A Solution for Densely Annotated Large Scale Object Detection Task

Yuan Gao, Hui Shen, Donghong Zhong, Jian Wang, Zeyu Liu, Ti Bai, Xiang Long and Shilei Wen
# Object 365 Dataset

<table>
<thead>
<tr>
<th>Pretrain</th>
<th># Class</th>
<th># Image</th>
<th># Box in Total</th>
<th>Box Num Avg</th>
<th>Image Height Avg</th>
<th>Image Width Avg</th>
<th>Box Area Avg (Pixel)</th>
<th>Max # Box</th>
</tr>
</thead>
<tbody>
<tr>
<td>COCO17 (Train)</td>
<td>80</td>
<td>118287</td>
<td>0.86M</td>
<td>7.27</td>
<td>484</td>
<td>577</td>
<td>12025</td>
<td>93</td>
</tr>
<tr>
<td>Object 365 (Train)</td>
<td>365</td>
<td>608606</td>
<td>9.62M</td>
<td>15.81</td>
<td>536</td>
<td>662</td>
<td>14074</td>
<td>835</td>
</tr>
</tbody>
</table>
R50 Cascade RCNN

Object365 Validation (mAP)

Baseline R50

22.73
Neural Architecture Search

- An RL based Neural Architecture Search is adopted.
- The NAS-FPN module is directly cascaded behind the original FPN module.
- A strong architecture found by prior knowledge\cite{Ghiasi2019} is used to initialize the NAS-FPN searching procedure.

Neural Architecture Search

The architecture graph of original FPN and NAS-FPN after ~400 episodes
Neural Architecture Search

Object365 Validation (mAP)

Baseline R50: 22.73
NASFPN: 23.66
Class Diversity Sensitive Sampling

15 Classes

4 Classes

Sampling probability equally is not appropriate.
The $i$th image contains $15$ classes.

$$W_i = \ln(C_i + \varepsilon) \sum_{c=1}^{N} P_c H_c$$

$W_i$ = Sampling weight of the $i$th image.

$C_i$ = Total number of the classes of the $i$th image.

$N$ = Total number of the classes of the dataset.

$P_c$ = The $c$th class prior probability, according to the total box number of the dataset.

$H_c = 1$ if the images contains class $c$ or 0.
Class Diversity Sensitive Sampling

Object365 Validation (mAP)

![Graph showing the comparison between Random and Class Diversity Sensitive sampling methods over iterations.](image-url)
Class Diversity Sensitive Sampling

Object365 Validation (mAP)

<table>
<thead>
<tr>
<th>Method</th>
<th>mAP</th>
</tr>
</thead>
<tbody>
<tr>
<td>Baseline R50</td>
<td>22.73</td>
</tr>
<tr>
<td>NASFPN</td>
<td>23.66</td>
</tr>
<tr>
<td>CDSS</td>
<td>25.01</td>
</tr>
<tr>
<td>SENet154+GN</td>
<td>30.1</td>
</tr>
<tr>
<td>OHEM+Deformable</td>
<td>30.7</td>
</tr>
</tbody>
</table>

Baseline R50, NASFPN, CDSS, SENet154+GN, OHEM+Deformable
Large Resolution Box Head

Roll-Align

7X7

Box

Class

9X9

Box

Class
Large Resolution Box Head

Object365 Validation (mAP)

- Baseline R50: 22.73
- NASFPN: 23.66
- CDSS: 25.01
- SPNet154+GN: 30.1
- OHEM+Deformable: 30.7
- Large Resolution Head: 30.9
• Use the predicted bbox of the 2nd stage to extract the feature. (Standard Cascade RCNN)

Cascade RCNN Adaptive Testing

- Use the predicted bbox of each stage itself to extract the feature.
Cascade RCNN Adaptive Testing

Object365 Validation (mAP)

- Baseline R50: 22.73
- NASFPN: 23.66
- CDSS: 25.01
- SENet154+GN: 30.1
- OHEM+Deformable: 30.7
- Large Resolution Head: 30.9
- Adaptive Cascade Testing: 31.1
- MS Training and Testing: 33.1

Validation
• Use COCO Pretrained model, mAP 52.9 on COCO17 minival.
• Training multiscale size (400, 1400), max size 1600.
• Testing multiscale size (400, 1400), max size 2100.
• 8 V100(32GB) x 2 for 7 days.
• Weight Standardization brings model diversity.
• SoftNMS is adopted.
Implementation Details

Object365 Validation (mAP)

- Baseline R50
- NASFPN
- CDSS
- SENet154+GN
- OHEM+Deformable
- Large Resolution Head
- Adaptive Cascade Testing
- MS Training and Testing
- Ensemble 5 models

Validation
Visualization
Visualization
Tiny Track
## Full Track Pretrain

<table>
<thead>
<tr>
<th>Pretrain</th>
<th>Full Val mAP</th>
<th>Tiny Val mAP</th>
<th>Gain</th>
<th>Tiny Test mAP</th>
</tr>
</thead>
<tbody>
<tr>
<td>COCO Pretrain</td>
<td>-</td>
<td>28.9</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>Obj365 Full Pretrain</td>
<td>30.7</td>
<td>33.0</td>
<td>+4.1</td>
<td>-</td>
</tr>
<tr>
<td>Obj365 Full Pretrain</td>
<td>32.9</td>
<td>34.8</td>
<td>+5.9</td>
<td>-</td>
</tr>
<tr>
<td>Ensemble 8 models</td>
<td>-</td>
<td>37.6</td>
<td>+8.7</td>
<td>29.0</td>
</tr>
</tbody>
</table>

- Multiscale input with flip in Training and Testing
Visualization

- Full Track Pretrained
- COCO17 Pretrained
Paddle Paddle Detection

• Fast/Faster R-CNN, FPN, Mask RCNN, Cascade R-CNN, Yolo v3, RetinaNet, SSD ......

• GN, SyncBN, Deformable Conv v1/v2 ......

• https://github.com/PaddlePaddle/models/tree/develop/PaddleCV/object_detection

• Training framework will be released soon.
Thank you!

Please feel free to contact us, if you have any questions.

gaoyuan18@baidu.com