Semantic Segmentation Using RGB Aerial Images

Perception and Learning for Robotics

Hrishikesh Gupta

Supervised by: Timo Hinzmann

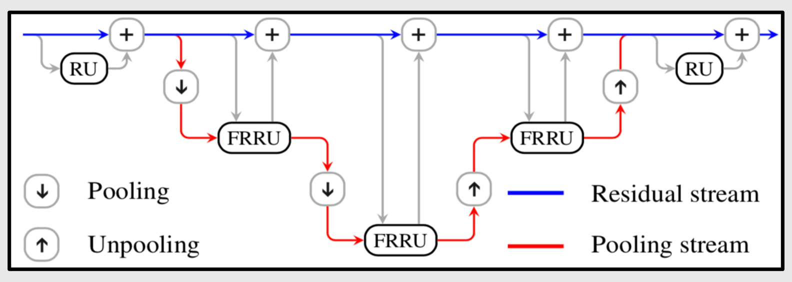
Autonomous Systems Lab

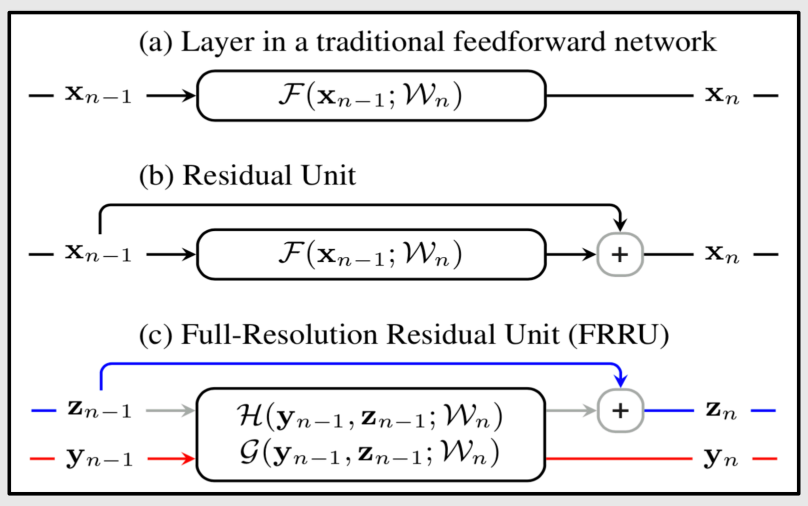
Problem Definition and Motivation

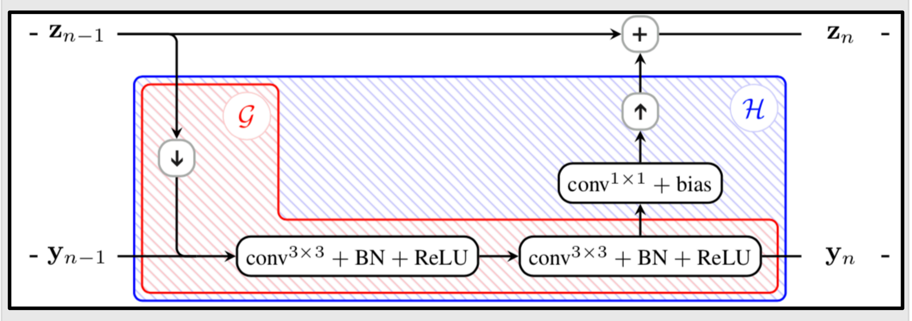
- Semantic Segmentation of Aerial Images for UAVs.
- Application for suitable landing spot for emergency landing of UAVs, Semantic localization and Autonomous landing site selection.
- Development of efficient solution for deployment on UAV hardware.
- CNN based pixel-level segmentation and classification model.

Network Architecture

- Thorough literature search was done, Many architectures were considered and narrowed down to one.
- Table 1 shows State-of-Art semantic segmentation methods, and its performance compared to chosen method FRRN.
- **Full Resolution Residual Network**: New novel architecture, Combines strong recognition performance and precise localization capabilities.
- **Two Streams**: Sequence of pooling operations → Largescale understanding of image elements. Feature maps at full image resolution → Precise boundary adherence.







- Comparison of FRRN to other networks:

Method	Subsample	Coarse	Pretrained	Mean	Road	Sidewalk	Building	Wall	Fence	Pole	Traf. Light	Traf. Sign	Vegetation	Terrain	Sky	Person	Rider	Car	Truck	Bus	Train	Motorcycle	Bicycle
SegNet [2]	×4		✓	57.0	96.4	73.2	84.0	28.5	29.0	35.7	39.8	45.2	87.0	63.8	91.8	62.8	42.8	89.3	38.1	43.1	44.2	35.8	51.9
FRRN A	×4			63.0	97.6	79.1	88.3	32.0	36.4	51.7	57.1	62.5	90.9	69.5	93.3	75.2	51.3	91.6	30.2	43.1	39.2	46.0	62.6
ENet [44]	×2			58.3	96.3	74.2	85.0	32.2	33.2	43.5	34.1	44.0	88.6	61.4	90.6	65.5	38.4	90.6	36.9	50.5	48.1	38.8	55.4
DeepLab [43]	×2	✓	✓	64.8	97.4	78.3	88.1	***47.5***	44.2	29.5	44.4	55.4	89.4	67.3	92.8	71.0	49.3	91.4	***55.9***	66.6	**56.7**	48.1	58.1
FRRN B	×2			**71.8**	**98.2**	**83.3**	**91.6**	45.8	***51.1***	**62.2**	**69.4**	***72.4***	**92.6**	70.0	***94.9***	**81.6**	**62.7**	**94.6**	49.1	***67.1***	***55.3***	53.5	69.5
Dilation [55]	×1		✓	67.1	97.6	79.2	89.9	37.3	47.6	53.2	58.6	65.2	91.8	69.4	93.7	78.9	55.0	93.3	45.5	53.4	47.7	52.2	66.0
Adelaide [34]	×1		✓	***71.6***	***98.0***	***82.6***	90.6	44.0	50.7	51.1	65.0	71.7	92.0	**72.0**	94.1	***81.5***	***61.1***	***94.3***	**61.1**	65.1	53.8	**61.6**	**70.6**
LRR [20]	×1		✓	69.7	97.7	79.9	90.7	44.4	48.6	58.6	***68.2***	72.0	***92.5***	69.3	94.7	**81.6**	60.0	94.0	43.6	56.8	47.2	54.8	***69.7***
LRR [20]	×1	✓	✓	**71.8**	97.9	81.5	***91.4***	**50.5**	**52.7**	***59.4***	66.8	**72.7**	***92.5***	***70.1***	**95.0**	81.3	60.1	***94.3***	51.2	**67.7**	54.6	***55.6***	69.6

Table 1

Dataset Collection

- **Microsoft AirSim** and ICGV TU **Graz Semantic Drone Dataset**.

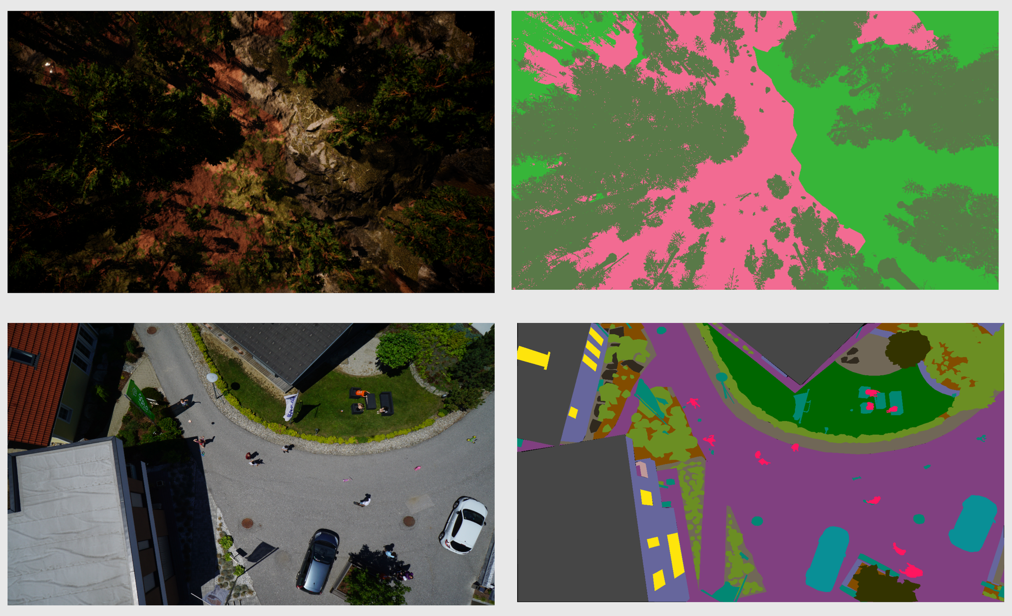
- **Semantic Drone Dataset**: 400 publicly available images, 5-30 meters above ground, Urban scenes.
- **Microsoft AirSim** : Forest and Mountainous Terrain, 300 images each, Simulated drone images.

Train	Test	Validation	Total Labels
4240	480	480	27

- **Segmentation Classes/Labels:**

Background	Ar Marker
Bicycle	Bald-Tree
Car	Conflicting
Dirt	Dog
Door	Fence
Pole-Fence	Grass
Gravel	Landscape
Obstacle	Paved Area
Person	Pool
Rocks	Roof
Tree	Forest Tree
Vegetation	Wall
Water	Windmill
Window	

- **Dataset Examples** :



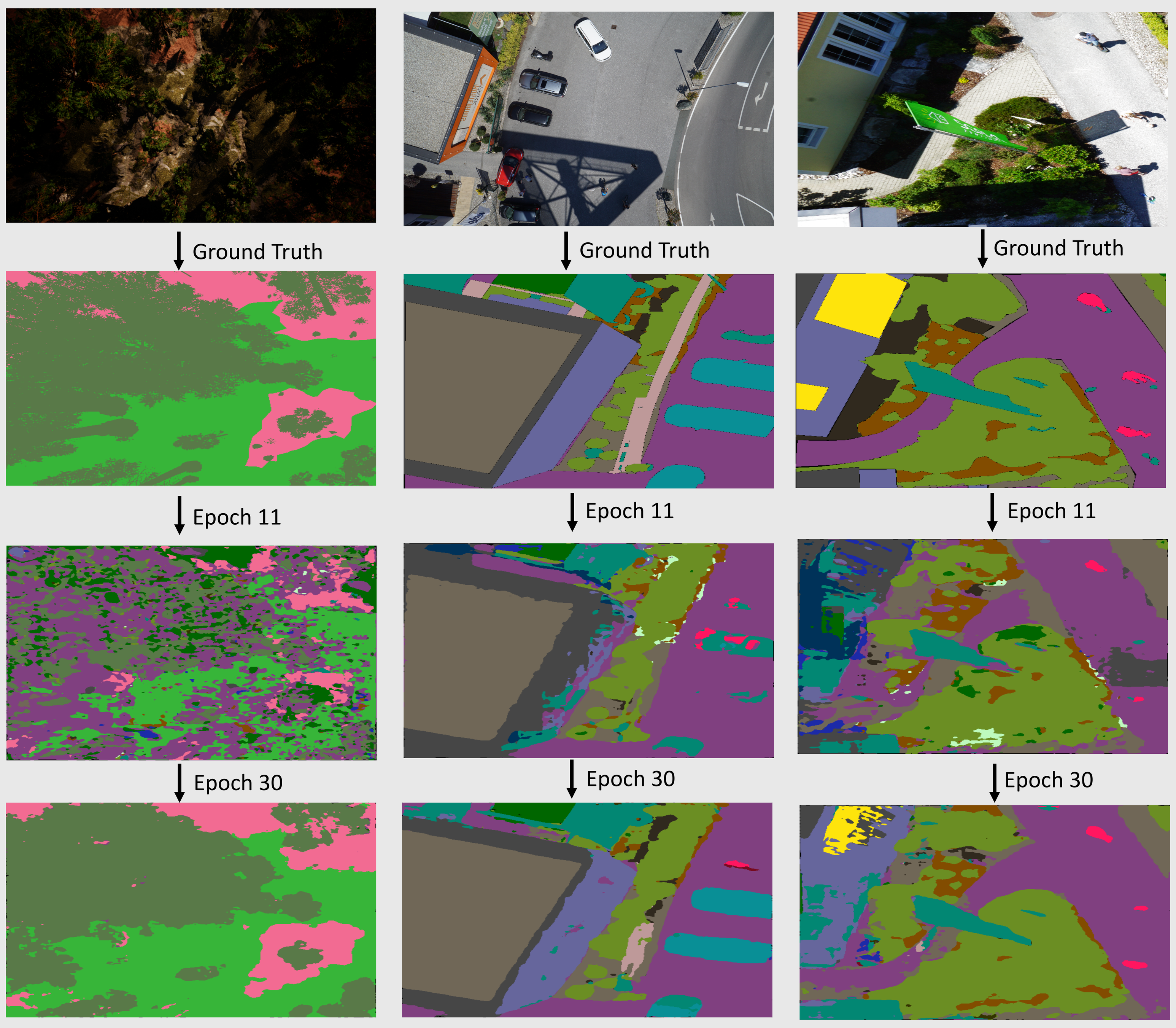
Visual Results for Different Epochs

- Per Class Accuracy for Test set [30 Epochs]:

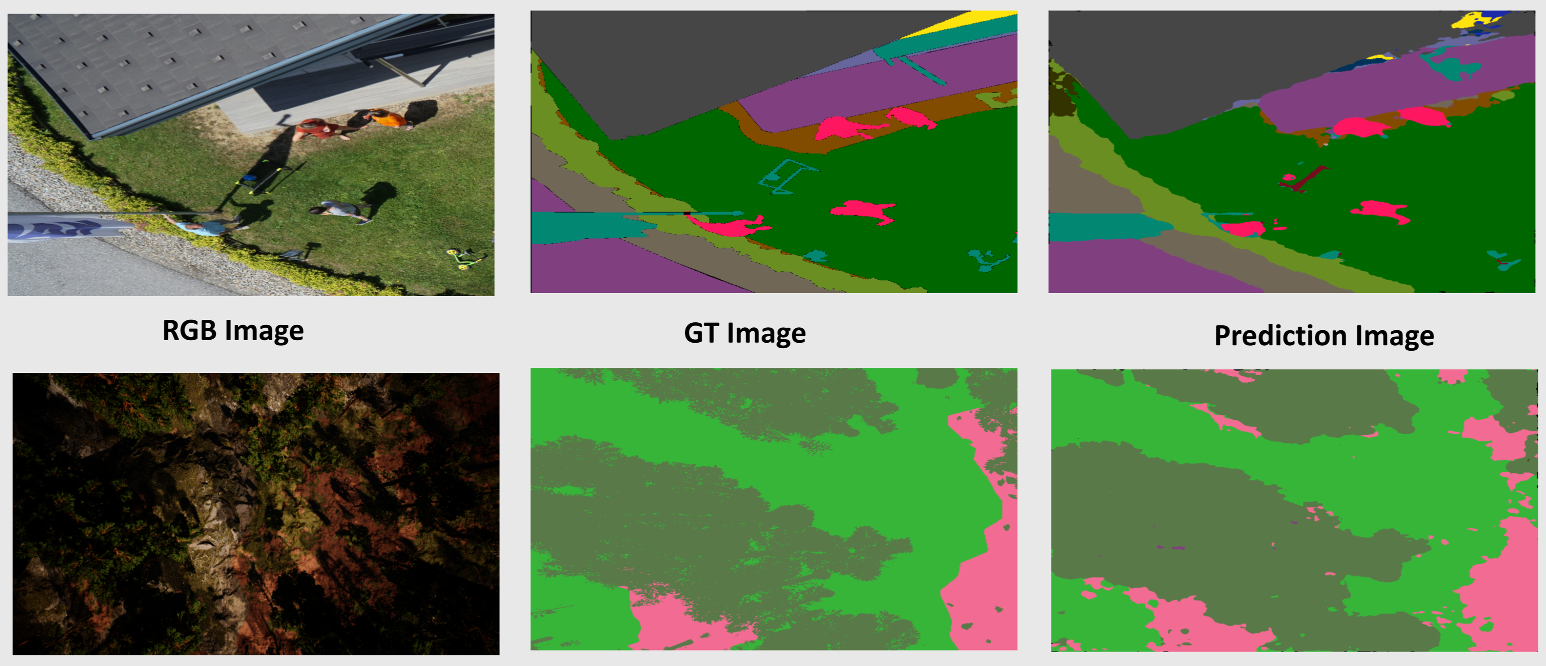
Avg. Test Accuracy	0.7463	Avg. F1 Score	0.7396	Avg. Mean IoU	0.3851	Avg. Pixel Accuracy	0.7477

Unlabeled	0.1358	Wall	0.7453	Bald-Tree	0.9498
Paved-area	0.9290	Window	0.8503	Ar-Marker	0.9106
Dirt	0.6766	Door	0.9666	Obstacle	0.5594
Grass	0.8995	Fence	0.7600	Conflicting	1.0000
Gravel	0.8460	Fence-Pole	0.8259	Windmill	0.9796
Water	0.9567	Person	0.7830	Landscape	0.9297
Rocks	0.8281	Dog	0.9944	Tree	0.9896
Pool	0.9891	Car	0.9803	Background	0.9858
Vegetation	0.8198	Bicycle	0.8861		
Roof	0.9225	Tree	0.8297		

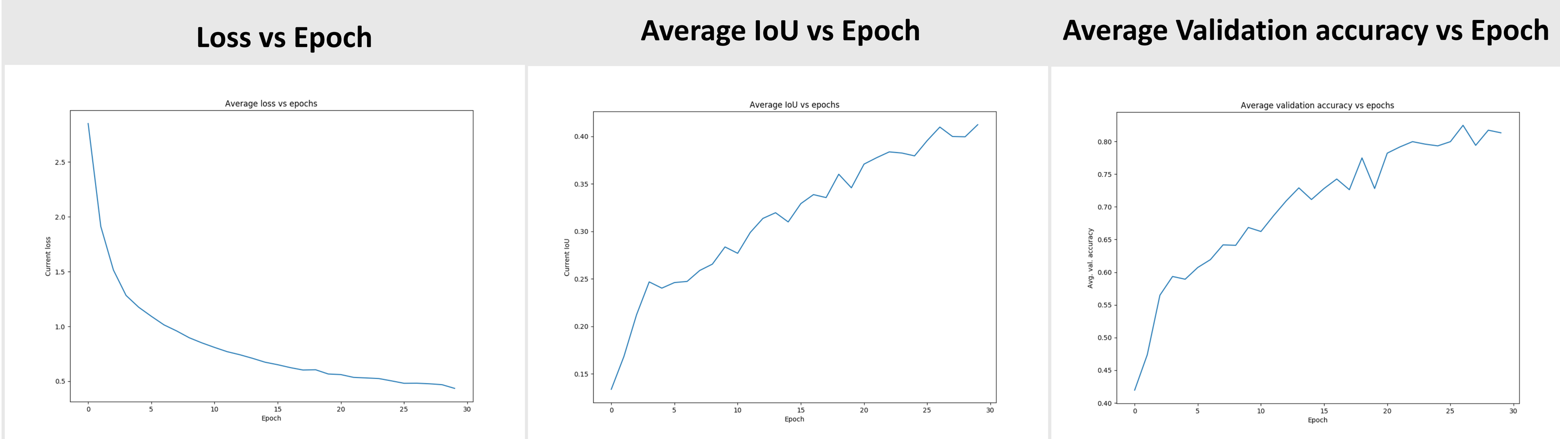
Visual Results for Different Epochs



- Test set Visual Results [19 Epochs]:



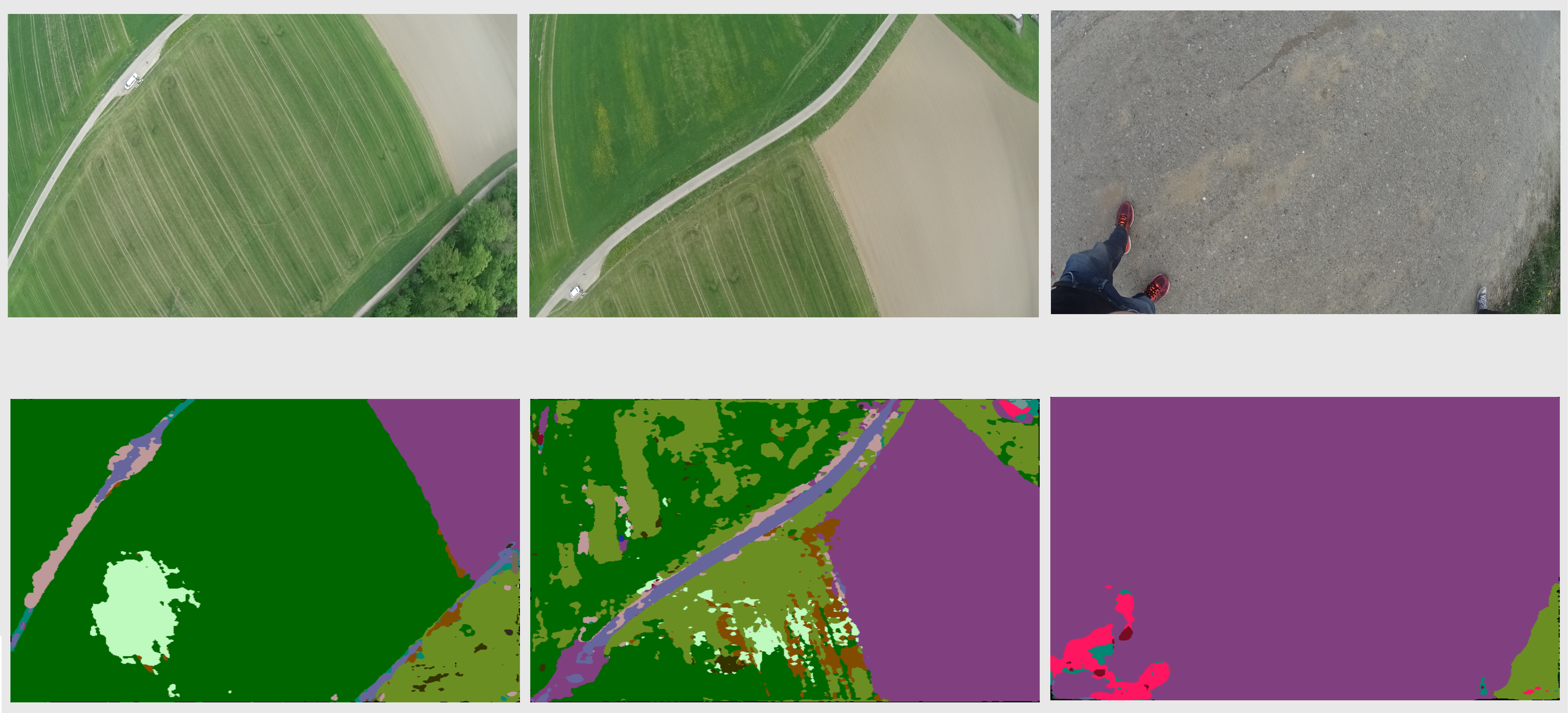
Loss, Accuracy, Evaluation Plots



- Plots for Training data, 30 epochs → 135066 sec.

Visual Prediction for ASL Drone Dataset

- Inference using trained model on the dataset of Real ASL UAV.



References

- J. Long, E. Shelhamer and T. Darrell. Fully Convolutional Networks for Semantic Segmentation. In CVPR, 2015.
- T. Pohlen, A. Hermans, M. Mathias and B. Leibe. Full-Resolution Residual Networks for Semantic Segmentation in Street Scenes. In CVPR, 2016.