

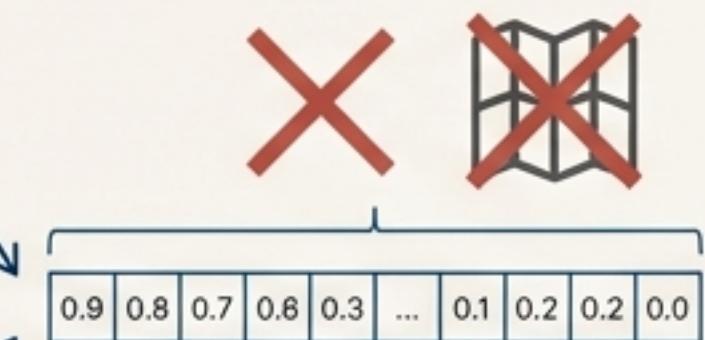
Slike so več kot le seznam številk

Gosto povezane nevronske mreže (Dense Neural Networks) obravnavajo vhodne podatke kot nepovezan seznam. Pri slikah s tem uničijo ključno informacijo: prostorsko razporeditev pikslov.

- Računalnik vidi sliko kot matriko števil, kjer vsaka vrednost predstavlja barvo enega piksla.
- Položaj vsakega piksla je ključnega pomena za razumevanje vsebine slike (npr. oko je nad nosom, ne pod njim).
- Če sliko 'sploščimo' v enodimenzionalen vektor za gosto povezano mrežo, kjer je vsak nevron povezan z vsemi, se ta ključna geometrijska struktura izgubi.



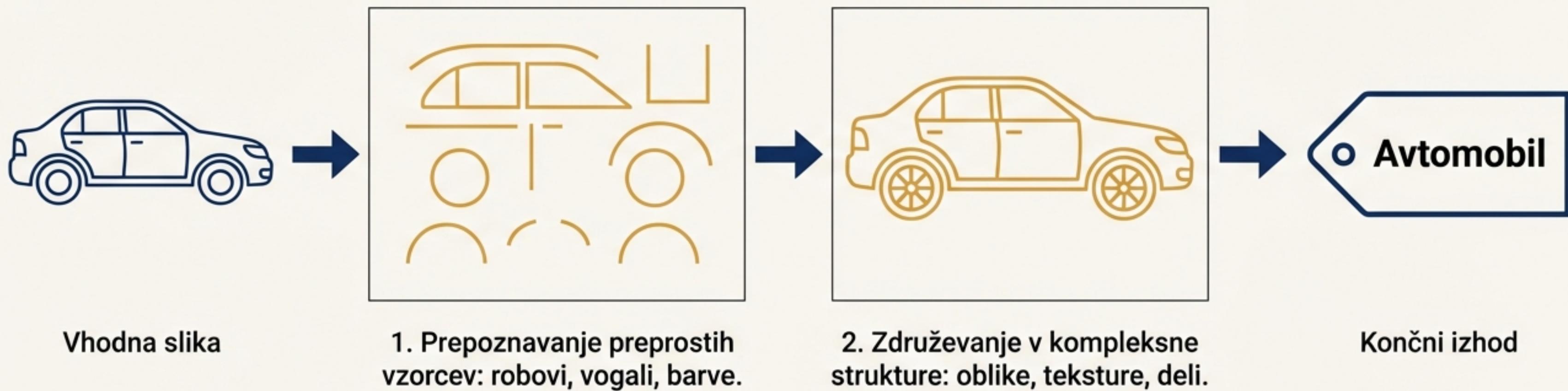
0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.2	0.0
0.1	0.1	0.9	0.8	0.8	0.8	0.2	0.1	
0.1	0.9	0.7			0.7	0.6	0.2	
0.1	0.6					0.6	0.2	
0.2							0.2	
0.1	0.8				0.6	0.0		
0.1	0.8				0.6	0.0		
0.1	0.5	0.7	0.6	0.6	0.6	0.5	0.0	
0.1	0.1	0.1	0.1	0.0	0.0	0.0	0.0	
0.0	0.0	0.0	0.0	0.2	0.0	0.0	0.0	0.0



Gosto povezana mrež -> Izguba prostorske informacije

Konvolucijske nevronske mreže: Arhitektura, ki 'vidi'

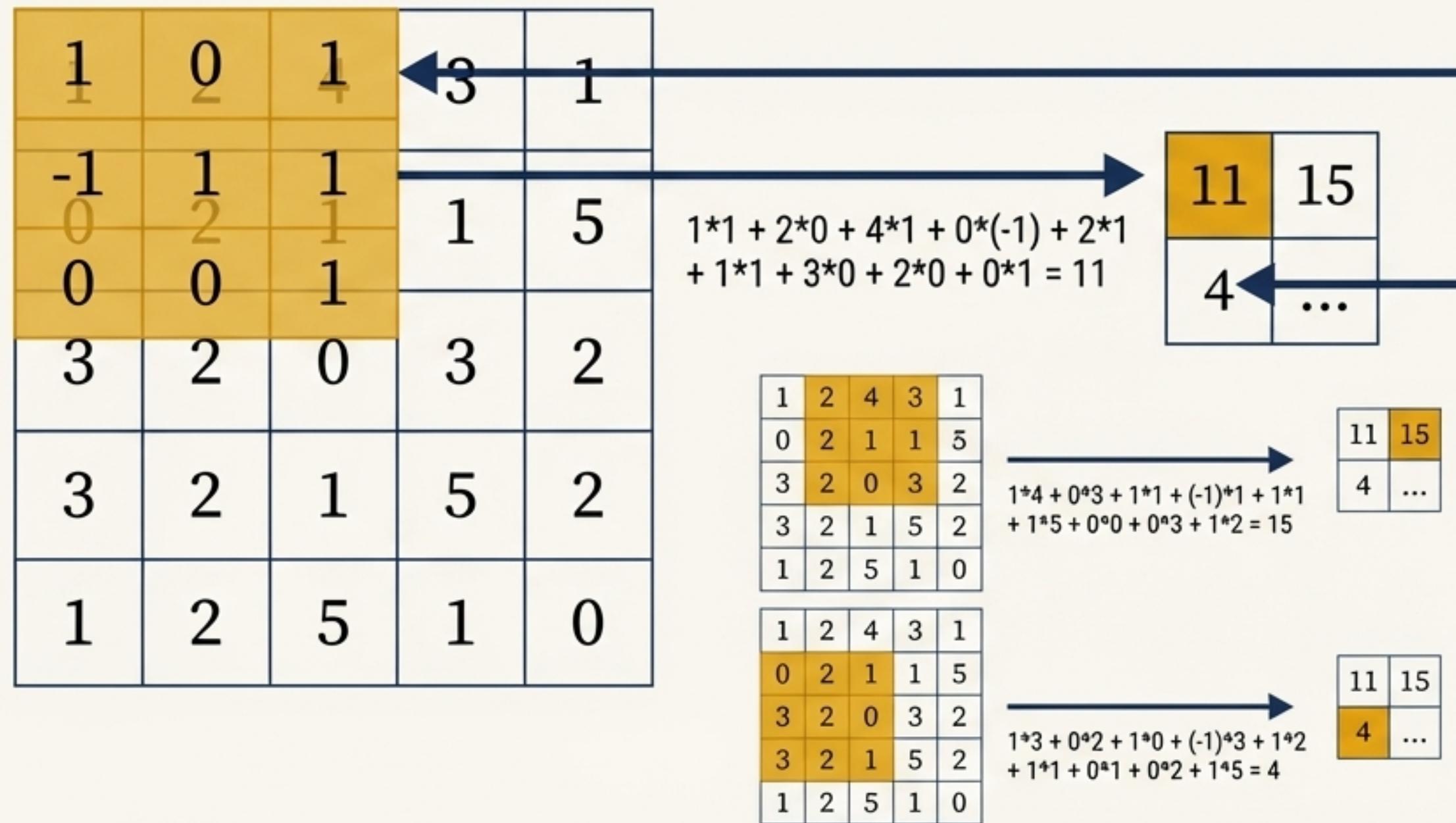
CNN-i so zasnovani tako, da ohranjajo in izkoriščajo prostorsko strukturo slike. Delujejo tako, da se postopoma učijo prepoznavati vizualne vzorce – od preprostih do kompleksnih.



CNN ohranja geometrijsko obliko slike in se uči hierarhično, od detajlov do celote.

Srce operacije: Konvolucijska plast

Konvolucijska plast uporablja majhne filtre (jedra), ki drsijo po sliki in prepoznavajo specifične vzorce. To doseže z dvema ključnima lastnostma: **lokalno povezanostjo** in **deljenjem uteži**.



Filter (jedro): Majhna matrika uteži (npr. 3x3), ki drsi po vhodni sliki.

Lokalna povezanost: Vsak nevron v izhodni plasti je povezan le z majhnim, lokalnim območjem vhodne slike.

Deljenje uteži: Isti filter (z istimi utežmi) se uporablja po celotni sliki. To omogoča mreži, da prepozna isti vzorec (npr. navpično črto) ne glede na to, kje na sliki se pojavi.

Kako ohraniti robove slike? Rešitev je "padding"

'Padding' (dodajanje robov) rešuje problem zmanjševanja dimenzij in zanemarjanja robnih pikslov z dodajanjem dodatnih pikslov okoli slike.

Brez paddinga se izhodna slika zmanjša, robni piksli pa so v izračun vključeni manjkrat kot sredinski.

Brez paddinga se izhodna slika zmanjša, robni piksli pa so v izračun vključeni manjkrat kot sredinski.

Strategije glede na velikost izhoda:

- **Valid padding:** Brez dodajanja robov; izhod se skrči.
- **Same padding:** Doda se toliko robov, da je izhodna dimenzija enaka vhodni.
- **Full padding:** Zagotovi, da je vsak vhodni piksel obdelan enako število krat.

Načini dodajanja pikslov:

- **Zero padding:** Najpogostejsi način; okoli slike se dodajo ničle.
- **Mirror/Reflect padding:** Robovi slike se zrcalijo.
- **Circular padding:** Slika se 'zavije' – piksli z desnega roba se pojavijo na levem.

Zero padding

0	0	0	0	0	0
0	3	1	5	2	0
0	7	9	4	8	0
0	2	6	0	1	0
0	9	3	7	5	0
0	0	0	0	0	0

Mirror padding

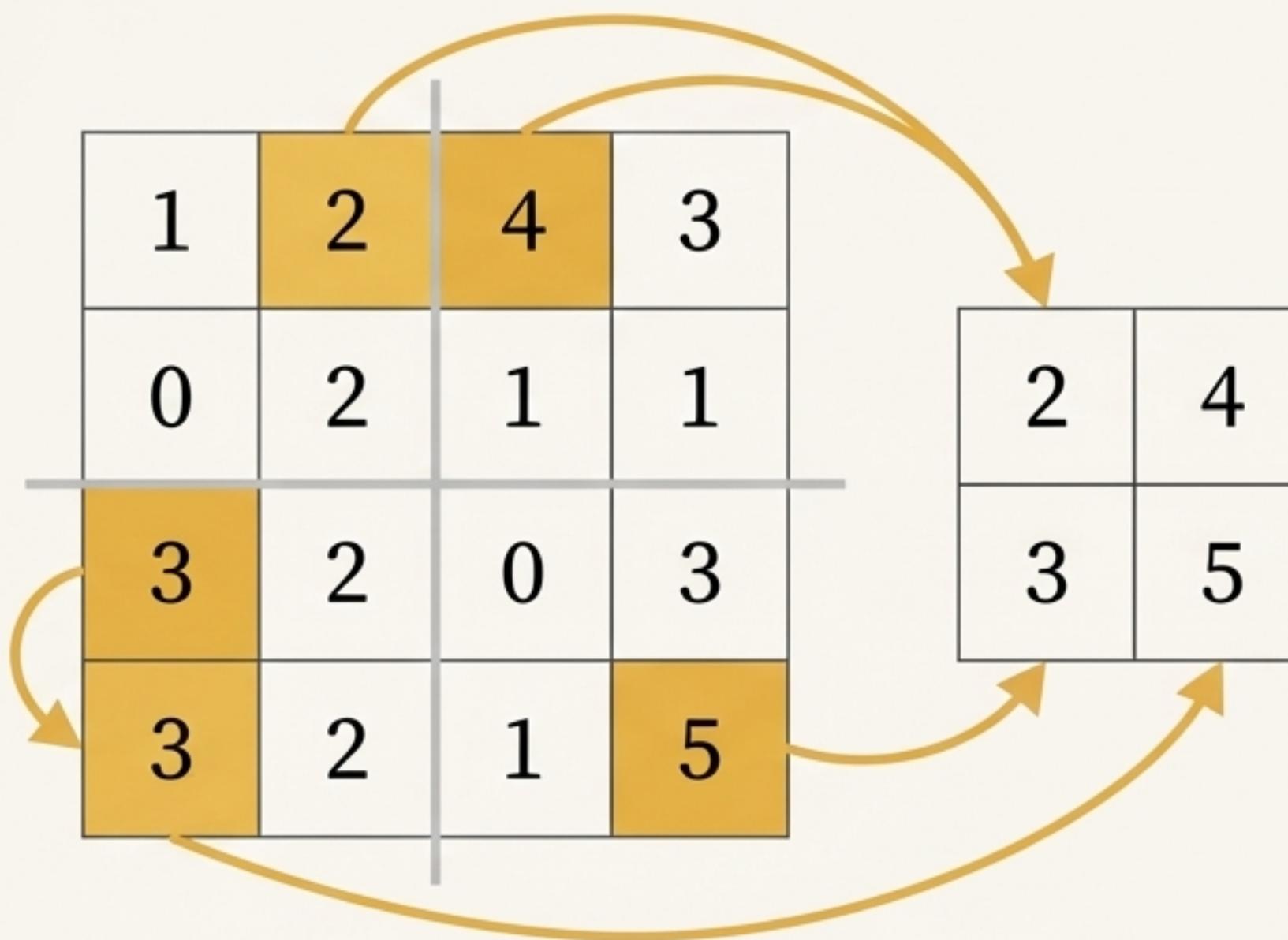
3	3	1	5	2	2
3	3	1	5	2	2
7	7	9	4	8	8
2	2	6	0	1	1
9	9	3	7	5	5
9	9	3	7	5	5

Circular padding

5	9	3	7	5	9
2	3	1	5	2	3
8	7	9	4	8	7
1	2	6	0	1	2
5	9	3	7	5	9
2	3	1	5	2	3

Združevanje in povzemanje: Moč 'pooling' plasti

Namen 'pooling' plasti je postopno zmanjševanje dimenzijske podatkov (downsampling), s čimer se zmanjša računska zahtevnost in ohranijo le najpomembnejše značilnosti.



'Pooling' deluje na majhnih območjih (npr. 2x2 piksla) in jih združi v eno samo vrednost.

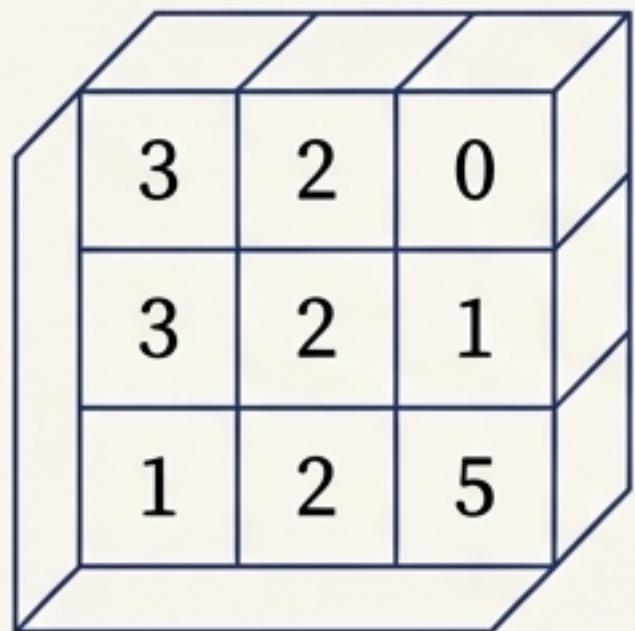
S tem mreža postane manj občutljiva na natančno lokacijo značilnosti v sliki.

Dve najpogosteji vrsti:

- **Max pooling:** Izbere največjo vrednost v območju. Poudari najmočnejši signal oziroma najbolj izrazito značilnost.
- **Average pooling:** Izračuna povprečno vrednost. Zgladi podatke in zajame splošno informacijo o območju.

Od zemljevida značilnosti do končne odločitve

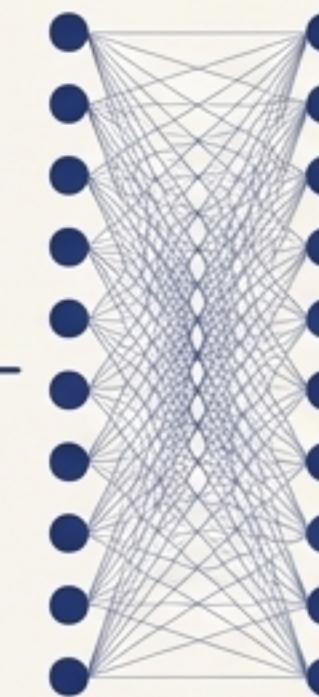
Po ekstrakciji ključnih vizualnih značilnosti je treba podatke pripraviti za končno klasifikacijo. To storita sploščevalna in gosto povezana plast.



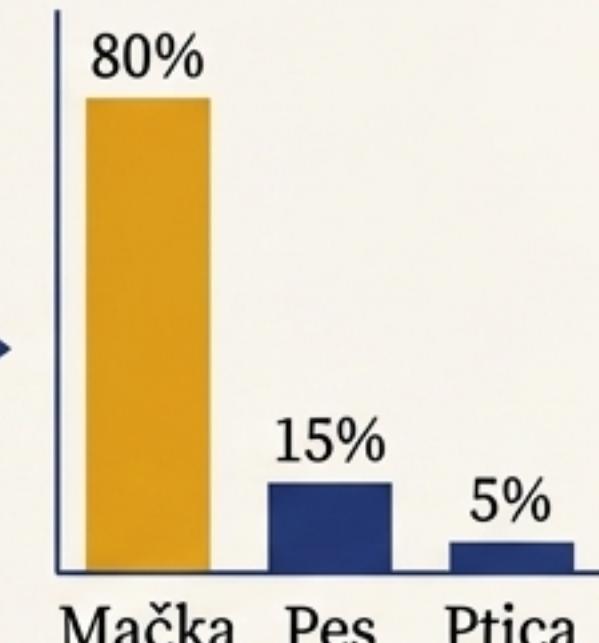
Zemljevid značilnosti



(3, 2, 0, 3, 2, 1, 1, 2, 5)



Enodimensionalen vektor



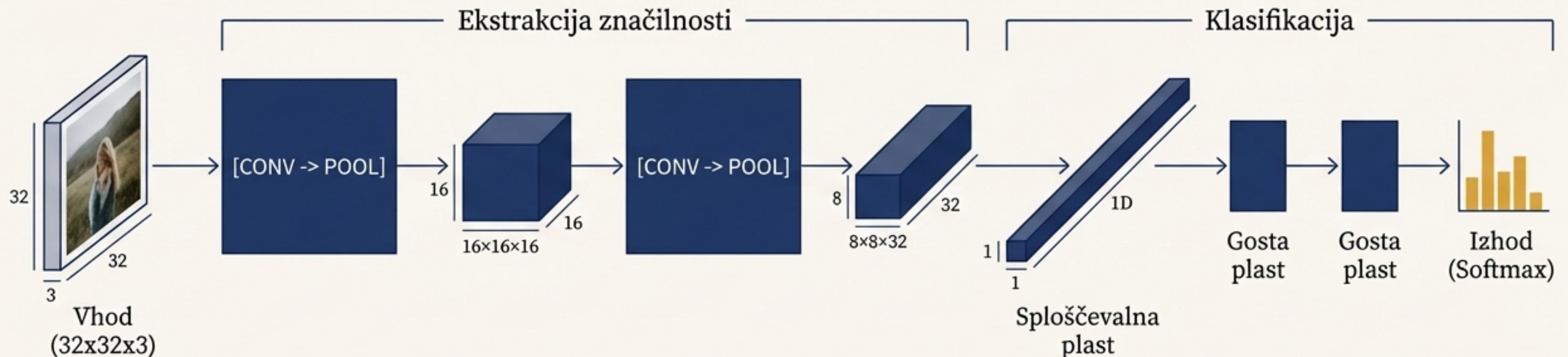
Izhodne verjetnosti

Sploščevalna plast (Flatten layer): Njena edina naloga je, da večdimenzionalno matriko (zemljevid značilnosti) pretvori v enodimensionalen vektor. S tem pripravi podatke za vnos v standardno gosto povezano mrežo.

Gosto povezane plasti (Dense layers): Na koncu arhitekture so klasične, polno povezane plasti. Te plasti na podlagi naučenih značilnosti izračunajo verjetnosti za vsako kategorijo.

Sestavljanka: Tipična arhitektura CNN mreže

Standardna CNN arhitektura je zaporedje izmenjujočih se konvolucijskih in ‘pooling’ plasti, ki se zaključi s sploščitvijo in gosto povezanimi plastmi za klasifikacijo.



1. → **Vhod***: Vhodna slika.
2. → **Ekstrakcija značilnosti***: Zaporedje blokov [Konvolucijska plast -> Plast za združevanje].
3. → **Klasifikacija***: Zaključek s ploskvijo [Sploščevalna plast -> Goste plasti].

VGG: Šampion, ki je definiral globino

Modela VGG16 in VGG19, razvita na Univerzi v Oxfordu, sta leta 2014 zmagala na prestižnem tekmovanju ImageNet. Njihov uspeh temelji na preprosti, a zelo globoki arhitekturi.

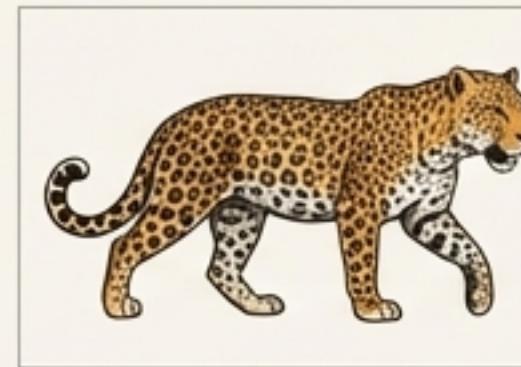


Razvoj: Visual Geometry Group (VGG)



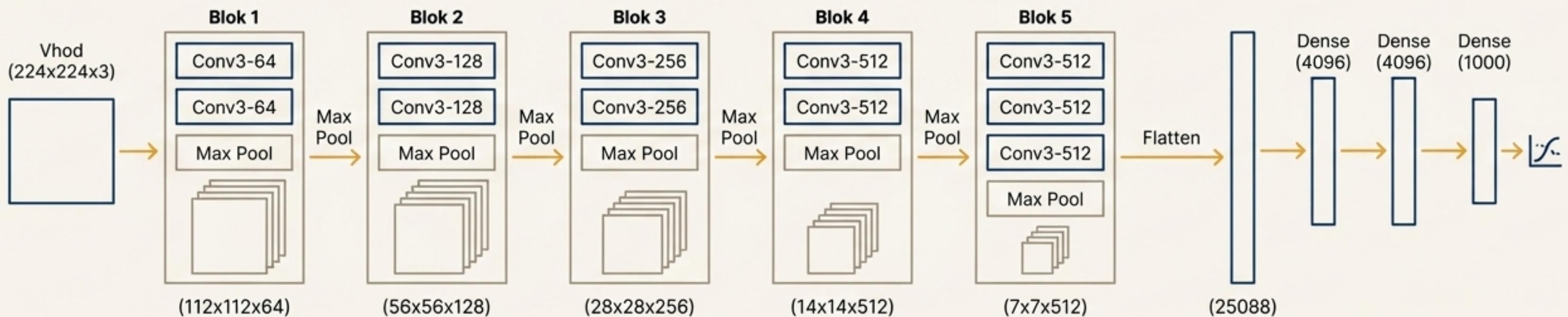
Tekmovanje: Zmagovalca tekmovanja ImageNet 2014

- **Naloga:** Klasifikacija več kot milijon slik v 1000 kategorij.
- **Ključne odlike:** Globoka arhitektura, uporaba preprostih in ponavljajočih se gradnikov, velika razširjenost v praksi.



Anatomija modela VGG

Arhitektura VGG dosledno uporablja majhne 3×3 konvolucijske filtre, ki jim sledijo 2×2 max-pooling plasti. Število filtrov se z globino mreže postopoma podvaja.



Konvolucijski del:

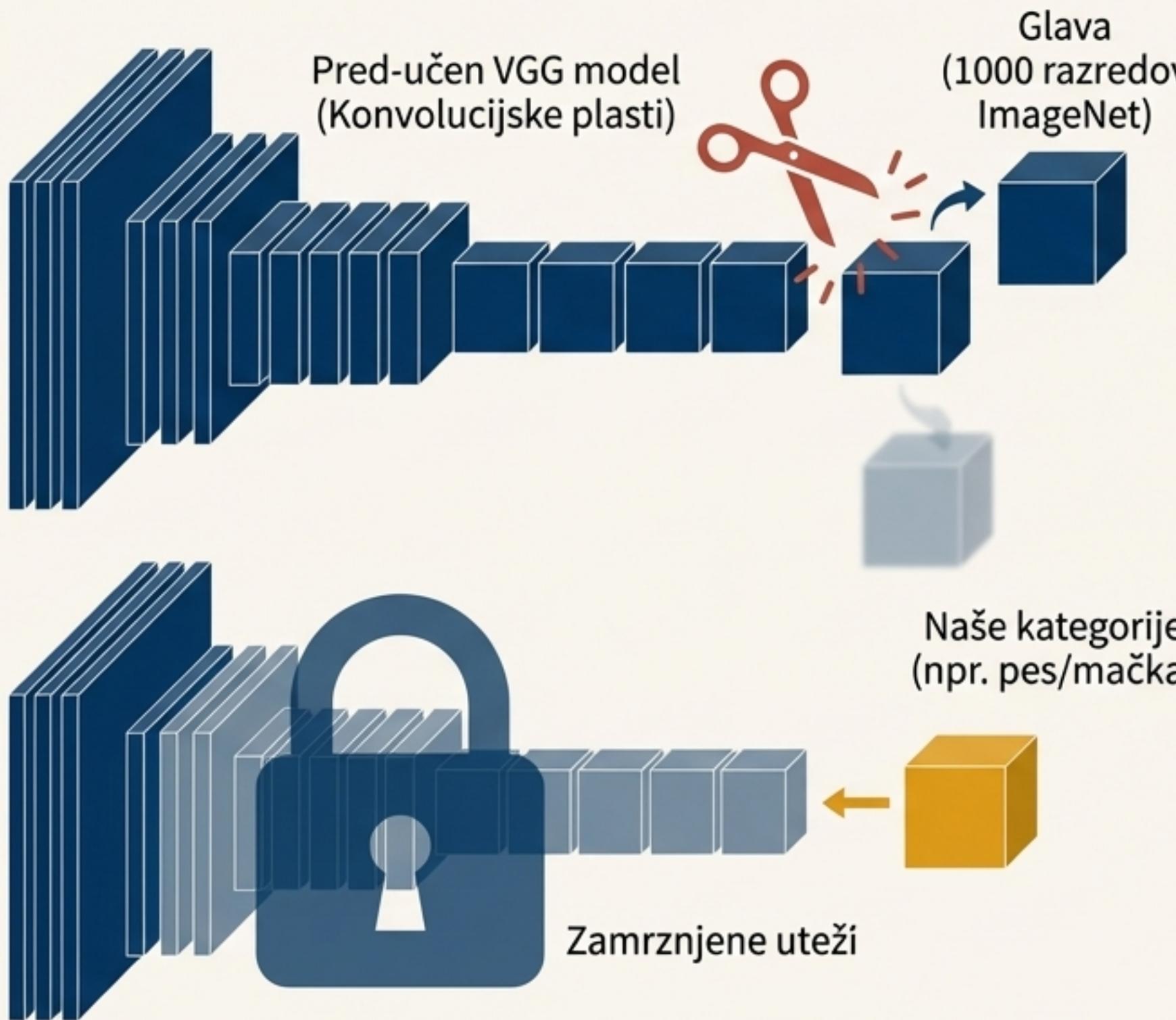
- Izključno 3×3 konvolucijske plasti (korak 1, `padding="same"`).
- Vsakima 2-3 konvolucijskima plastema sledi 2×2 `max pooling` plast.
- Število filtrov se sistematično podvaja: **$64 \rightarrow 128 \rightarrow 256 \rightarrow 512 \rightarrow 512$** .

Gosto povezane plasti (Dense):

- Dve gosti plasti, vsaka s po 4096 nevroni (z ReLU aktivacijo).
- Izhodna plast: `Softmax` plast za klasifikacijo v 1000 kategorij.

Ne začenjajte iz nič: Uporabite moč prenosnega učenja

Znanje modelov, kot je VGG, lahko prenesemo in prilagodimo za reševanje lastnih problemov z veliko manj podatki in računskimi viri.



Postopek prilagoditve VGG v Kerasu:

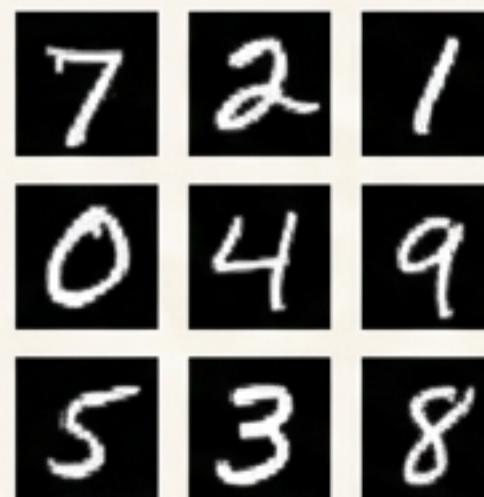
- Uvoz modela:** Uvozimo pred-učen model VGG, brez njegove zadnje klasifikacijske plasti.
- Zamenjava glave:** Odstranimo originalno izhodno plast in jo nadomestimo z novo, prilagojeno našemu številu kategorij.
- Zamrznitev uteži:** ‘Zamrznemo’ uteži v zgodnjih konvolucijskih plasteh, da ohranimo naučene splošne značilnosti.
- Ponovno učenje (Fine-tuning):** Mrežo (predvsem nove in nekaj zadnjih starih plasti) treniramo na našem specifičnem naboru podatkov.

Gorivo za vaše modele: Ključni podatkovni nabori

Za učenje modelov potrebujemo veliko kakovostnih podatkov. Na srečo obstaja veliko javno dostopnih, standardiziranih naborov podatkov.



TensorFlow Datasets (TFDS):
Knjižnica, integrirana v Keras.



MNIST: "Hello, world!" računalniškega vida.
28x28 črno-bele slike ročno napisanih števk.

```
from tensorflow.keras.datasets  
import mnist
```

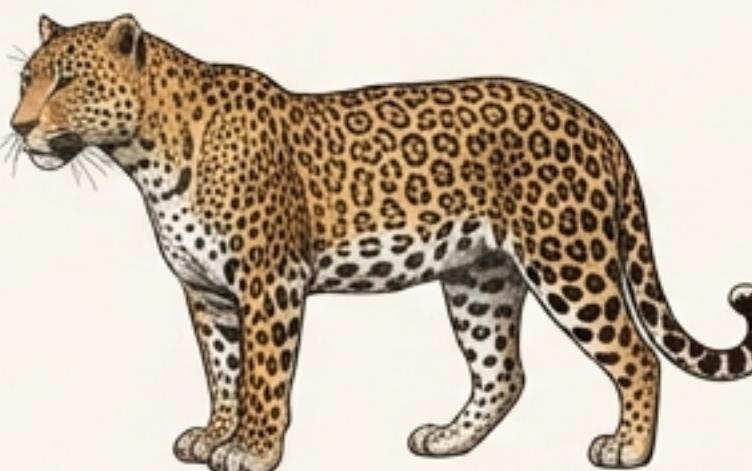


CIFAR-10 & CIFAR-100:
32x32 barvne slike, razvrščene v 10 ali 100 kategorij.

```
from tensorflow.keras.datasets  
import cifar10
```



ImageNet: Ogromen nabor (>14 mio. slik, >20.000 kategorij), ki je standard za primerjavo modelov.



Zahteva registracijo za prenos.



Specifični nabori (promet/vozila)

- **KITTI Dataset:** Zaznavanje vozil in drugih objektov v prometu.
- **Udacity Self-Driving Car Dataset:** Slike avtomobilov, pešcev, prometnih znakov.
- **Vehicle Classification Dataset (Kaggle):** Nabori za klasifikacijo vozil.

Onkraj slik: Prihodnost je v razumevanju konteksta s Transformerji

Medtem ko so CNN-i mojstri prepoznavanja prostorskih vzorcev, so **Transformerji** revolucionirali obdelavo sekvenčnih podatkov. Njihova supermoč je **mehanizem pozornosti**.



Ključna inovacija: Plast pozornosti (Attention Layer):

Ta mehanizem omogoča modelu, da samodejno ugotovi, katere besede v stavku so med seboj pomensko povezane, ne glede na njihovo oddaljenost.

Pomen:

Transformerji so hrbtenica velikih jezikovnih modelov (LLM), kot je ChatGPT (črka 'T' v GPT pomeni Transformer).