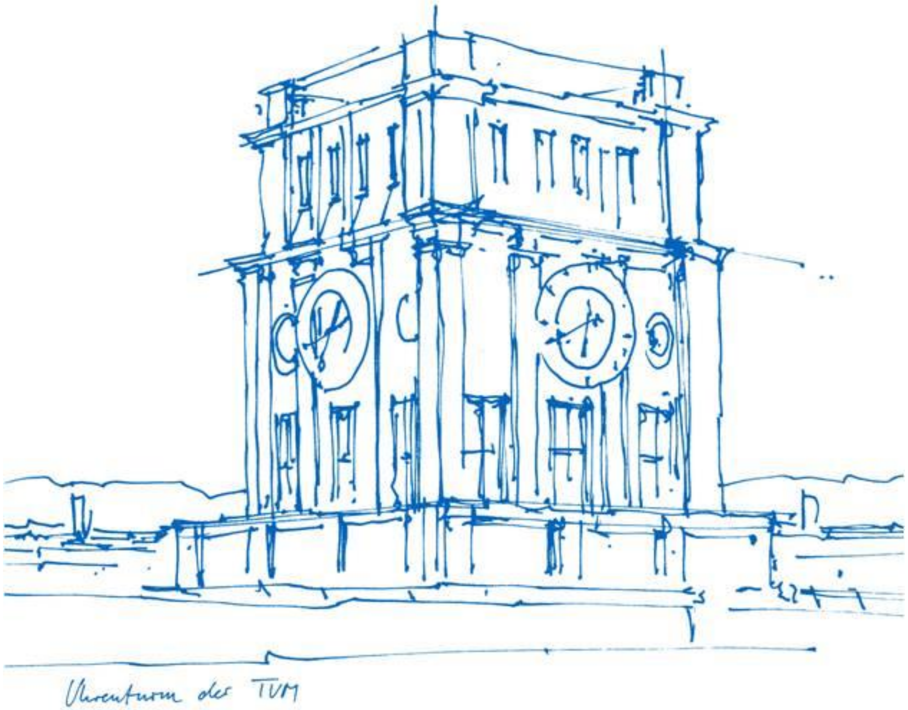
ClusterNet: unsupervised generic feature learning for fast interactive satellite image segmentation

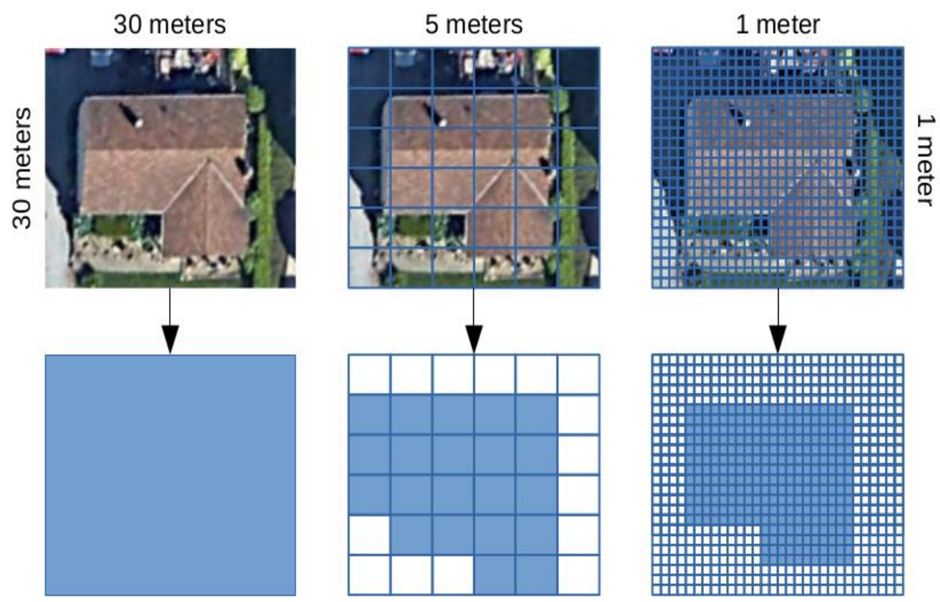


Andrii Zhygallo
September 10, 2019

Agenda

1. Introduction
2. Related Work
3. Proposed Method
4. Future Work

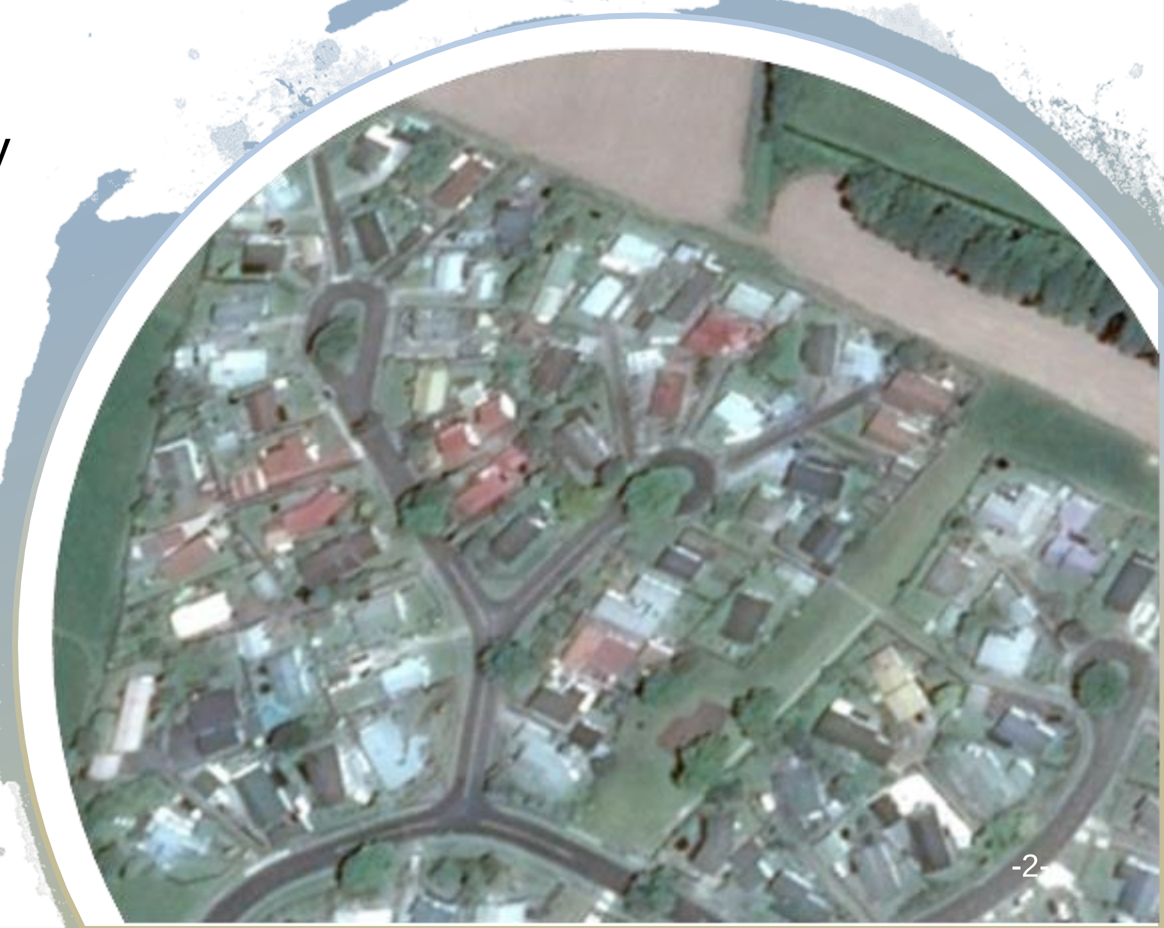
Introduction
Remote Sensing Data



- Aerial Images
  - high spatial resolution
  - expensive; a little of data
- **Satellite Images**
  - low spatial resolution
  - Cheap; a lot of data

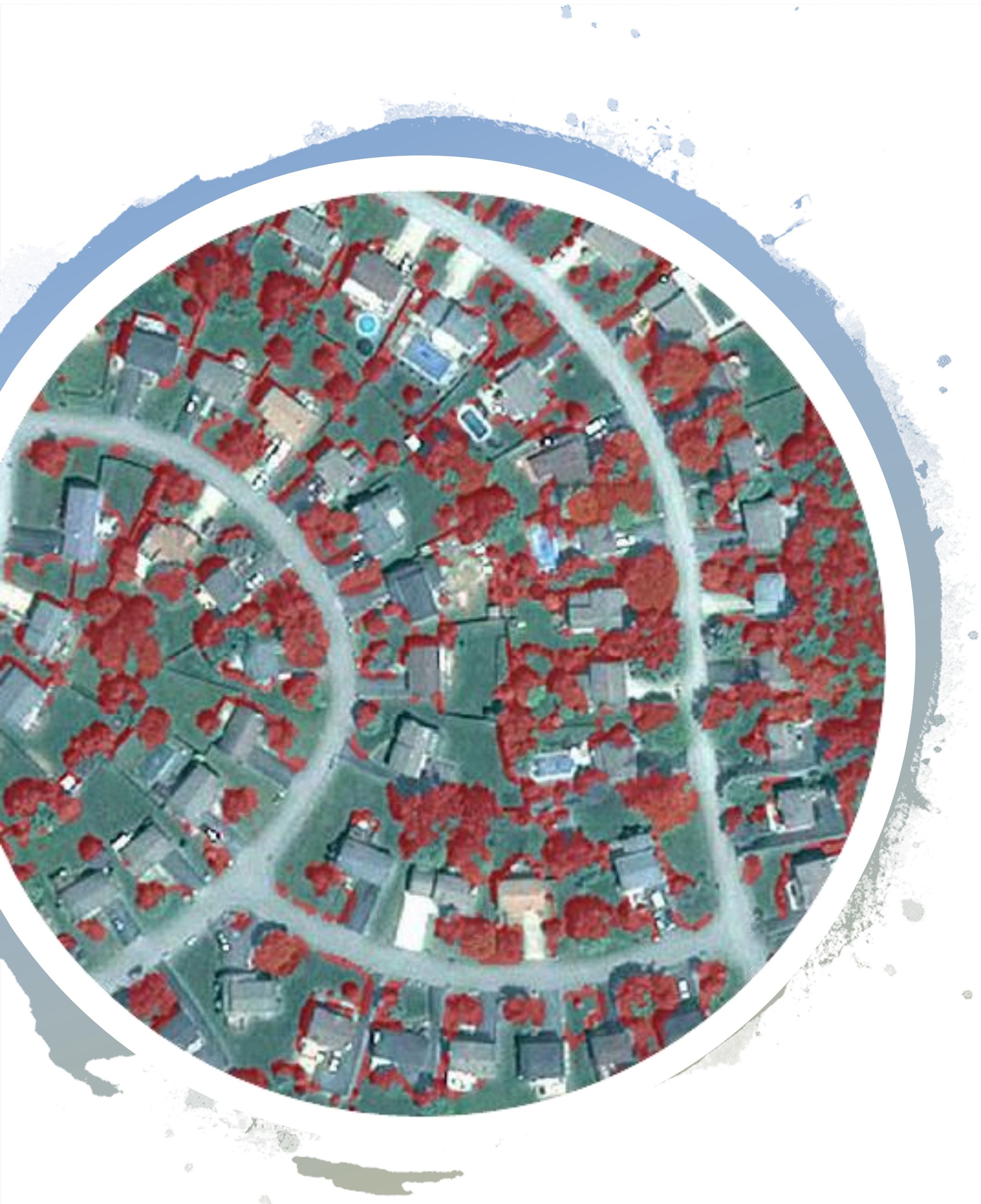
Introduction
Application

- Disaster Monitoring
- Estimating Urban Areas
- Autonomous Driving
- Agriculture planning
- Optimizing solar panel energy
- Others



Introduction
Task

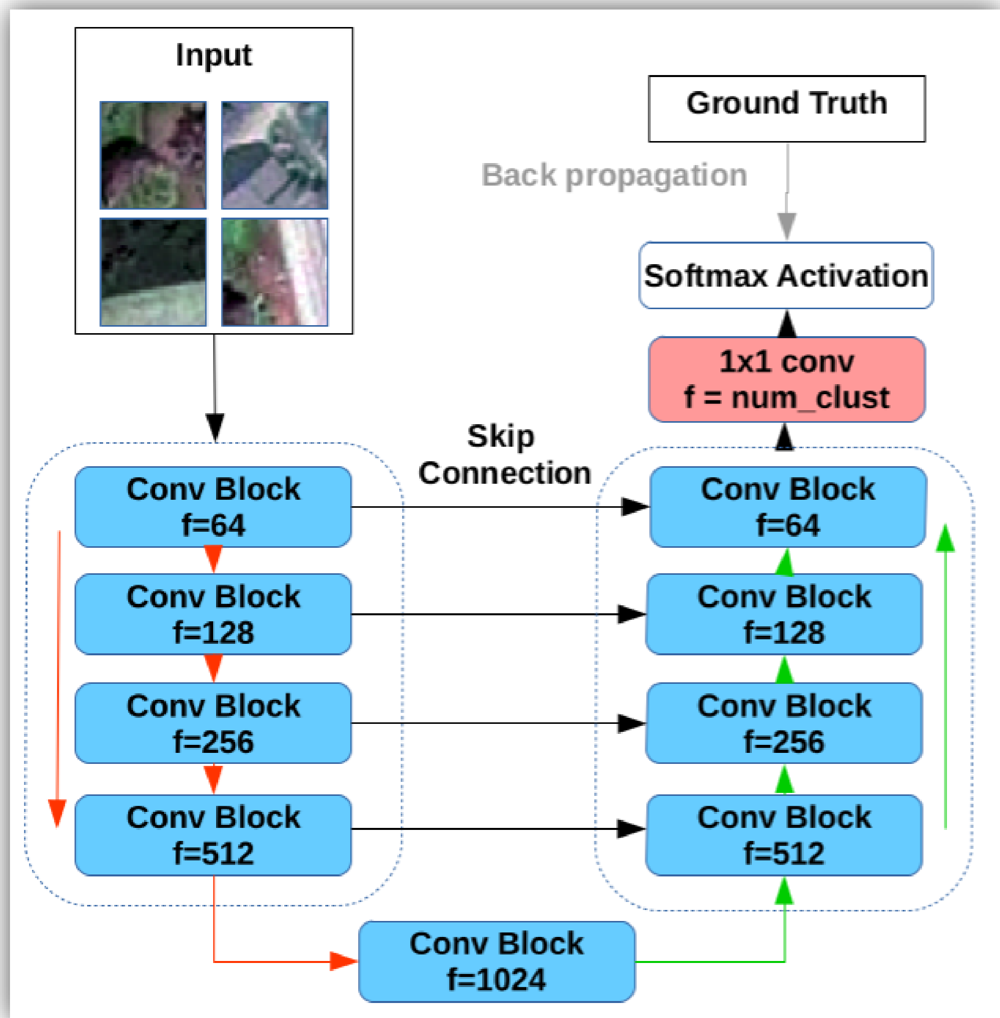
Automatically segment objects of interest (in our case trees) on satellite images preserving their geometric structure.



Introduction
Challenges

- Complicated Objects
- Low resolution
- Difference in data distribution
- **No ground-truth**
- **Bad ground-truth**

Related Work
Supervised U-Net*



Advantages:

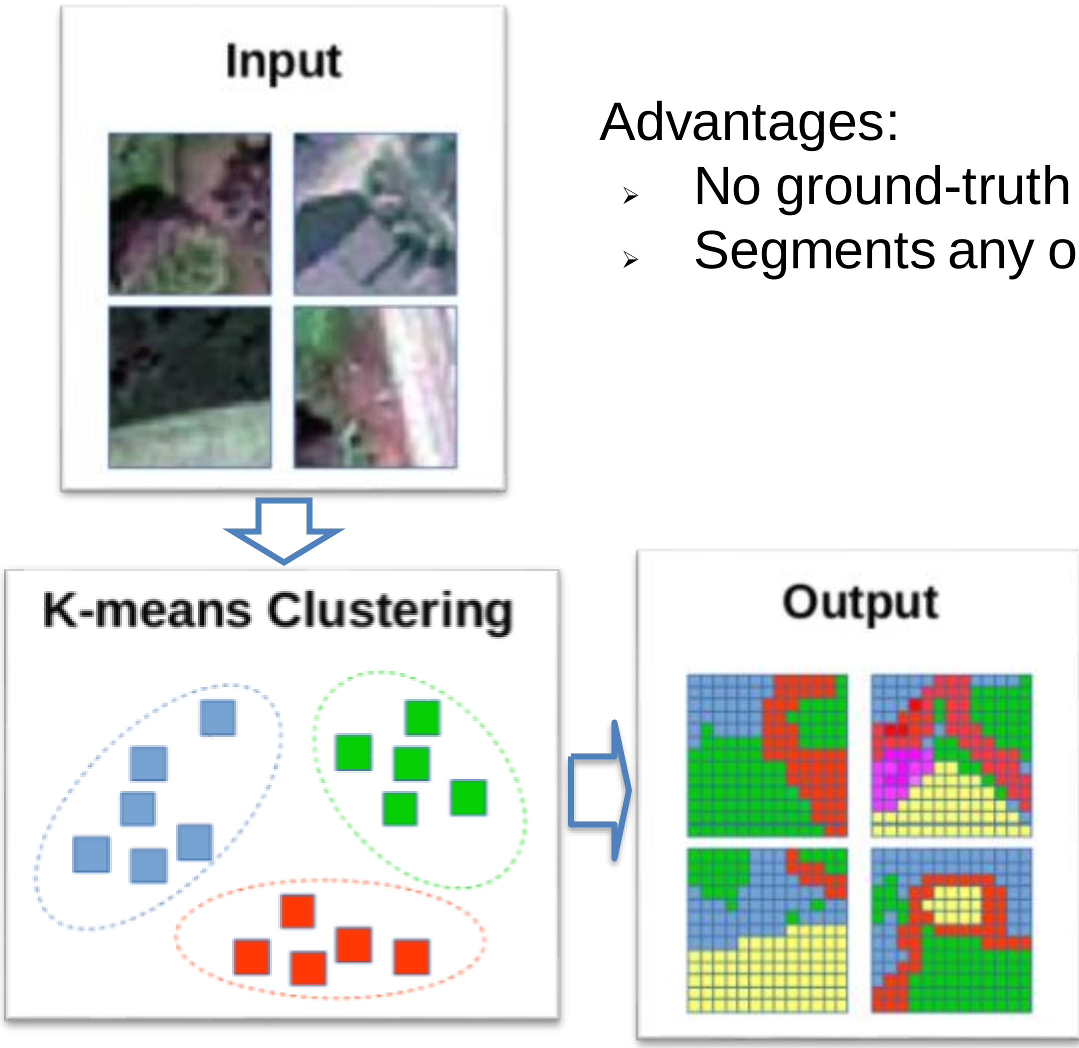
- State of the art performance

Disadvantages:

- Requires a lot of ground-truth
- Segments only the objects labeled by the ground-truth

* Olaf Ronneberger, Philipp Fischer, Thomas Brox. "U-Net: Convolutional Net-works for Biomedical Image Segmentation." In: arXiv:1505.04597 (2015).

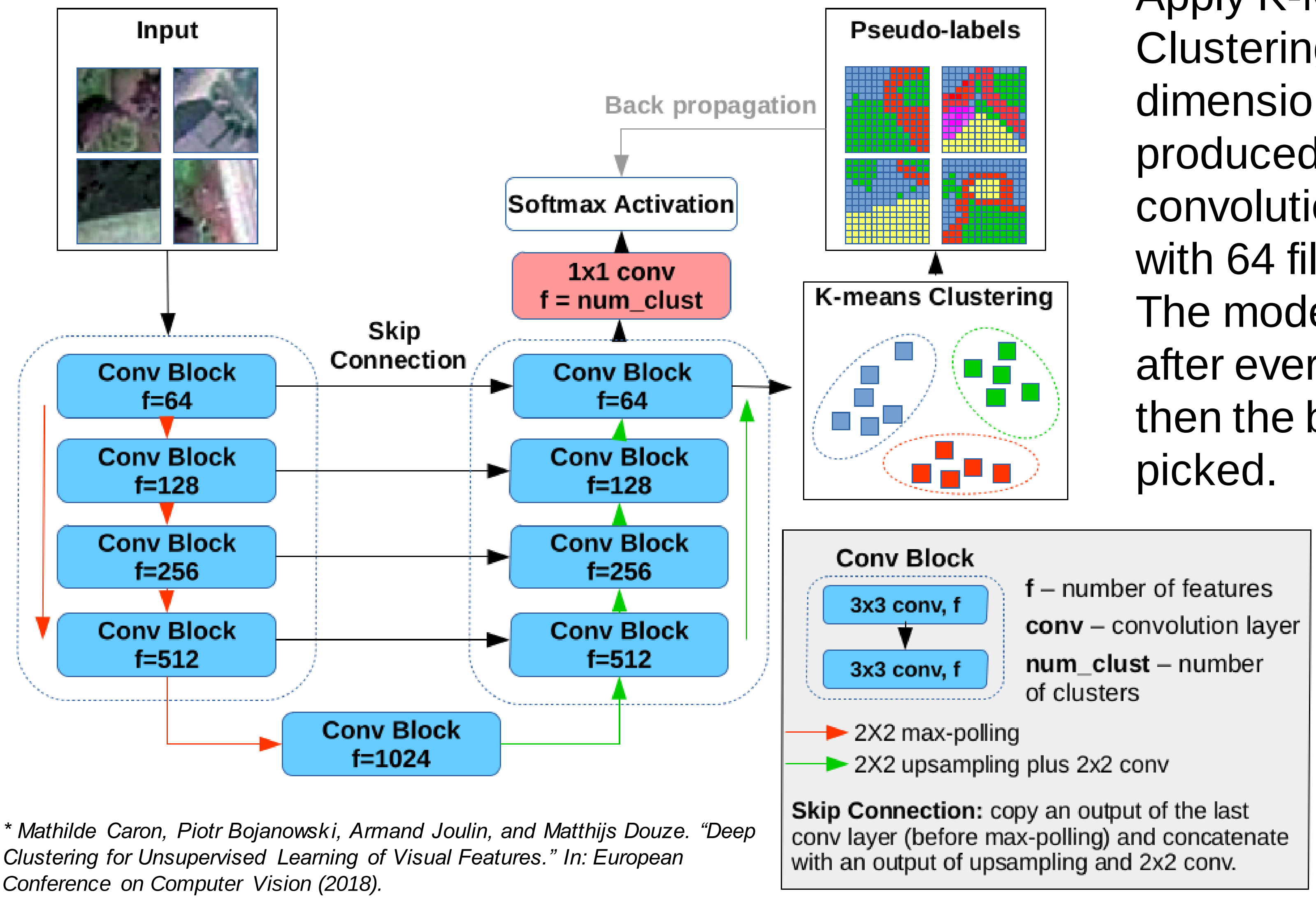
Related Work
Unsupervised K-Means



Advantages:
- No ground-truth required
- Segments any object

Disadvantages:
- Low performance
- Confuses different objects similar in color (trees – grass)
- Doesn’t account the geometric structure of an object
- Object class is unknown

Proposed Method
ClusterNet

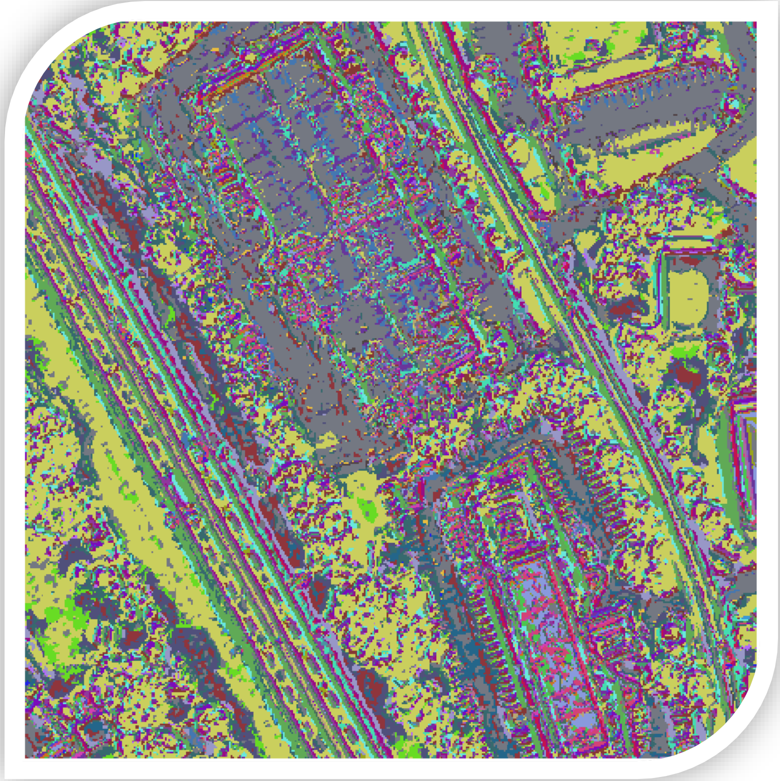


Apply K-Means Clustering to the high-dimensional features produced by the last convolutional layer with 64 filters.
The model is saved after every epoch, then the best one is picked.

* *Mathilde Caron, Piotr Bojanowski, Armand Joulin, and Matthijs Douze. "Deep Clustering for Unsupervised Learning of Visual Features." In: European Conference on Computer Vision (2018).*

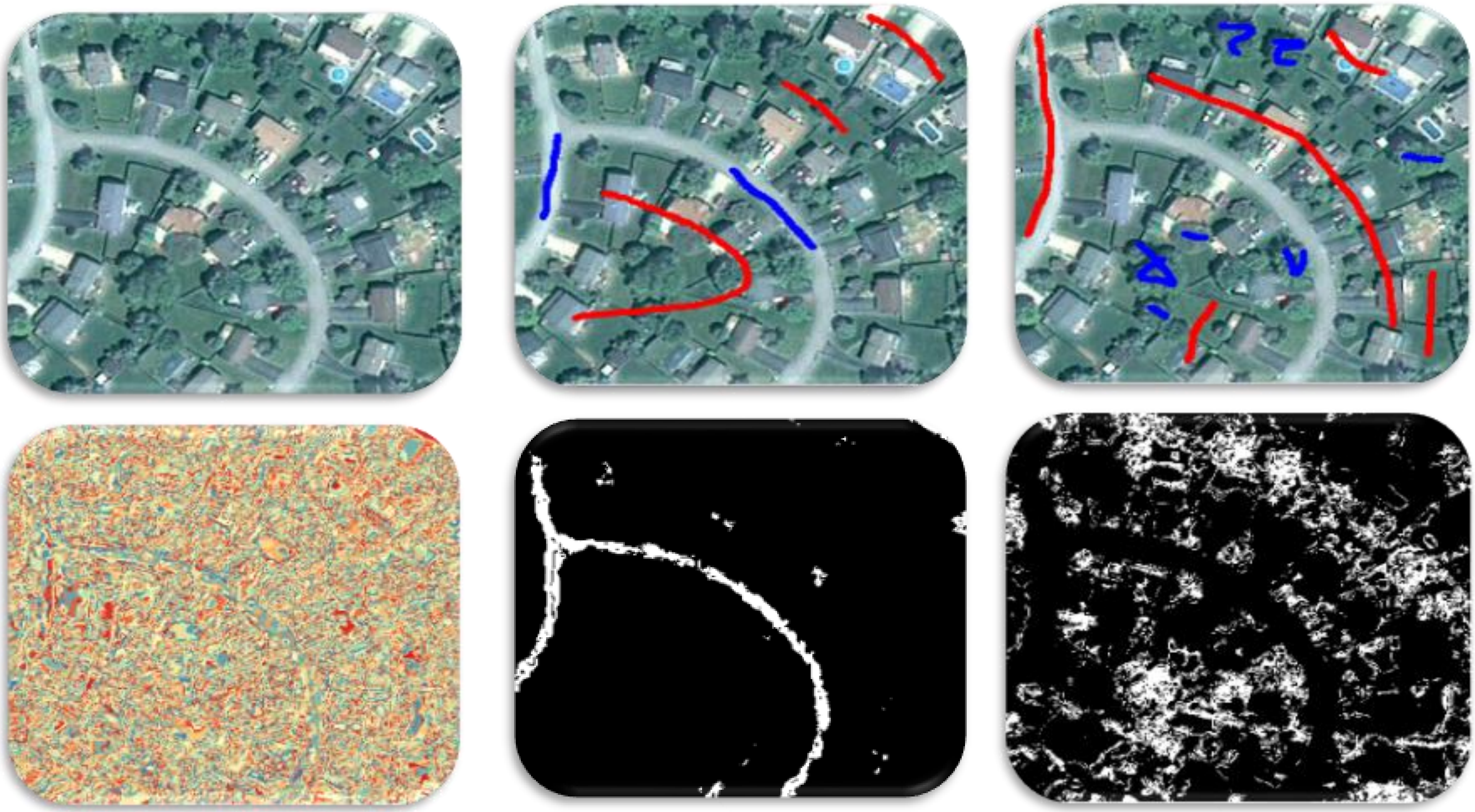
Proposed Method
Output

The output is a mask where every pixel is assigned to one of the possible clusters (a hyper-parameter chosen by user)



Proposed Method
Classes from Clusters

Draw Positive/Negative Strokes to identify foreground and background.
Clusters which overlap will be splited which makes the process interative.



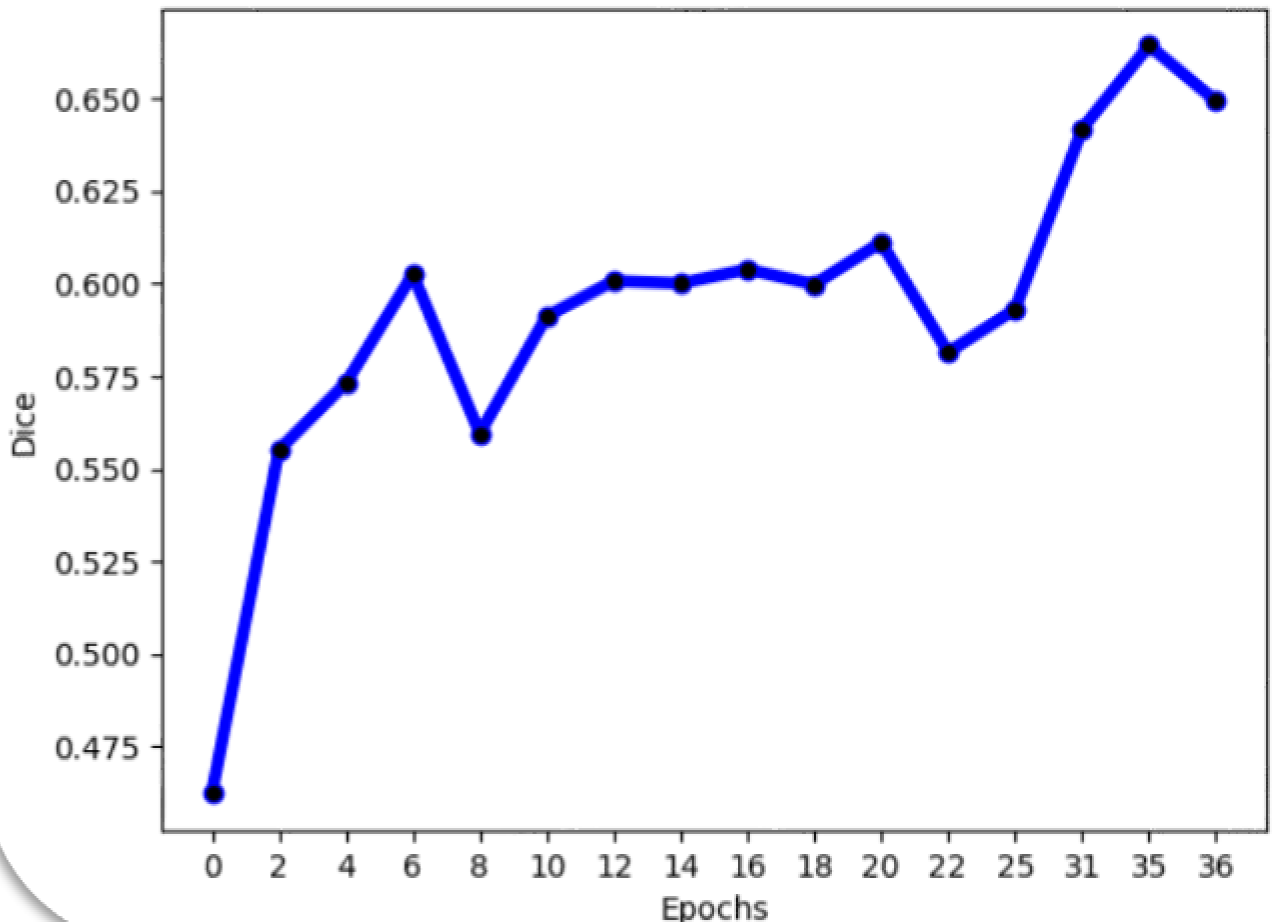
Proposed Method
Dataset & Evaluation

Data: LuxCarta sattelite images of 9 cities worldwide with 50 cm/pixel spatial resolution (721 patches for training, 24 patches for testing).

Evaluation: Dice similarity coefficient (DSC):

$$DSC = \frac{2|Prediction \cdot GroundTruth|}{|Prediction|^2 + |GroundTruth|^2}$$

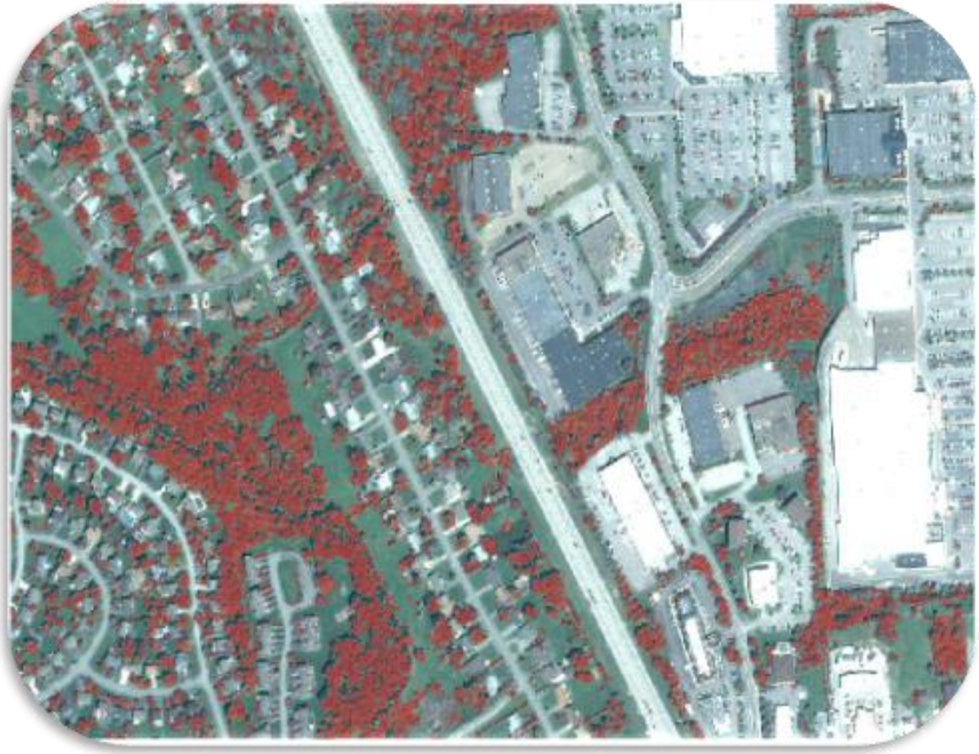
Proposed Method
Results



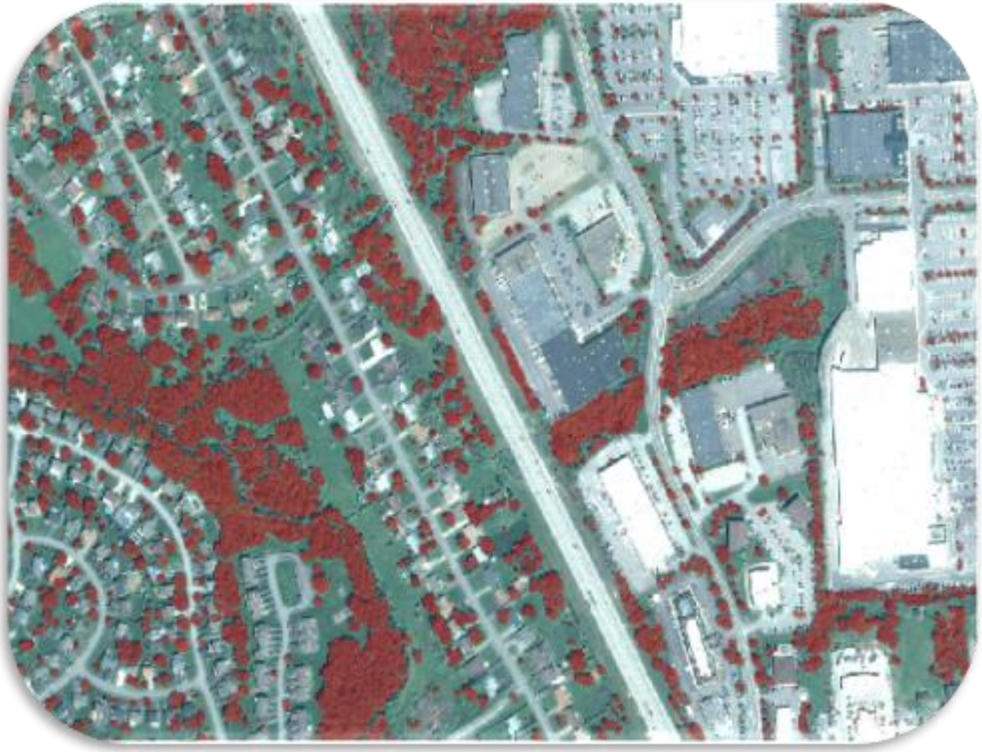
Performance is growing, hence training on pseudo-labels helps learning spatial information

Performance after the first epoch is close to the performance of K-Means applied to 3 channels

Proposed Method
Results



ClusterNet
66% Dice



Ground Truth

Proposed Method
Results

ClusterNet
66% Dice

Ground Truth

Proposed Method Comparison

	Dice	Accuracy
ClusterNet	**0.66**	**0.86**
Supervised U-Net	0.61	0.84
K-Means Clustering	0.48	0.76
SVM	0.18	0.69
SVM + GraphCut	0.19	0.71
SuperPixel GraphCut	0.37	0.68

Future Work

- Different clustering techniques
- Pretraining with Autoencoders
- Multitask Learning(clustering + autoencoders)
- Experiments on Data with higher spatial resolution

Thank You for Attention