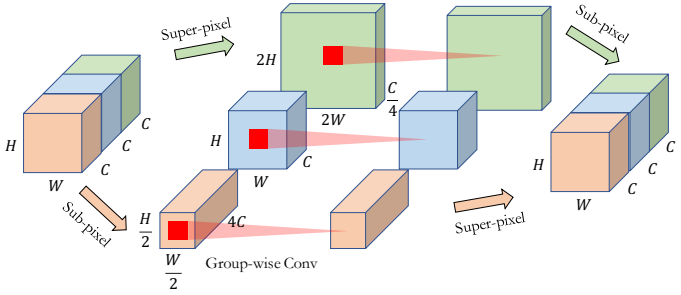
MUXConv: Information Multiplexing in Convolutional Neural Networks

Zhichao Lu, Kalyanmoy Deb and Vishnu Naresh Boddeti
Michigan State University
{luzhicha, kdeb, vishnu}@msu.edu
https://github.com/human-analysis/MUXConv

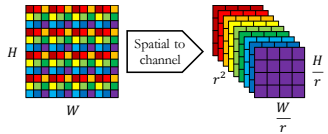
MUXConv

- New layer: Multiplexed Convolutions (Spatial + Channel Multiplexing)
- Idea: increase flow of information between space and channels
- Goal: smaller model size, increased efficiency, maintain/increase performance

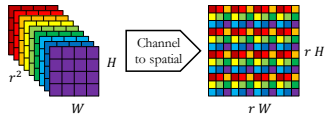
Spatial Multiplexing: Idea



Spatial Multiplexing

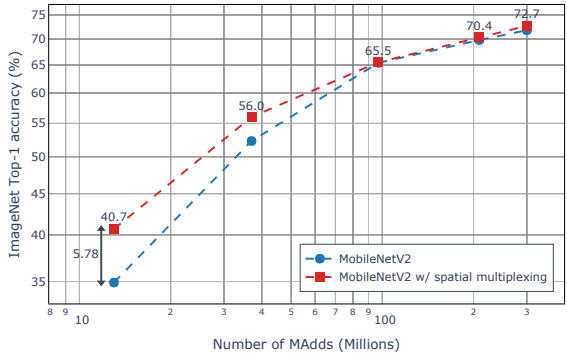


Subpixel



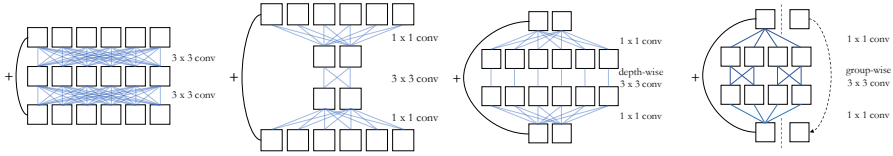
Superpixel

Spatial Multiplexing: Evaluation



- Consistent improvement on accuracy over the original *depth-wise separable convolution*
- Particularly effective in low MAdds Regime

Channel Multiplexing:Idea

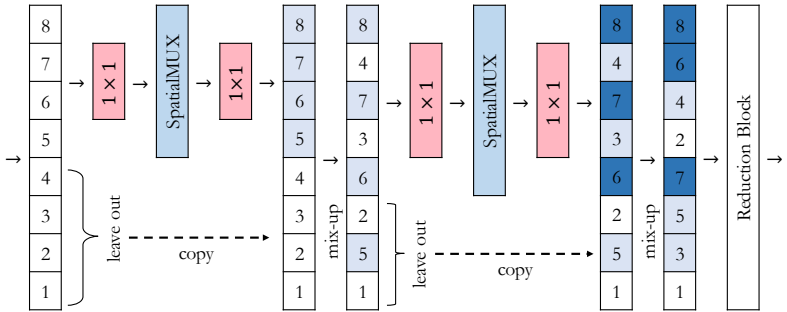


(a) Original residual block (ResNet-18, ResNet-34)

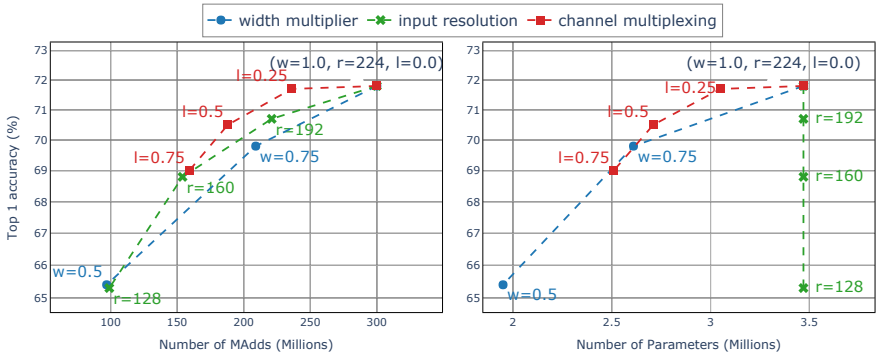
(b) Bottleneck (ResNet-50, DenseNet-BC)

(c) Inverted bottleneck (MobileNetV2/V3, MNASNet)

(d) Proposed

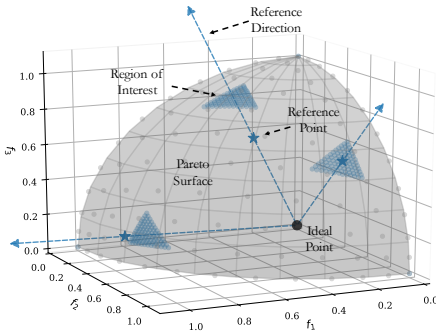


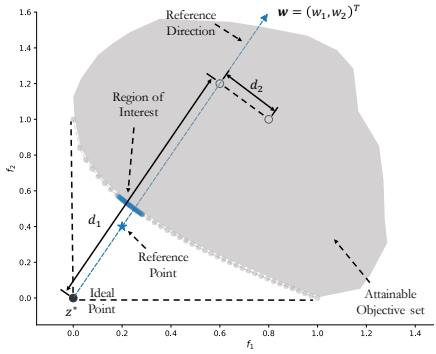
Channel Multiplexing: Evaluation



- Consistently outperforming existing scaling methods

Tri-Objective Search: Idea

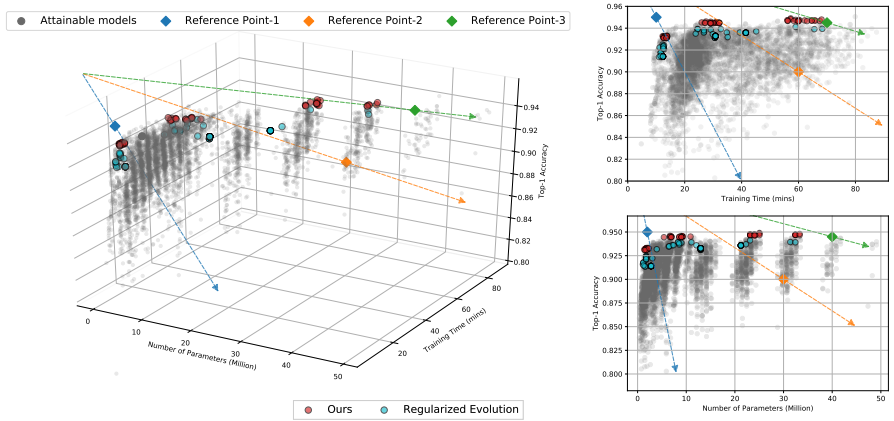




- Simultaneously optimize for accuracy (↑), #Params (↓), and #MAdds (↓)
- User preference guided search through PBI[1] decomposition

[1] Qingfu Zhang and Hui Li. Moea/d: A multiobjective evolutionary algorithm based on decomposition. *IEEE Transactions on Evolutionary Computation*, 11(6):712731, 2007

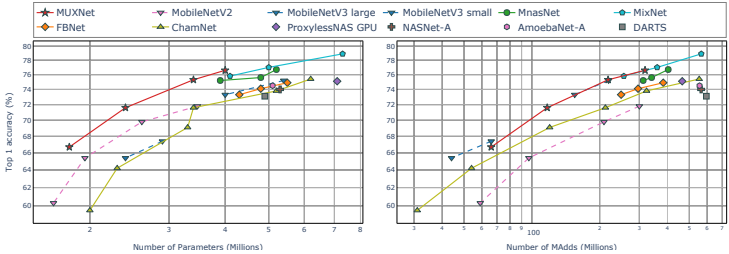
Tri-Objective Search: Evaluation



- NASBench101: our search is more efficient than regularized evolution[2]

[2] Esteban Real, Alok Aggarwal, Yanping Huang, and Quoc VLe. Regularized evolution for image classifier architecture search. In *AAAI*, 2019

ImageNet-1K Classification

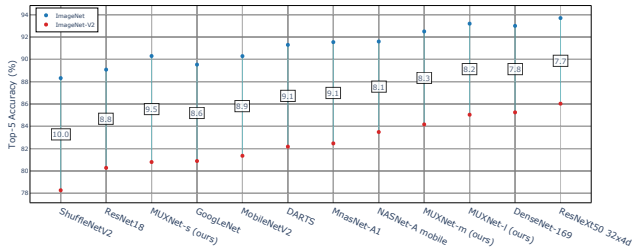


Model	Type	#MAdds	Ratio	#Params	Ratio	CPU(ms)	GPU(ms)	Top-1 (%)	Top-5 (%)
MUXNet-xs (ours)	auto	**66M**‡	1.0x	**1.8M**‡	1.0x	6.8	18	66.7	86.8
MobileNetV2_0.5	manual	97M	1.5x	2.0M	1.1x	**6.2**	17	65.4	86.4
MobileNetV3 small	combined	**66M**	1.0x	2.9M	1.6x	**6.2**‡	**14**	**67.4**	-
MUXNet-s (ours)	auto	**117M**‡	1.0x	**2.4M**‡	1.0x	9.5	25	**71.6**	90.3
MobileNetV1	manual	575M	4.9x	4.2M	1.8x	7.3	20	70.6	89.5
ShuffleNetV2	manual	146M	1.3x	-	-	**6.8**	**11**‡	69.4	-
ChamNet-C	auto	212M	1.8x	3.4M	1.4x	-	-	**71.6**	-
MUXNet-m (ours)	auto	**218M**‡	1.0x	**3.4M**‡	1.0x	14.7	42	**75.3**	92.5
MobileNetV2	manual	300M	1.4x	3.4M	1.0x	**8.3**‡	23	72.0	91.0
ShuffleNetV2 2×	manual	591M	2.7x	7.4M	2.2x	11.0	**22**‡	74.9	-
MnasNet-A1	auto	312M	1.4x	3.9M	1.1x	9.3‡	32	75.2	92.5
MobileNetV3 large	combined	219M	1.0x	5.4M	1.6x	10.0‡	33	75.2	-
MUXNet-l (ours)	auto	**318M**‡	1.0x	**4.0M**‡	1.0x	19.2	74	76.6	93.2
MnasNet-A2	auto	340M	1.1x	4.8M	1.2x	-	-	75.6	92.7
FBNet-C	auto	375M	1.2x	5.5M	1.4x	**9.1**‡	**31**	74.9	-
EfficientNet-B0	auto	390M‡	1.2x	5.3M	1.3x	14.4	46	76.3	93.2
MixNet-M	auto	360M‡	1.1x	5.0M	1.2x	24.3	79	**77.0**	93.3

‡ indicates the objective that the method explicitly optimizes through NAS.

Additional Experiments

Generalization to ImageNet-V2



PASCAL VOC2007 Detection

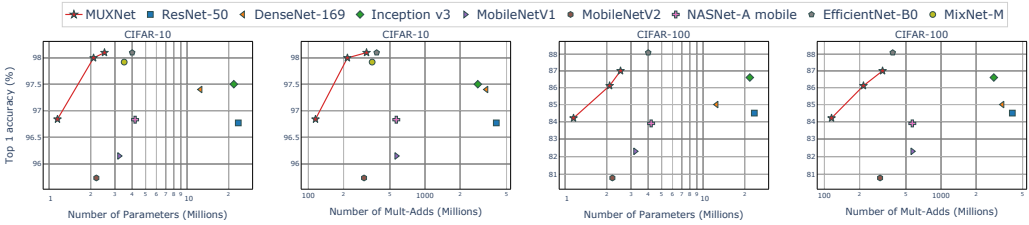
Network	#MAdds	#Params	mAP (%)
VGG16 + SSD	35B	26.3M	74.3
MobileNet + SSD	1.6B	9.5M	67.6
MobileNetV2 + SSDLite	0.7B	3.4M	67.4
MobileNetV2 + SSD	1.4B	8.9M	73.2
MUXNet-m + SSDLite	**0.5B**	**3.2M**	**68.6**
MUXNet-l + SSD	**1.4B**	**9.9M**	**73.8**

ADE20K Semantic Segmentation

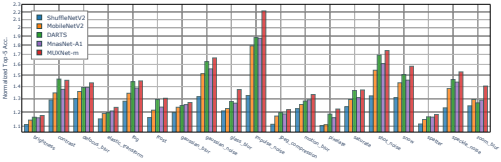
Network	#MAdds	#Params	mIoU (%)	Acc (%)
ResNet18 + C1	1.8B	11.7M	33.82	76.05
MobileNetV2 + C1	0.3B	3.5M	34.84	75.75
MUXNet-m + C1	**0.2B**	**3.4M**	**32.42**	**75.00**
ResNet18 + PPM	1.8B	11.7M	38.00	78.64
MobileNetV2 + PPM	0.3B	3.5M	35.76	77.77
MUXNet-m + PPM	**0.2B**	**3.4M**	**35.80**	**76.33**

Additional Experiments

Transfer Learning on CIFAR



Robustness to Degradations



Visualization on Segmentation Results