Emergency Response with Deep Neural Networks and Satellite Imagery

Benjamin Bischke, Damian Borth

Talk-ID: 23479
What can be done with it?

Agriculture: Water Utilization
What can be done with it?

Business: Mining Output
What can be done with it?

Business: Mining Output
What can be done with it?

Energy: Pollution Monitoring
What can be done with it?

Environment: Deforestation
Satellite Analysis - What can be done with it?
Satellite Analysis - What can be done with it?

1. **No Poverty**
2. **No Hunger**
3. **Good Health**
4. **Quality Education**
5. **Gender Equality**
6. **Clean Water and Sanitation**
7. **Renewable Energy**
8. **Good Jobs and Economic Growth**
9. **Innovation and Infrastructure**
10. **Reduced Inequalities**
11. **Sustainable Cities and Communities**
12. **Responsible Consumption**
13. **Climate Action**
14. **Life Below Water**
15. **Life on Land**
16. **Peace and Justice**
17. **Partnerships for the Goals**

**Pollution Monitoring (Oil Spills, Smok, Pipelines)**

**Infrastructure & Traffic Monitoring**

**Water utilisation**

**Ice Caps**

**Carbon Stocks**

**Climate Change**

**Alge-Bloom**

**Natural & Manmade Disasters**

**Wildlife Monitoring**

**Deforestation**

**Social Events**

**Agriculture**
Satellite Analysis - What can be done with it?

Pollution Monitoring (Oil Spills, Smok, Pipelines)
- Water utilisation
- Infrastructure & Traffic Monitoring

1. NO POVERTY
2. NO HUNGER
3. GOOD HEALTH
4. QUALITY EDUCATION
5. GENDER EQUALITY
6. CLEAN WATER AND SANITATION

7. RENEWABLE ENERGY
8. GOOD JOBS AND ECONOMIC GROWTH
9. INNOVATION AND INFRASTRUCTURE
10. REDUCED INEQUALITIES
11. SUSTAINABLE CITIES AND COMMUNITIES
12. RESPONSIBLE CONSUMPTION

13. CLIMATE ACTION
14. LIFE BELOW WATER
15. LIFE ON LAND
16. PEACE AND JUSTICE
17. PARTNERSHIPS FOR THE GOALS

- Ice Caps
- Carbon Stocks
- Climate Change
- Alge-Bloom
- Natural & Manmade Disasters
- Social Events
- Wildlife Monitoring
- Deforestation
- Agriculture

DFKI – KM - DLCC
Situation Awareness with Remote Sensing

**Earthquakes**


**Wildfires**

NASA, Oct. 2007

**Flooding**


**Dam Break**

NASA, Nov. 2015

Earthquakes, Wildfires, Flooding, and Dam Break are areas where remote sensing technology can be applied to improve situation awareness.
Situation Awareness with Remote Sensing

Earthquakes

Wildfires
NASA, Oct. 2007

Flooding

Dam Break
NASA, Nov. 2015
Is Satellite Imagery enough?

NASA, May 2016
- Satellite scenes from Landsat8-satellite
- Wildfire only visible in composite of certain bands
“DeepEye” Visualization
Situation Awareness with Remote Sensing

Earthquakes

Flooding

Wildfires

Dam Break


NASA, Oct. 2007

NASA, Nov. 2015
Current Natural Disasters

- Cold Wave: 73
- Drought: 122
- Earthquake: 376
- Epidemic: 158
- Extratropical Cyclone: 19
- Fire: 5
- Flash Flood: 222
- Flood: 1,241
- Heat Wave: 6
- Insect Infestation: 8
- Land Slide: 204
- Mud Slide: 39
- Other: 11
- Severe Local Storm: 141
- Snow Avalanche: 19
- Storm Surge: 11
- Technological Disaster: 42
- Tropical Cyclone: 438
- Tsunami: 15
- Volcano: 102
- Wild Fire: 54
Current Natural Disasters

How Houston's layout may have made its flooding worse

<table>
<thead>
<tr>
<th>Type</th>
<th>Count</th>
</tr>
</thead>
<tbody>
<tr>
<td>Cold Wave</td>
<td>73</td>
</tr>
<tr>
<td>Drought</td>
<td>122</td>
</tr>
<tr>
<td>Earthquake</td>
<td>378</td>
</tr>
<tr>
<td>Epidemic</td>
<td>158</td>
</tr>
<tr>
<td>Extratropical Cyclone</td>
<td>19</td>
</tr>
<tr>
<td>Fire</td>
<td>5</td>
</tr>
<tr>
<td>Flash Flood</td>
<td>222</td>
</tr>
<tr>
<td>Flood</td>
<td>1,241</td>
</tr>
<tr>
<td>Heat Wave</td>
<td>6</td>
</tr>
<tr>
<td>Insect Infestation</td>
<td>8</td>
</tr>
<tr>
<td>Land Slide</td>
<td>294</td>
</tr>
<tr>
<td>Mud Slide</td>
<td>39</td>
</tr>
<tr>
<td>Other</td>
<td>11</td>
</tr>
<tr>
<td>Severe Local Storm</td>
<td>141</td>
</tr>
<tr>
<td>Snow Avalanche</td>
<td>19</td>
</tr>
<tr>
<td>Storm Surge</td>
<td>11</td>
</tr>
<tr>
<td>Technological Disaster</td>
<td>42</td>
</tr>
<tr>
<td>Tropical Cyclone</td>
<td>438</td>
</tr>
<tr>
<td>Tsunami</td>
<td>15</td>
</tr>
<tr>
<td>Volcano</td>
<td>102</td>
</tr>
<tr>
<td>Wild Fire</td>
<td>54</td>
</tr>
</tbody>
</table>
How Houston's layout may have made its flooding worse
Is Satellite Imagery enough?

DigitalGlobe, October 2017
Is Satellite Imagery enough?

Limited Perspective (Clouds, etc.)
Low Temporal Resolution

DigitalGlobe, October 2017
Contextual Enrichment of Satellite Imagery (Idea)

DigitalGlobe, October 2017
The Multimedia Satellite Task at MediaEval 2017
Multimedia Satellite Task - Overview

- **Goal:** Combine Satellite Imagery with Social Multimedia

- **Focus on** Flooding Events

- Two Subtasks:
  - Disaster Image Retrieval from Social Media (DIRSM)
  - Flood Detection in Satellite Imagery (FDSI)
Multimedia Satellite Task - Overview

• **Goal:** Combine Satellite Imagery with Social Multimedia

• **Focus on** **Flooding Events**

• Two Subtasks:
  • **Disaster Image Retrieval from Social Media (DIRSM)**
  • **Flood Detection in Satellite Imagery (FDSI)**
Disaster Image Retrieval from Social Media

We're now also partnering with @TheChurchTX to provide relief in the #houstonflood Please continue to pray! #HurricaneHarvey

@realDonaldTrump showed grace & class in #HoustonFloods - not that the #Left gives him any credit.

Accuweather Concerned System May Develop Southwestern GOM This Weekend accuweather.com/en/weather-new ... ... #GulfofMexico #Storm #HoustonFloods #Harvey

This is Tex, We rescued him from #HoustonFlood

Trump's presidential grace and class in Texas
President Trump's visit to flooded parts of southern Texas went off without a hitch yet he has been besieged with scathing attacks by rabid left-wingers and their ...
bombthrowers.com

Near Port Arthur in #texas #TexasStrong #texasflood @VISITFLORIDA @viletlauderdale #HoustonFloods #HoustonFloods

At 10 neighborhood waiting to cleanup @abc13houston #Harvey2017 #HoustonFloods abc13eyewitness

This isn't an old picture. I just took it. The water isn't going anywhere. #houstonstrong HoustonFloods
Disaster Image Retrieval from Social Media

We're now also partnering with @TheChurchTX to provide relief in the houstonfloods. Please continue to pray! #HurricaneHarvey

@realDonaldTrump showed grace & class in HoustonFloods - not that the Left gives him any credit.

Accuweather Concerned System May Develop Southwestern GOM This Weekend accuweather.com/en/weather-new ... #GulfOfMexico #Storm #HoustonFloods #Harvey

This is Tex. We rescued him from HoustonFloods

Near Port Arthur in #texas #TexasStrong #texasflood @VISITFLORIDA @visitlauderdale #HoustonFloods

At 10 neighborhood waiting to clear up @abc13houston #Harvey2017 #HoustonFloods #abc13eyewitness

Trump's presidential grace and class in Texas:

President Trump's visit to flooded parts of southern Texas went off without a hitch yet he has been besieged with scathing attacks by rabid left-wingers and their ... bombthrowers.com

Rain's power flooding due to Tropical Concern.

VERITAS

 surroundings
Multimedia Satellite Task - Overview

• **Goal:** Combine Satellite Imagery with Social Multimedia

• **Focus on** Flooding Events

• Two Subtasks:
  • Disaster Image Retrieval from Social Media (DIRSM)
  • Flood Detection in Satellite Imagery (FDSI)
Flood Detection in Satellite Imagery
Run Submissions and Evaluation

• Up to 5 runs for each subtask:
  • **DIRSM**: three required runs only with the provided dev set (Visual, textual, both)
  • **FDSI**: three required runs only with the provided dev set

• Standard Evaluation Metrics:
  • **DIRSM**: Average Precision at different cutoffs
  • **FDSI**: Intersection over Union
Task Dataset

- **DIRSM-Dataset:**
  - 6.6k images from YFCC100M + metadata (under CC-licence)
  - Basic set of precomputed features
  - Two labels (Flooding/no Flooding)

- **FDSI-Dataset:**
  - High resolution satellite scenes of seven flooding events provided by PlanetLabs
  - Cropped Image Patches (320x320px)
  - Segmentation masks for each patch
Ground Truth

- **DIRSM:**
  - Image rating according to the strength of the evidence of flooding (1-5) on Crowdflower (crowdsourcing)
  - Label as flooding if annotators rate with 4, 5 and non flooding for 1, 2
  - Additional distractor images

- **FDSI:**
  - Segmentation masks extracted by human annotators

---

**Image Rating**

- **1:** There is no image / only an error message.
- **2:** Also mark this checkboxes if a placeholder image is shown instead of the original image, (indicated by a text such as This photo is no longer available.)

**How strongly does this image show evidence of a flooding event? (required)**

<table>
<thead>
<tr>
<th>No Flooding</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>Flooding</th>
</tr>
</thead>
</table>

- **3:** The term water includes water-bodies (e.g. lake, river, sea) and water visible due to the flooding (flooded street, muddy road, flooded field)

- **4:** Is there water visible in the image? (required)
  - **Yes**
  - **No**
Task Participation

• 15 Teams registered, 11 submitted runs:
  • from all the world (Brasil, Australia, Greece, Brunei, Italy, UK, Germany, Netherlands, Norway, Pakistan)
  • 11 teams submitted for the DIRSM subtask
  • 6 teams additionally for the FDSI subtask

• In total **63** submissions:
  • **44** submission for the first task
  • **19** submission for the second subtask
Disaster Image Retrieval From Social Media - Visual

- Retrieval only based on Image-Information
DIRSM - Classification using Features from IR

<table>
<thead>
<tr>
<th>SVM (RBF)</th>
<th>ACC</th>
<th>CEDD</th>
<th>CL</th>
<th>EH</th>
<th>FCTH</th>
<th>Gabor</th>
<th>JCD</th>
<th>SC</th>
<th>Tamara</th>
</tr>
</thead>
<tbody>
<tr>
<td>Precision</td>
<td>68</td>
<td>75</td>
<td>75</td>
<td>76</td>
<td>72</td>
<td>57</td>
<td>74</td>
<td>64</td>
<td>65</td>
</tr>
<tr>
<td>Recall</td>
<td>69</td>
<td>75</td>
<td>75</td>
<td>76</td>
<td>73</td>
<td>62</td>
<td>74</td>
<td>65</td>
<td>66</td>
</tr>
<tr>
<td>F1-Score</td>
<td>68</td>
<td>75</td>
<td>75</td>
<td>76</td>
<td>72</td>
<td>54</td>
<td>73</td>
<td>60</td>
<td>62</td>
</tr>
</tbody>
</table>

- **Features:**
  - ACC - Auto-Color-Correlogram
  - CEDD - Color and Edge Directivity Descriptor
  - CL - Color Layout
  - EH - Edge Histogram
  - FCTH - Fuzzy Color and Texture Histogram
  - Gabor
  - JCD - Joint Composite Descriptor
  - SC - Scalable color
  - Tamara

- **Training**
  - 80/20 Split Training/Test
  - Baseline 63.26%
• Low-Level features are not expressive enough
  • Similar color, texture, water
Low-Level features are not expressive enough
- Similar color, texture, water

Can we leverage high-level features from pre-trained-CNN?
- Fine-tuning of CNNs
- Extracting Layers of pre-trained CNNs
• Low-Level features are not expressive enough
  • Similar color, texture, water

• Can we leverage high-level features from pre-trained-CNN?
  • Fine-tuning of CNNs
  • Extracting Layers of pre-trained CNNs
DIRSM - Classification using Features of pre-trained CNNs

<table>
<thead>
<tr>
<th>ImageNet based FeatureExtractor</th>
<th>lin. SVM (F1-Score)</th>
<th>RBF SVM (F1-Score)</th>
</tr>
</thead>
<tbody>
<tr>
<td>AlexNet (ImageNet) - FC6</td>
<td>79</td>
<td>85</td>
</tr>
<tr>
<td>AlexNet (ImageNet) - FC7</td>
<td>77</td>
<td>83</td>
</tr>
<tr>
<td>SqueezeNet 1.0 (ImageNet) - Fire 10</td>
<td>83</td>
<td>87</td>
</tr>
<tr>
<td>SqueezeNet 1.1 (ImageNet) - Fire10</td>
<td>83</td>
<td>86</td>
</tr>
<tr>
<td>ResNet18 (ImageNet) - FC1000</td>
<td>72</td>
<td>70</td>
</tr>
<tr>
<td>ResNet50 (ImageNet) - FC1000</td>
<td>75</td>
<td>71</td>
</tr>
<tr>
<td>ResNet101 (ImageNet) - FC1000</td>
<td>77</td>
<td>74</td>
</tr>
<tr>
<td>ResNet150 (ImageNet) - FC1000</td>
<td>76</td>
<td>74</td>
</tr>
</tbody>
</table>
DIRSM - Classification using Features of pre-trained CNNs

<table>
<thead>
<tr>
<th>ImageNet based FeatureExtractor</th>
<th>lin. SVM (F1-Score)</th>
<th>RBF SVM (F1-Score)</th>
</tr>
</thead>
<tbody>
<tr>
<td>AlexNet (ImageNet) - FC6</td>
<td>79</td>
<td>85</td>
</tr>
<tr>
<td>AlexNet (ImageNet) - FC7</td>
<td>77</td>
<td>83</td>
</tr>
<tr>
<td>SqueezeNet 1.0 (ImageNet) - Fire 10</td>
<td>83</td>
<td>87</td>
</tr>
<tr>
<td>SqueezeNet 1.1 (ImageNet) - Fire10</td>
<td>83</td>
<td>86</td>
</tr>
<tr>
<td>ResNet18 (ImageNet) - FC1000</td>
<td>72</td>
<td>70</td>
</tr>
<tr>
<td>ResNet50 (ImageNet) - FC1000</td>
<td>75</td>
<td>71</td>
</tr>
<tr>
<td>ResNet101 (ImageNet) - FC1000</td>
<td>77</td>
<td>74</td>
</tr>
<tr>
<td>ResNet150 (ImageNet) - FC1000</td>
<td>76</td>
<td>74</td>
</tr>
</tbody>
</table>

• Are there more suitable CNN-models for this problem?
• Borth et. al. introduced Adjective Noun Pairs with the Visual Sentiment Ontology (VSO)
• DeepSentiBank uses state-of-the-art CNNs to predict Adjective-Noun Pairs (ANPs)
  • AlexNet
  • VGG
  • GoogleNet
  • X-ResNet

• TODO: Reference
DeepSentiBank Model X-ResNet

- MultiTask prediction for Adjectives, Nouns and Adjective-Noun Pairs
- Building upon Residual Networks
- Cross-Residual Connections

TODO: Reference
DeepSentiBank Predictions

- Describe the scene with ANPs (*wet road, damaged building, stormy sky*)
- Less domain change (compared to ImageNet)
### DIRSM - Classification using Features of pre-trained CNNs

**ImageNet based FeatureExtractor**

<table>
<thead>
<tr>
<th>Model</th>
<th>lin. SVM (F1-Score)</th>
<th>RBF SVM (F1-Score)</th>
</tr>
</thead>
<tbody>
<tr>
<td>AlexNet (ImageNet) - FC6</td>
<td>79</td>
<td>85</td>
</tr>
<tr>
<td>AlexNet (ImageNet) - FC7</td>
<td>77</td>
<td>83</td>
</tr>
<tr>
<td>SqueezeNet 1.0 (ImageNet) - Fire 10</td>
<td><strong>83</strong></td>
<td><strong>87</strong></td>
</tr>
<tr>
<td>SqueezeNet 1.1 (ImageNet) - Fire10</td>
<td>83</td>
<td>86</td>
</tr>
<tr>
<td>ResNet18 (ImageNet) - FC1000</td>
<td>72</td>
<td>70</td>
</tr>
<tr>
<td>ResNet50 (ImageNet) - FC1000</td>
<td>75</td>
<td>71</td>
</tr>
<tr>
<td>ResNet101 (ImageNet) - FC1000</td>
<td>77</td>
<td>74</td>
</tr>
<tr>
<td>ResNet150 (ImageNet) - FC1000</td>
<td>76</td>
<td>74</td>
</tr>
</tbody>
</table>

**VSO based FeatureExtractor**

<table>
<thead>
<tr>
<th>Model</th>
<th>lin. SVM (F1-Score)</th>
<th>RBF SVM (F1-Score)</th>
</tr>
</thead>
<tbody>
<tr>
<td>DeepSentiBank (AlexNet) - FC6</td>
<td>86.8</td>
<td>90.2</td>
</tr>
<tr>
<td>DeepSentiBank (AlexNet) - FC7</td>
<td>86.2</td>
<td>90.0</td>
</tr>
<tr>
<td>DeepSentiBank (GoogLeNet) - Pool5/7x7_s1</td>
<td>85.4</td>
<td>92.3</td>
</tr>
<tr>
<td>DeepSentiBank (XResNet) - Anptask/pool5</td>
<td><strong>90.9</strong></td>
<td><strong>93.7</strong></td>
</tr>
<tr>
<td>DeepSentiBank (XResNet) - Adjtask/pool5</td>
<td>89.2</td>
<td><strong>93.5</strong></td>
</tr>
<tr>
<td>DeepSentiBank (XResNet) - Nountask/pool5</td>
<td>90.2</td>
<td>93.2</td>
</tr>
</tbody>
</table>
DIRSM - Ranking using CNN-Features + SVM

- Comparing CNN-Vectors to mean class vector suboptimal
- Ranking based on distance to decision boundary of SVM with RBF-Kernel

<table>
<thead>
<tr>
<th></th>
<th>DSB (XResNet) - Anptask/pool5</th>
<th>DSB (XResNet) - Adjtask/pool5</th>
<th>DSB (GoogLeNet) - Pool5/7x7_s1</th>
</tr>
</thead>
<tbody>
<tr>
<td>Average Precision at 50</td>
<td>1.0</td>
<td>1.0</td>
<td>0.9790</td>
</tr>
<tr>
<td>Average Precision at 100</td>
<td>1.0</td>
<td>1.0</td>
<td>0.9847</td>
</tr>
<tr>
<td>Average Precision at 250</td>
<td>0.9949</td>
<td>0.9964</td>
<td>0.9791</td>
</tr>
<tr>
<td>Average Precision at 500</td>
<td>0.9869</td>
<td>0.9875</td>
<td>0.9683</td>
</tr>
<tr>
<td>Average Precision at 1156</td>
<td>0.9547</td>
<td>0.9538</td>
<td>0.9239</td>
</tr>
<tr>
<td>Average Precision at mean K</td>
<td>0.9873</td>
<td>0.9875</td>
<td>0.9687</td>
</tr>
</tbody>
</table>

- Only 40% (2111) training, 60% (3168) test
Disaster Image Retrieval

- Sample Predictions & Results
Disaster Image Retrieval From Social Media - Metadata

• Retrieval only based on Meta-Information

```json
{
  "images": [
    {
      "image_id": "12328463323",
      "image_url": "http://www.flickr.com/photos/9752474@N07/12328463323/",
      "image_extension_original": "jpg",
      "date_taken": "2014-01-30 10:18:12.0",
      "date_uploaded": "1391631137",
      "user_nsid": "9752474@N07",
      "user.nickname": "SurferJoe88",
      "title": "Emma Wood State Beach - campsites",
      "description": "Emma Wood State Beach - flooded campsites",
      "user_tags": ["flooding"],
      "license_name": "Attribution-NonCommercial-ShareAlike License",
      "license_url": "http://creativecommons.org/licenses/by-nc-sa/2.0/",
      "capture_device": "SAMSUNG PL70 / VLUU PL70 / SAMSUNG SL720",
      "latitude": 34.287154000000002,
      "longitude": -119.3298929999999844
    },
    ...
  ]
}
```
Disaster Image Retrieval - Metadata

- Evaluations of Title, Description, UserTags
- Word embeddings with word2Vec (200 dimensions) on UserTags
- Document representation as mean word embeddings
- Importance Weighting of Word Embeddings with term frequency document inverse frequency (TF-IDF)
- 63.41% AP on Testset
• Fusion of Embeddings and Visual Info on Feature-level
• Concatenation of Image Feature and Word Embedding

<table>
<thead>
<tr>
<th></th>
<th>AP@480</th>
<th>MAP@[50, 100, 150, 240, 480]</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Visual (XResNet)</strong></td>
<td>86.64</td>
<td>95.71</td>
</tr>
<tr>
<td><strong>Metadata</strong></td>
<td>63.41</td>
<td>77.64</td>
</tr>
<tr>
<td><strong>Visual &amp; Metadata</strong></td>
<td>90.45</td>
<td>97.40</td>
</tr>
<tr>
<td><strong>Visual (ImageNet)</strong></td>
<td>74.08</td>
<td>64.50</td>
</tr>
</tbody>
</table>
Flood Detection from Satellite Images

- Many approaches extract pixel information from Remote Sensing and traditional classifiers

- Can we apply networks used for Semantic Segmentation?

- Can we incorporate additional channels?
From Classification to Semantic Segmentation

- Fully Convolutional Networks
- Pixel-wise predictions by changing fully connected to fully convolutional layers

Todo: Reference
Network Architectures for Semantic Segmentation

- Fully Convolutional Network (RGB)
- Fully Convolutional Network (RGB + IR)
- SegNet (RGB + IR)

TODO: Reference
FDSI - Data Processing

- 90 Deg. rotations with vertical and horizontal flipping
- Normalisation based on location
  - mean per channel
  - conversion of 16 bit (geoTiffs) to 8 bit
• Evaluation Results on the Testset
  • Importance of the Infrared Channel for Flood Detection
  • Importance of the Decoder

<table>
<thead>
<tr>
<th>Model</th>
<th>Same Locations</th>
<th>New Location</th>
</tr>
</thead>
<tbody>
<tr>
<td>FCN (RGB)</td>
<td>73.56</td>
<td>69.32</td>
</tr>
<tr>
<td>FCN (RGB+IR)</td>
<td>84.27</td>
<td>70.87</td>
</tr>
<tr>
<td>SegNet (RGB+IR)</td>
<td>84.36</td>
<td>74.13</td>
</tr>
</tbody>
</table>
From Flood Detection to Emergency Response with Satellite Data
From Flood-Detection to Emergency Response
From Flood-Detection to Emergency Response
Problems with many Build Datasets

- Many publications on Building Segmentation from Satellite Data
  - Low generalisation to new places and cities
  - Many Variations (Building Types, Atmospheric Correction, Density)
Segmentation of Building Footprints - Overview

- Large-Scale Inria Aerial Image Labeling Dataset released in May 2017 with High Resolution Imagery
  - Specifically addressing Variations
- Open Challenge with Leaderboard
- Evaluated different CNN-architectures for Sem. Segmentation

TODO: Reference
Segmentation of Building Footprints - „Blobby“-Predictions!
Segmentation of Building Footprints - “Blobby”-Predictions!
Segmentation of Building Footprints - Baseline Architectures

- Different Architectures
  - FCN
  - Unpooling-Net (SegNet)
  - Skip-Net (U-Net)
  - FCN + MLP

- Results

<table>
<thead>
<tr>
<th></th>
<th>Austin</th>
<th>Chicago</th>
<th>Kitsap Co.</th>
<th>West Tyrol</th>
<th>Vienna</th>
<th>Overall</th>
</tr>
</thead>
<tbody>
<tr>
<td>FCN</td>
<td>47.66</td>
<td>53.62</td>
<td>33.70</td>
<td>46.86</td>
<td>60.60</td>
<td>53.82</td>
</tr>
<tr>
<td>Acc.</td>
<td>92.22</td>
<td>88.59</td>
<td>98.58</td>
<td>95.83</td>
<td>88.72</td>
<td>92.79</td>
</tr>
<tr>
<td>Skip</td>
<td>57.87</td>
<td>61.13</td>
<td>46.43</td>
<td>54.91</td>
<td>70.51</td>
<td>62.97</td>
</tr>
<tr>
<td>Acc.</td>
<td>93.85</td>
<td>90.54</td>
<td>98.84</td>
<td>96.47</td>
<td>91.48</td>
<td>94.24</td>
</tr>
<tr>
<td>MLP</td>
<td>61.20</td>
<td>61.30</td>
<td>51.50</td>
<td>57.95</td>
<td>72.13</td>
<td>64.67</td>
</tr>
<tr>
<td>Acc.</td>
<td>94.20</td>
<td>90.43</td>
<td>98.92</td>
<td>96.66</td>
<td>91.87</td>
<td>94.42</td>
</tr>
</tbody>
</table>
Using Multiple Networks for a Better Prediction

Class Boundary Network
Using Multiple Networks for a Better Prediction

Class Boundary Network

Model Ensemble of multiple Networks

- TODO: Reference
Multi-Task Learning

- Use complementary tasks to improve the original task
  - Implicit Data Augmentation
  - Attention Focusing
  - Representation Bias
  - Regularization

* TODO: Reference
Standard Output Representations

Input

Semantic Information
Additional Output Representations

Input

\[ D(p) = \delta_p \min(\min_{q \in Q} d(p, q), R), \]

\[ \delta_p = \begin{cases} 
+1 & \text{if } p \in C_{\text{building}} \\
-1 & \text{if } p \notin C_{\text{building}} 
\end{cases} \]

\[ D(p) \sum_{k=1}^{K} r_n b_k(p) \sum_{k=1}^{K} b_k(p) = 1 \]

Semantic Information

Geometric Information (Distance)
Cascaded Multi-Task Network

\[ L_{total}(x; \theta, \sigma_{dist}, \sigma_{seg}) = L_{dist}(x; \theta, \sigma_{dist}) + L_{seg}(x; \theta, \sigma_{seg}) \]
Uncertainty Based Task Weighting

- Model Task weights as Uncertainty of each Task
  - Optimize network parameters and uncertainty parameters together

\[
L_t(x, \theta, \sigma_t) = \sum_{c=1}^{C} -C_c \log P(C_c = 1|x, \theta, \sigma_t)
\]

\[
= \sum_{c=1}^{C} -C_c \log \left( \exp \left( \frac{1}{\sigma_t^2} f_c(x) \right) \right) + \log \sum_{c'=1}^{C} \exp \left( \frac{1}{\sigma_t^2} f_{c'}(x) \right)
\]

Applying the same assumption as in [25]:

\[
\frac{1}{\sigma^2} \sum_{c'} \exp \left( \frac{1}{\sigma^2} f_{c'}(x) \right) \approx \left( \sum_{c'} \exp(f_{c'}(x)) \right)^{\frac{1}{\sigma^2}}
\]

allows to simplify Eq. 6 to:

\[
L_t(x, \theta, \sigma_t) \approx \frac{1}{\sigma_t^2} \sum_{c=1}^{C} -C_c \log P(C_c = 1|x, \theta) + \log(\sigma_t^2)
\]
### Quantitative Results

<table>
<thead>
<tr>
<th>Model</th>
<th>mean IoU</th>
<th>Acc. (Pixel)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Baseline FCN [2]</td>
<td>53.82%</td>
<td>92.79 %</td>
</tr>
<tr>
<td>Baseline FCN + MLP[2]</td>
<td>64.67%</td>
<td>94.42 %</td>
</tr>
<tr>
<td>FCN (VGG16 encoder)</td>
<td>66.21%</td>
<td>94.54 %</td>
</tr>
<tr>
<td>FCN + MLP (VGG16 encoder)</td>
<td>68.17%</td>
<td>94.95 %</td>
</tr>
<tr>
<td>SegNet (VGG16 encoder)</td>
<td>70.14%</td>
<td>95.17 %</td>
</tr>
</tbody>
</table>
## Quantitative Results

<table>
<thead>
<tr>
<th>Model</th>
<th>mean IoU</th>
<th>Acc. (Pixel)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Baseline FCN [2]</td>
<td>53.82%</td>
<td>92.79 %</td>
</tr>
<tr>
<td>Baseline FCN + MLP[2]</td>
<td>64.67%</td>
<td>94.42 %</td>
</tr>
<tr>
<td>FCN (VGG16 encoder)</td>
<td>66.21%</td>
<td>94.54 %</td>
</tr>
<tr>
<td>FCN + MLP (VGG16 encoder)</td>
<td>68.17%</td>
<td>94.95 %</td>
</tr>
<tr>
<td>SegNet (VGG16 encoder)</td>
<td><strong>70.14%</strong></td>
<td><strong>95.17 %</strong></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Model</th>
<th>Austin</th>
<th>Chicago</th>
<th>Kitsap Co.</th>
<th>West Tyrol</th>
<th>Vienna</th>
<th>Overall</th>
</tr>
</thead>
<tbody>
<tr>
<td>FCN + MLP (Baseline)</td>
<td>IoU</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>61.20</td>
<td>61.30</td>
<td>51.50</td>
<td>57.95</td>
<td>72.13</td>
<td>64.67</td>
</tr>
<tr>
<td></td>
<td>Acc.</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>94.20</td>
<td>90.43</td>
<td>98.92</td>
<td>96.66</td>
<td>91.87</td>
<td>94.42</td>
</tr>
<tr>
<td>SegNet (Single-Loss)</td>
<td>IoU</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>74.81</td>
<td>52.83</td>
<td>68.06</td>
<td>65.68</td>
<td>72.90</td>
<td>70.14</td>
</tr>
<tr>
<td>NLL-Loss for Seg. Classes</td>
<td>Acc.</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>92.52</td>
<td>98.65</td>
<td>97.28</td>
<td>91.36</td>
<td>96.04</td>
<td>95.17</td>
</tr>
<tr>
<td>SegNet (Single-Loss)</td>
<td>IoU</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>76.49</td>
<td>66.77</td>
<td>72.69</td>
<td>66.35</td>
<td>76.25</td>
<td>72.57</td>
</tr>
<tr>
<td>NLL-Loss for Dist. Classes</td>
<td>Acc.</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>93.12</td>
<td>99.24</td>
<td>97.79</td>
<td>91.58</td>
<td>96.55</td>
<td>95.66</td>
</tr>
<tr>
<td>SegNet + MultiTask-Loss (Equally Weighted)</td>
<td>IoU</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>76.22</td>
<td>66.64</td>
<td>71.70</td>
<td><strong>67.03</strong></td>
<td><strong>76.68</strong></td>
<td>72.65</td>
</tr>
<tr>
<td></td>
<td>Acc.</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>93.03</td>
<td>99.24</td>
<td>97.71</td>
<td>91.66</td>
<td>96.60</td>
<td>95.65</td>
</tr>
<tr>
<td>SegNet + MultiTask-Loss (Uncertainty Weighted)</td>
<td>IoU</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td><strong>76.76</strong></td>
<td><strong>67.06</strong></td>
<td><strong>73.30</strong></td>
<td>66.91</td>
<td><strong>76.68</strong></td>
<td><strong>73.00</strong></td>
</tr>
<tr>
<td></td>
<td>Acc.</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td><strong>93.21</strong></td>
<td><strong>99.25</strong></td>
<td><strong>97.84</strong></td>
<td><strong>91.71</strong></td>
<td><strong>96.61</strong></td>
<td><strong>95.73</strong></td>
</tr>
</tbody>
</table>
Qualitative Results
Qualitative Results
Acknowledgments

[Image of NVIDIA and Planet logos]