1st place Solution for BDD100K MOT/MOTS/SSMOT/SSMOTS Challenges

Kaer Huang[1], Kanokphan Lertniphonphan[1], Feng Chen[1], Tao Zhang[2], Jun Xie[1], Huabing Liu[3], Qigang Wang[1], Zhepeng Wang[1]

[1]Lenovo Research [2]Tsinghua University [3] LCFC (Hefei) Electronics Technology Co., Ltd.

23/10/2022

Vehicle Computing In Lenovo Research

Lenovo aims to become the leader and enabler of Intelligent Transformation

Lenovo has been a recognized leader in standardized mass computing devices/units…

#1 WW PC

#3 WW Tablet

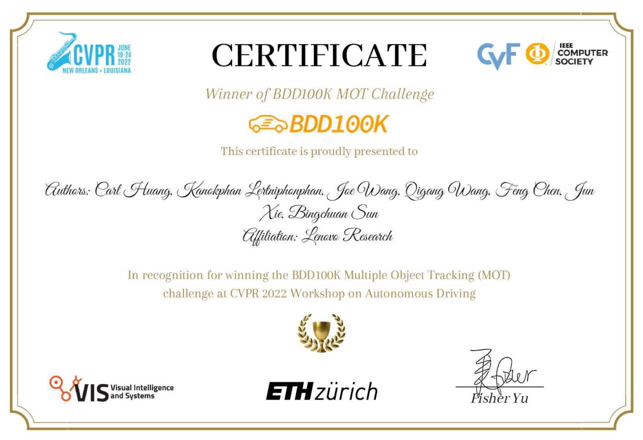
#2 LA SP

#4 WW x86 Server

→

The next opportunity to win big

#? Vehicle Computing



CVPR JUNE 19-24 2022 NEW ORLEANS • LOUISIANA

CERTIFICATE

CVF IEEE COMPUTER SOCIETY

Winner of BDD100K MOT Challenge

BDD100K

This certificate is proudly presented to

Authors: Carl Huang, Kanokphan Lertniphonphan, Joe Wang, Qigang Wang, Feng Chen, Jun Xie, Bingchuan Sun

Affiliation: Lenovo Research

In recognition for winning the BDD100K Multiple Object Tracking (MOT) challenge at CVPR 2022 Workshop on Autonomous Driving

VIS Visual Intelligence and Systems

ETH zürich

Fisher Yu

CES 2022 Demo

Contributors & Speakers

Competition Track	Authors	Affiliations
MOT	Kanokphan Lertniphonphan, Kaer Huang, Feng Chen, Jun Xie, Qigang Wang, Zhepeng Wang	Lenovo Research
MOTS	Kaer Huang, Kanokphan Lertniphonphan, Feng Chen, Jun Xie, Zhepeng Wang	Lenovo Research
SSMOT	Feng Chen[1], Kaer Huang[1], Huabing Liu[2], Kanokphan Lertniphonphan[1], Jun Xie[1], Zhepeng Wang[1]	[1]Lenovo Research, [2]LCFC (Hefei) Electronics Technology Co., Ltd.
SSMOTS	Zhepeng Wang[1], Kaer Huang[1], Feng Chen[1], Kanokphan Lertniphonphan[1], Jun Xie[1] , Tao Zhang[2]	[1]Lenovo Research, [2]Tsinghua University

Framework

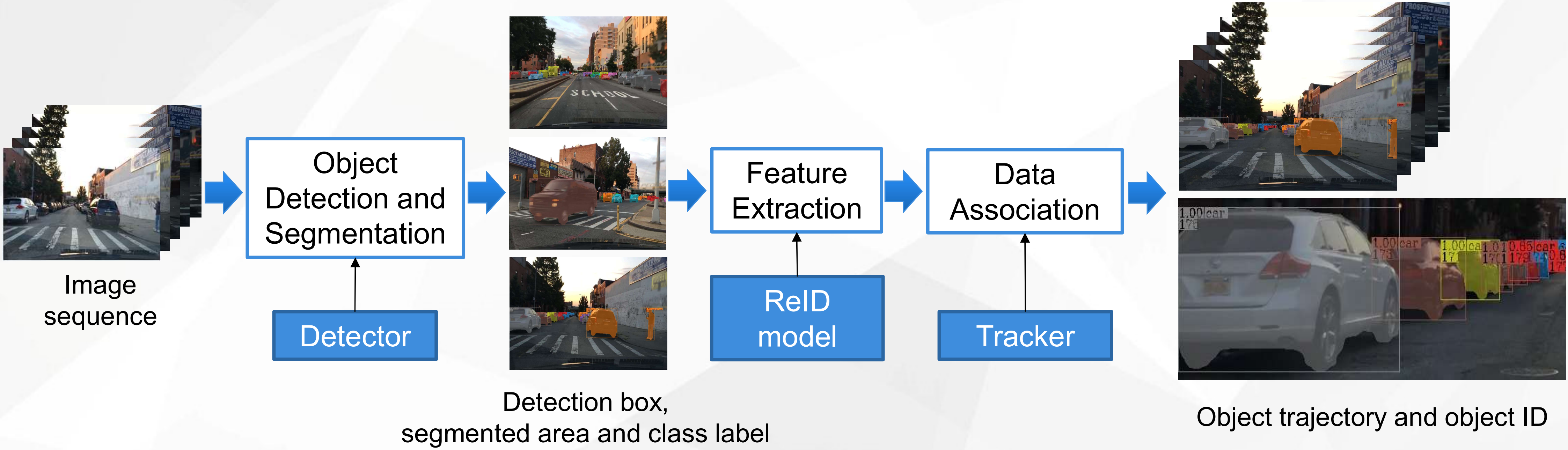


Image sequence

Object Detection and Segmentation

Detector

Detection box, segmented area and class label

Feature Extraction

ReID model

Data Association

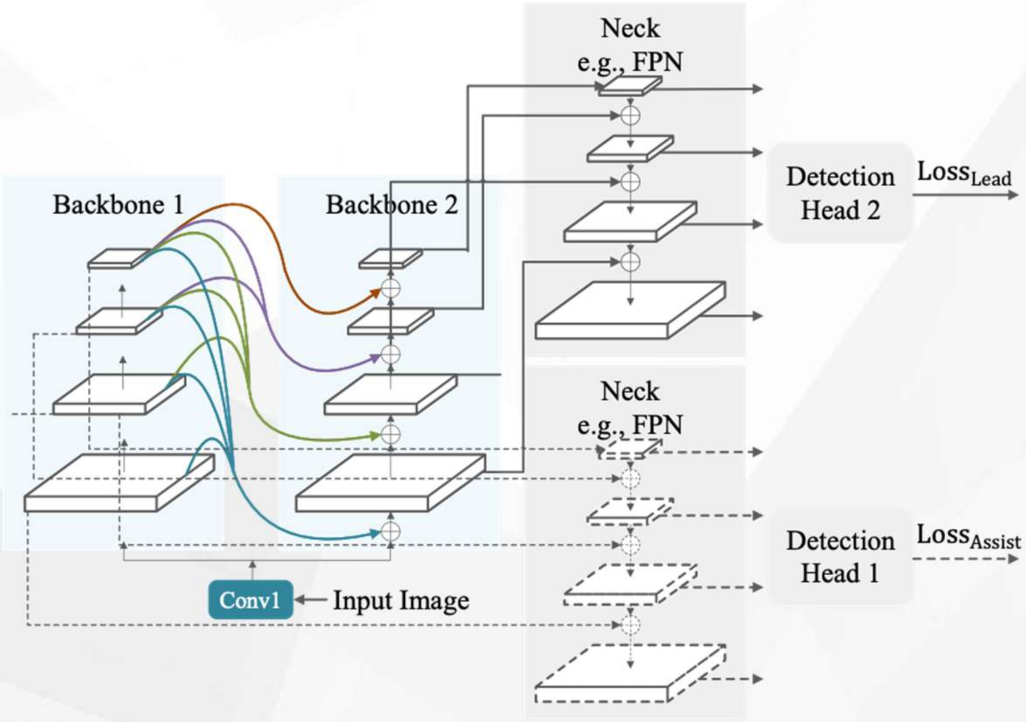
Tracker

Object trajectory and object ID

Object Detection

- **Detector**
  - CBNetV2
  - Experimental setting:
    - Backbone – Swin-L Transformer
    - Neck – FPN
    - Detection Head – HTC
    - Bbox_loss – GIoU Loss
    - Multi-threshold NMS

Backbone Pretrain: ImageNet22K



CBNetV2

Tingting Liang et al "CBNetV2: A Composite Backbone Network Architecture for Object Detection ", arXiv:2107.00420, 2021

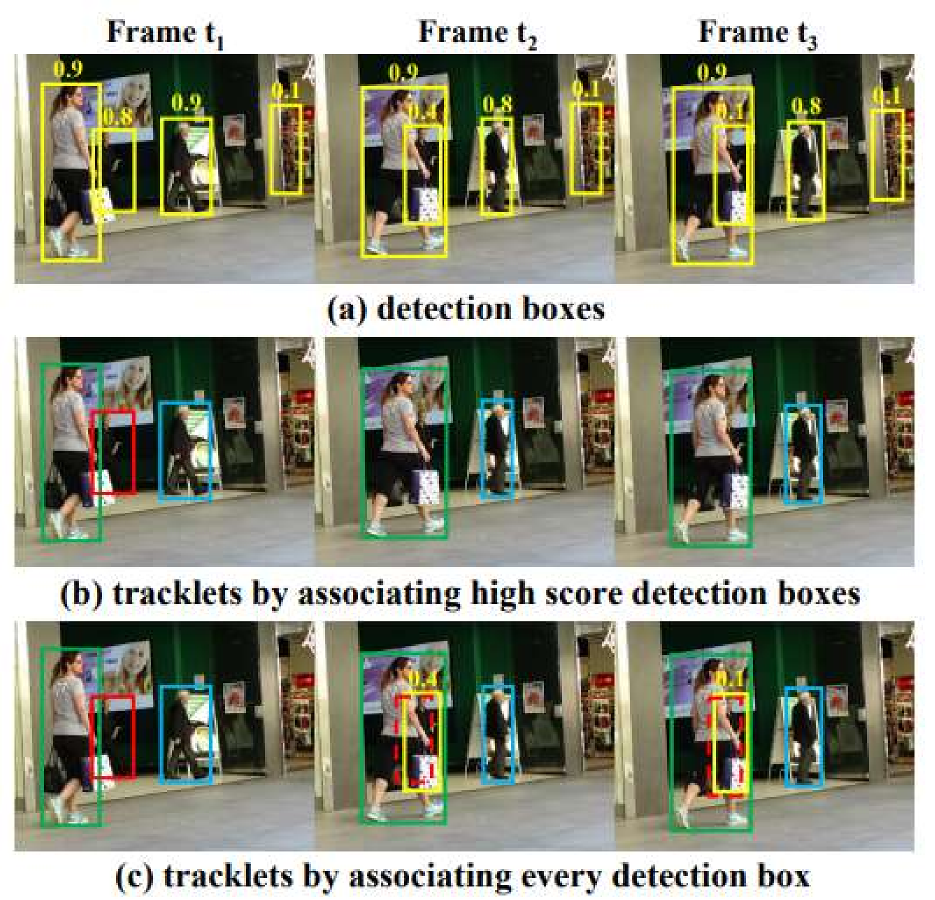
Tracking

- **Tracker**
  - ByteTrack: Multi-Object Tracking by Associating Every Detection Box
  - Tracking feature are combined by a detection score weighted sum

$$\hat{e}_j = \frac{\sum_{t=1}^{T} e_j^t \times s_j^t}{\sum_{t=1}^{T} s_j^t}$$

Similarity#1	Similarity#2	Similarity#3
High score detection box	Low score detection box	Tentative box

  - Similarity distance is based on ReID



(a) detection boxes

(b) tracklets by associating high score detection boxes

(c) tracklets by associating every detection box

ByteTrack method which associates every detected box

Zhang ,Yifu et al “ByteTrack: Multi-Object Tracking by Associating Every Detection Box”
Wu, Junfeng et al “In Defense of Online Models for Video Instance Segmentation”

MOT → MOTS

MOT（**Kanokphan Lertniphonphan**）

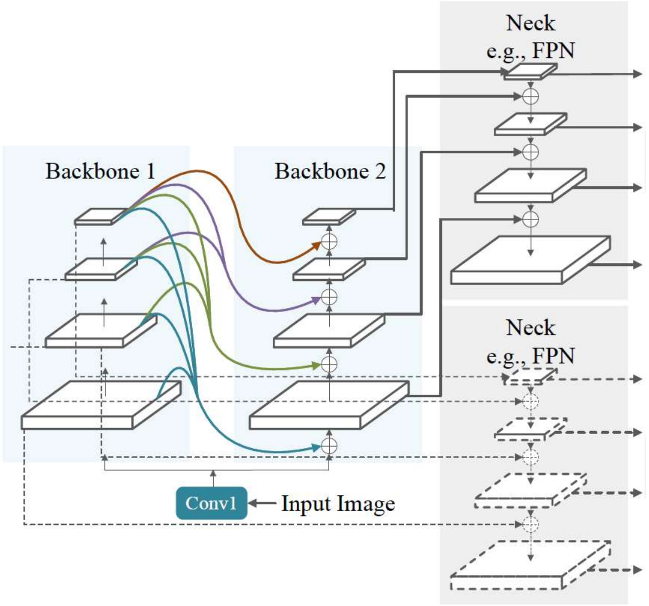
MOTS **(Carl Huang)**

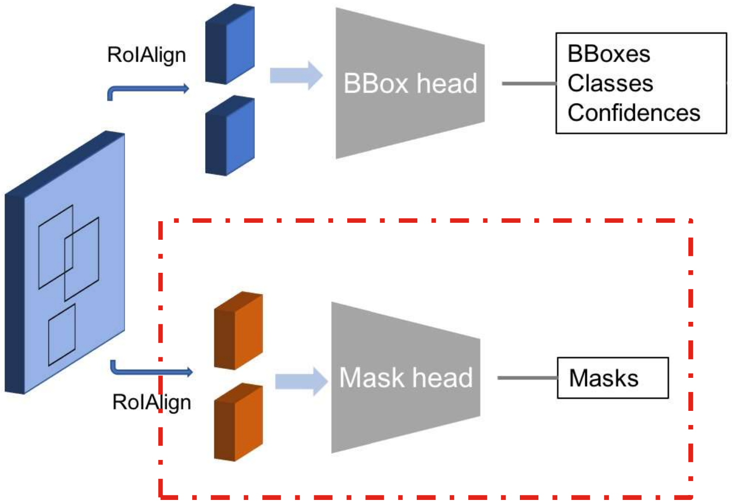
MOTS Solution

Detection

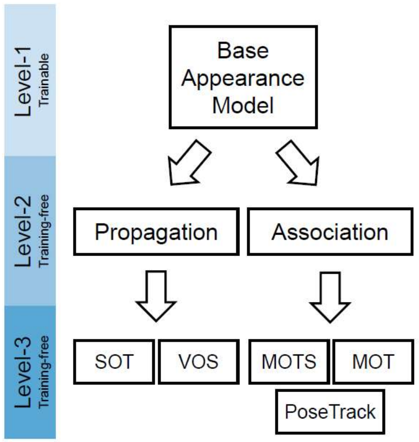
Re-ID

CBNetV2





Open/Add Mask Network



Zhang ,Yifu et al "ByteTrack: Multi-Object Tracking by Associating Every Detection Box"

Wang, Zhongdao et al "Do different tracking tasks require different appearance models? "

Tracking

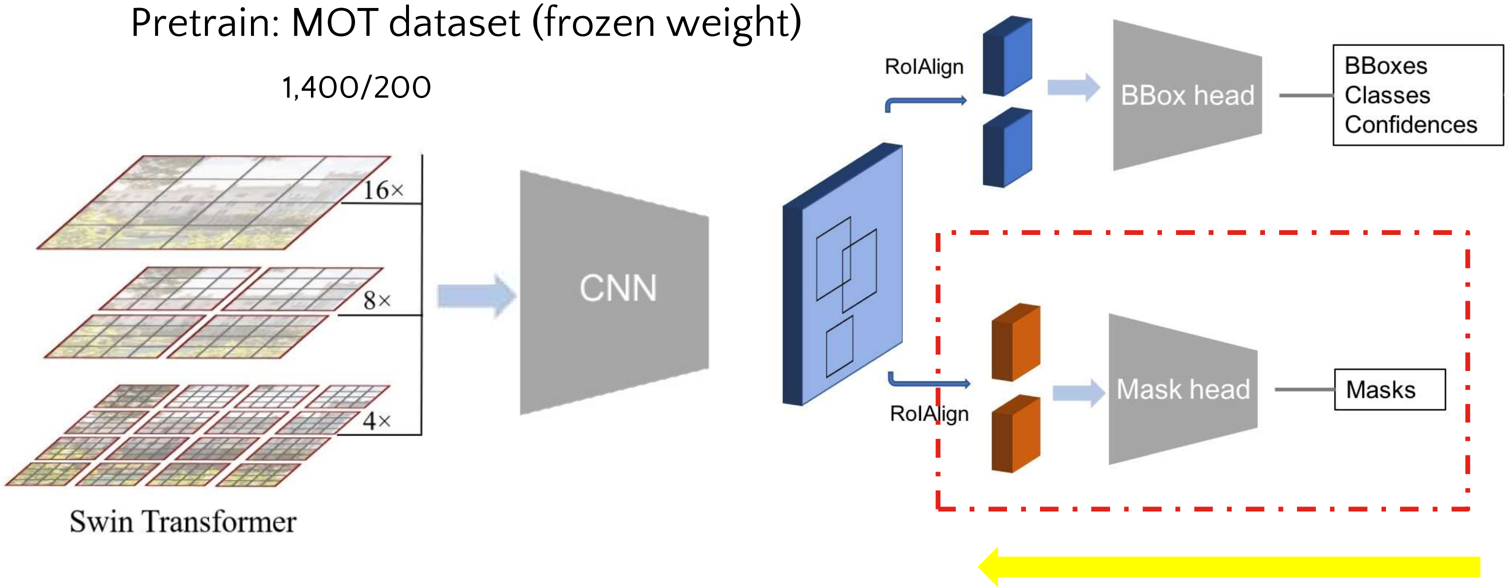
Detector With Mask(Frozen others & training mask)

MOTS dataset

154/32/37

Pretrain: MOT dataset (frozen weight)

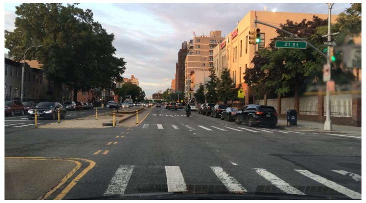
1,400/200



Training Detector with Mask (from frozen to fine tuning)

Data Distribution Change

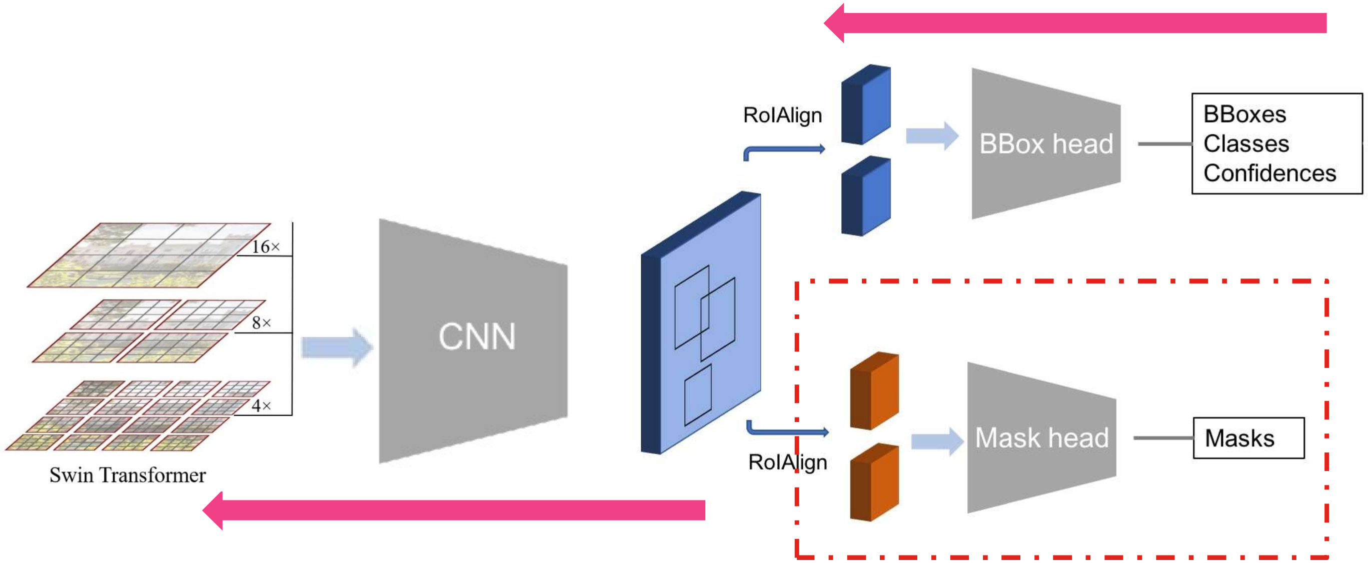
Normal weather(cloudy)



Snowy

Nighty

- Pretrain: MOT dataset

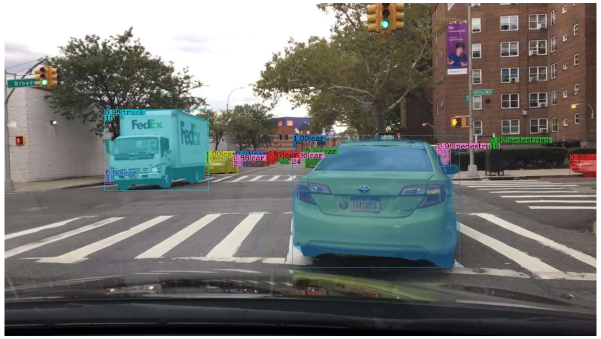


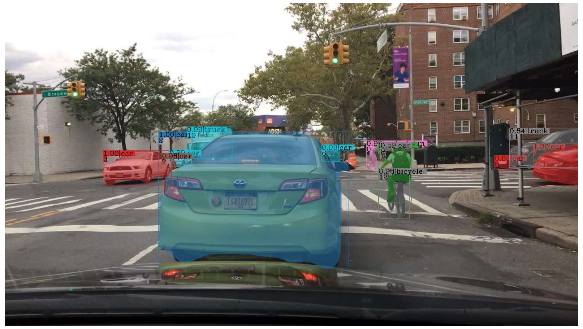
- MOTS dataset(fine tuning data)

154/32/37

Association(MOTS)

Box ReID





MOT/MOTS→SSMOT/SSMOTS

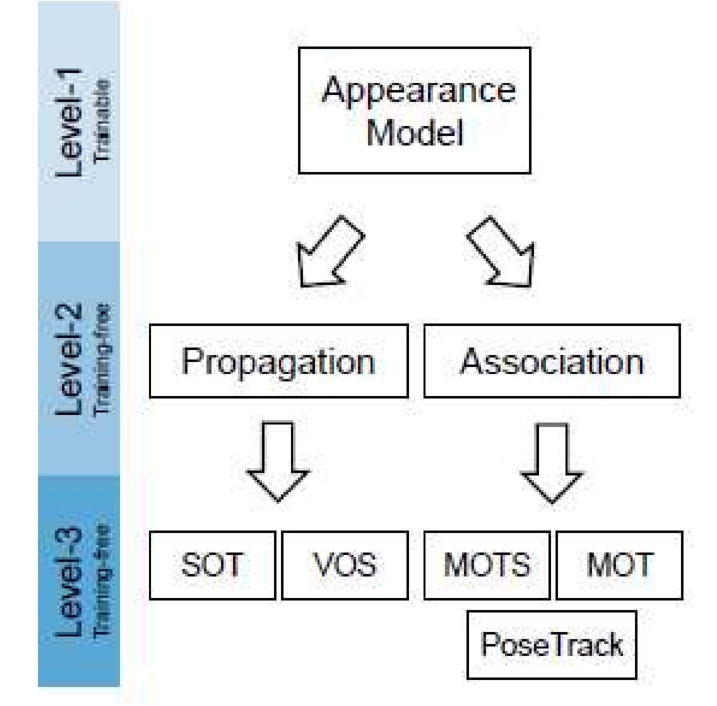
MOT/MOTS（**Kanokphan, Carl**）

SSMOT/SSMOTS **(Feng Chen)**

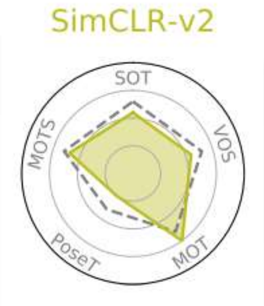
Re-Identification

- UniTrack: Do Different Tracking Tasks Require Different Appearance Models?
  - Appearance model
    - SimCLR-v2: Big Self-Supervised Models are Strong Semi-Supervised Learners
  - Association algorithm
    - By Class Hungarian Matching

Pretrain: ImageNet1K



UniTrack Framework

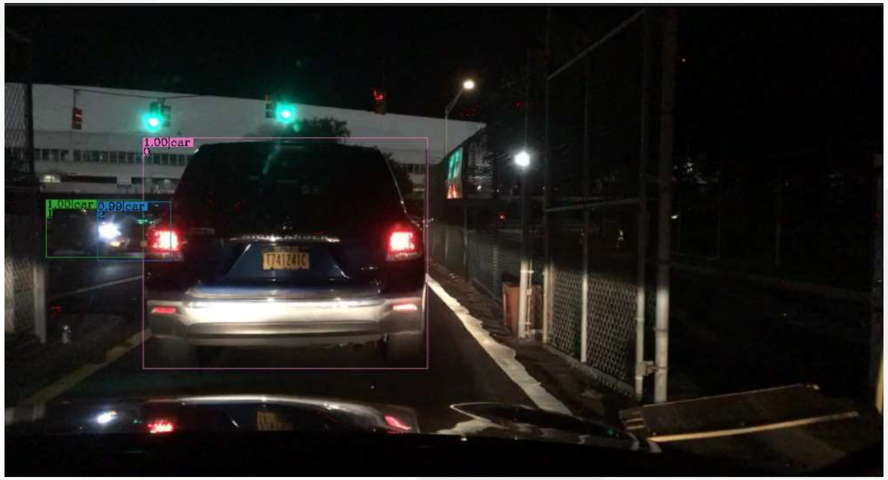


SimCLR-v2 performance on five tracking tasks

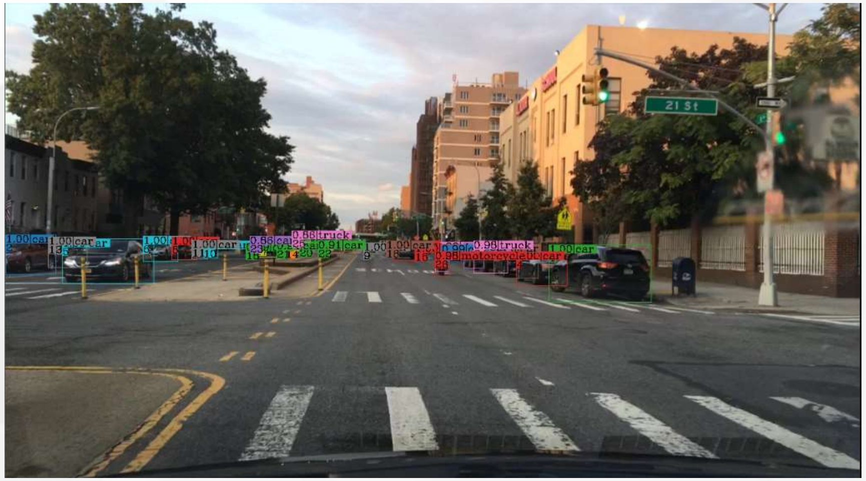
Wang, Zhongdao et al “Do different tracking tasks require different appearance models? “, NeruIPS 2021

Challenges Of ReID Model

- **Domain discrepancy** between BDD100K and ImageNet:
  - Night, diverse weather conditions
  - Small objects (~100 pixels)
- Teaser Track - **Self-Supervised Tracking**：no tracking annotations



Nighty

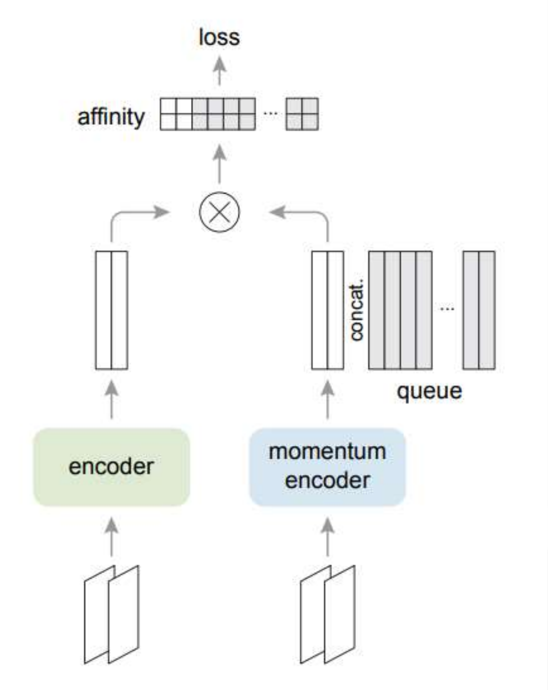


Small objects

Self-Supervised Learning For ReID Model

MoCo v2: Momentum Contrastive Learning

- Backbone: Resnet50

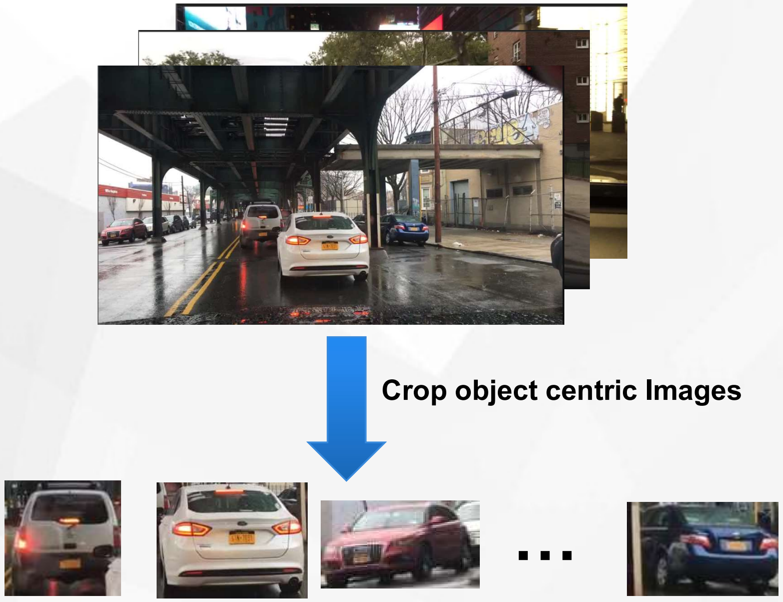


MoCo v2

Contrastive learning dataset generation:

- BDD100K images have multiple instances
- Crop object-wise images according to bounding box labels & pseudo labels.

Image-wise



Object-wise

- Kaiming He, et al. “Momentum contrast for unsupervised visual representation learning”, CVPR 2020
- Xinlei Chen, et al. “Improved baselines with momentum contrastive learning”, arXiv:2003.04297, 2020

Ablation Test

Ablation test on MOT validation dataset, and then applied the same configuration to SSMOT, MOTS and SSMOTS

Configuration	mHOTA	mMOTA
Baseline (CBNetV2+ ByteTrack + ReID)	48.8	45.0
+ Weighted ReID features	49.2 (+0.4)	45.3 (+0.4)
+ Contrastive Learning ReID Model	50.0 (+0.7)	45.8 (+0.5)
+ Tuning Matching Threshold	**50.0**	**45.9 (+0.1)**

Results

MOT and SSMOT

Split	mHOTA	mMOTA	mIDF1	mDetA	mAssA	mMOTP
Val	**50.0**	45.9	60.5	45.1	56.6	82.9
Test	**49.2**	43.0	59.5	43.9	56.4	81.4

MOTS and SSMOTS

Split	mHOTA	mMOTA	mIDF1	mDetA	mAssA	mMOTP
Val	**38.2**	37.8	47.0	35.6	42.8	70.5
Test	**44.0**	41.1	54.9	39.3	50.8	69.7

Summary

- Our framework is based on tracking by detection which consists of
  - Detector: CBNetV2
  - Re-ID: Unitrack (**MoCo-v2)**
    - Self-supervised learning on BDD dataset
  - Tracker: ByteTrack **(all ReID)**
    - Multi-class NMS
    - Weighted ReID features
  - **Mask head**

References

- Yu, F., Chen, H., Wang, X., Xian, W., Chen, Y., Liu, F., Madhavan, V., & Darrell, T. (2020). BDD100K: A Diverse Driving Dataset for Heterogeneous Multitask Learning. 2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), 2633-2642.
- Liang, T., Chu, X., Liu, Y., Wang, Y., Tang, Z., Chu, W., Chen, J., & Ling, H. (2021). CBNetV2: A Composite Backbone Network Architecture for Object Detection. *ArXiv, abs/2107.00420*.
- Ge, Z., Liu, S., Wang, F., Li, Z., & Sun, J. (2021). YOLOX: Exceeding YOLO Series in 2021. ArXiv, abs/2107.08430.
- Zhang, Y., Sun, P., Jiang, Y., Yu, D., Yuan, Z., Luo, P., Liu, W., & Wang, X. (2021). ByteTrack: Multi-Object Tracking by Associating Every Detection Box. *ArXiv, abs/2110.06864*.
- Wang, Z., Zhao, H., Li, Y., Wang, S., Torr, P.H., & Bertinetto, L. (2021). Do Different Tracking Tasks Require Different Appearance Models? *NeurIPS*.
- Chen, T., Kornblith, S., Swersky, K., Norouzi, M., & Hinton, G.E. (2020). Big Self-Supervised Models are Strong Semi-Supervised Learners. ArXiv, abs/2006.10029.
- Liu, Z., Lin, Y., Cao, Y., Hu, H., Wei, Y., Zhang, Z., Lin, S., & Guo, B. (2021). Swin Transformer: Hierarchical Vision Transformer using Shifted Windows. 2021 IEEE/CVF International Conference on Computer Vision (ICCV), 9992-10002.
- Kuhn, H.W. (1955). The Hungarian method for the assignment problem. Naval Research Logistics Quarterly, 2, 83-97.
- He, K., Fan, H., Wu, Y., Xie, S., and Girshick, R. Momentum contrast for unsupervised visual representation learning. CVPR 2020.
- Chen, X., Fan, H., Girshick, R., He, K.: Improved baselines with momentum contrastive learning. arXiv preprint arXiv:2003.04297
- MMTracking: OpenMMLab video perception toolbox and benchmark, https://github.com/open-mmlab/mmtracking
- MMSelfsup: OpenMMLab self-supervised representation learning toolbox, https://github.com/open-mmlab/mmselfsup

thanks.

Different is better

Contact information: Carl Huang 黄卡尔 huangke1@lenovo.com