Vision Meets Drones 2021: A Challenge

Luc Van Gool Junwei Han Steven Hoi Zhijian He Qinghua Hu Ming Liu
Lujia Wang Wenguan Wang Yixuan Yuan Dingwen Zhang Jinglin Zhang Pengfei Zhu

Outline

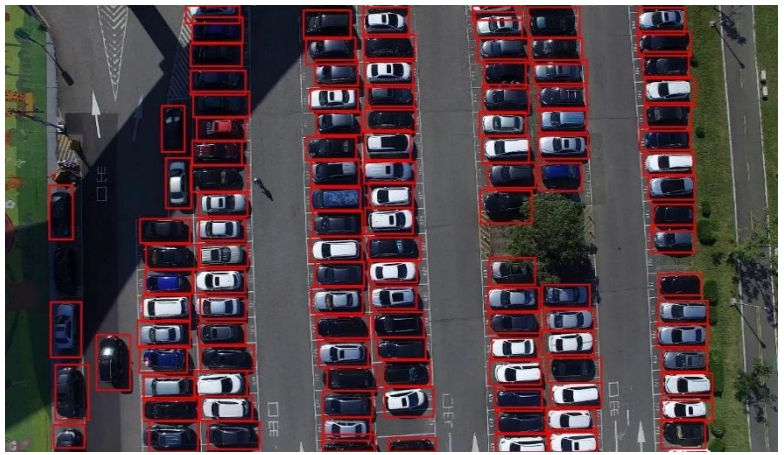
- Application of Drones
- A Review of VisDrone 2021 Challenge
- Keynote
- Winner Announcement
- Oral Presentation

Outline

- **Application of Drones**
- A Review of VisDrone 2021 Challenge
- Keynote
- Winner Announcement
- Oral Presentation

Applications of Drones

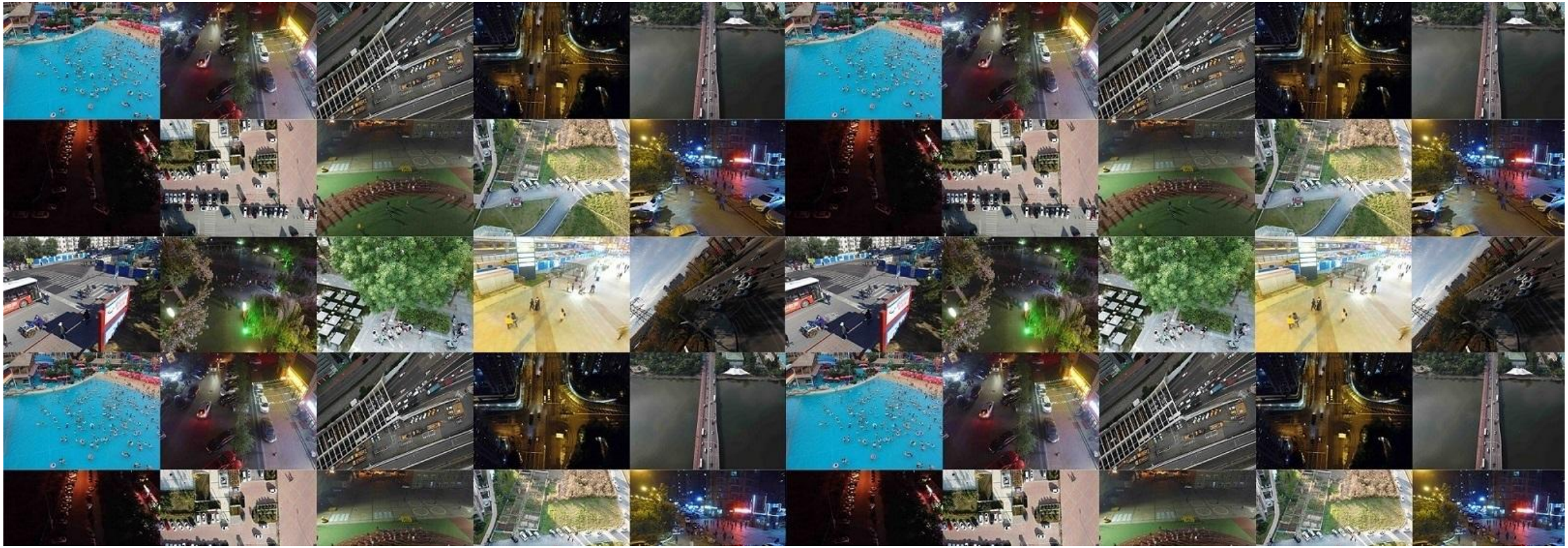
Applications of Drones



Outline

- Application of Drones
- **A Review of VisDrone 2021 Challenge**
- Keynote
- Winner Announcement
- Oral Presentation

VisDrone 2021



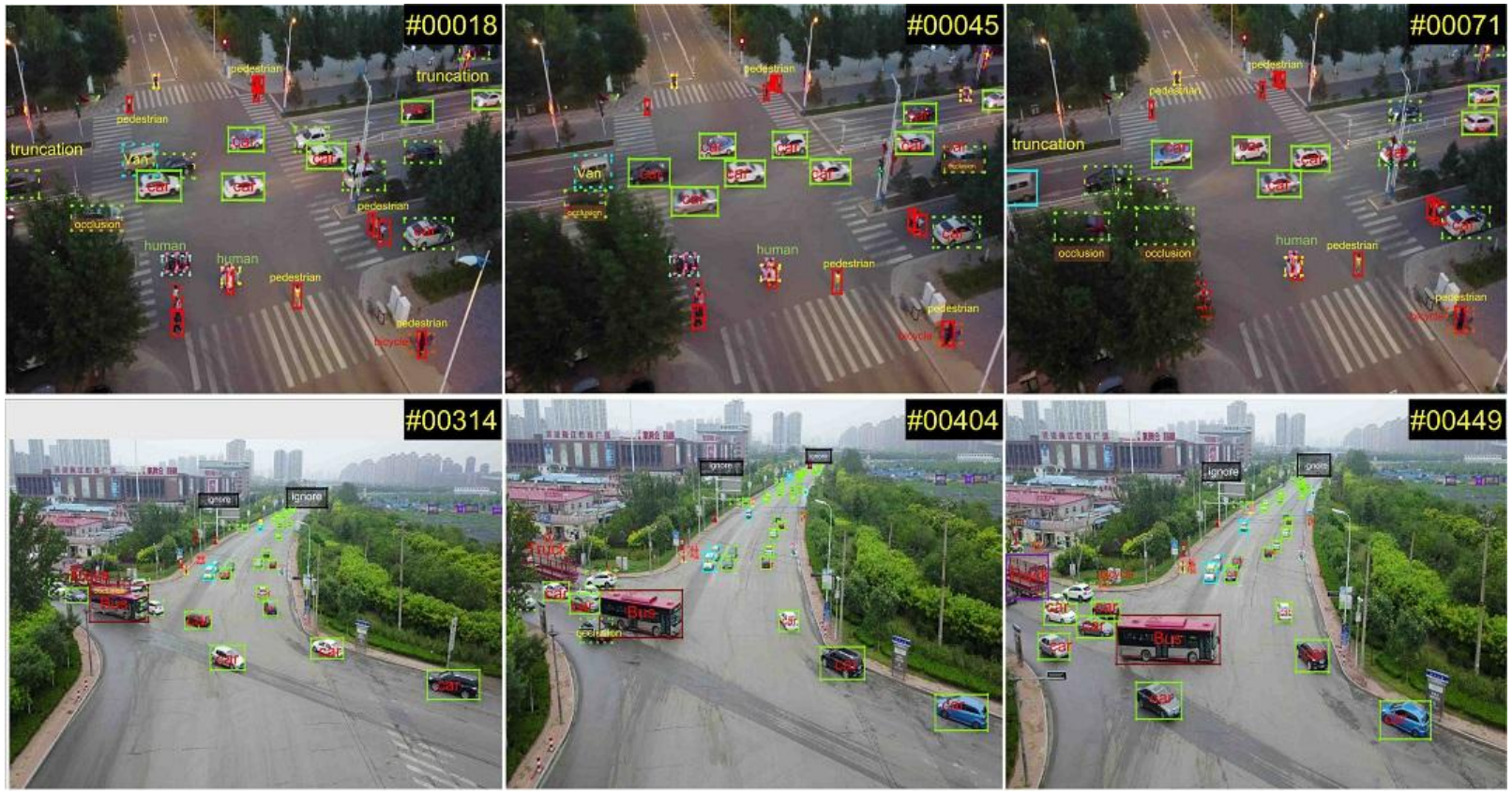
400 video clips formed by *265,228* frames and *10,209* static images

Three tasks: (1) object detection in images, (2) multi-object tracking, (3) crowd counting

Task 1: object detection in images

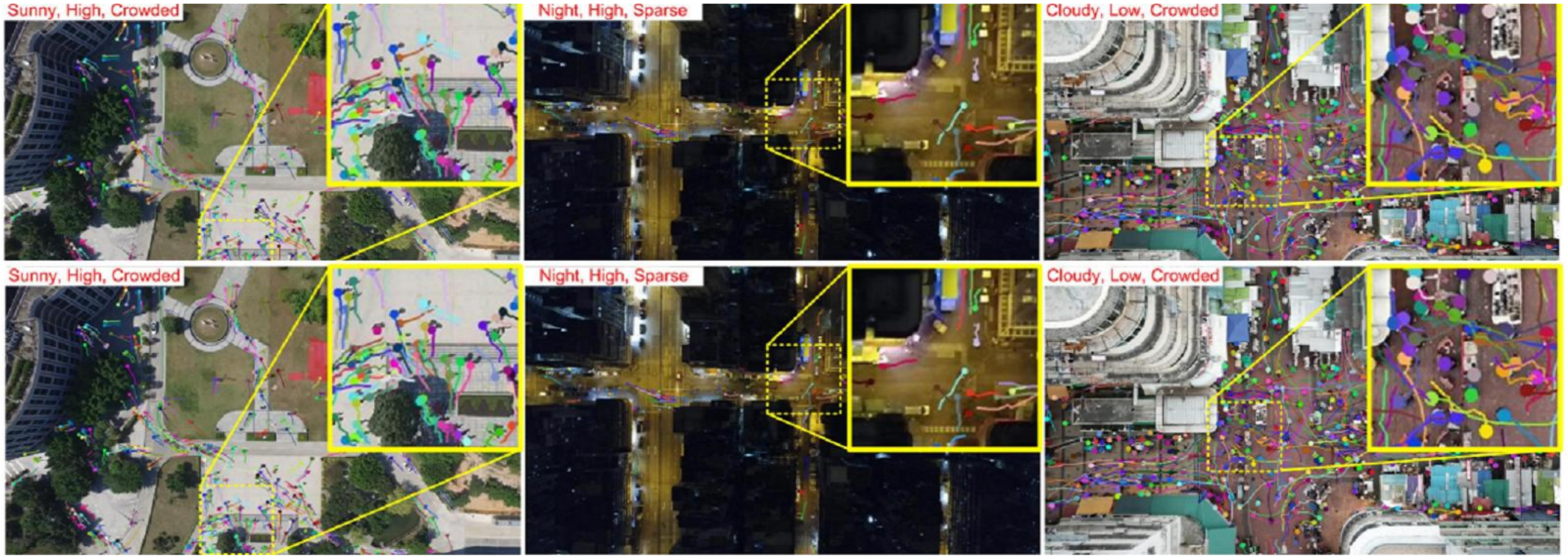
The task aims to detect objects of ten predefined categories (e.g., cars and pedestrians) from individual images taken from drones.

Task 2: multi-object tracking



The task aims to recover the trajectories of five object categories in each video frame.

Task 3: crowd counting



The task aims to estimate the number of people heads from sequential images taken from drones.

Our workshops in four years (2018~2021)

- Datasets: 263 videos -> 288 videos -> 400 videos -> 400 videos
- Participation: 66 Teams -> 95 Teams -> 169 Teams -> 208 Teams
- Accepted Papers: 0 Paper -> 21 Papers -> 8 Papers -> 9 Papers

Outline

- Application of Drones
- A Review of VisDrone 2021 Challenge
- **Keynote**
- Winner Announcement
- Oral Presentation

Invited Speaker

Martin Danelljan

ETH Zurich

Invited Speaker

Martin Danelljan is a researcher in the Computer Vision Lab at ETH Zurich, Switzerland. His main research interests are meta and online learning, deep probabilistic models, and generative methods. His research includes applications to visual tracking, video object segmentation, dense correspondence estimation, and super-resolution. He is among top 12 reviewers for ECCV 2020.

Invited Keynote

Discriminative Appearance-based Tracking and Segmentation

Martin Danelljan

Invited Speaker

Jingdong Wang

Baidu Inc.

Invited Speaker

Jingdong Wang is a Chief Architect for computer vision with the the Artifical Intelligence Group at Baidu Inc.. His areas of interest include neural architecture design, human pose estimation, semantic segmentation, image classification, object detection, large-scale indexing, and salient object detection. He has been serving/served as an Associate Editor of IEEE TPAMI, IJCV, IEEE TMM, and IEEE TCSVT, and an area chair of leading conferences in vision, multimedia, and AI, such as CVPR, ICCV, ECCV, ACM MM, IJCAI, and AAAI. He is an ACM Distinguished Member and a Fellow of IAPR.

Invited Keynote

Semantic Segmentation, Object Detection, and High-Resolution Backbone with Transformers

Jingdong Wang

Outline

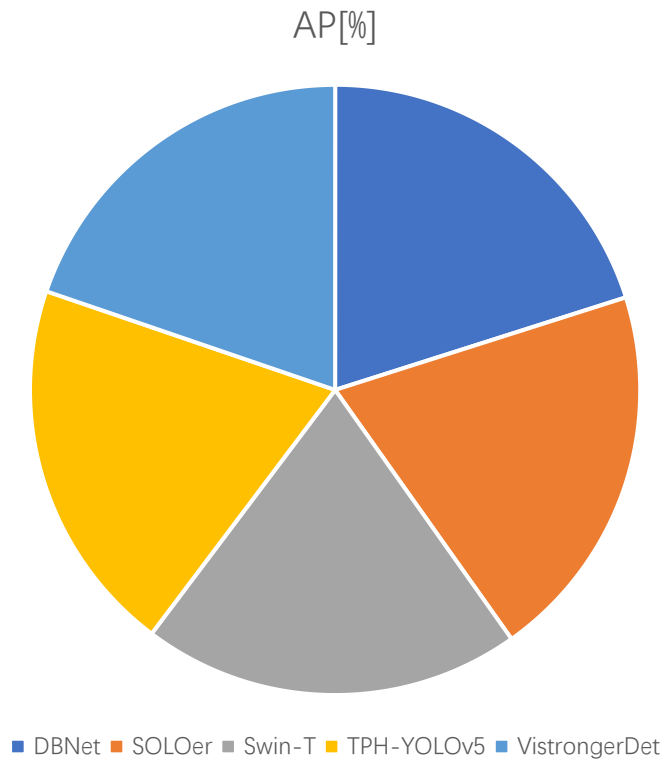
- Application of Drones
- A Review of VisDrone 2021 Challenge
- Keynote
- **Winner Announcement**
- Oral Presentation

Task 1: object detection in images

113 Teams

1 Winner

2 Honorable Mention



Method	AP[%]	AP50[%]	AP75[%]	AR1[%]	AR10[%]	AR100[%]	AR500[%]
DBNet(A.1)	39.43	65.34	41.07	0.29	2.03	12.13	55.36
SOLOer (A.2)	39.42	63.91	40.87	1.75	10.94	44.69	55.91
Swin-T(A.3)	39.40	63.91	40.87	1.76	10.96	44.65	56, 83

Task 1: object detection in images

Winner

Detection in Images Challenge

Zhe Wang, Jianye He, Zhenyu Xu, Zhimin Zhang, Zhiguang Zhang, Zhipeng Luo

DeepBlue Technology (Shanghai) Co., Ltd, China

Honorable Mention

Scaled-yolov4 with Transformer and BiFPN (SOLOer)

Xiaoqiang Lu, Guojin Cao,Zixiao Zhang,Yuting Yang

Xidian University, China

Swin-transformer Object Detection with Coarse Segmentation (Swin-T)

Hongkai Wang

Xi'an University of Technology, China

Task 1: winner talk

Object Detection in Images Challenge

Zhe Wang, Jianye He, Zhenyu Xu, Zhimin Zhang, Zhiguang Zhang, Zhipeng Luo
DeepBlue Technology (Shanghai) Co., Ltd, China

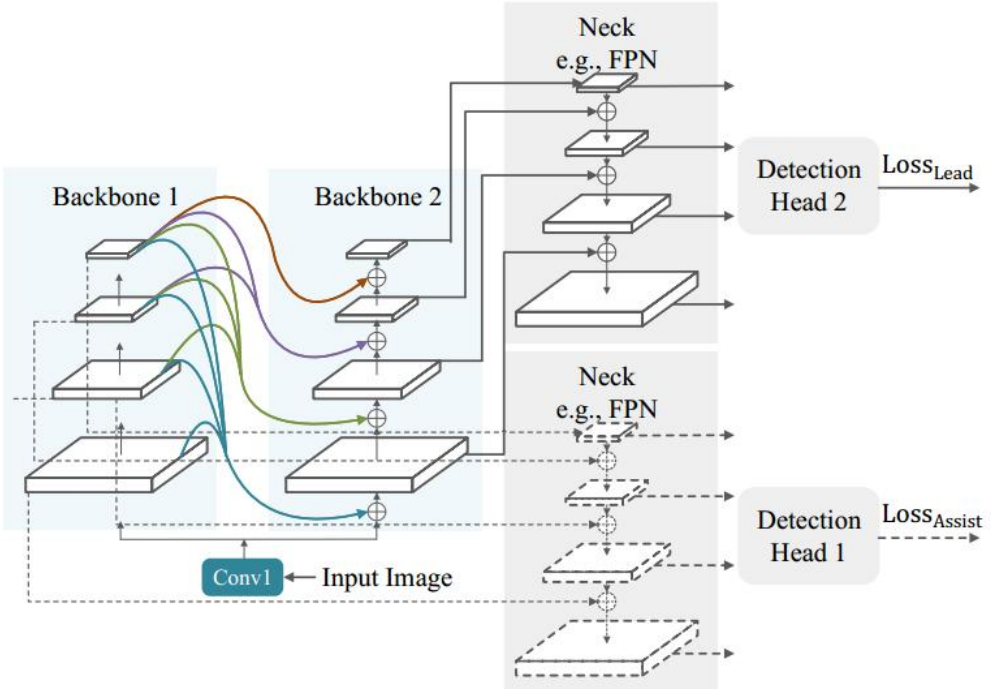


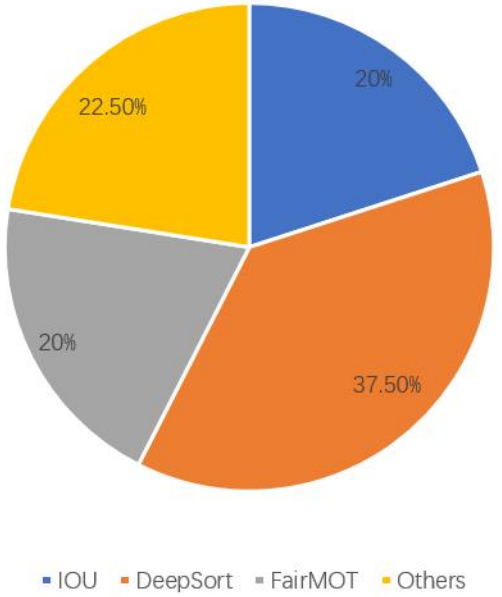
Figure: Architecture

Task 2: multi-object tracking

18 Teams

1 Winner

1 Honorable Mention



Algorithm	AP	$AP_{0.25}$	$AP_{0.5}$	$AP_{0.75}$	AP_{car}	AP_{bus}	AP_{trk}	AP_{ped}	AP_{van}
SOMOT	58.61	70.75	61.26	43.84	69.18	63.46	48.45	55.64	56.34
GIAOTracker-Fusion	54.18	63.41	55.35	43.78	69.33	51.05	43.20	55.06	52.26

Task 2: multi-object tracking

Winner

Simple Online Multi-Object-Tracking

Zhipeng Luo, Yuehan Yao, Zhenyu Xu

DeepBlue Technology(Shanghai) Co., Ltd, China

Honorable Mention

GIAOTracker: A Comprehensive Framework or MCMOT with Global Information and Optimizing Strategies in VisDrone 2021

Yunhao Du

Beijing University of Posts and Telecommunications, China

Task 2: winner talk

Simple Online Multi-Object-Tracking

Zhipeng Luo, Yuehan Yao, Zhenyu Xu
DeepBlue Technology(Shanghai) Co., Ltd, China

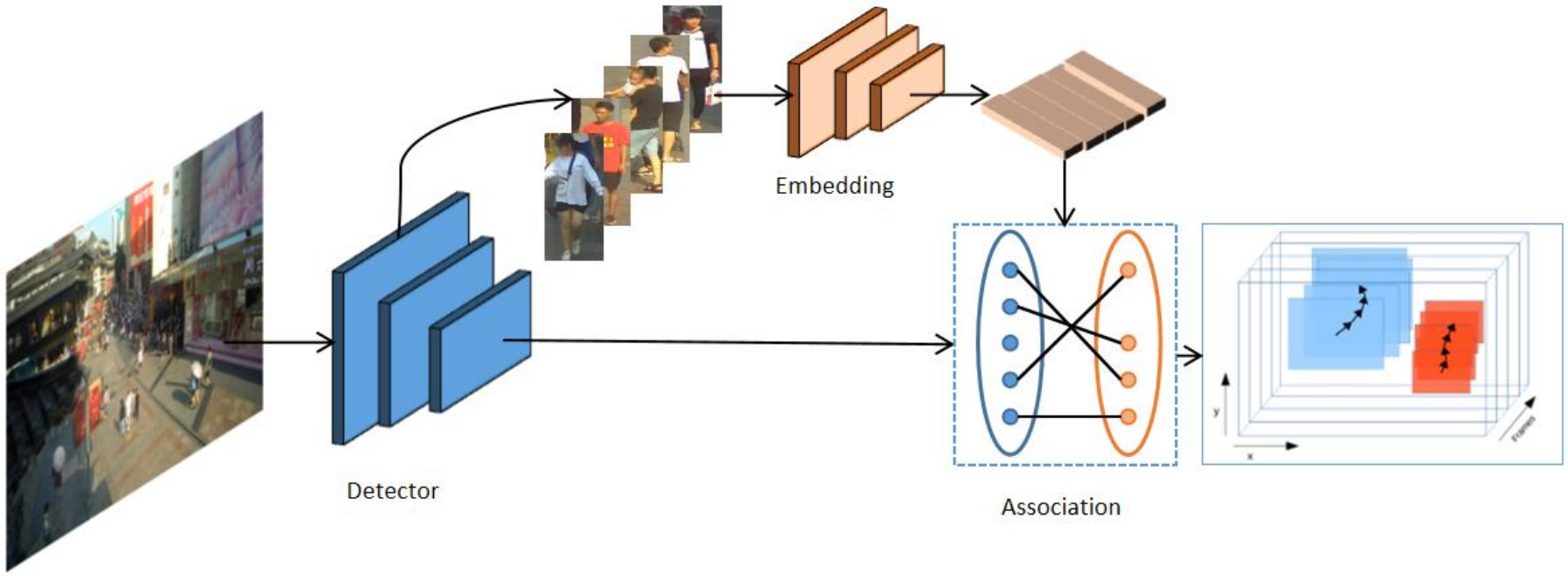


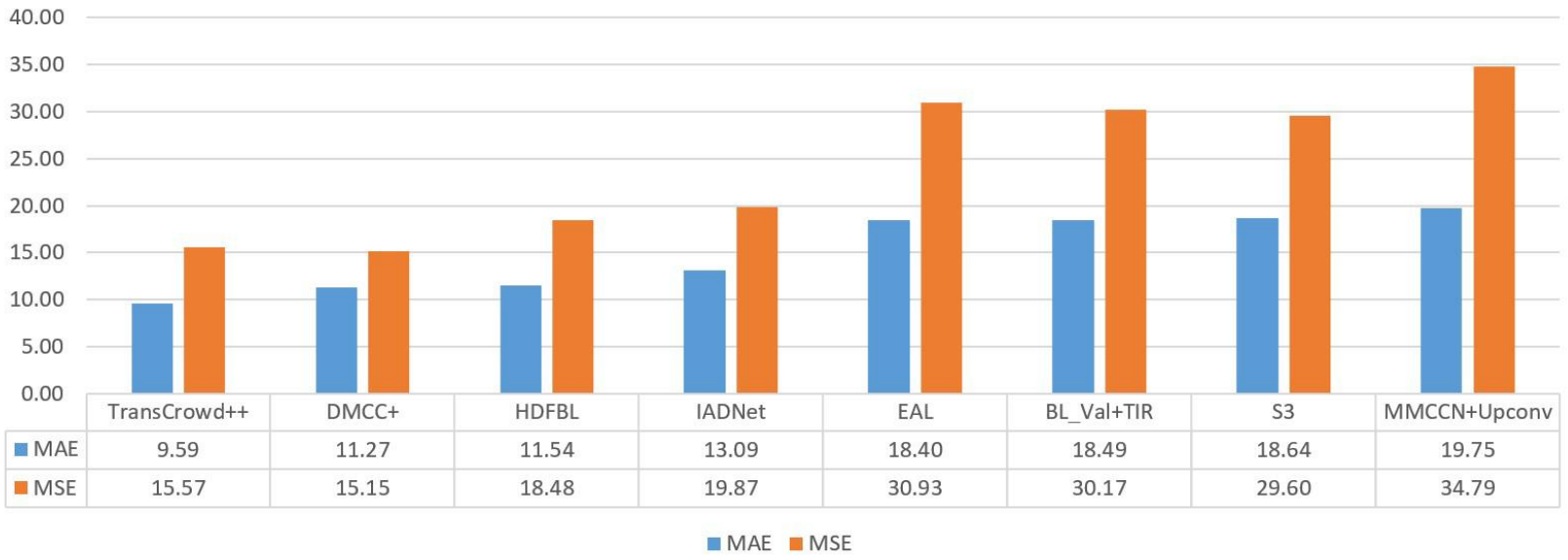
Figure: Architecture

Task 3: crowd counting

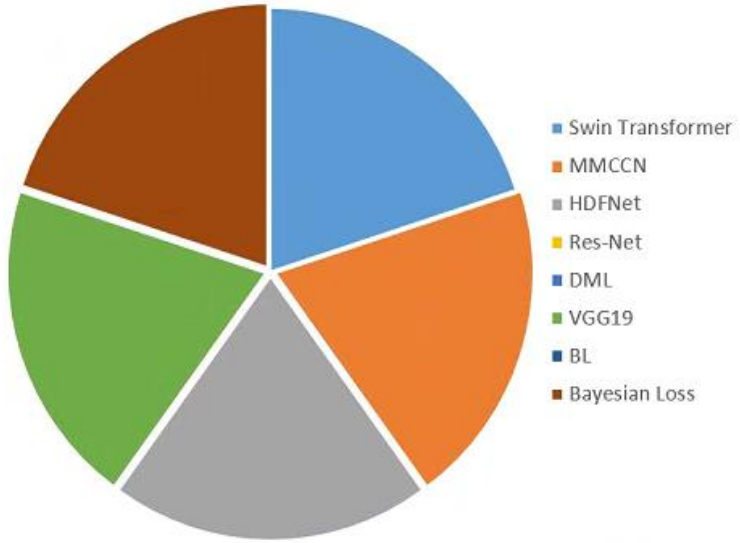
77 Teams

1 Winner

2 Honorable Mention



	TransCrowd++	DMCC+	HDFBL	IADNet	EAL	BL_Val+TIR	S3	MMCCN+Upconv
MAE	9.59	11.27	11.54	13.09	18.40	18.49	18.64	19.75
MSE	15.57	15.15	18.48	19.87	30.93	30.17	29.60	34.79



Task 3: crowd counting

Winner

RGBT Images Crowd Counting Using Swin Transformer (TransCrowd++)

Dingkang Liang, Xiwu Chen, Wei Xu, Xiang Bai

Huazhong University of Science and Technology, China

Honorable Mention

Density Map Learning based Crowd Counting Method for VisDrone RGBT Images (DMCC+)

Guanchen Ding, Lin Zhou, Ding Ding, Wenwei Han, Yiran Tao, Jingyuan Chen, Zhenzhong Chen

Wuhan University, China

Hierarchical Dynamic Filtering Network with Bayesian Loss for RGBTIR Crowd Counting (HDFBL)

Yabin Wang

Xi'an Jiaotong University, China

Task 3: winner talk

RGB-T Images Crowd Counting Using Swin Transformer (TransCrowd++)

Dingkang Liang, Xiwu Chen, Wei Xu, Xiang Bai
Huazhong University of Science and Technology, China

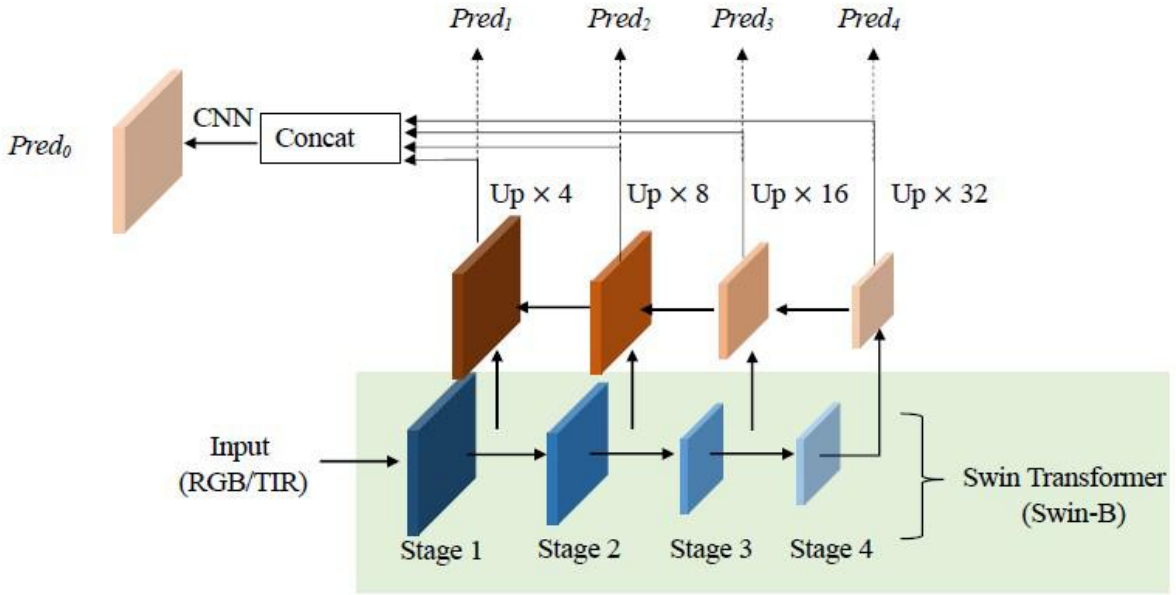


Figure: The pipeline of TransCrowd++.

Outline

- Application of Drones
- A Review of VisDrone 2021 Challenge
- Keynote
- Winner Announcement
- **Oral Presentation**

Paper Talk

Tackling the Background Bias in Sparse Object Detection via Cropped Windows

Leon Varga
University of Tuebingen, Germany

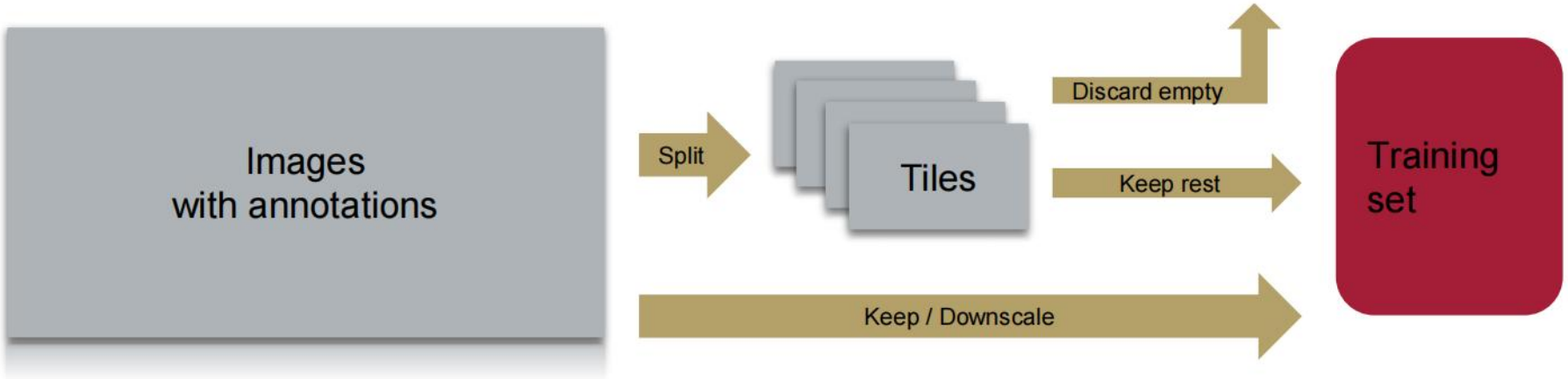


Figure: Training pipeline with Cropped Windows (CroW)

Paper Talk

TPH-YOLOv5: Improved YOLOv5 Based on Transformer Prediction Head for Object Detection on Drone-captured Scenarios

Xingkui Zhu
Beihang University, China

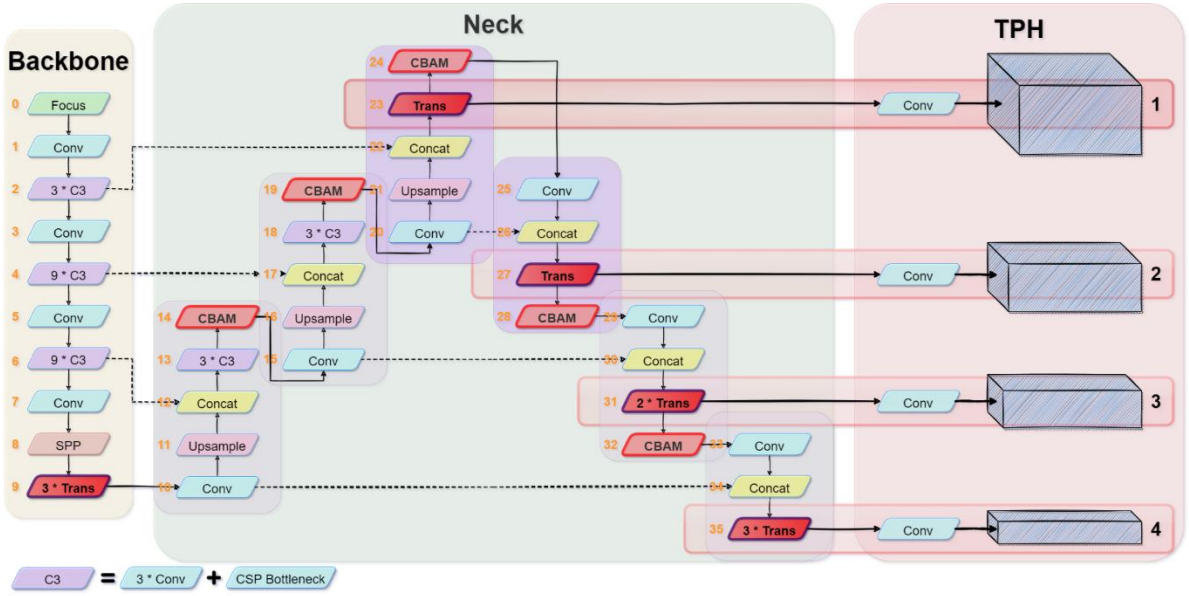


Figure: Architecture

Paper Talk

Coarse-grained Density Map Guided Object Detection in Aerial Images

Hongpeng Wang
Harbin Institute of Technology, China

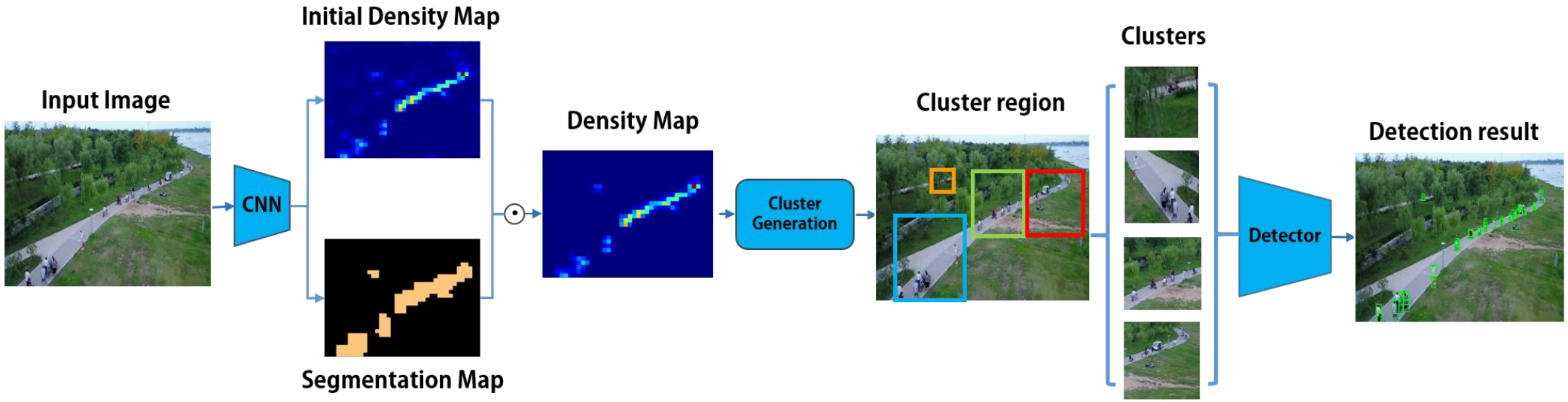


Figure: Architecture

Paper Talk

ViT-YOLO: Transformer-based YOLO for Object Detection

Xiaoqiang Lu
Xidian University, China

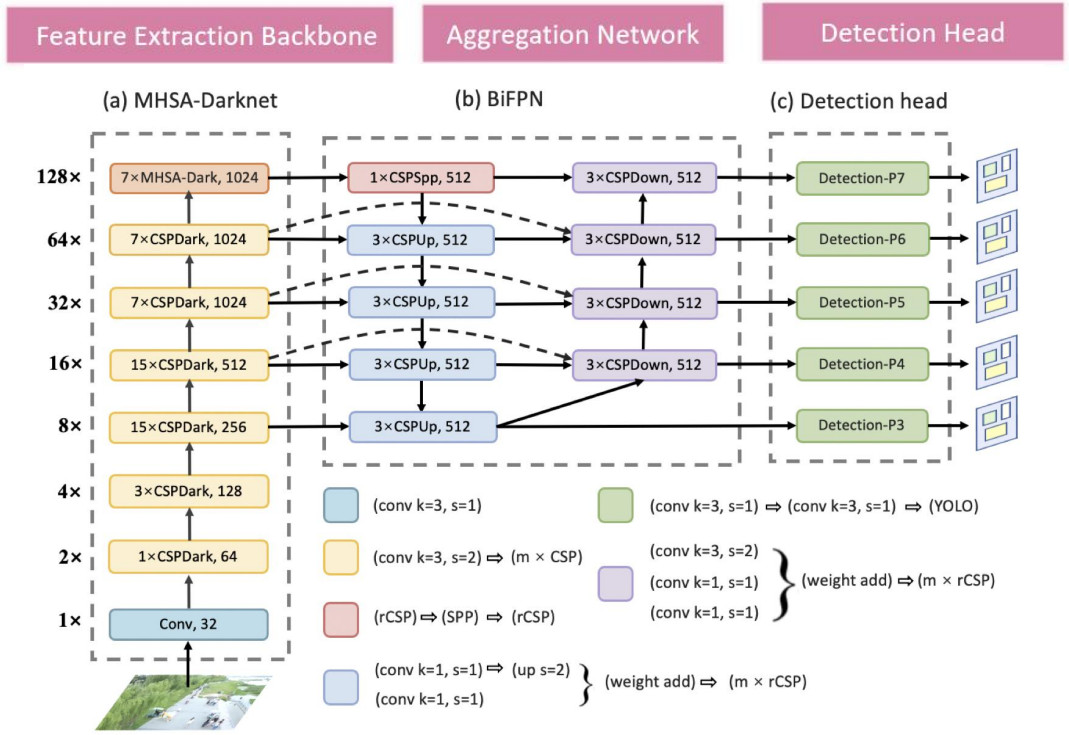


Figure: Architecture

Paper Talk

VistrongerDet: Stronger Visual Information for Object Detection in VisDrone Images

Junfeng Wan
Beijing University of Posts and Telecommunications, China

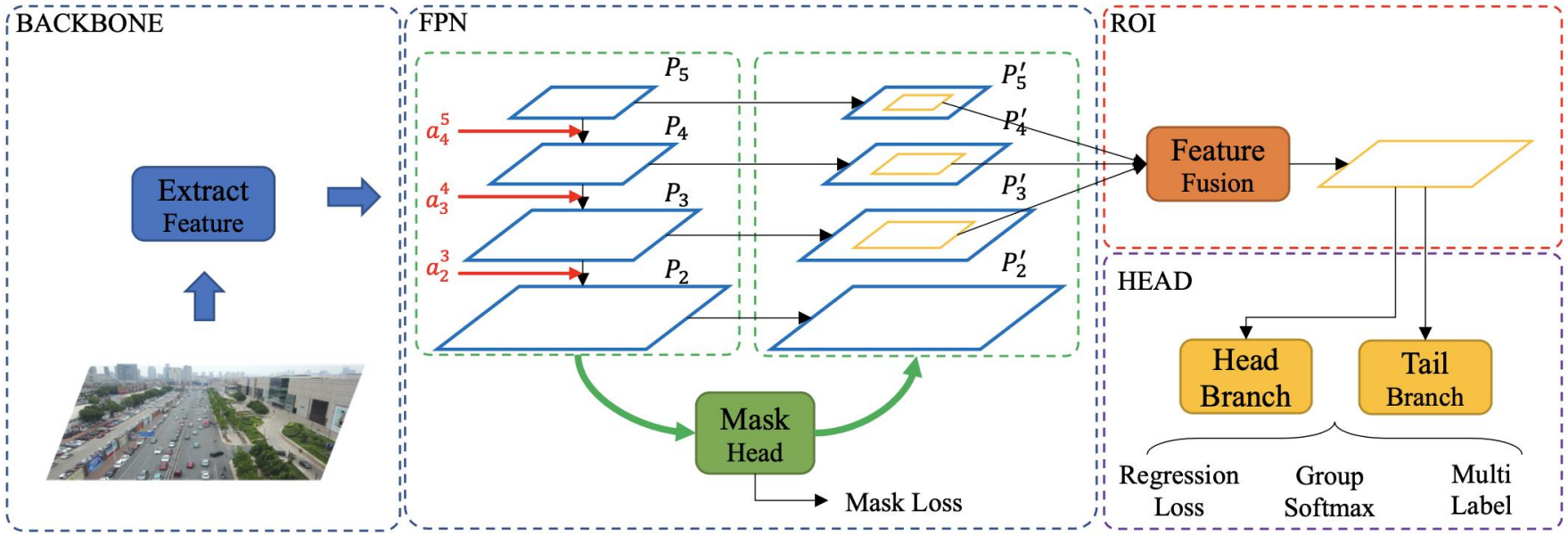


Figure: Architecture

Paper Talk

GIAOTracker: A Comprehensive Framework or MCMOT with Global Information and Optimizing Strategies in VisDrone 2021

Yunhao Du
Beijing University of Posts and Telecommunications, China

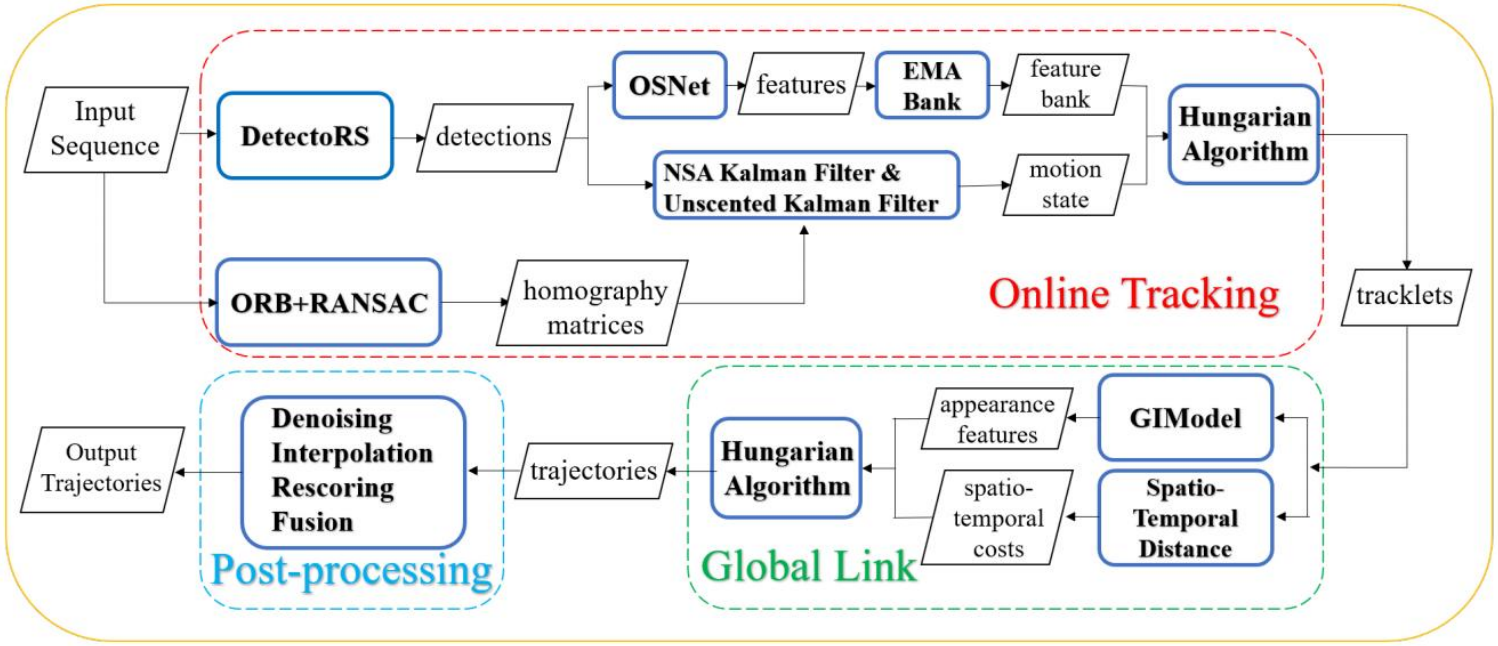


Figure: Overview of our proposed GIAOTracker pipeline for MCMOT

Advisory Committee

- Ming Liu (The Hong Kong University of Science and Technology, China)
- Qinghua Hu (Tianjin University, China)
- Steven Hoi (Singapore Management University, Singapore)
- Junwei Han (Northwestern Polytechnical University, China)
- Luc Van Gool (ETH Zurich, Switzerland)

Thank You

Enjoy ICCV!