Multi-person Joint Detection and Grouping (MJDG)

Sheng Jin, Wentao Liu
Tsinghua University

Outline

- Introduction
- Review
- Joint Detection and Grouping
- Visualizations
- Experiments

Outline

- **Introduction**
- Review
- Joint Detection and Grouping
- Visualizations
- Experiments

Introduction

Close Proximity

Occlusion

Rare Poses

Outline

- Introduction
- **Review**
- Joint Detection and Grouping
- Visualizations
- Experiments

Top-Down

Human Detection + Single Person Pose Estimation

Pros:

1. Use global context

Cons:

1. imperfect human detector
2. Isolated limbs or occlusion
3. Two-stage pipeline

Bottom-up

Body part detection + Grouping

Pros:

1. Robust to occlusion and rare poses.
2. Enable single-stage end-to-end prediction

Cons:

1. Lacking in global constraints

Top-Down

[1] G-RMI

[2] Mask-RCNN

[3] RMPE

Bottom-up

[4] DeeperCut

[5] PAF (CMU-Pose)

[6] Associative Embedding

[1] G. Papandreou, T. Zhu, N. Kanazawa, A. Toshev, J. Tompson, C. Bregler, and K. Murphy. Towards accurate multi-person pose estimation in the wild. arXiv preprint arXiv:1701.01779, 2017

[2] K. He, G. Gkioxari, P. Dollar, and R. Girshick. Mask r-cnn. arXiv preprint arXiv:1703.06870, 2017

[3] H. Fang, S. Xie, Y. Tai, and C. Lu. Rmpe: Regional multi-person pose estimation. arXiv preprint arXiv:1612.00137, 2016

[4] E. Insafutdinov, L. Pishchulin, B. Andres, M. Andriluka, and B. Schiele. Deepercut: A deeper, stronger, and faster multi-person pose estimation model. In European Conference on Computer Vision (ECCV), 2016

[5] Z. Cao, T. Simon, S. Wei, and Y. Sheikh. Realtime multi-person 2d pose estimation using part affinity fields. arXiv preprint arXiv:1611.08050, 2016

[6] A. Newell and J. Deng. Associative embedding: End-to-end learning for joint detection and grouping. In NIPS, 2017

Outline

- Introduction
- Review
- **Joint Detection and Grouping**
- Visualizations
- Experiments

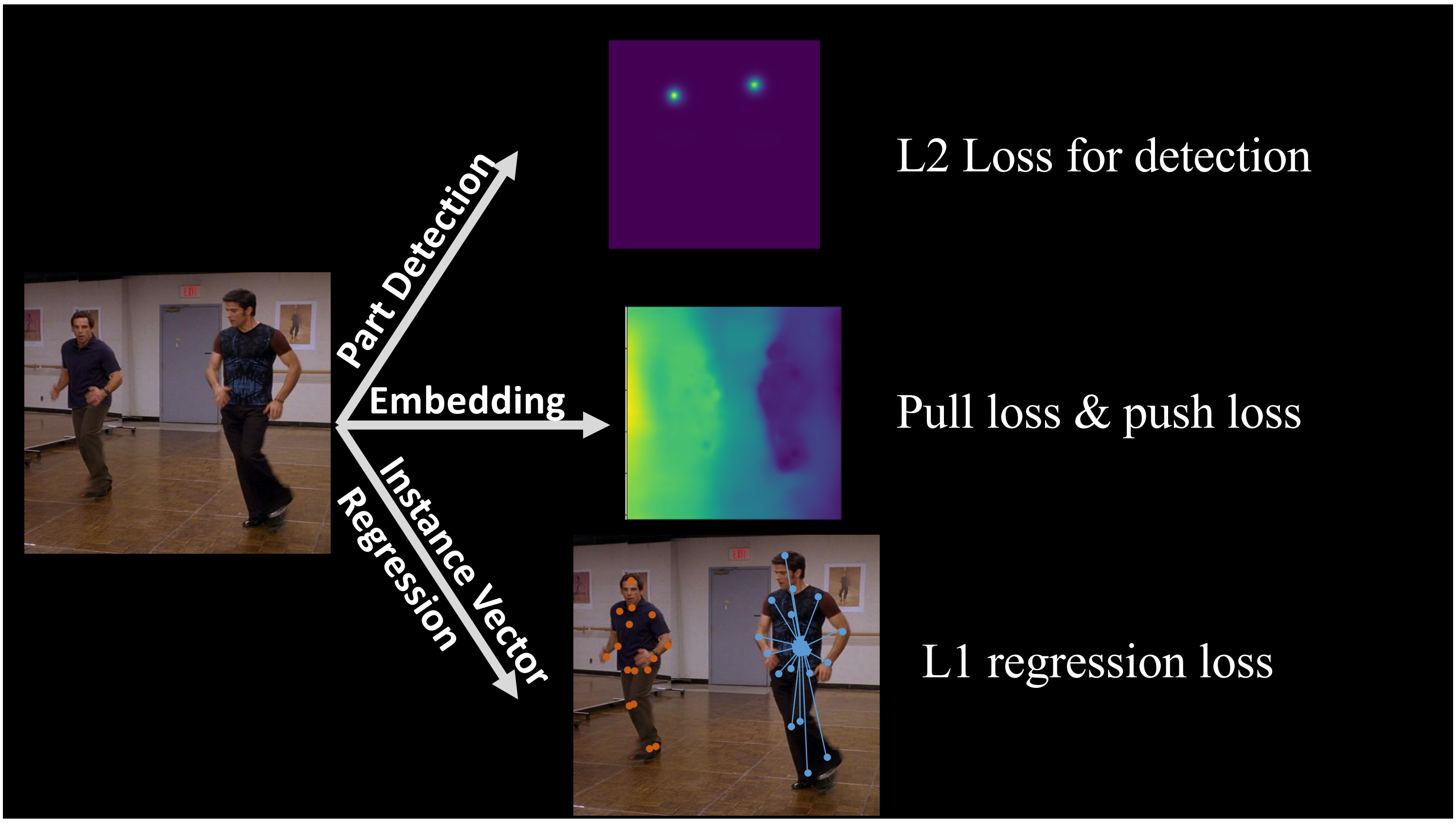
Joint Detection and Grouping

MJDG = Associative Embedding + Weakly-supervised instance segmentation

“weakly-supervised” human instance segmentation ----- using only keypoint supervision

Pose Estimation ←→ instance segmentation [7]

[7] A. Kendall, Y. Gal and R. Cipolla. Multi-Task Learning Using Uncertainty to Weigh Losses for Scene Geometry and Semantics. arXiv preprint arXiv:1705.07115, 2017


Part Detection
Embedding
Instance Vector Regression
L2 Loss for detection
Pull loss & push loss
L1 regression loss

Detection Loss

$$L_d = \sum_{j=1}^{J} \sum_{p} M(p) \left\| S_j(p) - S_j^*(p) \right\|_2^2$$

Grouping Loss

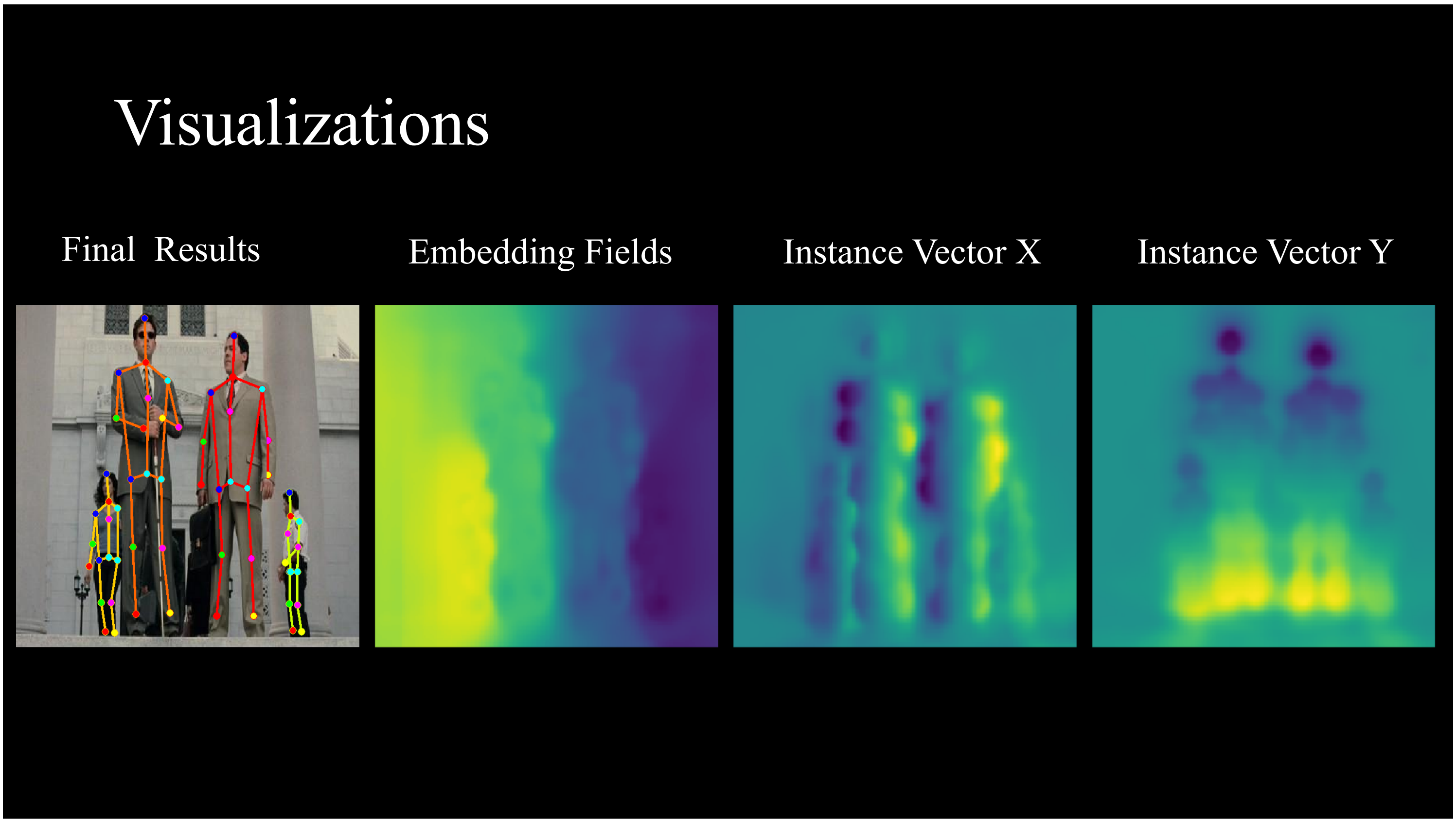
$$L_g = \frac{1}{N} \sum_{n} \sum_{j} \left(\bar{h}_n - h_j\left(x_{nj},\right) \right)^2 + \frac{1}{N^2} \sum_{n} \sum_{n'} \exp\{-\frac{1}{2\sigma^2}\left(\bar{h}_n - \bar{h}_{n'}\right)^2\}$$

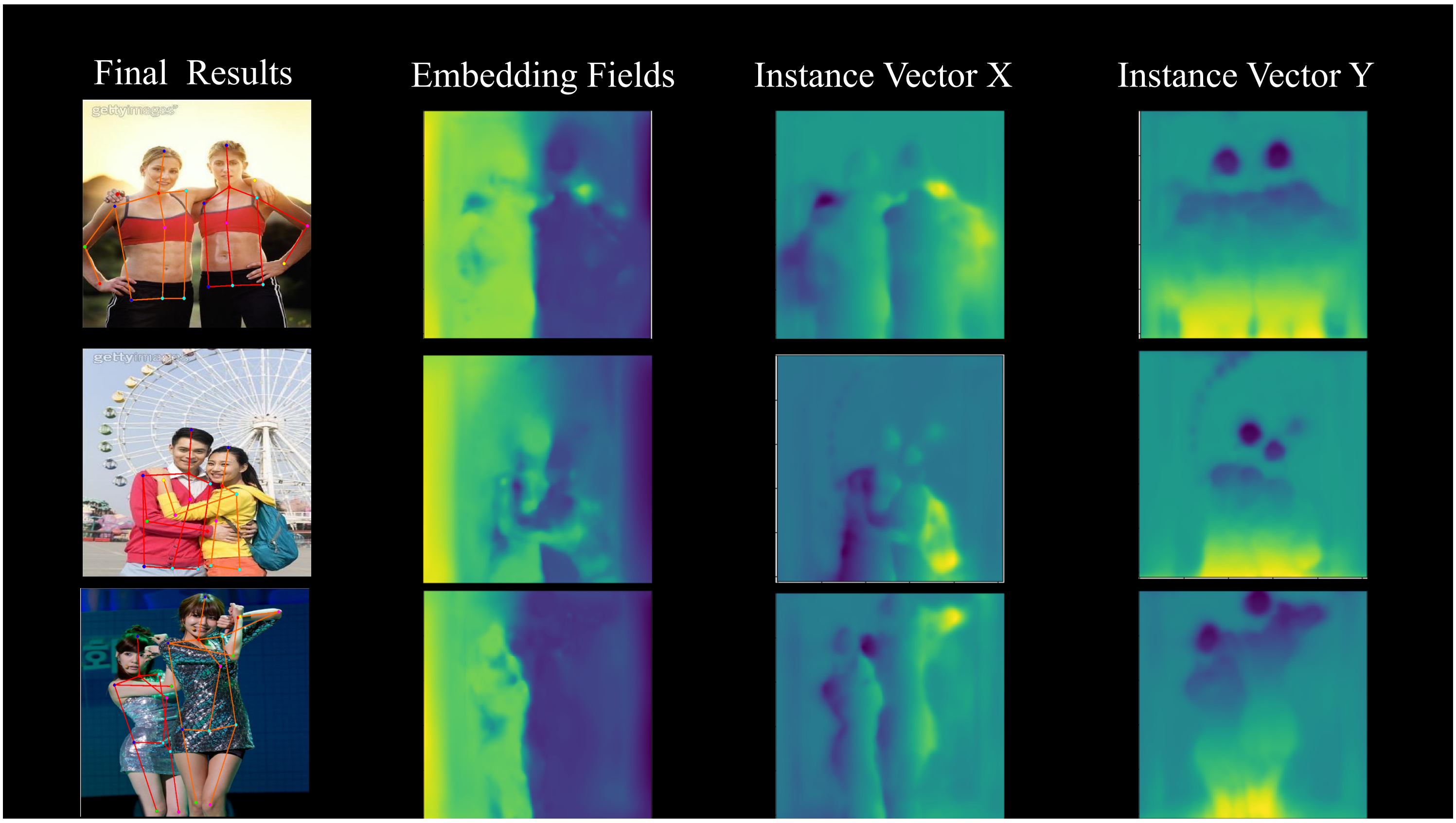
Instance Vector Regression Loss

$$L_s = \sum_{n} \sum_{j=1}^{J} \left\| R_{n,j} - R_n^* \right\|_1$$

Outline

- Introduction
- Review
- Joint Detection and Grouping
- **Visualizations**
- Experiments


Visualizations
Final Results
Embedding Fields
Instance Vector X
Instance Vector Y


Final Results
Embedding Fields
Instance Vector X
Instance Vector Y

Outline

- Introduction
- Review
- Joint Detection and Grouping
- Visualizations
- **Experiments**

Implementation Details

Note:

(1) Backbone network --- 4-stage stacked hourglass.

(2) Train from scratch using MHP dataset only.

(3) Single model without extra refinement.

(4) Multi-scale testing (2, 1.5, 1.25, 1, 0.75, 0.5)

(5) We use a weighted sum of losses. The weight of detection, grouping and instance vector regression loss is 1:1e-3:1e-4 respectively.

Table1. Pose Estimation Results on MHP validation dataset

Methods	Head	Shoulder	Elbow	Wrist	Hip	Knee	Ankle	Total
MJDG-seg	82.0	87.3	79.6	71.0	60.4	73.5	74.0	75.9
MJDG	82.3	87.6	79.7	71.5	61.2	74.0	74.6	76.3

Table2. Pose Estimation Results on MHP test dataset

Methods	Head	Shoulder	Elbow	Wrist	Hip	Knee	Ankle	Total
Baseline	68.1	69.2	61.9	58.3	38.6	51.2	43.5	55.8
RNG	29.1	71.9	71.0	67.2	44.1	63.0	58.3	57.8
OSU-Human	73.2	67.0	62.8	63.9	45.2	55.3	47.2	59.2
MJDG	85.8	78.5	74.2	73.9	54.1	64.0	58.7	69.9
JDAI-Human	85.0	79.2	76.0	75.3	59.2	68.3	62.3	72.2

Multi-person Pose Estimation on Other Datasets

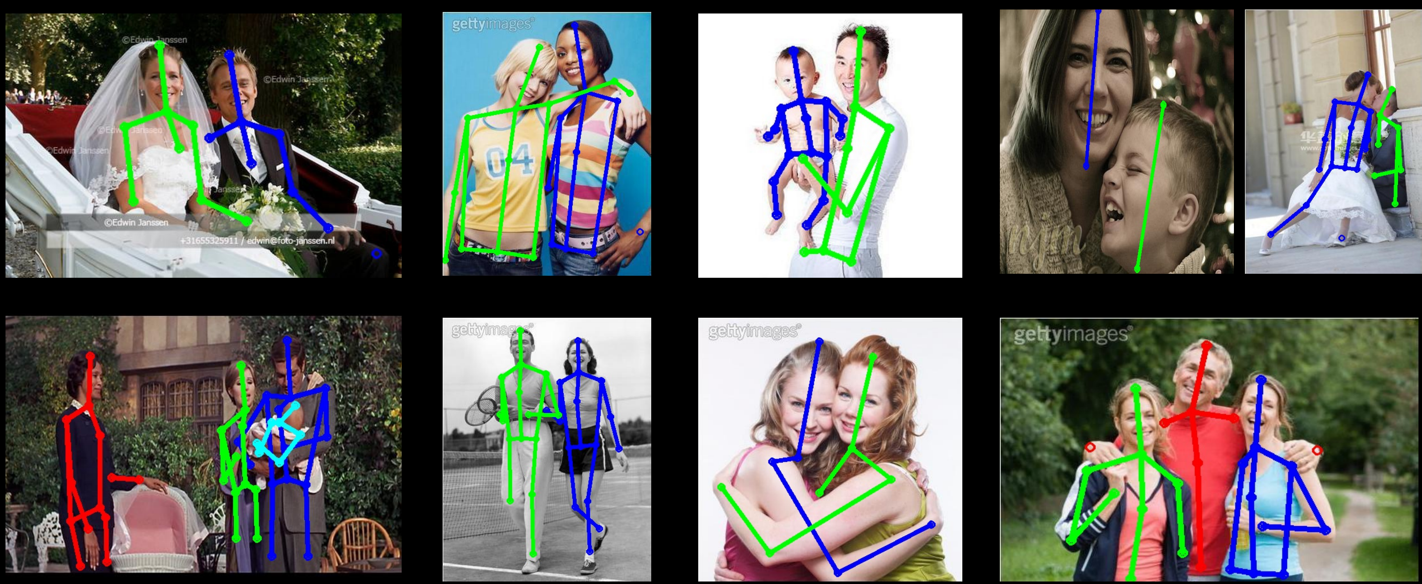
Table3. Pose Estimation on COCO mini-val

Methods	mAP
PAF (our implementation)	61.1
Mask R-CNN (our implementation)	63.1
MJDG	68.9

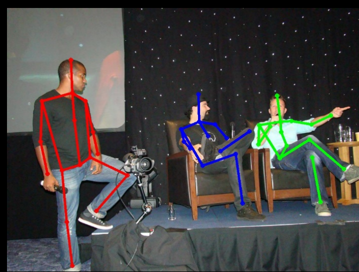
Table4. Pose Estimation on PoseTrack-val

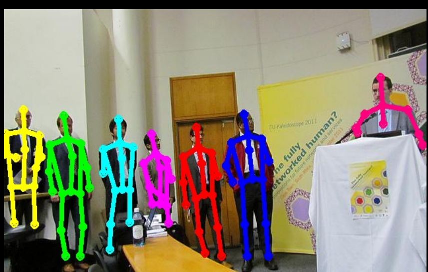
Methods	mAP
Mask R-CNN (our implementation)	57.4
PAF (our implementation)	67.8
MJDG	72.3

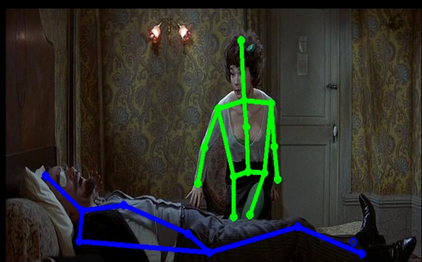
Close Proximity

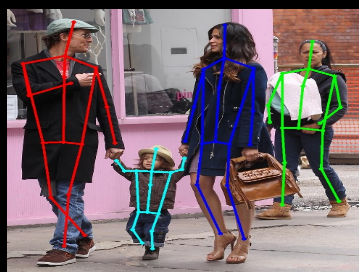


Occlusion



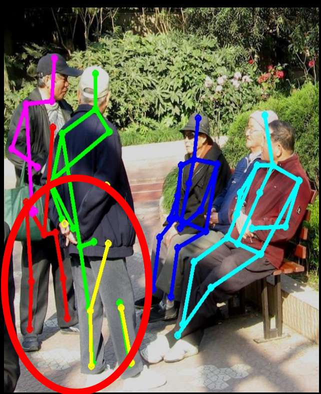






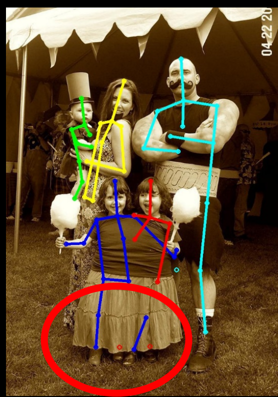
Rare Poses

Failure Cases



(a)

(b)



(c)

(d)

Thanks for your attention!