Neural networks:

The used DCN (Detection Classification Networks) models must not be real-time capable. The emphasis here is on the detection and classification performance of the models of the ground objects in the list of 18 classes

1. Deformable DETR: Deformable Transformers for end-to-end Object Detection

- **Papers with code:** https://paperswithcode.com/paper/deformable-detr-deformable-transformers-for-1
- **Deformable DETR Github:** https://github.com/fundamentalvision/Deformable-DETR
- **Paper:** https://arxiv.org/pdf/2010.04159.pdf
- **Results (COCO validation set, 2017):**

Deformable DETR (single scale):

Test: Total time: 0:06:05 (0.1464 s / it)
Averaged stats: class_error: 0.00 loss: 6.2165 (6.9726) loss_ce: 0.3699 (0.4111) loss_bbox: 0.2184 (0.2248) loss_giou: 0.4874 (0.4804) loss_ce_0: 0.4391 (0.4977) loss_bbox_0: 0.2403 (0.2575) loss_giou_0: 0.4890 (0.5311) loss_ce_1: 0.3887 (0.4480) loss_bbox_1: 0.2509 (0.2359) loss_giou_1: 0.4687 (0.4983) loss_ce_2: 0.3979 (0.4262) loss_bbox_2: 0.2277 (0.2296) loss_giou_2: 0.4776 (0.4883) loss_ce_3: 0.3749 (0.4165) loss_bbox_3: 0.2427 (0.2263) loss_giou_3: 0.4879 (0.4835) loss_ce_4: 0.3774 (0.4103) loss_bbox_4: 0.2177 (0.2259) loss_giou_4: 0.4834 (0.4813) loss_ce_unscaled: 0.1850 (0.2055) class_error_unscaled: 4.0000 (9.4654) loss_bbox_unscaled: 0.0437 (0.0450) loss_giou_unscaled: 0.2437 (0.2402) cardinality_error_unscaled: 293.5000 (291.9334) loss_ce_0_unscaled: 0.2195 (0.2488) loss_bbox_0_unscaled: 0.0481 (0.0515) loss_giou_0_unscaled: 0.2445 (0.2656) cardinality_error_0_unscaled: 292.5000 (292.3278) loss_ce_1_unscaled: 0.1944 (0.2240) loss_bbox_1_unscaled: 0.0502 (0.0472) loss_giou_1_unscaled: 0.2343 (0.2492) cardinality_error_1_unscaled: 293.5000 (292.1338) loss_ce_2_unscaled: 0.1990 (0.2131) loss_bbox_2_unscaled: 0.0455 (0.0459) loss_giou_2_unscaled: 0.2388 (0.2442) cardinality_error_2_unscaled: 293.5000 (292.2482) loss_ce_3_unscaled: 0.1875 (0.2082) loss_bbox_3_unscaled: 0.0485 (0.0453) loss_giou_3_unscaled: 0.2440 (0.2417) cardinality_error_3_unscaled: 293.5000 (292.2414) loss_ce_4_unscaled: 0.1887 (0.2052) loss_bbox_4_unscaled: 0.0435 (0.0452) loss_giou_4_unscaled: 0.2417 (0.2406) cardinality_error_4_unscaled: 293.5000 (292.0508)
Accumulating evaluation results...
DONE (t=9.70s).
IoU metric: bbox
Average Precision (AP) @[IoU=0.50:0.95 | area= all | maxDets=100] = 0.394
Average Precision (AP) @[IoU=0.50 | area= all | maxDets=100] = 0.597
Average Precision (AP) @[IoU=0.75 | area= all | maxDets=100] = 0.422
Average Precision (AP) @[IoU=0.50:0.95 | area= small | maxDets=100] = 0.207
Average Precision (AP) @[IoU=0.50:0.95 | area=medium | maxDets=100] = 0.430
Average Precision (AP) @[IoU=0.50:0.95 | area= large | maxDets=100] = 0.559
Average Recall (AR) @[IoU=0.50:0.95 | area= all | maxDets= 1] = 0.326
Average Recall (AR) @[IoU=0.50:0.95 | area= all | maxDets= 10] = 0.534
Average Recall (AR) @[IoU=0.50:0.95 | area= all | maxDets=100] = 0.571
Average Recall (AR) @[IoU=0.50:0.95 | area= small | maxDets=100] = 0.328

Average Recall (AR) @[IoU=0.50:0.95 | area=medium | maxDets=100] = 0.624
Average Recall (AR) @[IoU=0.50:0.95 | area= large | maxDets=100] = 0.800

Deformable DETR (single scale, DC5):

Test: Total time: 0:07:14 (0.1738 s / it)
Averaged stats: class_error: 0.00 loss: 6.2433 (6.6677) loss_ce: 0.3569 (0.4032) loss_bbox: 0.2149 (0.2195) loss_giou: 0.4354 (0.4505) loss_ce_0: 0.4772 (0.4924) loss_bbox_0: 0.2385 (0.2409) loss_giou_0: 0.4216 (0.4931) loss_ce_1: 0.4343 (0.4318) loss_bbox_1: 0.2544 (0.2253) loss_giou_1: 0.4325 (0.4620) loss_ce_2: 0.3789 (0.4119) loss_bbox_2: 0.2256 (0.2235) loss_giou_2: 0.4132 (0.4571) loss_ce_3: 0.3626 (0.4072) loss_bbox_3: 0.2110 (0.2206) loss_giou_3: 0.4484 (0.4533) loss_ce_4: 0.3565 (0.4049) loss_bbox_4: 0.2194 (0.2199) loss_giou_4: 0.4374 (0.4508) loss_ce_unscaled: 0.1785 (0.2016) class_error_unscaled: 4.0000 (9.0135) loss_bbox_unscaled: 0.0430 (0.0439) loss_giou_unscaled: 0.2177 (0.2252) cardinality_error_unscaled: 293.5000 (292.0380) loss_ce_0_unscaled: 0.2386 (0.2462) loss_bbox_0_unscaled: 0.0477 (0.0482) loss_giou_0_unscaled: 0.2108 (0.2465) cardinality_error_0_unscaled: 293.5000 (292.2852) loss_ce_1_unscaled: 0.2171 (0.2159) loss_bbox_1_unscaled: 0.0509 (0.0451) loss_giou_1_unscaled: 0.2162 (0.2310) cardinality_error_1_unscaled: 293.5000 (292.0064) loss_ce_2_unscaled: 0.1894 (0.2059) loss_bbox_2_unscaled: 0.0451 (0.0447) loss_giou_2_unscaled: 0.2066 (0.2285) cardinality_error_2_unscaled: 293.5000 (291.9220) loss_ce_3_unscaled: 0.1813 (0.2036) loss_bbox_3_unscaled: 0.0422 (0.0441) loss_giou_3_unscaled: 0.2242 (0.2266) cardinality_error_3_unscaled: 293.5000 (291.9970) loss_ce_4_unscaled: 0.1782 (0.2024) loss_bbox_4_unscaled: 0.0439 (0.0440) loss_giou_4_unscaled: 0.2187 (0.2254) cardinality_error_4_unscaled: 293.5000 (291.9752)
Accumulating evaluation results...
DONE (t=9.27s).
IoU metric: bbox
Average Precision (AP) @[IoU=0.50:0.95 | area= all | maxDets=100] = 0.414
Average Precision (AP) @[IoU=0.50 | area= all | maxDets=100] = 0.618
Average Precision (AP) @[IoU=0.75 | area= all | maxDets=100] = 0.449
Average Precision (AP) @[IoU=0.50:0.95 | area= small | maxDets=100] = 0.237
Average Precision (AP) @[IoU=0.50:0.95 | area=medium | maxDets=100] = 0.453
Average Precision (AP) @[IoU=0.50:0.95 | area= large | maxDets=100] = 0.560
Average Recall (AR) @[IoU=0.50:0.95 | area= all | maxDets= 1] = 0.340
Average Recall (AR) @[IoU=0.50:0.95 | area= all | maxDets= 10] = 0.556
Average Recall (AR) @[IoU=0.50:0.95 | area= all | maxDets=100] = 0.595
Average Recall (AR) @[IoU=0.50:0.95 | area= small | maxDets=100] = 0.373
Average Recall (AR) @[IoU=0.50:0.95 | area=medium | maxDets=100] = 0.646
Average Recall (AR) @[IoU=0.50:0.95 | area= large | maxDets=100] = 0.803

Deformable DETR:

Test: Total time: 0:11:01 (0.2645 s / it)
Averaged stats: class_error: 0.00 loss: 5.8611 (6.2284) loss_ce: 0.3304 (0.3914) loss_bbox: 0.1845 (0.2033) loss_giou: 0.3440 (0.4123) loss_ce_0: 0.4236 (0.4710) loss_bbox_0: 0.2006 (0.2192) loss_giou_0: 0.3595 (0.4394) loss_ce_1: 0.3551 (0.4228) loss_bbox_1: 0.1916 (0.2070) loss_giou_1: 0.3454 (0.4193) loss_ce_2: 0.3260 (0.4026) loss_bbox_2: 0.2022 (0.2053) loss_giou_2: 0.3509 (0.4147) loss_ce_3: 0.3647 (0.3954) loss_bbox_3: 0.1839 (0.2037) loss_giou_3: 0.3445 (0.4129) loss_ce_4: 0.3290 (0.3924) loss_bbox_4: 0.1871 (0.2030) loss_giou_4: 0.3449 (0.4127) loss_ce_unscaled: 0.1652 (0.1957) class_error_unscaled: 4.7619

(8.7249) loss_bbox_unscaled: 0.0369 (0.0407) loss_giou_unscaled: 0.1720 (0.2061) cardinality_error_unscaled: 293.5000 (292.1010) loss_ce_0_unscaled: 0.2118 (0.2355) loss_bbox_0_unscaled: 0.0401 (0.0438) loss_giou_0_unscaled: 0.1798 (0.2197) cardinality_error_0_unscaled: 293.5000 (292.3060) loss_ce_1_unscaled: 0.1776 (0.2114) loss_bbox_1_unscaled: 0.0383 (0.0414) loss_giou_1_unscaled: 0.1727 (0.2096) cardinality_error_1_unscaled: 293.5000 (292.2054) loss_ce_2_unscaled: 0.1630 (0.2013) loss_bbox_2_unscaled: 0.0404 (0.0411) loss_giou_2_unscaled: 0.1754 (0.2074) cardinality_error_2_unscaled: 293.5000 (292.1714) loss_ce_3_unscaled: 0.1823 (0.1977) loss_bbox_3_unscaled: 0.0368 (0.0407) loss_giou_3_unscaled: 0.1723 (0.2064) cardinality_error_3_unscaled: 293.5000 (292.1456) loss_ce_4_unscaled: 0.1645 (0.1962) loss_bbox_4_unscaled: 0.0374 (0.0406) loss_giou_4_unscaled: 0.1724 (0.2064) cardinality_error_4_unscaled: 293.5000 (292.1110)
Accumulating evaluation results...
DONE (t=9.55s).
IoU metric: bbox
Average Precision (AP) @[IoU=0.50:0.95 | area= all | maxDets=100] = 0.445
Average Precision (AP) @[IoU=0.50 | area= all | maxDets=100] = 0.635
Average Precision (AP) @[IoU=0.75 | area= all | maxDets=100] = 0.487
Average Precision (AP) @[IoU=0.50:0.95 | area= small | maxDets=100] = 0.268
Average Precision (AP) @[IoU=0.50:0.95 | area=medium | maxDets=100] = 0.477
Average Precision (AP) @[IoU=0.50:0.95 | area= large | maxDets=100] = 0.595
Average Recall (AR) @[IoU=0.50:0.95 | area= all | maxDets= 1] = 0.353
Average Recall (AR) @[IoU=0.50:0.95 | area= all | maxDets= 10] = 0.587
Average Recall (AR) @[IoU=0.50:0.95 | area= all | maxDets=100] = 0.629
Average Recall (AR) @[IoU=0.50:0.95 | area= small | maxDets=100] = 0.416
Average Recall (AR) @[IoU=0.50:0.95 | area=medium | maxDets=100] = 0.673
Average Recall (AR) @[IoU=0.50:0.95 | area= large | maxDets=100] = 0.819

Deformable DETR + iterative bounding box refinement:

Averaged stats: class_error: 0.00 loss: 6.2634 (6.3376) loss_ce: 0.3980 (0.3963) loss_bbox: 0.2170 (0.2078) loss_giou: 0.3444 (0.4101) loss_ce_0: 0.3935 (0.4610) loss_bbox_0: 0.2571 (0.2464) loss_giou_0: 0.4184 (0.4832) loss_ce_1: 0.3858 (0.4316) loss_bbox_1: 0.2158 (0.2113) loss_giou_1: 0.3547 (0.4210) loss_ce_2: 0.3943 (0.4139) loss_bbox_2: 0.2185 (0.2072) loss_giou_2: 0.3471 (0.4121) loss_ce_3: 0.4044 (0.4021) loss_bbox_3: 0.2142 (0.2076) loss_giou_3: 0.3472 (0.4109) loss_ce_4: 0.3956 (0.3962) loss_bbox_4: 0.2170 (0.2081) loss_giou_4: 0.3443 (0.4105) loss_ce_unscaled: 0.1990 (0.1982) class_error_unscaled: 0.0000 (7.5326) loss_bbox_unscaled: 0.0434 (0.0416) loss_giou_unscaled: 0.1722 (0.2051) cardinality_error_unscaled: 293.5000 (292.1730) loss_ce_0_unscaled: 0.1967 (0.2305) loss_bbox_0_unscaled: 0.0514 (0.0493) loss_giou_0_unscaled: 0.2092 (0.2416) cardinality_error_0_unscaled: 293.5000 (292.2916) loss_ce_1_unscaled: 0.1929 (0.2158) loss_bbox_1_unscaled: 0.0432 (0.0423) loss_giou_1_unscaled: 0.1773 (0.2105) cardinality_error_1_unscaled: 293.5000 (292.2470) loss_ce_2_unscaled: 0.1971 (0.2069) loss_bbox_2_unscaled: 0.0437 (0.0414) loss_giou_2_unscaled: 0.1736 (0.2060) cardinality_error_2_unscaled: 293.5000 (292.2042) loss_ce_3_unscaled: 0.2022 (0.2011) loss_bbox_3_unscaled: 0.0428 (0.0415) loss_giou_3_unscaled: 0.1736 (0.2055) cardinality_error_3_unscaled: 293.5000 (291.8936) loss_ce_4_unscaled: 0.1978 (0.1981) loss_bbox_4_unscaled: 0.0434 (0.0416) loss_giou_4_unscaled: 0.1722 (0.2053) cardinality_error_4_unscaled: 293.5000 (292.1126)
Accumulating evaluation results...
DONE (t=9.51s).
IoU metric: bbox

Average Precision (AP) @[IoU=0.50:0.95 | area= all | maxDets=100] = 0.463
Average Precision (AP) @[IoU=0.50 | area= all | maxDets=100] = 0.650
Average Precision (AP) @[IoU=0.75 | area= all | maxDets=100] = 0.501
Average Precision (AP) @[IoU=0.50:0.95 | area= small | maxDets=100] = 0.285
Average Precision (AP) @[IoU=0.50:0.95 | area=medium | maxDets=100] = 0.492
Average Precision (AP) @[IoU=0.50:0.95 | area= large | maxDets=100] = 0.615
Average Recall (AR) @[IoU=0.50:0.95 | area= all | maxDets= 1] = 0.365
Average Recall (AR) @[IoU=0.50:0.95 | area= all | maxDets= 10] = 0.598
Average Recall (AR) @[IoU=0.50:0.95 | area= all | maxDets=100] = 0.640
Average Recall (AR) @[IoU=0.50:0.95 | area= small | maxDets=100] = 0.430
Average Recall (AR) @[IoU=0.50:0.95 | area=medium | maxDets=100] = 0.685
Average Recall (AR) @[IoU=0.50:0.95 | area= large | maxDets=100] = 0.831

Deformable DETR + iterative bounding box refinement ++ two-stage Deformable DETR:

Test: Total time: 0:12:22 (0.2972 s / it)
Averaged stats: class_error: 0.00 loss: 6.6696 (7.0185) loss_ce: 0.3309 (0.3976) loss_bbox: 0.1784 (0.1943) loss_giou: 0.3366 (0.3688) loss_ce_0: 0.5190 (0.5896) loss_bbox_0: 0.1826 (0.1827) loss_giou_0: 0.3699 (0.3607) loss_ce_1: 0.4045 (0.4659) loss_bbox_1: 0.1723 (0.1887) loss_giou_1: 0.3437 (0.3646) loss_ce_2: 0.3578 (0.4218) loss_bbox_2: 0.1786 (0.1932) loss_giou_2: 0.3422 (0.3686) loss_ce_3: 0.3669 (0.4083) loss_bbox_3: 0.1822 (0.1949) loss_giou_3: 0.3368 (0.3702) loss_ce_4: 0.3366 (0.3995) loss_bbox_4: 0.1850 (0.1933) loss_giou_4: 0.3367 (0.3680) loss_ce_enc: 0.3916 (0.4304) loss_bbox_enc: 0.1863 (0.1913) loss_giou_enc: 0.3760 (0.3663) loss_ce_unscaled: 0.1655 (0.1988) class_error_unscaled: 0.0000 (7.0752) loss_bbox_unscaled: 0.0357 (0.0389) loss_giou_unscaled: 0.1683 (0.1844) cardinality_error_unscaled: 293.5000 (292.0032) loss_ce_0_unscaled: 0.2595 (0.2948) loss_bbox_0_unscaled: 0.0365 (0.0365) loss_giou_0_unscaled: 0.1849 (0.1803) cardinality_error_0_unscaled: 293.5000 (291.7504) loss_ce_1_unscaled: 0.2023 (0.2330) loss_bbox_1_unscaled: 0.0345 (0.0377) loss_giou_1_unscaled: 0.1719 (0.1823) cardinality_error_1_unscaled: 293.5000 (291.6832) loss_ce_2_unscaled: 0.1789 (0.2109) loss_bbox_2_unscaled: 0.0357 (0.0386) loss_giou_2_unscaled: 0.1711 (0.1843) cardinality_error_2_unscaled: 293.5000 (291.7440) loss_ce_3_unscaled: 0.1834 (0.2041) loss_bbox_3_unscaled: 0.0364 (0.0390) loss_giou_3_unscaled: 0.1684 (0.1851) cardinality_error_3_unscaled: 293.5000 (291.7824) loss_ce_4_unscaled: 0.1683 (0.1997) loss_bbox_4_unscaled: 0.0370 (0.0387) loss_giou_4_unscaled: 0.1684 (0.1840) cardinality_error_4_unscaled: 293.5000 (291.9870) loss_ce_enc_unscaled: 0.1958 (0.2152) loss_bbox_enc_unscaled: 0.0373 (0.0383) loss_giou_enc_unscaled: 0.1880 (0.1831) cardinality_error_enc_unscaled: 20091.0000 (22116.8064)
Accumulating evaluation results...
DONE (t=8.28s).
IoU metric: bbox
Average Precision (AP) @[IoU=0.50:0.95 | area= all | maxDets=100] = 0.469
Average Precision (AP) @[IoU=0.50 | area= all | maxDets=100] = 0.657
Average Precision (AP) @[IoU=0.75 | area= all | maxDets=100] = 0.511
Average Precision (AP) @[IoU=0.50:0.95 | area= small | maxDets=100] = 0.296
Average Precision (AP) @[IoU=0.50:0.95 | area=medium | maxDets=100] = 0.503
Average Precision (AP) @[IoU=0.50:0.95 | area= large | maxDets=100] = 0.616
Average Recall (AR) @[IoU=0.50:0.95 | area= all | maxDets= 1] = 0.363
Average Recall (AR) @[IoU=0.50:0.95 | area= all | maxDets= 10] = 0.610
Average Recall (AR) @[IoU=0.50:0.95 | area= all | maxDets=100] = 0.659
Average Recall (AR) @[IoU=0.50:0.95 | area= small | maxDets=100] = 0.460
Average Recall (AR) @[IoU=0.50:0.95 | area=medium | maxDets=100] = 0.708

Average Recall (AR) @[IoU=0.50:0.95 | area= large | maxDets=100] = 0.832

2. **InternImage: Exploring Large-Scale Vision Foundation Models with Deformable Convolutions**

   - **Papers with code (Object Detection):** https://paperswithcode.com/paper/internimage-exploring-large-scale-vision
   - **InternImage github:** https://github.com/opengvlab/internimage
   - **Paper:** https://arxiv.org/pdf/2211.05778v4.pdf
   - **Results:**

Backbone: InternImage-B / Method: Mask R-CNN / schd: 3x

load checkpoint from local path: ./checkpoints/mask_rcnn_internimage_b_fpn_3x_coco.pth
[>>] 5000/5000, 3.2 task/s, elapsed: 1573s, ETA: 0s
Evaluating bbox...
Loading and preparing results...
DONE (t=0.38s)
creating index...
index created!
Running per image evaluation...
Evaluate annotation type *bbox*
DONE (t=18.18s).
Accumulating evaluation results...
DONE (t=2.95s).

```
 Average Precision  (AP) @[ IoU=0.50:0.95 | area=   all | maxDets=100 ] = 0.503
 Average Precision  (AP) @[ IoU=0.50      | area=   all | maxDets=1000 ] = 0.714
 Average Precision  (AP) @[ IoU=0.75      | area=   all | maxDets=1000 ] = 0.553
 Average Precision  (AP) @[ IoU=0.50:0.95 | area= small | maxDets=1000 ] = 0.353
 Average Precision  (AP) @[ IoU=0.50:0.95 | area=medium | maxDets=1000 ] = 0.535
 Average Precision  (AP) @[ IoU=0.50:0.95 | area= large | maxDets=1000 ] = 0.646
 Average Recall     (AR) @[ IoU=0.50:0.95 | area=   all | maxDets=100 ] = 0.620
 Average Recall     (AR) @[ IoU=0.50:0.95 | area=   all | maxDets=300 ] = 0.620
 Average Recall     (AR) @[ IoU=0.50:0.95 | area=   all | maxDets=1000 ] = 0.620
 Average Recall     (AR) @[ IoU=0.50:0.95 | area= small | maxDets=1000 ] = 0.463
 Average Recall     (AR) @[ IoU=0.50:0.95 | area=medium | maxDets=1000 ] = 0.653
 Average Recall     (AR) @[ IoU=0.50:0.95 | area= large | maxDets=1000 ] = 0.765
```

category	AP	category	AP	category	AP
person	0.6	**bicycle**	0.4	**car**	0.51
motorcycle	0.52	**airplane**	0.71	**bus**	0.73
train	0.71	**truck**	0.45	**boat**	0.35
traffic light	0.32	**fire hydrant**	0.74	**stop sign**	0.72
parking meter	0.54	**bench**	0.34	**bird**	0.43
cat	0.75	**dog**	0.71	**horse**	0.66
sheep	0.61	**cow**	0.64	**elephant**	0.72
bear	0.8	**zebra**	0.69	**giraffe**	0.73

backpack	0.24	**umbrella**	0.47	**handbag**	0.26
tie	0.44	**suitcase**	0.49	**frisbee**	0.73
skis	0.33	**snowboard**	0.5	**sports ball**	0.49
kite	0.48	**baseball bat**	0.46	**baseball glove**	0.45
skateboard	0.61	**surfboard**	0.49	**tennis racket**	0.61
bottle	0.48	**wine glass**	0.45	**cup**	0.52
fork	0.52	**knife**	0.34	**spoon**	0.3
bowl	0.5	**banana**	0.27	**apple**	0.29
sandwich	0.44	**orange**	0.37	**broccoli**	0.25
carrot	0.27	**hot dog**	0.45	**pizza**	0.56
donut	0.56	**cake**	0.47	**chair**	0.38
couch	0.5	**potted plant**	0.36	**bed**	0.5
dining table	0.34	**toilet**	0.66	**tv**	0.64
laptop	0.7	**mouse**	0.68	**remote**	0.47
keyboard	0.58	**cell phone**	0.46	**microwave**	0.63
oven	0.42	**toaster**	0.54	**sink**	0.43
refrigerator	0.66	**book**	0.21	**clock**	0.51
vase	0.43	**scissors**	0.47	**teddy bear**	0.56
hair drier	0.24	**toothbrush**	0.4	**None**	None

Backbone: InternImage-L / Method: Cascade / schd: 3x

load checkpoint from local path: ./checkpoints/cascade_internimage_l_fpn_3x_coco.pth
[>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>] 5000/5000, 1.4 task/s, elapsed: 3619s, ETA: 0s
Evaluating bbox...
Loading and preparing results...
DONE (t=0.49s)
creating index...
index created!
Running per image evaluation...
Evaluate annotation type *bbox*
DONE (t=19.77s).
Accumulating evaluation results...
DONE (t=3.17s).

```
 Average Precision  (AP) @[ IoU=0.50:0.95 | area=   all | maxDets=100 ] = 0.561
 Average Precision  (AP) @[ IoU=0.50      | area=   all | maxDets=1000 ] = 0.748
 Average Precision  (AP) @[ IoU=0.75      | area=   all | maxDets=1000 ] = 0.607
 Average Precision  (AP) @[ IoU=0.50:0.95 | area= small | maxDets=1000 ] = 0.398
 Average Precision  (AP) @[ IoU=0.50:0.95 | area=medium | maxDets=1000 ] = 0.606
 Average Precision  (AP) @[ IoU=0.50:0.95 | area= large | maxDets=1000 ] = 0.720
 Average Recall     (AR) @[ IoU=0.50:0.95 | area=   all | maxDets=100 ] = 0.685
```

Average Recall (AR) @[IoU=0.50:0.95 | area= all | maxDets=300] = 0.685
Average Recall (AR) @[IoU=0.50:0.95 | area= all | maxDets=1000] = 0.685
Average Recall (AR) @[IoU=0.50:0.95 | area= small | maxDets=1000] = 0.525
Average Recall (AR) @[IoU=0.50:0.95 | area=medium | maxDets=1000] = 0.726
Average Recall (AR) @[IoU=0.50:0.95 | area= large | maxDets=1000] = 0.827

category	AP	category	AP	category	AP
person	0.65	bicycle	0.45	car	0.56
motorcycle	0.57	airplane	0.76	bus	0.76
train	0.77	truck	0.5	boat	0.4
traffic light	0.34	fire hydrant	0.77	stop sign	0.75
parking meter	0.58	bench	0.39	bird	0.48
cat	0.82	dog	0.78	horse	0.7
sheep	0.65	cow	0.69	elephant	0.78
bear	0.8	zebra	0.76	giraffe	0.78
backpack	0.29	umbrella	0.55	handbag	0.3
tie	0.5	suitcase	0.57	frisbee	0.78
skis	0.39	snowboard	0.54	sports ball	0.53
kite	0.54	baseball bat	0.55	baseball glove	0.51
skateboard	0.66	surfboard	0.56	tennis racket	0.66
bottle	0.53	wine glass	0.49	cup	0.58
fork	0.6	knife	0.41	spoon	0.39
bowl	0.55	banana	0.35	apple	0.34
sandwich	0.51	orange	0.39	broccoli	0.3
carrot	0.33	hot dog	0.53	pizza	0.63
donut	0.64	cake	0.52	chair	0.44
couch	0.53	potted plant	0.4	bed	0.59
dining table	0.39	toilet	0.72	tv	0.69
laptop	0.78	mouse	0.69	remote	0.54
keyboard	0.64	cell phone	0.55	microwave	0.72
oven	0.48	toaster	0.57	sink	0.51
refrigerator	0.77	book	0.25	clock	0.58
vase	0.47	scissors	0.53	teddy bear	0.65
hair drier	0.36	toothbrush	0.49	None	None

Backbone: InternImage-XL / Method: Cascade / schd: 3x

load checkpoint from local path: ./checkpoints/cascade_internimage_xl_fpn_3x_coco.pth
[>>] 5000/5000, 1.0 task/s, elapsed: 4800s, ETA: 0s
Evaluating bbox...
Loading and preparing results...
DONE (t=0.37s)
creating index...
index created!
Running per image evaluation...

Evaluate annotation type *bbox*
DONE (t=18.20s).
Accumulating evaluation results...
DONE (t=2.69s).

```
 Average Precision  (AP) @[ IoU=0.50:0.95 | area=   all | maxDets=100 ] = 0.562
 Average Precision  (AP) @[ IoU=0.50      | area=   all | maxDets=1000 ] = 0.750
 Average Precision  (AP) @[ IoU=0.75      | area=   all | maxDets=1000 ] = 0.612
 Average Precision  (AP) @[ IoU=0.50:0.95 | area= small | maxDets=1000 ] = 0.401
 Average Precision  (AP) @[ IoU=0.50:0.95 | area=medium | maxDets=1000 ] = 0.605
 Average Precision  (AP) @[ IoU=0.50:0.95 | area= large | maxDets=1000 ] = 0.726
 Average Recall     (AR) @[ IoU=0.50:0.95 | area=   all | maxDets=100 ] = 0.677
 Average Recall     (AR) @[ IoU=0.50:0.95 | area=   all | maxDets=300 ] = 0.677
 Average Recall     (AR) @[ IoU=0.50:0.95 | area=   all | maxDets=1000 ] = 0.677
 Average Recall     (AR) @[ IoU=0.50:0.95 | area= small | maxDets=1000 ] = 0.518
 Average Recall     (AR) @[ IoU=0.50:0.95 | area=medium | maxDets=1000 ] = 0.718
 Average Recall     (AR) @[ IoU=0.50:0.95 | area= large | maxDets=1000 ] = 0.823
```

category	AP	category	AP	category	AP
person	0.65	**bicycle**	0.44	**car**	0.56
motorcycle	0.56	**airplane**	0.76	**bus**	0.77
train	0.78	**truck**	0.5	**boat**	0.39
traffic light	0.35	**fire hydrant**	0.79	**stop sign**	0.76
parking meter	0.6	**bench**	0.39	**bird**	0.48
cat	0.82	**dog**	0.8	**horse**	0.7
sheep	0.65	**cow**	0.7	**elephant**	0.77
bear	0.81	**zebra**	0.76	**giraffe**	0.78
backpack	0.29	**umbrella**	0.54	**handbag**	0.3
tie	0.5	**suitcase**	0.57	**frisbee**	0.78
skis	0.4	**snowboard**	0.55	**sports ball**	0.54
kite	0.54	**baseball bat**	0.55	**baseball glove**	0.51
skateboard	0.66	**surfboard**	0.56	**tennis racket**	0.66
bottle	0.54	**wine glass**	0.5	**cup**	0.58
fork	0.61	**knife**	0.42	**spoon**	0.41
bowl	0.55	**banana**	0.35	**apple**	0.35
sandwich	0.52	**orange**	0.42	**broccoli**	0.3
carrot	0.33	**hot dog**	0.55	**pizza**	0.62
donut	0.64	**cake**	0.51	**chair**	0.44
couch	0.56	**potted plant**	0.4	**bed**	0.57
dining table	0.38	**toilet**	0.74	**tv**	0.69
laptop	0.77	**mouse**	0.7	**remote**	0.54
keyboard	0.65	**cell phone**	0.54	**microwave**	0.71
oven	0.48	**toaster**	0.46	**sink**	0.52
refrigerator	0.74	**book**	0.26	**clock**	0.58
vase	0.46	**scissors**	0.55	**teddy bear**	0.66
hair drier	0.36	**toothbrush**	0.48	**None**	None

3. **A Strong and Reproducible Object Detector with Only Public Datasets**

   - **Papers with code (Object Detection):** https://paperswithcode.com/paper/a-strong-and-reproducible-object-detector
   - **FocalNet github**: https://github.com/microsoft/FocalNet
   - **Paper:** https://arxiv.org/pdf/2304.13027v1.pdf
   - **Results:**

4. EVA: Exploring the Limits of Masked Visual Representation Learning at Scale

- **Papers with code (Object Detection):** https://paperswithcode.com/paper/eva-exploring-the-limits-of-masked-visual
- **EVA github:** https://github.com/baaivision/EVA/tree/master
- **Paper:** https://arxiv.org/pdf/2211.07636v2.pdf
- **Results:**

eva02_L_coco_bsl.pth

[06/01 18:42:19 d2.evaluation.evaluator]: Total inference time: 0:59:19.944545 (0.712702 s / iter per device, on 1 devices)
[06/01 18:42:19 d2.evaluation.evaluator]: Total inference pure compute time: 0:57:37 (0.692171 s / iter per device, on 1 devices)
[06/01 18:42:19 d2.evaluation.coco_evaluation]: Preparing results for COCO format ...
[06/01 18:42:19 d2.evaluation.coco_evaluation]: Evaluating predictions with unofficial COCO API...
Loading and preparing results...
DONE (t=0.06s)
creating index...
index created!
[06/01 18:42:19 d2.evaluation.fast_eval_api]: Evaluate annotation type *bbox*
[06/01 18:42:24 d2.evaluation.fast_eval_api]: COCOeval_opt.evaluate() finished in 4.57 seconds.
[06/01 18:42:24 d2.evaluation.fast_eval_api]: Accumulating evaluation results...
[06/01 18:42:24 d2.evaluation.fast_eval_api]: COCOeval_opt.accumulate() finished in 0.39 seconds.
Average Precision (AP) @[IoU=0.50:0.95 | area= all | maxDets=100] = 0.592
Average Precision (AP) @[IoU=0.50 | area= all | maxDets=100] = 0.787
Average Precision (AP) @[IoU=0.75 | area= all | maxDets=100] = 0.641
Average Precision (AP) @[IoU=0.50:0.95 | area= small | maxDets=100] = 0.419
Average Precision (AP) @[IoU=0.50:0.95 | area=medium | maxDets=100] = 0.645
Average Precision (AP) @[IoU=0.50:0.95 | area= large | maxDets=100] = 0.754
Average Recall (AR) @[IoU=0.50:0.95 | area= all | maxDets= 1] = 0.419
Average Recall (AR) @[IoU=0.50:0.95 | area= all | maxDets= 10] = 0.668
Average Recall (AR) @[IoU=0.50:0.95 | area= all | maxDets=100] = 0.696
Average Recall (AR) @[IoU=0.50:0.95 | area= small | maxDets=100] = 0.537
Average Recall (AR) @[IoU=0.50:0.95 | area=medium | maxDets=100] = 0.750
Average Recall (AR) @[IoU=0.50:0.95 | area= large | maxDets=100] = 0.847

[06/01 18:42:24 d2.evaluation.coco_evaluation]: Evaluation results for bbox:

AP	AP50	AP75	APs	APm	APl
59.165	78.66	64.1	41.92	64.46	75.38

AP	AP50	AP75	APs	APm	APl
59.165	78.660	64.103	41.920	64.458	75.383

[06/01 18:42:24 d2.evaluation.coco_evaluation]: Per-category bbox AP:
category	AP	category	AP	category	AP
person	65.714	bicycle	48.665	car	55.813
motorcycle	60.242	airplane	78.599	bus	78.619
train	80.043	truck	55.784	boat	43.413
traffic light	34.975	fire hydrant	81.447	stop sign	74.934
parking meter	57.540	bench	45.907	bird	49.346
cat	81.536	dog	82.393	horse	75.051
sheep	66.680	cow	70.742	elephant	79.746
bear	85.807	zebra	78.713	giraffe	79.813
backpack	34.018	umbrella	57.545	handbag	34.688
tie	52.759	suitcase	60.696	frisbee	77.919
skis	42.007	snowboard	54.114	sports ball	57.150
kite	59.881	baseball bat	59.542	baseball glove	51.239
skateboard	69.340	surfboard	59.724	tennis racket	71.186
bottle	54.362	wine glass	50.757	cup	60.596
fork	62.290	knife	45.251	spoon	39.668
bowl	58.454	banana	36.902	apple	34.995
sandwich	57.465	orange	41.407	broccoli	32.191
carrot	34.524	hot dog	60.339	pizza	65.505
donut	65.346	cake	56.940	chair	49.325
couch	59.608	potted plant	42.383	bed	63.225
dining table	41.959	toilet	73.572	tv	71.349
laptop	80.525	mouse	72.490	remote	57.435
keyboard	66.265	cell phone	57.493	microwave	75.826
oven	50.109	toaster	55.742	sink	51.618
refrigerator	77.926	book	28.466	clock	61.042
vase	51.557	scissors	61.884	teddy bear	71.724
hair drier	44.387	toothbrush	56.985		

category	AP	category	AP	category	AP
person	65.714	**bicycle**	48.665	**car**	55.813
motorcycle	60.242	**airplane**	78.599	**bus**	78.619
train	80.043	**truck**	55.784	**boat**	43.413
traffic light	34.975	**fire hydrant**	81.447	**stop sign**	74.934
parking meter	57.540	**bench**	45.907	**bird**	49.346
cat	81.536	**dog**	82.393	**horse**	75.051
sheep	66.680	**cow**	70.742	**elephant**	79.746
bear	85.807	**zebra**	78.713	**giraffe**	79.813
backpack	34.018	**umbrella**	57.545	**handbag**	34.688
tie	52.759	**suitcase**	60.696	**frisbee**	77.919
skis	42.007	**snowboard**	54.114	**sports ball**	57.150
kite	59.881	**baseball bat**	59.542	**baseball glove**	51.239

skateboard	69.340	**surfboard**	59.724	**tennis racket**	71.186
bottle	54.362	**wine glass**	50.757	**cup**	60.596
fork	62.290	**knife**	45.251	**spoon**	39.668
bowl	58.454	**banana**	36.902	**apple**	34.995
sandwich	57.465	**orange**	41.407	**broccoli**	32.191
carrot	34.524	**hot dog**	60.339	**pizza**	65.505
donut	65.346	**cake**	56.940	**chair**	49.325
couch	59.608	**potted plant**	42.383	**bed**	63.225
dining table	41.959	**toilet**	73.572	**tv**	71.349
laptop	80.525	**mouse**	72.490	**remote**	57.435
keyboard	66.265	**cell phone**	57.493	**microwave**	75.826
oven	50.109	**toaster**	55.742	**sink**	51.618
refrigerator	77.926	**book**	28.466	**clock**	61.042
vase	51.557	**scissors**	61.884	**teddy bear**	71.724
hair drier	44.387	**toothbrush**	56.985		

eva02_L_coco_sys:

[06/02 12:07:13 d2.evaluation.evaluator]: Total inference time: 2:46:07.088074 (1.995413 s / iter per device, on 1 devices)
[06/02 12:07:13 d2.evaluation.evaluator]: Total inference pure compute time: 2:38:53 (1.908665 s / iter per device, on 1 devices)
[06/02 12:07:13 d2.evaluation.coco_evaluation]: Preparing results for COCO format ...
[06/02 12:07:13 d2.evaluation.coco_evaluation]: Evaluating predictions with unofficial COCO API...
Loading and preparing results...
DONE (t=0.43s)
creating index...
index created!
[06/02 12:07:14 d2.evaluation.fast_eval_api]: Evaluate annotation type *bbox*
[06/02 12:07:22 d2.evaluation.fast_eval_api]: COCOeval_opt.evaluate() finished in 7.98 seconds.
[06/02 12:07:22 d2.evaluation.fast_eval_api]: Accumulating evaluation results...
[06/02 12:07:23 d2.evaluation.fast_eval_api]: COCOeval_opt.accumulate() finished in 1.11 seconds.
Average Precision (AP) @[IoU=0.50:0.95 | area= all | maxDets=100] = 0.623
Average Precision (AP) @[IoU=0.50 | area= all | maxDets=100] = 0.808
Average Precision (AP) @[IoU=0.75 | area= all | maxDets=100] = 0.681
Average Precision (AP) @[IoU=0.50:0.95 | area= small | maxDets=100] = 0.459
Average Precision (AP) @[IoU=0.50:0.95 | area=medium | maxDets=100] = 0.667
Average Precision (AP) @[IoU=0.50:0.95 | area= large | maxDets=100] = 0.780
Average Recall (AR) @[IoU=0.50:0.95 | area= all | maxDets= 1] = 0.430
Average Recall (AR) @[IoU=0.50:0.95 | area= all | maxDets= 10] = 0.722
Average Recall (AR) @[IoU=0.50:0.95 | area= all | maxDets=100] = 0.783
Average Recall (AR) @[IoU=0.50:0.95 | area= small | maxDets=100] = 0.668

Average Recall (AR) @[IoU=0.50:0.95 | area=medium | maxDets=100] = 0.822
Average Recall (AR) @[IoU=0.50:0.95 | area= large | maxDets=100] = 0.903
[06/02 12:07:23 d2.evaluation.coco_evaluation]: Evaluation results for bbox:

AP	AP50	AP75	APs	APm	APl
62.286	80.8	68.1	45.86	66.74	78.01

AP	AP50	AP75	APs	APm	APl
62.286	80.801	68.096	45.862	66.739	78.005

[06/02 12:07:23 d2.evaluation.coco_evaluation]: Per-category bbox AP:

category	AP	category	AP	category	AP
person	68.951	**bicycle**	52.066	**car**	59.994
motorcycle	63.230	**airplane**	82.877	**bus**	81.049
train	82.718	**truck**	57.952	**boat**	48.034
traffic light	39.244	**fire hydrant**	83.396	**stop sign**	76.519
parking meter	59.545	**bench**	47.449	**bird**	53.605
cat	84.159	**dog**	83.152	**horse**	77.768
sheep	69.102	**cow**	74.783	**elephant**	80.875
bear	85.708	**zebra**	81.385	**giraffe**	82.911
backpack	37.248	**umbrella**	59.672	**handbag**	37.829
tie	57.438	**suitcase**	64.546	**frisbee**	80.495
skis	45.776	**snowboard**	56.836	**sports ball**	61.903
kite	62.345	**baseball bat**	68.087	**baseball glove**	55.997
skateboard	73.468	**surfboard**	63.282	**tennis racket**	74.084
bottle	56.771	**wine glass**	55.317	**cup**	63.423
fork	65.951	**knife**	50.055	**spoon**	45.232
bowl	59.871	**banana**	41.184	**apple**	36.768
sandwich	62.666	**orange**	44.128	**broccoli**	34.325
carrot	37.035	**hot dog**	66.393	**pizza**	67.521
donut	68.848	**cake**	60.778	**chair**	51.970
couch	62.814	**potted plant**	42.976	**bed**	65.272
dining table	44.396	**toilet**	76.765	**tv**	74.093
laptop	83.057	**mouse**	74.747	**remote**	61.502
keyboard	67.001	**cell phone**	62.054	**microwave**	76.657
oven	52.937	**toaster**	50.565	**sink**	54.551
refrigerator	79.490	**book**	32.335	**clock**	63.126

vase	54.493	**scissors**	71.137	**teddy bear**	75.430
hair drier	52.041	**toothbrush**	61.702		

category	AP	category	AP	category	AP
person	68.951	bicycle	52.066	car	59.994
motorcycle	63.230	airplane	82.877	bus	81.049
train	82.718	truck	57.952	boat	48.034
traffic light	39.244	fire hydrant	83.396	stop sign	76.519
parking meter	59.545	bench	47.449	bird	53.605
cat	84.159	dog	83.152	horse	77.768
sheep	69.102	cow	74.783	elephant	80.875
bear	85.708	zebra	81.385	giraffe	82.911
backpack	37.248	umbrella	59.672	handbag	37.829
tie	57.438	suitcase	64.546	frisbee	80.495
skis	45.776	snowboard	56.836	sports ball	61.903
kite	62.345	baseball bat	68.087	baseball glove	55.997
skateboard	73.468	surfboard	63.282	tennis racket	74.084
bottle	56.771	wine glass	55.317	cup	63.423
fork	65.951	knife	50.055	spoon	45.232
bowl	59.871	banana	41.184	apple	36.768
sandwich	62.666	orange	44.128	broccoli	34.325
carrot	37.035	hot dog	66.393	pizza	67.521
donut	68.848	cake	60.778	chair	51.970
couch	62.814	potted plant	42.976	bed	65.272
dining table	44.396	toilet	76.765	tv	74.093
laptop	83.057	mouse	74.747	remote	61.502
keyboard	67.001	cell phone	62.054	microwave	76.657
oven	52.937	toaster	50.565	sink	54.551
refrigerator	79.490	book	32.335	clock	63.126
vase	54.493	scissors	71.137	teddy bear	75.430
hair drier	52.041	toothbrush	61.702		