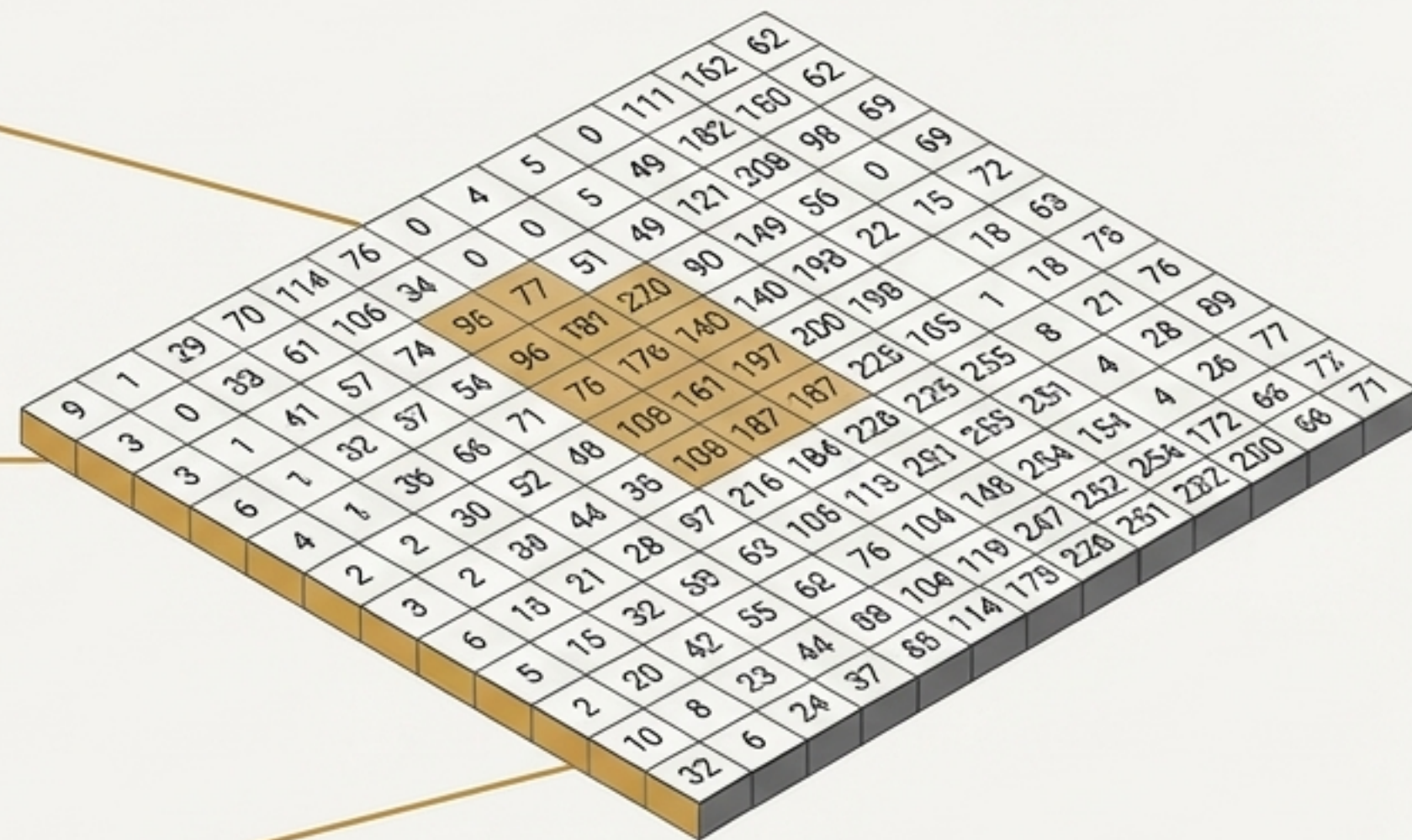


Kako mašina uči da vidi?



Za nas, ovo je slika mačke.

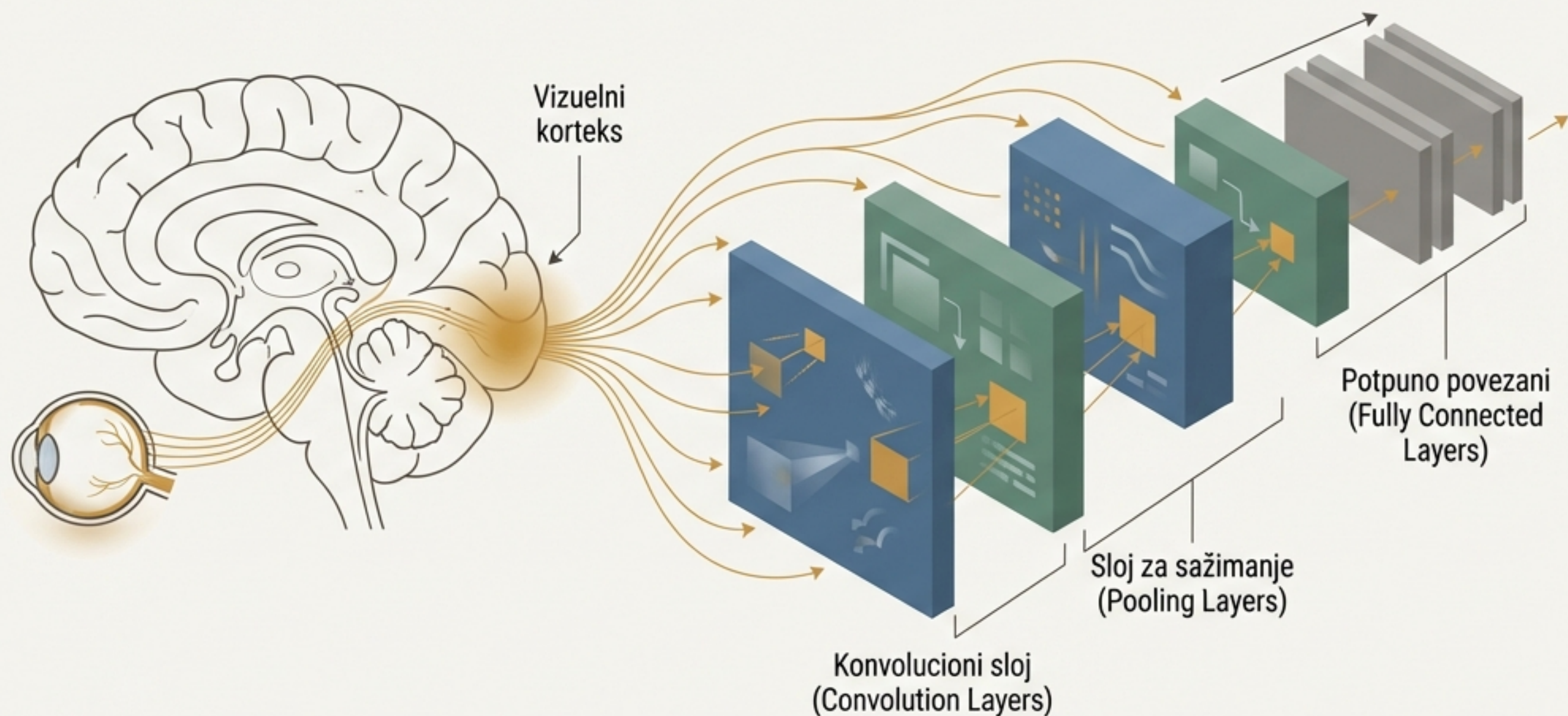


Za računar, to je samo mreža brojeva. Svaki broj predstavlja intenzitet jednog piksela, od 0 (crna) do 255 (bela). Bez konteksta, ova matrica je besmislena.

Kako premostiti jaz između sirovih piksela i smislenog razumevanja? Kako učimo mašinu da prepozna mačku unutar brojeva?

Konvolucione neuronske mreže: Digitalno oko inspirisano mozgom

Konvolucione neuronske mreže (CNN) su specijalizovana vrsta neuronskih mreža dizajnirana za obradu podataka u obliku mreže, kao što su slike. Njihova arhitektura je inspirisana vizuelnim korteksom mozga, gde neuroni reaguju na male pod-regione vidnog polja.

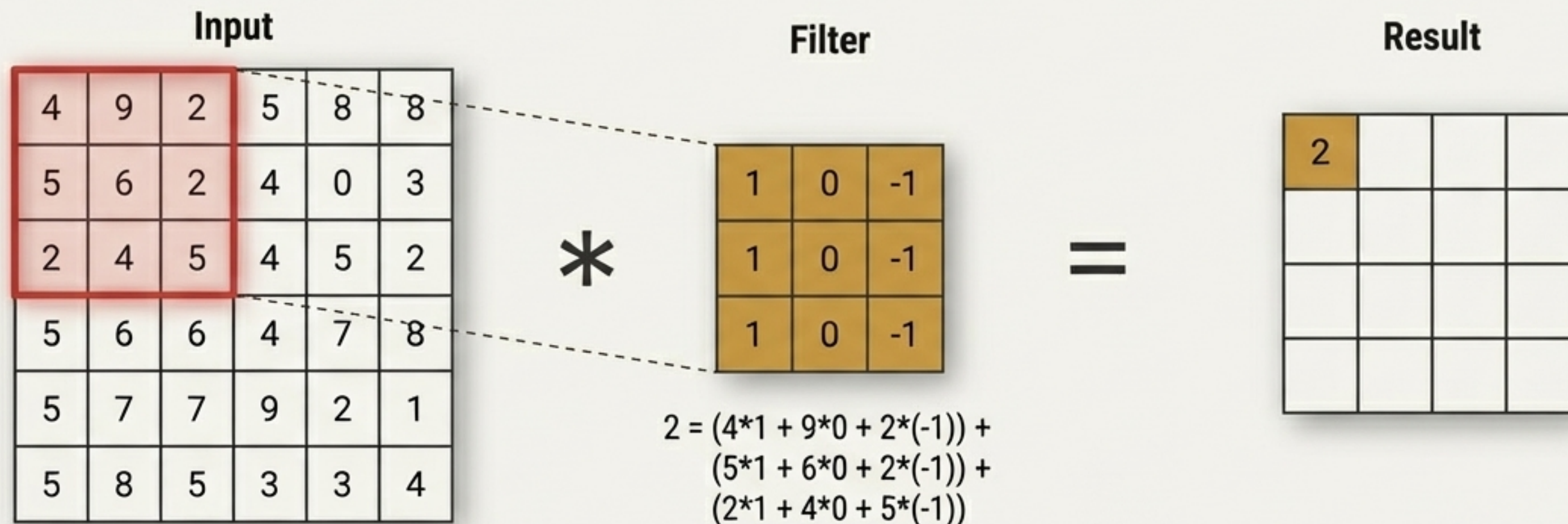


Ključne karakteristike

- **Specijalizovane za vizuelne podatke:** Primarno se koriste za rad sa slikama i video-sadržajima.
- **Automatska ekstrakcija atributa:** Mreža sama uči da prepozna relevantne karakteristike (ivice, teksture, oblike) direktno iz piksela, eliminišući potrebu za ručnim definisanjem.
- **Revolucija u računarskom vidu:** Njihov razvoj je omogućio prodor u zadacima kao što su klasifikacija slika, detekcija objekata i segmentacija.

Konvolucija: Proces „skeniranja“ slike u potrazi za obrascima

Osnovna operacija u CNN-u je konvolucija. Ona uključuje pomeranje male matrice, poznate kao **filter** ili **kernel**, preko ulazne slike. Na svakoj poziciji, filter se preklopi sa delom slike, izvrši se množenje elemenata, a njihov zbir se upisuje u novu matricu koja se naziva **mapa karakteristika** (feature map).



Filter (Kernel)

Mala matrica težina koja je dizajnirana da detektuje specifičan obrazac (npr. vertikalnu ivicu). Vrednosti u filteru se uče tokom procesa treninga.

Pomeranje (Stride)

Hiper-parametar koji definiše za koliko piksela se filter pomera u svakom koraku.

Rezultat

Mapa karakteristika koja naglašava delove slike gde je detektovan obrazac za koji je filter zadužen.

Šta filteri zapravo „vide“? Izdvajanje osnovnih elemenata

Svaki filter je specijalizovan za detekciju jedne vrste karakteristike. Na primer, filter prikazan ovde je dizajniran da pronade vertikalne ivice.



Originalna slika

+1	0	-1
+1	0	-1
+1	0	-1

- **Leva kolona (pozitivne vrednosti):** Poklapa se sa svetlim pikselima.
- **Desna kolona (negativne vrednosti):** Poklapa se sa tamnim pikselima.
- **Srednja kolona (nule):** Ignoriše piksele u sredini.

Kada ovaj filter pređe preko vertikalne ivice (prelaz iz svetlog u tamno), rezultat konvolucije je velika pozitivna vrednost. Na drugim delovima slike, rezultat je blizu nule.

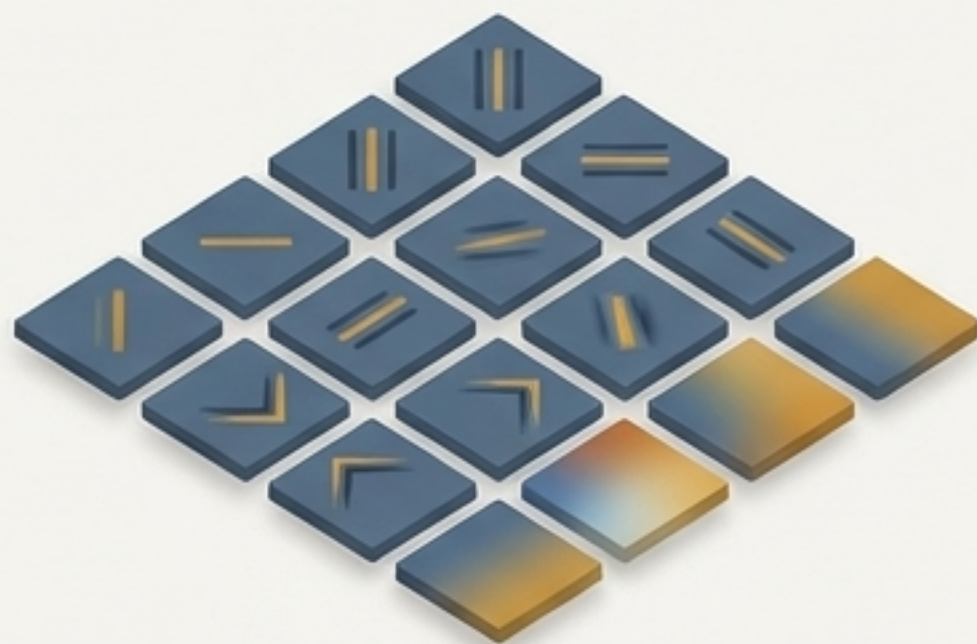
Rezultat: Istaknute vertikalne ivice



Primena ovog filtera na sliku mačke stvara novu sliku na kojoj su istaknute samo vertikalne konture.

Hijerarhija karakteristika: Od linija do lica

Moć CNN-a leži u njihovoj sposobnosti da grade **hijerarhiju karakteristika** kroz više slojeva. Svaki sledeći sloj uči složenije obrasce kombinujući karakteristike iz prethodnog sloja.



Sloj 1: Detekcija ivica, uglova i prelaza boja.



Sloj 2: Kombinovanje elemenata u delove objekata.

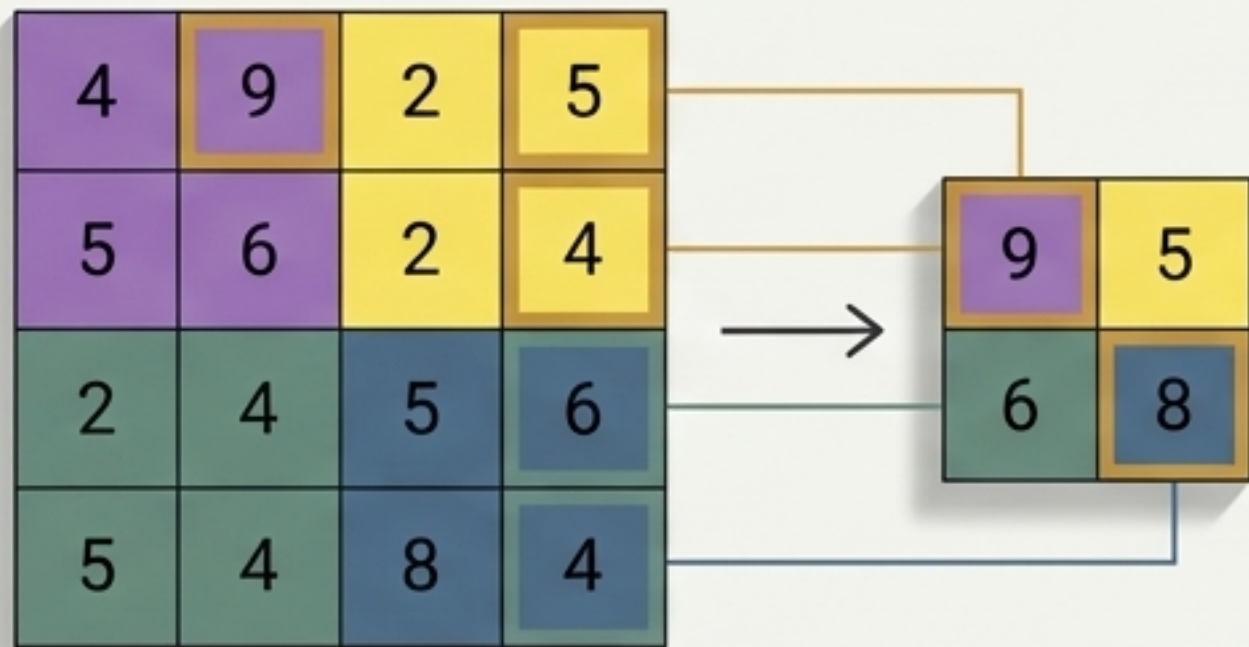


Sloj 3: Sastavljanje delova u prepoznatljive celine.

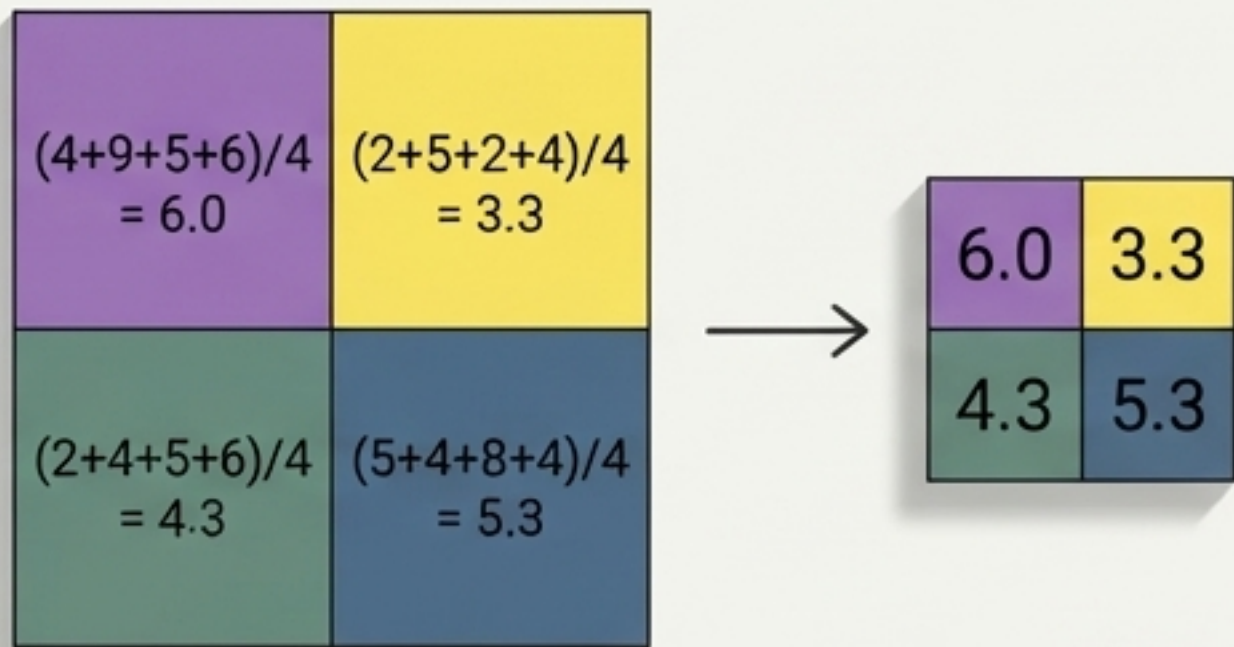
Ovaj proces omogućava mreži da nauči apstraktne i robusne reprezentacije ulaznih podataka.

Agregacija (Pooling): Sažimanje informacija i smanjenje kompleksnosti

Nakon konvolucionih slojeva, često se primenjuje operacija **agregacije** (pooling) sa ciljem smanjenja prostornih dimenzija mapa karakteristika. Ovo pomaže u **smanjenju broja parametara** i računskih troškova, a takođe čini model otpornijim na male translacije objekta na slici.



Max Pooling

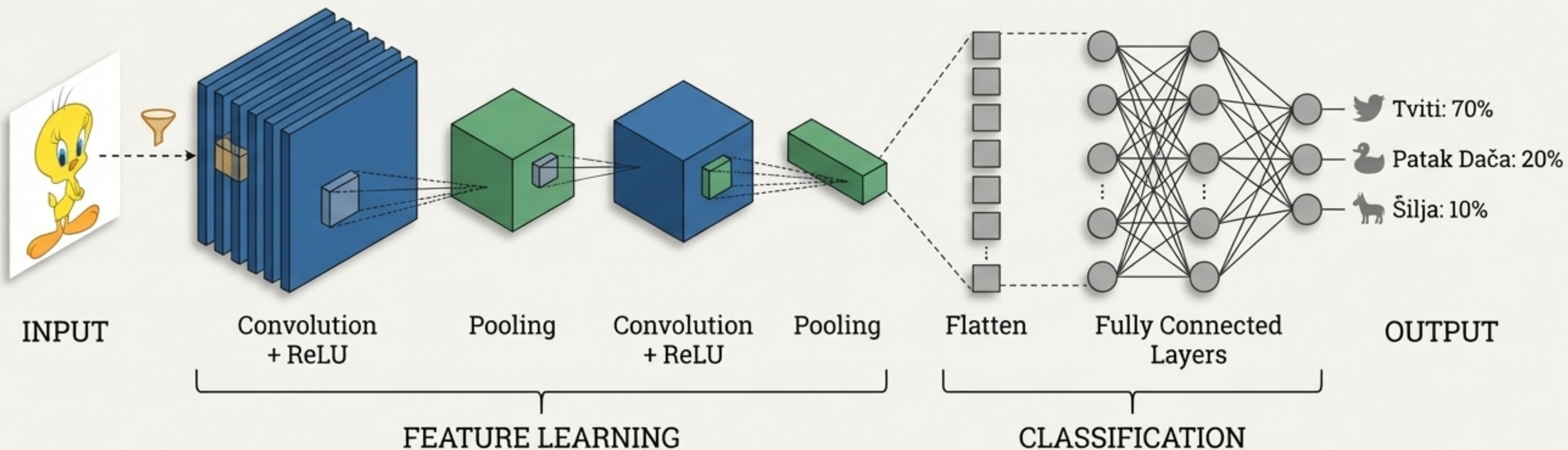


Average Pooling

Prednost 1: Smanjenje dimenzija	Prednost 2: Invarijantnost na translaciju	Prednost 3: Prevencija preprilagođavanja (overfitting)
Manje računarske zahtevnosti.	Male promene u poziciji karakteristika ne utiču značajno na izlaz.	Smanjuje kompleksnost modela.

Sklapanje arhitekture: Put podataka od piksela do predikcije

Tipična CNN arhitektura se sastoji od naizmeničnih blokova konvolucionih i agregacionih slojeva, koji zajedno čine deo za **učenje karakteristika**, nakon čega sledi potpuno povezana mreža za **klasifikaciju**.



Galerija Titana: Arhitekture koje su definisale moderni računarski vid

Nakon što smo razumeli osnovne gradivne blokove, upoznaćemo se sa najpoznatijim CNN arhitekturama. Svaka od njih je u svoje vreme predstavljala proboj, rešavajući ključne izazove i pomerajući granice mogućeg.

Ove arhitekture nisu samo nizovi slojeva; one predstavljaju različite filozofije dizajna koje su uvele fundamentalne inovacije:

- ✓ **ReLU** aktivacione funkcije
- **Dropout** regularizacija
- ▣ Duboke mreže sa malim filterima
- ↺ **Rezidualne** konekcije
- ↻ Paralelne konvolucione putanje

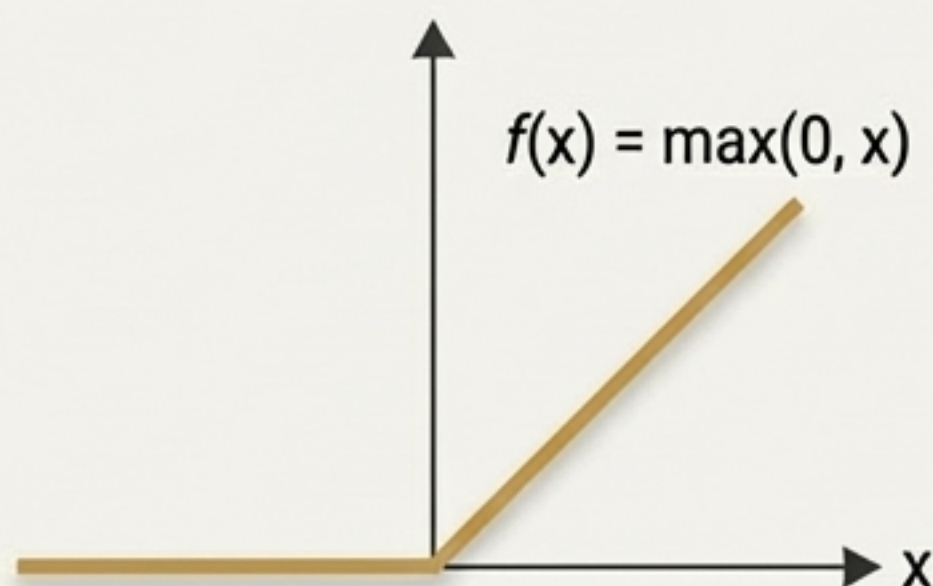
Upoznaćemo **AlexNet**, **VGGNet**, **ResNet** i **Inception (GoogLeNet)**.

AlexNet (2012): Revolucija koja je pokrenula sve

Na takmičenju ImageNet 2012. godine, AlexNet je postigao stopu greške od 15.4%, što je bilo za preko 10% bolje od drugoplasiranog. Ovaj rezultat je bio prekretnica koja je dokazala superiornost dubokih CNN-ova i pokrenula revoluciju u veštačkoj inteligenciji.

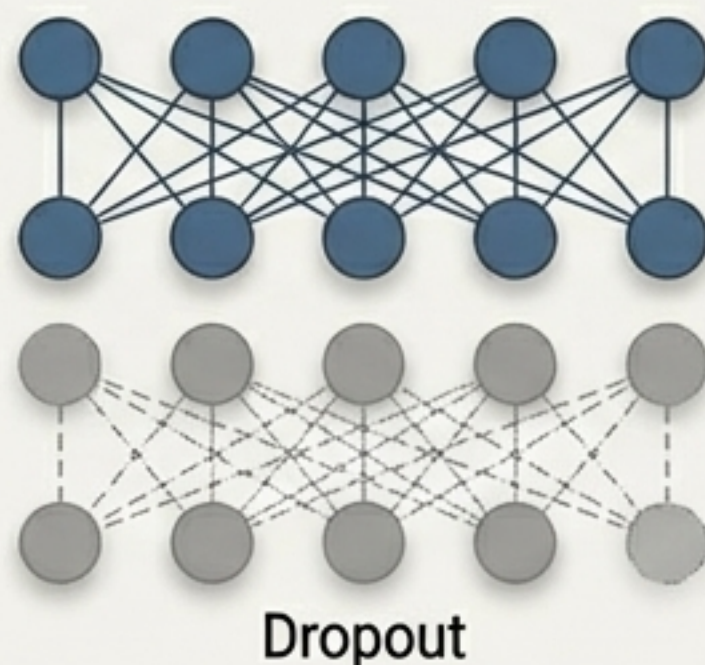
ReLU aktivaciona funkcija

Ubrzala je trening i smanjila problem nestajućeg gradijenta, zamenivši tradicionalne sigmoidne funkcije.



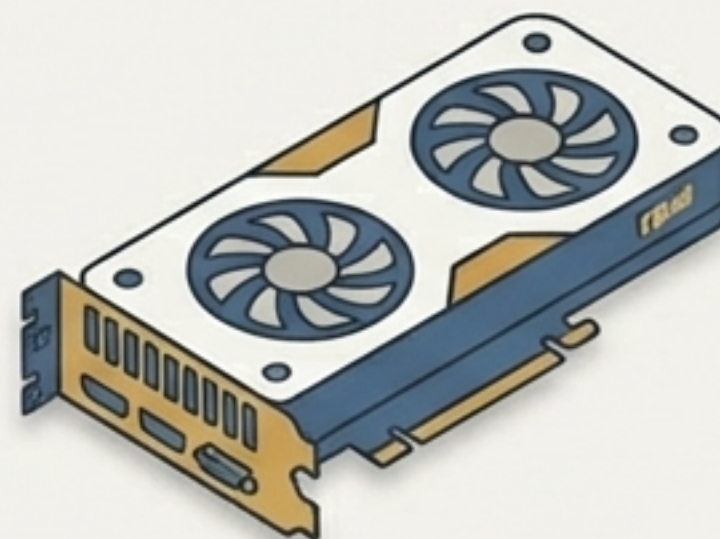
Dropout regularizacija

Tokom treninga, nasumično deaktivira neurone, čime se mreža primorava da uči robusnije karakteristike i sprečava preprilagođavanje.



Implementacija na GPU-u

Korišćenje NVIDIA GTX 580 GPU-a omogućilo je treniranje znatno većih i dubljih modela na ogromnom ImageNet skupu podataka, što do tada nije bilo izvodljivo.



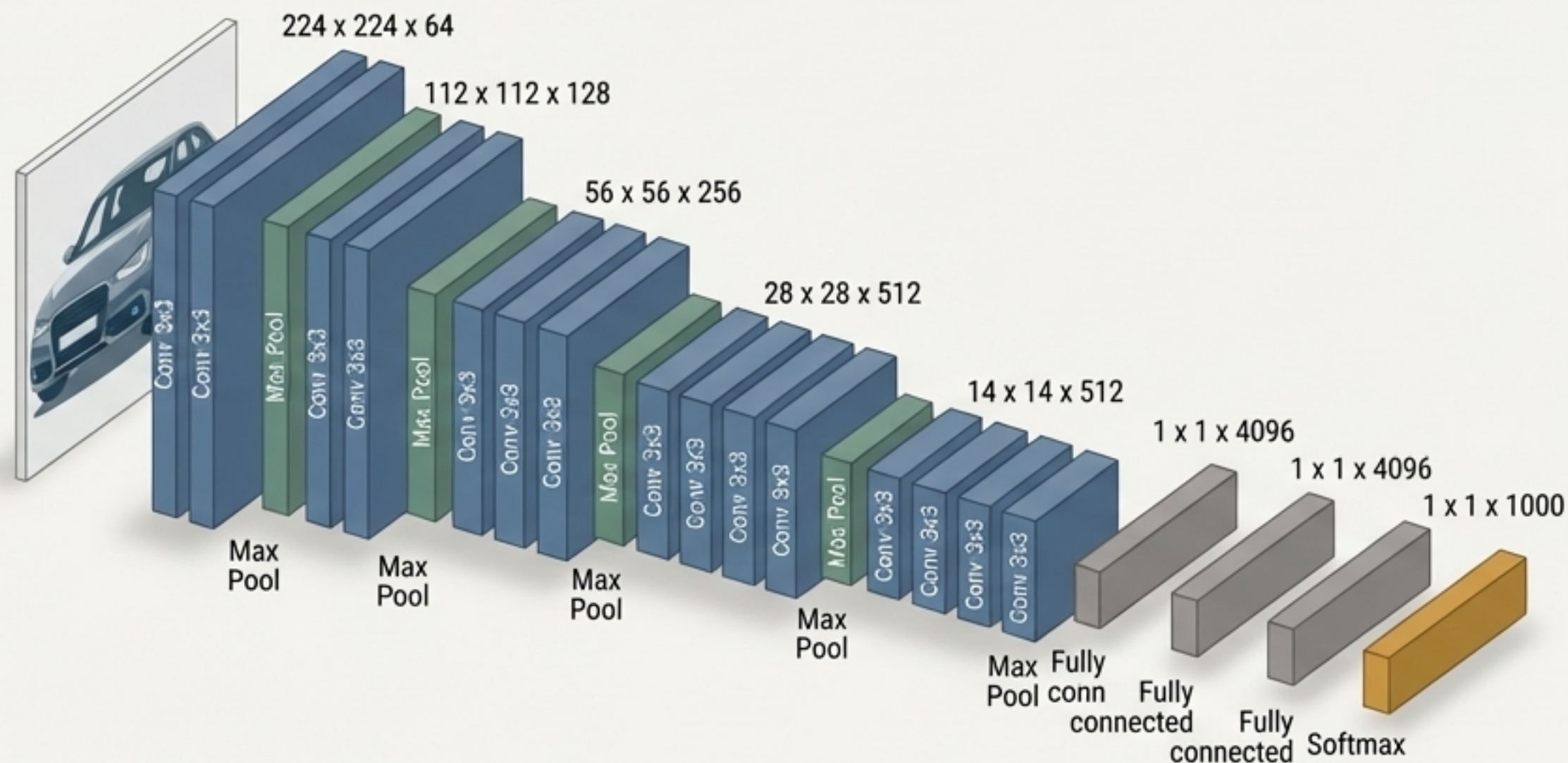
AlexNet je uspostavio obrazac za buduće CNN arhitekture i dokazao da je dubina mreže ključna za visoke performanse.

VGGNet: Arhitekta elegancije i dubine

Razvijen od strane Visual Geometry Group (VGG) sa Univerziteta u Oksfordu, VGGNet je pokazao da se izvanredni rezultati mogu postići korišćenjem veoma uniformne i duboke arhitekture.

Principi dizajna

- **Isključivo mali 3x3 filteri:** Umesto velikih filtera, VGG koristi nizove konvolucionih slojeva sa malim 3x3 filterima. Više uzastopnih 3x3 slojeva ima efektivno receptivno polje kao jedan veći filter, ali sa više nelinearnosti i manje parametara.
- **Dubina kao primarni faktor:** Sa varijantama od 16 (VGG-16) i 19 (VGG-19) slojeva, VGG je demonstrirao sirovu snagu povećanja dubine mreže.
- **Jednostavnost i ponavljanje:** Arhitektura se sastoji od ponavljajućih blokova (nekoliko konvolucionih slojeva praćenih max pooling slojem), što je čini jednostavnom za razumevanje i implementaciju.



Skala: VGG-16 mreža ima oko **138 miliona parametara**, što ilustruje računarsku zahtevnost ovih modela.

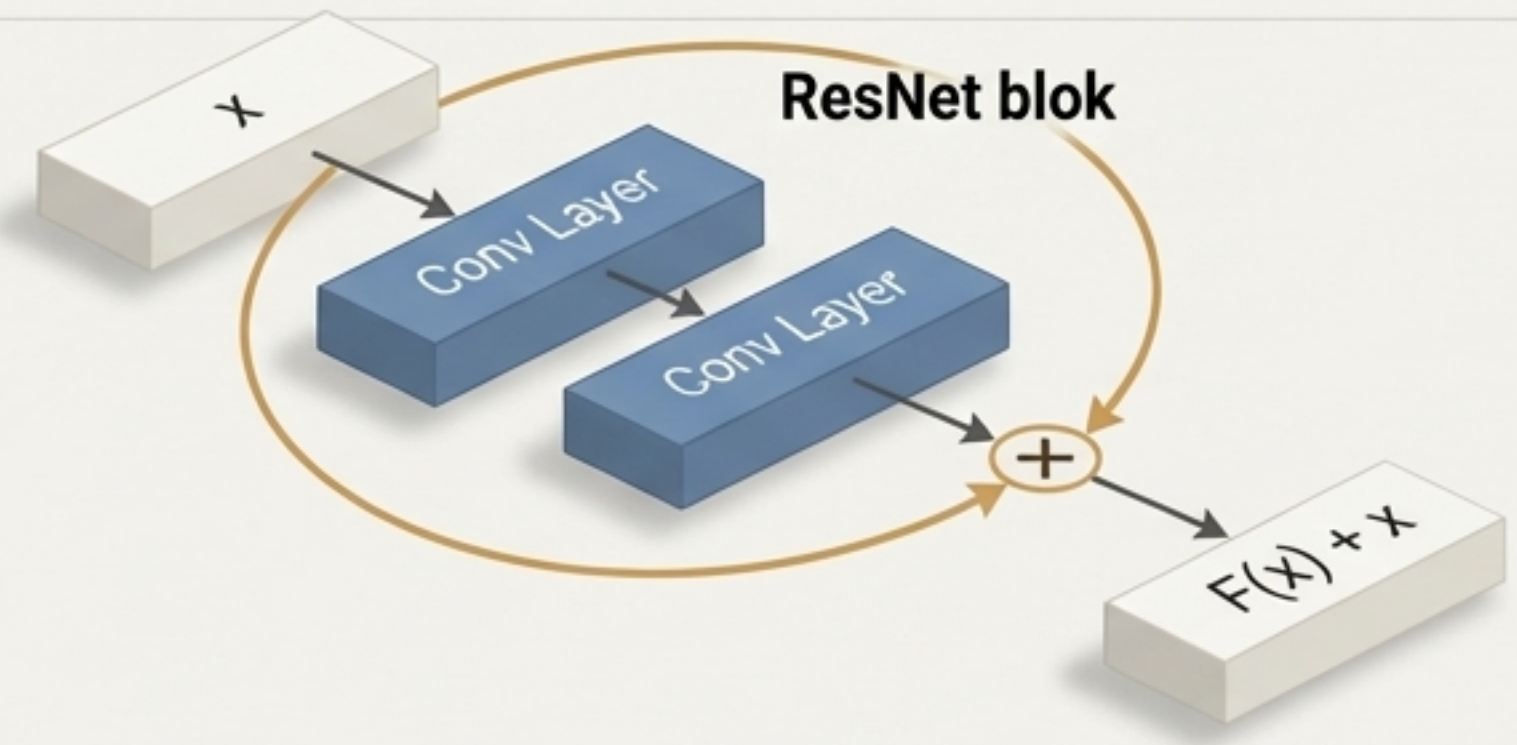
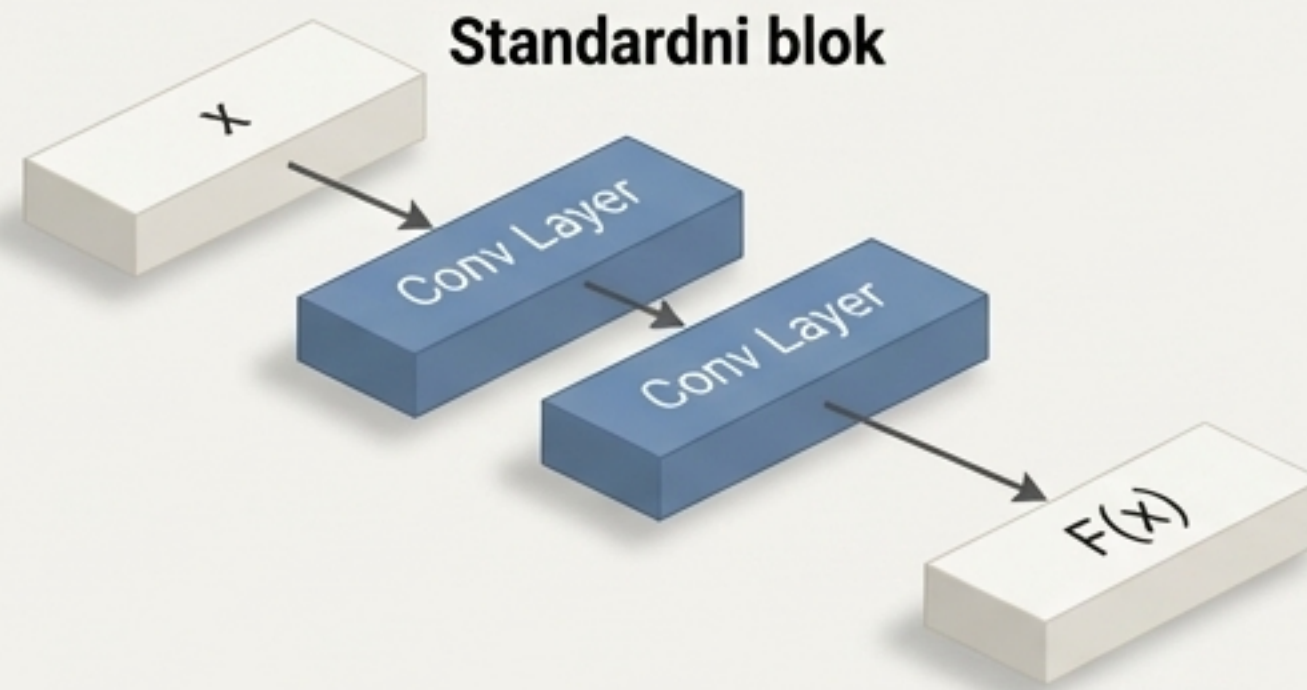
ResNet: Otključavanje nezamislive dubine

Problem

Kako mreže postaju sve dublje, javlja se problem **degradacije**: performanse počinju da opadaju jer je protok gradijenta kroz toliko slojeva otežan. Veoma duboke mreže postaju teške za treniranje.

Rešenje

ResNet uvodi revolucionarnu ideju „prečica“ ili **skip konekcija**. Umesto da slojevi uče direktno preslikavanje $F(x)$, oni uče **rezidualnu funkciju** $F(x) = y - x$. Izlaz bloka se definiše kao $y = F(x) + x$.



Kako pomaže?

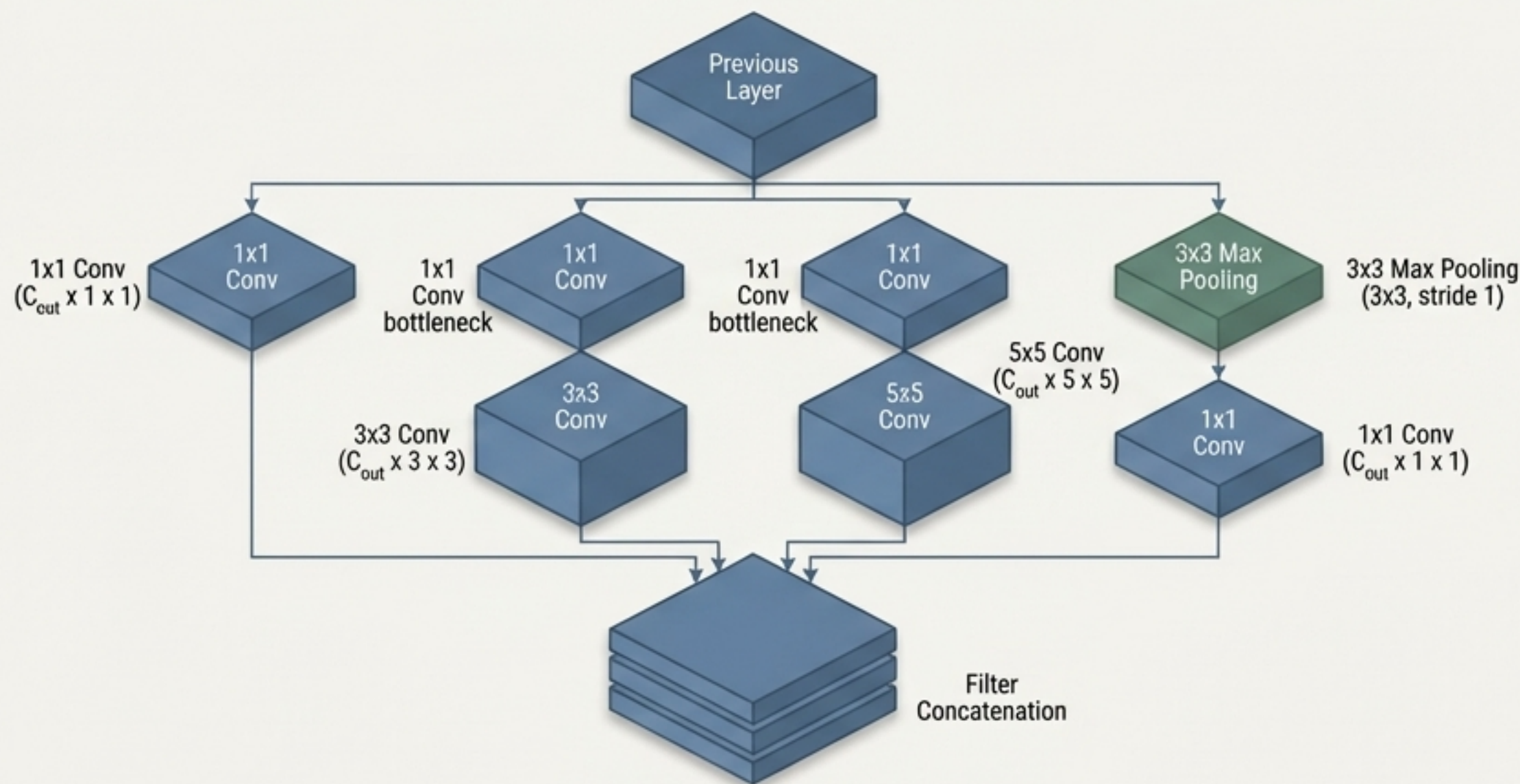
- **Lakša optimizacija:** Ove prečice omogućavaju gradijentu da teče direktno kroz mrežu, zaobilazeći nekoliko slojeva. Ovo drastično olakšava trening i rešava problem nestajućeg gradijenta.
- **Učenje identiteta:** Ako neki sloj nije koristan, mreža može lako naučiti da ga „preskoči“ tako što će $F(x)$ postaviti na nulu.

Uticaj

- **Uticaj:** ResNet je omogućio uspešno treniranje mreža sa 100, 152, pa čak i preko 1000 slojeva, postizujući performanse koje su po prvi put nadmašile ljudske na ImageNet klasifikaciji.

Inception/GoogLeNet: Efikasnost kroz paralelizam

Umesto da samo povećava broj slojeva, GoogLeNet je uveo **Inception modul**, arhitektonski blok koji ide „u širinu“. Ideja je da se na istom nivou mreže izvrši više različitih konvolucionih operacija paralelno.



Prednosti

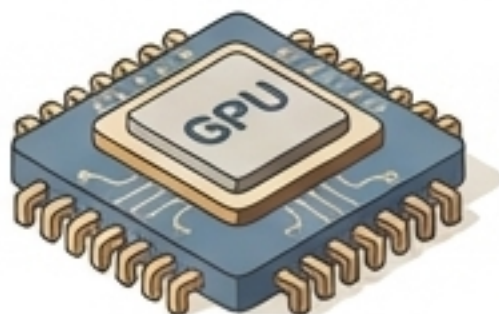
- **Detekcija karakteristika na više skala:** Mreža istovremeno uči da prepoznaje karakteristike različitih veličina.
- **Računarska efikasnost:** Upotreba **1x1 konvolucija** pre većih 3x3 i 5x5 filtera služi kao „bottleneck“ sloj koji smanjuje dubinu (broj kanala), drastično smanjujući broj proračuna.

Veoma duboka (22 sloja) i precizna mreža sa značajno manjim brojem parametara od VGGNet-a (oko 4 miliona).

Nasleđe inovacija koje je oblikovalo veštačku inteligenciju

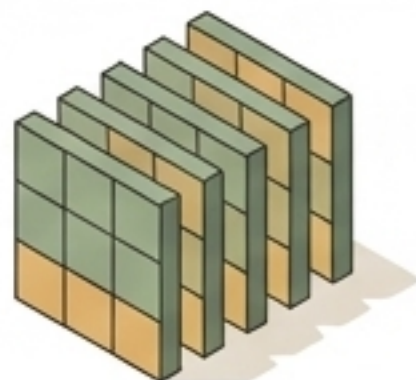
Svaka od ovih revolucionarnih arhitektura nije samo poboljšala preciznost, već je uvela fundamentalne koncepte koji su postali standard u dizajnu modernih neuronskih mreža.

AlexNet



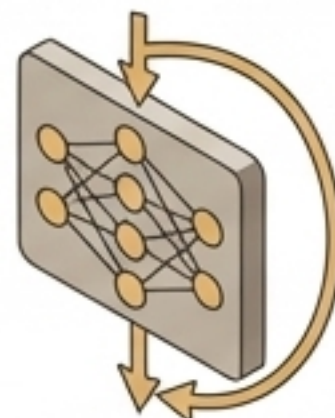
Dokazao je moć **dubine**, **GPU** ubrzanja i uveo **ReLU** i **Dropout** kao standard.

VGGNet



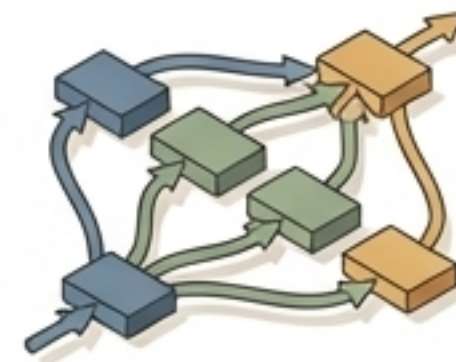
Demonstrirao je efikasnost **uniformnosti** i ekstremne dubine postignute korišćenjem isključivo malih **3x3 filtera**.

ResNet



Rešio je problem degradacije u veoma dubokim mrežama pomoću genijalnih **rezidualnih konekcija**, omogućavajući mrežama da pređu granicu od 100 slojeva.

Inception

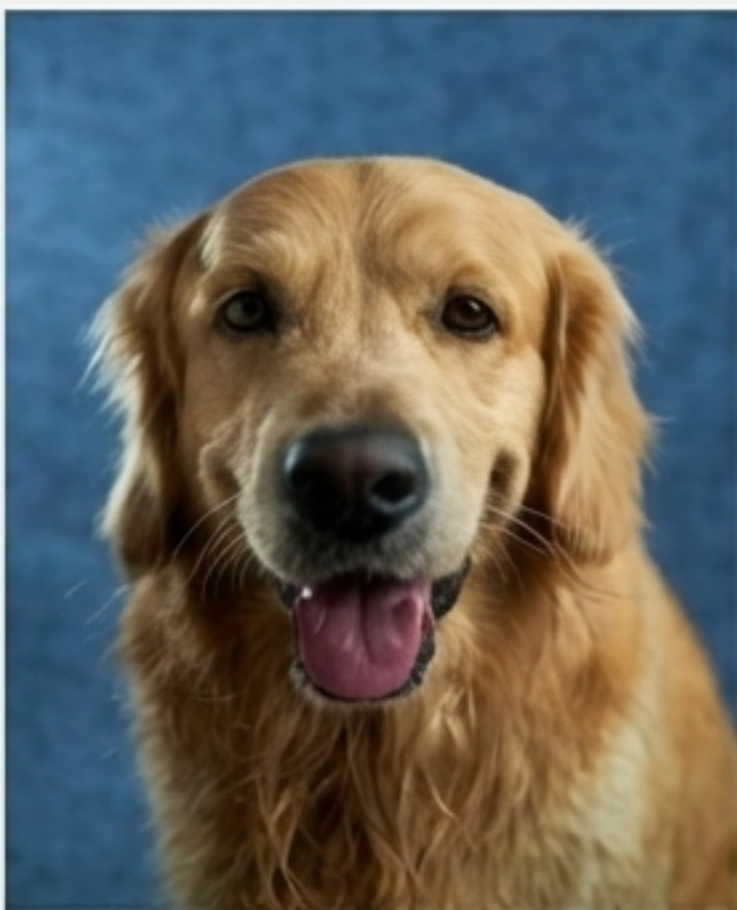


Uveo je koncept **paralelizma** i računarske efikasnosti kroz **Inception module** i pametnu upotrebu **1x1 konvolucija**.

Zajedno, ove inovacije su stvorile temelj na kojem počivaju današnji najnapredniji sistemi za računarski vid.

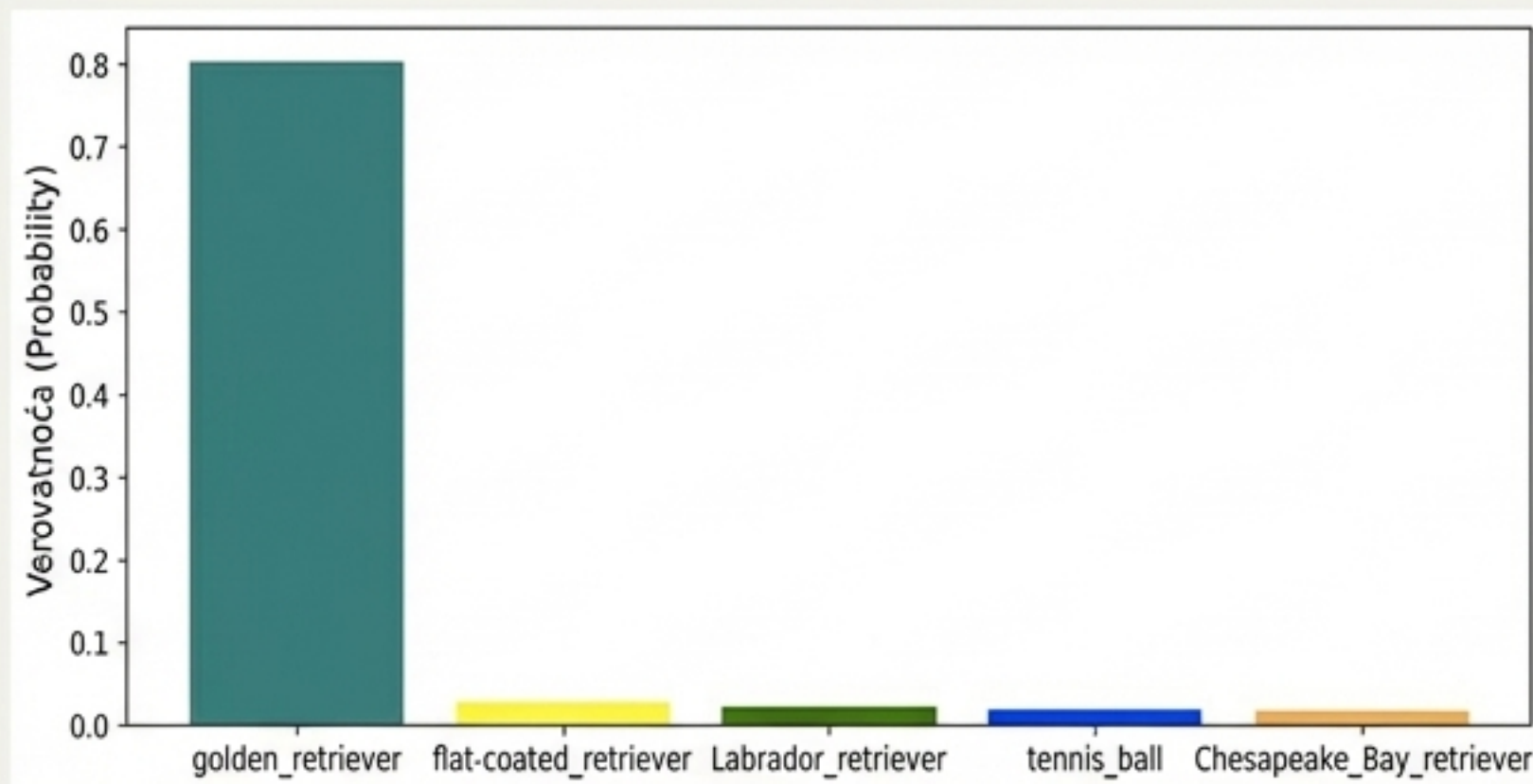
Od teorije do prakse: CNN na delu

Jednom istreniran, moćan model kao što je VGG-16 može se koristiti za klasifikaciju novih slika. Korišćenjem pre-treniranog modela sa težinama naučenim na ImageNet skupu podataka (koji sadrži 1000 klasa), možemo analizirati sliku zlatnog retrivera.



1. **Priprema slike:** Slika se skalira na dimenzije koje model očekuje (224x224 piksela) i normalizuje.
2. **Predikcija:** Slika se prosleđuje kroz VGG-16 mrežu.
3. **Rezultat:** Model na izlazu daje verovatnoće pripadnosti svakoj od 1000 poznatih klasa.

Analiza izlaza



Kao što vidimo, model je sa verovatnoćom od ~80% tačno identifikovao sliku kao zlatnog retrivera. Zanimljivo je da su druge najverovatnije klase takođe vrste retrivera, što pokazuje sposobnost modela da prepozna fine razlike.