

CUAI Advanced Track Multimodal팀

2023.03.28





발표자 : 이하윤

구성원 소개 및 만남 인증

2월 2023
일 월 화 수 목 금 토
1 2 3 4
5 6 7 8 9 10 11
12 13 14 15 16 17 18
19 20 21 22 23 24 25
26 27 28

3월 2023

4월 2023
일 월 화 수 목 금 토
1
2 3 4 5 6 7 8
9 10 11 12 13 14 15
16 17 18 19 20 21 22
23 24 25 26 27 28 29
30

일요일	월요일	화요일	수요일	목요일	금요일	토요일
			1	2	3	4
5	6	7	8	9	10	11
12	13	14 	15	16	17 	18
19	20	21	22	23	24 	25
26	27	28	29	30	31 	

7calendar.com/kr/

소프트웨어학부 이하윤
소프트웨어학부 김동영
융합공학부 김벼리


THOR

구성원 소개 및 만남 인증

3월 2023
일 월 화 수 목 금 토
1 2 3 4
5 6 7 8 9 10 11
12 13 14 15 16 17 18
19 20 21 22 23 24 25
26 27 28 29 30 31

4월 2023

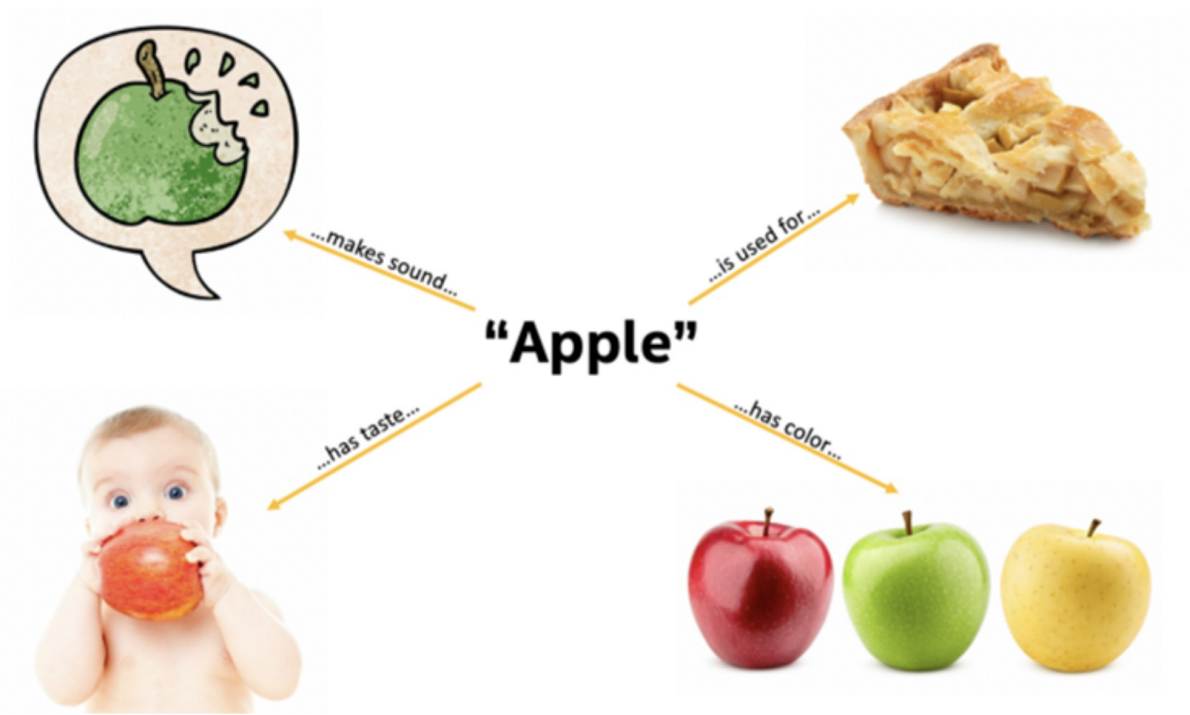
5월 2023
일 월 화 수 목 금 토
1 2 3 4 5 6
7 8 9 10 11 12 13
14 15 16 17 18 19 20
21 22 23 24 25 26 27
28 29 30 31

일요일	월요일	화요일	수요일	목요일	금요일	토요일
						1
2	3	4	5	6		8
9	10	11	12	13	14	15
16	17	18	19	20	21	22
23	24	25	26	27	28	29
30						

7calendar.com/kr/

소프트웨어학부 이하윤
소프트웨어학부 김동영
융합공학부 김벼리

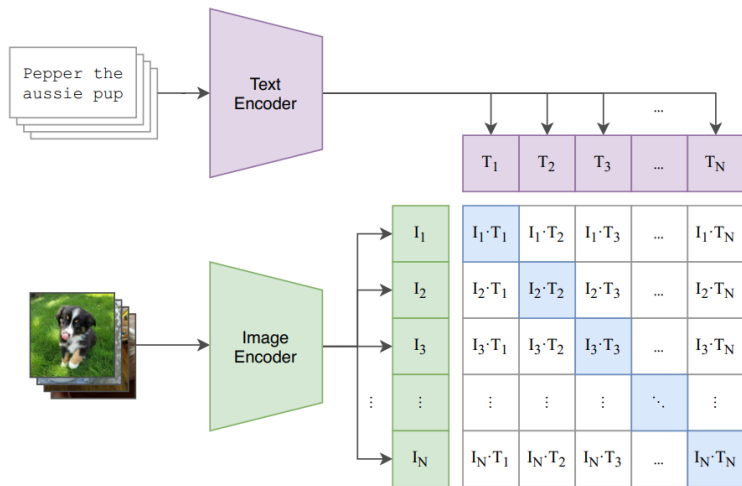
멀티모달이란?



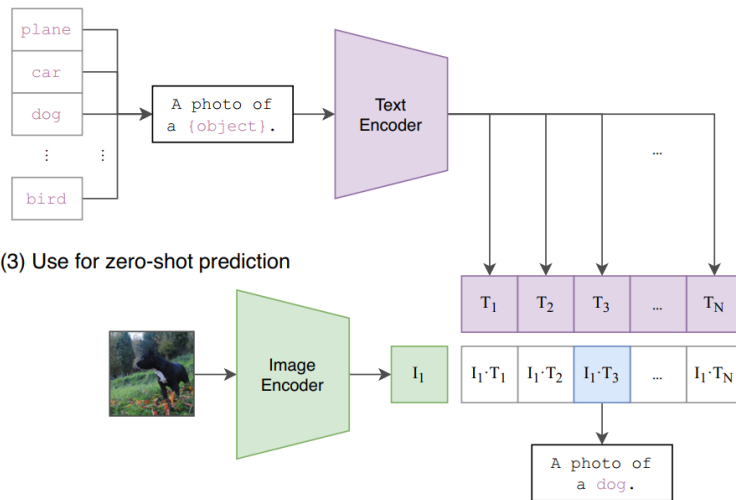
인간이 사과를 이해하는 방식은 다양해요. (출처: Intel Labs)

논문 리딩 - CLIP

(1) Contrastive pre-training

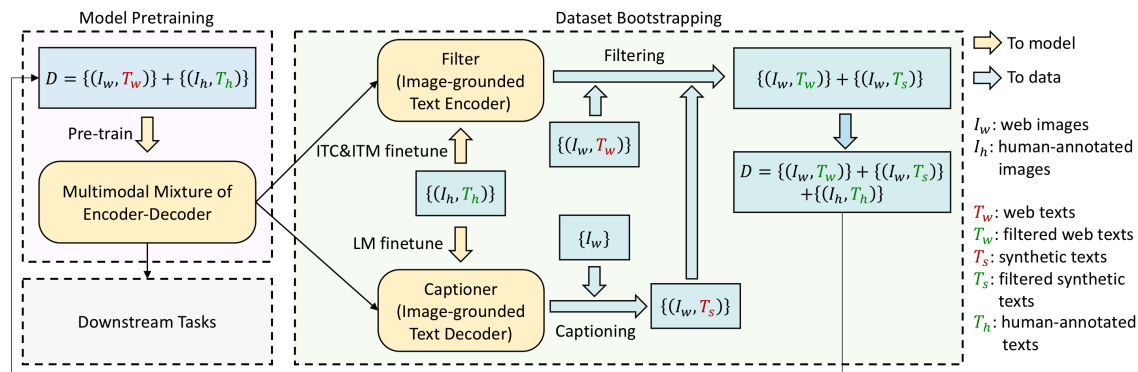
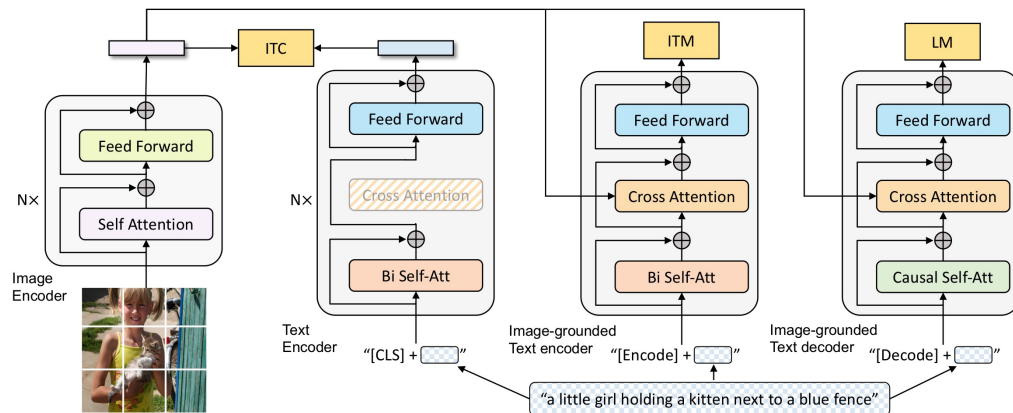


(2) Create dataset classifier from label text



(3) Use for zero-shot prediction

논문 리딩 - BLIP



참여 대회

Featured Code Competition

Stable Diffusion - Image to Prompts

Deduce the prompts that generated our "highly detailed, sharp focus, illustration, 3d renders of majestic, epic" images

Kaggle · 754 teams · 2 months to go (a month to go until merger deadline)

Submissions

Submit Predictions

...

Overview

Description

Evaluation

Timeline

Prizes

Code Requirements

Goal of the Competition

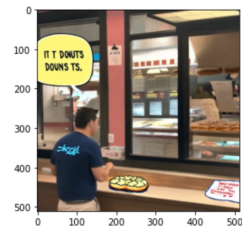
The goal of this competition is to reverse the typical direction of a generative text-to-image model: instead of generating an image from a text prompt, can you create a model which can predict the text prompt given a generated image? You will make predictions on a dataset containing a wide variety of (prompt, image) pairs generated by Stable Diffusion 2.0, in order to understand how reversible the latent relationship is.

Context

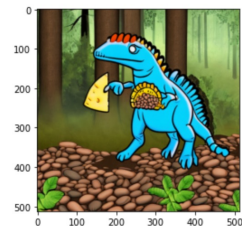
The popularity of text-to-image models has spawned an entire new field of prompt engineering. Part art and part unsettled science, ML practitioners and researchers are rapidly grappling with understanding the relationships between prompts and the images they generate. Is adding "4k" to a prompt the best way to make it more photographic? Do small perturbations in prompts lead to highly divergent images? How does the order of prompt keywords impact the resulting generated scene? This competition tasks you with creating a model that can reliably invert the diffusion process that generated to a given image.

In order to calculate prompt similarity in a robust way—meaning that "epic cat" is scored as similar to "majestic kitten" in spite of character-level differences—you will submit embeddings of your predicted prompts. Whether you model the embeddings directly or first predict prompts and then convert to embeddings is up to you! Good luck, and may you create "highly quality, sharp focus, intricate, detailed, in the style of unreal robust cross validation" models herein.

This is a Code Competition. Refer to Code Requirements for details.



a man standing in front of a counter with two donuts on it



a blue dinosaur eating a piece of cheese in a forest