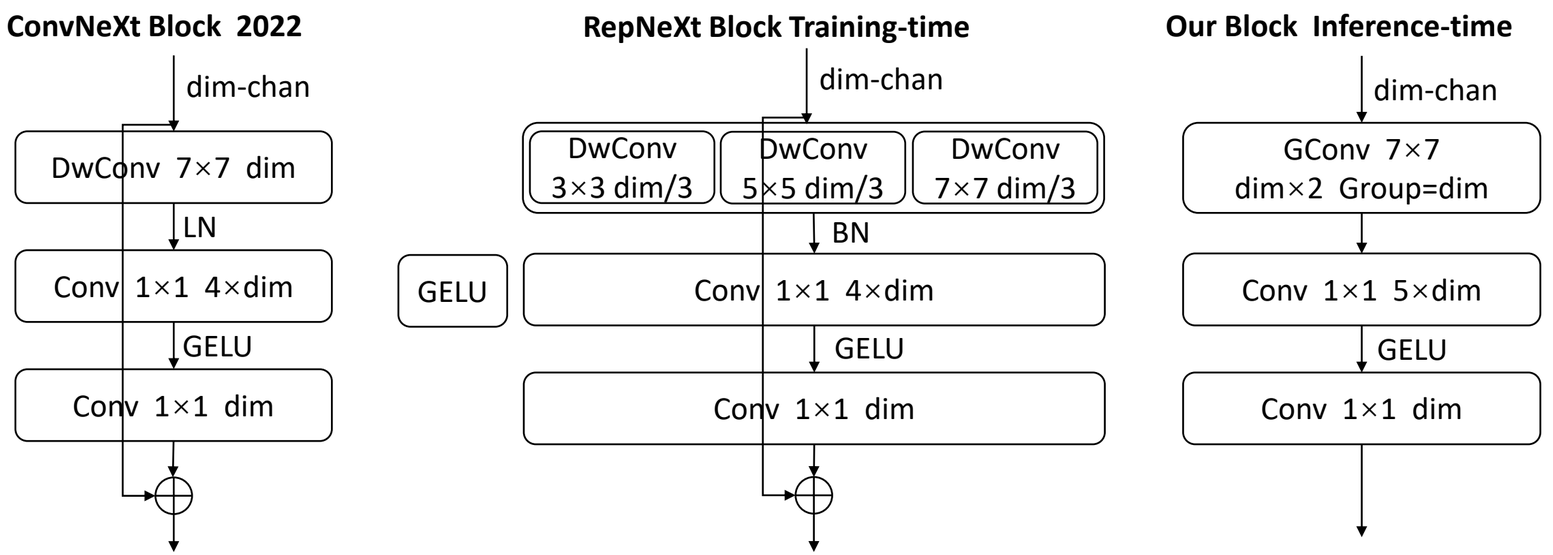
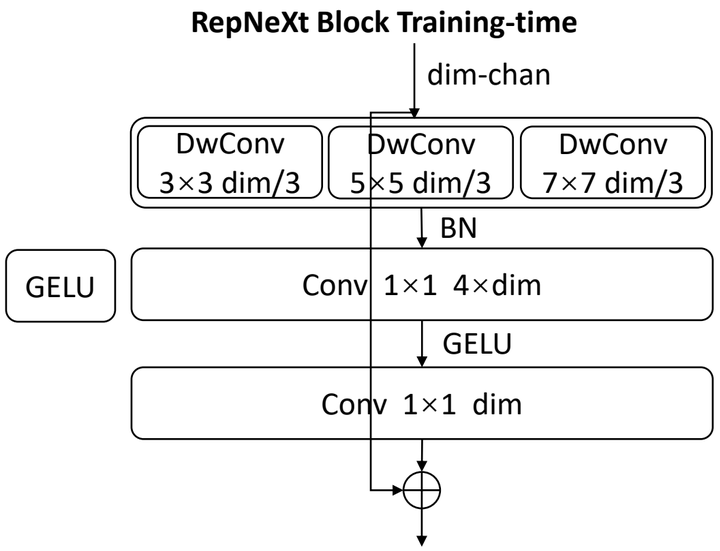
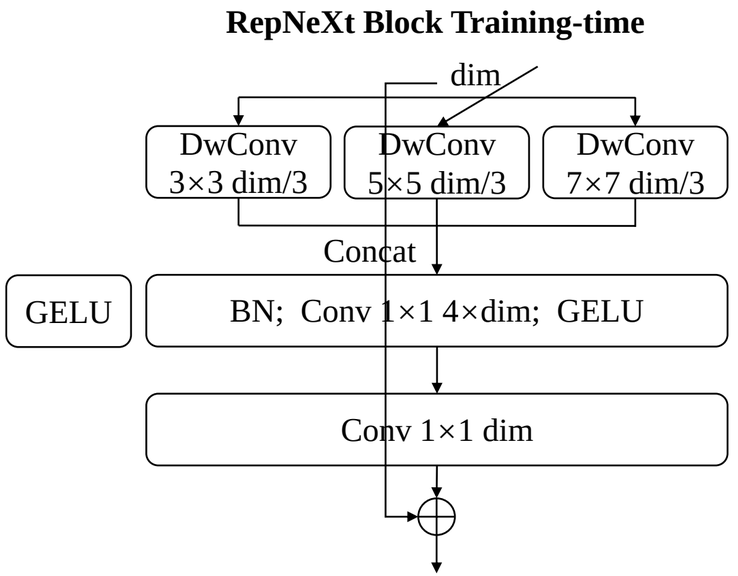

ConvNeXt Block 2022
dim-chan
DwConv 7×7 dim
LN
Conv 1×1 4×dim
GELU
Conv 1×1 dim
RepNeXt Block Training-time
dim-chan
DwConv 3×3 dim/3
DwConv 5×5 dim/3
DwConv 7×7 dim/3
BN
GELU
Conv 1×1 4×dim
GELU
Conv 1×1 dim
Our Block Inference-time
dim-chan
GConv 7×7 dim×2 Group=dim
Conv 1×1 5×dim
GELU
Conv 1×1 dim

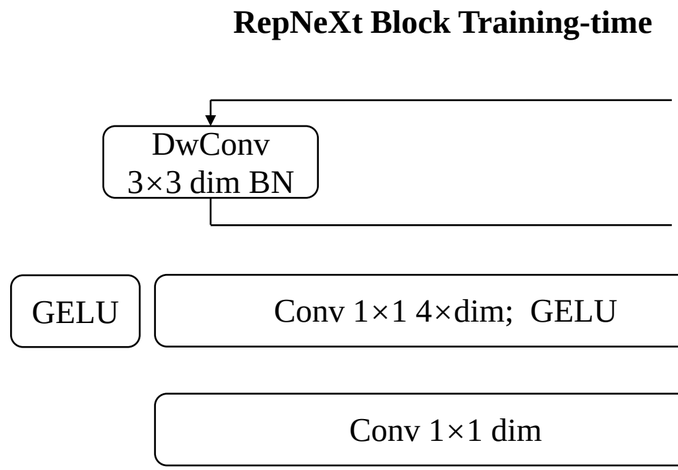
(224, 224)	stage2	stage3	stage4	stage5	FLOPs	# params.
Output	56×56	28×28	14×14	7×7		
Training	$\begin{bmatrix} \mathrm{d}7\times 7, 96 \\ 1\times 1, 384 \\ 1\times 1, 96 \end{bmatrix}\times 3$	$\begin{bmatrix} \mathrm{d}7\times 7, 192 \\ 1\times 1, 768 \\ 1\times 1, 192 \end{bmatrix}\times 3$	$\begin{bmatrix} \mathrm{d}7\times 7, 384 \\ 1\times 1, 1536 \\ 1\times 1, 384 \end{bmatrix}\times 9$	$\begin{bmatrix} \mathrm{d}7\times 7, 768 \\ 1\times 1, 3072 \\ 1\times 1, 768 \end{bmatrix}\times 3$		
Inference	$\begin{bmatrix} g7\times 7, 192 \\ 1\times 1, 480 \\ 1\times 1, 96 \end{bmatrix}\times 3$	$\begin{bmatrix} g7\times 7, 384 \\ 1\times 1, 960 \\ 1\times 1, 192 \end{bmatrix}\times 3$	$\begin{bmatrix} g7\times 7, 768 \\ 1\times 1, 1920 \\ 1\times 1, 384 \end{bmatrix}\times 9$	$\begin{bmatrix} g7\times 7, 1536 \\ 1\times 1, 3840 \\ 1\times 1, 768 \end{bmatrix}\times 3$		

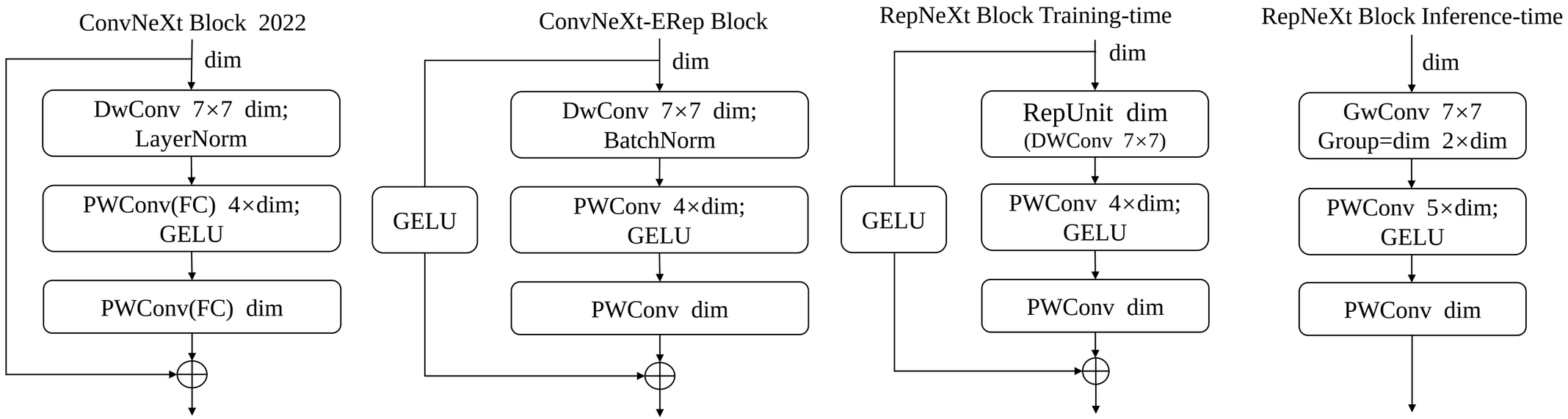
stem
$\begin{bmatrix} 4\times 4, s=4, 96 \\ \mathrm{BN} \\ \mathrm{Relu} \end{bmatrix}$
downsample
$\begin{bmatrix} \mathrm{BN} \\ 2\times 2, s=2, \times 2 \end{bmatrix}$

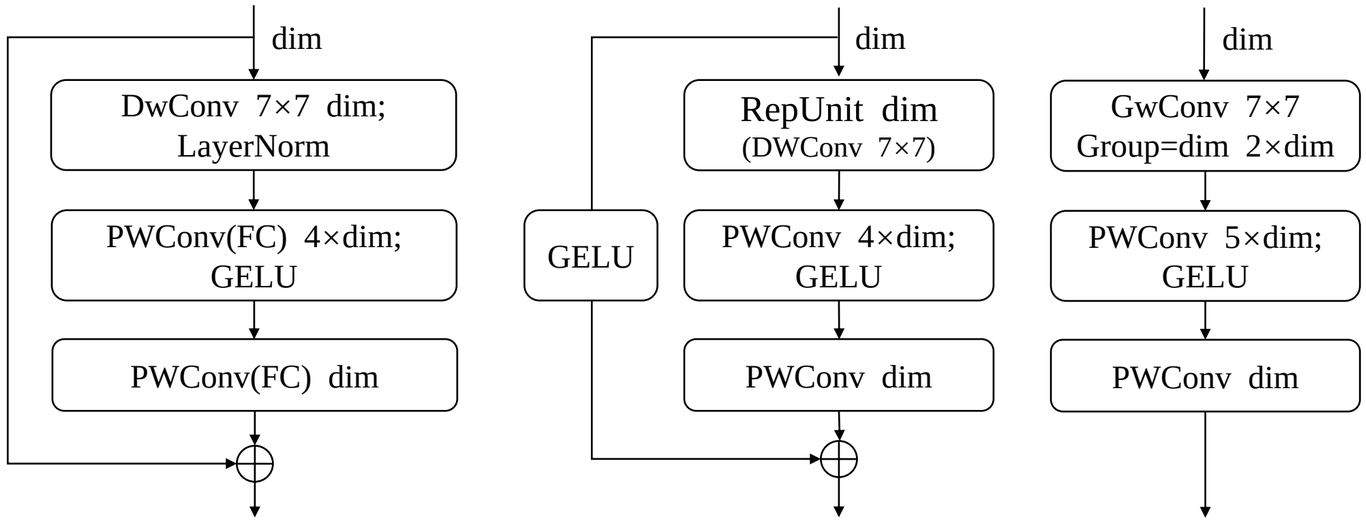
Structure	Output size	RepNeXt training-time	RepNeXt inference-time
Stem	56 × 56	$\begin{bmatrix} 4\times4, s=4, 96 \\ \text{BN} \end{bmatrix}$	$4\times4, s=4, 96$
Stage1	56 × 56	$\begin{bmatrix} RepUnit, dw, 96 \\ 1\times1, 384 \\ 1\times1, 96 \end{bmatrix}\times3$	$\begin{bmatrix} 7\times7, 192, group=96 \\ 1\times1, 480 \\ 1\times1, 96 \end{bmatrix}\times3$
Stage2	28 × 28	$\begin{bmatrix} RepUnit, dw, 192 \\ 1\times1, 768 \\ 1\times1, 192 \end{bmatrix}\times3$	$\begin{bmatrix} 7\times7, 384, group=192 \\ 1\times1, 960 \\ 1\times1, 192 \end{bmatrix}\times3$
Stage3	14 × 14	$\begin{bmatrix} RepUnit, dw, 384 \\ 1\times1, 1536 \\ 1\times1, 384 \end{bmatrix}\times9$	$\begin{bmatrix} 7\times7, 768, group=384 \\ 1\times1, 1920 \\ 1\times1, 384 \end{bmatrix}\times9$
Stage4	7 × 7	$\begin{bmatrix} RepUnit, dw, 768 \\ 1\times1, 3072 \\ 1\times1, 768 \end{bmatrix}\times3$	$\begin{bmatrix} 7\times7, 1536, group=768 \\ 1\times1, 3840 \\ 1\times1, 768 \end{bmatrix}\times3$
FLOPs		800M~850M	1.5B
# params.		27M~29M	51.6M
throughput		78.1	126.8

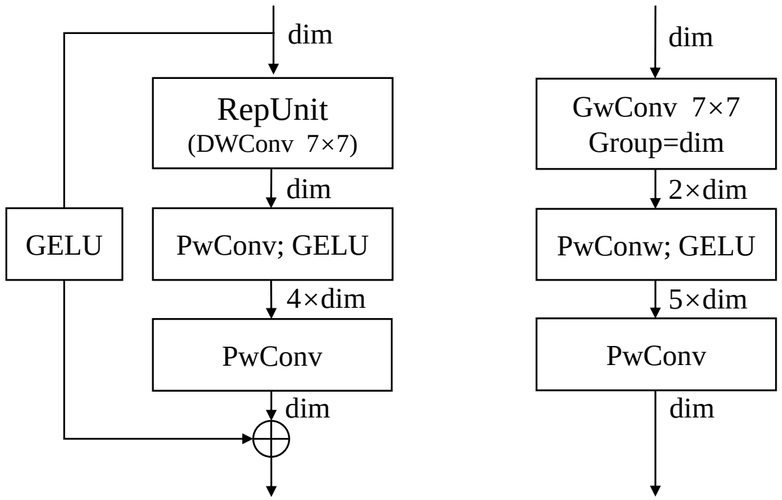

RepNeXt Block Training-time
dim-chan
DwConv 3×3 dim/3
DwConv 5×5 dim/3
DwConv 7×7 dim/3
BN
GELU
Conv 1×1 4×dim
GELU
Conv 1×1 dim

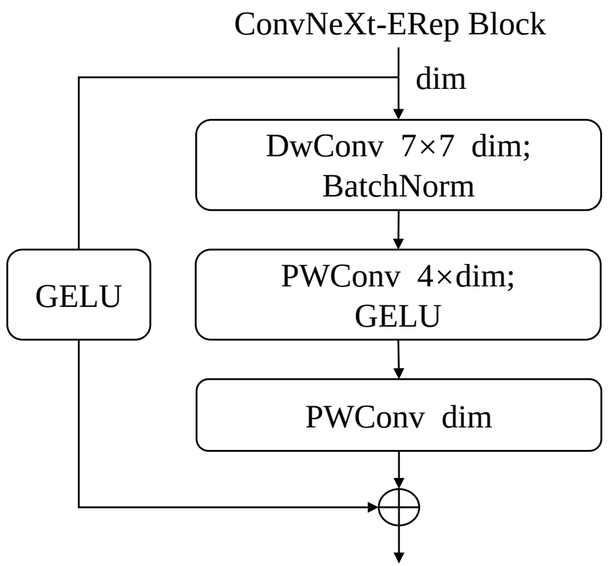

RepNeXt Block Training-time
dim
DwConv 3×3 dim/3
DwConv 5×5 dim/3
DwConv 7×7 dim/3
Concat
GELU
BN; Conv 1×1 4×dim; GELU
Conv 1×1 dim

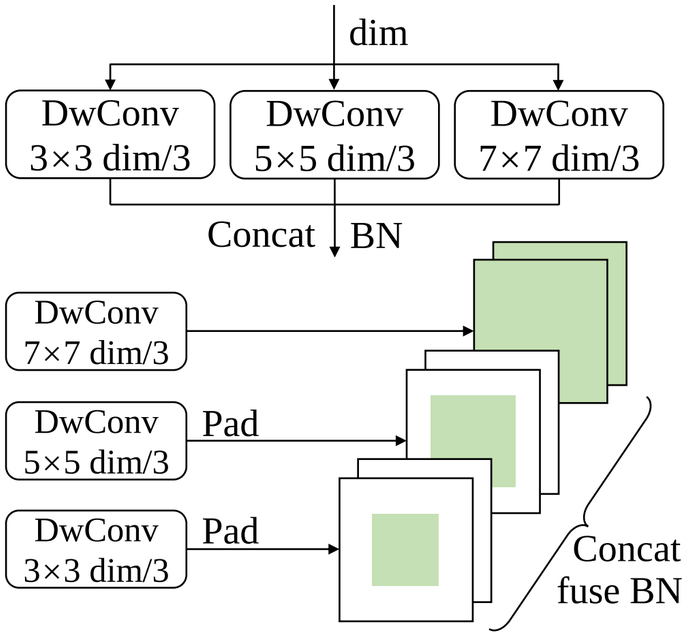

RepNeXt Block Training-time
DwConv 3×3 dim BN
GELU
Conv 1×1 4×dim; GELU
Conv 1×1 dim

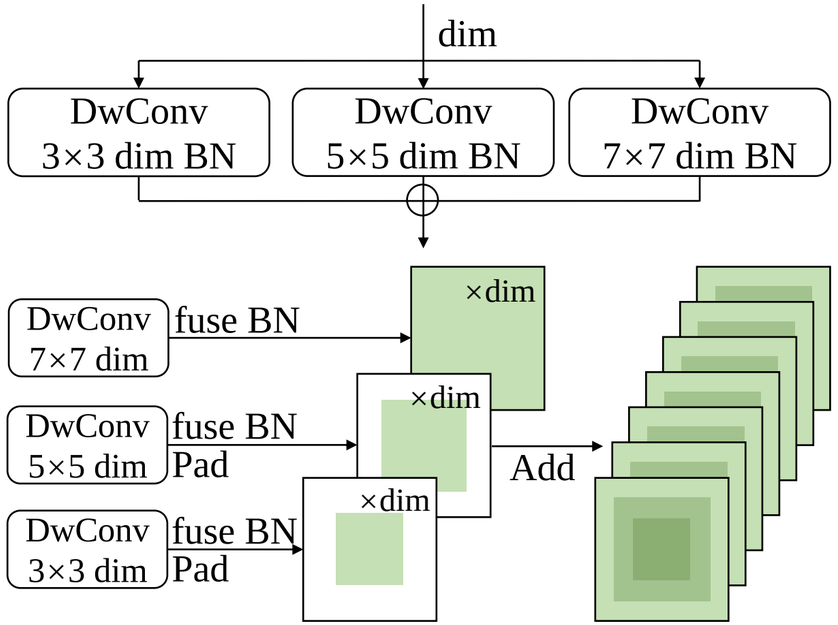

ConvNeXt Block 2022
dim
DwConv 7×7 dim;
LayerNorm
PWConv(FC) 4×dim;
GELU
PWConv(FC) dim
ConvNeXt-ERep Block
dim
DwConv 7×7 dim;
BatchNorm
GELU
PWConv 4×dim;
GELU
PWConv dim
RepNeXt Block Training-time
dim
RepUnit dim
(DWConv 7×7)
GELU
PWConv 4×dim;
GELU
PWConv dim
RepNeXt Block Inference-time
dim
GwConv 7×7
Group=dim 2×dim
PWConv 5×dim;
GELU
PWConv dim

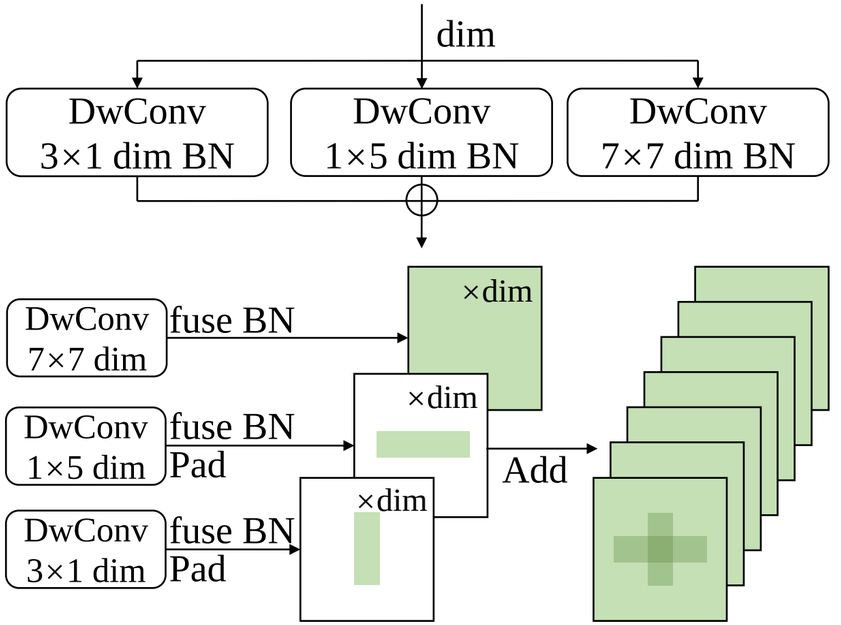

dim
DwConv 7×7 dim;
LayerNorm
PWConv(FC) 4×dim;
GELU
PWConv(FC) dim
dim
RepUnit dim
(DWConv 7×7)
GELU
PWConv 4×dim;
GELU
PWConv dim
dim
GwConv 7×7
Group=dim 2×dim
PWConv 5×dim;
GELU
PWConv dim

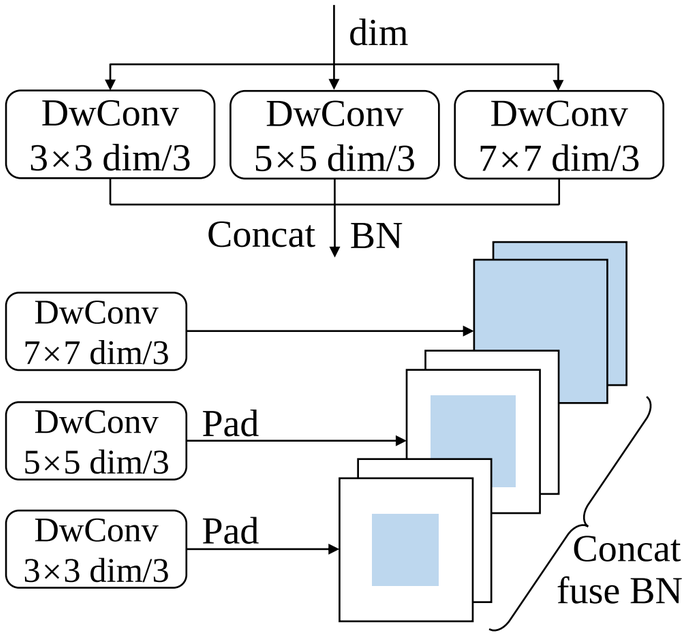

dim
RepUnit
(DWConv 7×7)
dim
GELU
PwConv; GELU
4×dim
PwConv
dim
dim
GwConv 7×7
Group=dim
2×dim
PwConw; GELU
5×dim
PwConv
dim

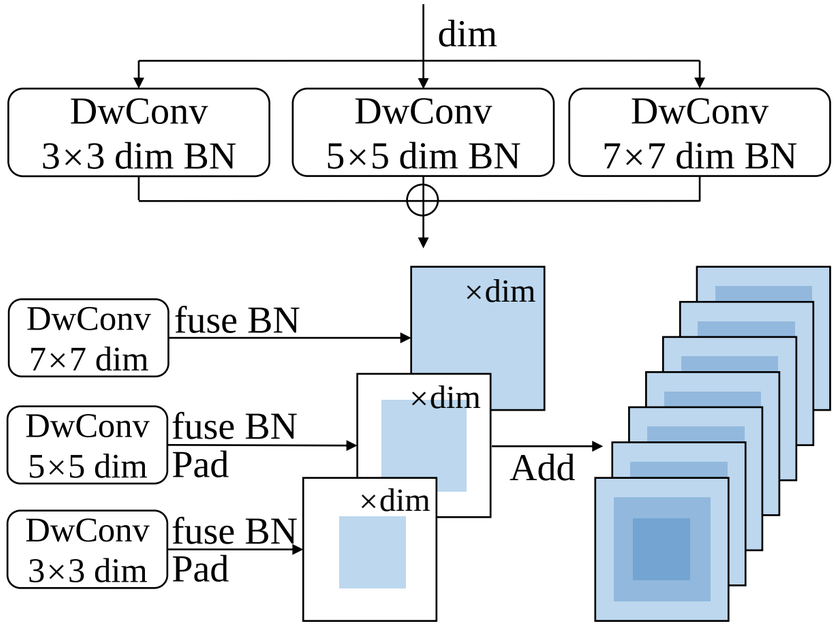

ConvNeXt-ERep Block
dim
DwConv 7×7 dim;
BatchNorm
PWConv 4×dim;
GELU
PWConv dim
GELU

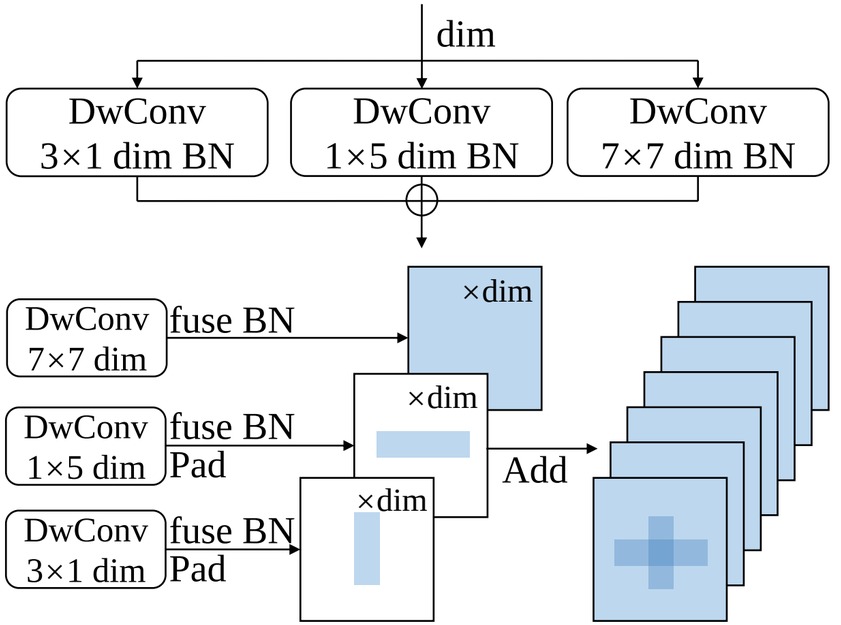

dim
DwConv
3×3 dim/3
DwConv
5×5 dim/3
DwConv
7×7 dim/3
Concat
BN
DwConv
7×7 dim/3
DwConv
5×5 dim/3
Pad
DwConv
3×3 dim/3
Pad
Concat
fuse BN

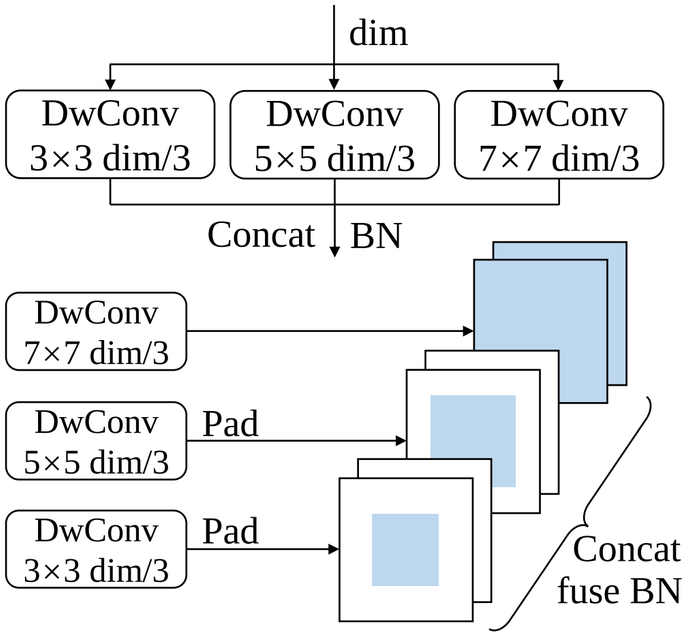

dim
DwConv
3×3 dim BN
DwConv
5×5 dim BN
DwConv
7×7 dim BN
DwConv
7×7 dim
fuse BN
×dim
DwConv
5×5 dim
fuse BN
Pad
×dim
DwConv
3×3 dim
fuse BN
Pad
×dim
Add

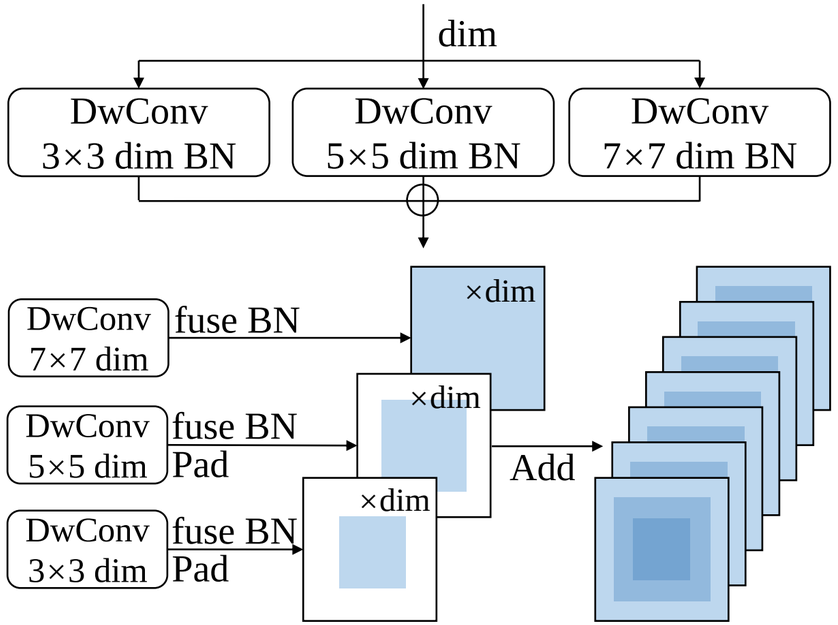

dim
DwConv
3×1 dim BN
DwConv
1×5 dim BN
DwConv
7×7 dim BN
DwConv
7×7 dim
fuse BN
×dim
DwConv
1×5 dim
fuse BN
Pad
×dim
DwConv
3×1 dim
fuse BN
Pad
×dim
Add

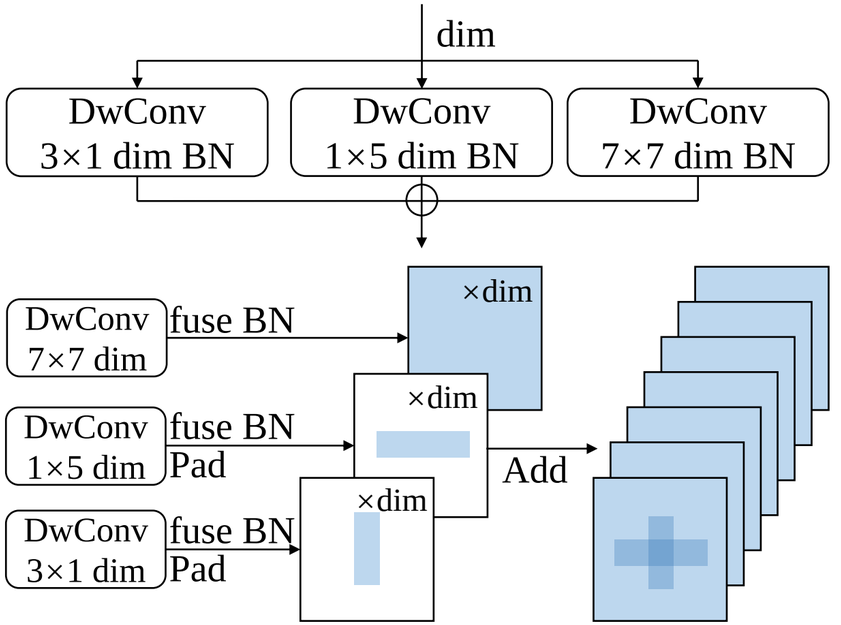

dim
DwConv
3×3 dim/3
DwConv
5×5 dim/3
DwConv
7×7 dim/3
Concat
BN
DwConv
7×7 dim/3
DwConv
5×5 dim/3
Pad
DwConv
3×3 dim/3
Pad
Concat
fuse BN

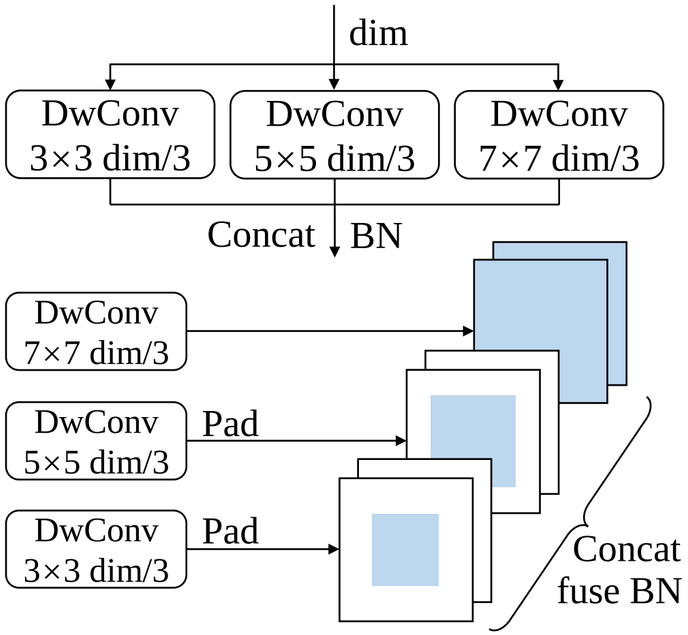

dim
DwConv
3×3 dim BN
DwConv
5×5 dim BN
DwConv
7×7 dim BN
DwConv
7×7 dim
fuse BN
×dim
DwConv
5×5 dim
fuse BN
Pad
×dim
DwConv
3×3 dim
fuse BN
Pad
×dim
Add

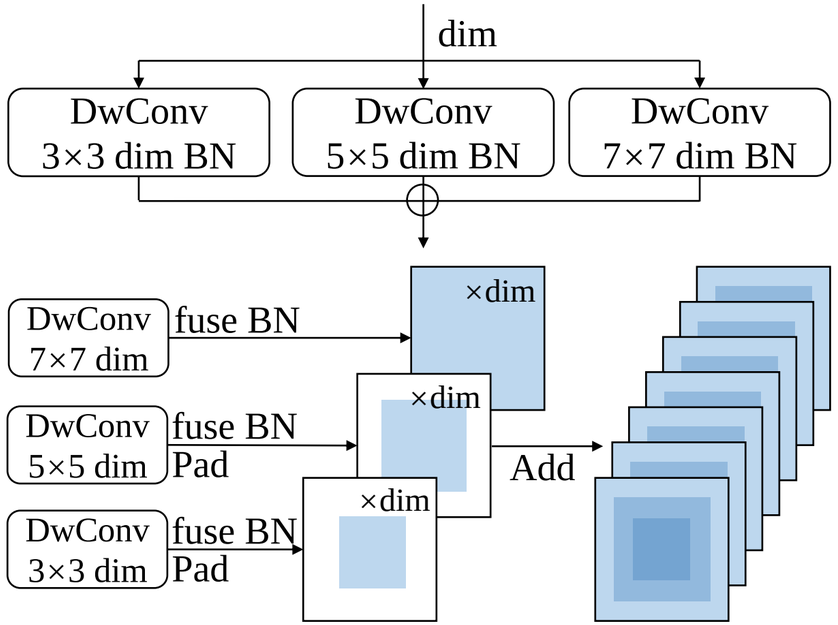

dim
DwConv
3×1 dim BN
DwConv
1×5 dim BN
DwConv
7×7 dim BN
DwConv
7×7 dim
fuse BN
×dim
DwConv
1×5 dim
fuse BN
Pad
×dim
DwConv
3×1 dim
fuse BN
Pad
×dim
Add

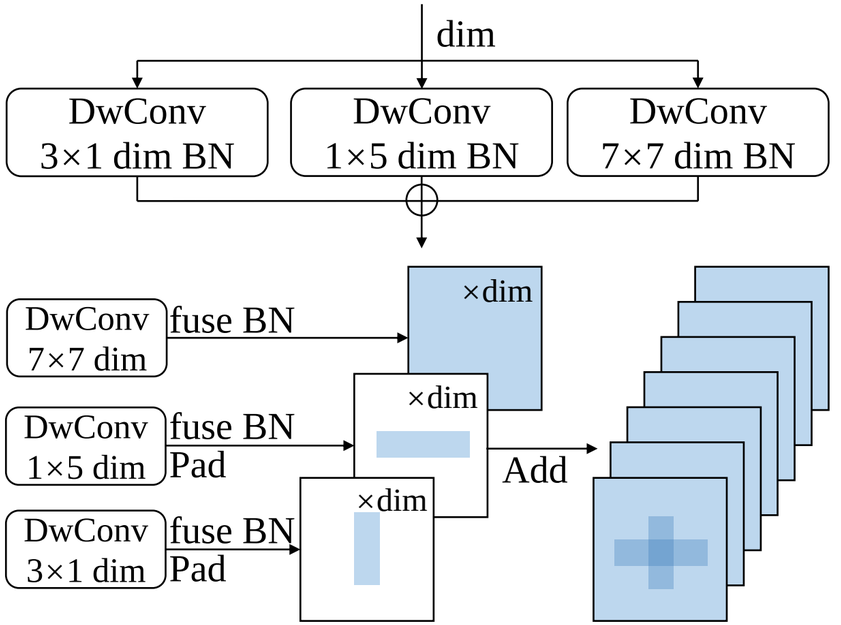

dim
DwConv
3×3 dim/3
DwConv
5×5 dim/3
DwConv
7×7 dim/3
Concat
BN
DwConv
7×7 dim/3
DwConv
5×5 dim/3
Pad
DwConv
3×3 dim/3
Pad
Concat
fuse BN

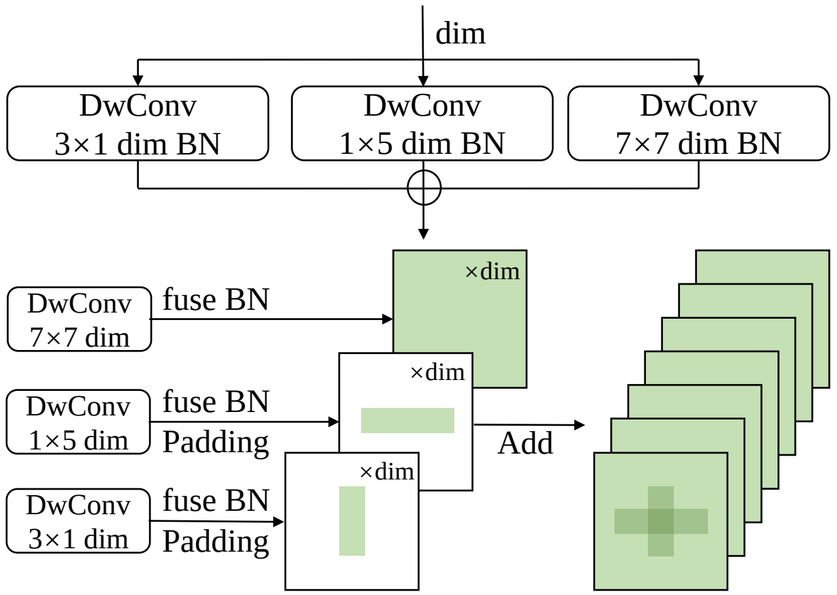

dim
DwConv
3×3 dim BN
DwConv
5×5 dim BN
DwConv
7×7 dim BN
DwConv
7×7 dim
fuse BN
×dim
DwConv
5×5 dim
fuse BN
Pad
×dim
DwConv
3×3 dim
fuse BN
Pad
×dim
Add

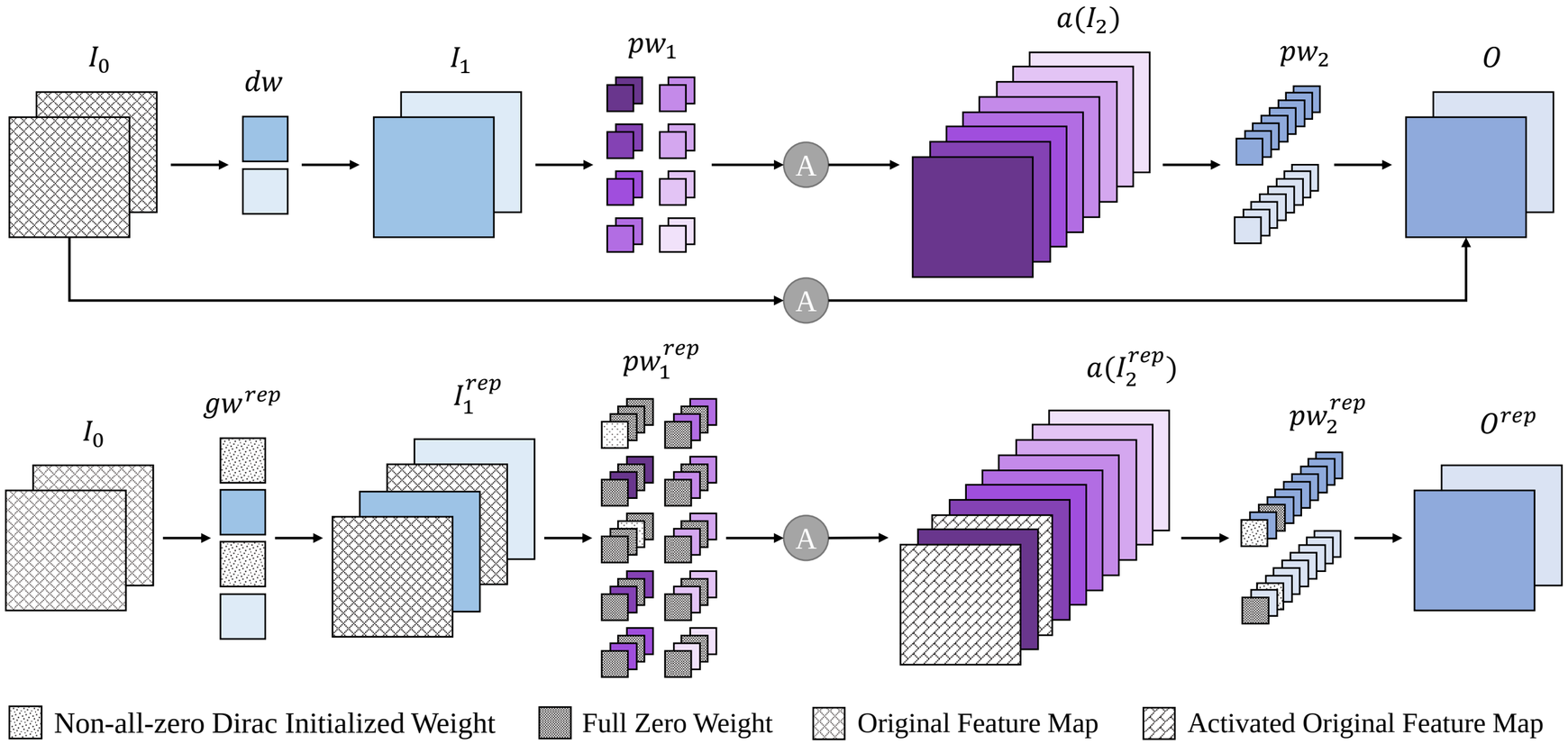

dim
DwConv
3×1 dim BN
DwConv
1×5 dim BN
DwConv
7×7 dim BN
DwConv
7×7 dim
fuse BN
×dim
DwConv
1×5 dim
fuse BN
Pad
×dim
DwConv
3×1 dim
fuse BN
Pad
×dim
Add

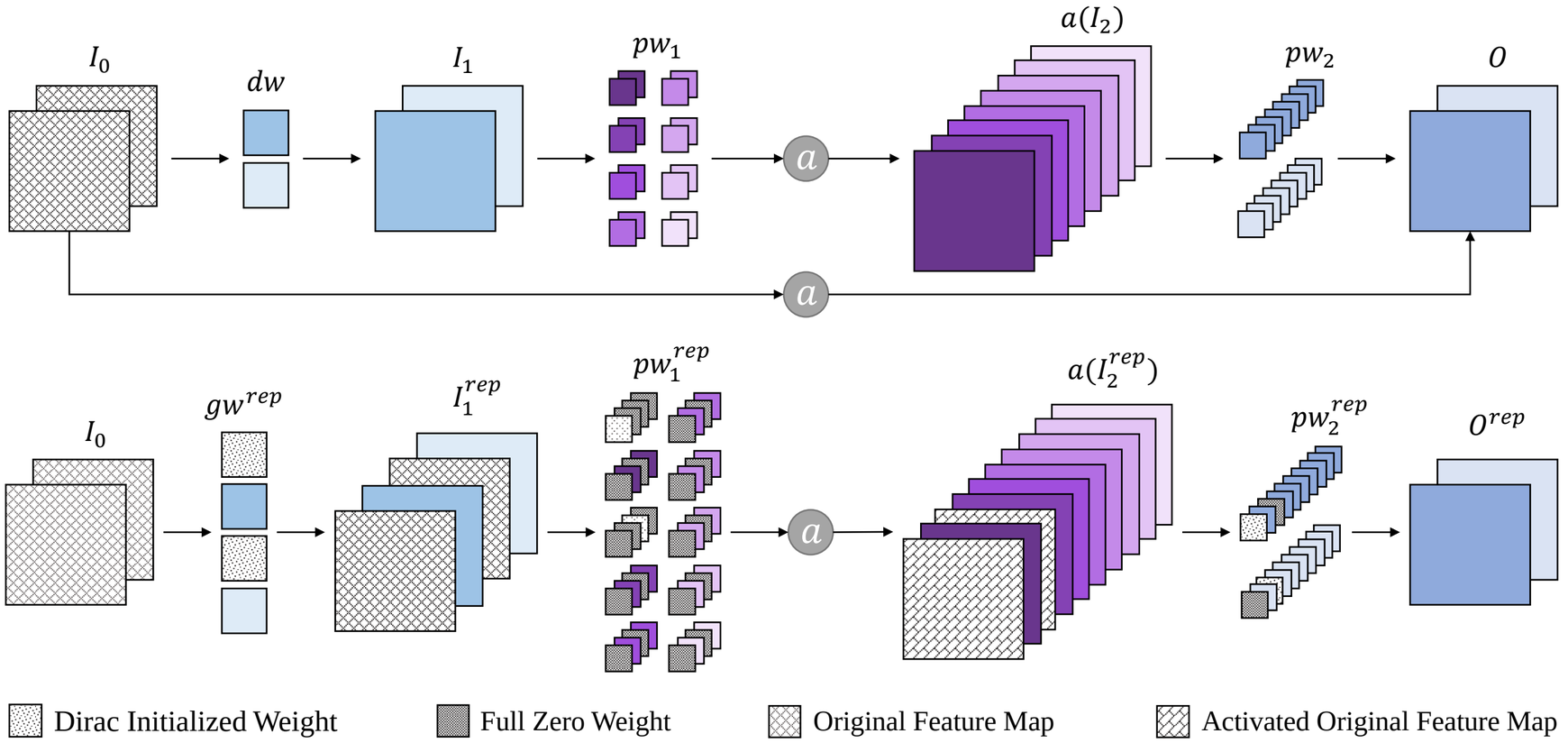

dim
DwConv
3×3 dim/3
DwConv
5×5 dim/3
DwConv
7×7 dim/3
Concat
BN
DwConv
7×7 dim/3
DwConv
5×5 dim/3
Pad
DwConv
3×3 dim/3
Pad
Concat
fuse BN

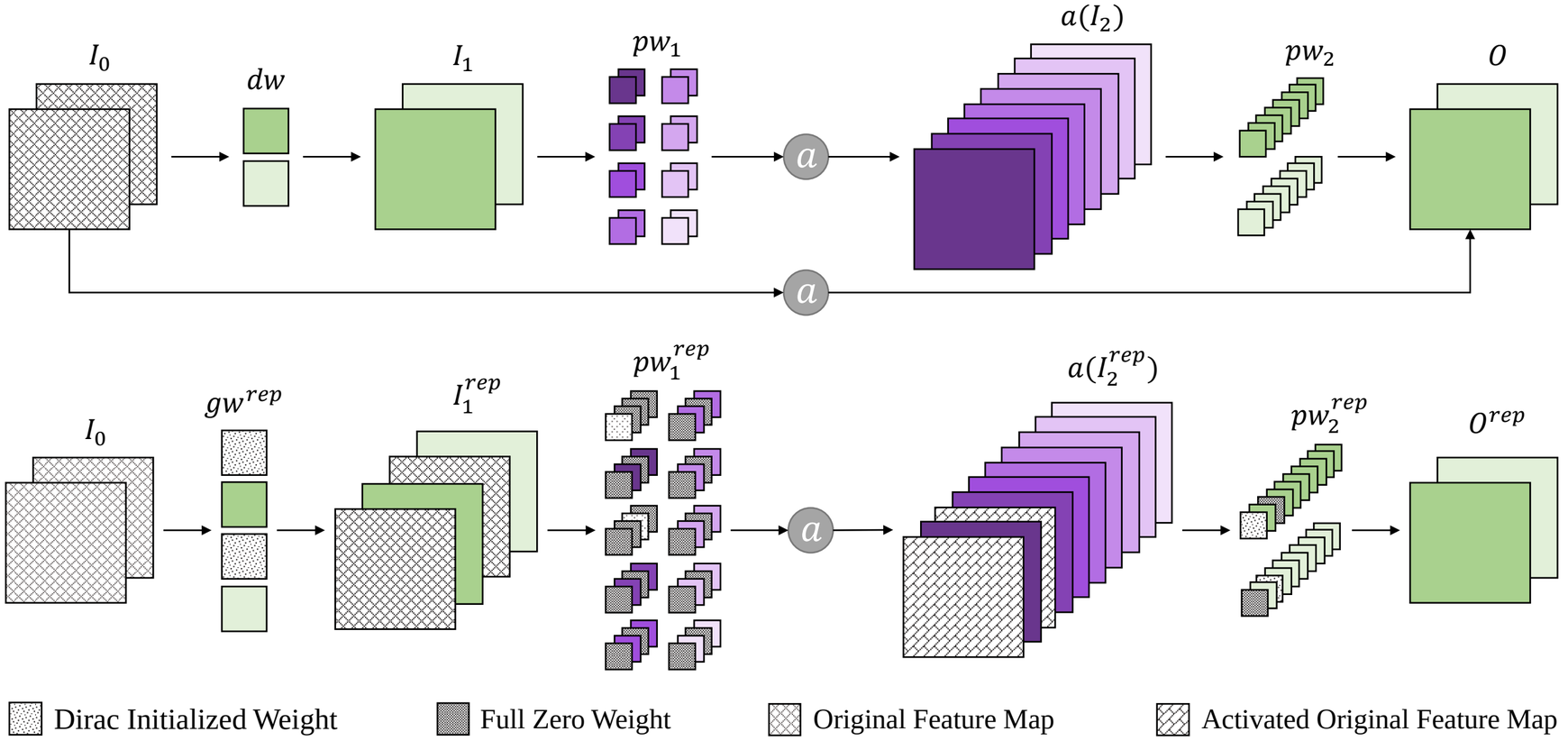

dim
DwConv
3×3 dim BN
DwConv
5×5 dim BN
DwConv
7×7 dim BN
DwConv
7×7 dim
fuse BN
×dim
DwConv
5×5 dim
fuse BN
Pad
×dim
DwConv
3×3 dim
fuse BN
Pad
×dim
Add

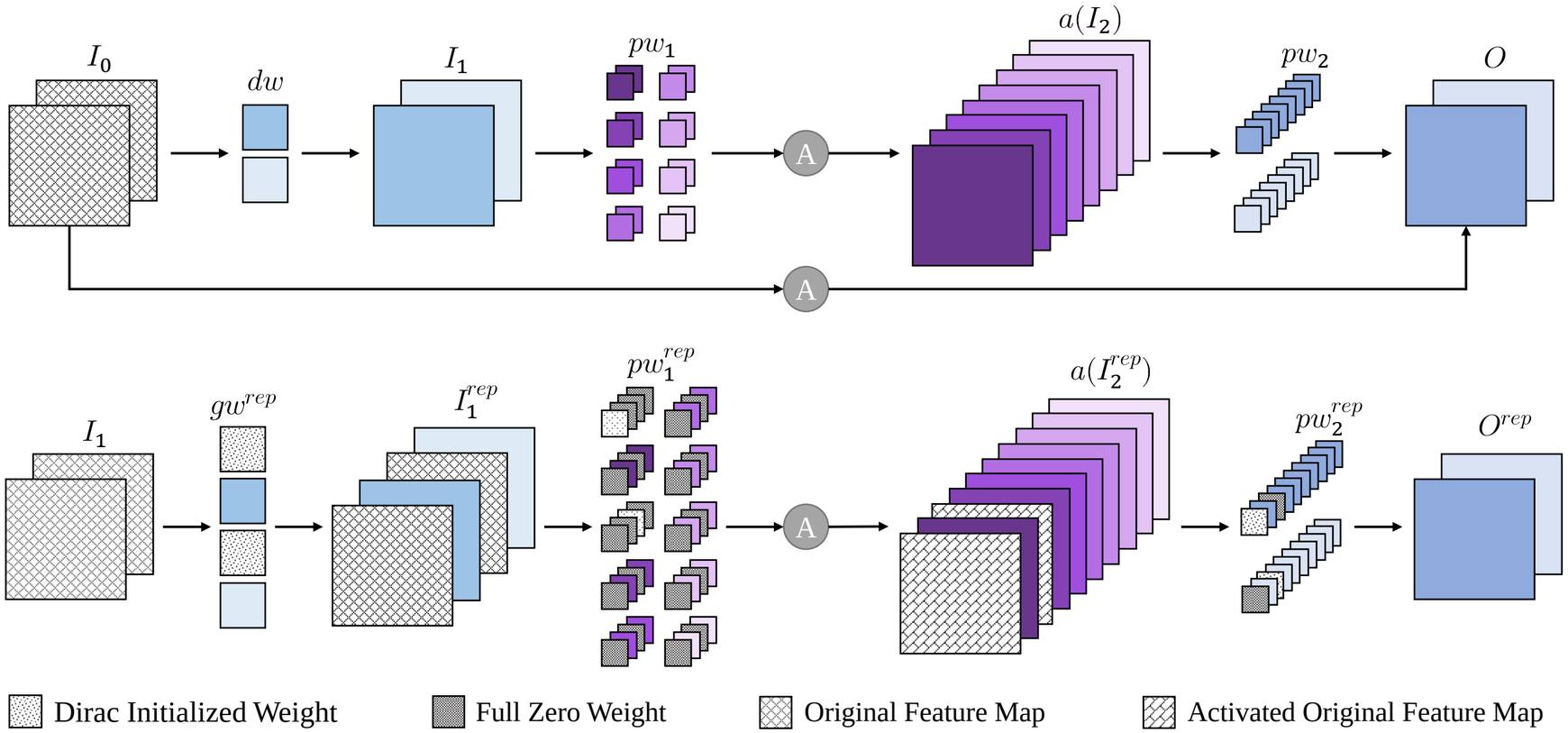

dim
DwConv
3×1 dim BN
DwConv
1×5 dim BN
DwConv
7×7 dim BN
DwConv
7×7 dim
fuse BN
×dim
DwConv
1×5 dim
fuse BN
Pad
×dim
DwConv
3×1 dim
fuse BN
Pad
×dim
Add

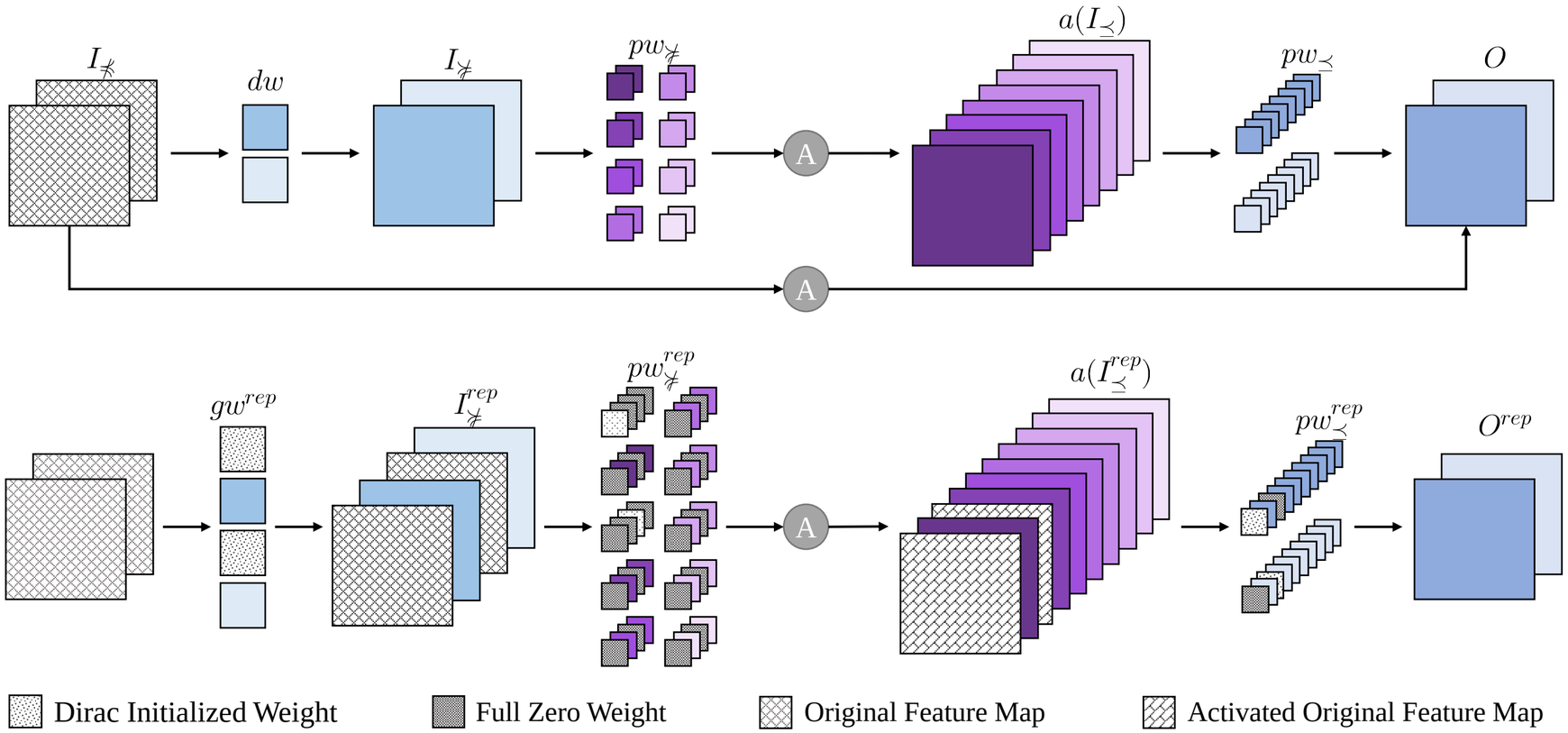

dim
DwConv
3×1 dim BN
DwConv
1×5 dim BN
DwConv
7×7 dim BN
DwConv
7×7 dim
fuse BN
×dim
DwConv
1×5 dim
fuse BN
Padding
×dim
DwConv
3×1 dim
fuse BN
Padding
×dim
Add


I_0
dw
I_1
pw_1
$a(I_2)$
pw_2
O
A
A
I_0
gw^{rep}
I_1^{rep}
pw_1^{rep}
$a(I_2^{rep})$
pw_2^{rep}
O^{rep}
A
Non-all-zero Dirac Initialized Weight
Full Zero Weight
Original Feature Map
Activated Original Feature Map


I_0
dw
I_1
pw_1
a
$a(I_2)$
pw_2
O
a
I_0
gw^{rep}
I_1^{rep}
pw_1^{rep}
a
$a(I_2^{rep})$
pw_2^{rep}
O^{rep}
Dirac Initialized Weight
Full Zero Weight
Original Feature Map
Activated Original Feature Map


I_0
dw
I_1
pw_1
a
$a(I_2)$
pw_2
O
a
I_0
gw^{rep}
I_1^{rep}
pw_1^{rep}
a
$a(I_2^{rep})$
pw_2^{rep}
O^{rep}
Dirac Initialized Weight
Full Zero Weight
Original Feature Map
Activated Original Feature Map


I_0
dw
I_1
pw_1
A
$a(I_2)$
pw_2
O
A
I_1
gw^{rep}
I_1^{rep}
pw_1^{rep}
A
$a(I_2^{rep})$
pw_2^{rep}
O^{rep}
Dirac Initialized Weight
Full Zero Weight
Original Feature Map
Activated Original Feature Map



$I_{\not\prec\preceq\succeq\lhd\rhd}$