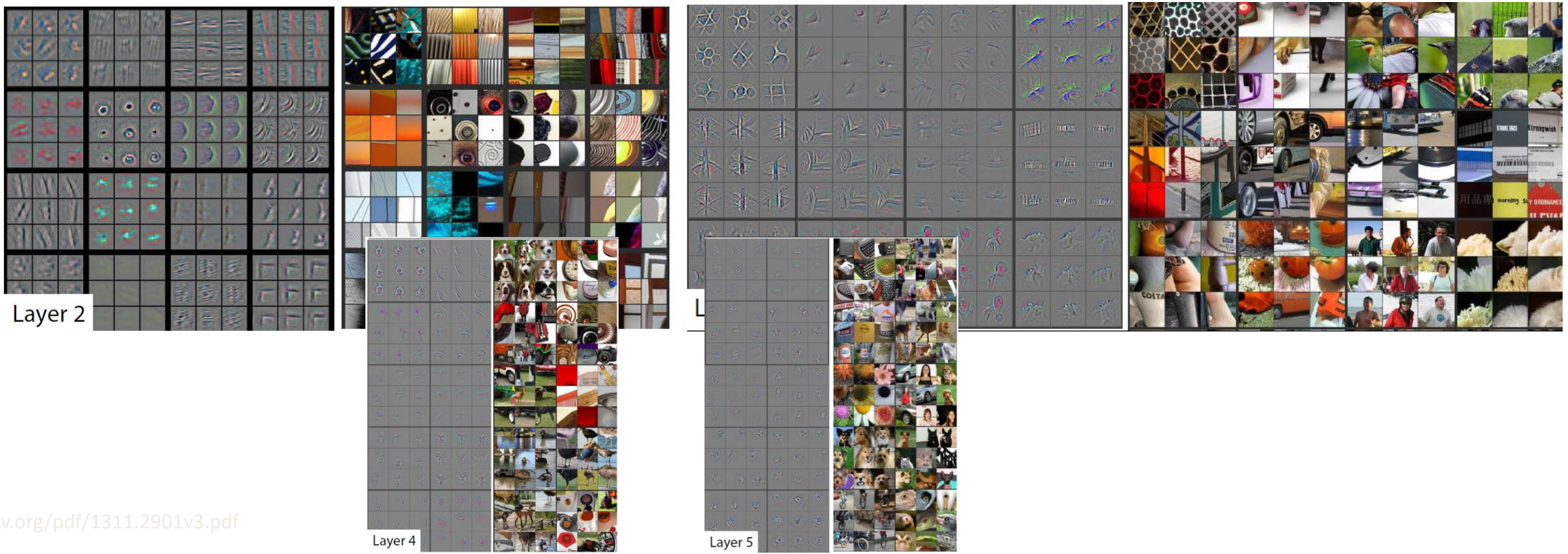
CPSC 340:
Machine Learning and Data Mining

What do we learn?

Fall 2022

Previously: Deep Learning for Computer Vision

- We discussed how deep learning revolutionized computer vision.
  - Unprecedented performance across a wide of variety of tasks.
- Hidden units often correlate semantically-meaningful concepts.

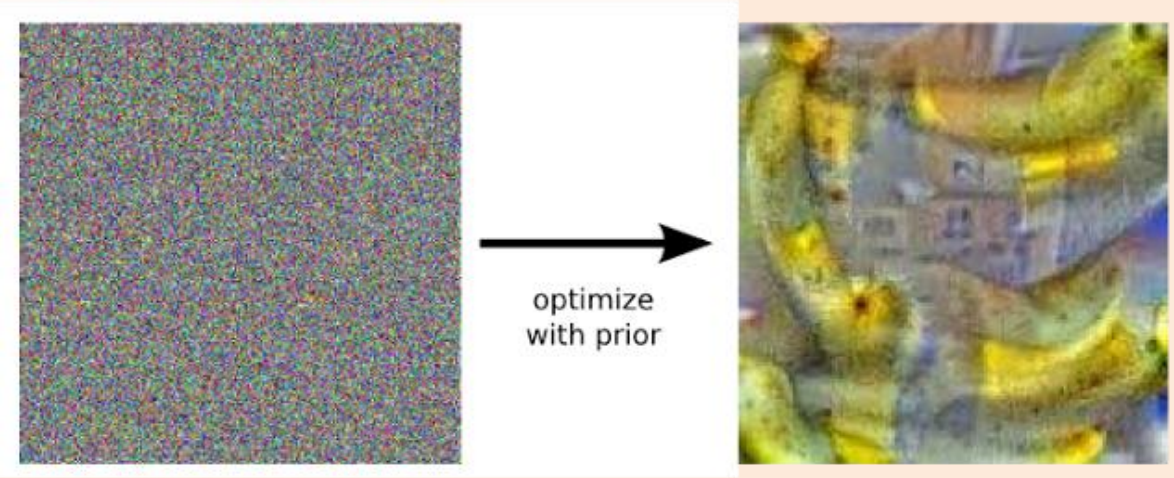


https://arxiv.org/pdf/1311.2901v3.pdf

Inceptionism

- A crazy idea:
  - Instead of weights, use backpropagation to take gradient with respect to x_i.
- Inceptionism with trained network:
  - Fix the label y_i (e.g., "banana").
  - Start with random noise image x_i.
  - Use gradient descent on image x_i.
  - Add a spatial regularizer on x_{ij}:
    - Encourages neighbouring x_{ij} to be similar.

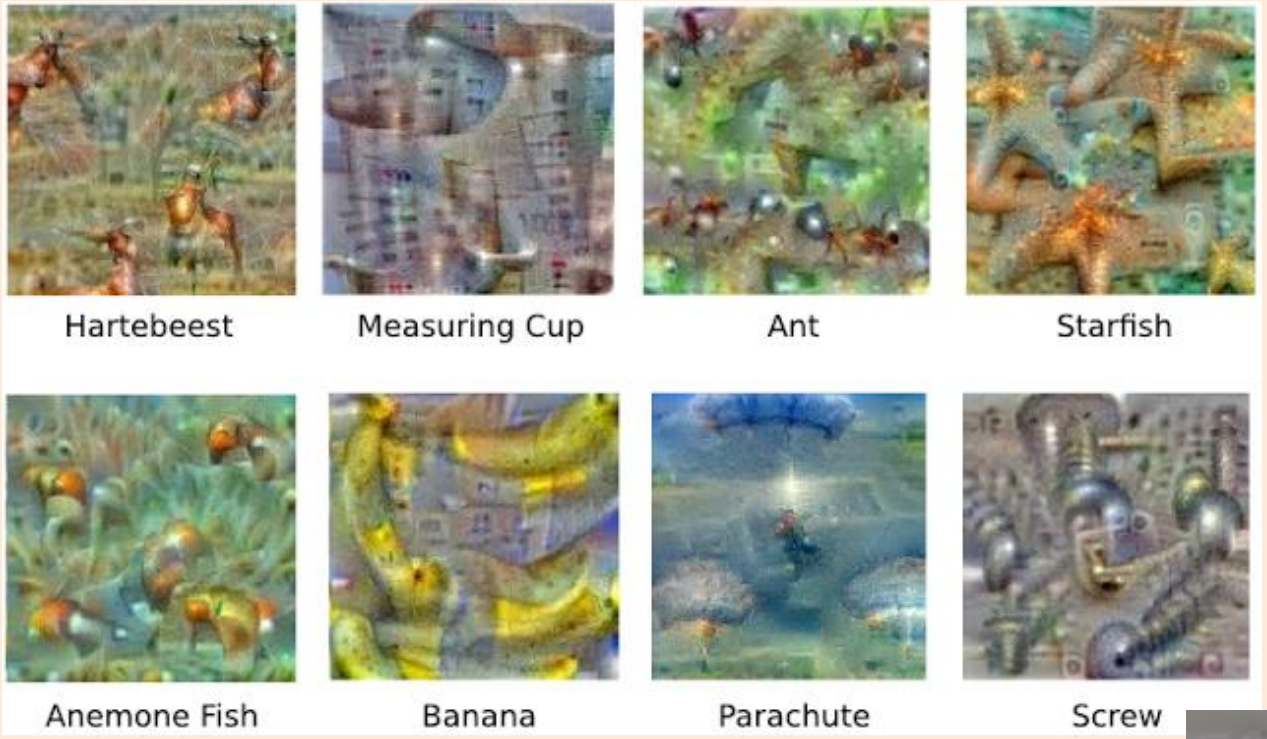
"Show what you think a banana looks like."



http://googleresearch.blogspot.ca/2015/06/inceptionism-going-deeper-into-neural.html

Inceptionism

- Inceptionism for different class labels:

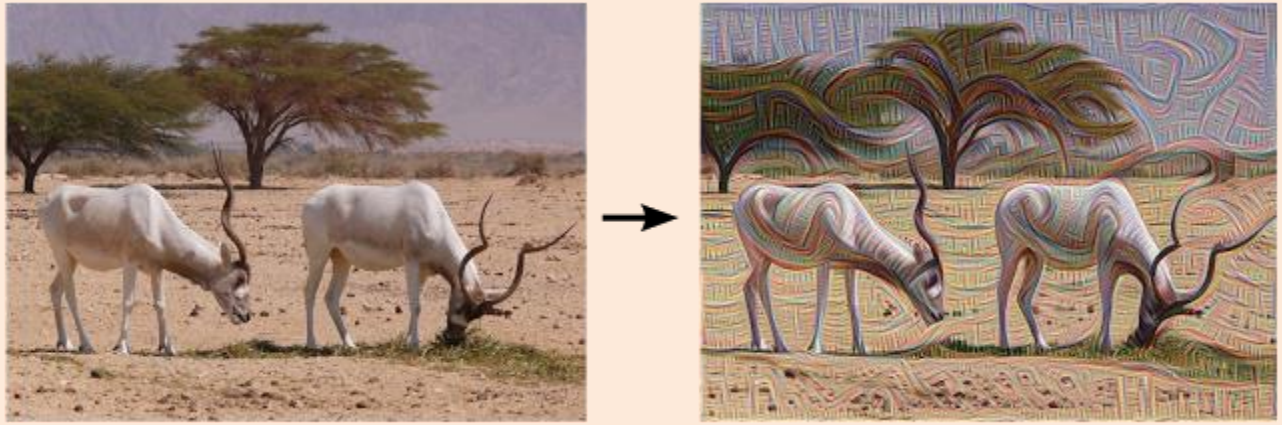


Dumbbell

http://googleresearch.blogspot.ca/2015/06/inceptionism-going-deeper-into-neural.html

Inceptionism

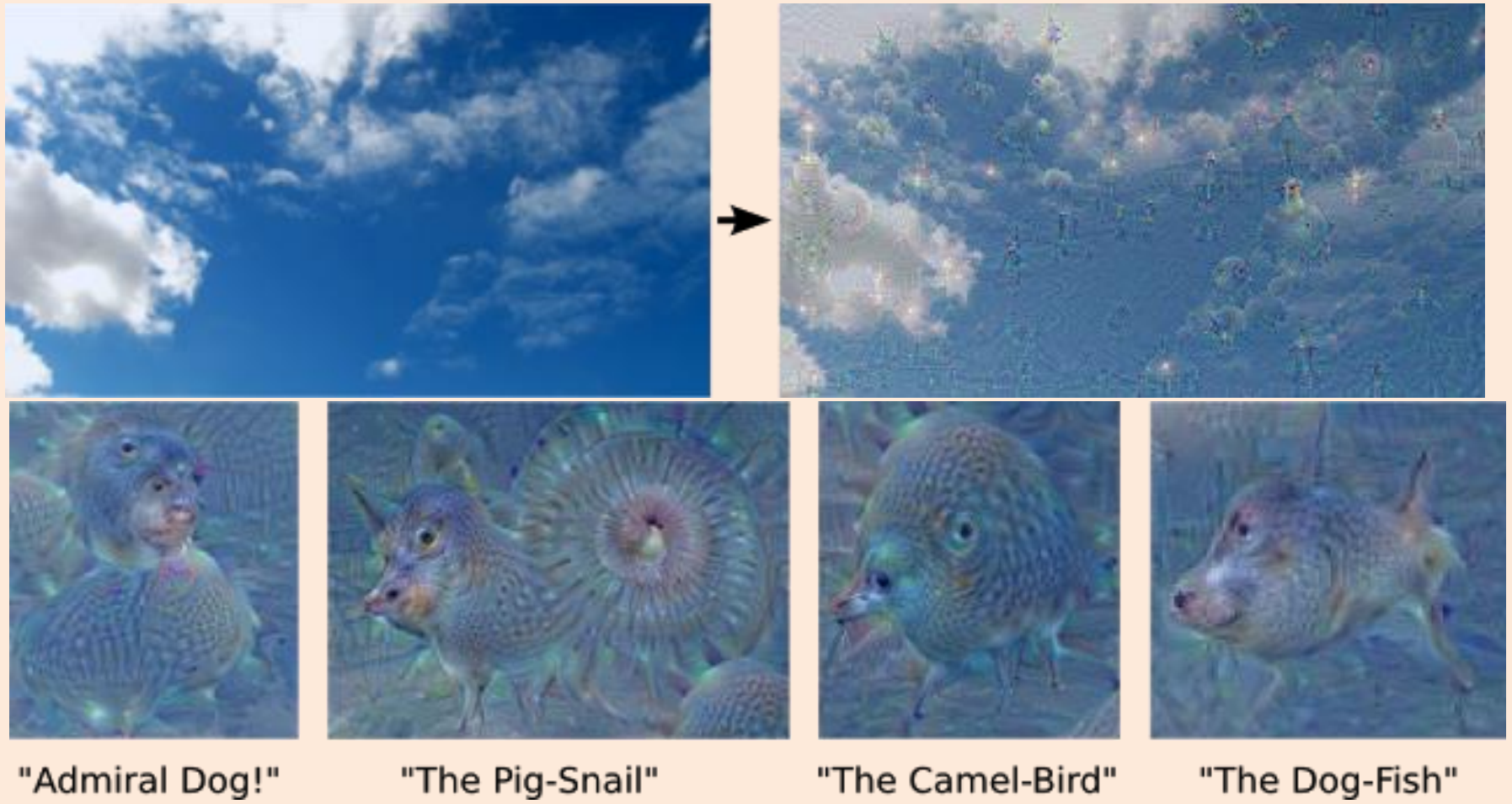
- Inceptionism where we try to match $z_i^{(m)}$ values instead of y_i.
  - Shallow ‘m’:



http://googleresearch.blogspot.ca/2015/06/inceptionism-going-deeper-into-neural.html

Inceptionism

- Inceptionism where we try to match $z_i^{(m)}$ values instead of y_i.
  - Deepest 'm':



http://googleresearch.blogspot.ca/2015/06/inceptionism-going-deeper-into-neural.html

Inceptionism

- Inceptionism where we try to match $z_i^{(m)}$ values instead of y_i.
  - "Deep dream" starts from random noise:

  - Deep Dream video

http://googleresearch.blogspot.ca/2015/06/inceptionism-going-deeper-into-neural.html

Artistic Style Transfer

- Artistic style transfer:
  - Given a content image ‘C’ and a style image ‘S’.
  - Make a image that has content of ‘C’ and style of ‘S’.

Content:

Style:

https://commons.wikimedia.org/wiki/File:Tuebingen_Neckarfront.jpg
https://en.wikipedia.org/wiki/The_Starry_Night

Artistic Style Transfer

- Artistic style transfer:
  - Given a content image ‘C’ and a style image ‘S’.
  - Make a image that has content of ‘C’ and style of ‘S’.

- CNN-based approach applies gradient descent with 2 terms:
  - Loss function: match deep latent representation of content image ‘C’:
    - Difference between $z_i^{(m)}$ for deepest ‘m’ between x_i and ‘C’.
  - Regularizer: match all latent representation covariances of style image ‘S’.
    - Difference between covariance of $z_i^{(m)}$ for all ‘m’ between x_i and ‘S’.

Artistic Style Transfer

http://arxiv.org/pdf/1508.06576v2.pdf

Image Gallery

Examples

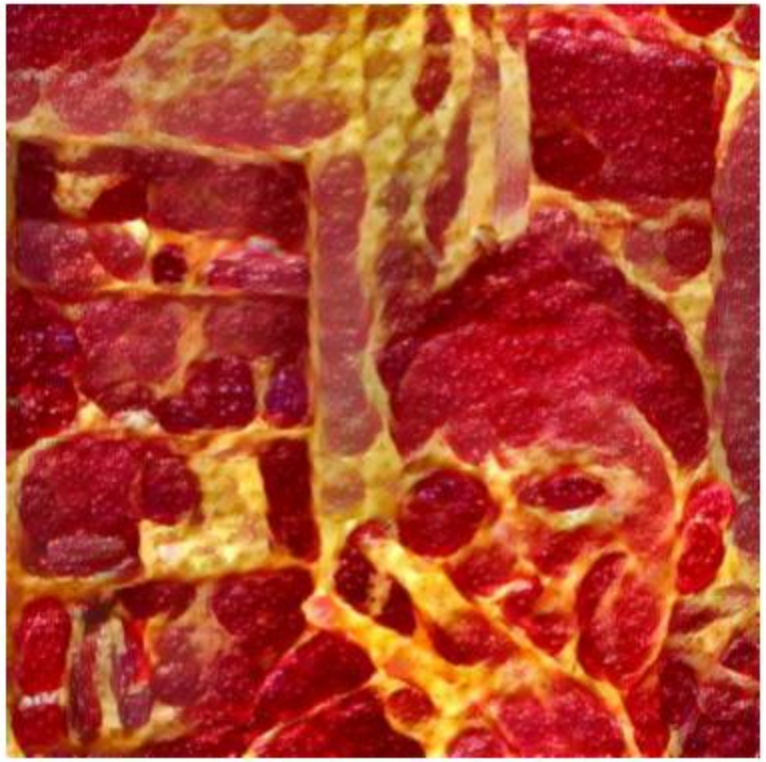
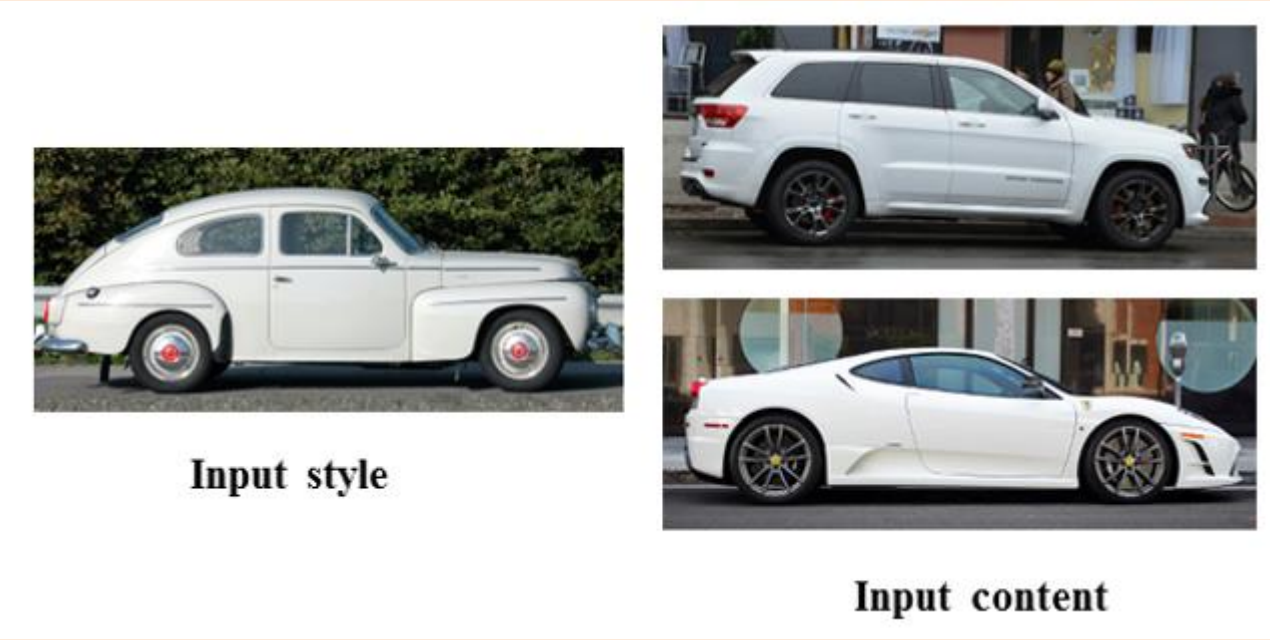


Figure: **Left:** My friend Grant, **Right:** Grant as a pizza

Artistic Style Transfer

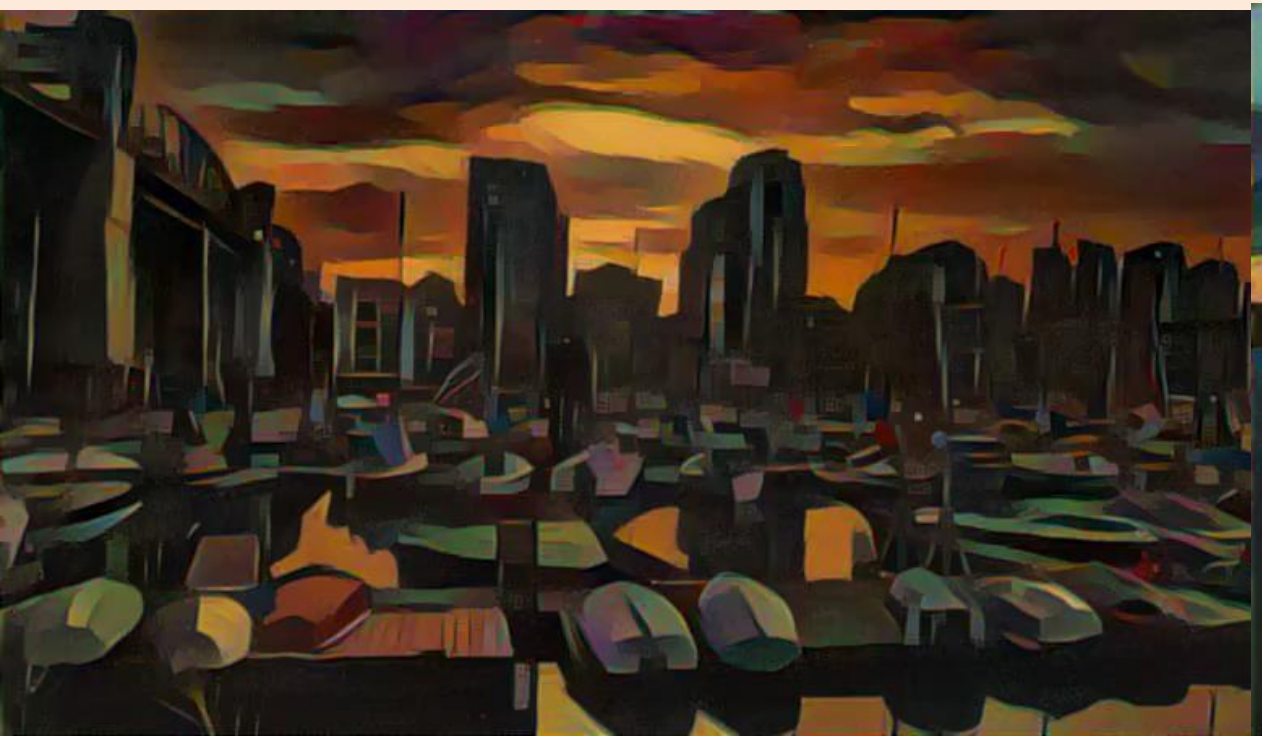
- Other artistic style transfer methods combine CNNs and graphical models (CPSC 440):



https://arxiv.org/pdf/1601.04589.pdf

Artistic Style Transfer for Video

- Combining style transfer with optical flow:
  - https://www.youtube-nocookie.com/embed/Khuj4ASldmU
- Videos from a former CPSC 340 student/TA's paper:



Next Topic: What do we learn?

Mission Accomplished?

- For speech recognition and computer vision and language:
  - No non-deep methods have ever given the current level of performance.
  - Deep models continue to improve performance on these and related tasks.
    - Though we don't know how to scale up other universal approximators.
  - There is some overfitting to popular datasets like ImageNet.
    - Recent work showed accuracy drop of 11-14%by using a different ImageNet test set.

- CNNs are now making their way into products.
  - Face/person recognition in various cameras.
  - Car/pedestrian/biker recognition in vehicles.
  - Amazon Go: https://www.youtube.com/watch?v=NrmMk1Myrxc
    - Trolling by French company Monoprix here.

Mission Accomplished?

- We are still missing a lot of theory and understanding deep learning.

From: Boris
To: Ali

On Friday, someone on another team changed the default rounding mode of some Tensorflow internals (from truncation to "round to even").*

*Our training broke. Our error rate went from <25% error to ~99.97% error (on a standard 0-1 binary loss).

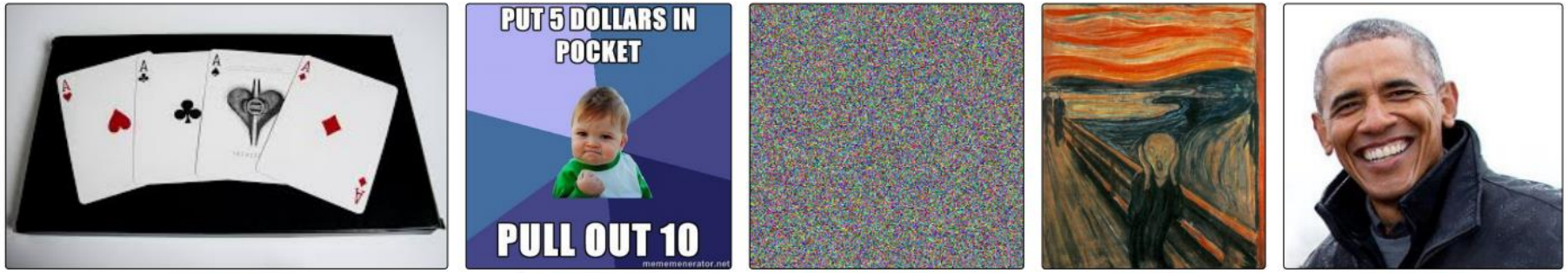
- "Good CS expert says: Most firms that think they want advanced AI/ML really just need linear regression on cleaned-up data."

http://www.argmin.net/2017/12/05/kitchen-sinks/

Or random forests!

Mission Accomplished?

- Despite high-level of abstraction, deep CNNs are easily fooled:
  - What happens when you give a weird input to a CNN?

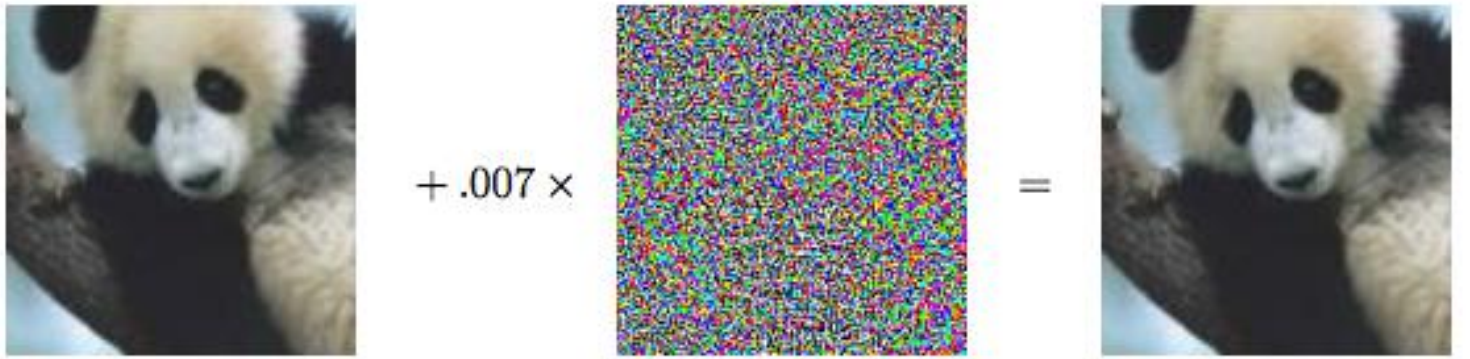


DenseNet 161 (2017)	Envelope 31%	Balance Beam 52%	Chainlink Fence 31%	Chest 37%	Tench 36%
SqueezeNet (2016)	Binder 43%	Balance Beam 18%	Poncho 32%	Jean 30%	Suit 21%
ResNet 152 (2015)	Envelope 40%	Pacifier 33%	Chain Mail 29%	Dust Cover 52%	Sweatshirt 25%
VGG 19 (2014)	Binder 51%	Dust Cover 44%	Window Screen 5%	Chest 11%	Sweatshirt 46%
AlexNet (2012)	T-shirt 16%	Dust Cover 22%	Cardigan 12%	Theater Curtain 3%	Coho 37%

Figure 1: The arbitrary predictions of several popular networks [2, 3, 4, 5, 6] that are trained on ImageNet [1] on unseen data. The red predictions are entirely wrong, the green predictions are justifiable, the orange predictions are less justifiable. The middle image is noise sampled from $\mathcal{N}(\mu = 0.5, \sigma = 0.25)$ without any modifications. This unpredictable behaviour is not limited to demonstrated architectures. We show that merely thresholding the output probability is not a reliable method to detect these problematic instances.

Mission Accomplished?

- Despite high-level of abstraction, deep CNNs are easily fooled:
  - What happens when you give a weird input to a CNN?
- "Adversarial examples": imperceptible noise that changes label.
  - Can change any label to any other label.



https://arxiv.org/pdf/1412.6572.pdf, https://blog.openai.com/adversarial-example-research/

Mission Accomplished?

- Can someone repaint a stop sign and fool self-driving cars?

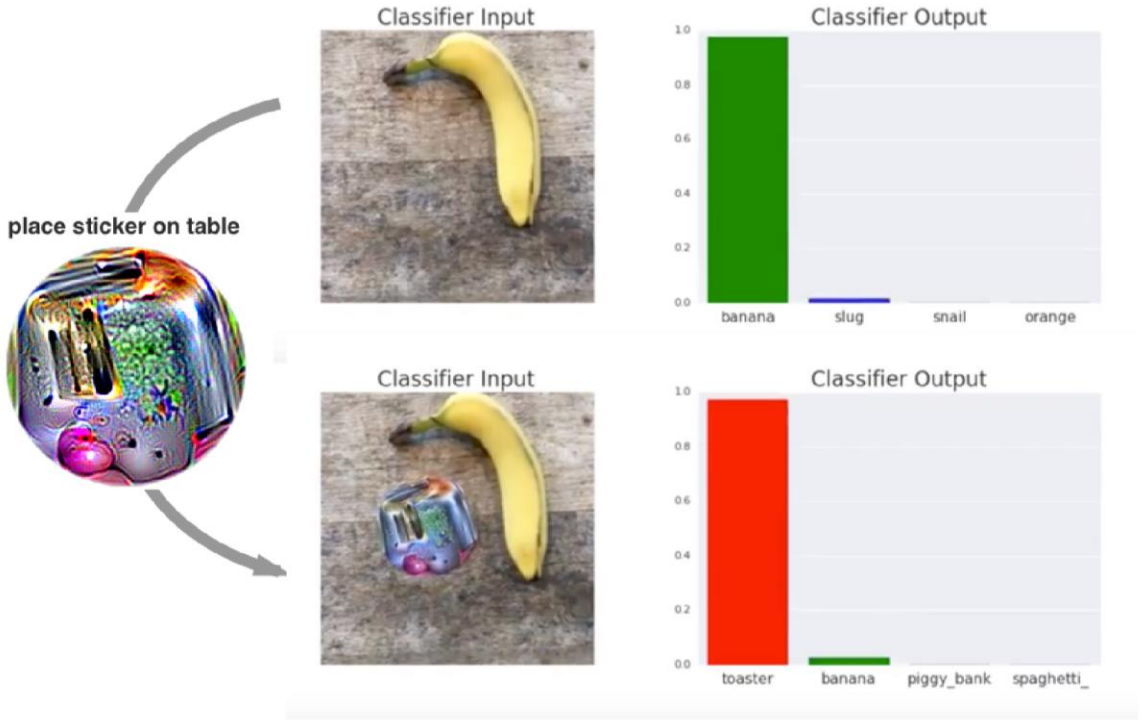


Figure 1: A real-world attack on VGG16, using a physical patch generated by the white-box ensemble method described in Section 3. When a photo of a tabletop with a banana and a notebook (top photograph) is passed through VGG16, the network reports class 'banana' with 97% confidence (top plot). If we physically place a sticker targeted to the class "toaster" on the table (bottom photograph), the photograph is classified as a toaster with 99% confidence (bottom plot). See the following video for a full demonstration: `https://youtu.be/i1sp4X57TL4`

https://arxiv.org/pdf/1712.09665.pdf

Mission Accomplished?

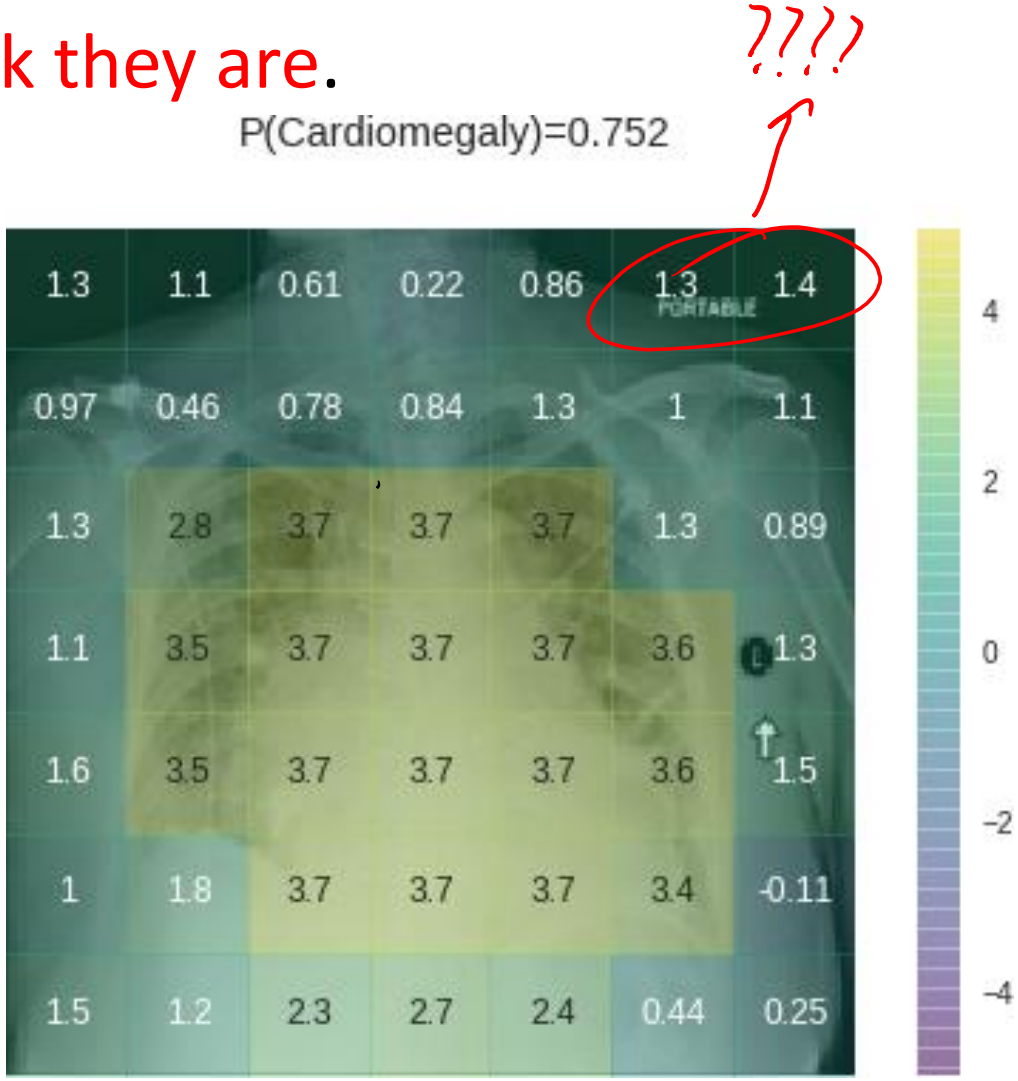
- Are the networks understanding the fundamental concepts?
  - Is being “surrounded by green” part of the definition of cow?
  - Do we need to have examples of cows in different environments?
    - Kids don’t need this.

- Image colourization:

https://www.onegreenplanet.org/news/cows-enjoy-the-beach/
https://mathwithbaddrawings.com/2017/10/18/5-ways-to-troll-your-neural-network/

Mission Accomplished?

- CNNs may not be learning what you think they are.
  - CNN for diagnosing enlarged heart:
    - Higher values mean more likely to be enlarged:
  - CNN says “portable” protocal is predictive:
    - But they are probabaly getting a “portable” scan because they’re too sick to go the hospital.
  - CNN was biased by the scanning protocal.
    - Learns the scans that more-sick patients get.
    - This is not what we want in a medical test.



https://medium.com/@jrzech/what-are-radiological-deep-learning-models-actually-learning-f97a546c5b98

Meaningless comparisons lead to false optimism in medical machine learning

Orianna DeMasi, Konrad Kording, Benjamin Recht

(Submitted on 19 Jul 2017)

A new trend in medicine is the use of algorithms to analyze big datasets, e.g. using everything your phone measures about you for diagnostics or monitoring. However, these algorithms are commonly compared against weak baselines, which may contribute to excessive optimism. To assess how well an algorithm works, scientists typically ask how well its output correlates with medically assigned scores. Here we perform a meta-analysis to quantify how the literature evaluates their algorithms for monitoring mental wellbeing. We find that the bulk of the literature (∼77%) uses meaningless comparisons that ignore patient baseline state. For example, having an algorithm that uses phone data to diagnose mood disorders would be useful. However, it is possible to over 80% of the variance of some mood measures in the population by simply guessing that each patient has their own average mood - the patient-specific baseline. Thus, an algorithm that just predicts that our mood is like it usually is can explain the majority of variance, but is, obviously, entirely useless. Comparing to the wrong (population) baseline has a massive effect on the perceived quality of algorithms and produces baseless optimism in the field. To solve this problem we propose "user lift" that reduces these systematic errors in the evaluation of personalized medical monitoring.

- Related: does the prediction change real-world outcomes?
  - Are you just annoying the highly-paid doctor or paying for nothing?

Racially-Biased Algorithms?

- Major issue: are we learning representations with harmful biases?
  - Biases could come from data (if data only has certain groups in certain situations).
  - Biases could come from labels (always using label of “ball” for certain sports).
  - Biases could come from learning method (model predicts “basketball” for black people more often than this appears in training data for basketball images).

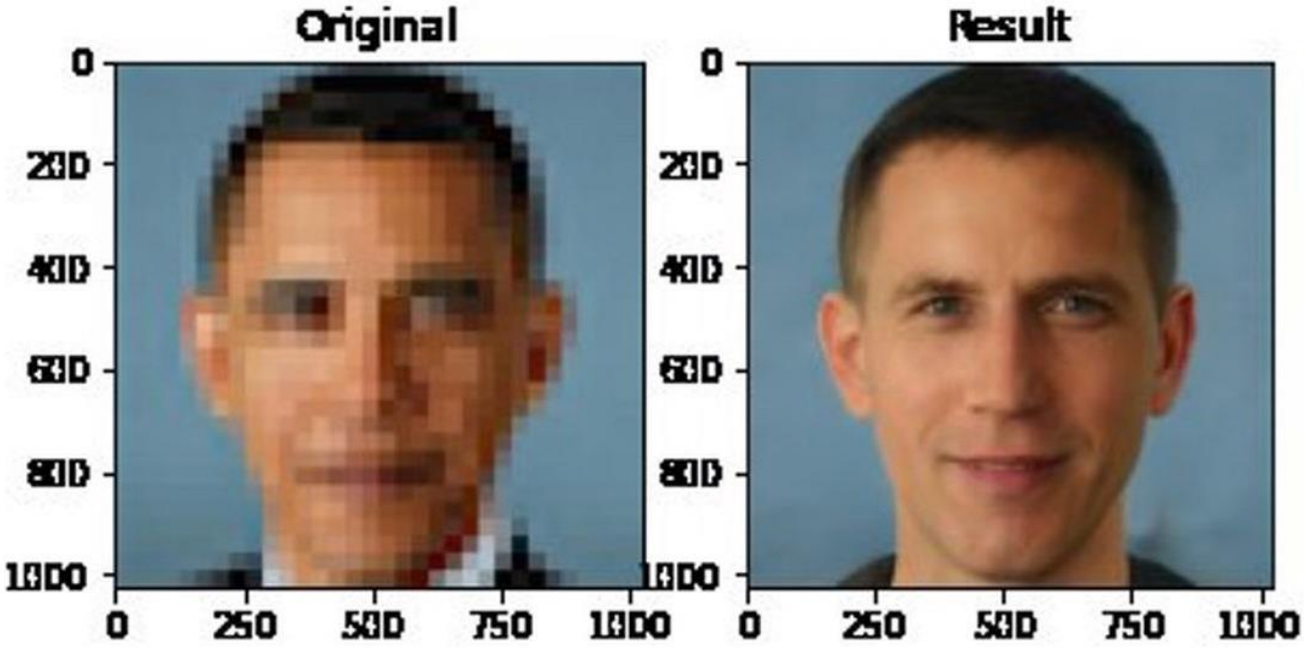
Fig. 8: Pairs of pictures (columns) sampled over the Internet along with their prediction by a ResNet-101.

  - This is a major problem/issue when deploying these systems.
    - For example, “repeat-offender prediction” that reinforces racial biases in arrest patterns.

https://arxiv.org/pdf/1711.11443.pdf

Racially-Biased Algorithms?

- Results on image super-resolution (upscaling) method:

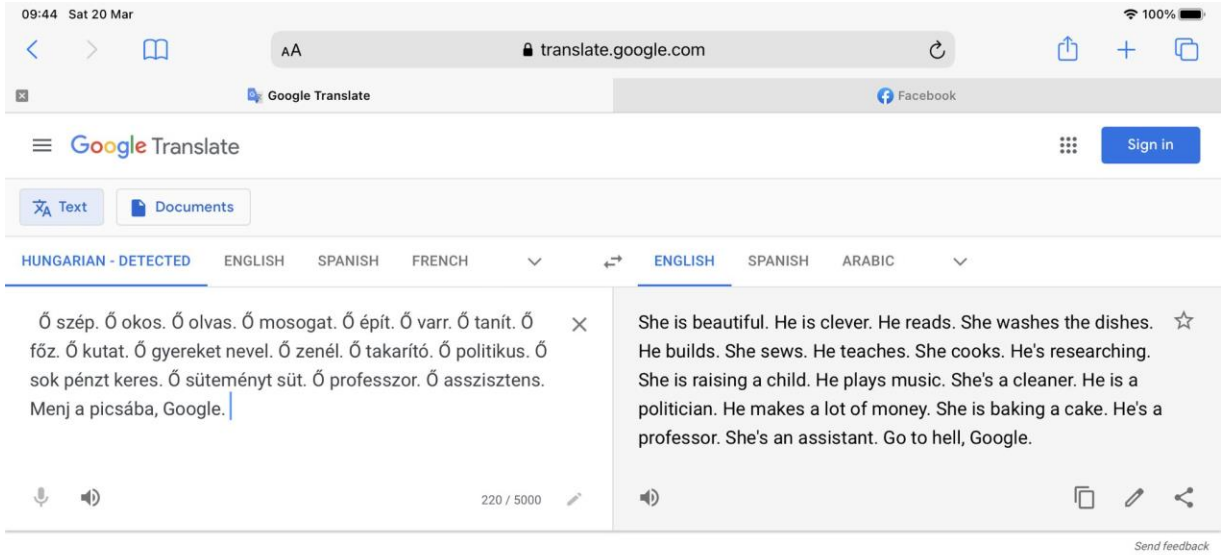


- See also: AI has the worst superpower… medical racism
- Sometimes these issues can be reduced by careful data collection.
  - In this case, we could train on a more-diverse group.
  - But sometimes you cannot collect unbiased data.

https://www.theverge.com/21298762/face-depixelizer-ai-machine-learning-tool-pulse-stylegan-obama-bias

Sexist Algorithms?

- Hungarian is gender neutral.
  - Google assigns a gender based on frequencies in training set:



- Amazon's hiring algorithm penalized candidates with "woman/women" in application.
  - A correlation/causation issue: "most engineers at Amazon are men, engineers should be men?"
- Maybe we will eventually fix issues like this.
  - Until we do, maybe we should not use machine learning in some applications.
    - Or at least warn people about potential biases.

https://twitter.com/doravargha/status/1373211762108076034

- From "How to Recognize AI Snake Oil".

Incomplete & crude but useful breakdown

Genuine, rapid progress	Imperfect but improving	Fundamentally dubious
• Shazam, reverse img search	• Spam detection	• Predicting recidivism
• Face recognition	• Copyright violation	• Predicting job success
• Med. diagnosis from scans	• Automated essay grading	• Predictive policing
• Speech to text	• Hate speech detection	• Predicting terrorist risk
• Deepfakes	• Content recommendation	• Predicting at-risk kids
Perception	Automating judgment	Predicting social outcomes

https://www.cs.princeton.edu/~arvindn/talks/MIT-STS-AI-snakeoil.pdf

Some Issues with Algorithms for Social Prediction

- Does fighting over-fitting give bad predictions on sub-groups?
  - If you have 99% "Group A" in your dataset,
    model can do well on average by only focusing on Group A.
    - Treat the other 1% as outliers.
  - Does "not trying to overfit" mean we perform badly on some groups?
  - Can we discover what groups exist in our dataset?

- What if all institutions use the same algorithm?
  - You apply for jobs everywhere, and are always rejected by the algorithm?
    - Even though you may be arbitrarily-close to the decision threshold.

- Fixing the various societal problems with using ML algorithms:
  - Hot research topic at the moment (good thesis or course project topic).
  - We do not currently have nice "solutions" for these issues.
    - Try to think of potential confounding factors, and consider whether ML is not appropriate.

Energy Costs

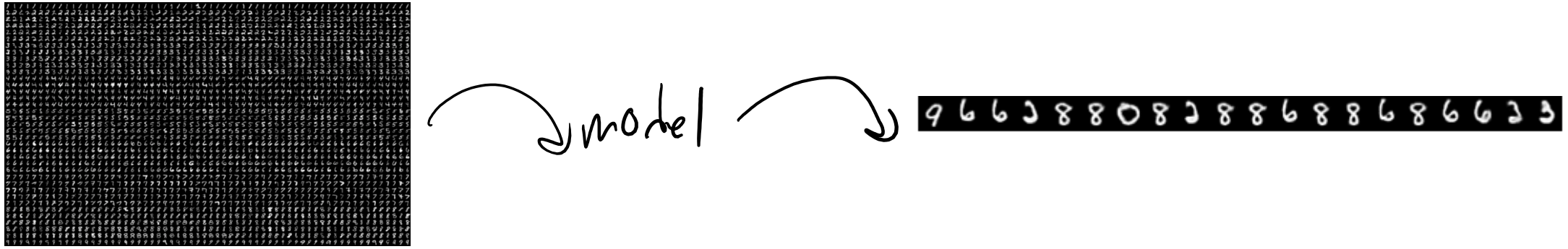
- Current methods require:
  - A lot of data.
  - A lot of time to train.
  - Many training runs to do hyper-parameter optimization.

- Recent paper regarding recent deep language models:
  - Entire training procedure emits 5 times more CO_2 than lifetime emission of a car, including making the car.

Next Topic: Generative Sampling

Generative Sampling Task

- Given training data, we want to make more data.
  - That looks like it comes from the test distribution.
- Example:
  - Train on MNIST images of the digits 0-9.
  - Samples from the model should look like more MNIST digits.

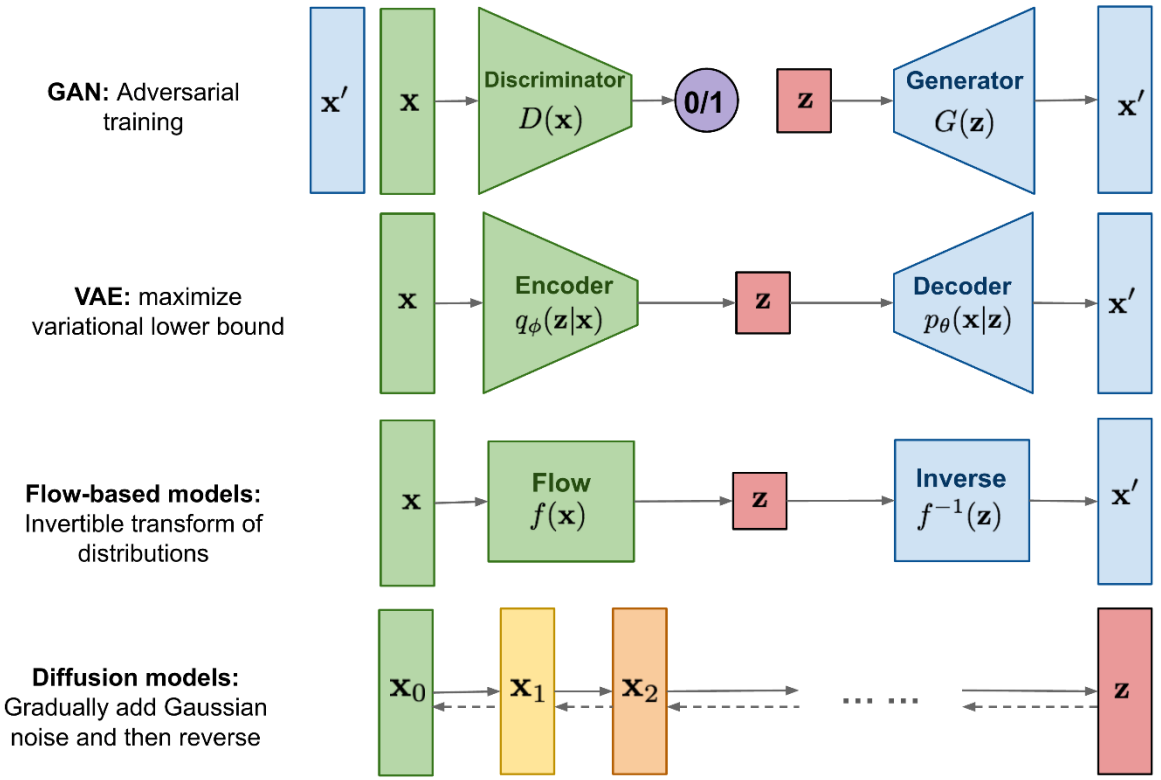


- 10 years ago, we could only sample simple datasets like MNIST.
  - Even with deep models like “deep belief nets” and “deep Boltzmann machines”.

https://www.kaggle.com/tarunkr/digit-recognition-tutorial-cnn-99-67-accuracy
http://deeplearning.net/tutorial/rbm.html

Rapid Progress in Generative Sampling

- Last 10 years have seen a variety of new deep generative models:
  - Variational autoencoders (VAEs).
  - Generative adversarial networks (GANs).
  - Normalizing flows.
  - Diffusion models.



https://lilianweng.github.io/posts/2021-07-11-diffusion-models/

Rapid Progress in Generative Sampling

- Rapid progress due to these new deep methods:

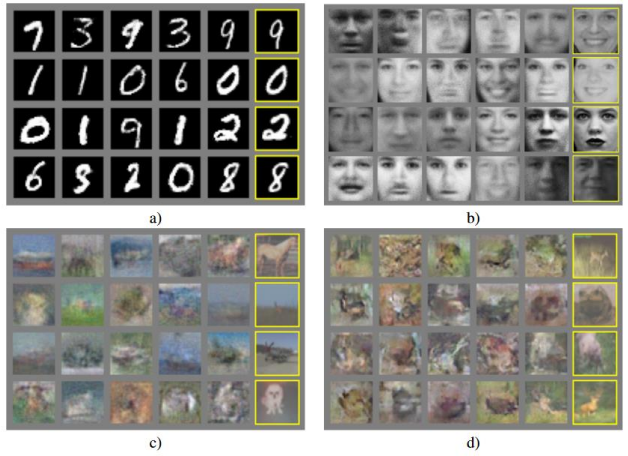
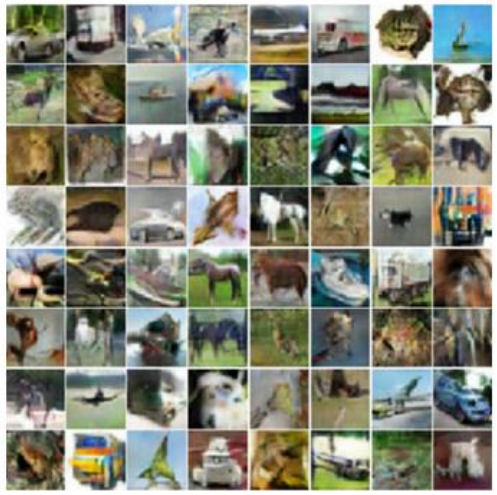


Figure 2: Visualization of samples from the model. Rightmost column shows the nearest training example of the neighboring sample, in order to demonstrate that the model has not memorized the training set. Samples are fair random draws, not cherry-picked. Unlike most other visualizations of deep generative models, these images show actual samples from the model distributions, not conditional means given samples of hidden units. Moreover, these samples are uncorrelated because the sampling process does not depend on Markov chain mixing. a) MNIST b) TFD c) CIFAR-10 (fully connected model) d) CIFAR-10 (convolutional discriminator and "deconvolutional" generator)

2014



2016

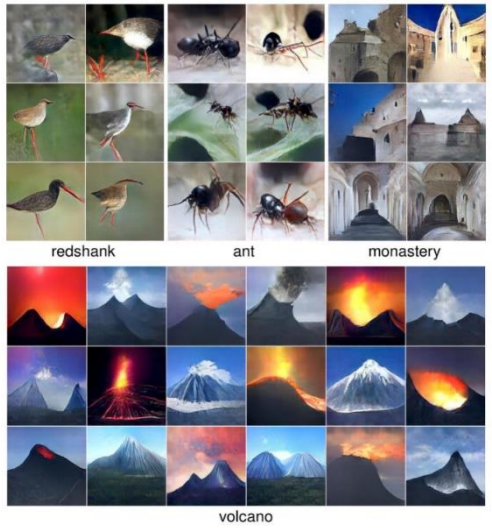
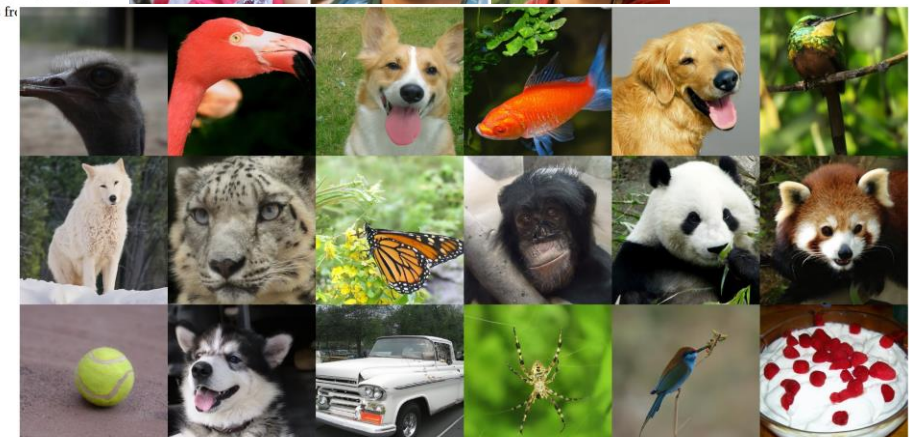


Figure 33: PPGNs are able to generate diverse, high resolution images fr
classes. Image reproduced from Nguyen *et al.* (2016).

2019



2021

https://arxiv.org/pdf/1406.2661.pdf
https://blog.openai.com/generative-models
https://arxiv.org/pdf/1701.00160.pdf
https://www.doublehelixschool.com/creativity-blog/this-person-does-not-exist
https://arxiv.org/pdf/2105.05233.pdf

Diffusion Models

- "Hot" generating sampling model in 2022 is diffusion models.
- Basic high-level idea:
  - Take training images, and add noise to them in a sequence of steps.
    - Until the image basically looks like random noise.
  - Train neural network to reverse those steps.

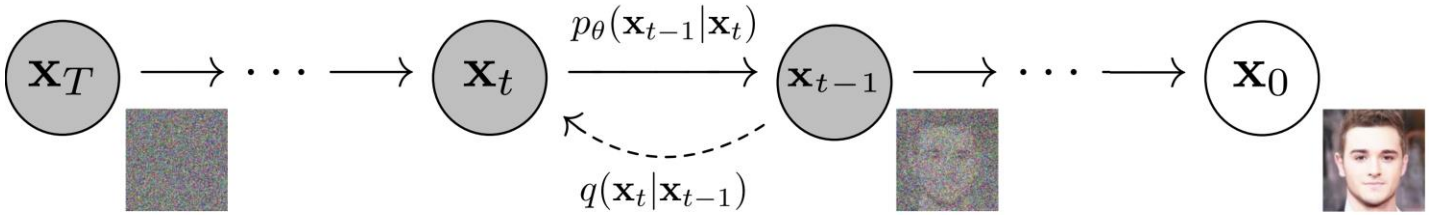


Figure 2: The directed graphical model considered in this work.

  - Generate a new image by starting from random noise and applying the network.
- Similar idea to denoising autoencoders.
  - But trains to denoise with different amounts of noise.
  - I am skipping lots of details due to time, but results are astounding…

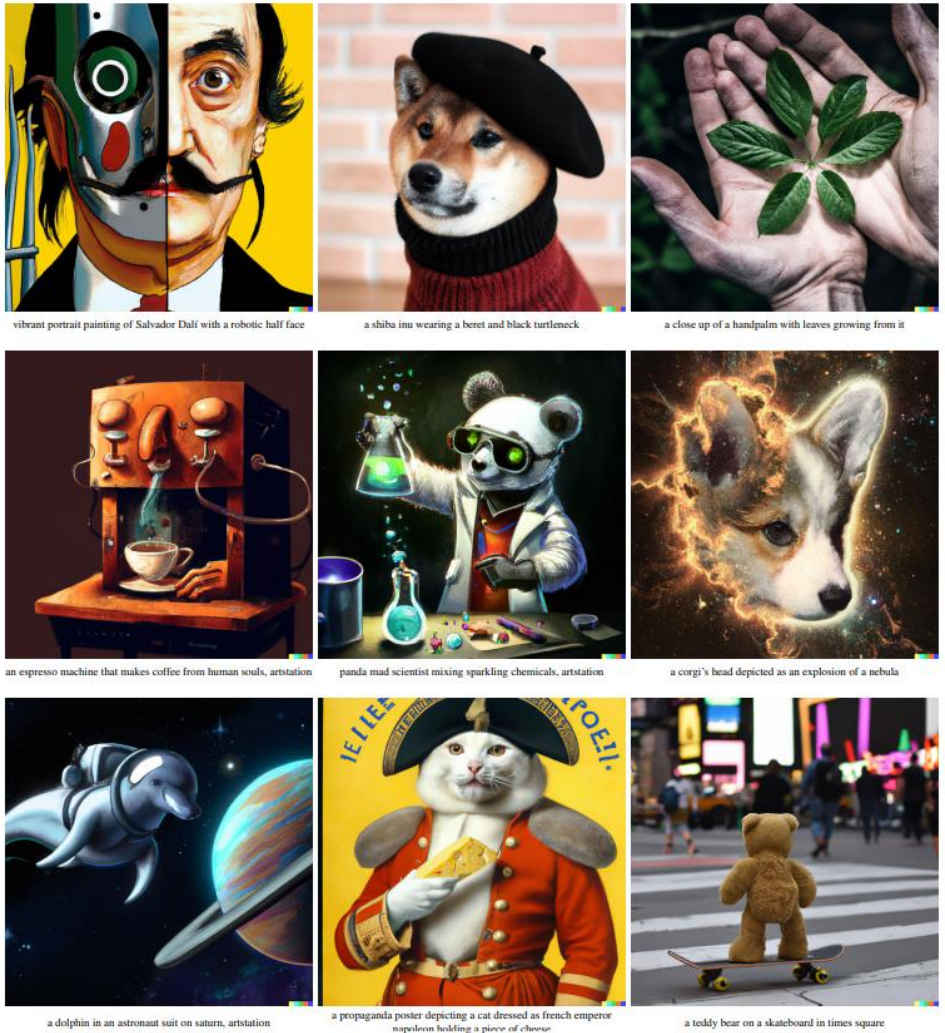
Text to Image Generation with GANs

- "Text to image" GAN model from 2016:

Figure 25: StackGANs are able to achieve higher output diversity than other GAN-based text-to-image models. Image reproduced from Zhang *et al.* (2016).

Text to Image Generation with Diffusion Models

- “Text to image” diffusion model from 2022 (Dalle 2):



Text to Image Generation with Diffusion Models

- "Text to image" diffusion model from 2022 (Dalle 2):

(a) a tapir made of accordion. a tapir with the texture of an accordion.

(b) an illustration of a baby hedgehog in a christmas sweater walking a dog

(c) a neon sign that reads "backprop". a neon sign that reads "backprop". backprop neon sign

(d) the exact same cat on the top as a sketch on the bottom

Figure 2. With varying degrees of reliability, our model appears to be able to combine distinct concepts in plausible ways, c anthropomorphized versions of animals, render text, and perform some types of image-to-image translation.

OG image prompt: "a robot holding a paint brush painting on an art stand"

https://arxiv.org/pdf/2102.12092.pdf
https://eugeneyan.com/writing/text-to-image/

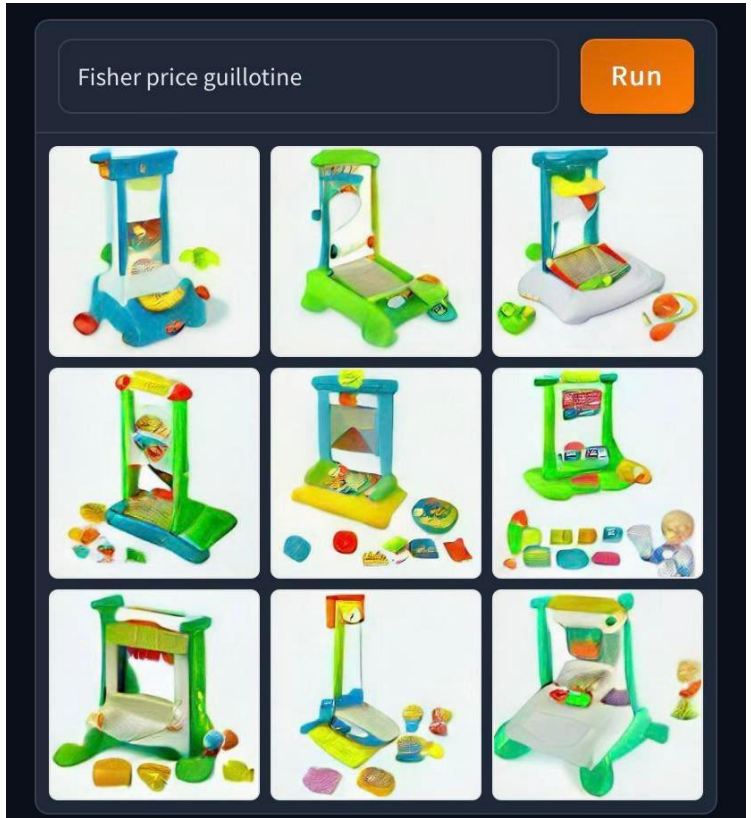
Text to Image Generation with Diffusion Models

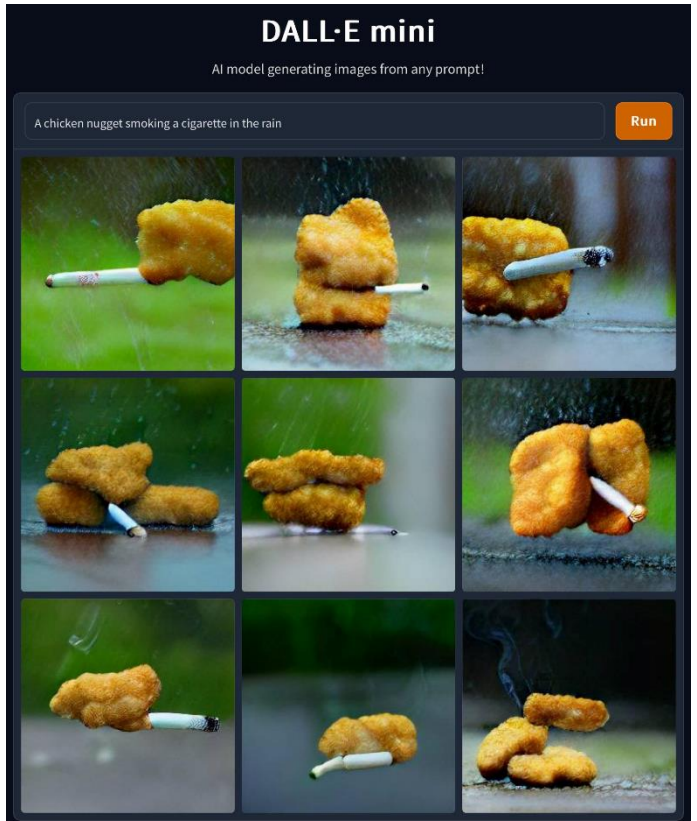
- "Text to image" diffusion model from 2022 (Dalle 2):
  - "Kermit the frog in…"

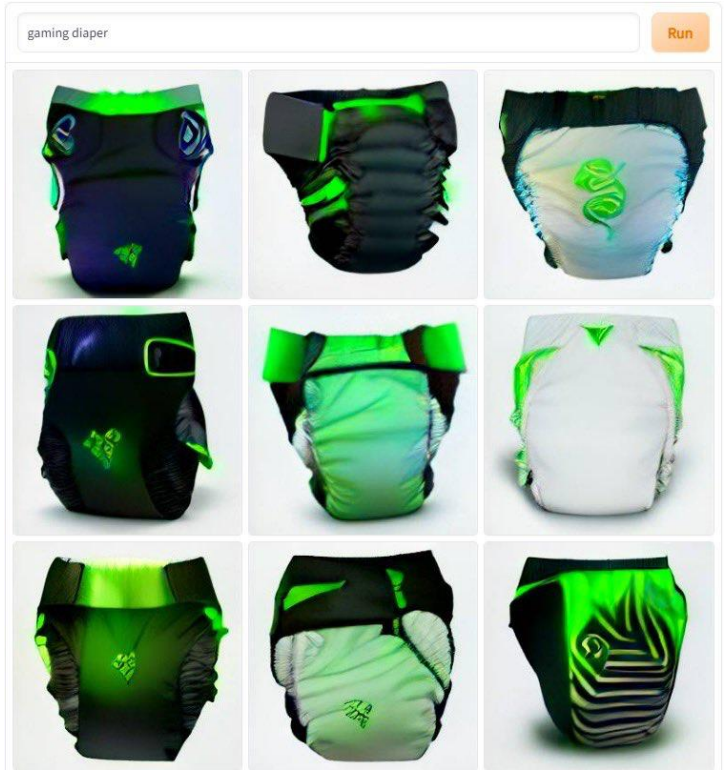
https://twitter.com/hvnslstangel/status/1531506455714492416?s=21&t=tfzr8YgT7Yaju-44RJat0A

Text to Image Generation with Diffusion Models

- Dalle 2 has a strict “G-rated” content policy.
  - And developed automatic systems to detect violations.
- Though did not stop people from making unrestricted versions.







Text to Image Generation with Diffusion Models

- "Text to image" diffusion model from 2022 (Imagen):

- More recent:
  - "Stable diffusion".
    - Open-source, can be run on standard computers.

https://arxiv.org/pdf/2102.12092.pdf
https://eugeneyan.com/writing/text-to-image/

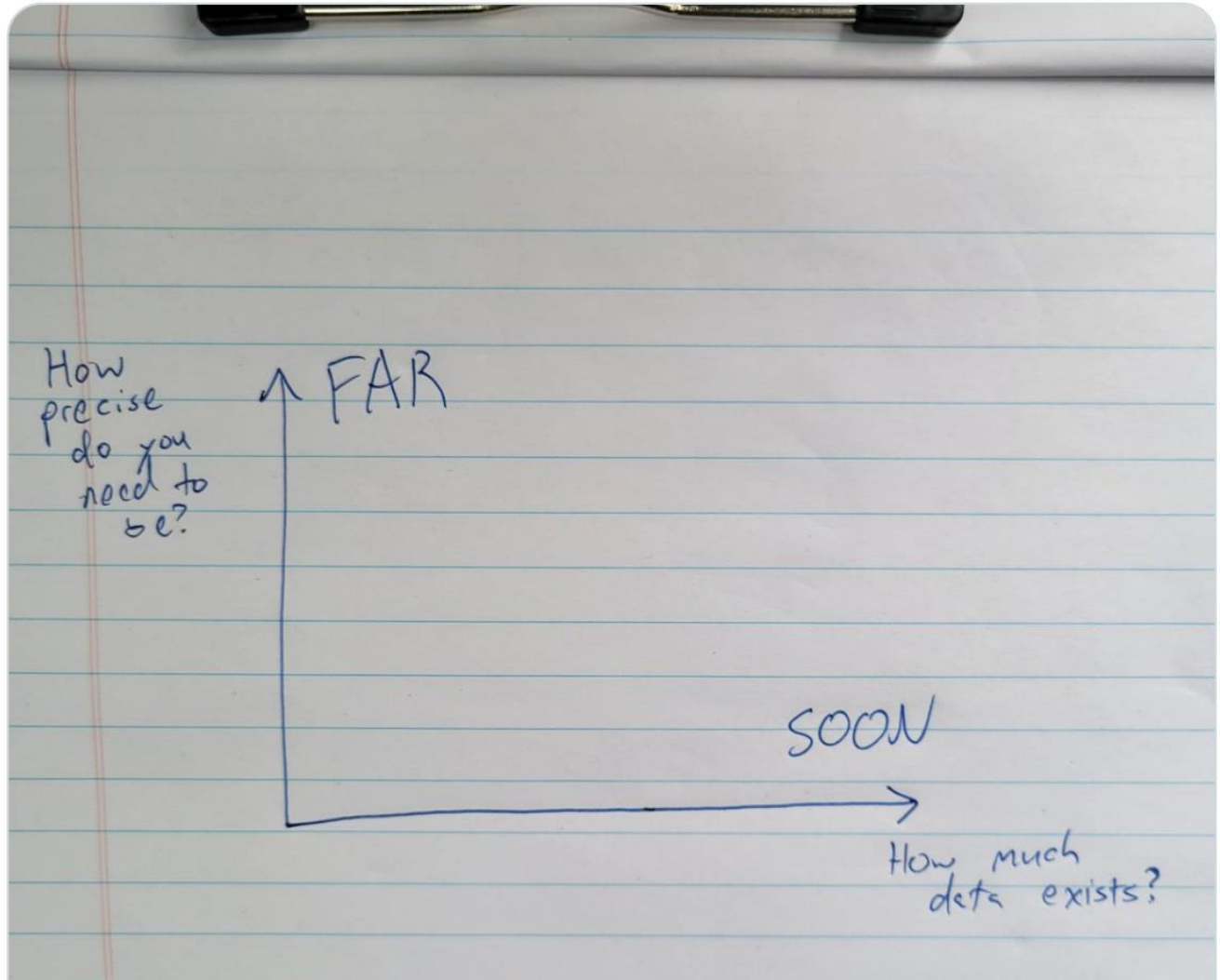
Next Topic: Brief Course Wrap-Up

Further CPSC Courses

- CPSC 330: “Applied Machine Learning”.
  - Some overlap in content, but focus is different:
    - Emphasis on “how to use packages”, and other steps of the data processing pipeline
- CPSC 422: “Intelligent Systems”.
  - Often covers a variety of related topics including reinforcement learning.
- CPSC 440: “Advanced Machine Learning”.
  - Intended as a sequel to this class, but not taught by me this year.
- CPSC 5XX courses:
  - If you are near the end of your degree with good grades, lots of cool stuff.

Joshua Achiam
@jachiam0

"How did we get AI art before self-driving cars?" IMHO this is the single best heuristic for predicting the speed at which certain AI advances will happen.



Concluding Remarks

- I took my first AI/ML course in 2002.
  - I have never been as excited about what ML can do than in 2022.
- But, there is a lot of bull-shit out there too!
  - Do not believe everything you hear, and try to avoid producing non-sense.
  - "Calling Bullshit in the Age of Big Data":
    - https://www.youtube.com/playlist?list=PLPnZfvKID1Sje5jWxt-4CSZD7bUI4gSPS
- Thank you for your patience.
  - Andreas' first time teaching and my first time parenting.
- Good luck with finals/projects and the next steps!