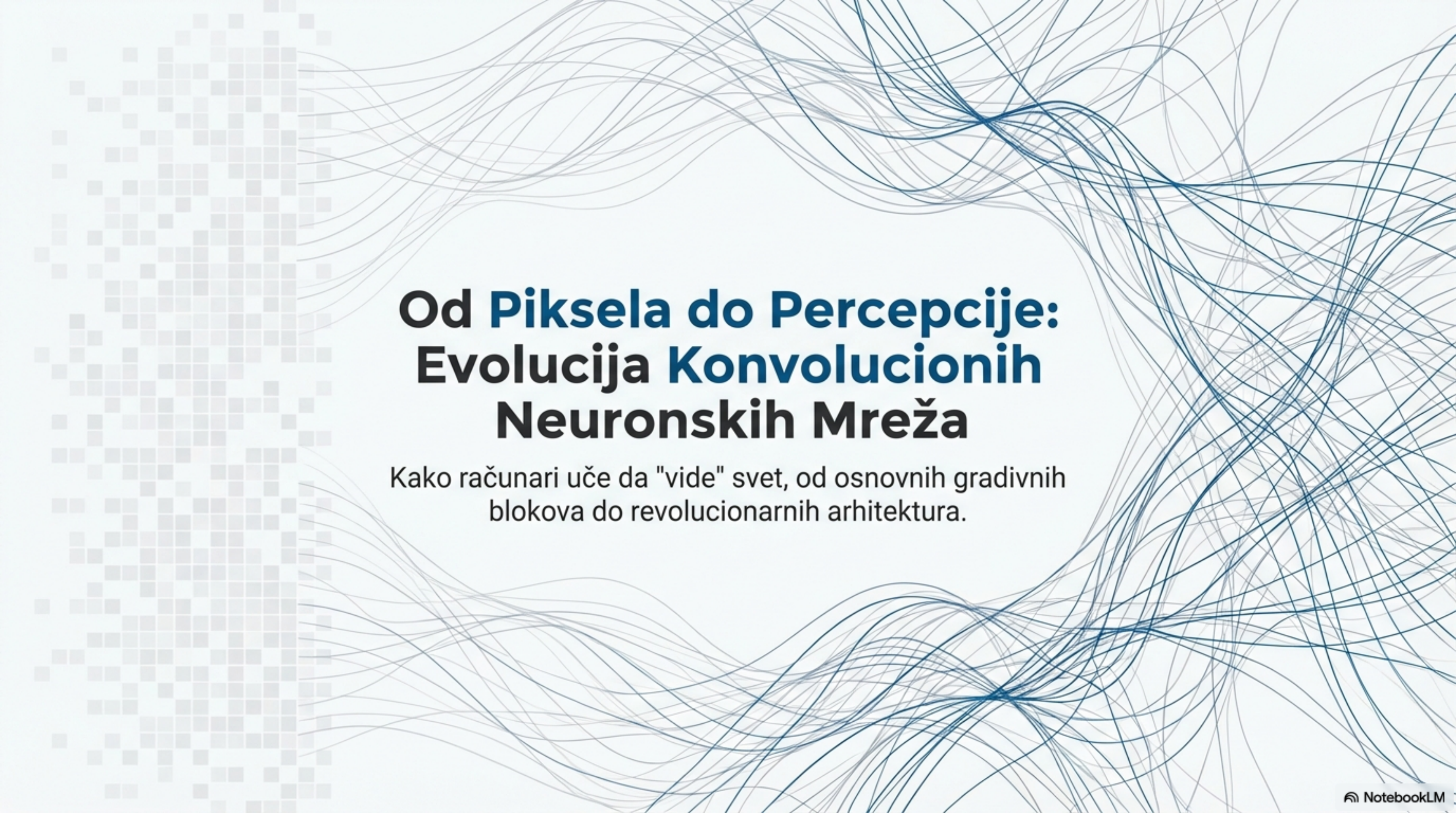




Od Piksela do Percepcije: Evolucija Konvolucionih Neuronskih Mreža

Kako računari uče da "vide" svet, od osnovnih gradivnih blokova do revolucionarnih arhitektura.

Možete li videti mačku?

Za čoveka, prepoznavanje objekata je intuitivno. Za računar, slika je samo matrica brojeva koji predstavljaju vrednosti piksela. Crno-bele slike su matrice gde 0 predstavlja crnu, 255 belu, a vrednosti između su nijanse sive. Izazov računarskog vida je prevesti ovu numeričku rešetku u smislen koncept.

```
[[ 9  1 29 70 114 76  0  8  4  5  5  0 111 162  9  8 62 62]
 [ 3  0 33 61 102 106 34  0  0  0  0 49 182 150  1 12 65 62]
 [ 1  0 40 54 123 90 72 77 52 51 49 182 150 98  0 15 67 62]
 [ 3  1 41 57 74 54 96 181 220 170 90 149 208 56  0 16 69 59]
 [ 6  1 32 36 47 81 85 96 176 200 140 171 186 22  3 15 72 63]
 [ 4  1 31 39 66 71 71 97 147 214 203 190 198 22  6 17 72 63]
 [ 2  3 15 30 52 57 68 123 161 167 207 200 179  8  8 18 73 66]
 [ 2  2 17 37 34 40 78 103 148 187 205 225 165  1  8 19 76 66]
 [ 2  3 20 44 37 34 35 36 78 186 214 145 200 38  2 21 76 69]
 [ 2  2 20 34 21 43 70 21 43 139 395 93 311 70  0 23 78 72]
 [ 3  4 16 24 14 21 102 175 120 130 226 222 236 75  0 23 78 72]
 [ 6  5 13 21 28 28 97 216 184 90 196 255 255 84  4 24 79 74]
 [ 6  5 15 15 30 30 63 105 140 66 113 251 251 74  4 26 79 75]
 [ 5  5 16 32 38 57 69 85 93 104 148 251 251 154 19 26 80 77]
 [ 6  5 20 42 55 62 66 76 86 104 145 242 254 241 83 26 78 77]
 [ 2  3 20 38 55 64 69 90 78 109 195 247 252 255 172 40 78 77]
 [10  8 23 34 44 64 88 104 119 173 234 247 253 254 227 66 74 74]
 [32  6 24 37 45 63 85 114 154 196 226 245 251 252 250 112 66 71]]
```

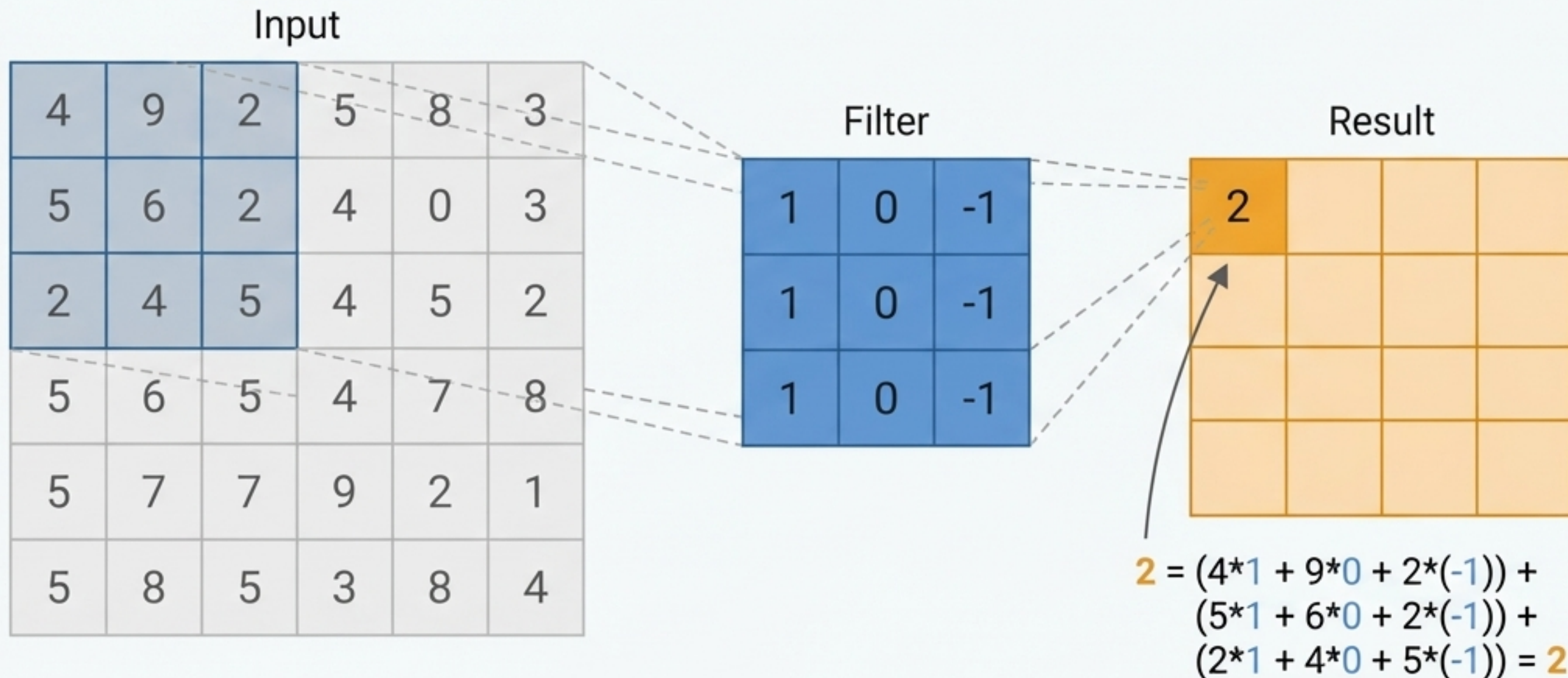
Matrica brojeva od 0 do 255.



Slika koju ljudski mozak trenutno prepoznaje.

Prvi Gradivni Blok: Magija Konvolucije

Osnovna operacija u CNN-u je konvolucija. Mali prozor, nazvan **filter** ili **kernel**, klizi preko ulazne slike. Na svakoj poziciji, izračunava se zbir proizvoda preklopljenih elemenata filtera i ulaza. Ovaj proces stvara novu matricu, takozvanu **mapu karakteristika** (feature map), koja naglašava određene obrasce sa slike.



Od Matematike do Značenja: Kako Filteri Detektuju Karakteristike

Različiti filteri su dizajnirani da detektuju različite karakteristike. Na primer, filter prikazan ovde je specijalizovan za pronalaženje vertikalnih ivica. Kada se primeni na sliku mačke, rezultat naglašava sve vertikalne linije, poput kontura njeng lica i ušiju. Mreža tokom obučavanja sama uči vrednosti filtera koje su najkorisnije za rešavanje zadatka.



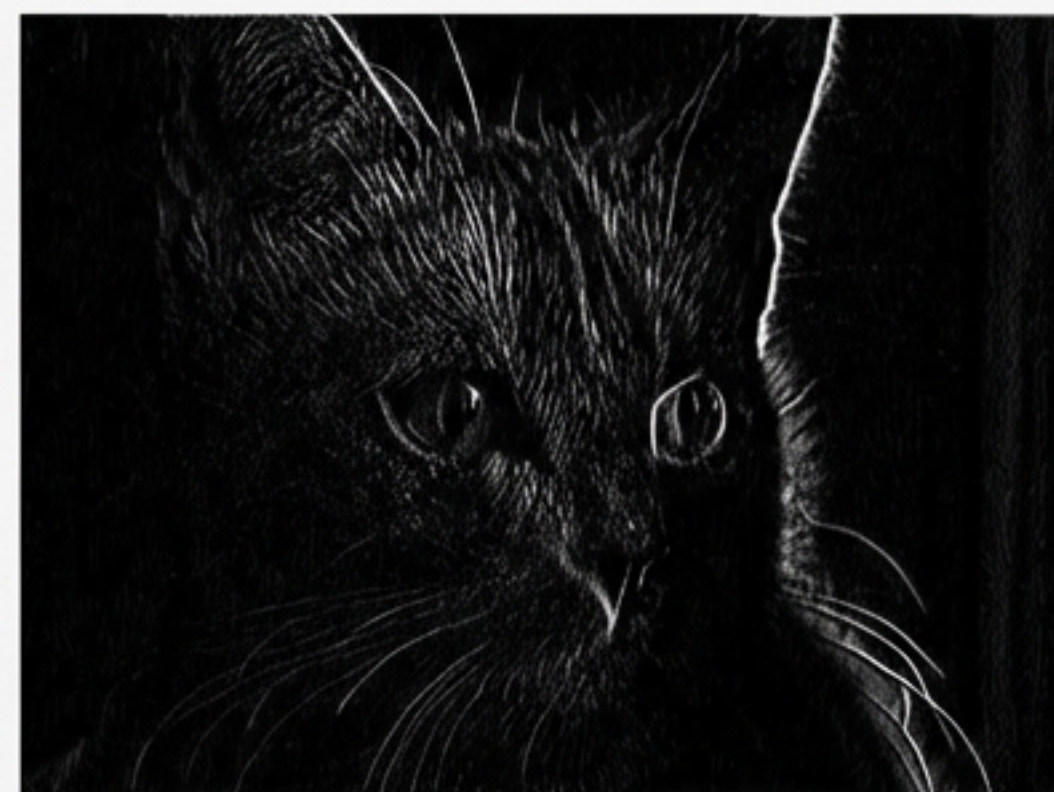
Originalna Slika

+

Filter za vertikalne ivice

1	0	-1
1	0	-1
1	0	-1

=

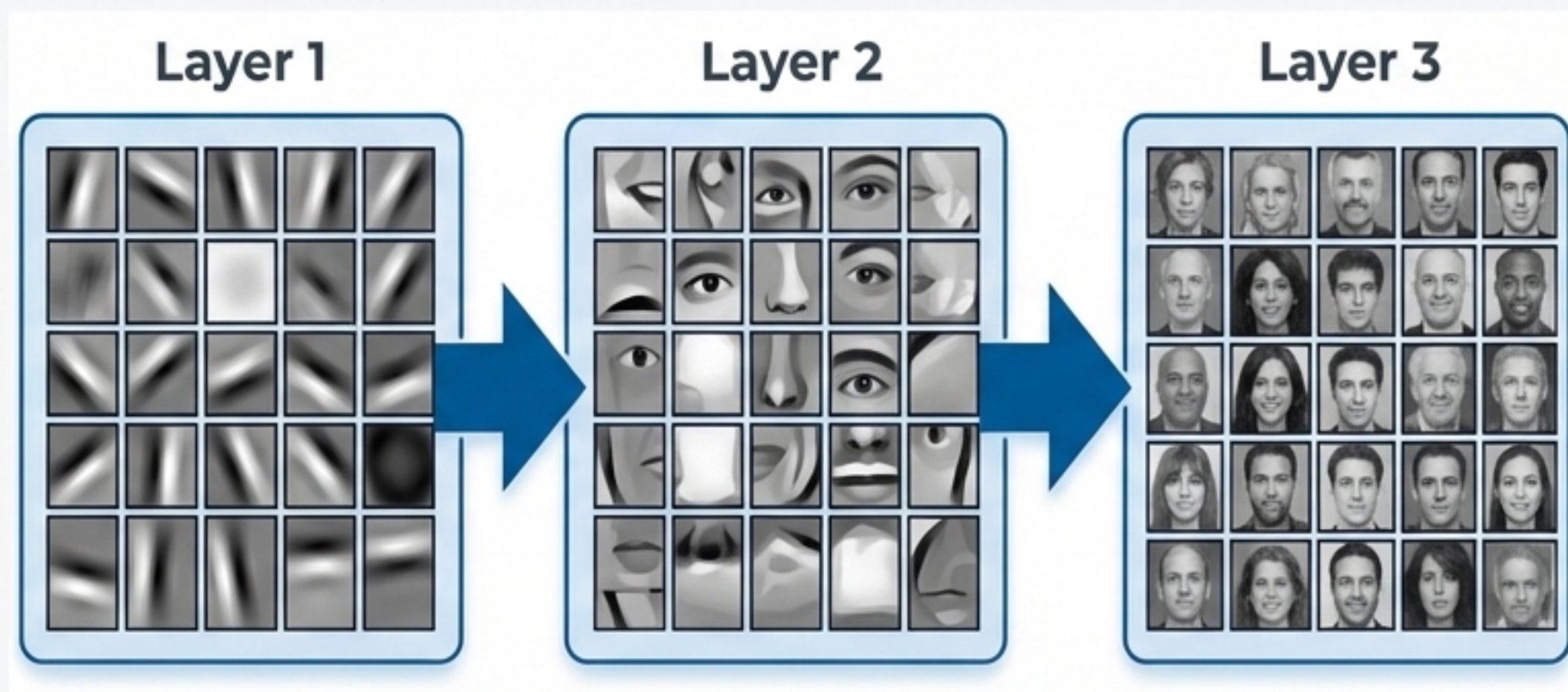


Rezultat: Detektovane Vertikalne Ivice

Hijerarhija Znanja: Od Ivica do Lica

Snaga dubokih CNN-ova leži u njihovoj slojevitoj strukturi. Svaki sloj gradi na znanju prethodnog, stvarajući hijerarhiju karakteristika.

- **Rani slojevi (Layer 1)** uče da prepoznaju jednostavne oblike: ivice, uglove i gradijente boja.
- **Srednji slojevi (Layer 2)** kombinuju ove jednostavne oblike u složenije delove: oči, noseve, usta.
- **Duboki slojevi (Layer 3)** sastavljaju ove delove u kompletne objekte, kao što su lica.



Drugi Gradivni Blok: Agregacija (Pooling) za Efikasnost

Pored konvolucije, **agregacija (pooling)** je ključna operacija. Njena svrha je da smanji prostorne dimenzije mapa karakteristika, čime se smanjuje broj parametara i računskih operacija. Ovo čini model bržim i pomaže u sprečavanju prekomernog prilagođavanja (overfitting).

- **Max Pooling:** Bira maksimalnu vrednost iz regiona. Naglašava najdominantnije karakteristike.
- **Average Pooling:** Računa prosečnu vrednost. Daje generalizovaniju reprezentaciju.

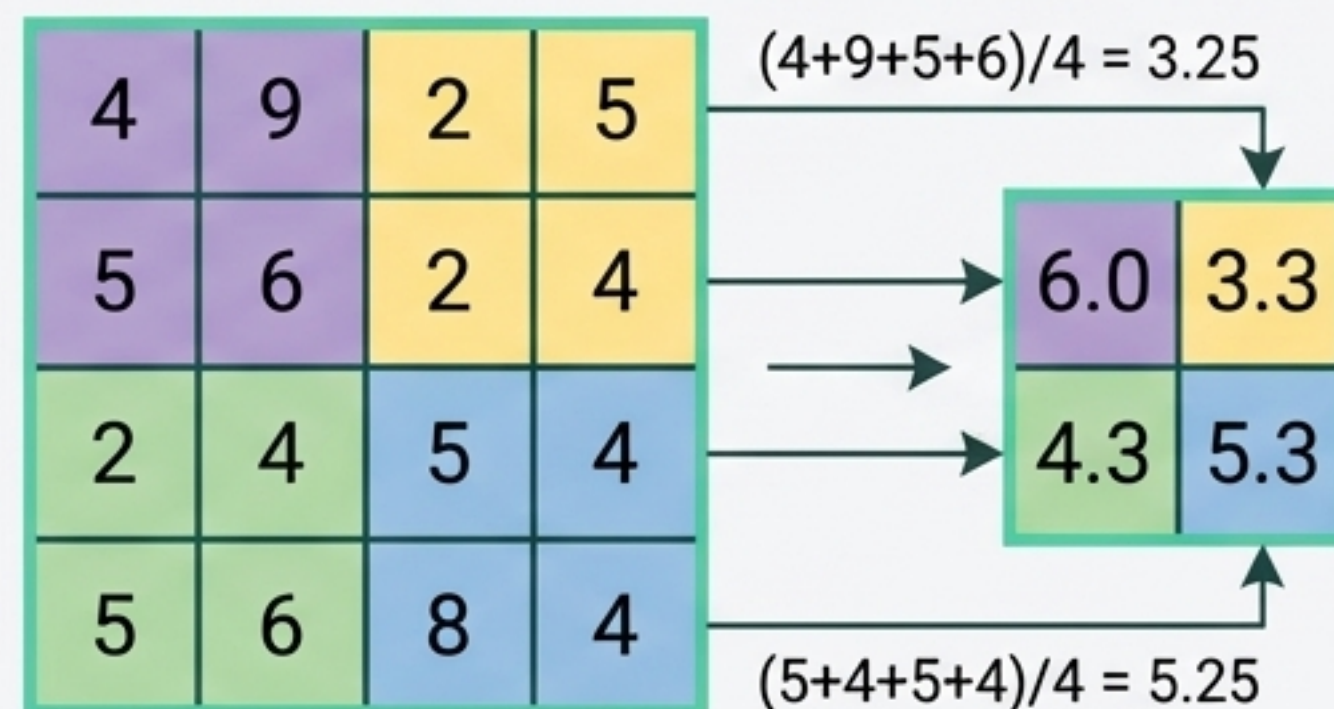
Ključne Prednosti Agregacije:

1. **Smanjenje dimenzionalnosti:** Ubrzava proračune.
2. **Invarijantnost na translaciju:** Pomaže mreži da prepozna objekat bez obzira na njegovu blagu promenu pozicije.

Max Pooling



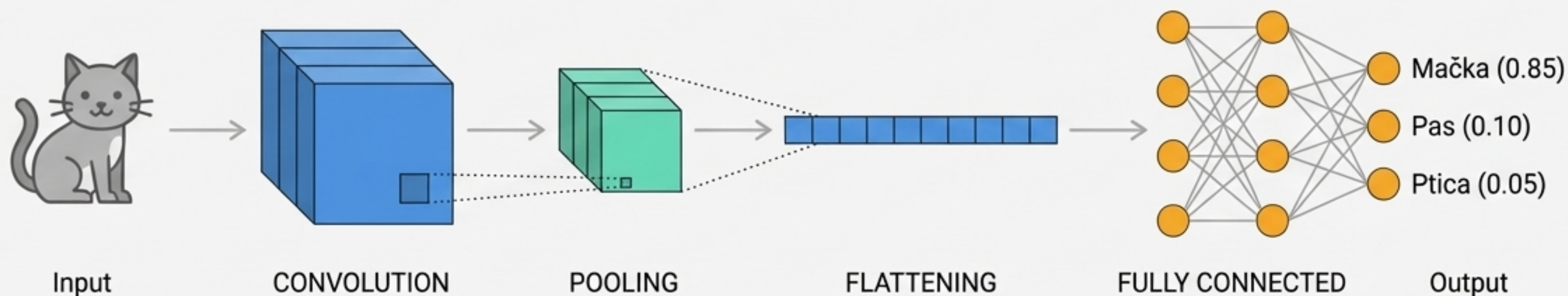
Average Pooling



Sastavljanje Arhitekture: Od Ulaza do Klasifikacije

Konvolucionni slojevi i slojevi agregacije se kombinuju u **sekvence** kako bi formirali **duboku mrežu**. Tipična arhitektura se sastoji od dva glavna dela:

1. **Učenje Karakteristika (Feature Learning)**: Niz Konvolucionih i Pooling slojeva progresivno ekstrahuje sve složenije karakteristike iz ulazne slike.
2. **Klasifikacija (Classification)**: Nakon "ispravljanja" (flattening) poslednje mape karakteristika u vektor, potpuno povezani slojevi (fully connected layers) analiziraju ove karakteristike i donose konačnu odluku.



Revolucija u Računarskom Vidu

Sa uspostavljenim osnovnim principima i snagom GPU-a, polje računarskog vida je doživelo eksplozivan rast. Nekoliko ključnih arhitektura je definisalo modernu eru, svaka rešavajući specifičan problem i pomerajući granice mogućeg. U nastavku ćemo istražiti četiri arhitekture koje su oblikovale duboko učenje:

- **AlexNet**: Pionir dubokih mreža na velikoj skali.
- **VGG**: Demonstracija moći ekstremne dubine.
- **ResNet**: Omogućavanje još dubljih mreža bez degradacije.
- **Inception**: Uvođenje efikasnosti kroz paralelizam.

AlexNet

VGG

Inception

ResNet

2012

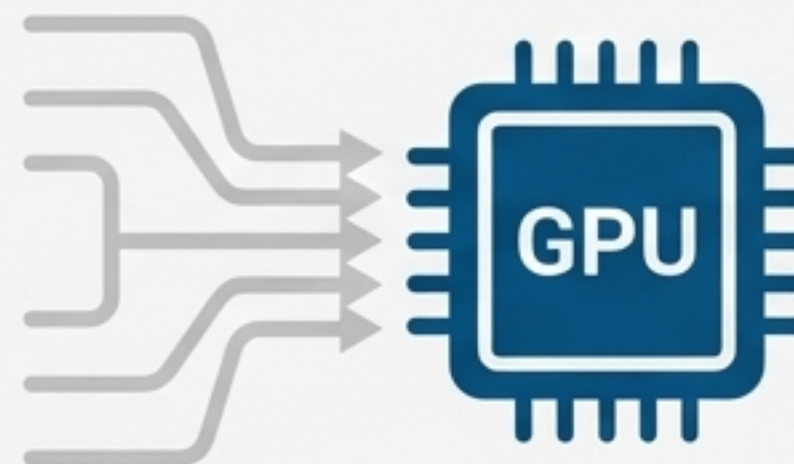
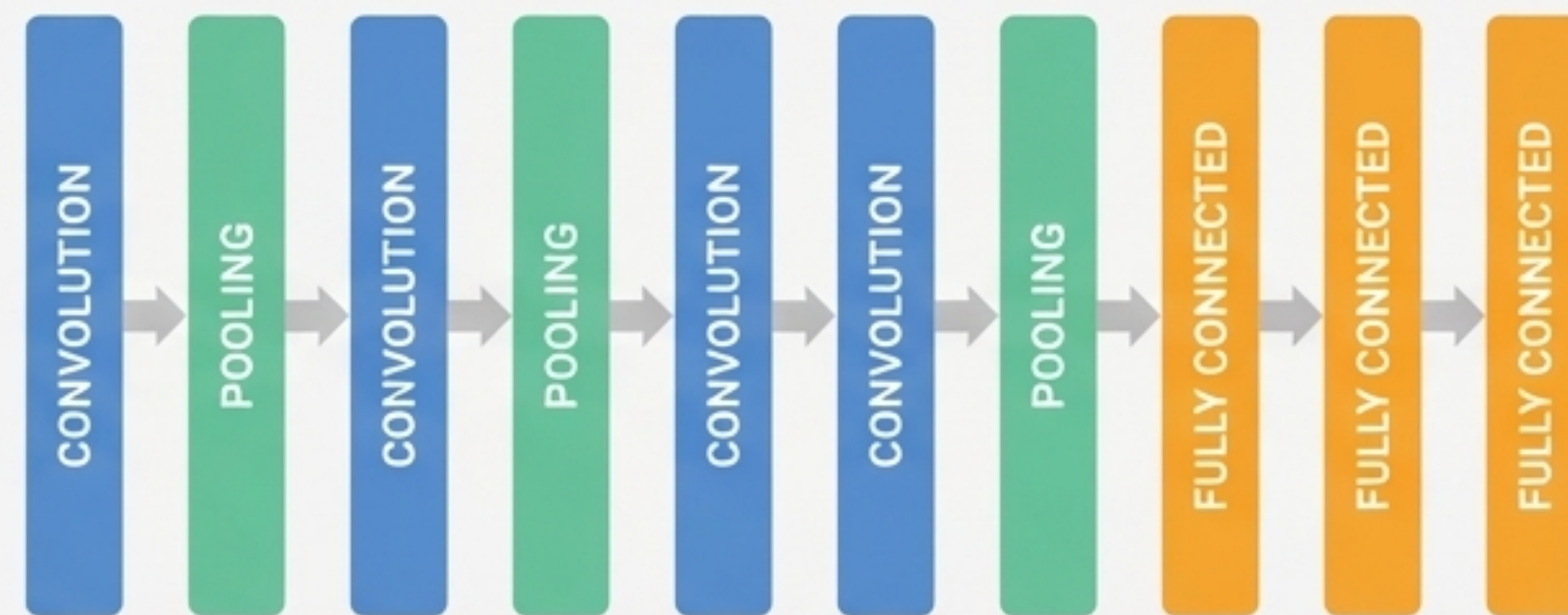
AlexNet: Iskra Koja je Pokrenula Revoluciju (2012)

Velika Ideja

AlexNet je dramatično nadmašila konkurenciju na ImageNet takmičenju 2012. godine, dokazujući da su duboke konvolucione mreže, trenirane na velikim skupovima podataka uz pomoć GPU-a, budućnost računarskog vida.

Ključne Inovacije

- **ReLU Aktivaciona Funkcija:** Ubrzala je trening i smanjila problem nestajućeg gradijenta. Formula: $f(x) = \max(0, x)$.
- **Korišćenje GPU-a:** Omogućilo je treniranje znatno većih i dubljih modela na masivnom ImageNet skupu podataka.
- **Dropout Regularizacija:** Tehnika za smanjenje prekomernog prilagođavanja nasumičnim "isključivanjem" neurona tokom treninga.



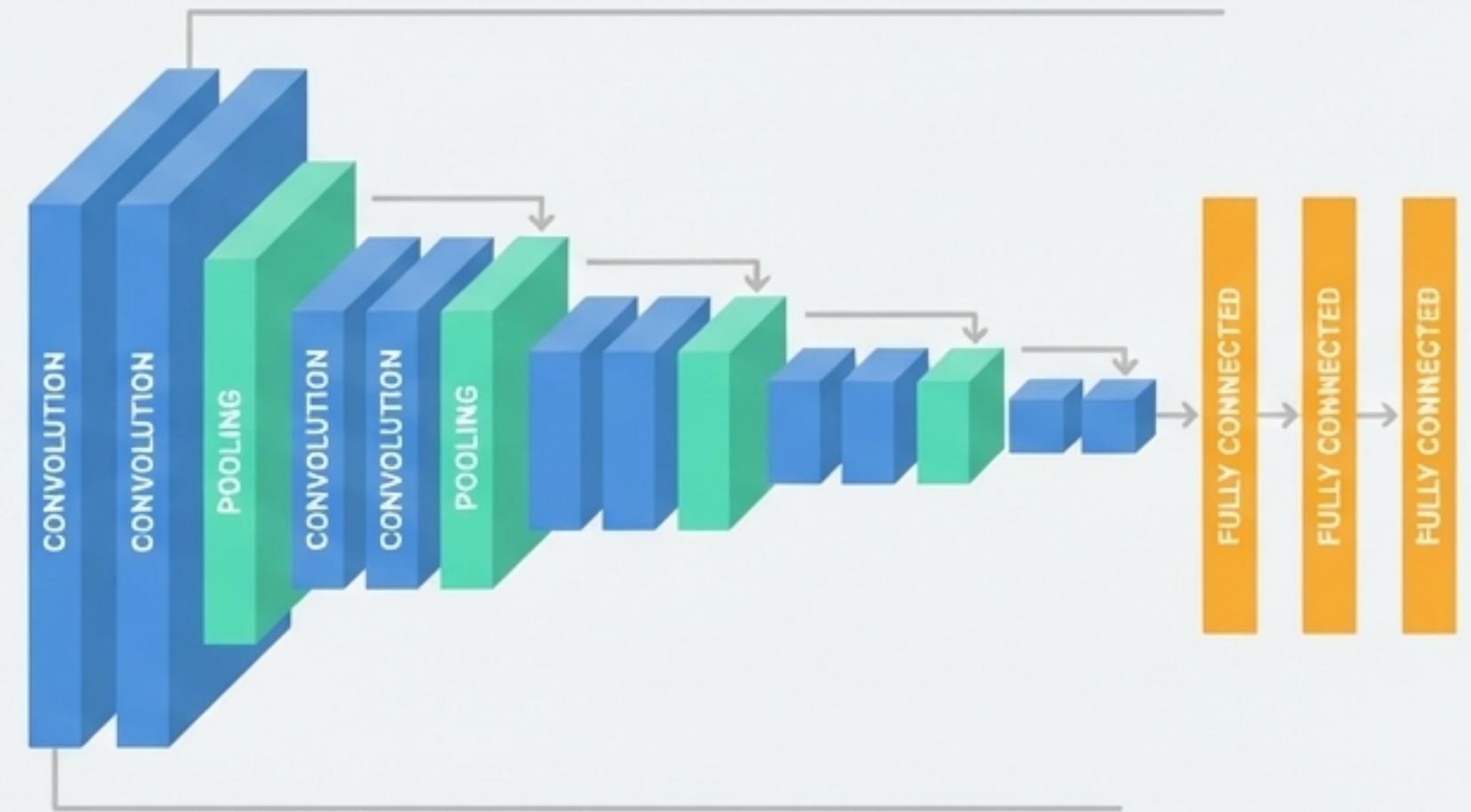
VGG: Dubina je Ključna (2014)

Velika Ideja

VGG arhitektura je pokazala da se izvanredne performanse mogu postići povećanjem dubine mreže, koristeći izuzetno uniformnu i jednostavnu strukturu.

Ključne Inovacije

- **Ekskluzivna upotreba 3x3 filtera:** Umesto velikih filtera, VGG koristi nizove malih 3x3 filtera. Dva uzastopna 3x3 filtera imaju efektivno receptivno polje kao jedan 5x5 filter, ali sa više nelinearnosti i manje parametara.
- **Ekstremna dubina:** Modeli kao što su VGG-16 i VGG-19 (sa 16 i 19 slojeva) postavili su nove standarde za dubinu mreža.
- **Jednostavnost i uniformnost:** Ponavljajući blokovi (konvolucija + konvolucija + max pooling) učinili su arhitekturu lakom za razumevanje i implementaciju.



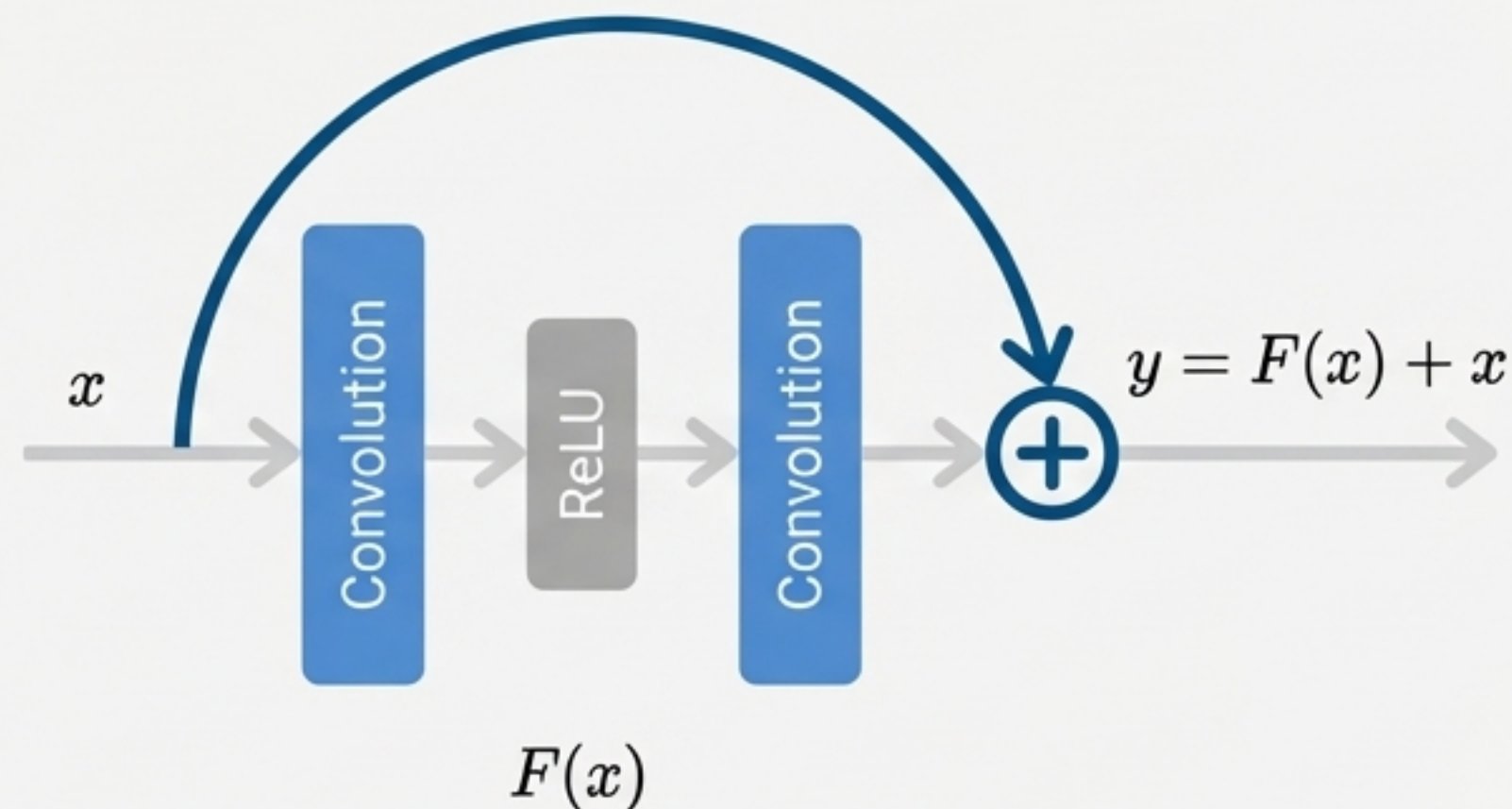
ResNet: Pobeđivanje Problema Degradacije (2015)

Velika Ideja

Kako mreže postaju sve dublje, javlja se problem degradacije—tačnost dostiže vrhunac i zatim počinje da opada. ResNet je ovo rešio uvođenjem 'prečica' koje omogućavaju treniranje mreža sa preko 100, pa čak i 1000 slojeva.

Ključne Inovacije

- **Rezidualne 'Skip' Konečije:** Omogućavaju da informacija zaobiđe jedan ili više slojeva. Formula $y = F(x) + x$ omogućava slojevima da uče samo rezidualnu (preostalu) informaciju, što olakšava optimizaciju.
- **Olakšan protok gradijenta:** Prečice pomažu u borbi protiv problema nestajućeg gradijenta u veoma dubokim mrežama.
- **Bottleneck Arhitektura:** Koristi 1x1 konvolucije za smanjenje i vraćanje dimenzija oko skupljih 3x3 konvolucija, čineći duboke mreže računski efikasnijim.



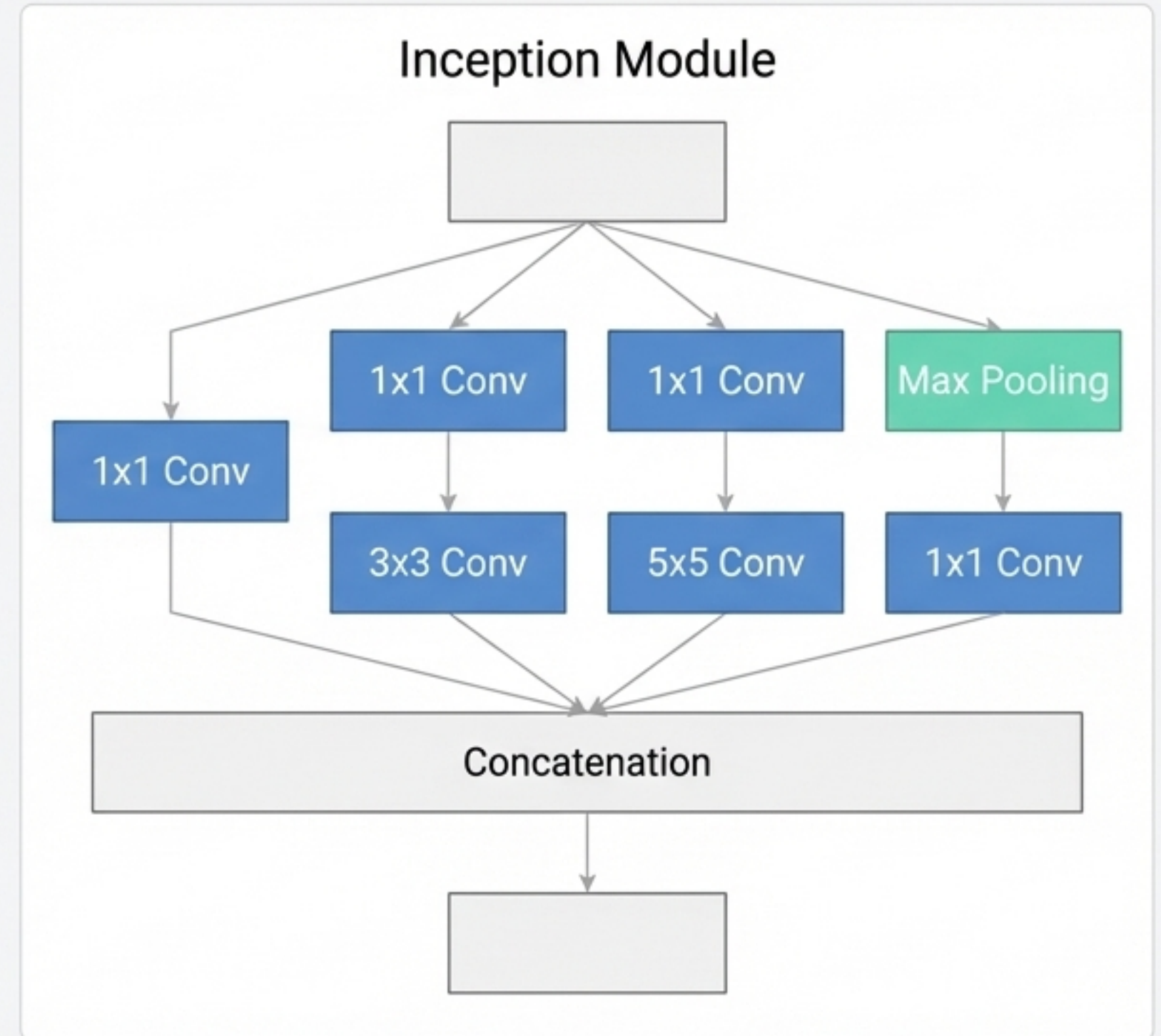
Inception: Efikasnost kroz Paralelne Puteve (2014)

Velika Ideja

Umesto da samo dodaje slojeve u dubinu, Inception arhitektura (poznata i kao GoogLeNet) uvodi ideju "širine". U jednom sloju, mreža paralelno primenjuje filtere različitih veličina kako bi efikasno uhvatila karakteristike na više skala.

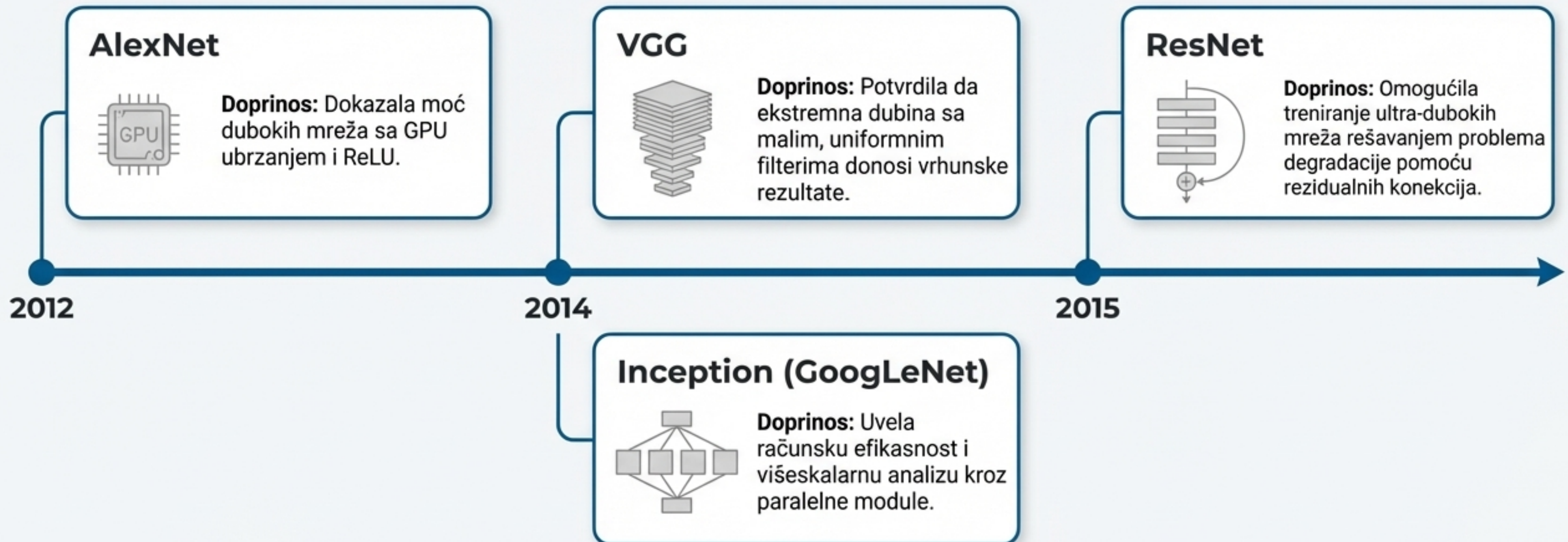
Ključne Inovacije

- **Inception Modul:** Osnovni gradivni blok koji vrši konvolucije sa 1x1, 3x3 i 5x5 filterima, kao i max pooling, unutar istog modula. Rezultati se zatim spajaju (konkateniraju).
- **1x1 Konvolucije za Smanjenje Dimenzija:** Pametno se koriste pre skupljih 3x3 i 5x5 konvolucija kako bi se smanjio broj računskih operacija.
- **Global Average Pooling:** Zamenjuje skupe potpuno povezane slojeve na kraju mreže, značajno smanjujući ukupan broj parametara i rizik od overfitting-a.



Vremenska Linija Inovacija

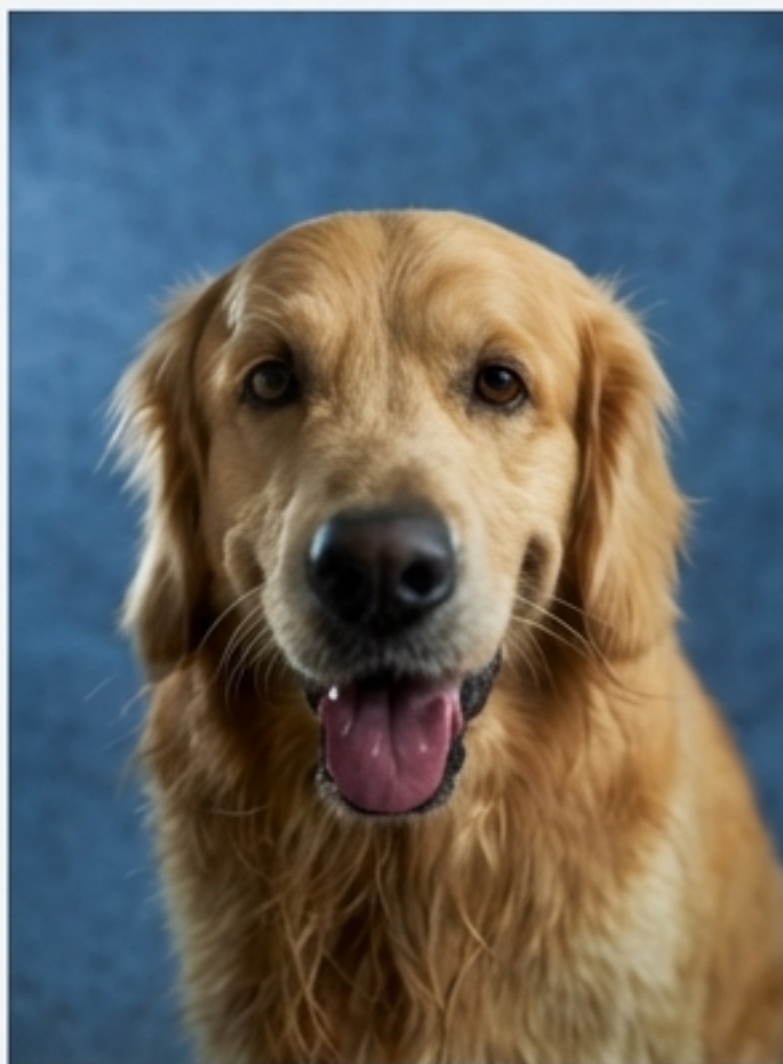
Svaka arhitektura je nadogradila prethodne, uvodeći nove ideje koje su postale standard u polju.



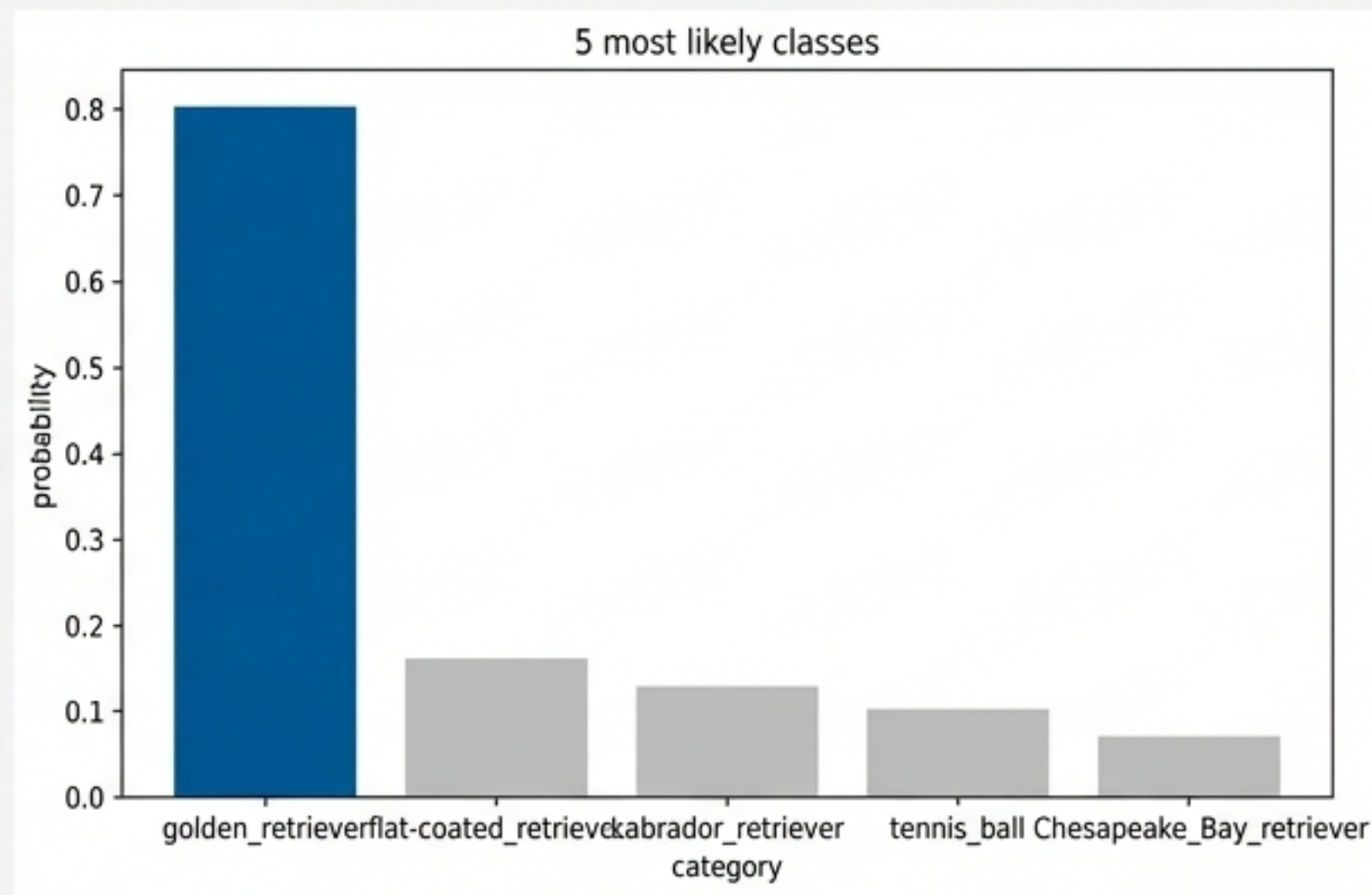
Od Teorije do Prakse: VGG-16 u Akciji

Snaga ovih arhitektura nije samo teorijska. Unapred istrenirani modeli, poput VGG-16, dostupni su u bibliotekama kao što je **Keras** i mogu se koristiti za rešavanje stvarnih problema sa samo nekoliko linija koda.

Primer: Testiramo VGG-16 model, treniran na ImageNet skupu podataka, na slici zlatnog retrivera.



```
from tensorflow.keras.applications import VGG16
model = VGG16(weights = 'imagenet')
# ... code to preprocess image and predict
```



Model je sa verovatnoćom od ~80% tačno identifikovao sliku kao "golden_retriever". Zanimljivo je da su druge predikcije srodne rase pasa, pa čak i "teniska loptica", verovatno zbog česte asocijacije u podacima za trening.

Evolucija se Nastavlja

Priča o konvolucionim mrežama se ne završava ovde. Arhitekture nastavljaju da se razvijaju, sa novim varijacijama kao što su:

- **Proširena konvolucija (Dilated convolution)** za veće receptivno polje bez dodatnih troškova.
- **Dubinski odvojiva konvolucija (Depthwise separable convolution)** za ekstremnu efikasnost na mobilnim uređajima.
- **Transformeri** koji donose nove paradigme u računarski vid.

Ipak, fundamentalni principi—**hijerarhijsko učenje karakteristika** kroz slojeve **konvoluciji** slojeve **konvolucije** i **agregacije**—ostaju temelj na kojem počiva celokupna moderna računarska vizija. Od jednostavne matrice piksela, ove mreže grade složeno razumevanje, dajući mašinama sposobnost da percipiraju naš svet.

