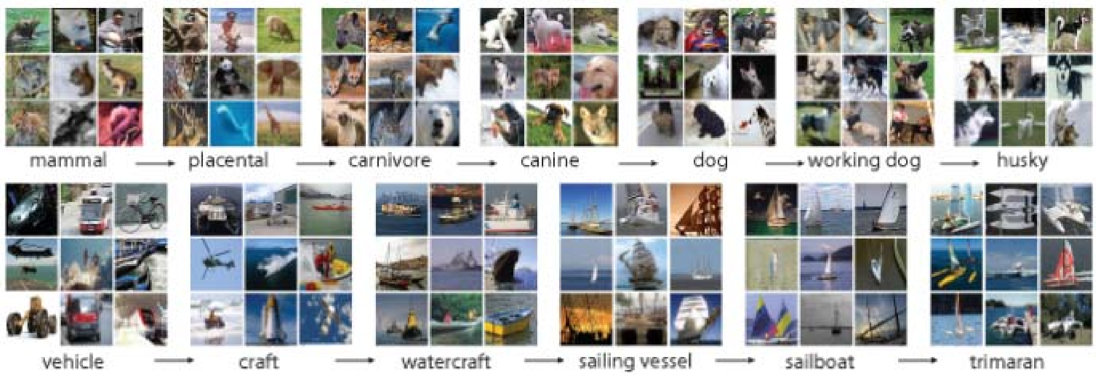
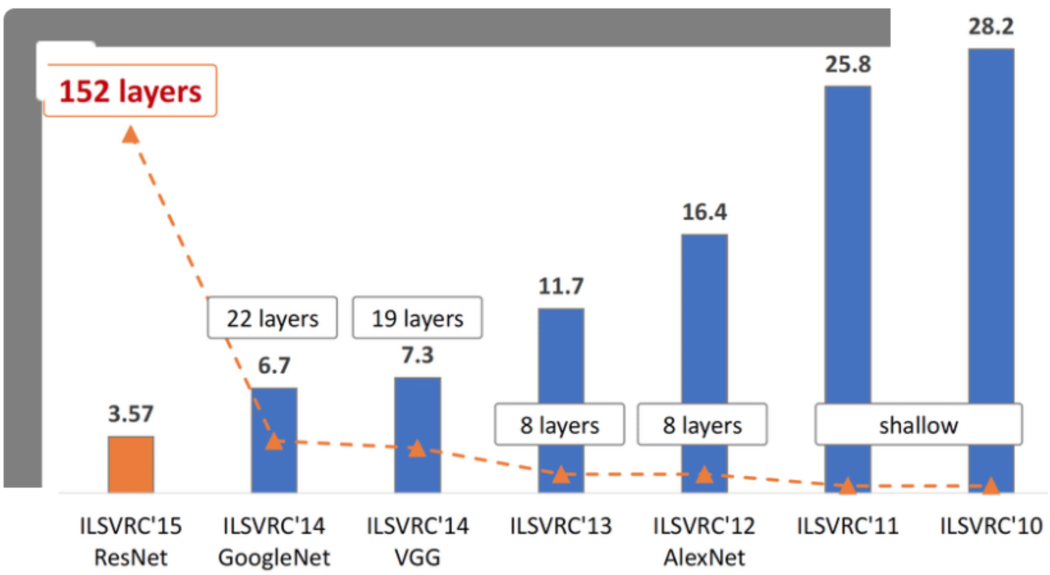
VGGNet

ImageNet ILSVRC competition deep NN
Ref “VERY DEEP CONVOLUTIONAL NETWORKS FOR LARGE-SCALE IMAGE RECOGNITION”
訓練資料：

ImageNet 是繼早期小型(例如 MNIST)後的大型影響辨識 Dataset，包含超過 1500 萬張有正確標記的影像，包含類別超過 22,000 種。ImageNet Large Scale Visual Recognition Challenge (ILS-VRC) 挑選出 ImageNet 當中一小部分的影像做為競賽的訓練樣本。共有 1000 種類別、每種類別有 1000 張左右的影像，總計 Training Set 約 120 萬張、Validation Set 5 萬張、Testing Set 1.5 萬張。

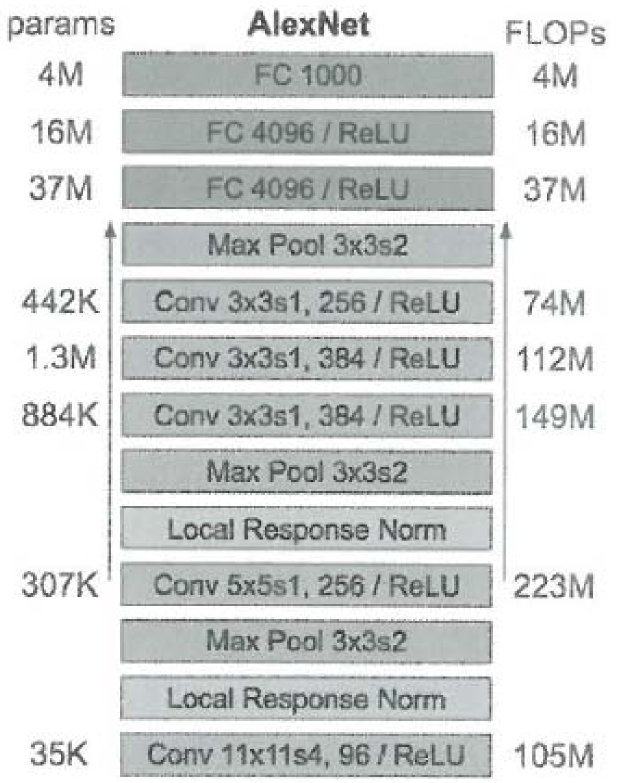


Winners of this competition Top 5 error rate

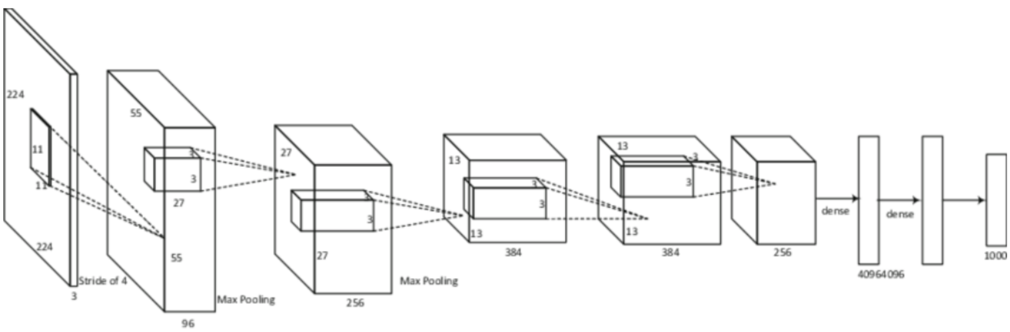


2012 AlexNet.
Reference to the Website for the details:
https://medium.com/@WhoYoung99/alexnet-%E6%9E%B6%E6%A7%8B%E6%A6%82%E8%BF%B0-988113c06b4b

Text p. 131-132
Structure of AlexNet: 5 convolution layers & 3 FC layers.



▲ 圖 6-5 AlexNet 每層的超參數及參數數量



VGG

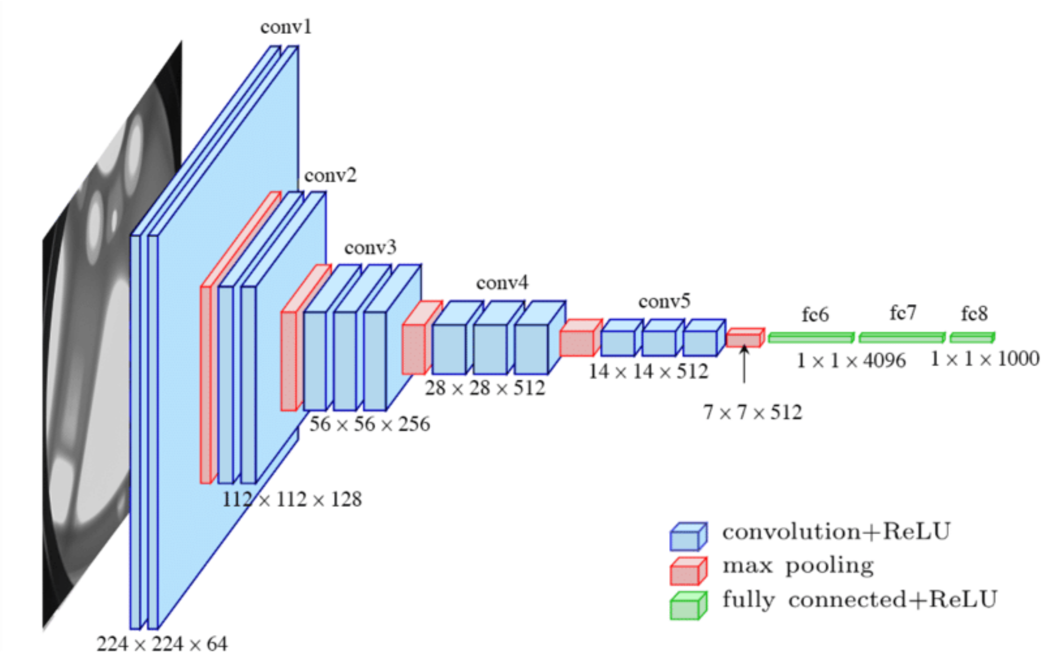
Table 1: **ConvNet configurations** (shown in columns). The depth of the configurations increases from the left (A) to the right (E), as more layers are added (the added layers are shown in bold). The convolutional layer parameters are denoted as "conv(receptive field size)-(number of channel)". The ReLU activation function is not shown for brevity.

ConvNet Configuration					
A	A-LRN	B	C	D	E
11 weight layers	11 weight layers	13 weight layers	16 weight layers	16 weight layers	19 weight layers
input (224 × 224 RGB image)					
conv3-64	conv3-64 **LRN**	conv3-64 **conv3-64**	conv3-64 conv3-64	conv3-64 conv3-64	conv3-64 conv3-64
maxpool					
conv3-128	conv3-128	conv3-128 **conv3-128**	conv3-128 conv3-128	conv3-128 conv3-128	conv3-128 conv3-128
maxpool					
conv3-256 conv3-256	conv3-256 conv3-256	conv3-256 conv3-256	conv3-256 conv3-256 **conv1-256**	conv3-256 conv3-256 **conv3-256**	conv3-256 conv3-256 conv3-256 **conv3-256**
maxpool					
conv3-512 conv3-512	conv3-512 conv3-512	conv3-512 conv3-512	conv3-512 conv3-512 **conv1-512**	conv3-512 conv3-512 **conv3-512**	conv3-512 conv3-512 conv3-512 **conv3-512**
maxpool					
conv3-512 conv3-512	conv3-512 conv3-512	conv3-512 conv3-512	conv3-512 conv3-512 **conv1-512**	conv3-512 conv3-512 **conv3-512**	conv3-512 conv3-512 conv3-512 **conv3-512**
maxpool					
FC-4096					
FC-4096					
FC-1000					
soft-max					

Table 2: **Number of parameters** (in millions).

Network	A,A-LRN	B	C	D	E
Number of parameters	133	133	134	138	144

1. Structure of VGG during training:
(1) the input to our ConvNets is a fixed-size 224 × 224 RGB image.
(2) subtracting the mean RGB value, computed on the training set, from each pixel.
(3) passed through a stack of convolutional (conv.) layers, where we use filters with a very
small receptive field: 3 × 3 (which is the smallest size to capture the notion of left/right, up/down,
center). Kernel is increasing 64 – 128 – 256 -512.
(4) Spatial pooling is carried out by five max-pooling layers, which follow some of the conv. layers (not all the conv. layers are followed by max-pooling). Max-pooling is performed over a 2 × 2 pixel window, with stride 2.



2. CLASSIFICATION EXPERIMENTS:

Dataset: The dataset includes images of 1000 classes, and is split into three sets: training (1.3M images), validation (50K images), and testing (100K images with held-out class labels).

- Performance evaluation (multiscale evalution):

(1) The classification performance is evaluated using two measures: the top-1 and top-5 error.
Top-1 classification error: the proportion of incorrectly classified images.
Top-5 classification error: the proportion of images such that the ground-truth category is outside the top-5 predicted categories. ILSVRC evaluation.

(2) running a model over several rescaled versions of a test image (corresponding to different values of Q), followed by averaging the resulting class posteriors.

(3) the models trained with fixed S were evaluated over three test image sizes, close to the training one: Q = {S − 32, S, S + 32}.

(4) At the same time, scale jittering at training time allows the network to be applied to a wider range of scales at test time, so the model trained with variable S ∈ [Smin; Smax] was evaluated over a larger range of sizes Q = {Smin, 0.5(Smin + Smax), Smax}.

Table 4: **ConvNet performance at multiple test scales.**

ConvNet config. (Table 1)	smallest image side		top-1 val. error (%)	top-5 val. error (%)
	train (S)	test (Q)		
B	256	224,256,288	28.2	9.6
C	256	224,256,288	27.7	9.2
	384	352,384,416	27.8	9.2
	[256; 512]	256,384,512	26.3	8.2
D	256	224,256,288	26.6	8.6
	384	352,384,416	26.5	8.6
	[256; 512]	256,384,512	**24.8**	**7.5**
E	256	224,256,288	26.9	8.7
	384	352,384,416	26.7	8.6
	[256; 512]	256,384,512	**24.8**	**7.5**