

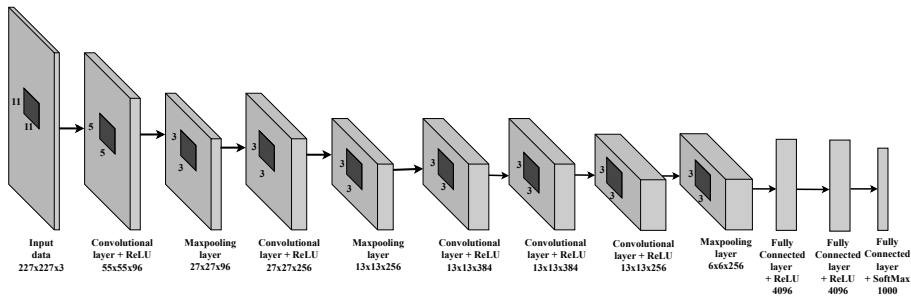
## 4. Convolutional Neural networks

### 4.3. Advanced CNN structures

Manel Martínez-Ramón  
Meenu Ajith  
Aswathy Rajendra Kurup

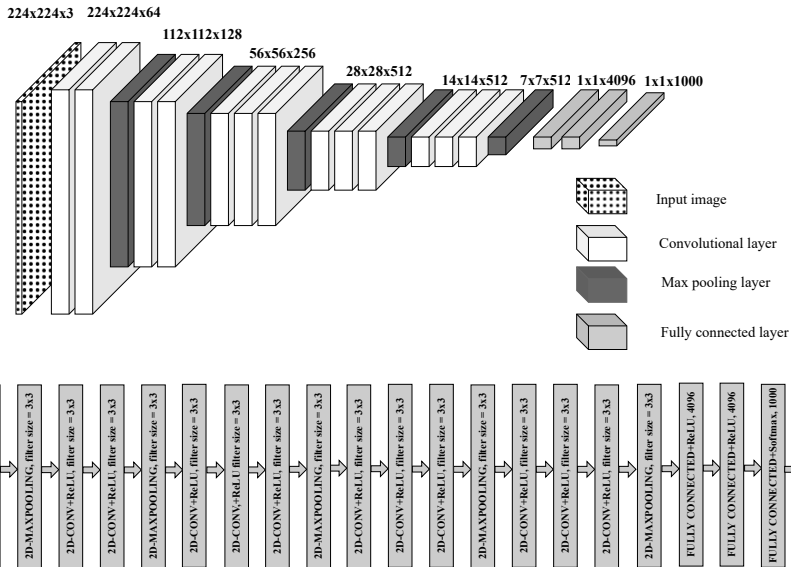
- Alex Krizhevsky, Ilya Sutskever and Geoffrey Hinton (2012).
- Able to solve the Imagenet image classification problem with 1000 classes.
- Won the 2-12 ImageNet Large Scale Visual Recognition Challenge (LSVRC) competition.
- Structure:
  - 8 layers and 60M learnable parameters.
  - 5 are convolutional layers along with max-pooling layers and 3 fully connected layers and ReLU non-linearity.
  - SoftMax with 1000 outputs.

# Alexnet



- RGB images,  $224 \times 224 \times 3$
- The convolutions include a stride of 4 pixels.
- A padding of 3 pixels are possibly used.
- The fifth convolutional layer contains 256 kernels of size  $3 \times 3 \times 192$
- Each of the hidden fully connected layers has 4096 neurons.
- Novelties:
  - Data augmentation by a factor of 2048 by extracting random patches from the images and altering the intensities of the RGB channels.
  - Dropout, ReLU, overlapping pooling, multi-GPU.

- VGG stands for Visual Geometric Group. This network was developed in the year 2014 by Karen Simonyan and Andrew Zisserman (Oxford University, 2014).
- Second place, 2014 ImageNet ILSVRC competition.
- Main idea: increase the depth of the convolutional network in large-scale image recognition settings.
- Smaller convolutional filters were used for increased depth.



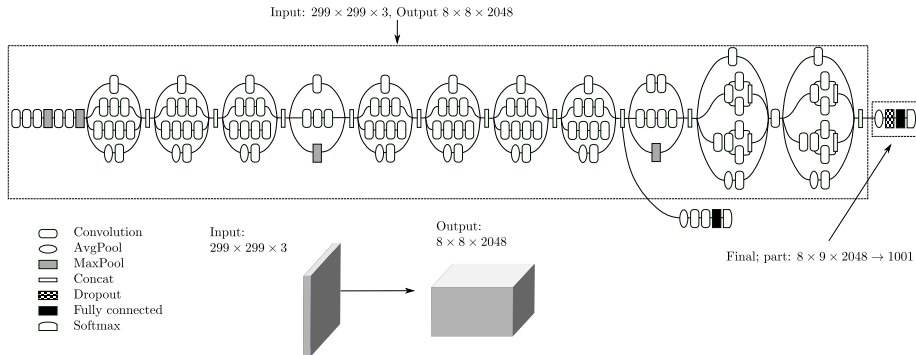
- Depth is achieved through a stack of convolutional layers.
- Convolutional kernels of smaller size ( $3 \times 3$ ).
- five max-pooling layers with a window size of  $2 \times 2$  and a stride 2.
- ReLU activations.
- Three fully connected layers with 4096, 4096 and 1000 nodes, with softmax activation.
- Incorporates  $1 \times 1$  convolutions, that are just linear transformations.

- First to use a preprocessing block that crops the image to size  $224 \times 224$ , extracts the mean RGB value.
- Shows that  $3 \times 3$  kernels stacked several times is as effective as a  $5 \times 5$  or  $7 \times 7$  kernel.
- 6 versions with 11 layers, 16 layers (VGG16) and 19 layers (VGG19), containing different number of convolution layers, but all of them kept the same number of fully connected layers and nodes.



- Inception-v1/GoogleNet has 22 layers and 5 million parameters.
- Winner of the 2014 ImageNet LSVR Competition.
- Previous structures are focused on increasing the depth. Inception has recurring blocks of convolutional designs called Inception modules.
- Around 100 layers.
- ReLU activations.
- Auxiliary classifiers in order to prevent the vanishing gradient problem.

# Inception

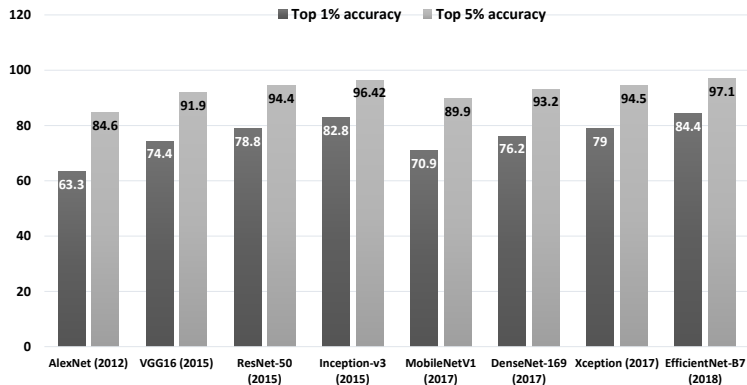


- Winner of the 2015 ImageNet localization, ImageNet detection, in the LSCR competition and segmentation and detection challenges in the 2015 Common Objects in COntext (COCO) competition.



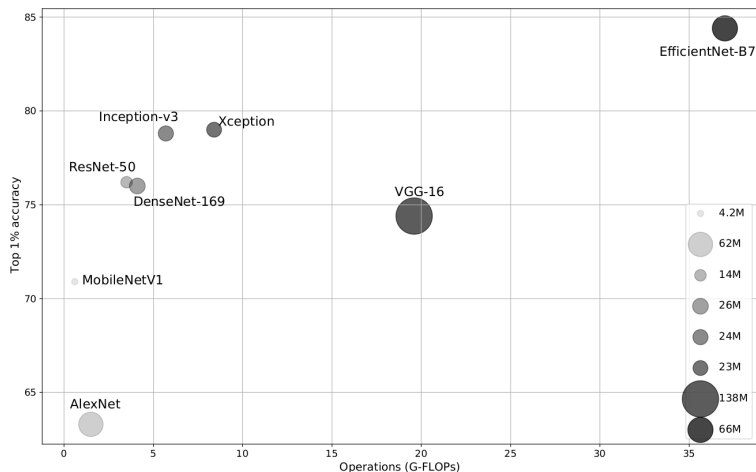
- Xception, François Chollet (2017).
- MobileNet, Google (2017).
- Densenet, Cornell University, Tsinghua University, Facebook, 2017.
- EfficientNet, Mingxing Tan and Quoc Le (2019).

# Comparisons



Comparison of top 1% and top 5% accuracy of different CNNs.

# Comparisons



Comparison of top 1% accuracy, Number of parameters and operations (G-FLOPs) of different CNN architectures for image classification.