

Deep Learning Applications

Tiago Vieira

Institute of Computing
Universidade Federal de Alagoas

Outline

Introduction

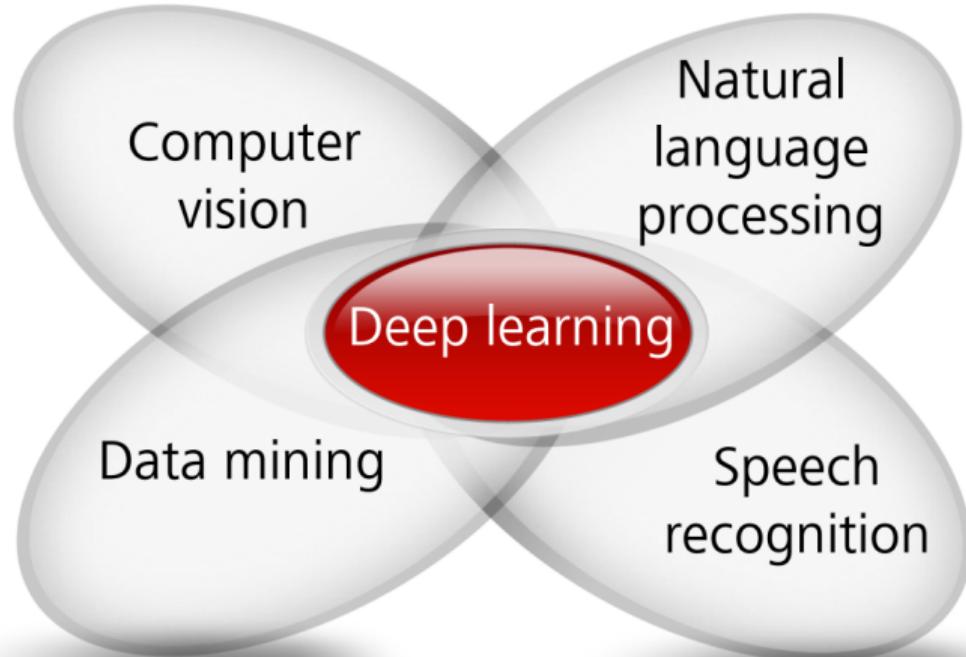
Computer Vision

Natural Language Processing (NLP)

Generative Adversarial Networks

Text-to-Image

Introduction



Computer Vision

- ▶ Image classification.
- ▶ Scene recognition.
- ▶ Face detection and recognition.
- ▶ Object detection and recognition.
- ▶ Semantic segmentation.
- ▶ Style transfer.
- ▶ Super-resolution.
- ▶ Optical Character Recognition (OCR).

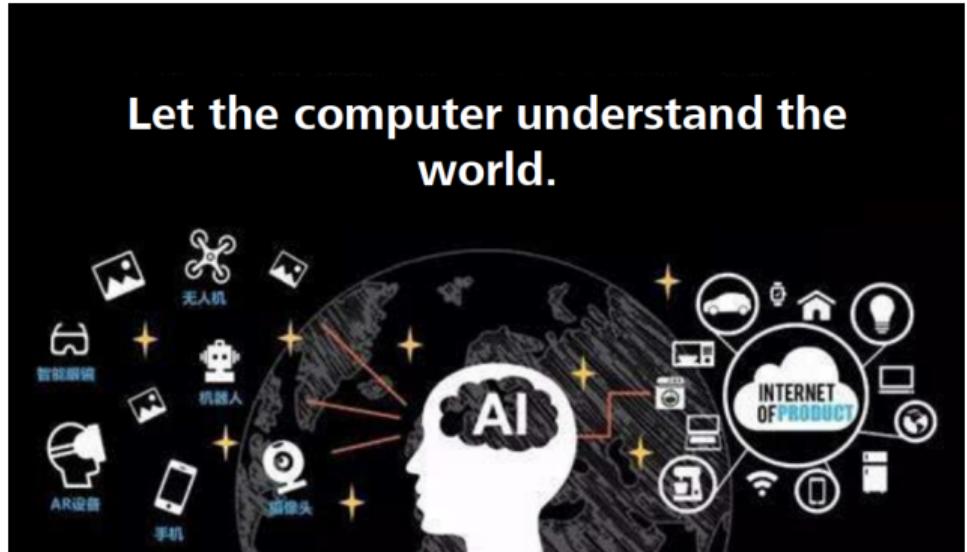


Image Classification

Maps images to different class sets, which can be used for:

- ▶ image retrieval, and;
- ▶ image archiving.

			
mite	container ship	motor scooter	leopard
mite	container ship	motor scooter	leopard
black widow	lifeboat	go-kart	jaguar
cockroach	amphibian	moped	cheetah
tick	fireboat	bumper car	snow leopard
starfish	drilling platform	golfcart	Egyptian cat

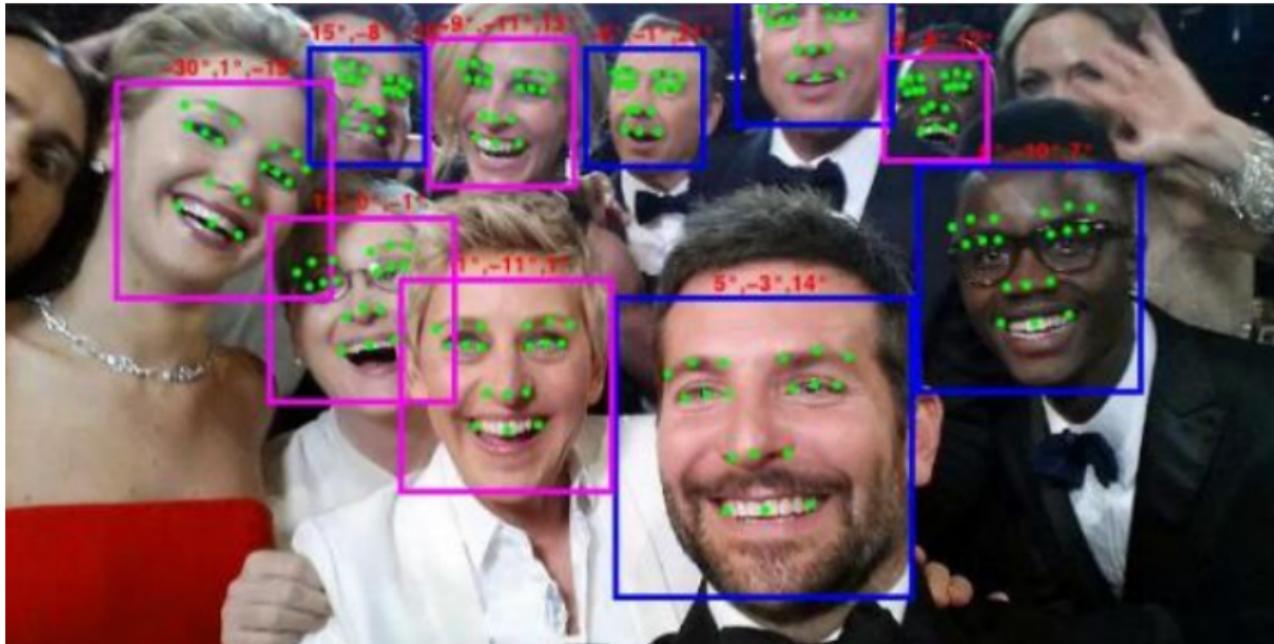
Scene Recognition

Classifies images based on scenes and environments, which can be used for situation awareness and intelligent AI for photographing.



Face Detection

Discovers and locates faces and facial features in images, which can be used for face focusing, polishing up, and augmented reality.



Face recognition

Differentiates people, which can be used for identity authentication and privacy protection.



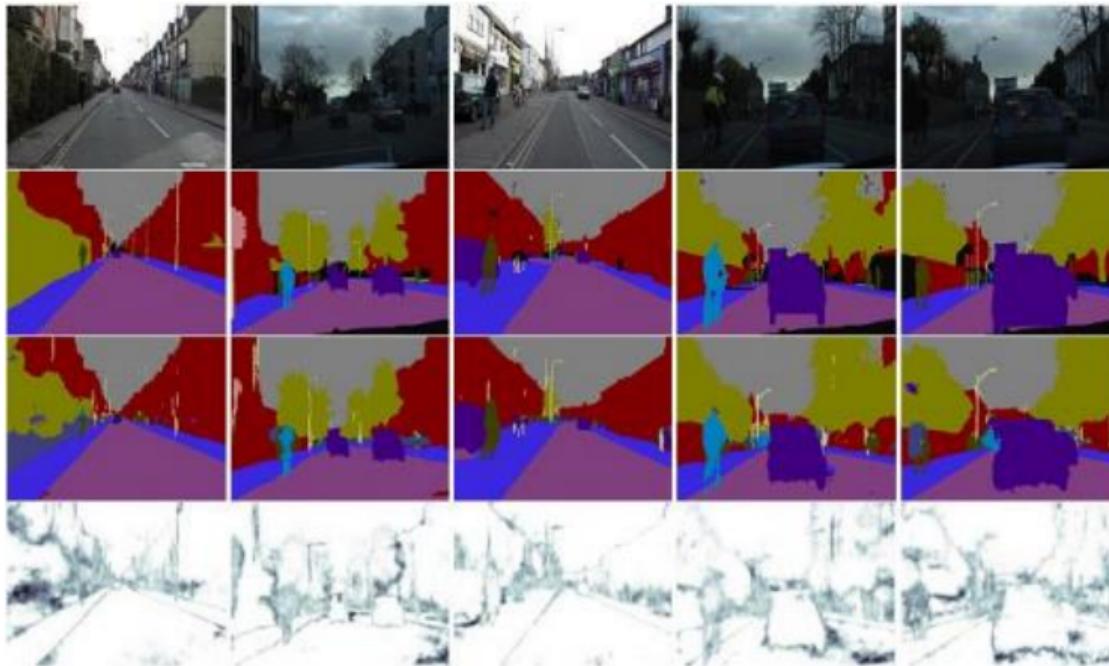
Object Detection and Recognition

Detects, locates, and identifies different objects in images, including digital, text, and pedestrian detection. This function can be used for OCR, unmanned driving, and intelligent image composition.



Semantic Segmentation

Predicts the label of each pixel in the image, including segmentation and recognition.
It can be used for unmanned driving, augmented reality, and situation awareness.

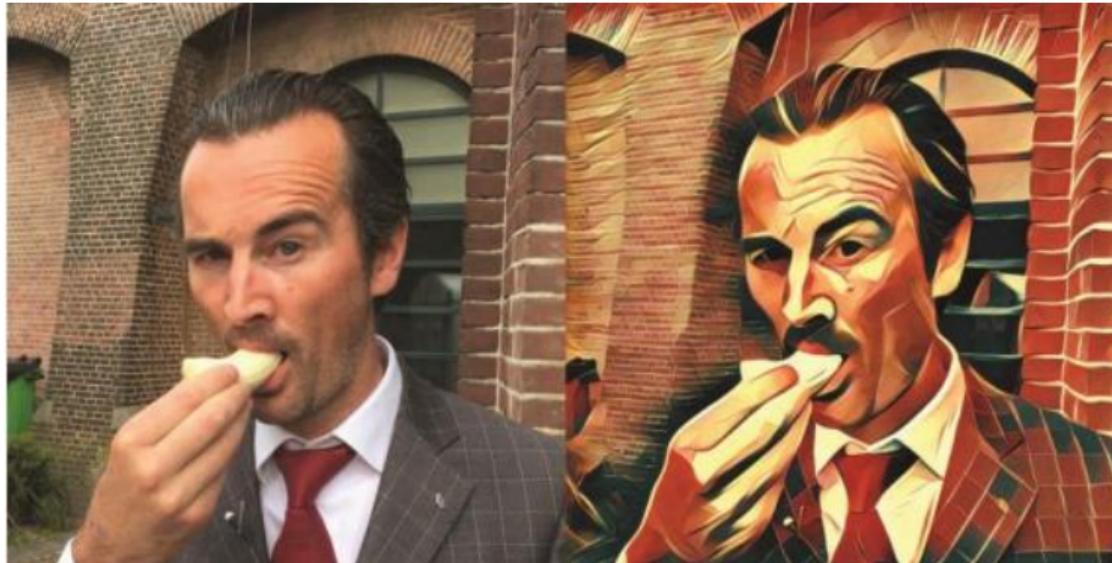


Google's AR core

- ▶ [https://www.youtube.com/c/googlevr.](https://www.youtube.com/c/googlevr)

Style Transfer (Stylization)

Changes the image style while retaining the content of the image, which can be used for stylized image processing.



Super-resolution

Generates high-resolution images from low-resolution images, which can be used for image processing, security surveillance, and medical imaging.



Ground Truth	Bicubic	Ours (ℓ_{pixel})	SRCNN [1]	Ours (ℓ_{feat})
This image	31.78 / 0.8577	31.47 / 0.8573	32.99 / 0.8784	29.24 / 0.7841
Set5 mean	28.43 / 0.8114	28.40 / 0.8205	30.48 / 0.8628	27.09 / 0.7680



Ground Truth	Bicubic	Ours (ℓ_{pixel})	SRCNN [1]	Ours (ℓ_{feat})
This Image	21.69 / 0.5840	21.66 / 0.5881	22.53 / 0.6524	21.04 / 0.6116
Set14 mean	25.99 / 0.7301	25.75 / 0.6994	27.49 / 0.7503	24.99 / 0.6731
BSD100 mean	25.96 / 0.682	25.91 / 0.6680	26.90 / 0.7101	24.95 / 0.6317

OCR

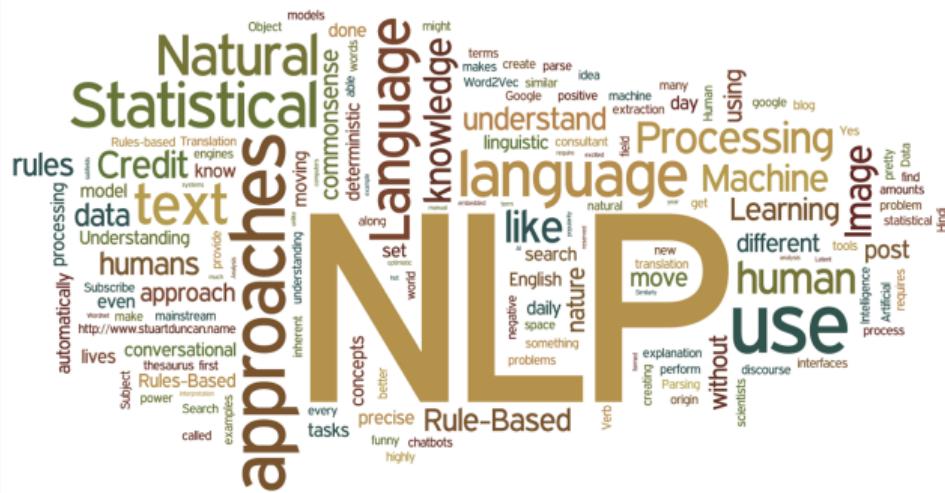
Identifies information such as numbers and texts in images to digitize images or paper documents, which can be used to automatically identify business cards or scan invoices.



Natural Language Processing (NLP)

Natural Language Processing

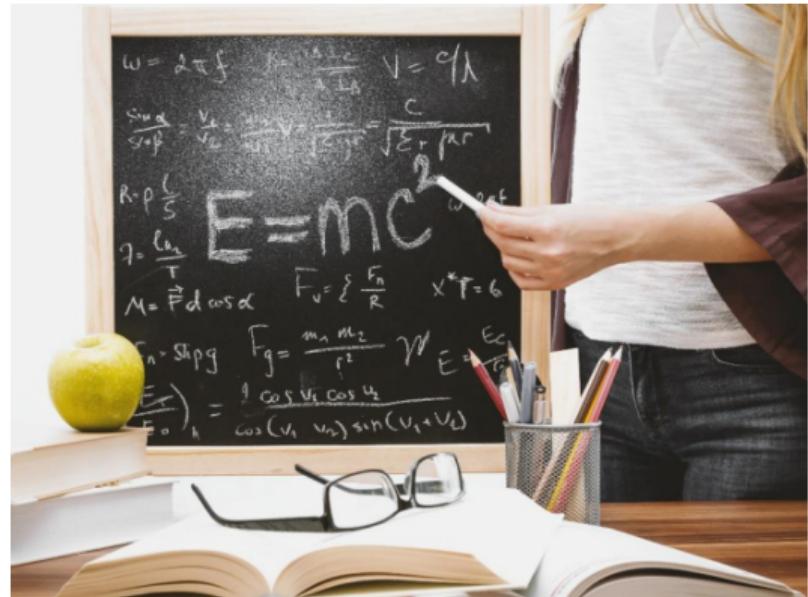
- ▶ Turing's test considered Natural Language.
 - ▶ Word segmentation.
 - ▶ Knowledge exploration.
 - ▶ Machine translation.
 - ▶ Sentiment analysis.



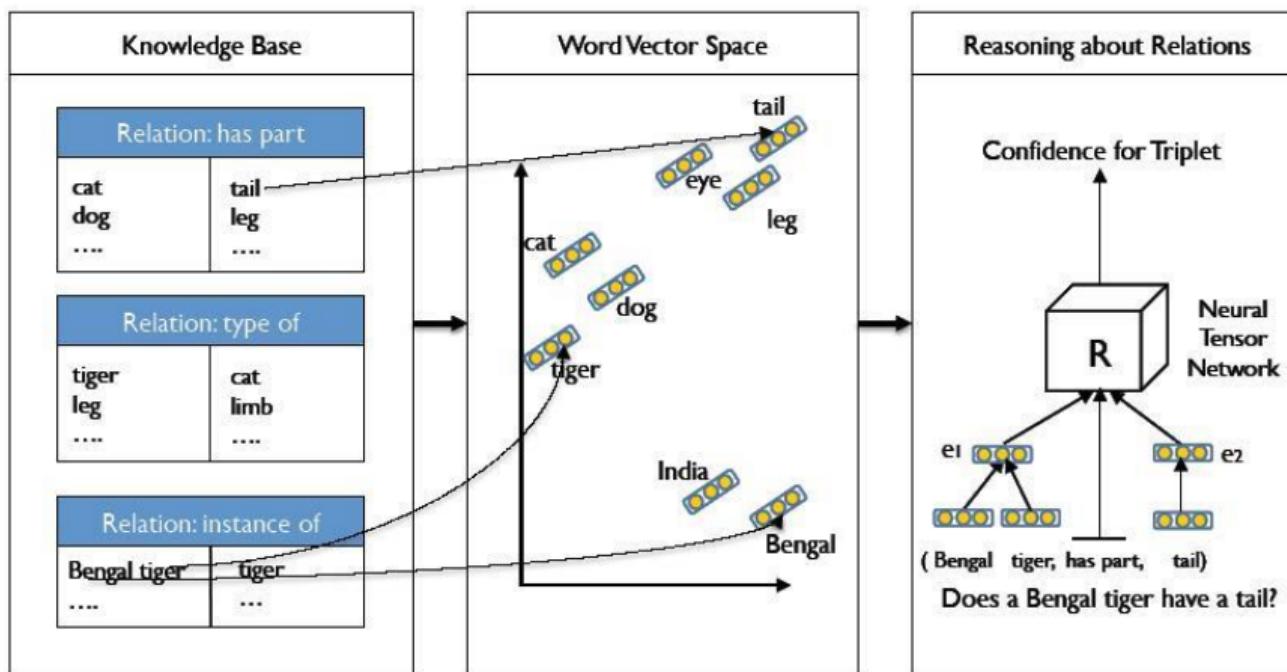
Knowledge Exploration

Two types of problems:

- ▶ New knowledge reasoning in the existing knowledge base
 - CYC, WordNet, FreeNet, etc.
 - Current studies use similar approaches.
 - Known entities are represented by word embeddings.
 - The entity relationship is modeled using tensor networks.
 - Backpropagation + SGD training.
- ▶ Mining structured knowledge from free text
 - Text mining
 - TF-IDF



New Knowledge Reasoning in the Existing Knowledge Base



Mining Structured Knowledge from Free Text (1)

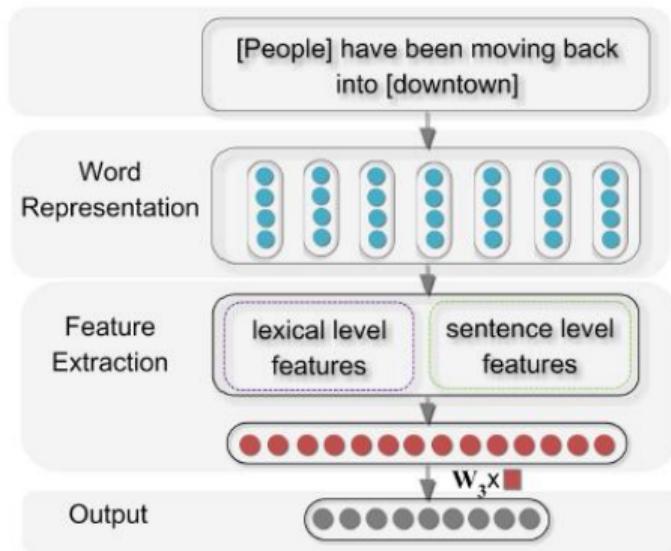


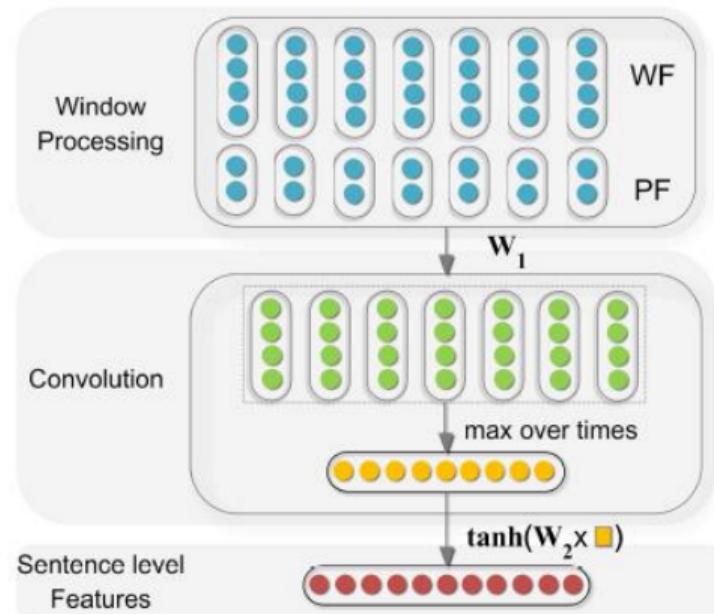
Figure: Overall structure

Features	Remark
L1	Noun 1
L2	Noun 2
L3	Left and right tokens of noun 1
L4	Left and right tokens of noun 2
L5	WordNet hypernyms of nouns

Figure: Lexical feature

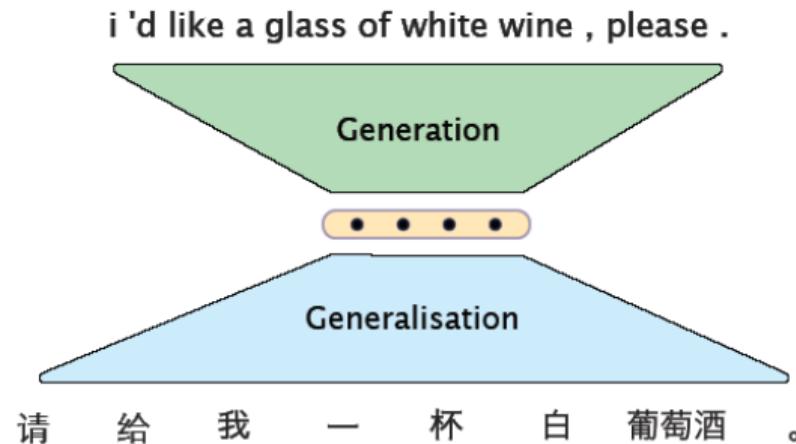
Mining Structured Knowledge from Free Text (2)

Sentence-level feature extraction: convolutional network.



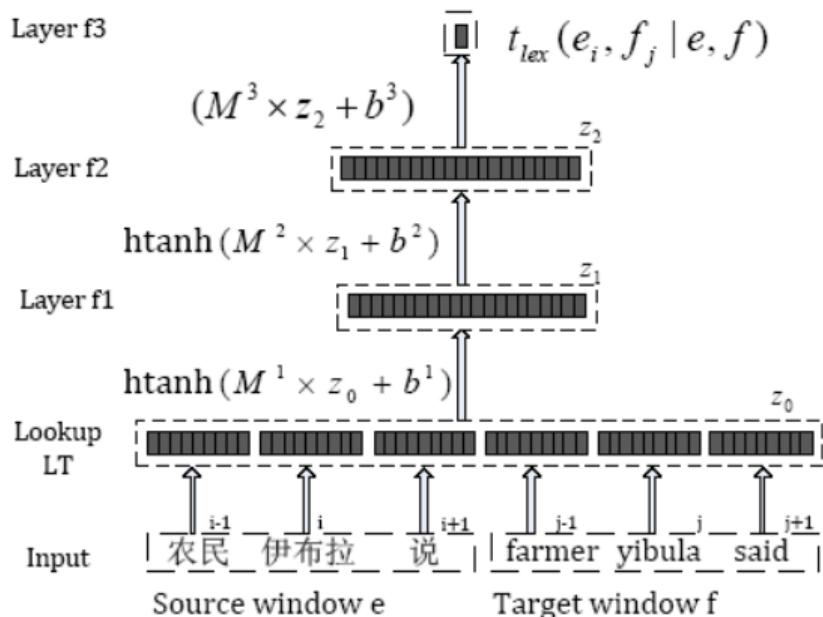
Machine Translation (Common Model)

- ▶ Decoder.
- ▶ Semantic vector.
- ▶ Encoder.



Machine Translation: Deep Learning Applied to Multiple Areas

- ▶ Word alignment
- ▶ Phrase alignment
- ▶ Phrase reordering
- ▶ Language model
- ▶ Translation model
- ▶ Conjunctive model
- ▶ Translation result reordering



Sentiment Analysis

Two core issues:

1. Sentence-level word embedding representation
2. How to encode emotional tendencies to word embedding at all levels
 - Semi-supervised or supervised learning: The emotion tendency is encoded into the WE structure through the training process.

What is Sentiment Analysis?

A linguistic analysis technique that identifies opinion early in a piece of text.

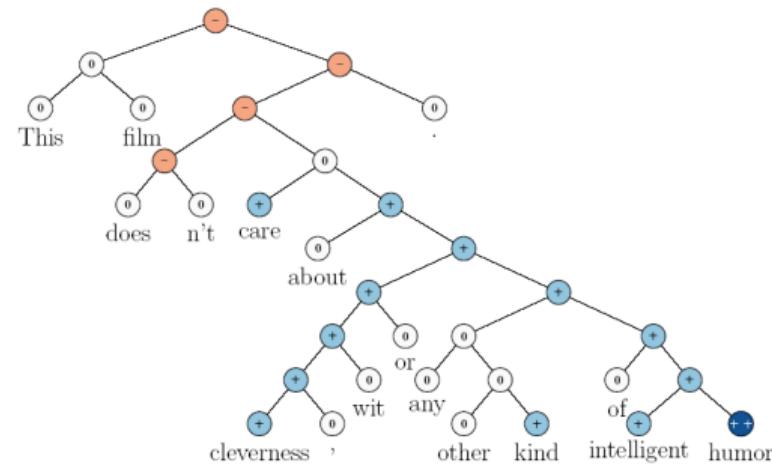
The movie is great.



The movie stars Mr. X



The movie is horrible.



Speech Recognition

Speech can have many roles in mobile apps



Recommendation System

IMDb Find Movies, TV shows, Celebrities and more... All

Movies TV News Showtimes Community IMDbPro



Spider-Man (2002)  Top 500

(PG-13) 121 min - Action | Fantasy - 3 May 2002 (USA)

Your rating:  7.3 /10 Ratings: 7.3/10 from 322,552 users Metascore: 73/100 Reviews: 1,976 user | 276 critic | 37 from Metacritic.com

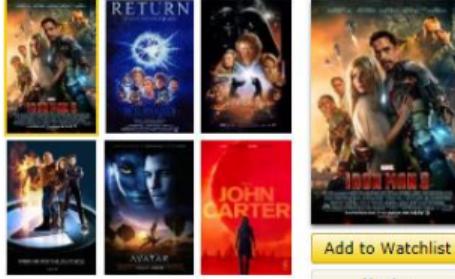
When bitten by a genetically modified spider, a nerdy, shy, and awkward high school student gains spider-like abilities that he eventually must use to fight evil as a superhero after tragedy befalls his family.

Director: Sam Raimi
Writers: Stan Lee (Marvel comic book), Steve Ditko (Marvel comic book), 1 more credit »
Stars: Tobey Maguire, Willem Dafoe, Kirsten Dunst | See full cast and crew

+ Watchlist Watch Trailer Share...

Learn more

People who liked this also liked...

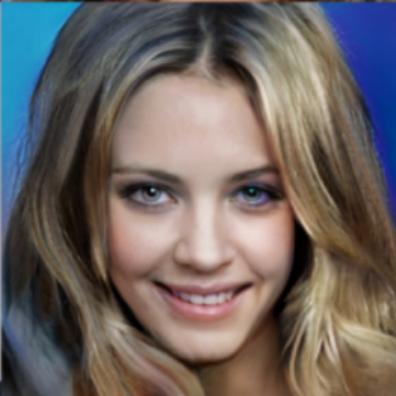
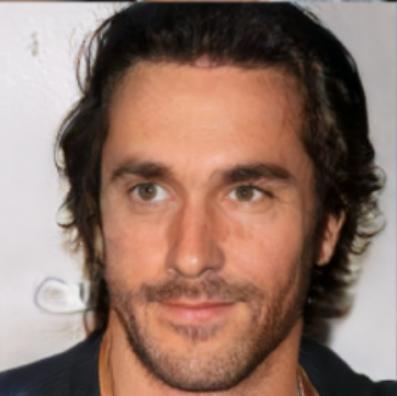
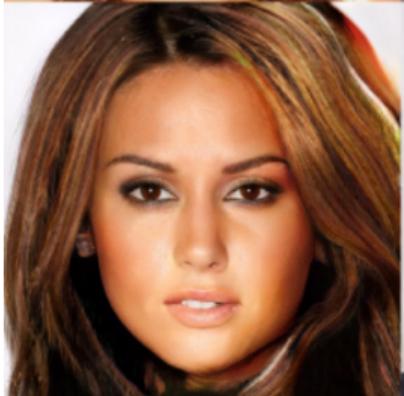


Add to Watchlist Next > ◀ Prev 6 ▶

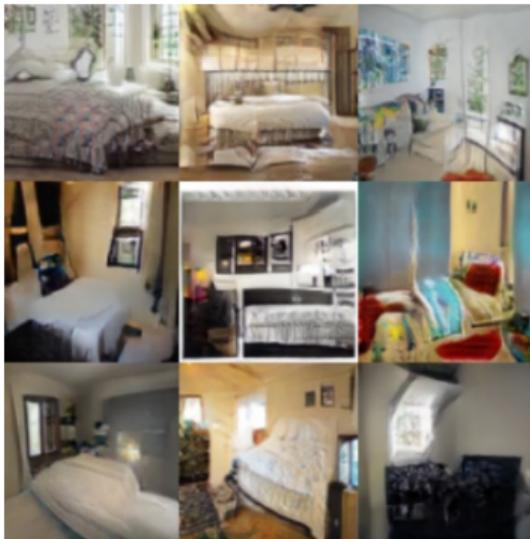
Iron Man 3 (2013)
(PG-13) Action | Adventure | Sci-Fi  7.7/10

When Tony Stark's world is torn apart by a formidable terrorist called the Mandarin, he starts an odyssey of rebuilding and retribution.

Director: Shane Black
Stars: Robert Downey Jr., Gwyneth...











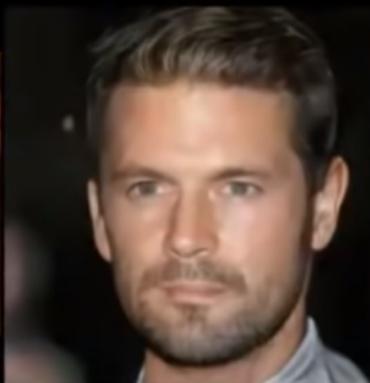
2014



2015



2016



2017



2018



Odena et al
2016



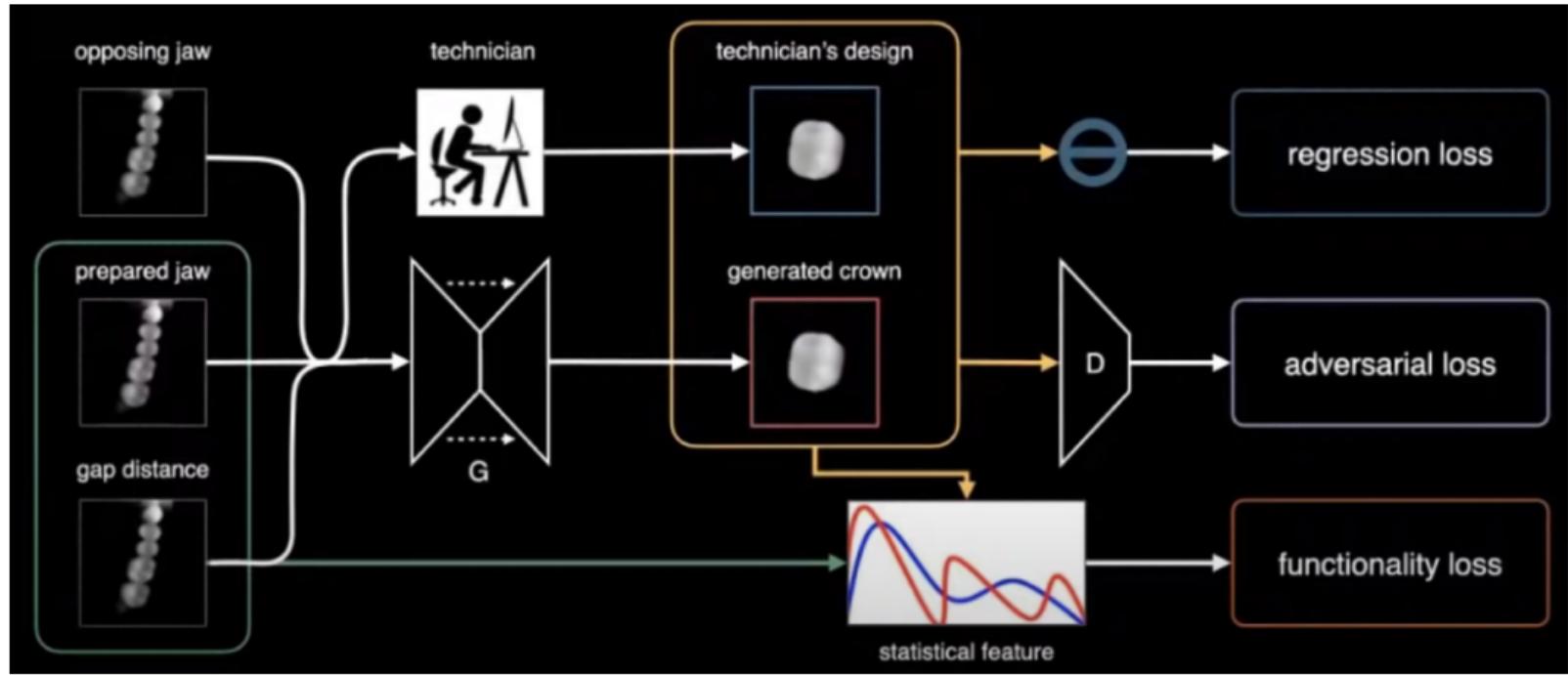
Miyato et al
2017

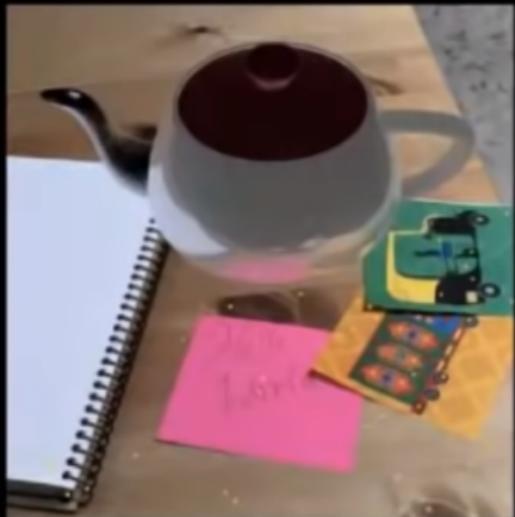


Zhang et al
2018

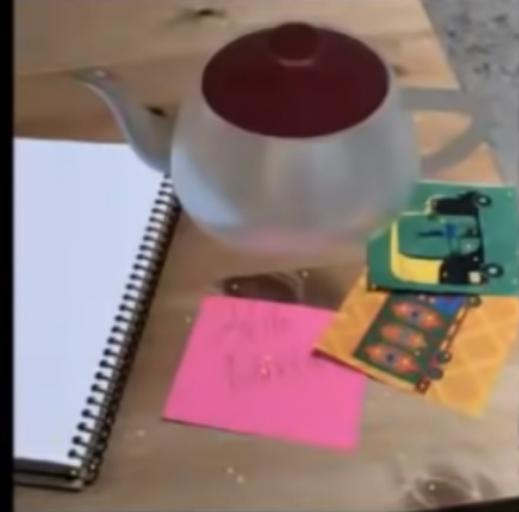
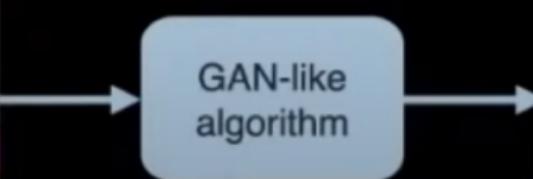


Brock et al
2018



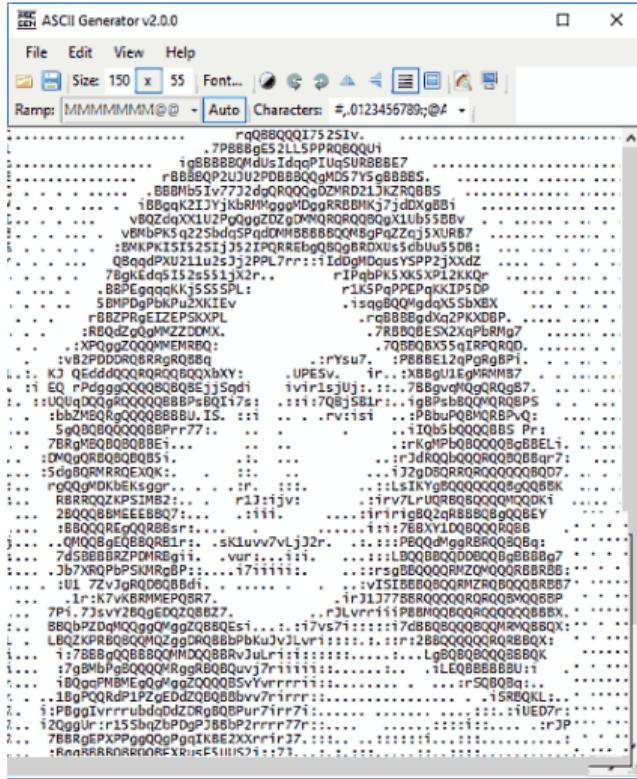


Incomplete Environment Map



Output Environment Map

Text-to-Image: Before



Text-to-Image: Now

- ▶ Input a natural language description.
- ▶ Output a high-resolution, realistic (and/or artistic) image.

OpenAI's DALL-E

- ▶ Created by commercial AI lab OpenAI.
- ▶ Create realistic images and art from a description in natural language.

Medieval painting of the wifi not working.

Medieval painting of the wifi not working.



References

- ▶ Twitter discussion about Dall-E limitations.
- ▶ The Verge article about *Imagen*: ALL THESE IMAGES WERE GENERATED BY GOOGLE'S LATEST TEXT-TO-IMAGE AI

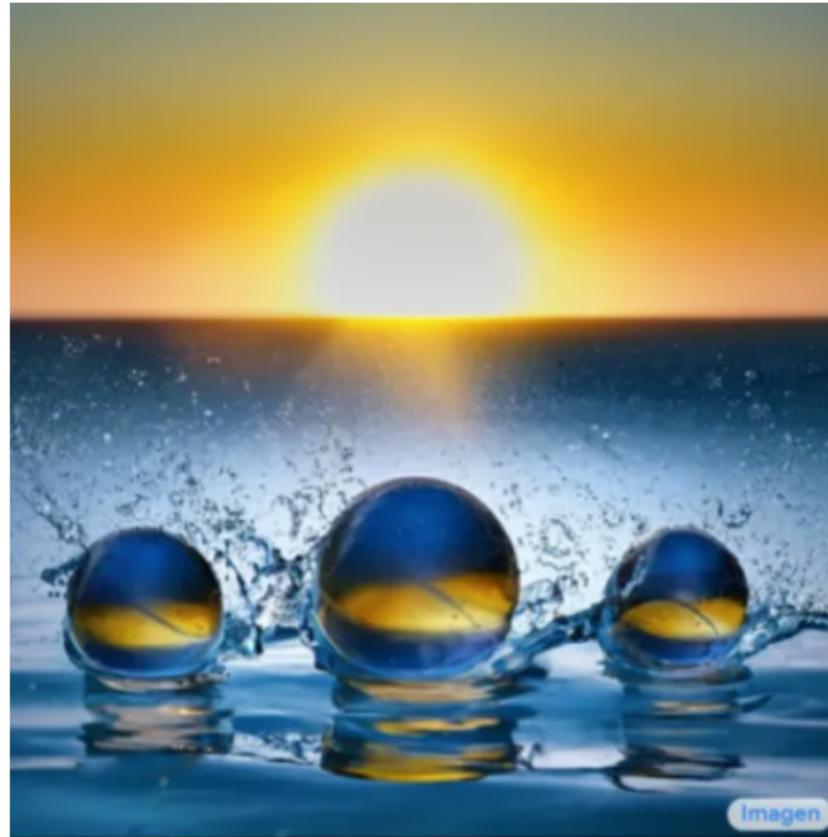
Dall-E's limitations

1. It can take a lot of adjusting prompts to get something good.
2. Sometimes it misunderstands what you meant, especially with long prompts.
3. At times DALL-E just doesn't know what you're asking for at all.
4. DALL-E can't deal with lots of extras or very long descriptions.
5. Sometimes things bleed over.
6. It can't do weird things.
7. Unsurprisingly, DALL-E is for images. So it's not great at text.
8. It's also often not brilliant at faces.

Google's "Imagen"

Three spheres
made of glass
falling into the
ocean. Water is
splashing. Sun is
setting.

Three spheres made of glass falling into the ocean. Water is splashing. Sun is setting.



A bald eagle
made of
chocolate
powder, mango
and whipped
cream.

A bald eagle
made of
chocolate
powder, mango
and whipped
cream.



Imagen

A photo of a
corgi dog riding
a bike in Times
Square. It is
wearing
sunglasses and a
beach hat.

A photo of a corgi dog riding a bike in Times Square. It is wearing sunglasses and a beach hat.



A bucket bag
made of blue
suede. The bag
is decorated with
intricate golden
paisley patterns.

The handle of
the bag is made
of rubies and
pearls.

A bucket bag made of blue suede. The bag is decorated with intricate golden paisley patterns.

The handle of the bag is made of rubies and pearls.



Imagen

A photo of a
raccoon wearing
an astronaut
helmet, looking
out of the
window at night.

A photo of a raccoon wearing an astronaut helmet, looking out of the window at night.



The Toronto
skyline with
Google brain
logo written in
fireworks.

The Toronto skyline with Google brain logo written in fireworks.



A zombie fly
suffering from a
fungus attacking
its body.

A zombie fly
suffering from a
fungus attacking
its body.

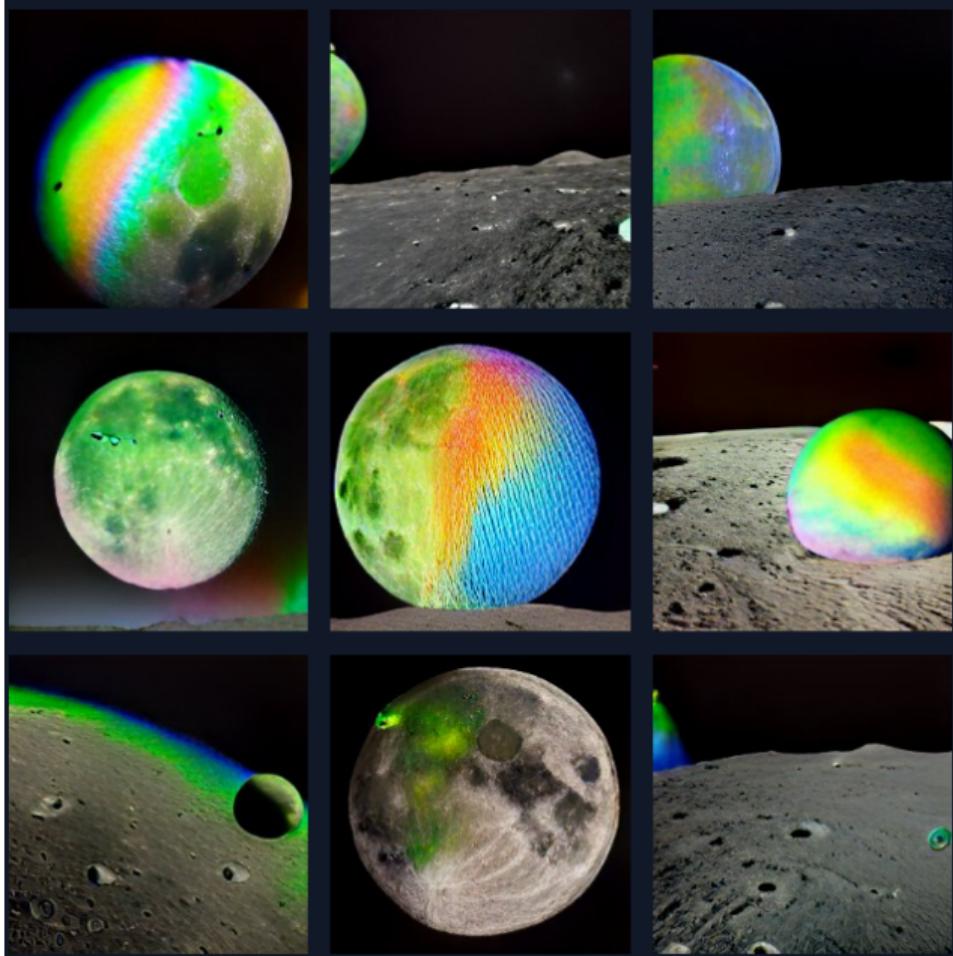


Craiyon

<https://tinyurl.com/craiyon-ai>

a rainbow
collored ball
bouncing on
moon's surface
beyond a small
green martian

a rainbow
collored ball
bouncing on
moon's surface
beyond a small
green martian



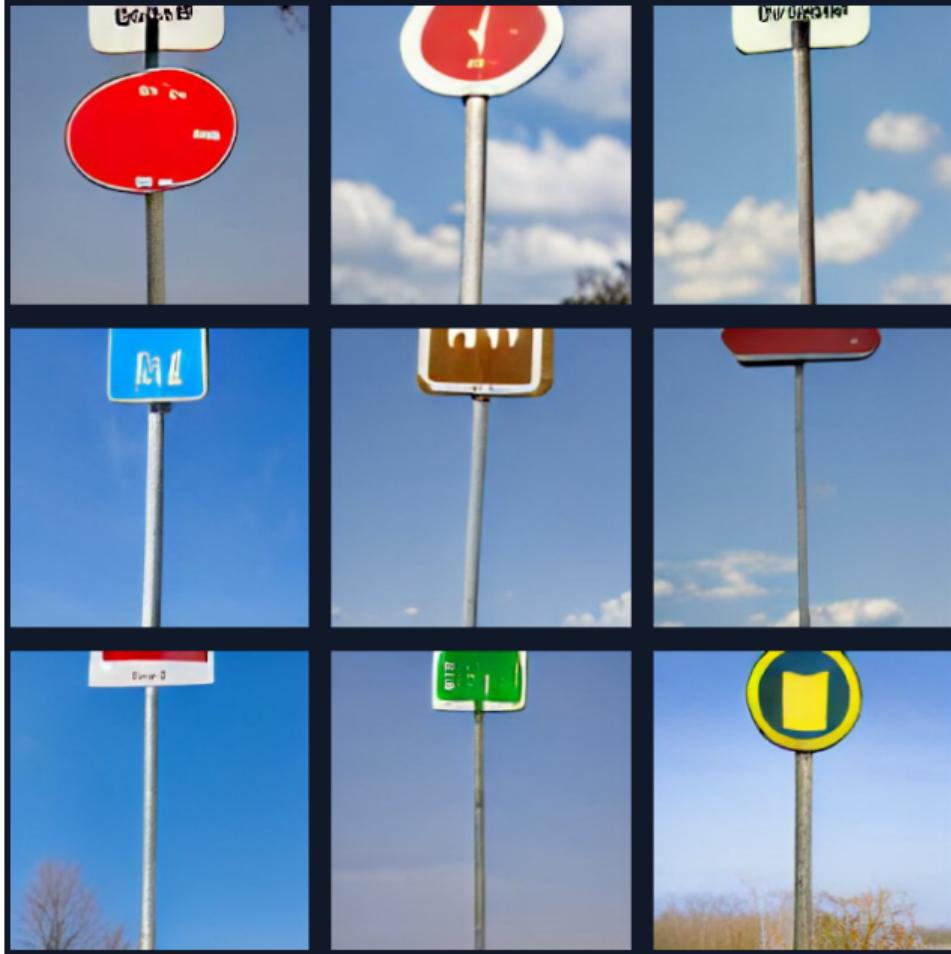
a green allien

a green alien



a sign saying:
“stop”

a sign saying:
“stop”



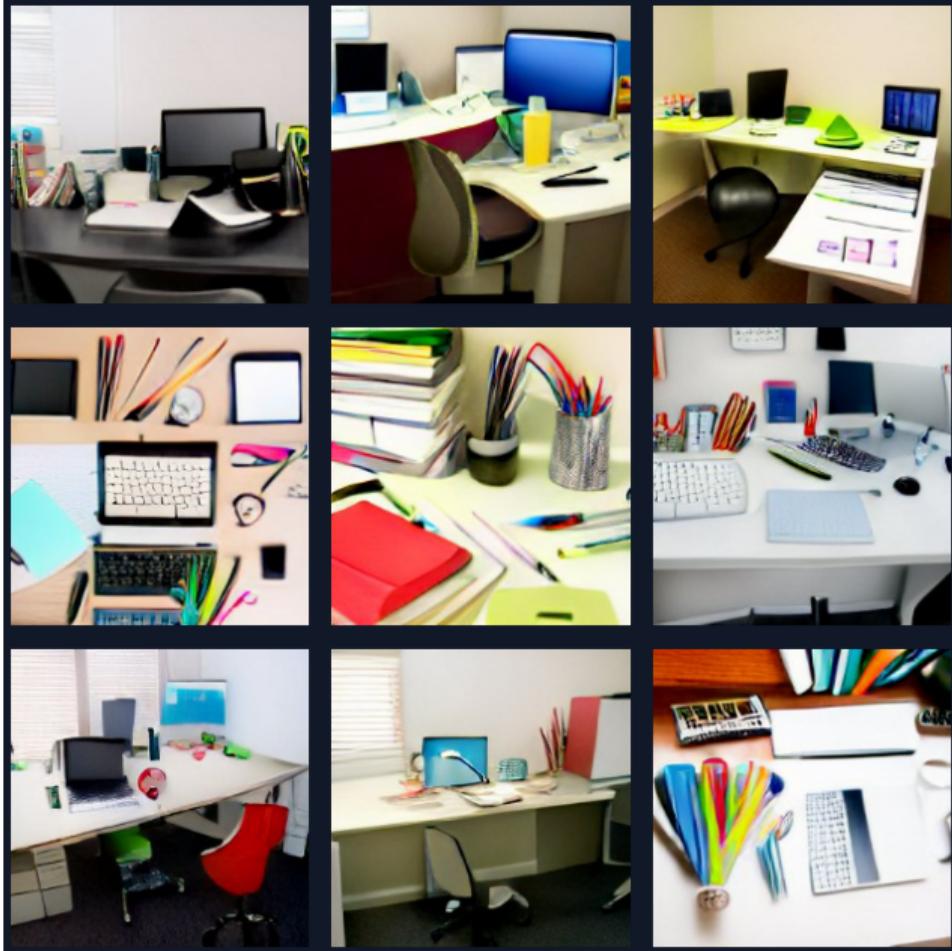
an unorganized
work office

an unorganized
work office



organized work
office

organized work
office



Stable Diffusion

<https://stability.ai/blog/stable-diffusion-public-release>

ChatGPT

<https://openai.com/blog/chatgpt/>

Thank you!
tvieira@ic.ufal.br