Design principles of convolutional networks

Use striding, increase channels

- Trade spatial resolution for channels

- Balance computation
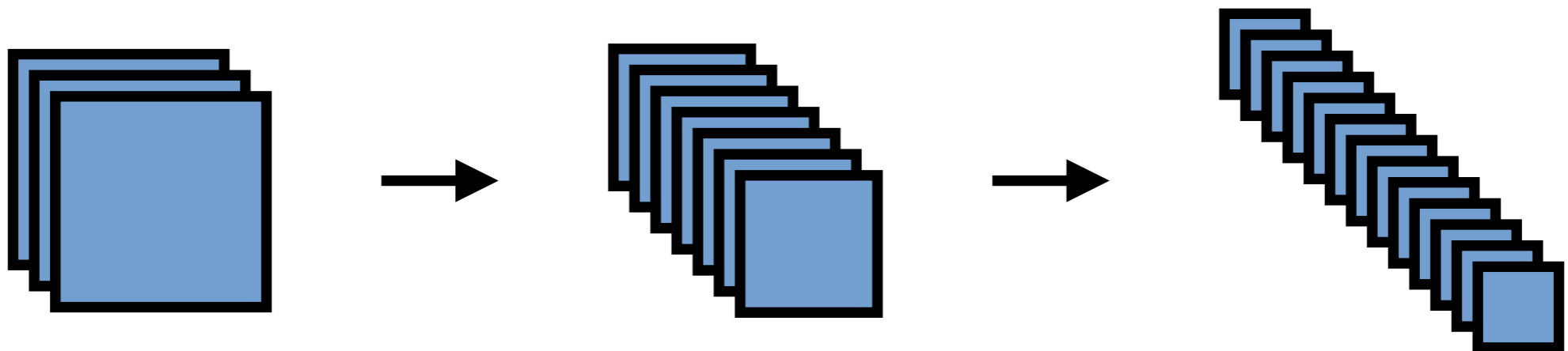
Keep kernels small

- 3x3 kernels almost everywhere

- exception:

  - first layer up to 7x7
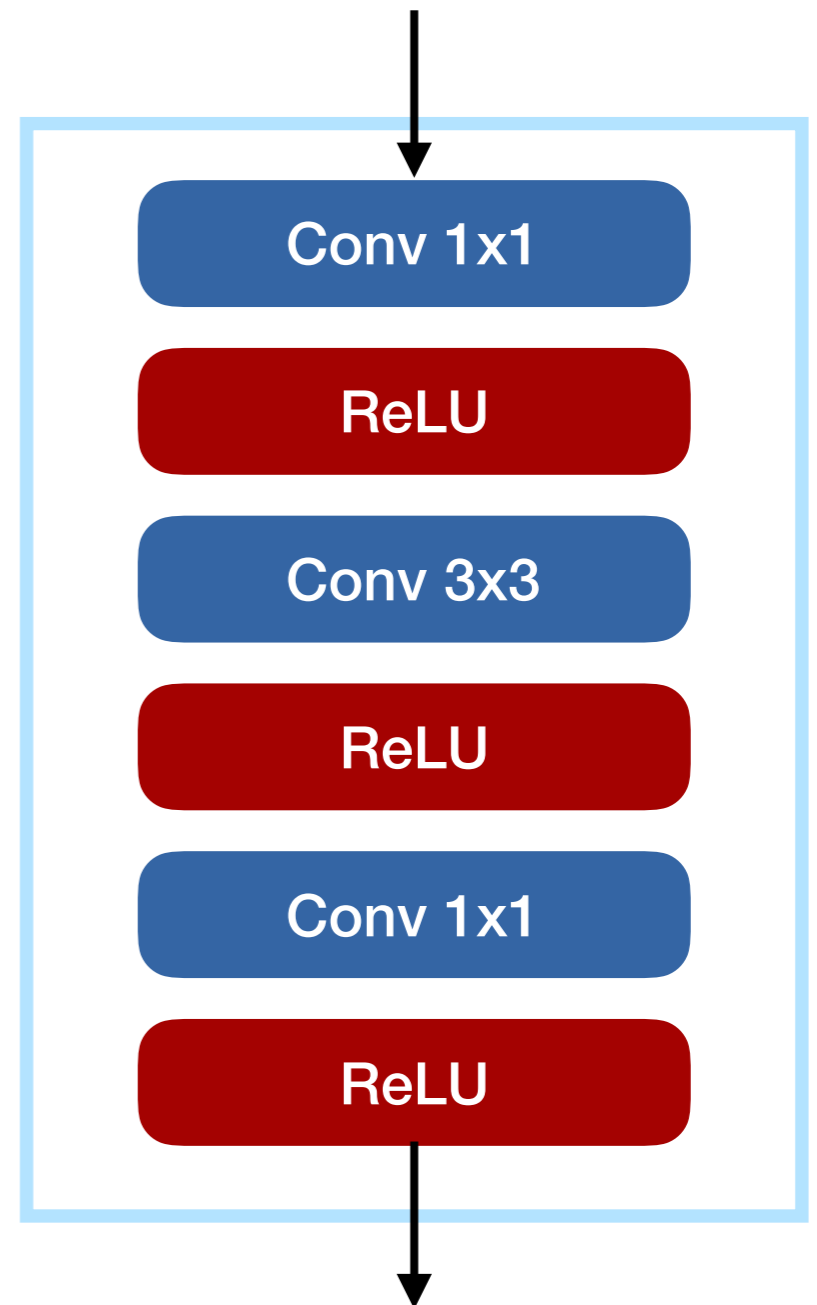
Repeat patterns

- First layer or two are special and not repeated

- All others usually follow a fixed pattern

All-convolutional

- Average in the end

- Fewer parameters

- Better training signal

- "Ensemble"/voting effect for testing