Object tracking and segmentation with deep learning approach for Digital Still Camera(DSC)
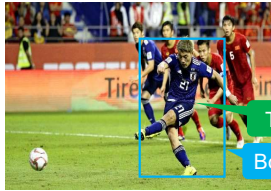
Applying SiamMask network with backbone MobileNet for object tracking and object segmentation. This approach is used to improve performance of current object tracking system on DSC.

Background

Current object tracking system

Detector creates Bounding Box (BB) surrounding the subject.
Focus on the subject inside the BB.

※The subject is selected by user

The subject
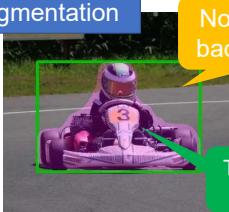Bounding Box

[Problem]

Background area
The subject (Motor)
Bounding Box

Created BB include the Background.
→Sometimes focus on the background.

Solution and application

Segmentation
Not include background
The subject (cart)

Solution: The subject area must be segmented from the background area for tracking

⇒ combination of DL-based object tracking and segmentation ⇒ **SiamMask**

What is SiamMask?
SiamMask is a variant of Siamese neural network which is used for both object tracking and segmentation tasks simultaneously.

Challenges during implementation
Heavy processing makes it impossible to run in real time on embedded systems.
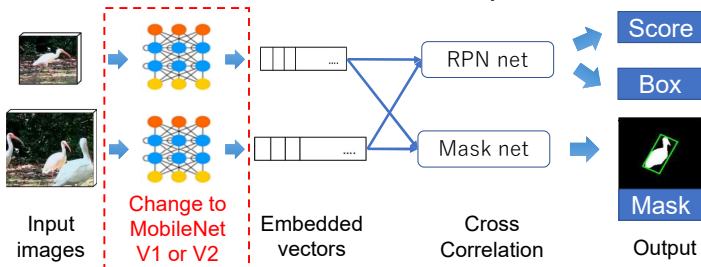→ **Need to lightweight the SiamMask model**

Strength	Lightweight SiamMask model by changing backbone network from ResNet50 to MobileNet V1, V2.
Goals	Possible to run in real time on embedded system.

Activities

1. Lightweight (Change and modify backbone network from ResNet50 to MobileNet V1 or V2)

Input images
Change to MobileNet V1 or V2
Embedded vectors
Cross Correlation
RPN net → Score, Box
Mask net → Mask
Output

2. Training and evaluating the modified model

Datasets
Training dataset: Youtube-VOS
Validation dataset: ImageNet-VID 2015
Test datasets: VOT2018

Motor cross   Racing   Singer

Evaluated results on VOT2018

Measurements	With ResNet50	With MobileNet V1	With MobileNet V2
Expected average overlap (EAO)	0.381	0.298	0.312
Accuracy (A)	0.609	0.563	0.591
Robustness (R)	0.281	0.335	0.309
Mean speed (FPS) ※tested on GPU	56 FPS	**121.71 FPS**	101.55 FPS

- **With ResNet50:** The model from original paper (*Fast Online Object Tracking and Segmentation: A Unifying Approach – CVPR 2019*)
- **Accuracy**: The accuracy metric measures the intersection over union (**IoU**) between the predicted bounding box and the ground truth bounding box for a particular sequence. *High accuracy means good result*.
- **Robustness**: The robustness metric measures the number of times when there is no overlap between the predicted bounding box and the ground truth box. *Low robustness means good*.
- **Expected average overlap (EAO)**: is an estimator of the average overlap a tracker is expected to attain on a large collection of short-term sequences with the same visual properties as the given dataset. *High EAO means good result*.

Future plan

1. Continuing modify and training the model to improve performance.
2. Evaluating segmentation performance of SiamMask model with backbone MobileNet_V1 and MobileNet_V2.
3. Deploying the model on simulation board (edge device)
4. Applying other lightweight techniques like quantization, pruning to the model.
5. Compressing the model and improving performance (decrease occlusion rate, distinguishing from similar objects, etc.)