

Convolutional Neural Networks

Definition

Computer Vision este unul din domeniile Inteligenței Artificiale. Acest domeniu a apărut din dorința de a oferi dispozitivelor, în special calculatorului, posibilitatea de a observa lumea precum oamenii, de a o percepe într-o manieră similară și de a folosi un set de cunoștințe obținute pe baza unui antrenament pentru o gamă largă de sarcini (exemplu: clasificarea imaginilor, detecția facială, sistemele de recomandare, procesarea limbajului natural, etc.).

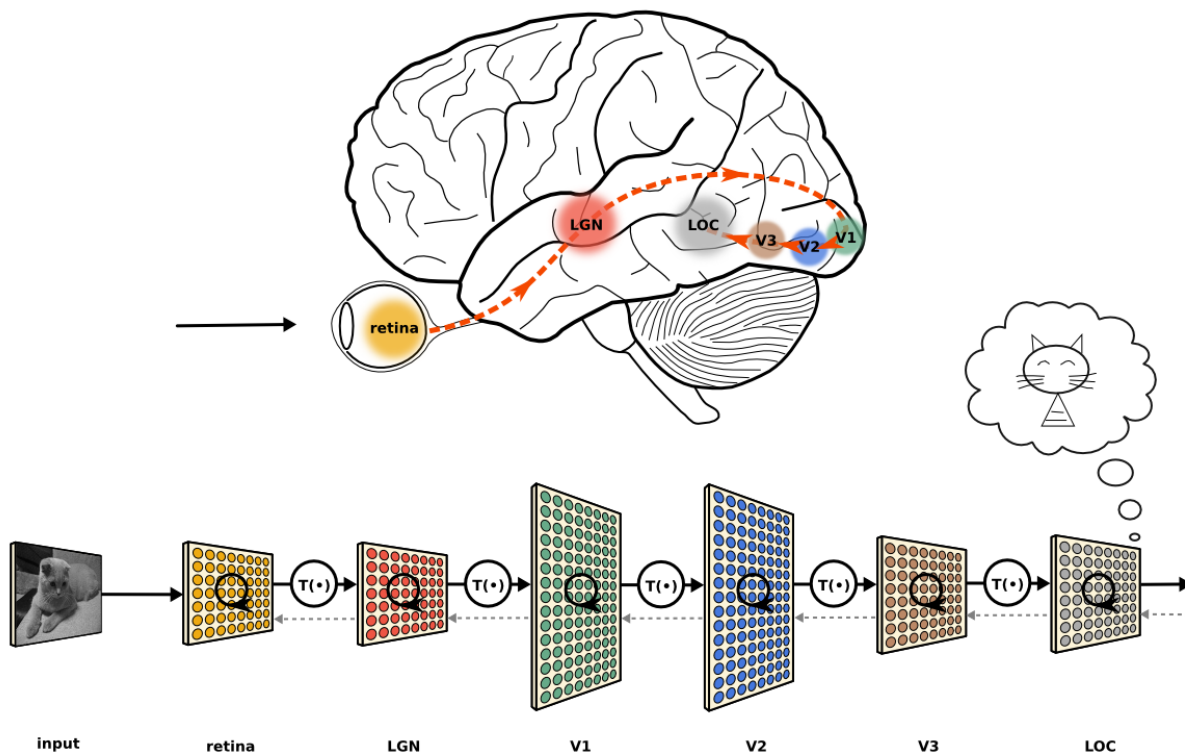


Figura 1: analog între procesul de recunoaștere uman al unei pisici dintr-o imagine și cel al unei rețele neuronale convoluționale.

Un **algorithm** este un set de pași bine definiți ce conferă calculatorului posibilitatea de a rezolva o problemă. Algoritmul este prezent în viața de zi cu zi a oamenilor; un exemplu sugestiv ar putea fi făcutul unei prăjituri unde ai niște instrucțiuni clare ce trebuie urmate.

O **rețea neuronală convoluțională** (CNN or ConvNet) este un algoritm ce primește o imagine și pe baza informațiilor acesteia, atribuie importanță diferitelor obiecte din imagine (weight-urilor și bias-urilor).

Arhitectura rețelei se aseamănă cu modelul de conectivitate al neuronilor din creierul uman și au fost inspirate de organizarea Cortexului Vizual (Figura 1).

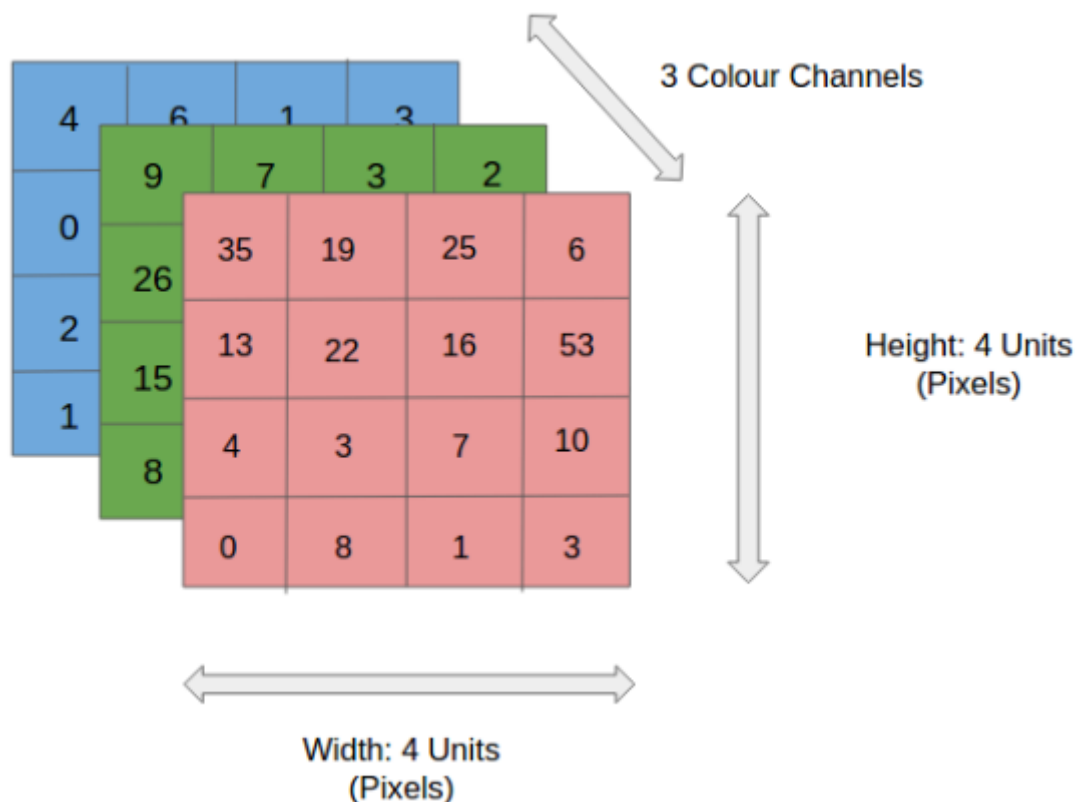


Figura 2: imagine care a fost separată pe cele 3 planuri de culori (roșu, verde și albastru)

În cazul unei imagini cu o calitate foarte mare, cum ar fi 8K (7680 X 4320), forța de calcul ar crește și ea. Rolul CNN-ului este de a reduce dimensiunea imaginilor la una mai simplă de procesat, fără a se pierde din calitatea și caracteristicile acesteia.

Kernel

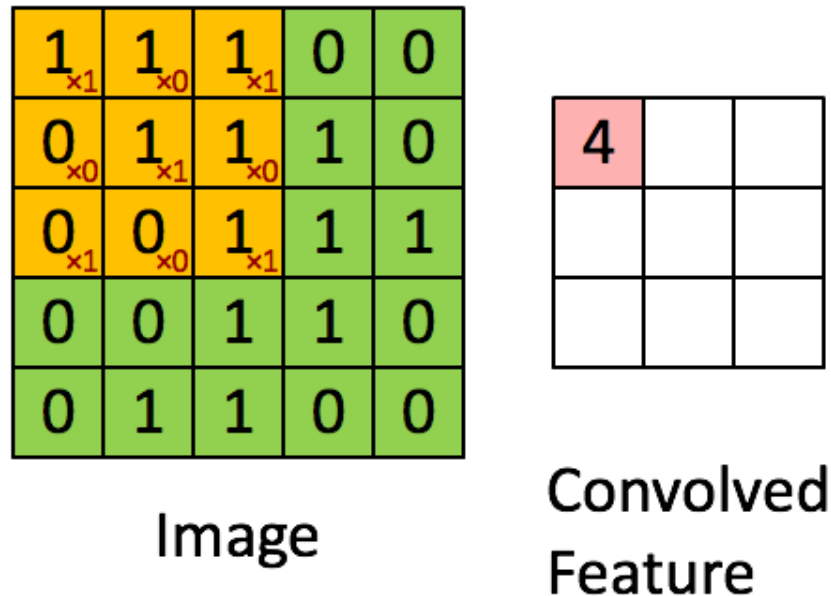


Figura 3: (dublu click pentru animație)

În Figura 3, imaginea de input este matricea verde, de dimensiune 5x5x1. Elementul implicat în procesul de convoluție se numește **kernel** sau **filtru.**, reprezentat cu galben și scris roșu. În acest exemplu, kernelul este o matrice de 3x3x1.

Acest kernel se deplasează spre dreapta pe imagine cu un pas, numit **Stride Length** (lungimea pasului), până analizează întreaga lățime. Din cele 9 căsuțe peste care trece, obține una singură. Ajuns la capătul lățimii imaginii, se deplasează cu o căsuță în jos. Pașii se repetă până e traversată imaginea în totalitate. Astfel dintr-o matrice de 5x5x1, s-a obținut o matrice de 3x3x1. Este ca și cum trece peste căsuțe și le împăturește în ceva mai mic.

Fiecare strat convoluțional are un număr de filtre. Filtrele detectează pattern-urile.

Privire în ansamblu

Ne putem gândi la o rețea neuronală convoluțională ca la o rețea specializată în detecția de tipare și să extragă informații despre acestea. Această detecție face rețelele atât de folositoare în analiza datelor.

Ceea ce o diferențiază de un perceptron cu mai multe straturi sunt layerurile hidden numite layeruri convoluționale. Un CNN poate avea și layeruri nonconvoluționale, dar la baza unei rețele convoluționale sunt convoluțiile.

O convoluție primește un input, o informație, îl transformă într-un mod anume și îl pasează următorului layer de convoluții (Figura 4).

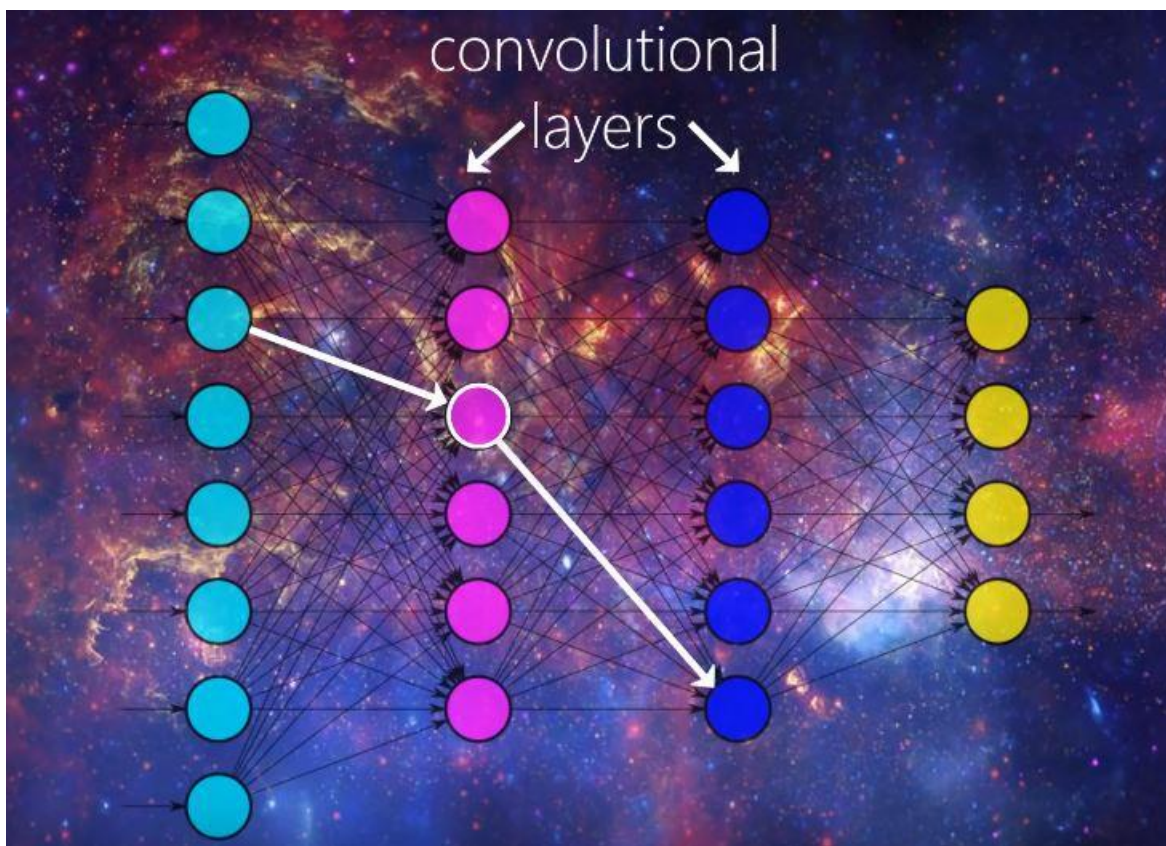


Figura 4: procesul de primire de input, procesare și pasare ca output următorului strat

Din punctul de vedere al șabloanelor, o imagine dispune de o multitudine de șabloane: muchii (linii, colțuri), forme (cercuri, triunghiuri), texturi, forme, etc. Deci dacă un filtru este specializat în detecția de muchii, acesta va avea denumire de edge detector.

Cu cât rețeaua este mai mare în adâncime, cu atât poate face o detecție mai fină asupra lucrurilor, precum ochi, pene, anumite obiecte, etc. Aceste layeruri sunt spre

finalul rețelei, față de layerele cu filtre ce detectează forme mai simple care se află la început. Mai multe filtre înseamnă detecții mai complexe, ca și pisici, câini, oameni (Figura 5).



Figura 5: a convolutional neural network's dream

În Figura 6 este reprezentată trecerea unei imagini printr-un strat convoluțional, de la input, procesare și obținerea outputului.

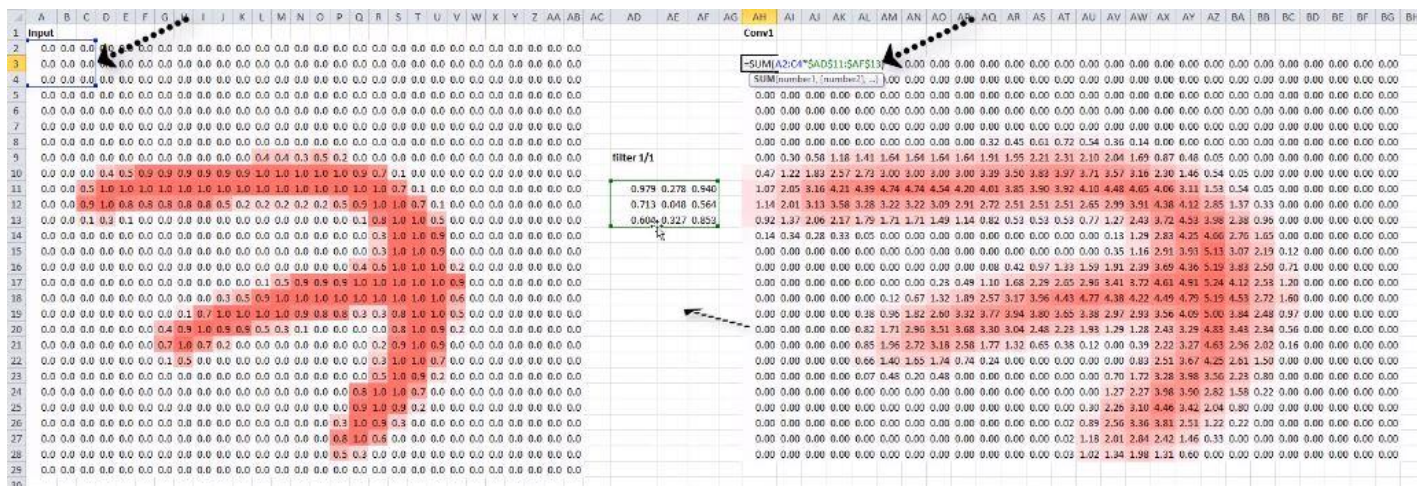


Figura 6: procesarea unei imagini de la un strat X, la stratul X+1

După trecerea filtrului prin întregul input, va rămâne o nouă reprezentare ce va fi transmisă stratului următor ca input (imaginea din dreapta). Filtrele pot fi văzute ca

un detector de modele (Figura 7). Cu fiecare strat mai adânc, sunt detectate forme mai complexe (Figura 8).

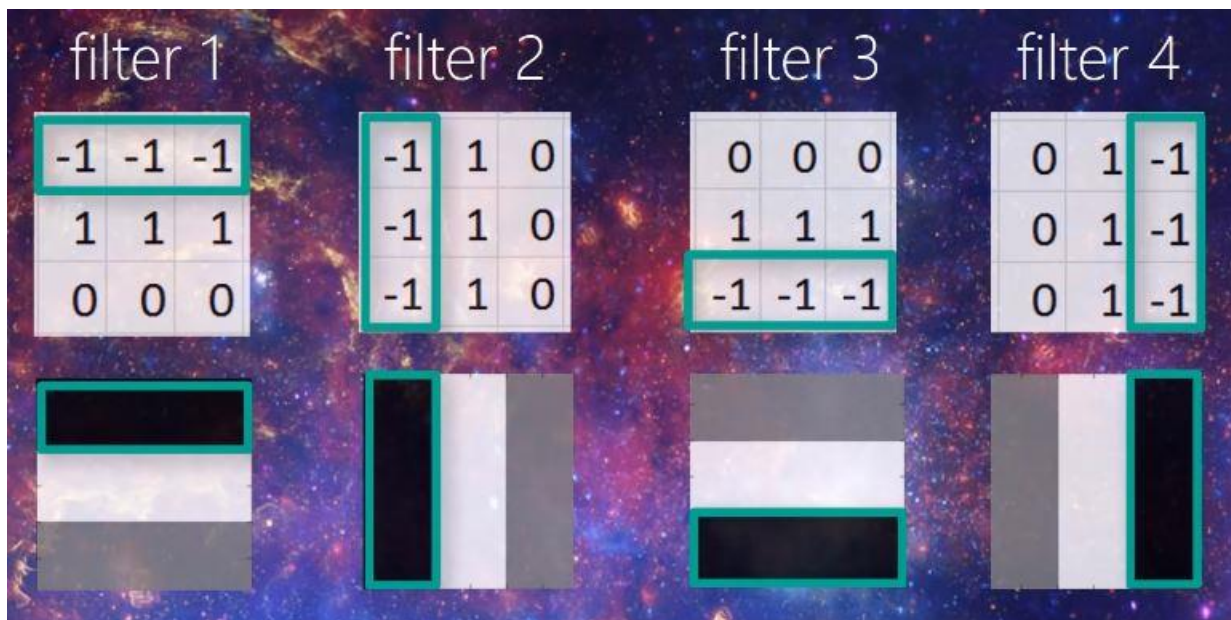


Figura 7: filtrele determină anumite patterns



Figure 8: detecția câinilor din imagine