

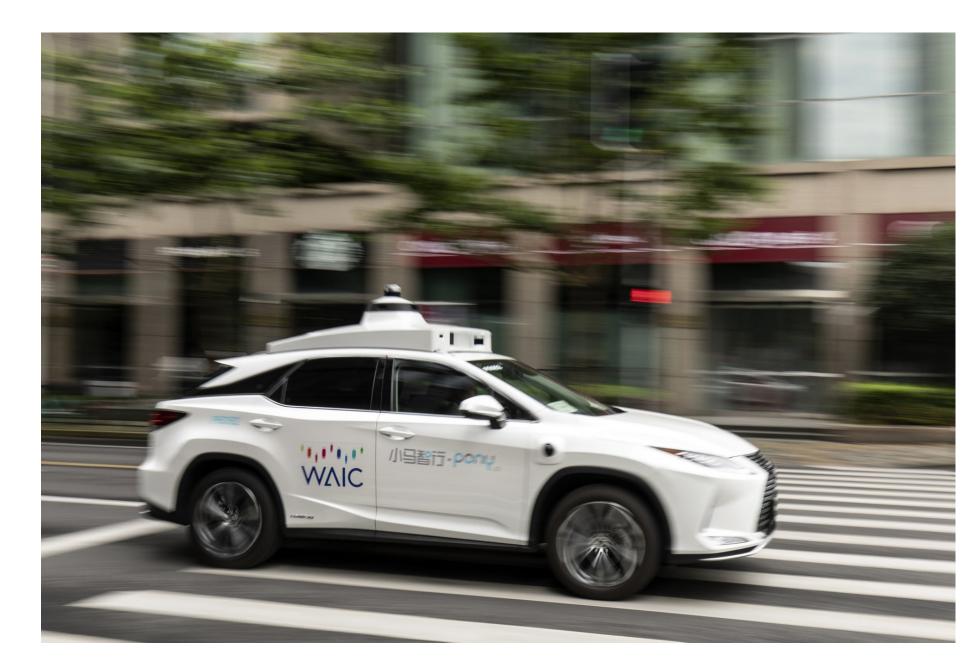
Non-Deep Networks

Ankit Goyal^{1,2} Alexey Bochkovskiy¹ Jia Deng² Vladlen Koltun¹

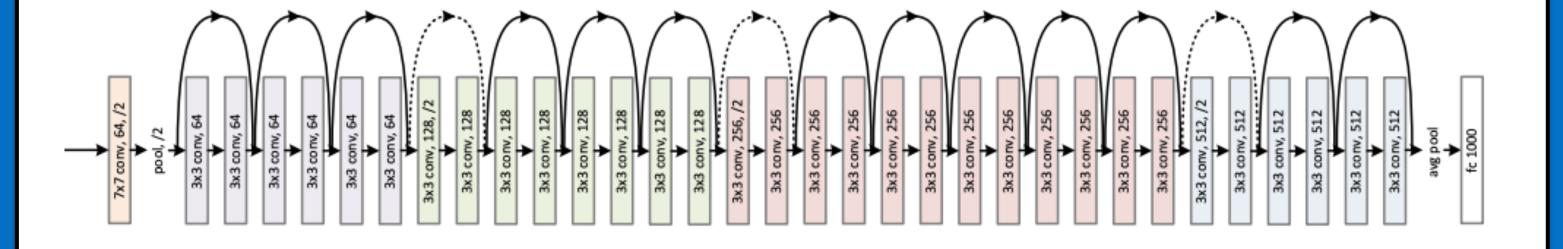
¹Intel ²Princeton University



Why non-deep networks?

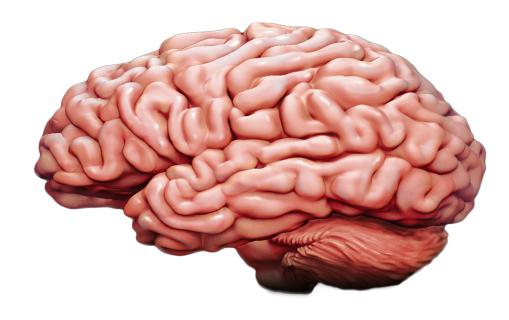


Latency is crucial for time-sensitive applications

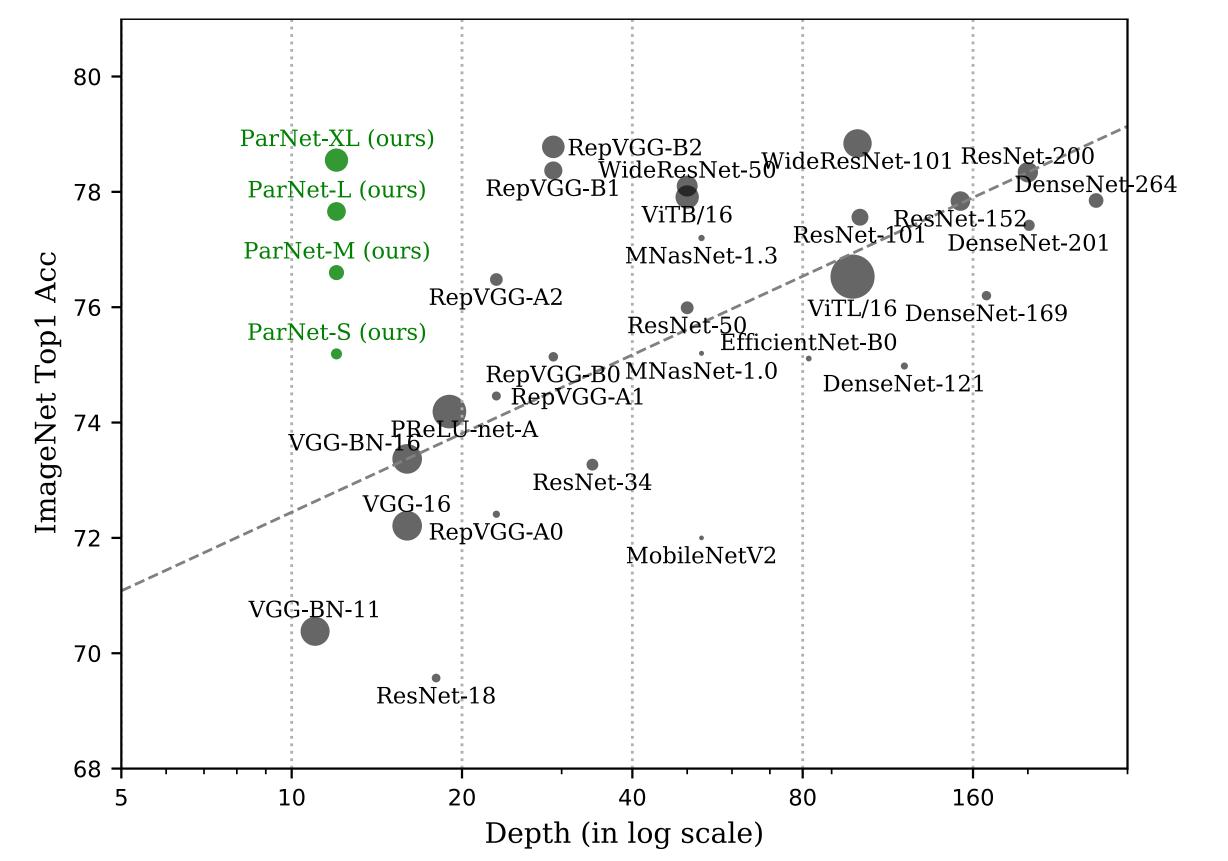


Minimum latency = Depth of $NN \times Clock$ speed of processor Large depth ~ Fundamental barrier to latency

Can non-deep nets perform well?

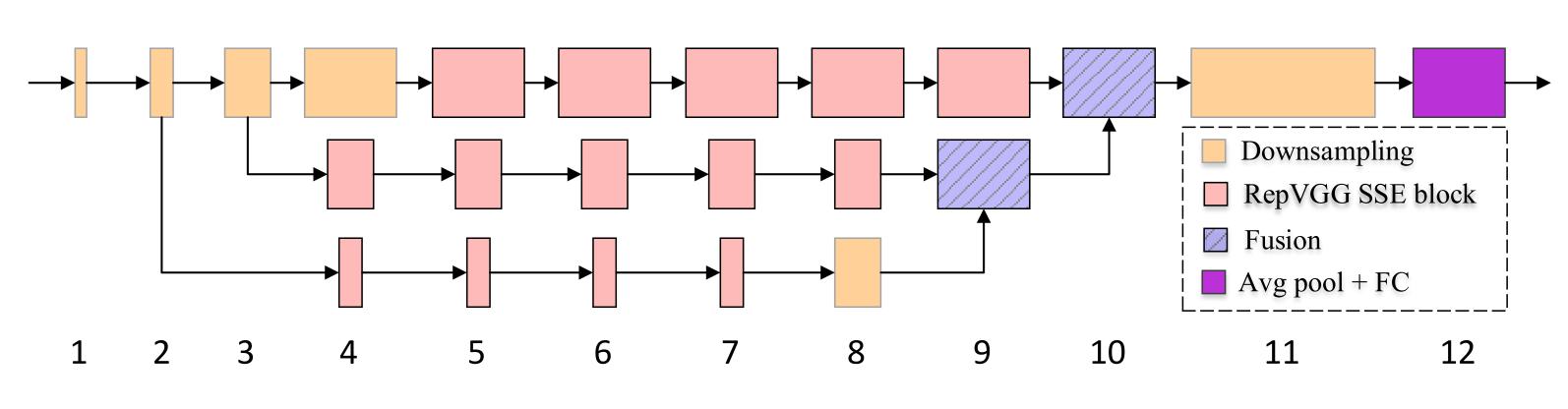


Human brain is not "very deep"



ParNet (our model) with a depth of 12 performs well

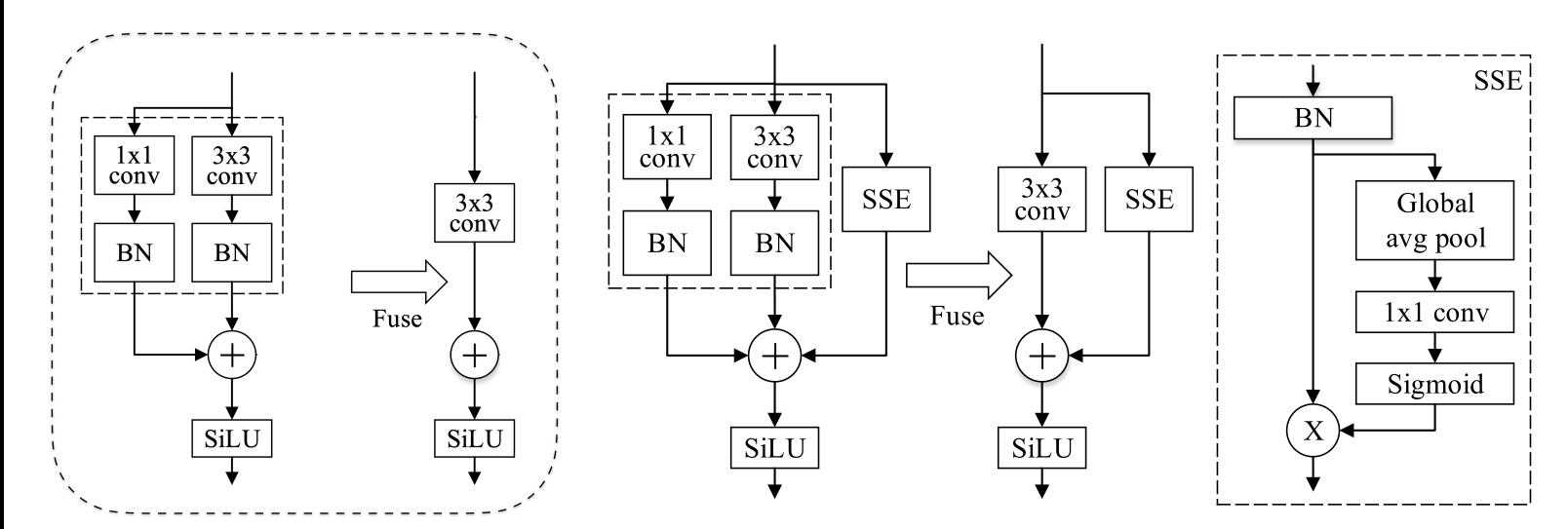
Non-deep Networks



ParNet Structure

Key ingredients:

- Parallel substructures instead of serial ones
- VGG style blocks instead of ResNet style
- Global context with Skip-Squeeze-and-Excitation



RepVGG [1] Re-parametrization

ParNet Block

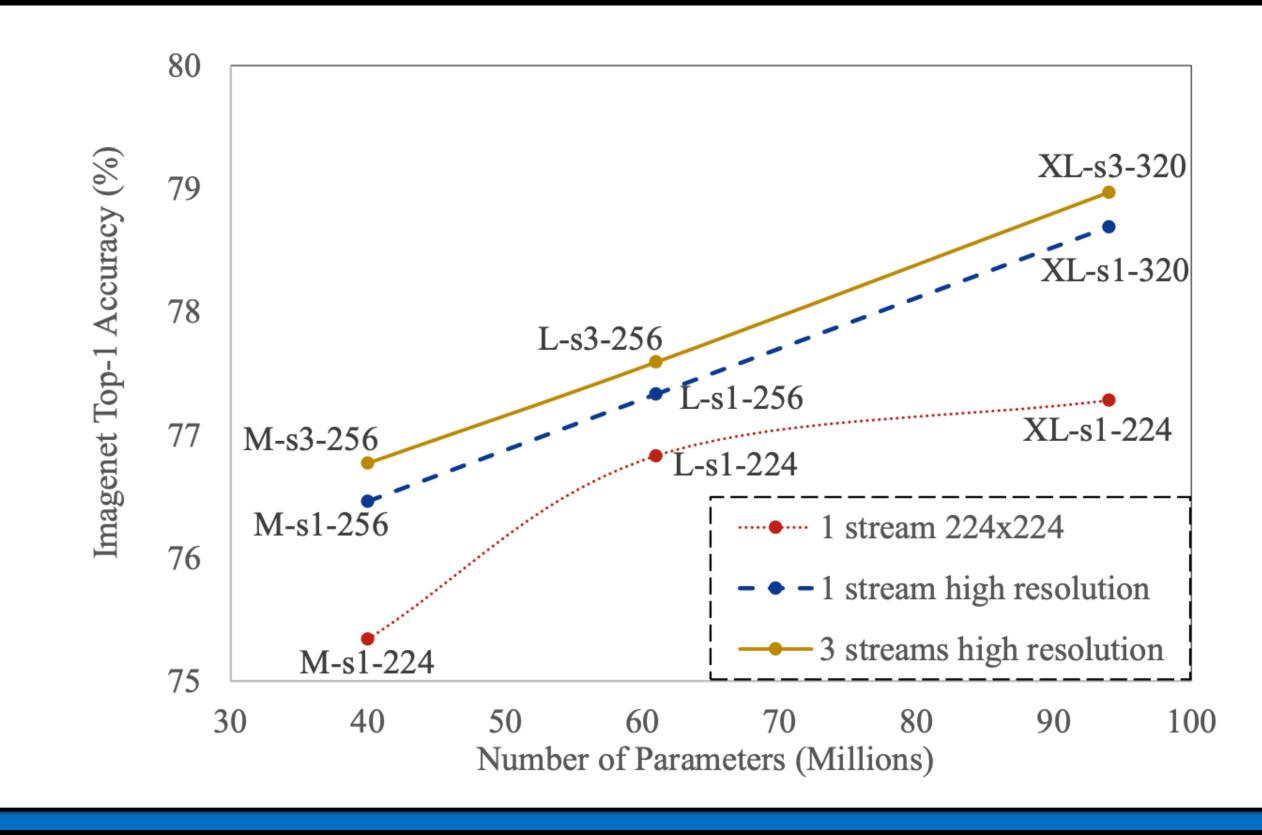
Results on CIFAR

Model	Depth	Params	CIFAR10	CIFAR100
ResNet (Stochastic Depth)	110	1.7	5.23	24.58
DenseNet	40	1.0	5.24	<u>24.42</u>
DenseNet (Bottleneck)	100	0.8	4.51	22.27
ParNet (ours)	12	1.3	<u>5.0</u>	24.62
ResNet (Stochastic depth)	1202	10.2	4.91	_
WideResNet	40	8.9	4.53	21.18
FractalNet	41	22.9	5.21	21.49
DenseNet	100	7	4.10	20.20
DenseNet (Bottleneck)	250	15.3	3.62	17.60
ParNet (ours)	12	15.5	<u>3.90</u>	20.02
WideResNet	28	36.5	3.89	18.85
FractalNet	21	36.8	5.11	22.85
DenseNet	100	27.2	<u>3.74</u>	19.25
DenseNet (Bottleneck)	190	25.6	3.46	17.18
ParNet (ours)	12	35	3.88	<u>18.65</u>

Results on ImageNet

Model	Depth	Top-1 Acc.	Speed (samples/s)	# Params (in M)	Flops (in B)
ResNet34	34	74.12	306	21.8	7.3
ResNet50	50	77.53	222	25.6	8.2
RepVGG-b1g4	29	77.59	376	36.1	14.6
RepVGG-b2g4	29	79.38	300	55.8	22.7
ParNet-S	12	75.19	280	19.2	9.7
ParNet-M	12	76.60	265	35.6	17.2
ParNet-L	12	77.66	249	54.9	26.7
ParNet-XL	12	78.55	230	85.0	41.5

Scaling ParNet



Conclusion

- Non-deep networks can perform surprisingly well
- We explore architectural changes required to build non-deep networks
- We show non-deep networks can be scaled without increasing depth SCAN ME

References:

- [1] Repvgg: Making vgg-style convnets great again, Ding et al.
- [2] Squeeze-and-Excitation Networks, Hu et al.

