NTU DLCV (Autumn, 2022) HW3 Report

R10921069　沈郁鈞

Problem 1

1

Methods analysis (3%)

The main reason is that CLIP changes the task from **image classification** to **text-image pairing**.
Thus, CLIP has higher performance than other *image classification based* models on **unseen class** task.

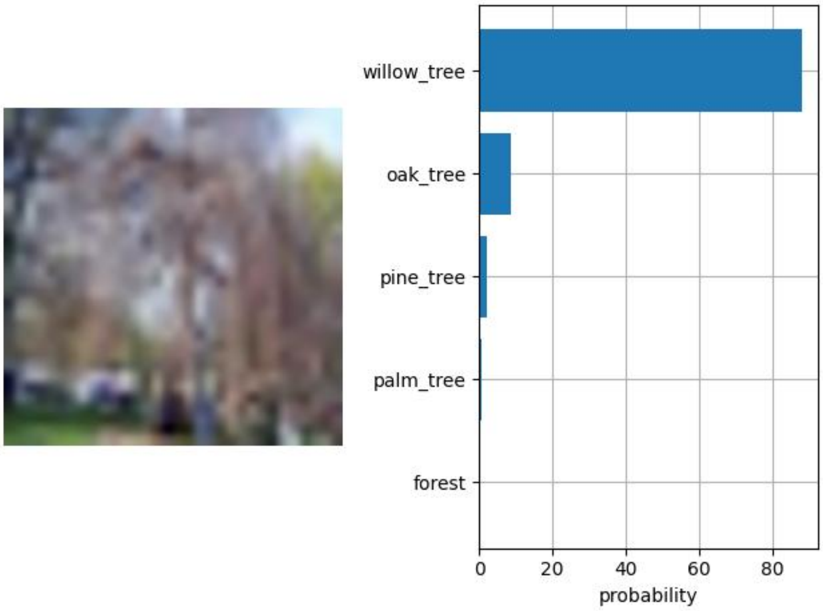
2

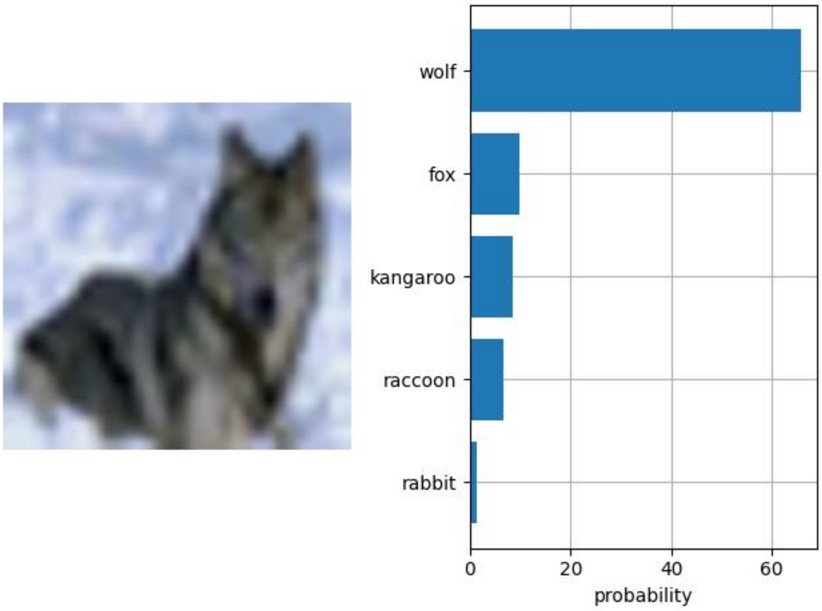
Prompt templates	Accuracy	Rank
This is a photo of {object}	0.6876	1
This is a {object} image.	0.6824	2
No {object}, no score.	0.5644	3

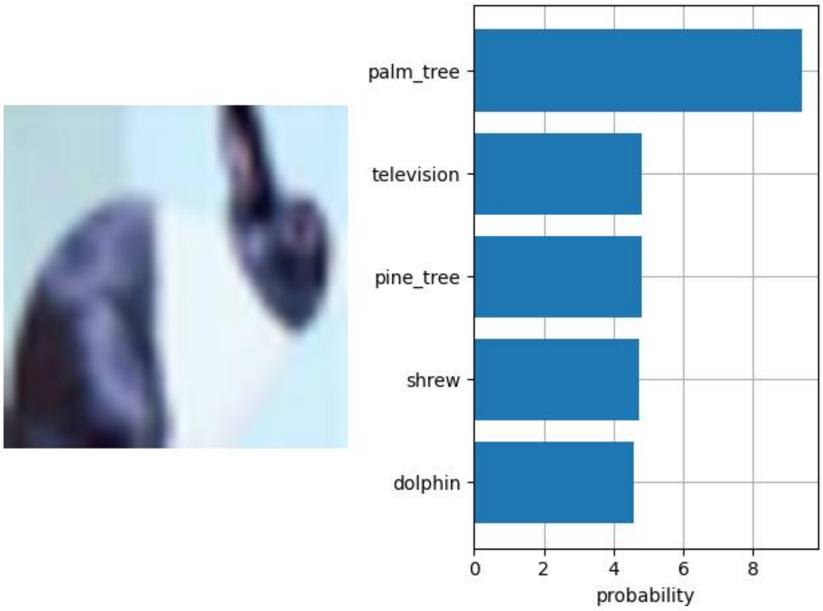
Discussion

I think the prompt template should be as simple and accurate as we could. Also, the template with positive sentence pattern would improve the performance.

3


willow_tree
oak_tree
pine_tree
palm_tree
forest
0
20
40
60
80
probability


wolf
fox
kangaroo
raccoon
rabbit
0
20
40
60
probability


palm_tree
television
pine_tree
shrew
dolphin
0
2
4
6
8
probability

Problem 2

1

Decoding Strategy: Greedy (k=1)
learning rate: 10^{-5}
epoch: 6

CIDEr	CLIPScore
0.959487	0.731331

2

Setting 1

Decoding Strategy: Greedy (k=1)
learning rate: $2 * 10^{-5}$
epoch: 6

CIDEr	CLIPScore
0.86034	0.73010

Setting 2

Decoding Strategy: beam search (k=3)
learning rate: $2 * 10^{-5}$
epoch: 6

CIDEr	CLIPScore
0.87933	0.71645

Setting 3

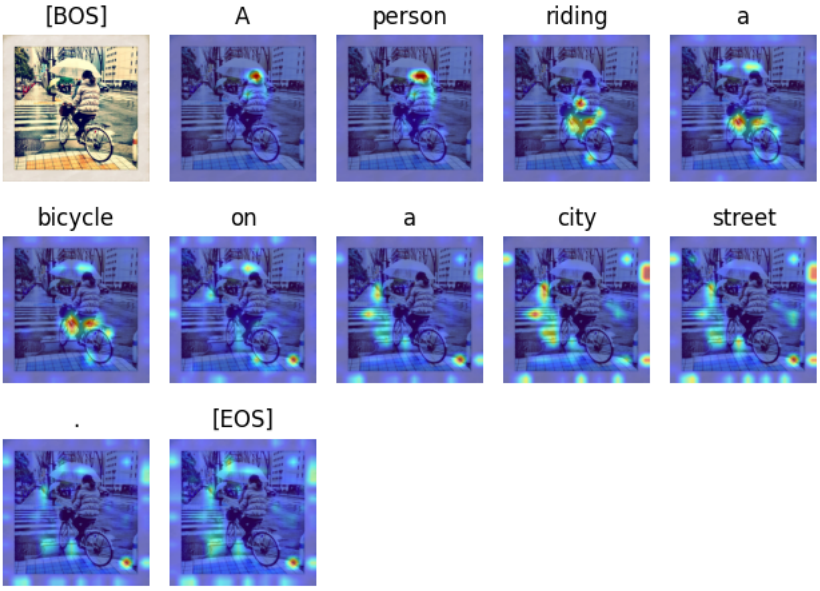
Decoding Strategy: Greedy (k=1)
learning rate: 10^{-5}
epoch: 3

CIDEr	CLIPScore
0.87437	0.72860

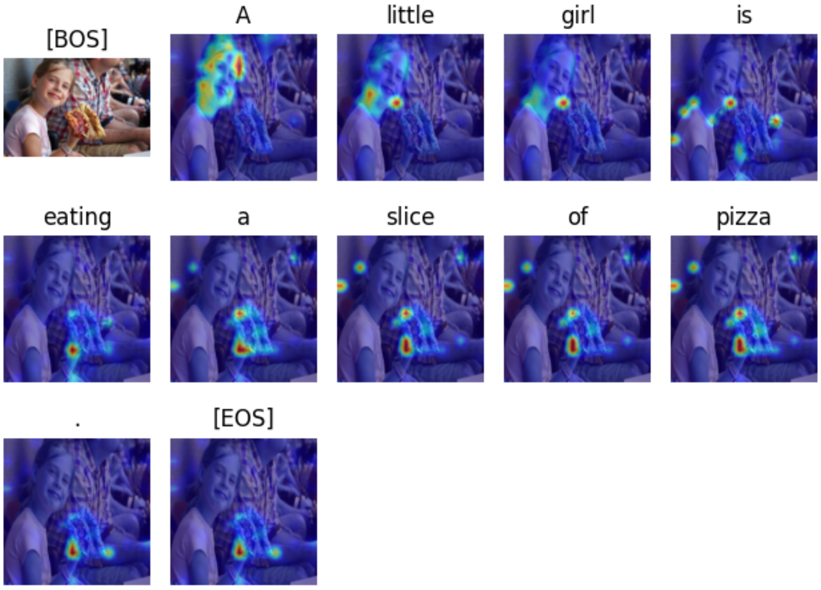
Problem 3

1

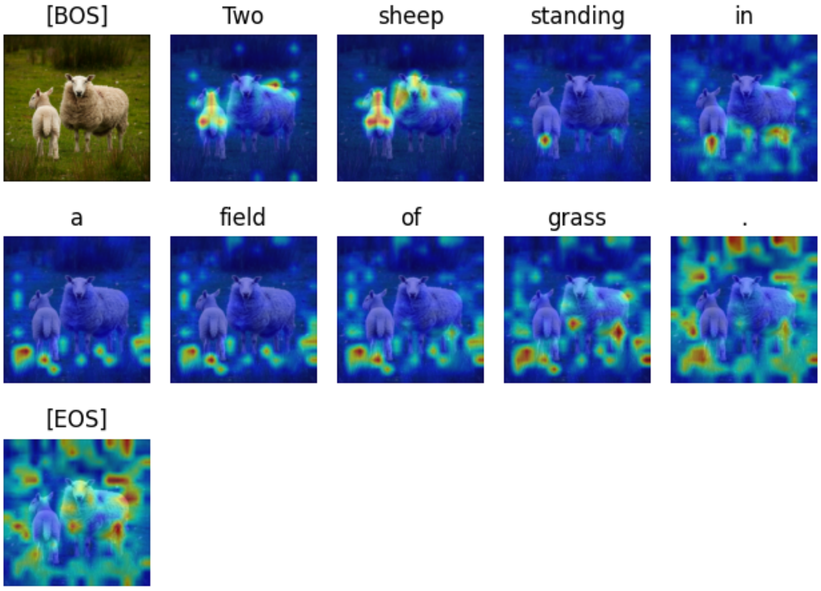
`BIKE.PNG`



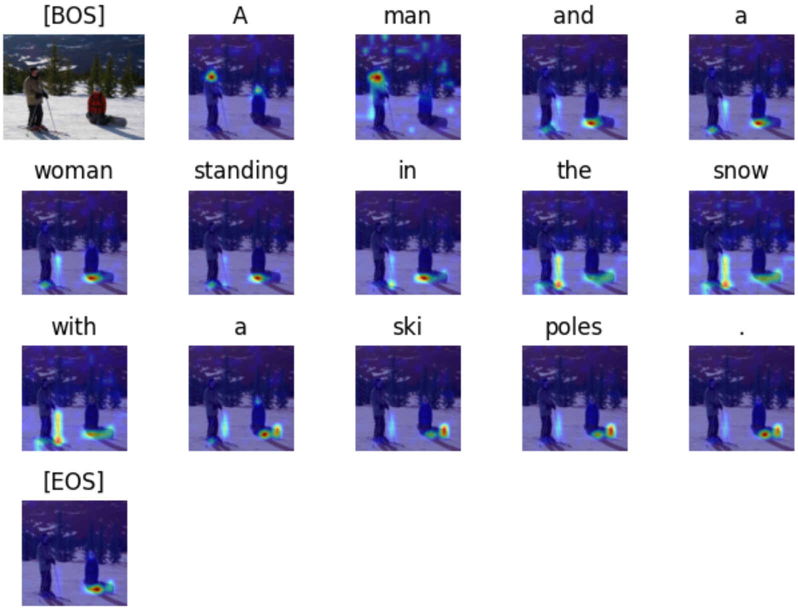
GIRL.PNG



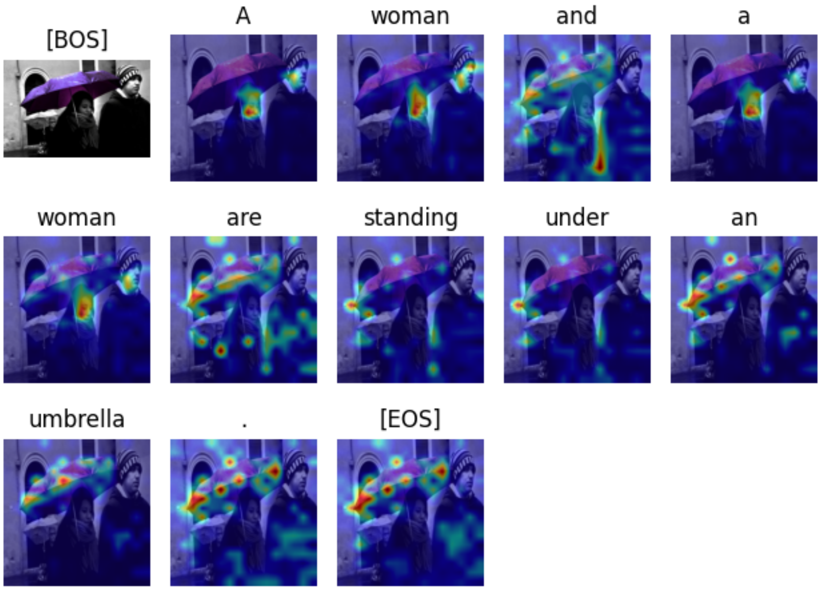
SHEEP.PNG



SKI.PNG



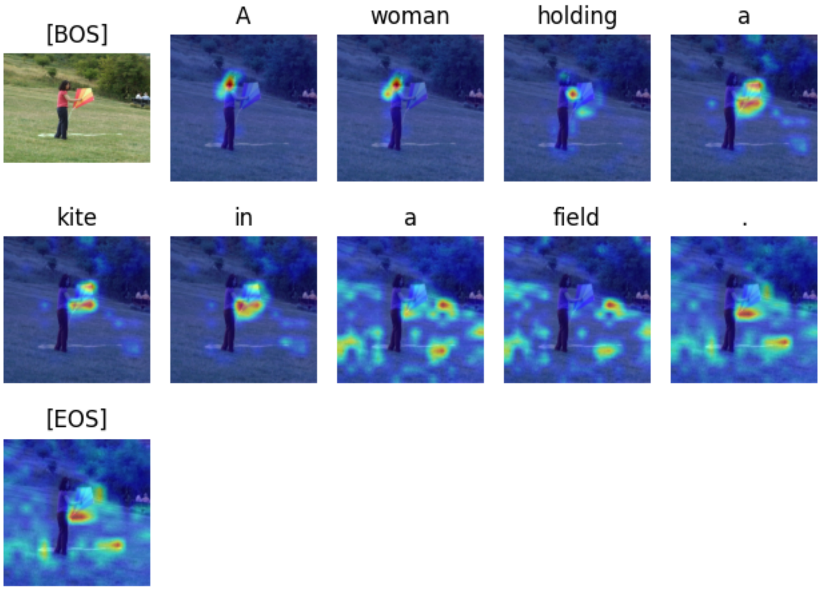
UNBRELLA.PNG



2

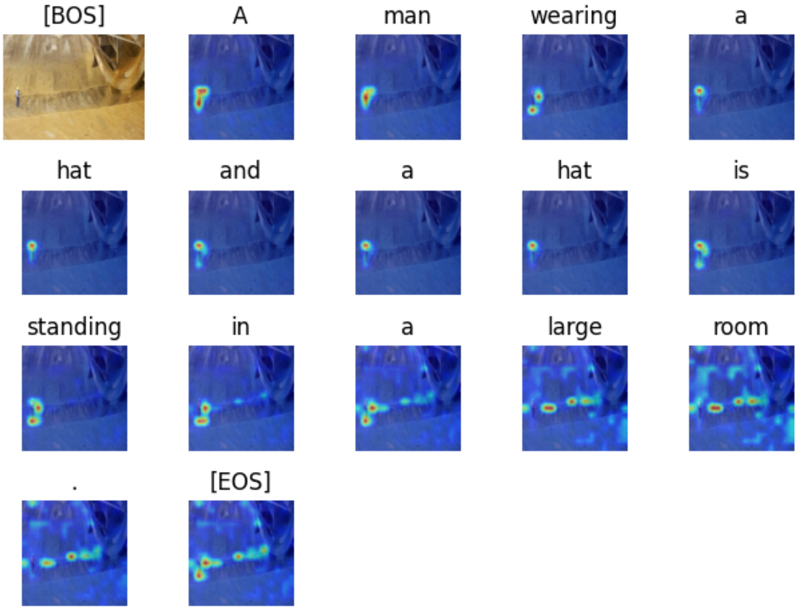
TOP-1

CLIPScore:0.977172



LAST-1

CLIPScore:0.427246



3

TOP-1 PREDICTION

- The predicted caption is reasonable for the picture.
    - (The kite is really on the woman's hands!!!)
- I found that **"kite"** is almost precisely marked on the position in the picture.

LAST-1 PREDICTION

- The predicted caption is partially correct and partially unreasonable.
    - **O** It's true that the man is wearing a helmet(hat) on his head.
    - **X** It's impossible for a hat to stand.
    - **?** I don't know the exact place where the man stands in.