Monocular 3D Object Detection by using Attention

Eng. Diana-Alexandra Sas

PhD Student Computer Science @UTCN
Computer Vision Engineer @Bosch
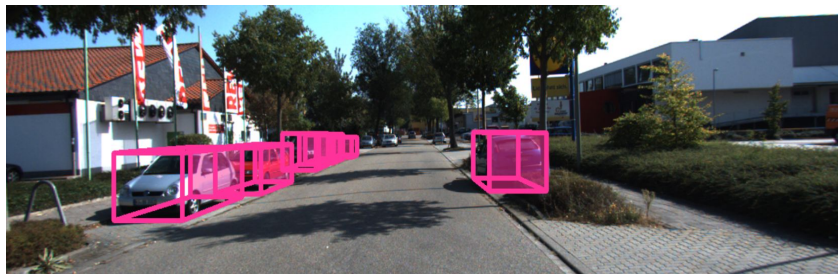sas.cr.diana@student.utcluj.ro

AI synergy!

ai forum
U T C N

Monocular 3D Object Detection

How can a single camera be used to understand 3D space?

- Autonomous driving context

- **Input**: single digital image
- **Output**: 3D bounding boxes
- **Categories**: cars, pedestrians, cyclists

2D Object Detection

3D Object Detection
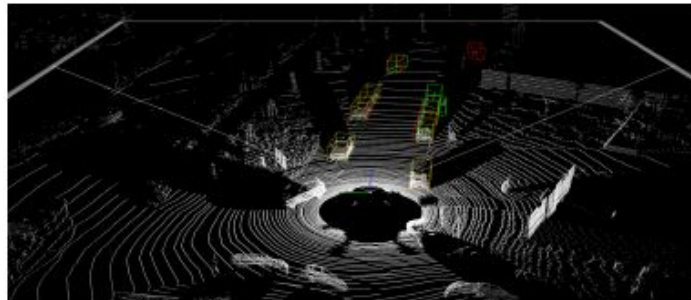
Monocular 3D Object Detection

The Challenge: Understanding 3D from 2D

Camera-based detection

LiDAR-based detection

Depth information

[1] Yin Zhou and Oncel Tuzel. VoxelNet: End-to-end learning for point cloud-based 3D Object detection, arXiv preprint arXiv:1711.06396, 2017.

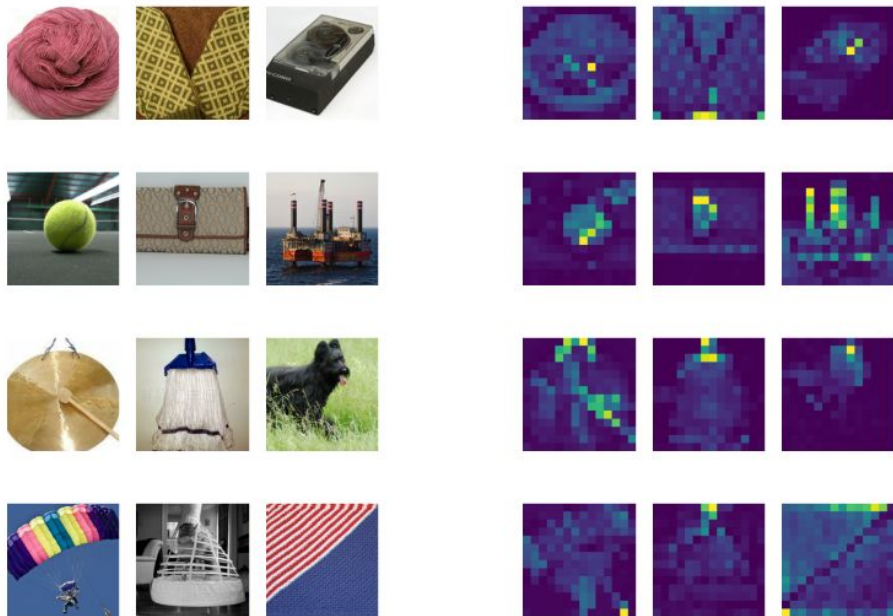*Can AI help cameras "see" in 3D without extra sensors?*

YES

How can we better understand the scene from a single image?

- Learn local spatial patterns.
- **Capture global context.**

 Attention and Transformer architectures

[2] Xiangning Chen and Cho-Jui Hsieh and Boqing Gong, When Vision Transformers Outperform ResNets without Pre-training or Strong Data Augmentations, arXiv preprint arXiv:1711.06396, 2022

Monocular 3D Object Detection by using Attention

Why add more sensors when we can still use information from the existing data?

- Advantages:
  - Robustness
  - Reliability

Figure 2. Segmentation overlays for cars, pedestrians and cyclists

Monocular 3D Object Detection by using Attention

Impact

- More affordable 🚗 -> Reduces dependency on LiDAR sensors.
- Improved performance 👁 -> Better detection for small, occluded and truncated objects.
- Beyond cars 🤖 -> Can be used in robotics or AR applications.
- Next steps 🔬 -> Testing in real-world conditions to move beyond TRL 3.

Collaboration offer

Open for

research collaborations, **joint research proposals, knowledge transfer**

on Monocular 3D Object Detection in the autonomous driving context

AI synergy!