# Exercise 3: Computer Vision

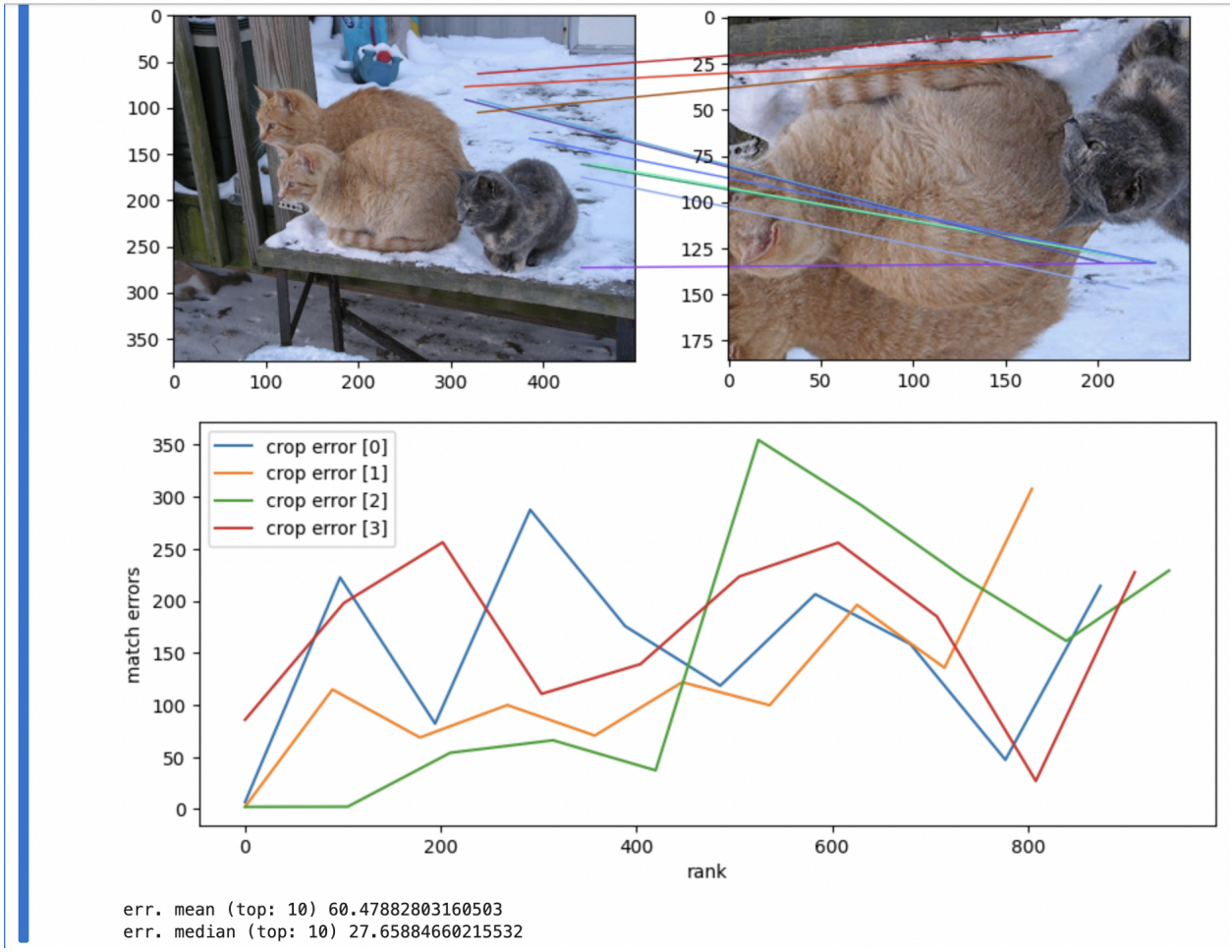
Student: Erlis Lushtaku
Matriculation number: 5772476

## Introduction

This report presents the results of experiments conducted as part of the Deep Learning Lab Course on semantic segmentation, image captioning and image-text retrieval.

## 1 Self-Supervised Learning

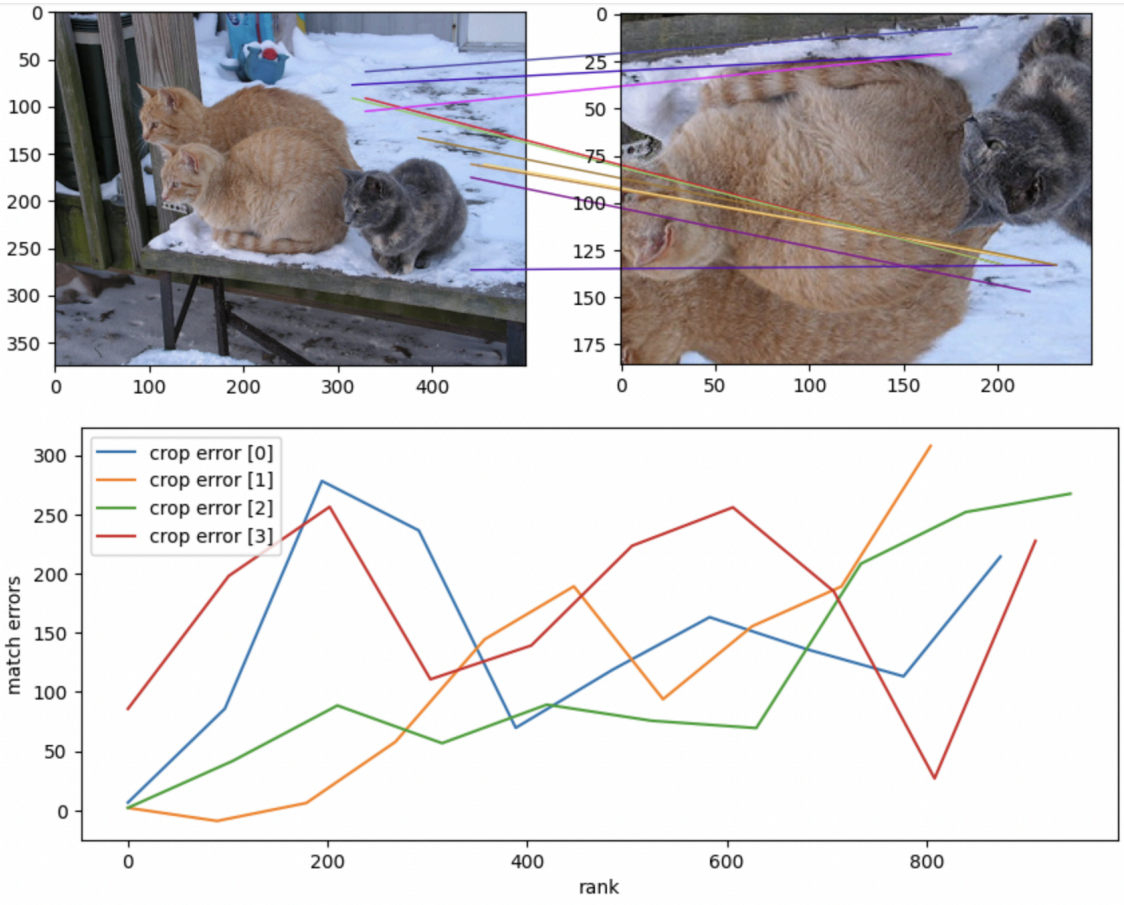
### 1.2 Visualize Nearest Neighbors and Calculate Crop Error with Feature Distance Ranking




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

Question: Describe the plot with your own words.
Answer: The x-axis represents the ranking of the feature matches based on their similarity scores. Each line labeled "crop error [0-3]" corresponds to a different image pair, where each pair consists of a reference image and its cropped and flipped version.
The crop error represents the Euclidean distance between the matched points in the original and modified images. A lower crop error indicates that the features in the modified image resemble more those in the original image which means that they match better.

## 1.4 Visualize Nearest Neighbors and Calculate Crop Error with Cycle Distance Ranking



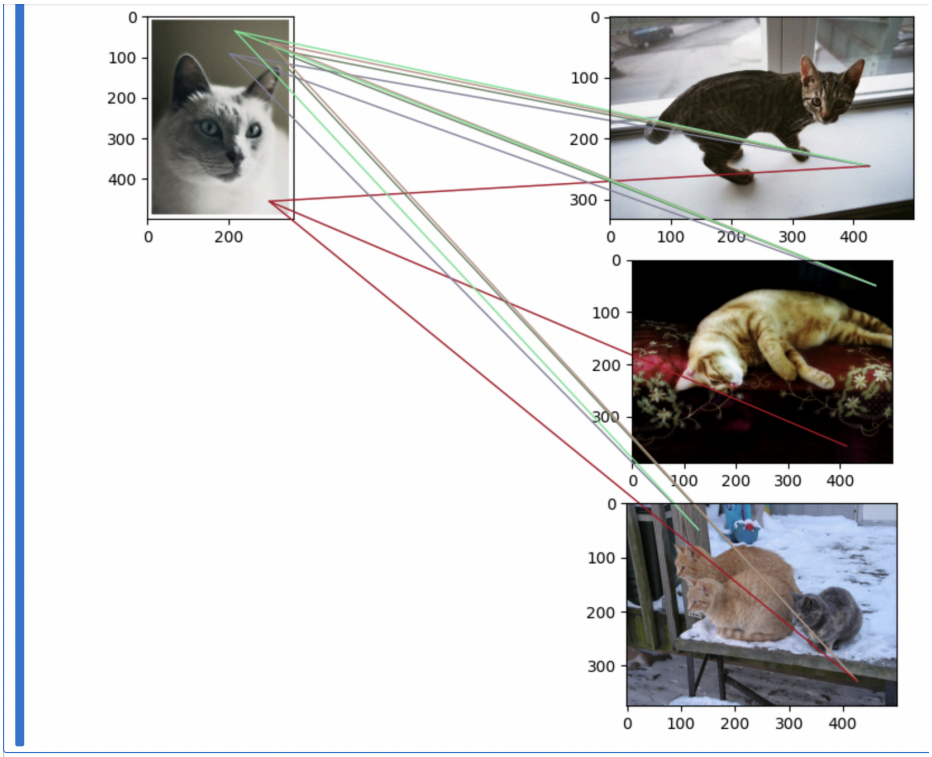

```
err. mean (top: 10) 51.53408997162253
err. median (top: 10) 24.422324041771013
```

err. mean (top: 10) 51.53408997162253
err. median (top: 10) 24.422324041771013

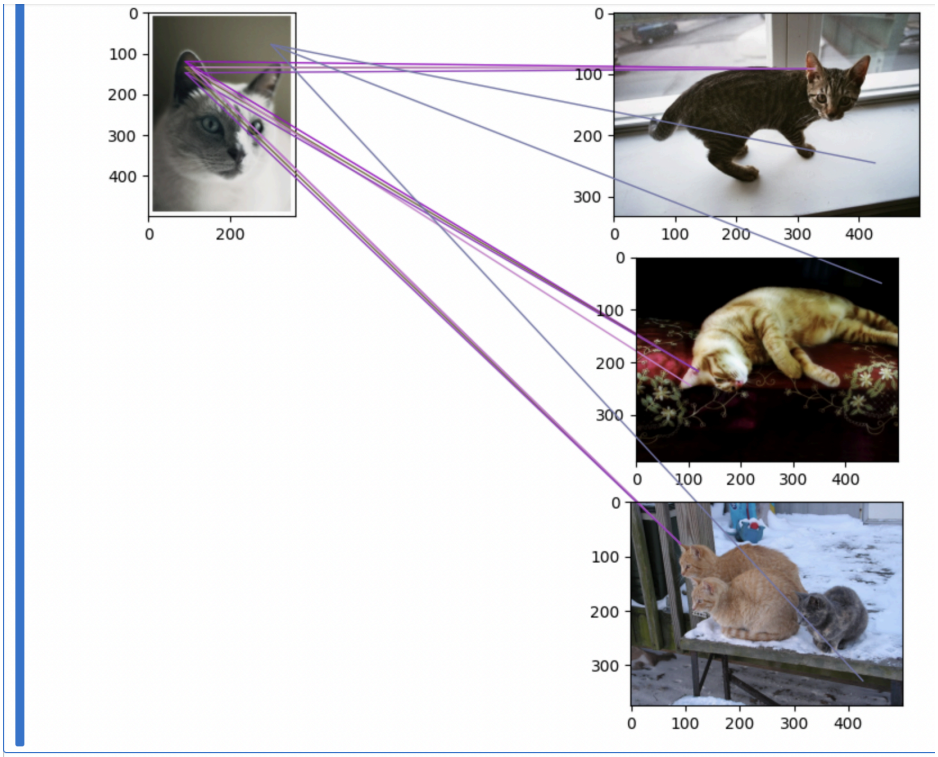
Question: How does the cycle distance score performs against the feature distance score? (2 Points)
Answer: The cycle distance score is slightly better than the feature distance score.

## 1.6 Visualize Nearest Neighbors Matches for 1-to-Many Frames

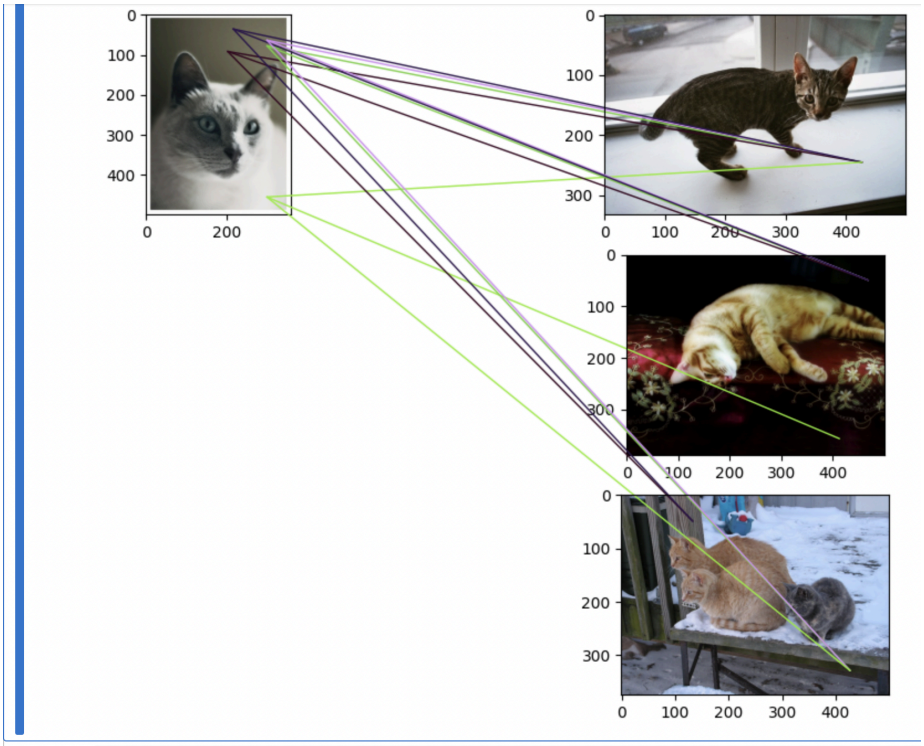
A) get_matches_feats_A_to_B_nn without PCA.



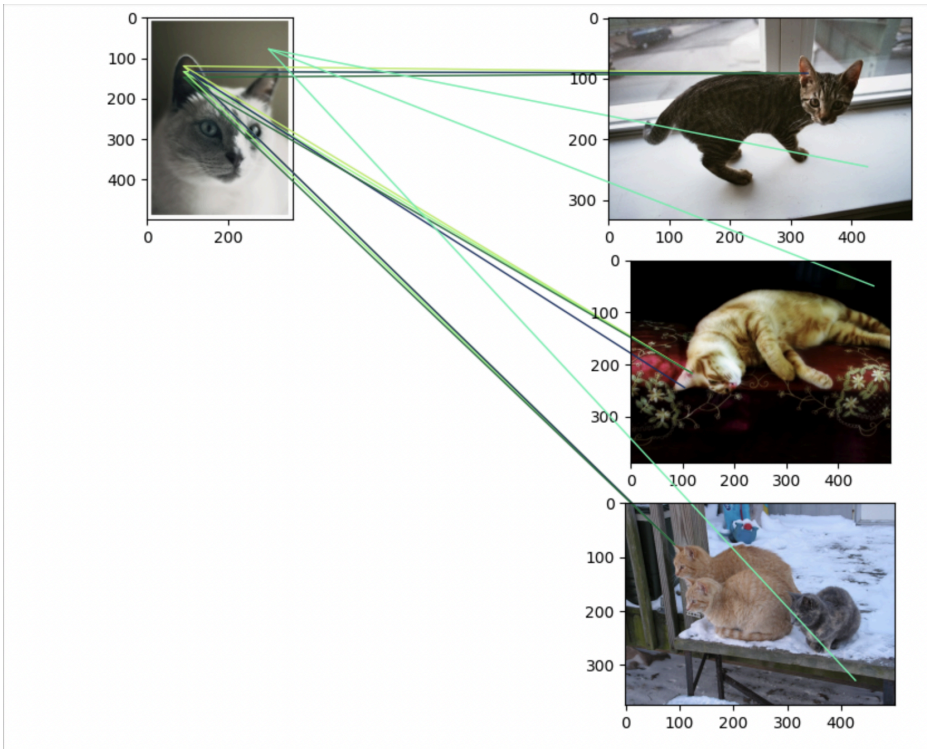
B) get_matches_feats_A_to_B_nn with PCA (10 components).



C) get_matches_feats_A_to_B_nn_with_score_cycle_dist without PCA



D) get_matches_feats_A_to_B_nn_with_score_cycle_dist with PCA (10 components).



Question: What is the advantage of using PCA?
Answer: Principal Component Analysis (PCA) is a dimensionality reduction technique that transforms the data into a lower-dimensional space while retaining most of the variability in the data. The advantages are:

- Noise Reduction: Focuses on the most significant features, reducing noise.
- Computational Efficiency: Fewer dimensions lead to faster computations.
- Improved Generalization: Better performance across different images.
- Reduced Overfitting: Limits the feature space to essential components.

Question: What is the advantage of using the negative cycle distance as score? (2 Points)
Answer:

- Consistency Check: Ensures matches are consistent in both directions.
- Robustness to Outliers: Identifies and discards incorrect matches more effectively.
- Improved Feature Discrimination: Provides better precision in matching features.

# 2 Image-Text

## 2.1 Image Captioning

### 2.1.1 Complete Caption Generation with Greedy Search

Final BLEU@4: 27.95%

### 2.1.2 Complete Caption Generation with Sampling

Temperature=1.0 - Final BLEU@4: 6.56%
Temperature=0.7 - Final BLEU@4: 12.90%

Question: Why do the results improve with lower temperature?
Answer: Lowering the temperature makes the probability distribution sharper. This means the model is more likely to pick high-probability tokens, reducing randomness. While higher temperature can introduce diversity, it can reduce precision, leading to captions that might not be as relevant, thereby reducing the score.

### 2.1.3 Prompt Engineering

| **Prompts** | **BLEU Score** |
|---|---|
| a picture of | 12.90% |
| an image showing | 9.55% |
| a depiction of | 7.62% |
| a scene showing | 9.38% |

### 2.1.4 Student Hyperparameter Search

| **Params** | **BLEU Score** |
|---|---|
| use_topk_sampling=True, topk=50, temperature=0.7, prompt="a picture of " | 12.90% |
| use_topk_sampling=False, prompt="this is a photo of " | 23.76% |
| use_topk_sampling=True, topk=50, temperature=0.5, prompt="a picture of " | 15.68% |
| use_topk_sampling=True, topk=100, temperature=0.7, prompt="a picture of " | 12.74% |

## 2.2 Image-Text Retrieval

### 2.2.1 Complete Forward Pass and Evaluate

image-to-text R@1: 53.34%
Validation results: {'i_r1': 0.533471359558316, 'i_r5': 0.8067632850241546, 'i_r10': 0.906832298136646, 'i_medr': 1.0, 'i_meanr': 4.0690131124913735, 't_r1': 0.5355417529330573, 't_r5': 0.8184955141476881, 't_r10': 0.8923395445134575, 't_medr': 1.0, 't_meanr': 4.680469289164941}

### 2.2.2 Complete Loss and Train from Scratch

image-to-text R@1: 44.92%
Validation results: {'i_r1': 0.4492753623188406, 'i_r5': 0.7556935817805382, 'i_r10': 0.8702553485162181, 'i_medr': 2.0, 'i_meanr': 5.715665976535542, 't_r1': 0.42581090407177363, 't_r5': 0.7342995169082126, 't_r10': 0.8454106280193237, 't_medr': 2.0, 't_meanr': 6.358178053830228}