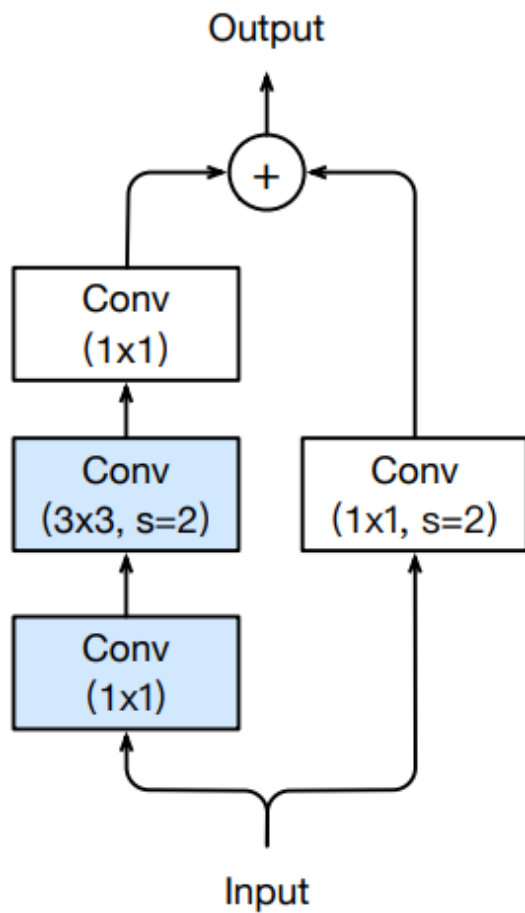


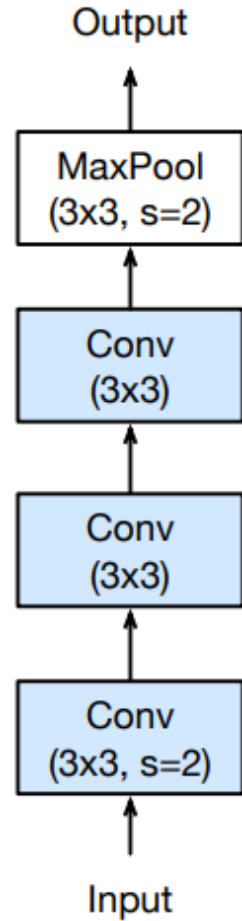
Bag of Tricks

- Model tweaks ResNet - B,C,D
- Linear scaling learning rate: $lr = 0.1 \cdot \frac{b}{256}, b - batch_size$
- Learning rate warmup
- Cosine learning rate decay
- Label smoothing
- No bias decay
- Zero gamma

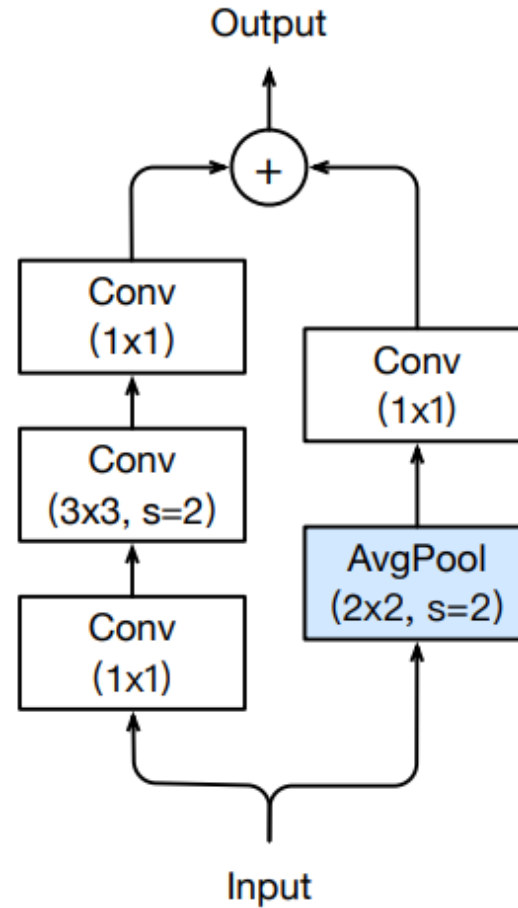
Model tweaks



(a) ResNet-B



(b) ResNet-C



(c) ResNet-D

ResNet-B: в остаточных блоках (down sampling) меняются местами значение stride у первых двух сверток (путь A).

ResNet-C: во входном блоке свертка 7×7 заменяется тремя сверточными слоями 3×3 , у первых двух сверток количество фильтров - 32, у третьей свертки - 64.

ResNet-D: в остаточных блоках (down sampling) добавляется average pooling слой перед сверткой 1×1 в пути B, а в самой свертке используйте шаг 1.