Deep Real-time Hand Detection Using CFPN on Embedded Systems

Pirdiansyah Hendri
Depart. of Com. Scie. Eng.
National Taiwan Ocean University
Keelung, Taiwan, R.O.C.
hendriplg@gmail.com

Jun-Wei Hsieh
College of AI.
National Chiao Tung University
Hsin Chu, Taiwan, R.O.C.
jwhsieh@nctu.edu.tw

Ping-Yang Chen
Dep. of C. S.
National Chiao Tung University
Hsin Chu, Taiwan, R.O.C.
pingyang.cs08g@nctu.edu.tw

M. Gochoo
Dep. of C. S.
United Arab Emirates University
United Arab Emirates
mgochoo@uaeu.ac.ae

Yong-Sheng Chen
Dep. of C. S.
National Chiao Tung University
Hsin Chu, Taiwan, R.O.C.
yscben@cs.nctu.edu.tw
One • Introduction
Two • Related Works
Three • Proposed Method
Four • Experimental Result
Five • Discussions & Conclusions
Introduction (1/2)

- Air writing
- Gesture Recognition
- Sign Language
- Gesture Recognition
- Home Control
- Action Recognition

Hand Detection
Introduction (2/2)

- Color Variation
- Occlusion
- Poses
- Lighting condition
- Deformable chances
Object Detection

Hand Detection

Related works
Related works
Proposal Method

Concatenated Feature Pyramid Network

Backbone
Proposal Method (1/6)

Feature Pyramid Network

Backbone

Feature Pyramid

Predictions

Conv. Layers

Scale 1

Scale 2

Scale 3

2x Up

1x1Conv. +
Yolo Architecture

Proposal Method (2/6)
Concatenated Feature Pyramid Network

Proposal Method (3/6)
Proposed Concatenation Block (CB)

- Deeper Layer
- Shallow layer
- 2x up
- Concat 1
- Conv. 1-3L
- Concat 2
- 1x1 Conv.
- Shallower Layer
- Prediction

Not applicable for Scale 1
Proposal Method (5/6)

- **Backbone**
  - Adopt darknet-19
  - Suitable to adopt deep backbone for better accuracy
  - Able to adopt a lighter backbone
Backbone of CFPN with CSP

Proposal Method (6/6)

CSP architecture

Part 1
Part 2

3x3 CONVOLUTION

CONCAT

New Feature
Experimental Result

- Dataset
- HandFlow Results
- COCO Dataset
Dataset

- HandFlow
- COCO
HandFlow Result
Results of hand detection using our CFPN-net
## Performance Comparisons on 416×416 Resolution Handflow Dataset on TX2 Embedded Device

<table>
<thead>
<tr>
<th>Model</th>
<th>Backbone</th>
<th>FP</th>
<th>CB</th>
<th>mAP:0.5</th>
<th>FPS</th>
</tr>
</thead>
<tbody>
<tr>
<td>YOLOv3 [38]</td>
<td>darknet53</td>
<td>3</td>
<td>-</td>
<td>84.0</td>
<td>5</td>
</tr>
<tr>
<td>YOLOv2 [36]</td>
<td>darknet19</td>
<td>-</td>
<td>-</td>
<td>78.5</td>
<td>6</td>
</tr>
<tr>
<td>YOLOv3-tiny [38]</td>
<td>tiny15</td>
<td>2</td>
<td>-</td>
<td>71.8</td>
<td>15</td>
</tr>
<tr>
<td>YOLO v2 SpeedUP [36]</td>
<td>darknet19</td>
<td>-</td>
<td>-</td>
<td>68.0</td>
<td>13</td>
</tr>
<tr>
<td>YOLT [49]</td>
<td>darknet19</td>
<td>-</td>
<td>-</td>
<td>85.6</td>
<td>5.5</td>
</tr>
<tr>
<td>FPN-tiny</td>
<td>tiny15</td>
<td>3</td>
<td>-</td>
<td>90.3</td>
<td>18</td>
</tr>
<tr>
<td>CFPN-1</td>
<td>~tiny15</td>
<td>2</td>
<td>1</td>
<td>91.3</td>
<td>33</td>
</tr>
<tr>
<td>CFPN-3</td>
<td>~tiny15</td>
<td>3</td>
<td>3</td>
<td>95.6</td>
<td>33</td>
</tr>
</tbody>
</table>
### Performance Comparisons on COCO Dataset

<table>
<thead>
<tr>
<th>Methods</th>
<th>Backbone</th>
<th>Train set</th>
<th>AP50</th>
<th>APS</th>
</tr>
</thead>
<tbody>
<tr>
<td>Faster R-CNN [31]</td>
<td>VGGNet-16</td>
<td>trainval</td>
<td>42.7</td>
<td>-</td>
</tr>
<tr>
<td>R-FCN [30]</td>
<td>ResNet-101</td>
<td>trainval</td>
<td>51.9</td>
<td>10.8</td>
</tr>
<tr>
<td>YOLOv2 [36]</td>
<td>DarkNet-19</td>
<td>trainval35k</td>
<td>44</td>
<td>5</td>
</tr>
<tr>
<td>YOLOv3-608[38]</td>
<td>DarkNet-53</td>
<td>trainval35k</td>
<td>50.2</td>
<td>16.9</td>
</tr>
<tr>
<td>SSD512 [32]</td>
<td>VGGNet-16</td>
<td>trainval35k</td>
<td>48.5</td>
<td>10.9</td>
</tr>
<tr>
<td>RefineDet320 [47]</td>
<td>ResNet-101</td>
<td>trainval35k</td>
<td>51.4</td>
<td>10.5</td>
</tr>
<tr>
<td>RefineDet512 [47]</td>
<td>VGGNet-16</td>
<td>trainval35k</td>
<td>54.5</td>
<td>16.3</td>
</tr>
<tr>
<td>PFPNet-S512 [34]</td>
<td>VGGNet-16</td>
<td>trainval35k</td>
<td>54.8</td>
<td>16.3</td>
</tr>
<tr>
<td>Proposed CFPN</td>
<td>Darknet-53</td>
<td>trainval35k</td>
<td>54.8</td>
<td>18.4</td>
</tr>
</tbody>
</table>
Discussion & Conclusion
- Outperforms the existing state-of-the-art models on HandFlow and COCO dataset
- The proposed CB reduces the computational cost but also improves the accuracy
- The model performance increases when the architecture integrates more CBs from the results of CFPN-1 and CFPN-3
Tha