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

Yongsheng Bai[1,2] Halil Sezen[2] Alper Yilmaz[1]
[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

25th International Conference on Pattern Recognition

➢ Scene level (scale) problem in cracking detection:

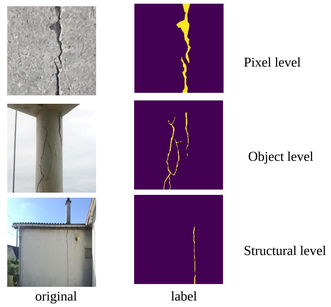
(a) pixel level (b) object level (c) structural level

Three scene levels (scales)

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

➢ Data preparation for training:

2,021 images with the size from 168×300 to 4600×3070 and with three scales are lablled.



Pixel level

Object level

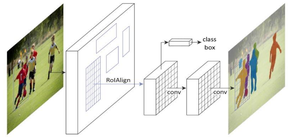
Structural level

original label

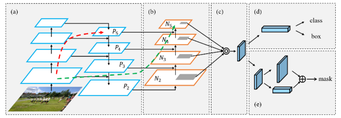
Some examples of training data at different scales

➢ Methodology:

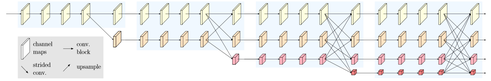
1. Basic network: Mask R-CNN



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 augmentation. (c) Adaptive feature pooling. (d) Box branch. (e) Fully connected fusion. Note that we omit channel dimension of feature maps in (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, fourth) stage repeats two-resolution (three-resolution, four-resolution) blocks [30]



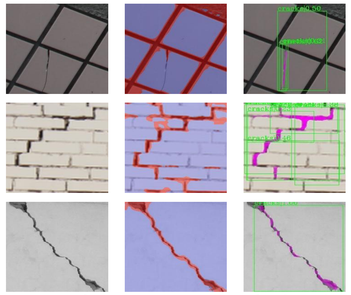
4. Evaluation of the Mask R-CNNs with validation data:

Methods	AP	AP50	AP75	AP_S	AP_M	AP_L
Mask R-CNN	21.7	54.9	16.6	28.6	41.2	23.1
Mask R-CNN + A-PANet	46.9	78.5	48.9	70.0	53.9	41.5
Mask R-CNN + HRNet	59.3	86.7	63.6	80.0	58.4	62.2

➢ Implementation

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

N = 4,663 images
Accuracy = 60.5% ; (U-Net)
Accuracy = **84.7%** (Mask R-CNN + A-PANet)

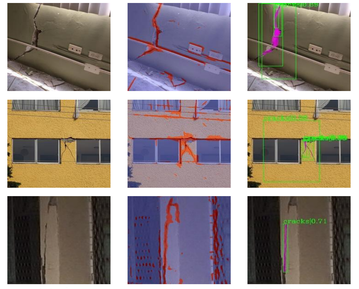


original U-Net Mask R-CNN + A-PANet

Some correct predictions of U-Net and Mask R-CNN + A-PANet at pixel level

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

N = 5,713 images
Accuracy = 26,2% ; (U-Net)
Accuracy = **77.1%** (Mask R-CNN + A-PANet)



original U-Net Mask R-CNN + A-PANet

Some correct predictions of U-Net and Mask R-CNN + A-PANet at object level

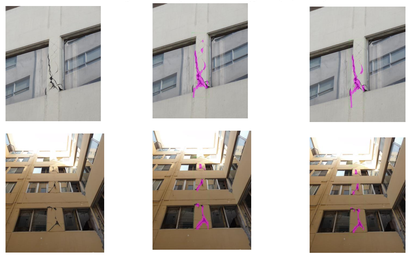
3) Structural scene level (scale) Task in Phi-Net:

N = 5,832 images
Accuracy = 8.9% ; (U-Net)
Accuracy = **81.9%** (Mask R-CNN + A-PANet)

original U-Net Mask R-CNN + A-PANet

Some correct predictions of U-Net and Mask R-CNN + A-PANet at object level

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



original Mask R-CNN + HRNet Mask R-CNN + A-PANet

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

Methods	Accuracy	Recall	Precision
Mask R-CNN + A-PANet	70.6%	53.6%	92.9%
Mask R-CNN + HRNet	73.0%	62.7%	90.5%

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

original Mask R-CNN + HRNet Mask R-CNN + A-PANet

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

Methods	Accuracy	Recall	Precision
Mask R-CNN + A-PANet	**74.1%**	56.9%	**94.7%**
Mask R-CNN + HRNet	74.0%	**63.6%**	88.3%

➢ 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 data and finding a way to speed up the prediction on high-resolution images.
3. We plan to try other networks in future.