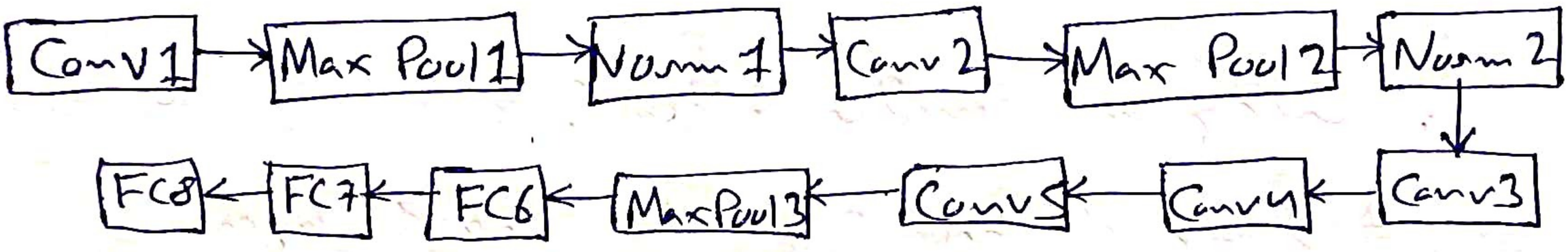
⑨ CNN Architectures

★ Case Study: Alex Net

[Krizhevsky et al. 2012]

⇒ First large scale CNN, that was able to do well on image net classification task.

Architecture



⇒ Input: 227 × 227 × 3 images

Conv1

96 11×11 filters applied at stride 4

Max Pool 1

3×3 filters applied at Stride 2

⇒ Details

- → First use of ReLU
- → Used Norm layers. (not common anymore)
- → dropout 0.5
- → batch size 128
- → SGD Momentum 0.9
- → Learning rate $1e^{-2}$ reduced by 10 manually when val accuracy plateaus
- → L2 weight decay $5e^{-4}$
- → 7 CNN ensemble 18.2% → 15.4%
- → heavy data augmentation.

★ GoogleNet and VGG

↓ 22 layers

↘ 19 layers

⊛ VGG Net (Case Study) { VGG 16 VGG 19 }

→ Small filters, Deeper networks

→ Only 3×3 Conv stride 1, pad 1
and 2×2 Max pool stride 2

⑦ Why use Smaller filters?

⇒ Stack of three 3×3 Conv (stride 1) layers
have same effective receptive field as one
7×7 Conv layer

⇒ But deeper, more non linearities

⇒ And fewer parameters: $3 * (3^2 C^2)$ vs $7^2 C^2$
for C channels per layer

⊛ GoogleNet (Case Study)

→ Deeper networks, with computational efficiency

→ 22 layers

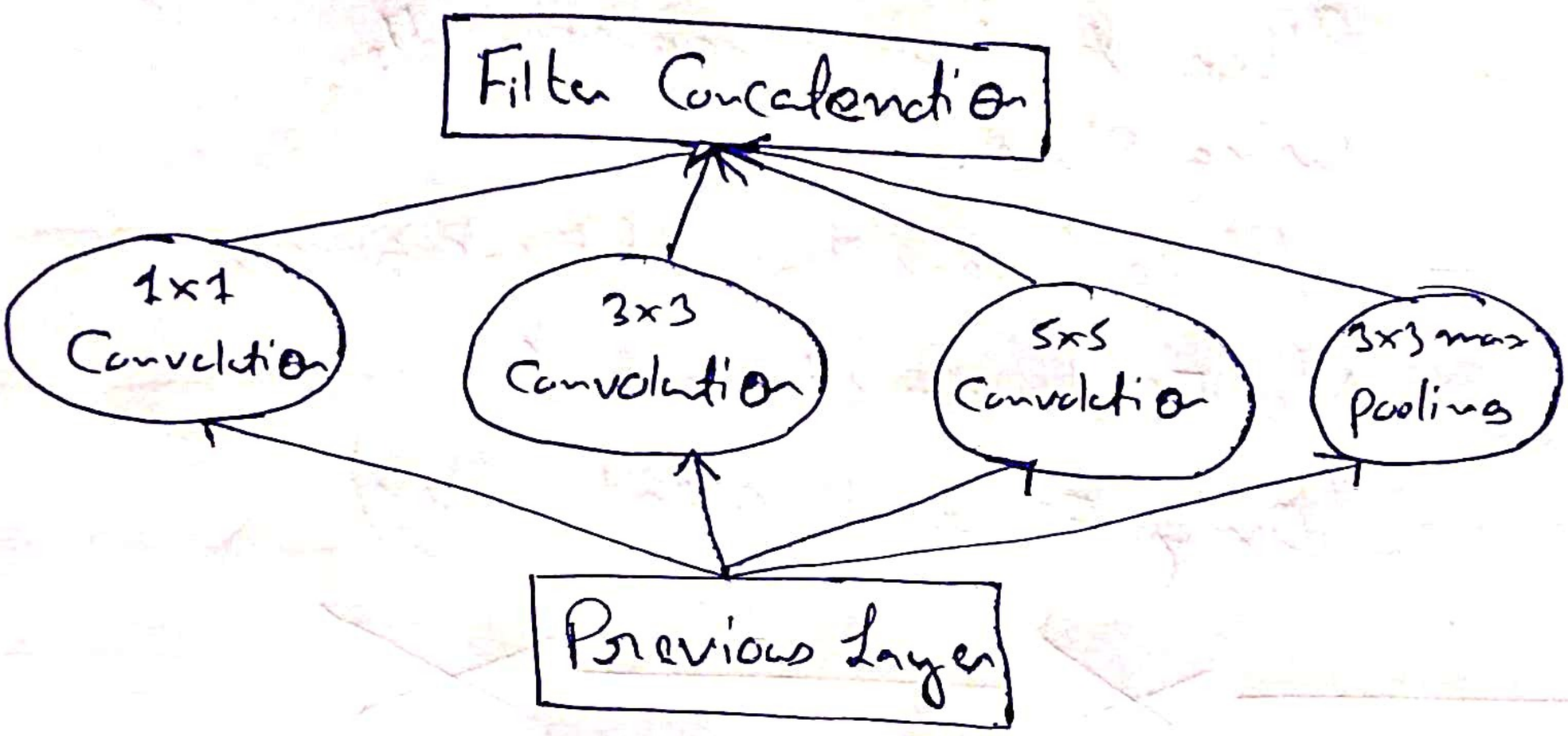
→ Efficient "Inception" module

→ No FC layers

→ Only 5 million parameters

Inception module

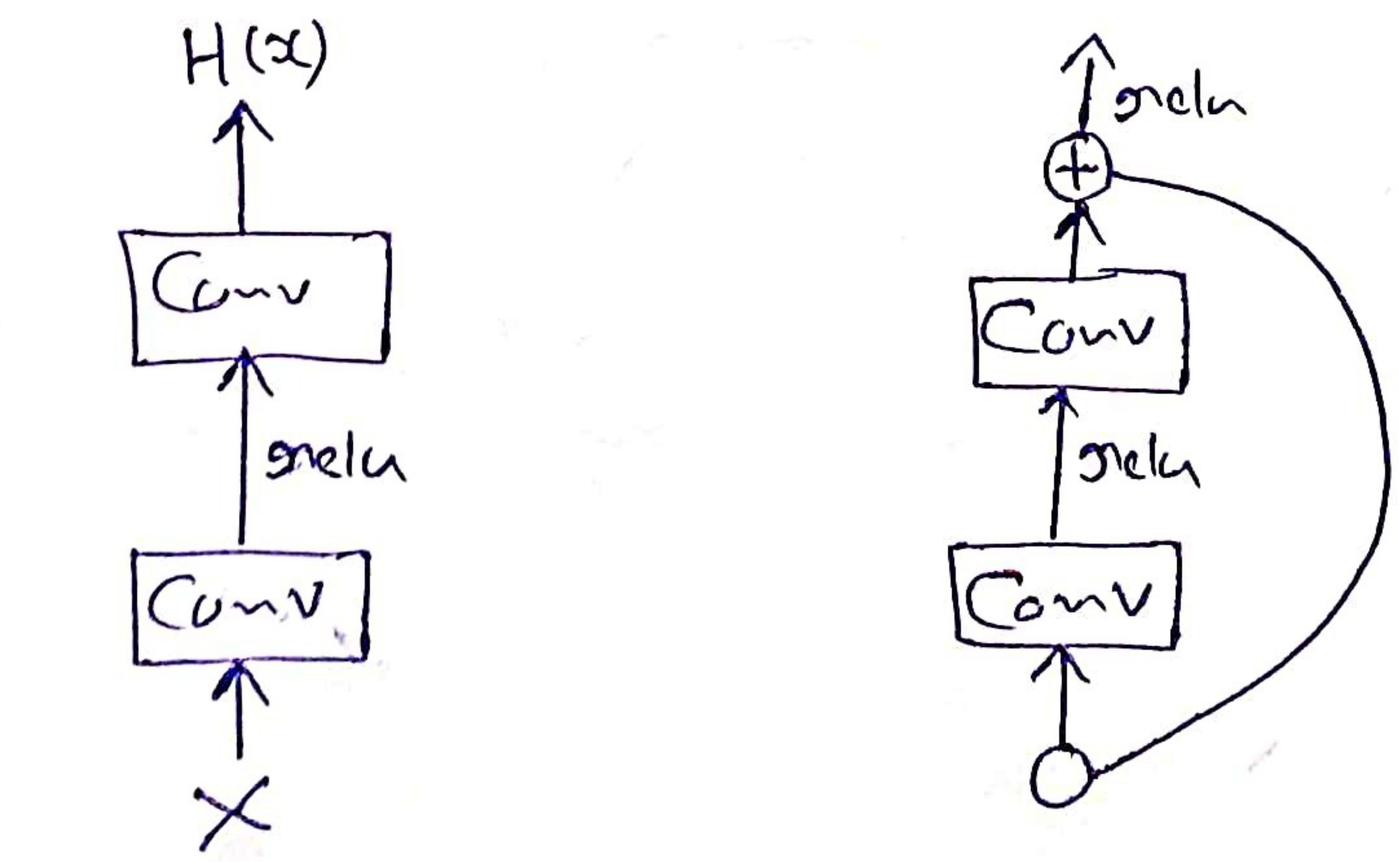
→ design good local network topology and then stack these modules on top of each other.



★ ResNet (Case Study)

→ 152 layers deep.

{ Very deep network using Residual Connection }



Plain Layer

Residual block

⇒ Full ResNet Architecture:

→ Stack residual blocks
→ Every residual block has two 3×3 conv layer.
→ Periodically, double the # of filters & downsample spatially using stride 2.
→ Additional conv layer at the beginning
→ No FC layers at the end
(Only FC 1000 to output classes)