A Lightweight Model Design Trend

The most recent lightweight CNNs for mobile platforms consist of multiple inverted residual (MBConv) blocks, following the design convention inspired by MobileNetV2 [1], for their computational efficiency. We summarize the specification of building blocks of modern mobile-efficient models in Table A.

Model	Stem (Out)	Inverted Residual (MBConv) Block Specification (Exp-Out,Stride,Kernel)						
		Stage 1	Stage 2	Stage 3	Stage 4	Stage 5	Stage 6	Stage 7
MobileNetV2 [1]	(32)	(32-16,1,3)	(96-24,2,3) (144-24,1,3)	(144-32,2,3) (192-32,1,3) (192-32,1,3)	(192-64,2,3) (384-64,1,3) (384-64,1,3) (384-64,1,3)	(384-96,1,3) (576-96,1,3) (576-96,1,3)	(576-160,2,3) (960-160,1,3) (960-160,1,3)	(960-320,1,3)
MnasNet-A1 [2]	(32)	(32-16,1,3)	(96-24,2,3) (144-24,1,3)	(72-40,2,5) (120-40,1,5) (120-40,1,5)	(240-80,2,3) (480-80,1,3) (480-80,1,3) (480-80,1,3)	(480-112,1,3) (672-112,1,3)	(672-160,2,5) (960-160,1,5) (960-160,1,5)	(960-320,1,3)
MnasNet-B1 [2]	(32)	(32-16,1,3)	(48-24,2,3) (72-24,1,3) (72-24,1,3)	(72-40,2,5) (120-40,1,5) (120-40,1,5)	(240-80,2,5) (480-80,1,5) (480-80,1,5)	(480-96,1,3) (576-96,1,3)	(576-192,2,5) (1152-192,1,5) (1152-192,1,5) (1152-192,1,5)	(1152-320,1,3)
FBNet-B [3]	(16)	(16-16,1,3)	(96-24,2,3) (24-24,1,5) (24,24,1,3) (24,24,1,3)	(144-32,2,5) (96-32,1,5) (192-32,1,3) (192-32,1,5)	(192-64,2,5) (64-64,1,5) (192-64,1,5)	(384-112,1,5) (112-112,1,3) (112-112,1,5) (336-112,1,5)	(672-184,2,5) (184-184,1,5) (1104-184,1,5) (1104-184,1,5)	(1104-352,1,3)
FBNet-C [3]	(16)	(16-16,1,3)	(96-24,2,3) (24-24,1,5) (24,24,1,3)	(144-32,2,5) (96-32,1,5) (192-32,1,5) (192-32,1,3)	(192-64,2,5) (192-64,1,5) (384-64,1,5) (384-64,1,5)	(384-112,1,5) (672-112,1,5) (672-112,1,5) (336-112,1,5)	(672-184,2,5) (1104-184,1,5) (1104-184,1,5) (1104-184,1,5)	(1104-352,1,3)
Proxyless-R [4]	(32)	(32-16,1,3)	(48-32,2,5) (96-32,1,3)	(96-40,2,7) (120-40,1,3) (120-40,1,5) (120-40,1,5)	(240-80,2,7) (240-80,1,5) (240-80,1,5) (240-80,1,5)	(480-96,1,5) (288-96,1,5) (288-96,1,5) (288-96,1,5)	(576-192,2,7) (1152-192,1,7) (576-192,1,7) (576-192,1,7)	(1152-320,1,7)
Single-Path NAS [5]	(32)	(32-16,1,3)	(48-24,2,3) (72-24,1,3) (72-24,1,3)	(144-40,2,5) (120-40,1,3) (120-40,1,3) (120-40,1,3)	(240-80,2,5) (240-80,1,3) (240-80,1,3) (240-80,1,3)	(480-96,1,5) (288-96,1,5) (288-96,1,5) (288-96,1,5)	(576-192,2,5) (1152-192,1,5) (1152-192,1,5) (1152-192,1,5)	(1152-320,1,3)
MobileNetV3-Large [6]	(32)	(32-16,1,3)	(64-24,2,3) (72-24,1,3)	(72-40,2,5)* (120-40,1,5)* (120-40,1,5)*	(240-80,2,3) (200-80,1,3) (184-80,1,3) (184-80,1,3)	(480-112,1,3)* (672-112,1,3)*	(672-160,2,5)* (960-160,1,5)* (960-160,1,5)*	Conv2D (×-960,1,1)
EfficientNet-B0 [7]	(32)	(32-16,1,3)*	(96-24,2,3) (144-24,1,3)	(144-40,2,5) (240-40,1,5)	(240-80,2,3) (480-80,1,3)* (480-80,1,3)	(480-112,1,5)* (672-112,1,5)* (672-112,1,5)*	(672-192,2,5)* (1152-192,1,5)* (1152-192,1,5)* (1152-192,1,5)*	(1152-320,1,3)*
MixNet-M [8]	(24)	(24-24,1,3)	(144-32,2,3/5/7) (96-32,1,3)	(192-40,2,3/5/7/9)* (240-40,1,3/5)* (240-40,1,3/5)* (240-40,1,3/5)*	(240-80,2,3/5/7)* (240-80,1,3/5/7/9)* (240-80,1,3/5/7/9)* (240-80,1,3/5/7/9)*	(480-120,1,3)* (360-120,1,3/5/7/9)* (360-120,1,3/5/7/9)* (360-120,1,3/5/7/9)*	(720-200,2,3/5/7/9)* (1200-200,1,3/5/7/9)* (1200-200,1,3/5/7/9)*	(1200-200,1,3/5/7/9)*
ReXNet [9]	(32)	(32-16,1,3)	(96-27,2,3) (162-38,1,3)	(228-50,2,3)* (300-61,1,3)*	(366-72,2,3)* (432-84,1,3)* (504-95,1,3)*	(570-106,1,3)* (636-117,1,3)* (702-128,1,3)*	(768-140,2,3)* (840-151,1,3)* (906-162,1,3)* (972-174,1,3)*	(1044-185,1,3)

Table A: All MBConv blocks are grouped as *stages*, and their expanded channel *Exp* and output channel *Out* information is also provided, where the first block of each stage follows the last block of its previous stage. *Stride* and *Kernel* indicate the stride and kernel size of a depth-wise convolution in each block. Note that $*$ mark indicates that squeeze-and-excitation is applied. We omitted other details, such as nonlinearities, to highlight the general structure.

References

[1] Mark Sandler, Andrew Howard, Menglong Zhu, Andrey Zhmoginov, and Liang-Chieh Chen. "Mobilenetv2: Inverted residuals and linear bottlenecks". In: *Proceedings of the IEEE conference on computer vision and pattern recognition*. 2018, pp. 4510–4520.

[2] Mingxing Tan, Bo Chen, Ruoming Pang, Vijay Vasudevan, Mark Sandler, Andrew Howard, and Quoc V Le. "Mnasnet: Platform-aware neural architecture search for mobile". In: *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*. 2019, pp. 2820–2828.

[3] Bichen Wu, Xiaoliang Dai, Peizhao Zhang, Yanghan Wang, Fei Sun, Yiming Wu, Yuandong Tian, Peter Vajda, Yangqing Jia, and Kurt Keutzer. "Fbnet: Hardware-aware efficient convnet design via differentiable neural architecture search". In: *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*. 2019, pp. 10734–10742.

[4] Han Cai, Ligeng Zhu, and Song Han. "ProxylessNAS: Direct Neural Architecture Search on Target Task and Hardware". In: *International Conference on Learning Representations (ICLR)*. 2019.

[5] Dimitrios Stamoulis, Ruizhou Ding, Di Wang, Dimitrios Lymberopoulos, Bodhi Priyantha, Jie Liu, and Diana Marculescu. "Single-path nas: Designing hardware-efficient convnets in less than 4 hours". In: *Machine Learning and Knowledge Discovery in Databases: European Conference, ECML PKDD* (2019).

[6] Andrew Howard, Mark Sandler, Grace Chu, Liang-Chieh Chen, Bo Chen, Mingxing Tan, Weijun Wang, Yukun Zhu, Ruoming Pang, Vijay Vasudevan, Quoc V. Le, and Hartwig Adam. "Searching for mobilenetv3". In: *Proceedings of the IEEE/CVF International Conference on Computer Vision*. 2019, pp. 1314–1324.

[7] Mingxing Tan and Quoc Le. "Efficientnet: Rethinking model scaling for convolutional neural networks". In: *International Conference on Machine Learning*. 2019, pp. 6105–6114.

[8] Mingxing Tan and Quoc V Le. "Mixconv: Mixed depthwise convolutional kernels". In: *In Proceedings of the British Machine Vision Conference* (2019).

[9] Dongyoon Han, Sangdoo Yun, Byeongho Heo, and YoungJoon Yoo. "Rethinking channel dimensions for efficient model design". In: *Proceedings of the IEEE/CVF conference on Computer Vision and Pattern Recognition*. 2021, pp. 732–741.