

Supplementary Material for PositiveCoOp: Rethinking Prompting Strategies for Multi-Label Recognition with Partial Annotations

Samyak Rawlekar, Shubhang Bhatnagar, Narendra Ahuja
University of Illinois Urbana-Champaign
samyakr2, sb56, n-ahuja@illinois.edu

1. Additional Visual Results

1.1. Prompt-Image Embedding Similarity Map

In Fig. 1, we provide additional visualizations comparing the similarity maps generated by the calculating the cosine similarity between image features and positive prompt features and image features and negative prompts features for each class. Consistent with the main paper, for each class, both the positive and negative prompt similarity maps activate the same regions that are associated with the presence of the class.

1.2. LAION-400M Dataset Visualization

In Fig. 2, we provide examples of image-text pairs from the LAION-400M dataset [2]. We observe that text descriptions mainly focus on the presence of objects(class) and do not describe the absence of objects(class) not present in the image. This explains why CLIP struggles to effectively guide the learning of a negative prompt.

2. Details of Experiments in Sec. 5 of Paper

2.1. Text Encoders Focus on Features Associated with Presence of a Class

In Table. 1, we list all 85 prompts that were used to generate the quantitative results in Sec 5.2

2.2. Presence of Negative Prompts in LAION-400M

In this section, we provide the details for Sec. 5.1 of the main paper, where we show that less than 0.5% of the LAION400M dataset has negative words, and even fewer have negative words followed by a noun (not necessarily immediately). The list of negative words we used for the task is as follows:

```
Negative Words = {  
    'not', 'no', 'never',  
    'none', 'nothing', 'nobody',  
    'nowhere', 'neither', 'nor',  
    "can't", "cannot", "won't",  
    "don't", "doesn't", "didn't",  
    "isn't", "aren't", "wasn't",  
    "weren't", "hasn't", "haven't",  
    "hadn't", "shouldn't", "wouldn't",  
    "couldn't", "mustn't"  
}
```

Below, we provide some examples of text containing negative words and the phrases that include the negative word followed by a noun, referred to as 'Phrase'.

1. Post-it note saying I will not compare myself to a stranger on Instagram
Phrase: not compare myself to a stranger
2. I'm not getting any younger Magnet
Phrase: not getting any younger Magnet
3. Lionel I was told this cake was not cut. It was painted with butter Cream Icing right on the cake. No transfer.
Phrase: No transfer
4. Dennis Pitta was not targeted often, but made his presence felt against Denver.
Phrase: not targeted often, but made his presence
5. Introduction Video recommendation in YouTube Related to the current video topic and user profile but not visualized
Phrase: -
6. Arguing, Children, and Facts: AMERICAN SNIPER Follow NEMA ACTS @cinfacts for more content CHRIS KYLE IS DEPICTED IN THE MOVIE AS SHOOTING THE CHILD WITH THE GRENADE AND THEN THE MOTHER WHEN SHE PICKED IT UP TO THROW IT. BUT IN THE REAL INCIDENT CHRIS KYLE NEVER SHOT ANY CHILDREN. SHE WAS THE ONLY ONE SHOT WHEN

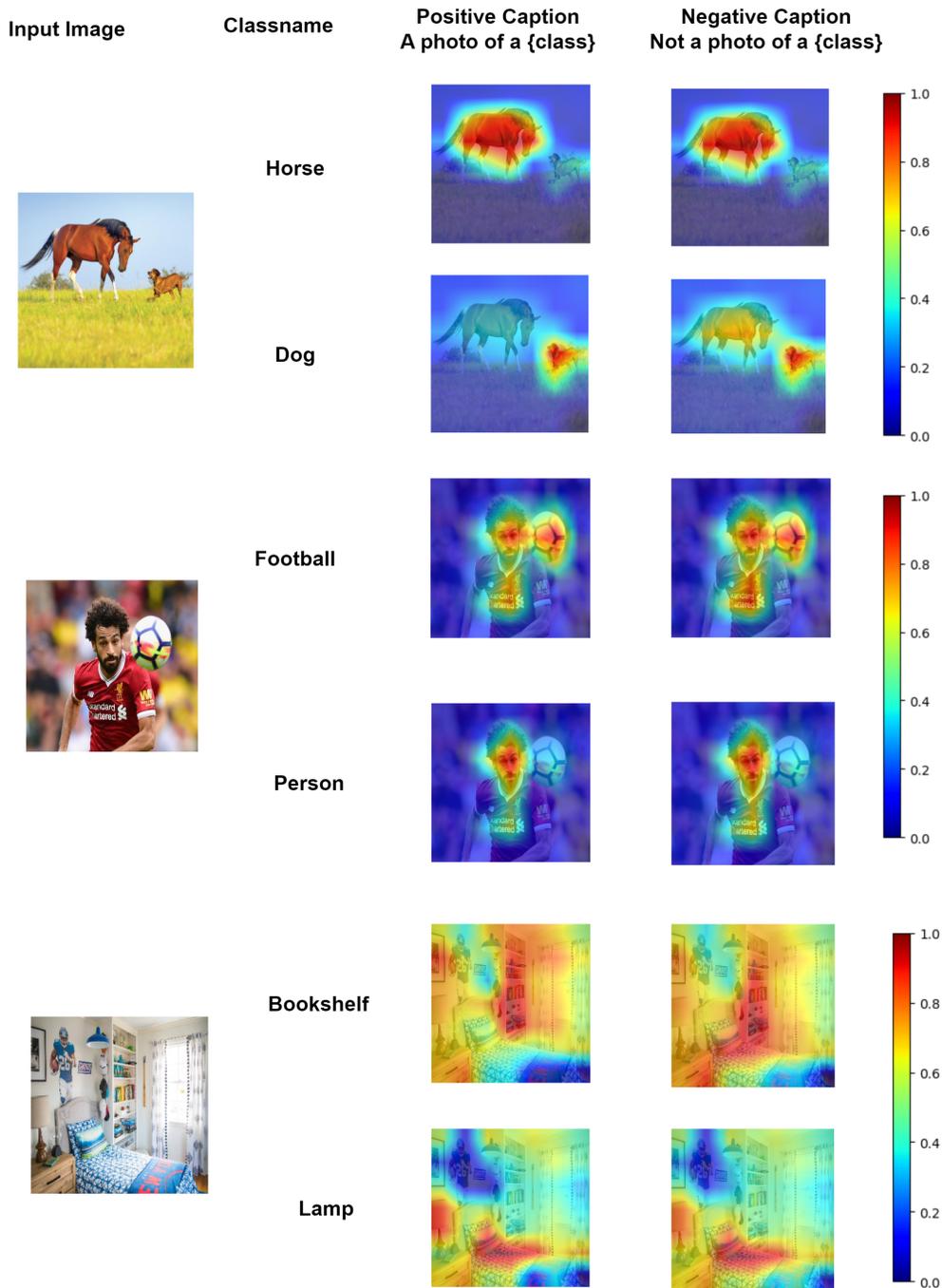


Figure 1. **Visualization of Similarity Maps.** We compare similarity maps obtained using cosine similarity between image features and positive prompt features versus image features and negative prompt features for each class. The activation of similar regions in both maps questions the effectiveness of CLIP’s guidance for learning a negative prompt.

SHE ATTEMPTED TO ATTACK U.S. FORCES. People will argue, ""It’s just a kid!"" Well, I’d agree with you but if Irín a soldier and some kid runs at me with a live explosive, Irín not going to hesitate. Your thoughts?
- Follow @cinfacts for more facts.

Phrase: NEVER SHOT

7. Yes i do.....but not with you ! aprons

Phrase: not with you ! aprons

8. Scary creatures: Brown’s monster-strewn Hollywood

a bad photo of a {}.	a photo of many {}.	a sculpture of a {}.
a photo of the hard to see {}.	a low resolution photo of the {}.	a rendering of a {}.
graffiti of a {}.	a bad photo of the {}.	a cropped photo of the {}.
a tattoo of a {}.	the embroidered {}.	a photo of a hard to see {}.
a bright photo of a {}.	a photo of a clean {}.	a photo of a dirty {}.
a dark photo of the {}.	a drawing of a {}.	a photo of my {}.
the plastic {}.	a photo of the cool {}.	a close-up photo of a {}.
a black and white photo of the {}.	a painting of the {}.	a painting of a {}.
a pixelated photo of the {}.	a sculpture of the {}.	a bright photo of the {}.
a cropped photo of a {}.	a plastic {}.	a photo of the dirty {}.
a jpeg corrupted photo of a {}.	a blurry photo of the {}.	a photo of the {}.
a good photo of the {}.	a rendering of the {}.	a {} in a video game.
a photo of one {}.	a doodle of a {}.	a close-up photo of the {}.
a photo of a {}.	the origami {}.	the {} in a video game.
a sketch of a {}.	a doodle of the {}.	a origami {}.
a low resolution photo of a {}.	the toy {}.	a rendition of the {}.
a photo of the clean {}.	a photo of a large {}.	a rendition of a {}.
a photo of a nice {}.	a photo of a weird {}.	a blurry photo of a {}.
a cartoon {}.	art of a {}.	a sketch of the {}.
a embroidered {}.	a pixelated photo of a {}.	itap of the {}.
a jpeg corrupted photo of the {}.	a good photo of a {}.	a plushie {}.
a photo of the nice {}.	a photo of the small {}.	a photo of the weird {}.
the cartoon {}.	art of the {}.	a drawing of the {}.
a photo of the large {}.	a black and white photo of a {}.	the plushie {}.
a dark photo of a {}.	itap of a {}.	graffiti of the {}.
a toy {}.	itap of my {}.	a photo of a cool {}.
a photo of a small {}.	a tattoo of the {}.	there is a {} in the scene.
there is the {} in the scene.	this is a {} in the scene.	this is the {} in the scene.
this is one {} in the scene.		

Table 1. **Prompt Templates:** Provides the complete list of prompts (85 default prompts from ImageNet [1]) that were used to find the similarity between prompt features.

Hills mansion not surprisingly distressed his wealthy neighbours

Phrase: not surprisingly distressed his wealthy neighbours

- An elderly monk meditates before the historic Shwedagon Paya, Yangon, Unuon of Myanmar (Burma), Nov. 26, 2009. The Shwedagon pagoda's central hti, an umbrella spire atop the giant zedi structure, sports a 76-karat diamond that casts red, green white beams to specific spots on the terrace as the sun rises or sets. The massive complex sits atop a 190-foot hill accessed by four stair-stepped walkways guarded by 30-foot-tall mythical half-lion half-dragon creatures called chinthe. The central 98-foot-tall zedi is surrounded by an incredible assortment of other smaller zedi, statues and temples...EDS: Not for syndication nor redistribution. Web slide show only. Please do not strip metadata for Web use.

Phrase: nor redistribution

- KEEP CALM AND REMEMBER You are not TUNA - Personalised iPhone 6 / 6S Case: Full Wrap White

Phrase: not TUNA

- Problems with the MMSE Mini-Mental State Exam – no psychometrics –Folstein et al., 1975 (antique) Considerable noise Several items do not provide adequate information Poor range for measuring change –Large standard error of measurement Poor power for assessing medication benefit Inadequate screening tool Better, shorter tests are available Now, copyright is being enforced (not free!!)

Phrase - no psychometrics

References

- [1] Jia Deng, Wei Dong, Richard Socher, Li-Jia Li, Kai Li, and Li Fei-Fei. Imagenet: A large-scale hierarchical image database.

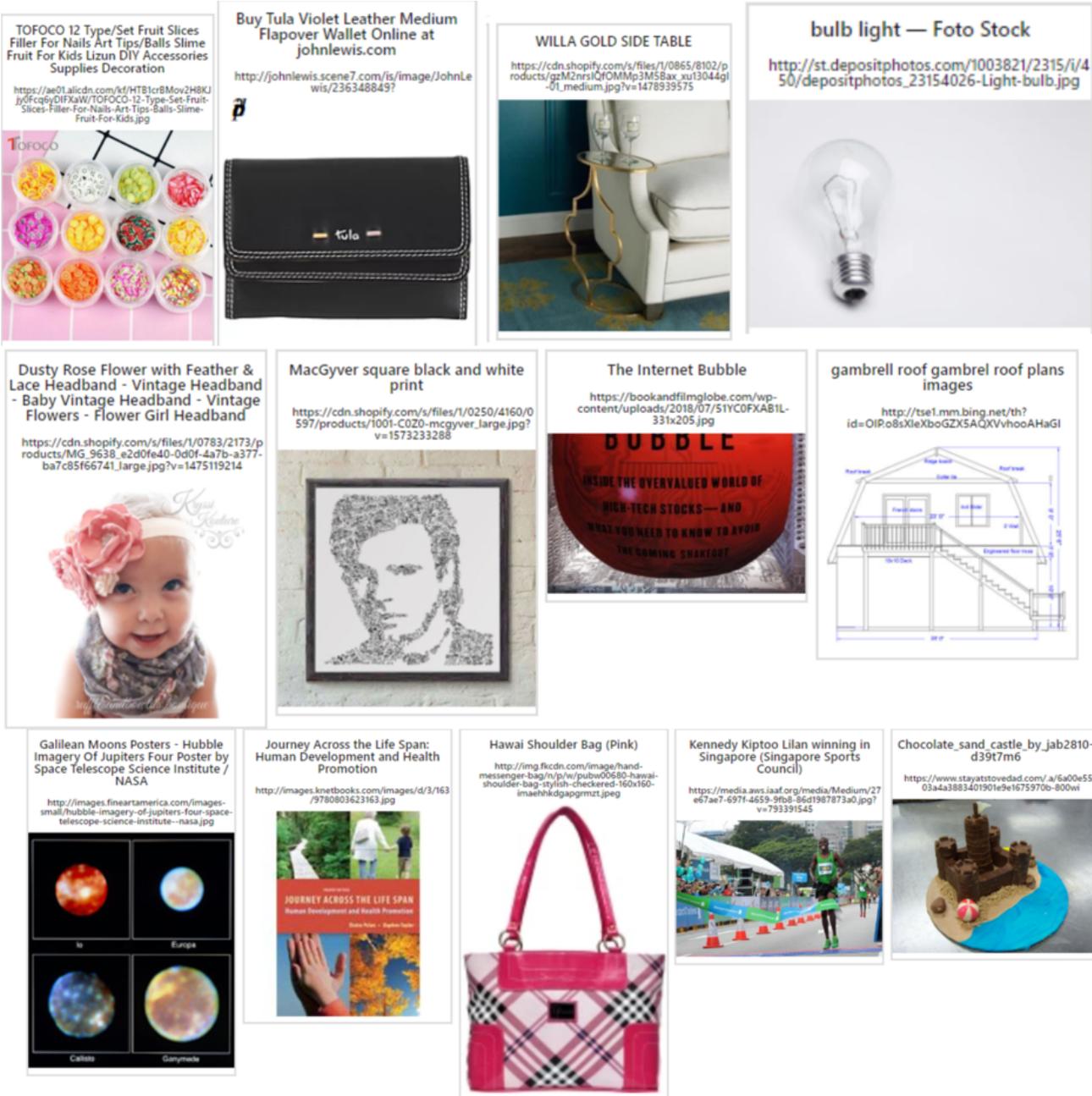


Figure 2. Image Text pairs from LAION400M dataset. The descriptions of the images mainly focus on the objects(classes) present in the image, and do not describe the absence of objects (classes).

In 2009 IEEE conference on computer vision and pattern recognition, pages 248–255. Ieee, 2009. 3

- [2] Christoph Schuhmann, Richard Vencu, Romain Beaumont, Robert Kaczmarczyk, Clayton Mullis, Aarush Katta, Theo Coombes, Jenia Jitsev, and Aran Komatsuzaki. Laion-400m: Open dataset of clip-filtered 400 million image-text pairs. *arXiv preprint arXiv:2111.02114*, 2021. 1