OpenVINO: Intel Model / Model Zoo

Jan 2019

Kai-Feng Chou/ Maple Chou
Specialist FAE Intel PSG

High-level outline

Open Model Zoo is distributed with OpenVINO™ and consists of the following:

- model_downloader
  - Downloads public networks from a predefined list
  - Contains mean values/scale info
- intel_models
  - Set of trained-by-Intel models in IR under ISSL covering a wide range of tasks
  - 1-page description for each model
- The above items are connected to the samples
  - Demonstrate the usage of intel_models

Samples & Models (+couple of live demos)

General samples

- Image Classification
- Image Segmentation
- Object Detection
- Object Detection for Single Shot Multibox Detector (SSD)
- Neural Style Transfer
- Validation Application

Pre-Trained Models

- Age – Gender
- Headpose
- Face detection
- Licence-plate recognition
- Vehicle Attributes
- …

https://software.intel.com/en-us/openvino-toolkit/documentation/pretrained-models

Pretrained Model	Supported Samples	CPU	Integrated Graphics	FPGA	VPU
face-detection-adas-0001	Interactive face detection	✓	✓	✓	✓
age-gender-recognition-retail-0013	Interactive face detection	✓	✓	✓	✓
head-pose-estimation-adas-0001	Interactive face detection	✓	✓	✓	
emotions-recognition-retail-0003	Interactive face detection	✓	✓	✓	✓
facial-landmarks-35-adas-0001	Interactive face detection	✓	✓		
vehicle-license-plate-detection-barrier-0106	Security barrier camera	✓	✓	✓	✓
vehicle-attributes-recognition-barrier-0039	Security barrier camera	✓	✓	✓	✓
license-plate-recognition-barrier-0001	Security barrier camera	✓	✓	✓	✓
person-detection-retail-0001	Object detection	✓	✓		
person-vehicle-bike-detection-crossroad-0078	Crossroad camera	✓	✓	✓	✓
person-attributes-recognition-crossroad-0200	Crossroad camera	✓	✓		
person-reidentification-retail-0076	Crossroad camera	✓	✓	✓	✓
person-reidentification-retail-0031	Crossroad camera pedestrian tracker	✓	✓	✓	✓

Example public models, more in the <INSTALL_DIR>/deployment_tools/model_downloader/list_topologies.yml

Network family	Model	Problem/ Dataset	URL
DenseNet	densenet-121 densenet-161 densenet-169 densenet-201	ImageNet	https://github.com/liuzhuang13/DenseNet
SqueezeNet	squeezenet1.0 squeezenet1.1	ImageNet	https://github.com/DeepScale/SqueezeNet
MTCNN	mtcnn-p mtcnn-r mtcnn-o	FDDB, AFLW	https://github.com/kpzhang93/MTCNN_face_detection_alignment
MobileNet-SSD	mobilenet-ssd	VOC0712	https://github.com/chuanqi305/MobileNet-SSD
VGG	vgg19 vgg16	ImageNet	https://gist.github.com/ksimonyan/3785162f95cd2d5fee77 https://gist.github.com/ksimonyan/211839e770f7b538e2d8
SSD	ssd512 ssd300	VOC0712	https://github.com/weiliu89/caffe

Object Detection

Model name	Complexity (GFLOPs)	Size (Mp)	Face	Person	Vehicle	Bike	License plate
face-detection-adas-0001	1.4	1.1	X				
face-detection-retail-0004	1.1	0.6	X				
face-person-detection-retail-0002	2.8	0.8	X	X			
person-detection-retail-0001	12.6	3.2		X			
person-detection-retail-0013	3.9	1.9		X			
pedestrian-detection-adas-0002	1.5	1.2		X			
pedestrian-and-vehicle-detector-adas-0001	4.0	1.6		X	X		
vehicle-detection-adas-0002	1.4	1.1			X		
person-vehicle-bike-detection-crossroad-0078	3.9	1.2		X	X	X	
vehicle-license-plate-detection-barrier-0007	3.0	1.1			X		X

Classification

Model name	Complexity (GFLOPs)	Size (Mp)
age-gender-recognition-retail-0013	0.09	2.1
head-pose-estimation-adas-0001	0.03	0.9
license-plate-recognition-barrier-0001	0.34	1.9
vehicle-attributes-recognition-barrier-0039	0.9	0.6
emotions-recognition-retail-0003	0.13	2.5
person-attributes-recognition-crossroad-0031	0.22	1.1

ReID

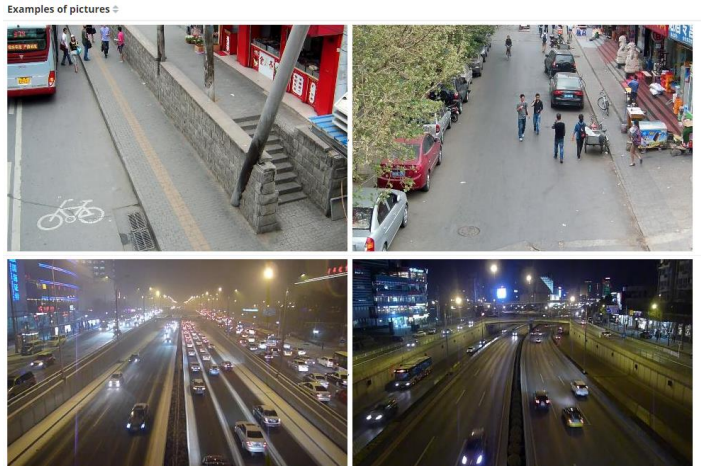
Model name	Complexity (GFLOPs)	Size (Mp)
person-reidentification-retail-0076	0.58	0.82
person-reidentification-retail-0079	0.12	0.82

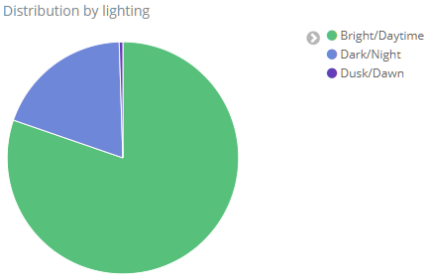
Semantic Segmentation

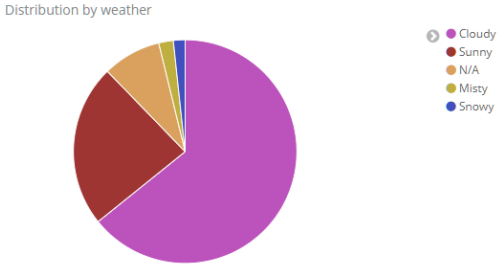
Model name	Complexity (GFLOPs)	Size (Mp)
road-segmentation-adas-0001	2.4	0.13
semantic-segmentation-adas-0001	29.04	6.6

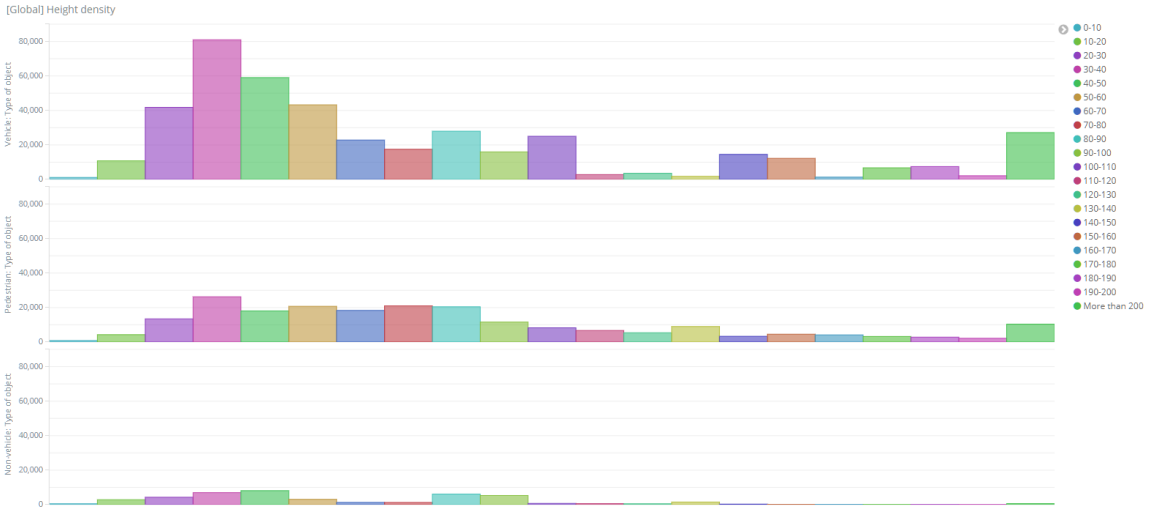
Crossroad scenario

Type of object	# of bounding boxes
Vehicle	426,734
Pedestrian	216,070
Non-vehicle	46,357



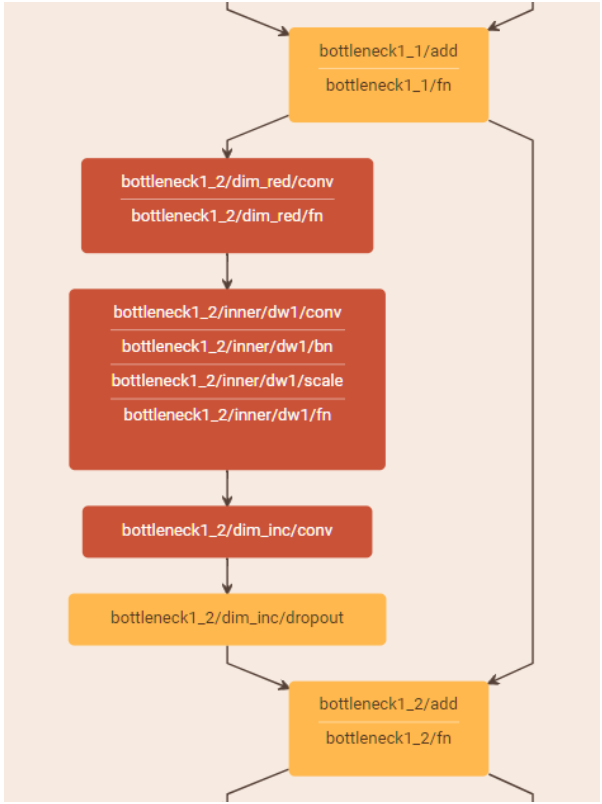




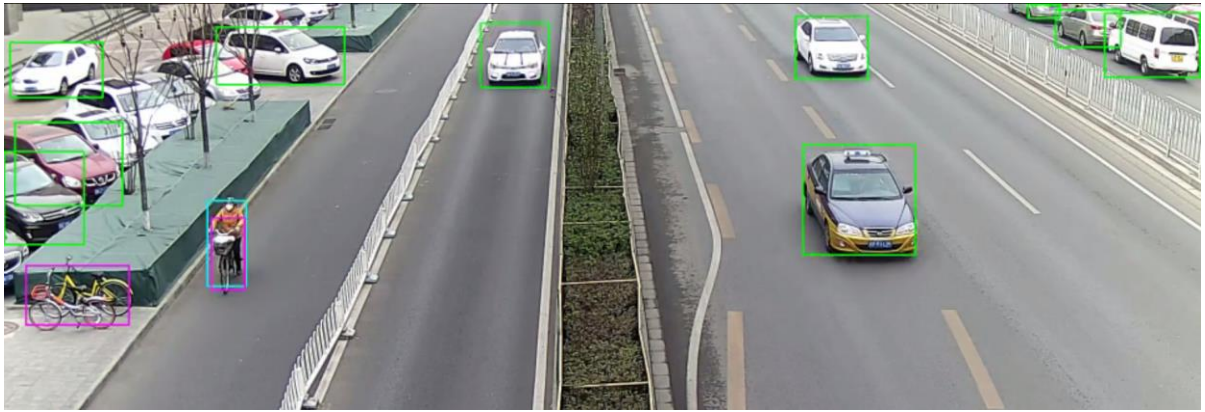


person-vehicle-bike-detection-crossroad-0078

- RMNet (MobileNetV2-like) + SSD-based network for Person/Vehicle/Bike detection in security surveillance applications. Works in a wide variety of scenes and weather/lighting conditions
- Thin and deep topology consisting of repeating blocks with depth-wise convolutions, ELU activations and residual connections



person-vehicle-bike-detection-crossroad-0078



person-vehicle-bike-detection-crossroad-0078

Metric	R1	R2
Mean Average Precision (mAP)	62.95%	72.23%
AP people	69.72%	79.82%
AP vehicles	72.68%	76.50%
AP bikes	46.45%	60.38%
Reasonable min width @ 1080p	32px	
Max objects to detect	200	
Number of train images	~30K	~170K
GFlops	4.71	6.31
MParams	1.18	TBD

(FPS) Core(TM) i5-6500 CPU

Caffe CPU MKL	IE MKLDNN	IE clDNN FP16	IE clDNN FP32	OpenCV CPU
1.12	24.12	15.00	11.90	5.78

Movidius M2 NCS/MA2450

IE M2 FP16
1.98

Person Detection

Demo:

https://videoportal.intel.com/media/Smart+Video+Demo+Algorithms+%E2%80%93+People+Counting/0_aqws7cpk

- RFCN and SSD variants with PVANet and MobileNet backbones, respectively
- Pre-trained with OpenImages, COCO and Pascal
- Fine-tuned on videos from purchased datasets
- Validated on videos from customers
- Core of People Counting algorithm
  - Part of Audience Analytics product
  - Was demoed on 4 APL cameras at Winter Olympics'18 for 33 days straight with over 100K visitors

person-detection-retail-0013

Metric	Value
AP	81.41%
Occlusion coverage	<50%
Min height @ 1080p	100 pixels
Max objects to detect	200
GFlops	3.89
MParams	1.94
Detection head	SSD

(FPS) Core(TM) i5-6500 CPU

Caffe CPU MKL	IE MKLDNN	IE clDNN FP16	IE clDNN FP32	OpenCV CPU
8.02	59.67	39.00	30.82	21.18

Movidius M2 NCS/MA2450

IE M2 FP16
3.10

Face Detection

Demo:

- https://videoportal.intel.com/media/IOTG+Computer+Vision+%28ICV%29+Face+Detection+on+Skylake+%2B+Arria+10+GX/0_7oounvyn
- https://videoportal.intel.com/media/Driver+Monitoring+%E2%80%93+Driver+Head+Pose+Estimation+%E2%80%93+Live+Demo+Q4+2017/0_y7wkl0lk

- Two “Pareto-optimal” networks for Face Detection. Originally, for Retail and Software Defined Cockpit (SDC) scenarios, but work well in a lot of “generic” scenes
- SSD variants with SqueezeNet and MobileNet-inspired backbones
- Pre-trained with OpenImages, COCO and Pascal
- Fine-tuned on WiderFace, several purchased datasets and 1 home-grown “hard negatives” dataset (bags, hands, gestures, etc.)
- Tested on WiderFace Validation

Face Detection

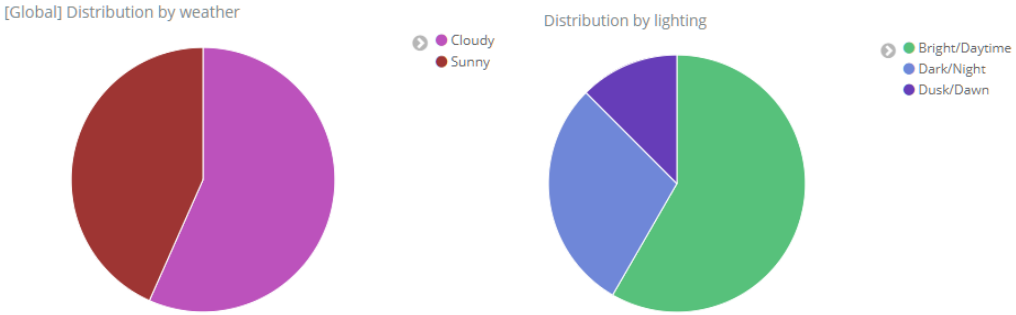
Metric	-adas-0001	-retail-0004
AP (head height >10px)	37.4%	29.71%
AP (head height >32px)	84.8%	71.41%
AP (head height >64px)	93.1%	84.77%
AP (head height >100px)	94.1%	88.82%
Backbone	MobileNet	SqueezeNet
GFlops	1.4	1.06
MParams	1.1	0.59

(FPS) Core(TM) i5-6500 CPU

	Caffe CPU MKL	IE MKLDNN	IE clDNN FP16	IE clDNN FP32	OpenCV CPU	IE M2 FP16
-retail-0004	31.59	215.62	151.13	131.77	116.14	19.41
-adas-0001	5.83	68.69	41.38	33.34	21.45	2.98
Ratio	x5.42	x3.14	x3.65	x3.95	x5.41	x6.51

Barrier scenario

Type of object	Number of bounding boxes
License plate	106,903
Vehicle	92,816



Demo:

https://videoportal.intel.com/media/Road+Barrier+on+Intel%C2%AE+CPU+GPU+%28NUC6i7KYB%29/0_7mkdq3nr

vehicle-license-plate-detection-barrier-0007

Metric	R1	R2
Mean Average Precision (mAP)	98.73%	98.49%
AP vehicles	98.36%	TBD
AP plates	99.10%	TBD
Car pose	Front facing cars	
Min plate width	96 pixels	
Max objects to detect	200	
GFlops	2.978	0.7
MParams	1.128	TBD

(FPS) Core(TM) i5-6500 CPU

Caffe CPU MKL	IE MKLDNN	IE clDNN FP16	IE clDNN FP32	OpenCV CPU
16.81	90.20	90.89	65.69	54.36

Movidius M2 NCS/MA2450

IE M2 FP16
11.12

Intel models – Object Recognition

Model name	Output type	Complexity (GFLOPs)	Size (Mp)
age-gender-recognition-retail-0013	category + regression	0.09	2.1
head-pose-estimation-adas-0001	3 x regression	0.03	0.9
license-plate-recognition-barrier-0001	string	0.34	1.9
vehicle-attributes-recognition-barrier-0039	2 x category	0.125	0.6

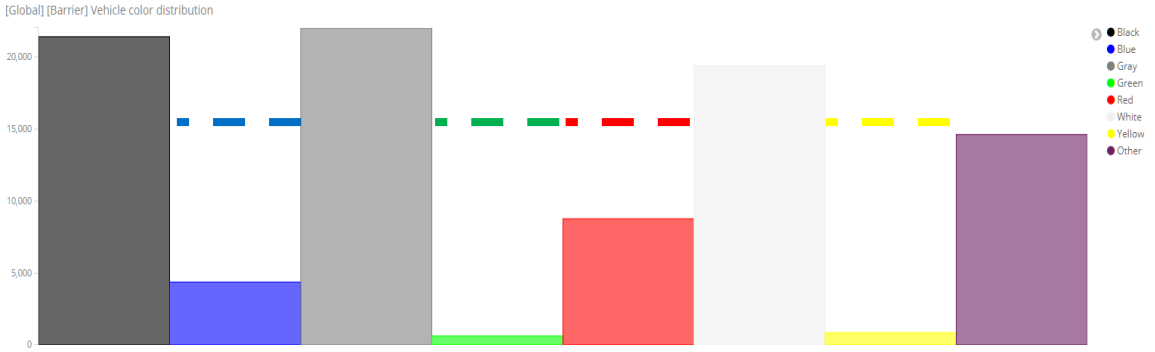
license-plate-recognition-barrier-0001

- Small-footprint network trained E2E to recognize Chinese License Plates in traffic scenarios
- Uses synthetic data generation to generate dataset with perfect character distribution
- ~10K plates to train style transfer network
- ~100K plates as style source

Rotation in-plane	±10˚
Rotation out-of-plane	Yaw: ±45˚ / Pitch: ±45˚
Min plate width	94 pixels
Ratio of correct reads	95%

vehicle-attributes-recognition-barrier-0039

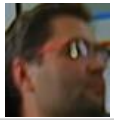
Car pose	Front facing cars
Occlusion coverage	<50%
Min object width	72 pixels
Supported colors	Blue, gray, yellow, green, black, white, red
Supported types	Car, van, truck, bus
Validation Dataset	BIT-Vehicle



	blue	gray	yellow	green	black	white	red
blue	75.95%	4.47%	0.00%	1.55%	18.04%	0.00%	0.00%
gray	1.97%	89.57%	0.00%	0.00%	1.66%	6.79%	0.00%
yellow	1.19%	9.52%	83.33%	2.38%	0.00%	2.38%	1.19%
green	10.46%	11.76%	1.96%	54.90%	12.42%	8.50%	0.00%
black	1.50%	1.47%	0.00%	0.41%	96.36%	0.25%	0.00%
white	1.67%	13.38%	0.00%	0.26%	0.00%	84.68%	0.00%
red	1.30%	0.26%	4.56%	0.13%	0.13%	0.13%	93.48%
						Average color accuracy:	
							82.61%

	car	van	truck	bus	
car	94.43%	3.61%	1.93%	0.04%	Average type accuracy:
van	5.78%	89.12%	4.08%	1.02%	87.85%
truck	1.71%	2.43%	92.86%	3.00%	
bus	0.00%	6.25%	18.75%	75.00%	

age-gender-recognition-0013

Input Image	Result
	Female, 18.97
	Male, 26.52
	Male, 33.41

Metric	Value
Rotation in-plane	±45˚
Rotation out-of-plane	Yaw: ±45˚ / Pitch: ±45˚
Min object width	62 pixels
GFlops	0.094
MParams	2.138

- Fully-convolutional network for simultaneous Age/Gender recognition
- Training dataset contains over 500K unique subjects with reasonable representation across different age, gender and ethnicity groups
- Validation dataset has ~20K subjects
- 96.66% gender recognition accuracy
- 6.07 years average age error

head-pose-estimation-adas-0001

Angle	Mean ± standard deviation
yaw	5.4 ± 4.4
pitch	5.5 ± 5.3
roll	4.6 ± 5.6

- Head pose estimation network based on trimmed SqueezeNet_v1.1
- Angle regression layers are convolutions (width: 128) + ReLU + batch norm + fully connected with one output
- 60x60px min face size
- Supported ranges:
  - Yaw: [-90, 90]
  - Pitch: [-70, 70]
  - Roll: [-70, 70]
- Trained and Validated on BIWI Kinect Head Pose

person-reidentification-retail-0076 (-0079)

Person re-identification models use whole body image as input and return 256-float embedding vectors as output. These can be used to match the pair of images by the Cosine distance.

Metric	Value
Pairwise accuracy	93.35% (92.93%)
Pose coverage	Standing upright
Validation dataset	10K pairs of 1.5K subjects
GFlops	0.58 (0.12)
MParams	0.82

semantic-segmentation-adas-0001

ICNet-based network for fine-grained multi-class segmentation in ADAS scenarios

Metric	Value
Input resolution	2048x1024
Validation dataset	2K images
Mean IOU	69.07%
Classes	road, sidewalk, building, wall, fence, pole, traffic light, traffic sign, vegetation, terrain, sky, person, rider, car, truck, bus, train, motorcycle, bicycle, ego-vehicle
GFlops	29.04
MParams	6.6