# Semantic Segmentation Learning for Autonomous UAVs using Simulators and Real Data

**Bianca-Cerasela-Zelia Blaga**
**Prof. Dr. Eng. Sergiu Nedevschi**

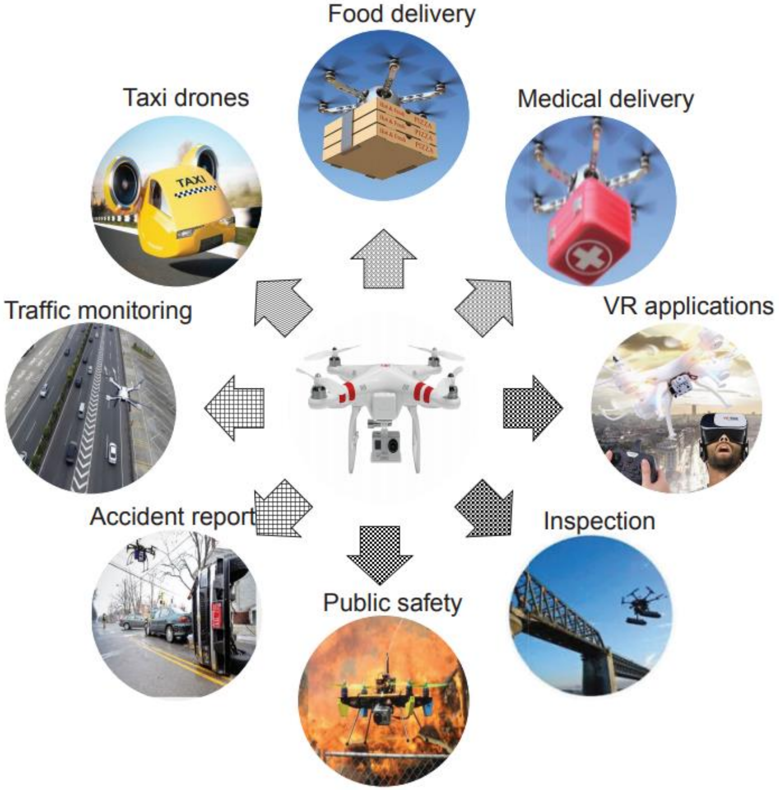
**Computer Science**

**Technical University of Cluj-Napoca**

# Contents

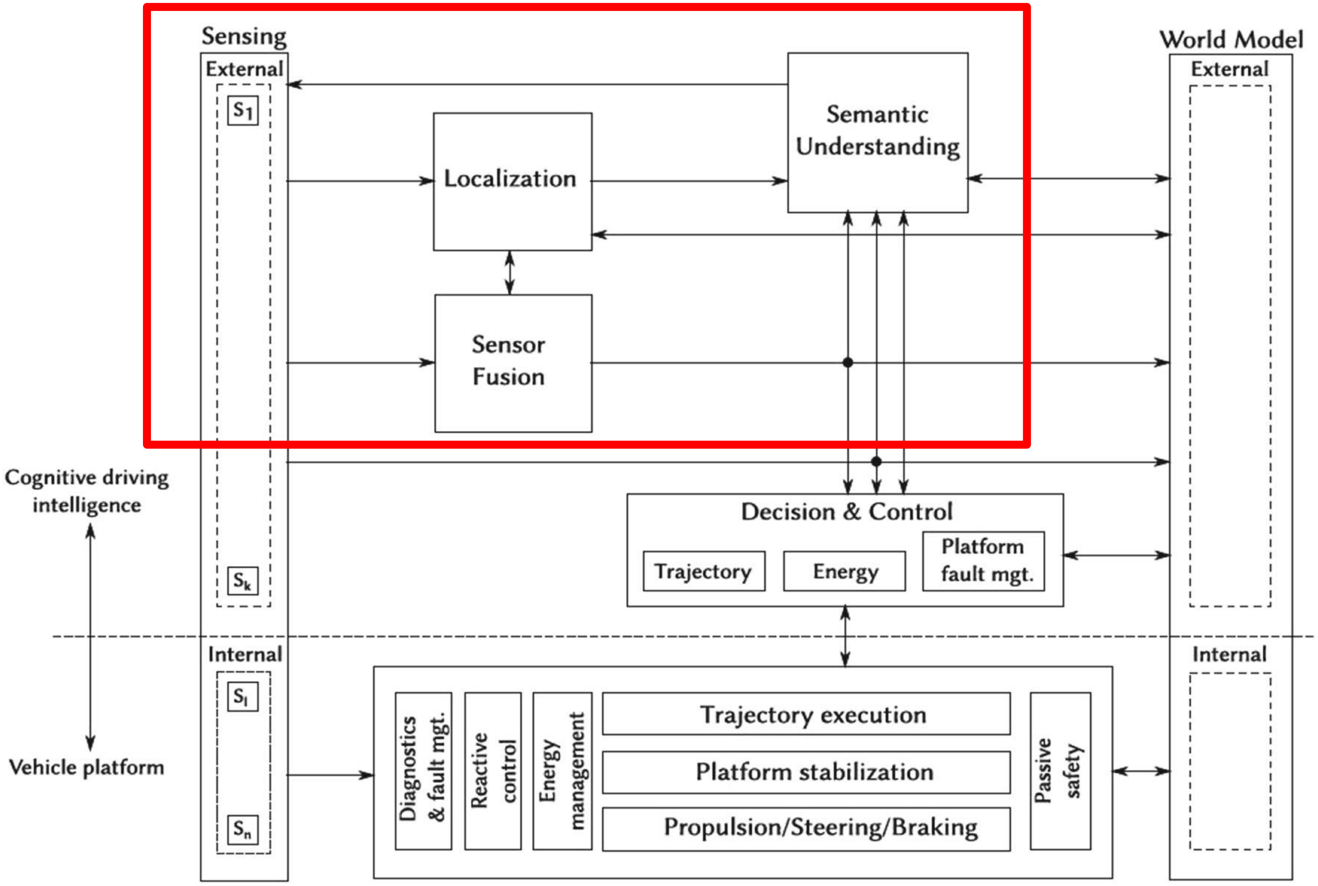
# 1. Introduction

- Nowadays, an increase in the use of UAVs has been noticed, for civilian applications like aerial photography, survey, inspection, mapping, package delivery or surveillance.



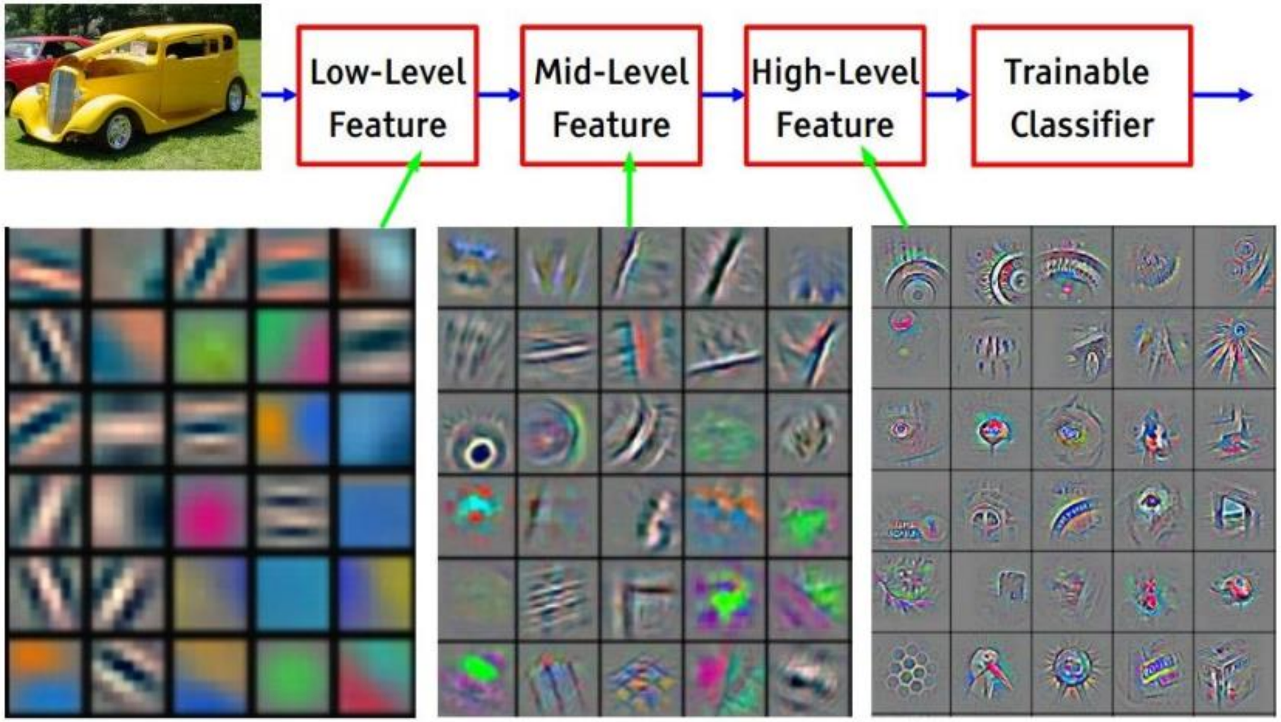
# 1. Introduction

- Perception – the first step to achieve autonomous navigation, be it for cars or Unmanned Aerial Vehicles (UAVs)



# 1. Introduction

- Deep learning – used to achieve a high level of accuracy through Convolutional Neural Networks (CNN), Recurrent Neural Networks (RNN) or Generative Adversarial Networks (GAN)



# 1. Introduction

- Application requirements for deep learning:
  - **depth and optical flow estimation** – stereo images, monocular image sequences, depth maps, optical flow ground truth,
  - **object detection and tracking** – semantically annotated 2D images, labeled 3D point clouds, object bounding boxes and classes,
  - **pedestrian detection and intention learning** – bounding boxes, semantic annotations, 3D human joint representations,
  - **scene understanding** – semantic annotations for 2D and 3D data, bounding boxes, action descriptors, object relationships,
  - **autonomous navigation** – steering wheel, throttle, and brake recordings, car trajectory, 3D maps, together with images and all the previous types of inputs.

# 2. Motivation and Objectives

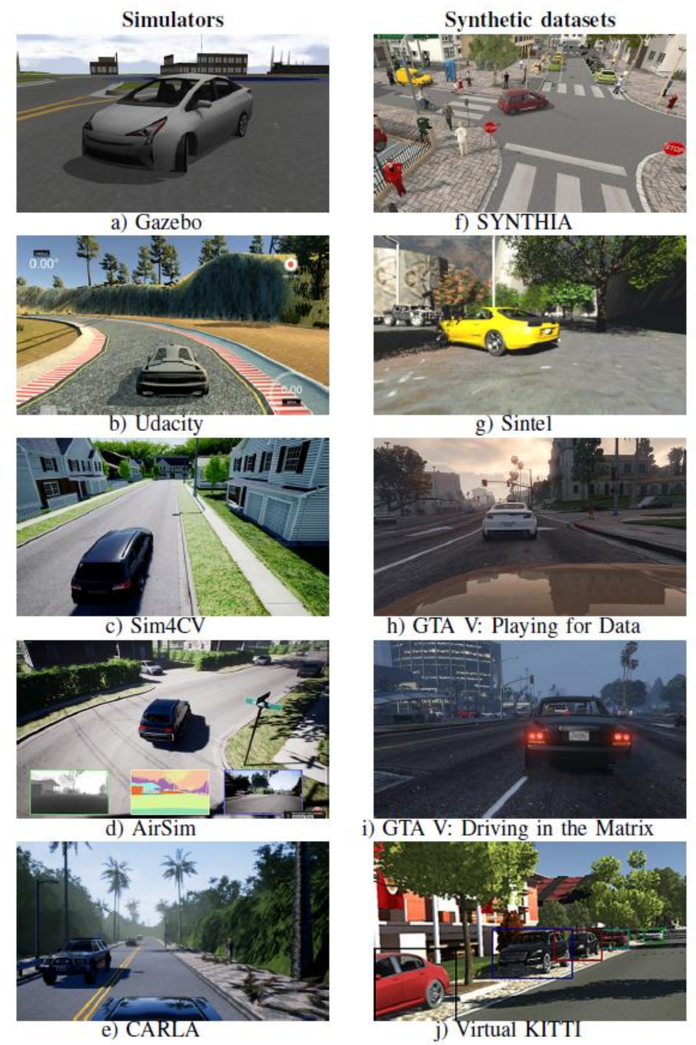
- Since:
  - massive amounts of data are required for training models,
  - the prediction accuracy depends on the quality and size of the input dataset,
  - manual annotation is time-consuming and difficult,
  - semantic segmentation of aerial images recorded from drones is a less researched topic (only two papers discuss it),
- We want to:
  - prove the importance of simulators for various computer vision tasks (depth and optical flow estimation, object detection and tracking, pedestrian detection and intention learning, scene understanding and autonomous navigation), as they can instantly create ground truth recordings for multiple sensors,
  - achieve a high level of accuracy through methods like deep learning for a case study on semantic segmentation for images taken from drones.

# 3. Contributions

- For the previously mentioned objectives, we bring the following contributions:
  - a survey of simulators and synthetic datasets,
  - introduction of an aerial camera in the CARLA simulator,
  - obtaining semantic segmentation on data from drones using deep learning,
  - generating a large and complex synthetic dataset from a UAV – which contains ground truth for both color and label images,
  - transitioning from virtual to real data by fine-tuning a network,
  - gathering a dataset that contains both real and synthetic images – which solves the issues noticed in both.

# 4. Survey of Simulators and Synthetic Datasets for Deep Learning

- **Simulators** – computer programs that model some aspects of the real world with the purpose of generating virtual recordings of scenarios that are scarce in existing data
- **Synthetic datasets** – artificially created and recorded from simulators
- **Advantages**:
  - ➢ large number of recordings
  - ➢ multiple sensors
  - ➢ ground truth data
  - ➢ various weather conditions and day times
  - ➢ accurate physics modeling
- **Disadvantages**:
  - ➢ low level of realism
  - ➢ simplistic scenarios
- **Future work directions**:
  - ➢ GANs – for style transfer
  - ➢ procedural environment generation



# 4. Survey of Simulators and Synthetic Datasets for Deep Learning

Contents and capabilities of the simulators and synthetic datasets

| Simulator | # images | Camera | Depth | Flow | Labeling | 3D data | Position |
|---|---|---|---|---|---|---|---|
| Gazebo | - | ✓ | ✓ | ✓ | ✗ | ✓ | ✓ |
| Udacity | - | ✓ | ✗ | ✗ | ✗ | ✗ | ✓ |
| Sim4CV | - | ✓ | ✓ | ✗ | ✗ | ✗ | ✓ |
| AirSim | - | ✓ | ✓ | ✗ | ✓ | ✓ | ✓ |
| CARLA | - | ✓ | ✓ | ✗ | ✓ | ✓ | ✓ |
| SYNTHIA | 20,000 | ✓ | ✓ | ✓ | ✓ | ✗ | ✗ |
| Sintel | 35,000 | ✓ | ✓ | ✓ | ✓ | ✗ | ✗ |
| GTA V v1 | 24,000 | ✓ | ✗ | ✗ | ✓ | ✗ | ✗ |
| GTA V v2 | 200,000 | ✓ | ✗ | ✗ | ✓ | ✗ | ✗ |
| Virtual Kitti | 25,000 | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ |

# 4. Survey of Simulators and Synthetic Datasets for Deep Learning

| CARLA | SYNTHIA | GTA V v1 | GTA V v2 | Virtual KITTI |
|---|---|---|---|---|
| Unlabeled | Sky | Road | Cat | Buidling |
| Building | Building | Buidling | Sofa | Car |
| Fence | Road | Sky | Sheep | Guard rail |
| Other | Sidewalk | Sidewalk | Boat | Misc |
| Pedestrian | Fence | Vegetation | Bus | Pole |
| Pole | Vegetation | Car | Motorbike | Sky |
| Road line | Pole | Terrain | Cow | Terrain |
| Road | Marking | Wall | Dog | Traffic light |
| Sidewalk | Car | Truck | Horse | Traffic sign |
| Vegetation | Sign | Pole | Car | Tree |
| Car | Pedestrian | Fence | Pottedplant | Truck |
| Wall | Cyclist | Bus | Tvmonitor | Van |
| Traffic Sign | | Person | Person | Vegetation |
| | | Traffic light | Aeroplane | |
| | | Traffic sign | Diningtable | |
| | | Train | Bicycle | |
| | | Motorcycle | Bird | |
| | | Rider | Train | |
| | | Bicycle | Bottle | |
| | | | Chair | |

# 5. Aerial Camera in the CARLA Simulator

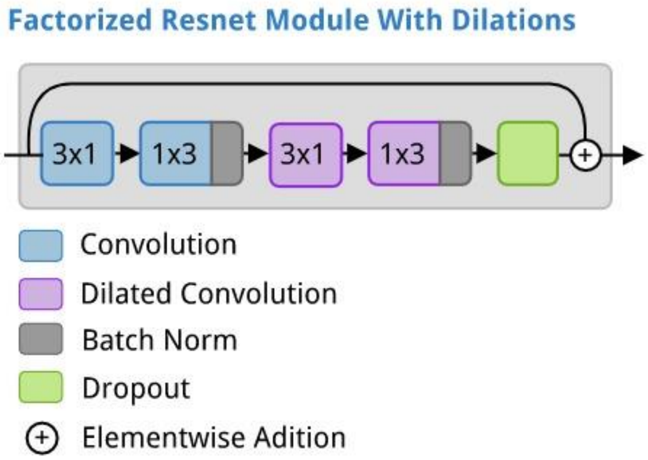
- Control update functions, where x-axis is forward, y-axis – to the left and the z-axis – upward, $\alpha$, $\beta$, $\gamma$ are yaw, pitch and roll angles, $c = 3$ and $a = 5$

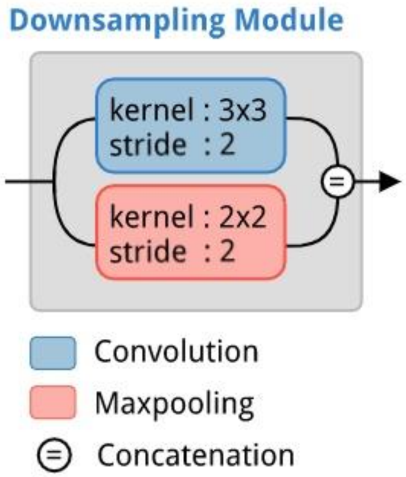
| Key | Control | Location or rotation update |
|---|---|---|
| T | Move forward | $x = x + c$ |
| G | Move backward | $x = x - c$ |
| F | Move left | $y = y - c$ |
| Y | Move right | $y = y + c$ |
| U | Move up | $z = z + c$ |
| J | Move down | $z = z - c$ |
| I | Rotate pitch forward | $\beta = \beta - a$ |
| K | Rotate pitch backward | $\beta = \beta + a$ |
| O | Rotate yaw left | $\alpha = \alpha - a$ |
| L | Rotate yaw right | $\alpha = \alpha + a$ |
| P | Rotate roll left | $\gamma = \gamma - a$ |
| ; | Rotate roll right | $\gamma = \gamma + a$ |

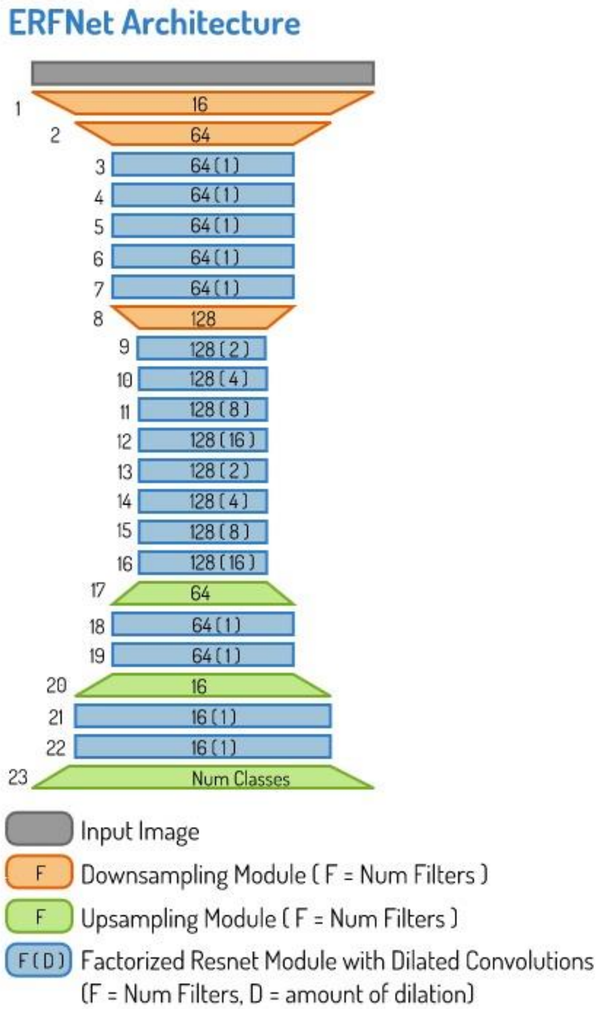
# 6. Semantic Segmentation on Images Taken from Drones

## 7.1. Semantic Segmentation

- ERFNet – Efficient Residual Factorized Network
- Components:
  - a factorized residual network module with dilations
  - a downsampling module inspired by an inception structure
  - an upsampling module



# 6. Semantic Segmentation on Images Taken from Drones

## 7.1. Semantic Segmentation

- Data augmentation techniques: shadow augmentation, random rotation, random crops, random brightness, random contrast, random blur, and random noise

- Weight class is assigned based on class probability

$$weight_{\text{class}} = \frac{1}{ln(c + probability_{\text{class}})}$$

- The output is computed using the softmax function, which assigns to each pixel the probability to belong in each class.

## 7.2. Testing

- **System specifications**: Ubuntu 18.04 operating system, Intel Core i7-6700K 4GHz CPU, NVIDIA GeForce GTX 1080Ti GPU, CUDA 10, TensorFlow 1.13
- **Metrics**: **precision** (the fraction of relevant instances among the retrieved ones), **recall** (the fraction of relevant instances that have been retrieved over the total amount of relevant instances) and **IoU** (Intersection over Union, Jaccard index)

$$precision = \frac{TP}{TP + FP}$$

$$recall = \frac{TP}{TP + FN}$$

$$IoU = \frac{TP}{TP + FP + FN}$$

- **Trained** for 800 epochs with decreasing learning rates
- **Validated** on 10% of total data

| Image size | Time (ms) |
|---|---|
| 600x400px | 10 |
| 1024x512px | 24 |

# 6. Semantic Segmentation on Images Taken from Drones

## 7.3. Dataset Gathering using CARLA

- Using the aerial camera inserted in the CARLA simulator we record:
  - 30,000 images
  - size 1024x512 px
  - scenarios: from towns with tall buildings to villages with small houses, that contain varying forms of vegetation, complex road structures, resembling real-life environments
  - 100 traffic participants in the form of cars, bicycles and motorcycles
  - dynamic weather conditions like sunny, cloudy, rainy, or dark
  - camera noise: vignette, grain jitter, bloom, auto exposure and lens flare
  - 13 semantic classes – unlabeled, building, fence, other, pedestrians, pole, road line, road, sidewalk, vegetation, car, wall, traffic sign

# 6. Semantic Segmentation on Images Taken from Drones

## 7.3. Dataset Gathering using CARLA

## 7.3. Evaluation on the CARLA Dataset

Precision, recall and IoU validation results on ERFNet trained on CARLA, for two image dimensions.

| Class | CARLA 512x512 | | | CARLA 1024x512 | | |
|---|---|---|---|---|---|---|
| | Prec. | Recall | IoU | Prec. | Recall | IoU |
| Unlabeled | 94.12 | 91.26 | 86.46 | **96.80** | **92.16** | **89.43** |
| Building | 92.98 | 96.45 | 90.69 | **96.03** | **97.36** | **92.75** |
| Fence | 65.40 | 65.98 | 48.91 | **67.67** | **77.22** | **56.41** |
| Other | 71.58 | 67.99 | 53.54 | **74.81** | **73.68** | **59.03** |
| Pedestrian | 0 | 0 | 0 | 0 | 0 | 0 |
| Pole | 60.72 | 39.84 | 33.52 | **67.87** | **62.86** | **44.69** |
| Road line | 68.96 | 93.50 | 65.81 | **78.46** | **93.61** | **74.47** |
| Road | 98.50 | 95.91 | 94.86 | **98.86** | **96.49** | **95.09** |
| Sidewalk | 94.83 | 93.77 | 89.23 | **94.85** | **95.25** | **90.54** |
| Vegetation | 87.60 | 92.04 | 81.45 | **87.62** | **93.91** | **82.89** |
| Car | 83.22 | 93.92 | 80.25 | **88.86** | **95.74** | **84.02** |
| Wall | 85.96 | 88.04 | 78.70 | **88.56** | **90.31** | **79.05** |
| Traffic sign | 50.93 | 51.73 | 35.37 | **58.24** | **53.66** | **37.73** |
| Average | 73.44 | 74.68 | 64.52 | **76.81** | **78.63** | **68.16** |

## 7.4. Real Drone Dataset Particularities

- **TUGRAZ** – 400 images of size 6000x4000px, with 24 classes, taken from birdeyes view at altitudes between 5 to 30 meters
- **senseFly University Campus** – 443 images of size 6000x4000px, maximum flight height of 285 meters
- **senseFly Village 1** – 37 images of 4000x3000px, 40 meters
- **senseFly Village 2** – 297 images of 4608x3456px, 162 meters

- **Downsides**:
  - ➢ the images do not contain noise
  - ➢ only one daytime and one weather condition (sunny)
  - ➢ the last 3 sets do not provide ground truth for semantic annotations

# 6. Semantic Segmentation on Images Taken from Drones

## 7.4. Evaluation on the TUGRAZ dataset

| Class | TUGRAZ 600x400 | | | TUGRAZ 1200x800 | | |
|---|---|---|---|---|---|---|
| | Precision | Recall | IoU | Precision | Recall | IoU |
| Unlabeled | 4.44 | 1.73 | 1.26 | **20.52** | **7.70** | **5.93** |
| Paved-area | **94.85** | 91.02 | 86.73 | 94.61 | **93.96** | **89.18** |
| Dirt | 63.47 | 64.21 | 46.88 | **67.87** | **75.18** | **55.45** |
| Grass | 93.71 | 94.09 | 88.49 | **95.78** | **94.88** | **91.07** |
| Gravel | 75.99 | 84.64 | 66.78 | **78.78** | **90.99** | **73.08** |
| Water | 94.41 | **98.17** | 92.78 | **97.84** | 95.41 | **93.44** |
| Rocks | 72.60 | **70.12** | **55.45** | **78.06** | 65.55 | **55.35** |
| Pool | 87.31 | **95.99** | 84.24 | **97.15** | 94.74 | **92.18** |
| Vegetation | **73.96** | 74.96 | 59.30 | **70.02** | **79.32** | 59.21 |
| Roof | 93.99 | **94.30** | **88.93** | **94.04** | 93.18 | 87.99 |
| Wall | 60.75 | **68.93** | 47.69 | **69.80** | 68.26 | **52.70** |
| Window | 73.48 | **71.57** | 56.88 | **80.10** | 68.84 | **58.78** |
| Door | 94.43 | 14.18 | 14.06 | **96.73** | **16.99** | **18.37** |
| Fence | **53.19** | 51.67 | 35.52 | 50.69 | **54.28** | **35.58** |
| Fence-pole | 14.64 | **7.31** | 5.13 | **33.75** | 5.23 | **4.74** |
| Person | 66.54 | 77.31 | 55.67 | **74.81** | **78.79** | **62.27** |
| Dog | 69.75 | 16.95 | 15.79 | **99.12** | **18.87** | **17.53** |
| Car | 93.90 | 90.87 | 85.80 | **96.49** | **96.39** | **87.32** |
| Bicycle | 78.03 | 85.11 | 68.66 | **79.05** | **90.31** | **72.87** |
| Tree | 78.42 | 60.53 | 51.88 | **84.43** | **64.09** | **59.18** |
| Bald-tree | 57.06 | **67.48** | 44.75 | **63.11** | 54.55 | **47.36** |
| AR-marker | 75.05 | 76.99 | 61.30 | **79.94** | **91.79** | **74.61** |
| Obstacle | 62.10 | 67.60 | 47.86 | **66.05** | **68.28** | **50.54** |
| Conflicting | 0 | 0 | 0 | 0 | 0 | 0 |
| Average | 68.00 | 63.57 | 52.58 | **73.70** | **65.32** | **56.03** |

## 7.5. Transitioning from Synthetic to Real Data

- Use a real dataset (TUGRAZ) to fine-tune the results obtained from training ERFNet on the synthetic images from CARLA
- Data variations:
  - Set 1: 800 images, half from CARLA, half from TUGRAZ
  - Set 2: 400 synthetic images, followed by 400 real ones
  - Set 3: 30,000 CARLA images intertwined with TUGRAZ ones
  - Set 4: 30,000 synthetic images, followed by 400 real ones

## 7.5. Evaluation after Fine-tuning

| Class | Set 1 | | | Set 2 | | | Set 3 | | | Set 4 | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | Prec. | Recall | IoU | Prec. | Recall | IoU | Prec. | Recall | IoU | Prec. | Recall | IoU |
| Unlabeled | 83.73 | 79.65 | 68.97 | 85.24 | 77.69 | 68.47 | 85.37 | 67.60 | 60.58 | **93.06** | **92.99** | **89.96** |
| Building | 73.27 | 52.08 | 43.76 | 57.54 | 41.49 | 31.77 | 40.59 | 94.72 | 39.69 | **92.84** | **96.39** | **88.68** |
| Fence | 49.33 | 41.22 | 28.96 | 55.95 | 39.36 | 30.05 | 43.30 | **51.62** | 30.80 | **63.82** | 51.21 | **38.49** |
| Other | 36.86 | 73.14 | 32.46 | 33.62 | **79.36** | 30.92 | 54.68 | 21.41 | 18.19 | **72.49** | 76.11 | **58.78** |
| Pedestrian | 50.83 | 84.69 | 46.56 | 41.12 | 84.79 | 38.30 | 48.93 | **85.05** | 50.15 | **94.01** | 79.51 | **71.56** |
| Pole | 20.11 | 31.10 | 27.91 | 24.95 | 33.71 | 33.12 | 53.07 | **53.36** | **36.25** | **67.01** | 51.82 | 29.78 |
| Road line | 12.96 | 18.65 | 14.84 | 14.01 | 18.39 | 16.76 | 64.60 | **95.67** | 62.76 | **77.98** | 73.92 | **74.23** |
| Road | 73.40 | 96.02 | 71.23 | 72.02 | 94.10 | 68.91 | 97.57 | 55.87 | 55.10 | **99.08** | **96.03** | **96.86** |
| Sidewalk | 38.52 | 43.83 | 40.98 | 40.35 | 48.52 | 44.76 | 52.70 | **97.51** | 52.00 | **90.45** | 89.74 | **88.91** |
| Vegetation | 70.78 | 82.35 | 61.46 | 67.99 | 77.86 | 56.97 | 56.44 | 61.00 | 41.47 | **84.18** | **85.35** | **73.34** |
| Car | 79.61 | 72.19 | 60.93 | 81.88 | 34.86 | 32.36 | 69.44 | 78.49 | 58.34 | **89.38** | **83.64** | **84.78** |
| Wall | 27.80 | 49.24 | 21.61 | 21.99 | 46.38 | 17.53 | **80.52** | 46.87 | 42.10 | 76.29 | **72.67** | **68.63** |
| Traffic Sign | 1.63 | 3.10 | 2.63 | 1.85 | 3.87 | 2.98 | **74.38** | 38.75 | 34.18 | 55.87 | **46.41** | **34.83** |
| Average | 47.60 | 55.94 | 40.18 | 46.04 | 52.34 | 36.38 | 63.20 | 65.22 | 44.74 | **81.27** | **76.60** | **69.14** |

# 6. Semantic Segmentation on Images Taken from Drones

**7.6. Creation of a Representative Dataset**

- Merge the two types of sets
- Established 17 representative classes that solve the problems noticed in the previous sections
- Manual re-annotation of 179 virtual and 64 real images
- Scenarios:
  - ➢ Case 1: trained and validated on CARLA
  - ➢ Case 2: trained on CARLA, validated on real dataset
  - ➢ Case 3: trained and validated on real dataset
  - ➢ Case 4: trained on merged dataset, validated on CARLA
  - ➢ Case 5: trained on merged dataset, validated on real dataset
  - ➢ Case 6: trained and validated on merged dataset

# 6. Semantic Segmentation on Images Taken from Drones

| CARLA | | TUGRAZ | | MERGED DATASET | |
|---|---|---|---|---|---|
| (0,0,0) | Unlabeled | (112,150,146) | AR-marker | (0,0,0) | Unlabeled |
| (70,70,70) | Building | (190,250,190) | Bald tree | (70,70,70) | Building |
| (190,153,153) | Fence | (119,11,32) | Bicycle | (190,153,153) | Fence |
| (250,170,160) | Other | (9,143,150) | Car | (250,170,160) | Other |
| (220,20,60) | Pedestrian | (255,0,0) | Conflicting | (220,20,60) | Pedestrian |
| (153,153,153) | Pole | (130,76,0) | Dirt | (153,153,153) | Pole |
| (157,234,50) | Road line | (102,51,0) | Dog | (157,234,50) | Road line |
| (128,64,128) | Road | (254,148,12) | Door | (128,64,128) | Road |
| (244,35,232) | Sidewalk | (190,153,153) | Fence | (244,35,232) | Sidewalk |
| (107,142,35) | Vegetation | (153,153,153) | Fence-pole | (107,142,35) | Vegetation |
| (0,0,142) | Car | (0,102,0) | Grass | (152,251,152) | Terrain |
| (102,102,156) | Wall | (112,103,87) | Gravel | (0,0,142) | Car |
| (220,220,0) | Traffic sign | (2,135,115) | Obstacle | (102,102,156) | Wall |
| | | (128,64,128) | Paved area | (220,220,0) | Traffic sign |
| | | (255,22,96) | Person | (32,95,255) | Water |
| | | (0,50,89) | Pool | (255,0,0) | Rider |
| | | (48,41,30) | Rocks | (119,11,32) | Bike |
| | | (70,70,70) | Roof | | |
| | | (51,51,0) | Tree | | |
| | | (0,0,0) | Unlabeled | | |
| | | (107,142,35) | Vegetation | | |
| | | (102,102,156) | Wall | | |
| | | (28,42,168) | Water | | |
| | | (254,228,12) | Window | | |

## 7.6. Evaluation on the Merged Dataset

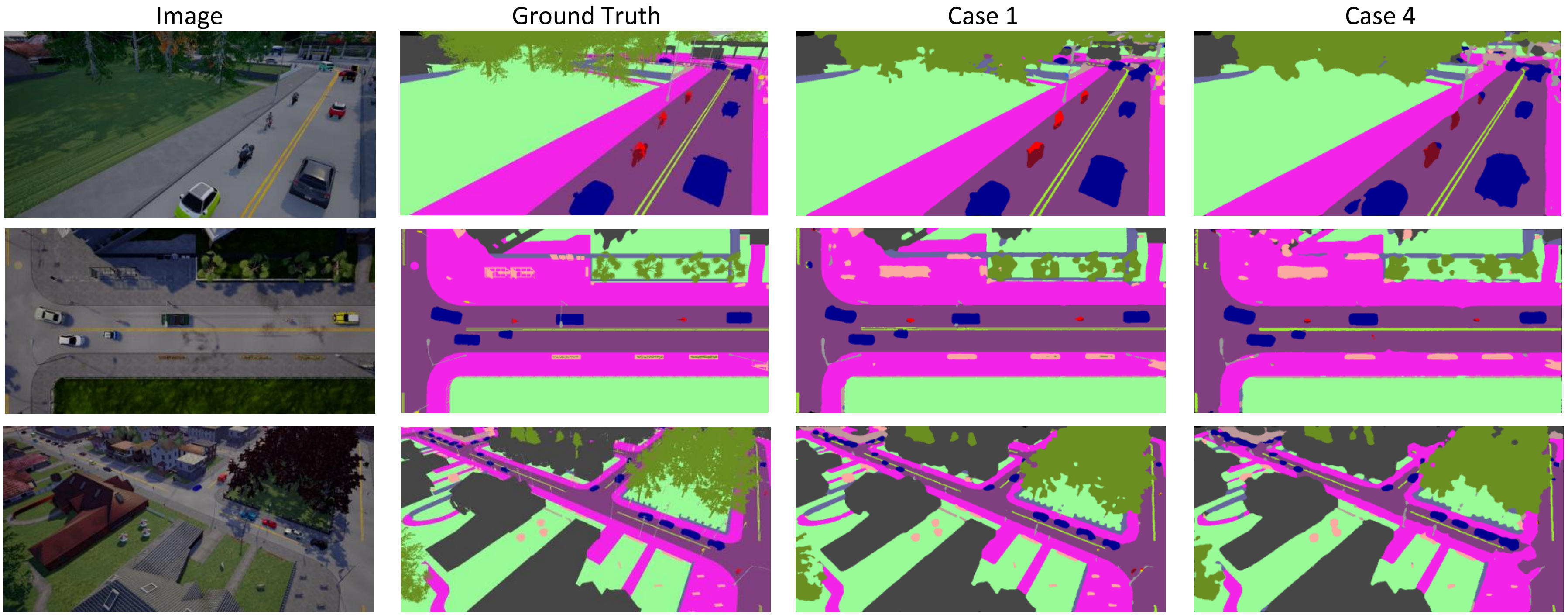
IoU results

| Class | Case 1 | Case 2 | Case 3 | Case 4 | Case 5 | Case 6 |
|---|---|---|---|---|---|---|
| Unlabeled | **76.76** | 2.28 | 30.66 | 71.77 | 29.39 | 54.78 |
| Building | 91.64 | 17.33 | 89.00 | 91.33 | **94.15** | 91.79 |
| Fence | 49.05 | 0.67 | **63.62** | 45.46 | 58.62 | 51.39 |
| Other | 68.18 | 3.85 | 66.08 | 64.99 | **70.12** | 67.41 |
| Pedestrian | 0.00 | 0.00 | 65.01 | 26.34 | 64.87 | **65.03** |
| Pole | **39.58** | 2.45 | 0.00 | 36.10 | 31.78 | 36.60 |
| Road line | 66.37 | 13.96 | 30.98 | **66.73** | 41.57 | 64.07 |
| Road | **93.55** | 24.27 | 91.06 | 92.90 | 92.54 | 92.75 |
| Sidewalk | **90.25** | 11.87 | 81.01 | 89.21 | 82.00 | 87.96 |
| Vegetation | 77.26 | 29.63 | 78.25 | 75.68 | **81.16** | 77.14 |
| Terrain | **88.15** | 19.20 | 83.45 | 87.40 | 84.23 | 86.24 |
| Car | 85.68 | 18.89 | 83.76 | 79.51 | **88.33** | 81.52 |
| Wall | **68.77** | 3.92 | 61.33 | 67.41 | 60.63 | 65.51 |
| Traffic sign | **39.63** | 0.00 | 0.00 | 33.48 | 0.00 | 32.00 |
| Water | **90.85** | 0.00 | 90.28 | 87.73 | 88.63 | 88.01 |
| Rider | 30.15 | 0.00 | 25.95 | 21.49 | **32.94** | 26.85 |
| Bicycle | 42.63 | 0.65 | 50.02 | 40.31 | **52.41** | 50.43 |
| Average | 64.62 | 8.76 | 58.28 | 63.40 | 61.96 | **65.85** |

# 6. Semantic Segmentation on Images Taken from Drones

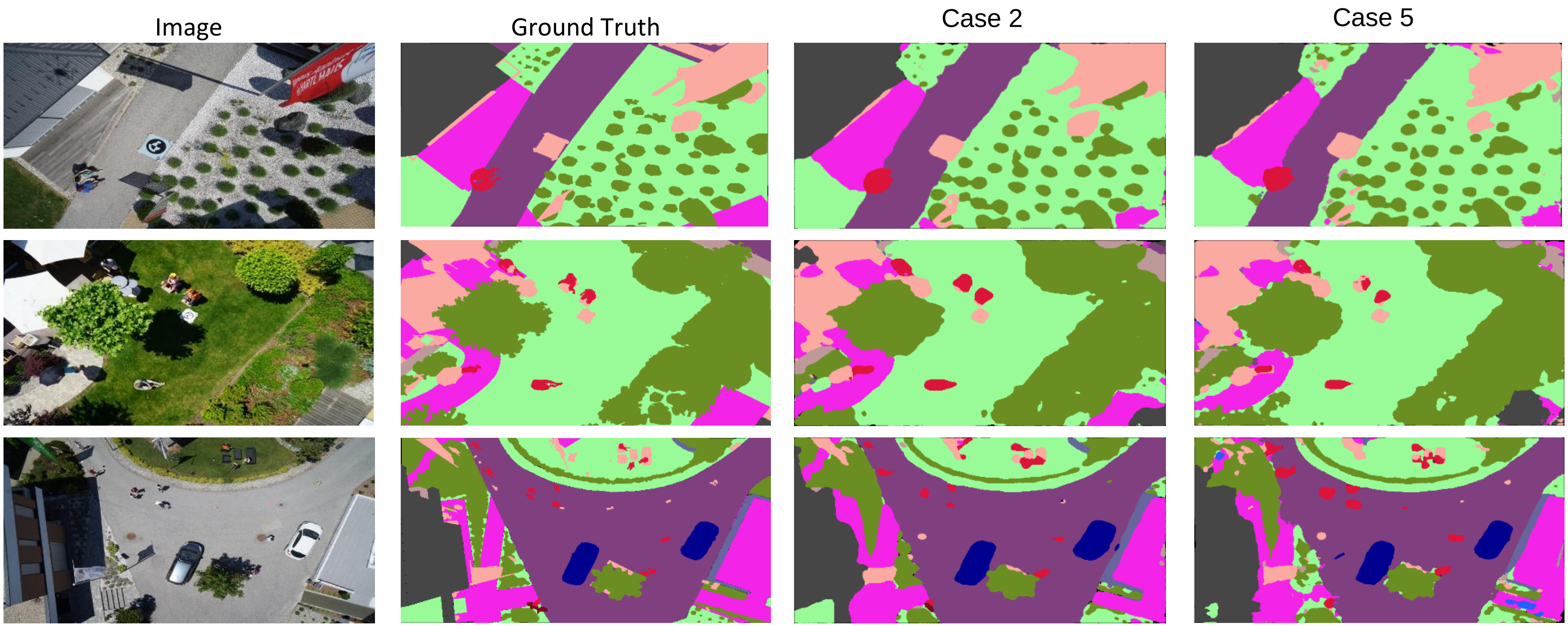
## 7.7. Qualitative Evaluation

ERFNet results on the synthetic dataset



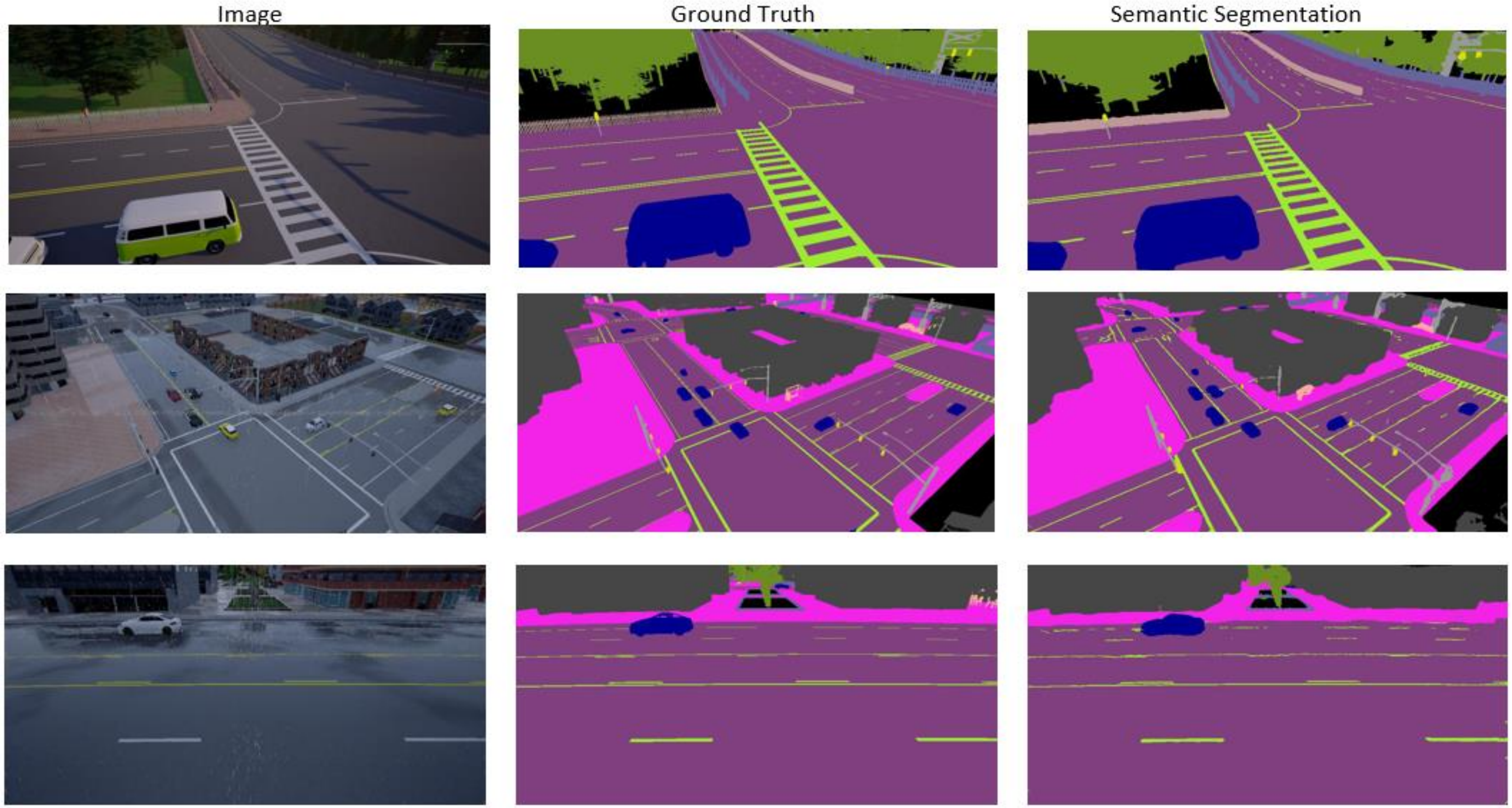
## 7.7. Qualitative Evaluation

ERFNet results on the real dataset



## 7.7. Qualitative Evaluation

Examples of road markings detection



# 7. Conclusions

- We studied the **state-of-the-art research** in the domain of computer vision for autonomous navigation perception tasks, highlighting the problem of **semantic segmentation** applied on 2D or 3D data.
- Because synthetic data was successfully used to improve the accuracy of detection systems, we performed a **survey** exploring the existing **simulators** and **synthetic datasets**.
- We propose an extension to the CARLA simulator by adding a **drone aerial camera**.
- We employed a **methodology for training and testing deep learning algorithms** on different types of inputs.

**Best results are obtained when the network is trained first on a large synthetic dataset and then fine-tuned with real data.**

# 7. Conclusions

- Future work:
  - use GANs for style transfer
  - improve the CARLA simulator by adding pedestrians and semantic class textures (terrain, rider, water, etc.)
  - employ GRUs and Spatial Transformers to propagate the semantic information from past frames to future ones

# 8. Bibliography


- Research was done as part of the program „**SEPCA - Perceptie Vizuala Semantica si Control Integrat pentru Sisteme Autonome**”



1. Bianca-Cerasela-Zelia Blaga and Sergiu Nedevschi, A Method for Automatic Pole Detection from Urban Video Scenes using Stereo Vision. In 2018 IEEE 14th International Conference on Intelligent Computer Communication and Processing (ICCP) (pp. 293-300).
2. E. Romera, J. M. Alvarez, L. M. Bergasa, and R. Arroyo, Erfnet: Efficient residual factorized convnet for real-time semantic segmentation, IEEE Transactions on Intelligent Transportation Systems, vol. 19, no. 1, pp. 263-272, 2017.
3. R. Restrepo. (2018) Erfnet semantic segmentation architecture in tensorflow. [Online]. Available: https://github.com/ronrest/erfnet_segmentation
4. U. Challita, A. Ferdowsi, M. Chen, and W. Saad, Machine learning for wireless connectivity and security of cellular-connected uavs, IEEE Wireless Communications, vol. 26, no. 1, pp. 28-35, 2019.
5. J. Li, H. Cheng, H. Guo, and S. Qiu, Survey on artificial intelligence for vehicles,Automotive Innovation, vol. 1, no. 1, pp. 2-14, 2018.
6. W. Geekly. (2019) A short course of machine learning or how to create a neural network to solve the scoring problem. [Online]. Available: https://weekly-geekly.github.io/articles/340792/index.html
7. D. Nilsson and C. Sminchisescu, Semantic video segmentation by gated recurrent ow propagation, in Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2018, pp. 6819{6828.
8. J. Laurmaa, A deep learning model for scene segmentation of images captured by drones, Ph.D. dissertation, Master's thesis, EPFL, Switzerland, 2016.
9. Y. Lyu, G. Vosselman, G. Xia, A. Yilmaz, and M. Y. Yang, The uavid dataset for video semantic segmentation, arXiv preprint arXiv:1810.10438, 2018.