P^2 Net: Augmented Parallel-Pyramid Net for Attention Guided Pose Estimation

Ran He

ran.he@ia.ac.cn

Luanxuan Hou

luanxuan.hou@cripac.ia.ac.cn

Outline

- Research Background
- Launched Applications
- Our Contributions

Outline

- **Research Background**
- Launched Applications
- Our Contributions

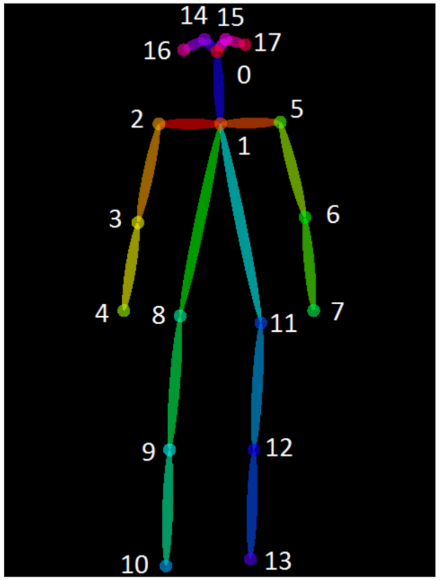
Research Background

Definition of Pose Estimation

- ✓ Articulated body **Pose Estimation** in computer vision is the study of algorithms and systems that **recover the pose of an articulated body**

Application

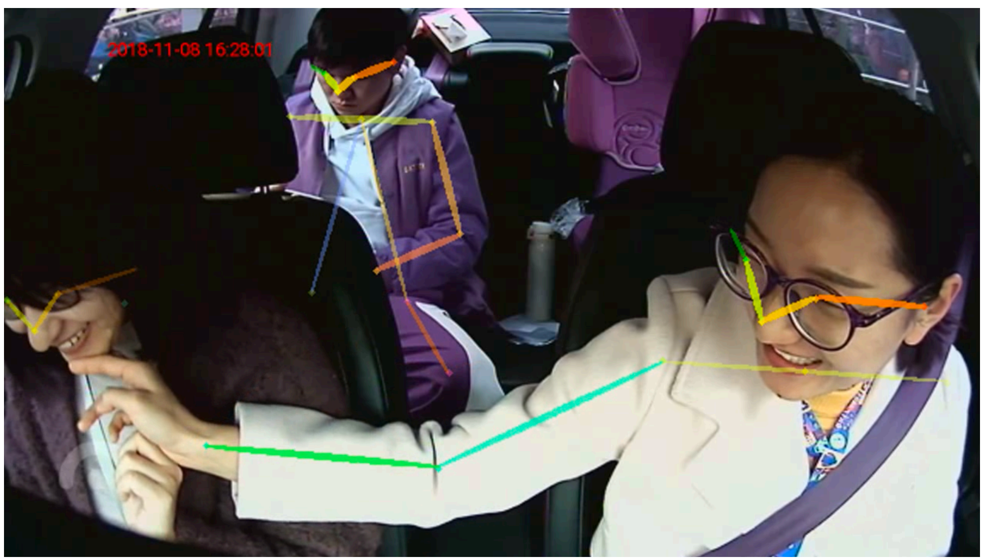
- ✓ Assisted living
- ✓ Character animation
- ✓ **Intelligent driver assisting system**
  - ➢ Car accidents account for about **2% of deaths globally each year**
  - ➢ Pedestrian detection
  - ➢ Distracted driving detection



Photograph taken from Pexels

Challenge on Application

- Change of **perspective** / **clothes**
- Multi-Person **blocking** each other
- Change in **morning** / **evening light**
- Complexity of **Model**



Data source from: Didi AI Labs

Outline

- Research Background
- **Launched Applications**
- Our Contributions

Overview of Contributions

- ✓ Data augmentation sequence for **Spontaneous Learning**
- ✓ Refining based on **Attention Module**
- ✓ **High-resolution** information compensation based on **parallel structures**

Data source from: MSCOCO test-dev

Overview of Contributions

Demo

Launched Applications

Conflict detection (Night)

Data source from: Didi AI Labs

Launched Applications

DiDi

Distracted driving detection

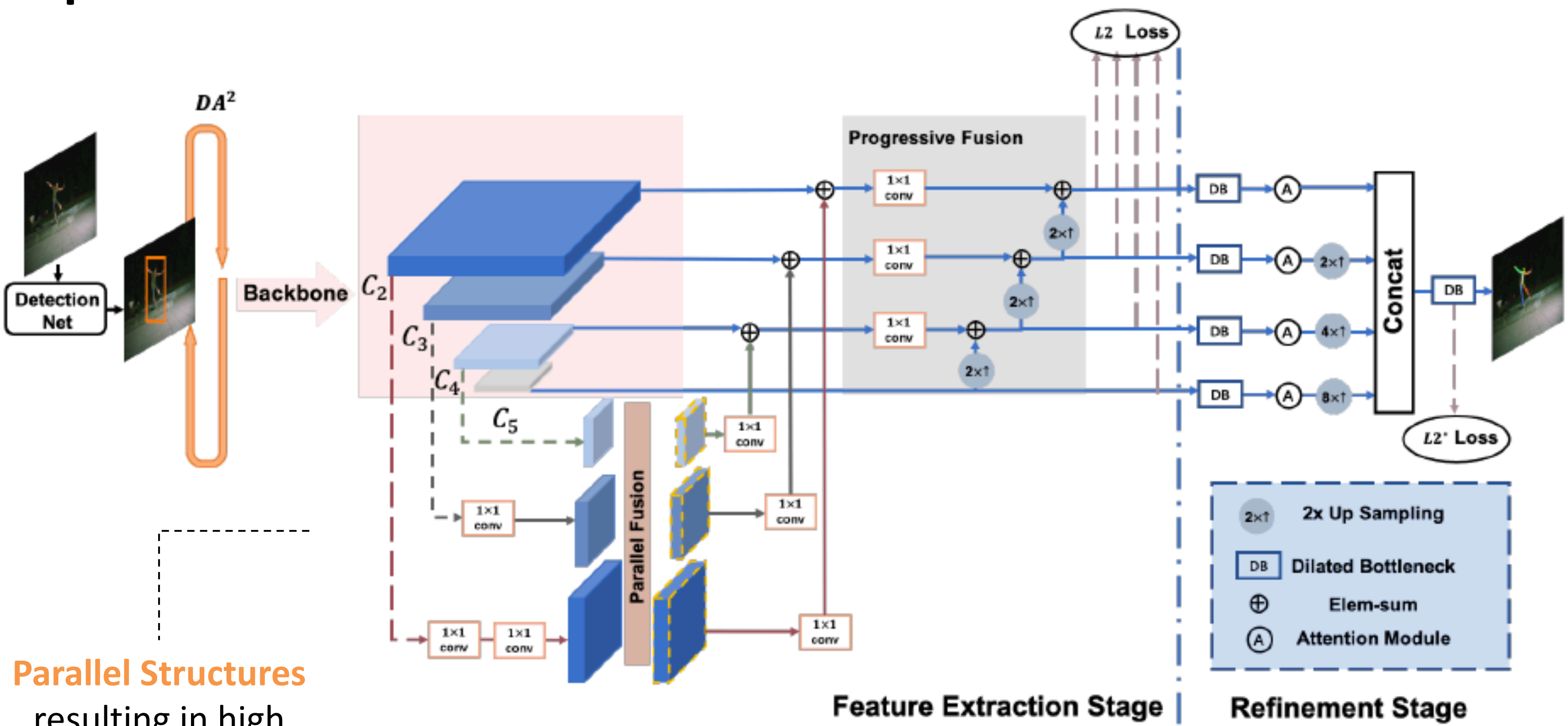
Data source from: Didi AI Labs

Outline

- Research Background
- Launched Applications
- **Our Contributions**

Our Contributions

Spontaneous FPN

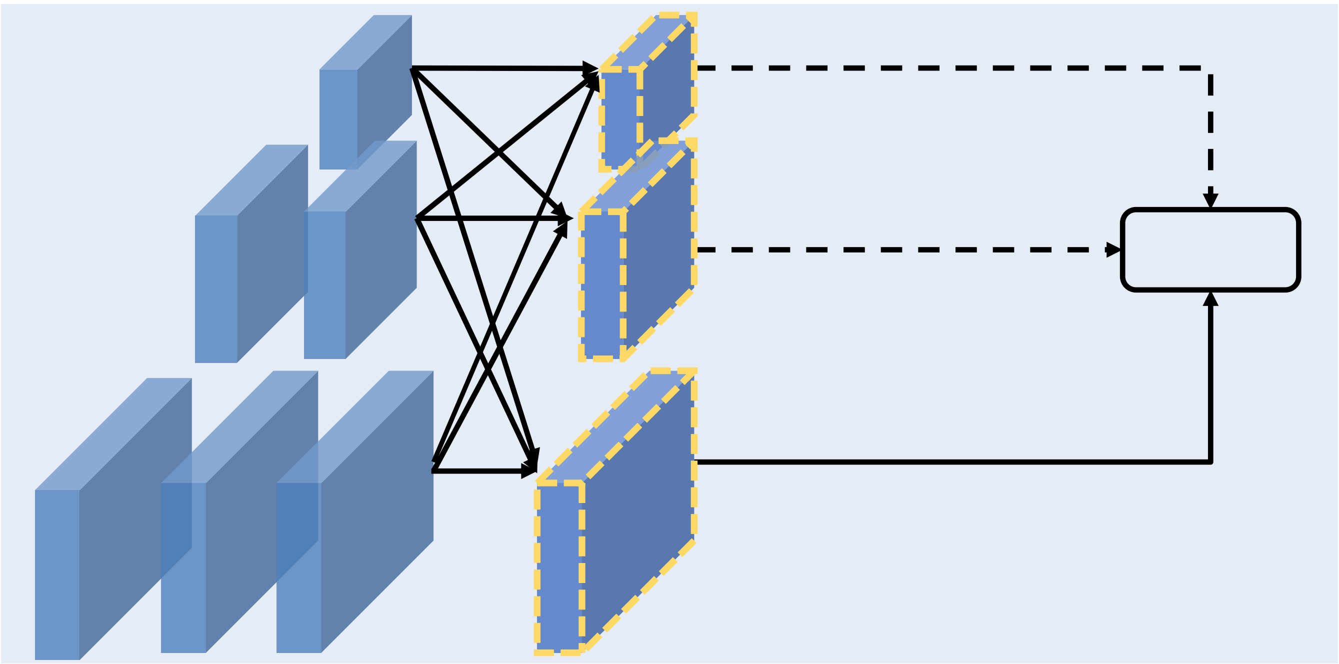


Parallel Structures resulting in high resolution representations

Our Contributions

Spontaneous FPN

- Parallel Structure



- **Pyramid structure** effectively retaining the *global and local* information.

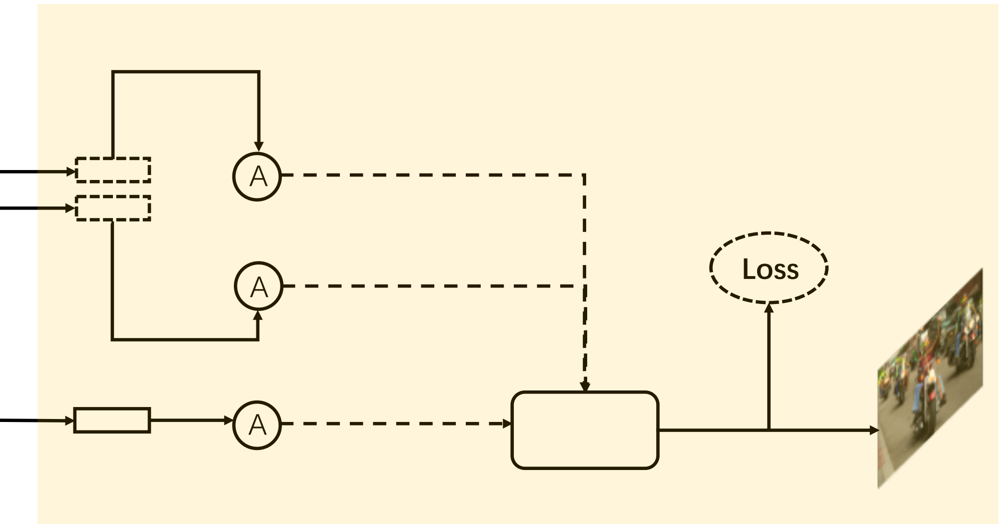
Making Up **Information loss**

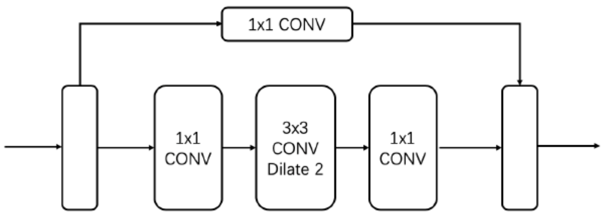
- Perform *multiscale* fusions, using **parallel structures**, resulting in high resolution representations.

Our Contributions

Spontaneous FPN

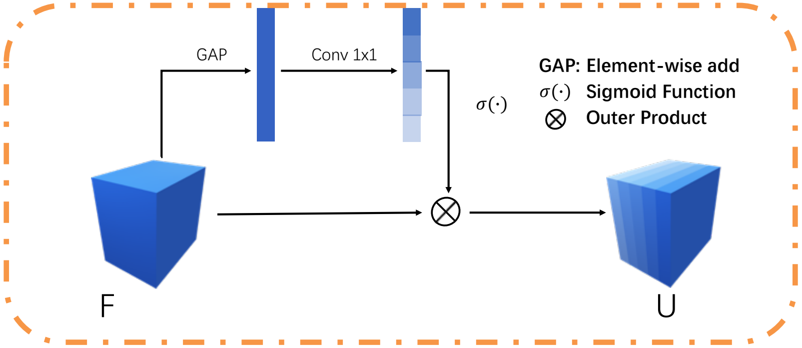
- RefineNet





- RefineNet adopt *L2*loss*

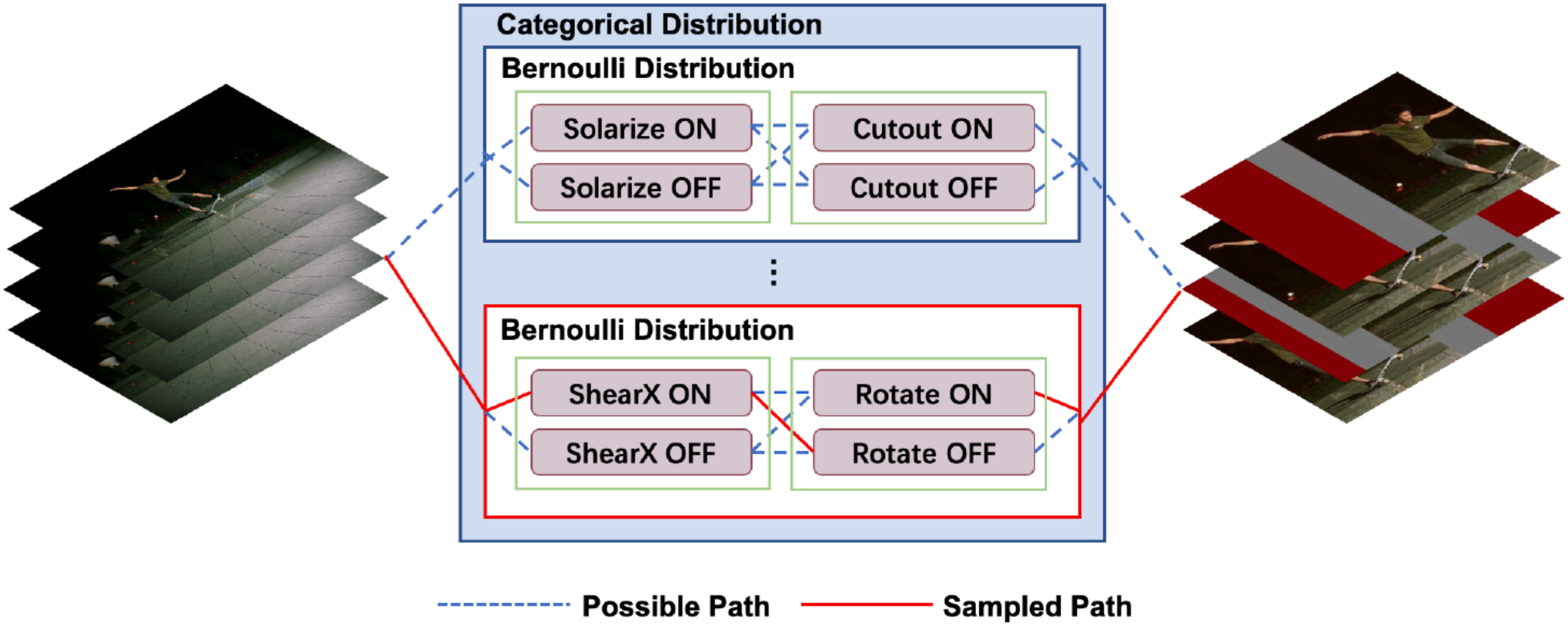
only keep the **top α** key points *loss* out of all N key points



- Using **Dilated Conv** to achieve a good trade-off between *Receptive Field and efficiency*

Our Contributions

Spontaneous Data Augmentation



Our Contributions

Spontaneous FPN

- Spontaneous Data Augmentation

➢ Operations

Operation Name	Description
TranslateX(Y)	Translate the image and the bounding boxes in the horizontal (vertical) direction by $magnitude$ number of pixels.
Rotate	Rotate the image and the bounding boxes $magnitude$ degrees.
Equalize	Equalize the image histogram.
Solarize	Invert all pixels above a threshold value of $magnitude$.
SolarizeAdd	For each pixel in the image that is less than 128, add an additional amount to it decided by the $magnitude$.
Brightness	Adjust the brightness of the image. A $magnitude$=0 gives a black image, whereas $magnitude$=1 gives the original image.
Sharpness	Adjust the sharpness of the image. A $magnitude$=0 gives a blurred image, whereas $magnitude$=1 gives the original image.
Cutout	Set a random square patch of side-length $magnitude$ pixels to gray.
Scale	Scale with this $magnitude$.

Table 1: Table of all the possible transformations that can be applied to an image. These are the transformations that are available to the controller during the search process. The range of magnitudes that the controller can predict for each of the transforms is listed in the third column. Some transformations do not have a magnitude associated with them (e.g. Equalize).

Our Contributions

Spontaneous FPN

- Spontaneous Data Augmentation

(Rotate,P, M)

(TranslateX,P, M)

(Brightness,P, M)

Our Contributions

Spontaneous FPN

- Experiments and Analysis (MSCOCO)

Method	Backbone	Input size	#Params	GFLOPs	AP	AP^{50}	AP^{75}	AP^{M}	AP^{L}	AR
Mask-RCNN[10]	ResNet-50-FPN	-	-	-	63.1	87.3	68.7	57.8	71.4	-
G-RMI[27]	ResNet-101	353×257	42.6M	57.0	64.9	85.5	71.3	62.3	70.0	69.7
G-RMI[27] + extra data	ResNet-101	353×257	42.6M	57.0	68.5	87.1	75.5	65.8	73.3	73.3
CPN[4]	ResNet-Inception	384×288	-	-	72.1	91.4	80.0	68.7	77.2	78.5
RMPE[9]	PyraNet	320×256	28.1M	26.7	72.3	89.2	79.1	68.0	78.6	-
CFN[14]	-	-	-	-	72.6	86.1	69.7	78.3	64.1	-
CPN[4] (ensemble)	ResNet-Inception	384×288	-	-	73.0	91.7	80.9	69.5	78.1	79.0
SimpleBaseline[37]	ResNet-152	384×288	68.6M	35.6	73.7	91.9	81.1	70.3	80.0	79.0
HRNet-W32[33]	HRNet-W32	384×288	28.5M	16.0	74.9	92.5	82.8	71.3	80.9	80.1
Ours	**ResNet101**	**384×288**	**42.5M**	**26.3**	**77.3**	**93.1**	**84.7**	**73.6**	**83.4**	**82.3**

Our Contributions

Spontaneous FPN

- Experiments and Analysis (MPII)

Method	Hea	Sho	Elb	Wri	Hip	Kne	Ank	Total
Stack Hourglass [25].	98.2	96.3	91.2	87.1	90.1	87.4	83.6	90.9
Sun et al [32].	98.1	96.2	91.2	87.2	89.8	87.4	84.1	91.0
Chu et al [6].	98.5	96.3	91.9	88.1	90.6	88.0	85.0	91.5
Chou et al [5].	98.2	96.8	92.2	88.0	91.3	89.1	84.9	91.8
Yang et al [38].	98.5	96.7	92.5	88.7	91.1	88.6	86.0	92.0
Ke et al[16].	98.5	96.8	92.7	88.4	90.6	89.3	86.3	92.1
Tang et al [35].	98.4	96.9	92.6	88.7	91.8	89.4	86.2	92.3
SimpleBaseline[37]	98.8	96.6	91.9	87.6	91.1	88.1	84.1	91.5
HRNet-W32[33]	98.6	96.9	92.8	89.0	91.5	89.0	85.7	92.3
Ours	**98.8**	**97.0**	**93.9**	**89.9**	**92.1**	**92.0**	**86.4**	**92.9**

Our Contributions

Spontaneous FPN

- Ablation study
  - Detector Net

Detector	Backbone	mAP	AP^{50}	AP^{S}	AP^{M}	AP^{L}	AP^{*}
Faster R-CNN	ResNet50	37.4	59.0	18.3	41.7	52.9	**77.2**
DetNet59	ResNet50	40.2	61.7	**23.9**	43.2	52.0	**77.3**
TridentNet	ResNet101	**40.6**	**61.8**	23.0	**45.5**	**55.9**	**77.3**

Table 5. Comparing with different detector networks. AP^{*} means the performance of P^2 Net based on the specific detector network.

  - Spontaneous Data Augmentation

Method	Backbone	AP	AP^{50}	AP^{75}	AP^{M}	AP^{L}	AR
Ours	**ResNet101**	**77.3**	**93.1**	**84.7**	**73.6**	**83.4**	**82.3**
$Ours^{*}$	ResNet101	75.8	91.7	83.1	72.2	81.5	82.2

Table 6. Comparing with the result without searching. $Ours^{*}$ means the method without searching, adopting the traditional data augmentations.

  - Attention Module

Method	Backbone	AP	AP^{50}	AP^{75}	AP^{M}	AP^{L}	AR
Ours	**ResNet101**	**77.3**	**93.1**	**84.7**	**73.6**	**83.4**	**82.3**
$Ours^{*}$	ResNet101	76.2	92.5	83.1	72.2	82.2	81.4

Table 7. Compared with the result without Attention Module. $Ours^{*}$ means the method without Attention Module, adopting the auto data augmentations.

CRIPAC
智能感知与计算研究中心
Center for Research on Intelligent Perception and Computing

CASIA
中国科学院自动化研究所
Institute of Automation
Chinese Academy of Sciences

DiDi

THANK YOU