VGGNET

It was proposed by Karen Simonyan & Andrew Zisserman and has two different architectures consisting of both 16 layers (known as VGG-16) and 19 layers (known as VGG-19) respectively.

"The key idea proposed in the paper was "Depth is a crucial factor. Stacking small filters gives better performance than having a few large filters. AlexNet used 11×11 filters which were large and thus lost spatial information early. Network of AlexNet was wide but not deep enough.

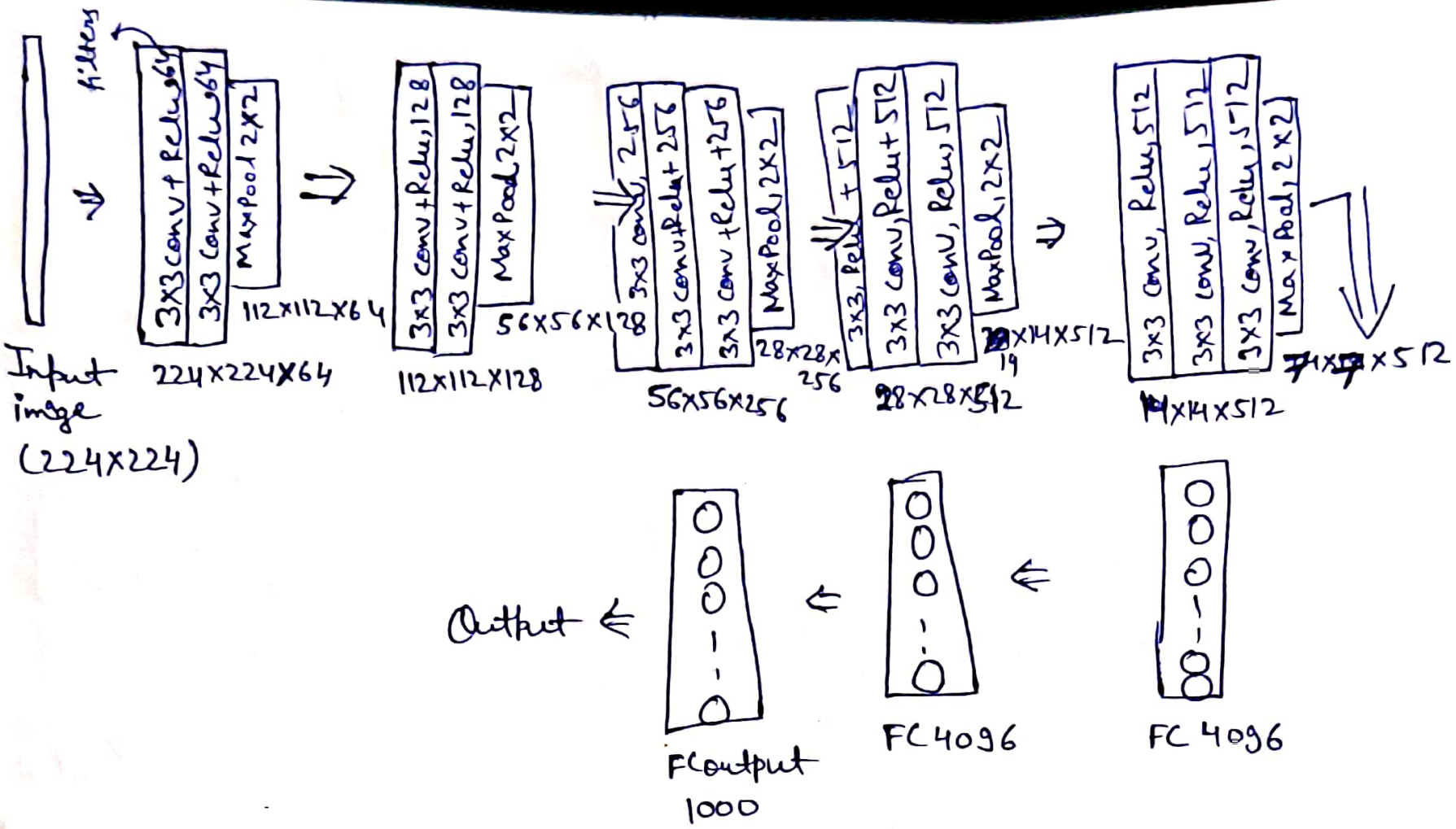
VGG Net architecture model achieved 2nd rank in ImageNet competition in 2014 (behind Google Net), but VGGNet was far more generalizable and influential in Transfer Learning.

Architecture ⇒ Core design principles:

1. Only 3×3 convolutions (smallest filter that captures directionality: up/down/left/right/center).
   * Stacking two filters (3×3) = receptive field of 5×5
   * Stacking three filters (3×3) = receptive field of 7×7
   * This replaces large filters (11×11) in AlexNet with deeper stacked small filters.
   * Advantage: More non-linearity, fewer parameters, better feature extraction.
2. Only 2×2 maxpooling (stride=2) for downsampling.
3. Depth: 16 or 19 (hence VGG-16/VGG-19)
4. Uniform Architecture: Same Conv-Pool blocks stacked repeatedly.
5. Fully connected layers at the end. (2×4096 + 1 softmax).

Main improvements of VGGNet:

→ Depth + Small filters gave much better feature extraction and generalization.
→ The uniform architecture made it scalable/easy to design deeper models).
→ Outperformed AlexNet significantly on ImageNet dataset.



VGG-16 Network has roughly 138M parameters (as compared to 60M in AlexNet). Because of these many parameters it has a large memory footprint and is slow in training and inference (needs high GPU resources).