Deep Learning Lab - Computer Vision

Ayisha Ryhana Dawood - 5653447

05 June 2024

1 Self-Supervised Learning

1.1 Nearest Neighbor Features with Negative Feature Space Distance as Score

This technique identifies the closest matching features between two sets by finding the nearest neighbors of features in one set (e.g., a feature map from image A) within another set. The score for each match is the negative of the distance calculated in a higher dimensional space, meaning closer features have higher scores.

```
err. mean (top: 10) : 60.47882803160503
err. median (top: 10) : 27.65884660215532
```

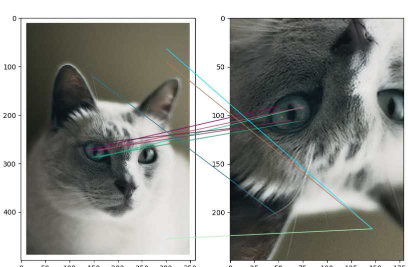

Figure 1: Image ID 78

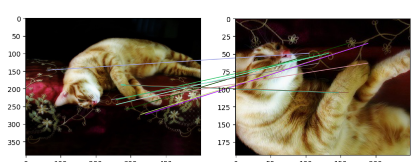

Figure 2: Image ID 133

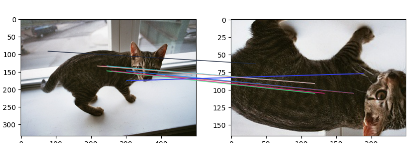

Figure 3: Image ID 288

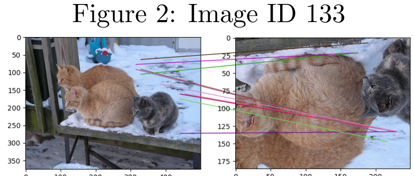

Figure 4: Image ID 415

Figure 5: Negative Feature Space Distance

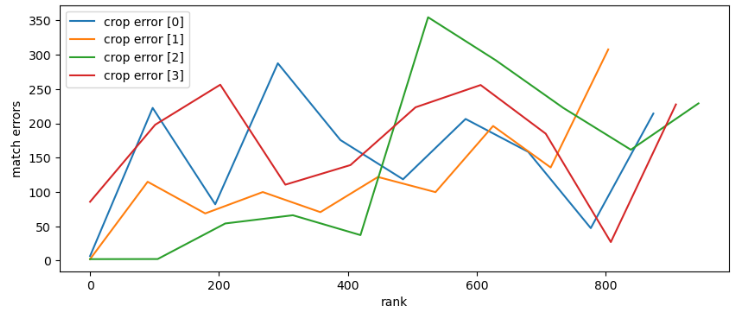

Figure 6: Crop error plot for Negative Feature Space Distance

1.2 Nearest Neighbor Features with Negative Cycle Distance as Score

This technique identifies the closest matching features between two sets by finding the nearest neighbors of features in one set (e.g., a feature map from image A) within another set and then verifying the match by checking the cycle distance, calculated by mapping the features back to the original set and measuring the distance. The score for each match is the negative of the distance calculated in a higher dimensional space, meaning closer features have higher scores.

```
err. mean (top: 10) : 57.07507840410403
err. median (top: 10) : 18.45058936262634
```

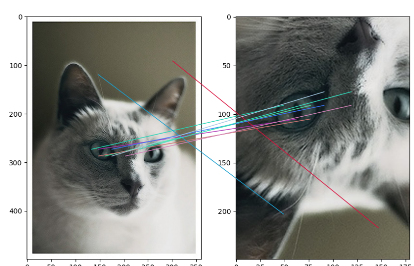

Figure 7: Image ID 78

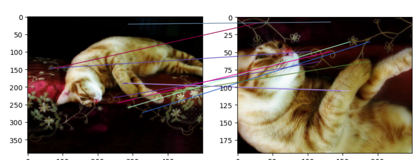

Figure 8: Image ID 133

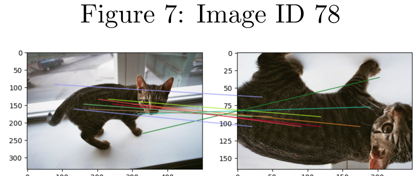

Figure 9: Image ID 288

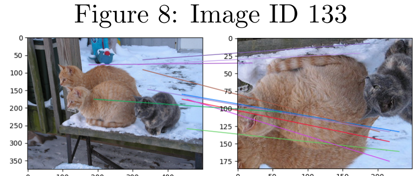

Figure 10: Image ID 415

Figure 11: Negative Cycle Distance

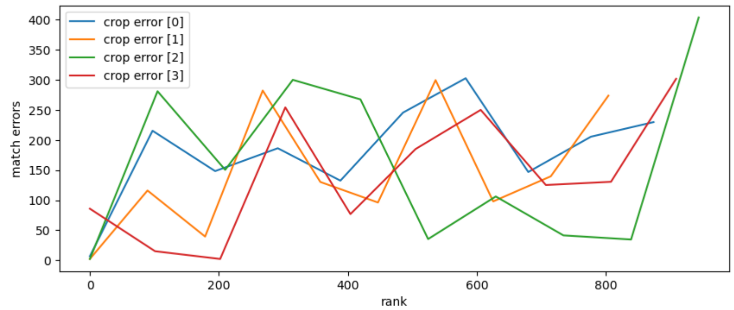

Figure 12: Crop error plot for Negative Cycle Distance

How does the cycle distance score performs against the feature distance score?

For the matching of image 288 to image 78, the stats of scores are

- Feature Space Distance
  - Number of -inf values: 651
  - Number of non -inf values: 154
  - Average of non -inf: -49.94686369462447
- Cyclic Space Distance
  - Number of -inf values: 734
  - Number of non -inf values: 71
  - Average of non -inf: -40.82621912889078

The cyclic distance method has more -inf values than the feature distance score because two-way matching is required.

2 PCA

What is the advantage of using PCA?

- Dimensionality reduction - reduces computational complexity
- Feature extraction - highlight discriminative features, reduces the impact of noise, improve generalization

What is the advantage of using the negative cycle distance as score?

Cyclic distance measures the consistency of feature matches by mapping features from one set to another and the back to their original set. This helps to prevent false positives in matching because the match has to exist in both directions.

3 Image Captioning

Why do the results improve with lower temperature?

Lower temperature causes the unit values to scale up, thus increasing the differences between them. On applying the softmax function, the increased differences result in a probability distribution that is more defined with extremes. This makes the outputs more coherant and aligned to reference phrases.

3.1 Prompt engineering

The fixed parameters for this search are `temperature=0.7` and `topk=50`.

Prompt	**BLEU score**
A picture of	11.81
A picture showing	9.5
An image of	**13.56**
The picture has	3.19
The picture shows	9.77
(empty prompt)	9.64

Table 1: BLEU Scores for Different Prompts

3.2 Hyperparameter Search

The best performance from prompt engineering was achieved with the phrase "An image of ".

Top-K	**Temperature**	**BLEU score**
10	0.1	**25.43**
10	0.4	18.44
10	0.7	13.89
10	1.0	9.56
25	0.1	24.79
25	0.4	18.75
25	0.7	13.48
25	1.0	7.40
50	0.1	24.22
50	0.4	18.78
50	0.7	13.56
50	1.0	7.98

Table 2: BLEU Scores for Different Top-K Values and Temperatures

4 Image-Text Retrieval

4.1 Finetune

What score do you get and how can you explain the difference to the score when training from scratch?

The R@1 score improves from 40.44% to 59.35% on 3 epochs using finetuning. The weights from the pretrained blip model instead of random weight initialization causes the improvement in performance.

4.2 Hyperparameter Search

Performing grid search by training with `finetuning=False` and different hyperparameters for 5 epochs.

Learning rate	**Temperature**	**Weight decay**	**R@1**
1e-4	0.05	1e-3	26.01
1e-4	0.05	1e-5	27.47
1e-4	0.1	1e-3	29.61
1e-4	0.1	1e-5	28.71
1e-3	0.05	1e-3	37.75
1e-3	0.05	1e-5	38.37
1e-3	0.1	1e-3	43.06
1e-3	0.1	1e-5	43.48

Table 3: R@1 Scores for Different Learning Rates, Temperatures, and Weight Decays