

Chapter 14: Summary

Diedrick Darrell Darmadi - 1103223031

Bab 14 membahas secara komprehensif bagaimana jaringan saraf konvolusional (CNN) bekerja serta mengapa arsitektur ini menjadi fondasi utama dalam bidang computer vision modern. Bab ini membuka dengan penjelasan mengapa tugas pengenalan visual yang tampak sederhana bagi manusia—seperti melihat anak anjing pada foto—ternyata sangat sulit untuk komputer. Hal ini karena proses persepsi manusia terjadi secara otomatis dalam modul visual otak, khususnya pada bagian visual cortex. Berdasarkan penelitian Hubel dan Wiesel, ditemukan bahwa neuron pada visual cortex memiliki receptive field kecil dan sensitif terhadap pola tertentu seperti garis horizontal atau vertikal. Temuan inilah yang menginspirasi desain arsitektur CNN, yang memproses gambar melalui lapisan-lapisan yang menangkap fitur sederhana lalu menyusunnya menjadi fitur kompleks.

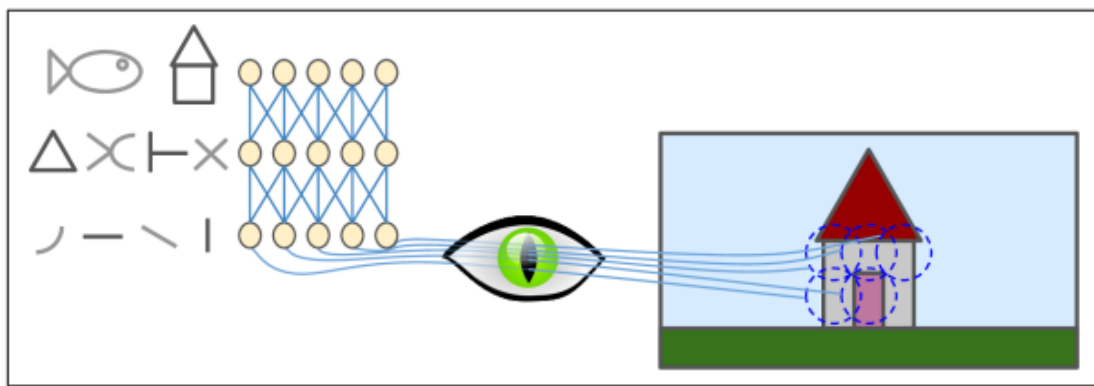


Figure 14-1. Local receptive fields in the visual cortex

Bagian berikutnya menjelaskan konsep fundamental dari **convolutional layer**, yaitu lapisan yang menggunakan receptive field lokal sehingga tiap neuron hanya terhubung dengan sebagian kecil piksel input, bukan seluruh gambar. Teknik **weight sharing** membuat semua neuron pada satu feature map menggunakan kernel (filter) yang sama, sehingga model lebih efisien dan mampu mengenali pola yang sama di lokasi berbeda. Bab ini menjelaskan detail tentang stride, padding (seperti *SAME* dan *VALID*), ukuran kernel, serta bagaimana stacking feature maps memungkinkan CNN mengekstraksi pola kompleks secara bertingkat. Konsep ini menjadi inti kekuatan CNN dalam memproses visual

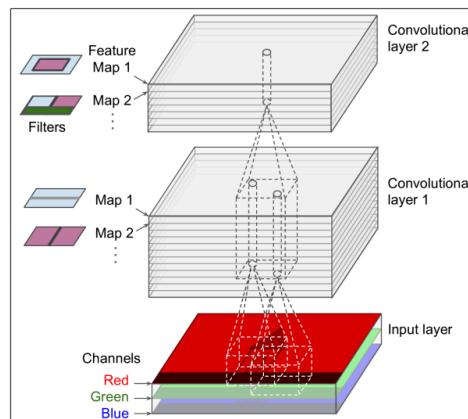


Figure 14-6. Convolution layers with multiple feature maps, and images with three color channels

Setelah memaparkan convolution, bab ini menjelaskan **pooling layer**, terutama max pooling, yang berfungsi mereduksi dimensi spatial sambil mempertahankan fitur paling informatif. Operasi pooling membuat jaringan lebih tahan terhadap translasi kecil, sekaligus mengurangi beban komputasi. Kemudian ditunjukkan implementasi CNN sederhana menggunakan TensorFlow, termasuk bagaimana mengatur jumlah filter, ukuran kernel, padding, serta penggunaan Dense layer setelah flattening. Bagian ini juga mengulas kebutuhan memori CNN, karena feature map yang besar akan memakan banyak RAM saat pelatihan dan inferensi.

Bab ini kemudian masuk ke pembahasan **arsitektur CNN klasik hingga modern**, dimulai dari LeNet-5, arsitektur tahun 1998 yang menjadi dasar banyak model modern. LeNet-5 memperkenalkan kombinasi convolution, pooling, dan fully connected layer untuk pengenalan digit tulisan tangan. Setelah itu, dibahas **AlexNet**, pemenang ImageNet 2012, yang memicu revolusi deep learning. AlexNet memperkenalkan penggunaan ReLU, dropout, data augmentation, serta memanfaatkan GPU untuk pelatihan model besar. Model ini jauh lebih dalam dibanding LeNet, dan menjadi titik loncatan CNN modern.

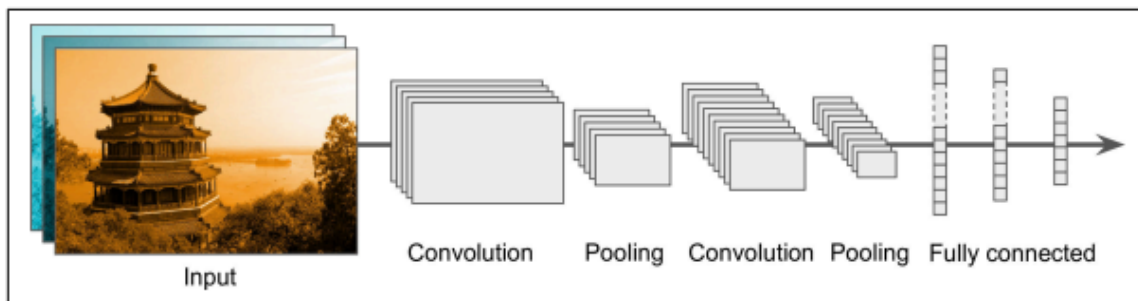


Figure 14-11. Typical CNN architecture