End-to-end Deep Learning Methods for Automated Damage Detection in Extreme Events at Various Scales

Yongsheng Bai1,2, Halil Sezer2, Alper Yilmaz1
1 Photogrammetric Computer Vision Lab, The Ohio State University, Columbus, USA
2 Department of Civil, Environmental and Geodetic Engineering, The Ohio State University, Columbus, USA

➢ Scene level (scale) problem in cracking detection:

(a) pixel level
(b) object level
(c) structural level

Can we find a deep learning method to detect cracks automatically and successfully on 2D images at various scene levels or scales?

➢ Methodology:

1. Basic network: Mask R-CNN

2. Mask R-CNN with Path Aggregation Network (PANet) and Spatial Attention Mechanisms (Mask R-CNN + A-PANet):
   - (a) FPN: Backbone (b) Bottom-up path implementation (c) Adaptive Attn (d) FPN branch
   - (d) Fully connected feature: Note the two exist channel dimensions of feature maps at (a) and (b) for brevity [27]

3. Mask R-CNN with High-resolution Network (Mask R-CNN + HRNet):
   - There are four stages. The first stage consists of high-resolution convolutions. The second, third, and fourth stages represent three resolutions, four resolution blocks [38]

4. Evaluation of the Mask R-CNNs with validation data:

<table>
<thead>
<tr>
<th>Methodology</th>
<th>Pixel Acc</th>
<th>Recall</th>
<th>F-measure</th>
<th>MRR</th>
<th>FPR</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mask R-CNN + A-FPN</td>
<td>86.9</td>
<td>76.5</td>
<td>80.6</td>
<td>70.0</td>
<td>53.6</td>
</tr>
<tr>
<td>Mask R-CNN + A-PANet</td>
<td>88.8</td>
<td>80.7</td>
<td>84.6</td>
<td>69.8</td>
<td>58.8</td>
</tr>
</tbody>
</table>

5) Testing on 2017 Pohang earthquake images (4,109):

- Prediction of Mask R-CNN with Attraction-Style PANet and HRNet for 2017 Pohang earthquake images

6) Testing on 2017 Mexico City earthquake images (4,136):

- Prediction of Mask R-CNN with Attraction-Style PANet and HRNet for 2017 Mexico City earthquake images

➢ Data preparation for training:

2,021 images with the size from 168×330 to 460×3070 and with three scales are labelled.

- Pixel level
- Object level
- Structural level

Some examples of training data at different scales

➢ Implementation:

1) Pixel scene level (scale) Task in Phi-Net:

N = 543 images
Accuracy = 60.5% (U-Net)
Accuracy = 94.7% (Mask R-CNN + A-PANet)

2) Object scene level (scale) Task in Phi-Net:

N = 573 images
Accuracy = 90.1% (U-Net)
Accuracy = 97.1% (Mask R-CNN + A-PANet)

3) Structural scene level (scale) Task in Phi-Net:

N = 5,032 images
Accuracy = 8.8% (U-Net)
Accuracy = 91.6% (Mask R-CNN + A-PANet)

4) Testing on 2017 Mexico City earthquake images (4,136):

- Prediction of Mask R-CNN + Attraction-Style PANet and HRNet for 2017 Mexico City earthquake images

5) Testing on 2017 Pohang earthquake images (4,109):

- Prediction of Mask R-CNN with Attraction-Style PANet and HRNet for 2017 Pohang earthquake images

➢ Conclusion:

1. With appropriate training data, end-to-end networks like the latest Mask R-CNN which can detect cracks at various scale precisely are possible.
2. We still need to collect more data for counteracting the imbalance among training and finding a way to speed up the prediction on high-resolution images.
3. We plan to try other networks in future.

---

1. Basic end Deep Learning Methods for Automated Damage Detection in Extreme Events at Various Scales
2. Mask R-CNN with Path Aggregation Network (PANet) and Spatial Attention Mechanisms (Mask R-CNN + A-PANet): (a) FPN Backbone (b) Bottom-up path implementation (c) Adaptive Attn (d) FPN branch
3. Mask R-CNN with High-resolution Network (Mask R-CNN + HRNet): There are four stages. The first stage consists of high-resolution convolutions. The second, third, and fourth stages represent three resolutions, four resolution blocks [38]
4. Evaluation of the Mask R-CNNs with validation data:

<table>
<thead>
<tr>
<th>Methodology</th>
<th>Pixel Acc</th>
<th>Recall</th>
<th>F-measure</th>
<th>MRR</th>
<th>FPR</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mask R-CNN + A-FPN</td>
<td>86.9</td>
<td>76.5</td>
<td>80.6</td>
<td>70.0</td>
<td>53.6</td>
</tr>
<tr>
<td>Mask R-CNN + A-PANet</td>
<td>88.8</td>
<td>80.7</td>
<td>84.6</td>
<td>69.8</td>
<td>58.8</td>
</tr>
</tbody>
</table>

---

20th International Conference on Pattern Recognition

1.
2.
3.

---

1. With appropriate training data, end-to-end networks like the latest Mask R-CNN which can detect cracks at various scale precisely are possible.
2. We still need to collect more data for counteracting the imbalance among training and finding a way to speed up the prediction on high-resolution images.
3. We plan to try other networks in future.