Density Map Guided Object Detection in Aerial Images

Changlin Li¹, Taojiannan Yang¹, Sijie Zhu¹, Chen Chen¹, Shanyue Guan²

1. University of North Carolina at Charlotte
2. East Carolina University
Object detection in nature images

- Majority of objects has medium/large scale
- No/Limited scale variation

Object detection in aerial images

- More small objects
- The distribution of objects are sparse and non-uniform
- Large scale variations


Current practice: generate image crops

Baseline: Random cropping and uniform cropping

Deep learning based: Cropping by identifying difficult regions of detections

Question: Can we do better?
Density map guided image cropping

**Key idea**: generating density crops to improve detection result

- Remove background pixels to reduce the difficulty of detection
- Increase the percentage of small object pixels to make it more recognizable
Density map guided object detection in aerial images

- Detection Pipeline
Density map generation modular
Image cropping modular
Thresholding

We run a sliding window (2x2 here for demo purpose) on an image and sum all pixels in the window.

If the sum is smaller than the threshold (0.06 here for demo purpose), we set all pixels in the corresponding regions of density mask with a constant value > 0. Else we assign 0 to those cells.
Object detection

Input Aerial Image → CNN → Density Map → Threshold → Density Mask → Cropping Connected Regions → Crops → Detector → Fusion → Crops Detection → Final Detection

Global Detection
Quantitative results on VisionDrone 2018 dataset

<table>
<thead>
<tr>
<th>Method</th>
<th>Backbone</th>
<th>Test data</th>
<th>#Image</th>
<th>AP</th>
<th>AP$_{50}$</th>
<th>AP$_{75}$</th>
<th>AP$_{small}$</th>
<th>AP$_{mid}$</th>
<th>AP$_{large}$</th>
</tr>
</thead>
<tbody>
<tr>
<td>DetecNet+CPNet+ScaleNet [26]</td>
<td>ResNet 50</td>
<td>Original+cluster</td>
<td>2716</td>
<td>26.7</td>
<td>50.6</td>
<td>24.7</td>
<td>17.6</td>
<td>38.9</td>
<td>51.4</td>
</tr>
<tr>
<td>DetecNet+CPNet+ScaleNet [26]</td>
<td>ResNet 101</td>
<td>Original+cluster</td>
<td>2716</td>
<td>26.7</td>
<td>50.4</td>
<td>25.2</td>
<td>17.2</td>
<td>39.3</td>
<td>54.9</td>
</tr>
<tr>
<td>DetecNet+CPNet+ScaleNet [26]</td>
<td>ResNeXt 101</td>
<td>Original+cluster</td>
<td>2716</td>
<td>28.4</td>
<td>53.2</td>
<td>26.4</td>
<td>19.1</td>
<td>40.8</td>
<td>54.4</td>
</tr>
<tr>
<td>DMNet</td>
<td>ResNet 50</td>
<td>Original+density</td>
<td>2736</td>
<td>28.2</td>
<td>47.6</td>
<td>28.9</td>
<td>19.9</td>
<td>39.6</td>
<td>55.8</td>
</tr>
<tr>
<td>DMNet</td>
<td>ResNet 101</td>
<td>Original+density</td>
<td>2736</td>
<td>28.5</td>
<td>48.1</td>
<td>29.4</td>
<td>20.0</td>
<td>39.7</td>
<td>57.1</td>
</tr>
<tr>
<td>DMNet</td>
<td>ResNeXt 101</td>
<td>Original+density</td>
<td>2736</td>
<td>29.4</td>
<td>49.3</td>
<td><strong>30.6</strong></td>
<td><strong>21.6</strong></td>
<td><strong>41</strong></td>
<td>56.9</td>
</tr>
</tbody>
</table>

- Improvement of **1-1.5** on AP, depending on the choice of backbone
- An improvement of nearly **4 AP** for small object detection with ResNeXt 101 backbone
Quantitative result on UAVDT

<table>
<thead>
<tr>
<th>Method</th>
<th>Backbone</th>
<th>#Image</th>
<th>AP</th>
<th>AP$_{50}$</th>
<th>AP$_{75}$</th>
<th>AP$_{small}$</th>
<th>AP$_{mid}$</th>
<th>AP$_{large}$</th>
</tr>
</thead>
<tbody>
<tr>
<td>R-FCN [6]</td>
<td>ResNet 50</td>
<td>15096</td>
<td>7.0</td>
<td>17.5</td>
<td>3.9</td>
<td>4.4</td>
<td>14.7</td>
<td>12.1</td>
</tr>
<tr>
<td>SSD [14]</td>
<td>N/A</td>
<td>15096</td>
<td>9.3</td>
<td>21.4</td>
<td>6.7</td>
<td>7.1</td>
<td>17.1</td>
<td>12.0</td>
</tr>
<tr>
<td>RON [8]</td>
<td>N/A</td>
<td>15096</td>
<td>5.0</td>
<td>15.9</td>
<td>1.7</td>
<td>2.9</td>
<td>12.7</td>
<td>11.2</td>
</tr>
<tr>
<td>FRCNN [17]</td>
<td>VGG</td>
<td>15096</td>
<td>5.8</td>
<td>17.4</td>
<td>2.5</td>
<td>3.8</td>
<td>12.3</td>
<td>9.4</td>
</tr>
<tr>
<td>ClusDet [26]</td>
<td>ResNet 50</td>
<td>25427</td>
<td>13.7</td>
<td>26.5</td>
<td>12.5</td>
<td>9.1</td>
<td>25.1</td>
<td>31.2</td>
</tr>
<tr>
<td>DMNet</td>
<td>ResNet 50</td>
<td>32764</td>
<td><strong>14.7</strong></td>
<td>24.6</td>
<td><strong>16.3</strong></td>
<td><strong>9.3</strong></td>
<td><strong>26.2</strong></td>
<td><strong>35.2</strong></td>
</tr>
</tbody>
</table>

Compared with SOTA method – ClusDet, we achieve 1 AP improvement on UAVDT.
Visual result – object detection examples

Visualization of our DMNet detection results on VisionDrone (first row) and UAVDT (second row)
Conclusion

- **Density map guided object detection in aerial images**
  
  - Utilize density map to help better detect small objects
  
  - Achieve state-of-the-art performance on two benchmark datasets
  
  - Overall design is simple and clean
  
  - Our code will be released soon
Thank you