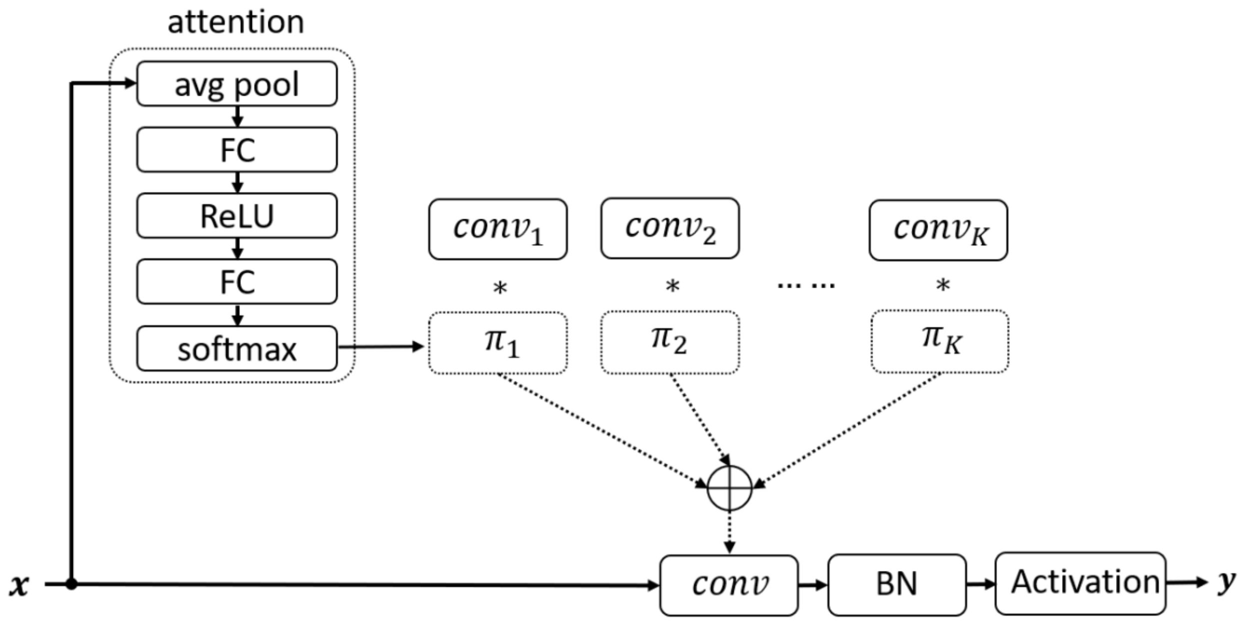
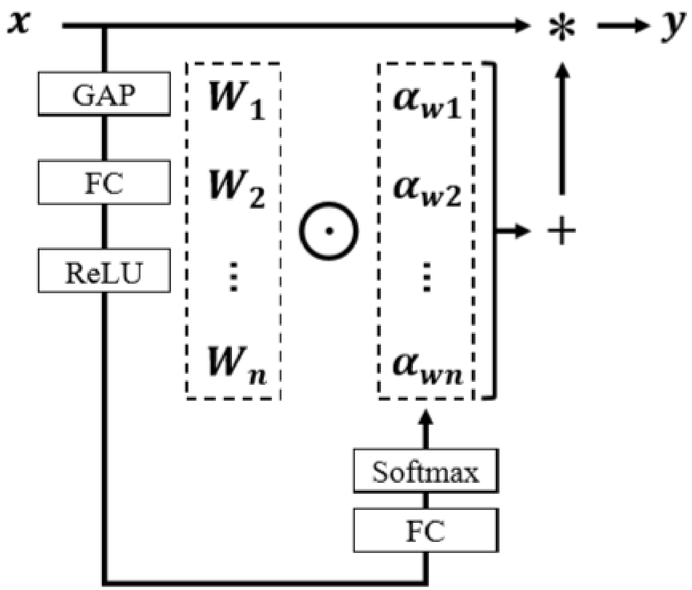
OMNI-DIMENSIONAL DYNAMIC CONVOLUTION

Chao Li[1,*], **Aojun Zhou**[2], **Anbang Yao**[1†]
[1]Intel Labs China, [2]CUHK-SenseTime Joint Lab, The Chinese University of Hong Kong
chao.li3@intel.com, aojun.zhou@gmail.com, anbang.yao@intel.com

ICLR2022

Review of dynamic convolution





learning a linear combination of N convolutional kernels weighted with their input-dependent attentions

$$y = (\alpha_{w1} W_1 + ... + \alpha_{wn} W_n) * x$$

Limitation

[N, in_channel, out_channel, k, k]

- Firstly

Only one dimension

regarding the convolutional kernel number

While the other three dimensions are overlooked

regarding the spatial size, the input channel number and the output channel number for each convolutional kernel

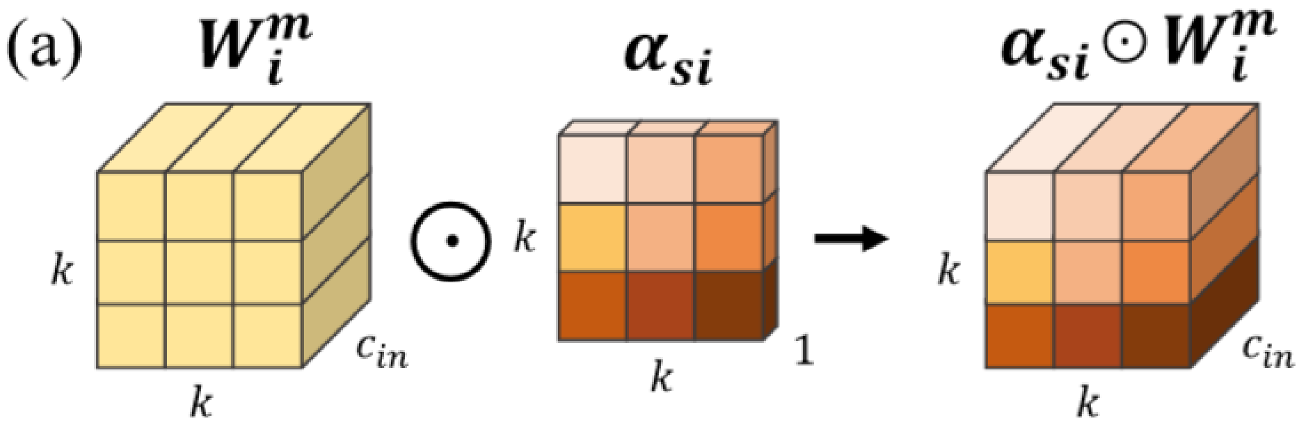
→ limiting their abilities to capture rich contextual cues

Limitation

- Secondly

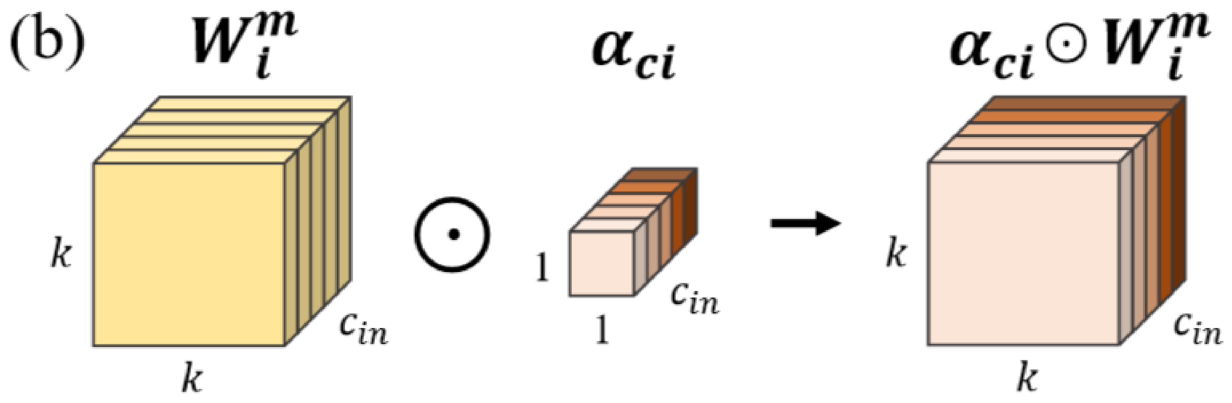
dynamic convolution increases the number of convolutional parameters by N times

[N, in_channel, out_channel, k, k]



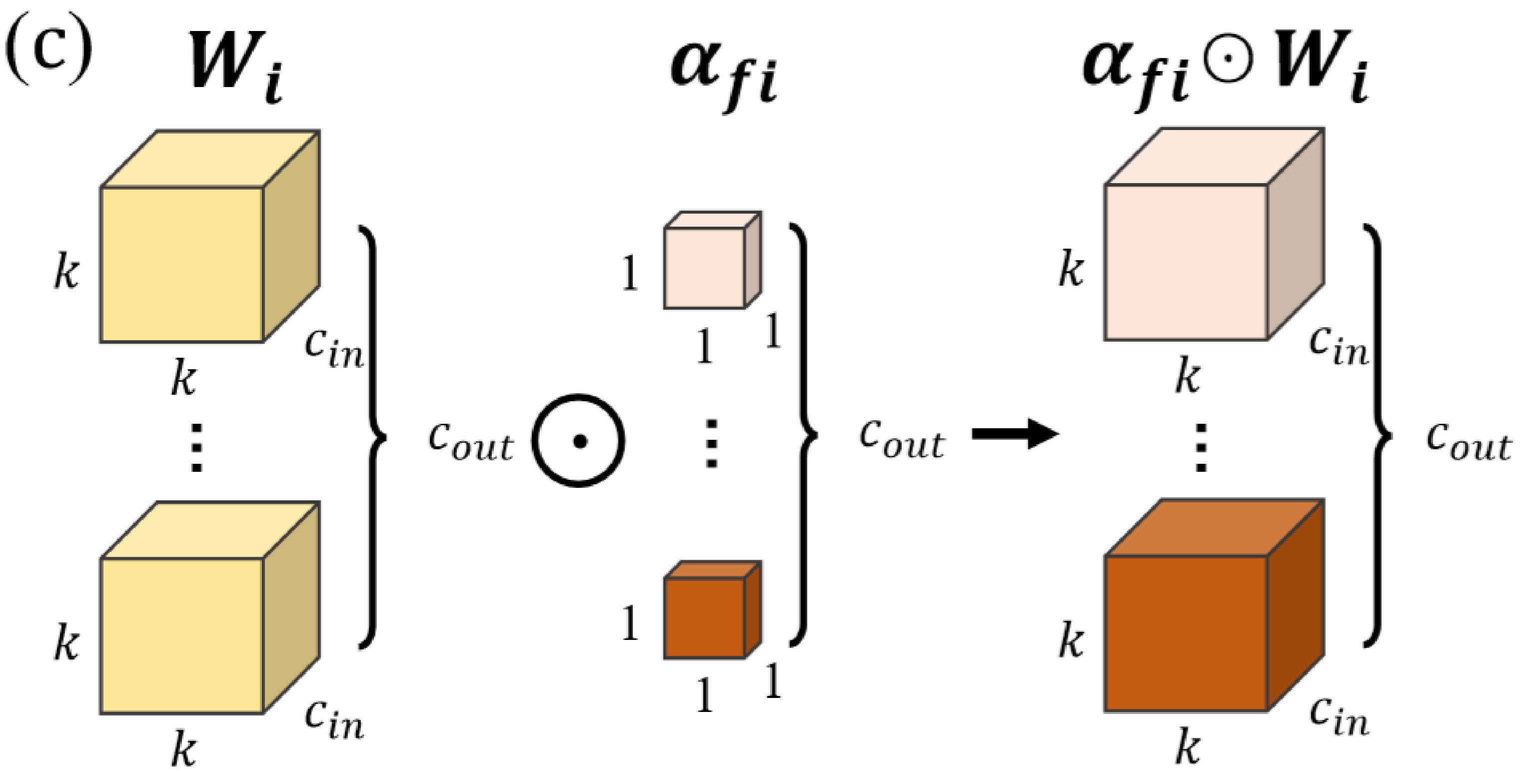
α_{si} assigns different attention scalars to convolutional parameters (per filter) at k * k spatial locations

[K, in_channel, out_channel, W, H]



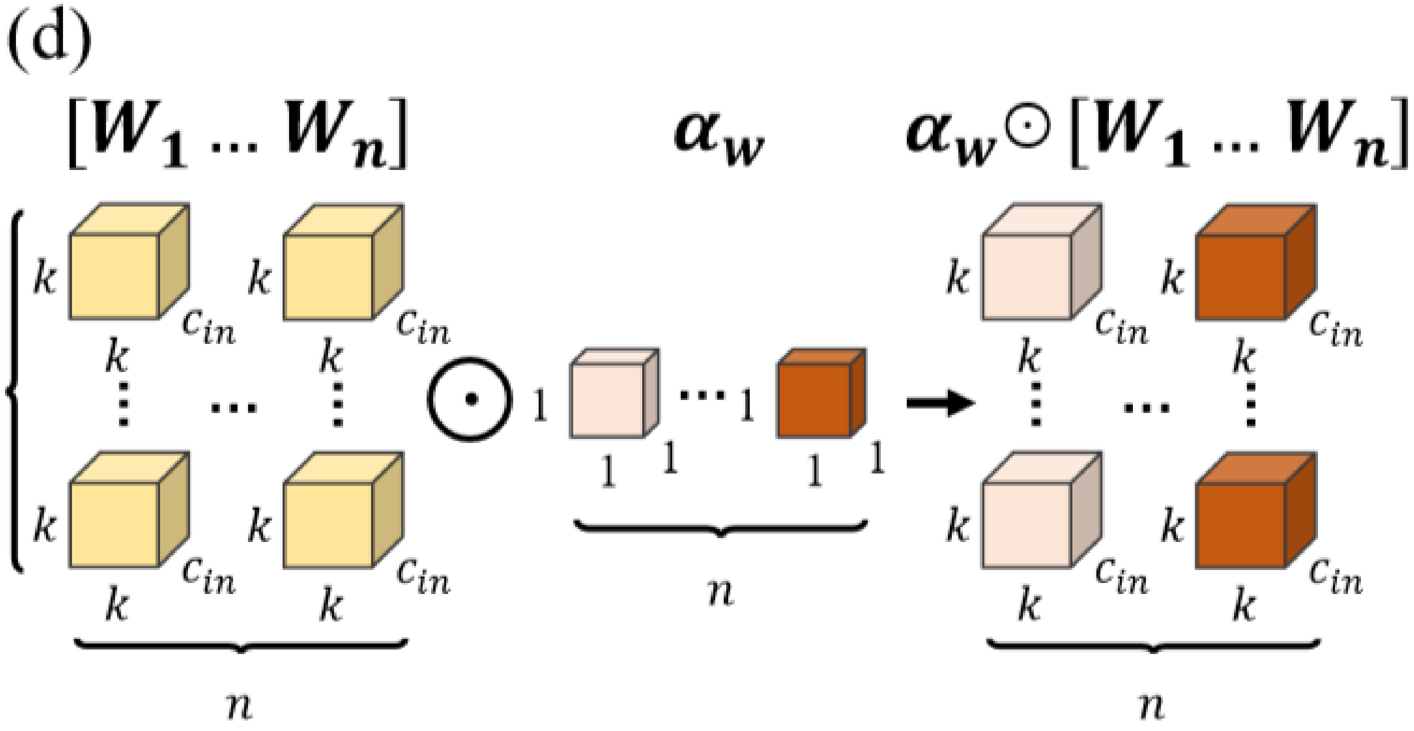
α_{ci} assigns different attention scalars to c_{in} channels of each convolutional filter W_i^m

[K, in_channel, out_channel, W, H]

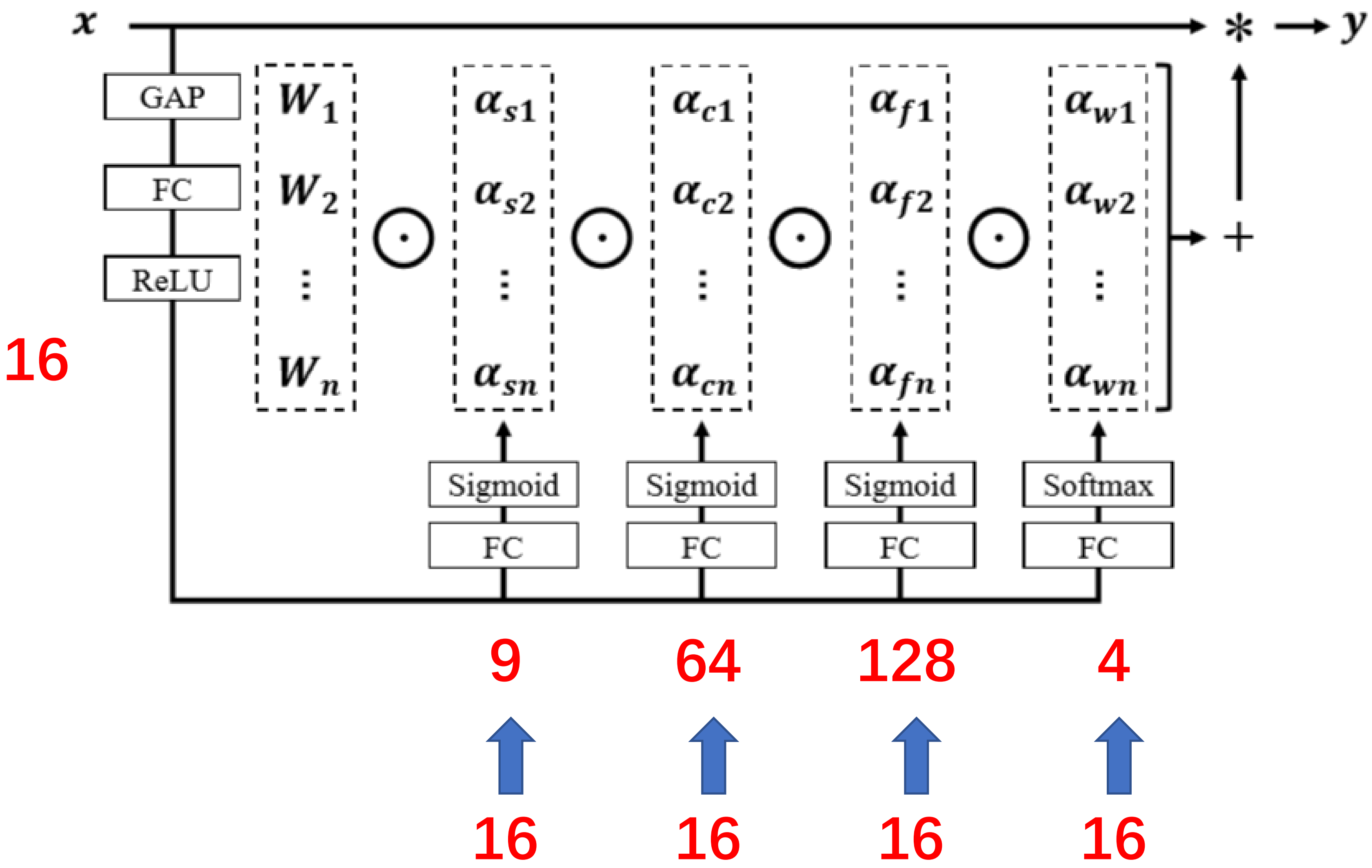


α_{fi} assigns different attention scalars to c_{out} convolutional filters

[K, in_channel, out_channel, W, H]



α_{wi} assigns an attention scalar to the whole convolutional kernel


x
y
GAP
FC
ReLU
16
W1
W2
Wn
αs1
αs2
αsn
αc1
αc2
αcn
αf1
αf2
αfn
αw1
αw2
αwn
Sigmoid
Sigmoid
Sigmoid
Softmax
FC
FC
FC
FC
9
64
128
4
16
16
16
16

Mobilenet

Models	Params	MAdds	Top-1 Acc (%)	Top-5 Acc (%)
MobileNetV2 (1.0×)	3.50M	300.8M	71.65	90.22
+ CondConv (8×)	22.88M	318.1M	74.13 (↑2.48)	91.67 (↑1.45)
+ DyConv (4×)	12.40M	317.1M	74.94 (↑3.29)	91.83 (↑1.61)
+ DCD	5.72M	318.4M	74.18 (↑2.53)	91.72 (↑1.50)
+ ODConv (1×)	4.94M	311.8M	74.84 (↑3.19)	92.13 (↑1.91)
+ ODConv (4×)	11.52M	327.1M	**75.42** (↑**3.77**)	**92.18** (↑**1.96**)
MobileNetV2 (0.75×)	2.64M	209.1M	69.18	88.82
+ CondConv (8×)	17.51M	223.9M	71.79 (↑2.61)	90.17 (↑1.35)
+ DyConv (4×)	7.95M	220.1M	72.75 (↑3.57)	90.93 (↑2.11)
+ DCD	4.08M	222.9M	71.92 (↑2.74)	90.20 (↑1.38)
+ ODConv (1×)	3.51M	217.1M	72.43 (↑3.25)	90.82 (↑2.00)
+ ODConv (4×)	7.50M	226.3M	**73.81** (↑**4.63**)	**91.33** (↑**2.51**)
MobileNetV2 (0.5×)	2.00M	97.1M	64.30	85.21
+ CondConv (8×)	13.61M	110.0M	67.24 (↑2.94)	87.51 (↑2.30)
+ DyConv (4×)	4.57M	103.2M	69.05 (↑4.75)	88.37 (↑3.16)
+ DCD	3.06M	105.6M	69.32 (↑5.02)	88.44 (↑3.23)
+ ODConv (1×)	2.43M	101.8M	68.26 (↑3.9.6)	87.98 (↑2.77)
+ ODConv (4×)	4.44M	106.4M	**70.01** (↑**5.71**)	**89.01** (↑**3.80**)

Resnet

Network	ResNet18				ResNet50			
Models	Params	MAdds	Top-1 Acc (%)	Top-5 Acc (%)	Params	MAdds	Top-1 Acc (%)	Top-5 Acc (%)
Baseline	11.69M	1.814G	70.25	89.38	25.56M	3.858G	76.23	93.01
+ CondConv (8×)	81.35M	1.894G	71.99 (↑1.74)	90.27 (↑0.89)	129.86M	3.978G	76.70 (↑0.47)	93.12 (↑0.11)
+ DyConv (4×)	45.47M	1.861G	72.76 (↑2.51)	90.79 (↑1.41)	100.88M	3.965G	76.82 (↑0.59)	93.16 (↑0.15)
+ DCD	14.70M	1.841G	72.33 (↑2.08)	90.65 (↑1.27)	29.84M	3.944G	76.92 (↑0.69)	93.46 (↑0.45)
+ ODConv (1×)	11.94M	1.838G	73.10 (↑2.85)	91.10 (↑1.72)	28.64M	3.916G	77.96 (↑1.73)	93.84 (↑0.83)
+ ODConv (4×)	44.90M	1.916G	**73.97** (↑**3.72**)	**91.35** (↑**1.97**)	90.67M	4.080G	**78.52** (↑**2.29**)	**94.01** (↑**1.00**)
+ SE	11.78M	1.816G	70.98 (↑0.73)	90.03 (↑0.65)	28.07M	3.872G	77.31 (↑1.08)	93.63 (↑0.62)
+ CBAM	11.78M	1.818G	71.01 (↑0.76)	89.85 (↑0.47)	28.07M	3.886G	77.46 (↑1.23)	93.59 (↑0.58)
+ ECA	11.69M	1.816G	70.60 (↑0.35)	89.68 (↑0.30)	25.56M	3.870G	77.34 (↑1.11)	93.64 (↑0.63)
+ CGC	11.69M	1.827G	71.60 (↑1.35)	90.35 (↑0.97)	25.59M	3.877G	76.79 (↑0.56)	93.37 (↑0.36)
+ WeightNet	11.93M	1.826G	71.56 (↑1.31)	90.38 (↑1.00)	30.38M	3.885G	77.51 (↑1.28)	93.69 (↑0.68)
+ WE (*)	11.90M	1.820G	71.00 (↑0.75)	90.00 (↑0.62)	28.10M	3.860G	77.10 (↑0.87)	93.50 (↑0.49)

Models	Params	MAdds	Top-1 Acc (%)	Top-5 Acc (%)
ResNet101	44.55M	7.570G	77.41	93.67
+ SE	49.29M	7.593G	78.42 (↑1.01)	94.15 (↑0.48)
+ CBAM	49.30M	7.617G	78.50 (↑1.09)	94.20 (↑0.53)
+ ECA	44.55M	7.590G	78.60 (↑1.19)	94.34 (↑0.67)
+ ODConv (1×)	50.82M	7.675G	78.98 (↑1.57)	94.38 (↑0.71)
+ ODConv (2×)	90.44M	7.802G	**79.27** (↑**1.86**)	**94.47** (↑**0.80**)

MS-COCO 2017

Backbone Models	Detectors	$AP(\%)$	$AP_{50}(\%)$	$AP_{75}(\%)$	$AP_S(\%)$	$AP_M(\%)$	$AP_L(\%)$	Params	MAdds
ResNet50	Faster R-CNN	37.2	57.8	40.4	21.5	40.6	48.0	43.80M (23.51M)	207.07G (76.50G)
+ CondConv (8×)		38.1	58.9	41.5	22.4	42.1	48.7	133.75M (113.46M)	207.08G (76.51G)
+ DyConv (4×)		38.3	59.7	41.6	22.6	42.3	49.4	119.12M (98.83M)	207.23G (76.66G)
+ DCD		38.1	59.3	41.3	21.9	42.0	49.5	48.08M (27.79M)	207.20G (76.63G)
+ ODConv (1×)		39.0	60.5	42.3	**23.4**	42.3	50.5	46.88M (26.59M)	207.18G (76.61G)
+ ODConv (4×)		**39.2**	**60.7**	**42.6**	23.1	**42.6**	**51.0**	108.91M (88.62M)	207.42G (76.85G)
MobileNetV2 (1.0×)		31.3	51.1	33.1	17.4	33.5	41.2	21.13M (2.22M)	122.58G (24.45G)
+ CondConv (8×)		33.7	54.9	35.6	19.3	36.4	43.7	31.54M (12.63M)	122.59G (24.46G)
+ DyConv (4×)		34.5	55.6	36.5	19.8	37.3	44.7	30.02M (11.12M)	123.01G (24.88G)
+ DCD		33.3	53.0	35.1	19.9	36.1	43.2	23.34M (4.44M)	123.01G (24.88G)
+ ODConv (1×)		34.3	55.6	36.5	**20.7**	37.3	44.5	22.56M (3.66M)	123.00G (24.87G)
+ ODConv (4×)		**35.1**	**56.7**	**37.0**	20.6	**38.0**	**45.2**	29.14M (10.24M)	123.02G (24.89G)
ResNet50	Mask R-CNN	38.0	58.6	41.5	21.6	41.5	49.2	46.45M (23.51M)	260.14G (76.50G)
+ CondConv (8×)		38.8	59.3	42.3	22.5	42.5	50.3	136.4M (113.46M)	260.15G (76.51G)
+ DyConv (4×)		39.2	60.3	42.5	23.0	42.9	51.4	121.77M (98.83M)	260.30G (76.66G)
+ DCD		38.8	59.8	42.2	23.1	42.7	49.8	50.73M (27.79M)	260.27G (76.63G)
+ ODConv (1×)		39.9	61.2	43.5	23.6	**43.8**	**52.3**	49.53M (26.59M)	260.25G (76.61G)
+ ODConv (4×)		**40.1**	**61.5**	**43.6**	**24.0**	43.6	**52.3**	111.56M (88.62M)	260.49G (76.85G)
MobileNetV2 (1.0×)		32.2	52.1	34.2	18.4	34.4	42.4	23.78M (2.22M)	175.66G (24.45G)
+ CondConv (8×)		34.4	55.4	36.6	19.8	36.9	44.6	34.19M (12.63M)	175.67G (24.46G)
+ DyConv (4×)		35.2	56.2	37.5	**20.7**	38.0	45.5	32.68M (11.12M)	176.09G (24.88G)
+ DCD		34.3	54.9	36.6	20.6	37.1	44.8	26.00M (4.44M)	176.09G (24.88G)
+ ODConv (1×)		35.0	56.1	37.3	19.9	37.7	**46.2**	25.22M (3.66M)	176.08G (24.87G)
+ ODConv (4×)		**35.8**	**57.0**	**38.1**	20.5	**38.5**	45.9	31.80M (10.24M)	176.10G (24.89G)

ABLATION STUDIES

different numbers of convolutional kernels n

Models	n	Params	MAdds	Top-1 Acc (%)	Top-5 Acc (%)
ResNet18	-	11.69M	1.814G	70.25	89.38
+ ODConv ($r = 1/16$)	1×	11.94M	1.838G	73.10	91.10
	2×	22.93M	1.872G	73.59	91.08
	3×	33.92M	1.894G	73.77	91.35
	4×	44.90M	1.916G	73.97	91.35
	8×	88.84M	2.006G	**74.08**	**91.44**

ABLATION STUDIES

Investigating the complementarity of four types of attentions

Models	α_{si}	α_{ci}	α_{fi}	α_{wi}	Params	MAdds	Top-1 Acc (%)	Top-5 Acc (%)
ResNet18	-	-	-	-	11.69M	1.814G	70.25	89.38
+ ODConv ($r = 1/4$)	✓	-	-	-	11.98M	1.827G	72.42	90.76
	-	✓	-	-	12.26M	1.827G	72.07	90.59
	-	-	✓	-	12.28M	1.827G	71.46	90.43
	✓	✓	-	-	12.27M	1.827G	73.13	91.14
	✓	-	✓	-	12.29M	1.827G	72.80	90.99
	-	✓	✓	-	12.57M	1.829G	72.20	90.67
	✓	✓	✓	-	12.58M	1.839G	73.41	91.29
	✓	✓	✓	✓	45.54M	1.953G	**74.33**	**91.53**

ABLATION STUDIES

inference speed

ResNet50	Speed on GPU	Speed on CPU
Baseline model	652.2	6.4
+ CondConv	425.1	4.0
+ DyConv	326.6	3.8
+ DCD	400.0	3.6
+ ODConv (1×)	293.9	3.9
+ ODConv (4×)	152.7	2.5

MobileNetV2 (1.0×)	Speed on GPU	Speed on CPU
Baseline model	1452.3	17.3
+ CondConv	1076.9	14.8
+ DyConv	918.2	11.9
+ DCD	875.7	9.5
+ ODConv (1×)	1029.2	12.8
+ ODConv (4×)	608.3	11.2

GPU： NVIDIA TITAN X (with batchsize 200)
CPU： Intel E5-2683 v3 (with batch size 1)