Multi-Person Pose Estimation using Part Affinity Fields

Zhe Cao, Shih-En Wei, Tomas Simon, Yaser Sheikh
Carnegie Mellon University
Earlier This Year…

CVPR’16 Demo
Problem of Shipment

CVPR’16  Demo
Top-down Approach: Person Detection + Pose Estimation

Top-down
Top-down Approach: Person Detection + Pose Estimation
Top-down Approach: Person Detection + Pose Estimation
Top-down Approach: Person Detection + Pose Estimation
Top-down Approach: Person Detection + Pose Estimation
Our Method: Parts Detection + Parts Association

Top-down

Ours
Our Method: Parts Detection + Parts Association
Our Method: Parts Detection + Parts Association
Our Method: Parts Detection + Parts Association

Top-down

Ours
Our Method: Parts Detection + Parts Association

Top-down
Our Method: Parts Detection + Parts Association

Top-down

Part Affinity Fields
Our Method: Parts Detection + Parts Association

Top-down
Our Method: Parts Detection + Parts Association

Top-down

Part Affinity Fields
Our Method: Parts Detection + Parts Association

Top-down

Ours
Novelty: Jointly Learning Parts Detection and Parts Association
Novelty: Jointly Learning Parts Detection and Parts Association
Sequential Prediction with Learned Spatial Context

Stage 1

Input Image

Right shoulder

Right wrist

Right knee

:\n
Convolutional Pose Machines, Wei, Ramakrishna, Kanade, Sheikh, CVPR 2016
Sequential Prediction with Learned Spatial Context

Stage 1

Convolutional Pose Machines, Wei, Ramakrishna, Kanade, Sheikh, CVPR 2016
Sequential Prediction with Learned Spatial Context

Stage 1

CNN

Right Wrist - Stage 1
Sequential Prediction with Learned Spatial Context

Stage 1

Stage 2

Right Wrist - Stage 1
Sequential Prediction with Learned Spatial Context

Stage 1

Stage 2

Input Image

Right Wrist - Stage 1

Right Wrist - Stage 2
Sequential Prediction with Learned Spatial Context

Stage 1

![CNN](image1)

Stage 2

![CNN](image2)

Stage T

![CNN](image3)

Input Image

Right Wrist - Stage 1

Right Wrist - Stage 2

Right Wrist - Stage T
Parts Score Maps Prediction from Image Sequence

Input

Nose

Neck

Right Shoulder

Right Elbow

Right Wrist
Jointly Learning Parts Detection and Parts Association
Part-Person Association for Multi-Person Pose Estimation

Part detections

Elbow
Wrist
Part-Person Association for Multi-Person Pose Estimation

Part detections

Pose
Part-to-Part Association for Multi-Person Pose Estimation

- Elbow
- Wrist
Part Affinity Score Guides the Connection
Part Affinity Score Guides the Connection
Part Affinity Score Guides the Connection
How to Obtain the Part Affinity Score

Elbow
Wrist
Part Affinity Score is Dependent on Visual Appearance
Part Affinity Score is Dependent on Visual Appearance
Key Idea: Encode the Part Affinity Score on the Image Plane

Part Affinity Fields encode **direction** and **position**
Key Idea: Encode the Part Affinity Score on the Image Plane

Part Affinity Fields encode **direction** and **position**
Midpoint Score Map for Part-to-Part Association

part1

part2
Midpoint Score Map for Part-to-Part Association
Midpoint Score Map for Part-to-Part Association

Affinity score between part1 and part2 = confidence score of the midpoint
Spatial Ambiguity of the Midpoint Representation — Correct Connection
Spatial Ambiguity of the Midpoint Representation

— Correct Connection
— Wrong Connection
Spatial Ambiguity of the Midpoint Representation

— Correct Connection
— Wrong Connection
Increasing Midpoint Number Cannot Solve The Problem

— Correct Connection
— Wrong Connection
Part Affinity Fields for Part-to-Part Association
Part Affinity Fields for Part-to-Part Association

Affinity score between $p_1$ and $p_2$

$= \text{sum}( \vec{v} \cdot p_1 \vec{p}_2 )$
Part Affinity Fields Avoid Spatial Ambiguity
Jointly Learning Parts Detection and Parts Association

Stage 1 → CNN → $P$

Stage 2 → CNN → $P$

Stage $T$ → CNN → $P$
Jointly Learning Parts Detection and Parts Association

Stage 1

2nd Branch
Part Affinity
Fields

Stage 2

Stage T

CNN

CNN

CNN

Stage 1, level 1

Stage n, (n²)

Stage 1, level 2

Stage n, level1

Loss

Loss

Loss

Jointly Learning Parts
Detection and Parts Association
Greedy Algorithm for Body Parts Association
Greedy Algorithm for Body Parts Association

Elbow

Wrist
Greedy Algorithm for Body Parts Association
Greedy Algorithm for Body Parts Association
## Results on COCO Challenge Validation Set

<table>
<thead>
<tr>
<th>Method</th>
<th>AP on val</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>GT bbox + CPM [1]</strong></td>
<td>63</td>
</tr>
<tr>
<td><strong>SSD [2] + CPM [1]</strong></td>
<td>53</td>
</tr>
<tr>
<td><strong>Our Method</strong></td>
<td>58.5</td>
</tr>
<tr>
<td><strong>Ours + Refinement</strong></td>
<td>61</td>
</tr>
</tbody>
</table>

Jointly Learning Parts Detection and Parts Association

Image → CNN

Parts Detection → CNN

Parts Association → CNN
Multi-Person Pose Estimation using Part Affinity Fields

Zhe Cao, Shih-En Wei, Tomas Simon, Yaser Sheikh
Carnegie Mellon University

Please come to our demo session at D-2A-49 (the 4th floor under oral paper posters) on Wednesday, October 12; 11:00 – 12:30