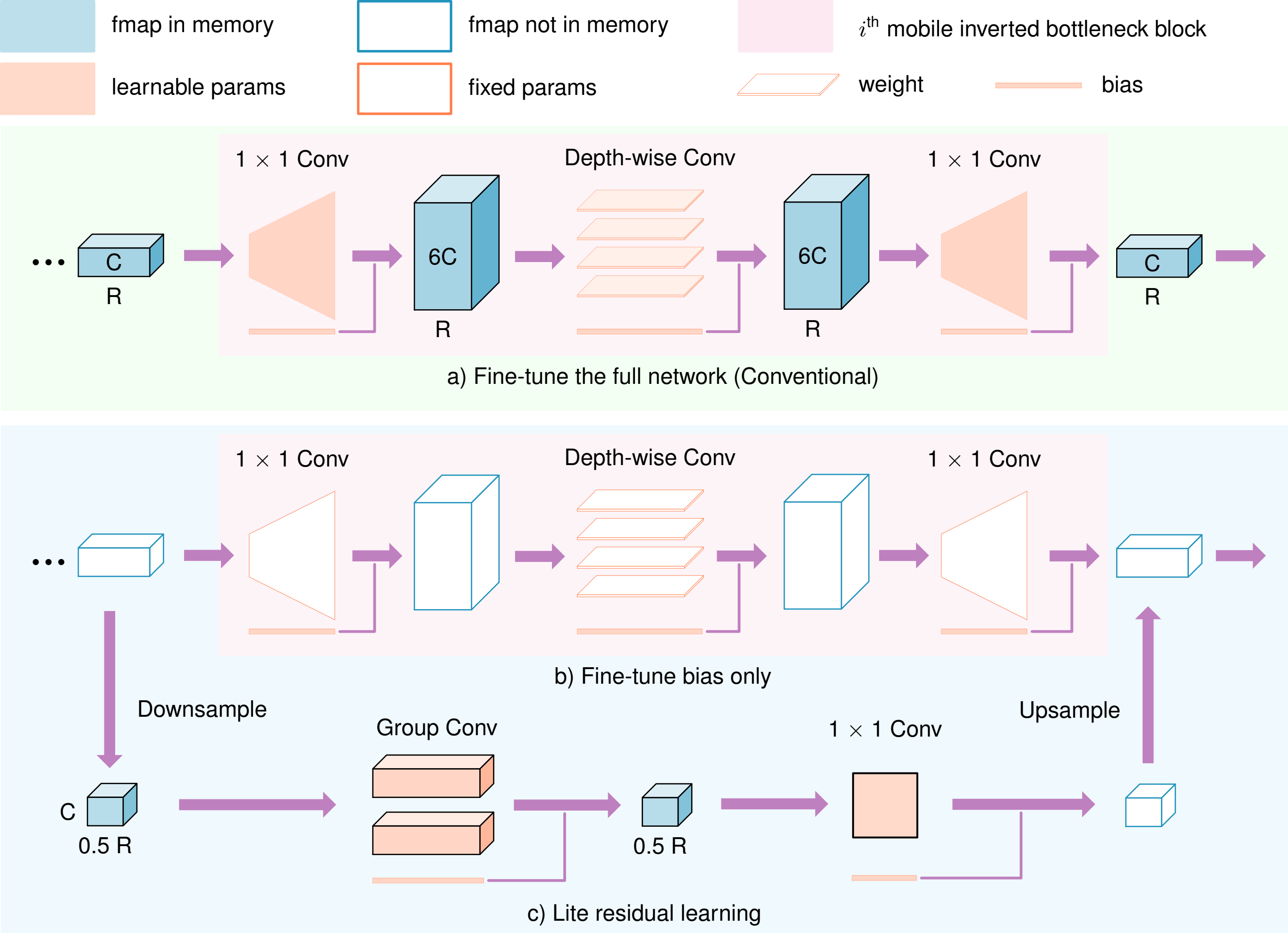
fmap in memory
fmap not in memory
i^{th} mobile inverted bottleneck block
learnable params
fixed params
weight
bias
1×1 Conv
Depth-wise Conv
1×1 Conv
C
R
6C
R
6C
R
C
R
a) Fine-tune the full network (Conventional)
1×1 Conv
Depth-wise Conv
1×1 Conv
b) Fine-tune bias only
Downsample
Group Conv
1×1 Conv
Upsample
C
0.5 R
0.5 R
c) Lite residual learning