



Master in
Computer Vision
Barcelona

M5 Project: Cross-modal Retrieval

Week 2

Introduction to Object detection and
Instance Segmentation with Detectron2

Rubèn Pérez Tito
rperez@cvc.uab.cat

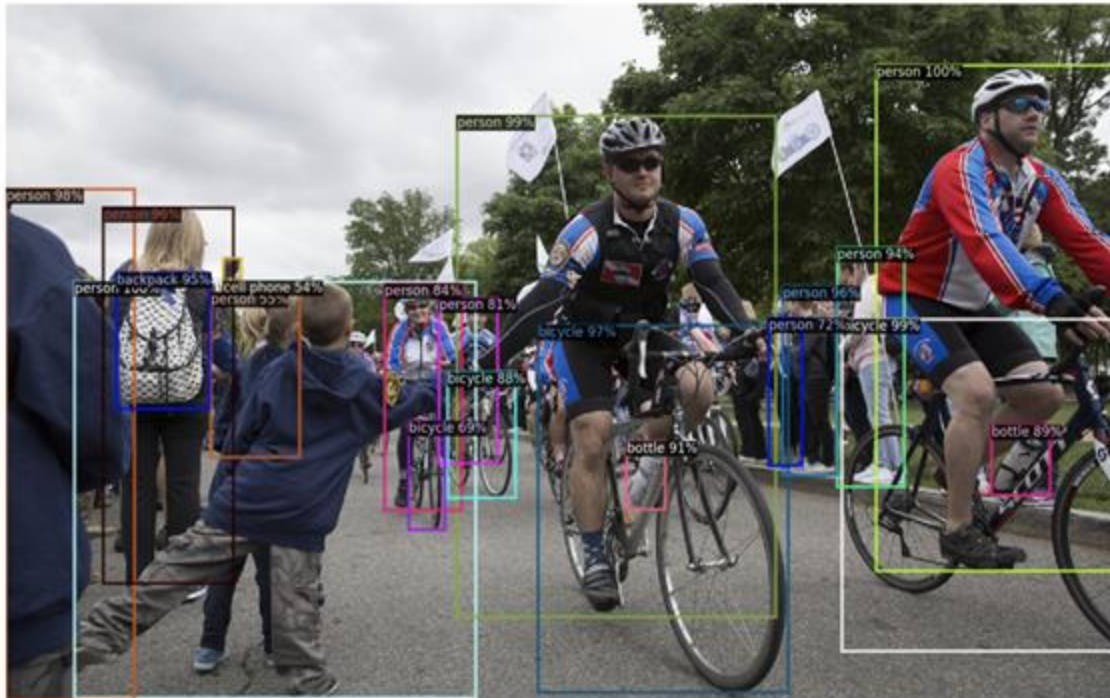
Ernest Valveny
ernest@cvc.uab.cat

P2 Introduction to Object Detection and Segmentation

In M5 project, we will use [Detectron2](#) framework from Facebook Artificial Intelligence Research (FAIR), which is a research platform for object detection and segmentation in Pytorch.



Detectron2 is Facebook AI Research's next generation software system that implements state-of-the-art object detection algorithms. It is a ground-up rewrite of the previous version, [Detectron](#), and it originates from [maskrcnn-benchmark](#).



P2 Introduction to Object Detection and Segmentation

In M5 project, we will use [Detectron2](#) framework from Facebook Artificial Intelligence Research (FAIR), which is a research platform for object detection and segmentation in Pytorch.

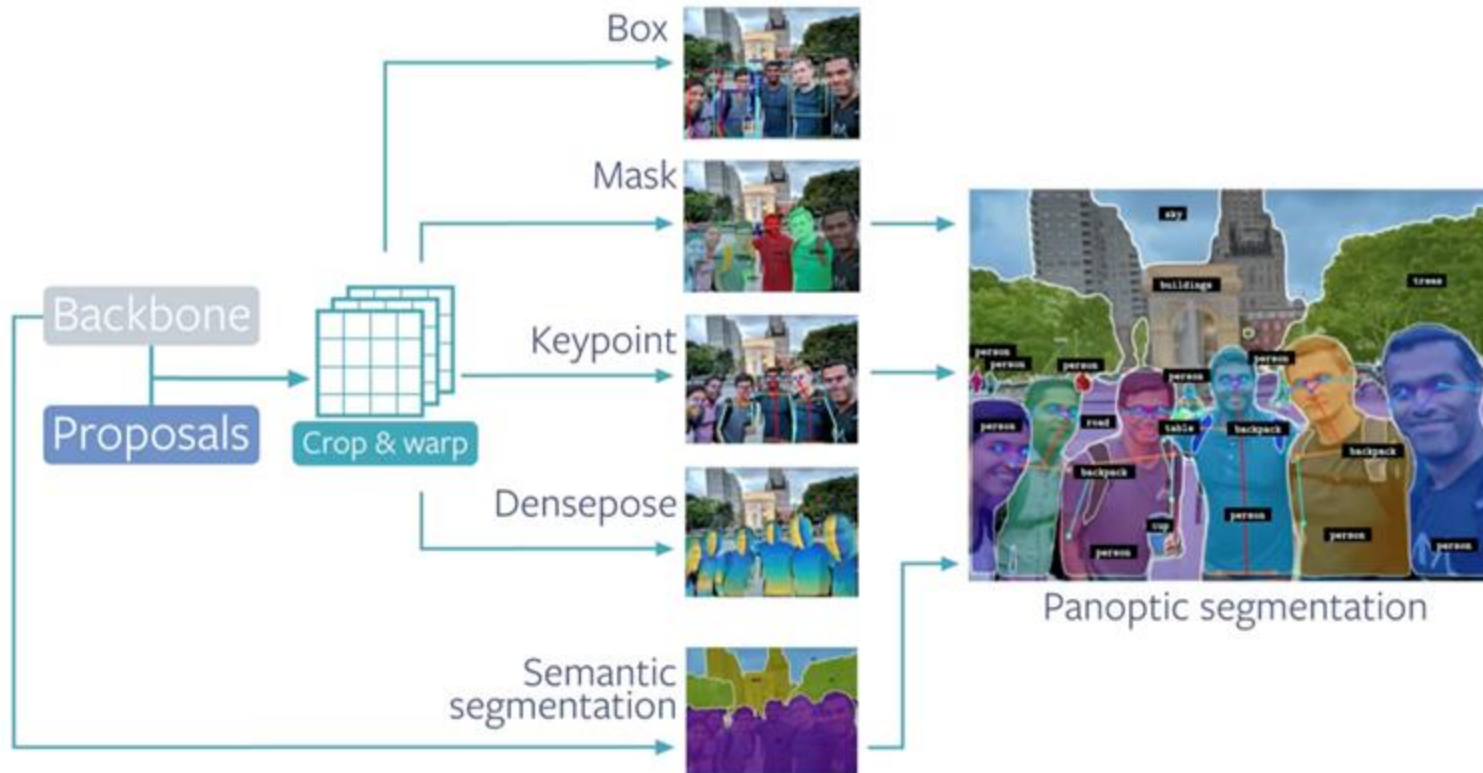
What can we find in Detectron2?

- Powered by PyTorch deep learning framework.
- It supports a range of tasks related to object detection:
 - **Object detection with boxes**
 - **Object detection with instance segmentation masks**
 - Human pose prediction
 - Semantic segmentation
 - Panoptic segmentation

P2 Introduction to Object Detection and Segmentation

In M5 project, we will use [Detectron2](#) framework from Facebook Artificial Intelligence Research (FAIR), which is a research platform for object detection and segmentation in Pytorch.

What can we find in Detectron2?



P2 Introduction to Object Detection and Segmentation

In M5 project, we will use [Detectron2](#) framework from Facebook Artificial Intelligence Research (FAIR), which is a research platform for object detection and segmentation in Pytorch.

What can we find in Detectron2?

- Powered by PyTorch deep learning framework.
- It supports a range of tasks related to object detection.
- It includes many models (Check [model zoo and baselines](#)):

- **Faster R-CNN**
- **Mask R-CNN**
- RetinaNet
- Dense Pose
- TensorMask
- Panoptic FPN
- Cascade R-CNN

COCO Object Detection Baselines

Faster R-CNN:

Name	lr sched	train time (s/iter)	inference time (s/im)	train mem (GB)	box AP	model id	download
R50-C4	1x	0.551	0.102	4.8	35.7	137257644	model metrics
R50-DC5	1x	0.380	0.068	5.0	37.3	137847829	model metrics
R50-FPN	1x	0.210	0.038	3.0	37.9	137257794	model metrics
R50-C4	3x	0.543	0.104	4.8	38.4	137849393	model metrics
R50-DC5	3x	0.378	0.070	5.0	39.0	137849425	model metrics
R50-FPN	3x	0.209	0.038	3.0	40.2	137849458	model metrics
R101-C4	3x	0.619	0.139	5.9	41.1	138204752	model metrics
R101-DC5	3x	0.452	0.086	6.1	40.6	138204841	model metrics
R101-FPN	3x	0.286	0.051	4.1	42.0	137851257	model metrics
X101-FPN	3x	0.638	0.098	6.7	43.0	139173657	model metrics

COCO Instance Segmentation Baselines with Mask R-CNN

Name	lr sched	train time (s/iter)	inference time (s/im)	train mem (GB)	box AP	mask AP	model id	download
R50-C4	1x	0.584	0.110	5.2	36.8	32.2	137259246	model metrics
R50-DC5	1x	0.471	0.076	6.5	38.3	34.2	137260150	model metrics
R50-FPN	1x	0.261	0.043	3.4	38.6	35.2	137260431	model metrics
R50-C4	3x	0.575	0.111	5.2	39.8	34.4	137849525	model metrics
R50-DC5	3x	0.470	0.076	6.5	40.0	35.9	137849551	model metrics
R50-FPN	3x	0.261	0.043	3.4	41.0	37.2	137849600	model metrics
R101-C4	3x	0.652	0.145	6.3	42.6	36.7	138363239	model metrics
R101-DC5	3x	0.545	0.092	7.6	41.9	37.3	138363294	model metrics
R101-FPN	3x	0.340	0.056	4.6	42.9	38.6	138205316	model metrics
X101-FPN	3x	0.690	0.103	7.2	44.3	39.5	139653917	model metrics

P2 Introduction to Object Detection and Segmentation

In M5 project, we will use [Detectron2](#) framework from Facebook Artificial Intelligence Research (FAIR), which is a research platform for object detection and segmentation in Pytorch.

What can we find in Detectron2?

- Powered by PyTorch deep learning framework.
- It supports a range of tasks related to object detection.
- It includes many models (Check [model zoo and baselines](#)).
- Detectron2 beginner's [tutorial](#).

Detectron2 Beginner's Tutorial



Welcome to detectron2! This is the official colab tutorial of detectron2. Here, we will go through some basics usage of detectron2, including the following:

- Run inference on images or videos, with an existing detectron2 model
- Train a detectron2 model on a new dataset

You can make a copy of this tutorial by "File -> Open in playground mode" and make changes there. **DO NOT** request access to this tutorial.

P2 Introduction to Object Detection and Segmentation

In M5 project, we will use [Detectron2](#) framework from Facebook Artificial Intelligence Research (FAIR), which is a research platform for object detection and segmentation in Pytorch.

What can we find in Detectron2?

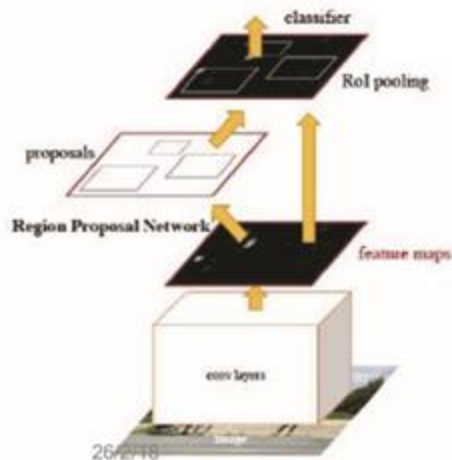
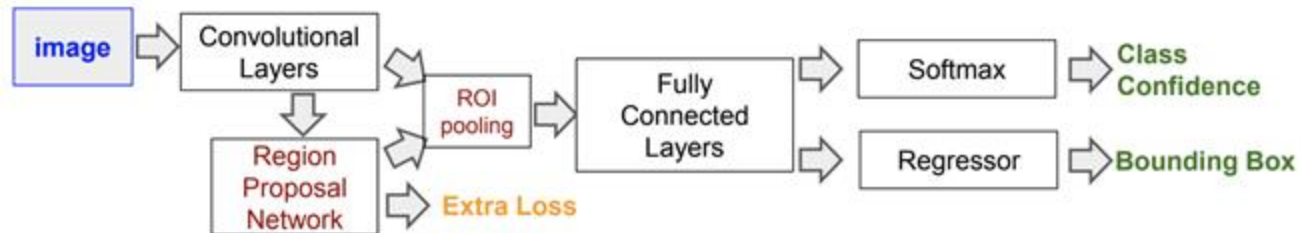
- Powered by PyTorch deep learning framework.
- It supports a range of tasks related to object detection.
- It includes many models (Check [model zoo and baselines](#)).
- Detectron2 beginner's [tutorial](#).
- Official installation [instructions](#).
 - M5 installation instructions
 - Note: Detectron2 beginner's tutorial is thought to be run in Google Colab, not on the server. Therefore:
 - You need to install opencv-python (included in instructions)
 - Check PIL version => If problems:

```
!pip install Pillow==5.3.0
```
 - You can't use cv2_imshow from google.colab.patches but

```
cv2.imwrite(img_filename, img)
```

P2 Introduction to Object Detection and Segmentation

Faster R-CNN [paper](#)

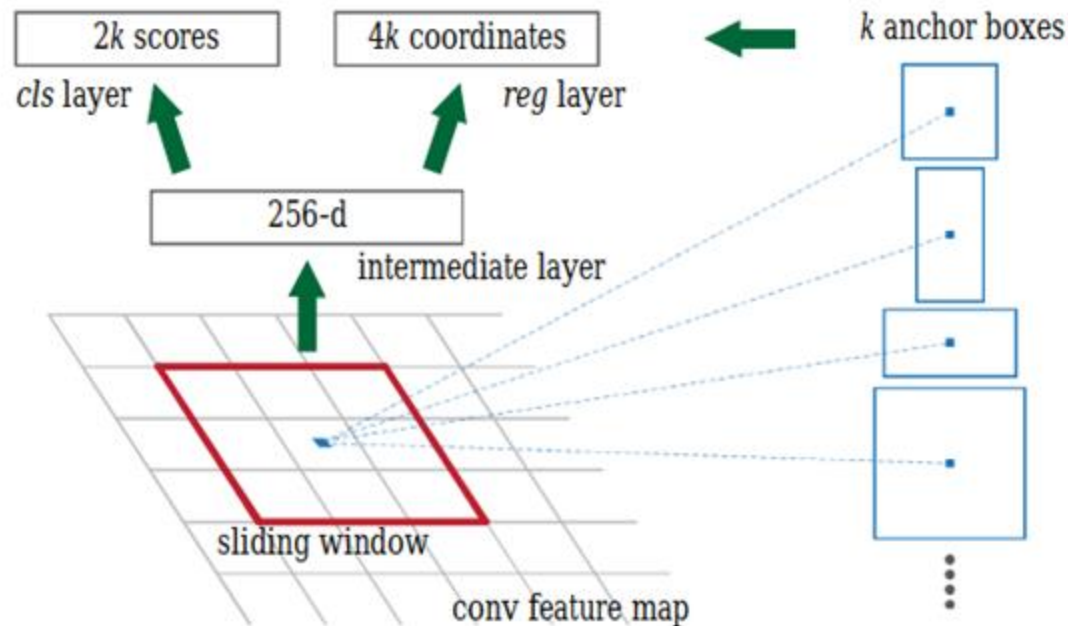


1. Extract feature map and region proposals
2. Infer class, confidence and bounding box for each proposal

P2 Introduction to Object Detection and Segmentation

Faster R-CNN [paper](#)

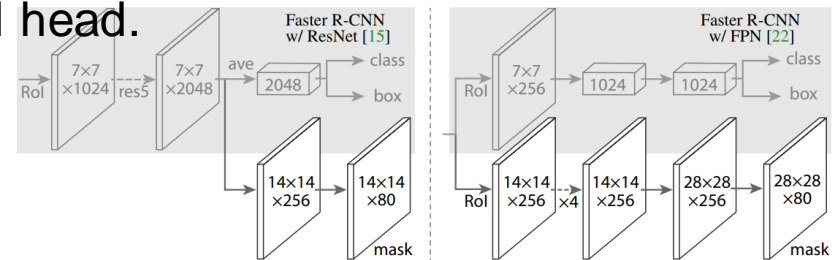
- Novel RPN



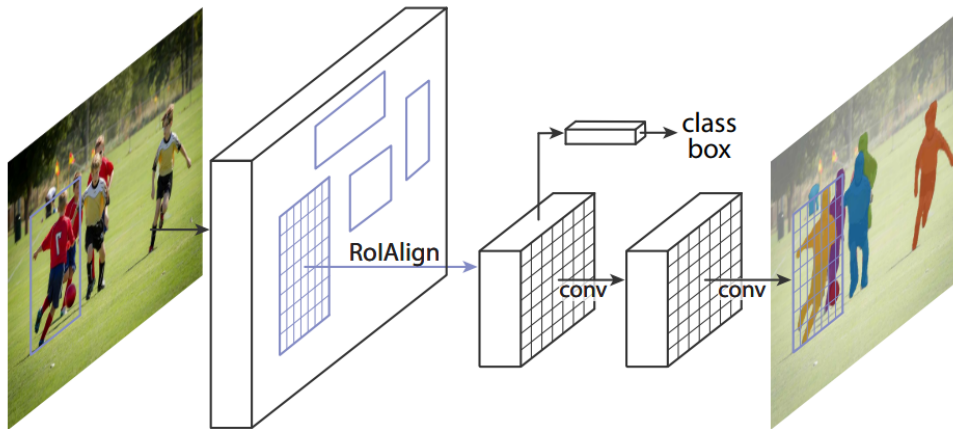
P2 Introduction to Object Detection and Segmentation

Mask R-CNN [paper](#)

- Extension of Faster R-CNN
- Predicts a binary mask for each RoI head.



Head architecture



Mask R-CNN

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 3	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 – P2 Tasks

Week 2: Introduction to Object Detection and Instance Segmentation

Tasks

- a. Get familiar with Detectron2 framework.
- b. Set up project.
- c. Run inference with pre-trained Faster R-CNN (detection) and Mask R-CNN (detection and segmentation) on KITTI-MOTS dataset.
- d. Evaluate pre-trained Faster R-CNN (detection) and Mask R-CNN (detection and segmentation) on KITTI-MOTS dataset.
- e. Fine-tune Faster R-CNN and Mask R-CNN on KITTI-MOTS
- f. Start writing paper

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 overlap about object detection and segmentation.

P2 Dataset: KITTI-MOTS

Dataset: KITTI-MOTS

Tracking and segmentation of CARS and PEDESTRIANS

- Training data
 - 12 sequences
 - 8,073 pedestrian masks + 18,831 car masks
- Validation data
 - 9 sequences
 - 3,347 pedestrian masks + 8,068 car masks
- Testing data: 29 sequences

(*) More details on training and validation split in the original paper: [link](#)



P2 Dataset: KITTI-MOTS

Dataset: KITTI-MOTS

- **Dataset structure**

- Instances (segmentation masks in png format)
- Instances_txt: Annotations in .txt format
- Training/testing: Original image splits.



P2 Dataset: KITTI-MOTS

Dataset: KITTI-MOTS

- **Annotation format**

- Class ids:
 - 1 --> car
 - 2 --> pedestrian
- **Class id** is obtained performing floor division by 1000
- **Instance id** is obtained by modulo 1000
- Car instances: 1000, 1001, 1002, etc.
- Pedestrian instances: 2000, 2001, 2002, etc.

M5 – P2 Tasks

Task (a): **Get familiar with Detectron2 framework**

- Installation of the framework.
- Follow Detectron2 beginner's tutorial.

Task (b): **Set up project**

- Review descriptions of the official challenge [link1](#), [link2](#) (KITTI-MOTS).
- Get familiar with how to read images and annotations.
- You will find KITTI-MOTS dataset on the server.
 - `/home/mcv/datasets/KITTI-MOTS/`

M5 – P2 Tasks

Task (c): Run inference with pre-trained Faster R-CNN (detection) and Mask R-CNN (detection and segmentation) on KITTI-MOTS dataset.

- Apply Faster R-CNN using Detectron2 framework with pretrained weights of COCO on KITTI-MOTS dataset.
- Apply Mask R-CNN using Detectron2 framework with pretrained weights of COCO on KITTI-MOTS dataset.
- **Project presentation:**
 - Dataset description
 - Qualitative results.

M5 – P2 Tasks

Task (d): **Evaluate pre-trained Faster R-CNN (detection) and Mask R-CNN (detection and segmentation) on KITTI-MOTS dataset**

- Apply Faster R-CNN using Detectron2 framework with pretrained weights of COCO on KITTI-MOTS dataset.
- Apply Mask R-CNN using Detectron2 framework with pretrained weights of COCO on KITTI-MOTS dataset.
- Use official validation partition of KITTI-MOTS as your test set.
- Don't use KITTI-MOTS evaluation metrics. Instead use official COCO metrics provided by Detectron2.
- You will have to map class labels of KITTI-MOTS to class labels of COCO. Modify MetadataCatalog: [Official documentation](#), [detectron2 thread](#).
- **Project presentation:**
 - Metric description
 - Quantitative results.

M5 – P2 Tasks

Task (e): **Fine-tune Faster R-CNN and Mask R-CNN on KITTI-MOTS.**

- Train Faster R-CNN and Mask R-CNN using Detectron2 framework on KITTI-MOTS dataset.
 - Split original training set into training and validation sets.
- Evaluate fine-tuned models on your test set (official validation partition) using COCO metrics.
- Compare results with pre-trained models without finetuning.
- Include quantitative and qualitative results in your presentation.
- **Project presentation:**
 - Quantitative and qualitative results.
 - Compare against non-fine-tuned!

M5 – P2 Tasks

Task (f): **Start writing paper.**

- Abstract
- **Introduction (½ page)**
- **State of the art (1 page)**
- Methodology (1 page with diagram)
 - Faster R-CNN & Mask R-CNN
- Experiments
 - Datasets
 - Metrics
- Results
- Conclusion

Max: 6 pages w/o references

M5 Project

General tips:

- The **weekly presentation** objective is to follow the **students' progress**. Therefore, extensive experiments are welcome. You should also include problems you faced and examples you find interesting.
- The final **report/paper** objective is to **summarize** your work and teach you how to write a paper. Only the most relevant experiments and qualitative results are expected, those from where you can get **relevant** conclusions.
- Don't limit the results section to show the results. You must compare and get some insights or conclusions of the results of your experiments.

M5 – P2: Introduction to Object Det. and Seg.

Due date

20th of March, 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.