NASNet

Cho Sung Man

Related Works

AutoML

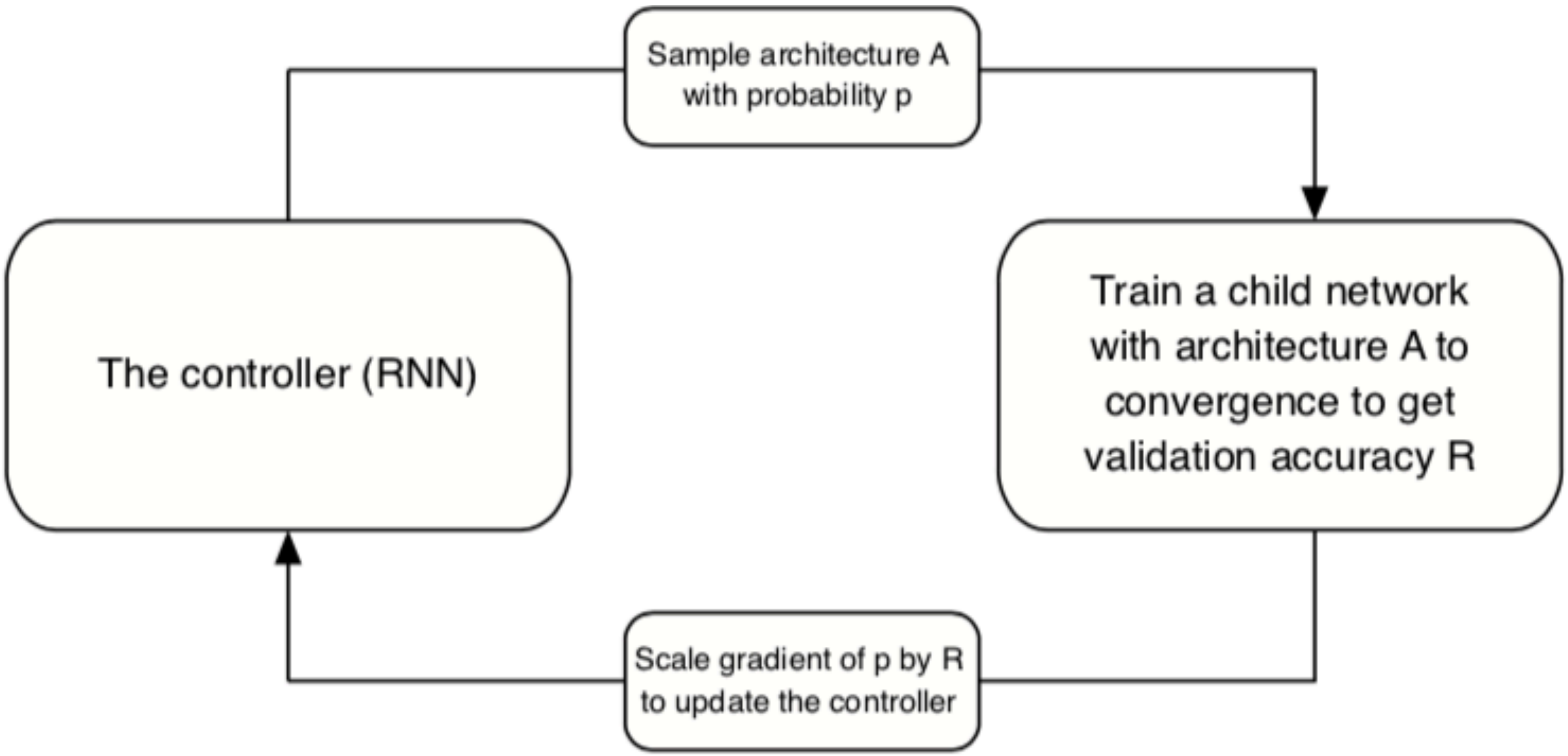
- Automated Feature Learning
- **Architecture Search**
- Hyperparameter Optimization

NAS

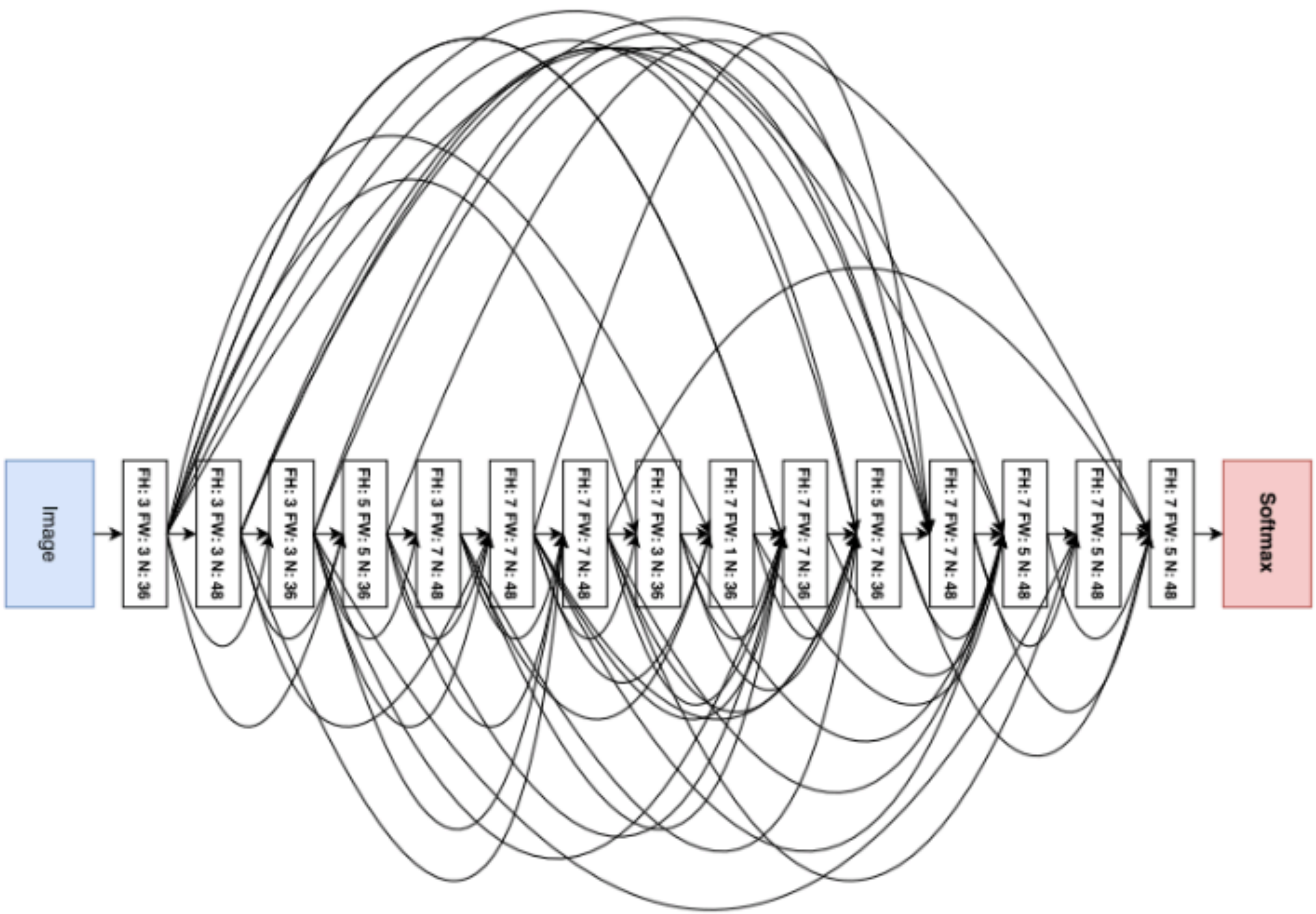
- Neural Architecture Search with reinforcement learning
- RNN predicts convolutional layer's filter size, stride.
- reward: RNN Controller's output (validation accuracy)
- training period : 800 GPU, 1 month.
- results : better than ResNET in **CIFAR-10.**

How about ImageNet ?

NAS


Sample architecture A
with probability p
The controller (RNN)
Train a child network
with architecture A to
convergence to get
validation accuracy R
Scale gradient of p by R
to update the controller

NAS



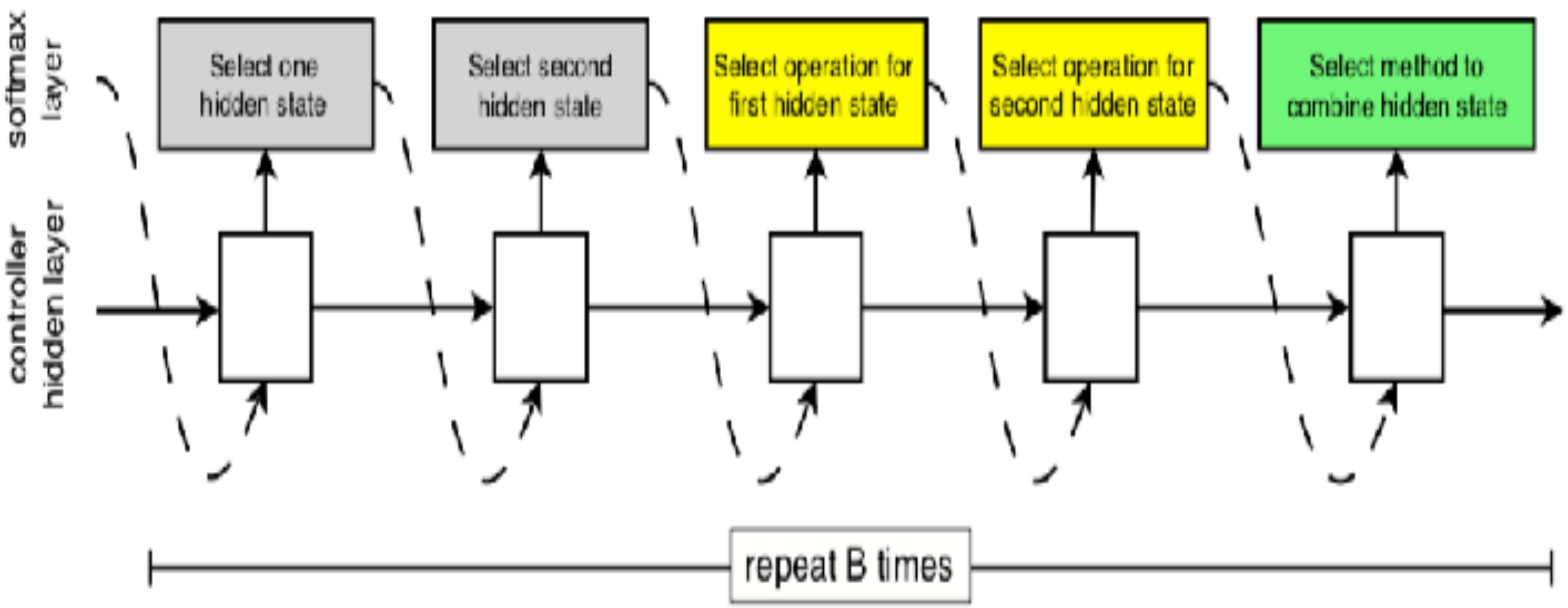
Too Large Search Space

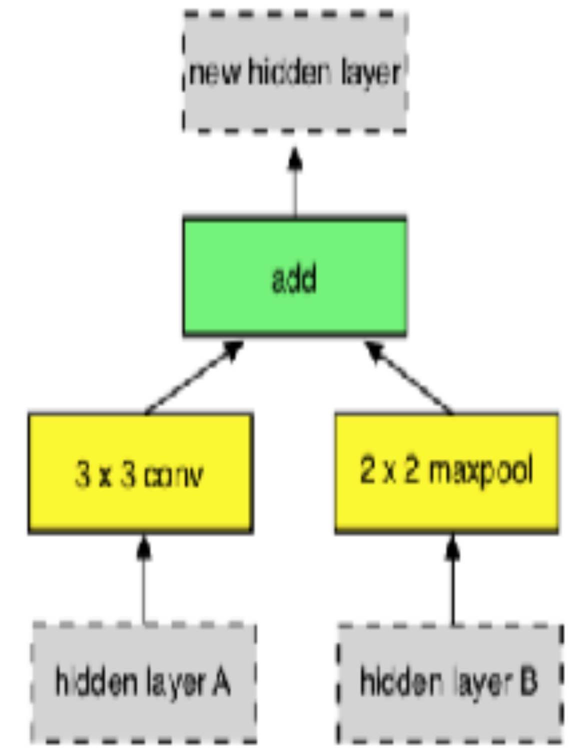
NASNet

NASNet

- Transferable Architecture Search.
- Small Search Space
- Inspired by Inception, ResNet. ‘CELL’
- 500 GPU, 4days

NASNet (Block)





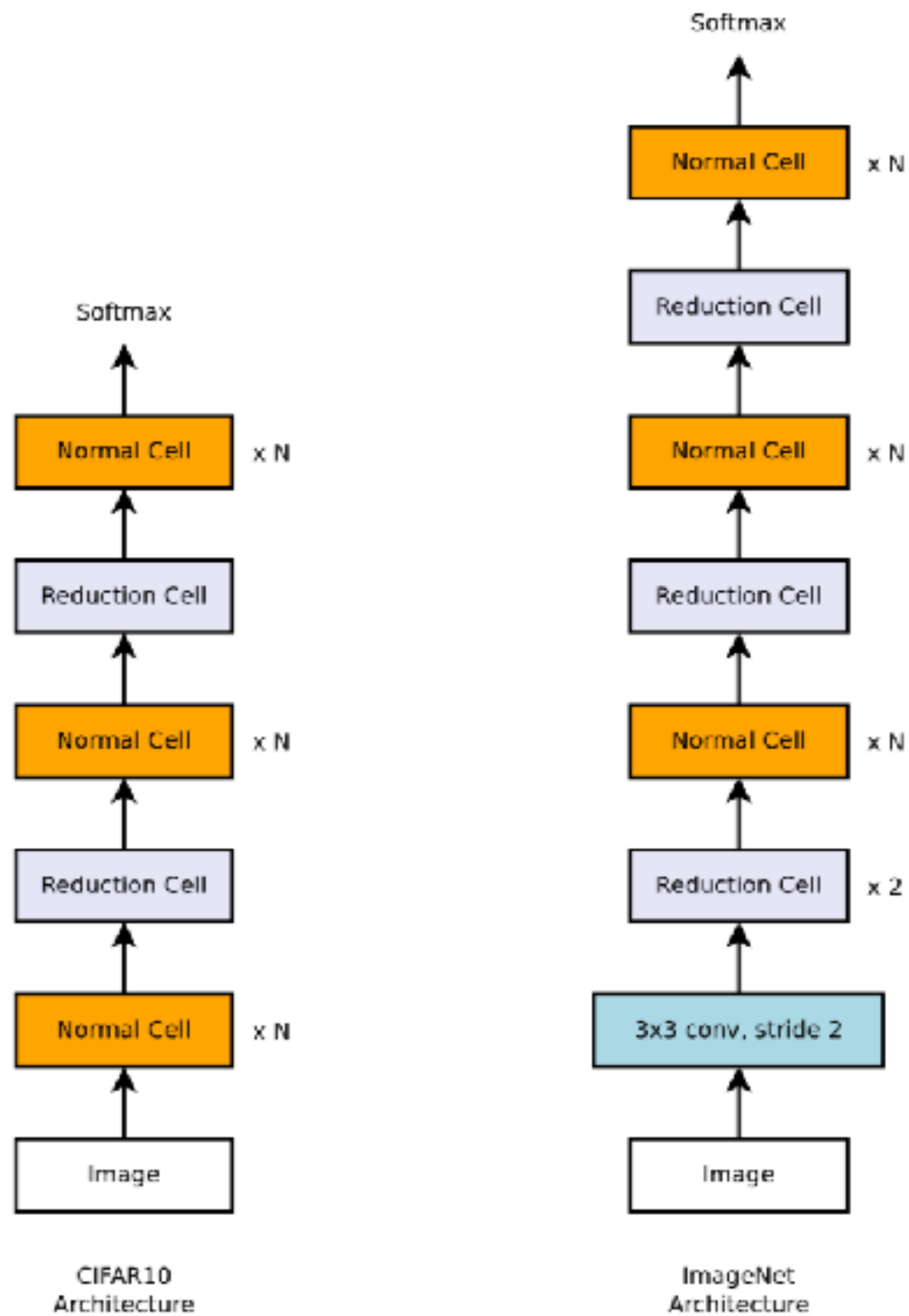
- h_i
- h_{i-1}
- $h_{i,0}$
- $h_{i,1}$
- $h_{i,2}$
- $h_{i,3}$

- Identity
- conv 1x7 + 7x1
- conv 1x3 + 3x1
- avg 3x3
- max 3x3
- max 5x5
- max 7x7
- conv 1x1
- conv 3x3
- sep 3x3
- sep 5x5
- sep 7x7
- dilated 3x3

- add
- concat

Results : #B blocks -> 1 Cell

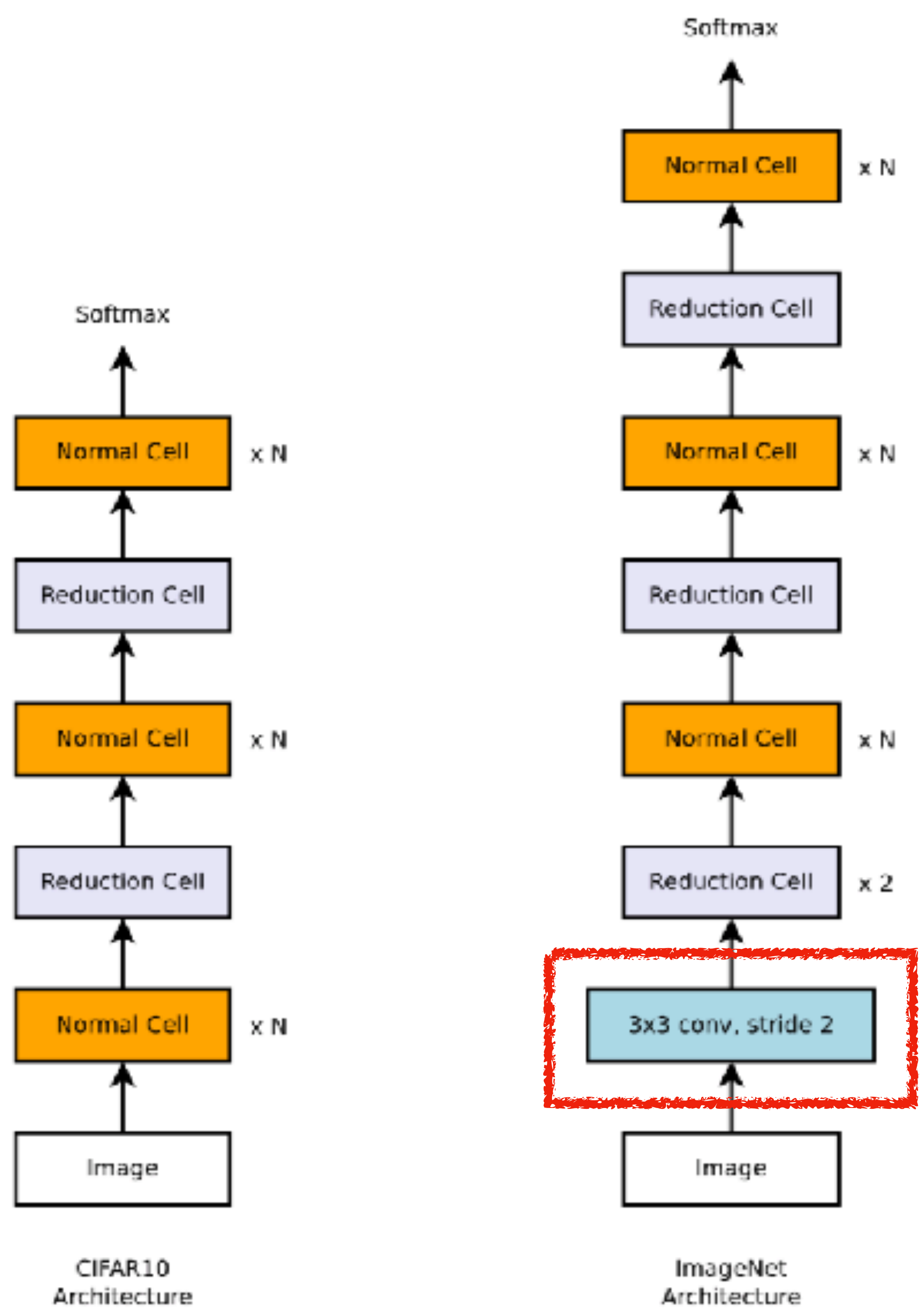
NASNet (Conv cell)



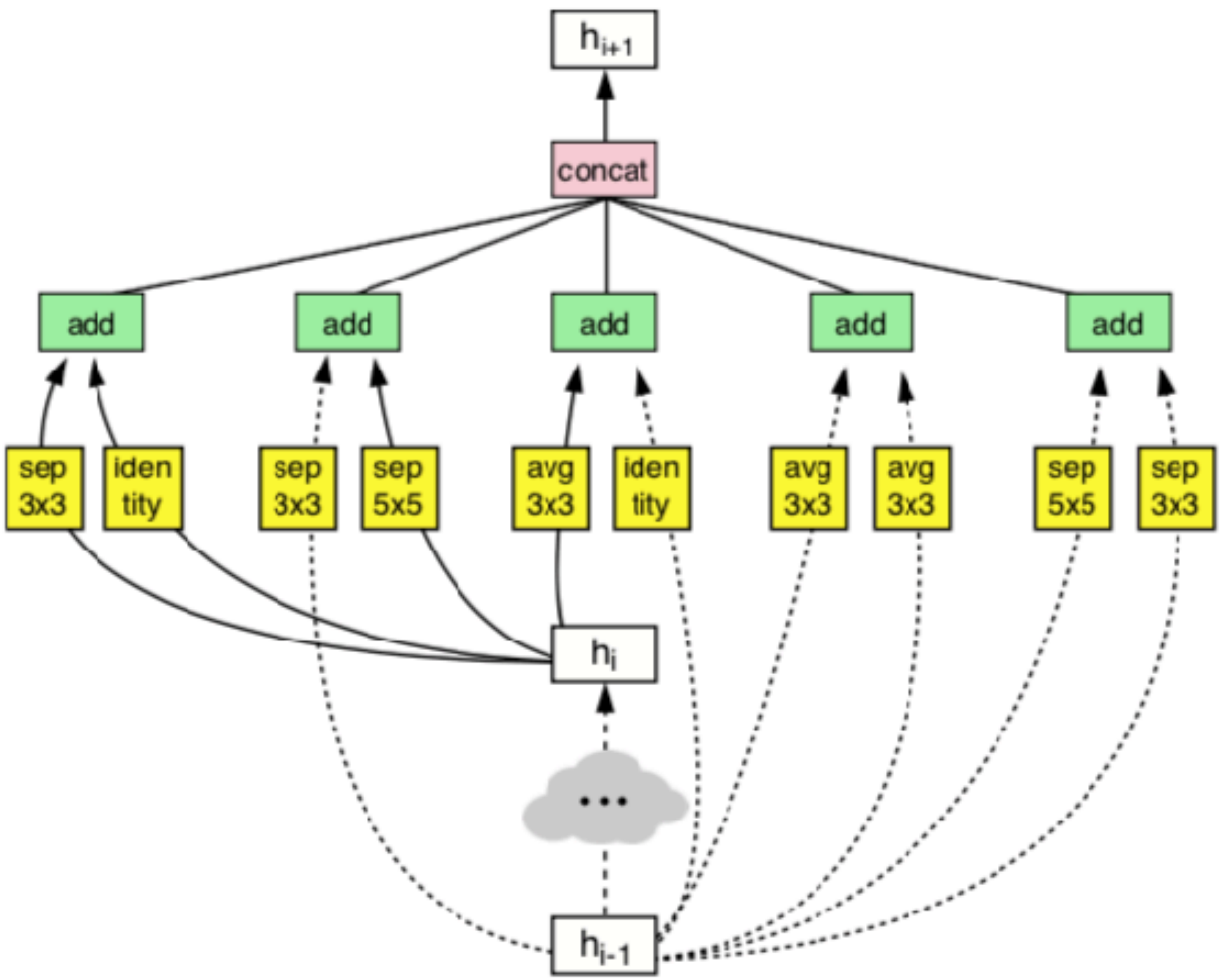
stride : 1 or 2

stride : 1

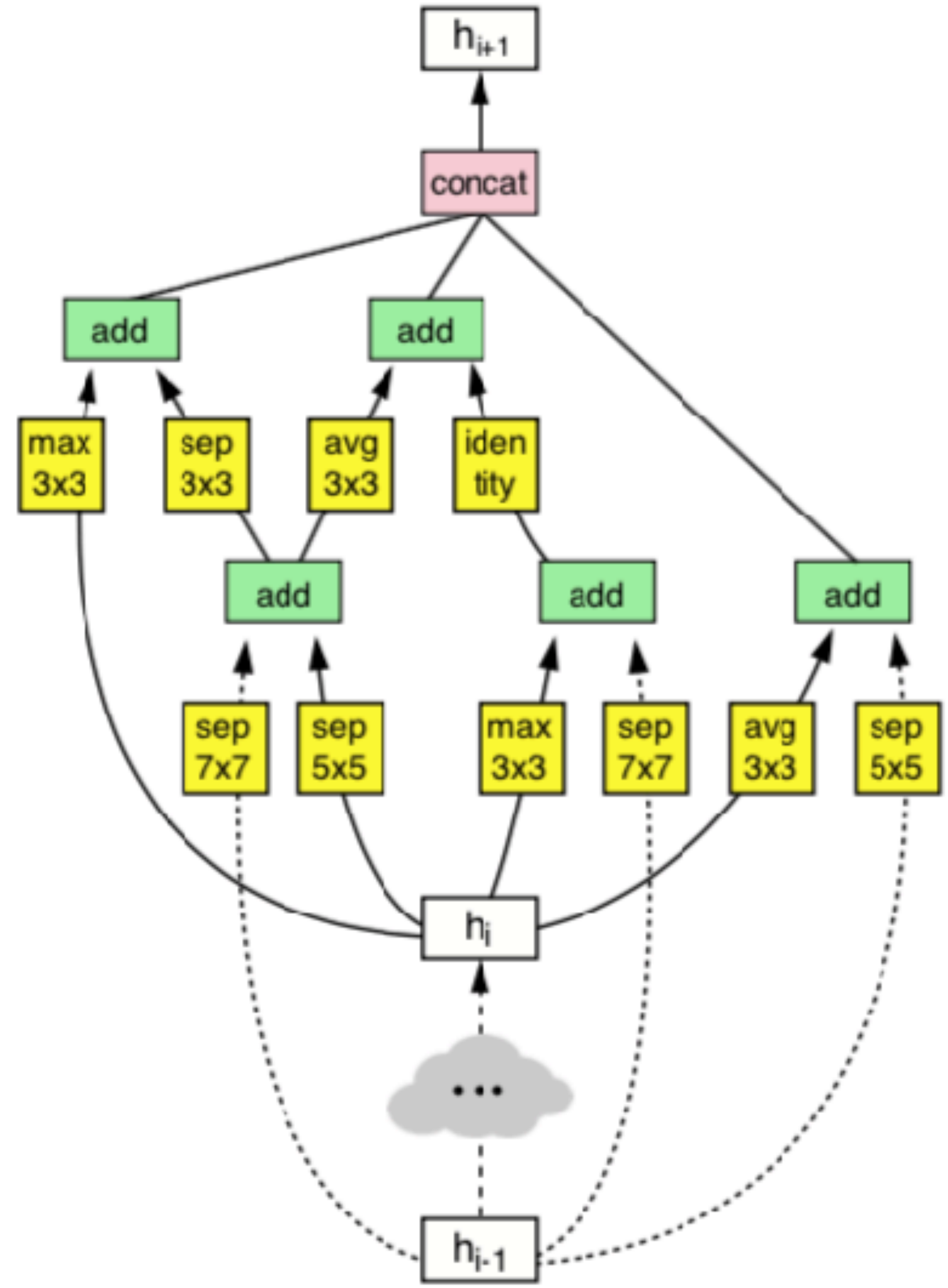
NASNet (Transfer)



NASNet



Normal Cell



Reduction Cell

Experiments & Results

Performance

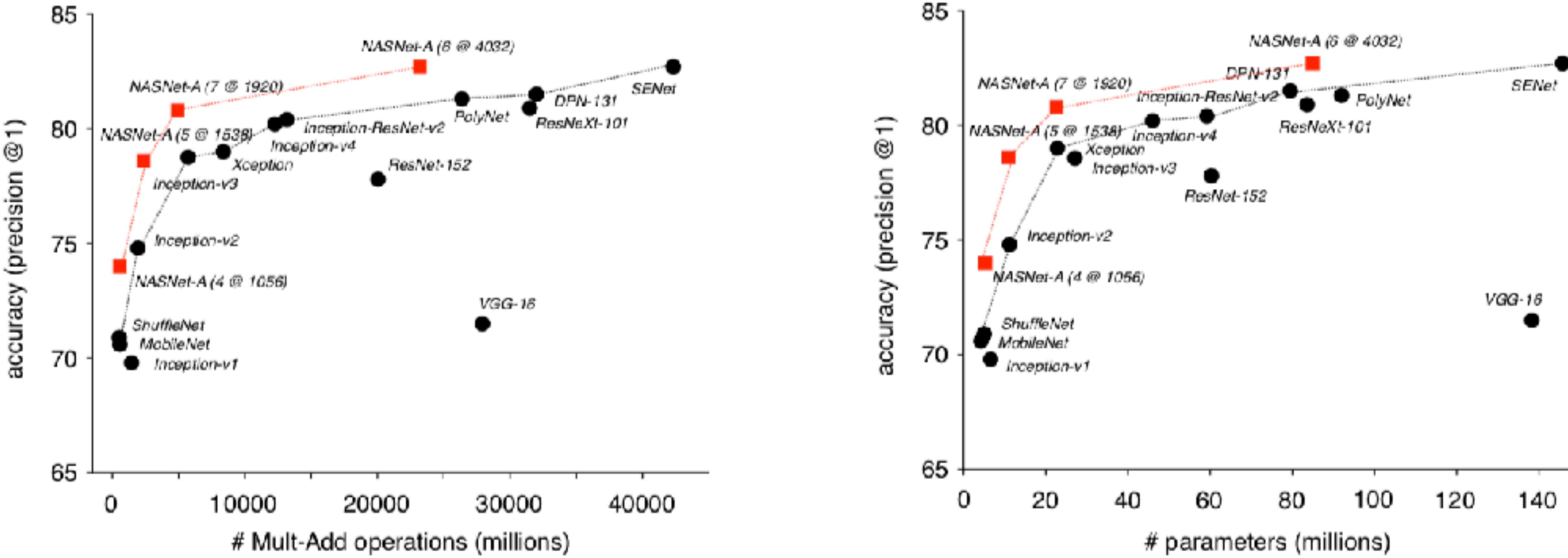


Figure 5. Accuracy versus computational demand (left) and number of parameters (right) across top performing published CNN architectures on ImageNet 2012 ILSVRC challenge prediction task. Computational demand is measured in the number of floating-point multiply-add operations to process a single image. Black circles indicate previously published results and red squares highlight our proposed models.

Performance

model	depth	# params	error rate (%)
DenseNet ($L = 40, k = 12$) [26]	40	1.0M	5.24
DenseNet($L = 100, k = 12$) [26]	100	7.0M	4.10
DenseNet ($L = 100, k = 24$) [26]	100	27.2M	3.74
DenseNet-BC ($L = 100, k = 40$) [26]	190	25.6M	3.46
Shake-Shake 26 2x32d [18]	26	2.9M	3.55
Shake-Shake 26 2x96d [18]	26	26.2M	2.86
Shake-Shake 26 2x96d + cutout [12]	26	26.2M	2.56
NAS v3 [71]	39	7.1M	4.47
NAS v3 [71]	39	37.4M	3.65
NASNet-A (6 @ 768)	-	3.3M	3.41
NASNet-A (6 @ 768) + cutout	-	3.3M	2.65
NASNet-A (7 @ 2304)	-	27.6M	2.97
NASNet-A (7 @ 2304) + cutout	-	27.6M	2.40
NASNet-B (4 @ 1152)	-	2.6M	3.73
NASNet-C (4 @ 640)	-	3.1M	3.59

Table 1. Performance of Neural Architecture Search and other state-of-the-art models on CIFAR-10. All results for NASNet are the mean accuracy across 5 runs.

Performance

Model	image size	# parameters	Mult-Adds	Top 1 Acc. (%)	Top 5 Acc. (%)
Inception V2 [29]	224×224	11.2 M	1.94 B	74.8	92.2
NASNet-A (5 @ 1538)	**299×299**	**10.9 M**	**2.35 B**	**78.6**	**94.2**
Inception V3 [60]	299×299	23.8 M	5.72 B	78.8	94.4
Xception [9]	299×299	22.8 M	8.38 B	79.0	94.5
Inception ResNet V2 [58]	299×299	55.8 M	13.2 B	80.1	95.1
NASNet-A (7 @ 1920)	**299×299**	**22.6 M**	**4.93 B**	**80.8**	**95.3**
ResNeXt-101 (64 x 4d) [68]	320×320	83.6 M	31.5 B	80.9	95.6
PolyNet [69]	331×331	92 M	34.7 B	81.3	95.8
DPN-131 [8]	320×320	79.5 M	32.0 B	81.5	95.8
SENet [25]	**320×320**	**145.8 M**	**42.3 B**	**82.7**	**96.2**
NASNet-A (6 @ 4032)	**331×331**	**88.9 M**	**23.8 B**	**82.7**	**96.2**

Table 2. Performance of architecture search and other published state-of-the-art models on ImageNet classification. Mult-Adds indicate the number of composite multiply-accumulate operations for a single image. Note that the composite multiple-accumulate operations are calculated for the image size reported in the table. Model size for [25] calculated from open-source implementation.

Performance (constrained)

Model	# parameters	Mult-Adds	Top 1 Acc. (%)	Top 5 Acc. (%)
Inception V1 [59]	6.6M	1,448 M	69.8 †	89.9
MobileNet-224 [24]	4.2 M	569 M	70.6	89.5
ShuffleNet (2x) [70]	$\sim$ 5M	524 M	70.9	89.8
NASNet-A (4 @ 1056)	**5.3 M**	**564 M**	**74.0**	**91.6**
NASNet-B (4 @ 1536)	5.3M	488 M	72.8	91.3
NASNet-C (3 @ 960)	4.9M	558 M	72.5	91.0

Table 3. Performance on ImageNet classification on a subset of models operating in a constrained computational setting, i.e., < 1.5 B multiply-accumulate operations per image. All models use 224x224 images. † indicates top-1 accuracy not reported in [59] but from open-source implementation.

Performance

Model	resolution	mAP (mini-val)	mAP (test-dev)
MobileNet-224 [24]	600×600	19.8%	-
ShuffleNet (2x) [70]	600×600	24.5%†	-
NASNet-A (4 @ 1056)	600×600	**29.6%**	-
ResNet-101-FPN [36]	800 (short side)	-	36.2%
Inception-ResNet-v2 (G-RMI) [28]	600×600	35.7%	35.6%
Inception-ResNet-v2 (TDM) [52]	600×1000	37.3%	36.8%
NASNet-A (6 @ 4032)	800×800	41.3%	40.7%
NASNet-A (6 @ 4032)	1200×1200	**43.2%**	**43.1%**
ResNet-101-FPN (RetinaNet) [37]	800 (short side)	-	39.1%

Table 4. Object detection performance on COCO on *mini-val* and *test-dev* datasets across a variety of image featurizations. All results are with the Faster-RCNN object detection framework [47] from a single crop of an image. Top rows highlight mobile-optimized image featurizations, while bottom rows indicate computationally heavy image featurizations geared towards achieving best results. All *mini-val* results employ the same 8K subset of validation images in [28].

Thank You.