分类模型性能对比

opencv load image costs 3.54 ms （imgsize 700*430）

系列	模型名字	Image Net Top1	Image Net Top5	FLOPS	其他速度描述	参数量	Github	是否支持caffe（or sensenet）finetune（BN未合并）	是否支持sdk cpu 打包（i7 速度）	是否支持sdk gpu 打包（1080 速度）	是否支持sdk tensorRT 打包（1080 速度，batch 为1 以及 batch>1最优平均速度）ms per image	是否支持sdk mobile 打包（xx 速度）
CondenseNet	CondenseNet-74 (C=G=4)	73.8	91.7	529M		4.8M						
	CondenseNet-74 (C=G=8)	71.0	90.0	274M		2.9M						
PeleeNet	ICLR2018	71.3										
	盂天健复现	71.6	90.3	510M		2.8M					0.379ms(batch256)	
VGG	VGG-16	75.3	92.5									
	VGG-16 (ref GoogleNet)		92.7									
	VGG-16 (ref ResNet)	71.9	90.7									
	VGG-16 (ref MobileNet)	71.5	90.1	15300M		138M						
SqueezeNet	SqueezeNet (ref MobileNet)	57.7		1700M		1.25M						
	SqueezeNet	57.5	80.3	833M		4.8M						
AlexNet	AlexNet	57.2	80.3	720M	高通骁龙820 224x224：184.0 ms 480x640：1156.7 ms 720x1280：3633.9 ms	60M						
GoogleNet	GoogleNet(ref MobileNet)	69.8		1550M		6.8M						
	GoogleNet		93.3									
	GoogleNet (by BVLC)	68.7		1500M								
Inception	Inception-V1	69.8	89.9	1448M		6.6M						
	Inception-V2	74.8	92.2	1940M		11.2M						
	Inception-V3	78.8	94.4									
	Inception-V3 (ref Xception)	78.2	94.1		每秒31次迭代	23.6M						
	Inception-V3-299	78.0	93.9	5720M		23.8M						
	Inception-ResNet-v2	80.1	95.1									
	Inception-ResNet-v2(ref SENet)	79.6	94.8	11750M								
Xception	Xception-299	79.0	94.5	8380M	每秒28次迭代	22.9M						
Fast	Fast-v1.0							支持sensenet合并bn	46.7ms	1.02ms (batch128)	0.47ms(batch128)	
	Fast-v1.1							支持sensenet合并bn	16.3ms	0.51ms (batch256)	0.24ms(batch256)	
ResNet	ResNet-50(ref ResNeXt)	76.1	92.9	4100M		25.5M		支持caffe合并batch norm	149.7ms	2.17ms (batch64)	1.25ms(batch64)	
	ResNet-101(ref ResNeXt)	78.0	94.0	7800M	在M40上0.70s							
	ResNet-152	77.0	93.3									
	ResNet-200	78.3	94.2									
	1/2 ResNet-50 (手动剪枝, ref 家敏)	71.8		¯1100M		7.9M		支持caffe合并batch norm	47.5ms	1.04ms (batch128)	0.47ms(batch128)	
	1/2 Res50(手动剪枝, ref诗涛)[1]	72.05										
	1/4 ResNet-50 (手动剪枝, ref 家敏)	66.9		¯300M		2M		支持caffe合并batch norm	17.7ms	0.51ms (batch256)	0.24ms(batch256)	
ResNeXt	ResNeXt-50 (2x40d)	77.0										
	ResNeXt-50 (4x24d)	77.4										
	ResNeXt-50 (8x14d)	77.7										
	ResNeXt-50(32x4d)	77.8		4100M		25.0M						

	ResNeXt-101 (2x40d)	78.3								
	ResNeXt-101 (2x64d)	79.3	94.5							
	ResNeXt-101 (4x24d)	78.6								
	ResNeXt-101 (8x14d)	78.7								
	ResNeXt-101 (32x4d)	78.8	94.4	7800M	在M40上 0.95s					
	ResNeXt-101 (64x4d)	79.6	94.7							
	ResNeXt-101 (64x4d)-320	80.9	95.6	31500M		83.6M				
Mobile Net	Conv MobileNet	71.7		4866M		29.3M				
	1.0 MobileNet-V1-224	70.6	89.5 (NASNet)	569M	手机 CPU123ms ———————————— 高通骁龙820 224x224:110.0 ms 480x640:612.0 ms 720x1280: 1879.2 ms 113ms (Google Pixel)*	4.2M	支持caffe合并batch norm	27.6ms	暂不支持	0.53ms（batch256）
	1.0 MobileNet-V1-192	69.1		418M		4.2M				
	1.0 MobileNet-V1-160	67.2		290M		4.2M				
	1.0 MobileNet-V1-128	64.4		186M		4.2M				
	0.75 MobileNet-V1-224	68.4		325M		2.6M				
	0.5 MobileNet-V1-224	63.7		149M		1.3M				
	0.5 MobileNet-V1-160	60.2		76M		1.32M				
	0.25 MobileNet-V1-224	50.6		41M		0.5M				
	Shallow MobileNet	65.3		307M		2.9M				
	MobileNet-V2	72.0		300M	手机 CPU80ms ———————————— 75ms (Google Pixel)*	3.4M		51.2ms	139.3ms （batch128）	0.72ms（batch256）
	MobileNet-V2 (1.4)	74.7		585M	手机 CPU149ms ———————————— 143ms (Google Pixel)*	6.9M				
Shuffl eNet	ShuffleNet 2x (arxiv:v1) (NasNet)	70.9	89.8	524M		5M				
	ShuffleNet 2x (g=3)	73.7		524M	高通骁龙820 224x224:108.8 ms 480x640:617.0 ms 720x1280: 1857.6 ms	5.4M				
	ShuffleNet 2x (g=3, SE)	75.3		527M						
	ShuffleNet 1.5x (g=3)	71.5		292M		3.4M				

Category	Model	Top-1	Top-5	FLOPs	Device / Timing	Params	GPU	Batch	Note
	ShuffleNet 1x (g=3)	67.4		140M	高通骁龙820 224x224：37.8ms 480x640：222.2ms 720x1280：684.5ms		17.2 ms	40.6ms（batch256）	不支持shuffle
	ShuffleNet 1x (g=8)	67.6		140M					
	ShuffleNet 0.5x (g=3)	56.8		38M	高通骁龙820 224x224：15.2ms 480x640：87.4ms 720x1280：260.1ms				
	ShuffleNet 0.5x (g=3,shallow)	57.2		40M					
	ShuffleNet 0.5x (g=4)	58.4		38M					
	ShuffleNet 0.5x (g=8)	57.7							
	ShuffleNet 0.25x (g=3)	45.0							
	ShuffleNet 0.25x (g=8)	47.3							
NASNet	NASNet-A (4@1056)	74.0	91.6	564M		5.3M			
	NASNet-B (4@1536)	72.8	91.3	488M		5.3M			
	NASNet-C (3@960)	72.5	91.0	558M		4.9M			
	NASNet-A (5@1538)-299	78.6	94.2	2350M		10.9M			
	NASNet-A (7@1920)	80.8	95.3	4930M		22.6M			
	NASNet-A (6@4032) 331x331	82.7	96.2	23800M		88.9M			
DenseNet	DenseNet-121	75.0	92.3						
	DenseNet-169	76.2	93.2						
	DenseNet-201	77.4	93.7						
	DenseNet-264	77.9	93.9						
SENet	SENet	81.3	95.5						
	SE-ResNet-50	76.7	93.4	3870M					
	SE-ResNet-101	77.6	93.9	7600M					
	SE-ResNet-152	78.4	94.3	11320M					
	SE-ResNeXt-50	78.9	94.5	4250M					
	SE-ResNeXt-101	79.3	95.0	8000M					
	SE-BN-Inception	75.8	92.9	2040M					
	SE-Inception-ResNet-V2	80.2	95.2	11760M					

1. 超参设置，SGD，momentum 0.9，weight_decay 1e-4，base_lr 0.1，每30个epoch降一次lr，data aug: random corp, random flip, color augment, normalize. 8 卡共256 batchsize