Master in Computer Vision Barcelona

M5 Project: Cross-modal Retrieval

Week 4

Image Retrieval

Rubèn Pérez Tito
rperez@cvc.uab.cat

Ernest Valveny
ernest@cvc.uab.cat

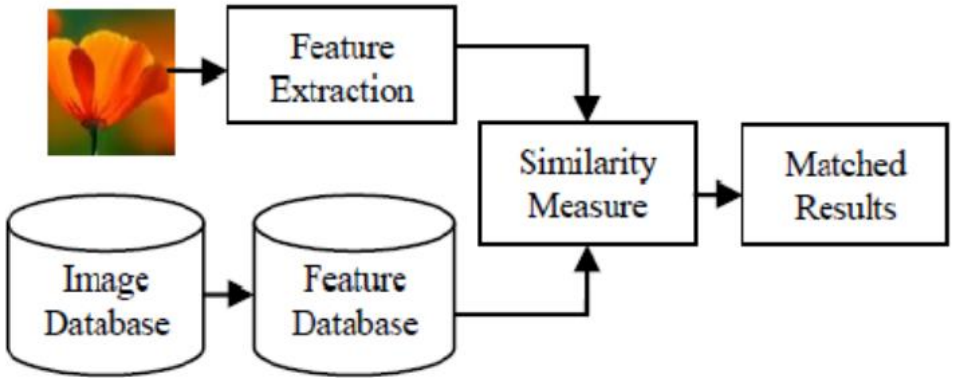
M5 Project Stages and Schedule

Week 1 March 6-12	**P1: Introduction to Pytorch - Image Classification**
Week 2 March 13-19	**P2 & P3: Object Detection, Recognition and Segmentation**
Week 3 Marh 20 - 26	
Week 4 March 27 – April 2	**P4: Image Retrieval**
EASTER	
Week 5 April 17 - 23	**P5: Cross-modal Retrieval** **Deliverable: Report on object Detection and Segmentation, first version**
Week 6 April 24	**Deliverable: Presentation** **Deliverable: Report on object Detection and Segmentation, final version**

M5 – Image retrieval

Application approach

- Extract features from database images (train set).
- Extract features of the query image (val/test set).
- Retrieve the most similar images from the database.

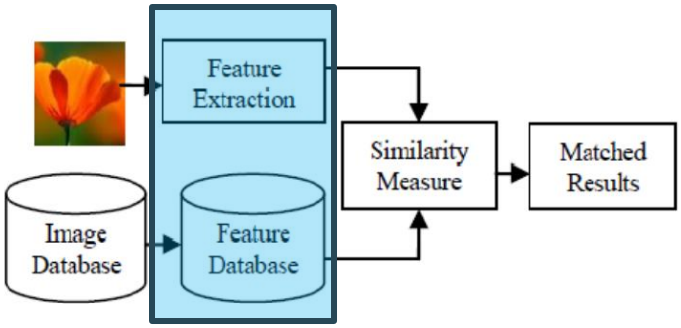


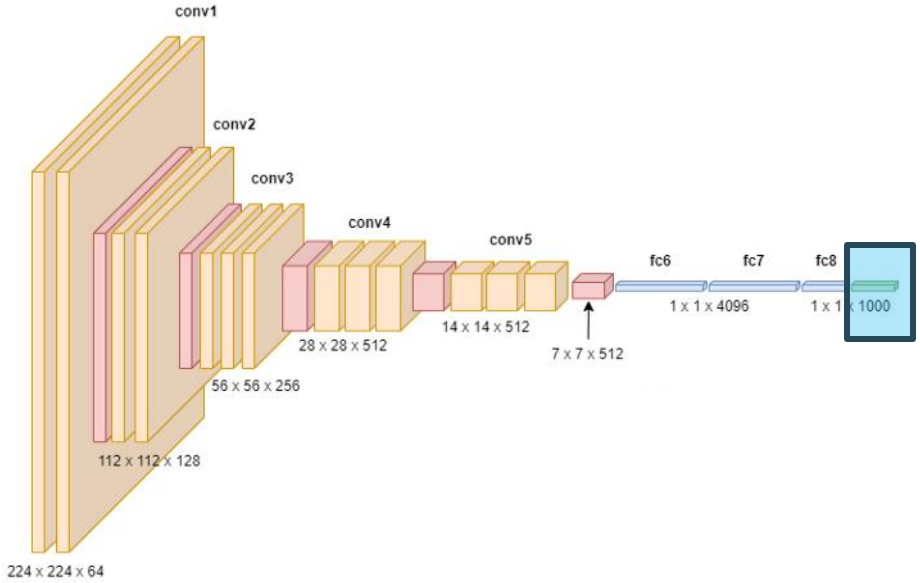
Notice that image retrieval is not a training methodology, but an application!

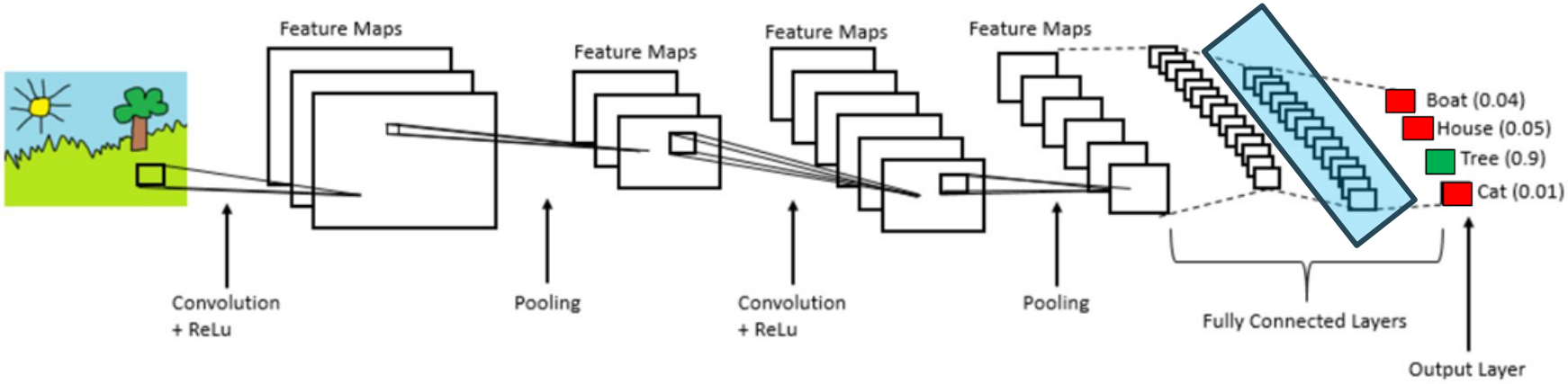
M5 – Image retrieval

Training strategies for image retrieval

- Classification





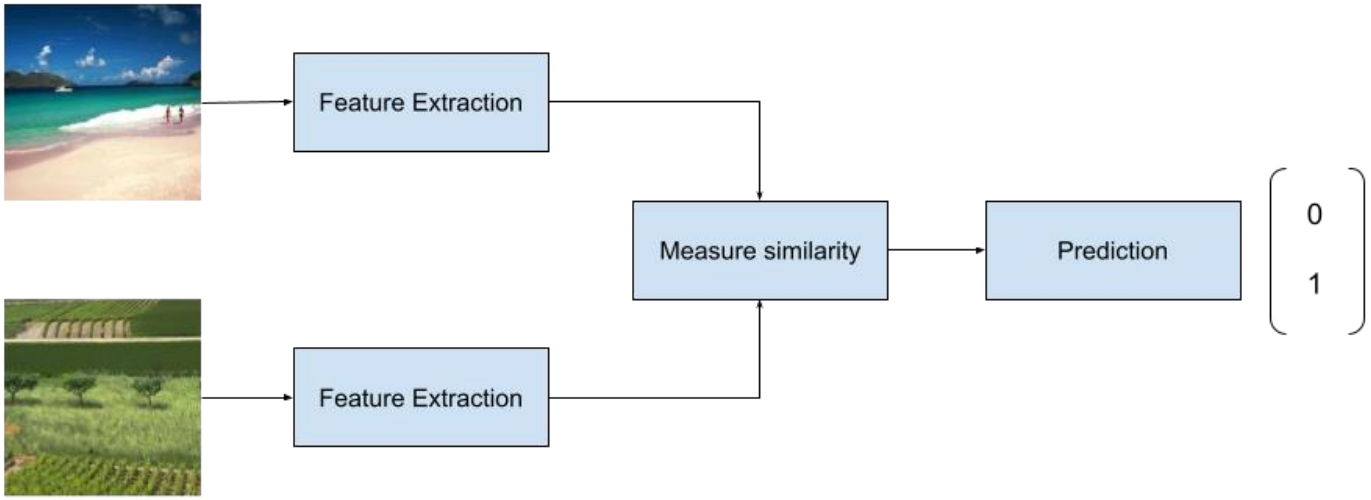


By training the to classify. It will implicitly learn an image representation that is representative to perform retrieval.

M5 – Image retrieval

Training strategies for image retrieval

- Classification
- Metric learning:
  - Siamese networks

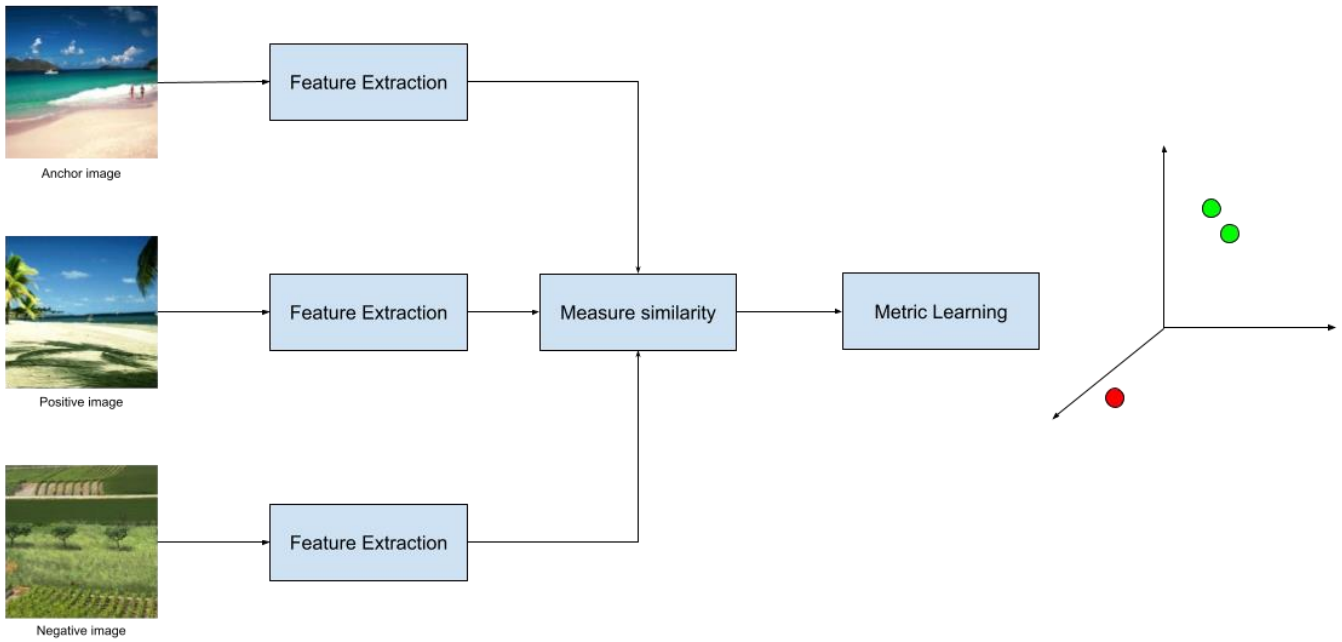


By performing metric learning, we explicitly learn a representation that facilitates the retrieval of the images.

M5 – Image retrieval

Training strategies for image retrieval

- Classification
- Metric learning:
  - Siamese networks
  - Triplet networks

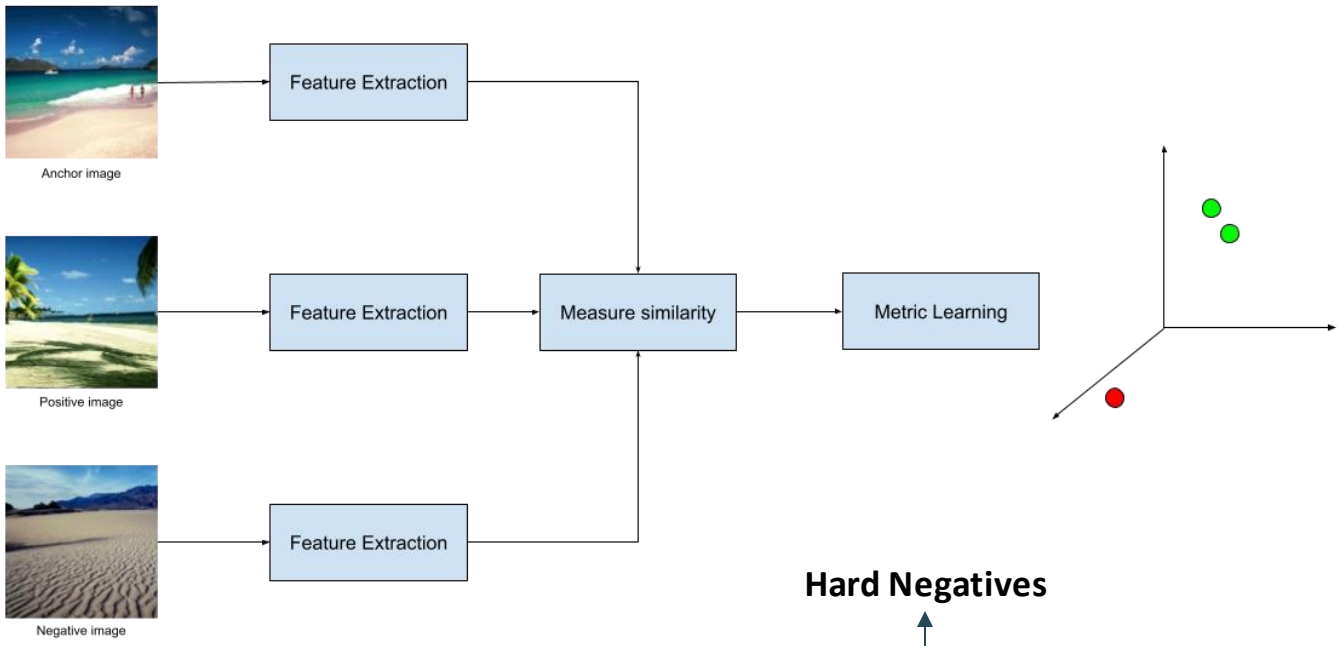


By performing metric learning, we explicitly learn a representation that facilitates the retrieval of the images.

M5 – Image retrieval

Training strategies for image retrieval

- Classification
- Metric learning:
  - Siamese networks
  - Triplet networks



By performing metric learning, we explicitly learn a representation that **facilitates** the retrieval of the images.

M5 – Image retrieval

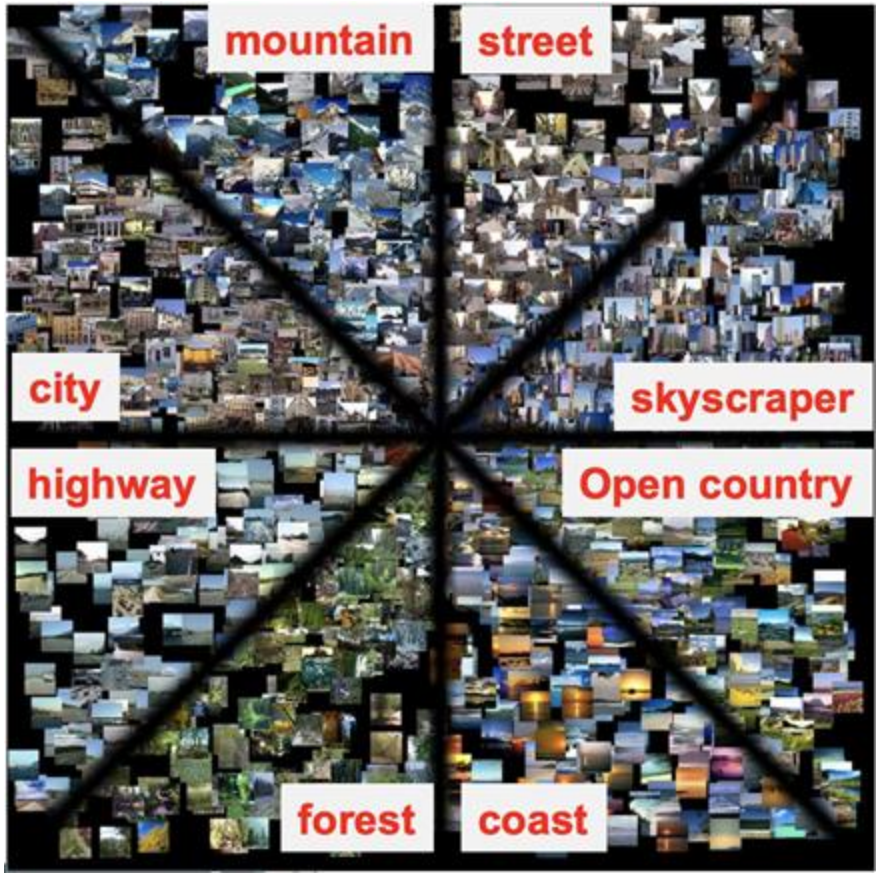
Training strategies for image retrieval

- Classification
- Metric learning:
  - Siamese networks
  - Triplet networks
  - Quadruplet Networks
  - Etc.

By performing metric learning. We explicitly learn a representation that facilitates the retrieval of the images.

M5 – Image retrieval

Dataset: MIT Split



M5 – Image retrieval

Training strategies for image retrieval

- Classification
- Metric learning:
  - Siamese networks
  - Triplet networks

Note: When you will read that models share parameters, you can use the same model.

```
1. img1_emb = model(img1)
2. img2_emb = model(img2)
3. loss = criterion(img1_emb, img2_emb)
```

M5 – Image retrieval

Retrieval process

- Extract features from database images (train set).
- Extract features of the query image (val/test set).
- Retrieve the most similar images from the database.
  - NN, KNN…
  - Facebook AI Similarity Search (FAIS), getting started documentation.

M5 – Image retrieval

Retrieval process

- Extract features from database images (train set) → use torch.no_grad()
- Extract features of the query image (val/test set) → use torch.no_grad()
- Retrieve the most similar images from the database.
  - NN, KNN…
  - Facebook AI Similarity Search (FAIS), getting started documentation.

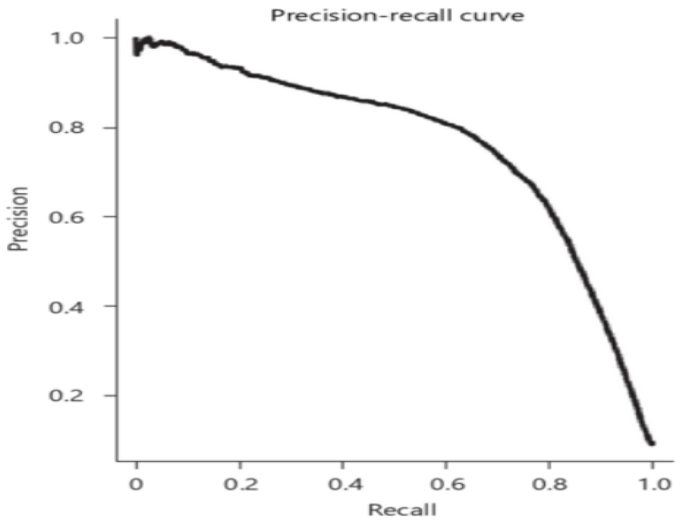
M5 – Image retrieval

Retrieval process

- Extract features from database images (train set).
- Extract features of the query image (val/test set).
- Retrieve the most similar images from the database.
  - NN, KNN…
  - Facebook AI Similarity Search (FAIS), getting started documentation.

Evaluation / Metrics

- Mean Average Precision (MAP)
- Precision@K
- Recall@K
- Difference between object detection
- and information retrieval metrics link.



M5 – P4 Tasks

Week 4: Image Retrieval

Tasks

a. Image retrieval with pre-trained image classification model.
b. Train the model on metric learning (Siamese network).
c. Train the model on metric learning (Triplet network).
d. Visualize the learned image representation of each of the previous tasks a-c
e. Image Retrieval on COCO with Faster R-CNN or Mask R-CNN
f. Finish the paper: Results and Conclusions.

Deliverable (for next week)

- **Github** repository with readme.md (code explanation & instructions)
- Presentation with all items listed in the tasks under the **Project presentation** title.
- **One summary slide** at the end of your presentation.
- **Report** on overlaf about object detection and segmentation.

M5 – P4 Tasks

Task (a): **Image retrieval with pre-trained image classification model.**

- Use P1 or standard Image Classification method (ResNet) pre-trained for Image Classification on the MIT_Split dataset.
  - You might need to remove the last linear layer where you project the hidden size into the output (num_classes) size.

- Show (and analyze) precision-recall curve.
- Show qualitative results in your presentation.
- Show quantitative results in your presentation.
  - At least MAP, Prec@1, Prec@5
  - For MAP use the average_precision_score() function from the Sklearn library
    - Sklearn: Metrics, Basic models (NN, KNN, K-Means, SVMs)…
    - You will have to turn your integer targets $[7, 3, 1, 3, \ldots]_{bs}$ to binary $[0, 1, 0, 1, \ldots]_{database\_size}$

- You can choose the retrieval method you prefer (NN, KNN, FAIS…)

M5 – P4 Tasks

Task (b): **Train the model on metric learning (Siamese network)**

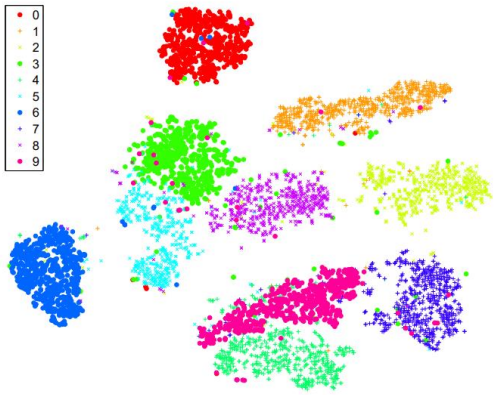
- Include precision-recall curve, quantitative and qualitative results in your presentation.

Task (c): **Train the model on metric learning (Triplet network)**

- Include precision-recall curve, quantitative and qualitative results in your presentation.

Task (d**): Visualize the learned image representation of each of the previous tasks a-c**

- You can use PCA, TSNE, UMAP or another you choose.
  - TSNE paper and implementation in sklearn.



(a) Visualization by t-SNE.

M5 – P4 Tasks

Task (e): **Image Retrieval on COCO with Faster R-CNN or Mask R-CNN**

- Perform image retrieval on subset of COCO with triplet networks.
- Dataset: COCO 2014
  - /home/mcv/datasets/COCO/
    - train2014
    - val2014
  - Annotations
    - **Train** (metric learning) ← Train set (82K images: 100 %)
    - **Database** (image retrieval DB) ← Train set (1.9K images: 2.5 %)
    - **Val** (queries) ← Val set (1.1K images: 2.9 %)
    - **Test** (queries) ← Val set (1.9K images: 4.8 %)
    - Format:
      - Obj_M: [$ImageId_0$, $ImageId_1$, …. $ImageId_N$]

M5 – P4 Tasks

Task (e): **Image Retrieval on COCO with Faster R-CNN or Mask R-CNN**

- Evaluating correct / wrongly retrieved images:
  - The retrieved image contains at least one object of the queried image.
    - Selection
  - The retrieved image contains same objects as the queried image.
    - Aggregation
  - The retrieved image contains similar objects with similar quantities as the queried image.
    - Weighted aggregation

M5 – P4 Tasks

Task (e): **Image Retrieval on COCO with Faster R-CNN or Mask R-CNN**

- Evaluating correct / wrongly retrieved images:
  - The retrieved image contains at least one object of the queried image.
    - Selection
  - The retrieved image contains same objects as the queried image.
    - Aggregation
  - The retrieved image contains similar objects with similar quantities as the queried image.
    - Weighted aggregation

M5 – P4 Tasks

Task (f): **Finish the paper.**

- Abstract
- Introduction (½ page)
- Related Work (1 page)
- Methodology (1 page with diagram)
  - Faster R-CNN & Mask R-CNN
- Experiments
  - Datasets
  - Metrics
  - **You can include a section on implementation details (detectron, backbones, hyperparameters, training, …)**
- **Results**
  - **Do not include all experiments and results. Only a summary of most relevant ones**
- **Conclusion**

Max: 6 pages w/o references

M5 – P4 Tasks

Interesting features to analyze

1. How different metric learning setups affect the results?
   - Different losses, different distances (Euclidean, Mahalanobis), different weights or margins.
   - Use of hard negative and different hard-negative mining strategies.

2. How different retrieval methods (NN, KNN, FAIS) affect the results for the same learned image representations?

3. How different visualization methods plot the same learned image representations?

M5 – P4 Tasks

General information requirements for the presentation

- Describe your method.
  - Was it necessary to perform any change? (remove the last fully connected layer).
- Describe the training strategies (loss function).
  - Did you use any hard negative strategy? Which one?
- Describe the retrieval method.
- Describe the visualization method.

Extra material

- Siamese, Triplet examples (AdamBielski)
- Pytorch-metric-learning library (Kevin Musgrave)
  - Oficial Github repository
  - CIFAR 10 examples

M5 – P4: Image Retrieval

Due date

17th of April, Monday, before 10:00 AM

Include **one** summary slide at the end of your presentation with main results and conclusions

- One member of the group members will have to present this slide in **1 minute** during the follow-up session next week.