

# 用于不同硬件的轻量神经网络

GhostNets on Heterogeneous Devices via Cheap Operations. IJCV 2022 (CCF-A)

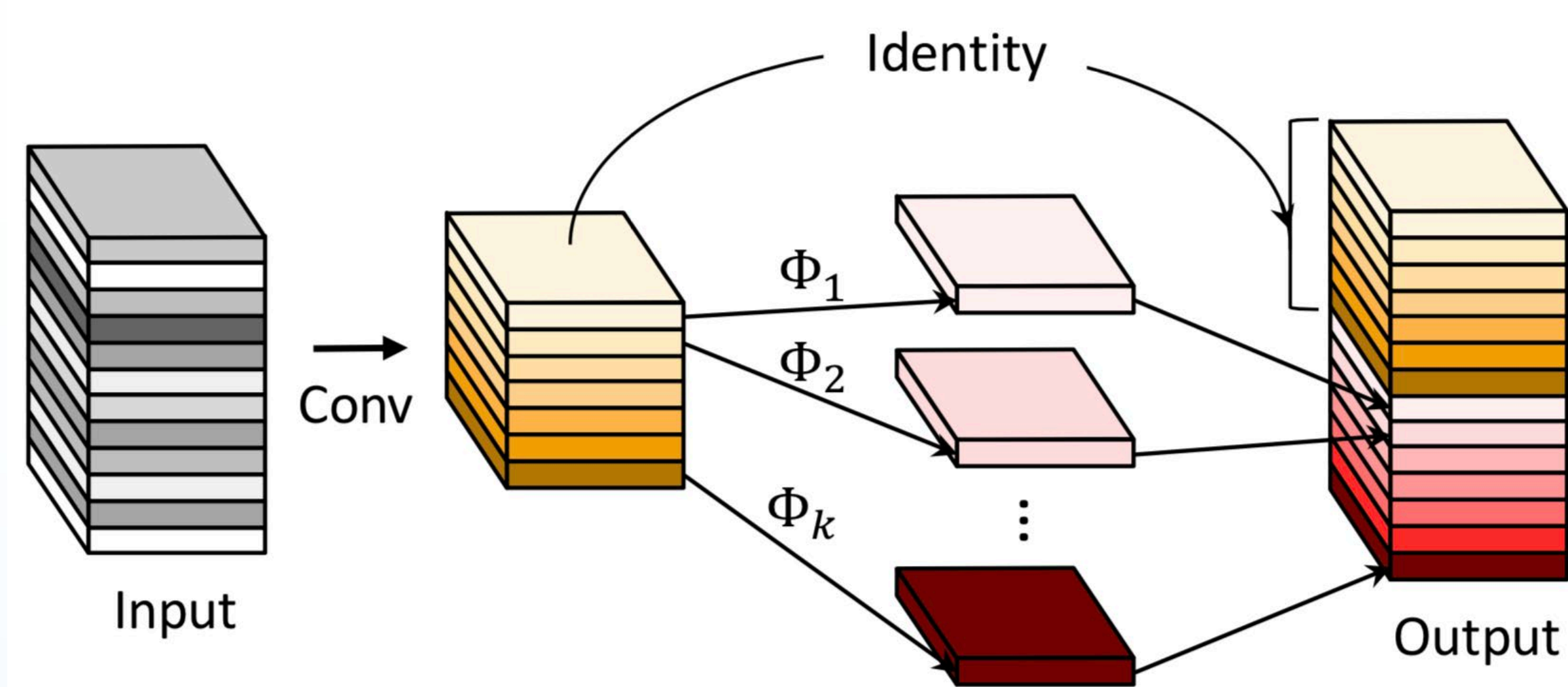
韩凯 吴恩华

## 摘要

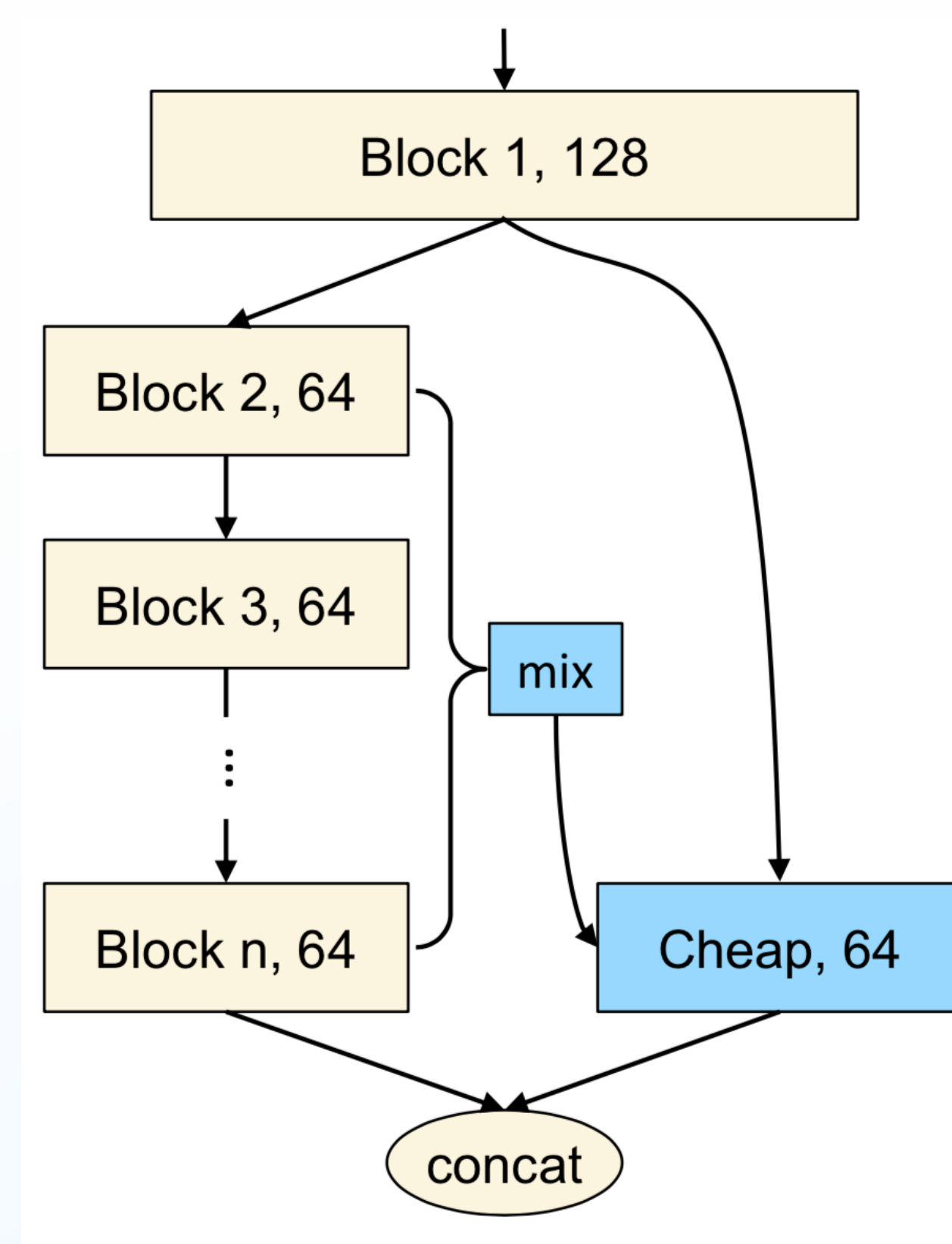
卷积神经网络模型在计算机视觉、图形学等领域应用广泛，但是卷积网络体积大速度慢。本文提出轻量级神经网络GhostNet，利用廉价操作生成更多特征。所提出GhostNet能够用于CPU/GPU等不同硬件平台，大幅降低卷积网络的计算量和时延。

## 方法

本文提出利用廉价操作生成更多特征，来代替传统卷积操作。在ARM/CPU这种流水线型硬件上，我们对每个卷积层做简化：首先用普通卷积生成一部分特征图，然后利用Depthwise卷积等廉价操作生成剩余的特征图，如图1所示。而对于GPU这类并行计算型硬件，我们以Stage结构为单元，其中一部分输出特征图用原始操作生成，另外一部分特征图使用廉价操作生成，如图2所示。



CPU版Ghost结构



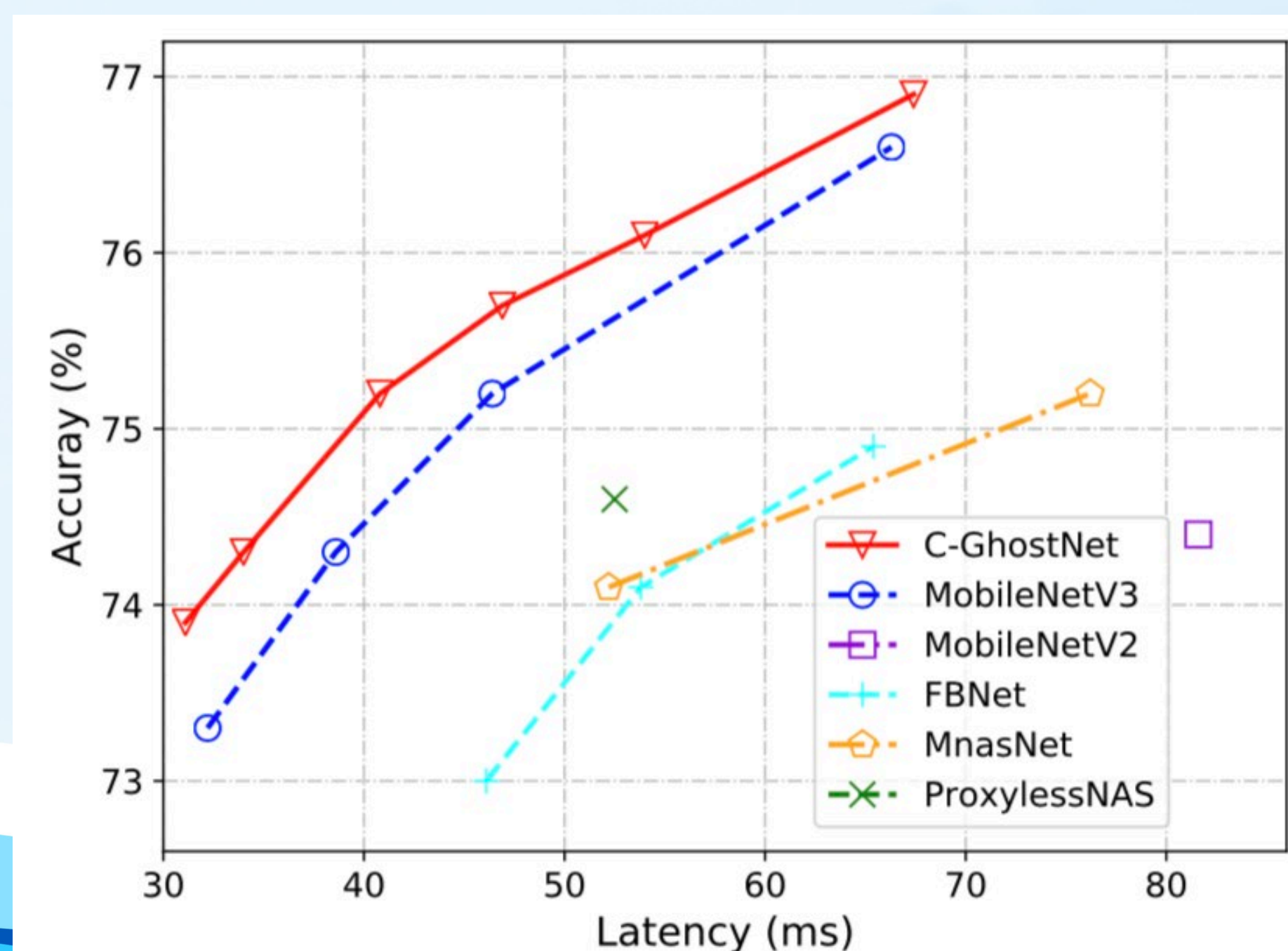
GPU版Ghost结构

## 实验

### ImageNet图像分类任务表现

Model	Params (M)	FLOPs (M)	ARM Latency (ms)	Top-1 (%)	Top-5 (%)
MobileNetV2 0.35x (Sandler et al., 2018)	1.7	59	14.5	60.3	82.9
ShuffleNetV2 0.5x (Ma et al., 2018)	1.4	41	-	61.1	82.6
MobileNetV3 Large 0.35x (Howard et al., 2019)	2.2	40	11.8	64.2	-
CondenseNetV2-A (Yang et al., 2021)	2.0	46	-	64.4	84.2
C-GhostNet 0.5x	2.6	42	10.8	66.2	86.6
MobileNetV1 0.5x (Howard et al., 2017)	1.3	150	19.7	63.3	84.9
ShuffleNetV1 1.0x (g=3) (Zhang et al., 2018)	1.9	138	-	67.8	87.7
MobileNetV2 0.75x (192x192) (Sandler et al., 2018)	2.6	153	28.6	68.7	-
ShuffleNetV2 1.0x (Ma et al., 2018)	2.3	146	-	69.4	88.9
MobileNetV3 Large 0.75x (Howard et al., 2019)	4.0	155	32.2	73.3	-
C-GhostNet 1.0x	5.2	141	31.1	73.9	91.4
MobileNetV2 1.0x (Sandler et al., 2018)	3.5	300	50.5	71.8	91.0
ShuffleNetV2 1.5x (Ma et al., 2018)	3.5	299	-	72.6	90.6
FE-Net 1.0x (Chen et al., 2019c)	3.7	301	-	72.9	-
FBNet-B (Wu et al., 2019)	4.5	295	53.8	74.1	-
MobileNeXt 1.0x (Zhou et al., 2020)	3.4	300	54.9	74.0	-
ProxylessNAS (Cai et al., 2019)	4.1	320	52.5	74.6	92.2
MnasNet-A1 (Tan et al., 2019)	3.9	312	76.2	75.2	92.5
MobileNetV3 Large 1.0x (Howard et al., 2019)	5.4	219	46.4	75.2	-
C-GhostNet 1.3x	7.3	226	46.9	75.7	92.7

### CPU速度



### GPU速度

