Università degli Studi di Modena e Reggio Emilia

LM Ingegneria Informatica – AI Engineering
a.a. 2023-2024

Computer Vision and Cognitive Systems

Project Tracks

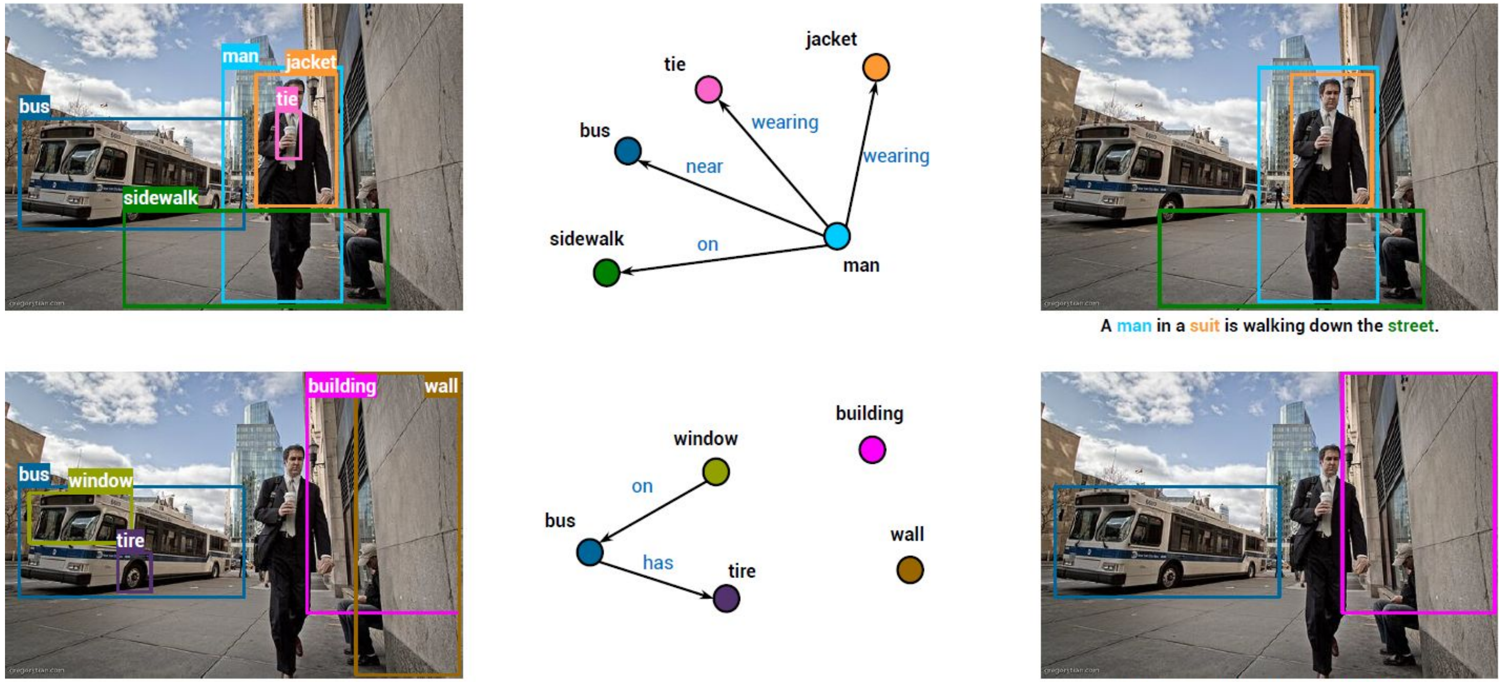
Lorenzo Baraldi, Vittorio Cuculo

Object detection and description

Given images containing a scene with different objects or people, build a pipeline which:

- Detects objects in the scene
- Identifies object positions and their spatial relationships, like “A is behind B”, “A is under B”, “A is next to B”, etc.
- Generates a natural language description of the scene and a quantitative description of the objects which also exploits the knowledge of their relative spatial positions (an LLM might be used)

Possible datasets to be used: Microsoft COCO, Conceptual Captions 3M



A man in a suit is walking down the street.

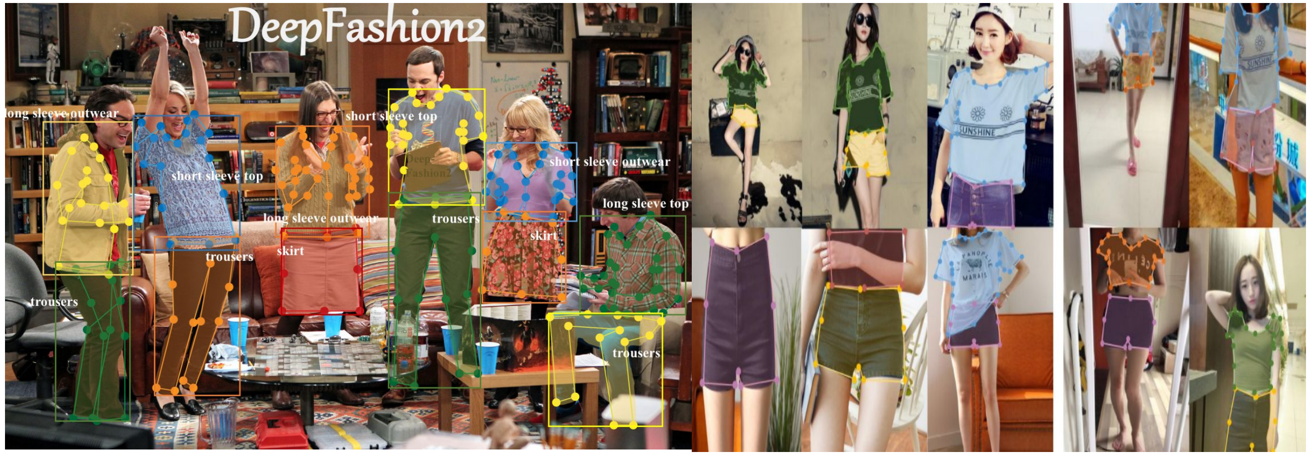
A bus is parked in front of a building.

Human and Fashion data

Given images containing a scene with different people, build a pipeline which:

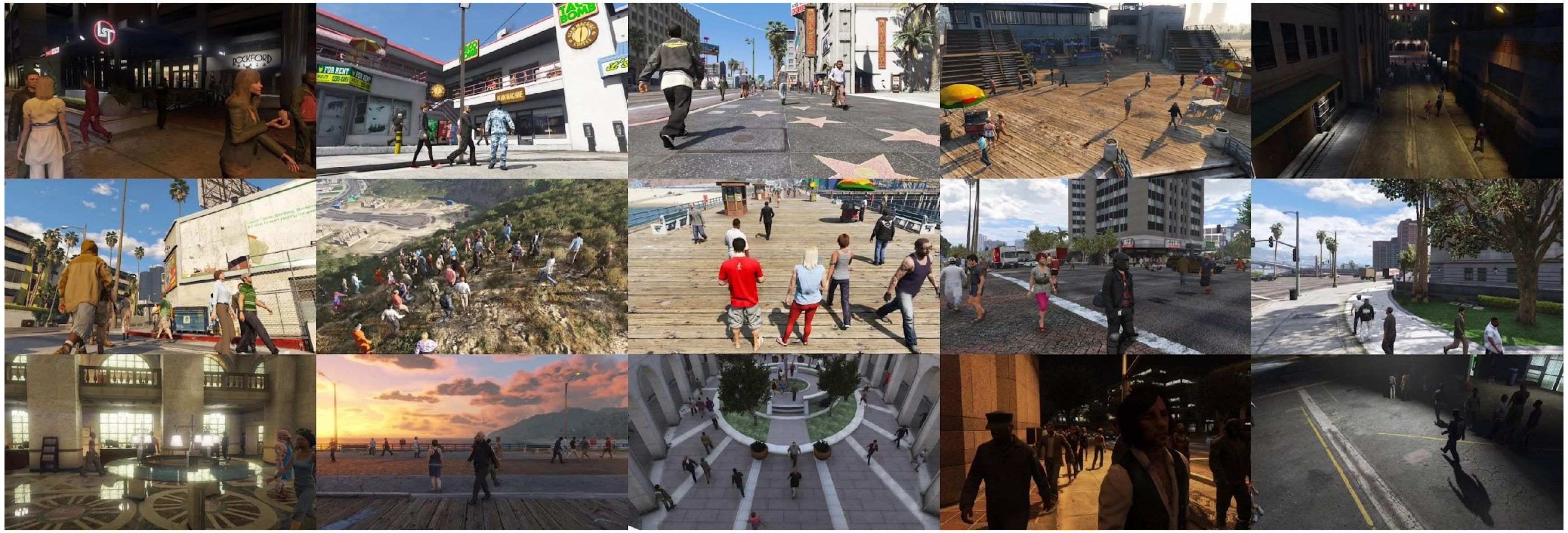
- Detects people on the scene;
- Segments their clothes and classifies them (e.g. shorts, trousers, skirt, outwear, sleeve dress, etc.)
- For each clothing item, retrieves similar items in a database of clothes (e.g. via color or low-level similarity or with a learnable CNN/Vision Transformer) – OR – given a person instance and a clothing item from a database of clothes, generates a "virtual try-on" view of that person wearing that clothing item (e.g. using a GAN).

Possible dataset to be used: UNIMORE Fashion Dataset or https://github.com/switchablenorms/DeepFashion2.



People counting

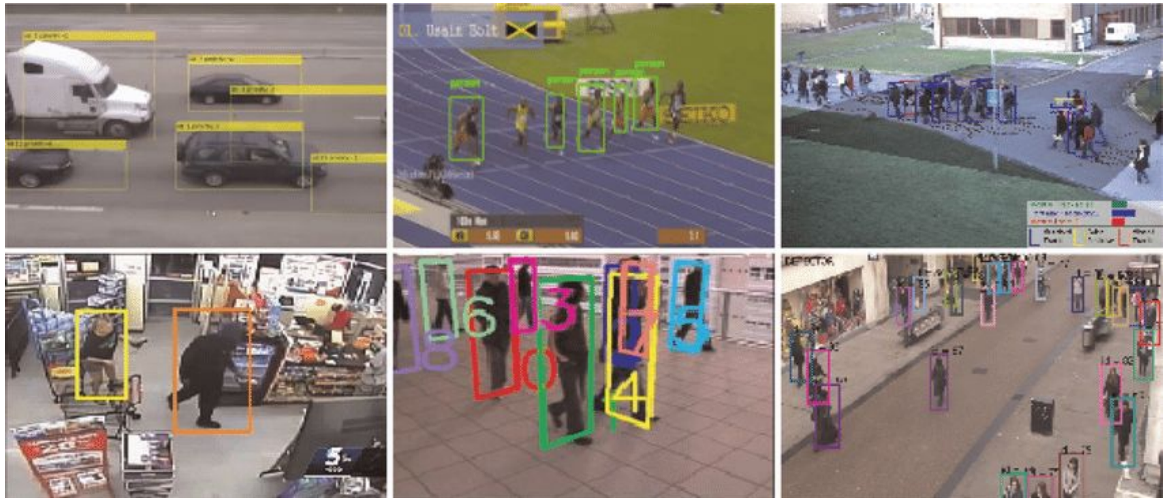
- **Count the number of individuals present in a location of interest**, within a given time frame (e.g., ranging from a few seconds to 1-2 minutes).
- The system must be designed principally for outdoor urban scenarios, which could be crowded with people strolling, cycling, or jogging.
- Other object categories may be present (e.g., vehicles), but the system should ignore them while counting.
- You might use a two-stage pipeline involving object detection and tracking; however, the investigation of alternative and more modern strategies is also encouraged.
- The proposed solution will be benchmarked on the synthetic dataset **MotSynth**, a dataset for pedestrian detection and tracking in urban scenarios created by exploiting the highly photorealistic video game Grand Theft Auto V.



Multi-Object Tracking

- **Tracking-by-detection** splits the problem of tracking objects into two main stages: **detect objects** and then **associate them over consecutive frames**.
- For each tracked object, these approaches usually integrate an established motion model (based on the Kalman Filter) with additional cues, reflecting the expected visual appearance of the object under consideration in the next frame.
- The project aims to provide an approach to adapt the deep features used by tracking-by-detection algorithms (e.g., Deep SORT [3]). In particular, the students will leverage an off-the-shelf appearance model, which has been already trained on the synthetic MOT Synth dataset [1].

[1] Fabbri, M., Brasó, G., Maugeri, G., Cetintas, O., Gasparini, R., Ošep, A., ... &amp; Cucchiara, R. (2021). Motsynth: How can synthetic data help pedestrian detection and tracking?. In Proceedings of the IEEE/CVF International Conference on Computer Vision (pp. 10849-10859).
[2] Seidenschwarz, J., Brasó, G., Serrano, V. C., Elezi, I., &amp; Leal-Taixé, L. (2023). Simple Cues Lead to a Strong Multi-Object Tracker. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (pp. 13813-13823).
[3] Wojke, N., Bewley, A., &amp; Paulus, D. (2017, September). Simple online and realtime tracking with a deep association metric. In 2017 IEEE international conference on image processing (ICIP) (pp. 3645-3649). IEEE.



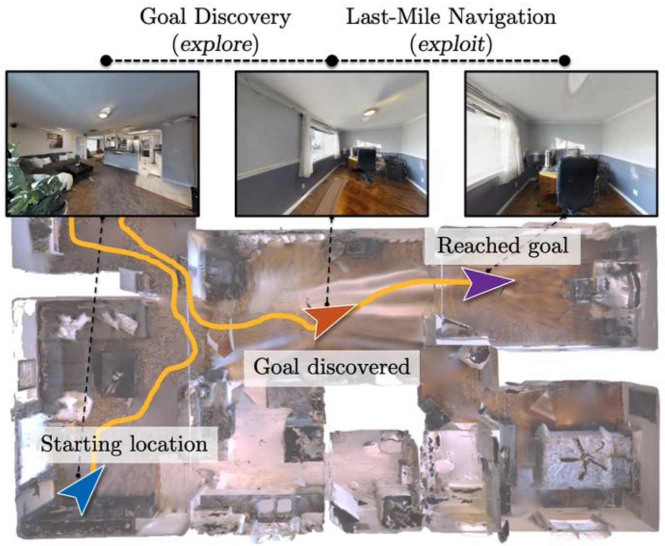
Embodied Vision

Given an image taken from the camera of a robot in an environment, build a pipeline that:

- Retrieves from a database of videos collected by the robot in the same environment (possibly under different conditions) image(s) taken in the same place as the query image
- Geometrically aligns the query image with the retrieved candidates
- Moves the robot towards the place in which the image was taken

Possible platform for simulation: Habitat (https://aihabitat.org/)

Possible datasets to be used: the IDOL2 dataset (indoor), the Saint Lucia dataset (urban), your own dataset collected by using one of the robots at the AImageLab.



Generative AI

- TBD

Foundation Models

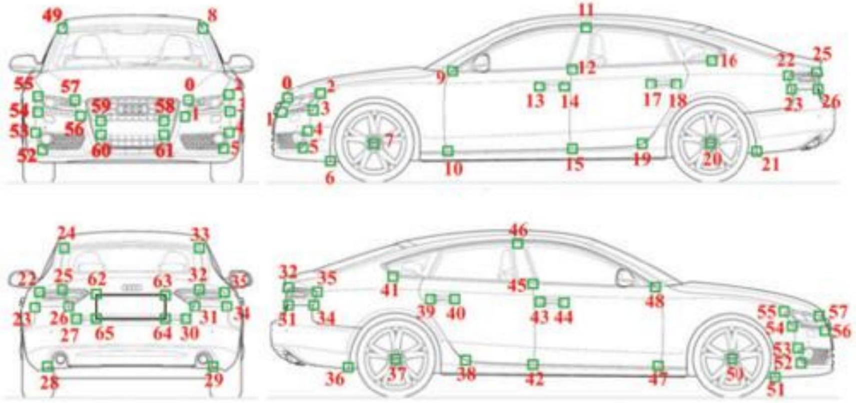
- TBD

Keypoint detection for Automotive

prometeia

UNIMORE UNIVERSITÀ DEGLI STUDI DI MODENA E REGGIO EMILIA

- Identify a set of 3D models representing a category of large vehicles (e.g., buses/trucks/vans).
- Identify a set of keypoints (> 20, < 60) common to all vehicles.
- Build a dataset of 2D captures (> 2000 images) of these vehicles from various perspectives enriched with keypoint annotations (hint: devise a strategy to avoid manually annotating each individual image).
- Train a keypoint detection model on the created dataset.
- Test the model on:
  - ○ New captures of the same 3D models used in training.
  - ○ Captures of 3D models never seen in training.



Websites for 3D models (examples):
https://sketchfab.com/feed, https://www.cgtrader.com/, https://free3d.com/
Repository gathering various examples of keypoint detection models:
https://github.com/openmmlab/mmpose

Satellite Images

Given a dataset of satellite images…

- Identify macro-classes of areas of interest (urban areas, forests, rivers/lakes, etc.) using existing or custom detection/segmentation models.
- Detect classes of objects in photos with known dimensions, such as cars, football fields, buildings, etc.
- Evaluate the distance (m, km) between defined points and areas of interest: this calculation can be done using the proportion with a known object or by using the satellite's spatial resolution
- Define a metric for assessing the impact of risk areas (such as rivers, lakes, forests) on areas of interest.

What to do now

- ASAP, declare your group (3 people) and the project of choice at: https://ailb-web.ing.unimore.it/courses/course/cvcs2024/
- Before starting, make sure to revise the “Final Project and Exam” slides on Moodle and especially the “What we evaluate” part.
- Request a tutor from one of the teachers. Contact him/her to grab relevant information, papers to read, additional datasets, and code. Start designing your project and prepare the project proposal.
- Request regular meetings with your tutor during project development.
- Once the project is ok, start writing your report. Submit it online before the exam.