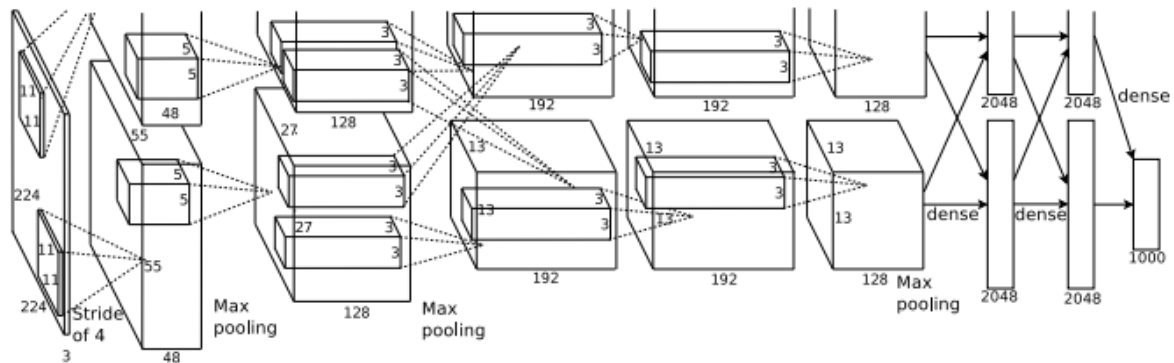


AlexNet (2012)

Pada tahun 2012, AlexNet secara signifikan mengungguli semua pesaing sebelumnya dan memenangkan tantangan dengan mengurangi kesalahan 5 besar dari 26% menjadi 15,3%. Tingkat kesalahan 5 teratas tempat kedua, yang bukan variasi CNN, adalah sekitar 26,2%.



Jaringan memiliki arsitektur yang sangat mirip dengan LeNet oleh Yann LeCun et al tetapi lebih dalam, dengan lebih banyak filter per lapisan, dan dengan lapisan konvolusional bertumpuk. Terdiri dari 11x11, 5x5, 3x3, convolutions, max pooling, dropout, data augmentation, aktivasi ULT, SGD dengan momentum. Itu melampirkan aktivasi ReLU setelah setiap lapisan konvolusional dan terhubung sepenuhnya. AlexNet dilatih selama 6 hari secara bersamaan pada dua GPU Nvidia Geforce GTX 580 yang menjadi alasan mengapa jaringan mereka dibagi menjadi dua pipeline. AlexNet dirancang oleh grup SuperVision, yang terdiri dari Alex Krizhevsky, Geoffrey Hinton, dan Ilya Sutskever.

GoogLeNet / Inception (2014)

Pemenang kompetisi ILSVRC 2014 adalah GoogLeNet (alias Inception V1) dari Google. Ini mencapai tingkat kesalahan 5 teratas sebesar 6,67%! Ini sangat mirip dengan kinerja tingkat manusia yang sekarang dipaksa untuk dievaluasi oleh penyelenggara tantangan. Ternyata, ini sebenarnya agak sulit dilakukan dan membutuhkan beberapa pelatihan manusia untuk mengalahkan akurasi GoogLeNets. Setelah beberapa hari pelatihan, ahli manusia (Andrej Karpathy) mampu mencapai tingkat kesalahan 5 teratas yaitu 5,1% (model tunggal) dan 3,6% (ansambel). Jaringan menggunakan CNN yang terinspirasi oleh LeNet tetapi menerapkan elemen baru yang disebut modul awal. Ini menggunakan normalisasi batch, distorsi gambar dan RMSprop. Modul ini didasarkan pada beberapa konvolusi yang sangat kecil untuk mengurangi jumlah parameter secara drastis. Arsitektur mereka terdiri dari CNN 22 lapisan tetapi mengurangi jumlah parameter dari 60 juta (AlexNet) menjadi 4 juta.

VGGNet (2014)

Runner-up kompetisi ILSVRC 2014 ini dijuluki VGGNet oleh komunitas dan dikembangkan oleh Simonyan dan Zisserman. VGGNet terdiri dari 16 lapisan konvolusional dan sangat menarik karena arsitekturnya yang sangat seragam. Mirip dengan AlexNet, hanya konvolusi 3x3, tetapi banyak filter. Dilatih dengan 4 GPU selama 2-3 minggu. Saat ini merupakan pilihan paling disukai di komunitas untuk mengekstraksi fitur dari gambar. Konfigurasi bobot VGGNet tersedia untuk

umum dan telah digunakan di banyak aplikasi dan tantangan lain sebagai ekstraktor fitur dasar. Namun, VGGNet terdiri dari 138 juta parameter, yang mungkin sedikit menantang untuk ditangani.