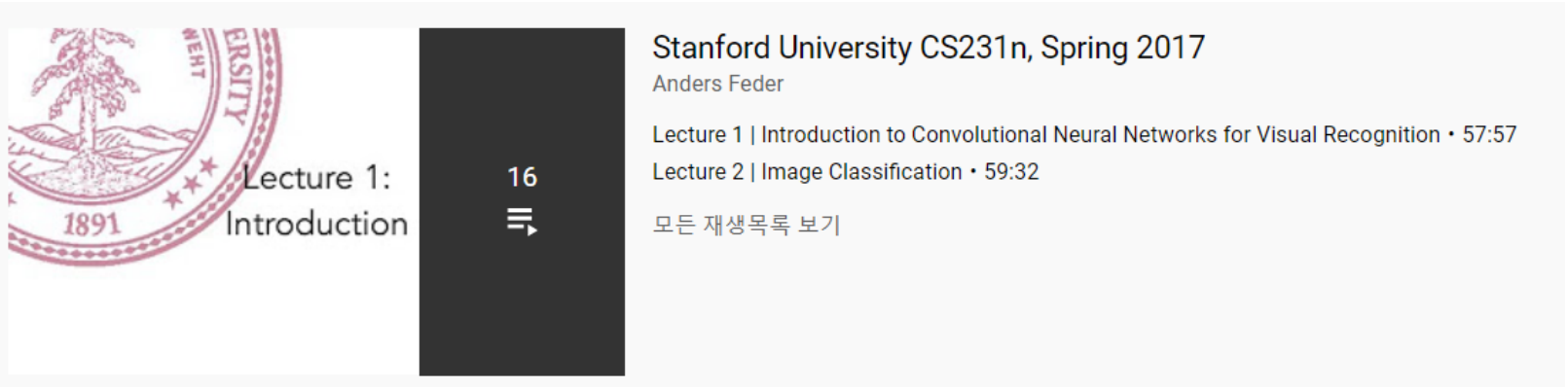
[1주차]
Introduction to Convolutional Neural Networks for Visual Recognition

1기 강다연
1기 김연수
1기 나정현

EURON

목차

Intro



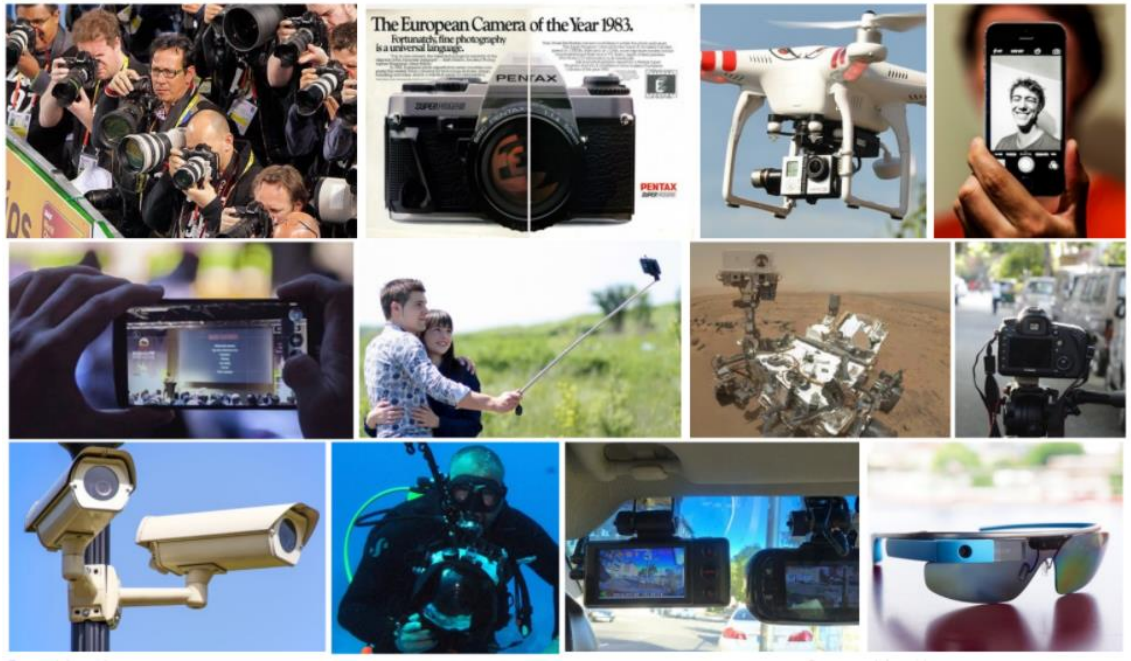
CS231n은 컴퓨터 비전에 관한 수업으로
Neual Network(신경망)과 CNN과 관련된 부분을 중점적으로 공부합니다.

Intro

대부분의 데이터가 영상

데이터 가공 및 이해해야 함

하지만 사람이 가공하기에는 한계 존재



History of Computer Vision

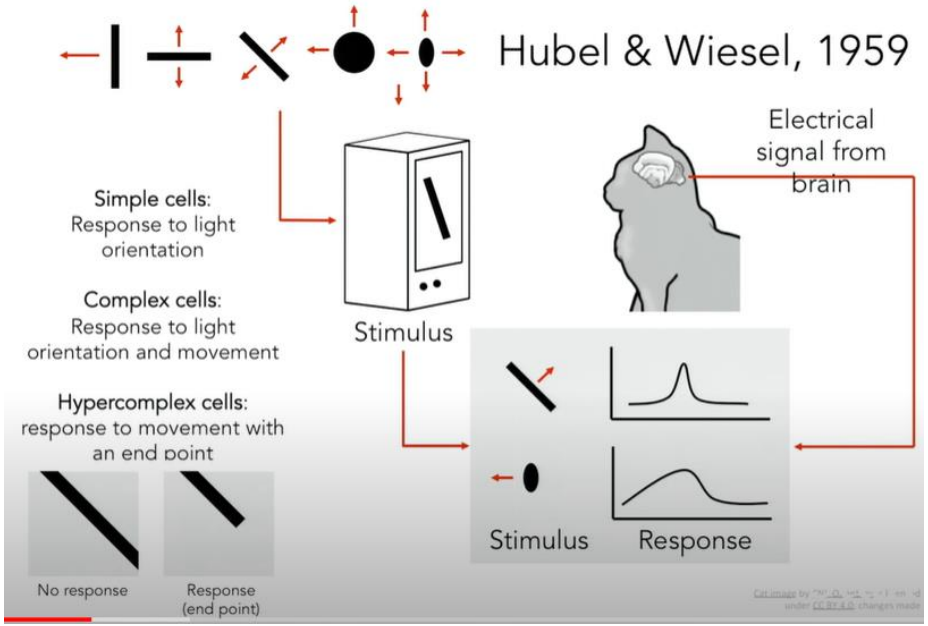
* 어디서 시작됐는가?

  진화의 시작은 항상 'vision'

  - 지능화된 종에서 가장 큰 sensory system

* 언제 시작됐는가?

* 현재(2017) 어느정도 와있는가?



vision의 매커니즘

*** 어떻게 발전했는가?**

가장 simple한 것부터 시작 : 방향

복잡

object recognition 등

이러한 visual process를 해체하여 하나씩 보면…

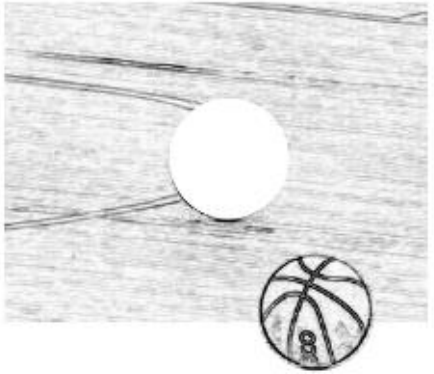
=> 간단한 단계를 모아 복잡한 결과를 창출

1. input
2. edge
3. 2+1/2D sketch : 물체를 분리해서
4. 3D model

Input image

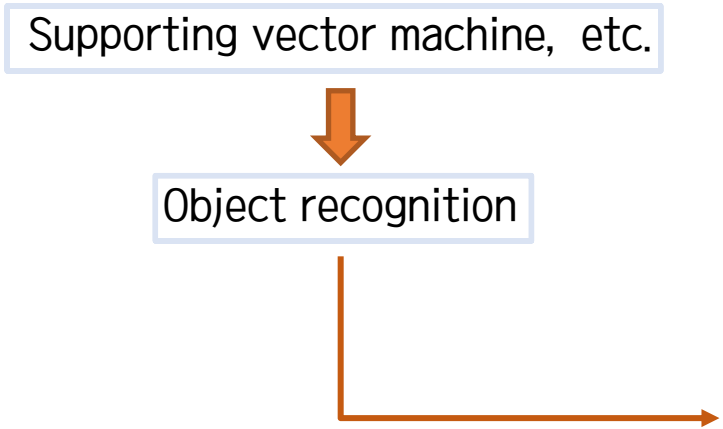
Edge image

2 ½-D sketch



3-D model

Algorithms and Fields



- object segmentation : Graph theory algorithm
- Face recognition
- Real-time face detection
- Feature-based object recognition
- Transfer / Match an entire object to entire object with different features

Dataset

Object recognition의 성능 평가를 위해 Benchmark dataset이 필요!

하지만… 시각 데이터는

- 복잡
- 고차원의 input
- 여러 hyper parameter 및 parameter의 조정이 필요
- 일반화하기 어려움

- PASCAL
- IMAGENET
  - 거의 모든 시각 데이터를 포함
  - object recognition 알고리즘이 발전하는데 가장 큰 역할을 함

👉 어떻게 생성되었는가?

: Data - WordNet - 분류/ 정리/ 라벨링

2012:

‘deep learning’ starts

SIFT Algorithm

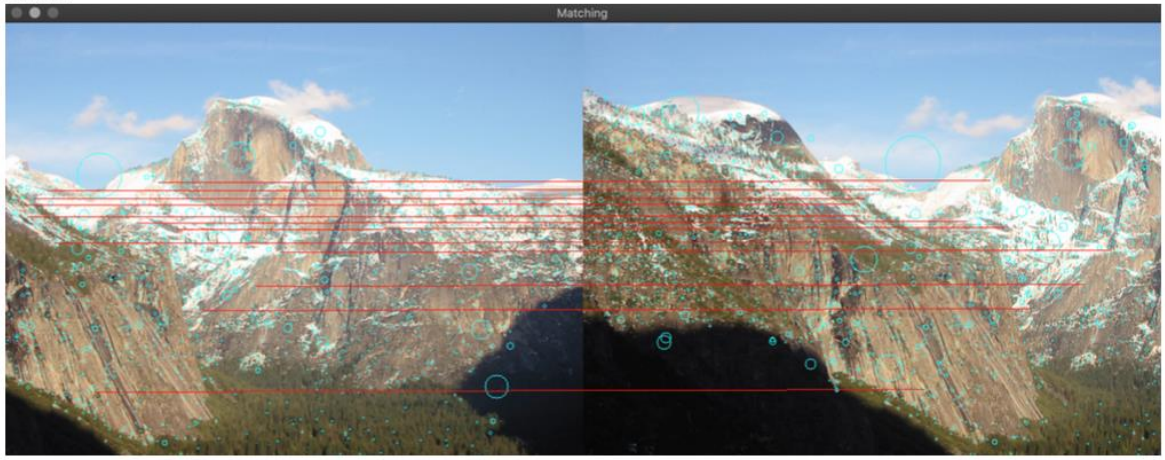
1990~2010년 : 특징 기반 객체인식 알고리즘↑

Image is public domain

Image is CC BY-SA 2.0

카메라 앵글, 조도 등의
특징들을 매칭해서 식별

Case3

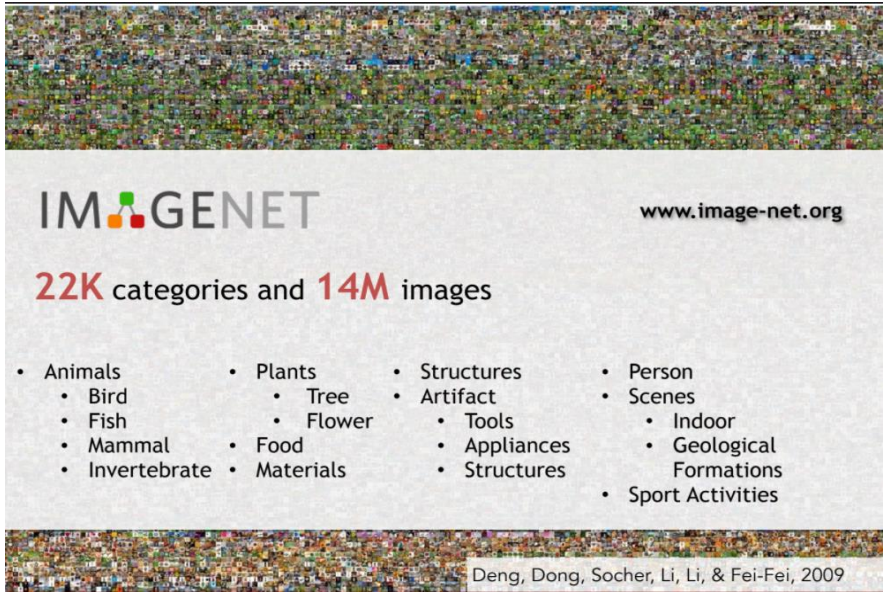


ImageNet_Image Classification

- 2000년대 초,
  CV 분야에서 풀어야 할 문제 정의

SVM, Adaboost
- training data set 부족
- 시각 데이터가 매우 복잡

overfitting 줄이기
+
세상의 모든 객체를 인식 해보자!



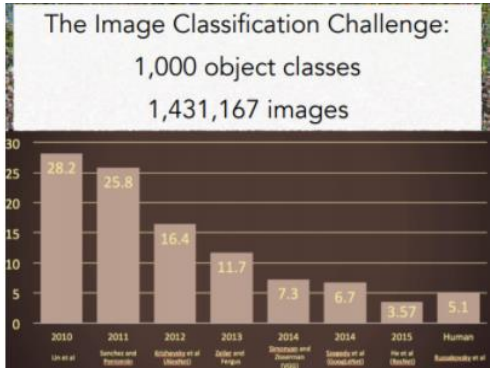


Image Classification 대회에서
CNN 기반 AlexNet 제안
-> 이를 계기로 CNN을 변형한 연구 ↑

Wrap up

*** 적용 분야**

- Medical diagnosis
- 로보틱스
- 자율주행 자동차

*** CV 분야의 목표**

- Build machine to see like humans
- 이미지에 담긴 story를 깊게 이해할 수 있게!

PT = 500ms

Some kind of game or fight. Two groups of two men? The man on the left is throwing something. Outdoors seemed like because i have an impression of grass and maybe lines on the grass? That would be why I think perhaps a game, rough game though, more like rugby than football because they pairs weren't in pads and helmets, though I did get the impression of similar clothing. maybe some trees? in the background. (Subject: SM)

1주차 과제

1. 2주차 과제를 하기 위한 setup하기

https://cs231n.github.io/assignments2020/assignment1/ 의 파일을 받아 구글드라이브에 옮겨주세요!

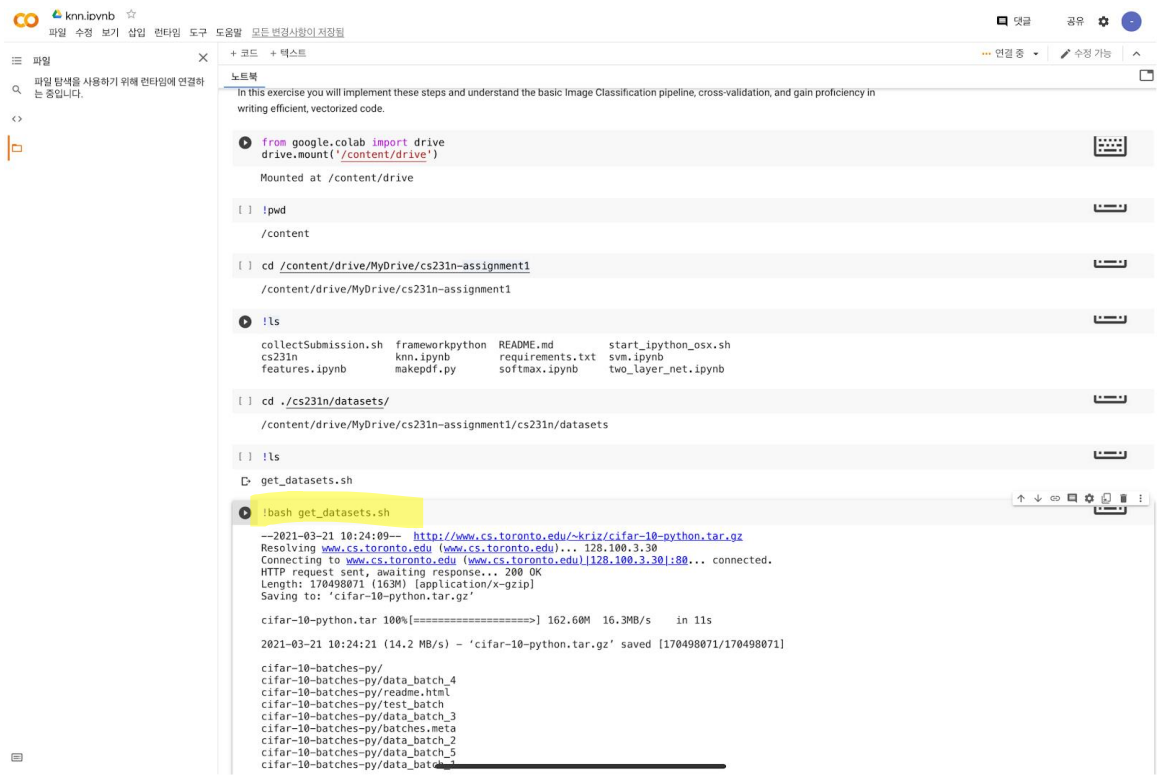
↓

코랩에서 해당 폴더 마운트 하고

↓

get_datasets.sh 실행한 뒤 캡처해서 인증

2. Python-numpy tutorial 실습



감사합니다

Q&A

요즘도?

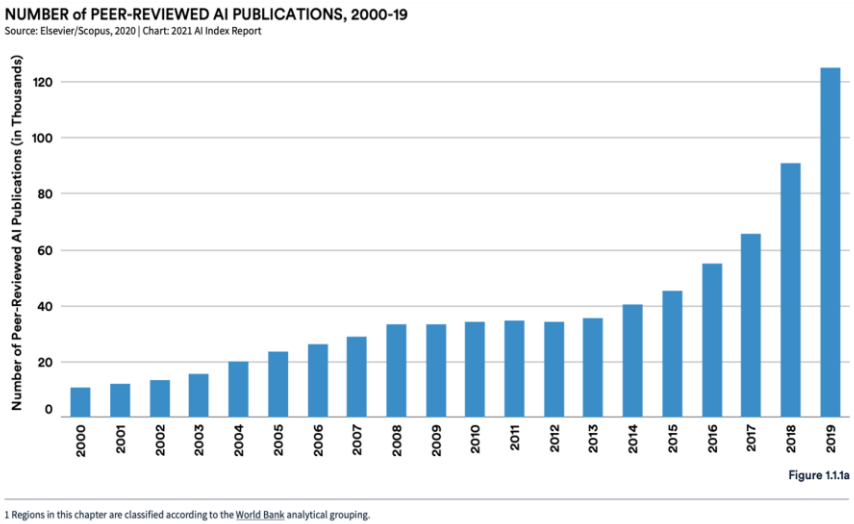
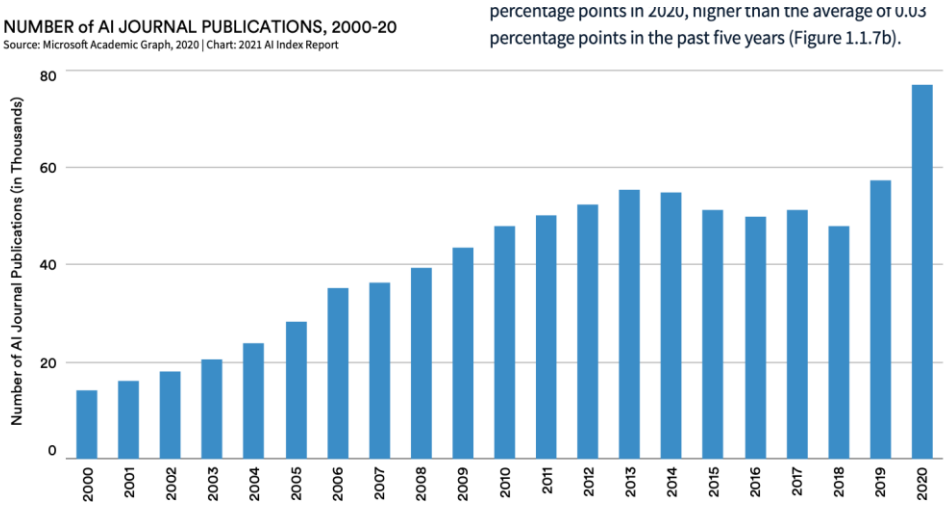


Figure 1.1.1a

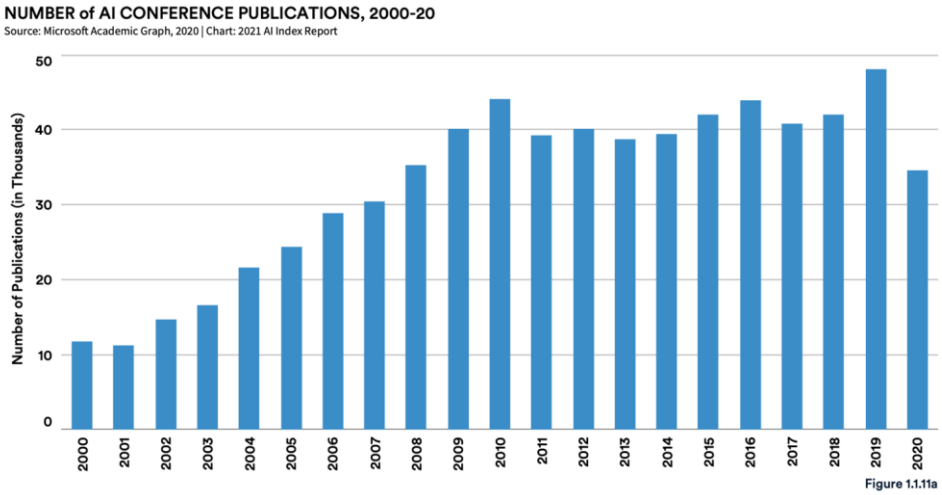
1 Regions in this chapter are classified according to the World Bank analytical grouping.

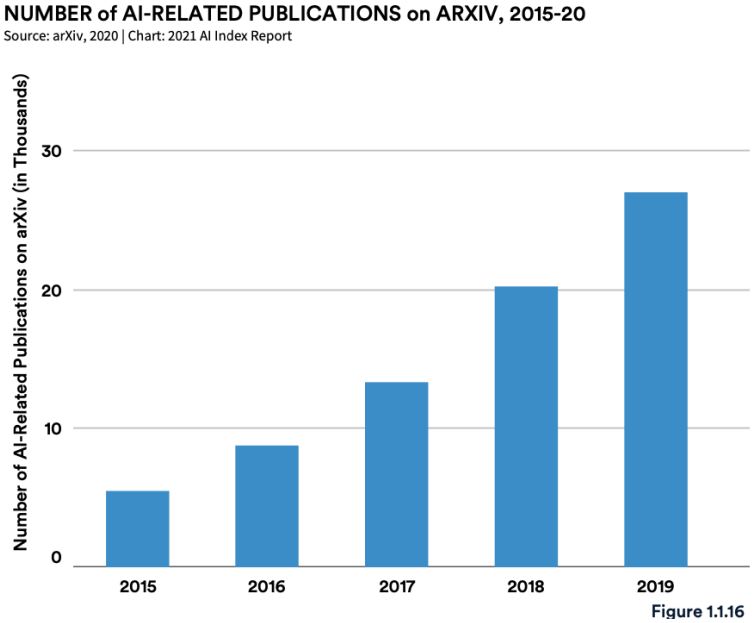


percentage points in 2020, higher than the average of 0.03 percentage points in the past five years (Figure 1.1.7b).

The 2021 AI Index

- Stanford Institute for Human-Centered Artificial Intelligence





The 2021 AI Index

- Stanford Institute for Human-Centered Artificial Intelligence

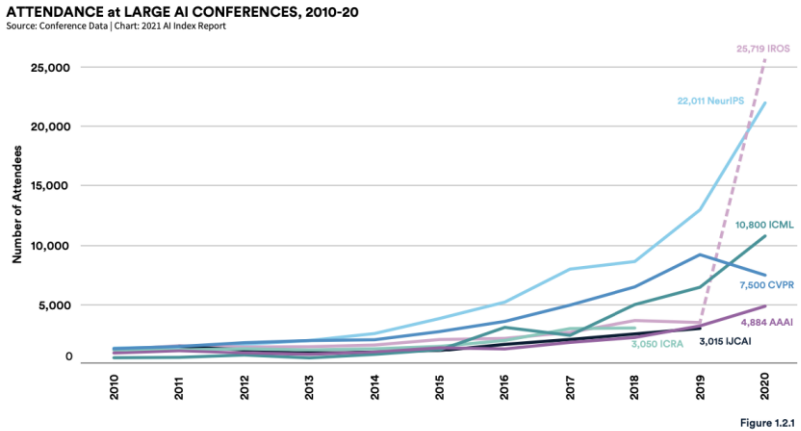


Figure 1.2.1

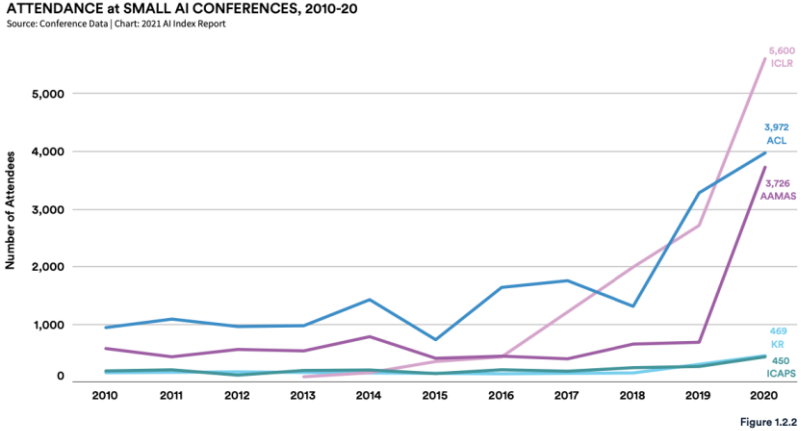
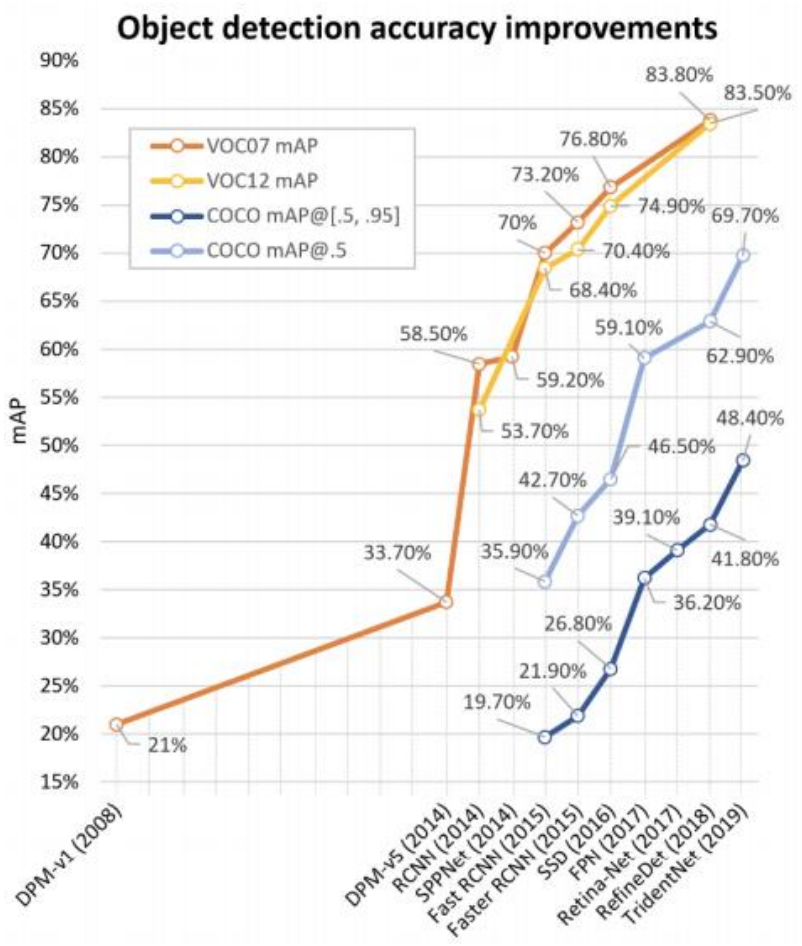


Figure 1.2.2

The 2021 AI Index

- Stanford Institute for Human-Centered Artificial Intelligence

Example : Object Detection Research



https://arxiv.org/pdf/1905.05055.pdf

Task

- Vision : Image Classification, Object Detection, Generation, Human Pose Estimation, Deepfake Detection, Semantic Segmentation, Activity Recognition
- Language : Machine Translation, Classification, Generation, QA, Summarization, NER
- Multimodal : VQA, Image Captioning, Image-text matching, Text-to-Image
- Speech : Speech Recognition, Voice Synthesis, Voice Conversion
- Recommendation
- Meta Learning
- Reinforcement Learning
- Graph Neural Network
- Domain Adaptation

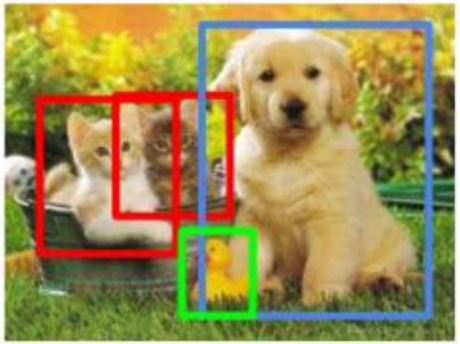
Task

- Vision : Image Classification, Object Detection, Generation, Human Pose Estimation, Deepfake Detection, Semantic Segmentation, Activity Recognition
- Language : Machine Translation, Classification, Generation, QA, Summarization, NER
- Multimodal : VQA, Image Captioning, Image-text matching, Text-to-Image
- Speech : Speech Recognition, Voice Synthesis, Voice Conversion
- Recommendation
- Meta Learning
- Reinforcement Learning
- Graph Neural Network
- Domain Adaptation

Classification

CAT

Object Detection



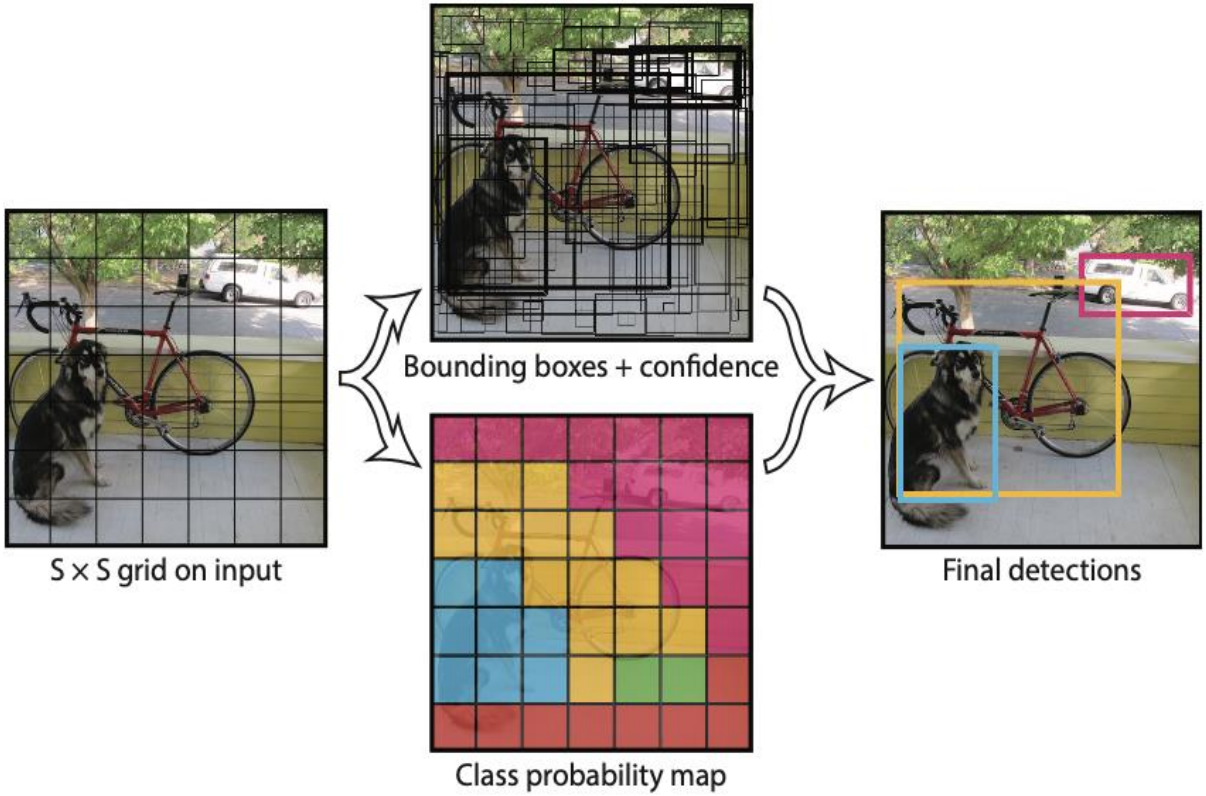
CAT, DOG, DUCK

Instance Segmentation

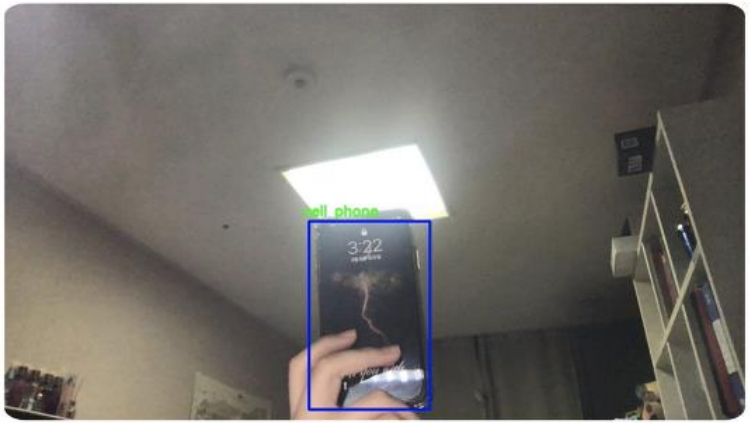
CAT, DOG, DUCK

https://hoya012.github.io/blog/Tutorials-of-Object-Detection-Using-Deep-Learning-what-is-object-detection/

YOLO


Bounding boxes + confidence
S × S grid on input
Class probability map
Final detections

🖥️ Video Streaming 🖥️

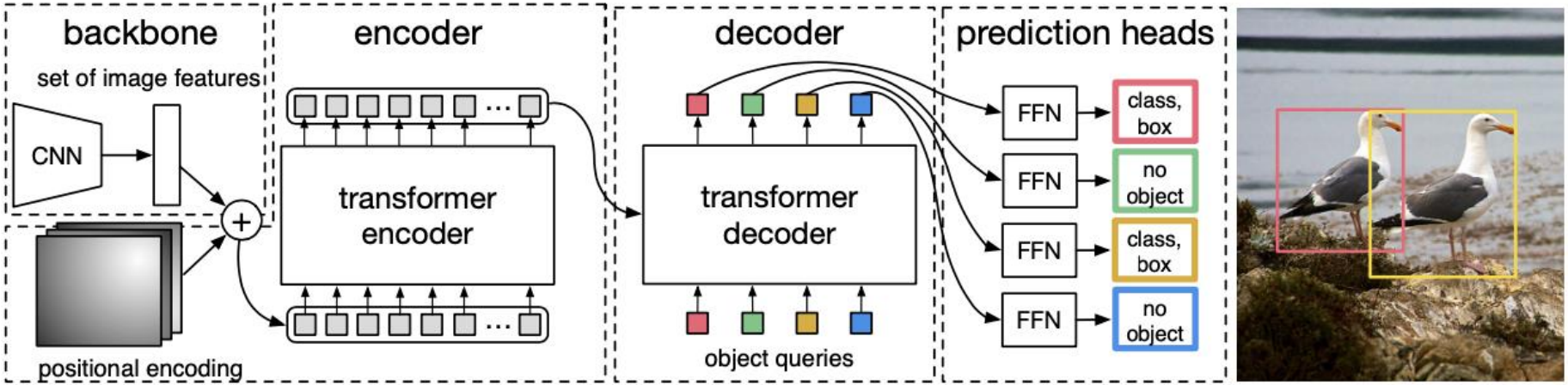


🤔 현재 predict label 🤔

예측을 시작하려면 Enter키↩️ 를 눌러주세요.

cell phone

DETR


backbone
set of image features
CNN
positional encoding
encoder
transformer encoder
decoder
transformer decoder
object queries
prediction heads
FFN
FFN
FFN
FFN
class, box
no object
class, box
no object

Vision Transformer(ViT)

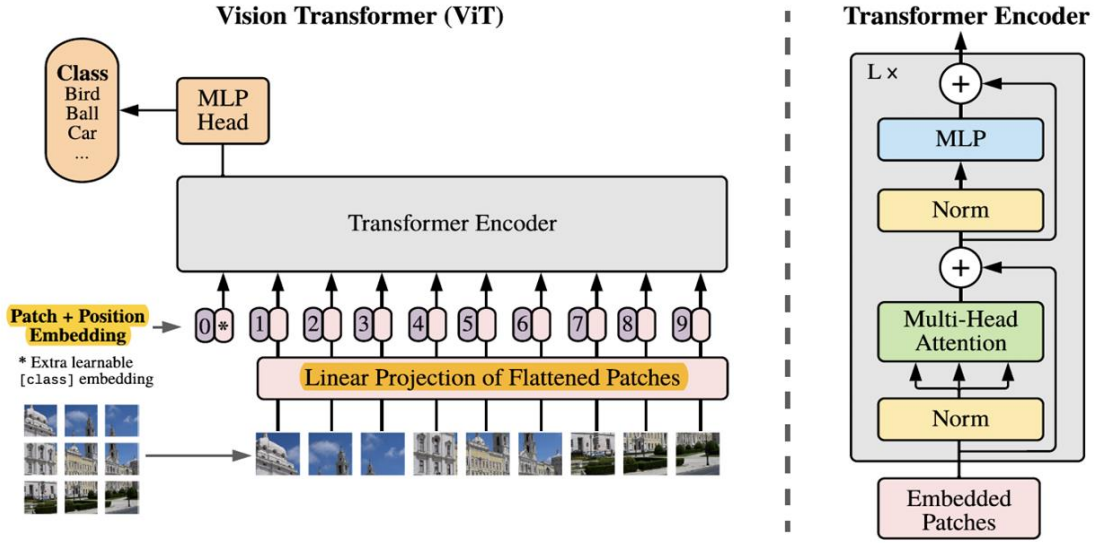


Figure 1: Model overview. We split an image into fixed-size patches, linearly embed each of them, add position embeddings, and feed the resulting sequence of vectors to a standard Transformer encoder. In order to perform classification, we use the standard approach of adding an extra learnable "classification token" to the sequence. The illustration of the Transformer encoder was inspired by Vaswani et al. (2017).

StarGAN

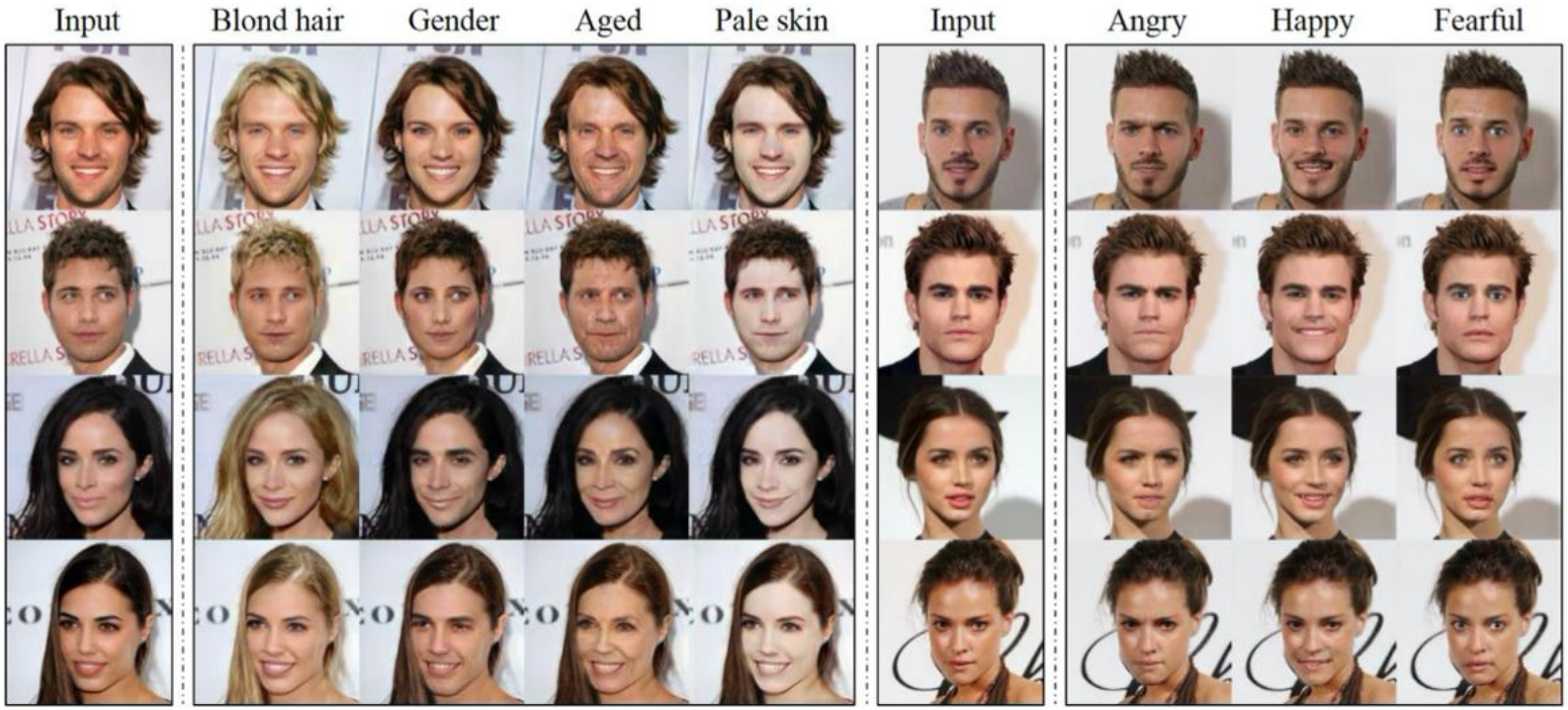
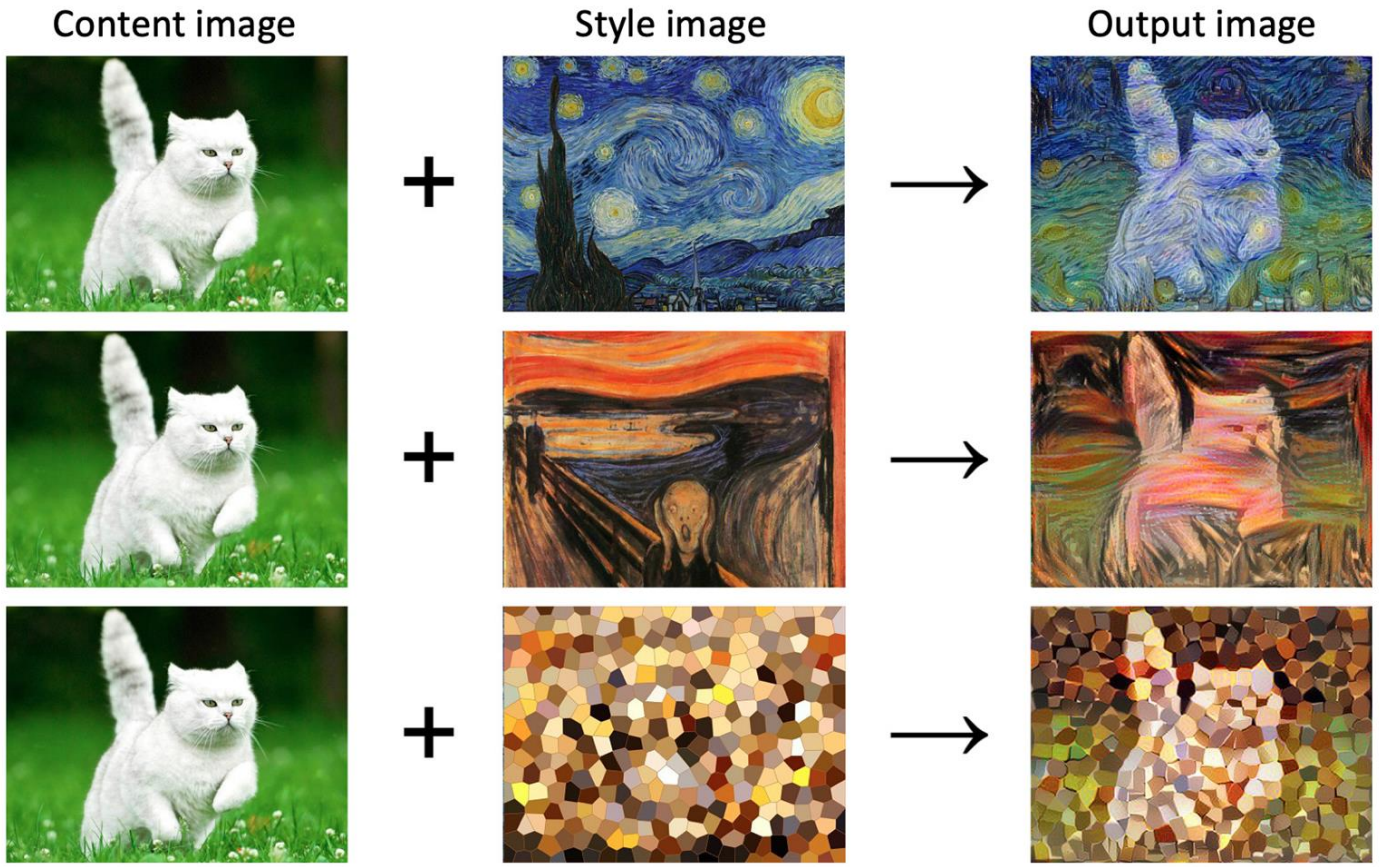
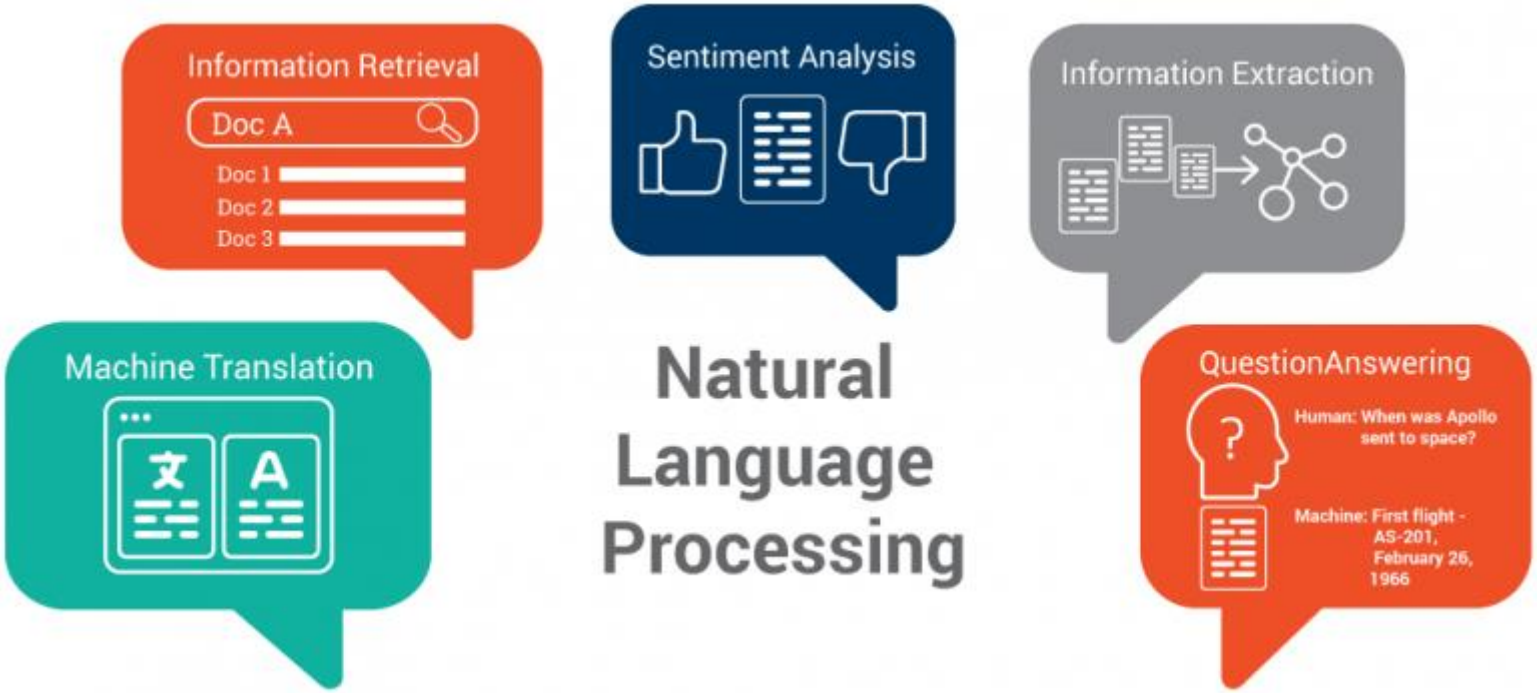


Figure 1. Multi-domain image-to-image translation results on the CelebA dataset via transferring knowledge learned from the RaFD dataset. The first and sixth columns show input images while the remaining columns are images generated by StarGAN. Note that the images are generated by a single generator network, and facial expression labels such as angry, happy, and fearful are from RaFD, not CelebA.

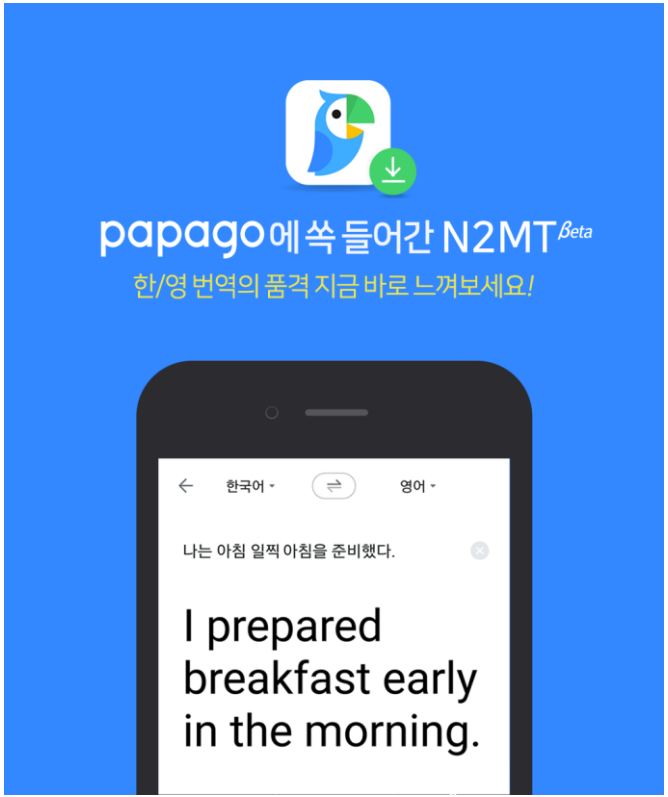
Style Transfer



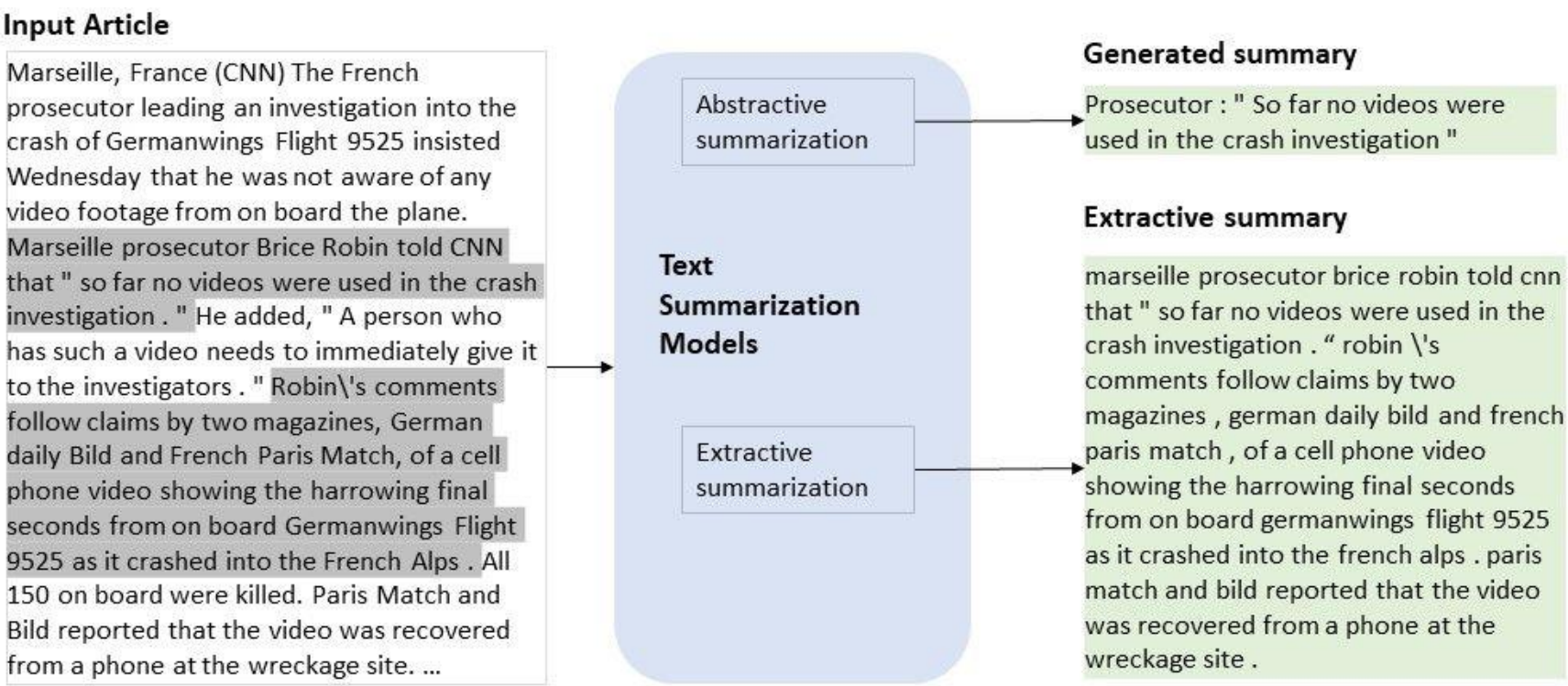


https://www.kdnuggets.com/2018/10/main-approaches-natural-language-processing-tasks.html

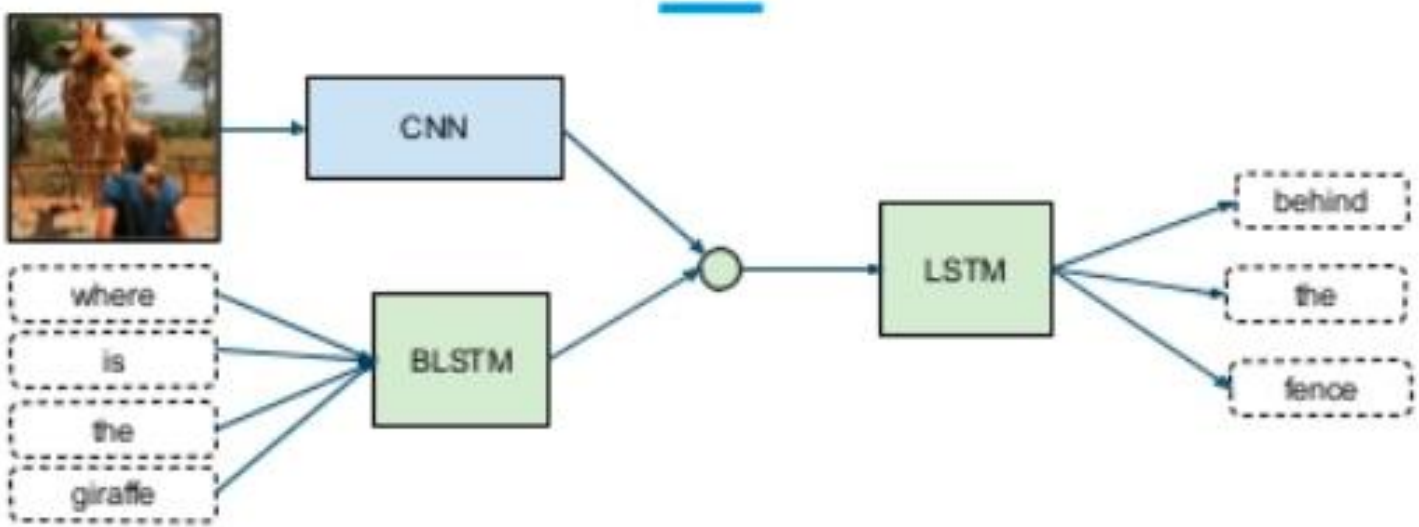
Machine Translation



Text Summarization



https://techcommunity.microsoft.com/t5/ai-customer-engineering-team/bootstrap-your-text-summarization-solution-with-the-latest/ba-p/1268809


CNN
where
is
the
giraffe
BLSTM
LSTM
behind
the
fence

Visual Question Answering (VQA)

Q: Does this foundation have any sunscreen?
A: yes

Q: What is this?
A: 10 euros

Q: What color is this?
A: green

Q: Please can you tell me what this item is?
A: butternut squash red pepper soup

Q: Is it sunny outside?
A: yes

Q: Is this air conditioner on fan, dehumidifier, or air conditioning?
A: air conditioning

Q: What type of pills are these?
A: unsuitable image

Q: What type of soup is this?
A: unsuitable image

Q: Who is this mail for?
A: unanswerable

Q: When is the expiration date?
A: unanswerable

Q: What is this?
A: unanswerable

Q: Can you please tell me what the oven temperature is set to?
A: unanswerable

https://vizwiz.org/tasks-and-datasets/vqa/

Visual Question Answering (VQA)

Q: Does this foundation have any sunscreen?
A: yes

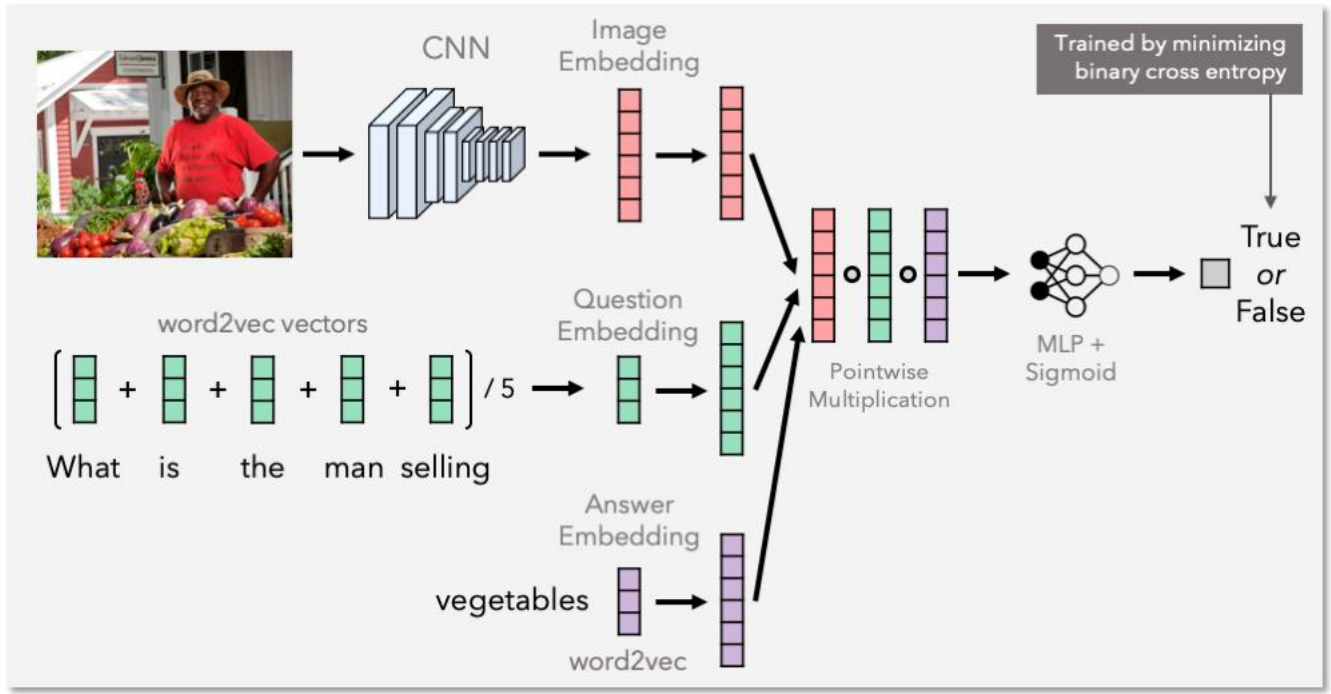
Q: What is this?
A: 10 euros

Q: What color is
A: green

Q: What type of pills are these?
A: unsuitable image

Q: What type of soup is this?
A: unsuitable image

Q: Who is this m
A: unanswerable



https://medium.com/ai2-blog/vanilla-vqa-adcaaaa94336

Image Captioning

A young boy is playing basketball. 	Two dogs play in the grass. 	A dog swims in the water. 	A little girl in a pink shirt is swinging. 
A group of people walking down a street. 	A group of women dressed in formal attire. 	Two children play in the water. 	A dog jumps over a hurdle. 

https://github.com/danieljl/keras-image-captioning

Text to Image Generation – Dall e

TEXT PROMPT

an illustration of a baby daikon radish in a tutu walking a dog

AI-GENERATED IMAGES

Edit prompt or view more images

TEXT PROMPT

an armchair in the shape of an avocado [...]

AI-GENERATED IMAGES

Edit prompt or view more images

TEXT PROMPT

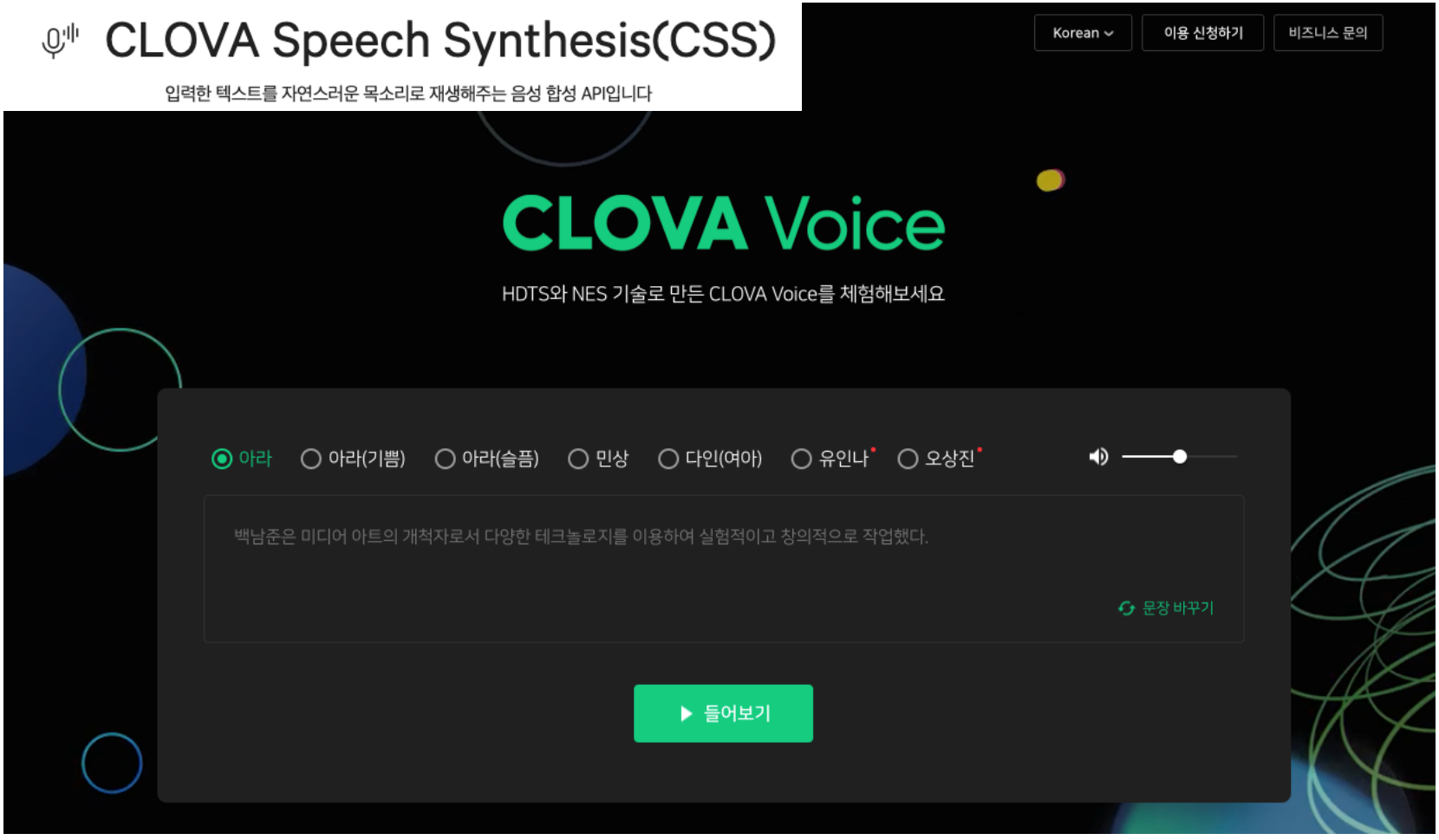
a store front that has the word 'openai' written on it [...]

AI-GENERATED IMAGES

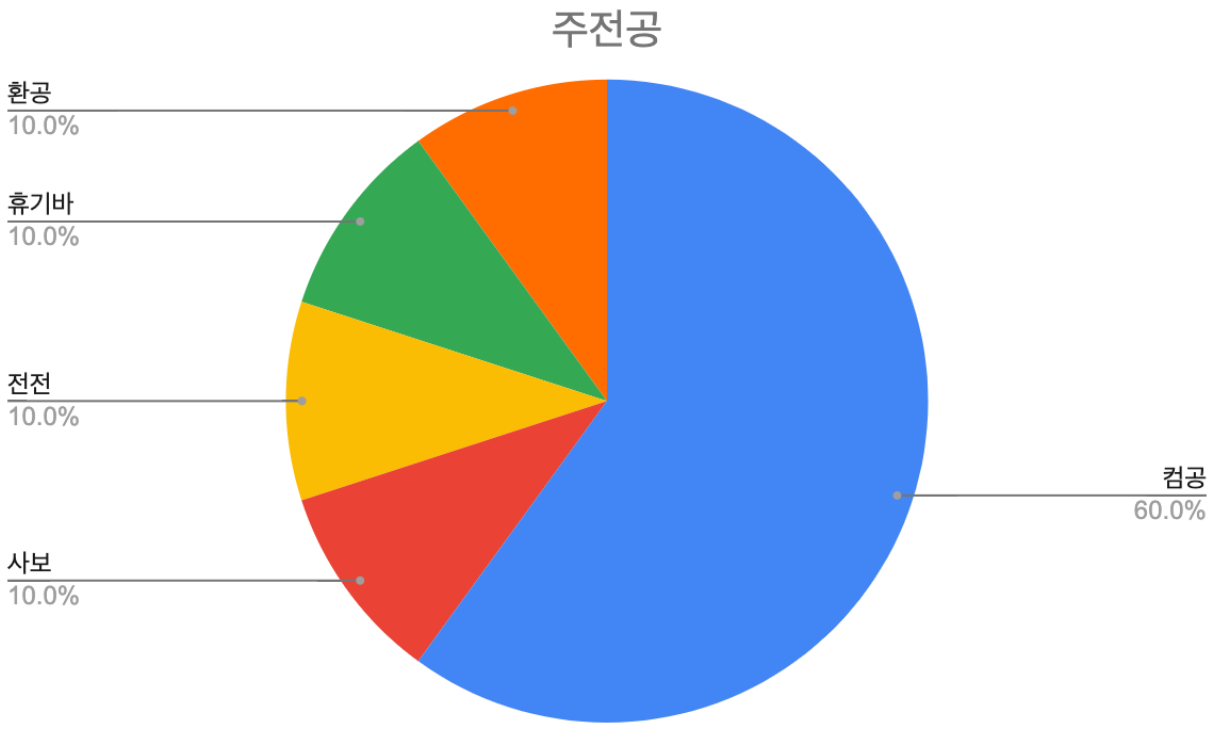
Edit prompt or view more images

https://openai.com/blog/dall-e/

Speech Synthesis



AI팀 벗들의 (주)전공 조사



Survey Result

Programming [0-5]			
Python	Numpy	Tensorflow	Pytorch
3.45	2.64	1.91	1.64
중급	초급		

Machine Learning [0-3]							
kNN	SVM	Tree	Ensemble	SGD	CNN	RNN	Transformer
1.27	1.18	1.36	1.27	1.27	1.91	1.36	1.00

CNN 빼고는 ㅜ.ㅜ