Deformable Convolutional Networks
-- MSRA COCO Detection & Segmentation Challenge 2017 Entry

Jifeng Dai
With Haozhi Qi*, Zheng Zhang, Bin Xiao, Han Hu, Bowen Cheng*, Yichen Wei
Visual Computing Group
Microsoft Research Asia
(*interns at MSRA)
Outline

• Deformable ConvNets idea

• Deformable ConvNets for COCO challenge
Highlights

• Enabling effective modeling of spatial transformation in ConvNets

• No additional supervision for learning spatial transformation

• Significant accuracy improvements on sophisticated vision tasks

Code is available at https://github.com/msracver/Deformable-ConvNets
Modeling Spatial Transformations

• A long standing problem in computer vision
  - Deformation:
  - Scale:
  - Viewpoint variation:
  - Intra-class variation:

(Some examples are taken from Li Fei-fei’s course CS223B, 2009-2010)
Traditional Approaches

• 1) To build training datasets with sufficient desired variations

• 2) To use transformation-invariant features and algorithms

Drawbacks: geometric transformations are assumed fixed and known, hand-crafted design of invariant features and algorithms
Spatial Transformations in CNNs

• Regular CNNs are inherently limited to model large unknown transformations
  • The limitation originates from the fixed geometric structures of CNN modules
Spatial Transformer Networks

- Learning a global, parametric transformation on feature maps
  - Prefixed transformation family, infeasible for complex vision tasks
Deformable Convolution

• Local, dense, non-parametric transformation
  • Learning to deform the sampling locations in the convolution/RoI Pooling modules
Deformable Convolution

Regular convolution
\[ y(p_0) = \sum_{p_n \in R} w(p_n) \cdot x(p_0 + p_n) \]

Deformable convolution
\[ y(p_0) = \sum_{p_n \in R} w(p_n) \cdot x(p_0 + p_n + \Delta p_n) \]

where \( \Delta p_n \) is generated by a sibling branch of regular convolution.
Deformable RoI Pooling

Regular RoI pooling
\[ y(i, j) = \sum_{p \in \text{bin}(i, j)} x(p_0 + p)/n_{ij} \]

Deformable RoI pooling
\[ y(i, j) = \sum_{p \in \text{bin}(i, j)} x(p_0 + p + \Delta p_{ij})/n_{ij} \]

where \( \Delta p_{ij} \) is generated by a sibling fc branch
Deformable ConvNets

• Same input & output as the plain versions
  • Regular convolution -> deformable convolution
  • Regular RoI pooling -> deformable RoI pooling

• End-to-end trainable without additional supervision
Sampling Locations of Deformable Convolution

(a) standard convolution

(b) deformable convolution
Part Offsets in Deformable RoI Pooling
Deformable ConvNets for Object Detection

- Regular object detectors
Deformable ConvNets for Object Detection

- Deformable object detectors
XCeption -> Aligned XCeption

- Proper feature alignment in XCeption
  - Efficient: 9.5 GFLOPS on 224*224 img (ResNet-101, 7.6 GFLOPS)
  - Accurate: mAP 2.8% better than ResNet-101 using FPN on COCO (det, test-dev)
Object Detection on COCO (Test-dev)

- MSRA 2017 Entry
  - ~3% mAP improvements by Deformable ConvNets
  - Best single model performance: 48.5%
Object Detection on COCO (Test-dev)

- Deformable ConvNets v.s. regular ConvNets
  - Noticeable improvements for varies baselines
  - Marginal parameter & computation overhead

<table>
<thead>
<tr>
<th>Model</th>
<th>mAP (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Deformable ConvNets</td>
<td></td>
</tr>
<tr>
<td>Regular ConvNets</td>
<td></td>
</tr>
<tr>
<td>CLASS-AWARE RPN (RESNET-101)</td>
<td>23.2</td>
</tr>
<tr>
<td>FASTER R-CNN, 2FC (RESNET-101)</td>
<td>25.8</td>
</tr>
<tr>
<td>R-FCN (RESNET-101)</td>
<td>32.1</td>
</tr>
<tr>
<td>R-FCN (ALIGNED-INCEPTION-RESNET)</td>
<td>34.5</td>
</tr>
<tr>
<td>FPN+OHEM (RESNET-101)</td>
<td>37.4</td>
</tr>
<tr>
<td>FPN+OHEM (ALIGNED-XCEPTION)</td>
<td>40.2</td>
</tr>
<tr>
<td>FPN++ (ALIGNED-XCEPTION)</td>
<td>45.2</td>
</tr>
<tr>
<td>FPN++ (ALIGNED-XCEPTION)</td>
<td>48.5</td>
</tr>
</tbody>
</table>
Conclusion

• Deformable ConvNets for dense spatial modeling
  • Simple, efficient, deep, and end-to-end
  • No additional supervision
  • Feasible and effective on sophisticated vision tasks for the first time

• Our team

Haozhi Qi*  Zheng Zhang  Bin Xiao  Han Hu  Bowen Cheng*  Yichen Wei  Jifeng Dai

* interns at MSRA