CVPR LONG BEACH CALIFORNIA June 16-20, 2019

Team ByteDance-SEU

CVPR LONG BEACH CALIFORNIA June 16-20, 2019

1st place

in the **Single-person Human Pose Estimation** Track of CVPR-19 LIP Challenge

Speaker: Kai Su

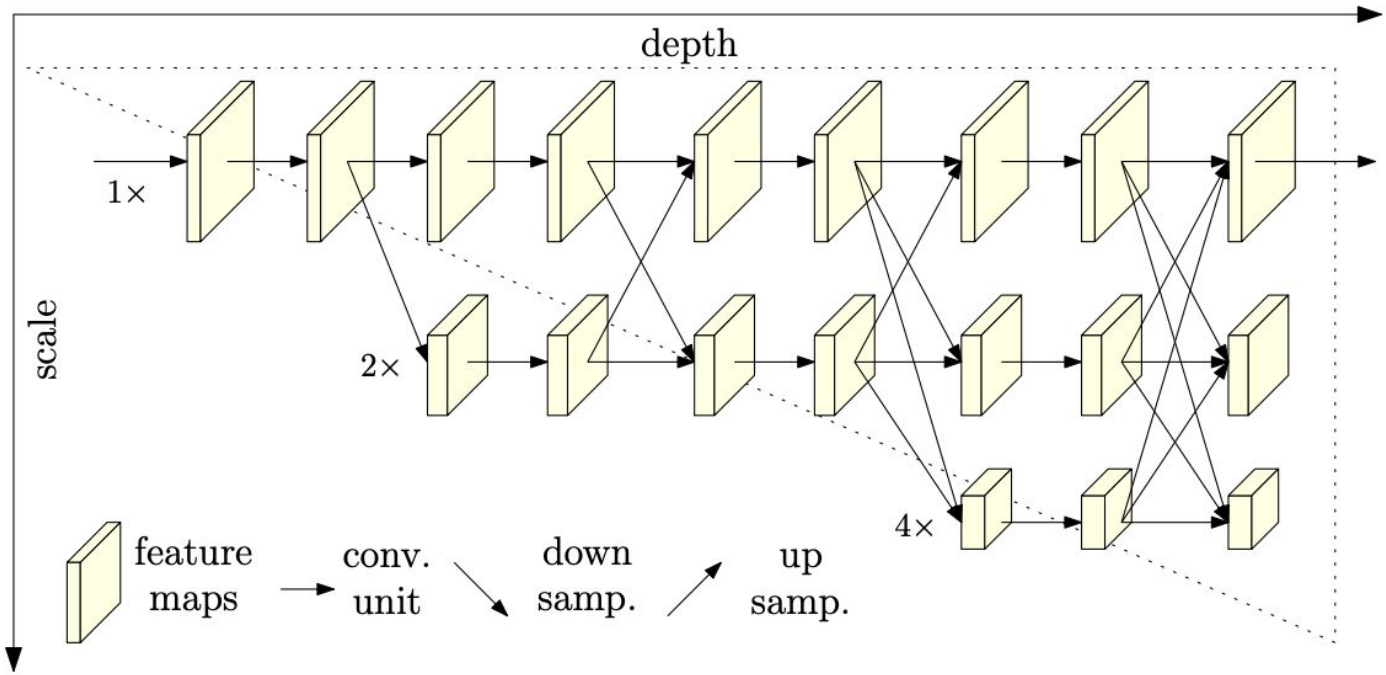
Kai Su[*,1,2], Dongdong Yu[*,1], Xin Geng[2], Changhu Wang[1]
[1]ByteDance AI Lab [2]Southeast University
(* Equal Contribution)

ByteDance

Outline

- Human Pose Estimation Networks
- Datasets
- Experiments
- Summary

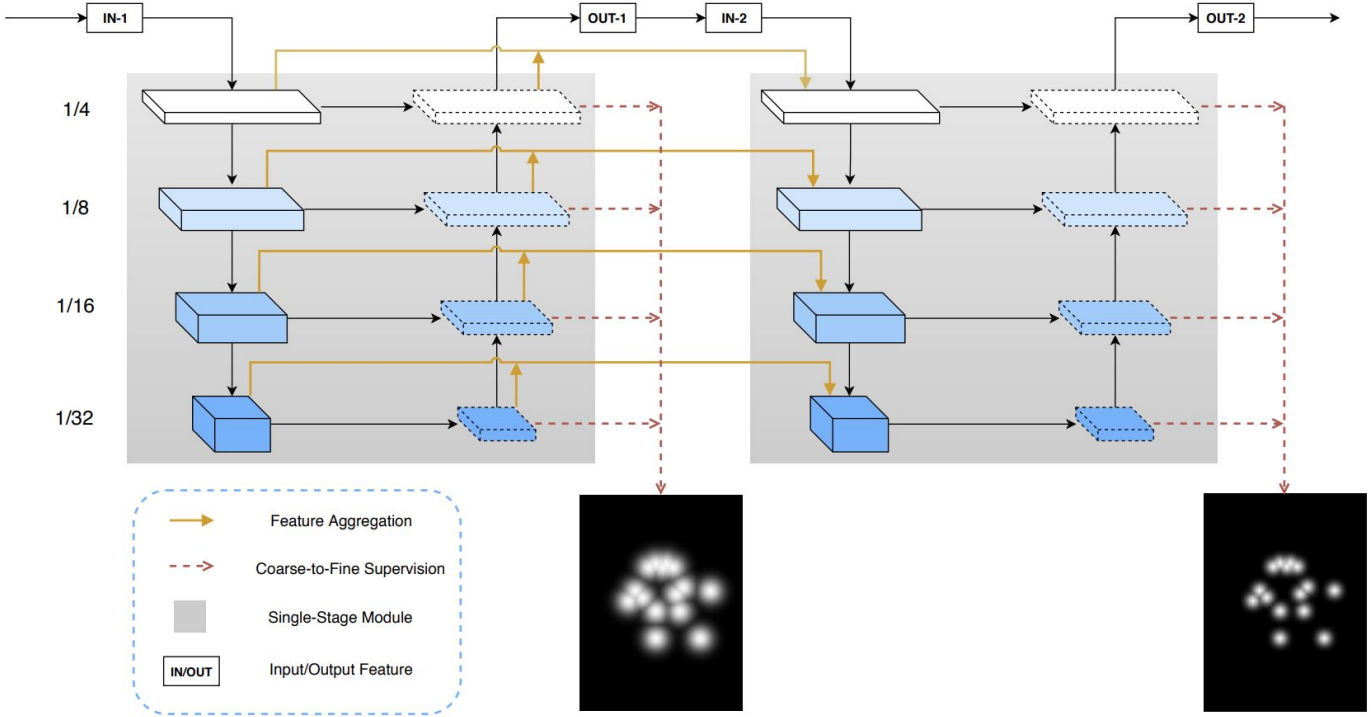
High-Resolution Net (HRNet)



Sun, Ke, et al. "Deep High-Resolution Representation Learning for Human Pose Estimation." arXiv preprint arXiv:1902.09212 (2019).

ByteDance

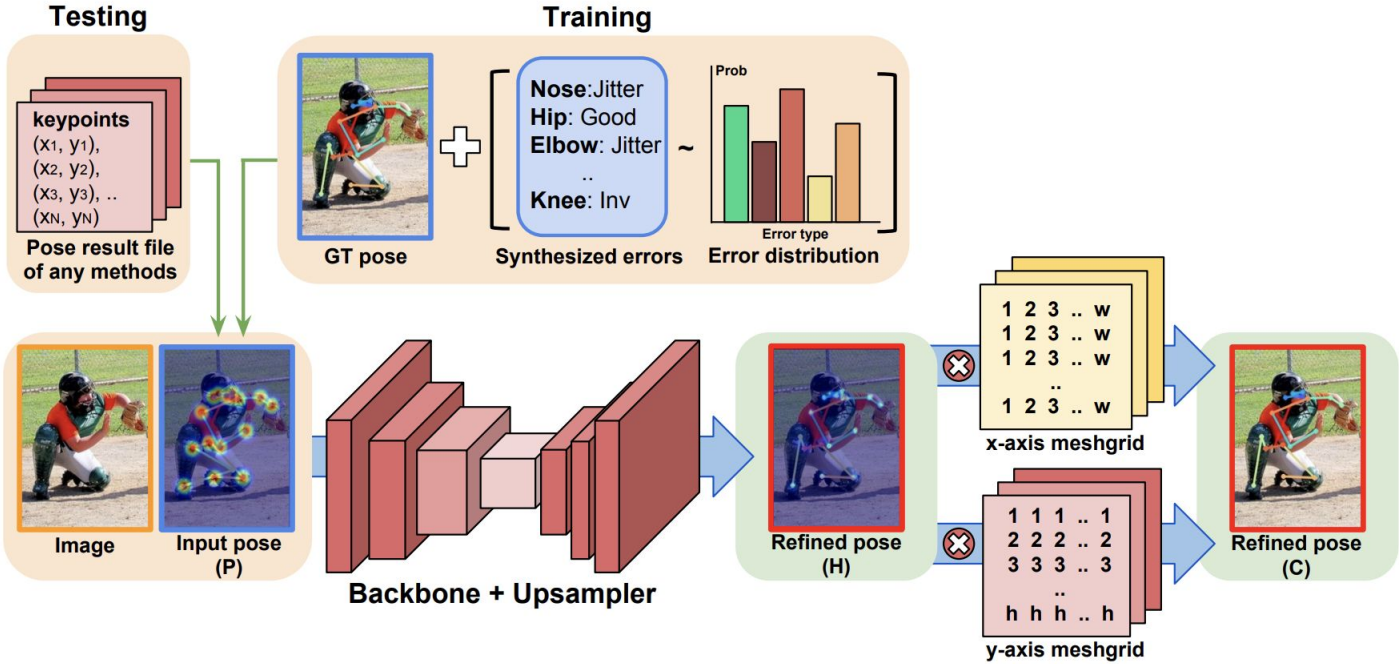
Multi-Stage Networks (MSPN)



Li, Wenbo, et al. "Rethinking on Multi-Stage Networks for Human Pose Estimation." arXiv preprint arXiv:1901.00148 (2019).

ByteDance

Human Pose Refinement Network (PoseFix)



Moon, Gyeongsik, Ju Yong Chang, and Kyoung Mu Lee. "PoseFix: Model-agnostic General Human Pose Refinement Network." arXiv preprint arXiv:1812.03595 (2018).

ByteDance

Outline

- Human Pose Estimation
- Datasets
- Experiments
- Summary

Datasets

Dataset	Number of images	Number of Keypoints
LIP	3w+ for training, 1w for validation, 1w for test	16
COCO	14w+ for training	17
AI Challenge	20w+ for training	14

Outline

- Human Pose Estimation
- Datasets
- Experiments
- Summary

Baseline

Method	Input Size	Training Datasets	PCKh on LIP test set
HRNet-w32	384*288	LIP train	**88.2**

External Dataset

Method (HRNet-w32)	Input Size	PCKh on LIP test set
with LIP train (baseline)	384*288	88.2
+ COCO	384*288	91.2 ↑3.0
+ AI Challenge	384*288	91.8 ↑0.6
+ LIP validation	384*288	**91.8** ↑0.0

External Dataset

Method (MSPN)	Input Size	PCKh on LIP test set
with COCO + LIP train (baseline)	384*288	91.4
+ AI Challenge	384*288	91.8 ↑0.4
+ LIP validation	384*288	**91.9** ↑0.1

Input Size

Method (HRNet-32)	Training Datasets	PCKh on LIP test set
384*288	All Available Datasets	91.8
+ 384*288 → 512*384	All Available Datasets	**92.0** ↑0.2

Method (MSPN)	Training Datasets	PCKh on LIP test set
384*288	All Available Datasets	91.9
+ 384*288 → 512*384	All Available Datasets	**92.0** ↑0.1

Ensemble

Method	Training Datasets	PCKh on LIP test set
HRNet-w32	All Available Datasets	92.0
+ MSPN	All Available Datasets	92.4 ↑0.4
+ PoseFix	LIP train/val + COCO	**92.6** ↑0.2

Outline

- Human Parsing Networks
- Datasets
- Experiments
- Summary

Summary

Baseline Result: 88.2

+ external dataset 88.2 -> 91.8 (+3.6)
+ enlarge the input size 91.8 -> 92.0 (+0.2)
+ ensemble with average heatmaps 92.0 -> 92.6 (+0.6)

Bad Cases

CVPR LONG BEACH CALIFORNIA June 16-20, 2019

2nd place

in the **Single-person Human Parsing** Track of CVPR-19 LIP Challenge

Speaker: Dongdong Yu

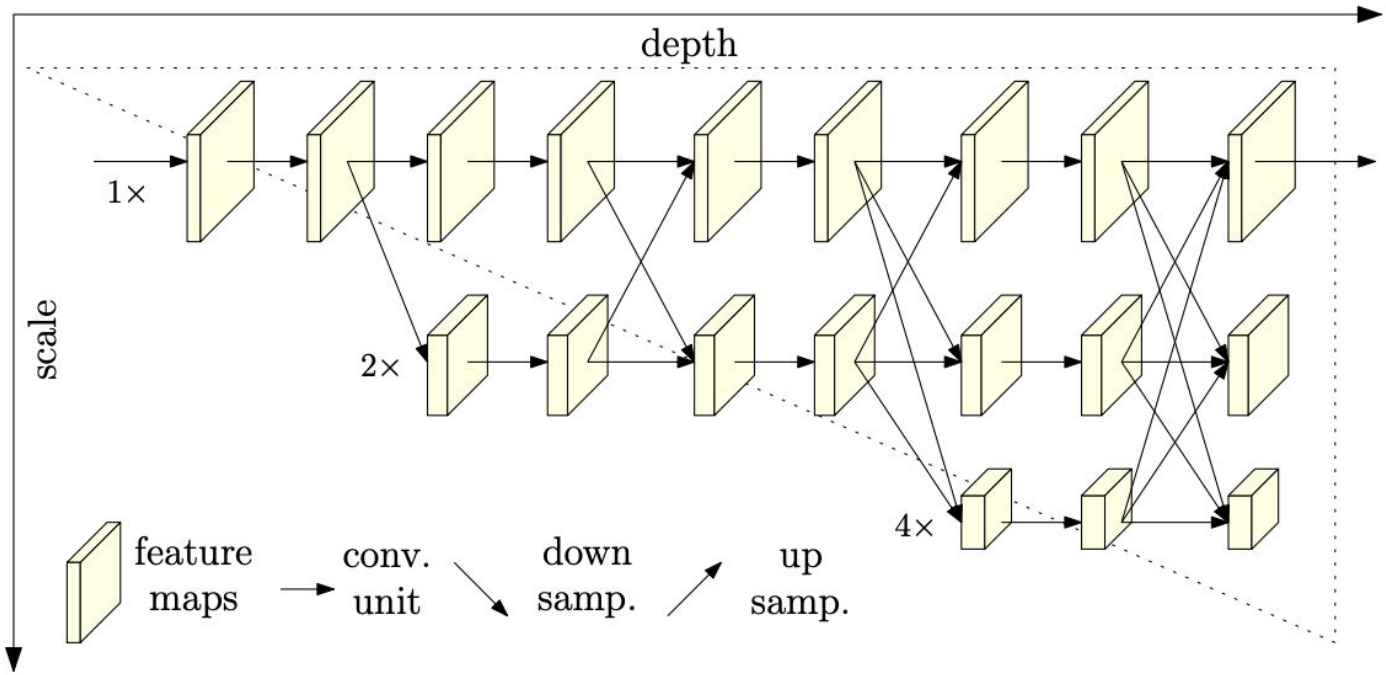
Dongdong Yu[*,1], Kai Su[*,1,2], Jian Wang[1], Kaihui Zhou[1], Xin Geng[2], Changhu Wang[1]
[1]ByteDance AI Lab [2]Southeast University
(* Equal Contribution)

ByteDance

Outline

- Human Parsing Networks
- Datasets
- Experiments
- Summary

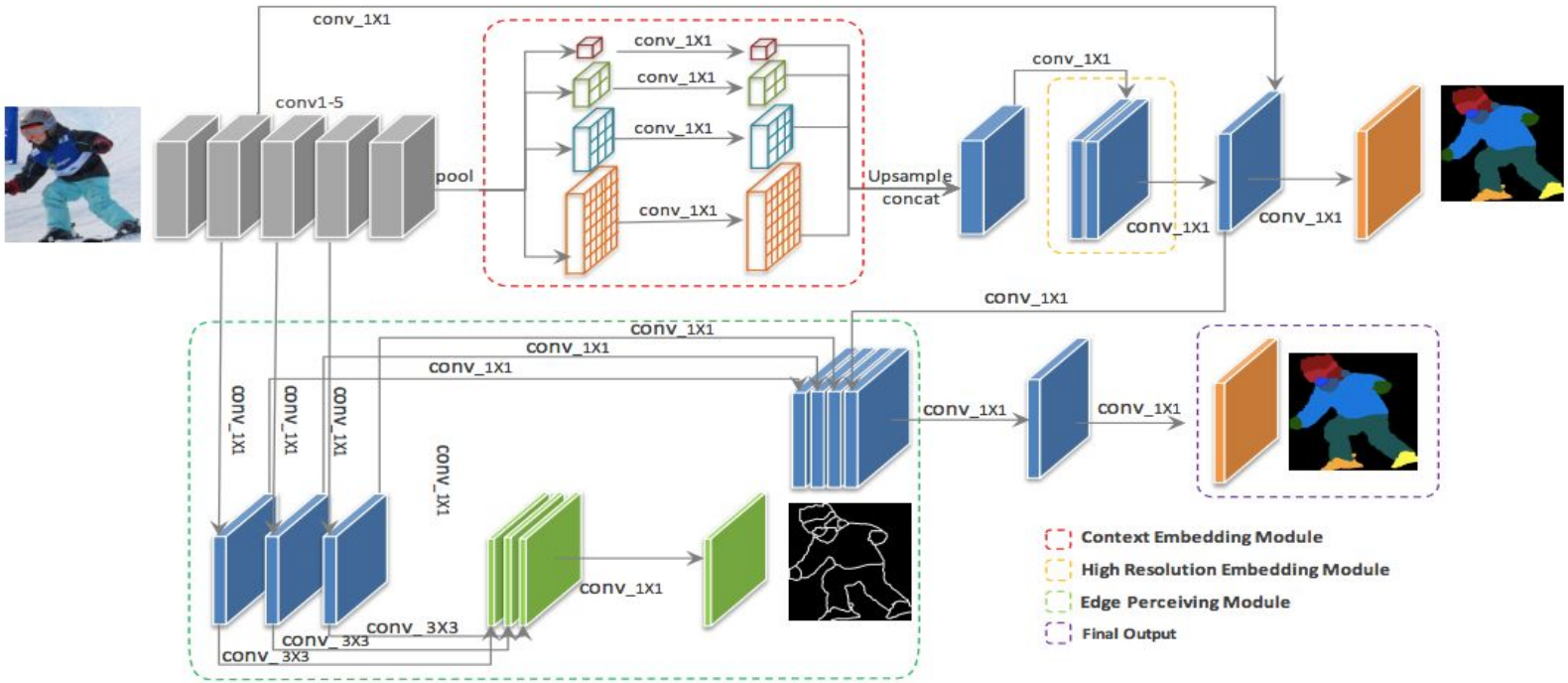
High-Resolution Net (HRNet)



Sun, Ke, et al. "High-Resolution Representations for Labeling Pixels and Regions." arXiv preprint arXiv:1904.04514 (2019).

ByteDance

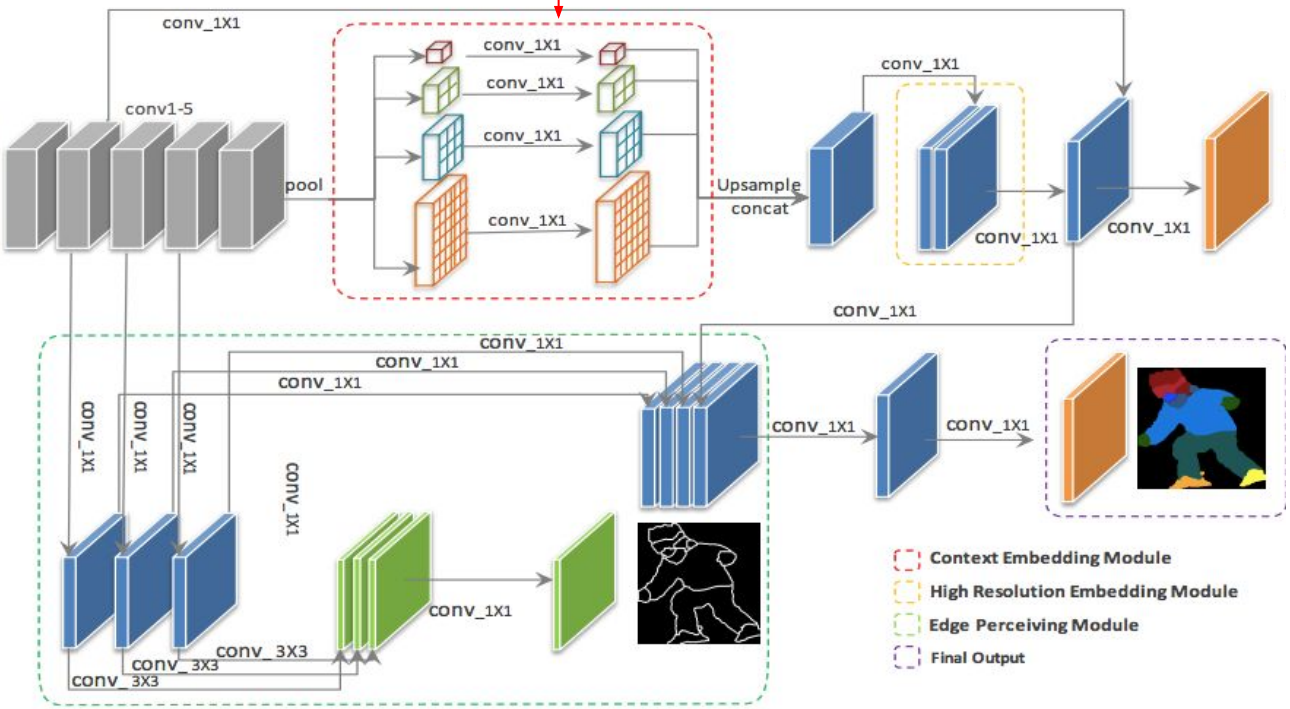
Context Embedding with Edge Perceiving (CE2P)



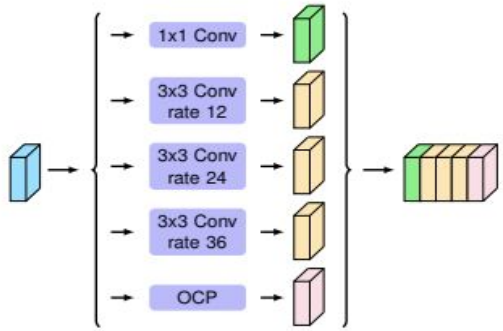
Ruan, Tao, et al. "Devil in the Details: Towards Accurate Single and Multiple Human Parsing." arXiv preprint arXiv:1809.09556 (2018).

ByteDance

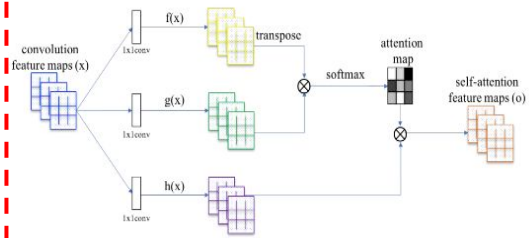
OCNet-CE2P



OCP ASPP:



OCP Module:



Yuan, Yuhui, et al. "OCNet: Object Context Network for Scene Parsing." arXiv preprint arXiv:1809.00916 (2018).

ByteDance

Outline

- Human Parsing Networks
- Datasets
- Experiments
- Summary

Datasets

Dataset	Number of instances	Class Definition
LIP	3w+ for training 1w for validation 1w for test	Background(0), Hat(1), Hair(2), Glove(3), Sunglasses(4), Upper-clothes(5), Dress(6),Coat(7),Socks(8),Pants(9), Jumpsuits(10), Scarf(11), Skirt(12), Face(13), Left-arm(14), Right-arm(15), Left-leg(16), Right-leg(17), Left-shoe(18), Right-shoe(19)
Multi-Person Human Parsing	8W+ human instance	Background(0), Hat(1), Hair(2), Glove(3), Sunglasses(4), Upper-clothes(5), Dress(6),Coat(7),Socks(8),Pants(9), tosor-skin(10), Scarf(11), Skirt(12), Face(13), Left-arm(14), Right-arm(15), Left-leg(16), Right-leg(17), Left-shoe(18), Right-shoe(19)

Outline

- Human Parsing Networks
- Datasets
- Experiments
- Summary

Baseline

Method	Input Size	Training Datasets	Miou on LIP test set
HRNetv2-w48	480*480	LIP training	**55.64**
CE2P-OCNet (ResNet101)	480*480	LIP training	54.20

External Dataset

Training Datasets	Method	Input Size	Miou on LIP test set
LIP training	CE2P-OCNet (ResNet101)	480*480	54.20
LIP training + val	CE2P-OCNet (ResNet101)	480*480	56.59 ↑2.39
LIP training + val Multi-Human	CE2P-OCNet (ResNet101)	480*480	**60.45** ↑6.25

Input Size

Input Size	Training Datasets	Method	Miou on LIP test set
480*480	LIP training + val Multi-Human	CE2P-OCNet (ResNet101)	60.45
576*576	LIP training + val Multi-Human	CE2P-OCNet (ResNet101)	**60.65** ↑0.20

Different Backbone

Method	Training Datasets	Input Size	Miou on LIP test set
CE2P-OCNet (ResNet101)	LIP training + val Multi-Human	576*576	60.65
CE2P-OCNet (SENet152)	LIP training + val Multi-Human	576*576	61.00 ↑0.35
CE2P-OCNet (SCNet101)	LIP training + val Multi-Human	576*576	61.23 ↑0.58
CE2P-OCNet (DeResNet101)	LIP training + val Multi-Human	576*576	**62.17** ↑1.52
HRNetv2-w48	LIP training + val Multi-Human	576*576	61.96 ↑1.31

Ensemble

Method	Training Datasets	Input Size	Miou on LIP test set
CE2P-OCNet (DeResNet101)	LIP training + val Multi-Human	576*576	62.17
Ensemble with Average Heatmaps	LIP training + val Multi-Human	576*576	64.00 ↑1.83
Ensemble with Average Softmax-Heat maps	LIP training + val Multi-Human	576*576	**64.13** ↑1.96

Outline

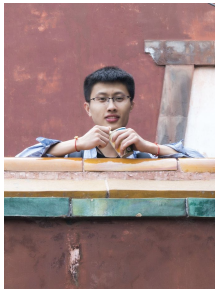
- Human Parsing Networks
- Datasets
- Experiments
- Summary

Summary

Baseline Result: 54.20

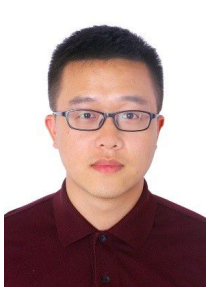
+ external dataset 54.20 -> 60.45 (+6.25)
+ enlarge the input size 60.45 -> 60.65 (+0.20)
+ strong backbone 60.65 -> 62.17 (+1.52)
+ ensemble with average heatmaps 62.17 -> 64.00 (+1.83)
+ ensemble with average softmax-heatmaps 64.00 -> 64.13 (+0.13)

Our Team



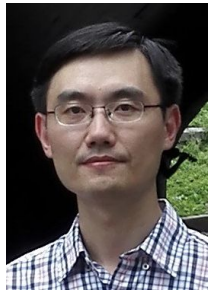
Kai Su

Dongdong Yu



Jian Wang

Kaihui Zhou



Xin Geng

Changhu Wang

Thanks & Questions

Slides are avaliable at
https://7color94.github.io/files/CVPR-19-LIP.pdf