
DA PALAVRA
À IMAGEM
WORKSHOP – CASA FIRJAN – FESTIVAL FUTUROS POSSÍVEIS 26/11 - 10h às 13h
MATTEO MORICONI
VFXRIO
LUIZ VELHO
VISGRAPH
BERNARDO ALEVATO
PUC-RIO
ACMSIGGRAPH
Casa Firjan apresenta:
Festival Futuros Possíveis
2022
IMAGINAÇÃO PARA CRIAR FUTUROS
25 e 26/11
On-line e com oficinas presenciais na Casa Firjan

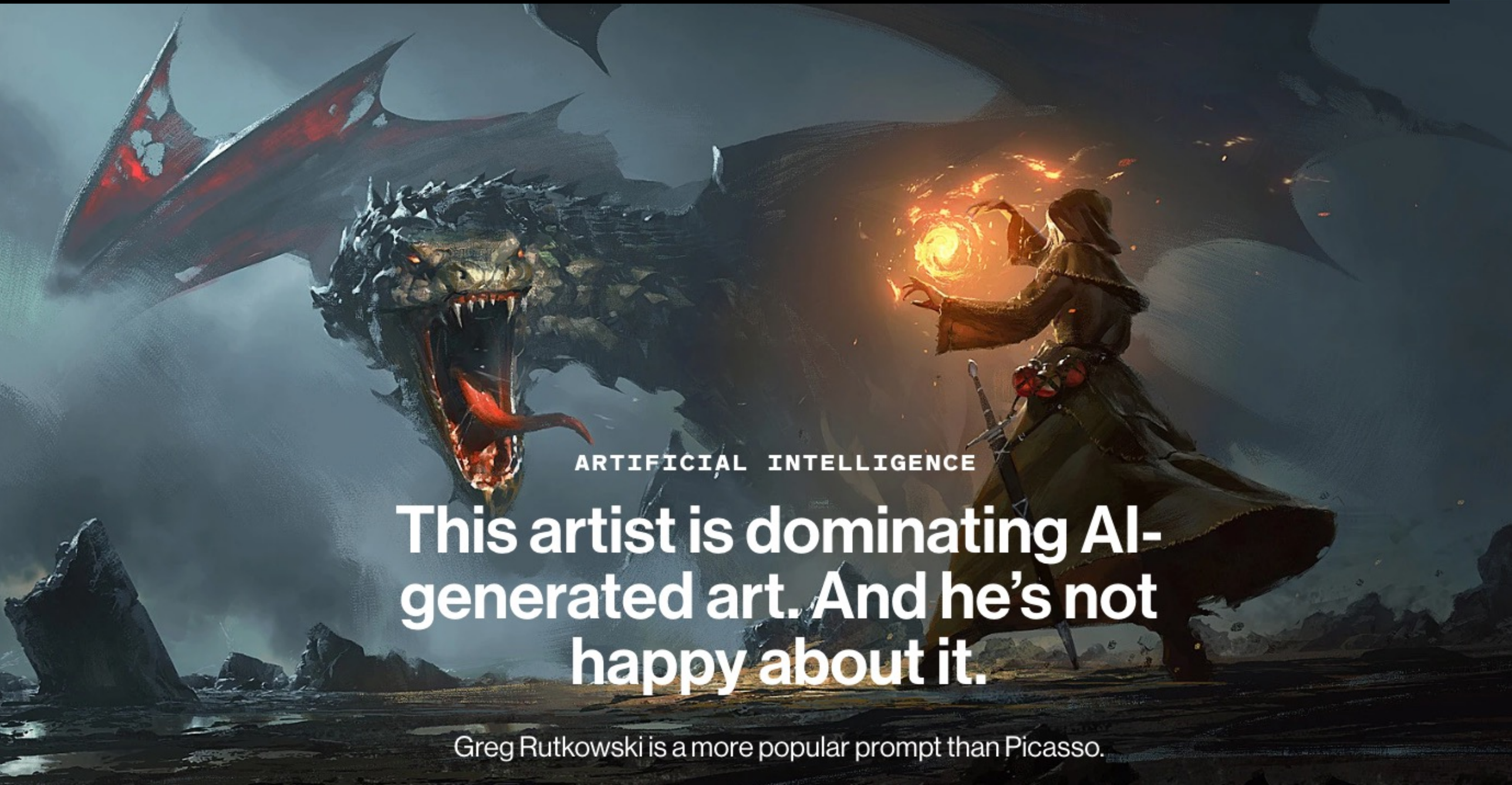
The Art in AI

Luiz Velho
IMPA

Midjourney Teaser

Jason Allen via Discord

Stable Diffusion Fuss



ARTIFICIAL INTELLIGENCE

This artist is dominating AI-generated art. And he's not happy about it.

Greg Rutkowski is a more popular prompt than Picasso.

Source of Inspiration

- The Secret Pass - Eagle Nest

Personal work featured in Rutkowski's ArtStation portfolio.

MS TECH VIA STABLE DIFFUSION

Rutkowski’s name has been used as a prompt around 93,000 times!

The Big Debate

Hot from the Press

WIRED

The Joy and Dread of AI Image Generators Without Limits

AI art looks way too European Vox

DALL-E and other models keep making art that ignores traditions from the rest

WILL KNIGHT BUSINESS SEP 21, 2022 7:00 AM WIRED

This Uncensored AI Art Tool Can Generate Fantasies—and Nightmares

The New York Times | https://www.nytimes.com/2022/10/21/technology/generative-ai.html

THE SHIFT

A Coming-Out Party for Generative A.I., Silicon Valley's New Craze

A celebration for Stability AI, the start-up behind the controversial Stable Diffusion image generator, re

TECHPERSONAL TECHPERSONAL TECHNOLOGY; JOANNA STERN THE WALL STREET JOURNAL.

Ask an AI Art Generator for Any Image. The Results Are Amazing —and Terrifying.

The New York Times | https://www.nytimes.com/2022/10/21/technology/ai-generated-art-jobs-

THE SHIFT

A.I.-Generated Art Is Already Transforming Creative Work

Tech Talk

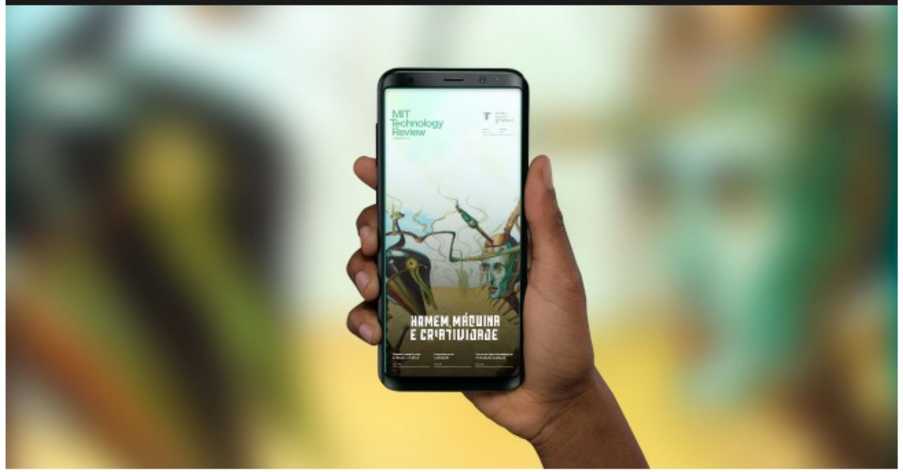
MIT Technology Review

Publicado por TEC

EDIÇÃO - OUT 2022

Edição Digital | Generative AI

Homem, máquina e criatividade.



MIT Technology Review

Featured Topics Newsletters Events Podcasts

ARTIFICIAL INTELLIGENCE

Get ready for the next generation of AI

MIT Technology Review

ARTIFICIAL INTELLIGENCE

The dark secret behind those cute AI-generated animal images

Google Brain has revealed its own image-making AI, called Imagen. But don't expect to see anything that isn't wholesome.

From: **MIT Technology Review Brasil**

Imagens geradas por IA são vetadas por empresas

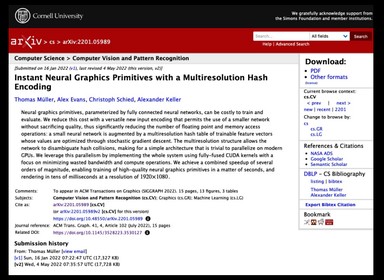
O QUE É?

Empresas vetam imagens geradas por inteligência artificial.

Midjourney e DALL-E são alguns dos programas usados.

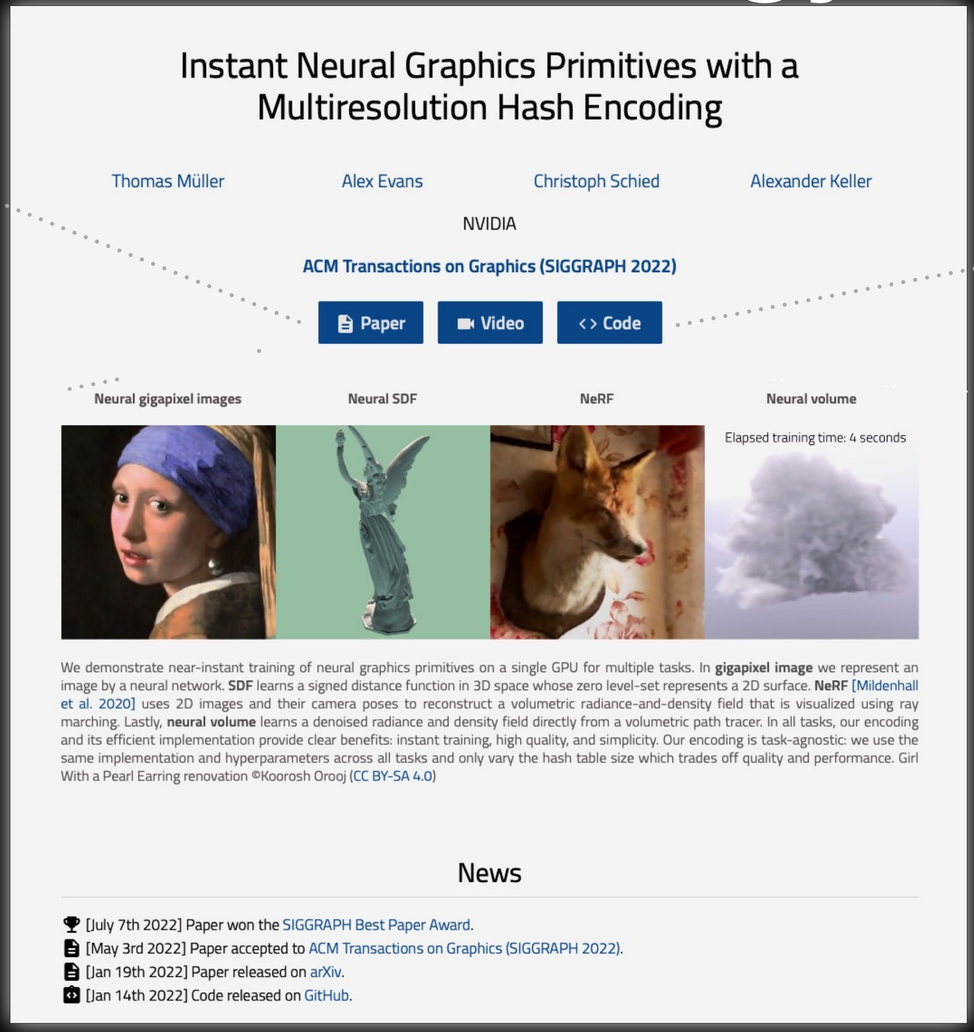
The Era of Machine Learning

A New Way for Science & Technology

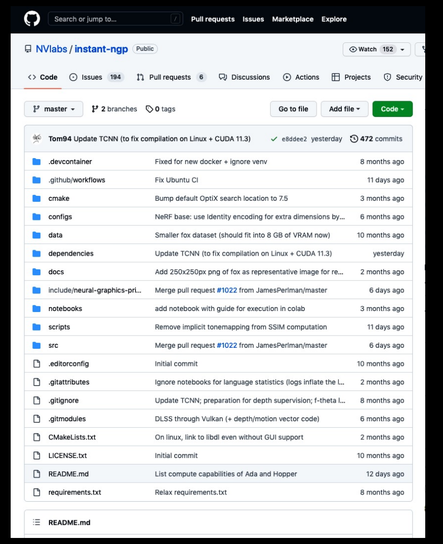


arXiv publication

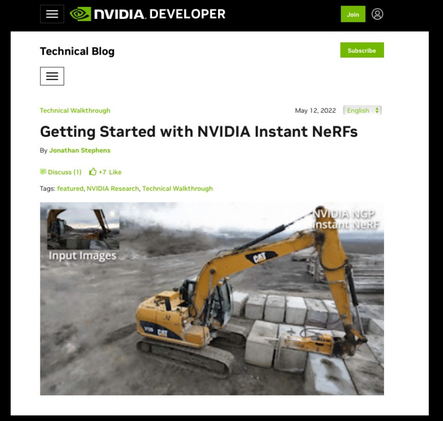
journal paper



project page



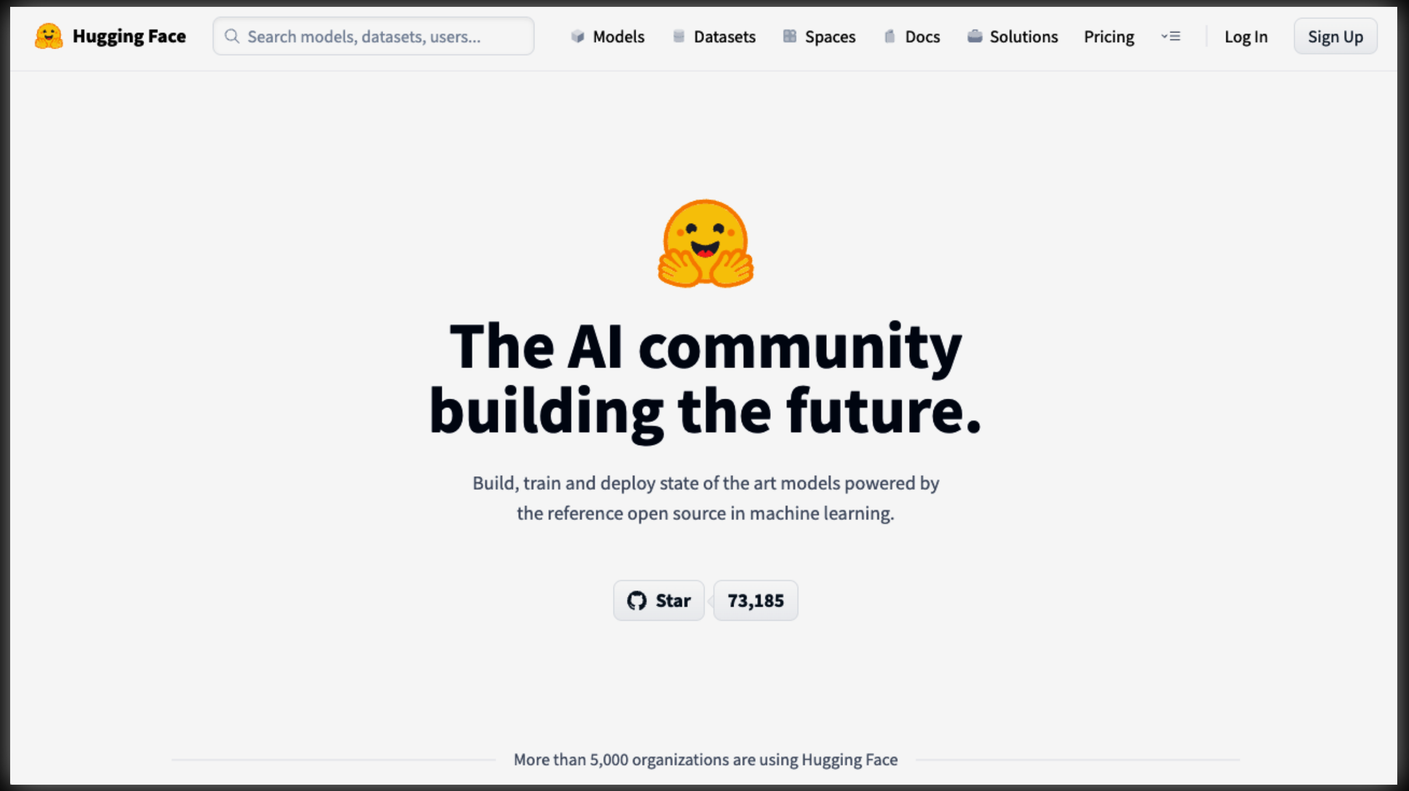
GitHub repository

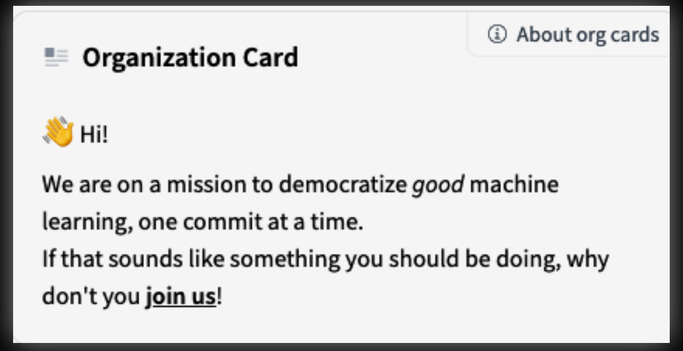


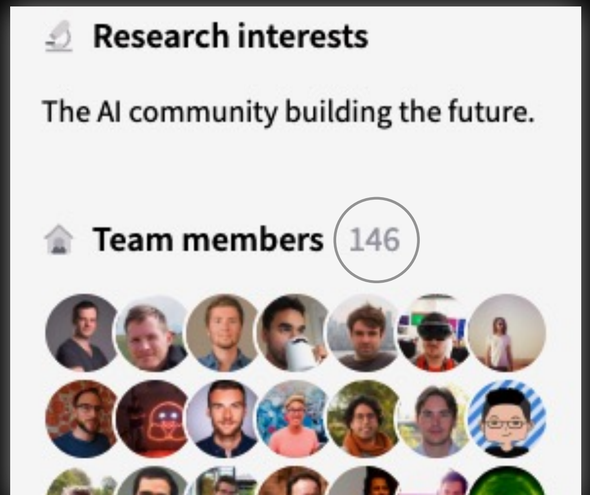
blog posts

Open Source Community

- One Good Example

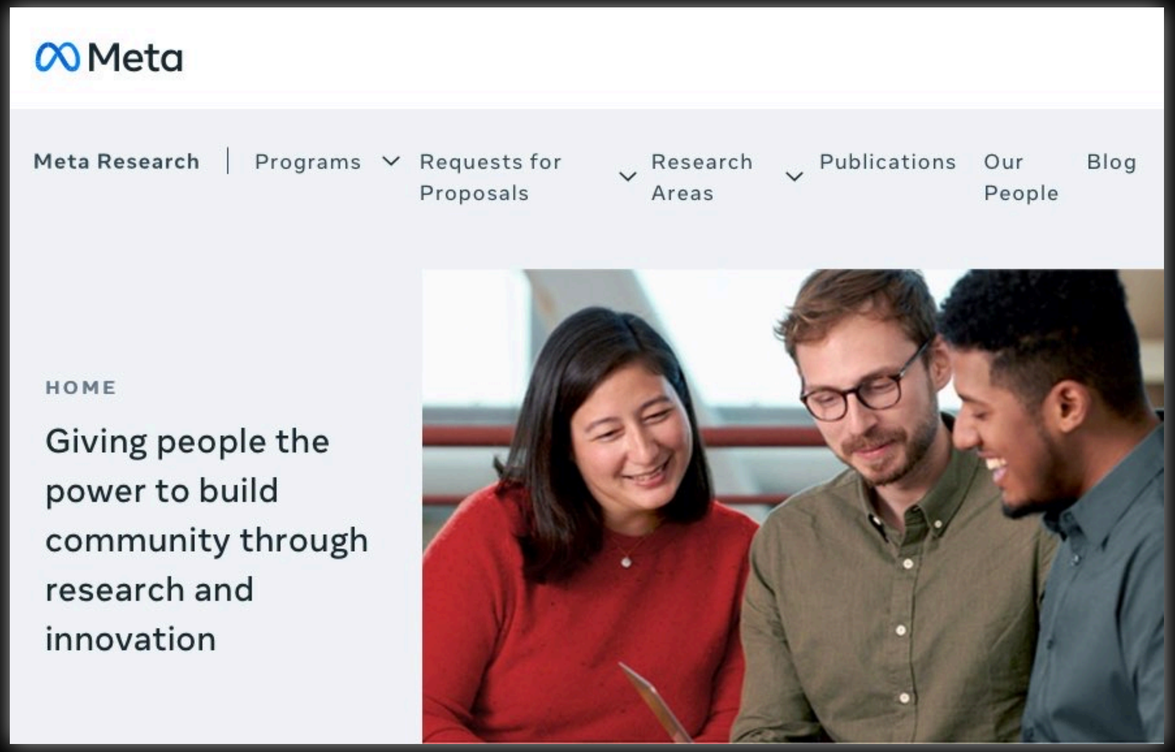


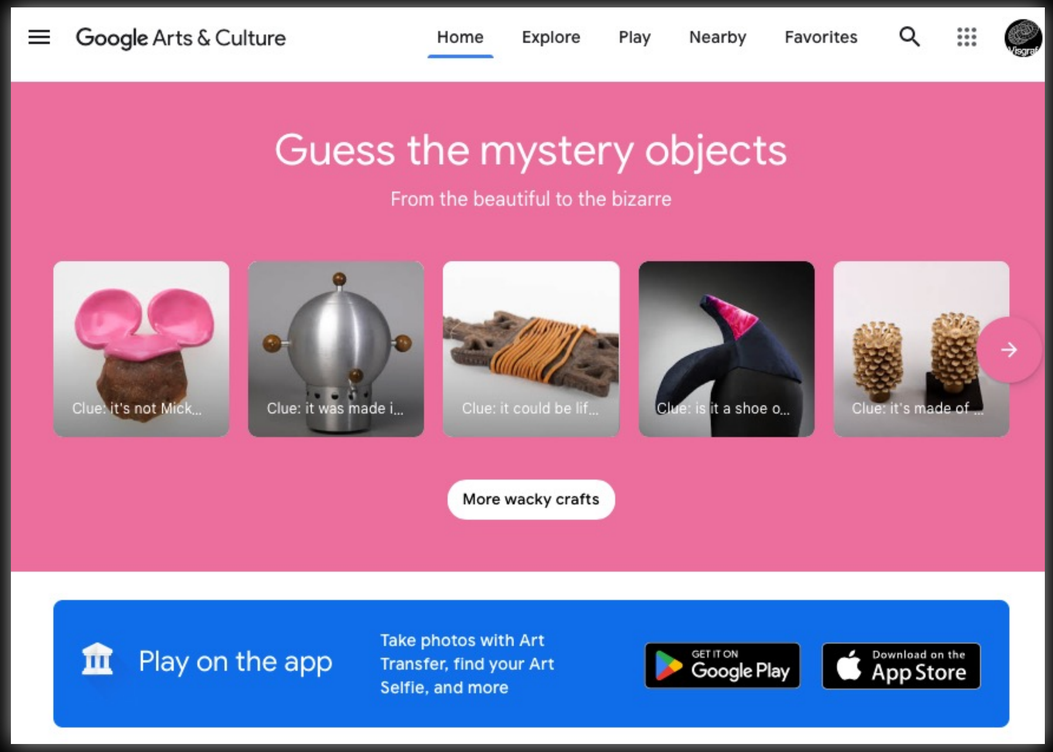




Outsourcing to Society

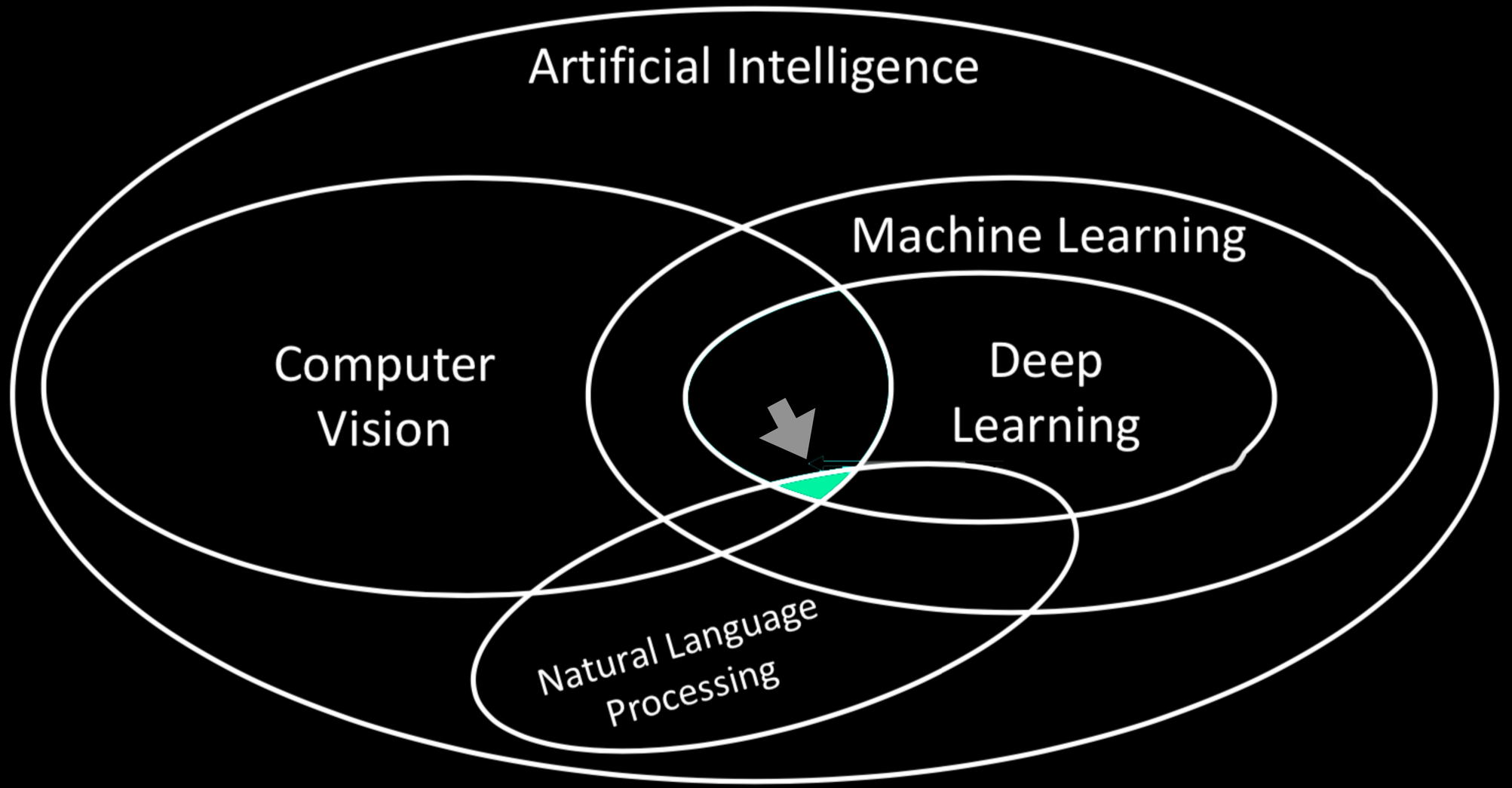
- Google, Meta and many more…





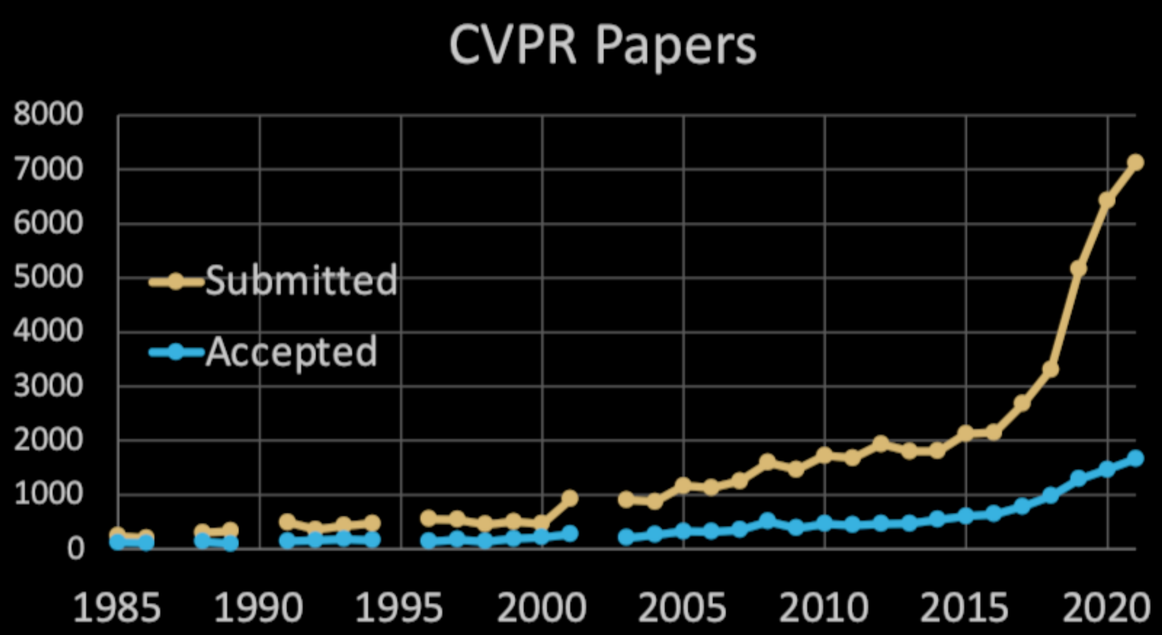
Deep Learning (R)evolution

Historical Context

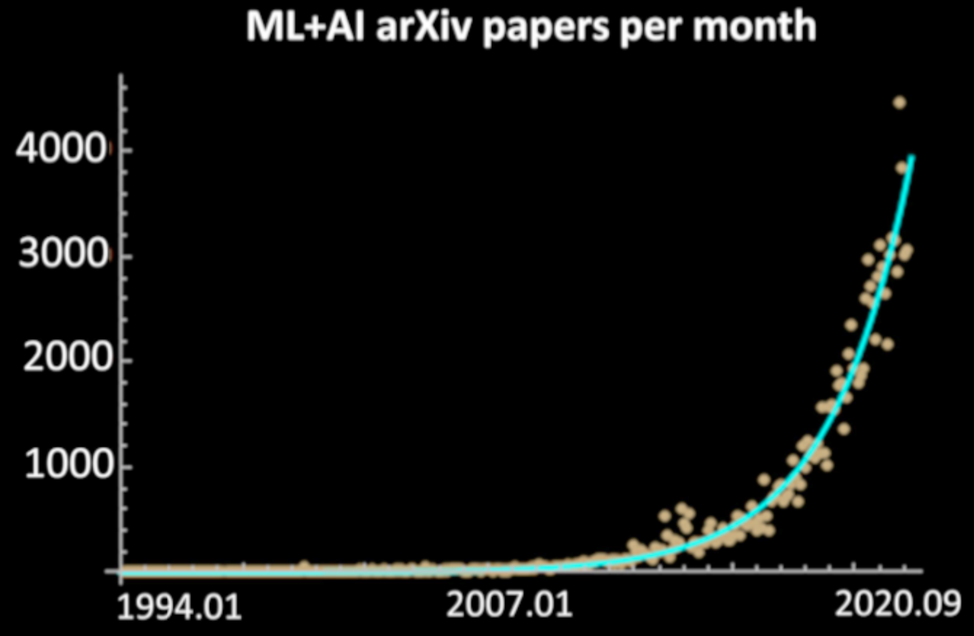


Exponential Growth

2012 to Present: Deep Learning Explosion



Publications at top Computer Vision conference



arXiv papers per month (source)

Image Serendipity

Metadata from Images

- Image Analisys

Image Classification

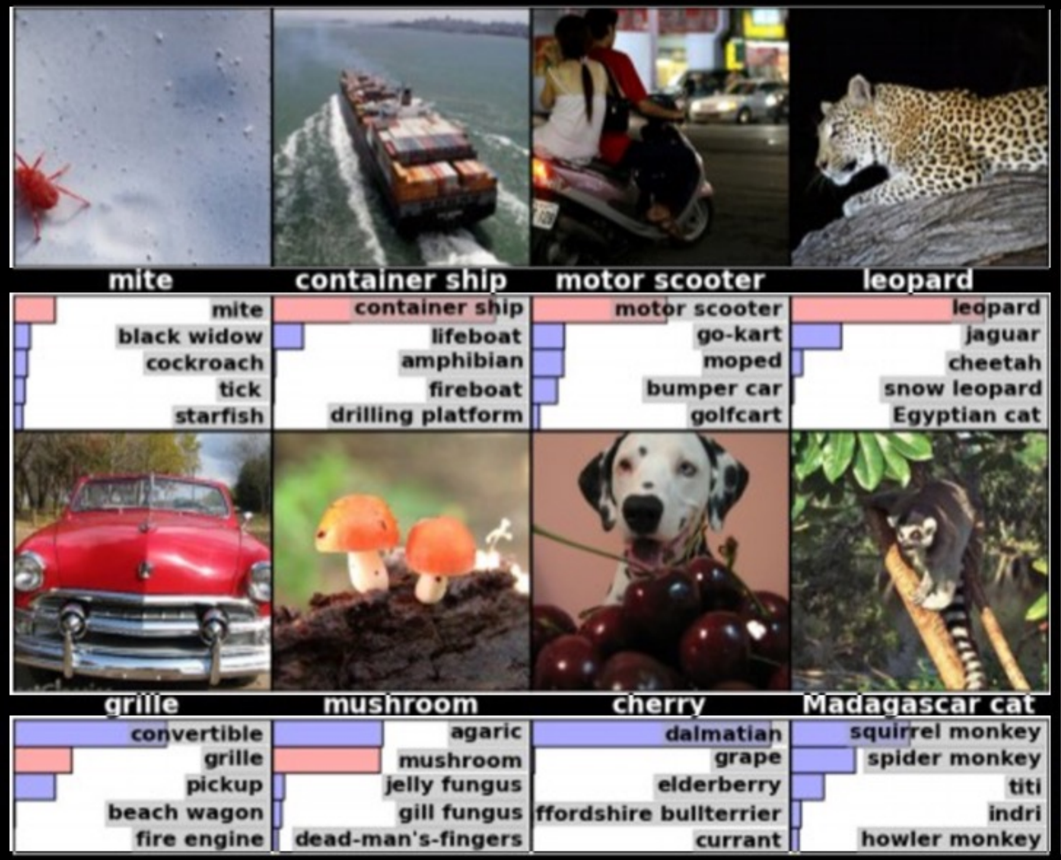
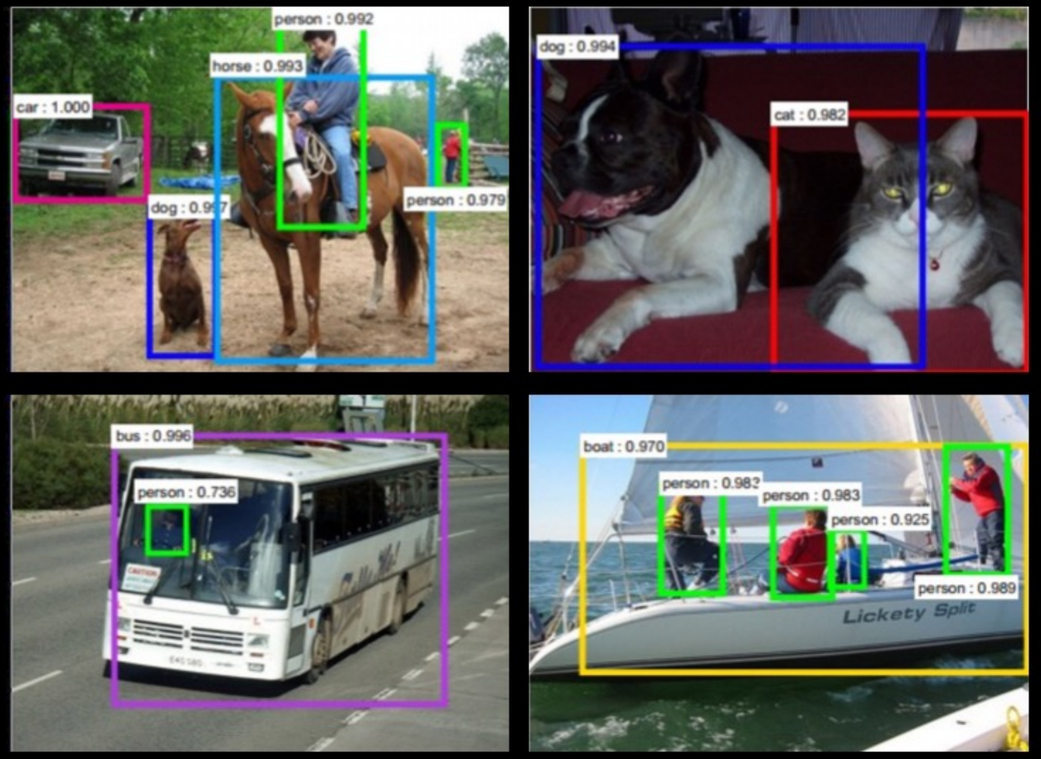


Image Retrieval

Features in Images

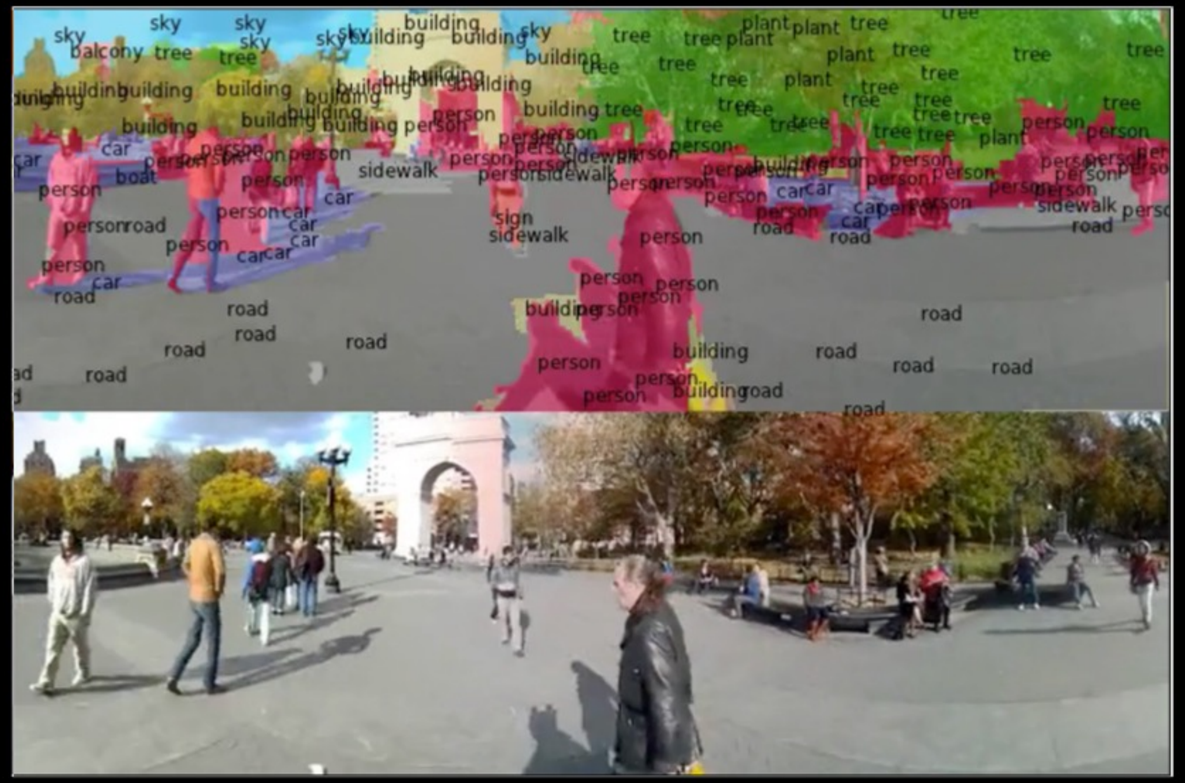
- Image Structuring

Object Detection



Ren, He, Girshick, and Sun, 2015

Image Segmentation

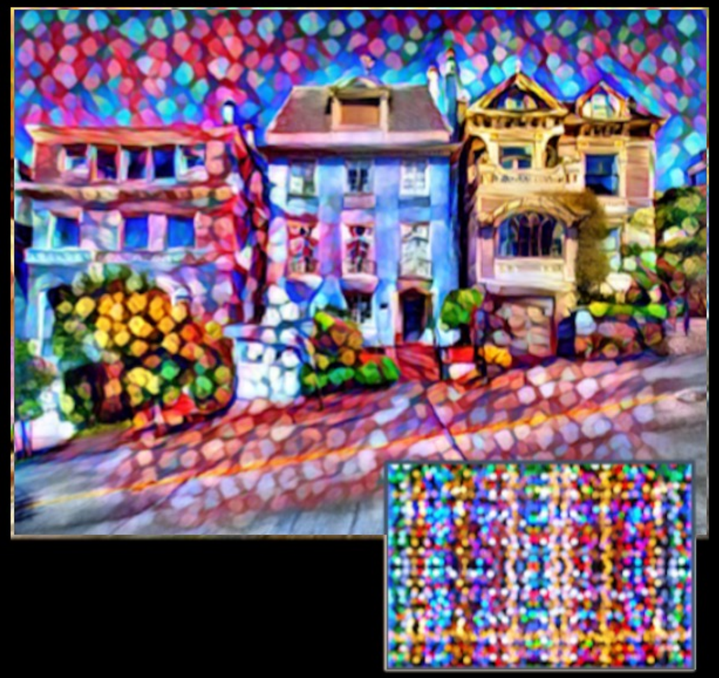


Fabaret et al, 2012

Image from Images

- Style Transfer

Original Image



Style A

Style B

Image-to-Text

- Image Captioning

A white teddy bear sitting in the grass

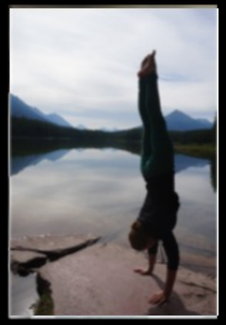
A man in a baseball uniform throwing a ball

A woman is holding a cat in her hand

Vinyals et al, 2015
Karpathy and Fei-Fei, 2015

A man riding a wave on top of a surfboard

A cat sitting on a suitcase on the floor



A woman standing on a beach holding a surfboard

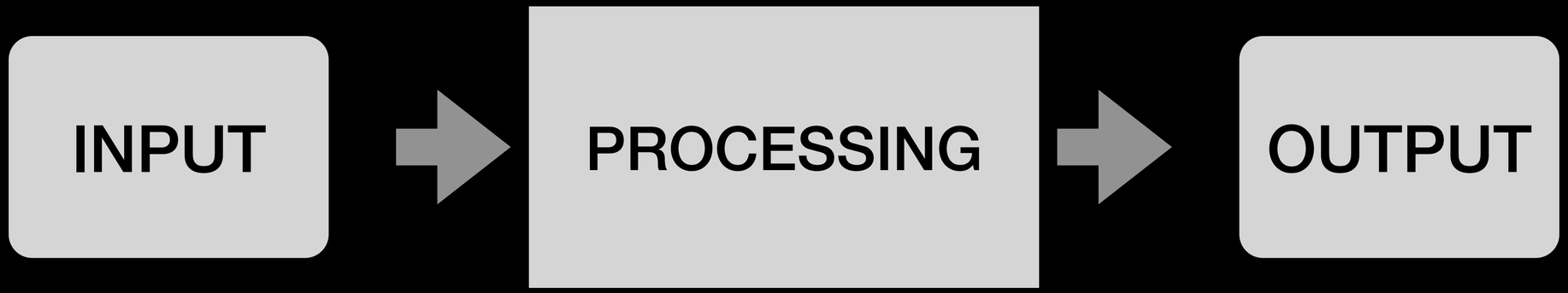
Text-to-Image

- Image Synthesis

The Mathematics of Machine Learning

Computational Applied Mathematics

- Computation Model

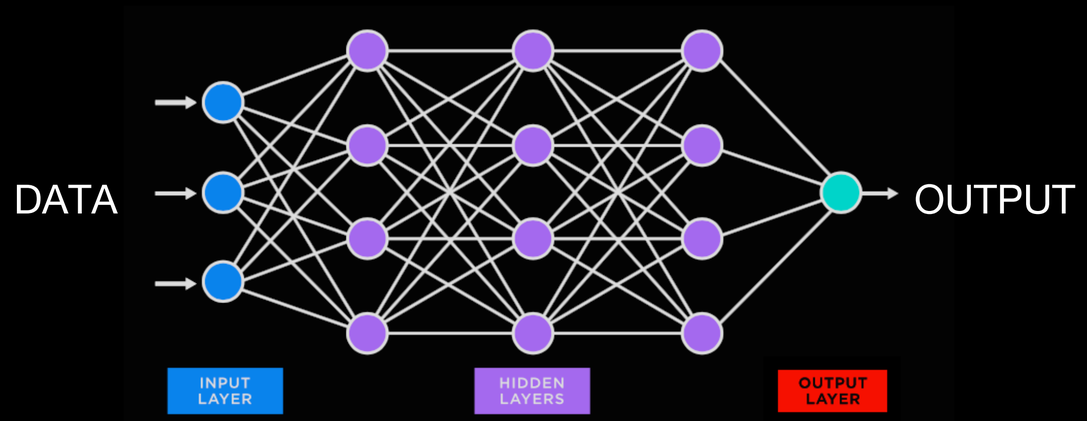


Old, New Way

- Algorithm Code

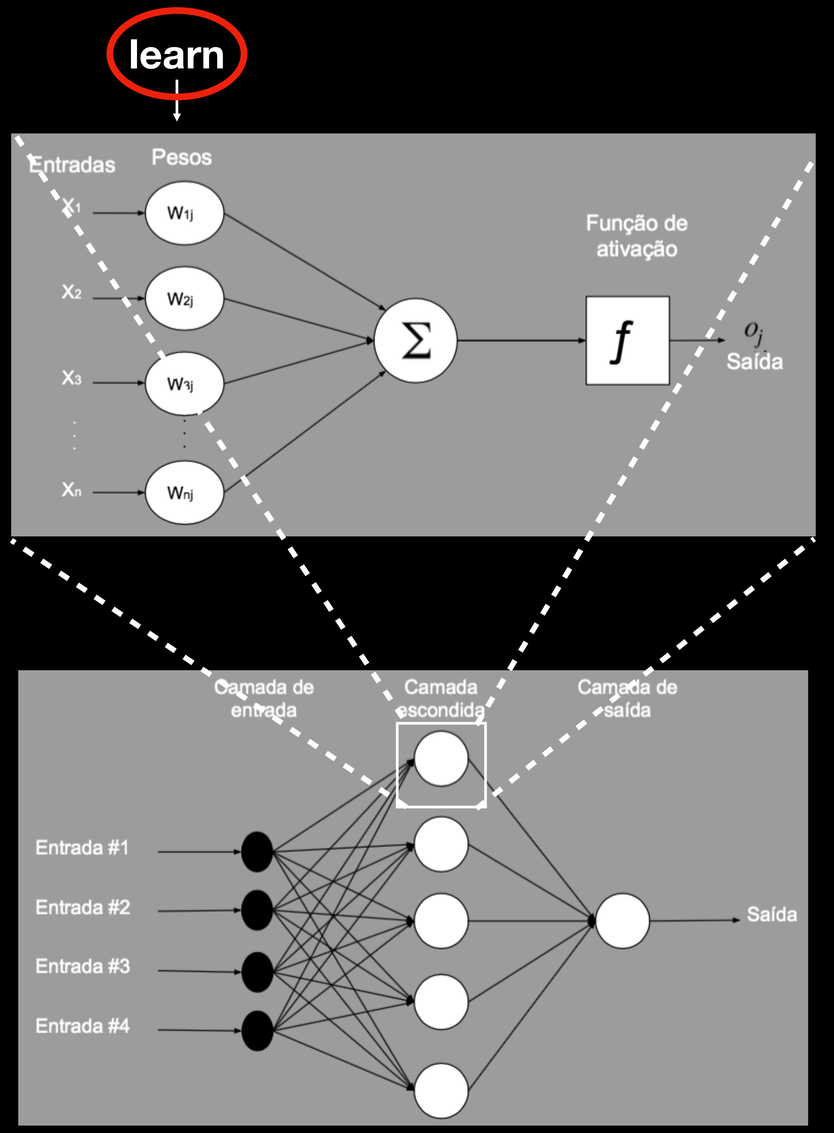
```
Step 1: Input M1,M2,M3,M4
Step 2: GRADE ← (M1+M2+M3+M4)/4
Step 3: if (GRADE <50) then
                Print "FAIL"
        else
                Print "PASS"
        endif
```

- Network Architecture



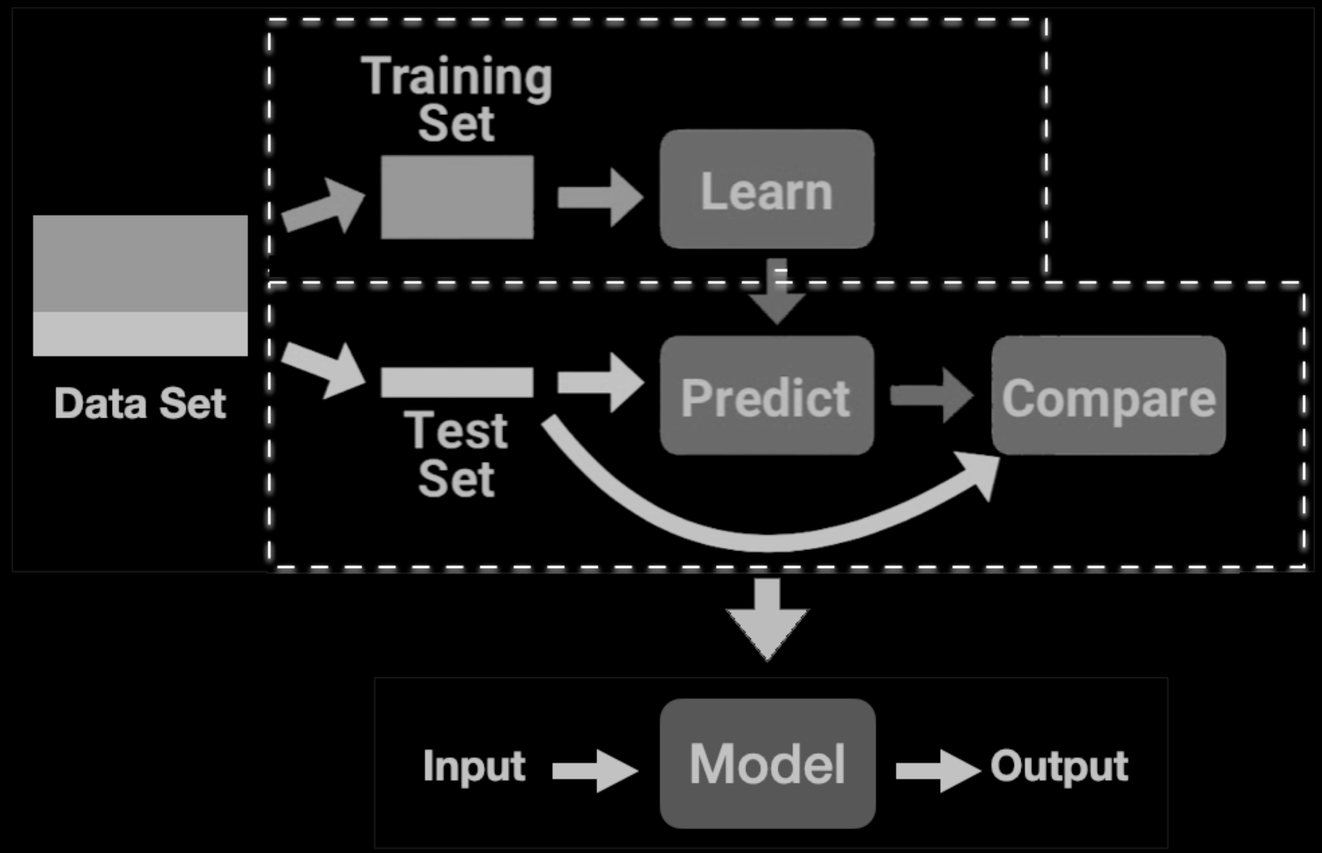
Neural Networks

- Perceptron (operator)
- Neural Nets (composition)



Learning from Data

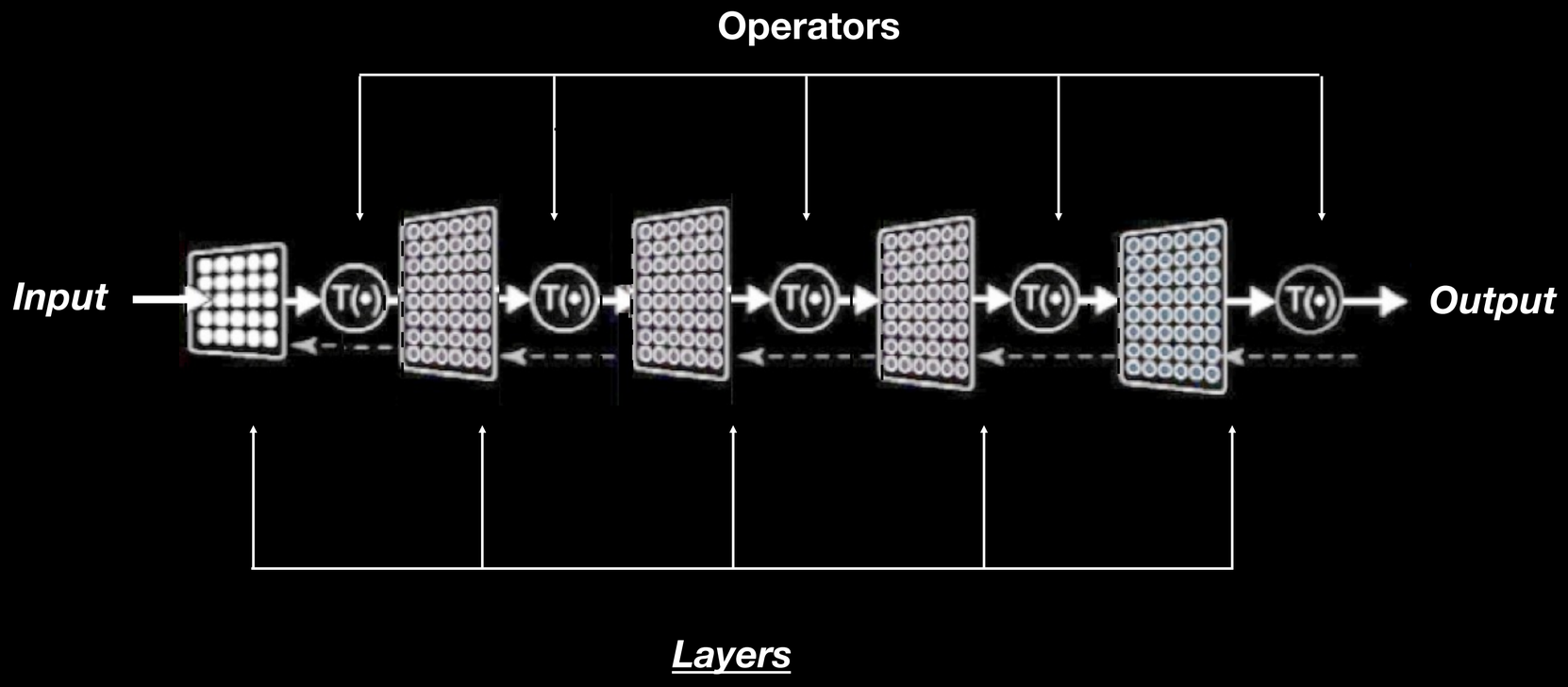
- Stages
  1. Train
  2. Validate
  3. Inference



Deep Neural Network Architectures

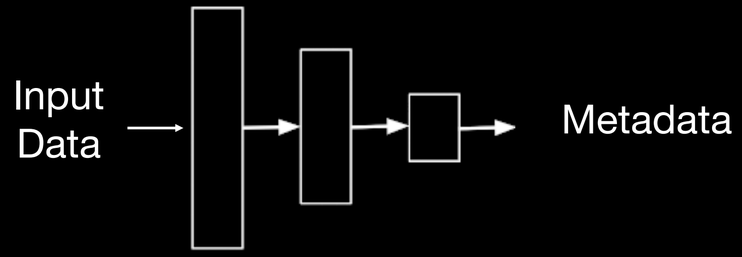
Deep Neural Networks

- Multi Layer Architecture

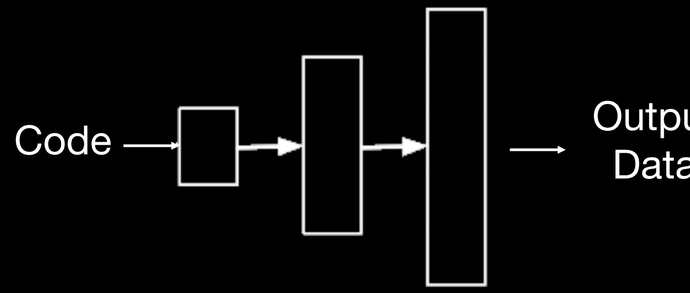


Types of Neural Networks

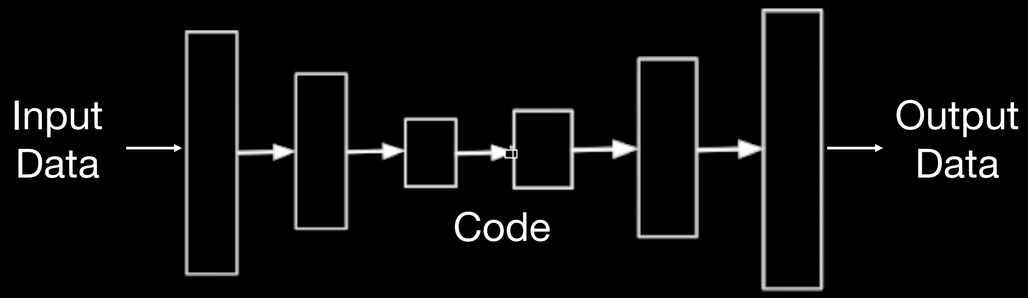
- Analysis



- Synthesis

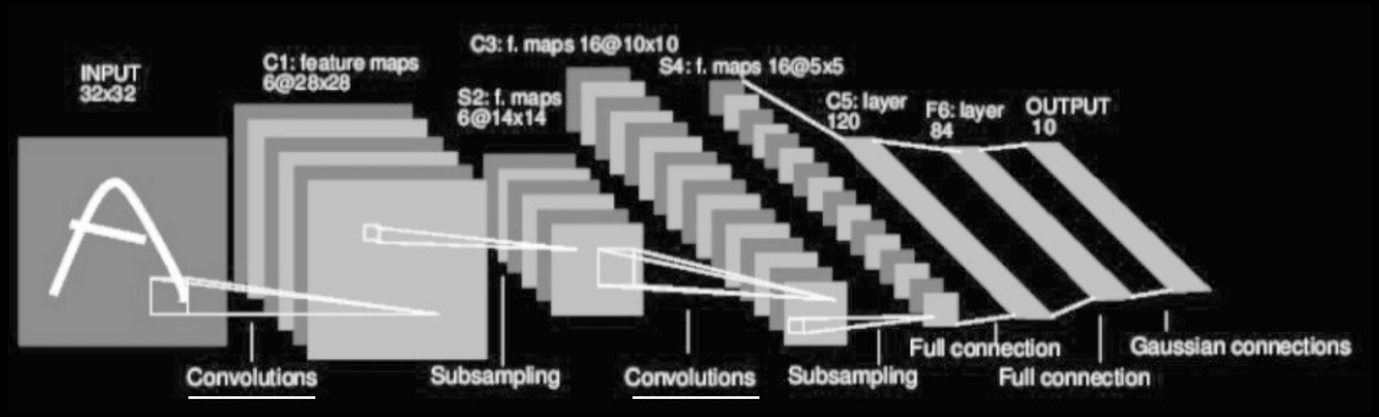


- Full

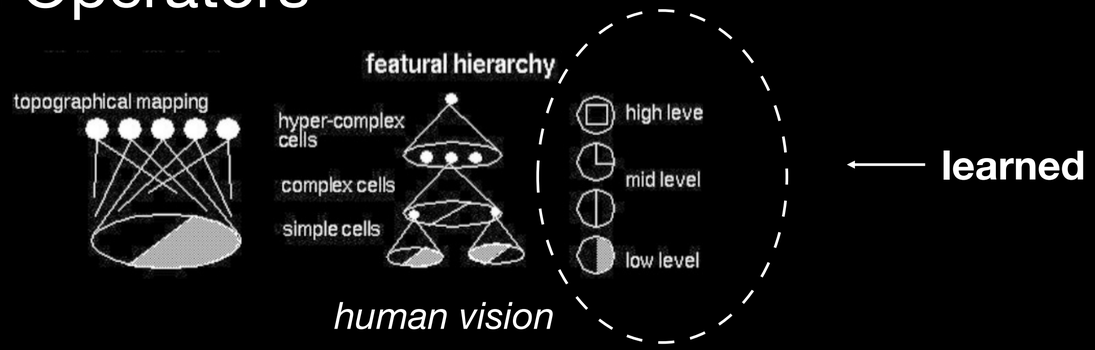


Convolutional Networks

- Learn Features (*Image*)
- Deep Convolutional Neural Networks

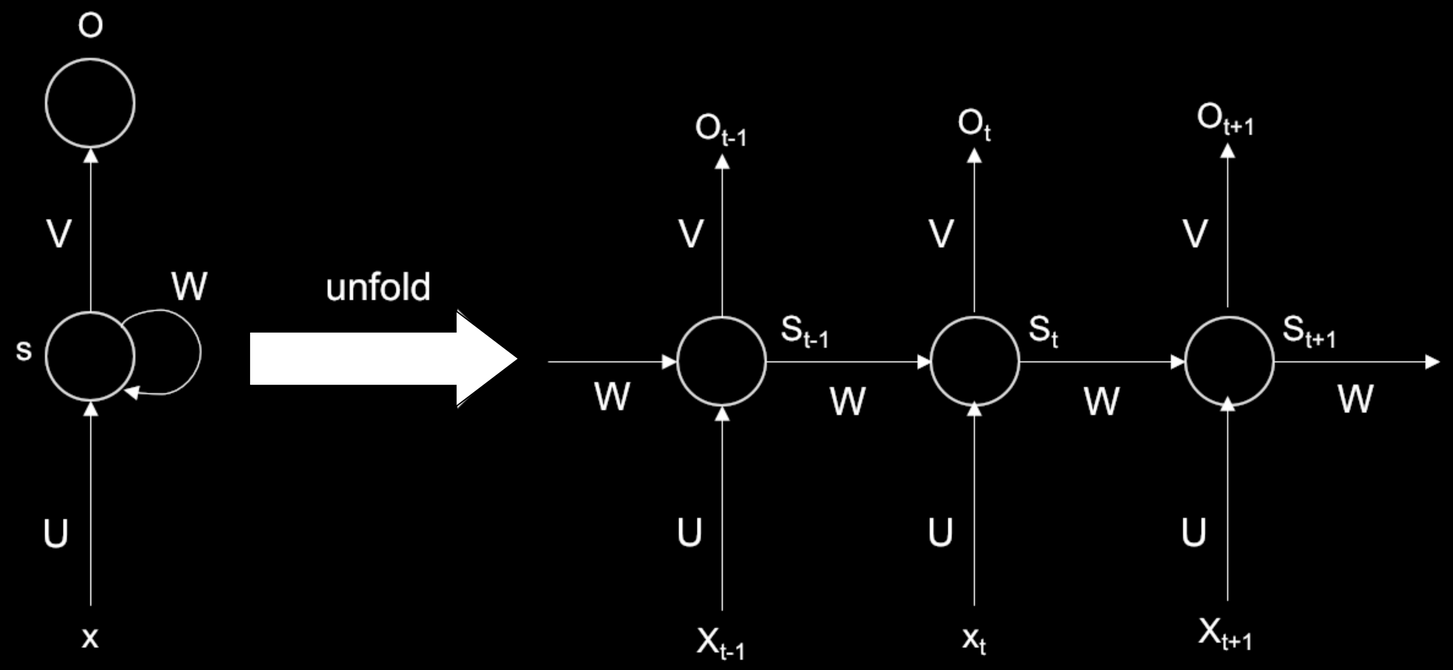


- Convolution Operators



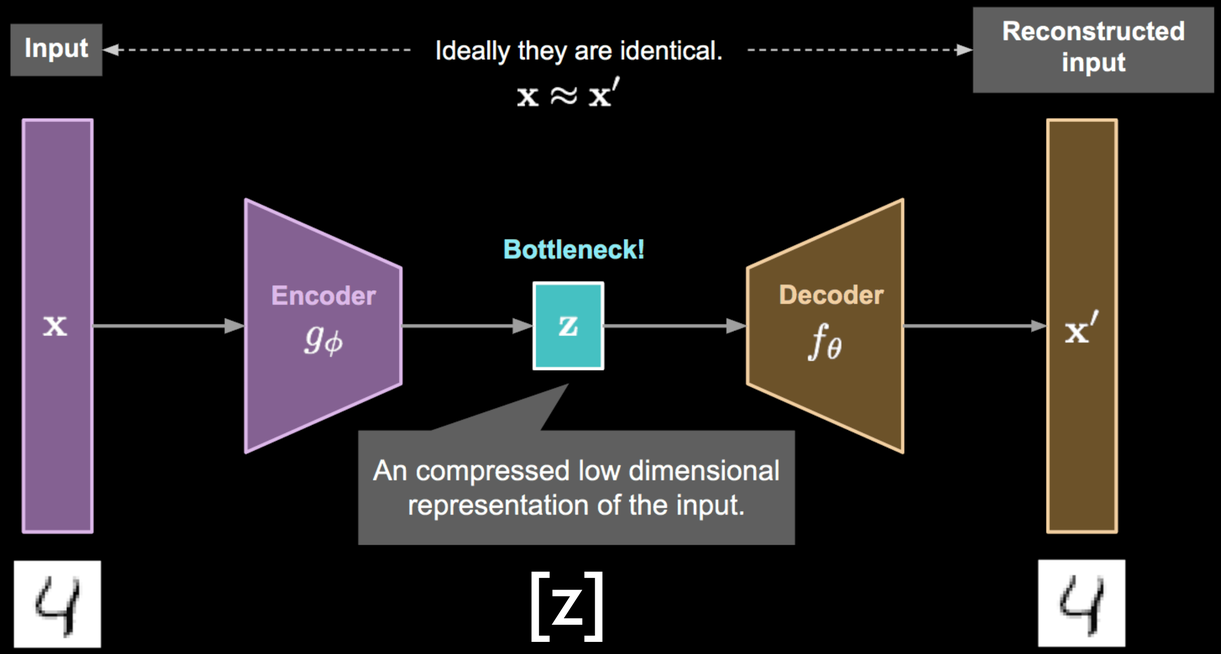
Recurrent Networks

- Learn a Sequence Process (*Text*)



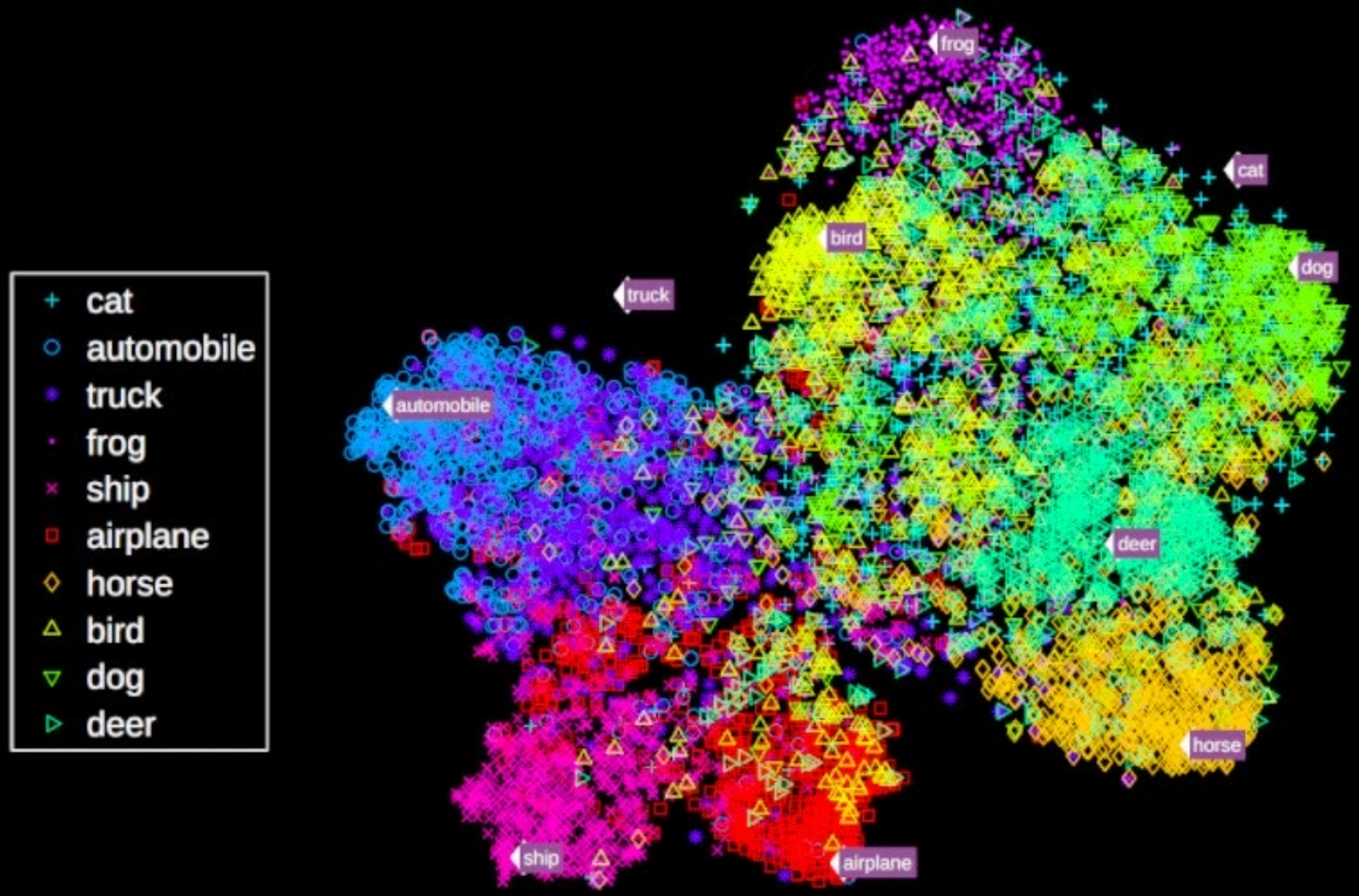
(Variational) Auto-Encoders

- Learn Representations (*General Probability Distributions*)



Natural Language Models

Model

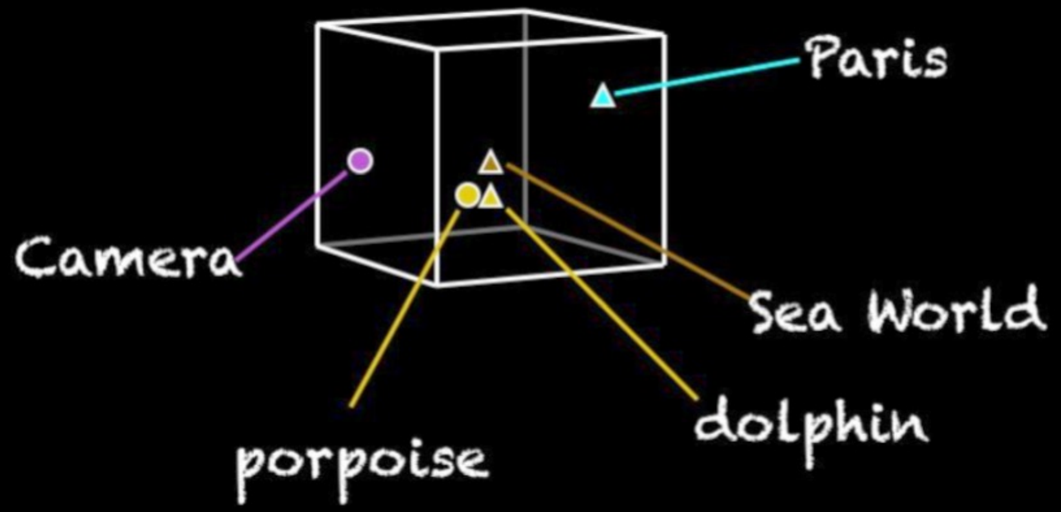


Semantic Structure

Model is a Subspace in a High Dimensional Space

Embeddings of Vector Spaces

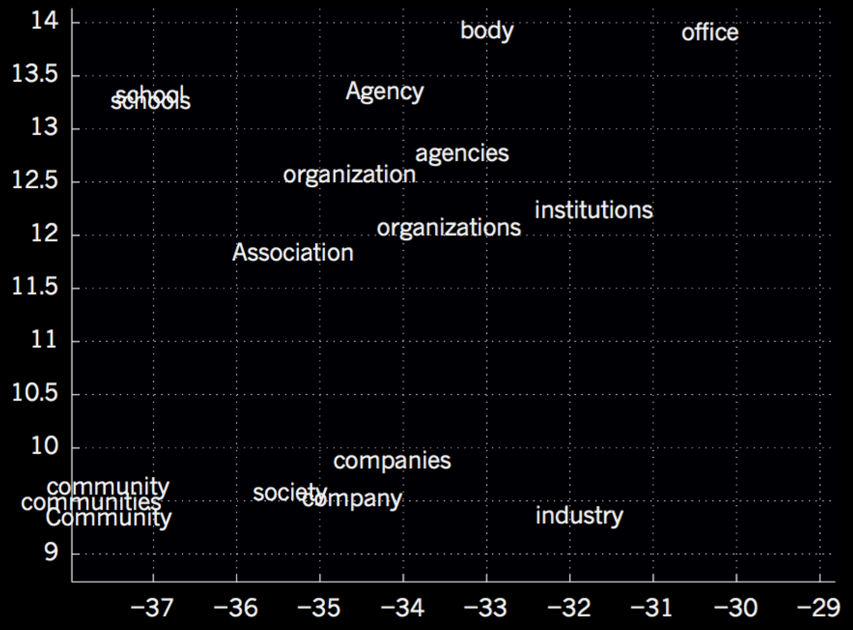
- Embedding Space (100D - 1000D)

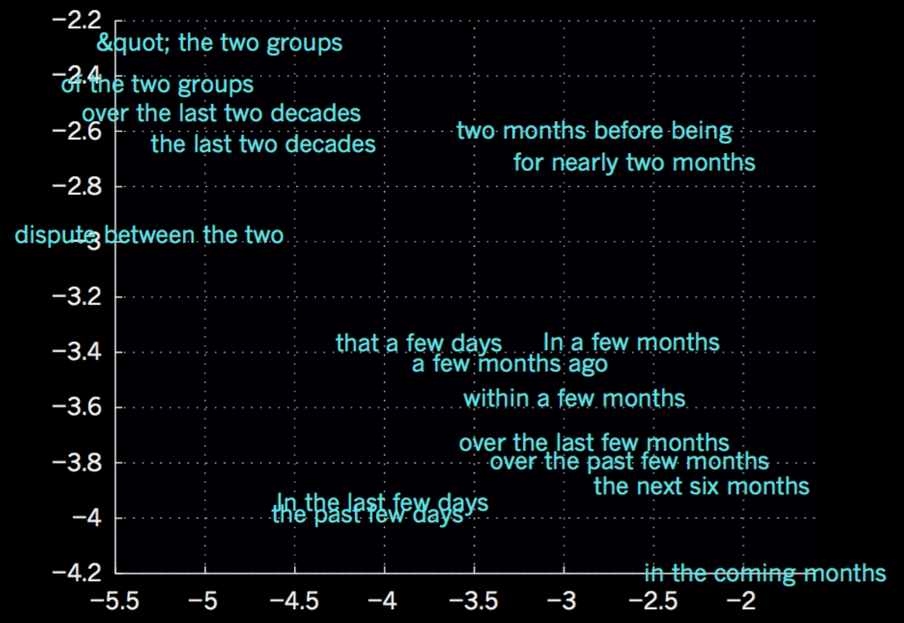


Embedding Function: A look-up-table that maps sparse features into dense floating point vectors.

Word - 2 - Vect

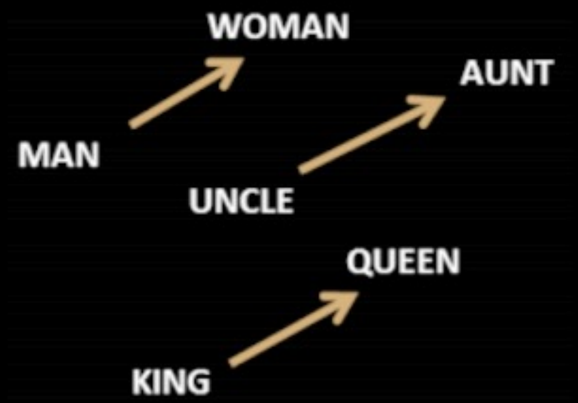
- Semantically related words are close in vector space





Analogies / Operations

- Differences between words ~ Structure



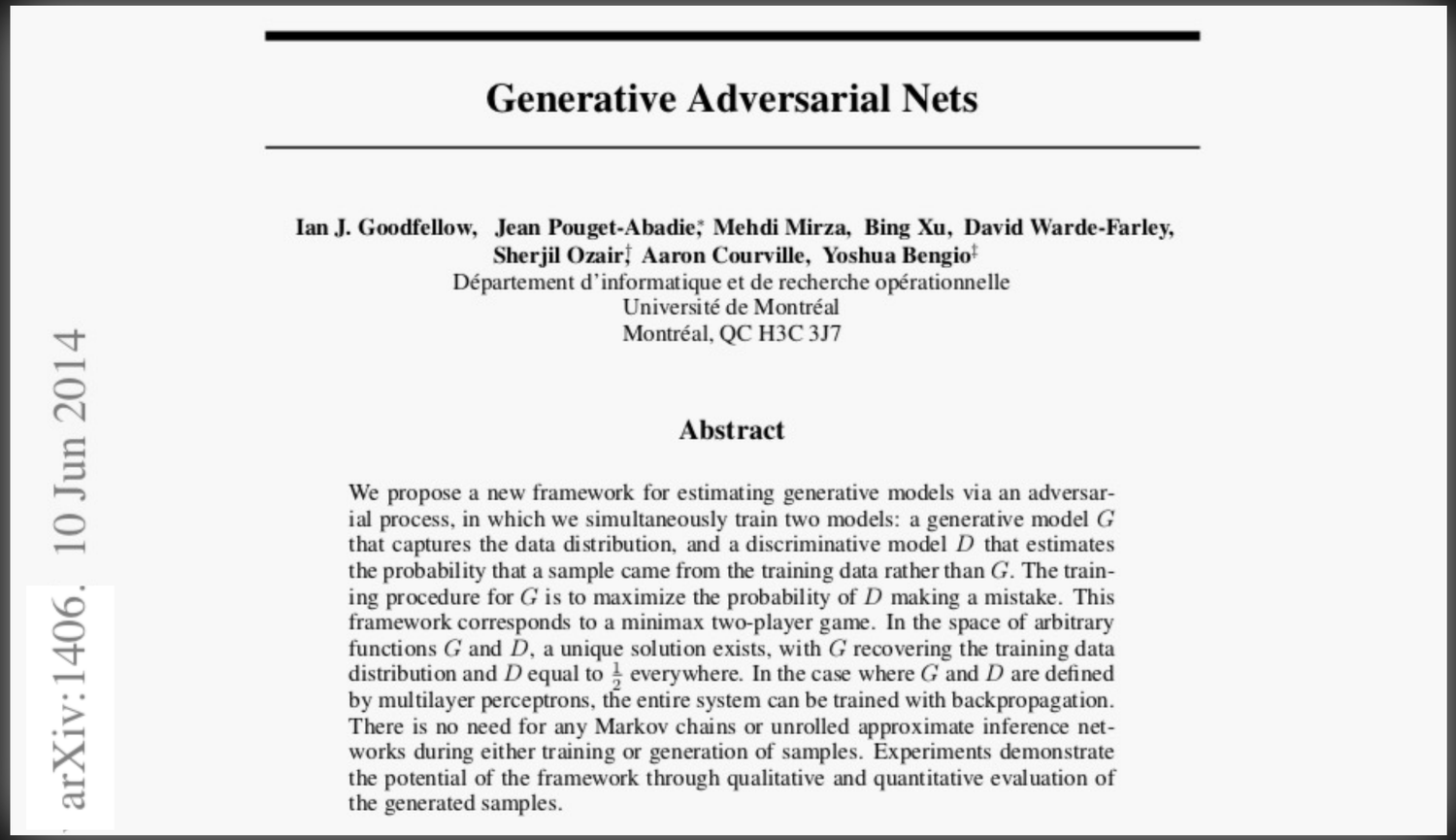
$$W(\text{``woman "}) - W(\text{``man "}) \simeq W(\text{``aunt "}) - W(\text{``uncle "})$$

$$W(\text{``woman "}) - W(\text{``man "}) \simeq W(\text{``queen "}) - W(\text{``king "})$$

Generative Models

The First Paper (2014)

- Ian Goodfellow



Generative Adversarial Nets

Ian J. Goodfellow, Jean Pouget-Abadie*, Mehdi Mirza, Bing Xu, David Warde-Farley, Sherjil Ozair†, Aaron Courville, Yoshua Bengio†
Département d'informatique et de recherche opérationnelle
Université de Montréal
Montréal, QC H3C 3J7

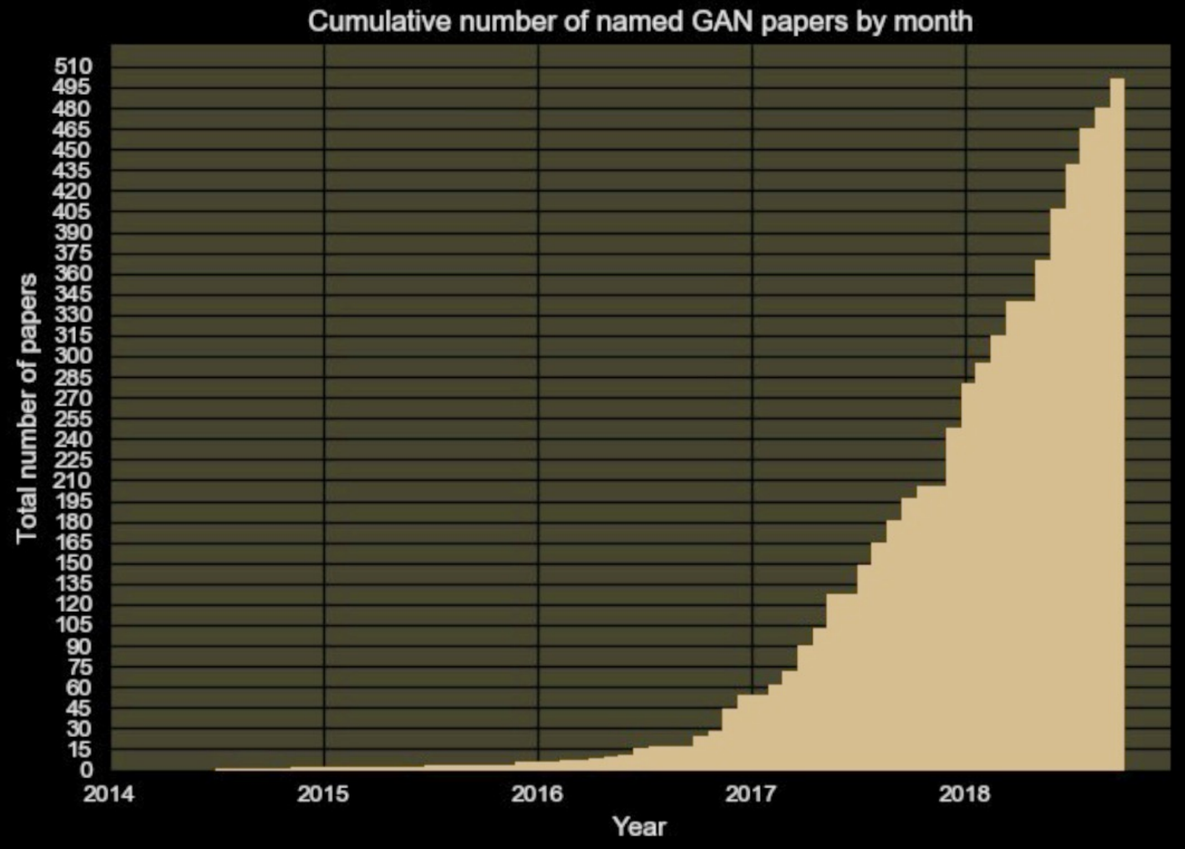
Abstract

We propose a new framework for estimating generative models via an adversarial process, in which we simultaneously train two models: a generative model G that captures the data distribution, and a discriminative model D that estimates the probability that a sample came from the training data rather than G. The training procedure for G is to maximize the probability of D making a mistake. This framework corresponds to a minimax two-player game. In the space of arbitrary functions G and D, a unique solution exists, with G recovering the training data distribution and D equal to $\frac{1}{2}$ everywhere. In the case where G and D are defined by multilayer perceptrons, the entire system can be trained with backpropagation. There is no need for any Markov chains or unrolled approximate inference networks during either training or generation of samples. Experiments demonstrate the potential of the framework through qualitative and quantitative evaluation of the generated samples.

GANs Explosion (2017)

“The GAN Zoo”

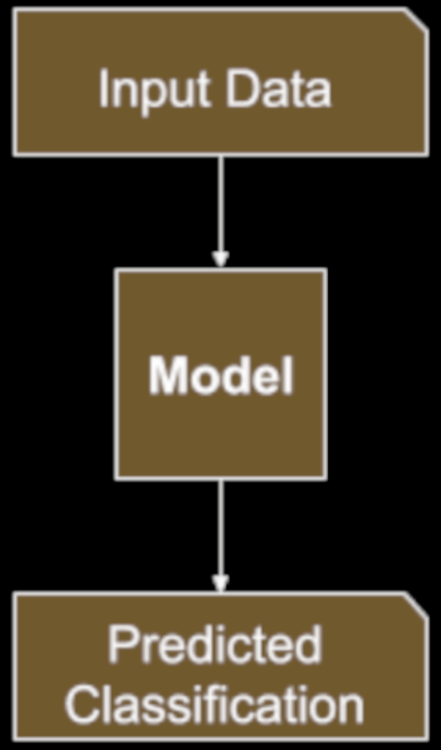
- GAN - Generative Adversarial Networks
- 3D-GAN - Learning a Probabilistic Latent Space of Object Shapes via 3D Generative-Adversarial Modeling
- acGAN - Face Aging With Conditional Generative Adversarial Networks
- AC-GAN - Conditional Image Synthesis With Auxiliary Classifier GANs
- AdaGAN - AdaGAN: Boosting Generative Models
- AEGAN - Learning Inverse Mapping by Autoencoder based Generative Adversarial Nets
- AffGAN - Amortised MAP Inference for Image Super-resolution
- AL-CGAN - Learning to Generate Images of Outdoor Scenes from Attributes and Semantic Layouts
- ALI - Adversarially Learned Inference
- AM-GAN - Generative Adversarial Nets with Labeled Data by Activation Maximization
- AnoGAN - Unsupervised Anomaly Detection with Generative Adversarial Networks to Guide Marker Discovery
- ArtGAN - ArtGAN: Artwork Synthesis with Conditional Categorial GANs
- b-GAN - b-GAN: Unified Framework of Generative Adversarial Networks
- Bayesian GAN - Deep and Hierarchical Implicit Models
- BEGAN - BEGAN: Boundary Equilibrium Generative Adversarial Networks
- BiGAN - Adversarial Feature Learning
- BS-GAN - Boundary-Seeking Generative Adversarial Networks
- CGAN - Conditional Generative Adversarial Nets
- CaloGAN - CaloGAN: Simulating 3D High Energy Particle Showers in Multi-Layer Electromagnetic Calorimeters with Generative Adversarial Networks
- CCGAN - Semi-Supervised Learning with Context-Conditional Generative Adversarial Networks
- CatGAN - Unsupervised and Semi-supervised Learning with Categorical Generative Adversarial Networks
- CoGAN - Coupled Generative Adversarial Networks

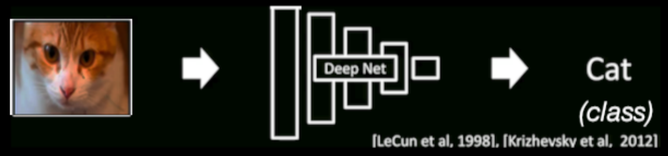


https://github.com/hindupuravinash/the-gan-zoo (*600+ papers*)

Analysis vs. Synthesis

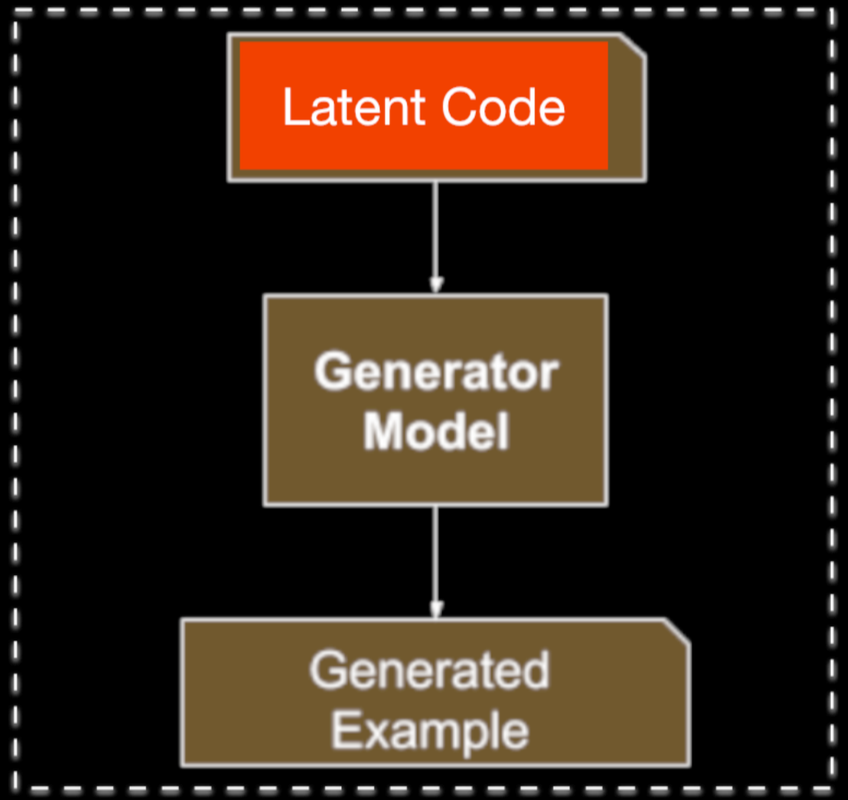
- Discriminative Model

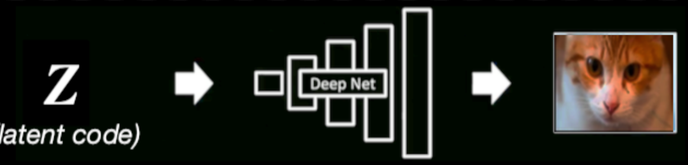




[LeCun et al, 1998], [Krizhevsky et al, 2012]

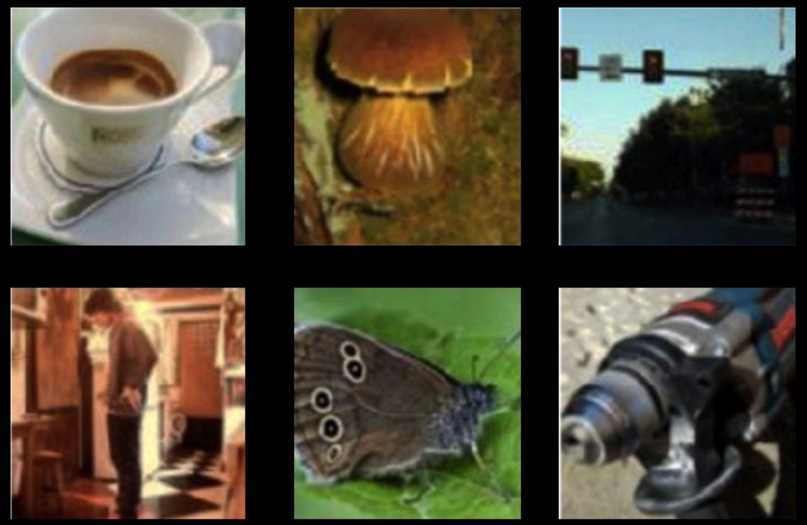
- Generative Model



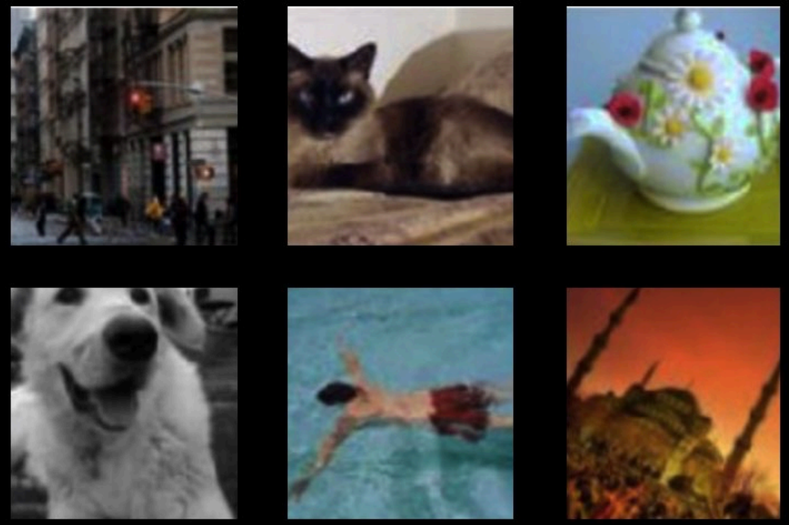


Modeling a Data Distribution

- Density estimation



Training examples

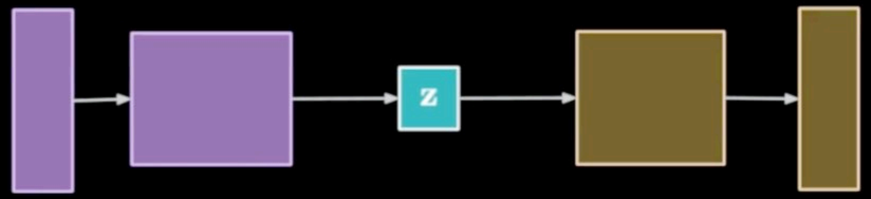


Model samples

(Goodfellow 2016)

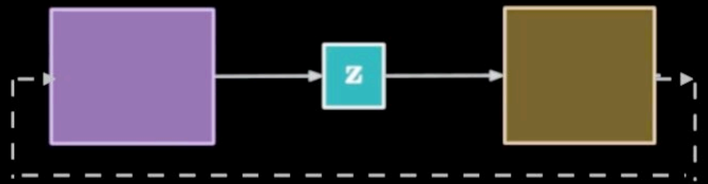
Training and Inference

- Network (Model)



- *Training the Model*

Training Data
(CelebA)

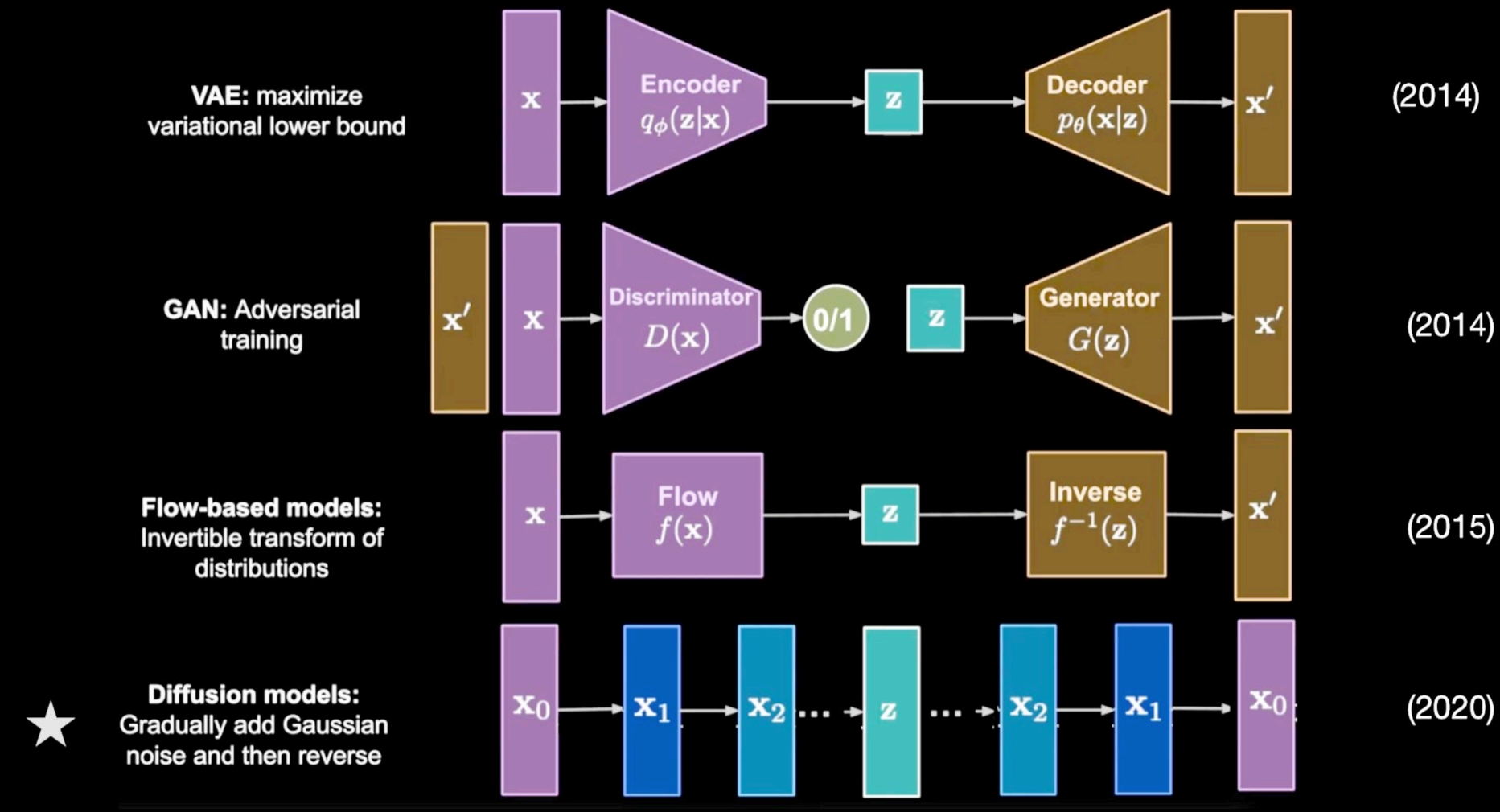


(self-supervised training)

- *Sample Generation*

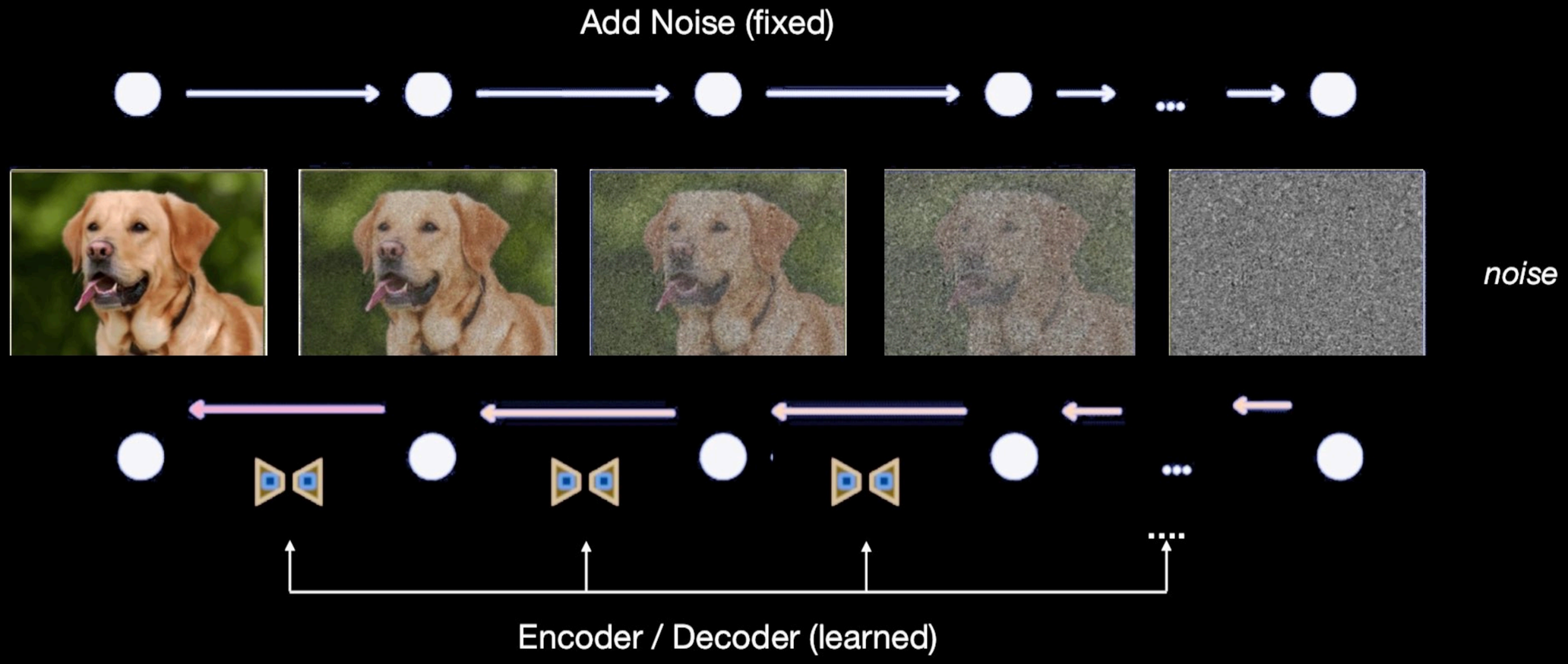
Sample Generator
(Karras et al, 2017)

Types of Generative Models



Diffusion Models

- Sequential Application of Denoising Auto-Encoders



Text & Images

Clip - unClip

- CLIP : *image to text*

"a photo of guacamole, a type of food"

(Image Caption)

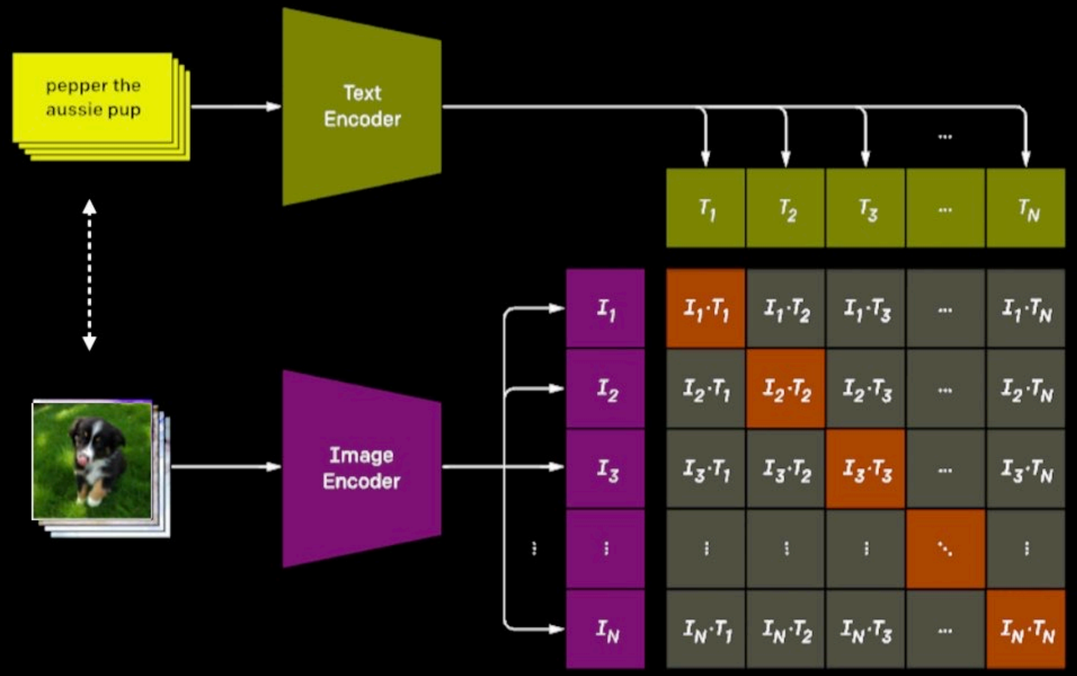
- unCLIP : *text* to *image*

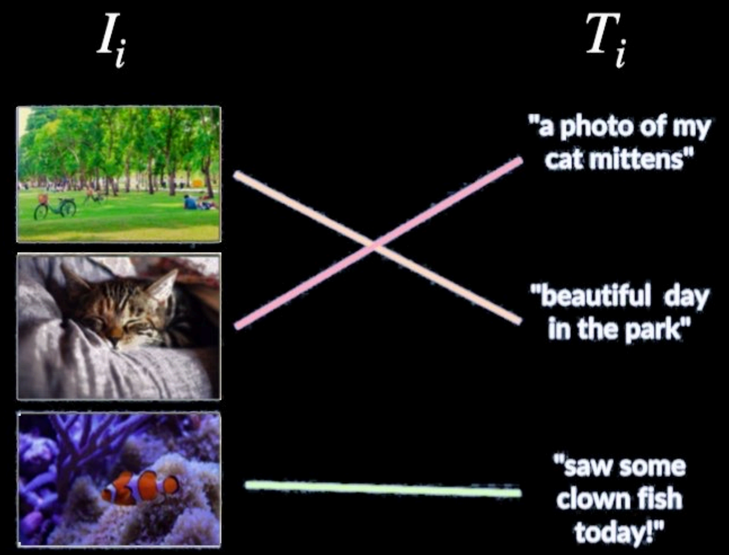
"An astronaut lounging in a tropical resort in space in a photorealistic style"

(Text-Image Generation) ★

Contrastive Language–Image Pre-training

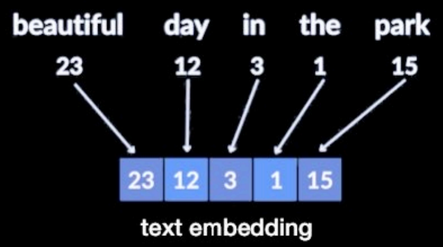
- Learn Text-Picture Correlations

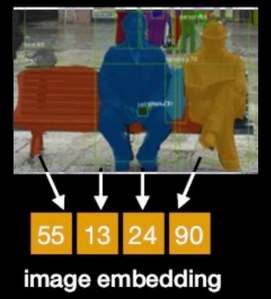




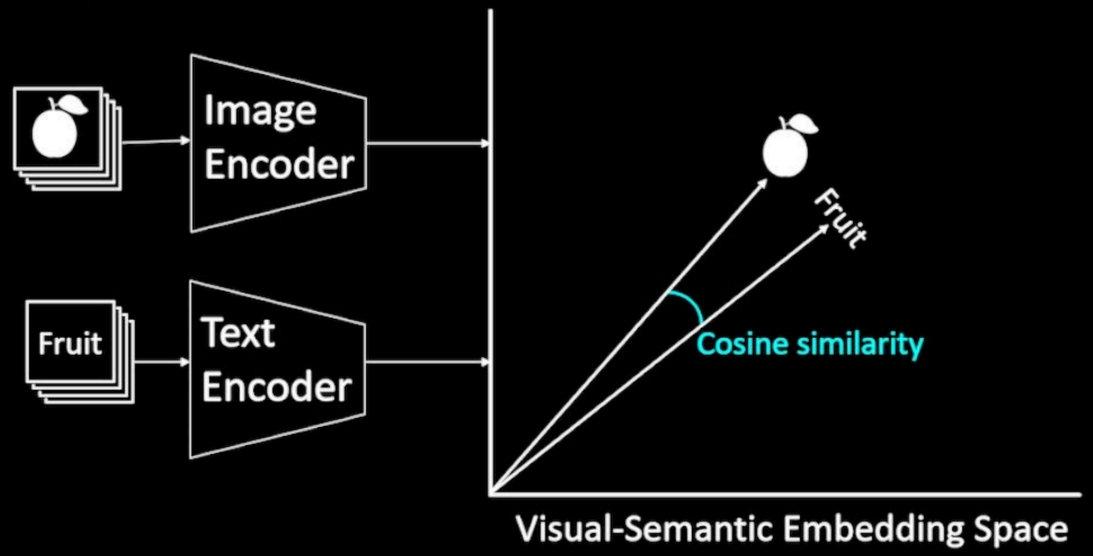
Visual Concept Prediction

- Embeddings

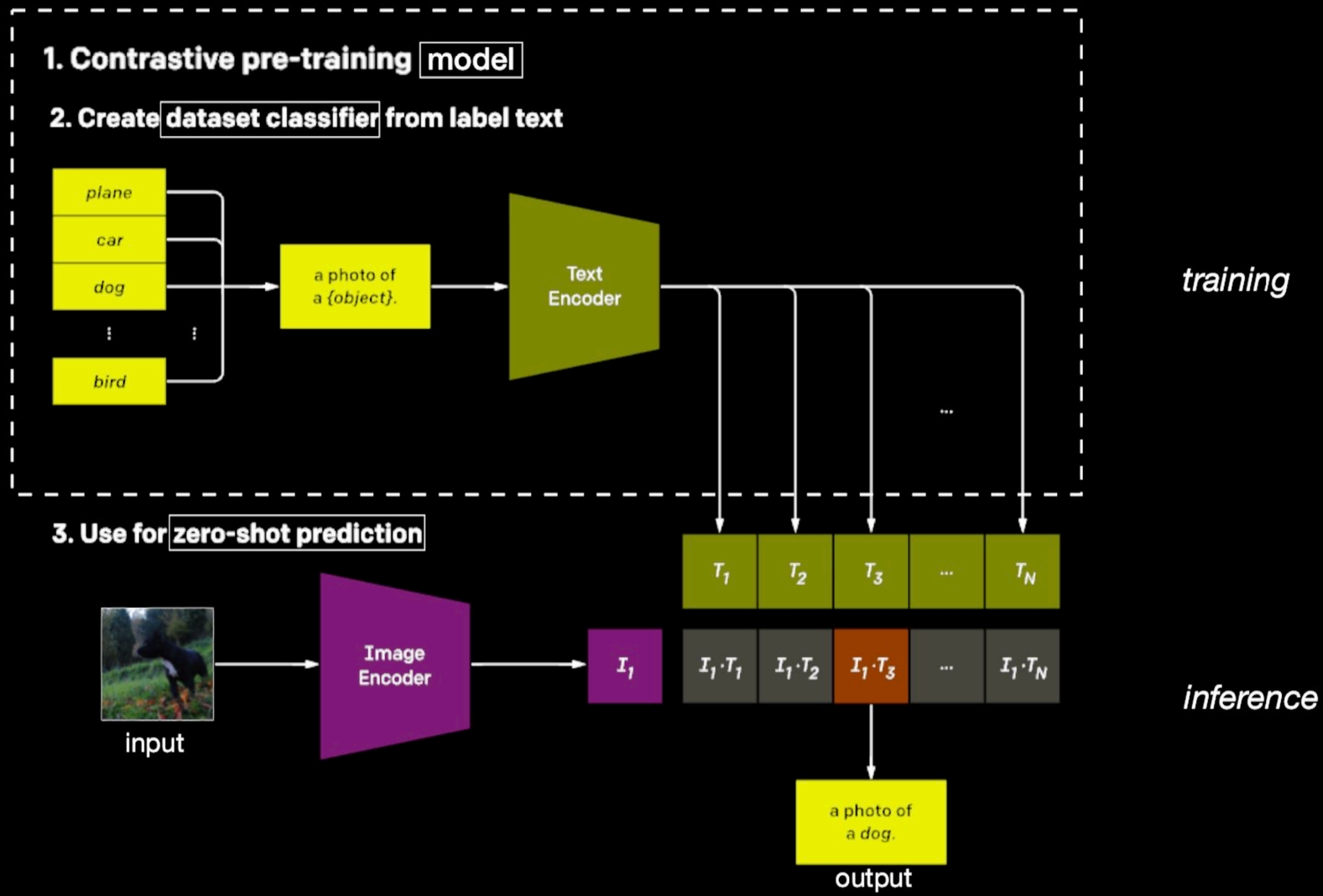




- Visual-Semantic Similarity



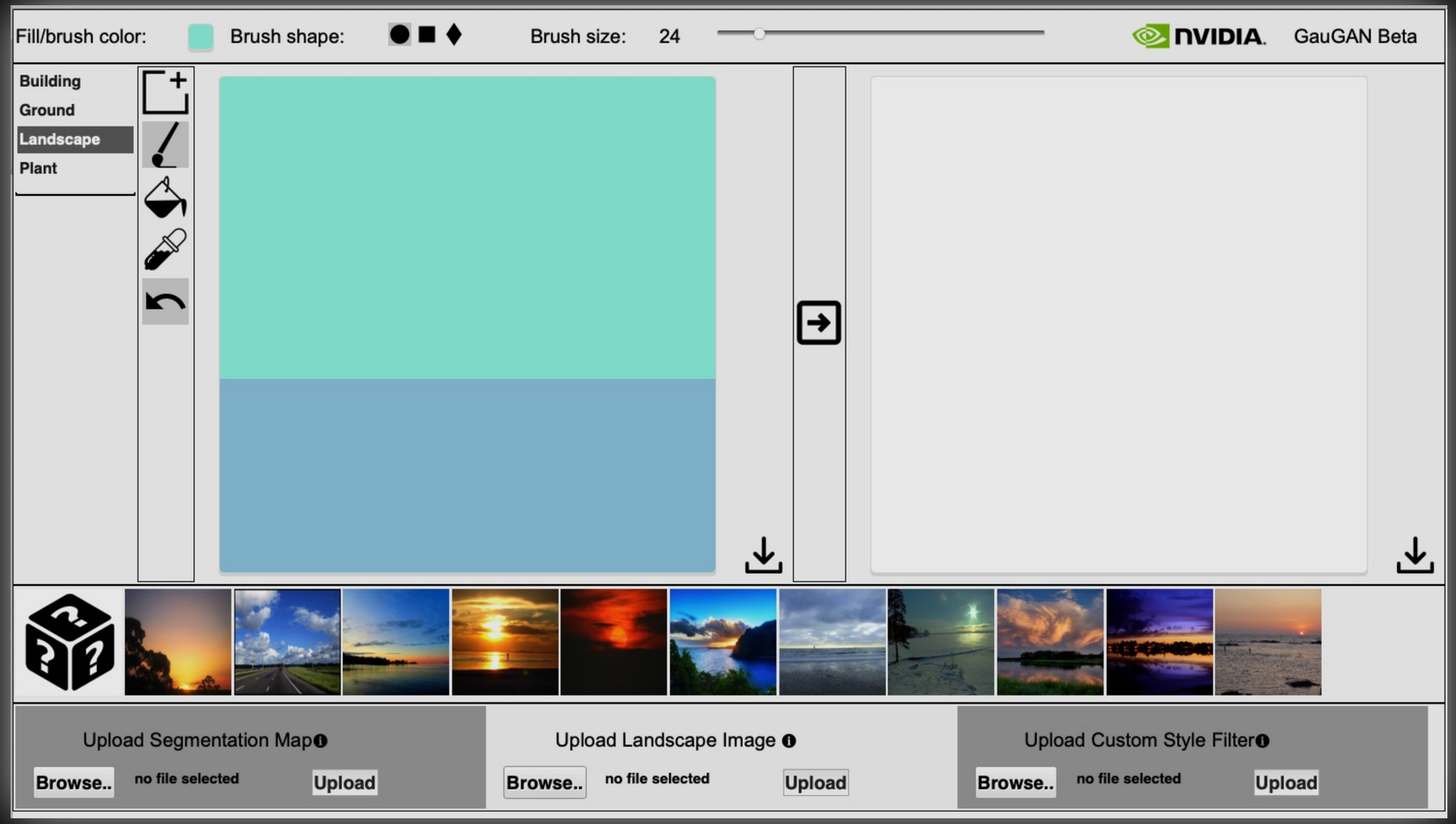
Clip

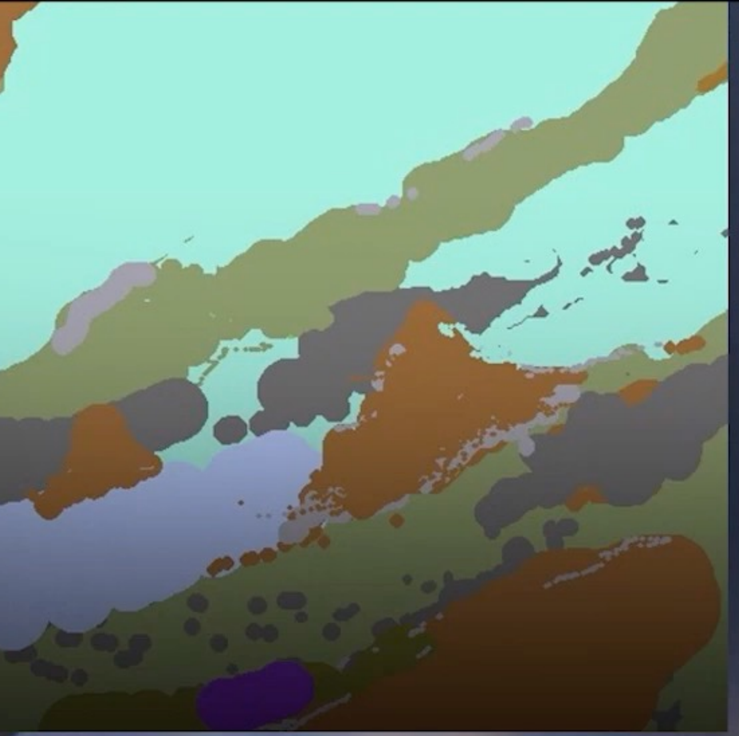

1. Contrastive pre-training model
2. Create dataset classifier from label text
plane
car
dog
bird
a photo of a {object}.
Text Encoder
...
training
3. Use for zero-shot prediction
T_1 T_2 T_3 ... T_N
input
Image Encoder
I_1
$I_1 \cdot T_1$ $I_1 \cdot T_2$ $I_1 \cdot T_3$ ... $I_1 \cdot T_N$
inference
a photo of a dog.
output

Creative Tools

GauGAN

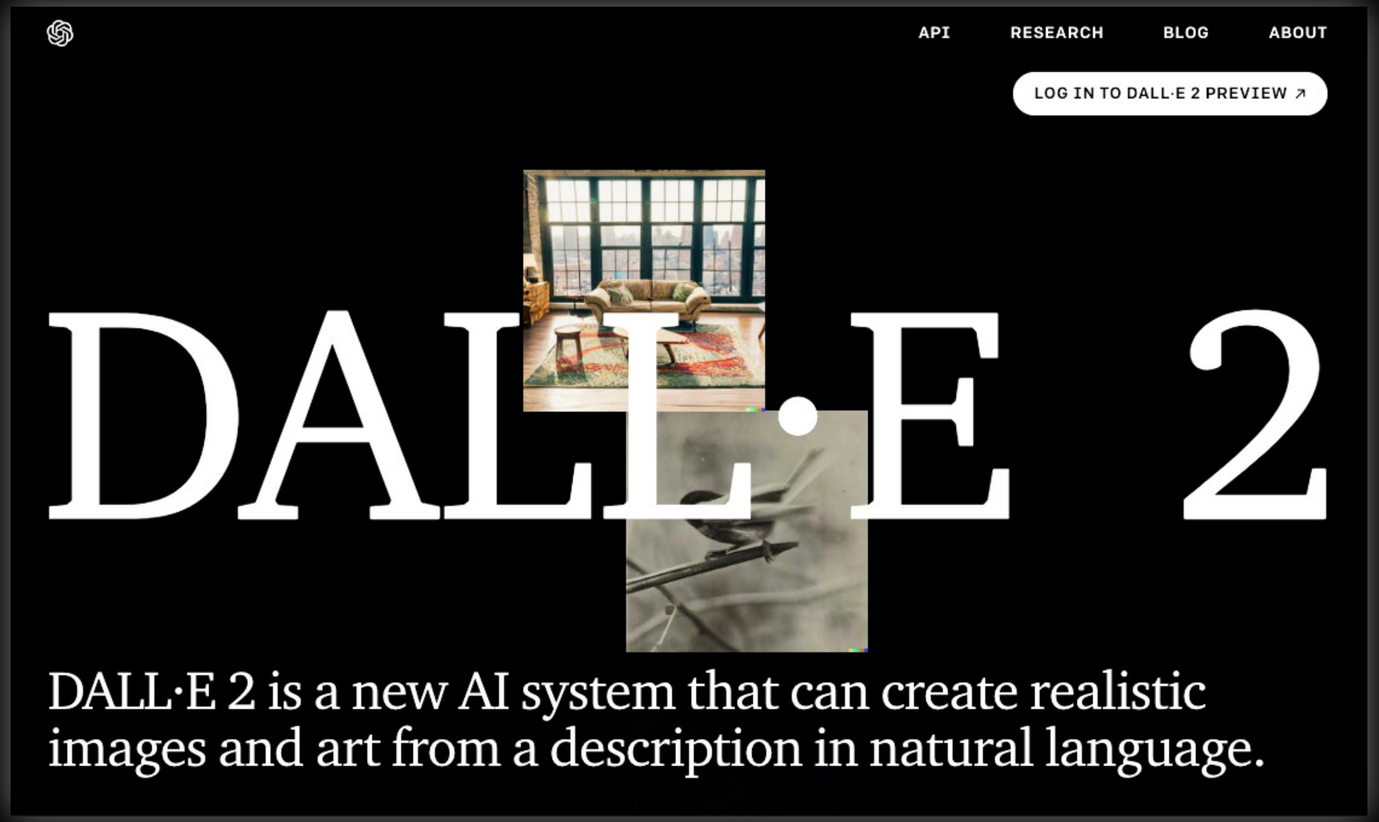
- Visual Semantics

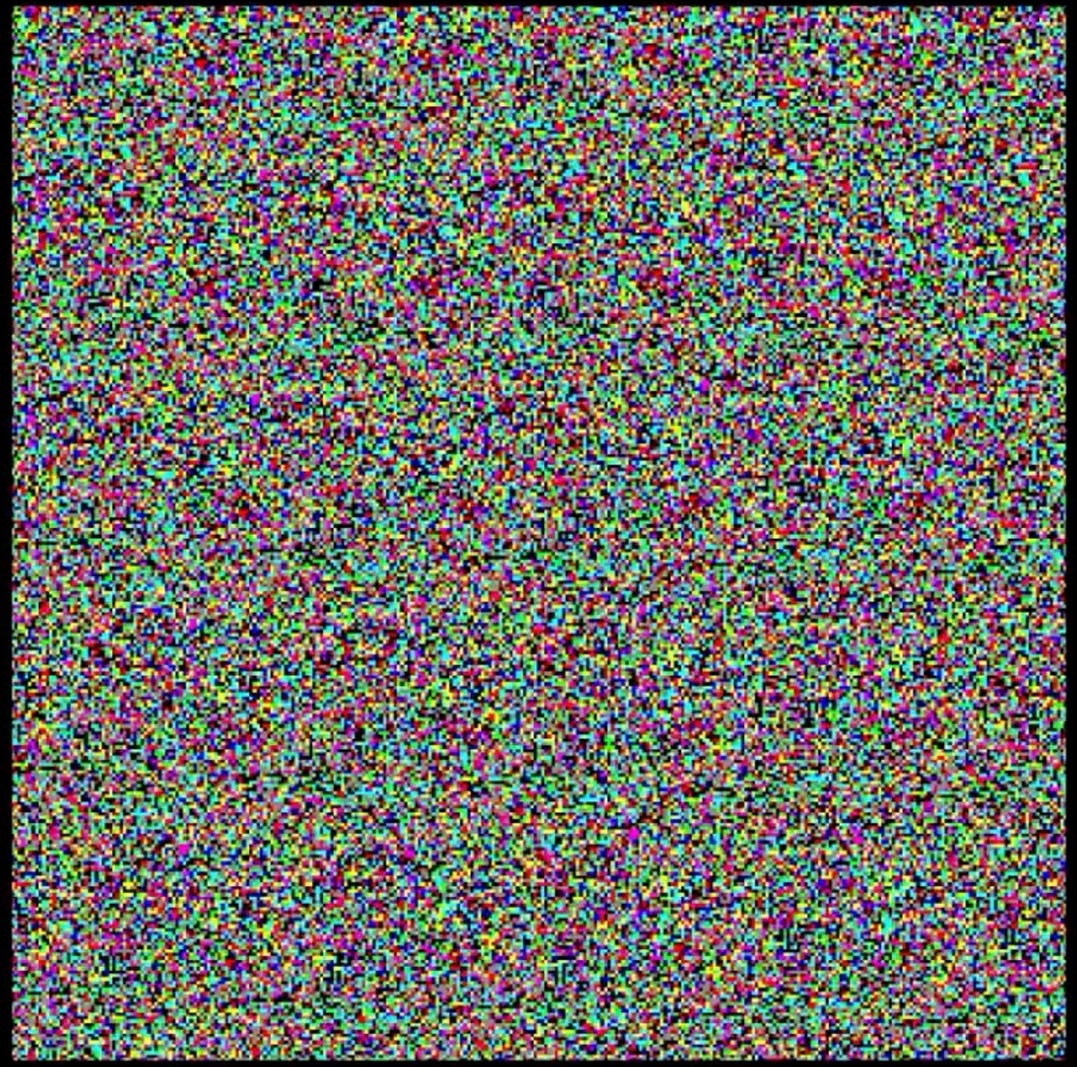




DALL-E 2

- Lexical-Visual Semantics

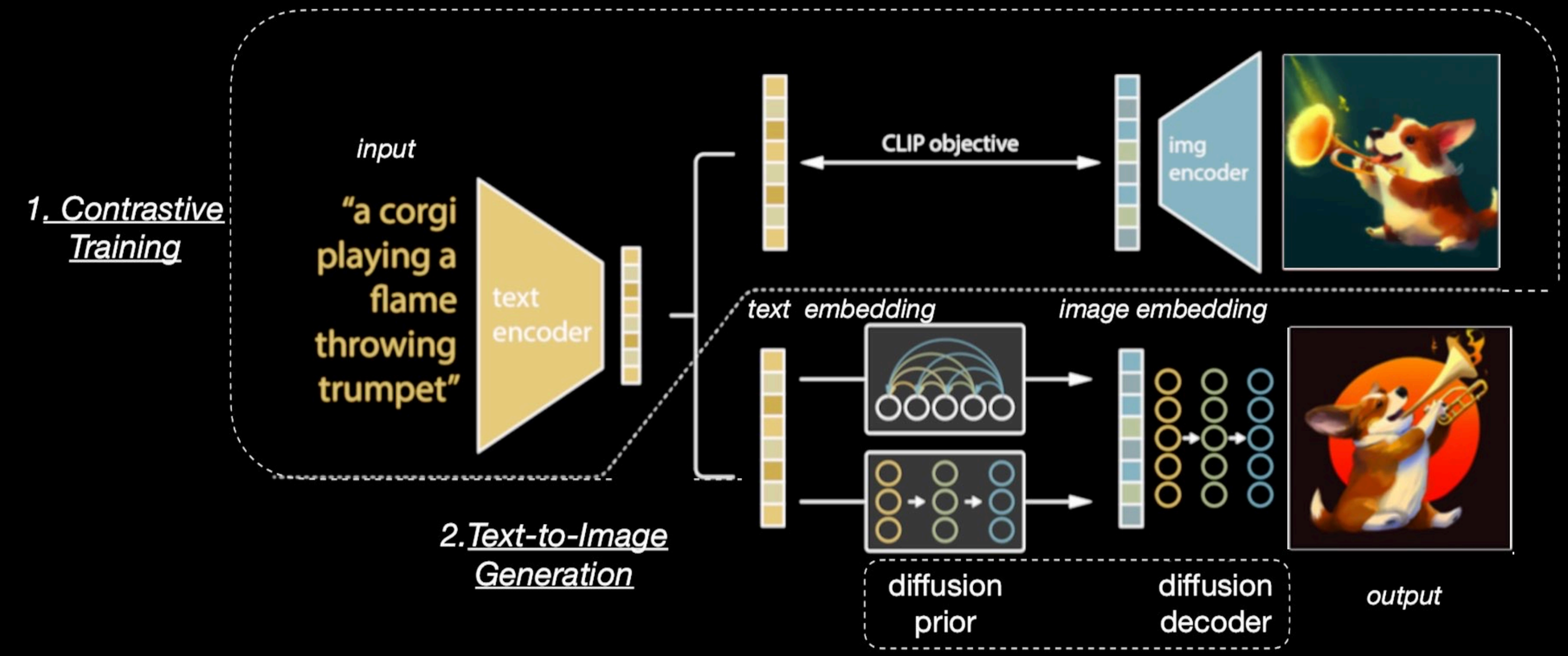




A Look Deep Inside:
DALL-E, Stable Diffusion, et. al.

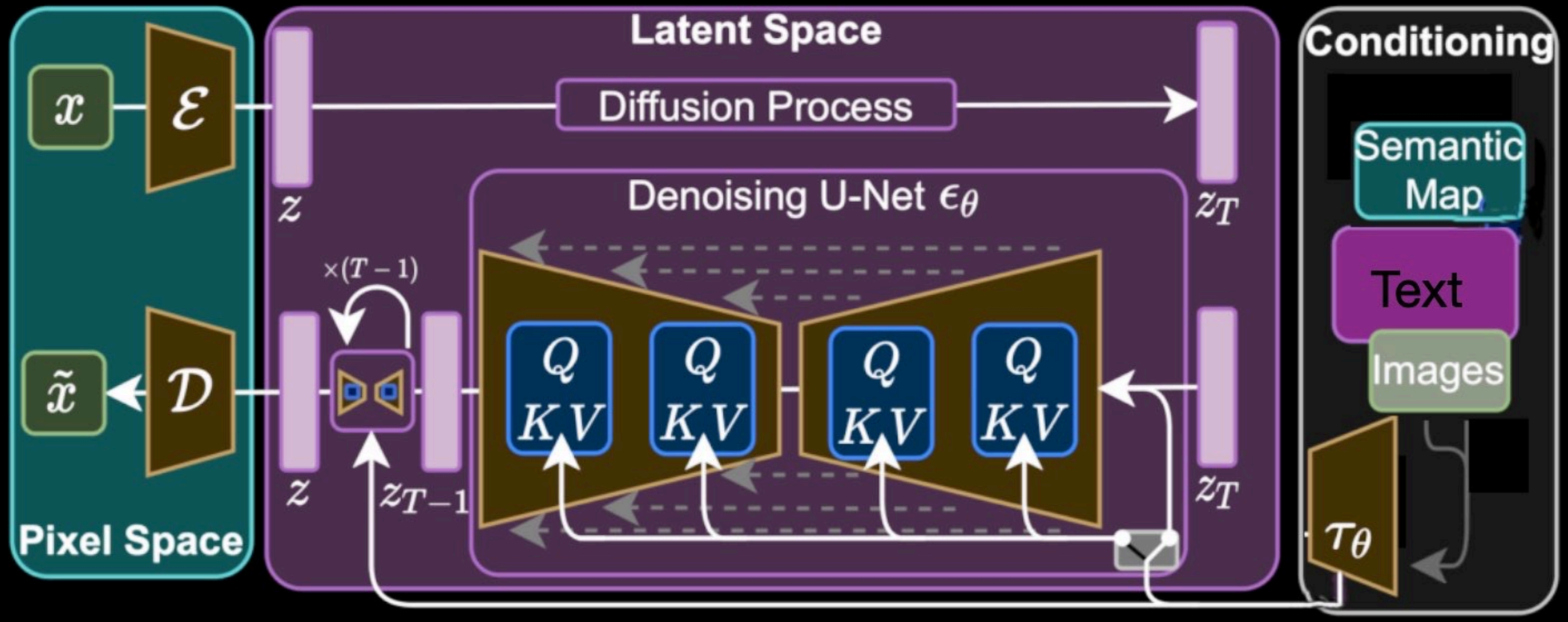
DALL-E 2 (unClip)

- OpenAI (April 2022)
  - Clip + Diffusion



Stable Diffusion

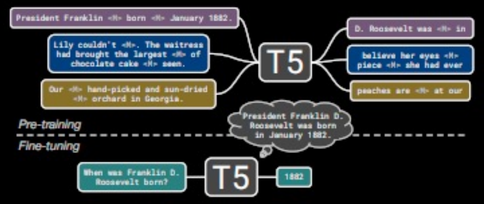
- LMU Munich (April 2022)
  - Latent Diffusion Model



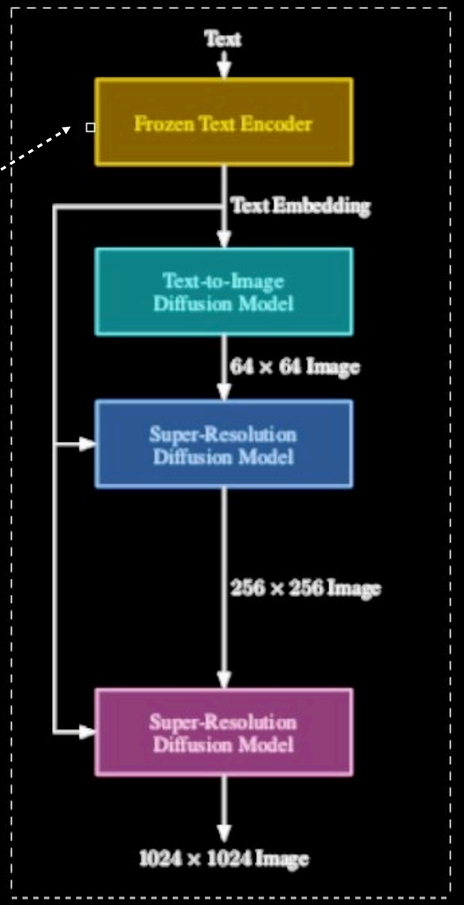
Imagen

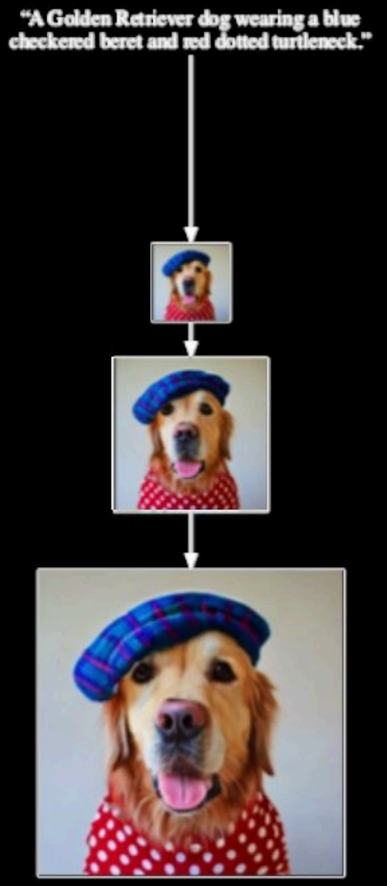
- Google Brain (June, 2022)

- Text Encoder (Transformer,T5-XXL)
  - maps text to sequence of embeddings



- Cascade of Conditional Diffusion Models
  - map embeddings to images of increasing resolutions





“That’s All Folks!"

"Stay Tuned for the Next Episodes.”

–L.V.