Real-time object detection

August 12, 2022

Sogang University
Vision & Display Systems Lab, Dept. of Electronic Engineering

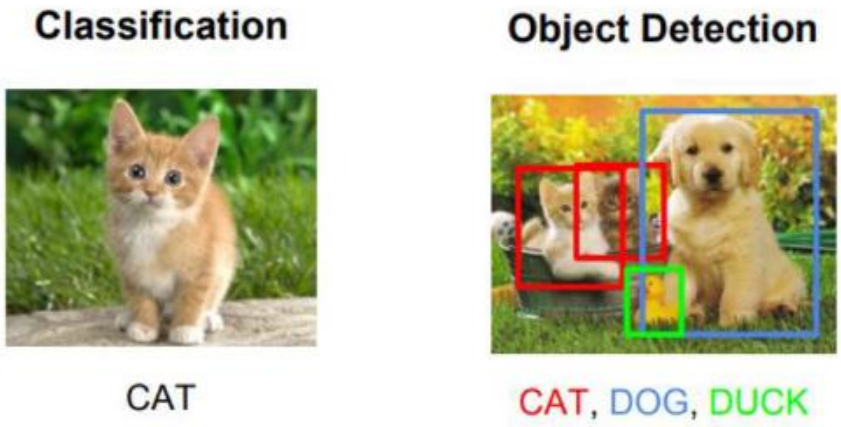
Presented By
이제임스

Outline

- Background
  - Object detection
- Papers
  - YOLOX: Exceeding YOLO Series in 2021
  - Real-time Object Detection for Streaming Perception (CVPR 2022)

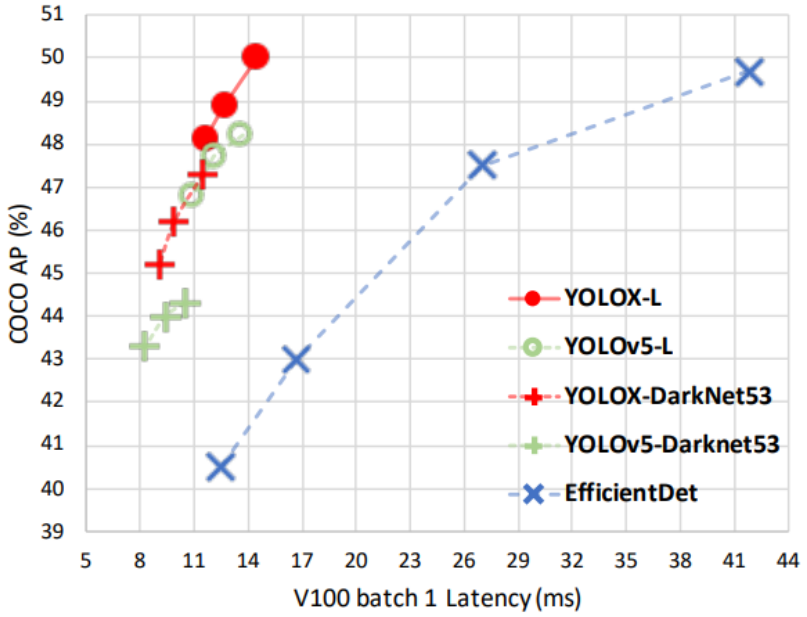
Background

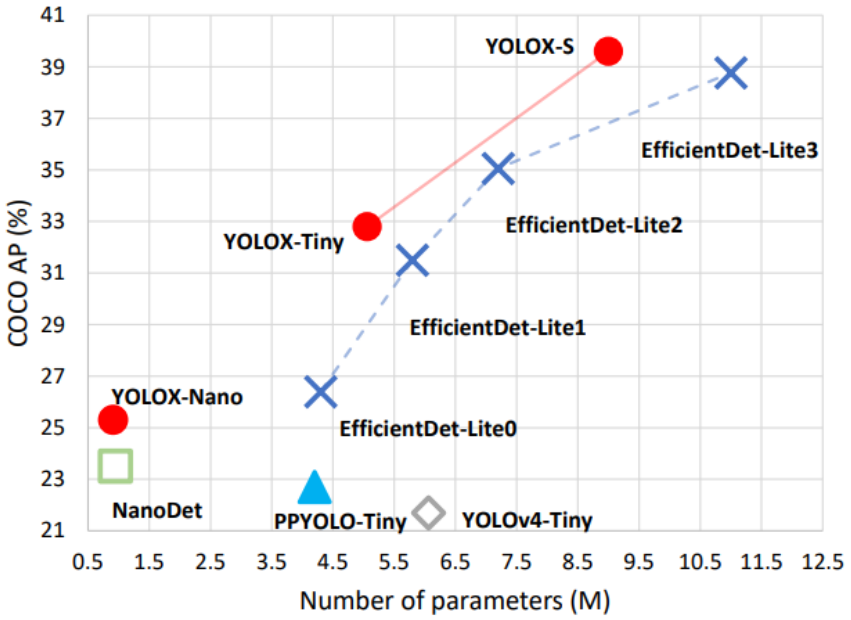
- Object detection
  - Detecting specific objects that are significant within images and videos
    - Localization : find where the object exists
    - Classification: determine what objects exist in the local
- Real-time object detection
  - Real-time detection with fast processing speed, giving up a bit of accuracy
    - 2-Stage Detector: regional proposal, classification ➔ sequential (R-CNN, Faster R-CNN)
    - 1-Stage Detector: regional proposal, classification ➔ simultaneously (YOLO, SSD)



YOLOX: Exceeding YOLO Series in 2021

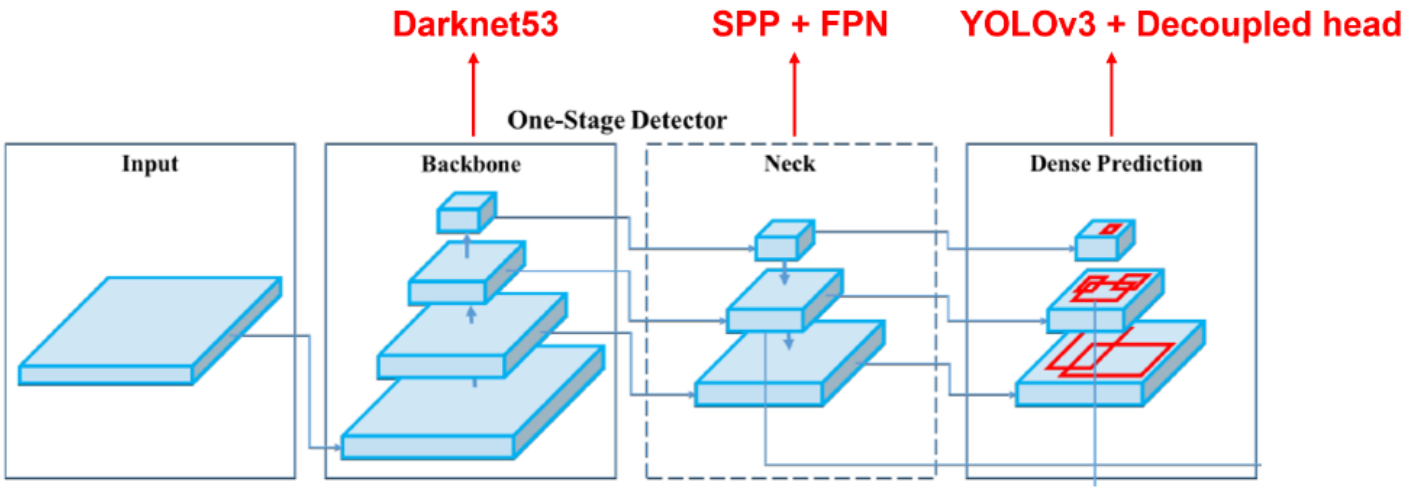
- Stream Perception Challenge (Workshop on Autonomous Driving at CVPR 2021)
  - Architecture
  - Anchor-free detection
  - Decoupled head
  - Multi-positive



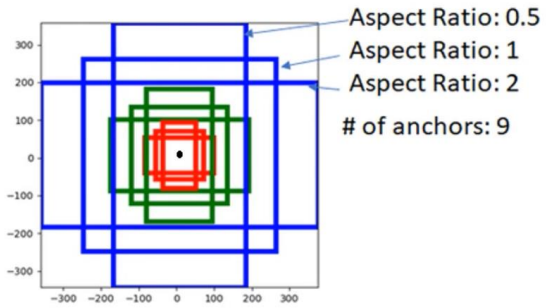


Architecture

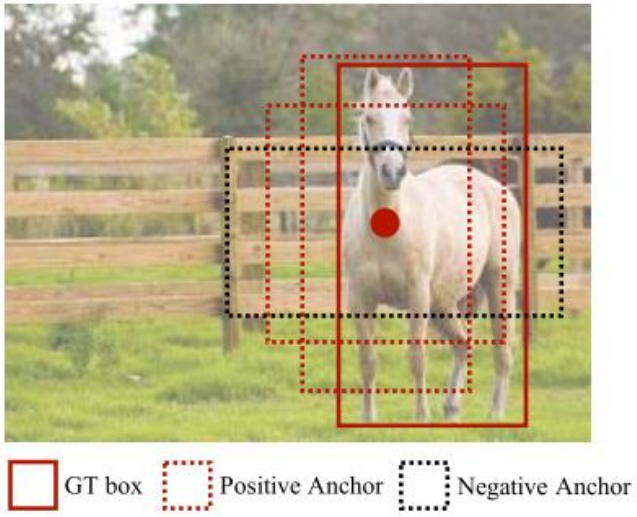
- Structure
  - DarkNet53 (backbone, YOLOv3) ➔ extract feature map
  - SPP (Spatial Pyramid Pooling layer)
    - Network with multiple layers of pooling
  - FPN (Feature Pyramid Network)
    - Add high level extraction information to row level feature map
  - Dense prediction
    - Classification, regression, objectness

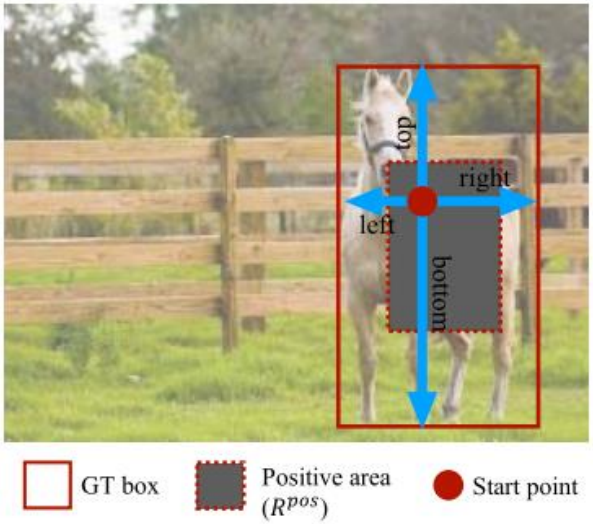


Method



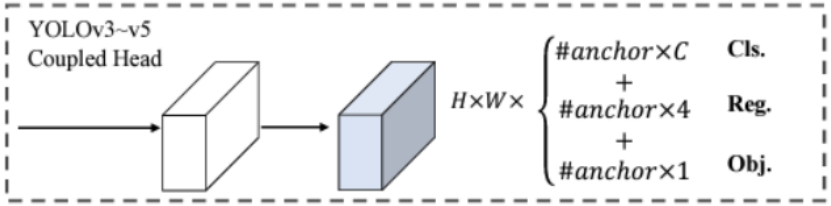
- Anchor-free detection
  - Anchor (YOLOv2-v5)
    - Settings for creating bounding box
    - 특정 cell에서 regression 진행 – IoU threshod이상이면 해당 anchor box는 positive
  - Anchor-free (YOLOX)
    - Ground Truth 안에 cell들이 모두 positive sample
    - 해당점에서 GT까지의 길이를 학습 ➔ bounding box 생성

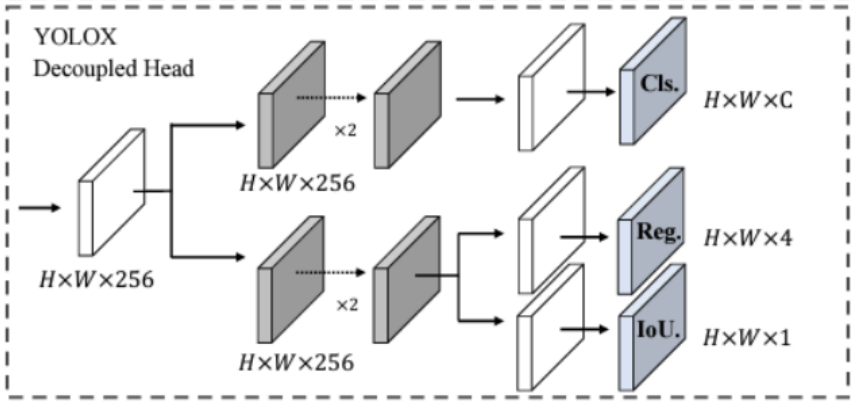




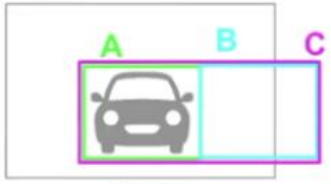
Method

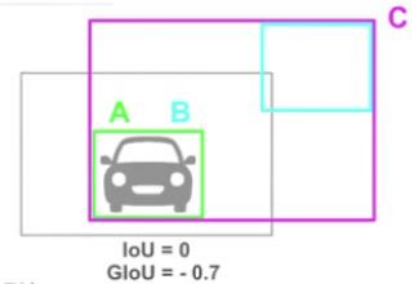
- Decoupled head
  - Classification and bounding box Regression have different characteristics
    - Classification : FC Head
    - Localization : Convolution Head





- IoU loss

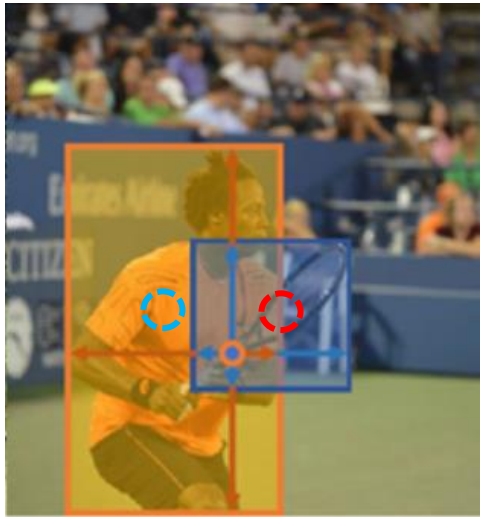


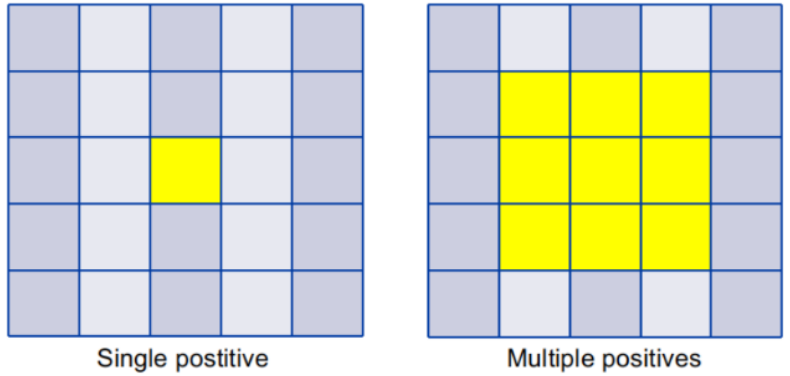


$$GIoU = IoU - \frac{|C \backslash (A \cup B)|}{|C|}$$

Method

- Multi-positive
  - Center-ness
    - If the Object label is unclear, only the center cell is set to positive among all cells in each area
  - Multi-positive
    - In this paper, some cells around the center are set to positive
  - SimOTA
    - Select only cells with low loss (top-k) ➔ positive

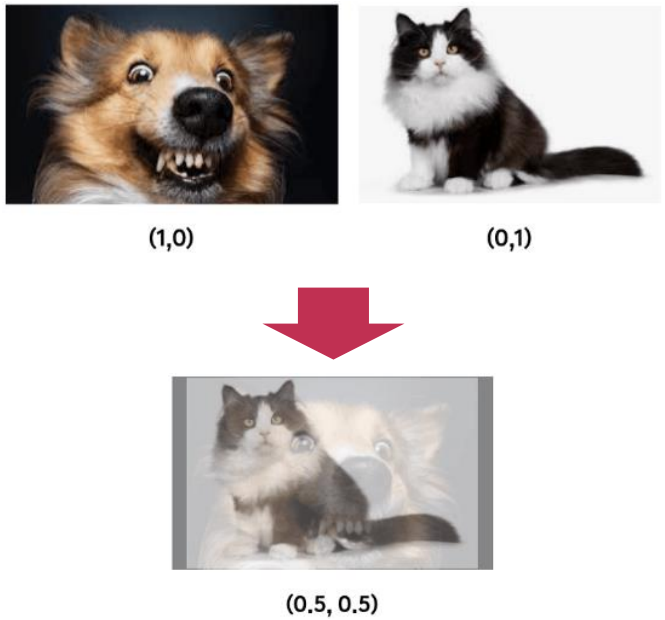




Method

- Strong augmentation
  - Mosaic
    - Augmentation Techniques for Making Four Images into One Sheet➔ detect small object
  - MixUP
    - Training image and label interpolation➔ Apply train data and label respectively to learn

<Mosaic>



<Mixup>

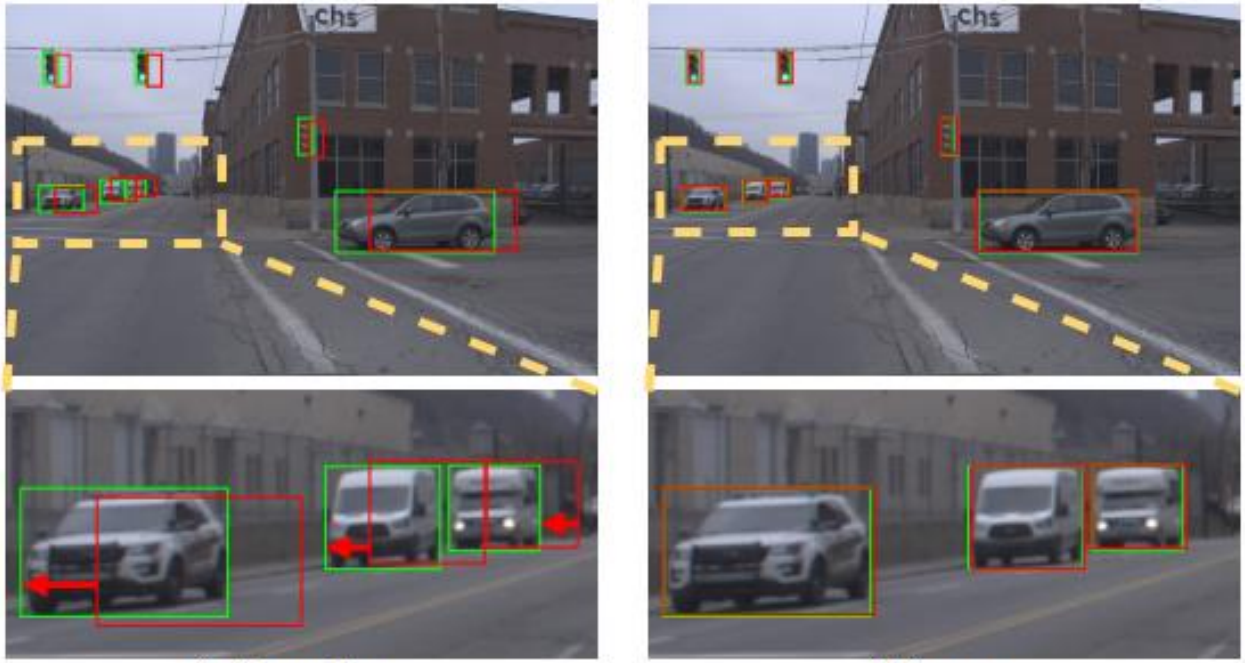
Experiments

- Quantitative results

Method	Backbone	Size	FPS (V100)	AP (%)	AP_{50}	AP_{75}	AP_S	AP_M	AP_L
YOLOv3 + ASFF* [18]	Darknet-53	608	45.5	42.4	63.0	47.4	25.5	45.7	52.3
YOLOv3 + ASFF* [18]	Darknet-53	800	29.4	43.9	64.1	49.2	27.0	46.6	53.4
EfficientDet-D0 [28]	Efficient-B0	512	98.0	33.8	52.2	35.8	12.0	38.3	51.2
EfficientDet-D1 [28]	Efficient-B1	640	74.1	39.6	58.6	42.3	17.9	44.3	56.0
EfficientDet-D2 [28]	Efficient-B2	768	56.5	43.0	62.3	46.2	22.5	47.0	58.4
EfficientDet-D3 [28]	Efficient-B3	896	34.5	45.8	65.0	49.3	26.6	49.4	59.8
PP-YOLOv2 [11]	ResNet50-vd-dcn	640	68.9	49.5	68.2	54.4	30.7	52.9	61.2
PP-YOLOv2 [11]	ResNet101-vd-dcn	640	50.3	50.3	69.0	55.3	31.6	53.9	62.4
YOLOv4 [1]	CSPDarknet-53	608	62.0	43.5	65.7	47.3	26.7	46.7	53.3
YOLOv4-CSP [30]	Modified CSP	640	73.0	47.5	66.2	51.7	28.2	51.2	59.8
YOLOv3-ultralytics[2]	Darknet-53	640	95.2	44.3	64.6	-	-	-	-
YOLOv5-M [7]	Modified CSP v5	640	90.1	44.5	63.1	-	-	-	-
YOLOv5-L [7]	Modified CSP v5	640	73.0	48.2	66.9	-	-	-	-
YOLOv5-X [7]	Modified CSP v5	640	62.5	50.4	68.8	-	-	-	-
YOLOX-DarkNet53	Darknet-53	640	90.1	47.4	67.3	52.1	27.5	51.5	60.9
YOLOX-M	Modified CSP v5	640	81.3	46.4	65.4	50.6	26.3	51.0	59.9
YOLOX-L	Modified CSP v5	640	69.0	50.0	68.5	54.5	29.8	54.5	64.4
YOLOX-X	Modified CSP v5	640	57.8	**51.2**	69.6	55.7	31.2	56.1	66.1

Real-time Object Detection for Streaming Perception (CVPR 2022)

- Streaming perception task
  - Take the model processing latency into account
    - Consider the online processing latency
    - Predict future results on the online setting



(a) baseline (b) ours

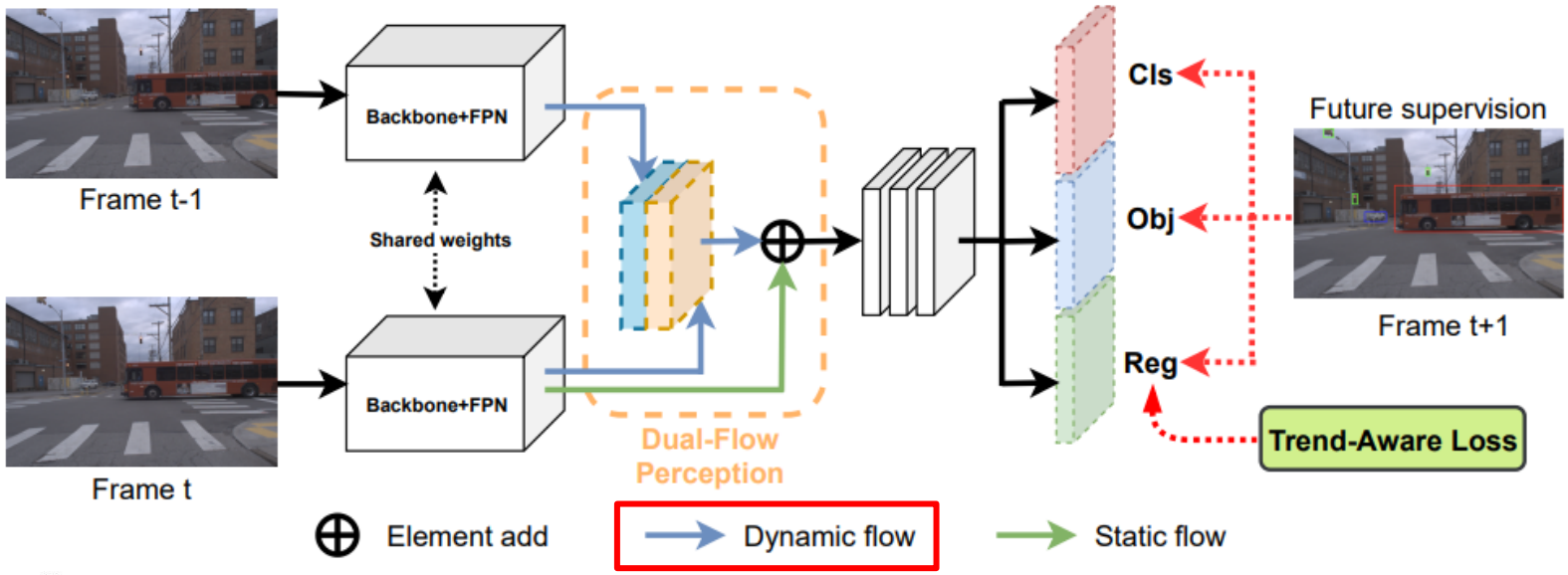
Method

- Training
  - Pipeline
    - Use YOLO-X as base detector
      - Remove TensorRT and change the input scale to the half resolution
  - Rebuild the training dataset : triplet (F_{t-1}, F_t, G_{t+1})
    - To know the moving status ➔ to predict the detection results of the next frame

  - Dual-Flow Perception Module (DFP) + Trend-Aware Loss (TAL)
    - To better capture the moving trend between two input frames

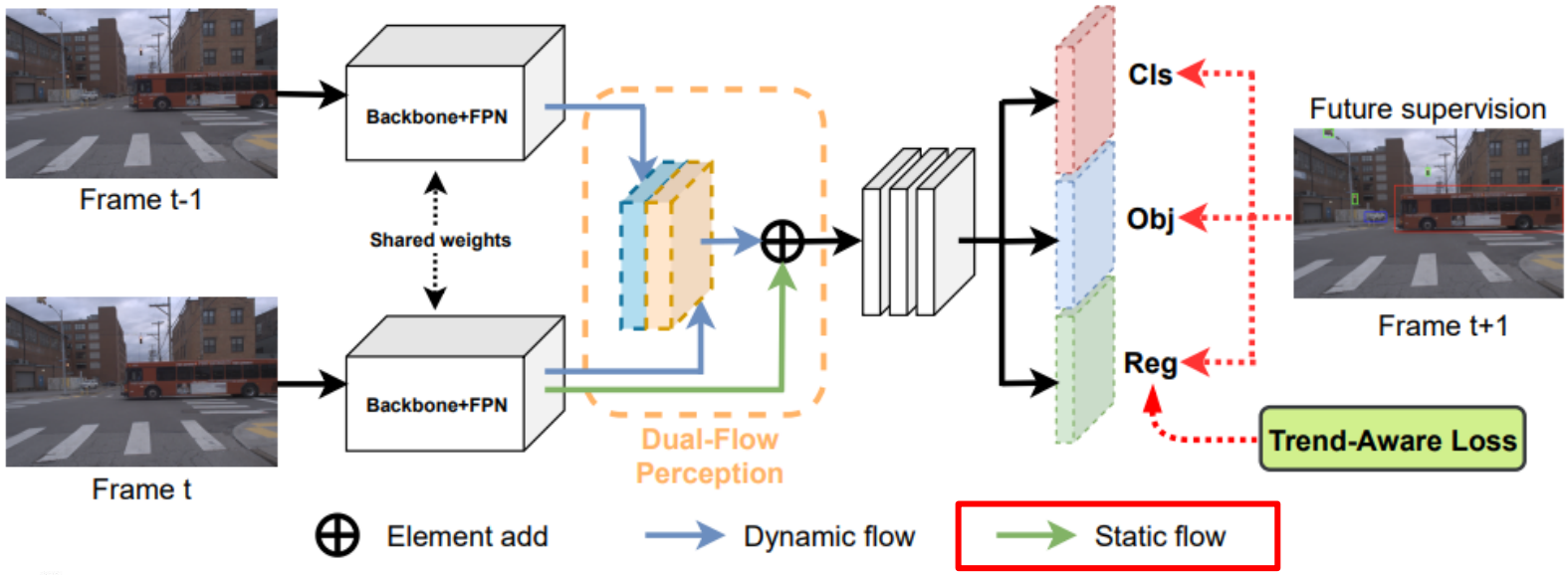
Method

- Dual-Flow Perception Module (DFP)
  - Dynamic flow
    - Fuse the FPN feature of two adjacent frames to learn the moving information
      - Shared weight reduces the channel to half numbers
      - Concatenate two reduced features to generate dynamic features



Method

- Dual-Flow Perception Module (DFP)
  - Static flow
    - Add the original feature of the current frame through a residual connection
      - Provide the basic information for detection
      - Improve the predicting robustness across different moving speeds



Method

- Trend-Aware Loss (TAL)
  - Adaptive weight for each object (different size, moving states)
    - Pay more attention to fast-moving objects
  - Define trend factor
    - Calculate IoU $\{F_{t-1}, F_t\}$ + max operation ➔ mIoU
      - The small value of the matching IoU means the fast-moving speed
      - If a new object comes ➔ matching IoU is much smaller ➔ threshold τ

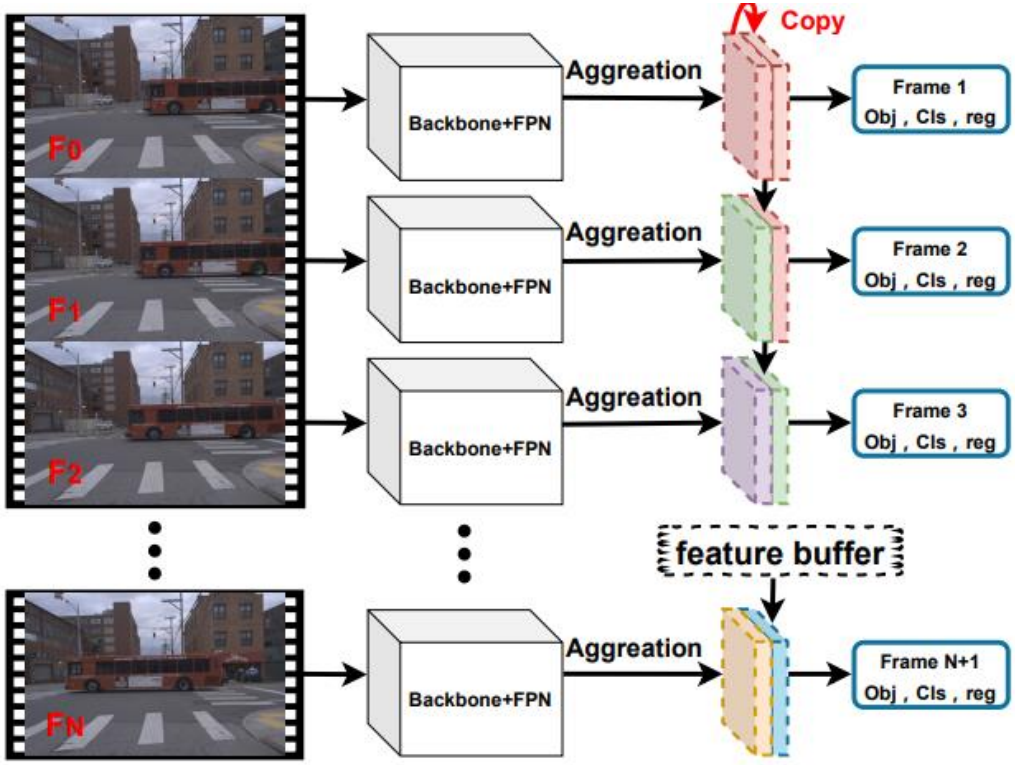
$$mIoU_i = \max_j(\{IoU(box_i^{t+1}, box_j^t)\})$$

$$\omega_i = \begin{cases} 1/mIoU_i & mIoU_i \geq \tau \\ 1/\nu & mIoU_i < \tau \end{cases}$$

$$\mathcal{L}_{total} = \sum_{i \in positive} \hat{\omega}_i \mathcal{L}_i^{reg} + \mathcal{L}_{cls} + \mathcal{L}_{obj}$$

Method

- Inference
  - Feature buffer
    - Store all the feature maps of previous frame
    - Duplicate the FPN feature maps as pseudo historical buffers (beginning frame F0)



Experiments

- Quantitative results

Model	Pipe.	DFP	TAL	Off AP	sAP	sAP_{50}	sAP_{75}
YOLOX-S				32.0	26.3	48.1	24.0
	✓				27.6 ↑ 1.3	48.3	26.1
	✓	✓			28.2 (+0.6)	49.4	27.4
	✓		✓		28.1 (+0.5)	49.1	27.0
	✓	✓	✓		28.8 (+1.2)	50.3	27.6
YOLOX-M				34.5	29.2	51.9	27.7
	✓				31.2 ↑ 2.0	51.1	31.9
	✓	✓			32.3 (+1.1)	52.9	32.5
	✓		✓		31.8 (+0.6)	53.1	31.8
	✓	✓	✓		32.9 (+1.7)	54.0	32.5
YOLOX-L				38.3	31.2	54.8	29.5
	✓				34.2 ↑ 3.0	54.6	34.9
	✓	✓			35.5 (+1.3)	56.4	35.3
	✓		✓		35.1 (+0.9)	55.5	35.6
	✓	✓	✓		36.1 (+1.9)	57.6	35.6

Experiments

- Visualization results

Thank you!