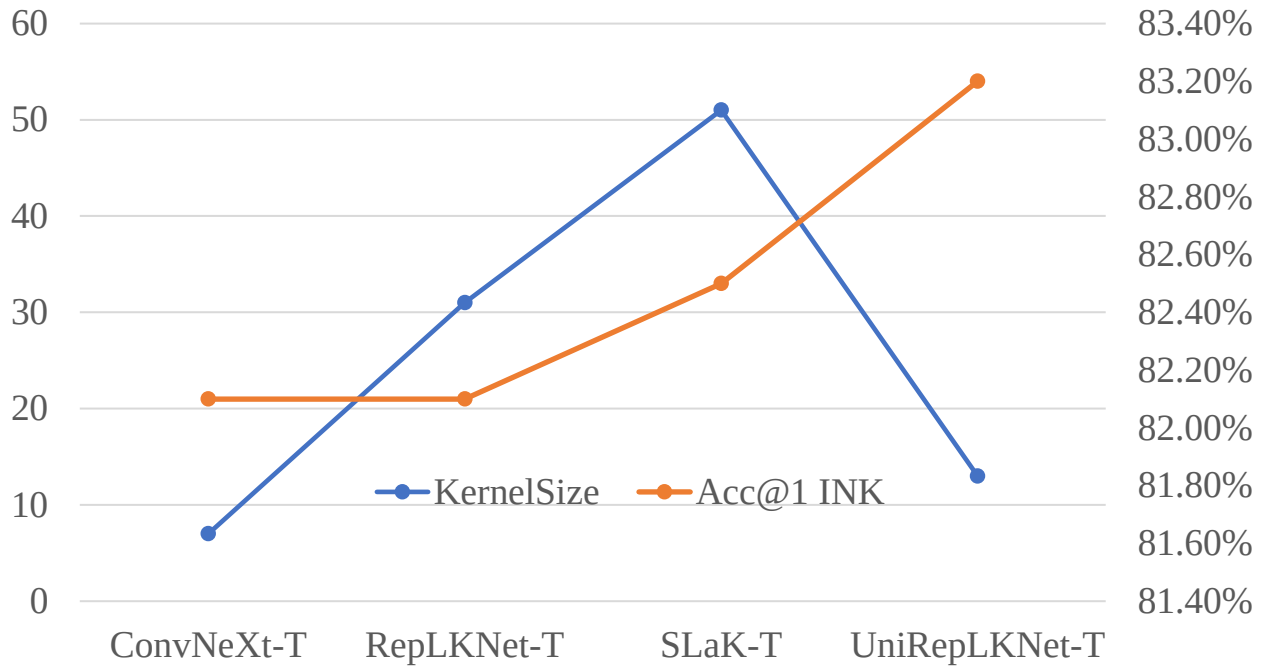
ShiftwiseConv: Small Convolutional Kernel with Large Kernel Effect

Dachong Li, Li Li, Zhuangzhuang Chen, Jianqiang Li

01 Introduction

深圳大学 SHENZHEN UNIVERSITY

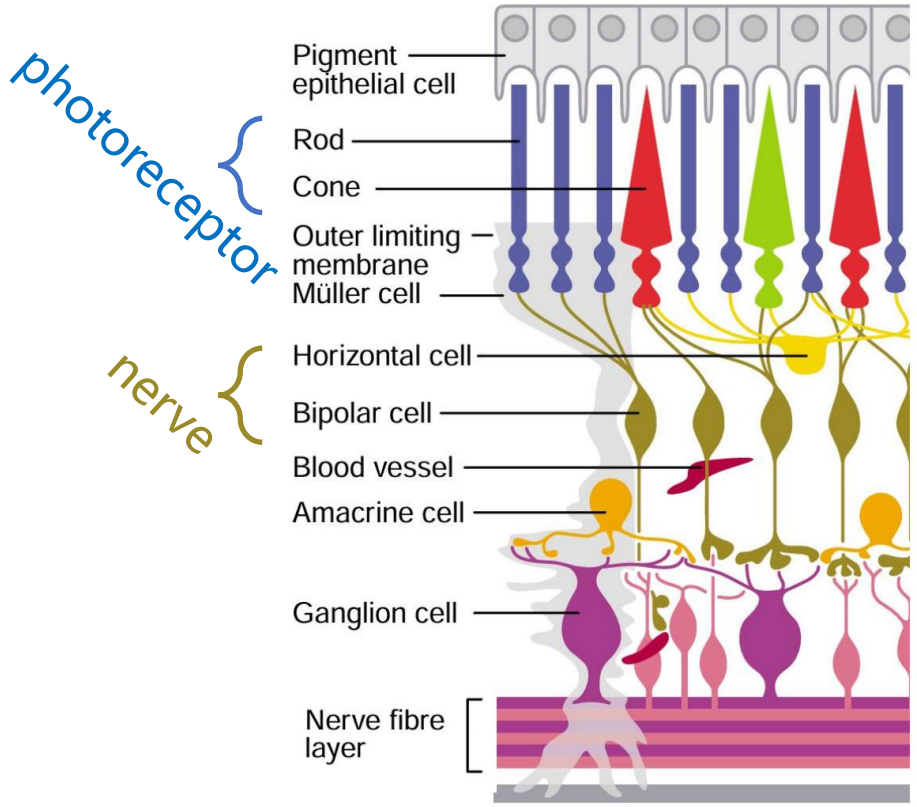
- ✓ Large convolutional kernels are crucial for modern CNN performance.
- ✓ Using kernel size adjustment to boost performance:
  1. Limited adjustment space.
  2. Diminishing returns.

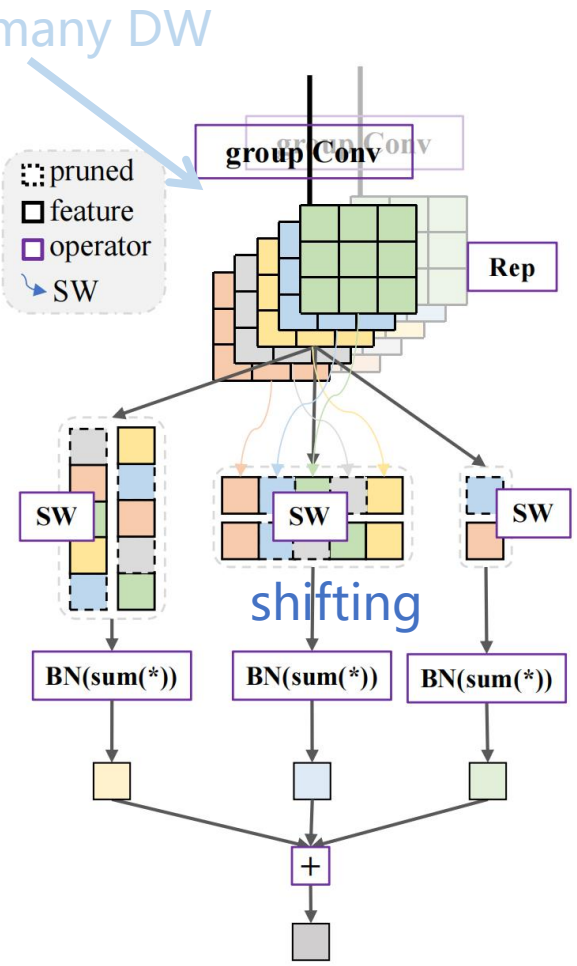


01 Introduction

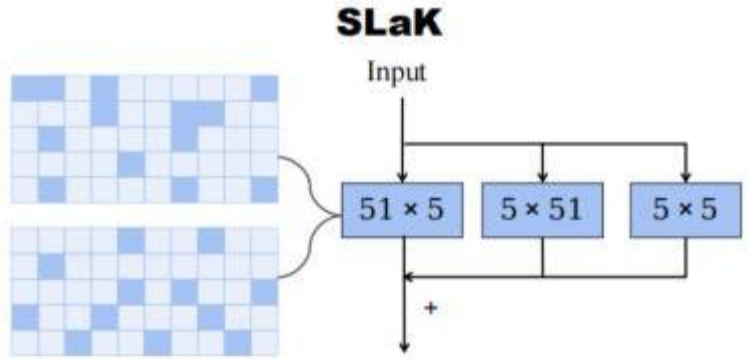
深圳大学 SHENZHEN UNIVERSITY

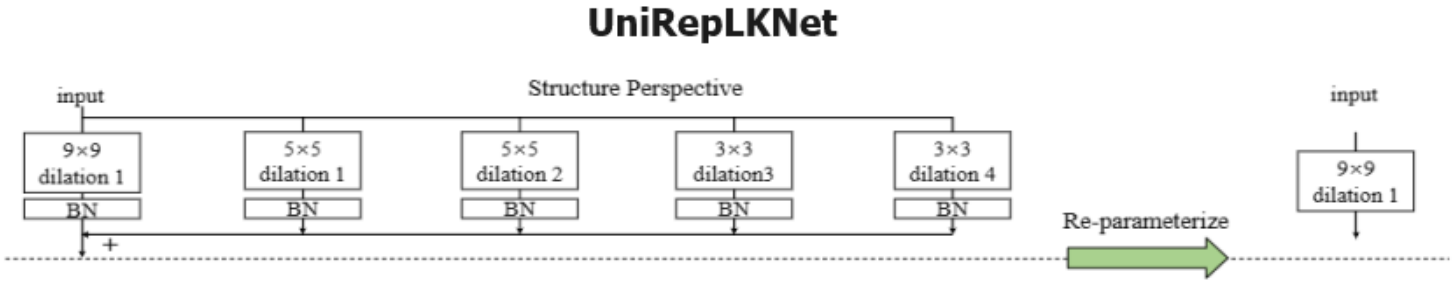
- ✓ Retinal functional decoupling
- ✓ ShiftWise:
- ✓ • One-to-many depth-wise separable convolutions
- ✓ • Feature shifting





02 Result





ADE20K *SS* and *MS* denote single-scale and multi-scale evaluations

Method	Crop size	Params (M)	FLOPs (G)	mIoU (SS)	mIoU (MS)
SW-tiny	512^2	62	948	**49.22**	**50.06**
ConvNeXt-T [38]	512^2	60	939	46.0	46.7
SLaK-T [35]	512^2	65	936	47.6	-
UniRepLKNet-T [17]	512^2	61	946	48.6	49.1
Swin-T [36]	512^2	60	945	44.5	45.8
InternImage-T [54]	512^2	59	944	47.9	48.1
SW-small	512^2	88	1039	49.79	50.83
ConvNeXt-S [38]	512^2	82	1027	48.7	49.6
SLaK-S [35]	512^2	91	1028	49.4	-
UniRepLKNet-S [17]	512^2	86	1036	50.5	51.0
Swin-S [36]	512^2	81	1038	47.6	49.5
InternImage-S [54]	512^2	80	1017	50.1	50.9

COCO. inputs (800*1280)

Method	Params (M)	FLOPs (G)	AP^{box}	AP^{mask}
SW-tiny	87	751	**52.21**	**45.19**
UniRepLKNet-T [17]	89	749	51.8	44.9
Swin-T [36]	86	745	50.4	43.7
ConvNeXt-T [38]	86	741	50.4	43.7
SLaK-T [35]	-	-	51.3	44.3
SW-small	110	839	52.7	45.67
UniRepLKNet-S [17]	113	835	53.0	45.9
Swin-S [36]	107	838	51.9	45.0
ConvNeXt-S [38]	108	827	51.9	45.0
SLaK-T [35]	-	-	51.3	44.3

Monocular 3D object detection on nuScenes.

Method	Lr schd	mAP
FCOS3D-SW-small	1x	31.42
FCOS3D-UniRepLKNet-S	1x	31.38
FCOS3D-ResNet101 w/ DCN	1x	29.8

ImageNet classification. "T/C" denote transformer/ConvNet.

Method	Type	Input size	Params (M)	FLOPs (G)	Acc (%)
SW-tiny	C	224^2	31	5.0	**83.4**
UniRepLKNet-T[17]	C	224^2	31	4.9	83.2
FastViT-SA24 [48]	T	256^2	21	3.8	82.6
PVTv2-B2 [53]	T	224^2	25	4.0	82.0
CoAtNet-0 [13]	T	224^2	25	4.2	81.6
DeiT III-S [47]	T	224^2	22	4.6	81.4
SwinV2-T/8 [37]	T	256^2	28	6	81.8
SLaK-T [35]	C	224^2	30	5.0	82.5
SW-small	C	224^2	56	9.4	**83.9**
UniRepLKNet-S[17]	C	224^2	56	9.1	83.9
ConvNeXt-S [38]	C	224^2	50	8.7	83.1
HorNet-T [44]	C	224^2	23	3.9	83.0
FastViT-SA36 [48]	T	256^2	30	5.6	83.6
CoAtNet-1 [13]	T	224^2	42	8.4	83.3
SLaK-S [35]	C	224^2	55	9.8	83.8
FastViT-MA36 [48]	T	256^2	43	7.9	83.9
SwinV2-S/8 [37]	T	256^2	50	12	83.7
RepLKNet-31B [16]	C	224^2	79	15.3	83.5
PVTv2-B5 [53]	T	224^2	82	11.8	83.8

83.46

自立 自律 自强

03 Operator inference

深圳大学 SHENZHEN UNIVERSITY

Inference throughput comparison. The symbol ⋆indicates the use of efficient large-kernel convolutions provided by RepLKNet. The symbol ® denotes Rep.

Method	Throughput (img/s)	IN-1K acc.
SW-tiny	378	83.4
SLaK-T ®★ [35]	638	82.5
UniRepLKNet-T [17]	1125	83.2
SLaK-T [35]	371	82.5
SLaK-T ® [35]	174	82.5
SW-small	243	83.9
SLaK-S ®★ [35]	385	83.9
UniRepLKNet-S [17]	689	83.8
SLaK-S [35]	217	83.8
SLaK-S ® [35]	108	83.8

Operator inference speed optimization:

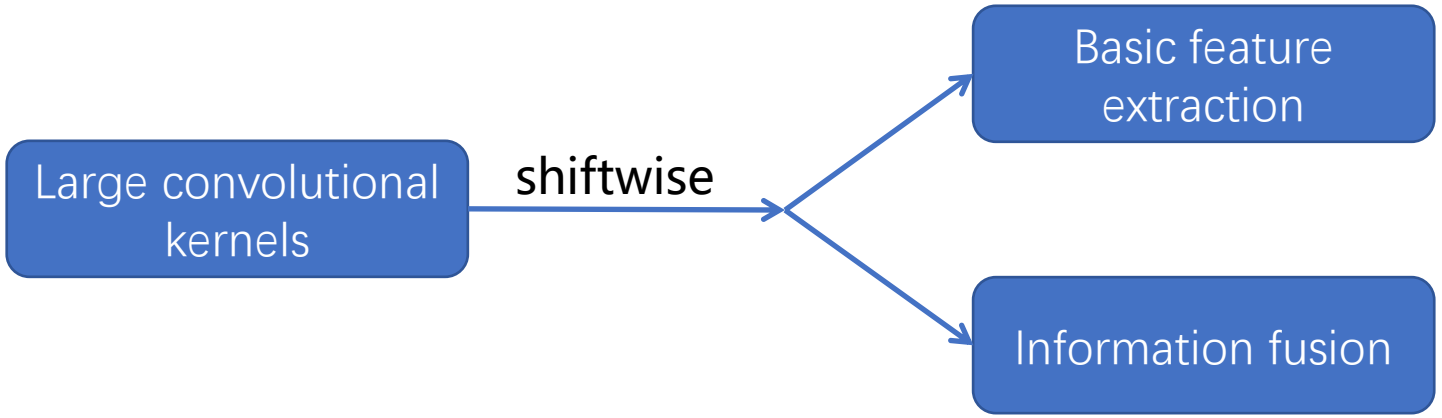
① Multi-path feature reuse

② Separable convolution + feature shifting

③ Coarse-grained sparsity

④ CUDA operator optimization techs

04 Summary


Large convolutional kernels
shiftwise
Basic feature extraction
Information fusion

深圳大学
SHENZHEN UNIVERSITY

Thank you!

自立 自律 自强