VLM 끄적끄적 4

Pretraining Task (3): Alignment Objectives

Alignment Objectives의 종류

1. (Global) Image-Text Matching

2. (Local) Region-Word Matching

1. (Global) Image-Text Matching

앞선 Section에서 살펴봤던 알고리즘들

예시

- FLAVA: Image & Text에 대한 binary classification

- FIBER: Pair-wise simliarity 바탕으로 mine hard negatives

2. (Local) Region-Word Matching

목적: 보다 "fine-grained"한 관계를 포착하자

=> Dense prediction task를 보다 더 잘 하기 위해

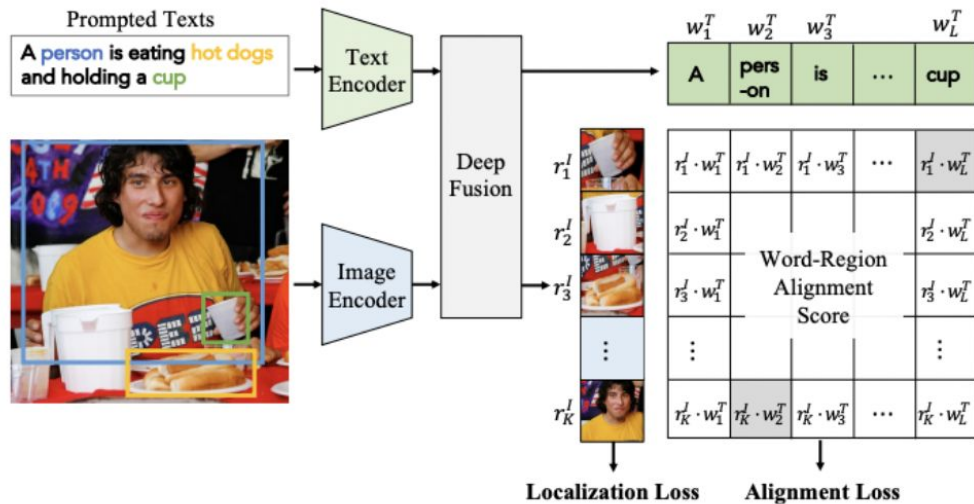- e.g., Object detection, Semantic segmentation

예시

- GLIP, FIBER, DetCLIP

2. (Local) Region-Word Matching

GLIP (CVPR 2022)

- Grounded + CLIP

- Object classification 태스크를

  region-word alignment로 치환!

Summary

(1) Contrastive Objectives

- 1-1) Image CL
- 1-2) Image-Text CL
- 1-3) Image-Text-Label CL

(2) Generative Objectives

- 2-1) MIM
- 2-2) MLM
- 2-3) MCM
- 2-4) Image-to-Text Generation

(3) Alignment Objectives

- 3-1) Image-Text Matching
- 3-2) Region-Word Matching