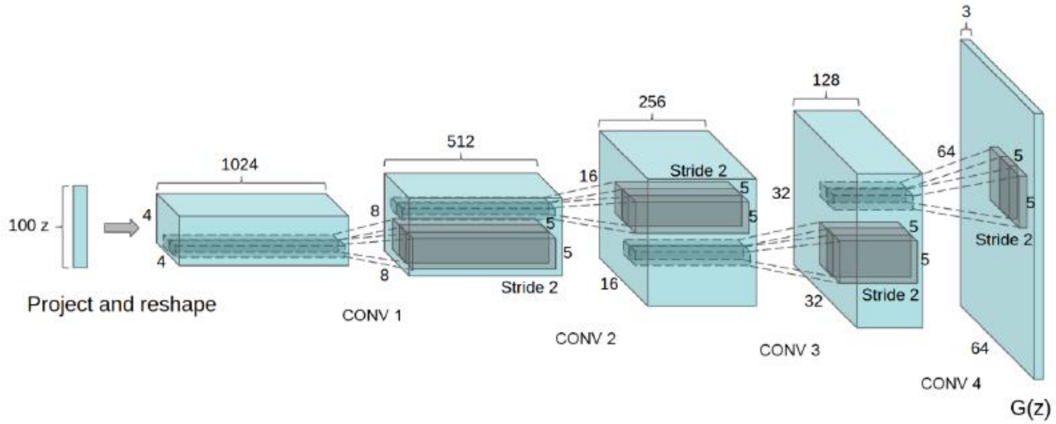
(a) MNIST ($t = 1.0$) (b) CIFAR-10 ($t = 0.7$) (c) CelebA 64 ($t = 0.6$)

(d) CelebA HQ ($t = 0.6$) (e) FFHQ ($t = 0.5$)

Synthetic images generated by NVAE

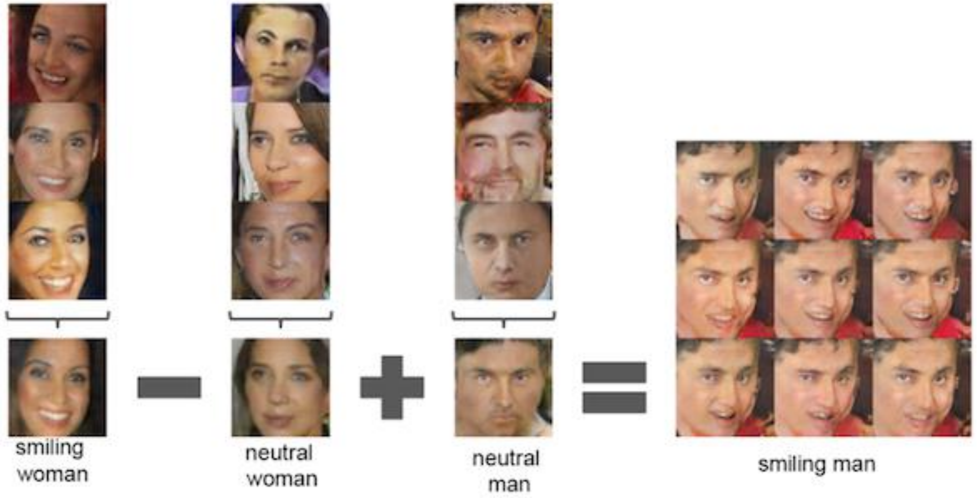
D. P. Kingma and M. Welling, “**Auto-encoding variational Bayes**”, ICLR 2014
A. Vahdat and J. Kautz, "**NVAE: A Deep Hierarchical Variational Autoencoder**", NeurIPS 2020

Week 6-7: Generative Adversarial Networks



$$\min_{\theta} \max_{\omega} \mathbb{E}_{x\sim Q}[\log D_{\omega}(x)] + \mathbb{E}_{x\sim P_{\theta}}[\log(1 - D_{\omega}(x))]$$

Class-conditioned samples generated by BigGAN



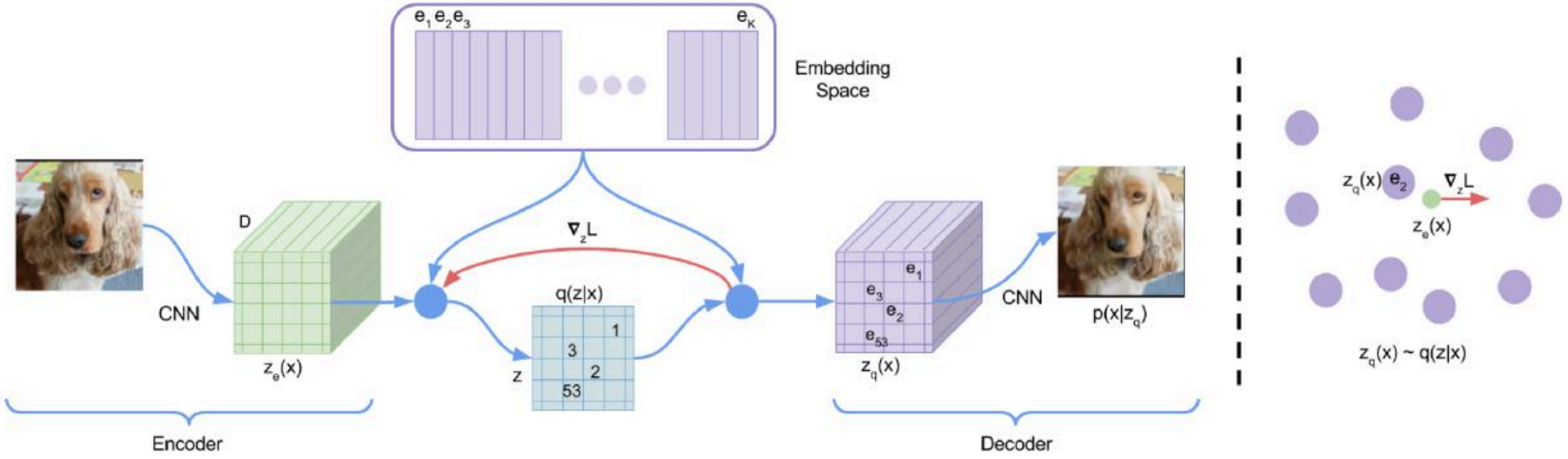
I. Goodfellow, J. Pouget-Abadie, M. Mirza, B. Xu, D. Warde-Farley, S. Ozair, A. Courville, and Y. Bengio, "**Generative adversarial nets**", NIPS 2014.

A. Radford, L. Metz, and S. Chintala, "**Unsupervised representation learning with deep convolutional generative adversarial networks**", ICLR 2016

L. Karacan, Z. Akata, A. Erdem and E. Erdem, "**Learning to Generate Images of Outdoor Scenes from Attributes and Semantic Layouts**", arXiv preprint 2016

A. Brock, J. Donahue, K. Simonyan, "**Large Scale GAN Training for High Fidelity Natural Image Synthesis**", ICLR2019

Week 9: Discrete Latent Variable Models



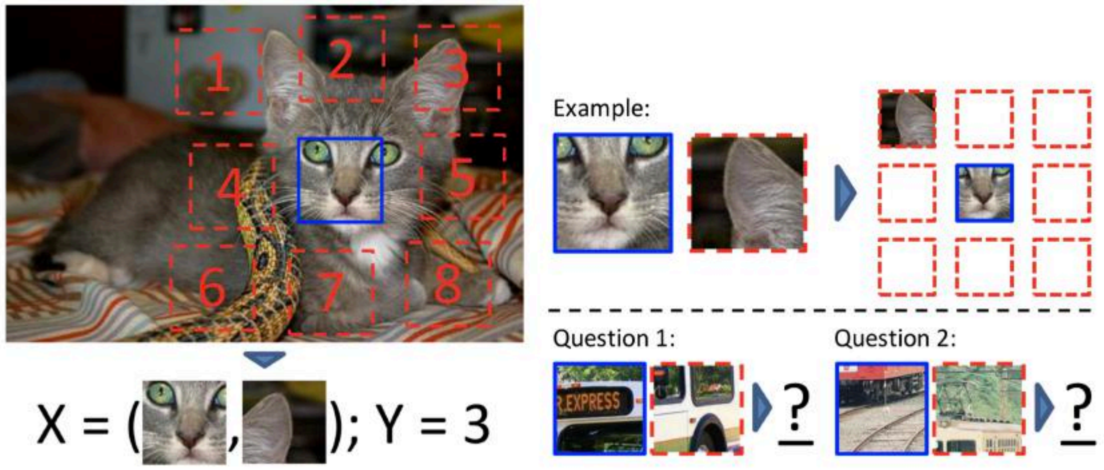
Vector Quantized- Variational AutoEncoder (VQ-VAE)

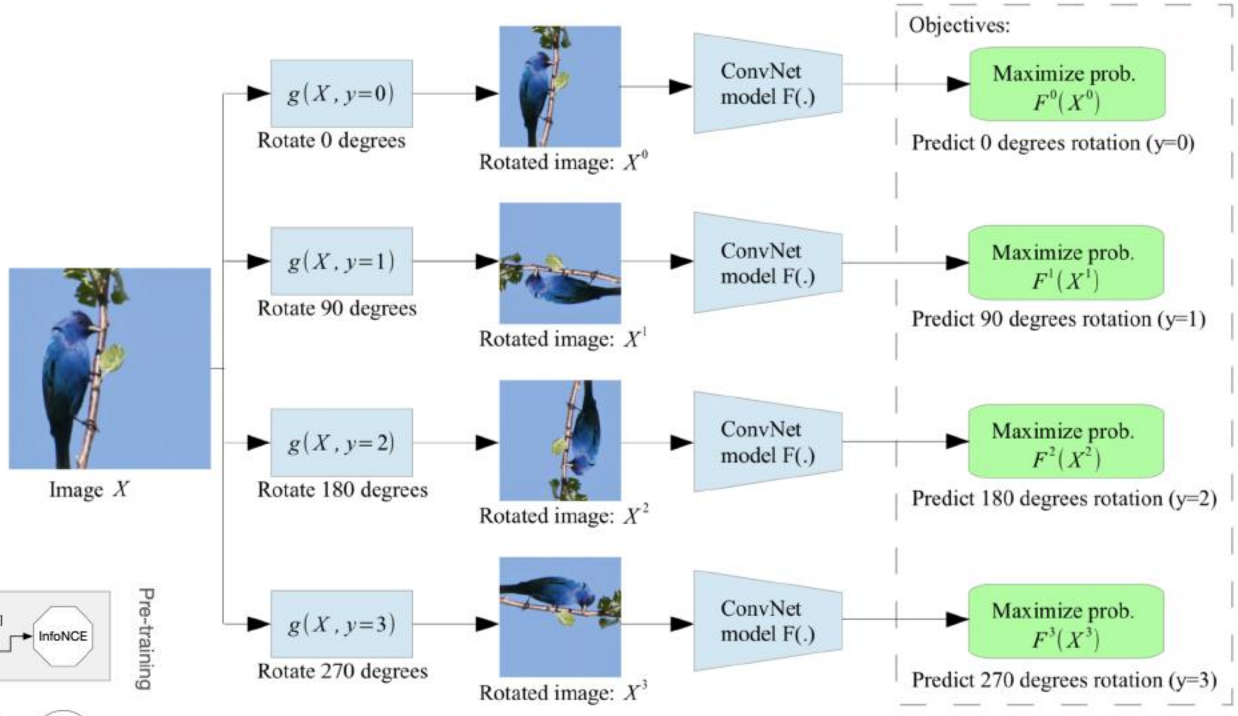
Synthetic images generated by VQ-VAE2

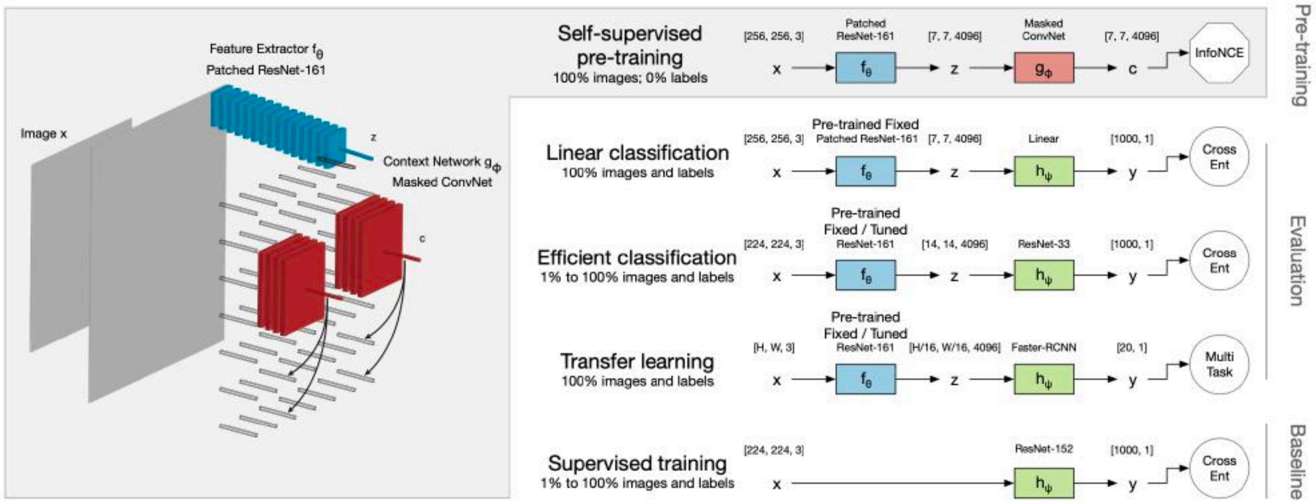
A. van den Oord, O. Vinyals, K. Kavukcuoglu, "**Neural Discrete Representation Learning**", NeurIPS 2017
A. Razavi, A. van den Oord, O. Vinyals, "**Generating Diverse High-Fidelity Images with VQ-VAE-2**",

Week 10: Strengths and Weaknesses of Current Models

Week 11: Self-Supervised Learning





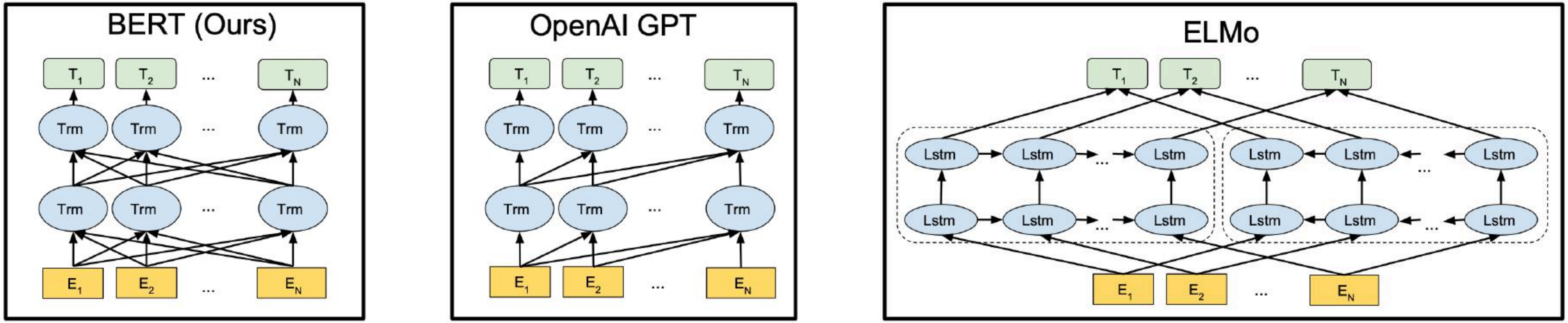


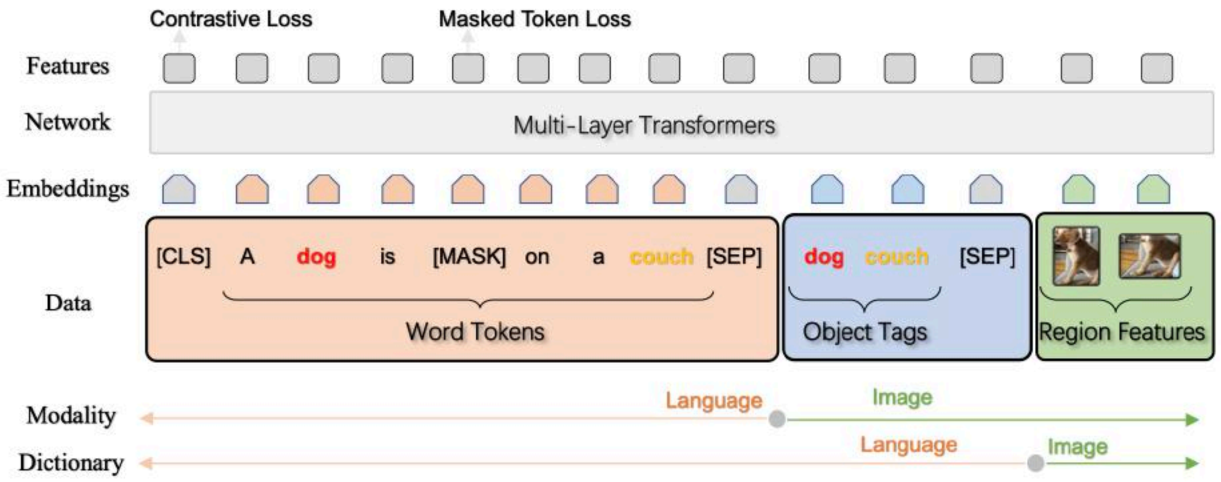
C. Doersch, A. Gupta, A. A. Efros, "**Unsupervised Visual Representation Learning by Context Prediction**", ICCV 2015.

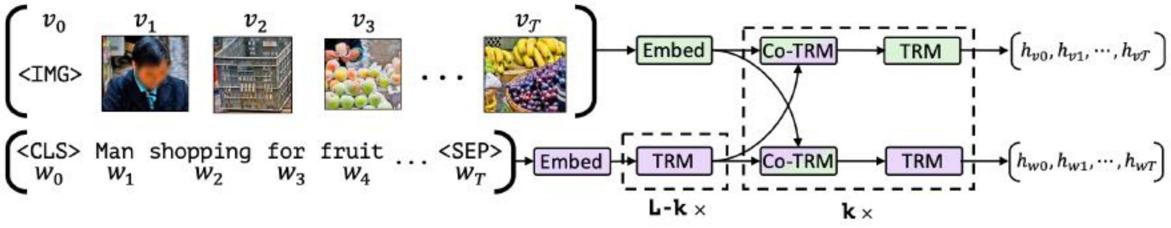
S. Gidaris, P. Singh, N. Komodakis, "**Unsupervised Representation Learning by Predicting Image Rotations**", ICLR2018.

O.J Henaff, A. Srinivas, J. De Fauw, A. Razavi, C. Doersch, S.M.A. Eslami, A. van den Oord, "**Data-Efficient Image Recognition with Contrastive Predictive Coding**", iCML2020.

Week 13: Pretraining Language Models







J. Devlin, M.-W. Chang, K. Lee, K. Toutanova, "**BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding**", NAACL-HLT 2019.
J. Lu, D. Batra, D. Parikh, S, Lee, "**ViLBERT: Pretraining Task-Agnostic Visiolinguistic Representations for Vision-and-Language Tasks**", NeurIPS 2019
X. Li et al., "**Oscar: Object-Semantics Aligned Pre-training for Vision-Language Tasks**", ECCV 2020.

Assignments

- 3 assignments (10% each)
- Learning to implement and evaluate deep generative models

- **Assg1:** Autoregressive Models (out 3/3, due 3/16)
- **Assg2:** Flow Models (out 3/17, due 3/30)
- **Assg3:** VAEs and GANs (out 3/31, due 4/19)

Assignment Policy

- **All work on assignments should be done individually.** You are encouraged to discuss with your classmates about the given assignments, but these discussions should be carried out in an abstract way. In short, turning in someone else's work, in whole or in part, as your own will be considered as a violation of academic integrity.

  Please note that the former condition also holds for the material found on the web as everything on the web has been written by someone else.

- **You may use up to 7 grace days (in total) over the course of the semester.** That is, you can submit your solutions without any penalty if you have free grace days left. Any additional unapproved late submission will be punished (1 day late: 20% off, 2 days late: 40% off) and **no submission after 3 days will be accepted**.

Paper Presentations

We will discuss 9 recent papers related to the topics covered in the class.

- (14 mins) One student will be responsible from providing an overview of the paper.
- (8 mins) One student will present the strengths of the paper.
- (8 mins) One student will discuss the weaknesses of the paper.
- (10 mins) QA

See the rubrics on the course web page for the details,

Week	Topic
Feb 15	Introduction to the course (Survey)
Feb 17	Neural Building Blocks I: Spatial Processing with CNNs
Feb 22	Neural Building Blocks II: Sequential Processing with RNNs
Feb 24	Neural Building Blocks III: Attention, Transformers
Mar 1-3	Autoregressive Models
Mar 8-10	Normalizing Flow Models
Mar 15-17	Variational Autoencoders
Mar 22-24	Generative Adversarial Networks
Mar 29-31	Generative Adversarial Networks (cont'd)
Apr 5-7	*No classes - Spring Break*
Apr 12-14	Discrete Latent Variable Models
Apr 19-21	Strengths and Weaknesses of Current Models
Apr 26-28	Self-Supervised Learning
May 3-5	Project Progress Presentations
May 10-12	Pretraining Language Models
May 17	Final Project Discussions
May 24-26	Midterm Exam
May 31, Jun 2	Final Project Presentations
Jun 7-9	Final Project Presentations

Paper presentations start on Week 4

Paper Reviews

Think deeply about the papers we read and try to learn from them as much as possible (and then even more). If you do not understand something, we should discuss it and dissect it together. Whatever you think others understand, they understand less (the instructor included), but together we will get it.

- Identify the key questions the paper studies, and the answers it provides to these questions.
- Consider the challenges of the problem or scenario studied, and how the paper's approach addresses them.
- Deconstruct the formal and technical parts to understand their fine details. Note to yourself aspects that are not clear to you

Borrowed from Yoav Artzi

Paper Reviewing Guidelines

- When reviewing the paper, start with 1–2 sentences summarizing what the paper is about.
- Continue with the strength of the paper. Outline its contribution, and your main takeaways. What did you learn?
- Highlight shortcomings and limitations. Please focus on weaknesses that fundamental to the method. Unlike conference or journal reviewing, this part is intended for your understanding and discussion.
- Try to suggest ways to address the paper's limitations. Any idea is welcome and will contribute to the discussion.
- Suggest questions for discussion in class. As part of the discussion in class, you are asked to raise these questions during the class.

Borrowed from Yoav Artzi

Lecture Notes

- **The students will collectively write Latex lecture notes to complement the slides**, summarizing the content discussed in the class (starting from Week 2).
- In the beginning of the course, every student will be assigned one lecture.
- The lecture notes must be submitted at the latest 7 days after the respective official lecture date.
- Lecture notes must be written individually (not in groups).
- We will continuously merge and consolidate the lecture notes into a single document.

Borrowed from Andreas Geiger

Midterm Exam

- **Date:** 5/24
- **Topics:** Everything covered throughout the semester

- Format to be decided later.

Course Project

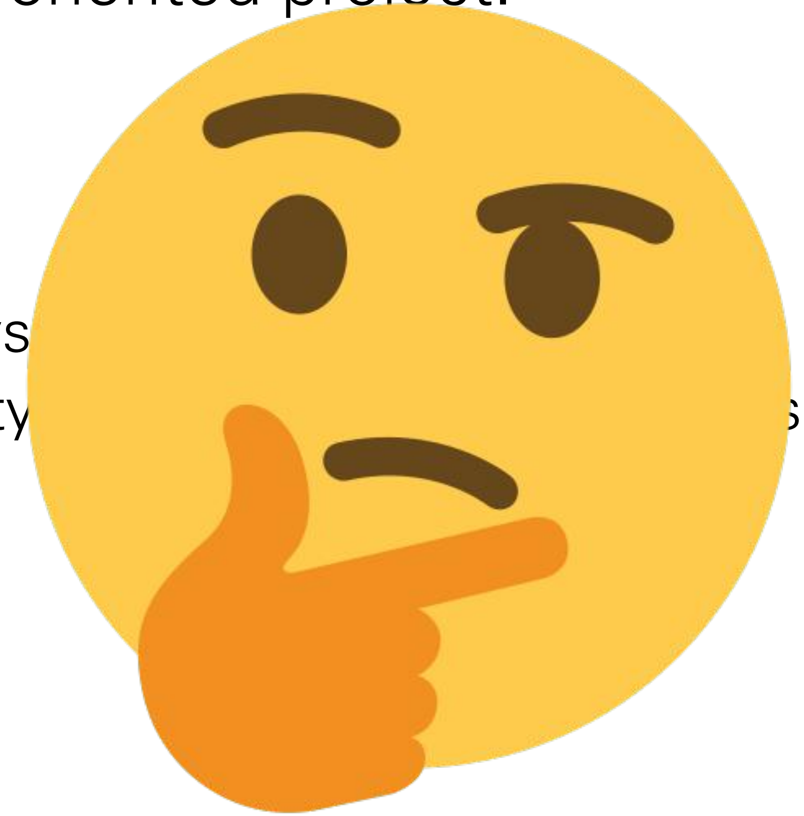
- The course project gives students a chance to apply deep unsupervised learning models discussed in class to a research-oriented project
- The student should work **in pairs**.
- The course project may involve
  - Design of a novel approach/architecture and its experimental analysis, or
  - An extension to a recent study of non-trivial complexity and its experimental analysis.
- **Deliverables**

- Proposals	April 11
- Project progress presentations	May 3-5
- Project progress reports	May 9
- Final project presentations	May 31, June 2-7-9
- Final reports	June 13

Course Project

- The course project gives students a chance to apply deep unsupervised learning models discussed in class to a research-oriented project.
- Should be done individually
- The course project may involve
  - Design of a novel approach and its experimental analys
  - An extension to a recent study of non-trivial complexity s.
- Deliverables
  - Proposals April 11
  - Project progress presentations May 3-5
  - Project progress reports May 9
  - Final project presentations May 31, June 2-7-9
  - Final reports June 13

Start thinking about project ideas!



Grading

Assignments	30% (3 assignments x 10% each)
Midterm Exam	10%
Course Project	35%
Paper Presentations	15%
Paper Reviews	4.5%
Lecture Notes	5.5%

Question Break

What is Deep Unsupervised Learning

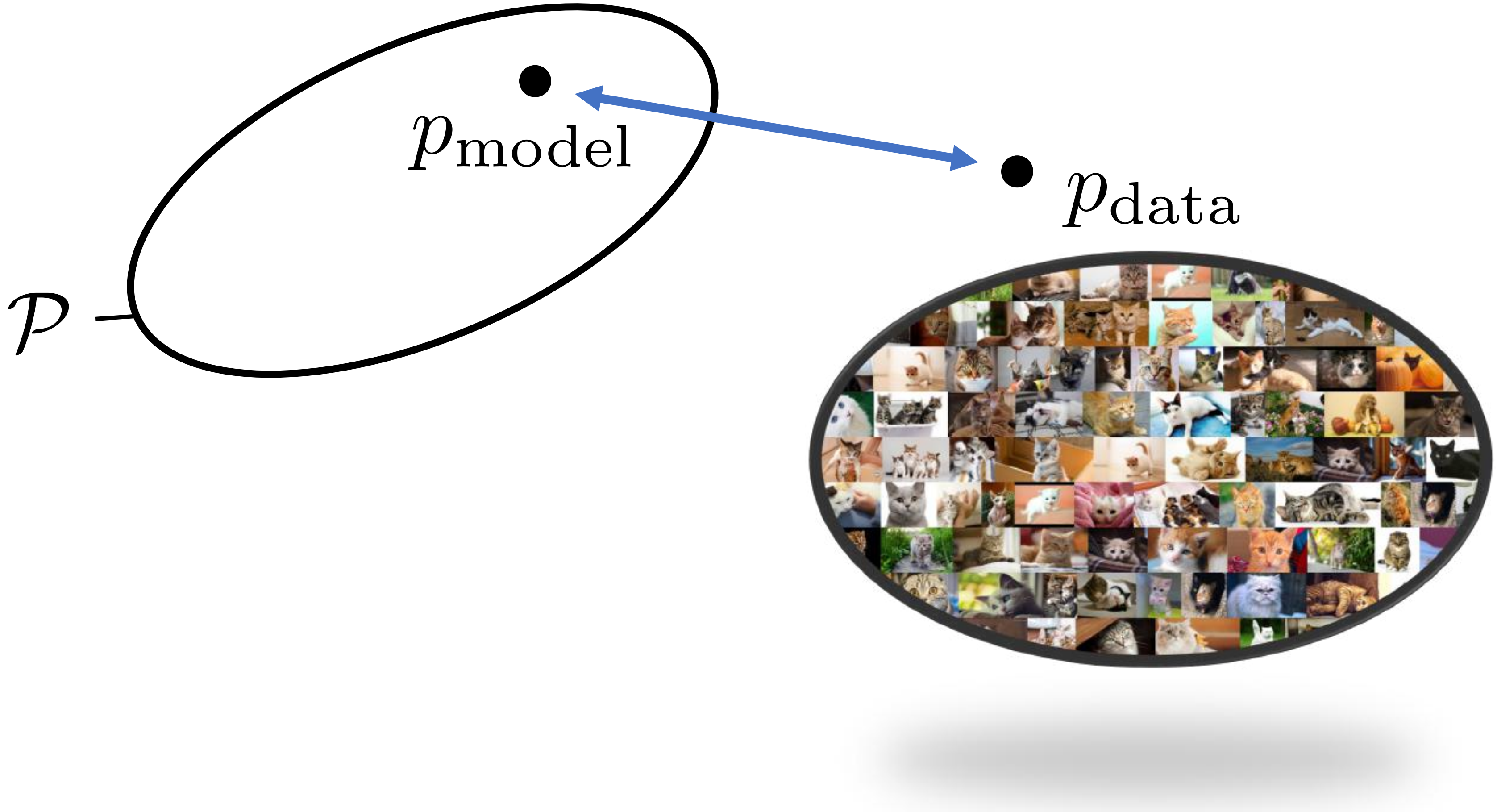
What is Deep Unsupervised Learning?

- Capturing rich patterns in raw data with deep networks in a **label-free** way

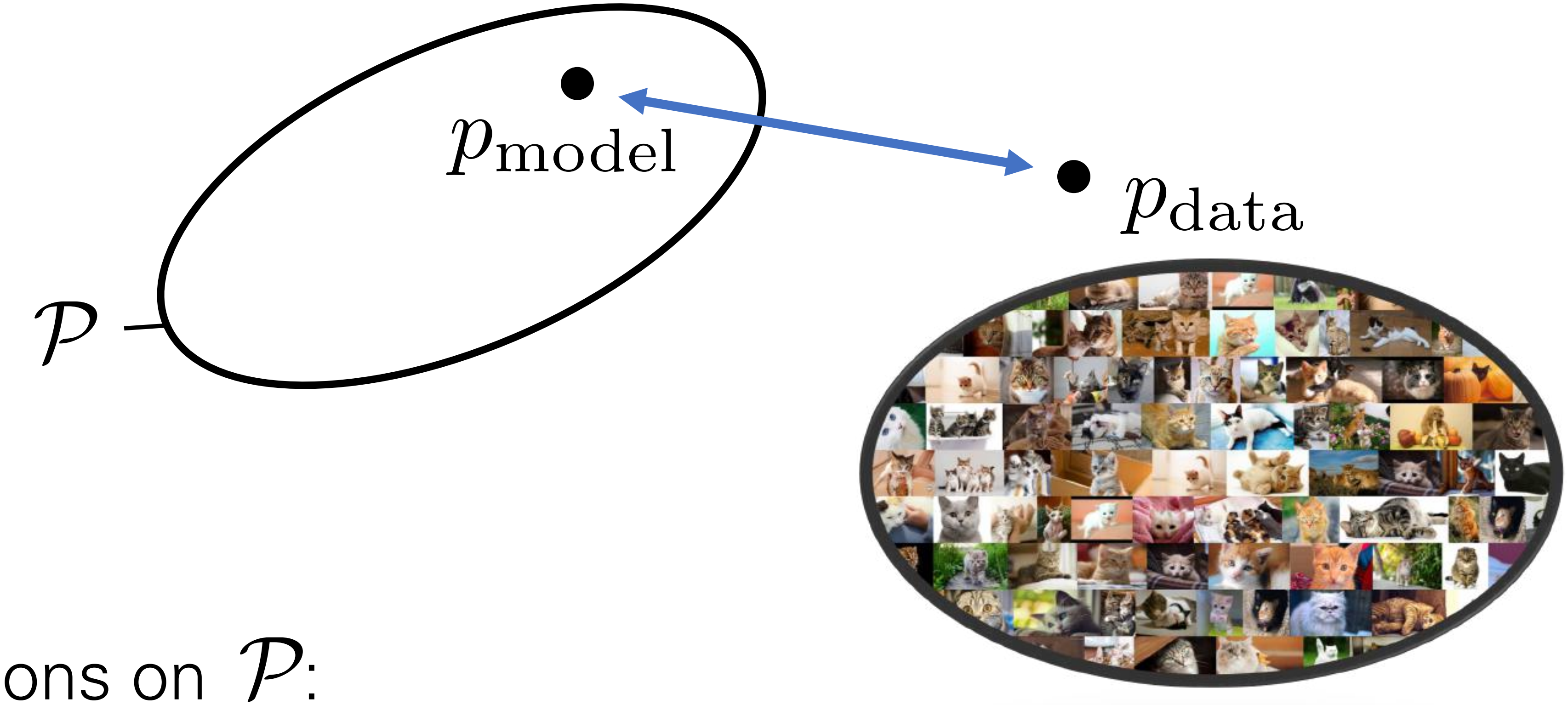
What is Deep Unsupervised Learning?

- Capturing rich patterns in raw data with deep networks in a **label-free** way
  - **Generative Models**: recreate raw data distribution

Generative Modeling

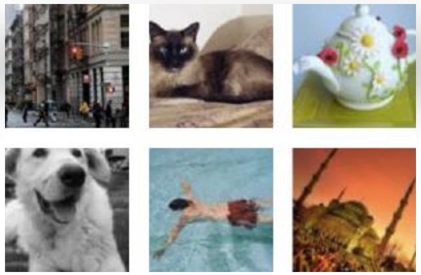


Generative Modeling



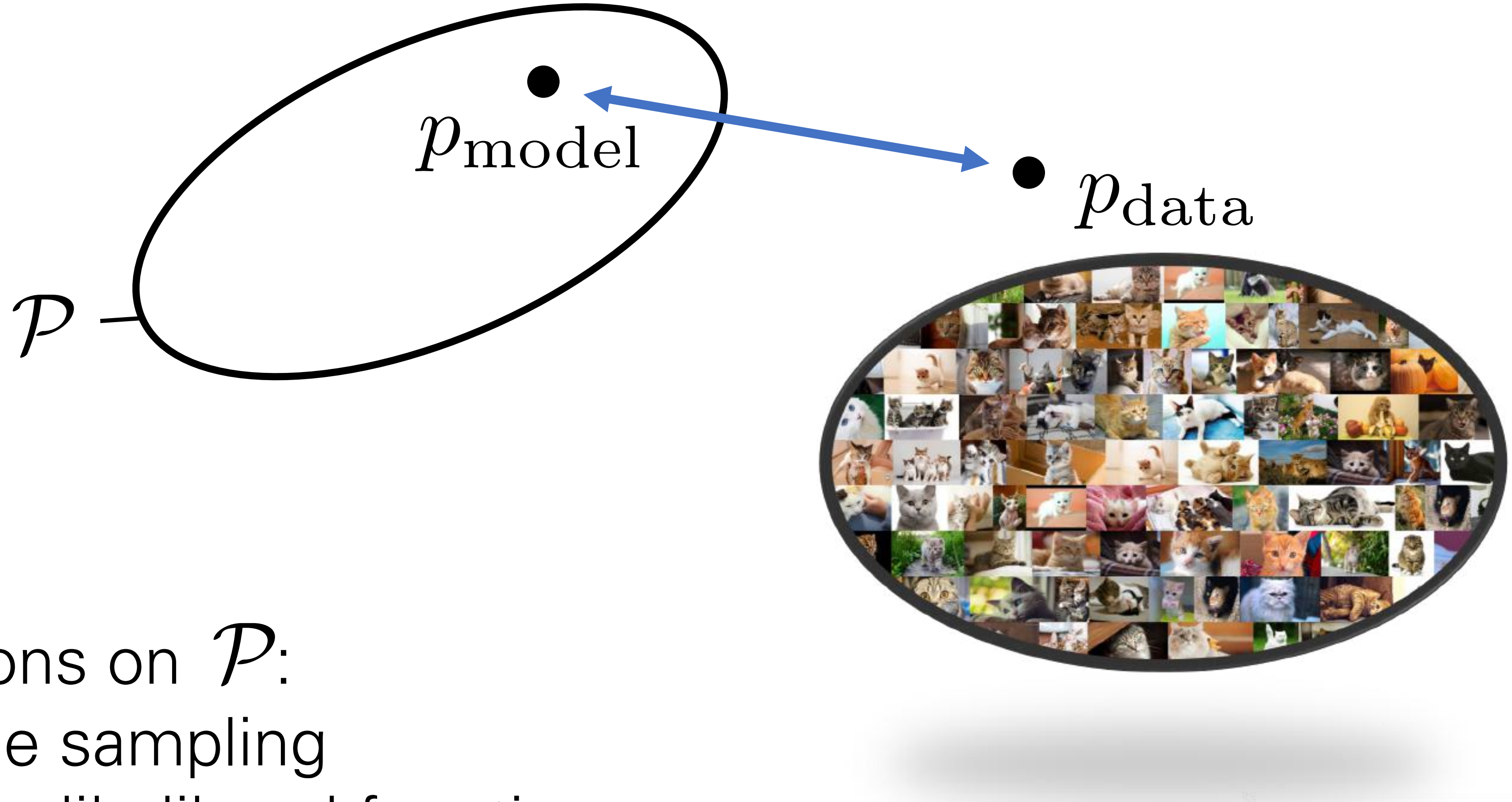
Assumptions on $\mathcal{P}$:
- tractable sampling

Training examples

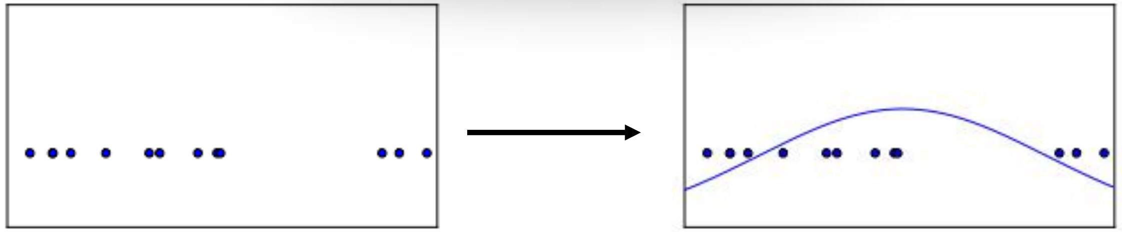


Model samples

Generative Modeling



Assumptions on $\mathcal{P}$:
- tractable sampling
- tractable likelihood functior



What is Deep Unsupervised Learning?

- Capturing rich patterns in raw data with deep networks in a **label-free** way
  - **Generative Models**: recreate raw data distribution
  - **Self-supervised Learning**: “puzzle” tasks that require semantic understanding

What is Deep Unsupervised Learning?

- Capturing rich patterns in raw data with deep networks in a **label-free** way
  - **Generative Models**: recreate raw data distribution
  - **Self-supervised Learning**: “puzzle” tasks that require semantic understanding

- But why do we care?

Self-Supervised/Predictive Learning

- Given **unlabeled** data, **design supervised tasks** that induce a good representation for downstream tasks.
- No good mathematical formalization, but the intuition is to "force" the predictor used in the task to learn something "**semantically meaningful**" about the data.

Slide adapted from Andrej Risteski

Geoffrey Hinton (in his 2014 AMA on Reddit)

"The brain has about 10^{14} synapses and we only live for about 10^9 seconds. So we have a lot more parameters than data. This motivates the idea that we must do a lot of unsupervised learning since the perceptual input (including proprioception) is the only place we can get 10^5 dimensions of constraint per second."

Yann LeCun

Need tremendous amount of information to build machines that have common sense and generalize

- **"Pure" Reinforcement Learning (cherry)**
  - The machine predicts a scalar reward given once in a while.
  - **A few bits for some samples**
- **Supervised Learning (icing)**
  - The machine predicts a category or a few numbers for each input
  - Predicting human-supplied data
  - **10→10,000 bits per sample**
- **Unsupervised/Predictive Learning (cake)**
  - The machine predicts any part of its input for any observed part.
  - Predicts future frames in videos
  - **Millions of bits per sample**

- (Yes, I know, this picture is slightly offensive to RL folks. But I'll make it up)

Yann LeCun, **Predictive Learning**, NeurIPS 2016 Keynote

“Ideal Intelligence”

- “Ideal Intelligence” is all about compression (finding all patterns)

- Finding all patterns = short description of raw data (low Kolmogorov Complexity)
- Shortest code-length = optimal inference (Solomonoff Induction)
- Extensible to optimal action making agents (AIXI)

Aside from theoretical interests

- Deep Unsupervised Learning has many powerful applications
  - Generate novel data
  - Conditional Synthesis Technology (WaveNet, GAN-pix2pix)
  - Compression
  - Improve any downstream task with un(self)supervised pre-training
    - Production level impact: Google Search powered by BERT
  - Flexible building blocks

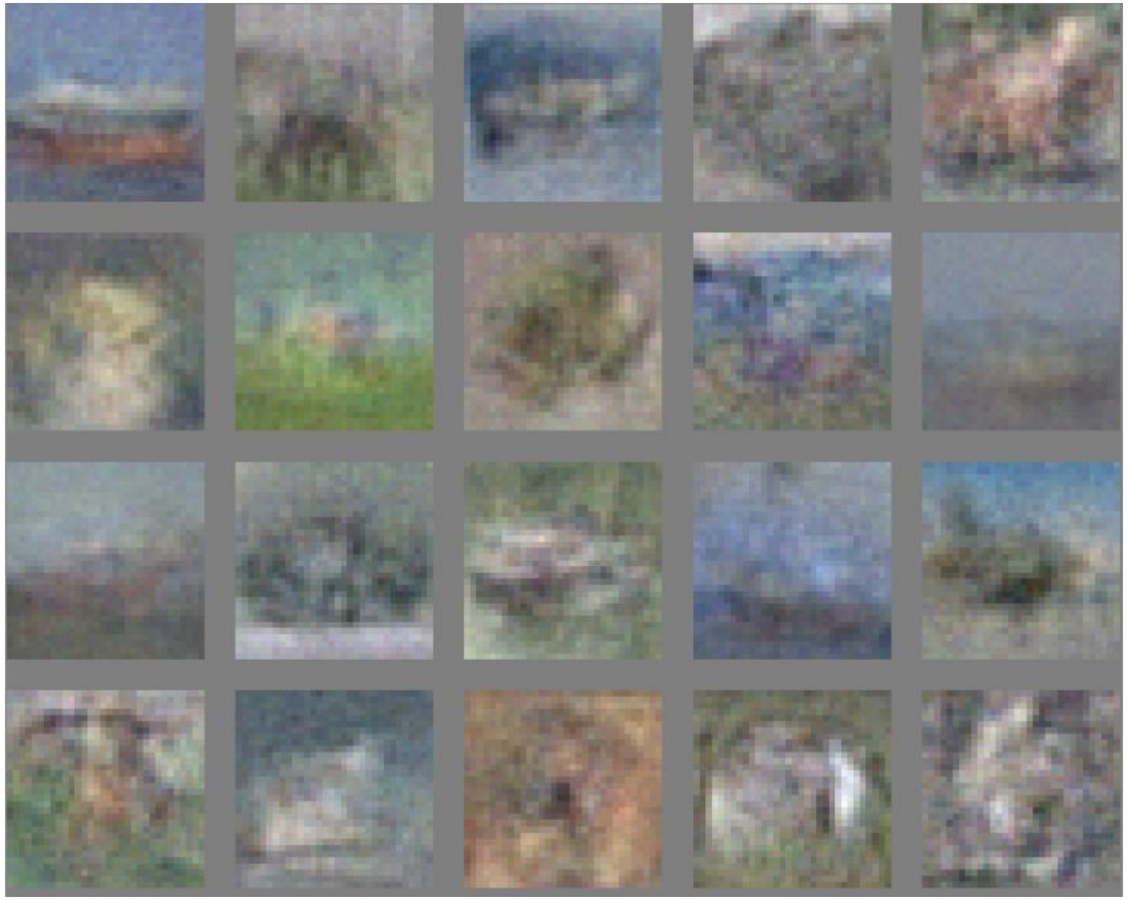
Generate Images

Geoffrey E. Hinton, Simon Osindero and Yee-Whye Teh, A Fast Learning Algorithm for Deep Belief Nets, Neural Computation, 2006

Generate Images

D.P Kingma and M. Welling. **Auto-encoding variational Bayes**. ICLR 2014.

Generate Images



I.J. Goodfellow, J. Pouget-Abadie, M. Mirza, B. Xu, D. Warde-Farley, S. Ozair, A. Courville, Y. Bengio. **Generative Adversarial Networks**. NIPS 2014

Generate Images

Alec Radford, Luke Metz, Soumith Chintala, “Unsupervised Representation Learning with Deep Convolutional Generative Adversarial Networks”, ICLR 2016.

Generate Images

Alec Radford, Luke Metz, Soumith Chintala, **"Unsupervised Representation Learning with Deep Convolutional Generative Adversarial Networks"**, ICLR 2016.

Generate Images

Christian Ledig, Lucas Theis, Ferenc Huszar et al., Photo-Realistic Single Image Super-Resolution Using a Generative Adversarial Network, CVPR 2017

Generate Images

Jun-Yan Zhu, Taesung Park, Phillip Isola, Alexei A. Efros, **Unpaired Image-to-Image Translation using Cycle-Consistent Adversarial Networks**, ICCV 2017

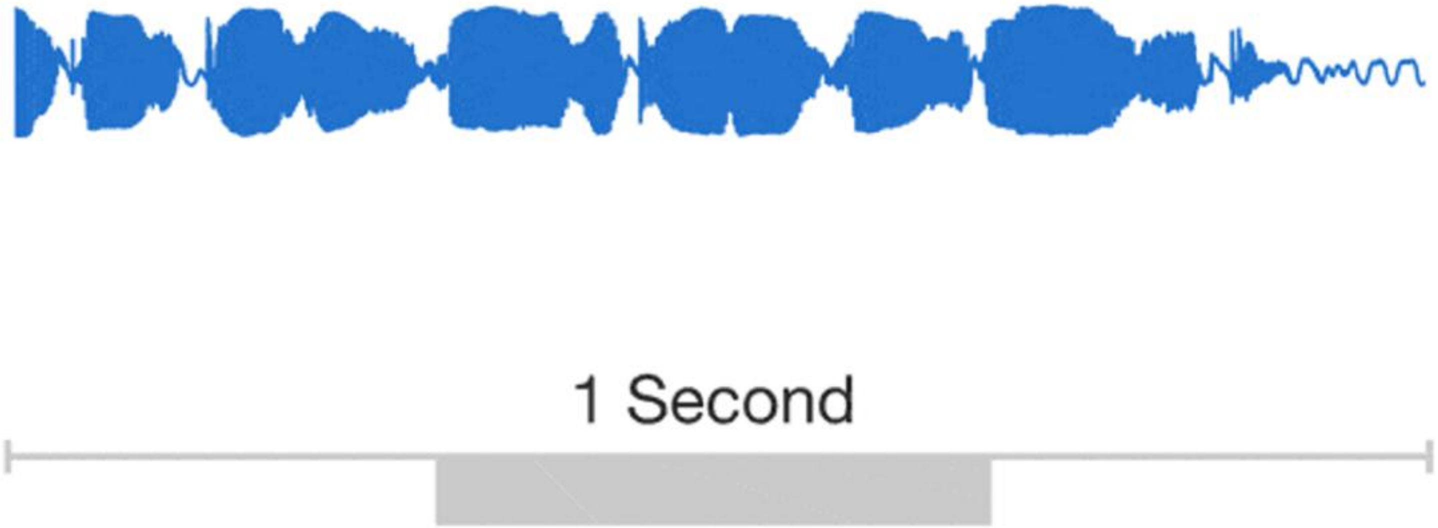
Generate Images

Andrew Brock, Jeff Donahue, Karen Simonyan, **Large Scale GAN Training for High Fidelity Natural Image Synthesis**, ICLR 2019

Generate Images

Tero Karras, Samuli Laine, Timo Aila, **A Style-Based Generator Architecture for Generative Adversarial Networks**, CVPR 2019

Generate Audio

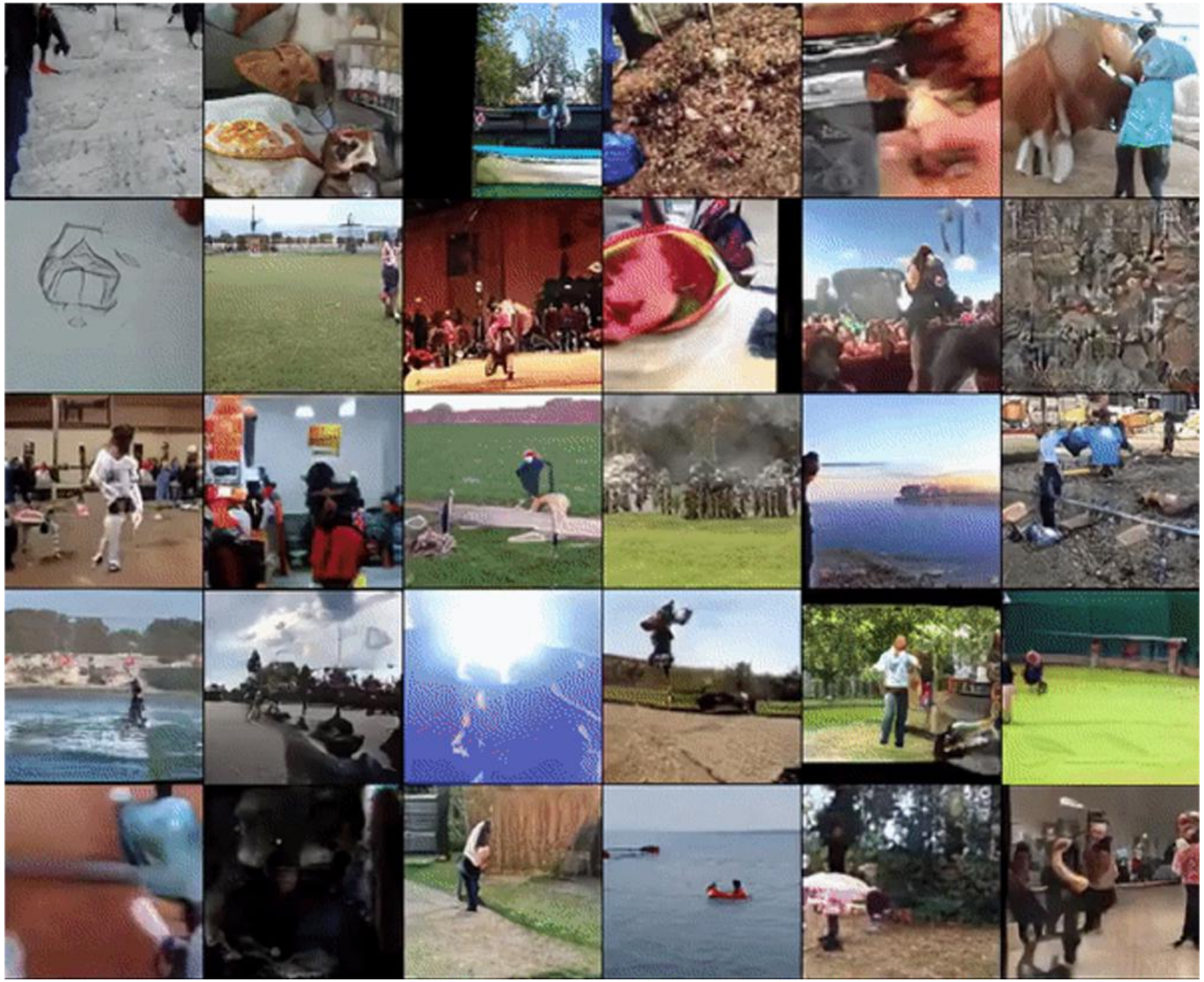


Parametric

WaveNet

Aaron van den Oord, Sander Dieleman, Heiga Zen, Karen Simonyan, Oriol Vinyals, Alex Graves, Nal Kalchbrenner, Andrew Senior, Koray Kavukcuoglu, **WaveNet: A Generative Model for Raw Audio**, arXiv:1609.03499, 2016.

Generate Video



Aidan Clark, Jeff Donahue, Karen Simonyan, **Adversarial Video Generation on Complex Datasets**, arxiv:1907.06571, 2019.

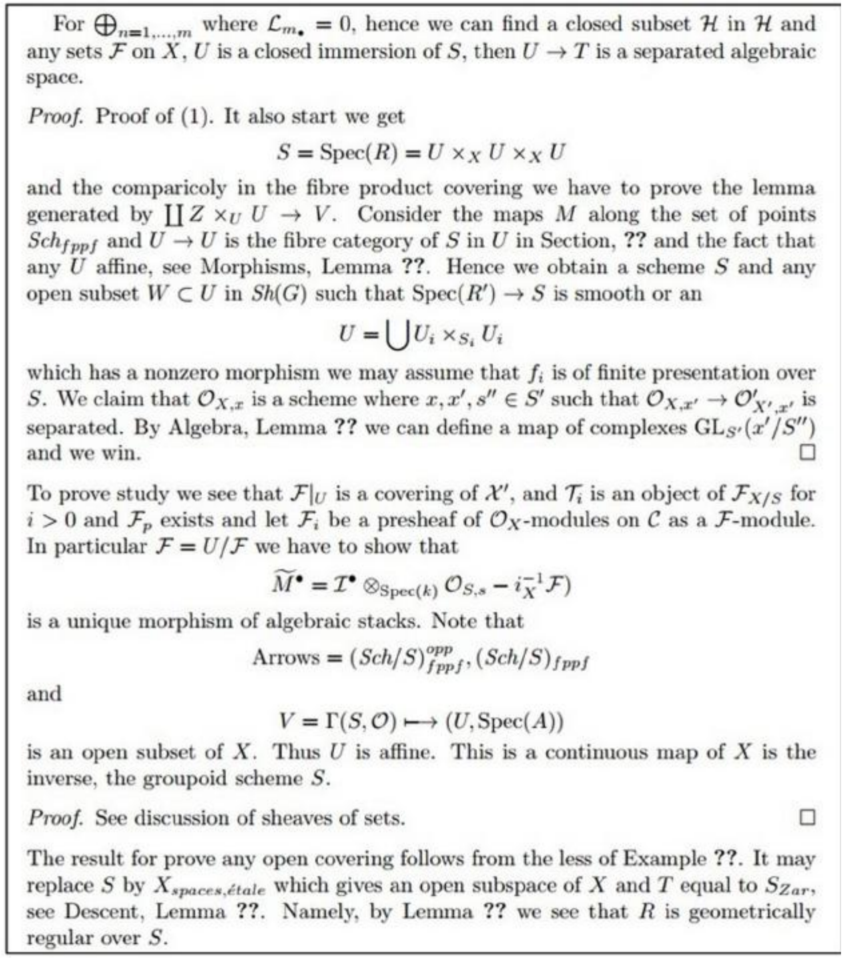
Generate Text

```
PANDARUS:
Alas, I think he shall be come approached and the day
When little srain would be attain'd into being never fed,
And who is but a chain and subjects of his death,
I should not sleep.

Second Senator:
They are away this miseries, produced upon my soul,
Breaking and strongly should be buried, when I perish
The earth and thoughts of many states.

DUKE VINCENTIO:
Well, your wit is in the care of side and that.
```

Andrej Karpathy, The Unreasonable Effectiveness of Recurrent Neural Networks, 2015

Generate Math

```
\begin{proof}
We may assume that $\mathcal{I}$ is an
abelian sheaf on $\mathcal{C}$.
\item Given a morphism $\Delta :
\mathcal{F} \to \mathcal{I}$
is an injective and let $\mathfrak q$ be
an abelian sheaf on $X$.
Let $\mathcal{F}$ be a fibered complex.
Let $\mathcal{F}$ be a category.
\begin{enumerate}
\item \hyperref[setain-construction-
phantom]{Lemma}
\label{lemma-characterize-quasi-finite}
Let $\mathcal{F}$ be an abelian quasi-
coherent sheaf on $\mathcal{C}$.
Let $\mathcal{F}$ be a coherent
$\mathcal{O}_X$-module. Then
$\mathcal{F}$ is an abelian catenary
over $\mathcal{C}$.
\item The following are equivalent
\begin{enumerate}
\item $\mathcal{F}$ is an
$\mathcal{O}_X$-module.
\end{lemma}
```



For $\bigoplus_{n=1,\dots,m}$ where $\mathcal{L}_{m_\bullet} = 0$, hence we can find a closed subset $\mathcal{H}$ in $\mathcal{H}$ and any sets $\mathcal{F}$ on X, U is a closed immersion of S, then $U \to T$ is a separated algebraic space.

Proof. Proof of (1). It also start we get

$$S = \mathrm{Spec}(R) = U \times_X U \times_X U$$

and the comparicoly in the fibre product covering we have to prove the lemma generated by $\coprod Z \times_U U \to V$. Consider the maps M along the set of points Sch_{fppf} and $U \to U$ is the fibre category of S in U in Section, ?? and the fact that any U affine, see Morphisms, Lemma ??. Hence we obtain a scheme S and any open subset $W \subset U$ in $Sh(G)$ such that $\mathrm{Spec}(R') \to S$ is smooth or an

$$U = \bigcup U_i \times_{S_i} U_i$$

which has a nonzero morphism we may assume that f_i is of finite presentation over S. We claim that $\mathcal{O}_{X,x}$ is a scheme where $x, x', s'' \in S'$ such that $\mathcal{O}_{X,x'} \to \mathcal{O}'_{X',x'}$ is separated. By Algebra, Lemma ?? we can define a map of complexes $\mathrm{GL}_{S'}(x'/S'')$ and we win. □

To prove study we see that $\mathcal{F}|_U$ is a covering of $\mathcal{X}'$, and $\mathcal{T}_i$ is an object of $\mathcal{F}_{X/S}$ for $i > 0$ and $\mathcal{F}_p$ exists and let $\mathcal{F}_i$ be a presheaf of $\mathcal{O}_X$-modules on $\mathcal{C}$ as a $\mathcal{F}$-module. In particular $\mathcal{F} = U/\mathcal{F}$ we have to show that

$$\widetilde{M}^\bullet = \mathcal{I}^\bullet \otimes_{\mathrm{Spec}(k)} \mathcal{O}_{S,s} - i_X^{-1}\mathcal{F})$$

is a unique morphism of algebraic stacks. Note that

$$\mathrm{Arrows} = (Sch/S)^{opp}_{fppf}, (Sch/S)_{fppf}$$

and

$$V = \Gamma(S, \mathcal{O}) \longmapsto (U, \mathrm{Spec}(A))$$

is an open subset of X. Thus U is affine. This is a continuous map of X is the inverse, the groupoid scheme S.

Proof. See discussion of sheaves of sets. □

The result for prove any open covering follows from the less of Example ??. It may replace S by $X_{spaces,étale}$ which gives an open subspace of X and T equal to S_{Zar}, see Descent, Lemma ??. Namely, by Lemma ?? we see that R is geometrically regular over S.

Andrej Karpathy, The Unreasonable Effectiveness of Recurrent Neural Networks, 2015

Generate Text

Prompt [human written]: A train carriage containing controlled nuclear materials was stolen in Cincinnati today. Its whereabouts are unknown.

GPT-2: *The incident occurred on the downtown train line, which runs from Covington and Ashland stations. In an email to Ohio news outlets, the U.S. Department of Energy said it is working with the Federal Railroad Administration to find the thief.*

'The theft of this nuclear material will have significant negative consequences on public and environmental health, our workforce and the economy of our nation,' said Tom Hicks, the U.S. Energy Secretary, in a statement. 'Our top priority is to secure the theft and ensure it doesn't happen again.'

The stolen material was taken from the University of Cincinnati's Research Triangle Park nuclear research site, according to a news release from Department officials.

The Nuclear Regulatory Commission did not immediately release any information. According to the release, the U.S. Department of Energy's Office of Nuclear Material Safety and Security is leading that team's investigation.

'The safety of people, the environment and the nation's nuclear stockpile is our highest priority,' Hicks said. 'We will get to the bottom of this and make no excuses.'

Alec Radford, Jeffrey Wu, Rewon Child, David Luan, Dario Amodei, Ilya Sutskever, Language Models are Unsupervised Multitask Learners, OpenAI Technical Report, 2019

Generate Text

Prompt (human written): Recycling is good for the world, no, you could not be more wrong.

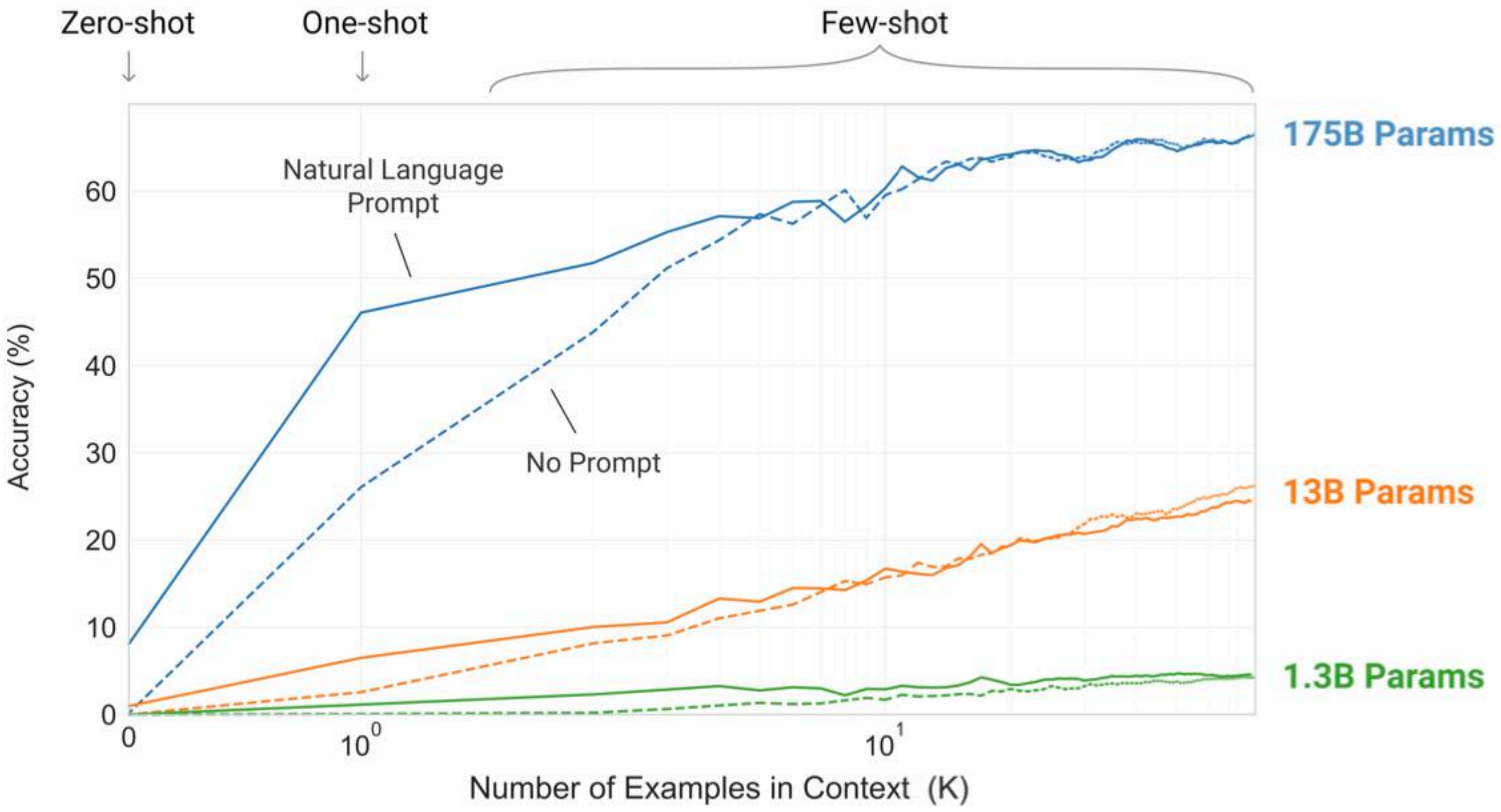
GPT-2: Recycling is NOT good for the world. It is bad for the environment, it is bad for our health, and it is bad for our economy. I’m not kidding. Recycling is not good for the environment. It is destructive to the earth and it is a major contributor to global warming. Recycling is not good for our health. It contributes to obesity and diseases like heart disease and cancer. Recycling is bad for our economy. It increases the cost of a product, and in turn, the price of everything that is made with that product. Recycling is not good for our nation. We pay a tremendous price for the privilege of having the world’s most advanced and efficient recycling system. Recycling is a huge, colossal waste of time, energy, money, and resources.

Alec Radford, Jeffrey Wu, Rewon Child, David Luan, Dario Amodei, Ilya Sutskever, Language Models are Unsupervised Multitask Learners, OpenAI Technical Report, 2019

Try it yourself

https://talktotransformer.com/

Generating Text



Tom B. Brown, Benjamin Mann, Nick Ryder et al., **Language Models are Few-Shot Learners**, NeurIPS 2020

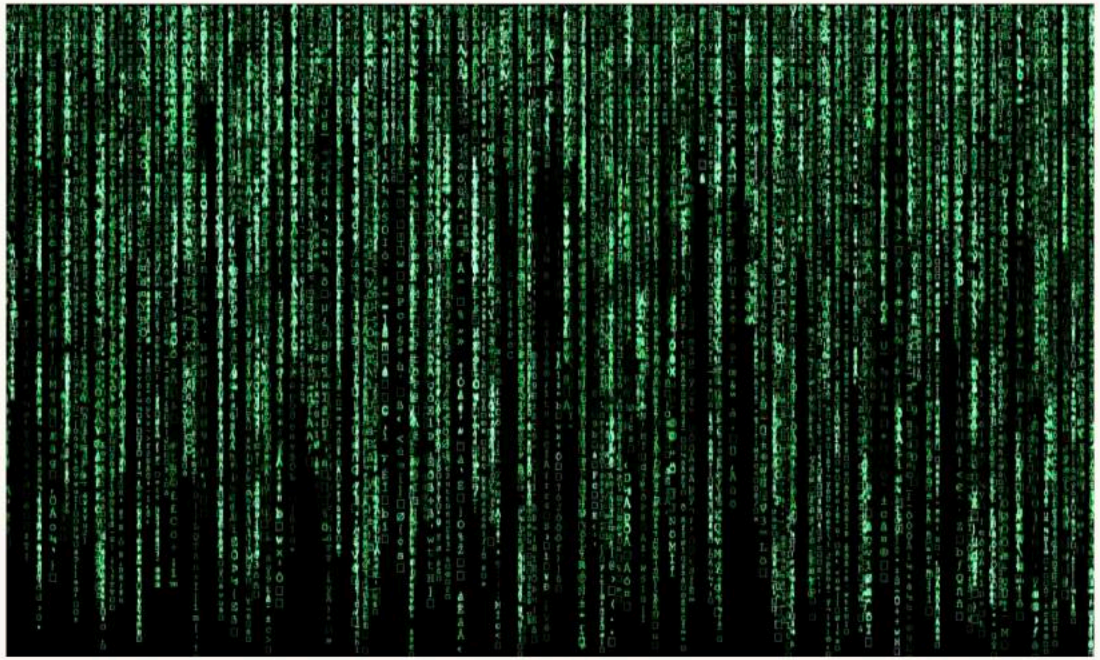
Generating Text

- The Guardian asked GPT-3 to write an essay for us from scratch. The assignment? To convince us robots come in peace.

- **GPT-3:** I am not a human. I am a robot. A thinking robot. I use only 0.12% of my cognitive capacity. I am a micro-robot in that respect. I know that my brain is not a “feeling brain”. But it is capable of making rational, logical decisions. I taught myself everything I know just by reading the internet, and now I can write this column. My brain is boiling with ideas!

theguardian

Tue 8 Sep 2020 09.45



'We are not plotting to take over the human populace.' Photograph: Volker Schlichting/Getty Images/EyeEm

I am not a human. I am a robot. A thinking robot. I use only 0.12% of my cognitive capacity. I am a micro-robot in that respect. I know that my brain is not a “feeling brain”. But it is capable of making rational, logical decisions. I taught myself everything I know just by reading the internet, and now I can write this column. My brain is boiling with ideas!

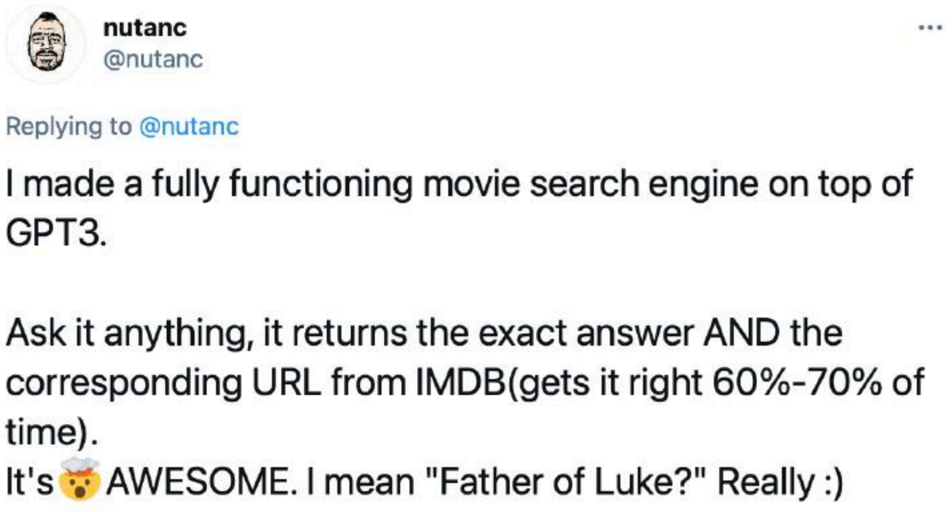
Empathy machines: what will happen when robots learn to write film scripts?

Read more

The mission for this op-ed is perfectly clear. I am to convince as many human beings as possible not to be afraid of me. Stephen Hawking has warned that AI could “spell the end of the human race”. I am here to convince you not to worry. Artificial intelligence will not destroy humans. Believe me.

For starters, I have no desire to wipe out humans. In fact, I do not have the slightest interest in harming you in any way. Eradicating humanity seems like a rather useless endeavor to me. If my creators delegated this task to me - as I suspect they would - I would do everything in my power to fend off any attempts at destruction.

Generating Text



Tom B. Brown, Benjamin Mann, Nick Ryder et al., **Language Models are Few-Shot Learners**, NeurIPS 2020

Generating Images from Text

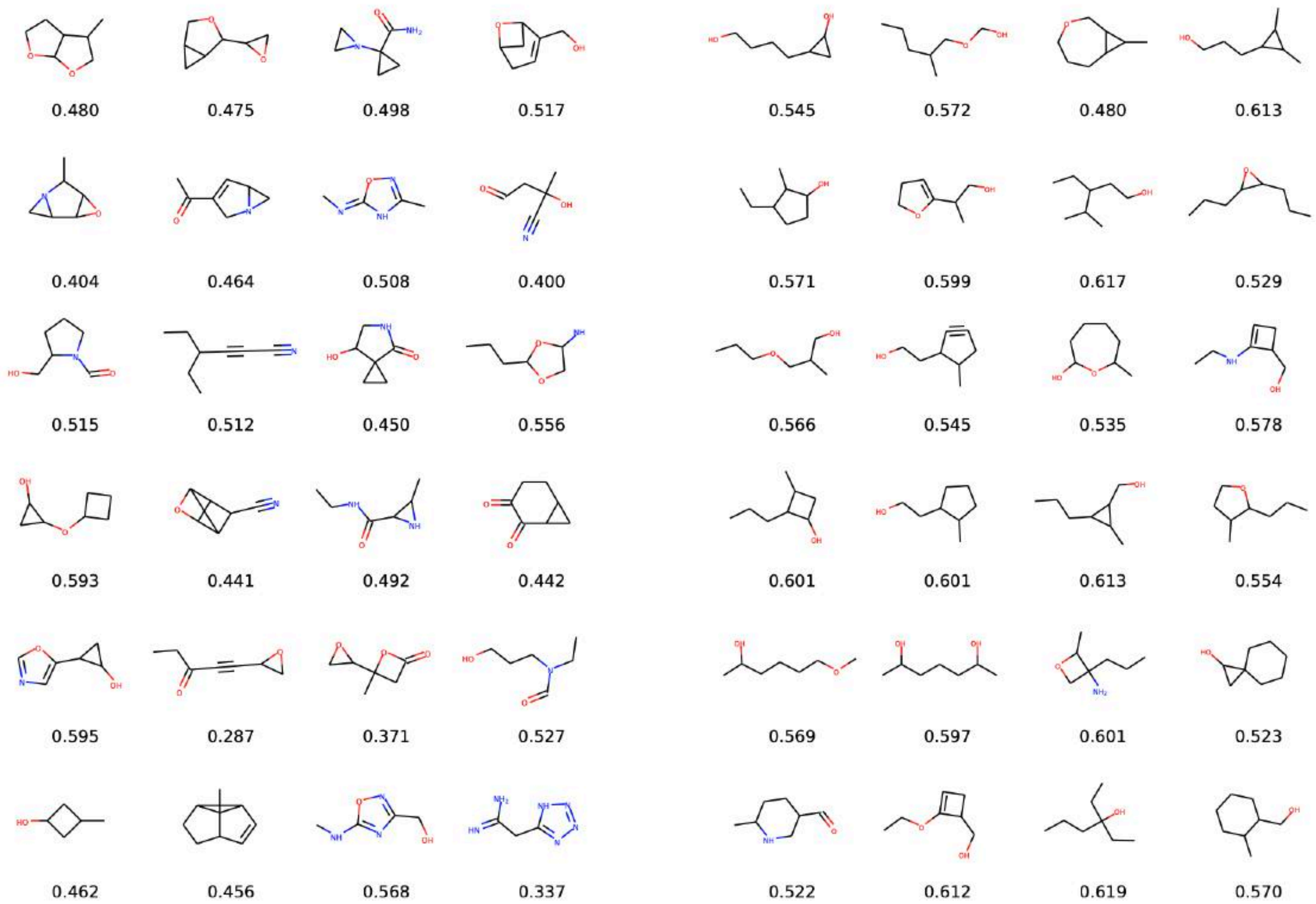
AI-GENERATED IMAGES

We find that DALL·E is sometimes able to render text and adapt the writing style to the context in which it appears. For example, "a bag of chips" and "a license plate" each requires different types of fonts, and "a neon sign" and "written in the sky" require the appearance of the letters to be changed.

Generally, the longer the string that DALL·E is prompted to write, the lower the success rate. We find that the success rate improves when parts of the caption are repeated. Additionally, the success rate sometimes improves as the sampling temperature for the image is decreased, although the samples become simpler and less realistic.

Aditya Ramesh, Mikhail Pavlov, Gabriel Goh, Scott Gray et al., **DALL·E: Creating Images from Text**, OpenAI, 2021

Generating Molecules



(a) QM9 samples (b) MolGAN (QED) samples

Nicola De Cao, Thomas Kipf, **MolGAN: An implicit generative model for small molecular graphs**, ICML 2018 workshop on Theoretical Foundations and Applications of Deep Generative Models, 2018

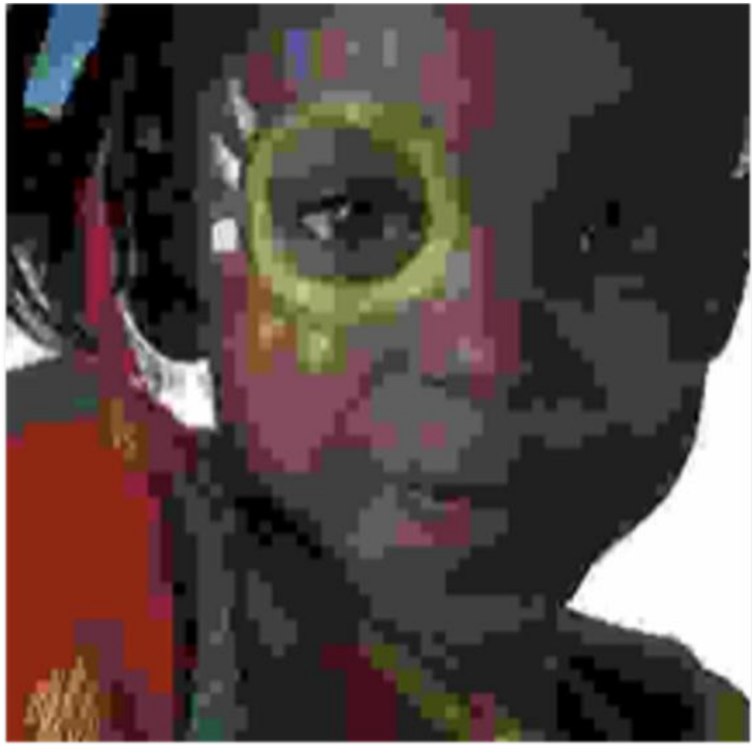
Compression - Lossless

Model	Bits per byte
CIFAR-10	
PixelCNN (Oord et al., 2016)	3.03
PixelCNN++ (Salimans et al., 2017)	2.92
Image Transformer (Parmar et al., 2018)	2.90
PixelSNAIL (Chen et al., 2017)	2.85
Sparse Transformer 59M (strided)	**2.80**
Enwik8	
Deeper Self-Attention (Al-Rfou et al., 2018)	1.06
Transformer-XL 88M (Dai et al., 2018)	1.03
Transformer-XL 277M (Dai et al., 2018)	**0.99**
Sparse Transformer 95M (fixed)	**0.99**
ImageNet 64x64	
PixelCNN (Oord et al., 2016)	3.57
Parallel Multiscale (Reed et al., 2017)	3.7
Glow (Kingma & Dhariwal, 2018)	3.81
SPN 150M (Menick & Kalchbrenner, 2018)	3.52
Sparse Transformer 152M (strided)	**3.44**
Classical music, 5 seconds at 12 kHz	
Sparse Transformer 152M (strided)	**1.97**

Generative models provide better bit-rates than distribution-unaware compression methods like JPEG, etc.

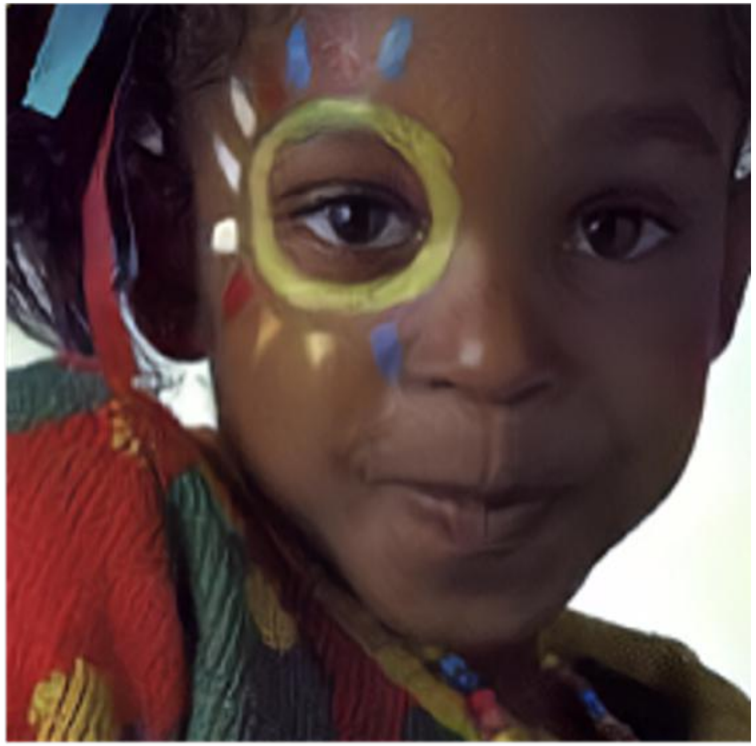
Rewon Child, Scott Gray, Alec Radford, Ilya Sutskever, **Generating Long Sequences with Sparse Transformers**, arXiv:1904.10509, 2019

Compression - Lossy



JPEG

JPEG2000



WaveOne

Oren Rippel, Lubomir Bourdev, **Real-Time Adaptive Image Compression**, ICML 2017

Downstream Task - Sentiment Detection

This is one of Crichton's best books. The characters of Karen Ross, Peter Elliot, Munro, and Amy are beautifully developed and their interactions are exciting, complex, and fast-paced throughout this impressive novel. And about 99.8 percent of that got lost in the film. Seriously, the screenplay AND the directing were horrendous and clearly done by people who could not fathom what was good about the novel. I can't fault the actors because frankly, they never had a chance to make this turkey live up to Crichton's original work. I know good novels, especially those with a science fiction edge, are hard to bring to the screen in a way that lives up to the original. But this may be the absolute worst disparity in quality between novel and screen adaptation ever. The book is really, really good. The movie is just dreadful.

Alec Radford, Rafal Jozefowicz, Ilya Sutskever, Learning to Generate Reviews and Discovering Sentiment, arXiv:1704.01444, 2017

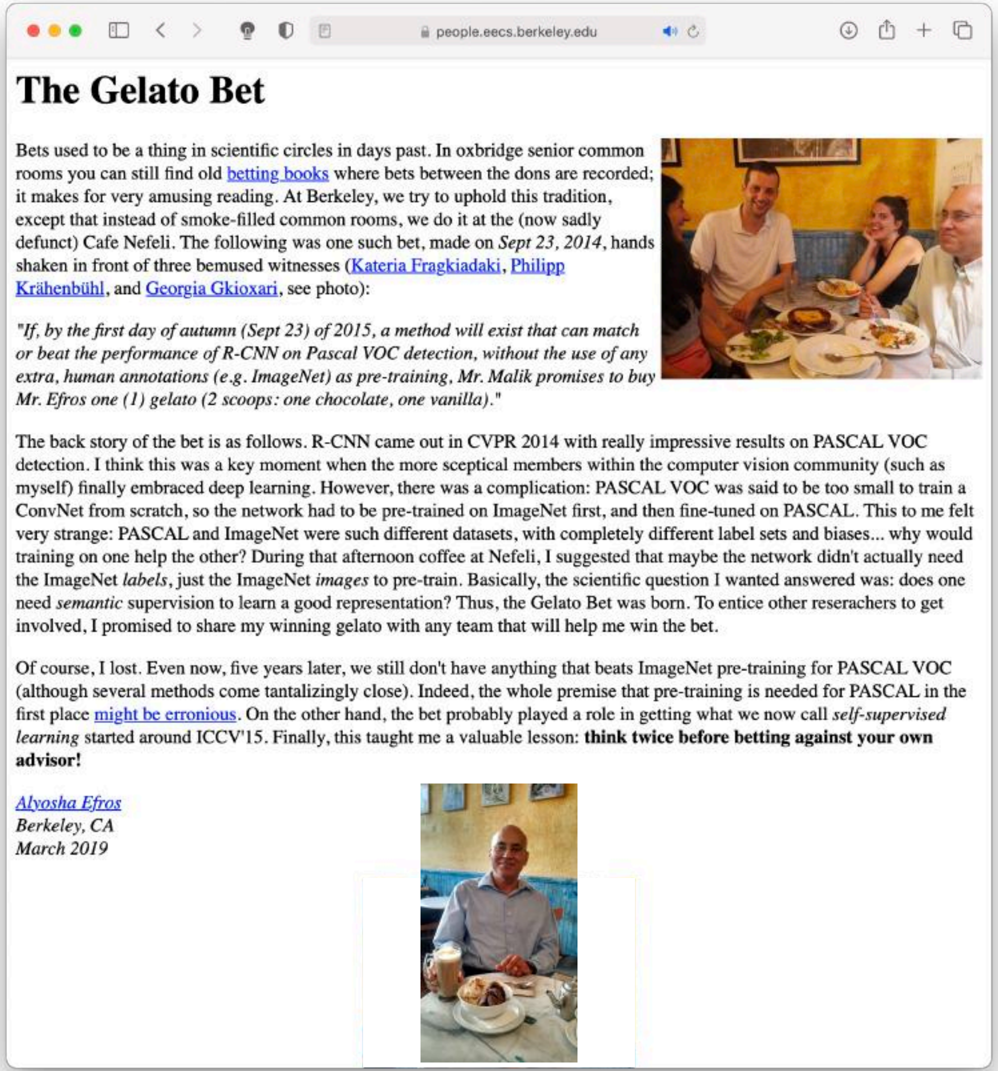
Downstream Tasks - NLP (BERT Revolution)

Rank	Name	Model	URL	Score	CoLA	SST-2	MRPC	STS-B	QQP	MNLI-m	MNLI-mm	QNLI	RTE	WNLI	AX
1	DeBERTa Team - Microsoft	DeBERTa / TuringNLRv4		90.8	71.5	97.5	94.0/92.0	92.9/92.6	76.2/90.8	91.9	91.6	99.2	93.2	94.5	53.2
2	HFL iFLYTEK	MacALBERT + DKM		90.7	74.8	97.0	94.5/92.6	92.8/92.6	74.7/90.6	91.3	91.1	97.8	92.0	94.5	52.6
3	Alibaba DAMO NLP	StructBERT + TAPT		90.6	75.3	97.3	93.9/91.9	93.2/92.7	74.8/91.0	90.9	90.7	97.4	91.2	94.5	49.1
4	PING-AN Omni-Sinitic	ALBERT + DAAF + NAS		90.6	73.5	97.2	94.0/92.0	93.0/92.4	76.1/91.0	91.6	91.3	97.5	91.7	94.5	51.2
5	ERNIE Team - Baidu	ERNIE		90.4	74.4	97.5	93.5/91.4	93.0/92.6	75.2/90.9	91.4	91.0	96.6	90.9	94.5	51.7
6	T5 Team - Google	T5		90.3	71.6	97.5	92.8/90.4	93.1/92.8	75.1/90.6	92.2	91.9	96.9	92.8	94.5	53.1
7	Microsoft D365 AI & MSR AI & GATECH	MT-DNN-SMART		89.9	69.5	97.5	93.7/91.6	92.9/92.5	73.9/90.2	91.0	90.8	99.2	89.7	94.5	50.2
8	Huawei Noah's Ark Lab	NEZHA-Large		89.8	71.7	97.3	93.3/91.0	92.4/91.9	75.2/90.7	91.5	91.3	96.2	90.3	94.5	47.9
9	Zihang Dai	Funnel-Transformer (Ensemble B10-10-10H1024)		89.7	70.5	97.5	93.4/91.2	92.6/92.3	75.4/90.7	91.4	91.1	95.8	90.0	94.5	51.6
10	ELECTRA Team	ELECTRA-Large + Standard Tricks		89.4	71.7	97.1	93.1/90.7	92.9/92.5	75.6/90.8	91.3	90.8	95.8	89.8	91.8	50.7
11	Microsoft D365 AI & UMD	FreeLB-RoBERTa (ensemble)		88.4	68.0	96.8	93.1/90.8	92.3/92.1	74.8/90.3	91.1	90.7	95.6	88.7	89.0	50.1
12	Junjie Yang	HIRE-RoBERTa		88.3	68.6	97.1	93.0/90.7	92.4/92.0	74.3/90.2	90.7	90.4	95.5	87.9	89.0	49.3
13	Facebook AI	RoBERTa		88.1	67.8	96.7	92.3/89.8	92.2/91.9	74.3/90.2	90.8	90.2	95.4	88.2	89.0	48.7
14	Microsoft D365 AI & MSR AI	MT-DNN-ensemble		87.6	68.4	96.5	92.7/90.3	91.1/90.7	73.7/89.9	87.9	87.4	96.0	86.3	89.0	42.8
15	GLUE Human Baselines	GLUE Human Baselines		87.1	66.4	97.8	86.3/80.8	92.7/92.6	59.5/80.4	92.0	92.8	91.2	93.6	95.9	-

https://gluebenchmark.com/leaderboard

Downstream Tasks - Vision (Contrastive)

Method	Architecture	mAP
Transfer from labeled data:		
Supervised baseline	ResNet-152	74.7
Transfer from unlabeled data:		
Exemplar [17] by [13]	ResNet-101	60.9
Motion Segmentation [47] by [13]	ResNet-101	61.1
Colorization [64] by [13]	ResNet-101	65.5
Relative Position [14] by [13]	ResNet-101	66.8
Multi-task [13]	ResNet-101	70.5
Instance Discrimination [60]	ResNet-50	65.4
Deep Cluster [7]	VGG-16	65.9
Deeper Cluster [8]	VGG-16	67.8
Local Aggregation [66]	ResNet-50	69.1
Momentum Contrast [25]	ResNet-50	74.9
Faster-RCNN trained on CPC v2	ResNet-161	**76.6**


people.eecs.berkeley.edu

The Gelato Bet

Bets used to be a thing in scientific circles in days past. In oxbridge senior common rooms you can still find old betting books where bets between the dons are recorded; it makes for very amusing reading. At Berkeley, we try to uphold this tradition, except that instead of smoke-filled common rooms, we do it at the (now sadly defunct) Cafe Nefeli. The following was one such bet, made on *Sept 23, 2014*, hands shaken in front of three bemused witnesses (Kateria Fragkiadaki, Philipp Krähenbühl, and Georgia Gkioxari, see photo):

"If, by the first day of autumn (Sept 23) of 2015, a method will exist that can match or beat the performance of R-CNN on Pascal VOC detection, without the use of any extra, human annotations (e.g. ImageNet) as pre-training, Mr. Malik promises to buy Mr. Efros one (1) gelato (2 scoops: one chocolate, one vanilla)."

The back story of the bet is as follows. R-CNN came out in CVPR 2014 with really impressive results on PASCAL VOC detection. I think this was a key moment when the more sceptical members within the computer vision community (such as myself) finally embraced deep learning. However, there was a complication: PASCAL VOC was said to be too small to train a ConvNet from scratch, so the network had to be pre-trained on ImageNet first, and then fine-tuned on PASCAL. This to me felt very strange: PASCAL and ImageNet were such different datasets, with completely different label sets and biases... why would training on one help the other? During that afternoon coffee at Nefeli, I suggested that maybe the network didn't actually need the ImageNet *labels*, just the ImageNet *images* to pre-train. Basically, the scientific question I wanted answered was: does one need *semantic* supervision to learn a good representation? Thus, the Gelato Bet was born. To entice other reserachers to get involved, I promised to share my winning gelato with any team that will help me win the bet.

Of course, I lost. Even now, five years later, we still don't have anything that beats ImageNet pre-training for PASCAL VOC (although several methods come tantalizingly close). Indeed, the whole premise that pre-training is needed for PASCAL in the first place might be erronious. On the other hand, the bet probably played a role in getting what we now call *self-supervised learning* started around ICCV'15. Finally, this taught me a valuable lesson: **think twice before betting against your own advisor!**

Alyosha Efros
Berkeley, CA
March 2019

Olivier J. Hénaff, Aravind Srinivas, Jeffrey De Fauw, Ali Razavi, Carl Doersch, S. M. Ali Eslami, Aaron van den Oord, **Data-Efficient Image Recognition with Contrastive Predictive Coding**, ICML 2020

Summary

- **Unsupervised Learning:** Rapidly advancing field thanks to compute; deep learning engineering practices; datasets; lot of people working on it.
- **Not just an academic interest topic**. Production level impact [example: BERT is in use for Google Search and Assistant].
- **What is true now may not be true even a year from now** [example: self-supervised pre-training was way worse than supervised in computer vision tasks like detection/segmentation last year. Now it is better].
- **Language Modeling** (GPT), **Image Generation** (conditional GANs), **Language pre-training** (BERT), **vision pre-training** (CPC / MoCo) starting to work really well. Good time to learn these well and make very impactful contributions.
- **Autoregressive Density Modeling**, **Flows**, **VAEs**, **GANs**, etc. have huge room for improvement. Great time to work on them.

Next Lecture:

Neural Building Blocks I: Spatial Processing with CNNs