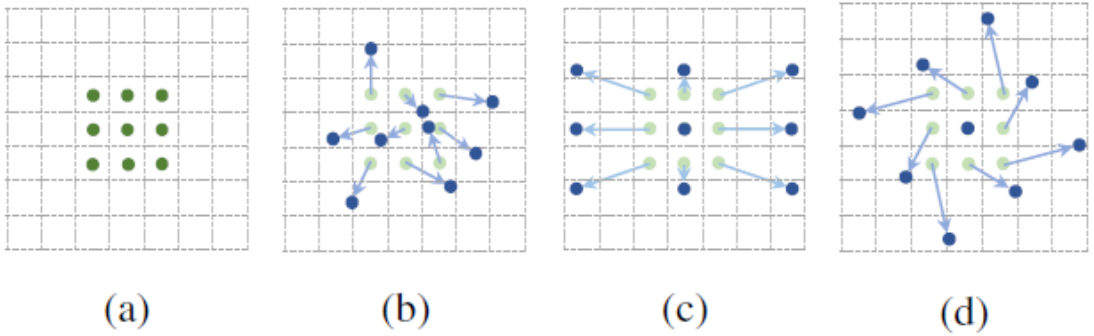
Deformable convolution and adaptive key points

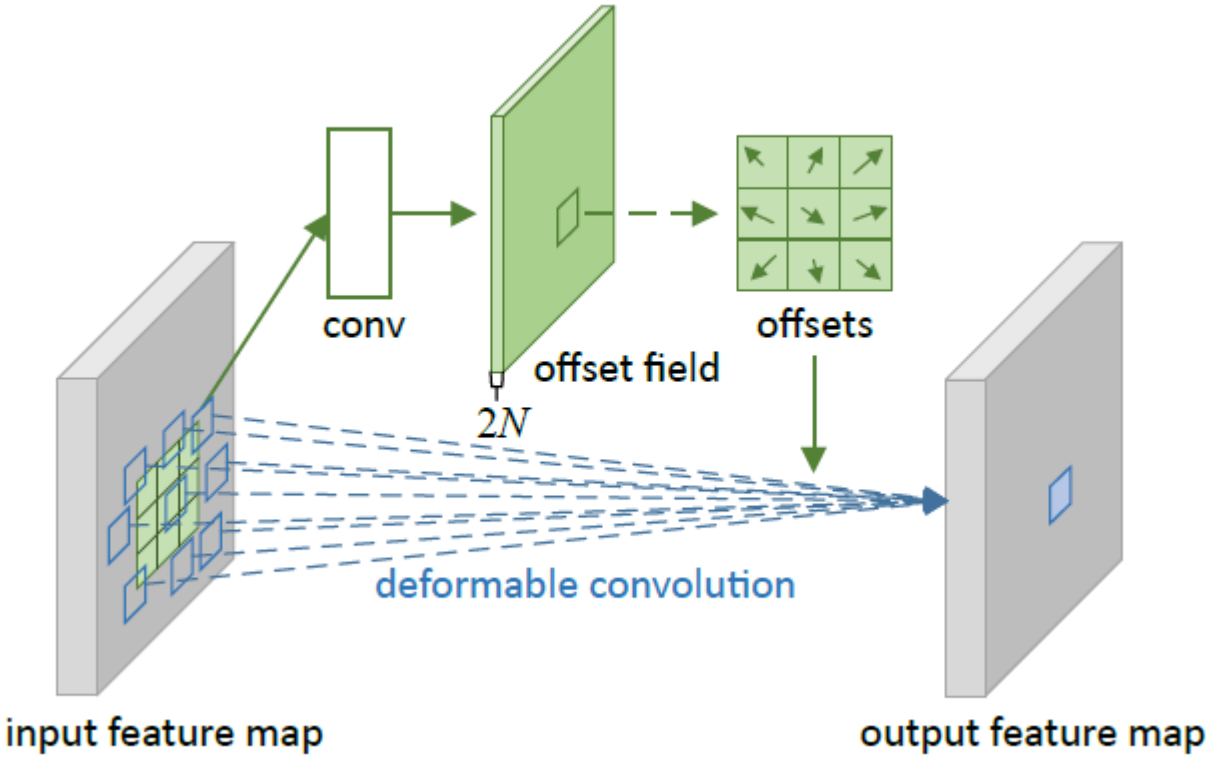
Outline:

- Deformable convolution network(DCN) and DCNv2
- RepPoint for object detection on image
- FGFA and STSN for video application

Deformable convolution



Regular convolution vs deformable convolution



Realization of deformable convolution

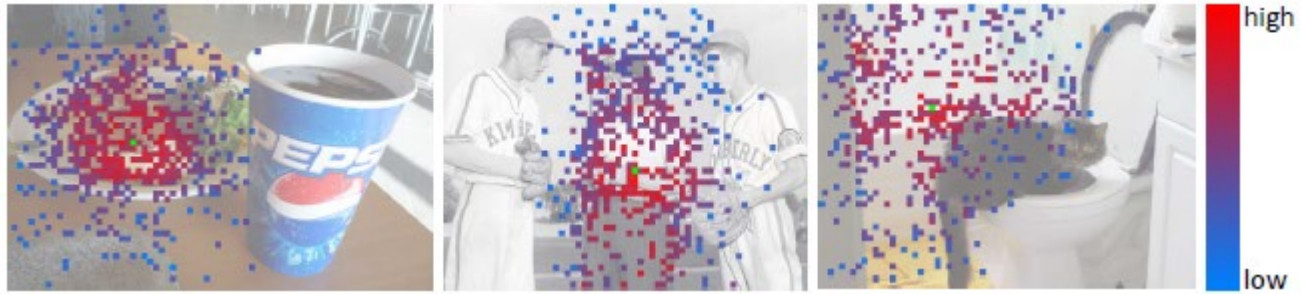
Reference:
1. Deformable Convolutional Networks
2. Deformable ConvNets v2: More Deformable, Better Results

Result of deformable convolution

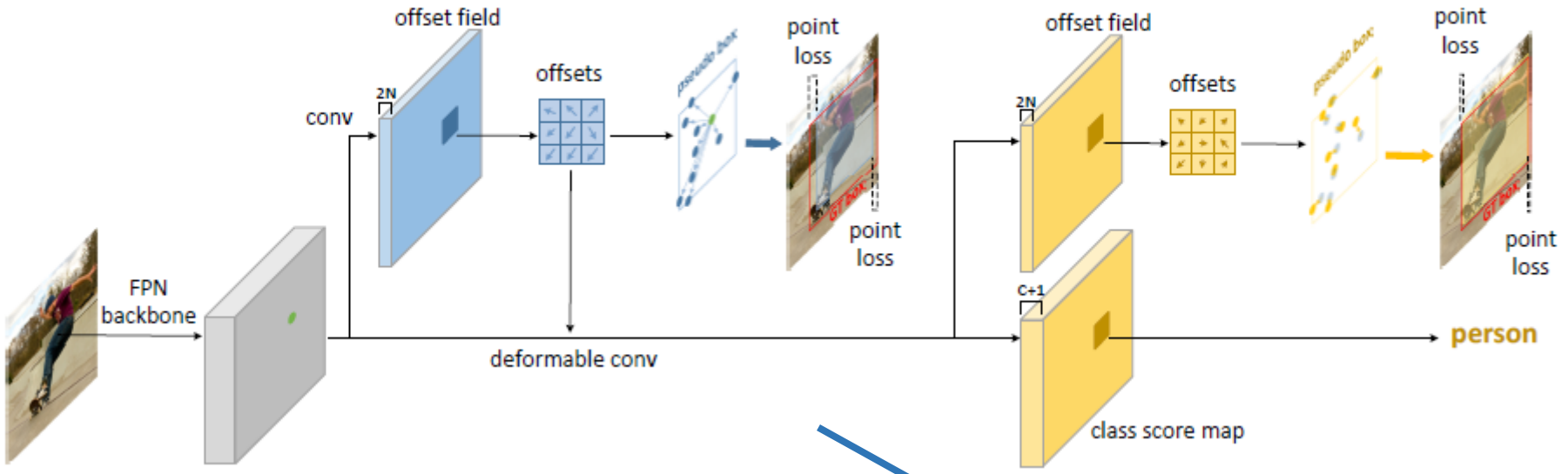
DCN v1 result

background forground forground background forground forground

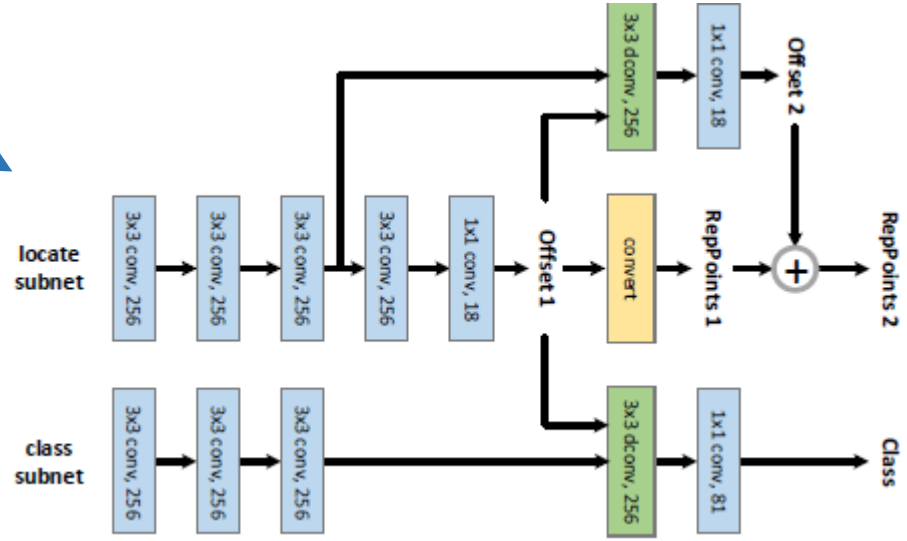
DCN v2 result



RepPoints



Realization of RepPoints

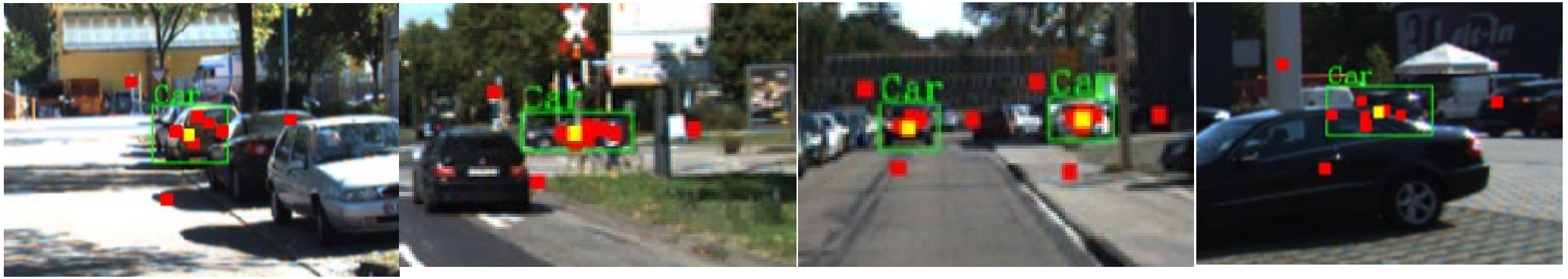


RepPoints Head

Reference: RepPoints: Point Set Representation for Object Detection

Result of RepPoints on COCO

RepPoints Result on KITTI

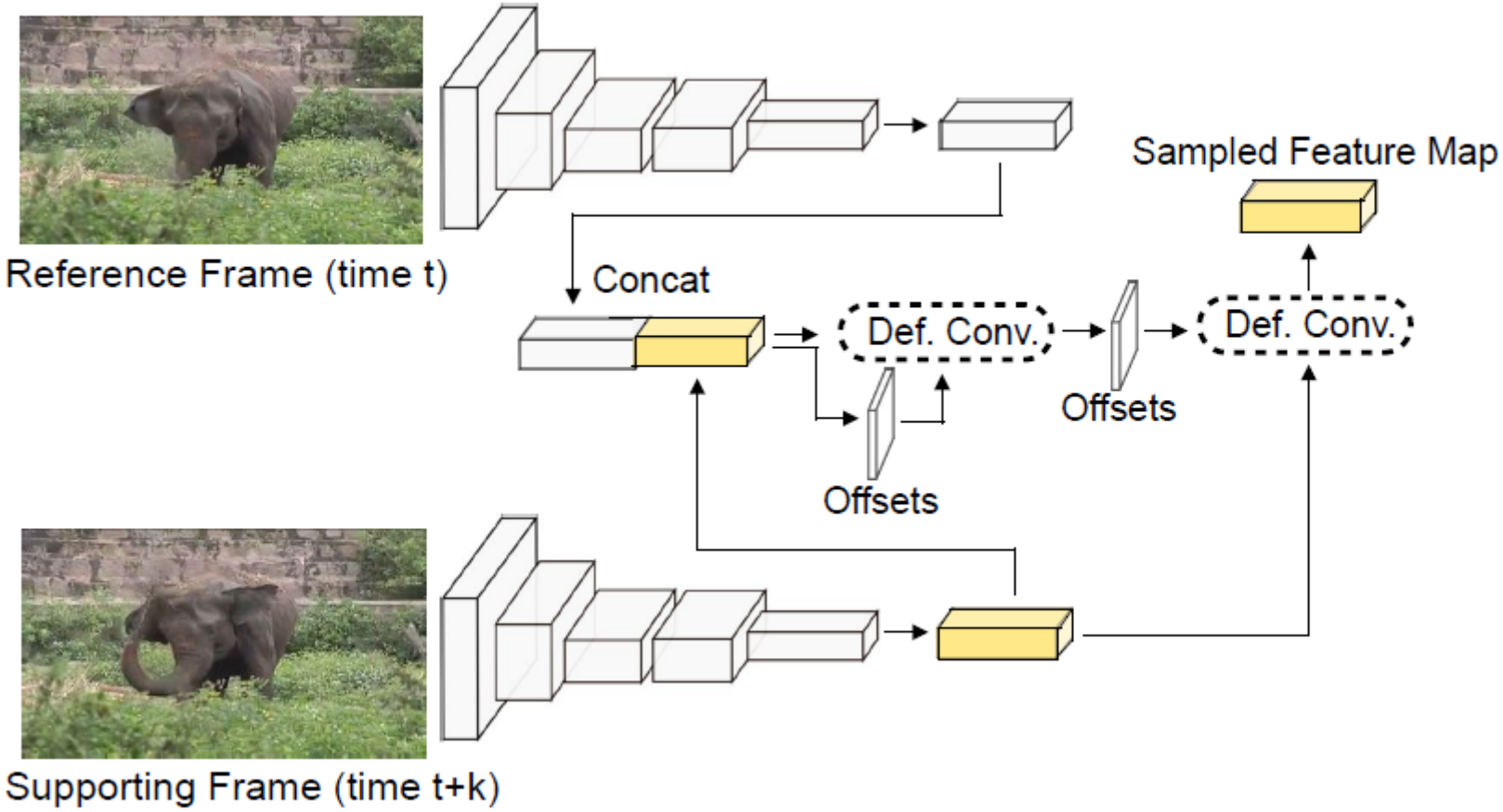


Car

Car

Person

Spatiotemporal Sampling Networks(STSN)



Reference: Object Detection in Video with Spatiotemporal Sampling Networks

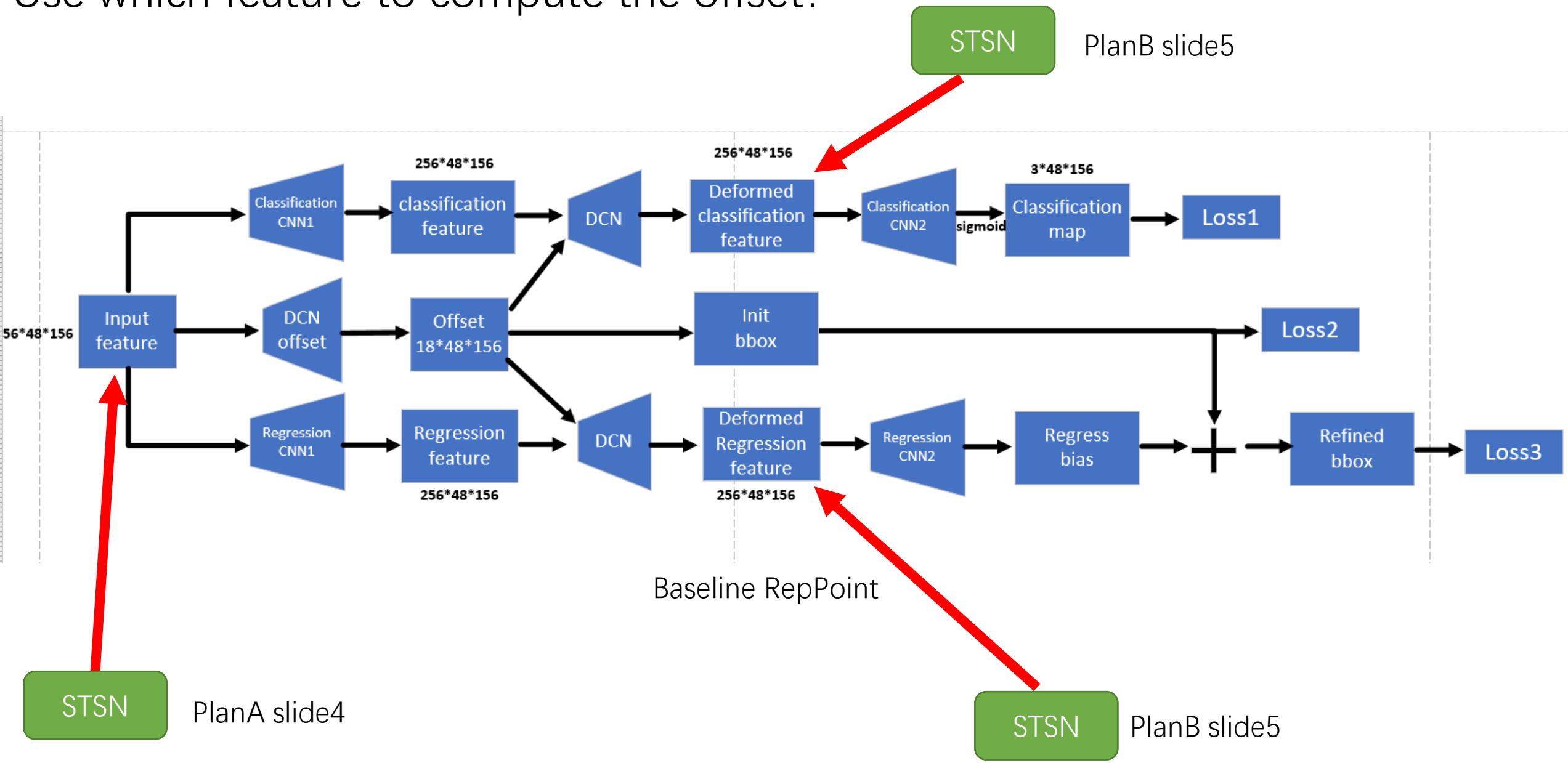
STSN result on VID

Reference Frame (t) Supporting Frame (t+9)

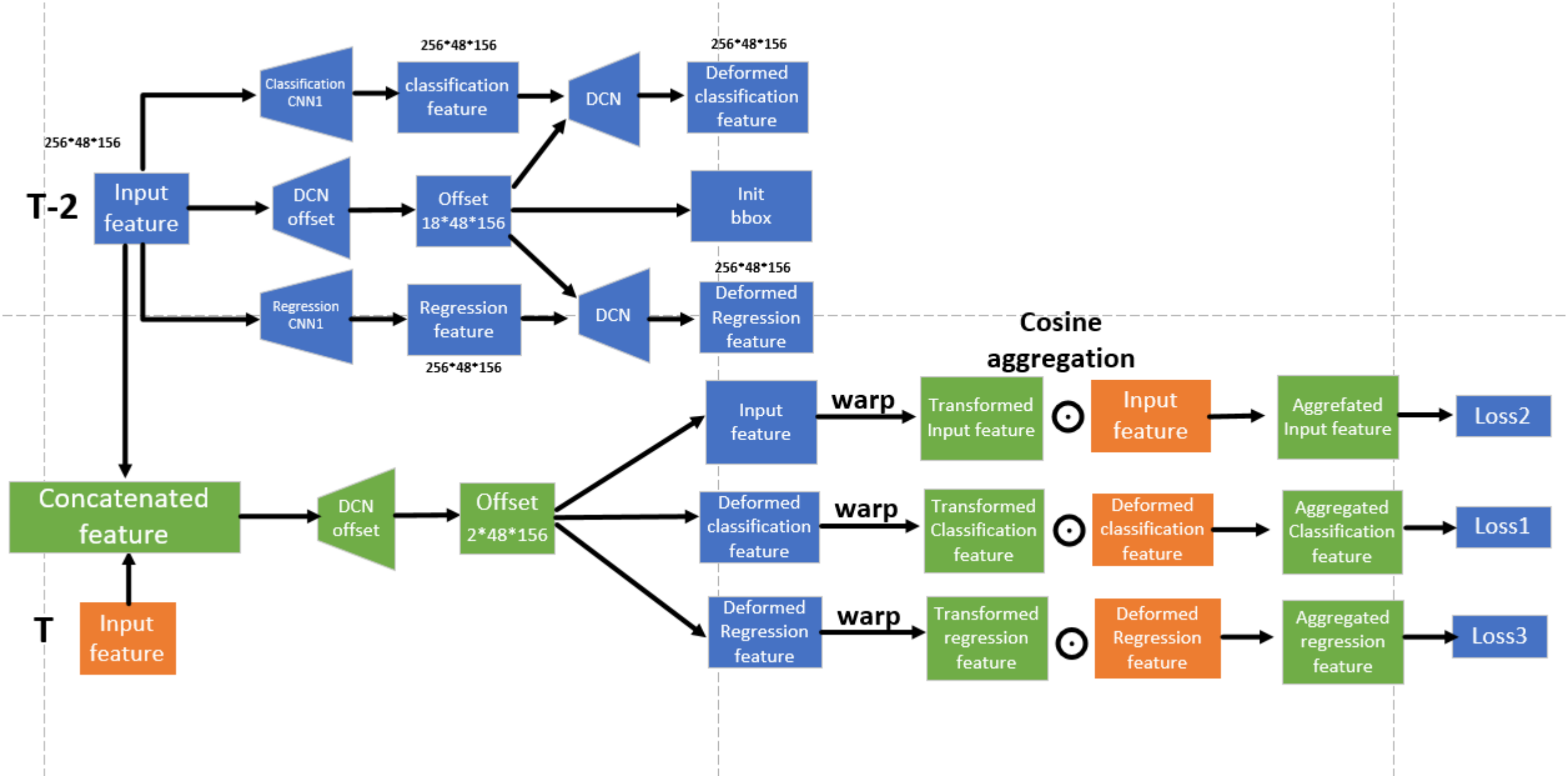
Reference Frame (t) Supporting Frame (t+9)

Supporting Frame (t-9) Supporting Frame (t-4) Reference Frame (t) Supporting Frame (t+4) Supporting Frame (t+9)

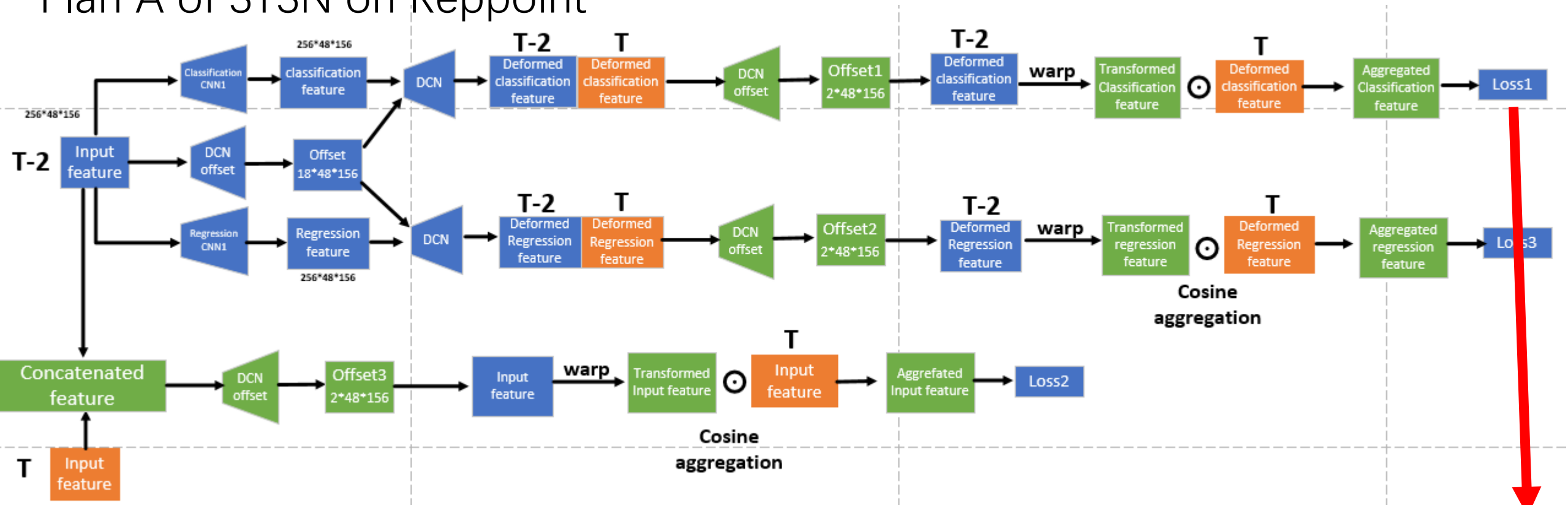
Use which feature to compute the offset?



Plan A of STSN on Reppoint, result on Slides 2



Plan A of STSN on Reppoint



Old Result Need update

evaluating result of refer

class	gts	dets	recall	precision	ap
Vehicle	10747	14870	0.794	0.574	0.743
Pedestrian	9429	11400	0.678	0.560	0.620
Cyclist	509	1441	0.497	0.176	0.187
mAP					0.517

baseline

evaluating result of agg epoch29 from the pretrained epoch9 optimize whole

class	gts	dets	recall	precision	ap
Vehicle	10747	26993	0.876	0.349	0.774
Pedestrian	9429	15771	0.675	0.404	0.598
Cyclist	509	4518	0.701	0.079	0.297
mAP					0.556

Only warp the classification feature

New KITTI split

Before

	Vehicle	Person	Cyclist
Train	20680	2740	#
test	9921	9406	#

```
evaluating result of  refer
+------------+-------+-------+--------+-----------+-------+
| class      | gts   | dets  | recall | precision | ap    |
+------------+-------+-------+--------+-----------+-------+
| Vehicle    | 10747 | 14870 | 0.794  | 0.574     | 0.743 |
| Pedestrian | 9429  | 11400 | 0.678  | 0.560     | 0.620 |
| Cyclist    | 509   | 1441  | 0.497  | 0.176     | 0.187 |
+------------+-------+-------+--------+-----------+-------+
| mAP        |       |       |        |           | 0.517 |
+------------+-------+-------+--------+-----------+-------+
```

Baseline

Now

	Vehicle	Person	Cyclist
Train	22543	8702	1354
test	8058	3444	584

```
evaluating result of  baseline
+------------+------+-------+--------+-----------+-------+
| class      | gts  | dets  | recall | precision | ap    |
+------------+------+-------+--------+-----------+-------+
| Vehicle    | 8058 | 10070 | 0.954  | 0.764     | 0.947 |
| Pedestrian | 3444 | 5463  | 0.825  | 0.520     | 0.789 |
| Cyclist    | 584  | 1089  | 0.777  | 0.417     | 0.739 |
+------------+------+-------+--------+-----------+-------+
| mAP        |      |       |        |           | 0.825 |
+------------+------+-------+--------+-----------+-------+
```

Baseline

train frames: 5976
test frames: 2032
all frames: 8008

This performance might be too high to improve. Lixin just talked to me today that he also need the BDD data, so I also re-split the BDD tracking data.

Training and test on the new split KITTI dataset

evaluating result of baseline

class	gts	dets	recall	precision	ap
Vehicle	8058	10070	0.954	0.764	0.947
Pedestrian	3444	5463	0.825	0.520	0.789
Cyclist	584	1089	0.777	0.417	0.739
mAP					0.825

Baseline
Figure is in
Slides 3

evaluating result of agg epoch19

class	gts	dets	recall	precision	ap
Vehicle	8058	11140	0.958	0.693	0.950
Pedestrian	3444	6706	0.826	0.424	0.781
Cyclist	584	1276	0.769	0.352	0.725
mAP					0.819

evaluating result of agg epoch29

class	gts	dets	recall	precision	ap
Vehicle	8058	10546	0.949	0.725	0.942
Pedestrian	3444	3998	0.713	0.614	0.654
Cyclist	584	927	0.740	0.466	0.707
mAP					0.767

evaluating result of agg epoch 39

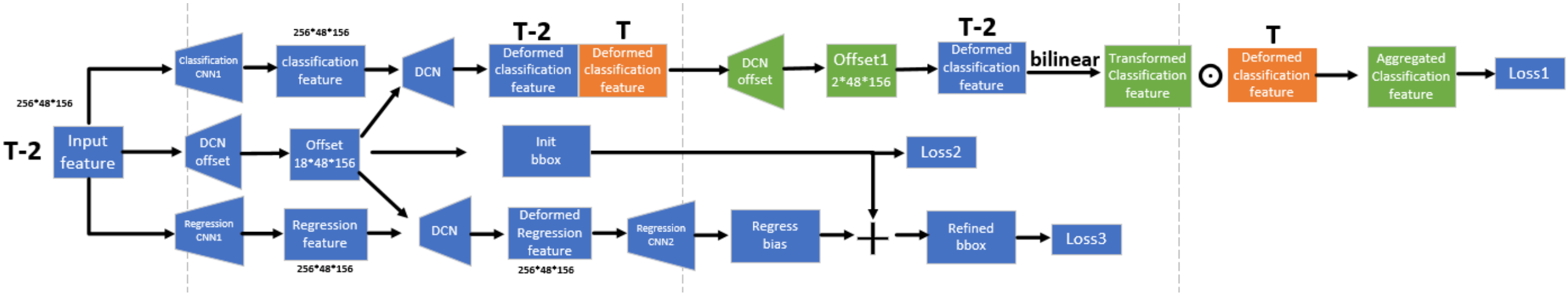
class	gts	dets	recall	precision	ap
Vehicle	8058	10799	0.962	0.717	0.953
Pedestrian	3444	6033	0.817	0.467	0.776
Cyclist	584	1243	0.765	0.360	0.731
mAP					0.820

evaluating result of agg epoch38

class	gts	dets	recall	precision	ap
Vehicle	8058	10188	0.957	0.757	0.949
Pedestrian	3444	5574	0.827	0.511	0.785
Cyclist	584	1056	0.779	0.431	0.742
mAP					0.825

STSN
Figure is in
Slides 4

One offset warp the classification feature



Comparison between different feature to compute the offset

evaluating result of baseline

class	gts	dets	recall	precision	ap
Vehicle	8058	10070	0.954	0.764	0.947
Pedestrian	3444	5463	0.825	0.520	0.789
Cyclist	584	1089	0.777	0.417	0.739
mAP					0.825

baseline

evaluation of agg classification baseline epoch19

class	gts	dets	recall	precision	ap
Vehicle	8058	11994	0.965	0.648	0.956
Pedestrian	3444	6000	0.823	0.472	0.786
Cyclist	584	1507	0.791	0.307	0.737
mAP					0.827

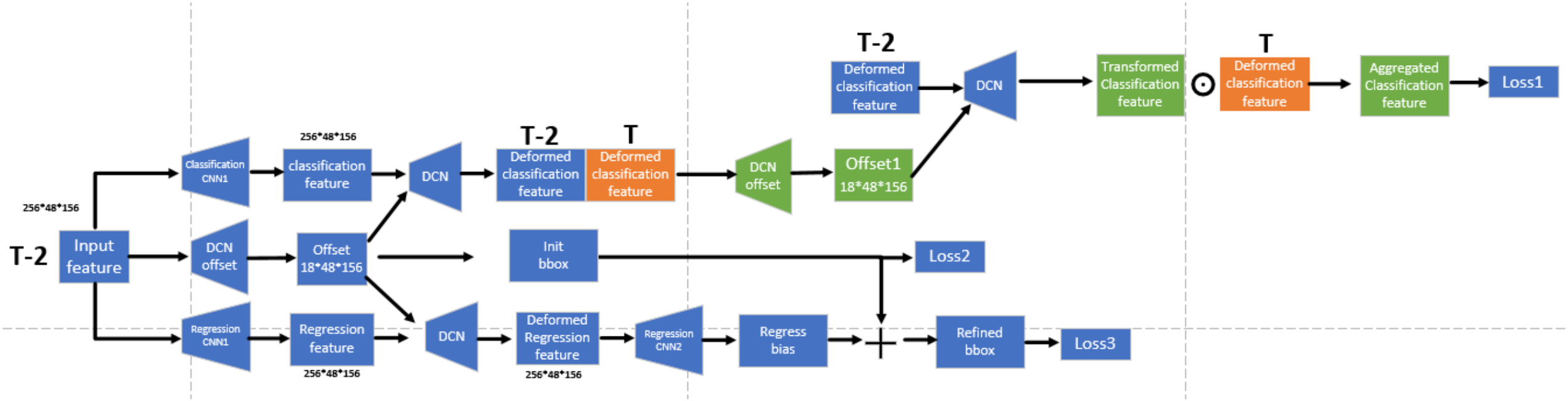
One offset warp the classification feature

evaluating result of agg epoch19

class	gts	dets	recall	precision	ap
Vehicle	8058	11140	0.958	0.693	0.950
Pedestrian	3444	6706	0.826	0.424	0.781
Cyclist	584	1276	0.769	0.352	0.725
mAP					0.819

Use backbone feature to compute the offset and warp the classification and regression feature.

Analyze of the fail case of stsn and compare with reppoint



The stsn version I analyze in this report

With learnable dcn to warp the support feature
Support frames are N-2 and N+2
Aggregation between N, N-2 and N+2
Fix the reppoint weight

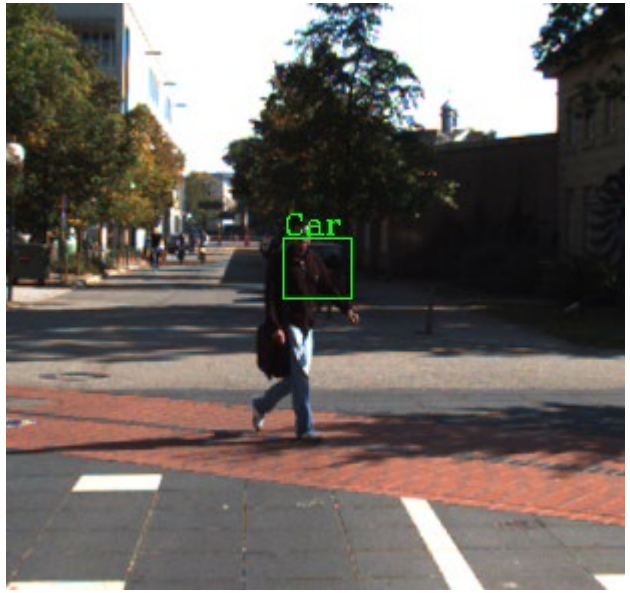
Car--not-detected bbox and offset

Green bbox is the ground bbox not-detected.
Yellow point is the center of the bbox
Red points are the offset of the yellow point.
The result is checked frame by frame, if the bbox only shows on frame 3 and frame 5, that means frame 4 detects all objects.
Vehicle class consists of car, van and trunk classes of kitti
Support1 is N-2 frame, support2 is N+2 frame
Reppoint+stsn is two dcn layers, so we get 81 offsets to show

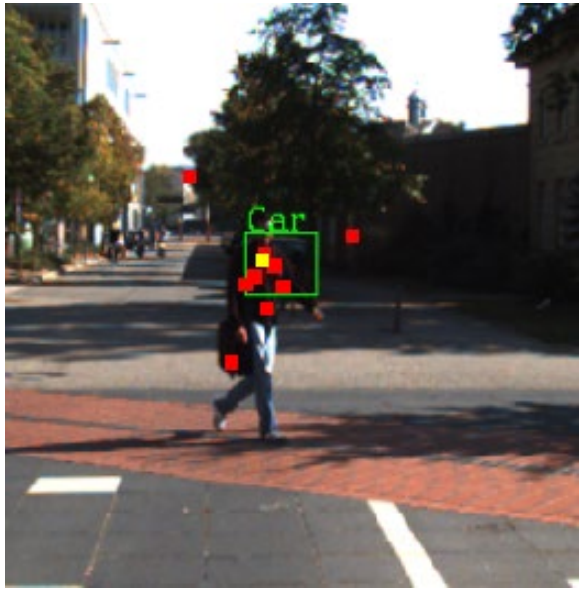
	Vehicle	Person	Cyclist
Train	22543	8702	1354
test	8058	3444	584

- First show failure case of reppoint but can be detected by stsn
- Then the failure case of stsn

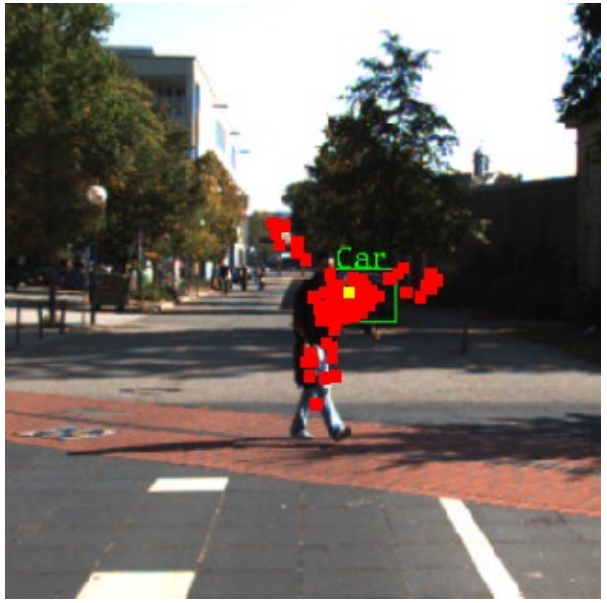
Stsn gets not-detected bbox of reppoint



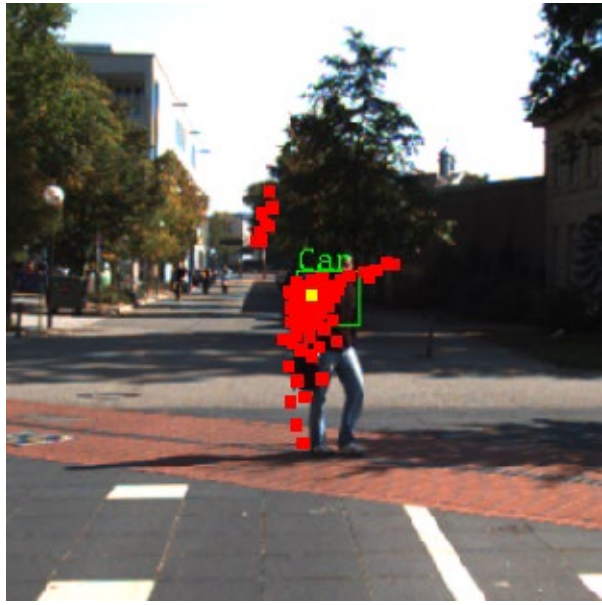
Reppoint not-detected



Reppoint offset



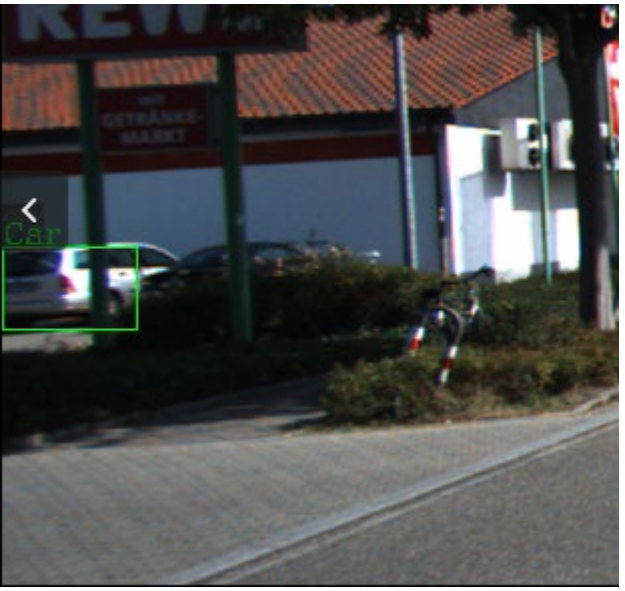
support1 offset

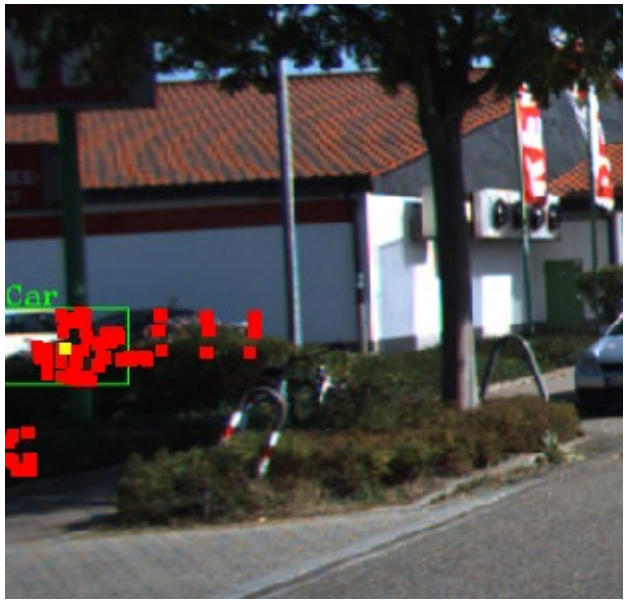


support2 offset

The car is highly occluded by the preson, most reppoints are on the person
Stsn adds more points on the car.

Stsn gets not-detected bbox of reppoint





Reppoint not-detected

Reppoint offset

support1 offset

support2 offset

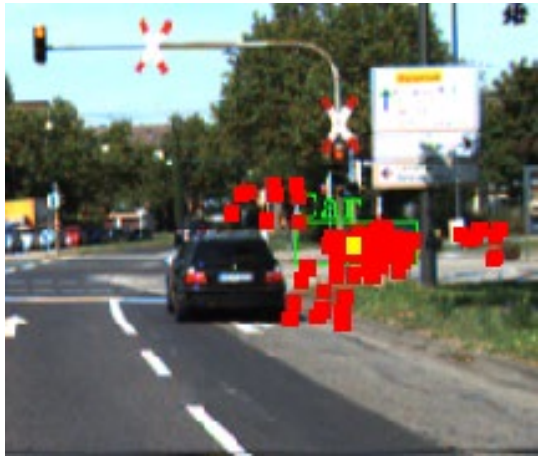
Stsn adds a denser sampling on the car from N-2 frame
It seems the reppoint allow offsets lie on background or other object, these offset may have very low weight, which are shown as grid on the image.

Stsn gets not-detected bbox of reppoint

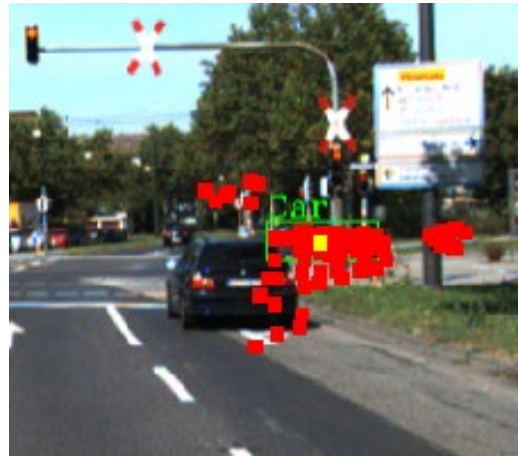
Frame 74

Reppoint not-detected

Reppoint offset

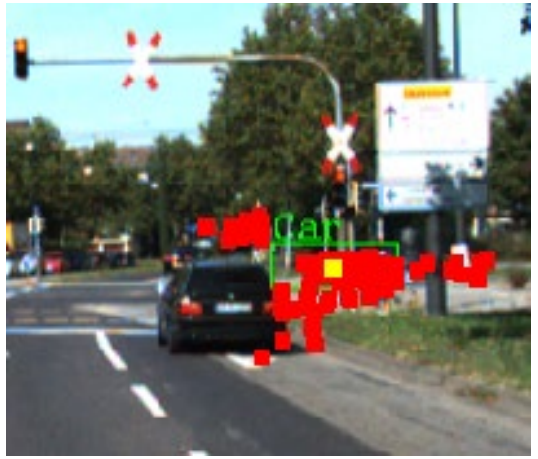


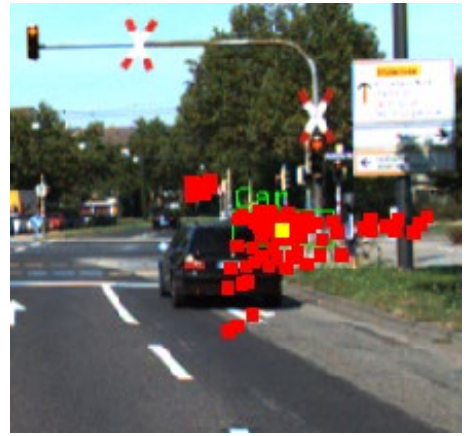
support1 offset



support2 offset

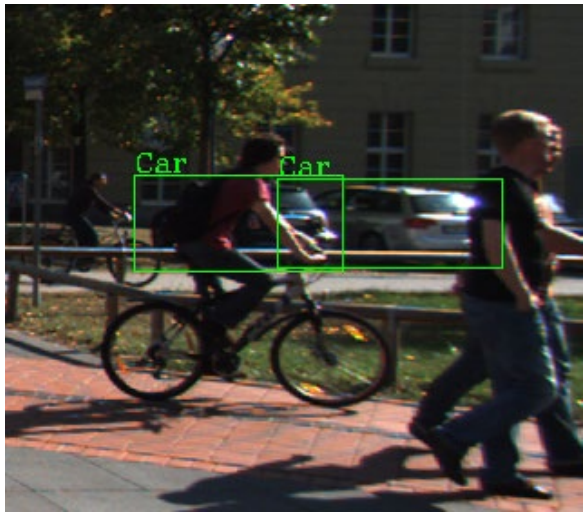
Frame 75



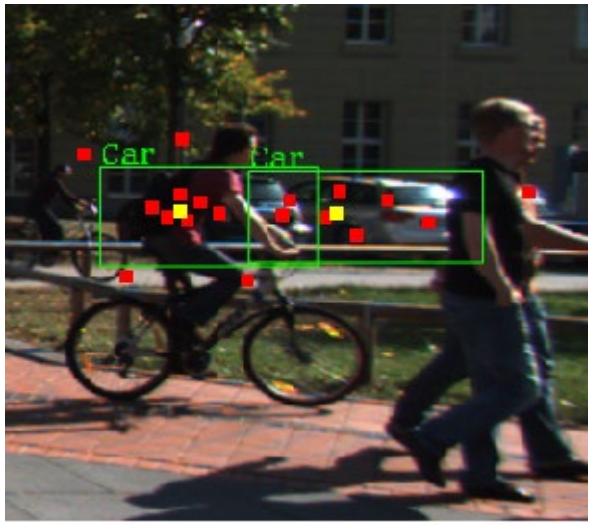


On small objects, the offset visualization is hard to see
Im working on the one offset warp model for the visualization.

Stsn get not-detected bbox of reppoint

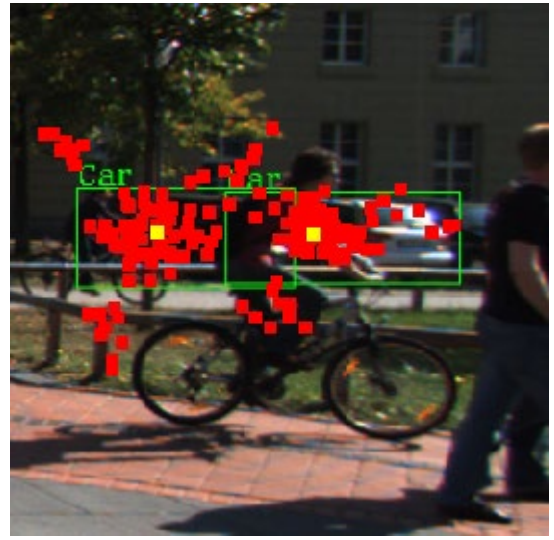


Reppoint not-detected



Reppoint offset

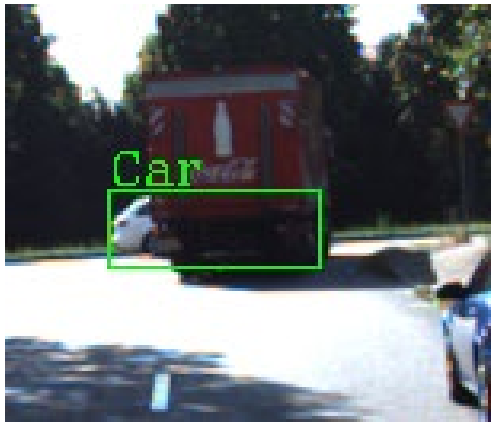
support1 offset



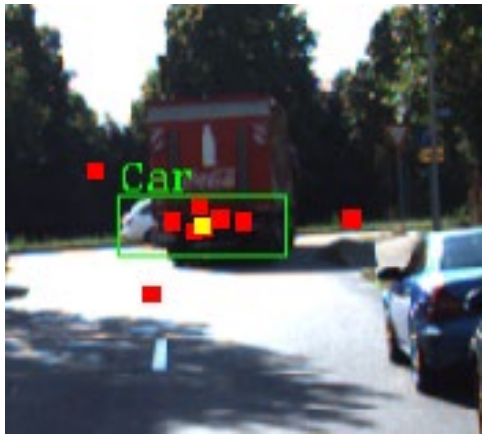
support2 offset

For the left car, on N+2, the stsn brings into the right signal

Stsn get not-detected bbox of reppoint

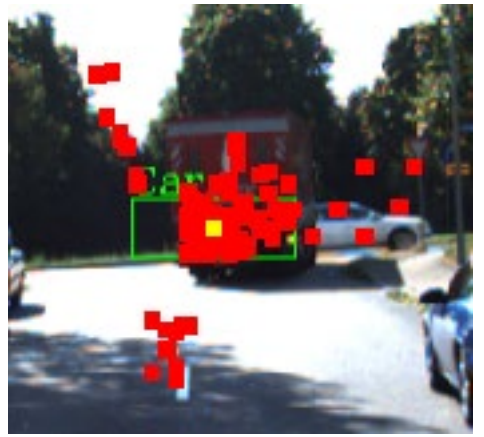


Reppoint not-detected



Reppoint offset

support1 offset

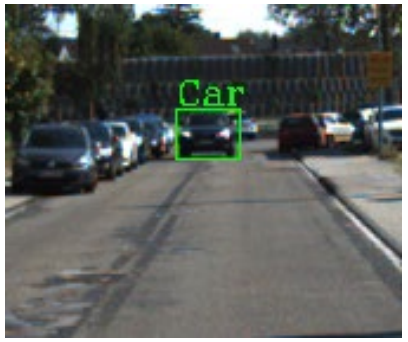


support2 offset

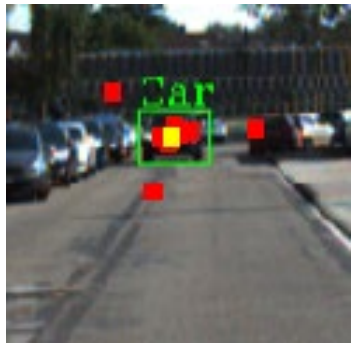
- It is hard to say why it works. Some points from N+2 frame is right, but I don't think they get a high weight.
- Maybe the offset I visualized has a bias, the yellow point should be on the left rather than on the center.
- On the N+2 frame, we can see the offset is trying to find the signal from while color.

Stsn get not-detected bbox of reppoint

Frame 26



Reppoint not-detected

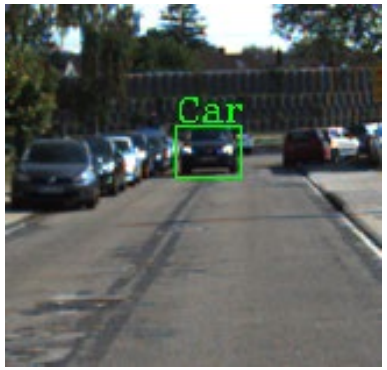


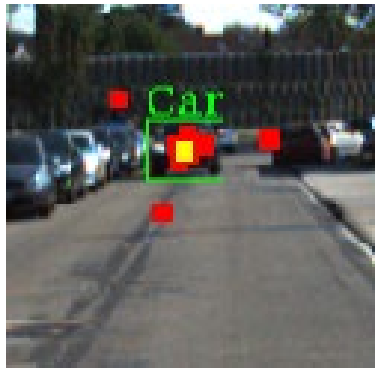
Reppoint offset

support1 offset

support2 offset

Frame 28

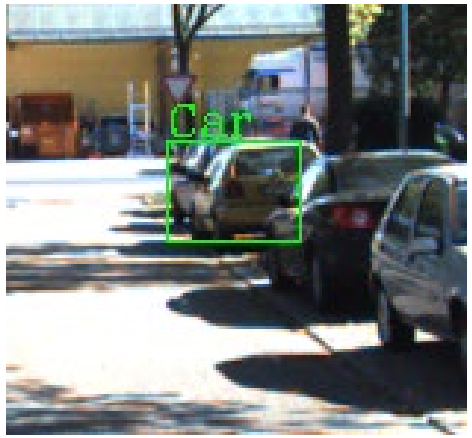




Signal is obviously enhanced by the stsn, the ratio of point on the car is higher than only using one frame.

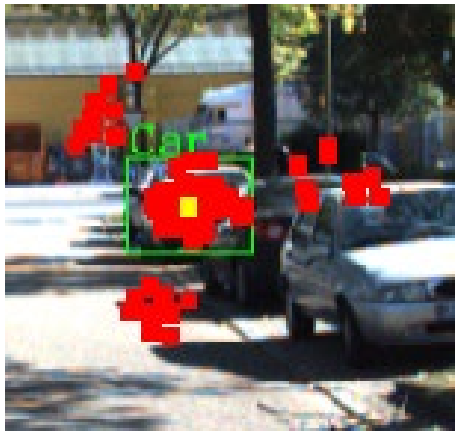
fail case of stsn

Frame 119



Stsn not-detected

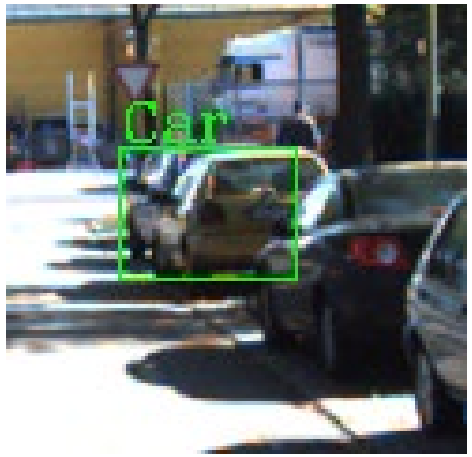
Reppoint offset

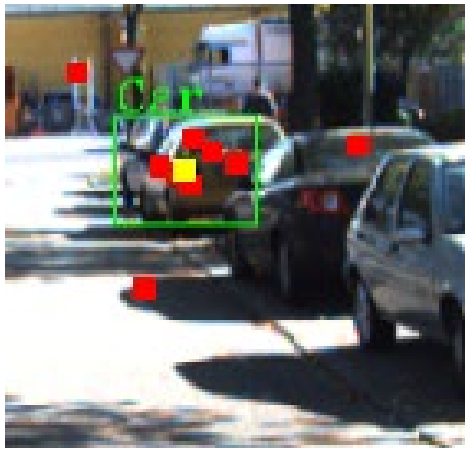


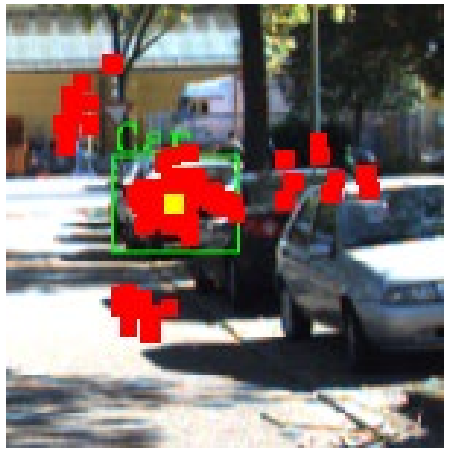
support1 offset

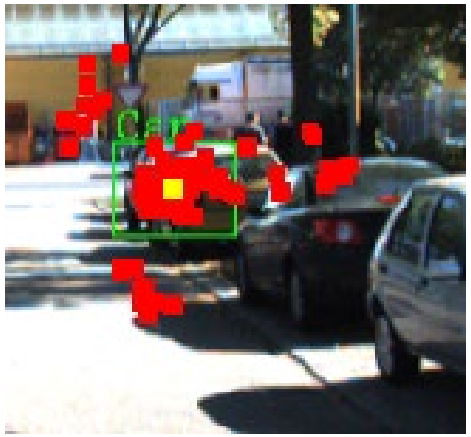
support2 offset

Frame 120



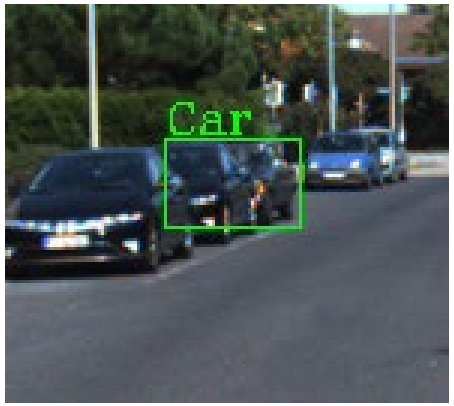




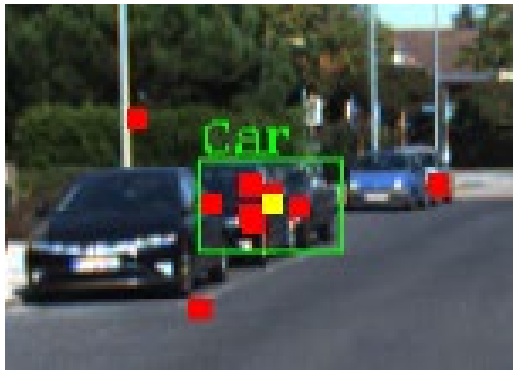


The offset seems right.
This high occlusion may not be the detection problem but the nms kill the occluded bbox.

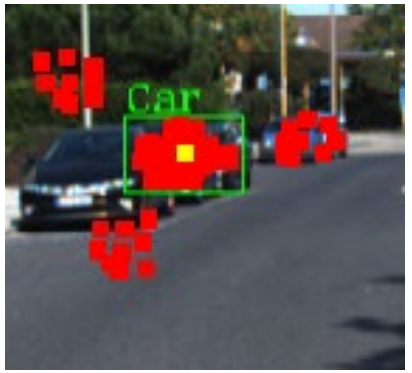
fail case of stsn



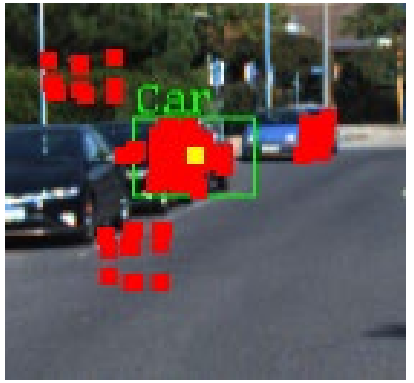
stsn not-detected



Reppoint offset

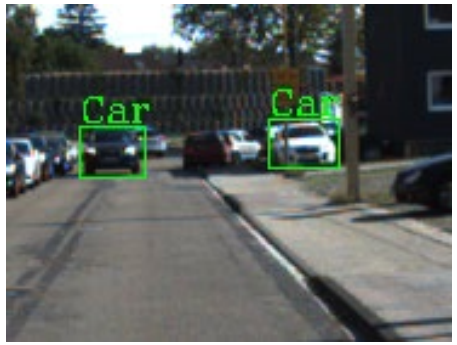


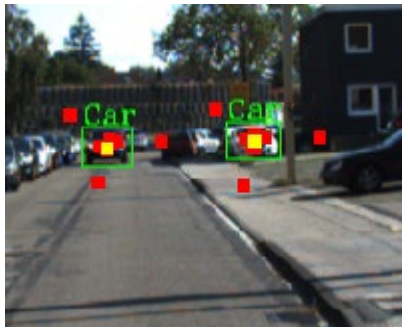
support1 offset

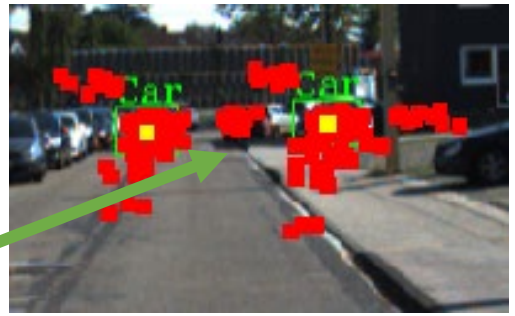


support2 offset

Maybe two cars fused as a object not a car?





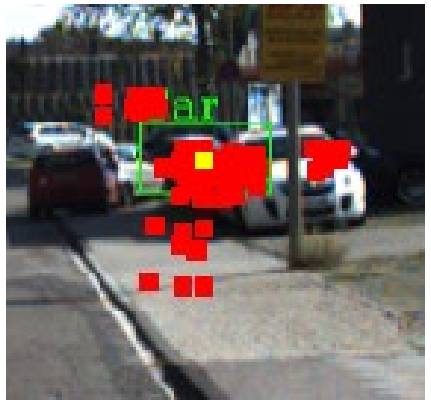


Again, for small object, too much signal is aggregated.
But we can see the offset trying to find the signal on cars.

fail case of stsn

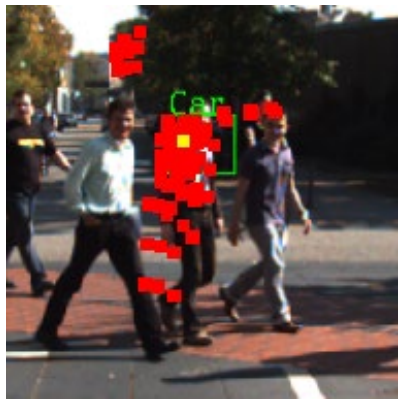
stsn not-detected

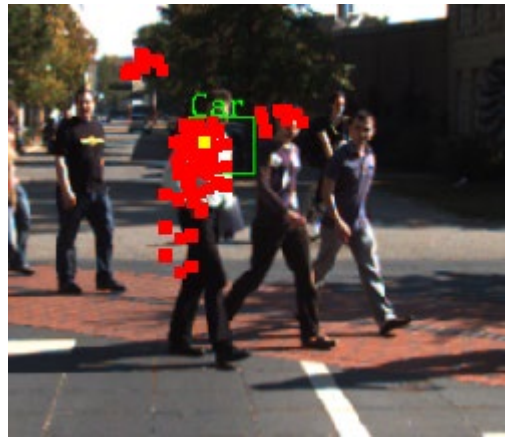
Reppoint offset



support1 offset

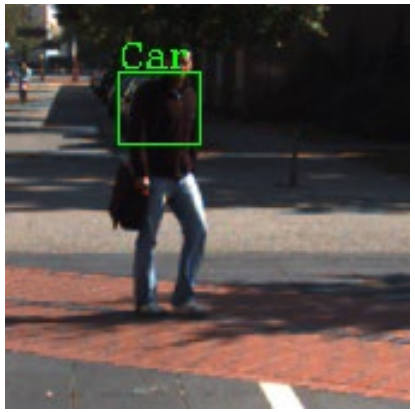
support2 offset



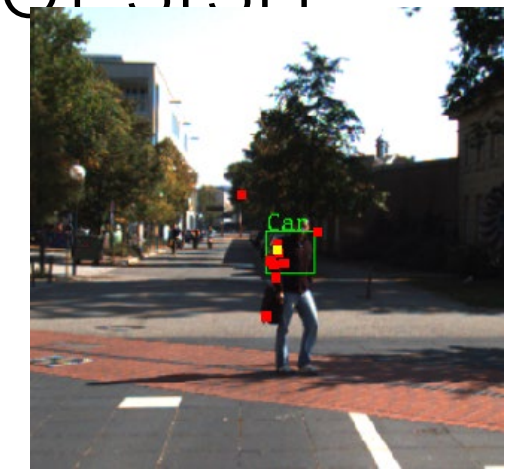


- For car occluded by car, the two cars should have similar color, this is also we can see from slides7.
- For car occluded by person, it will fail if the car is similar to the background and the person is good visible, offsets will lie on strong signal on the person.

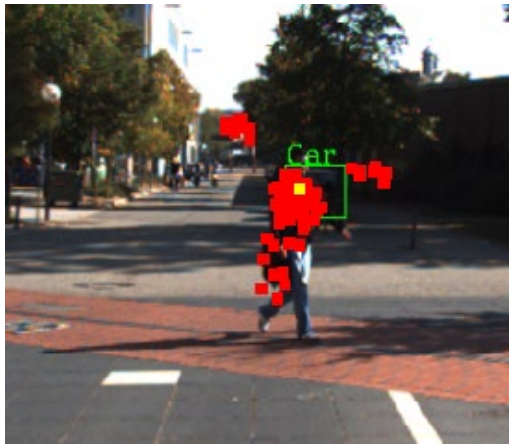
fail case of stsn



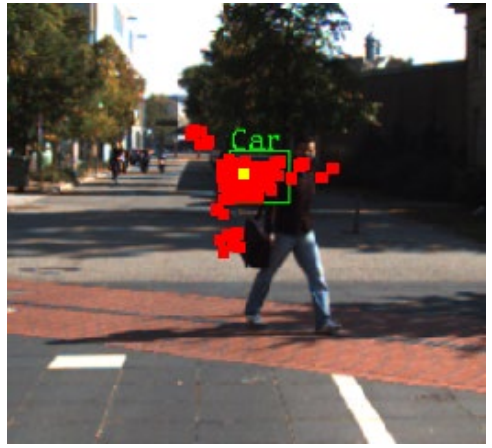
stsn not-detected



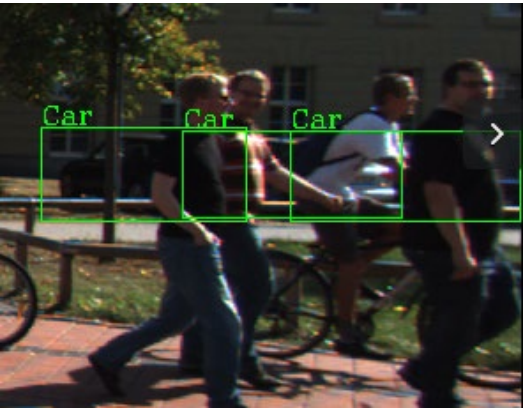
Reppoint offset

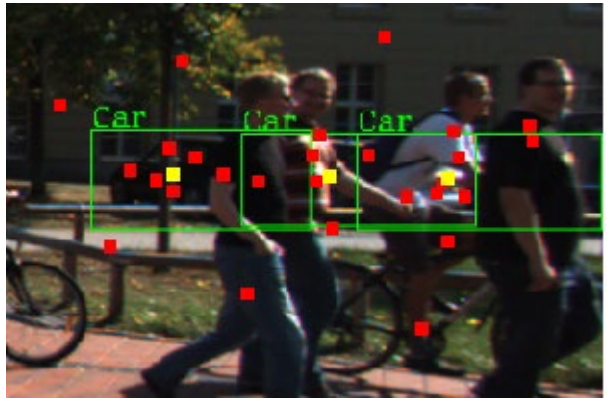


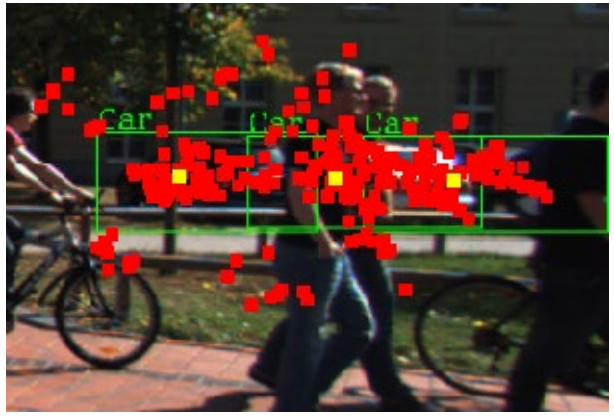
support1 offset



support2 offset







For the top case, the N+2 signal is right, then the problem could come from the aggregation, where we use N frame as anchor.
For the bottom case, signals are messed up, both the car and person signal is fused together, the reason might be the resnet and fpn have messed up the signal.

Person--not-detected bbox and offset

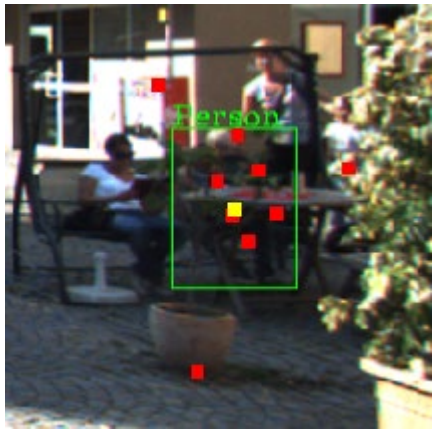
Green bbox is the ground bbox not-detected.
Yellow point is the center of the bbox
Red points are the offset of the yellow point.
The result is checked frame by frame, if the bbox only shows on frame 3 and frame 5, that means frame 4 detects all objects.
Person class consist of pedestrian and sitting person on kitti
Support1 is N-2 frame, support2 is N+2 frame
Reppoint+stsn is two dcn layers, so we get 81 offsets to show

	Vehicle	Person	Cyclist
Train	22543	8702	1354
test	8058	3444	584

- First show failure case of reppoint but can be detected by stsn
- Then the failure case of stsn

Stsn get not-detected bbox of reppoint

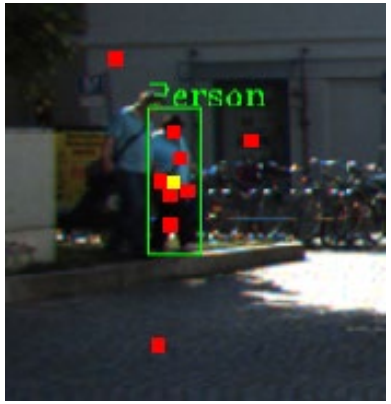
Reppoint not-detected

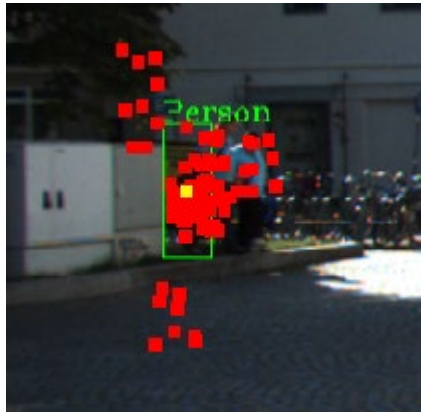


Reppoint offset

support1 offset

support2 offset

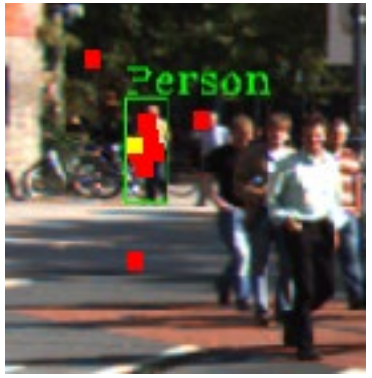




For the top case, N+2 frame brings in the good signal and the N frame is basically right on the person.
For the bottom case, the two persons are similar, the N-2 frame fuse the signal to make it more like a person.

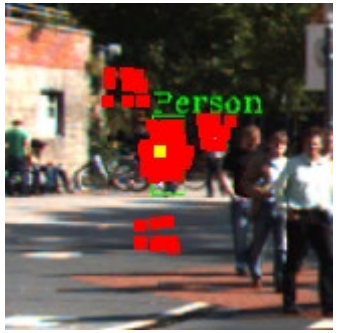
Stsn get not-detected bbox of reppoint

Reppoint not-detected

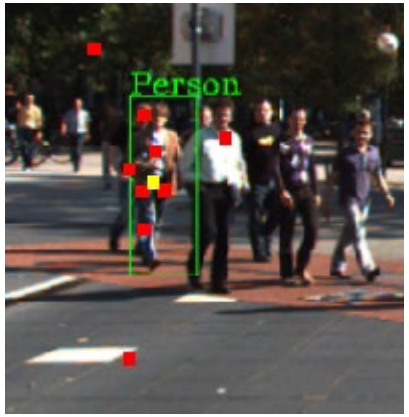


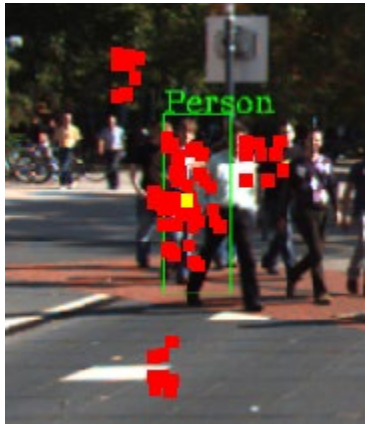
Reppoint offset

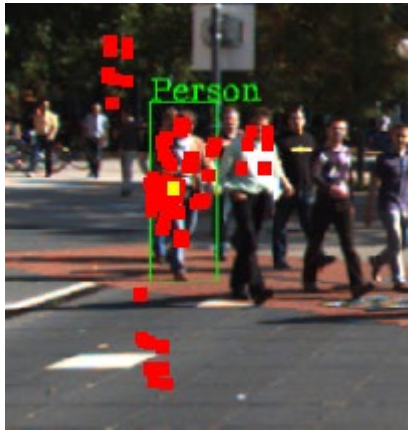
support1 offset



support2 offset





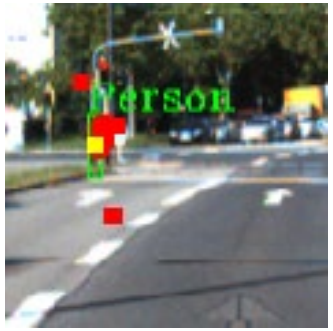


- For the top case, the stsn moves the person's head to the small person and enhance the body signl at the same time.
- For the bottom case, two persons all wear the bright shirt, N+2 frame enhances the signal and the head signal might be extremely important for the classification of person, so finding where the head is the both stsn and reppoint trying to do.

fail case of stsn

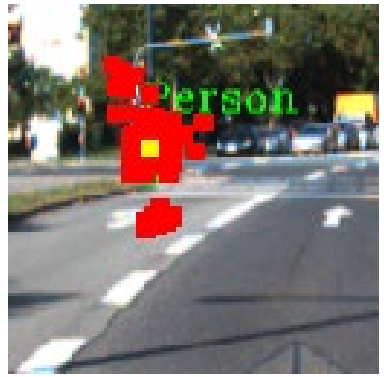
Frame 64

stsn not-detected



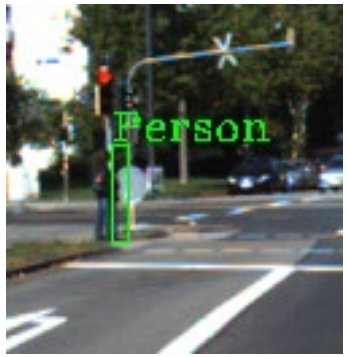
Reppoint offset

support1 offset

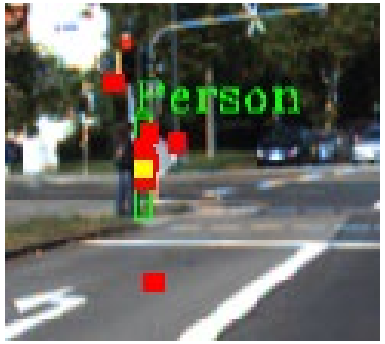


support2 offset

Frame 82

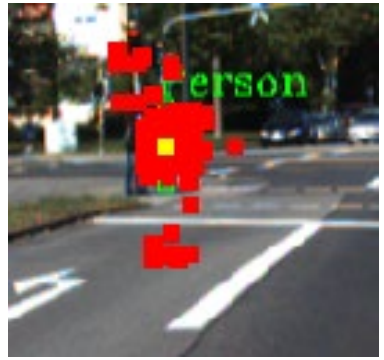


stsn not-detected



Reppoint offset

support1 offset

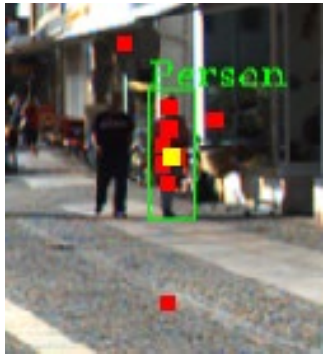


support2 offset

It is interesting the offset looks like a shape of person, that's we can say the feature to infer the offset has recognize the person, but the signal and scale is not suitable for classifier.
As the offset also is a signal, we may use it as a feature too, rather than just treat it a motion signal.

fail case of stsn

stsn not-detected



Reppoint offset

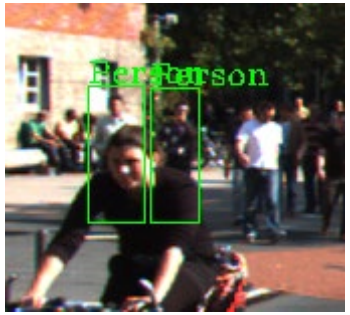
support1 offset

support2 offset

t

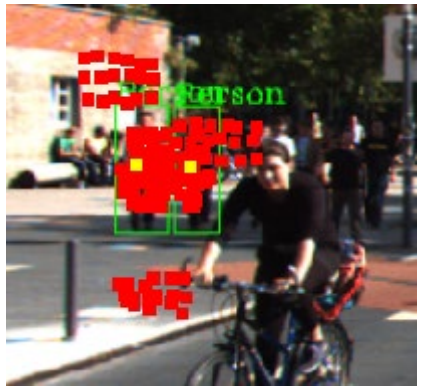
The top case is right might because of head is wrong and we can see the N+2 frame fuses the signal of two person but the similarity of the two person is high, this might confuse the aggregation.
For the bottom case, the two person's signal is messed up. The other reason is that the sitting person only gets 500 samples on the benchmark, and the pose is obviously different from the pedestrian.

fail case of stsn

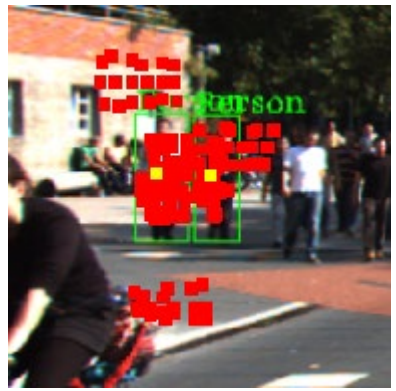


stsn not-detected

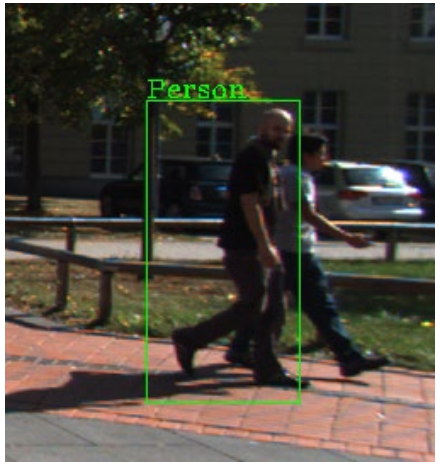
Reppoint offset



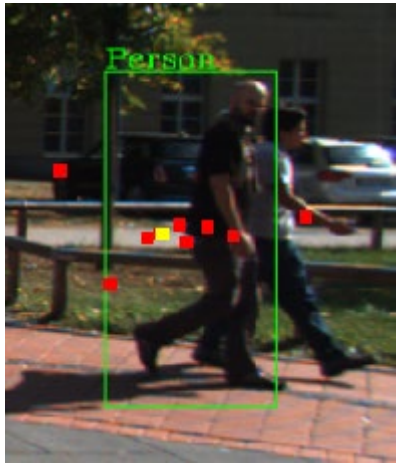
support1 offset



support2 offset

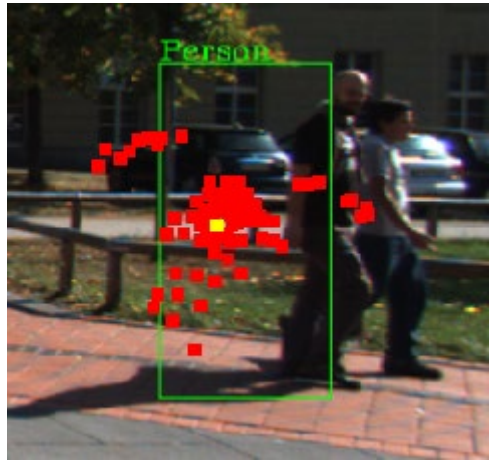


stsn not-detected



Reppoint offset

support1 offset

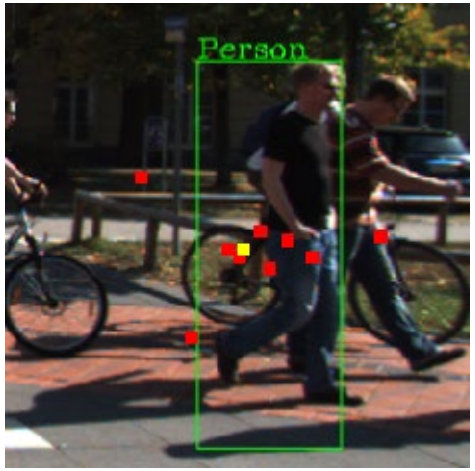


support2 offset

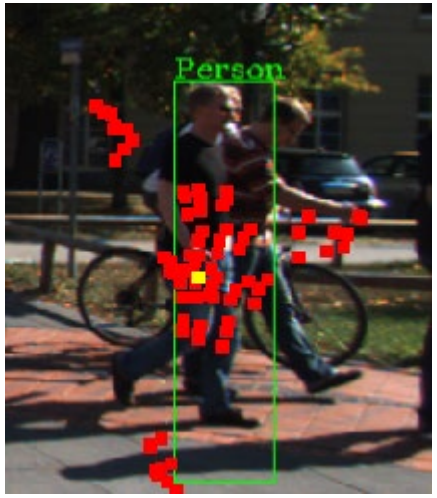
For the top case, N+2 and N-2 bring into right signal, the problem might be the N frame is not right, so the aggregation is failed because of the wrong anchor.
For the bottom case, I only visualize the offset of stage2, but we can see the offset on frame N is not right, but on frame N-2 is right, this might require the fusion of signal between different scales.

fail case of stsn

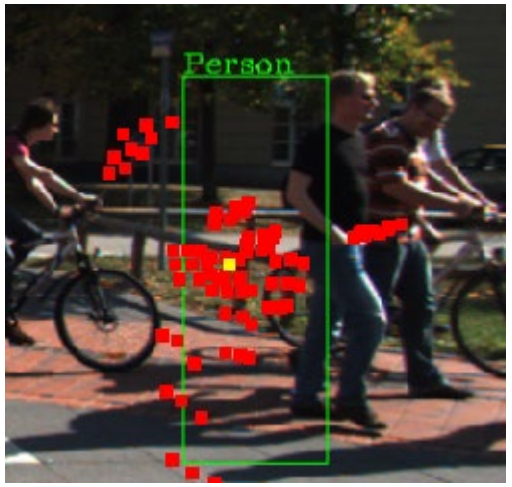
stsn not-detected



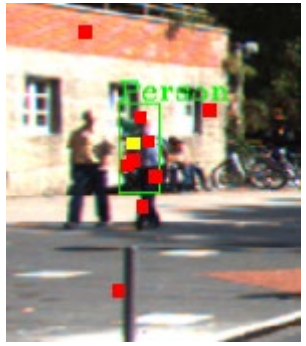
Reppoint offset

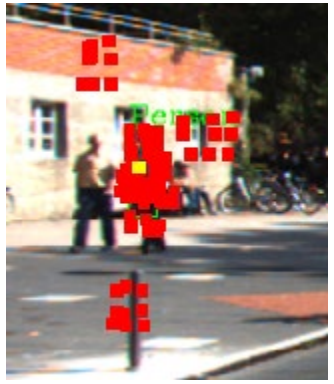


support1 offset



support2 offset





- For the top case, the offset on stage 2 is not right, the big object needs to visualize on the later stage. But on the N-2 frame, we can still see it is trying to find the right signal, using stsn may also explore the spatial space, so that we do not need to rely on the later stages.
- For the bottom case, too many signals are fused into the small objects.

cyclist--not-detected bbox and offset

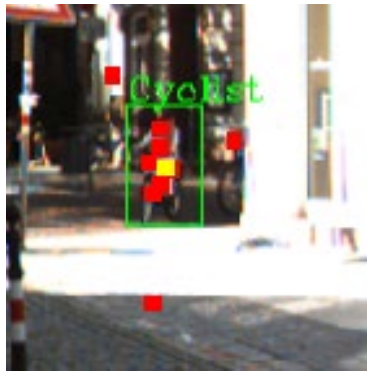
Green bbox is the ground bbox not-detected.
Yellow point is the center of the bbox
Red points are the offset of the yellow point.
The result is checked frame by frame, if the bbox only shows on frame 3 and frame 5, that means frame 4 detects all objects.
The training data for cyclist is real small, and the cyclist in the far distance is very easy to be classified as a person.
Support1 is N-2 frame, support2 is N+2 frame
Reppoint+stsn is two dcn layers, so we get 81 offsets to show

	Vehicle	Person	Cyclist
Train	22543	8702	1354
test	8058	3444	584

- First show failure case of reppoint but can be detected by stsn
- Then the failure case of stsn

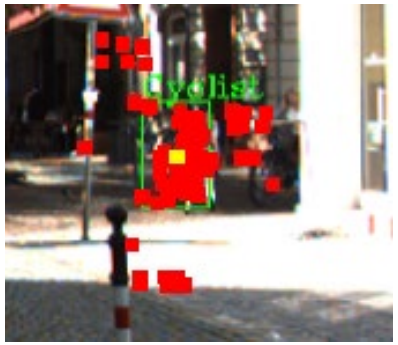
Fail case of stsn

stsn not-detected

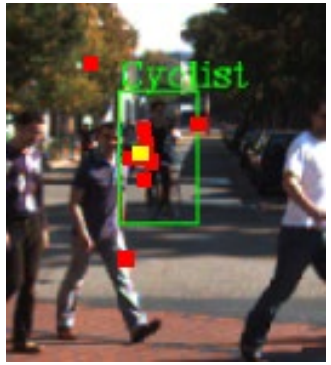


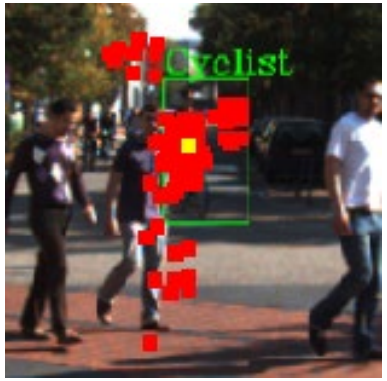
Reppoint offset

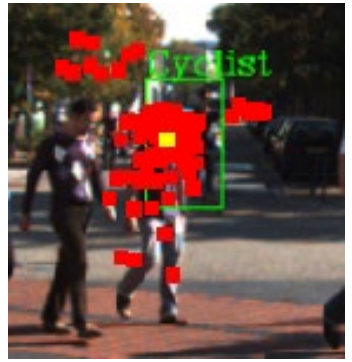
support1 offset



support2 offset



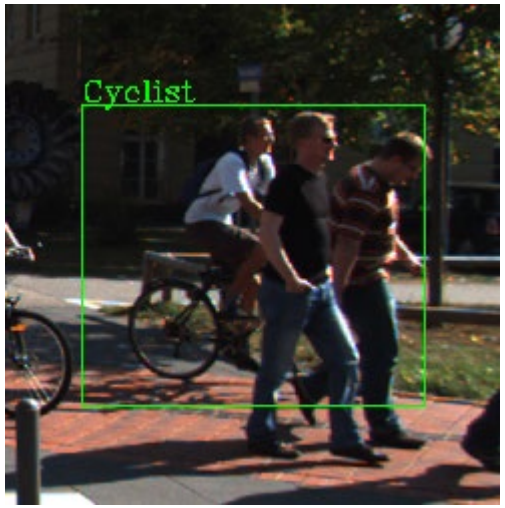


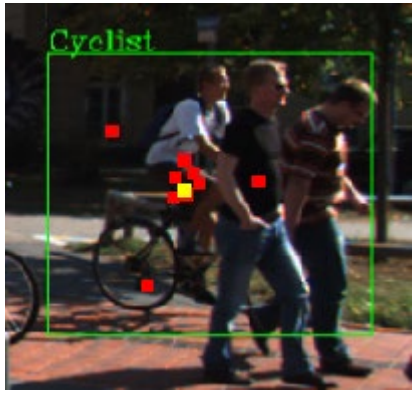


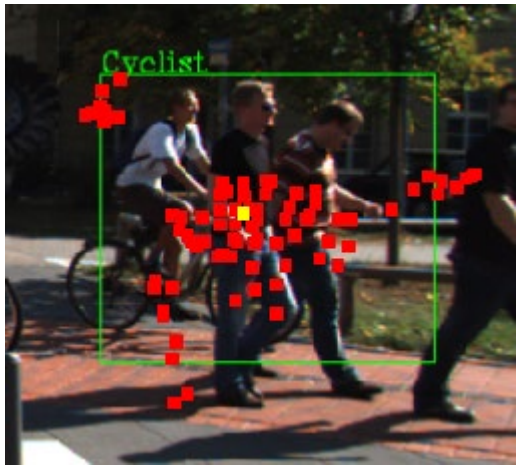
To be clear, most cyclist is classified as person because of the lack of training samples. They are detected but in the wrong class.
From bottom case, we can see the signal is easy to move on the person.

fail case of stsn

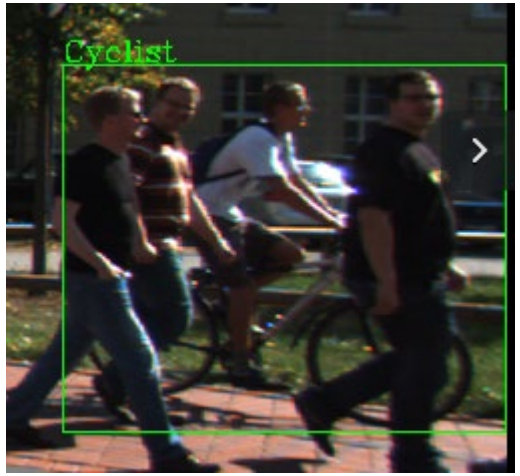
Cyclist occluded by person







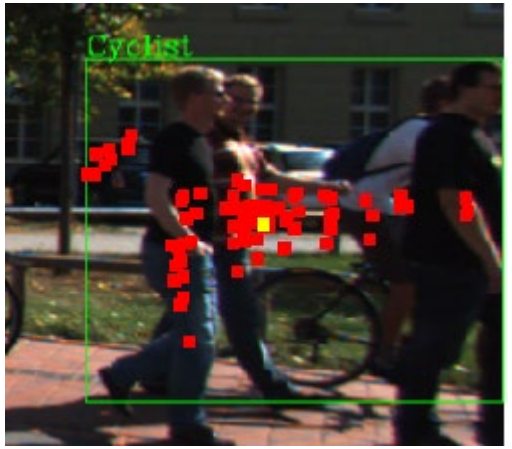
stsn not-detected



Reppoint offset

support1 offset

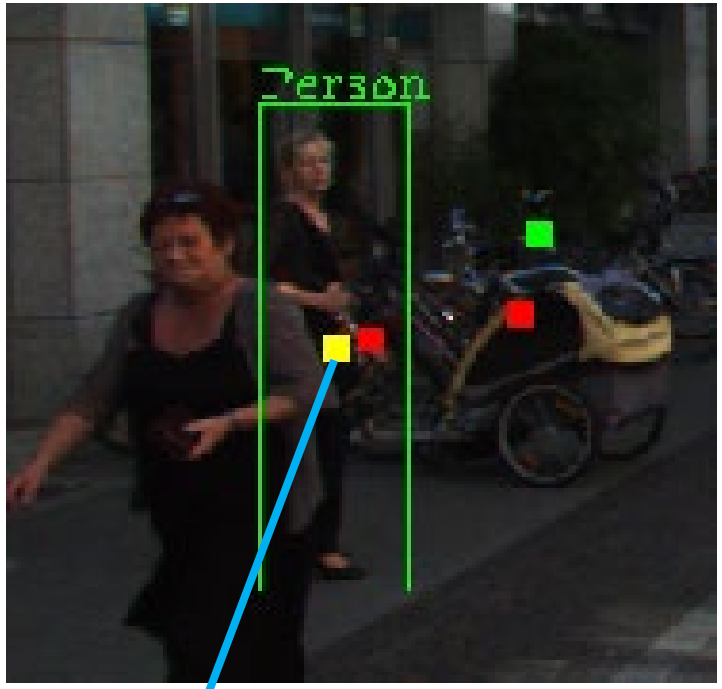
support2 offset



The stsn offset is wrong, it do not search for the person with white shirt but directly searching for person with black shirt. On the bottom base, even the signal of the rider on N-2 frame is more similar to frame N, the offsets are still on the black shirt person and the head is on the other person, the signal is confused for both aggregation and classification.

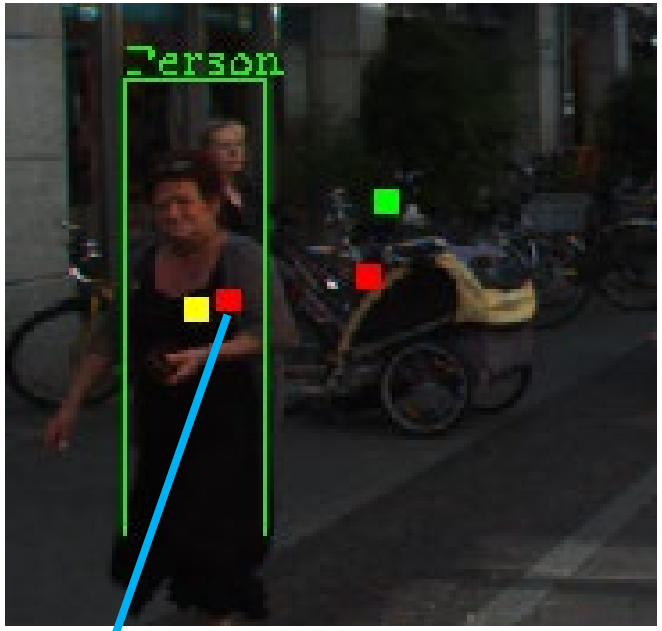
Init offset Success case

Frame N



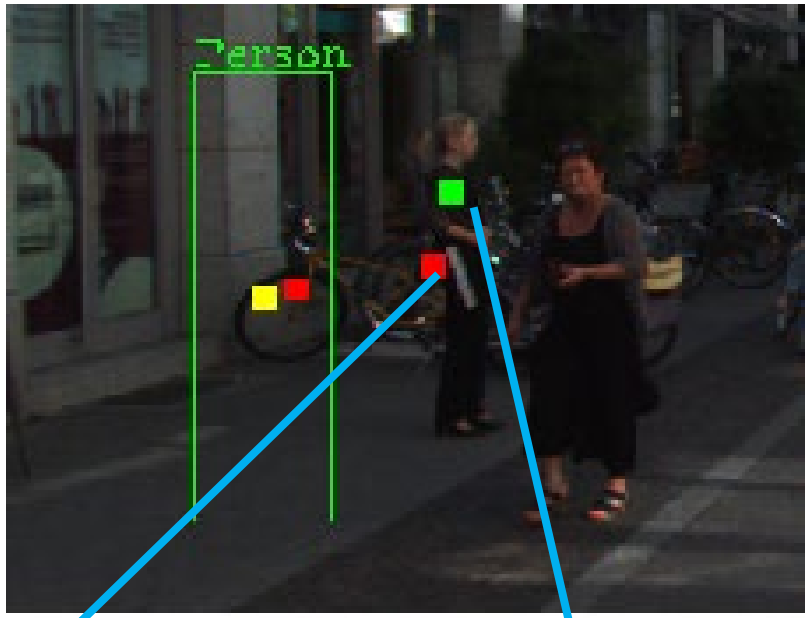
Occluded person at frame N

Frame N-1



Compute Stsn Offset on N-1
Offset is the step A
Compute the grad by linearization

Frame N-5



Init offset=5*A

- Warp reference feature by bilinear
- Linearization
  feature=grad*5A+reference

Final offset
Computed between support and warped feature

Quantity result of initialization

Training with random selected two frames among 8 frames,
Test with N-1 and N-5

Stsn one
Score 0.3, nms 0.5

evaluating result of refer

class	gts	dets	recall	precision	ap
Vehicle	29289	34230	0.752	0.643	0.700
Pedestrian	19463	32328	0.765	0.460	0.682
mAP					0.691

Use the pre-trained stsn one model.
Test with N-1 and N-5,
Use 4*(n-1 offset) as init for N-5 frame
Use bilinear to warp the reference to N-5

bilinear
Score 0.3, nms 0.5

evaluating result of agg

class	gts	dets	recall	precision	ap
Vehicle	29289	45518	0.768	0.494	0.687
Pedestrian	19463	45564	0.782	0.334	0.649
mAP					0.668

Test with N-1 and N-5,
Use 4*(n-1 offset) as step length, use grad from n-1 frame to compute the reference feature at N-5

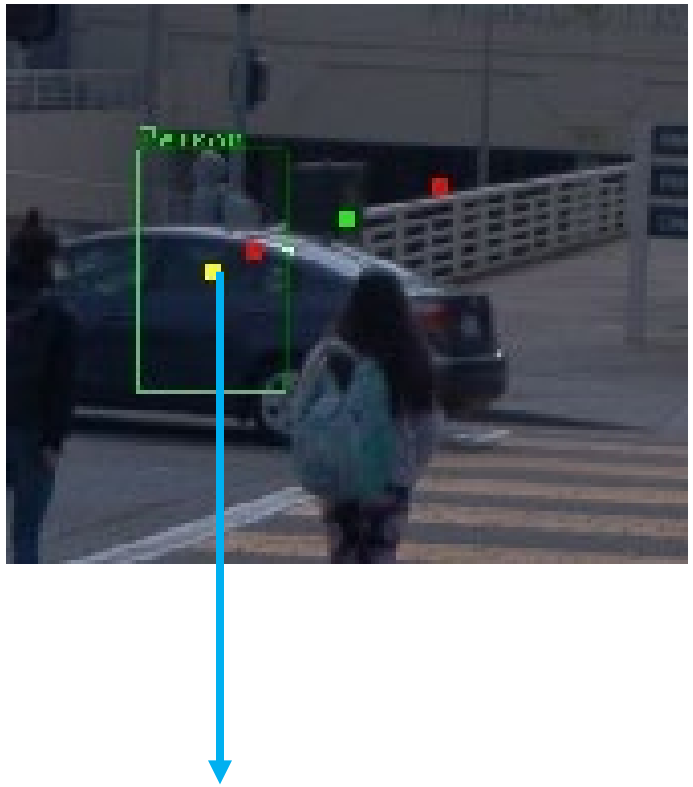
linear
Score 0.3, nms 0.5

evaluating result of agg

class	gts	dets	recall	precision	ap
Vehicle	29289	42730	0.756	0.518	0.657
Pedestrian	19463	38378	0.752	0.381	0.573
mAP					0.615

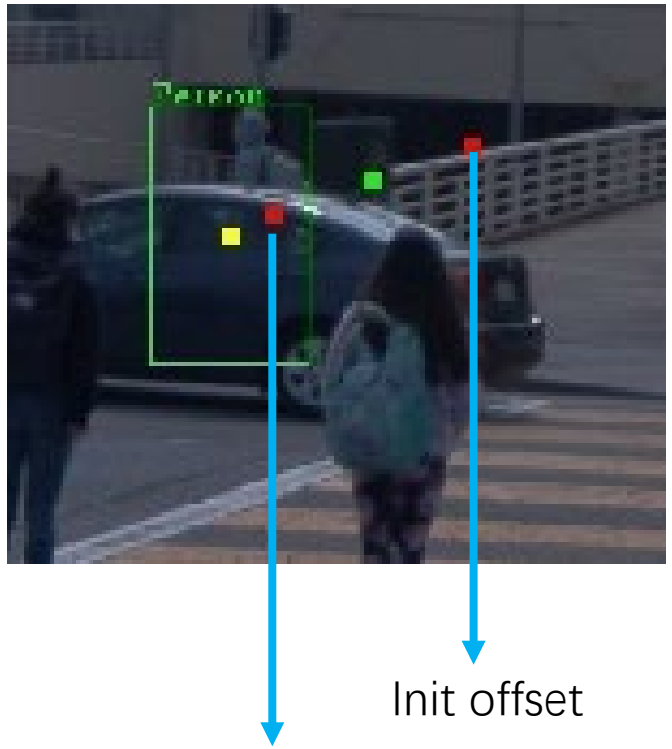
Success case

Frame N



Occluded person

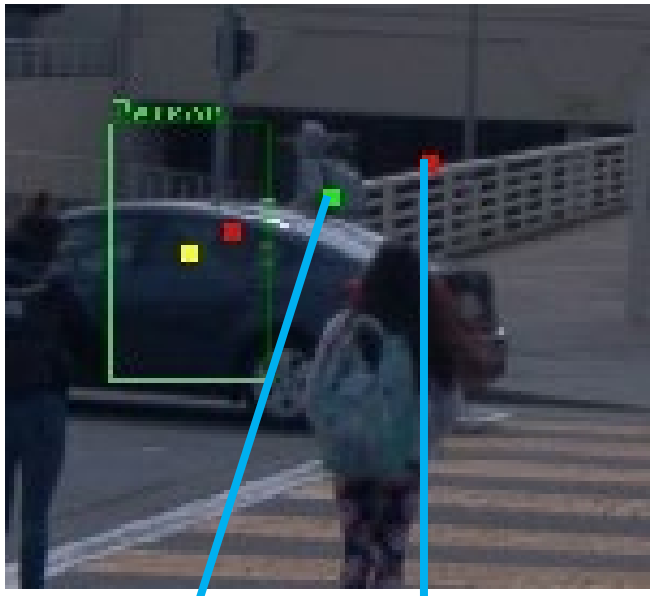
Frame N-1



Init offset

Offset on N-1

Frame N-5

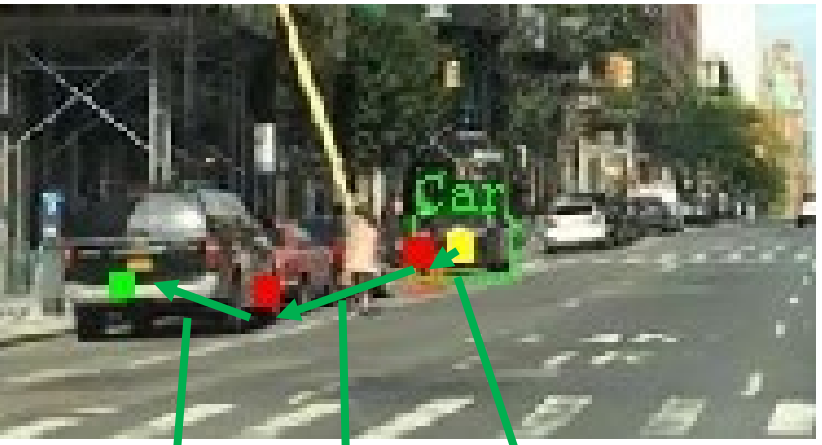


Init offset

Final offset

Fail case: most offset do not work as flow

Frame N

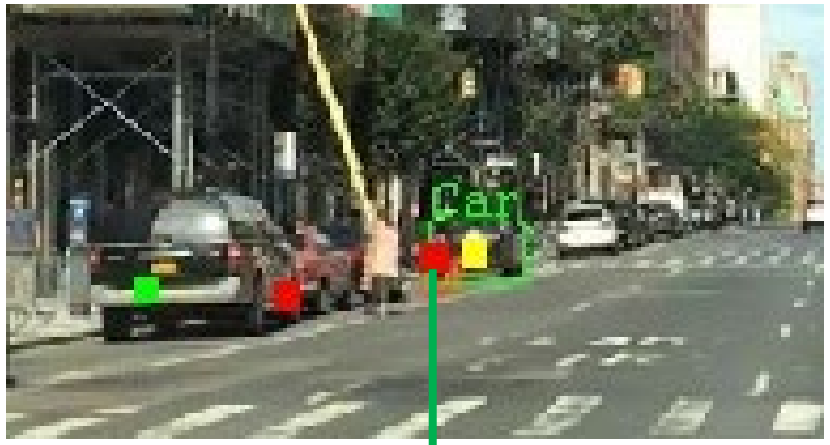


stsn

Initial offset

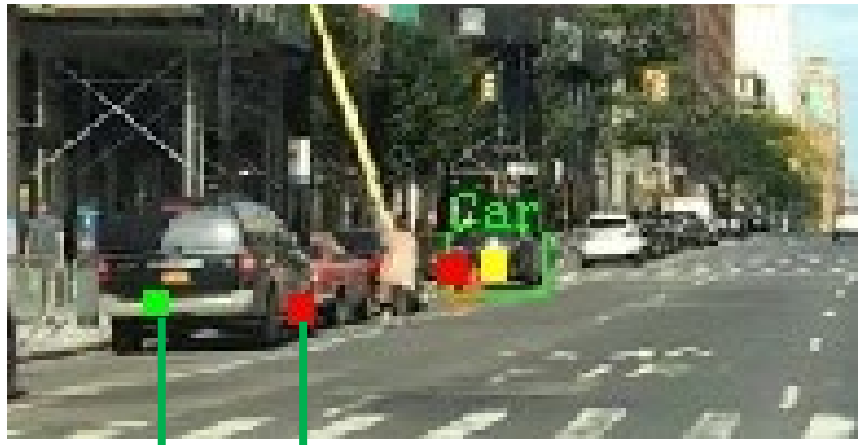
stsn

Frame N-1



Offset A between frame N and N-1
Compute grad and step=offset A

Frame N-5

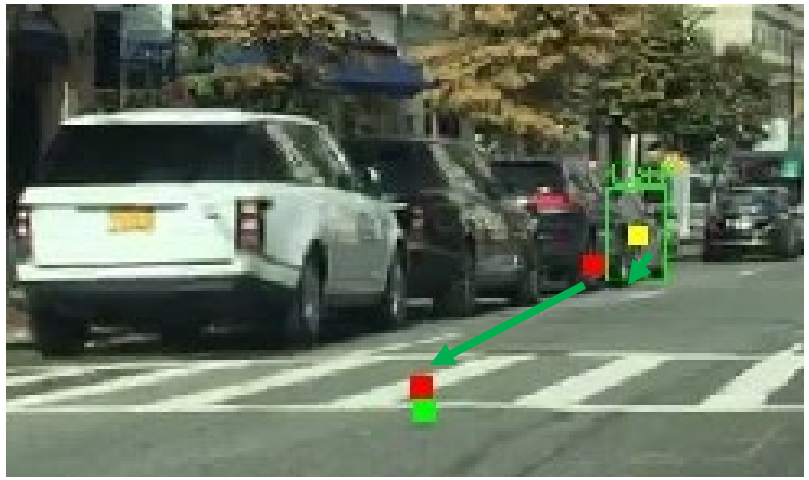


5*times of A to get the initial offset on frame 1

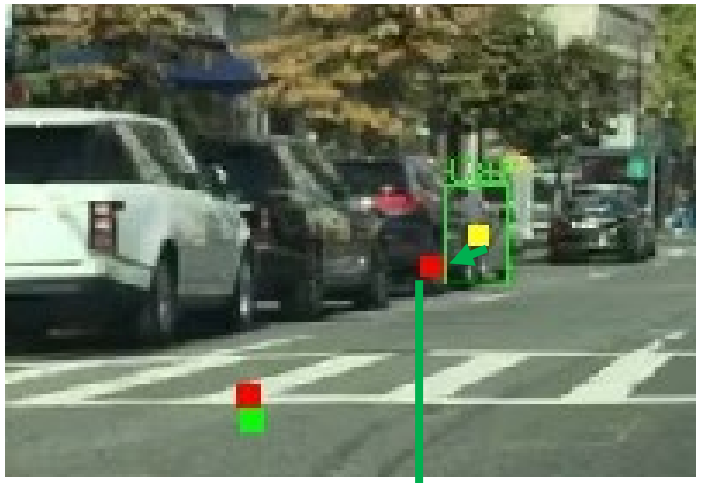
Offset found on frame 1
warped feature computed from linearization by grad*5*A +reference

The previous exps show the offset do not work as flow as we expected, so use the offset the initialize do not give the result as flow.

Frame N

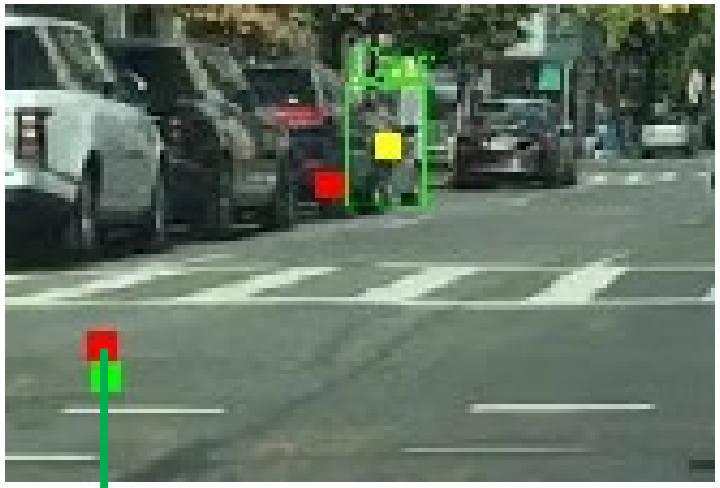


Frame N-1



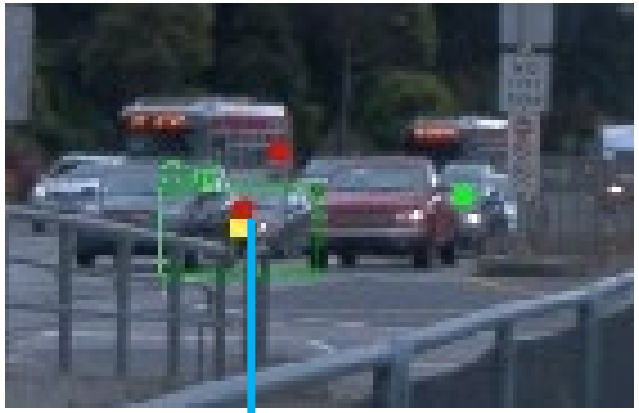
The offset is on the same class, the car itself is not moving

Frame N-5

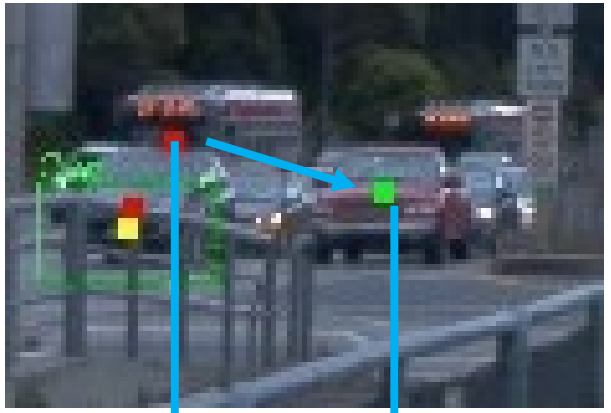


Initialize the offset by the N-1 result is failed as the offset is on the other object

Occluded car



Offset on N-1 frame

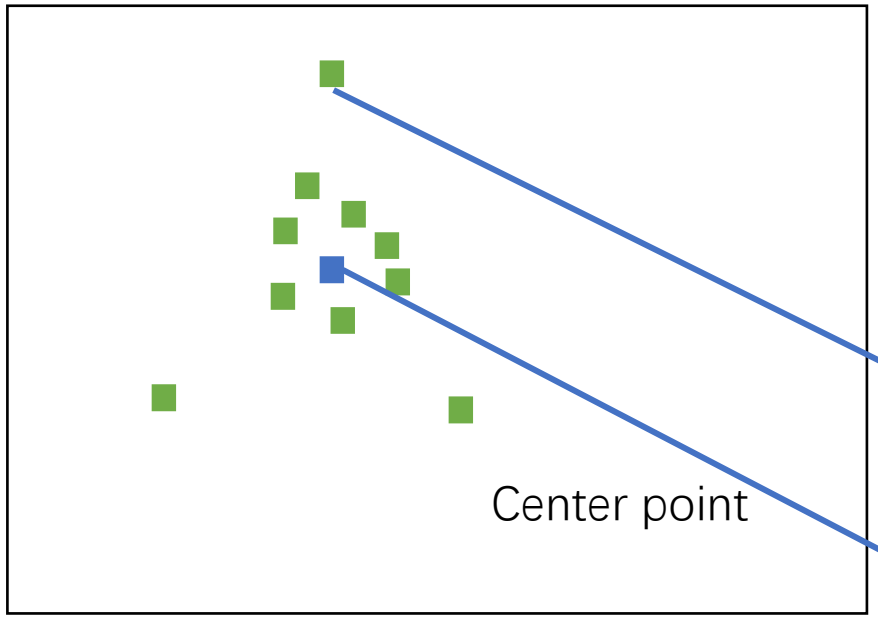


Init offset on N-5 frame

offset on N-5 frame

RANSAC realization1

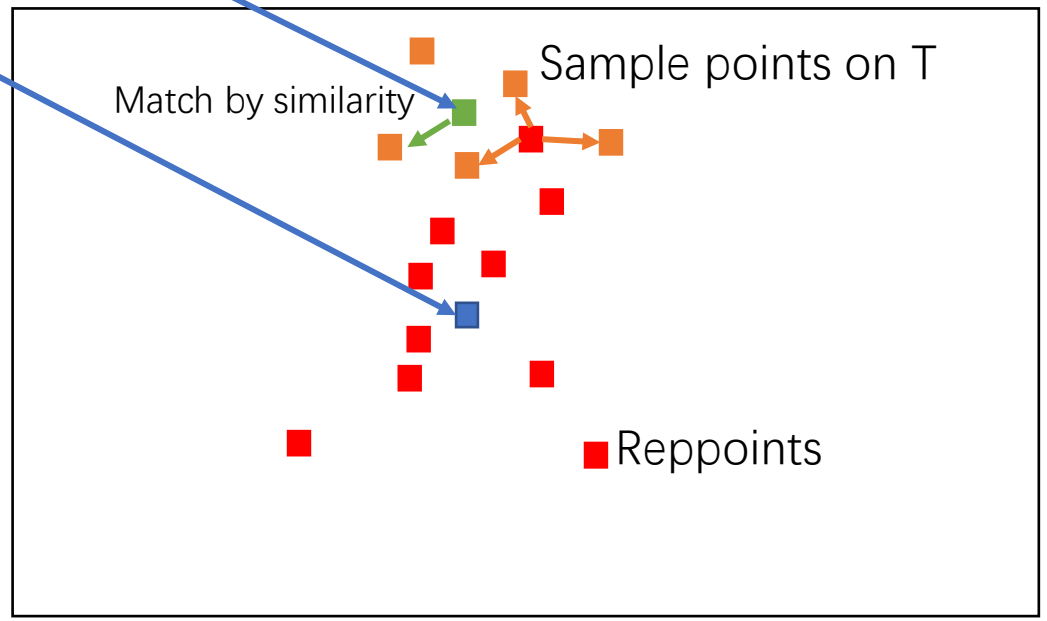
Step1: connect the center point by optical flow
Step2: warp the reppoints at T-1 to T
Step3: sample around the reppoints on T to compute the correspondence between reppoints at T-1 and T



Frame T-1

warp by flow

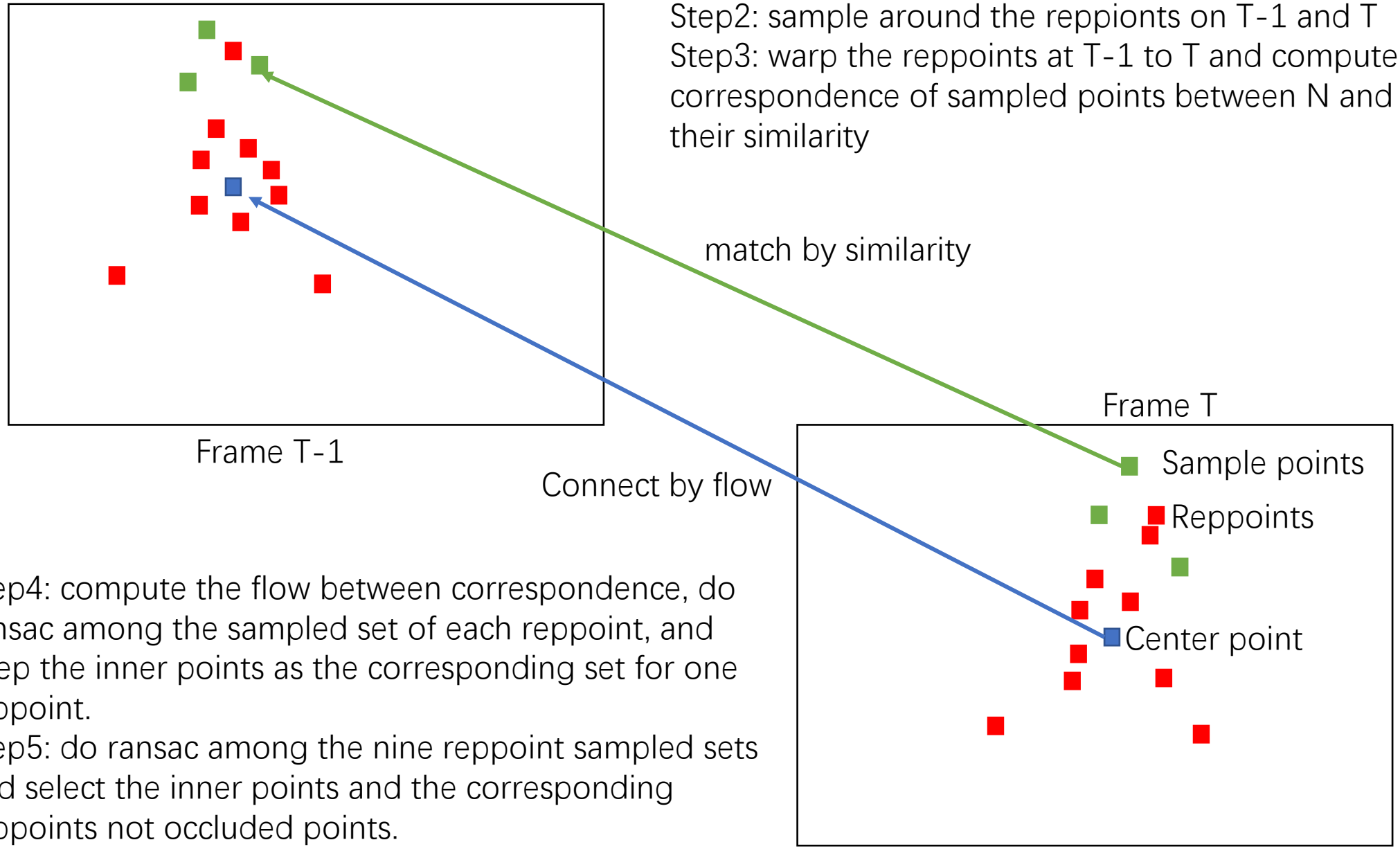
Frame T



Step4: compute the flow between correspondence, do ransac on the computed correspondence, and find the inner points to keep as the not occluded parts

RANSAC realization2

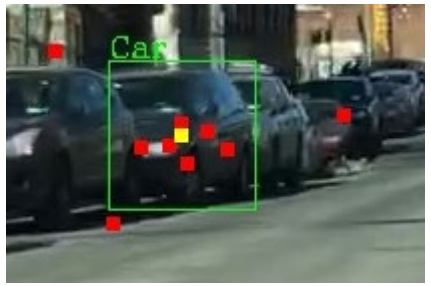
Step1: connect the center point by optical flow
Step2: sample around the reppionts on T-1 and T
Step3: warp the reppoints at T-1 to T and compute the correspondence of sampled points between N and N-1 by their similarity

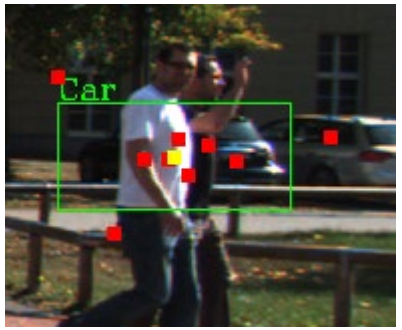


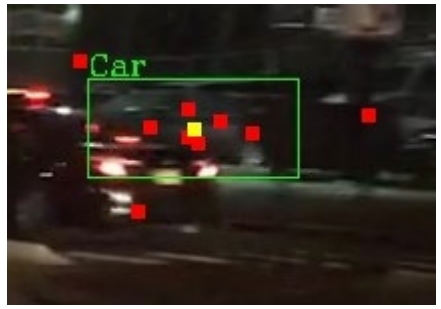
Step4: compute the flow between correspondence, do ransac among the sampled set of each reppoint, and keep the inner points as the corresponding set for one reppoint.
Step5: do ransac among the nine reppoint sampled sets and select the inner points and the corresponding reppoints not occluded points.

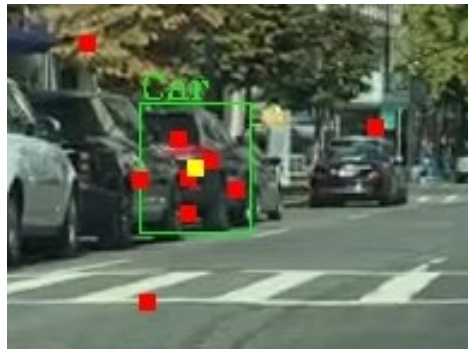
ransac

before

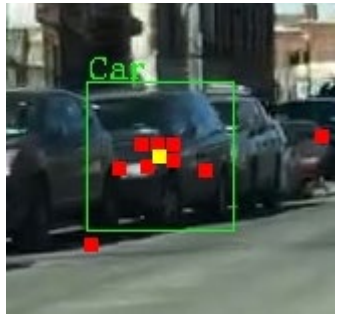


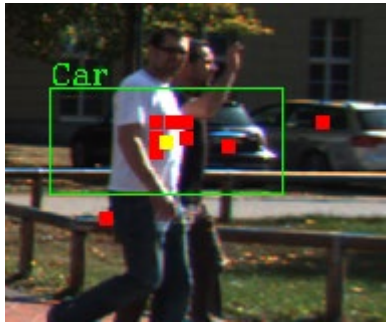


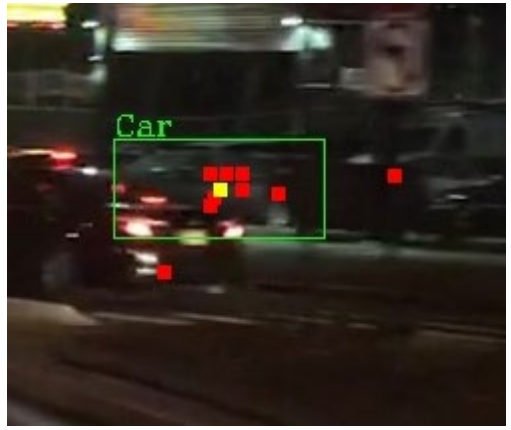


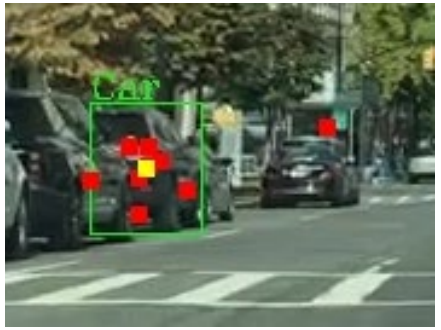


after









Points at grid postion is deactivate

Iterate the deactivate and reactivate 10 times

baseline

```
evaluating result of  /home/ld/RepPoints/ld_result/stsn_one_r
+------------+-------+-------+--------+-----------+-------+
| class      | gts   | dets  | recall | precision | ap    |
+------------+-------+-------+--------+-----------+-------+
| Vehicle    | 29289 | 36823 | 0.748  | 0.595     | 0.694 |
| Pedestrian | 19463 | 30665 | 0.765  | 0.486     | 0.691 |
+------------+-------+-------+--------+-----------+-------+
| mAP        |       |       |        |           | 0.693 |
+------------+-------+-------+--------+-----------+-------+
```

Deactivate+ reactivate

```
evaluating result of  /home/ld/RepPoints/ld_result/stsn_one_r
+------------+-------+-------+--------+-----------+-------+
| class      | gts   | dets  | recall | precision | ap    |
+------------+-------+-------+--------+-----------+-------+
| Vehicle    | 29289 | 35504 | 0.746  | 0.615     | 0.696 |
| Pedestrian | 19463 | 28945 | 0.764  | 0.513     | 0.698 |
+------------+-------+-------+--------+-----------+-------+
| mAP        |       |       |        |           | 0.697 |
+------------+-------+-------+--------+-----------+-------+
```

Iterate de-reactivate 5 times

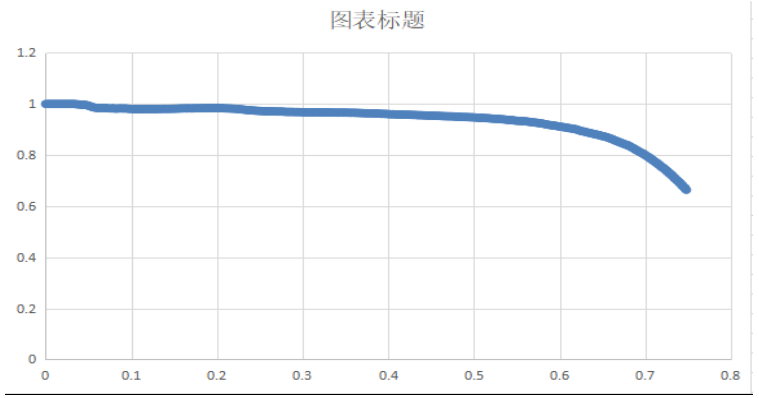
```
single test
evaluating result of  /home/ld/RepPoints/ld_result/stsn_one_r
+------------+-------+-------+--------+-----------+-------+
| class      | gts   | dets  | recall | precision | ap    |
+------------+-------+-------+--------+-----------+-------+
| Vehicle    | 29289 | 35597 | 0.745  | 0.613     | 0.695 |
| Pedestrian | 19463 | 29038 | 0.764  | 0.512     | 0.698 |
+------------+-------+-------+--------+-----------+-------+
| mAP        |       |       |        |           | 0.696 |
+------------+-------+-------+--------+-----------+-------+
```

Stsn one result trained end to end

Stsn one

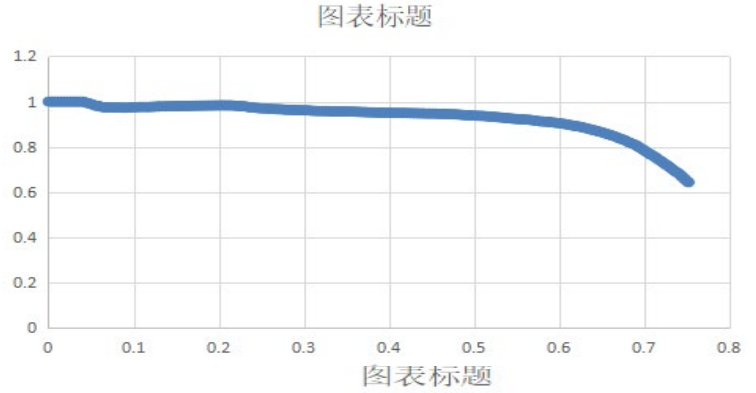
baseline

```
In [1]: kitti_eval_base(result_record, data_selected)
+--------+-------+-------+--------+-----------+-------+
| class  | gts   | dets  | recall | precision | ap    |
+--------+-------+-------+--------+-----------+-------+
| Car    | 29289 | 32998 | 0.748  | 0.664     | 0.703 |
| Person | 19463 | 29309 | 0.760  | 0.505     | 0.693 |
+--------+-------+-------+--------+-----------+-------+
| mAP    |       |       |        |           | 0.698 |
+--------+-------+-------+--------+-----------+-------+
```



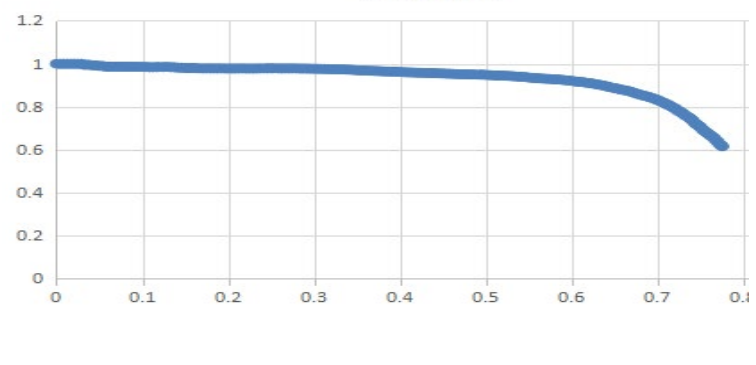
Fix weight

```
evaluating result of  refer
+------------+-------+-------+--------+-----------+-------+
| class      | gts   | dets  | recall | precision | ap    |
+------------+-------+-------+--------+-----------+-------+
| Vehicle    | 29289 | 34230 | 0.752  | 0.643     | 0.700 |
| Pedestrian | 19463 | 32328 | 0.765  | 0.460     | 0.682 |
+------------+-------+-------+--------+-----------+-------+
| mAP        |       |       |        |           | 0.691 |
+------------+-------+-------+--------+-----------+-------+
```



End to end

```
evaluating result of  refer
+------------+-------+-------+--------+-----------+-------+
| class      | gts   | dets  | recall | precision | ap    |
+------------+-------+-------+--------+-----------+-------+
| Vehicle    | 29289 | 36909 | 0.775  | 0.615     | 0.725 |
| Pedestrian | 19463 | 30404 | 0.754  | 0.483     | 0.679 |
+------------+-------+-------+--------+-----------+-------+
| mAP        |       |       |        |           | 0.702 |
+------------+-------+-------+--------+-----------+-------+
```



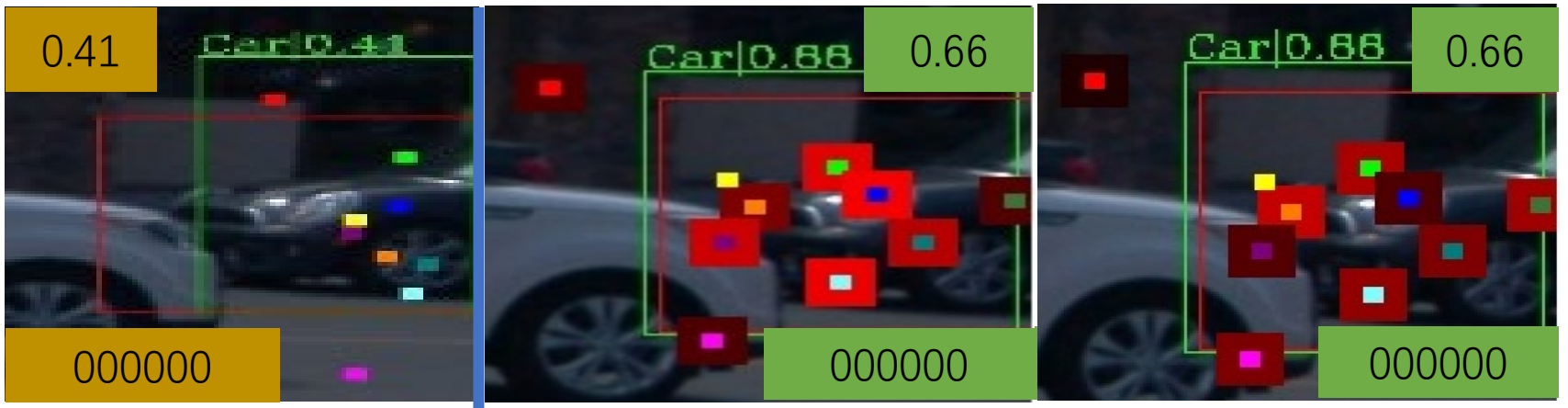
Result on Waymo

Reppoints Baseline

class	gts	dets	recall	precision	ap
Car	20825	22403	0.700	0.651	0.670
Person	7652	5539	0.492	0.680	0.450
mAP					0.560

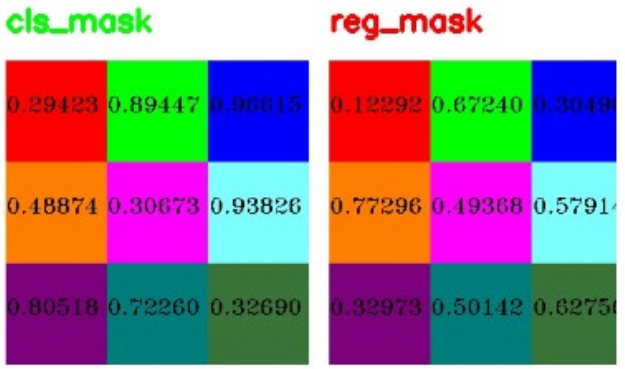
Reppoints with Mask

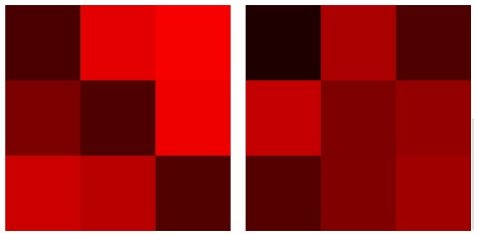
class	gts	dets	recall	precision	ap
Car	20825	20004	0.748	0.779	0.728
Person	7652	5797	0.565	0.745	0.527
mAP					0.628

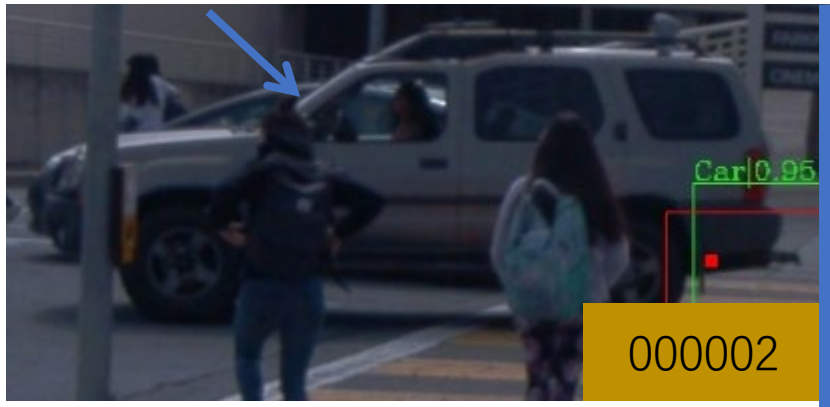


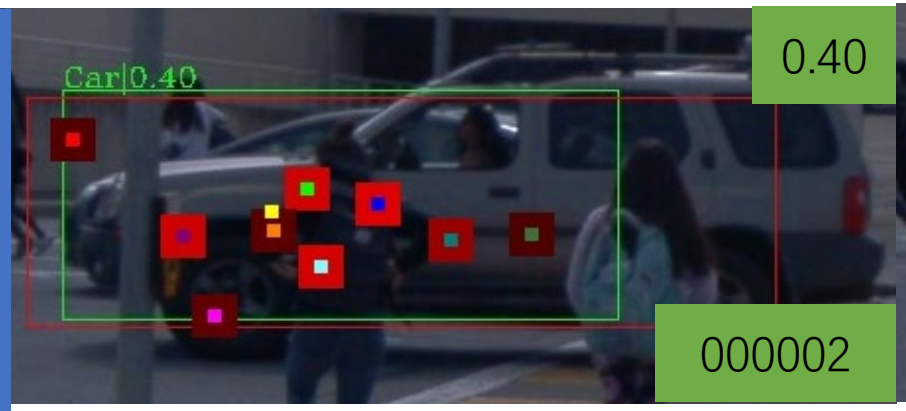
Reppoints Baseline

Reppoints With Mask



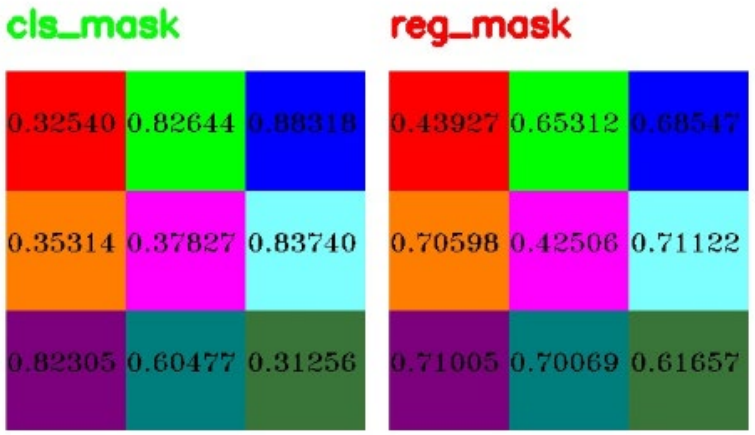


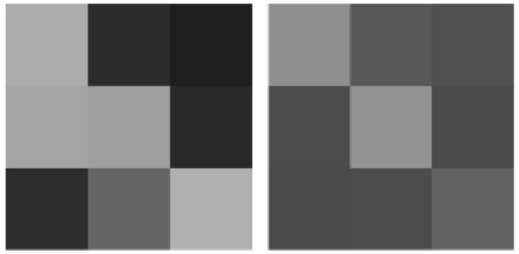




Reppoints Baseline

Reppoints With Mask



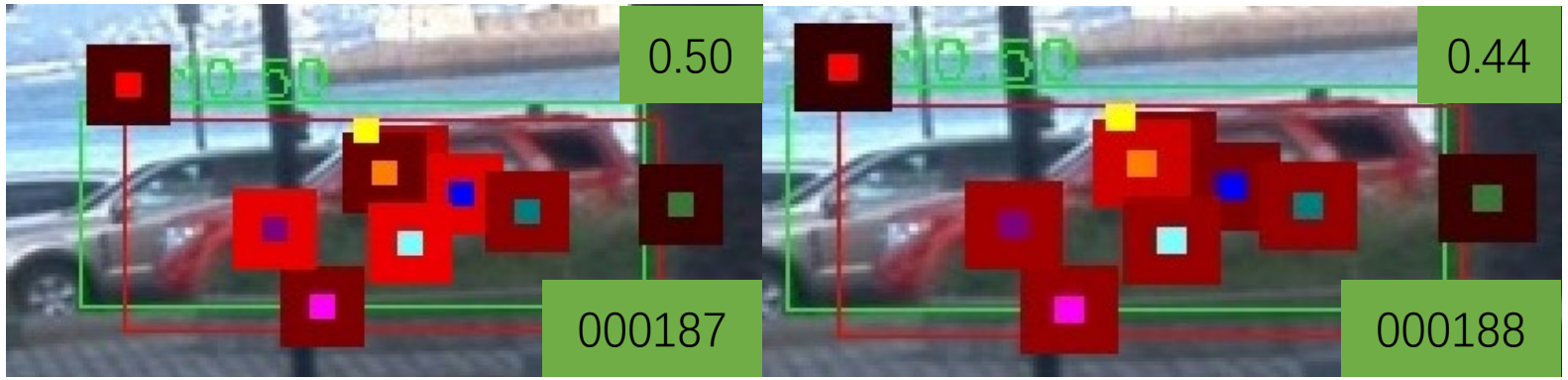


000187	000188	000189
000190	000191	000192
000193	000194	000195
000196	000197	000198

Reppoints Baseline

绿色代表能检测出红色车

红色代表能检测不出红色车

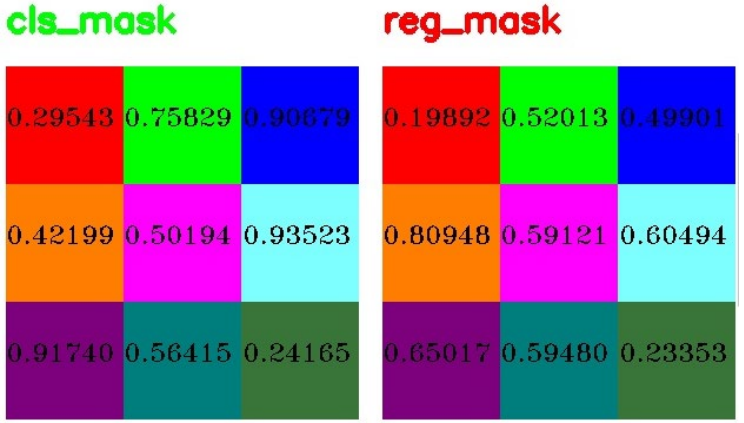


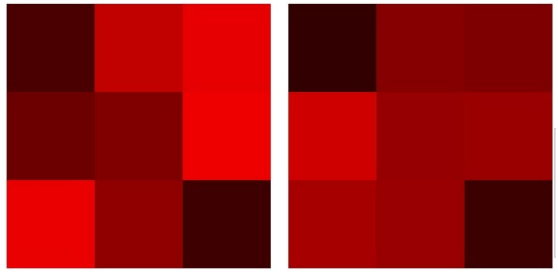
000187	000188	000189
000190	000191	000192
000193	000194	000195
000196	000197	000198

Reppoints With Mask

绿色代表能检测出红色车

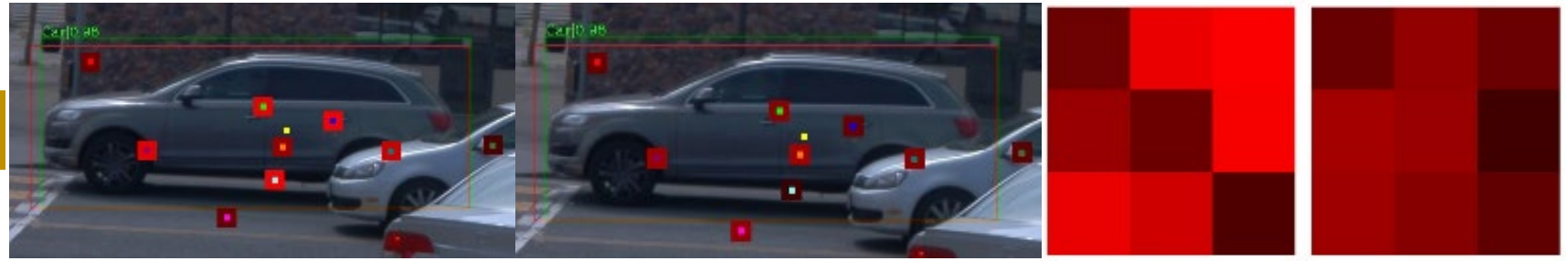
红色代表能检测不出红色车





000066

000068



000070

000072

reppoints baseline VS reppoints with mask

☑ 1	minmax	baseline		

```
In [1]: kitti_eval_base(result_record, data_selected)
+--------+-------+-------+--------+-----------+-------+
| class  | gts   | dets  | recall | precision | ap    |
+--------+-------+-------+--------+-----------+-------+
| Car    | 20825 | 22403 | 0.700  | 0.651     | 0.670 |
| Person | 7652  | 5539  | 0.492  | 0.680     | 0.450 |
+--------+-------+-------+--------+-----------+-------+
| mAP    |       |       |        |           | 0.560 |
+--------+-------+-------+--------+-----------+-------+
```

☑ 2	minmax	mask		

```
In [1]: kitti_eval_base(result_record, data_selected)
+--------+-------+-------+--------+-----------+-------+
| class  | gts   | dets  | recall | precision | ap    |
+--------+-------+-------+--------+-----------+-------+
| Car    | 20825 | 20004 | 0.748  | 0.779     | 0.728 |
| Person | 7652  | 5797  | 0.565  | 0.745     | 0.527 |
+--------+-------+-------+--------+-----------+-------+
| mAP    |       |       |        |           | 0.628 |
+--------+-------+-------+--------+-----------+-------+
```

☑ 3	minmax	mask	prediction	

```
evaluating result of  refer
+------------+-------+-------+--------+-----------+-------+
| class      | gts   | dets  | recall | precision | ap    |
+------------+-------+-------+--------+-----------+-------+
| Vehicle    | 20825 | 23240 | 0.758  | 0.679     | 0.728 |
| Pedestrian | 7652  | 7852  | 0.631  | 0.614     | 0.574 |
+------------+-------+-------+--------+-----------+-------+
| mAP        |       |       |        |           | 0.651 |
+------------+-------+-------+--------+-----------+-------+
```

☑ 4	moment	baseline		

```
In [1]: kitti_eval(result_record, dataset)
+------------+-------+-------+--------+-----------+-------+
| class      | gts   | dets  | recall | precision | ap    |
+------------+-------+-------+--------+-----------+-------+
| Vehicle    | 20825 | 22126 | 0.744  | 0.700     | 0.717 |
| Pedestrian | 7652  | 5807  | 0.537  | 0.708     | 0.478 |
+------------+-------+-------+--------+-----------+-------+
| mAP        |       |       |        |           | 0.597 |
+------------+-------+-------+--------+-----------+-------+
```

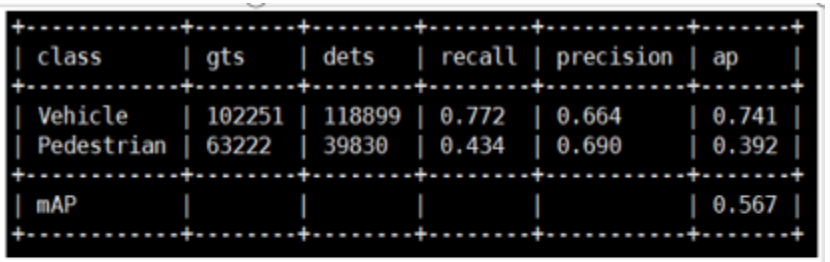
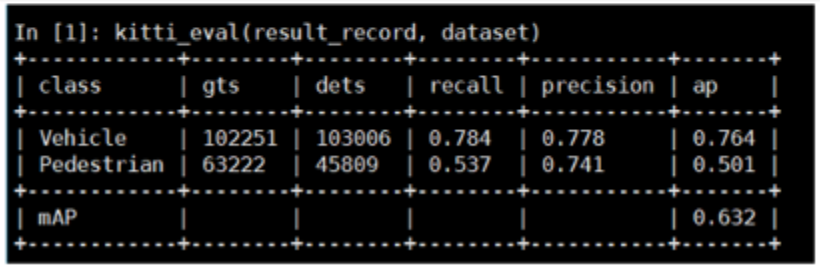
☑ 5	moment	mask		

```
+------------+-------+-------+--------+-----------+-------+
| class      | gts   | dets  | recall | precision | ap    |
+------------+-------+-------+--------+-----------+-------+
| Vehicle    | 20825 | 21921 | 0.756  | 0.718     | 0.726 |
| Pedestrian | 7652  | 7784  | 0.628  | 0.618     | 0.571 |
+------------+-------+-------+--------+-----------+-------+
| mAP        |       |       |        |           | 0.648 |
+------------+-------+-------+--------+-----------+-------+
```

☑ 6	moment	mask	prediction	

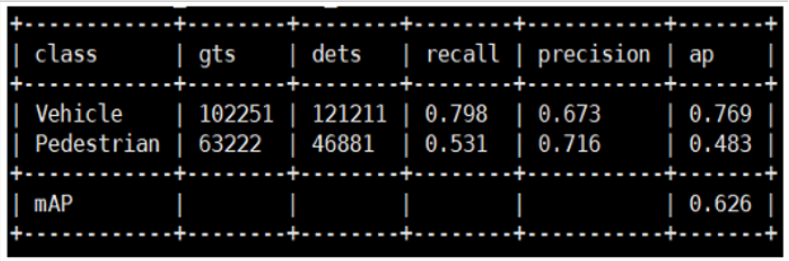
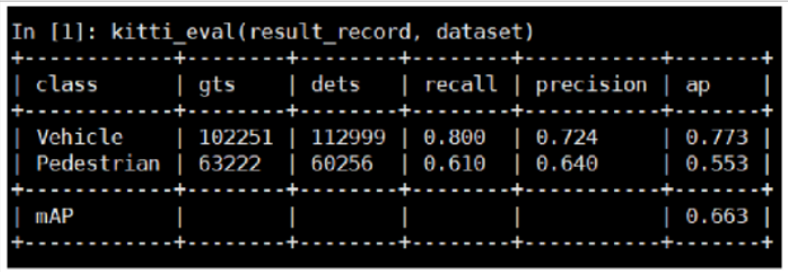
```
evaluating result of  refer
+------------+-------+-------+--------+-----------+-------+
| class      | gts   | dets  | recall | precision | ap    |
+------------+-------+-------+--------+-----------+-------+
| Vehicle    | 20825 | 22529 | 0.756  | 0.699     | 0.729 |
| Pedestrian | 7652  | 8301  | 0.644  | 0.594     | 0.580 |
+------------+-------+-------+--------+-----------+-------+
| mAP        |       |       |        |           | 0.655 |
+------------+-------+-------+--------+-----------+-------+
```

reppoints baseline VS reppoints with mask

☑ 1	minmax	baseline	
☑ 2	minmax	mask	

```
+------------+--------+--------+--------+-----------+-------+
| class      | gts    | dets   | recall | precision | ap    |
+------------+--------+--------+--------+-----------+-------+
| Vehicle    | 102251 | 118899 | 0.772  | 0.664     | 0.741 |
| Pedestrian | 63222  | 39830  | 0.434  | 0.690     | 0.392 |
+------------+--------+--------+--------+-----------+-------+
| mAP        |        |        |        |           | 0.567 |
+------------+--------+--------+--------+-----------+-------+
```

```
In [1]: kitti_eval(result_record, dataset)
+------------+--------+--------+--------+-----------+-------+
| class      | gts    | dets   | recall | precision | ap    |
+------------+--------+--------+--------+-----------+-------+
| Vehicle    | 102251 | 103006 | 0.784  | 0.778     | 0.764 |
| Pedestrian | 63222  | 45809  | 0.537  | 0.741     | 0.501 |
+------------+--------+--------+--------+-----------+-------+
| mAP        |        |        |        |           | 0.632 |
+------------+--------+--------+--------+-----------+-------+
```

☑ 3	moment	baseline	
☑ 4	moment	mask	

```
+------------+--------+--------+--------+-----------+-------+
| class      | gts    | dets   | recall | precision | ap    |
+------------+--------+--------+--------+-----------+-------+
| Vehicle    | 102251 | 121211 | 0.798  | 0.673     | 0.769 |
| Pedestrian | 63222  | 46881  | 0.531  | 0.716     | 0.483 |
+------------+--------+--------+--------+-----------+-------+
| mAP        |        |        |        |           | 0.626 |
+------------+--------+--------+--------+-----------+-------+
```

```
In [1]: kitti_eval(result_record, dataset)
+------------+--------+--------+--------+-----------+-------+
| class      | gts    | dets   | recall | precision | ap    |
+------------+--------+--------+--------+-----------+-------+
| Vehicle    | 102251 | 112999 | 0.800  | 0.724     | 0.773 |
| Pedestrian | 63222  | 60256  | 0.610  | 0.640     | 0.553 |
+------------+--------+--------+--------+-----------+-------+
| mAP        |        |        |        |           | 0.663 |
+------------+--------+--------+--------+-----------+-------+
```

reppoints baseline VS reppoints with mask

Baseline

```
DONE (t=8.33s).
 Average Precision  (AP) @[ IoU=0.50:0.95 | area=   all | maxDets=100 ] = 0.416
 Average Precision  (AP) @[ IoU=0.50      | area=   all | maxDets=100 ] = 0.620
 Average Precision  (AP) @[ IoU=0.75      | area=   all | maxDets=100 ] = 0.453
 Average Precision  (AP) @[ IoU=0.50:0.95 | area= small | maxDets=100 ] = 0.245
 Average Precision  (AP) @[ IoU=0.50:0.95 | area=medium | maxDets=100 ] = 0.463
 Average Precision  (AP) @[ IoU=0.50:0.95 | area= large | maxDets=100 ] = 0.541
 Average Recall     (AR) @[ IoU=0.50:0.95 | area=   all | maxDets=  1 ] = 0.342
 Average Recall     (AR) @[ IoU=0.50:0.95 | area=   all | maxDets= 10 ] = 0.547
 Average Recall     (AR) @[ IoU=0.50:0.95 | area=   all | maxDets=100 ] = 0.582
 Average Recall     (AR) @[ IoU=0.50:0.95 | area= small | maxDets=100 ] = 0.379
 Average Recall     (AR) @[ IoU=0.50:0.95 | area=medium | maxDets=100 ] = 0.631
 Average Recall     (AR) @[ IoU=0.50:0.95 | area= large | maxDets=100 ] = 0.742
(mmlab2)
```

epoch5

```
 Average Precision  (AP) @[ IoU=0.50:0.95 | area=   all | maxDets=100 ] = 0.374
 Average Precision  (AP) @[ IoU=0.50      | area=   all | maxDets=100 ] = 0.569
 Average Precision  (AP) @[ IoU=0.75      | area=   all | maxDets=100 ] = 0.405
 Average Precision  (AP) @[ IoU=0.50:0.95 | area= small | maxDets=100 ] = 0.210
 Average Precision  (AP) @[ IoU=0.50:0.95 | area=medium | maxDets=100 ] = 0.420
 Average Precision  (AP) @[ IoU=0.50:0.95 | area= large | maxDets=100 ] = 0.497
 Average Recall     (AR) @[ IoU=0.50:0.95 | area=   all | maxDets=  1 ] = 0.323
 Average Recall     (AR) @[ IoU=0.50:0.95 | area=   all | maxDets= 10 ] = 0.516
 Average Recall     (AR) @[ IoU=0.50:0.95 | area=   all | maxDets=100 ] = 0.546
 Average Recall     (AR) @[ IoU=0.50:0.95 | area= small | maxDets=100 ] = 0.344
 Average Recall     (AR) @[ IoU=0.50:0.95 | area=medium | maxDets=100 ] = 0.592
 Average Recall     (AR) @[ IoU=0.50:0.95 | area= large | maxDets=100 ] = 0.708
(mmlab2)
```

epoch15

```
Average Precision  (AP) @[ IoU=0.50:0.95 | area=   all | maxDets=100 ] = 0.370
Average Precision  (AP) @[ IoU=0.50      | area=   all | maxDets=100 ] = 0.561
Average Precision  (AP) @[ IoU=0.75      | area=   all | maxDets=100 ] = 0.399
Average Precision  (AP) @[ IoU=0.50:0.95 | area= small | maxDets=100 ] = 0.214
Average Precision  (AP) @[ IoU=0.50:0.95 | area=medium | maxDets=100 ] = 0.410
Average Precision  (AP) @[ IoU=0.50:0.95 | area= large | maxDets=100 ] = 0.488
Average Recall     (AR) @[ IoU=0.50:0.95 | area=   all | maxDets=  1 ] = 0.318
Average Recall     (AR) @[ IoU=0.50:0.95 | area=   all | maxDets= 10 ] = 0.511
Average Recall     (AR) @[ IoU=0.50:0.95 | area=   all | maxDets=100 ] = 0.543
Average Recall     (AR) @[ IoU=0.50:0.95 | area= small | maxDets=100 ] = 0.339
Average Recall     (AR) @[ IoU=0.50:0.95 | area=medium | maxDets=100 ] = 0.587
Average Recall     (AR) @[ IoU=0.50:0.95 | area= large | maxDets=100 ] = 0.695
```

Reppoints minmax

```
In [1]: kitti_eval_base(result_record, data_selected)
+--------+-------+-------+--------+-----------+-------+
| class  | gts   | dets  | recall | precision | ap    |
+--------+-------+-------+--------+-----------+-------+
| Car    | 20825 | 22403 | 0.700  | 0.651     | 0.670 |
| Person | 7652  | 5539  | 0.492  | 0.680     | 0.450 |
+--------+-------+-------+--------+-----------+-------+
| mAP    |       |       |        |           | 0.560 |
+--------+-------+-------+--------+-----------+-------+
```

Reppoints minmax Mask

```
In [1]: kitti_eval_base(result_record, data_selected)
+--------+-------+-------+--------+-----------+-------+
| class  | gts   | dets  | recall | precision | ap    |
+--------+-------+-------+--------+-----------+-------+
| Car    | 20825 | 20004 | 0.748  | 0.779     | 0.728 |
| Person | 7652  | 5797  | 0.565  | 0.745     | 0.527 |
+--------+-------+-------+--------+-----------+-------+
| mAP    |       |       |        |           | 0.628 |
+--------+-------+-------+--------+-----------+-------+
```

moment

```
In [1]: kitti_eval(result_record, dataset)
+------------+-------+-------+--------+-----------+-------+
| class      | gts   | dets  | recall | precision | ap    |
+------------+-------+-------+--------+-----------+-------+
| Vehicle    | 20825 | 22126 | 0.744  | 0.700     | 0.717 |
| Pedestrian | 7652  | 5807  | 0.537  | 0.708     | 0.478 |
+------------+-------+-------+--------+-----------+-------+
| mAP        |       |       |        |           | 0.597 |
+------------+-------+-------+--------+-----------+-------+
```

epoch 20

Moment+mask

```
In [1]: kitti_eval(result_record, dataset)
+------------+-------+-------+--------+-----------+-------+
| class      | gts   | dets  | recall | precision | ap    |
+------------+-------+-------+--------+-----------+-------+
| Vehicle    | 20825 | 19813 | 0.734  | 0.771     | 0.708 |
| Pedestrian | 7652  | 8693  | 0.624  | 0.549     | 0.559 |
+------------+-------+-------+--------+-----------+-------+
| mAP        |       |       |        |           | 0.633 |
+------------+-------+-------+--------+-----------+-------+
```

epoch 9

```
evaluating result of  refer
+------------+-------+-------+--------+-----------+-------+
| class      | gts   | dets  | recall | precision | ap    |
+------------+-------+-------+--------+-----------+-------+
| Vehicle    | 20825 | 23240 | 0.758  | 0.679     | 0.728 |
| Pedestrian | 7652  | 7852  | 0.631  | 0.614     | 0.574 |
+------------+-------+-------+--------+-----------+-------+
| mAP        |       |       |        |           | 0.651 |
+------------+-------+-------+--------+-----------+-------+
```

minmax+mask+prediction

```
evaluating result of  refer
+------------+-------+-------+--------+-----------+-------+
| class      | gts   | dets  | recall | precision | ap    |
+------------+-------+-------+--------+-----------+-------+
| Vehicle    | 20825 | 22529 | 0.756  | 0.699     | 0.729 |
| Pedestrian | 7652  | 8301  | 0.644  | 0.594     | 0.580 |
+------------+-------+-------+--------+-----------+-------+
| mAP        |       |       |        |           | 0.655 |
+------------+-------+-------+--------+-----------+-------+
```

epoch 20

Moment+mask+prediction

waymo8

Reppoints minmax

```
In [1]: kitti_eval_base(result_record, data_selected)
+--------+-------+-------+--------+-----------+-------+
| class  | gts   | dets  | recall | precision | ap    |
+--------+-------+-------+--------+-----------+-------+
| Car    | 20825 | 22403 | 0.700  | 0.651     | 0.670 |
| Person | 7652  | 5539  | 0.492  | 0.680     | 0.450 |
+--------+-------+-------+--------+-----------+-------+
| mAP    |       |       |        |           | 0.560 |
+--------+-------+-------+--------+-----------+-------+
```

Reppoints minmax Mask

```
In [1]: kitti_eval_base(result_record, data_selected)
+--------+-------+-------+--------+-----------+-------+
| class  | gts   | dets  | recall | precision | ap    |
+--------+-------+-------+--------+-----------+-------+
| Car    | 20825 | 20004 | 0.748  | 0.779     | 0.728 |
| Person | 7652  | 5797  | 0.565  | 0.745     | 0.527 |
+--------+-------+-------+--------+-----------+-------+
| mAP    |       |       |        |           | 0.628 |
+--------+-------+-------+--------+-----------+-------+
```

```
In [1]: kitti_eval(result_record, dataset)
+------------+-------+-------+--------+-----------+-------+
| class      | gts   | dets  | recall | precision | ap    |
+------------+-------+-------+--------+-----------+-------+
| Vehicle    | 20825 | 21921 | 0.756  | 0.718     | 0.726 |
| Pedestrian | 7652  | 7784  | 0.628  | 0.618     | 0.571 |
+------------+-------+-------+--------+-----------+-------+
| mAP        |       |       |        |           | 0.648 |
+------------+-------+-------+--------+-----------+-------+
```

moment

```
In [1]: kitti_eval(result_record, dataset)
+------------+-------+-------+--------+-----------+-------+
| class      | gts   | dets  | recall | precision | ap    |
+------------+-------+-------+--------+-----------+-------+
| Vehicle    | 20825 | 22126 | 0.744  | 0.700     | 0.717 |
| Pedestrian | 7652  | 5807  | 0.537  | 0.708     | 0.478 |
+------------+-------+-------+--------+-----------+-------+
| mAP        |       |       |        |           | 0.597 |
+------------+-------+-------+--------+-----------+-------+
```

epoch 20

Moment+mask

```
In [1]: kitti_eval(result_record, dataset)
+------------+-------+-------+--------+-----------+-------+
| class      | gts   | dets  | recall | precision | ap    |
+------------+-------+-------+--------+-----------+-------+
| Vehicle    | 20825 | 19813 | 0.734  | 0.771     | 0.708 |
| Pedestrian | 7652  | 8693  | 0.624  | 0.549     | 0.559 |
+------------+-------+-------+--------+-----------+-------+
| mAP        |       |       |        |           | 0.633 |
+------------+-------+-------+--------+-----------+-------+
```

epoch 9

```
evaluating result of  refer
+------------+-------+-------+--------+-----------+-------+
| class      | gts   | dets  | recall | precision | ap    |
+------------+-------+-------+--------+-----------+-------+
| Vehicle    | 20825 | 22529 | 0.756  | 0.699     | 0.729 |
| Pedestrian | 7652  | 8301  | 0.644  | 0.594     | 0.580 |
+------------+-------+-------+--------+-----------+-------+
| mAP        |       |       |        |           | 0.655 |
+------------+-------+-------+--------+-----------+-------+
```

epoch 20

Moment+mask+prediction

Moment+mask

```
+------------+-------+-------+--------+-----------+-------+
| class      | gts   | dets  | recall | precision | ap    |
+------------+-------+-------+--------+-----------+-------+
| Vehicle    | 20825 | 21921 | 0.756  | 0.718     | 0.726 |
| Pedestrian | 7652  | 7784  | 0.628  | 0.618     | 0.571 |
+------------+-------+-------+--------+-----------+-------+
| mAP        |       |       |        |           | 0.648 |
+------------+-------+-------+--------+-----------+-------+
```

epoch 20

waymo54

☑ 1	minmax	baseline	

```
+------------+--------+--------+--------+-----------+-------+
| class      | gts    | dets   | recall | precision | ap    |
+------------+--------+--------+--------+-----------+-------+
| Vehicle    | 102251 | 118899 | 0.772  | 0.664     | 0.741 |
| Pedestrian | 63222  | 39830  | 0.434  | 0.690     | 0.392 |
+------------+--------+--------+--------+-----------+-------+
| mAP        |        |        |        |           | 0.567 |
+------------+--------+--------+--------+-----------+-------+
```

☑ 2	minmax	mask	

```
In [1]: kitti_eval(result_record, dataset)
+------------+--------+--------+--------+-----------+-------+
| class      | gts    | dets   | recall | precision | ap    |
+------------+--------+--------+--------+-----------+-------+
| Vehicle    | 102251 | 103006 | 0.784  | 0.778     | 0.764 |
| Pedestrian | 63222  | 45809  | 0.537  | 0.741     | 0.501 |
+------------+--------+--------+--------+-----------+-------+
| mAP        |        |        |        |           | 0.632 |
+------------+--------+--------+--------+-----------+-------+
```

☑ 3	moment	baseline	

```
+------------+--------+--------+--------+-----------+-------+
| class      | gts    | dets   | recall | precision | ap    |
+------------+--------+--------+--------+-----------+-------+
| Vehicle    | 102251 | 121211 | 0.798  | 0.673     | 0.769 |
| Pedestrian | 63222  | 46881  | 0.531  | 0.716     | 0.483 |
+------------+--------+--------+--------+-----------+-------+
| mAP        |        |        |        |           | 0.626 |
+------------+--------+--------+--------+-----------+-------+
```

☑ 4	moment	mask	

```
In [1]: kitti_eval(result_record, dataset)
+------------+--------+--------+--------+-----------+-------+
| class      | gts    | dets   | recall | precision | ap    |
+------------+--------+--------+--------+-----------+-------+
| Vehicle    | 102251 | 112999 | 0.800  | 0.724     | 0.773 |
| Pedestrian | 63222  | 60256  | 0.610  | 0.640     | 0.553 |
+------------+--------+--------+--------+-----------+-------+
| mAP        |        |        |        |           | 0.663 |
+------------+--------+--------+--------+-----------+-------+
```

```
evaluating result of  refer
+------------+--------+--------+--------+-----------+-------+
| class      | gts    | dets   | recall | precision | ap    |
+------------+--------+--------+--------+-----------+-------+
| Vehicle    | 102251 | 119940 | 0.813  | 0.693     | 0.787 |
| Pedestrian | 63222  | 61177  | 0.583  | 0.603     | 0.520 |
+------------+--------+--------+--------+-----------+-------+
| mAP        |        |        |        |           | 0.654 |
+------------+--------+--------+--------+-----------+-------+
```

Prediction mimax

```
evaluating result of  refer
+------------+--------+--------+--------+-----------+-------+
| class      | gts    | dets   | recall | precision | ap    |
+------------+--------+--------+--------+-----------+-------+
| Vehicle    | 102251 | 117039 | 0.801  | 0.700     | 0.776 |
| Pedestrian | 63222  | 63292  | 0.610  | 0.609     | 0.552 |
+------------+--------+--------+--------+-----------+-------+
| mAP        |        |        |        |           | 0.664 |
+------------+--------+--------+--------+-----------+-------+
```

Prediction moment

Stsn waymo8

Reppoints minmax	In [1]: kitti_eval_base(result_record, data_selected) +--------+-------+-------+--------+-----------+-------+ | class | gts | dets | recall | precision | ap | +--------+-------+-------+--------+-----------+-------+ | Car | 20825 | 22403 | 0.700 | 0.651 | 0.670 | | Person | 7652 | 5539 | 0.492 | 0.680 | 0.450 | +--------+-------+-------+--------+-----------+-------+ | mAP | | | | | 0.560 | +--------+-------+-------+--------+-----------+-------+
Reppoints minmax Mask	In [1]: kitti_eval_base(result_record, data_selected) +--------+-------+-------+--------+-----------+-------+ | class | gts | dets | recall | precision | ap | +--------+-------+-------+--------+-----------+-------+ | Car | 20825 | 20004 | 0.748 | 0.779 | 0.728 | | Person | 7652 | 5797 | 0.565 | 0.745 | 0.527 | +--------+-------+-------+--------+-----------+-------+ | mAP | | | | | 0.628 | +--------+-------+-------+--------+-----------+-------+

```
evaluating result of  refer
+------------+-------+-------+--------+-----------+-------+
| class      | gts   | dets  | recall | precision | ap    |
+------------+-------+-------+--------+-----------+-------+
| Vehicle    | 20825 | 20625 | 0.725  | 0.732     | 0.702 |
| Pedestrian | 7652  | 7923  | 0.637  | 0.615     | 0.568 |
+------------+-------+-------+--------+-----------+-------+
| mAP        |       |       |        |           | 0.635 |
+------------+-------+-------+--------+-----------+-------+
/home/ld/RepPoints/ld_result/stsn_minmax_waymo/epoch_97_thres0.3_nms0.5_-1
```

```
evaluating result of  refer
+------------+-------+-------+--------+-----------+-------+
| class      | gts   | dets  | recall | precision | ap    |
+------------+-------+-------+--------+-----------+-------+
| Vehicle    | 20825 | 20629 | 0.725  | 0.732     | 0.702 |
| Pedestrian | 7652  | 7868  | 0.635  | 0.618     | 0.566 |
+------------+-------+-------+--------+-----------+-------+
| mAP        |       |       |        |           | 0.634 |
+------------+-------+-------+--------+-----------+-------+
/home/ld/RepPoints/ld_result/stsn_minmax_waymo/epoch_97_thres0.3_nms0.5_-5
```

```
evaluating result of  refer
+------------+-------+-------+--------+-----------+-------+
| class      | gts   | dets  | recall | precision | ap    |
+------------+-------+-------+--------+-----------+-------+
| Vehicle    | 20825 | 20647 | 0.725  | 0.731     | 0.702 |
| Pedestrian | 7652  | 7854  | 0.636  | 0.620     | 0.567 |
+------------+-------+-------+--------+-----------+-------+
| mAP        |       |       |        |           | 0.635 |
+------------+-------+-------+--------+-----------+-------+
/home/ld/RepPoints/ld_result/stsn_minmax_waymo/epoch_97_thres0.3_nms0.5_-10
```