# COCO Keypoint Challenge 2020

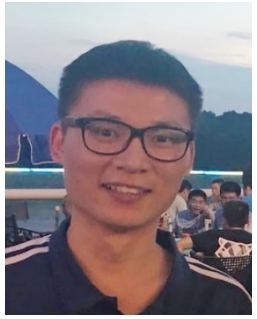
Xforward AI Technology Co., LTD.

# Team Member

Junjie Huang*

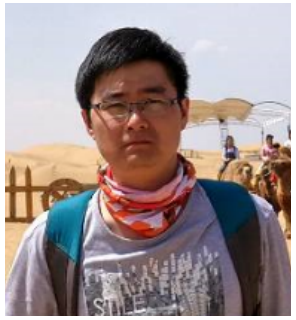
Zengguang Shan*

Yuanhao Cai*



Feng Guo

Yun Ye



Xinze Chen

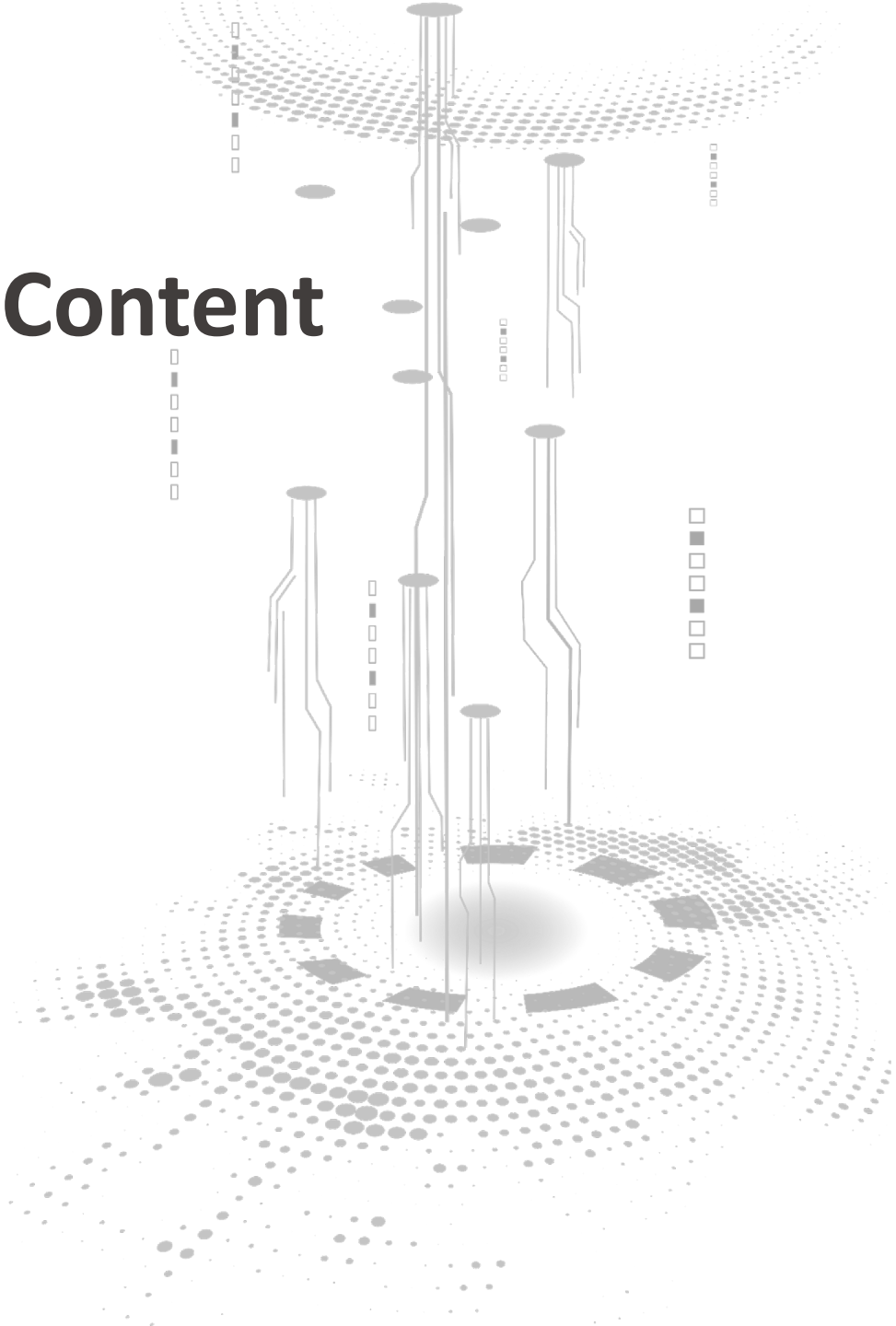
Zheng Zhu
Tsinghua University

Guan Huang

Jiwen Lu
Tsinghua University

Dalong Du

# Content

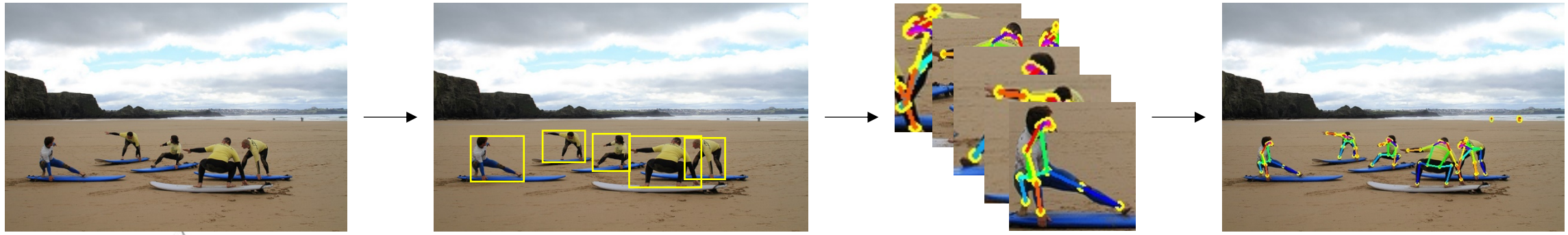



# PART 01
# Problem

# Multi-instance Keypoint Detection

XFORWARDAI 芯翌科技

Top-Down Pipeline



Human Detection  Keypoint Detection

# The main challenge/shortage

XFORWARDAI 芯豐科技

## The systemic bias in the data processing pipeline

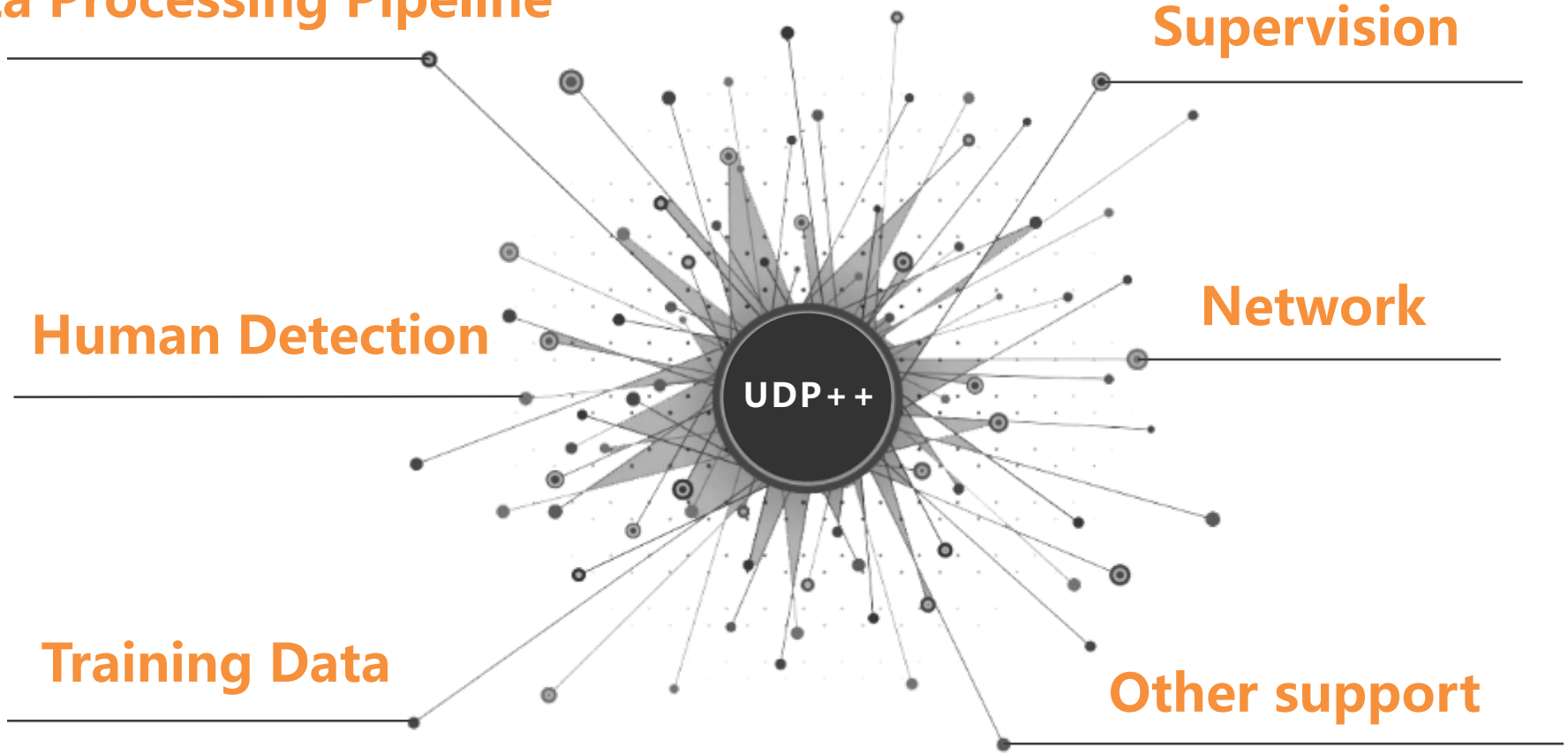
**Degrade the performance** and **obstruct the research**

## Hard pose where appearance cue is not sufficient

**Ceiling the performance of the baseline**


# PART 02
# Method

# Main Element



# Two Breakthrough

XFORWARDAI 芯翼科技

## Unbiased Data Processing Pipeline

Wipe out the systemic bias in data processing pipeline and provide a reliable baseline

## The Focus on the Constraint Cues 约束信息

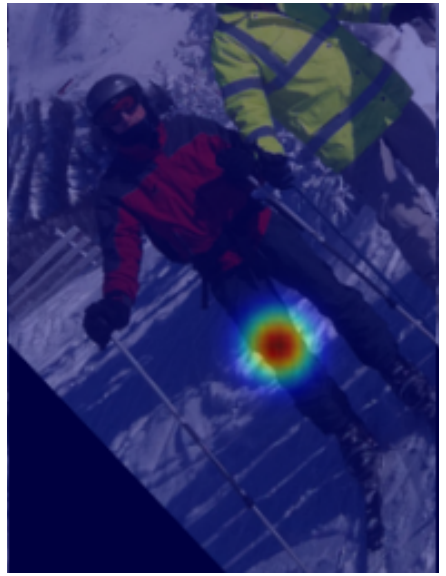
Locate keypoint under challenge situations

# Unbiased Data Processing Pipeline

Data Transformation

Encoding

Decoding

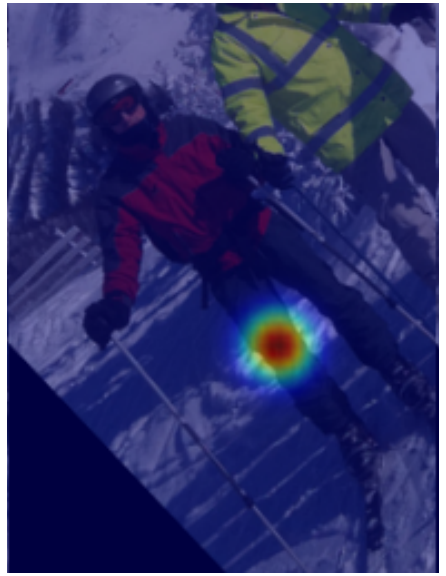


# Unbiased Data Processing Pipeline

Data
Transformation

Encoding

Decoding



Data processing pipeline: $\mathbf{k}' = T'(D(E(T(\mathbf{k}))))$

Unbiased Condition: $\mathbf{k}' = \mathbf{k}$

# Publication

Junjie Huang, Zheng Zhu, Feng Guo, and Guan Huang. The devil is in the details: Delving into unbiased data processing for human pose estimation. CVPR, June 2020.

# Project

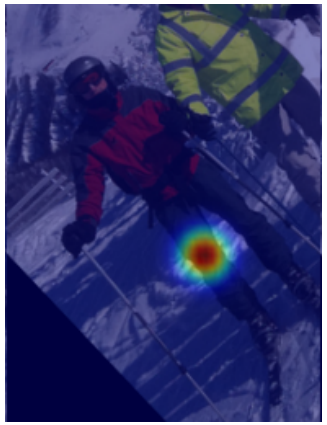
https://github.com/HuangJunJie2017/UDP-Pose

# The Focus on the Constraint Cues

Two main cues

Appearance

Supervision

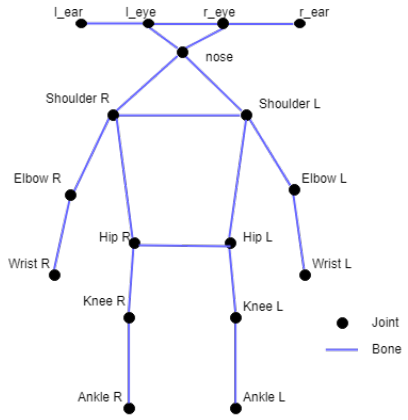


# The Focus on the Constraint Cues

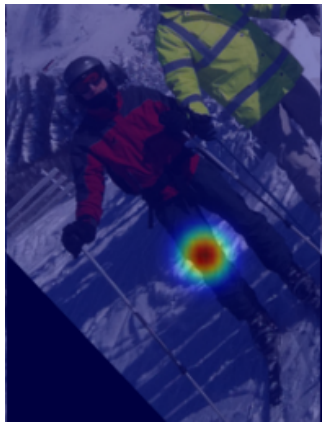
Two main cues

外观
Appearance

约束
Constraint




Supervision



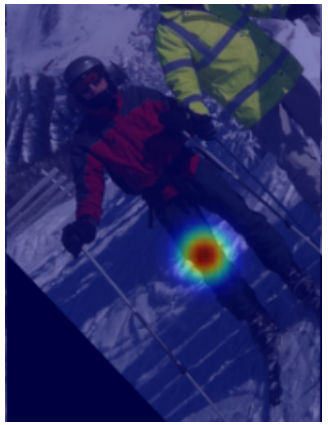
# The Focus on the Constraint Cues

Two main cues

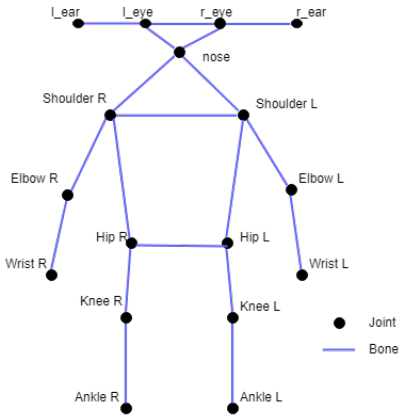
Appearance

Overfit

Supervision

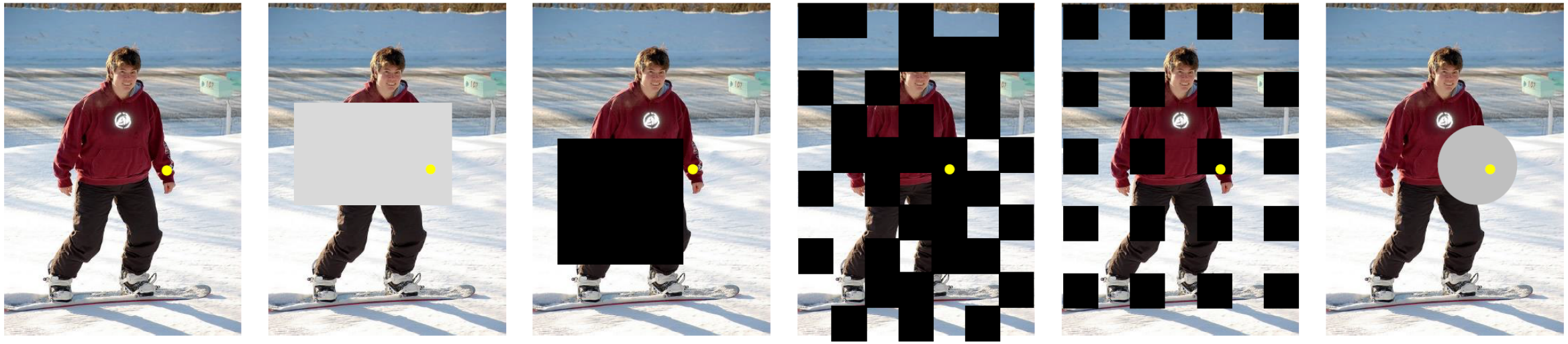


Constraint



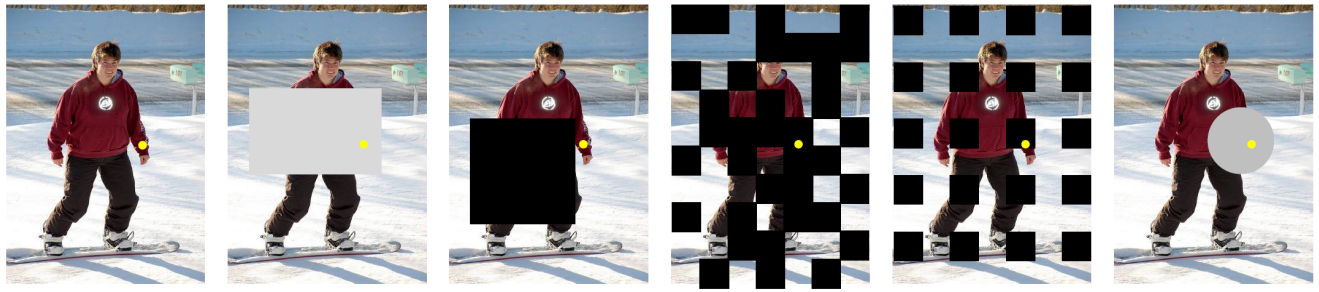

Overlook

# Information Dropping



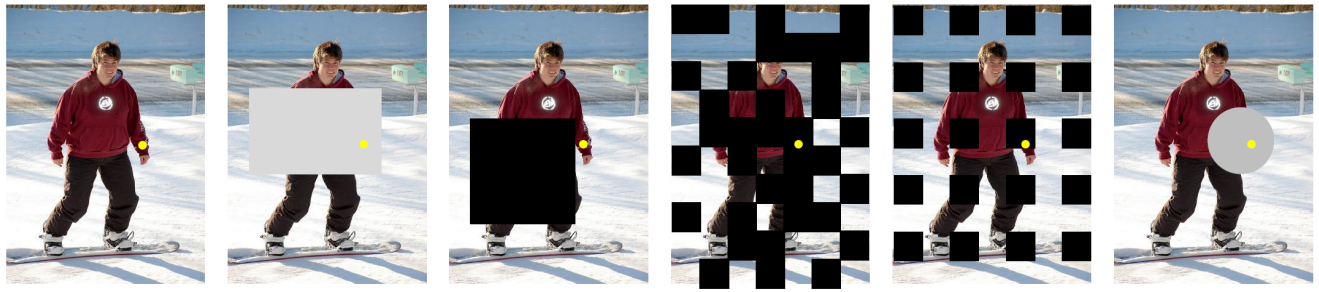
Origin Random Erase Cutout HaS Grid Mask Ours

# Keypoint-aware Occlusion Augmetation



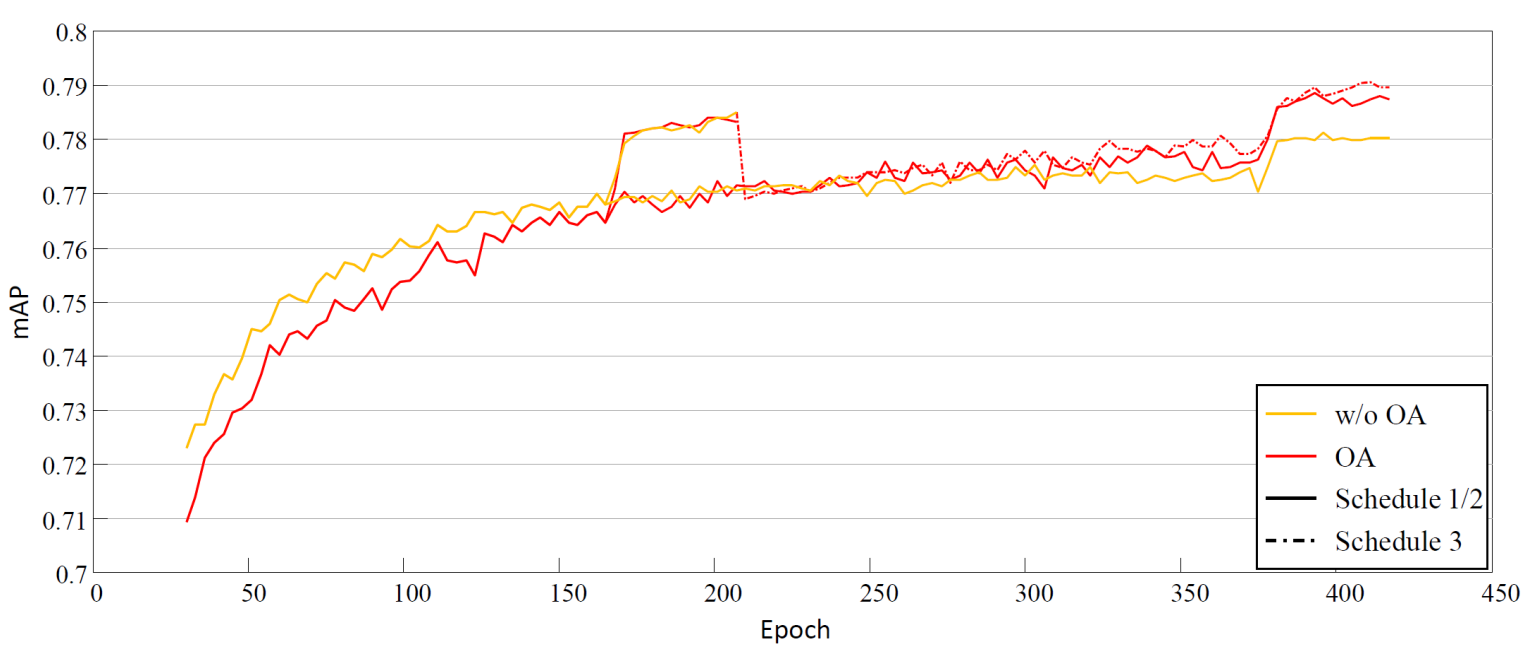
Origin Random Erase Cutout HaS Grid Mask Ours

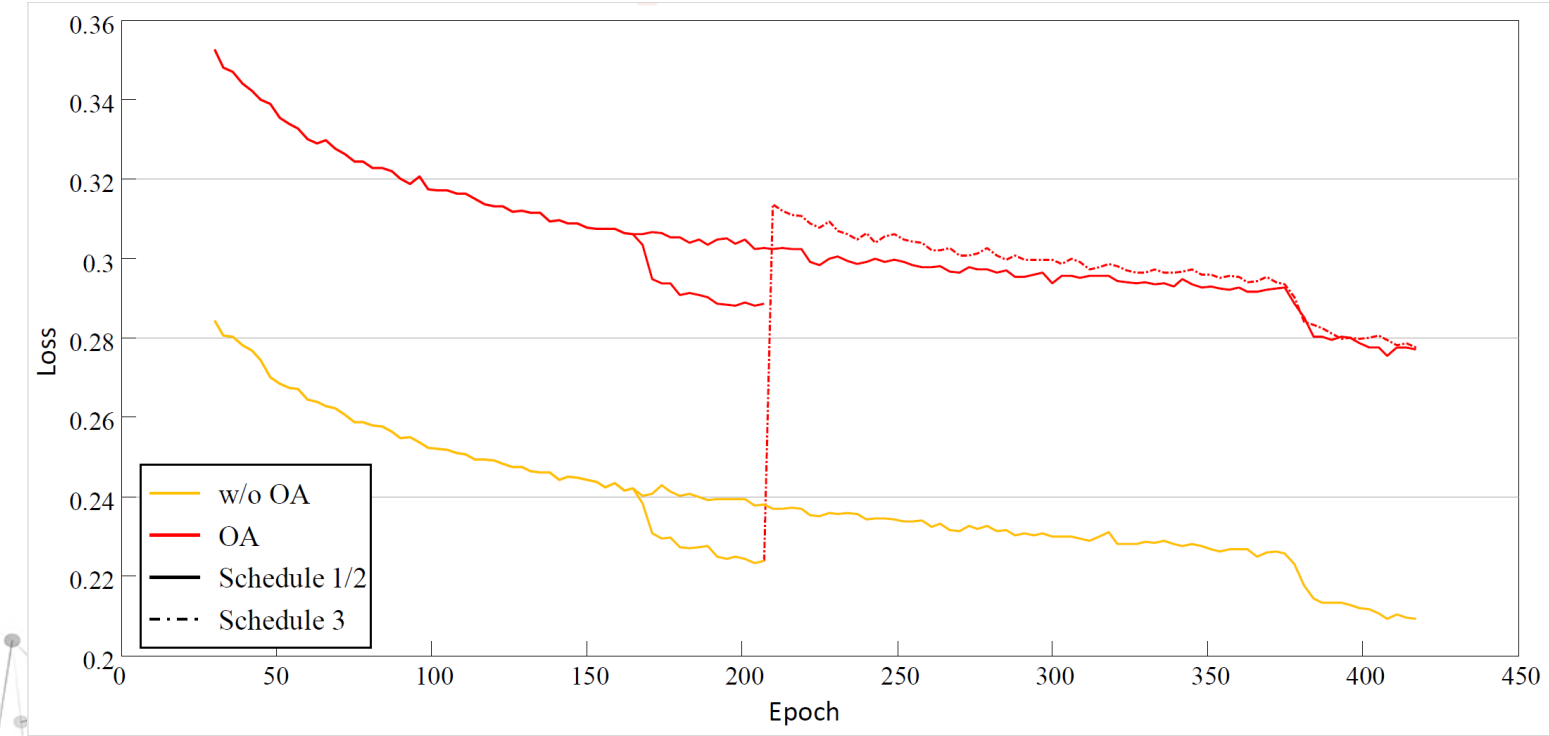
# Keypoint-aware Occlusion Augmetation



Origin Random Erase Cutout HaS Grid Mask Ours

# Training Schedule



# Experiment

## COCO val

| Method | Backbone | Input size | AP |
|---|---|---|---|
| UDPv1[16] | ResNet-50 | 256 × 192 | 74.6 |
| **+OA** | ResNet-50 | 256 × 192 | **75.3(+0.7)** |
| UDPv1[16] | 2xRSN-50 | 256 × 192 | 77.7 |
| **+OA** | 2xRSN-50 | 256 × 192 | **78.2(+0.5)** |
| UDPv1[16] | HRNet-W32 | 256 × 192 | 77.2 |
| **+OA** | HRNet-W32 | 256 × 192 | **77.8(+0.6)** |
| UDPv1[16] | HRNet-W48 | 256 × 192 | 77.8 |
| **+OA** | HRNet-W48 | 256 × 192 | **78.4(+0.6)** |
| UDPv1[16] | HRNet-W32 | 384 × 288 | 77.9 |
| **+OA** | HRNet-W32 | 384 × 288 | **78.7(+0.8)** |
| UDPv1[16] | HRNet-W48plus | 384 × 288 | 78.5 |
| **+OA** | HRNet-W48plus | 384 × 288 | **79.1(+0.6)** |

## COCO test-dev

| | | | |
|---|---|---|---|
| Simple+UDPv1[16] | ResNet-50 | 256 × 192 | 73.1 |
| **+OA** | ResNet-50 | 256 × 192 | **73.7(+0.6)** |
| RSN+UDPv1[16] | 2xRSN-50 | 256 × 192 | 76.0 |
| **+OA** | 2xRSN-50 | 256 × 192 | **76.6(+0.6)** |
| UDPv1[16] | HRNet-W32 | 256 × 192 | 75.6 |
| **+OA** | HRNet-W32 | 256 × 192 | **76.2(+0.6)** |
| UDPv1[16] | HRNet-W32 | 384 × 288 | 76.5 |
| **+OA** | HRNet-W32 | 384 × 288 | **77.0(+0.5)** |
| UDPv1[16] | HRNet-W48 | 256 × 192 | 76.1 |
| **+OA** | HRNet-W48 | 256 × 192 | **76.7(+0.6)** |
| UDPv1[16] | HRNet-W48plus | 384 × 288 | 76.8 |
| **+OA** | HRNet-W48plus | 384 × 288 | **77.5(+0.7)** |
| HRNet*[35] | HRNet-W48 | 384 × 288 | 77.0 |
| MSPN*[24] | MSPN | 384 × 288 | 77.1 |
| DARK*[45] | HRNet-W48 | 384 × 288 | 77.4 |
| UDPv1*[16] | HRNet-W48plus | 384 × 288 | 78.2 |
| +posefix [27] | HRNet-W48plus+ResNet-152 | 384 × 288 | 78.3 |
| **+OA** | HRNet-W48plus | 384 × 288 | **78.7(+0.5)** |
| **+OA+posefix** | HRNet-W48plus+ResNet-152 | 384 × 288 | **78.8** |

## Crowdpose test

| Method | Backbone | Input size | AP | $AP^{50}$ | $AP^{75}$ | AR | $AR^{50}$ | $AR^{75}$ |
|---|---|---|---|---|---|---|---|---|
| CrowdPose[22] | ResNet-101 | - | 61.1 | 82.0 | 65.1 | 67.8 | 87.2 | 73.1 |
| HRNet+UDPv1[16] | HRNet-W32 | 256 × 192 | 76.3 | 93.3 | 82.9 | 79.6 | 94.7 | 85.0 |
| **+OA** | HRNet-W32 | 256 × 192 | **77.9(+1.6)** | **94.4** | **84.1** | **80.9** | **85.4** | **86.4** |

# Publication

Junjie Huang, Zheng Zhu, Guan Huang, and Dalong Du. How to Train Your Robust Human Pose Estimator: Pay Attention to the Constraint Cue. arXiv preprint arXiv:2008.07139, 2020.


# PART 03
# Road Map

# Road Map

| Methods | AP | Improvement |
|---|---|---|
| Baseline | 73.5 | - |
| +1.Unbiased Data Processing | 75.5 | +2.0 |
| +2.Occlusion Augmentation | 76.2 | +0.7 |
| +3.other tricks | 76.8 | +0.6 |
| +4.larger input size | 77.6 | +0.8 |
| +5.larger backbone | 78.1 | +0.5 |
| +6.extra data | 79.4 | +1.3 |
| +7.refine with SyncBN | 79.6 | +0.2 |
| +8.stronger human detection | 80.2 | +0.6 |
| +9.model ensemble | 80.8 | +0.6 |

**Other augmentation/ hyper-parameters searching**

**256x192 -> 384x288**

**HRNetW32 -> HRNetW48plus**

**COCO -> COCO+AIChallenger**

**HumanAP 64.1 -> HumanAP 68.6**

**17models**

**Table 3.** Keypoint road maps of COCO 2020 challenge.


# PART 04
# Summary

XFORWARDAI 芯智科技

## A Reliable Baseline

**Boost the current performance** and **benefit the future research**

## More Breakthrough in Methodology

**Improve the performance with limit computing resource**

## A Human Detector with High Performance

**No shortage, to avoid Bucket Effect.**

XFORWARDAI 芯翌科技


# THANK YOU

Join us :

junjie.huang@ieee.org