FlagDetSeg: Multi-Nation Flag Detection and Segmentation in the Wild

Shou-Fang Wu\textsuperscript{1}, Ming-Ching Chang\textsuperscript{2}, Cheng-Shih Wong\textsuperscript{3}, Abhineet Kumar Pandey\textsuperscript{2}, Siwei Lyu\textsuperscript{4}, Po-Chi Su\textsuperscript{1}

\textsuperscript{1} National Central University, Taiwan
\textsuperscript{2} University at Albany, State University of New York, USA
\textsuperscript{3} Academic Sinica, Taiwan
\textsuperscript{4} University at Buffalo, State University of New York, USA
Outline

- Overview
- Prior Work
- Method
- Experiment
- Discussion
Overview

• Precise **instance segmentation for 200+ country flags**
• **Data-augmentation-based** methods for fine-tuning
• Experiments performed on several popular detectors
• **RELEASE** (https://github.com/sfstefanwu/FlagDetSeg.git)
  • Pre-trained multi-nation flag detector
  • Annotated multi-nation flag dataset (authentic images)
  • Synthetic multi-nation flag dataset
Overview - Quick Demo

Images are from Google Open Image v6
Overview - Challenge

- **Non-rigid** - tiled, rotated, elastified, etc.
- **Heavy occlusion** in many cases
- **Lack of data** - labor-intensive for production
  - 193 member states in UN
Prior Work

• **Binary Flag Detection**
  - HSV color texture analysis and gradient features (S. Jetley, et al.)
  - Color features and a fuzzy-neural algorithm with kNN classifier (E. Hart, et al.)
  - A 5-layer CNN but limited results (H. H. Duc, et al.)

• **Multi-class Flag Detection**
  - Deep CNN, yet is uncompetitive against RPN-based detectors (M. Gu, et al.)
  - Based on VGG16 FCN, local context network and Color-BRIEF features (T. Said, et al.)
Method - Overview

- Open Image V6 Flags
  - Data Aug. + Random Bg.
  - OVDv6 Segmentation Set
  - Binary Flag Detector

- Multi-Nation Flag Templates
  - Cropping + Data Aug.
  - Flags in Transformations
  - Random Bg. + Color Adaption
  - Synthetic Flag Aug. Dataset

- Real Flags with Annotations
  - Masking
  - Real Flag Instances
  - Real Flag Aug. Dataset
  - Mask-RCNN Flag Training

- Model training
  - Fine-tuning
Flags from Open Image Dataset v6
OIDv6 Segmentation Dataset
Real Image Set

- Collect **5** real images for each country
- Select flags with different poses to **maximize diversity**
Synthetic Image Set

- **Natural Backgrounds + (Template or Instance)**
- Generate **large** and **balanced** dataset with **ground truth**
- “Simple copy-paste is a strong data augmentation method for instance segmentation” (G. Ghiasi, et al., CVPR 2021)
- **Source of background images**
  - Human Made Scene Collection (Burge J., et al.)
  - Stanford Background Dataset (S. Gould, et al.)
  - In-house Collection
Samples of Background Image
Template in SVG
Samples of Transformed Instance

- **Safe Transformations** are applied to
  - (a) Templates in SVG format
  - (b) Segmented instances from our *Real Image Set*

  to hallucinate realistic-looking flag images
Unsafe Transformation

- Vertical flipping

![Indonesia Flag](image1)

![Poland Flag](image2)

- RGB channel shifting or aggressive hue change

![Romania Flag](image3)

![Ireland Flag](image4)

![Italy Flag](image5)

![Nigeria Flag](image6)

![Mali Flag](image7)
Samples of Synthetic Image Set

(a) template

(b) real
Training Pipeline

- Pre-trained Models
- FlagDetSeg OIDv6 Segmentation Set
- FlagDetSeg Binary Flag Detector
- FlagDetSeg Synthetic Image Set
- FlagDetSeg Real Image Set + Synthetic Image Set
- Multi-country Flag Detector

Parameters:
- LR = 1e-3
- Iter = 5e5
## Experiment

- **Multi-nation Flag Detector**

<table>
<thead>
<tr>
<th>Backbone</th>
<th>$AP$</th>
<th>$AP_{50}$</th>
<th>$AP_{75}$</th>
<th>$AP_s$</th>
<th>$AP_m$</th>
<th>$AP_l$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mask-RCNN ResNet-101-FPN</td>
<td>87.92</td>
<td>93.1</td>
<td>92.36</td>
<td>40.8</td>
<td>80.03</td>
<td>91.03</td>
</tr>
<tr>
<td>Mask-RCNN ResNeXt-101-FPN</td>
<td>85.81</td>
<td>90.75</td>
<td>90.11</td>
<td>44.71</td>
<td>81.58</td>
<td>88.84</td>
</tr>
<tr>
<td>PointRend ResNet-101-FPN</td>
<td>83.91</td>
<td>87.68</td>
<td>87.13</td>
<td>31.49</td>
<td>77.82</td>
<td>88.04</td>
</tr>
<tr>
<td>PointRend ResNeXt-101-FPN</td>
<td>82.05</td>
<td>85.23</td>
<td>84.94</td>
<td>32.96</td>
<td>79.22</td>
<td>85.34</td>
</tr>
</tbody>
</table>
Visual Result of Synthetic Images
Visual Result of Real Images
Visual Result

- PointRend predicts clearer segmentation masks
Discussion and Future Work

- **Limitation**: heavy occlusion causes mis-classification

- **3D engine** (b) to create realistic template and simulate deformation
- **Teacher-Student network** (semi-supervised learning) from flag images without annotation
Thank you for your listening.
Experiment

- **Binary Flag Detection**
  - *YOLACT++*
    - not ideal for delineating complex flag instances
  - *PointRend* has slightly better performance than *Mask-RCNN*

<table>
<thead>
<tr>
<th></th>
<th>Backbone</th>
<th>$AP$</th>
<th>$AP_{50}$</th>
<th>$AP_{75}$</th>
<th>$AP_s$</th>
<th>$AP_m$</th>
<th>$AP_l$</th>
</tr>
</thead>
<tbody>
<tr>
<td>YOLACT++</td>
<td>ResNet-101-FPN</td>
<td>17.26</td>
<td>20.32</td>
<td>18.27</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>Mask-RCNN</td>
<td>ResNet-50-FPN</td>
<td>69.90</td>
<td>87.72</td>
<td>78.16</td>
<td>18.00</td>
<td>45.62</td>
<td>77.07</td>
</tr>
<tr>
<td>Mask-RCNN</td>
<td>ResNet-101-FPN</td>
<td>73.85</td>
<td>89.13</td>
<td>80.35</td>
<td>17.02</td>
<td>47.32</td>
<td>81.96</td>
</tr>
<tr>
<td>Mask-RCNN</td>
<td>ResNeXt-101-FPN</td>
<td>72.30</td>
<td>88.81</td>
<td>79.02</td>
<td>32.65</td>
<td>48.76</td>
<td>79.36</td>
</tr>
<tr>
<td>PointRend</td>
<td>ResNet-101-FPN</td>
<td><strong>76.30</strong></td>
<td><strong>90.24</strong></td>
<td><strong>81.90</strong></td>
<td>20.40</td>
<td>47.68</td>
<td><strong>84.68</strong></td>
</tr>
<tr>
<td>PointRend</td>
<td>ResNeXt-101-FPN</td>
<td>76.21</td>
<td>89.63</td>
<td><strong>82.06</strong></td>
<td>25.91</td>
<td><strong>49.75</strong></td>
<td>84.30</td>
</tr>
</tbody>
</table>
Binary flag detection - YOLACT++

- Miss
- False alerts, i.e. parachutes, birds