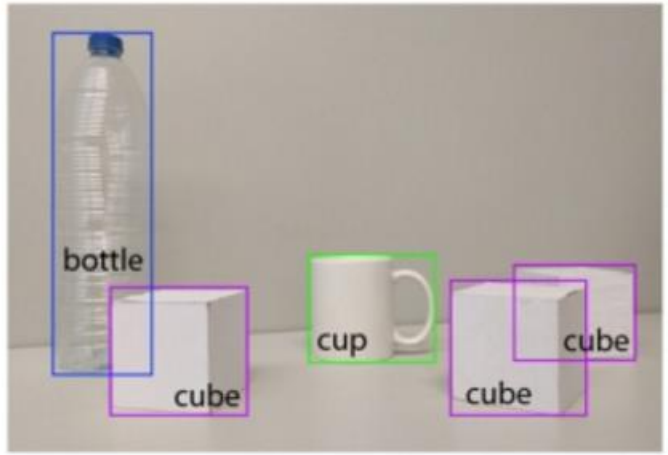
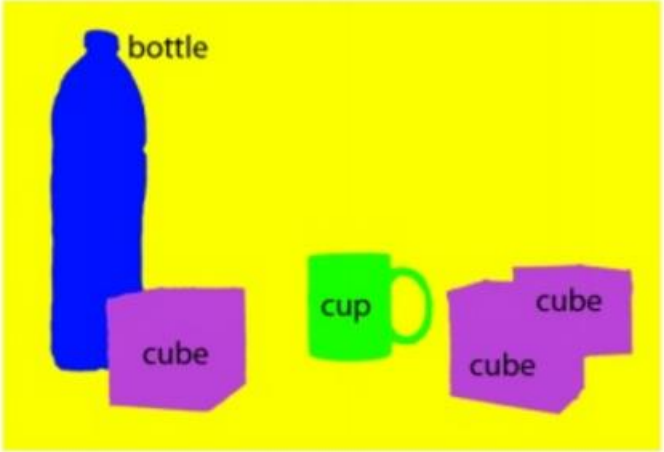
POPULAR DATASETS IN COMPUTER VISION

CONTENTS

Classification



Detection



Segmentation

1

Classification

Evaluation index

$$Accuracy = \frac{n_{correct}}{n_{total}}$$

percentage error=1-accuracy

MNIST

CIFAR 10

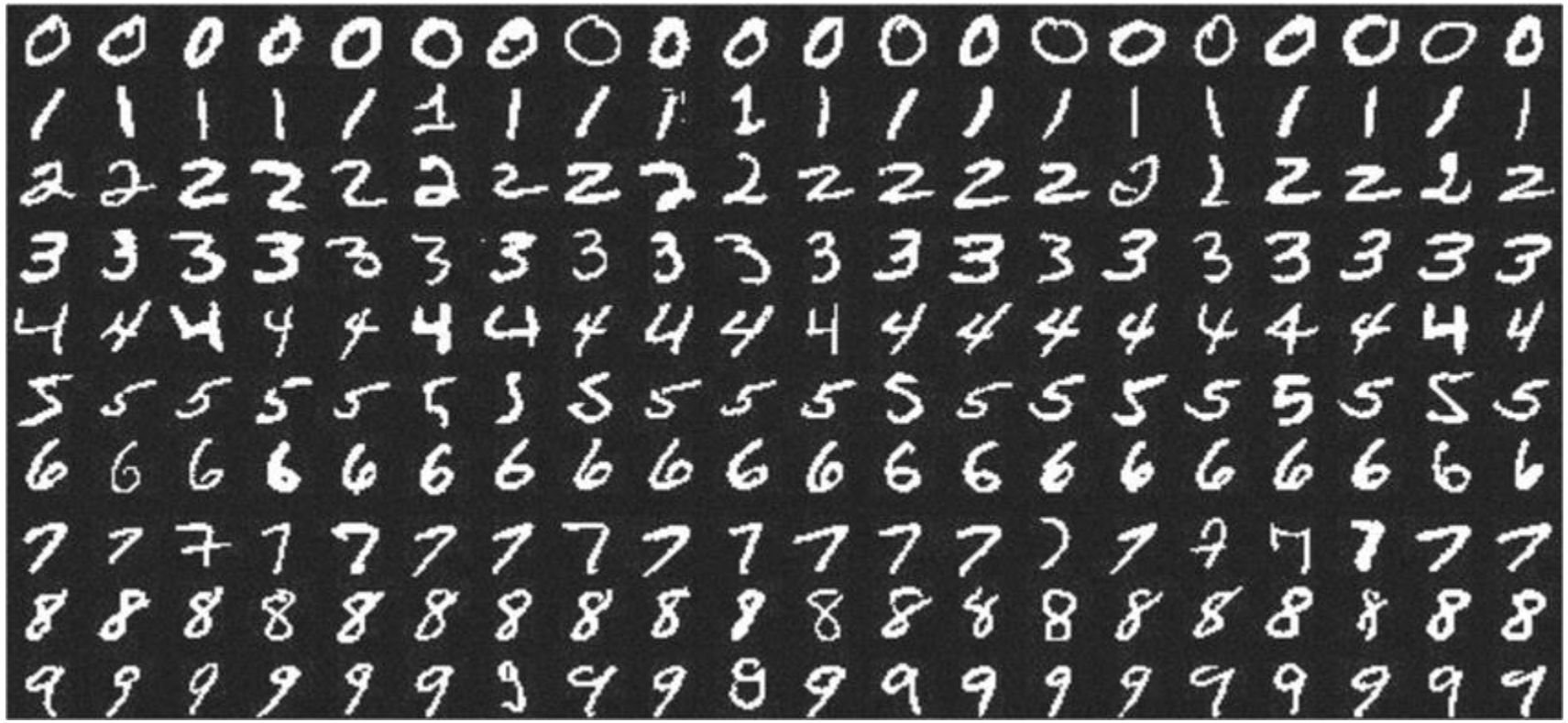
CIFAR 100

ILSVRC CLS-LOC

Places

MNIST

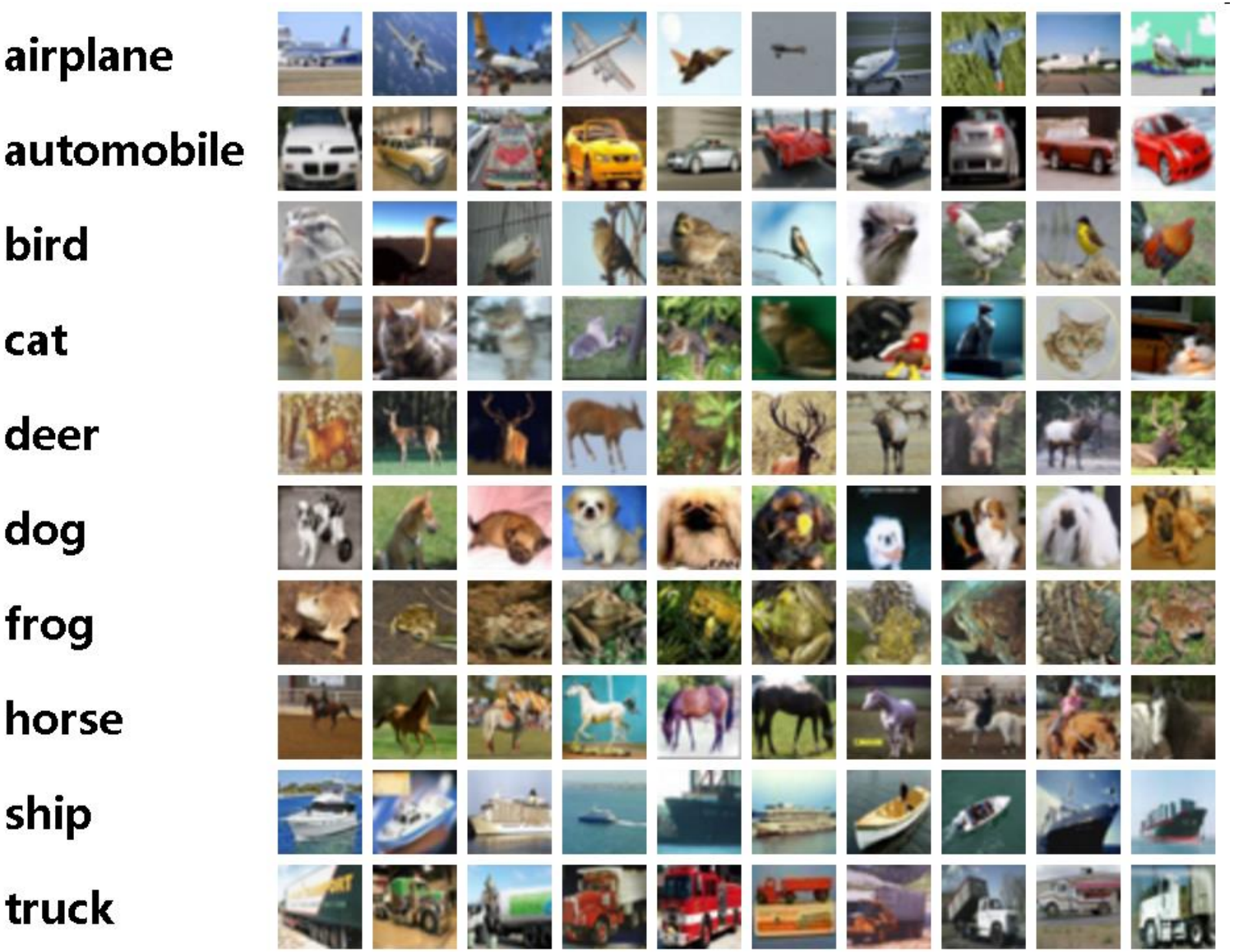
The MNIST database of handwritten digits has a training set of 60,000 examples, and a test set of 10,000 examples. It is a subset of a larger set available from NIST. The digits have been size-normalized and centered in a fixed-size image (28*28) .



RANK	MODEL	PERCENTAGE ERROR	ACCURACY	TRAINABLE PARAMETERS	PAPER	CODE	RESULT	YEAR
1	**Branching/Merging CNN + Homogeneous Filter Capsules**	0.16	99.84	**1,514,187**	A Branching and Merging Convolutional Network with Homogeneous Filter Capsules			2020
2	**EnsNet** (Ensemble learning in CNN augmented with fully connected subnetworks)	0.16	99.84					2020
3	**SOPCNN**	0.17	99.83	>1,400,000	Stochastic Optimization of Plain Convolutional Neural Networks with Simple methods			2020
4	**RMDL** (30 RDLs)	0.18			RMDL: Random Multimodel Deep Learning for Classification			2018
5	**DropConnect**	0.21			Regularization of Neural Networks using DropConnect			2013

CIFAR 10

- 60000 32x32 colour images
- 10 classes
- 6000 images per class.
- 50000 training images
- 10000 test images
- The dataset is divided into five training batches and one test batch, each with 10000 images.
- The test batch contains exactly 1000 randomly-selected images from each class.



RANK	MODEL	PERCENTAGE CORRECT	PERCENTAGE ERROR	FLOPS	PARAMS	EXTRA TRAINING DATA	PAPER	CODE	RESULT	YEAR
1	BiT-L (ResNet)	99.37	0.63			✓	Big Transfer (BiT): General Visual Representation Learning			2019
2	GPIPE + transfer learning	99	1			×	GPipe: Efficient Training of Giant Neural Networks using Pipeline Parallelism			2018
3	TResNet-XL	99				×	TResNet: High Performance GPU-Dedicated Architecture			2020
4	BiT-M (ResNet)	98.91	1.09			✓	Big Transfer (BiT): General Visual Representation Learning			2019
5	EfficientNet-B7	98.9			64M	×	EfficientNet: Rethinking Model Scaling for Convolutional Neural Networks			2019

CIFAR 100

100 classes
600 images each class
500 training images
100 testing images per class

20 superclasses.

Each image comes with a "fine" label (the class to which it belongs) and a "coarse" label (the superclass to which it belongs).

Superclass	Classes
aquatic mammals	beaver, dolphin, otter, seal, whale
fish	aquarium fish, flatfish, ray, shark, trout
flowers	orchids, poppies, roses, sunflowers, tulips
food containers	bottles, bowls, cans, cups, plates
fruit and vegetables	apples, mushrooms, oranges, pears, sweet peppers
household electrical devices	clock, computer keyboard, lamp, telephone, television
household furniture	bed, chair, couch, table, wardrobe
insects	bee, beetle, butterfly, caterpillar, cockroach
large carnivores	bear, leopard, lion, tiger, wolf
large man-made outdoor things	bridge, castle, house, road, skyscraper
large natural outdoor scenes	cloud, forest, mountain, plain, sea
large omnivores and herbivores	camel, cattle, chimpanzee, elephant, kangaroo
medium-sized mammals	fox, porcupine, possum, raccoon, skunk
non-insect invertebrates	crab, lobster, snail, spider, worm
people	baby, boy, girl, man, woman
reptiles	crocodile, dinosaur, lizard, snake, turtle
small mammals	hamster, mouse, rabbit, shrew, squirrel
trees	maple, oak, palm, pine, willow
vehicles 1	bicycle, bus, motorcycle, pickup truck, train
vehicles 2	lawn-mower, rocket, streetcar, tank, tractor

RANK	MODEL	PERCENTAGE CORRECT	PERCENTAGE ERROR	FLOPS	PARAMS	EXTRA TRAINING DATA	PAPER	CODE	RESULT	YEAR
1	BiT-L (ResNet)	93.51	6.49			✓	Big Transfer (BiT): General Visual Representation Learning			2019
2	BiT-M (ResNet)	92.17	7.83			✓	Big Transfer (BiT): General Visual Representation Learning			2019
3	EfficientNet-B7	91.7			64M	✓	EfficientNet: Rethinking Model Scaling for Convolutional Neural Networks			2019
4	TResNet-XL	91.5				✓	TResNet: High Performance GPU-Dedicated Architecture			2020
5	GPIPE	91.3				✓	GPipe: Efficient Training of Giant Neural Networks using Pipeline Parallelism			2018

Places&Places2

Places(Places1 or Places205), with 205 scene categories and 2.5 millions of images with a category label

Places2(Places365) contains more than 10 million images comprising 400+ unique scene categories. The dataset features 5000 to 30,000 training images per class, consistent with real-world frequencies of occurrence

Leaderboard of Places Database

Top1 accuracy and Top5 accuracy on the test set of Places205:

Display Name	Affiliation	Top1 Accuracy	Top5 Accuracy	Submission Date
SamExynos	Qian Zhang(Beijing Samsung Telecom R&D Center)	0.6410	0.9065	2016-04-17 07:18:06
SIAT_MMLAB	Limin Wang,Sheng Guo,Weilin Huang,Yu Qiao	0.6234	0.8966	2015-12-31 02:40:03
Residual-CNDS	Hussein Al-barazanchi, Hussam Qassim, Dr. Abhishek Verma (CSUF)	0.5703	0.8646	2016-09-16 02:03:02
ResNet-34	La Trobe University	0.5689	0.8591	2016-05-15 00:43:57
Places205_CNDS	Liwei Wang(UIUC),Chen-Yu Lee(UCSD)	0.5571	0.8575	2015-05-24 16:22:14
Places205-GoogLeNet	MIT	0.5550	0.8566	2015-05-22 10:35:00
reynoldscem	Digital Bridge	0.5309	0.8309	2016-07-11 05:37:04
blueblood22	xunlei	0.5237	0.8331	2017-06-21 04:36:07
Places205-AlexNet	MIT	0.5004	0.8110	2015-05-16 12:51:00
fdsafadfasdf	dfasdfasdf	0.5002	0.8109	2016-01-14 06:11:47
VAL-CDS	Indian Institute of Science, Bangalore	0.4769	0.7862	2016-06-09 02:59:54
dougal - baseline 1	CMU - Auton	0.4750	0.7983	2015-11-06 18:58:07
Shuai	Dalian University of Technology	0.4324	0.7505	2017-06-13 03:16:06

http://places.csail.mit.edu/user/leaderboard.php

Task A: Scene classification with provided training data

Team name	Entry description	Classification error
WM	Fusion with product strategy	0.168715
WM	Fusion with learnt weights	0.168747
WM	Fusion with average strategy	0.168909
WM	A single model (model B)	0.172876
WM	A single model (model A)	0.173527
SIAT_MMLAB	9 models	0.173605
SIAT_MMLAB	13 models	0.174645
SIAT_MMLAB	more models	0.174795
SIAT_MMLAB	13 models	0.175417
SIAT_MMLAB	2 models	0.175868
Qualcomm Research	Weighted fusion of two models. Top 5 validation error is 16.45%.	0.175978
Qualcomm Research	Ensemble of two models. Top 5 validation error is 16.53%.	0.176559
Qualcomm Research	Ensemble of seven models. Top 5 validation error is 16.68%	0.176766
Trimps-Soushen	score combine with 5 models	0.179824
Trimps-Soushen	score combine with 8 models	0.179997
Trimps-Soushen	top10 to top5, label combine with 9 models	0.180714
Trimps-Soushen	top10 to top5, label combine with 7 models	0.180984
Trimps-Soushen	single model, bn07	0.182357
ntu_rose	test_4	0.193367
ntu_rose	test_2	0.193645
ntu_rose	test_5	0.19397
ntu_rose	test_3	0.194262

Download Places365: http://places2.csail.mit.edu/download.html

Places Challenge 2015 result: http://places2.csail.mit.edu/results2015.html

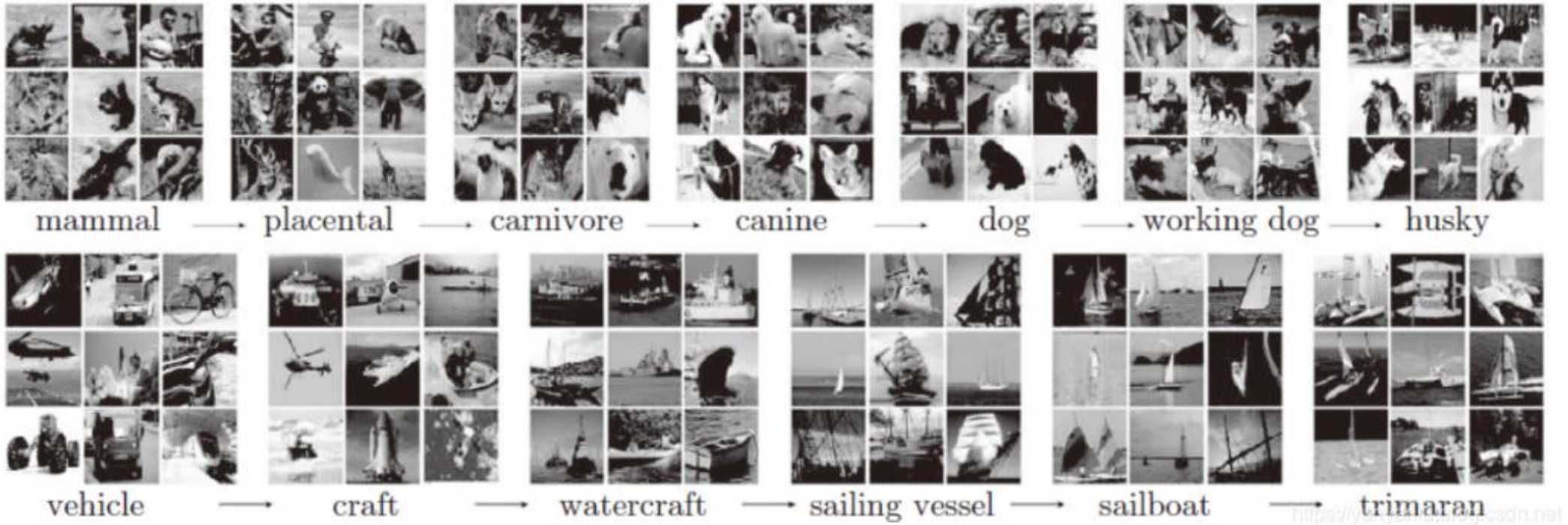
ImageNet

· 图像分类与目标定位（CLS-LOC）
· 目标检测（DET）
· 视频目标检测（VID）
· 场景分类（Scene）

ILSVRC（ImageNet Large Scale Visual Recognition Challenge）

ImageNet contains over 14 million full-size annotated images, 21,841 Synsets, WordNet

e.g.in 2012: train 1281167，val 50000，test 100000。



（1）图像分类与目标定位（CLS-LOC）

图像分类的任务是要判断图片中物体在1000个分类中所属的类别，主要采用**top-5错误率**的评估方式，即对于每张图给出5次猜测结果，只要5次中有一次命中真实类别就算正确分类，最后统计没有命中的错误率。

2012年之前，图像分类最好的成绩是26%的错误率，2012年AlexNet的出现降低了10个百分点，错误率降到16%。2016年，公安部第三研究所选派的“搜神”（Trimps-Soushen）代表队在这一项目中获得冠军，将成绩提高到仅有2.9%的错误率。

目标定位是在分类的基础上，**从图片中标识出目标物体所在的位置**，用**方框**框定，以错误率作为评判标准。目标定位的难度在于图像分类问题可以有5次尝试机会，而在目标定位问题上，每一次都需要框定的非常准确。

目标定位项目在2015年ResNet从上一年的最好成绩25%的错误率提高到了9%。2016年，公安部第三研究所选派的“搜神”（Trimps-Soushen）代表队的错误率仅为7%。

Start exploring here

Numbers in brackets: (the number of synsets in the subtree).

- ImageNet 2011 Fall Release (32326)
  - plant, flora, plant life (4486)
  - geological formation, formation (175)
  - natural object (1112)
  - sport, athletics (176)
  - artifact, artefact (10504)
  - fungus (308)
  - person, individual, someone, someboc
  - animal, animate being, beast, brute, c
  - Misc (20400)

Popular Synsets

Animal

fish
bird
mammal
invertebrate

Plant

tree
flower
vegetable

Activity

sport

Material

fabric

Instrumentation

utensil
appliance
tool
musical instrument

Scene

room
geological formation

Food

beverage

Image Classification on ImageNet

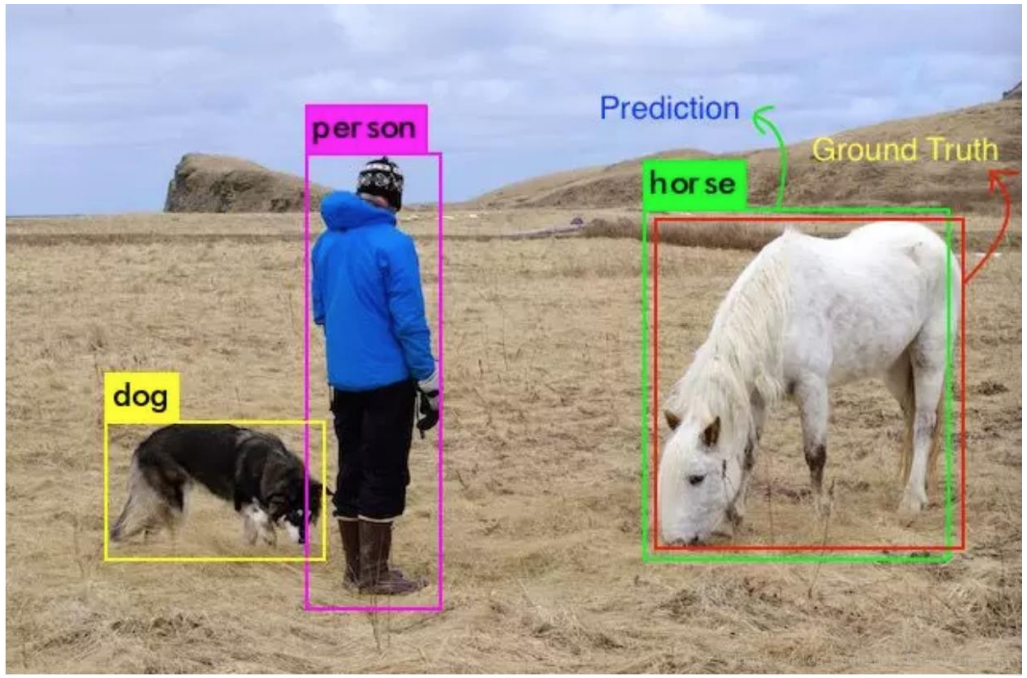
RANK	MODEL	TOP 1 ACCURACY	TOP 5 ACCURACY	NUMBER OF PARAMS	EXTRA TRAINING DATA	PAPER	CODE	RESULT	YEAR
1	FixEfficientNet-L2	88.5%	98.7%	480M	✓	Fixing the train-test resolution discrepancy: FixEfficientNet			2020
2	NoisyStudent (EfficientNet-L2)	88.4%	98.7%	480M	✓	Self-training with Noisy Student improves ImageNet classification			2020
3	BiT-L (ResNet)	87.54%	98.46%		✓	Big Transfer (BiT): General Visual Representation Learning			2019
4	FixEfficientNet-B7	87.1%	98.2%	66M	✓	Fixing the train-test resolution discrepancy: FixEfficientNet			2020
5	NoisyStudent (EfficientNet-B7)	86.9%	98.1%	66M	✓	Self-training with Noisy Student improves ImageNet classification			2019

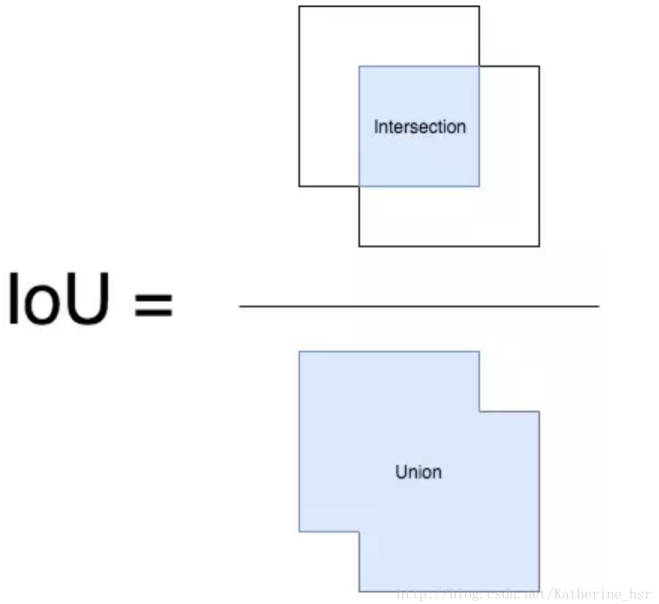
Dataset download: http://image-net.org/download-images

2

Detection

Evaluation index：mAP



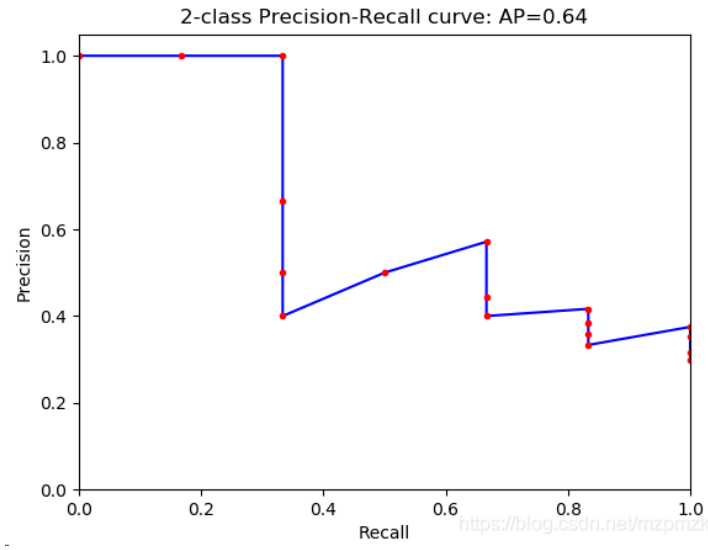


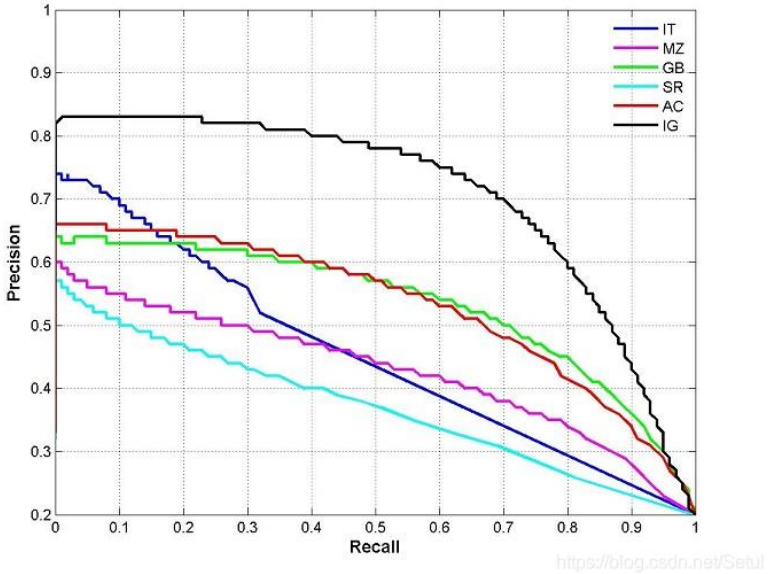
$$Precesion_C = \frac{N(TruePositives)_C}{N(TotalObjects)_C}$$

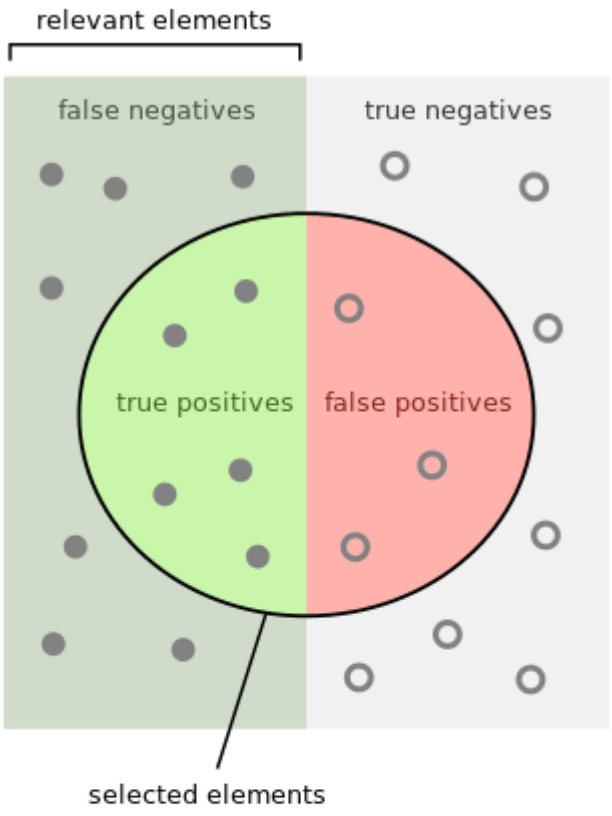
$$AveragePrecision_C = \frac{\sum Precision_C}{N(TotalImages)_C}$$

$$MeanAveragePrecision = \frac{\sum AveragePrecision_C}{N(Classes)}$$

P-R曲线

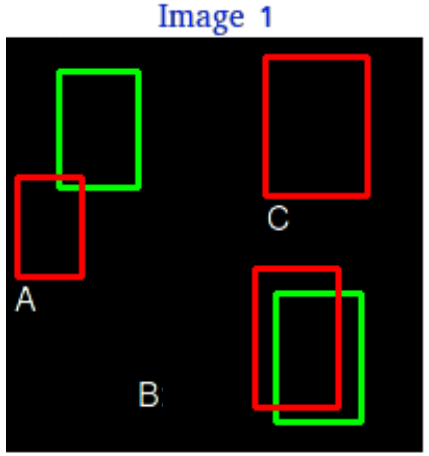


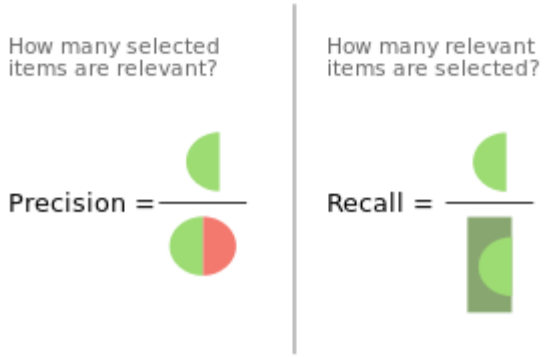




$$\text{Precision} = \frac{TP}{TP+FP} = \frac{TP}{\text{all detections}}$$

$$\text{Recall} = \frac{TP}{TP+FN} = \frac{TP}{\text{all ground truths}}$$





Pascal VOC

MS COCO

KITTI

ILSVRC-DET

1、PASCAL VOC的挑战任务

- Classification/Detection Competitions
  分类：对于每一个分类，判断该分类是否在测试照片上存在（共20类）；
  检测：检测目标对象在待测试图片中的位置并给出边界框坐标（bounding box)
- Segmentation Competition
  分割：Object Segmentation
- Action Classification Competition
  人体动作识别（Action Classification)
- ImageNet Large Scale Visual Recognition Competition
  ImageNet大型视觉识别大赛
- Person Layout Taster Competition
  人体布局（Human Layout)

Pascal VOC

VOC2007:
Include 20 classes:
Person: person
Animal: bird, cat, cow, dog, horse, sheep
Vehicle: aeroplane, bicycle, boat, bus, car, motorbike, train
Indoor: bottle, chair, dining table, potted plant, sofa, tv/monitor
Train/validation/test: 9,963 images containing 24,640 annotated objects.
Train：5011 Test：4952

20 classes

VOC2012: 20 classes. The train/val data has 11,530 images containing 27,450 ROI annotated objects and 6,929 segmentations.

Dataset download: http://host.robots.ox.ac.uk/pascal/VOC

- 训练集

aeroplane 238
bicycle 243
bird 330
boat 181
bottle 244
bus 186
car 713
cat 337
chair 445
cow 141
diningtable 200
dog 421
horse 287
motorbike 245
person 2008
pottedplant 245
sheep 96
sofa 229
train 261
tvmonitor 256

- 测试集

aeroplane 204
bicycle 239
bird 282
boat 172
bottle 212
bus 174
car 721
cat 322
chair 417
cow 127
diningtable 190
dog 418
horse 274
motorbike 222
person 2007
pottedplant 224
sheep 97
sofa 223
train 259
tvmonitor 229


2007_000027.jpg


2007_000032.jpg


2007_000033.jpg


2007_000039.jpg


2007_000042.jpg


2007_000061.jpg


2007_000063.jpg


2007_000068.jpg


2007_000121.jpg


2007_000123.jpg


2007_000129.jpg


2007_000170.jpg


2007_000175.jpg

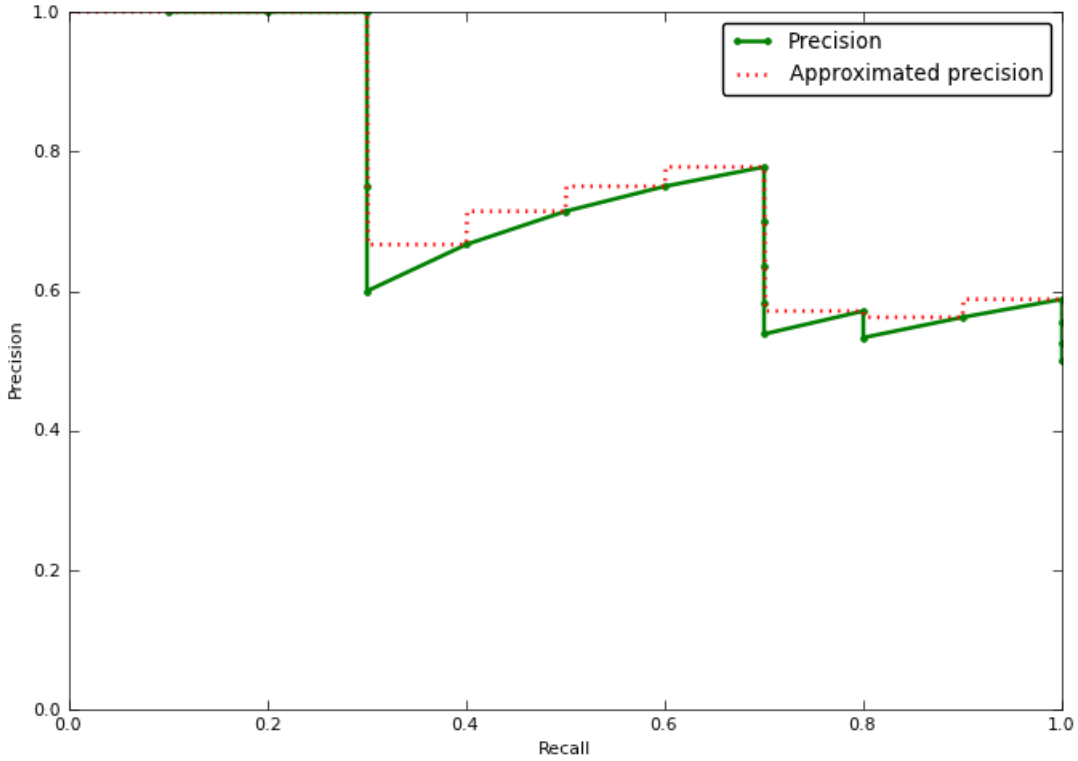

2007_000187.jpg

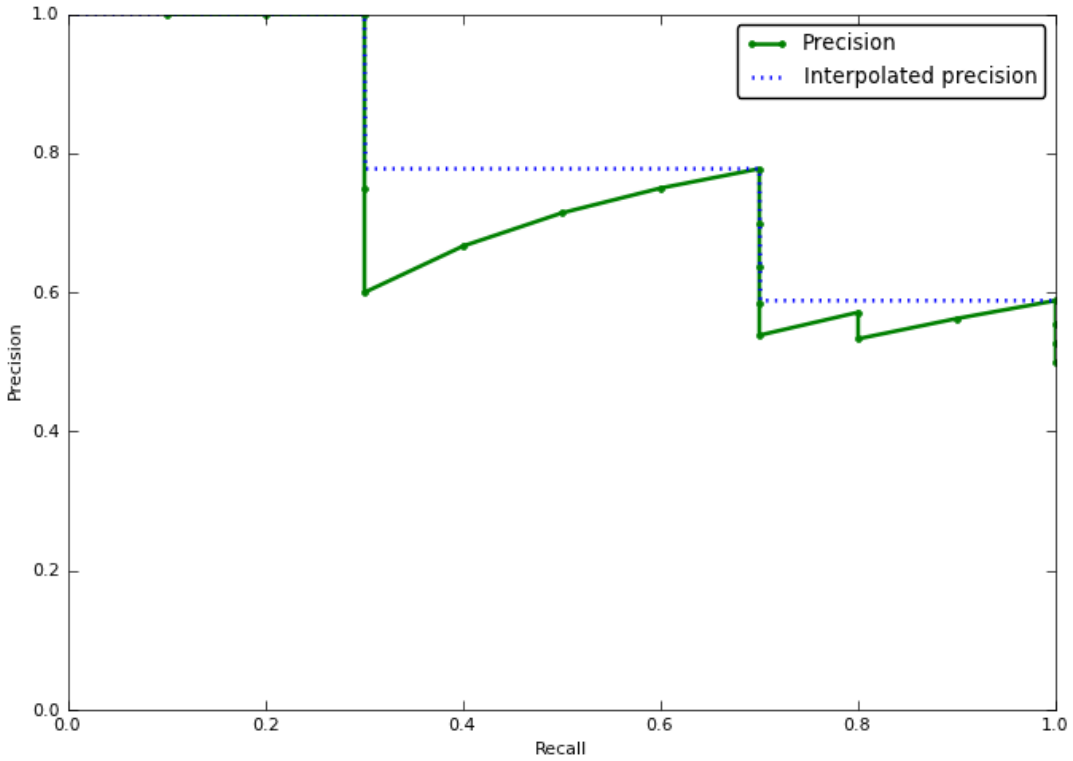

2007_000241.jpg


2007_000243.jpg


2007_000250.jpg


2007_000256.jpg


Precision
Approximated precision
Precision
Recall
0.0
0.2
0.4
0.6
0.8
1.0


Precision
Interpolated precision
Precision
Recall
0.0
0.2
0.4
0.6
0.8
1.0

RANK	MODEL	MAP	PAPER	CODE	RESULT	YEAR
1	RODEO (recon, n=12)	90.6%	RODEO: Replay for Online Object Detection			2020
2	SNIPER	86.9%	SNIPER: Efficient Multi-Scale Training			2018
3	RefineDet512+	83.8%	Single-Shot Refinement Neural Network for Object Detection			2017
4	YOLOv3 (sync. BN + rand. shapes + cos. lr + lbl. smoothing + mixup)	83.68%	Bag of Freebies for Training Object Detection Neural Networks			2019
5	InterNet (ResNet-101)	82.7%	Feature Intertwiner for Object Detection			2019
6	CoupleNet	82.7%	CoupleNet: Coupling Global Structure with Local Parts for Object Detection			2017
7	SPP (Overfeat-7)	82.44%	Spatial Pyramid Pooling in Deep Convolutional Networks for Visual Recognition			2014
8	SSD512 (07+12+COCO)	81.6%	SSD: Single Shot MultiBox Detector			2015
9	BlitzNet512 + seg (s8)	81.5%	BlitzNet: A Real-Time Deep Network for Scene Understanding			2017
10	Faster-RCNN (cos. lr, label smoothing, mixup)	81.32%	Bag of Freebies for Training Object Detection Neural Networks			2019

MS COCO

(1) Object segmentation
(2) Recognition in Context
(3) Multiple objects per image
(4) More than 300,000 images
(5) More than 2 Million instances
(6) 80 object categories
(7) 5 captions per image
(8) Keypoints on 100,000 people

Fig. 6: Samples of annotated images in the MS COCO dataset.

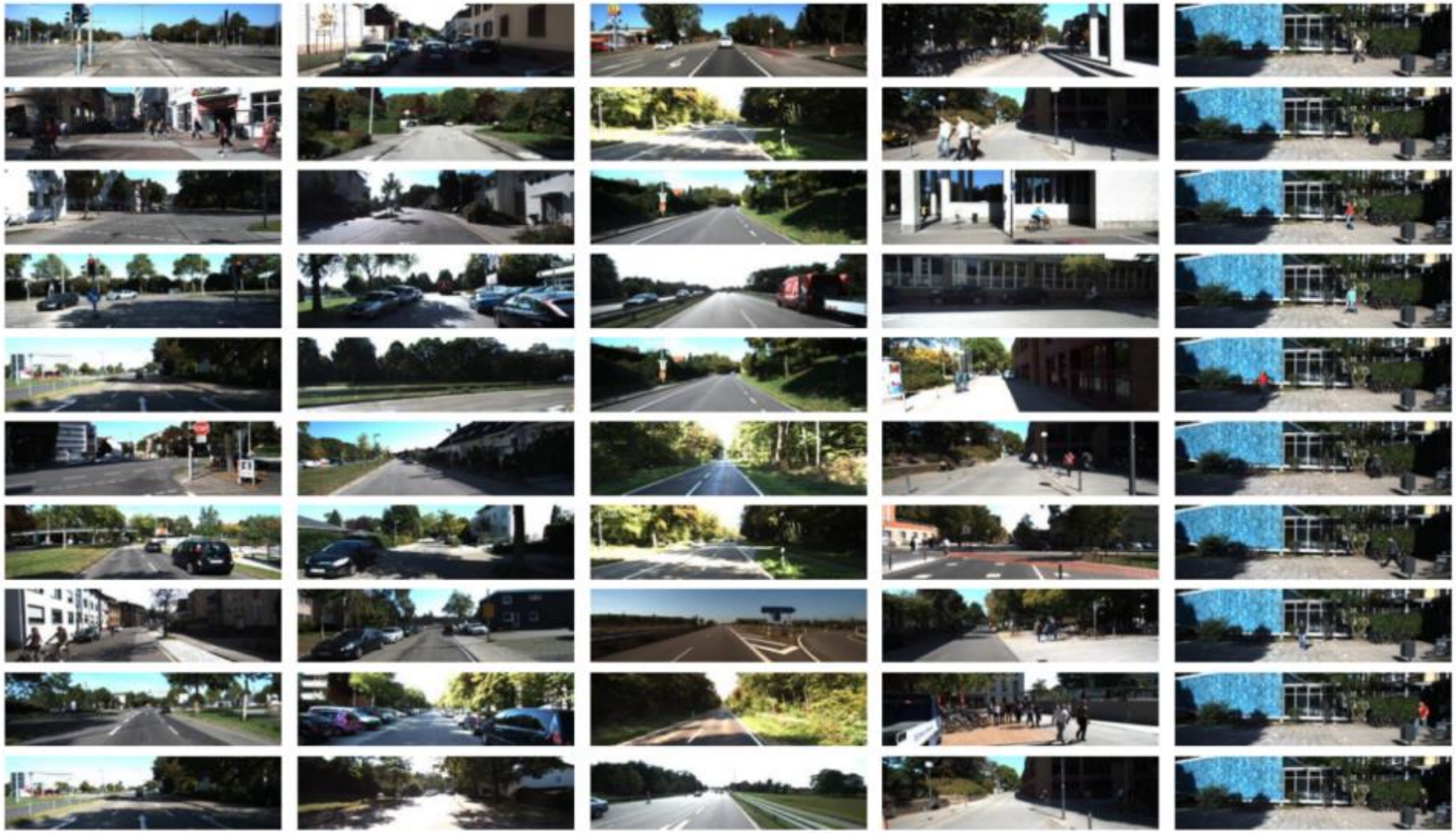
person(人)
bicycle(自行车) car(汽车) motorbike(摩托车) aeroplane(飞机) bus(公共汽车) train(火车) truck(卡车) boat(船)
traffic light(信号灯) fire hydrant(消防栓) stop sign(停车标志) parking meter(停车计费器) bench(长凳)
bird(鸟) cat(猫) dog(狗) horse(马) sheep(羊) cow(牛) elephant(大象) bear(熊) zebra(斑马) giraffe(长颈鹿)
backpack(背包) umbrella(雨伞) handbag(手提包) tie(领带) suitcase(手提箱)
frisbee(飞盘) skis(滑雪板双脚) snowboard(滑雪板) sports ball(运动球) kite(风筝) baseball bat(棒球棒) baseball glove(棒球手套) skateboard(滑板) surfboard(冲浪板) tennis racket(网球拍)
bottle(瓶子) wine glass(高脚杯) cup(茶杯) fork(叉子) knife(刀)
spoon(勺子) bowl(碗)
banana(香蕉) apple(苹果) sandwich(三明治) orange(橘子) broccoli(西兰花) carrot(胡萝卜) hot dog(热狗) pizza(披萨) donut(甜甜圈) cake(蛋糕)
chair(椅子) sofa(沙发) pottedplant(盆栽植物) bed(床) diningtable(餐桌) toilet(厕所) tvmonitor(电视机)
laptop(笔记本) mouse(鼠标) remote(遥控器) keyboard(键盘) cell phone(电话)
microwave(微波炉) oven(烤箱) toaster(烤面包器) sink(水槽) refrigerator(冰箱)
book(书) clock(闹钟) vase(花瓶) scissors(剪刀) teddy bear(泰迪熊) hair drier(吹风机) toothbrush(牙刷)

label

{"segmentation": [[392.87, 275.77, 402.24, 284.2, 382.54, 342.36, 375.99, 356.43, 372.23, 357.37, 372.23, 397.7, 383.48, 419.27,407.87, 439.91, 427.57, 389.25, 447.26, 346.11, 447.26, 328.29, 468.84, 290.77,472.59, 266.38], [429.44,465.23, 453.83, 473.67, 636.73, 474.61, 636.73, 392.07, 571.07, 364.88, 546.69,363.0]], "area": 28458.996150000003, "iscrowd": 0,"image_id": 503837, "bbox": [372.23, 266.38, 264.5,208.23], "category_id": 4, "id": 151109},

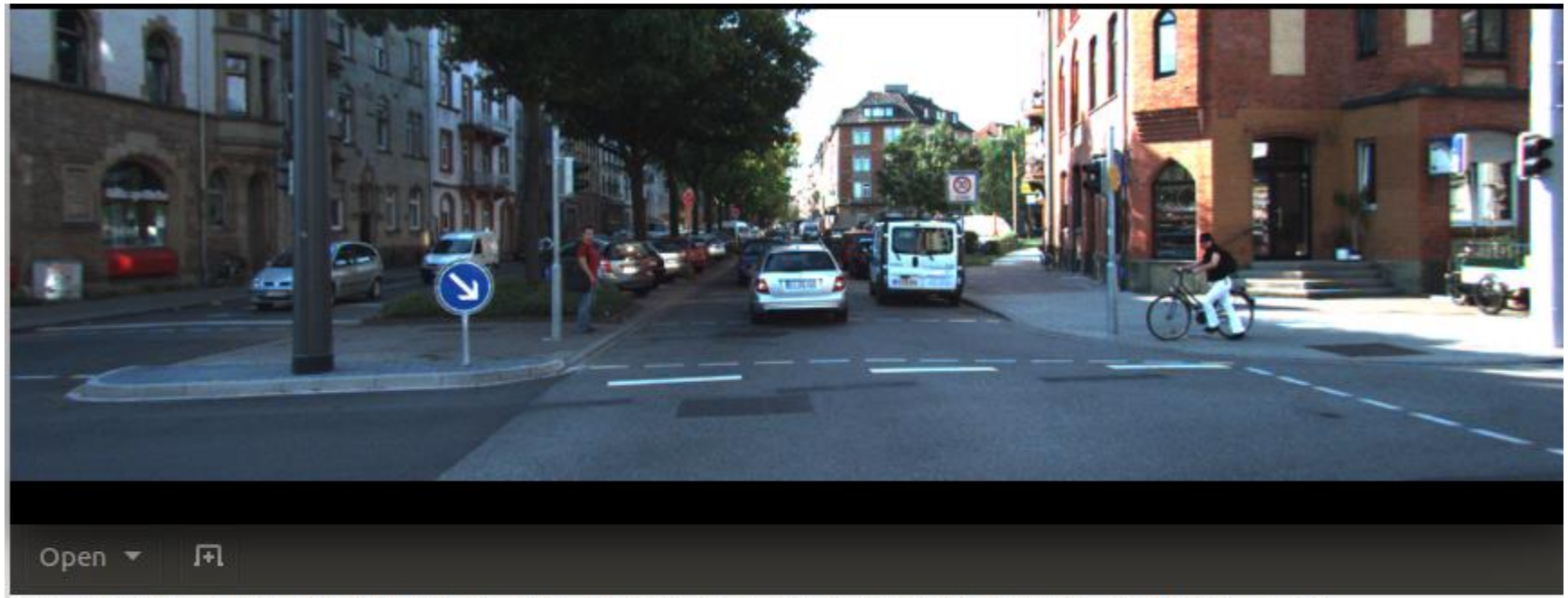
RANK	MODEL	BOX AP	AP50	AP75	APS	APM	APL	EXTRA TRAINING DATA	PAPER	CODE	RESULT	YEAR
1	EfficientDet-D7x (single-scale)	55.1	74.3	59.9	37.2	57.9	68.0	×	EfficientDet: Scalable and Efficient Object Detection			2020
2	DetectoRS (ResNeXt-101-32x4d, multi-scale)	54.7	73.5	60.1	37.4	57.3	66.4	×	DetectoRS: Detecting Objects with Recursive Feature Pyramid and Switchable Atrous Convolution			2020
3	SpineNet-190 (1280, with Self-training on OpenImages, single-scale)	54.3						✓	Rethinking Pre-training and Self-training			2020
4	CSP-p6 + Mish	53.8	71.4	59	38.3	58.2	67.7	×	Mish: A Self Regularized Non-Monotonic Activation Function			2019
5	EfficientDet-D7 (single-scale)	53.7	72.4	58.4		57.0	66.3	×	EfficientDet: Scalable and Efficient Object Detection			2019

KITTI



Object detection:

- Car
- Van
- Truck
- Pedestrian
- Person (sit- ting)
- Cyclist
- Tram
- Misc




```
Car 0.00 0 -1.59 589.01 187.21 668.42 253.27 1.36 1.69 3.38 0.35 1.73 17.14 -1.57
Car 0.00 1 2.04 185.19 184.44 302.47 240.64 1.59 1.72 3.86 -11.47 1.98 22.83 1.58
Cyclist 0.00 3 2.78 888.50 173.04 1019.87 266.61 1.68 0.86 2.01 6.34 1.70 13.46 -3.08
Van 0.00 3 -1.68 682.68 157.58 763.18 235.92 2.12 1.86 4.41 3.27 1.74 21.92 -1.54
Pedestrian 0.00 0 0.08 447.05 168.53 472.39 258.42 1.87 0.64 0.65 -3.25 1.78 15.37 -0.13
Van 0.00 3 1.89 325.53 175.96 390.57 216.45 1.71 1.56 4.12 -11.42 1.87 32.86 1.56
Car 0.00 0 -2.21 409.03 180.12 515.81 231.99 1.59 1.63 3.64 -5.01 1.85 24.07 -2.41
Car 0.00 2 -2.35 445.20 184.58 542.42 220.39 1.39 1.61 4.09 -4.95 1.91 30.30 -2.51
Car 0.00 2 -2.38 485.80 181.56 556.50 211.93 1.50 1.57 3.54 -4.68 1.97 37.56 -2.50
Car 0.00 2 -2.37 520.40 180.16 572.44 200.74 1.40 1.60 3.55 -4.55 1.94 51.13 -2.45
Car 0.00 2 -1.55 579.15 180.82 622.59 220.52 1.52 1.67 3.61 -0.38 1.85 29.71 -1.57
Car 0.00 2 1.96 329.94 179.62 388.01 205.99 1.47 1.77 4.25 -14.86 1.88 42.86 1.63
DontCare -1 -1 -10 555.40 164.60 601.27 188.60 -1 -1 -1 -1000 -1000 -1000 -10
DontCare -1 -1 -10 622.06 164.60 662.73 189.64 -1 -1 -1 -1000 -1000 -1000 -10
```

RANK	MODEL	AP	PAPER	CODE	RESULT	YEAR
1	Patches	87.87	Patch Refinement -- Localized 3D Object Detection			2019
2	PointRCNN Shi et al. (2019)	85.94	PointRCNN: 3D Object Proposal Generation and Detection from Point Cloud			2018
3	Roamet	83.71	RoarNet: A Robust 3D Object Detection based on RegiOn Approximation Refinement			2018
4	AVOD-FPN	81.94	Joint 3D Proposal Generation and Object Detection from View Aggregation			2017
5	PointPillars	79.05	PointPillars: Fast Encoders for Object Detection from Point Clouds			2018
6	VoxelNet	77.47	VoxelNet: End-to-End Learning for Point Cloud Based 3D Object Detection			2017

Image Detection on ImageNet

（2）目标检测（DET）

目标检测是在定位的基础上更进一步，在**图片中同时检测并定位多个类别的物体**。具体来说，是要在每一张测试图片中找到属于200个类别中的所有物体，如人、勺子、水杯等。**评判方式是看模型在每一个单独类别中的识别准确率**，在多数类别中都获得最高准确率的队伍获胜。**平均检出率mean AP（mean Average Precision）**也是重要指标，一般来说，平均检出率最高的队伍也会多数的独立类别中获胜，2016年这一成绩达到了66.2。

（3）视频目标检测（VID）

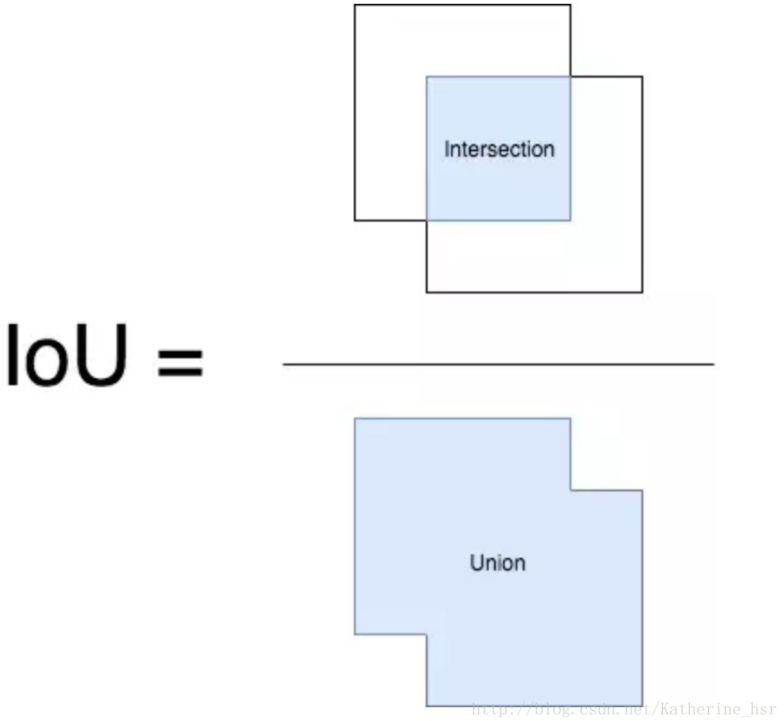
视频目标检测是要检测出**视频每一帧中包含的多个类别的物体**，与图片目标检测任务类似。要检测的目标物体有30个类别，是目标检测200个类别的子集。此项目的最大难度在于要求算法的检测效率非常高。评判方式是在独立类别识别最准确的队伍获胜。

2016年南京信息工程大学队伍在这一项目上获得了冠军，他们提供的两个模型分别在10个类别中胜出，并且达到了平均检出率超过80%的好成绩。

3

Segmentation

Evaluation index：mIoU(mean IoU)



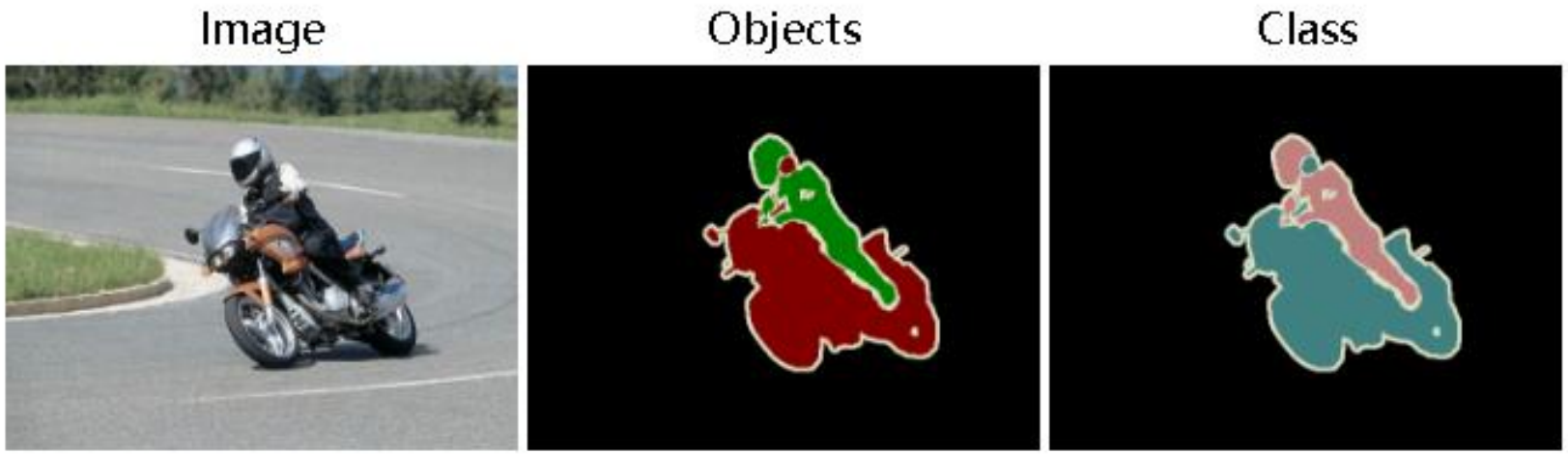
Pascal VOC

MS COCO

ADE20K

Cityscapes

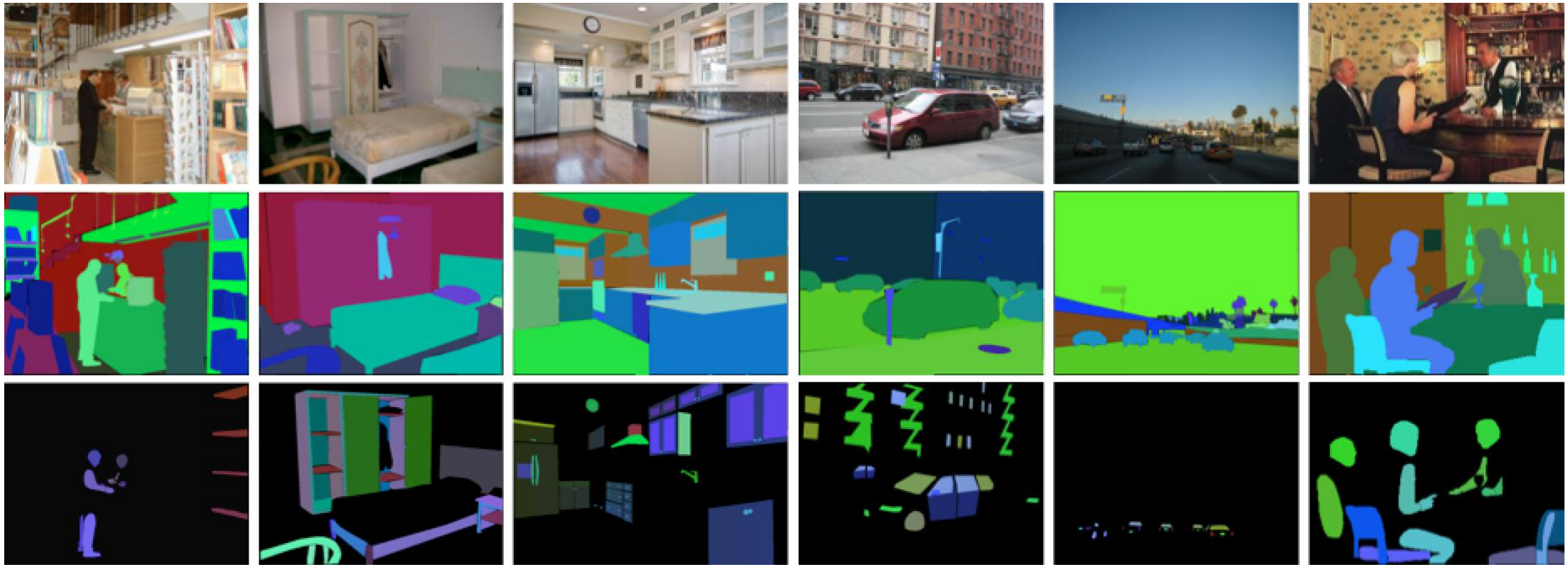
Pascal VOC



MS COCO

ADE20K

- Invlude over 25,000 images (20ktrain, 2k val, 3ktest)
- same scene categories than the Places Database
- each image has the object and part segmentations
- All object and part instances are annotated sparately



Dataset download: http://groups.csail.mit.edu/vision/datasets/ADE20K/

RANK	MODEL	VALIDATION MIOU	TEST SCORE	PAPER	CODE	RESULT	YEAR
1	ResNeSt-200	48.36		ResNeSt: Split-Attention Networks			2020
2	ResNeSt-269	47.60		ResNeSt: Split-Attention Networks			2020
3	ResNeSt-101	46.91		ResNeSt: Split-Attention Networks			2020
4	CPN (ResNet-101)	46.27		Context Prior for Scene Segmentation			2020
5	DRAN (ResNet-101)	46.18%		Scene Segmentation with Dual Relation-aware Attention Network			2019
6	PyConvSegNet-152	45.99	0.5652	Pyramidal Convolution: Rethinking Convolutional Neural Networks for Visual Recognition			2020
7	LaU-regression-loss	45.02	0.5632	Location-aware Upsampling for Semantic Segmentation			2019
8	PSPNet	44.94	0.5538	Pyramid Scene Parsing Network			2016
9	CFNet (ResNet-101)	44.89		Co-Occurrent Features in Semantic Segmentation			2019
10	EncNet	44.65	0.5567	Context Encoding for Semantic Segmentation			2018

Cityscapes

Group	Classes
flat	road · sidewalk · parking+ · rail track+
human	person* · rider*
vehicle	car* · truck* · bus* · on rails* · motorcycle* · bicycle* · caravan*+ · trailer*+
construction	building · wall · fence · guard rail+ · bridge+ · tunnel+
object	pole · pole group+ · traffic sign · traffic light
nature	vegetation · terrain
sky	sky
void	ground+ · dynamic+ · static+

Benchmark suite and evaluation server

•Pixel-level semantic labeling

•Instance-level semantic labeling

Features

•30 classes

•50 cities

•Several months (spring, summer, fall)

•Daytime

•Good/medium weather conditions

•Manually selected frames

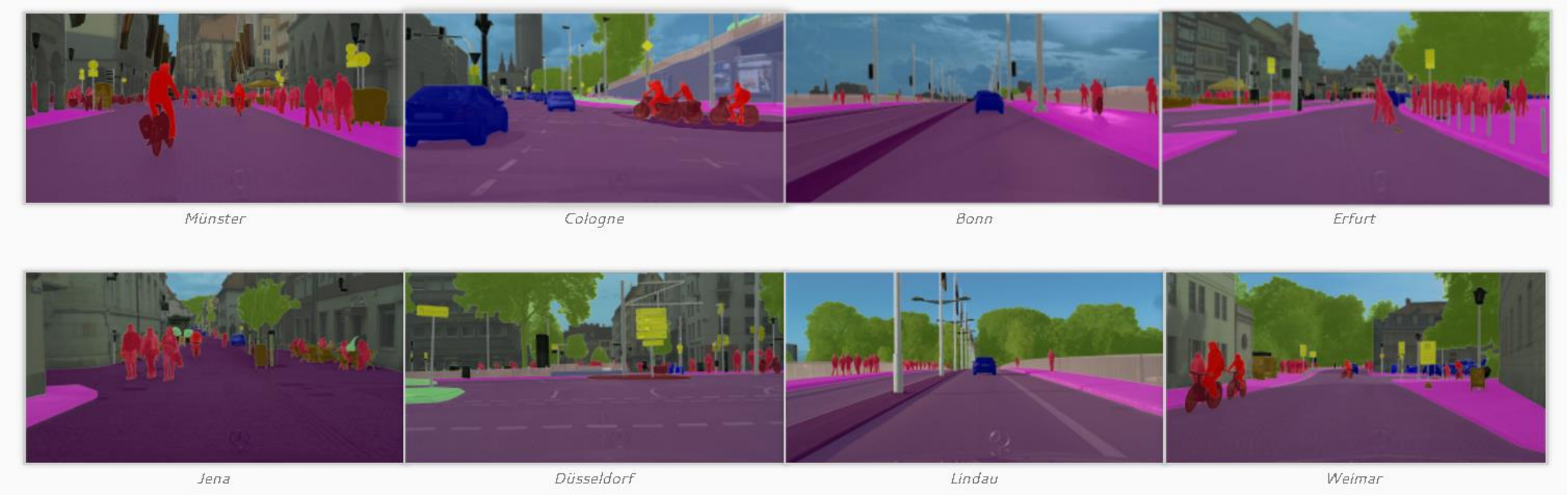
Volume

•5 000 annotated images with fine annotations (2975train， 500 val,1525test)

•20 000 annotated images with coarse annotations

Dataset download: https://www.cityscapes-dataset.com/

Fine annotations



Coarse annotations

RANK	MODEL	MEAN IOU (CLASS)	CATEGORY MIOU	GFLOPS	EXTRA TRAINING DATA	PAPER	CODE	RESULT	YEAR
1	HRNet-OCR (Hierarchical Multi-Scale Attention)	85.1%			✓	Hierarchical Multi-Scale Attention for Semantic Segmentation			2020
2	HRNetV2 + OCR +	84.5%			✓	Object-Contextual Representations for Semantic Segmentation			2019
3	EfficientPS	84.21%			✓	EfficientPS: Efficient Panoptic Segmentation			2020
4	Panoptic-DeepLab	84.2%			✓	Panoptic-DeepLab: A Simple, Strong, and Fast Baseline for Bottom-Up Panoptic Segmentation			2019
5	HRNetV2 + OCR (w/ ASP)	83.7%			✓	Object-Contextual Representations for Semantic Segmentation			2019
6	DCNAS	83.6%			✓	DCNAS: Densely Connected Neural Architecture Search for Semantic Image Segmentation			2020
7	DeepLabV3Plus + SDCNetAug	83.5%			✓	Improving Semantic Segmentation via Video Propagation and Label Relaxation			2018
8	GALDNet (+Mapillary)	83.3%			✓	Global Aggregation then Local Distribution in Fully Convolutional Networks			2019
9	ResNeSt200	83.3%			✓	ResNeSt: Split-Attention Networks			2020
10	HANet (Height-driven Attention Networks by LGE A&B)	83.2%			✓	Cars Can't Fly up in the Sky: Improving Urban-Scene Segmentation via Height-driven Attention Networks			2020

https://paperswithcode.com/sota

Search for papers, code and tasks

Browse State-of-the-Art Methods Trends About RC2020 Log In/Register

Browse State-of-the-Art

3,104 benchmarks • 1,694 tasks • 2,743 datasets • 27,179 papers with code

Follow on Twitter for updates

Computer Vision

Task	Benchmarks	Papers with code
Semantic Segmentation	61 benchmarks	1097 papers with code
Image Classification	151 benchmarks	917 papers with code
Object Detection	132 benchmarks	806 papers with code
Image Generation	111 benchmarks	384 papers with code
Pose Estimation	97 benchmarks	381 papers with code

See all 877 tasks

Natural Language Processing

Task	Benchmarks	Papers with code
Machine Translation	49 benchmarks	681 papers with code
Language Modelling	14 benchmarks	658 papers with code
Question Answering	56 benchmarks	611 papers with code
Sentiment Analysis	37 benchmarks	422 papers with code
Text Classification	66 benchmarks	259 papers with code

See all 312 tasks

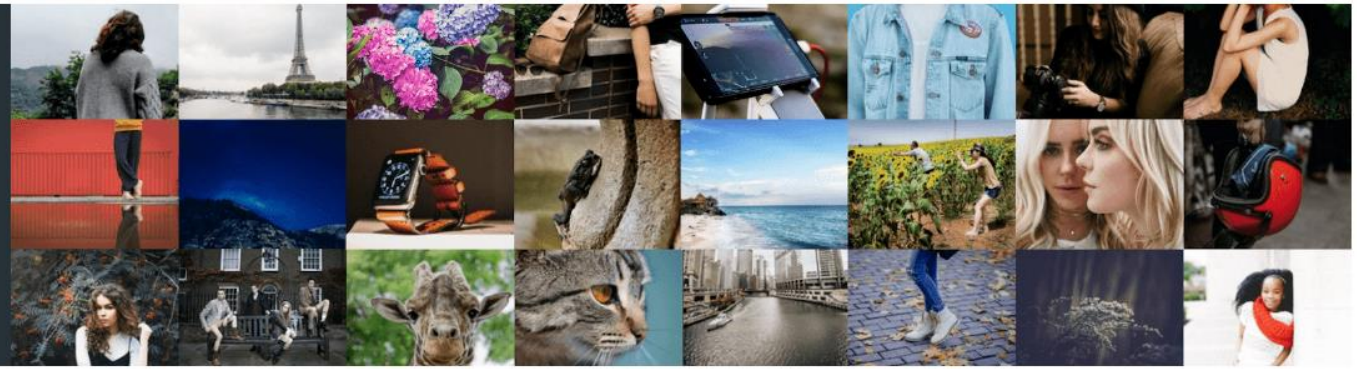
https://www.visualdata.io/discovery

VisualData Discovery Studio Preview Log in Sign Up

VisualData Discovery

Best place to find and share computer vision datasets

Subscribe for updates



MENTIONED IN

KDnuggets 机器之心 towards data science awesome McGill UNIVERSITY O'REILLY

Service cannot be reached at this moment. Please check again later.

Add My Dataset What are you looking for Sort by

Topics

Select topics

Filter by:

Top

Thanks for Listening