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
Advantages:
➢ State of the art performance

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

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
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.


Proposed Method
ClusterNet
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 split which makes the process interactive.
Proposed Method
Dataset & Evaluation

Data: LuxCarta satellite 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
<table>
<thead>
<tr>
<th>Method</th>
<th>Dice</th>
<th>Accuracy</th>
</tr>
</thead>
<tbody>
<tr>
<td>ClusterNet</td>
<td>0.66</td>
<td>0.86</td>
</tr>
<tr>
<td>Supervised U-Net</td>
<td>0.61</td>
<td>0.84</td>
</tr>
<tr>
<td>K-Means Clustering</td>
<td>0.48</td>
<td>0.76</td>
</tr>
<tr>
<td>SVM</td>
<td>0.18</td>
<td>0.69</td>
</tr>
<tr>
<td>SVM + GraphCut</td>
<td>0.19</td>
<td>0.71</td>
</tr>
<tr>
<td>SuperPixel GraphCut</td>
<td>0.37</td>
<td>0.68</td>
</tr>
</tbody>
</table>
Future Work

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

Andrii Zhygallo, 10.09.2019
Thank You for Attention