

Supervisor: Joul Skaf

J1 - Instance Segmentation on Indoor Objects Datasets

Description

The main objective of the project is to train/finetune, evaluate, test, and run inference on a model capable of performing instance-level segmentation on a pre-defined set of objects, a combination of different datasets could be used. The combined dataset should be trained on one or multiple state-of-the-art models. A comparison between the models is required regarding precision and execution speed (FPS). A theoretical comparison justifying the choice of the model is also acceptable.

Difficulties

- Requires intensive and careful data pre-processing, depending on the selected dataset.
- The training could take a long time, mainly if transfer learning is not properly utilized.
- Some state-of-the-art models offer different versions, in which some are smaller and faster, while others are slower and more accurate. Therefore a trade-off is required.
- Unbalanced dataset.

Datasets:

- [COCO instance segmentation dataset](#)
- [ADE20K](#)
- [SUNRGB-D](#) & NYU V2
- [NYU V1](#)
- [MIT IndoorScene](#)
- Other datasets that contain the objects in question

Categories

Wall, floor, ceiling, chair, table, bed, sofa, pillow, window, curtain, picture, monitor, cabinet, refrigerator, oven, trashcan, bathtub, sink, toilet, and "Object" which is a general class.

Resources

- <https://paperswithcode.com/paper/mask-dino-towards-a-unified-transformer-based-1>
- <https://paperswithcode.com/paper/exploring-plain-vision-transformer-backbones>
- <https://paperswithcode.com/paper/swin-transformer-v2-scaling-up-capacity-and>
- [Instance Segmentation Benchmark datasets](#)
- <https://github.com/alibaba/EasyCV>
- <https://github.com/facebookresearch/detectron2>

J2 - 3D Instance Segmentation on Indoor Objects Datasets

Description

The main objective of the project is to train/finetune, evaluate, test, and run inference on a model capable of performing 3D instance-level segmentation on a pre-defined set of objects, a combination of different datasets could be used. The combined dataset should be trained on one or multiple state-of-the-art models. A comparison between the models is required regarding precision and execution speed (FPS). A theoretical comparison justifying the choice of the model is also acceptable.

Difficulties

- Requires intensive and careful data pre-processing, depending on the selected dataset.
- The training could take a long time, mainly if transfer learning is not properly utilized.
- Some state-of-the-art models offer different versions, in which some are smaller and faster, while others are slower and more accurate. Therefore a trade-off is required.
- Unbalanced dataset.

Dataset

- <http://www.scan-net.org/>
- <http://buildingparser.stanford.edu/dataset.html>
- <https://hkust-vgd.github.io/scenenn/>
- Other datasets that contain the objects in question

Categories

Wall, floor, ceiling, chair, table, bed, sofa, pillow, window, curtain, picture, monitor, cabinet, refrigerator, oven, trashcan, bathtub, sink, toilet, and "Object" which is a general class.

Resources:

- [3D Instance Segmentation Benchmark datasets](#)

J3 - 3D point-cloud/Object Classification

Description:

The main objective of the project is to train/finetune, evaluate, test, and run inference on a model capable of classifying 3D mesh points. A combination of different datasets could be used. The combined dataset should be trained or fine-tuned using transfer learning on one or multiple state-of-the-art models. A comparison between the models is required regarding precision and speed (FPS), A theoretical comparison justifying the choice of the model is also acceptable.

Difficulties

- Requires intensive literature research.
- Unbalanced datasets

Dataset

- [ModelNet40](#)
- [ScanObjectNN](#)
- Other datasets.

Resources:

- <https://paperswithcode.com/task/3d-point-cloud-classification>
- <https://paperswithcode.com/task/3d-object-classification>

J4 - Multiple-object Tracking in Real-time

Description

The main objective of the project is to train/finetune, evaluate, test, and run inference on a state-of-the-art model capable of tracking multiple objects across frames in a given video. A comparison between the models is required regarding precision and execution speed (FPS), A theoretical comparison justifying the choice of the model is also acceptable. It's also required to reach real-time performance with minimal delay.

Difficulties

- Require optimization to achieve real-time performance.

Datasets

- <https://paperswithcode.com/dataset/dancetrack>
- <https://paperswithcode.com/sota/multi-object-tracking-on-mot20-1>

Resources

- [Benchmark datasets](#)

J5 - Speech Separation with visual Enhancement

Description:

The task is to train a speech separation model that utilizes the visual information of the speaker (facial expressions, and movements in general) to completely separate the speaker's speech from the background noise, and other people in the vicinity.

Difficulties:

- Requires intensive literature search.
- Requires search for English-based datasets, and pre-trained models.
- Advanced programming skills.

Resources:

- <https://github.com/danmic/av-se>
- https://paperswithcode.com/search?q_meta=&q_type=&q=Speech+Separation+with+visual+Enhancement
- https://scholar.google.com/scholar?as_ylo=2021&q=speech+separation+visual+enhancement&hl=en&as_sdt=0.5

J6 - MBTI Personality Classification

Description:

Train, evaluate, test, and run inference on a model capable of classifying text of varying lengths based on the Myers-Briggs Type Indicator (MBTI). Which contains 16 classes. The task could also be converted into 4 binary classification tasks. A comparison between different state-of-the-art models is required regarding the accuracy. A theoretical comparison justifying the choice of the model is also acceptable.

Difficulties:

- Requires literature search on state-of-the-art text classification models using pre-trained language models.

Dataset:

- [MBTI Personality Types 500](#)

Resources:

- <https://arxiv.org/abs/2207.04476>
- <https://paperswithcode.com/task/text-classification>
- <https://paperswithcode.com/sota/text-classification-on-20news>

J7 - Deep Clustering for Unsupervised Image Classification

Description

The purpose of the project is to utilize unsupervised image classification techniques on the German traffic sign dataset. A comparison between different state-of-the-art models is required regarding the accuracy, A theoretical comparison justifying the choice of the model is also acceptable.

Difficulties

- Relatively new concept that requires background knowledge in clustering techniques.

Dataset

- <https://www.kaggle.com/datasets/meowmeowmeowmeowmeow/gtsrb-german-traffic-sign>

Resource

- <https://paperswithcode.com/paper/invariant-information-distillation-for>
- <https://paperswithcode.com/task/unsupervised-image-classification>