

Strojové učení v dálkovém průzkumu Země

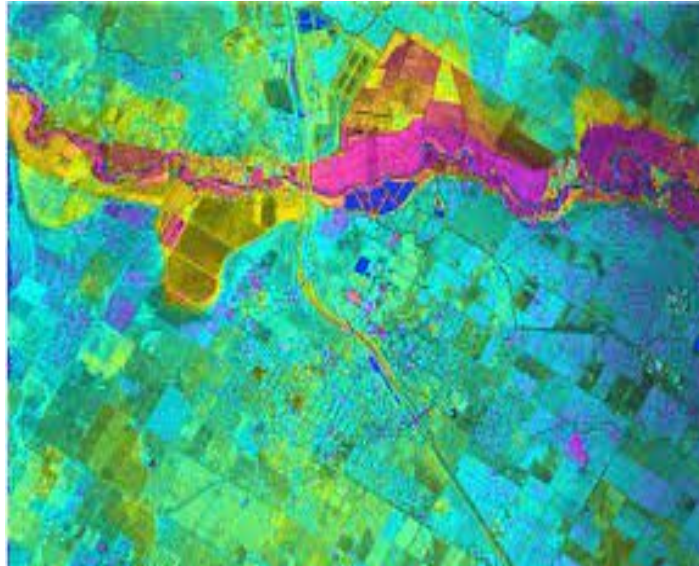
.....

Lekce 10: Hubokého učení: Konvoluční neuronové sítě (CNN)
[princip, komponenty, klasifikace, segmentace, architektury: FCN, ImageNet, SegNet, U-Net]

Přednášející: Lukáš Brodský lukas.brodsky@natur.cuni.cz, Ondřej Pešek
ondrej.pesek@fsv.cvut.cz a Martin Landa martin.landa@fsv.cvut.cz

Prostorová data

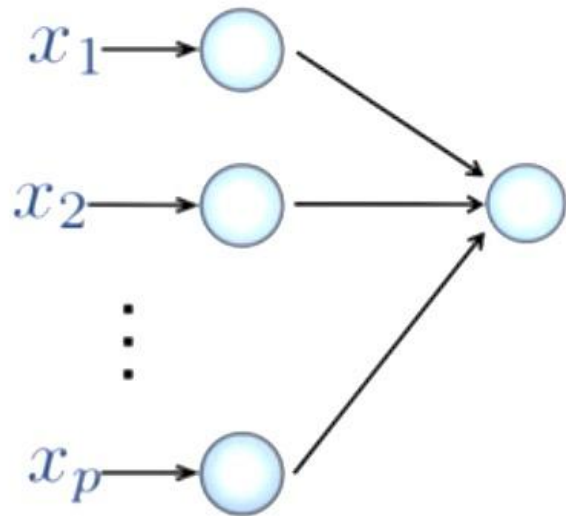
- Rastr



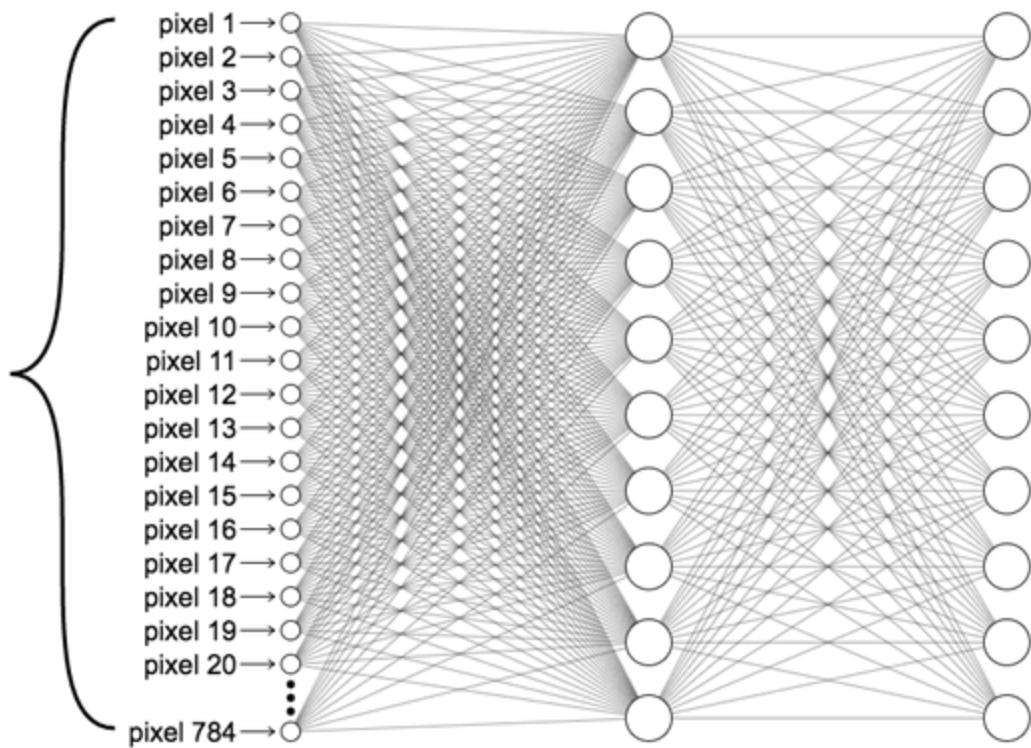
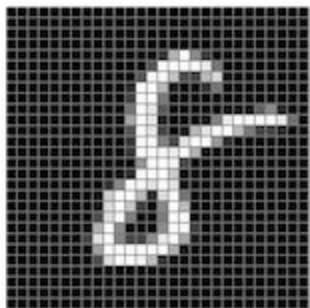
Prostorová data

Vstup: 2D (vektor vstupních hodnot pro všechny pixely)

- Plně propojená neuronová síť
- MLP: každý pixel je příznak (feature)



28x28



Prostorová data

- Přidáním jedné vrstvy přidáte parametry $(size_{l-1} + 1) * size_l$;
- Přidáním vrstvy o 1000 jednotkách do NN přidáte více než 1 milion dalších parametrů;

Konvoluční neuronové sítě

Convolution Neural Networks

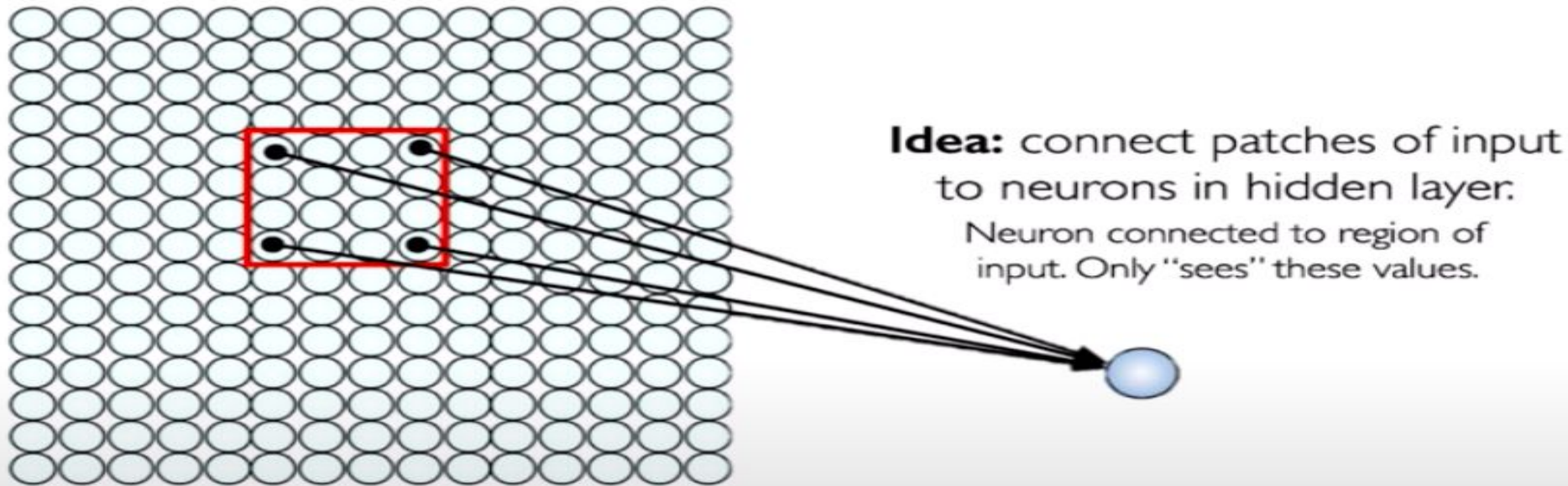
(CNN)

Myšlenka

**Prostorově blízké pixely spolu
pravděpodobně souvisejí a jsou
vzájemně korelovány;**

=> propojte pixely vstupu s neurony ve
skryté vrstvě!

- **Struktura jádra**
- **Jádro (*kernel*): *posuvné okno - spojení;***

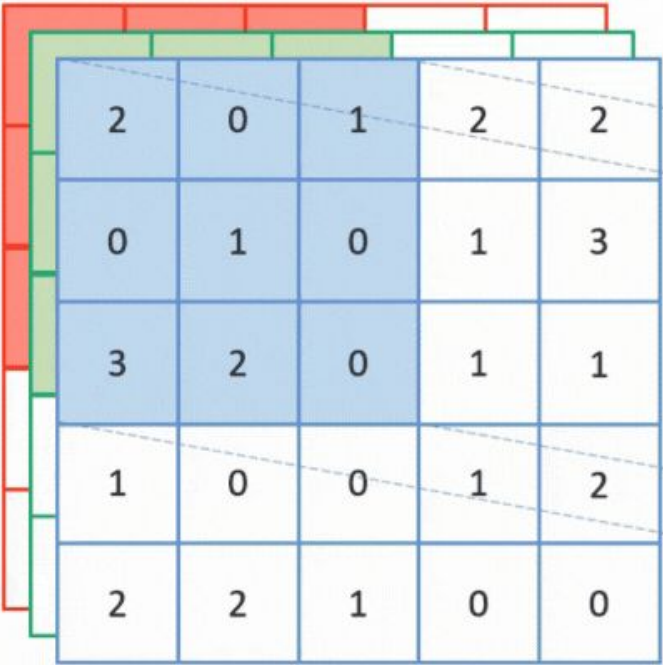


1 _{x1}	1 _{x0}	1 _{x1}	0	0
0 _{x0}	1 _{x1}	1 _{x0}	1	0
0 _{x1}	0 _{x0}	1 _{x1}	1	1
0	0	1	1	0
0	1	1	0	0

Image

4		

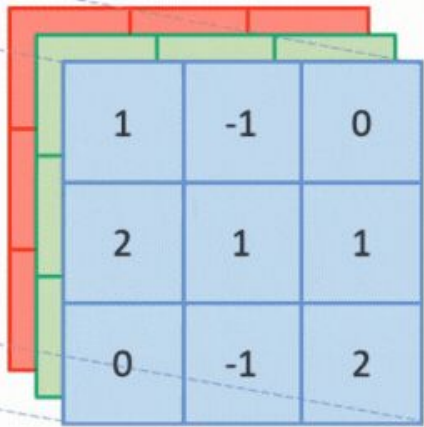
Convolved
Feature



2	0	1	2	2
0	1	0	1	3
3	2	0	1	1
1	0	0	1	2
2	2	1	0	0

Input Image

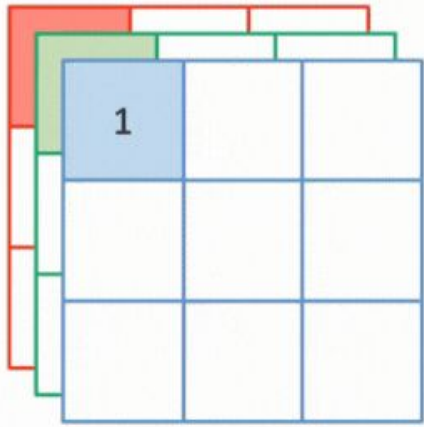
*



1	-1	0
2	1	1
0	-1	2

Filter/Kernel

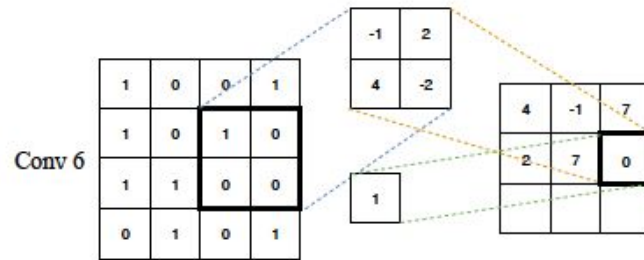
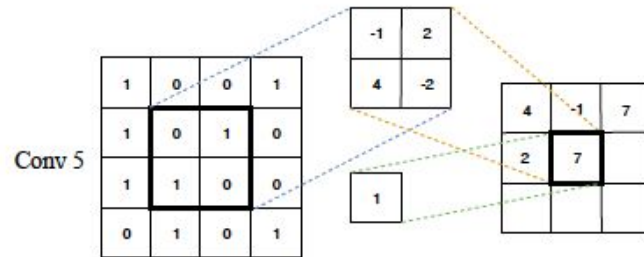
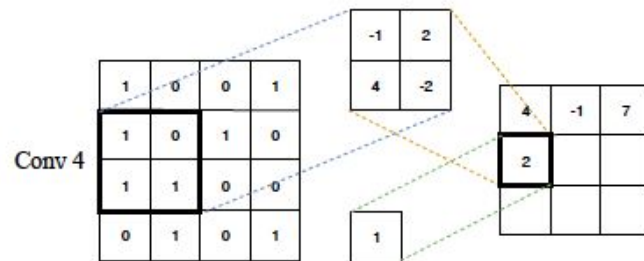
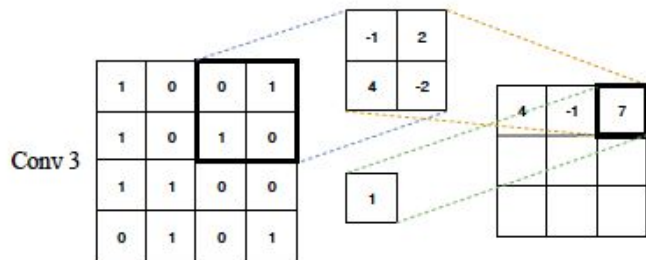
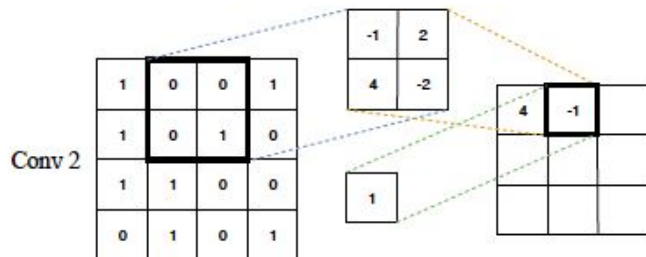
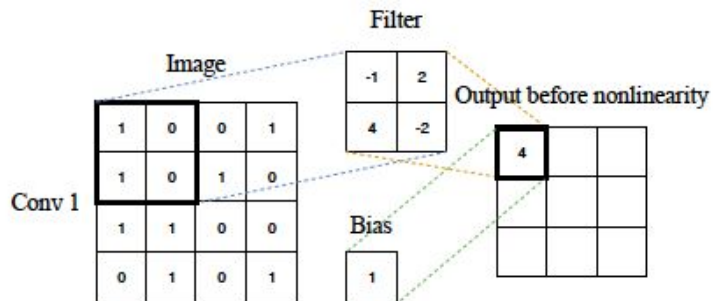
=



1		

Feature Map

Konvolute



Konvoluce



Filter Detecting a Curve Type
Visualization

0	0	0	0	0
40	40	40	0	0
0	0	0	40	0
0	0	0	40	0
0	0	0	0	0

Filter Detecting a Curve Type
Pixel Matrix

*

Element-wise
Multiplication



Receptive Field
Visualization

0	0	0	0	0
40	40	40	0	0
0	0	0	40	0
0	0	0	40	0
0	0	0	0	0

Receptive Field
Pixel Matrix

=

0	0	0	0	0
1600	1600	1600	0	0
0	0	0	1600	0
0	0	0	1600	0
0	0	0	0	0

Σ

Summation

8000

Konvoluce



Filter Detecting a Curve Type
Visualization

0	0	0	0	0
40	40	40	0	0
0	0	0	40	0
0	0	0	40	0
0	0	0	0	0

Filter Detecting a Curve Type
Pixel Matrix

*

Element-wise
Multiplication



Receptive Field
Visualization

0	0	0	0	0
0	0	0	0	0
40	40	40	0	0
0	40	0	0	0
0	0	0	0	0

Receptive Field
Pixel Matrix

=

0	0	0	0	0
0	0	0	0	0
0	0	0	0	0
0	0	0	0	0
0	0	0	0	0

Σ

Summation

0

Konvoluce

- Tradiční konvoluce: ručně navržené filtry
- Hluboké učení: učení filtrů => klíč pro CNN



Original



Sharpen



Edge Detect



"Strong" Edge
Detect

Konvoluce

- Konvoluce je matematická operace, která kombinuje dvě funkce a vytváří třetí funkci. Ve 2D konvoluci je filtrem 2D matice a operace konvoluce je aplikována na 2D vstup, jako je obrázek. Filtr se přesouvá přes obrázek a na každé pozici se vypočítá a sečte součin po jednotlivých prvcích.

Mathematically:

$$(f * g)(x, y) = \sum_i \sum_j f(i, j) \cdot g(x - i, y - j)$$

kde:

- f jsou vstupní obrazová data,
- g je konvoluční jádro,
- $(f*g)(x,y)$ je hodnota konvoluční operace na pozici (x,y) ve výstupní mapě příznaků.

Konvoluce

- . V konvolučních neuronových sítích (CNN) používají konvoluční vrstvy k extrakci funkcí ze vstupních dat naučitelné filtry. Tyto filtry jsou aktualizovány během tréninkového procesu.

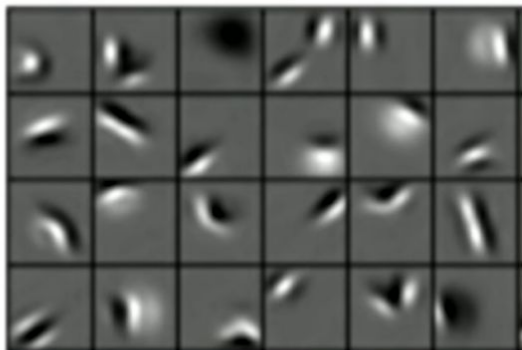
Konvoluční operátor

Operátor konvoluce:

- 1/ použije sadu vah - filtr - k extrakci lokálních rysů;
- 2/ použijte více filtrů k extrakci různých rysů;
- 3/ prostorově sdílí parametry jednotlivých filtrů;

Naučené příznkay

Low level features



Edges, dark spots

Mid level features



Eyes, ears, nose

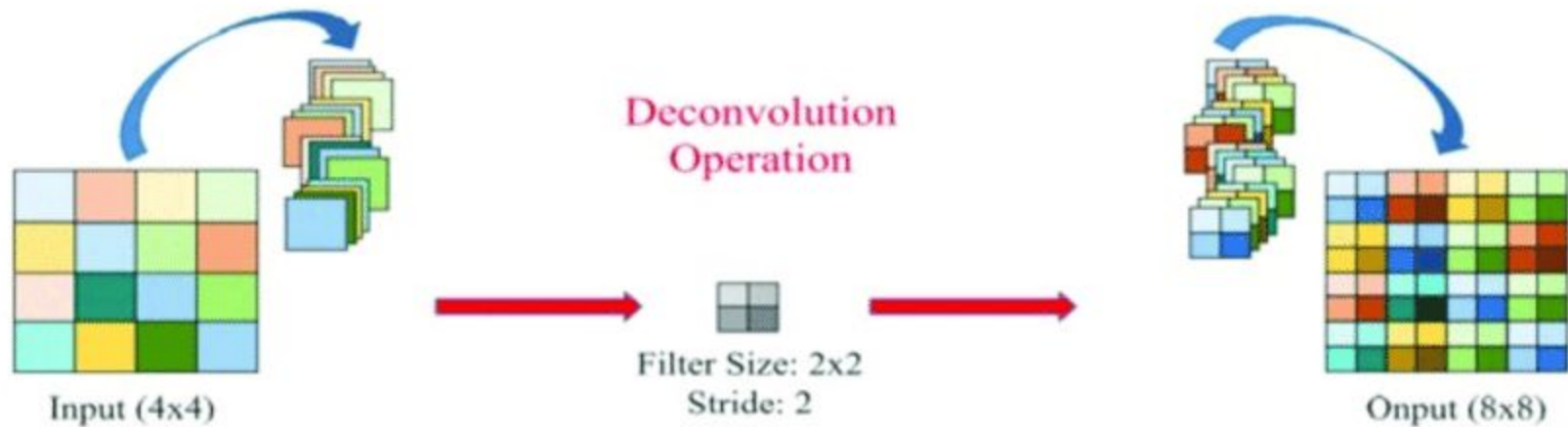
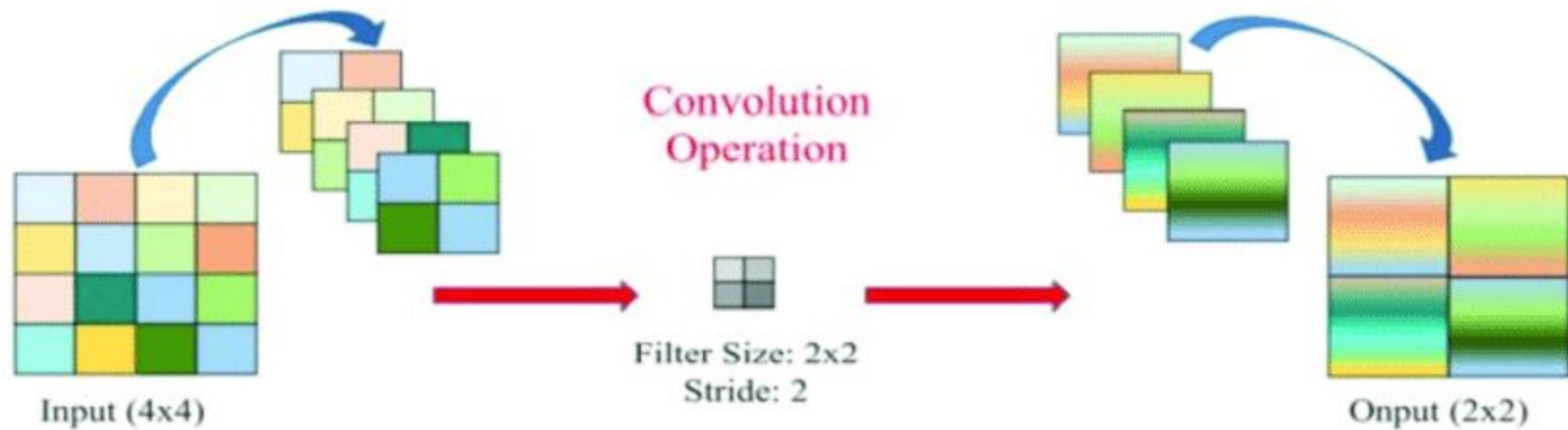
High level features



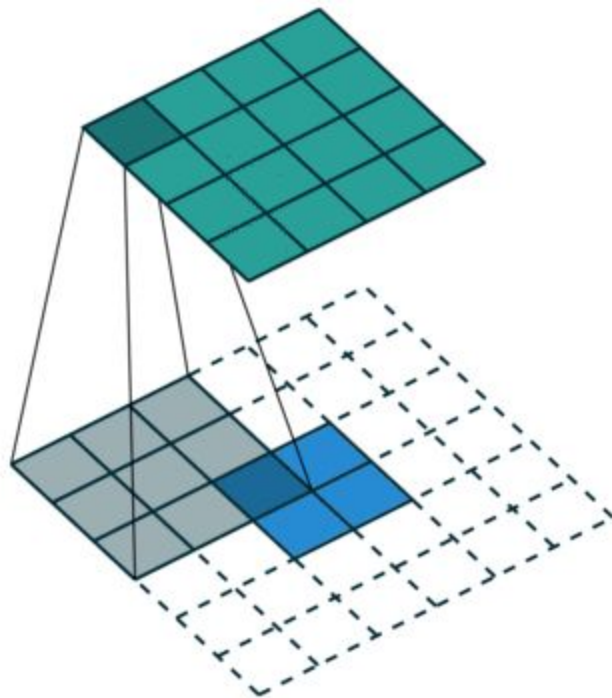
Facial structure

Dekonvoluce

- Dekonvoluce je výpočetní proces, jehož cílem je **invertovat konvoluční operaci**, rekonstruovat původní signál nebo obraz z jeho konvolvované reprezentace.
- Konvoluce zahrnuje matematickou integraci vstupního signálu s filtrem nebo jádrem, čímž vzniká transformovaný výstup. Dekonvoluce se používá k obrácení tohoto procesu konvoluce, zrušení účinků konvoluce a obnovení původního signálu nebo obrazu.



Dekonvoluce



Výplň (Padding)

Výplň

- Výplň v CNN udržuje během konvolucí konzistentní velikost vstupních dat, čímž zabraňuje ztrátě dat na okrajích.
- Tím se zachovávají prostorové detaily, které jsou klíčové pro přesnou reprezentaci rysů.
- Rovněž zabraňuje okrajovým efektům tím, že přidává kontext k okrajovým pixelům.

Nulová výplň

výstupní velikost < vstupní velikost

Stride = 1	Convolution Kernel = 3 x 3	Output Size (2x2) < Input Size (4x4)
VALID PADDING (Zero Padding)		

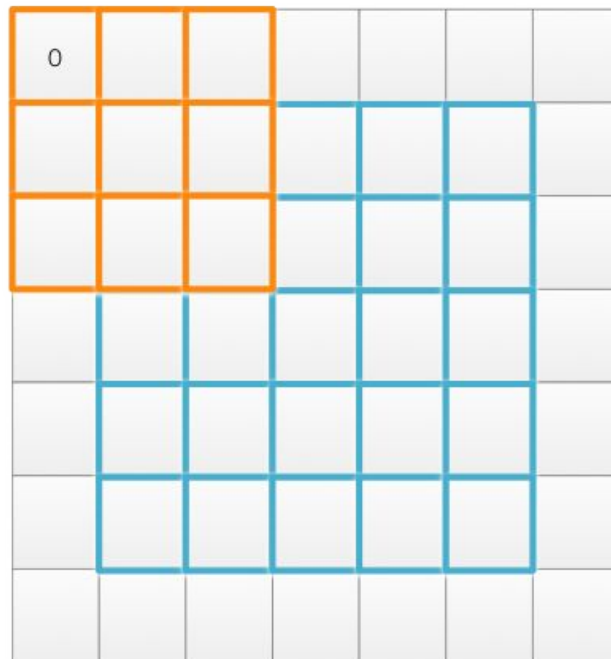
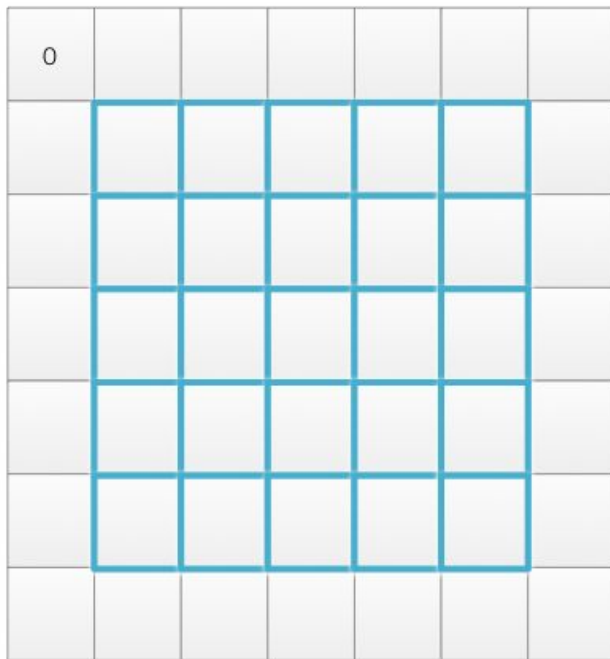
Kernel		
1	1	1
0	0	0
0	0	0

Input			
2	2	0	1
4	5	1	2
0	0	4	0
0	1	0	6



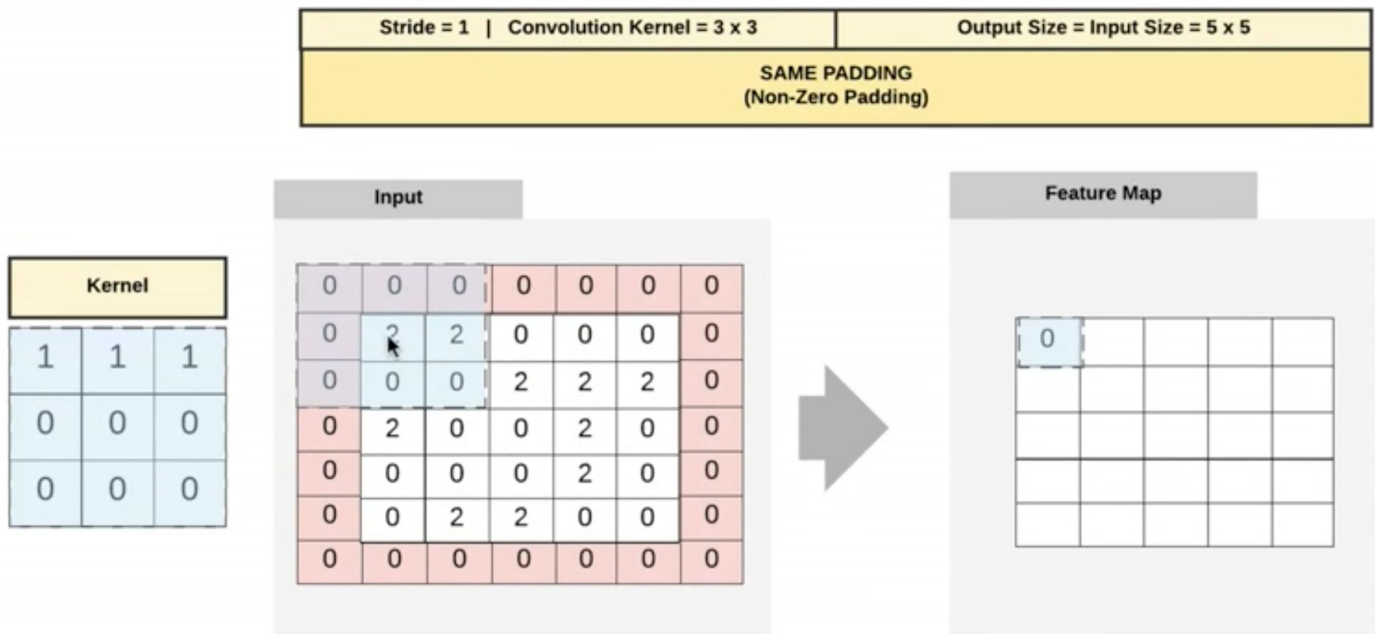
Feature Map	
4	

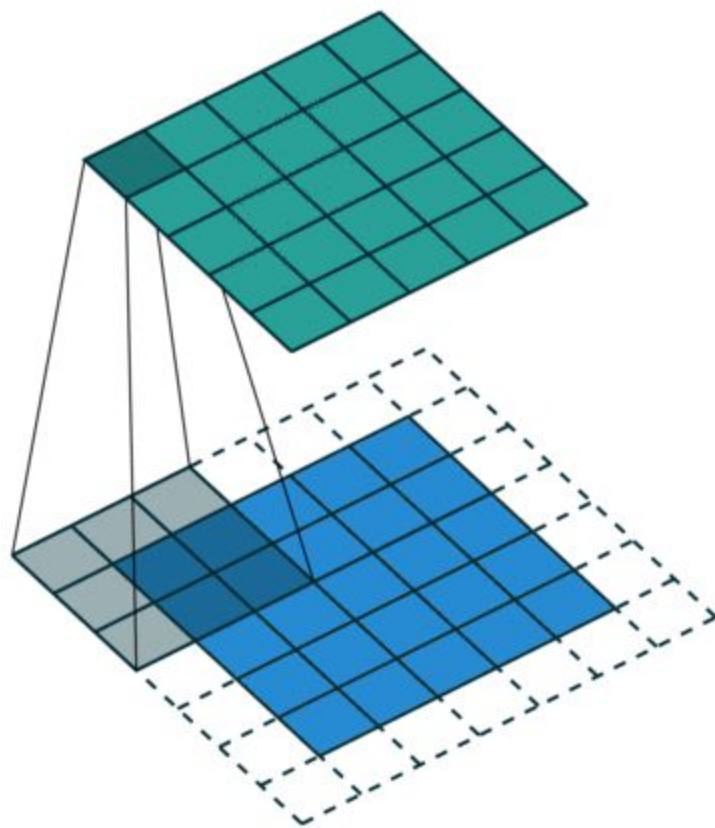
Přidání nenulové výplně



Přidání nenulové výplně

velikost vstupu = velikost výstupu



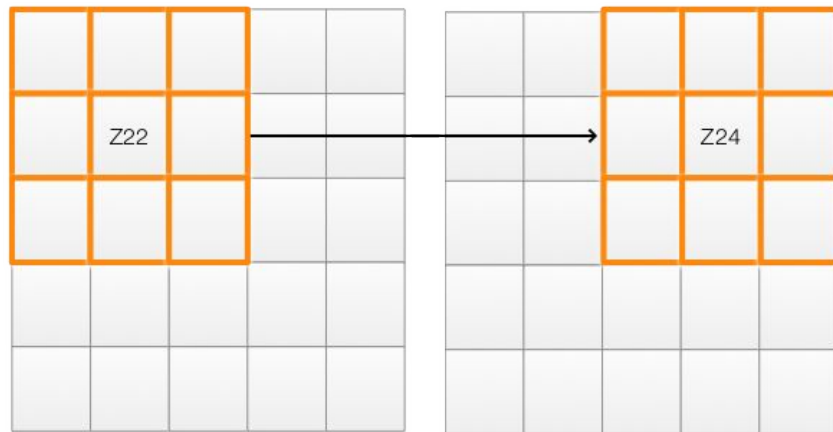


Krok (Stride)

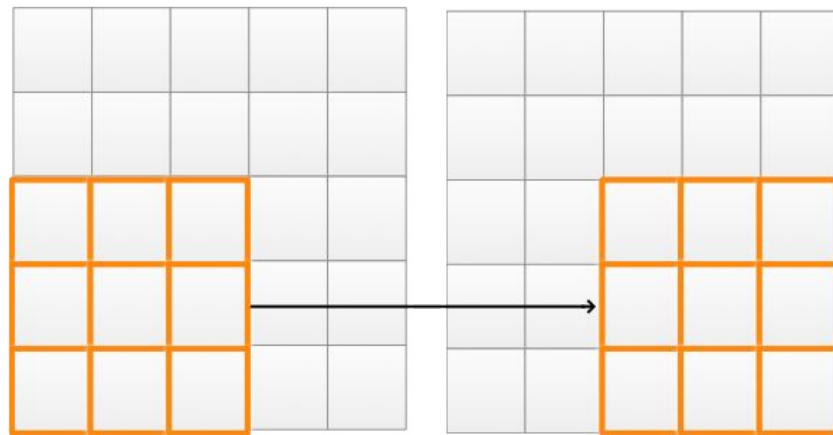
Krok

- Počet pixelů, které filtr najednou přesune přes vstupní obrázek
- Větší krok má za následek menší výstupní obraz.

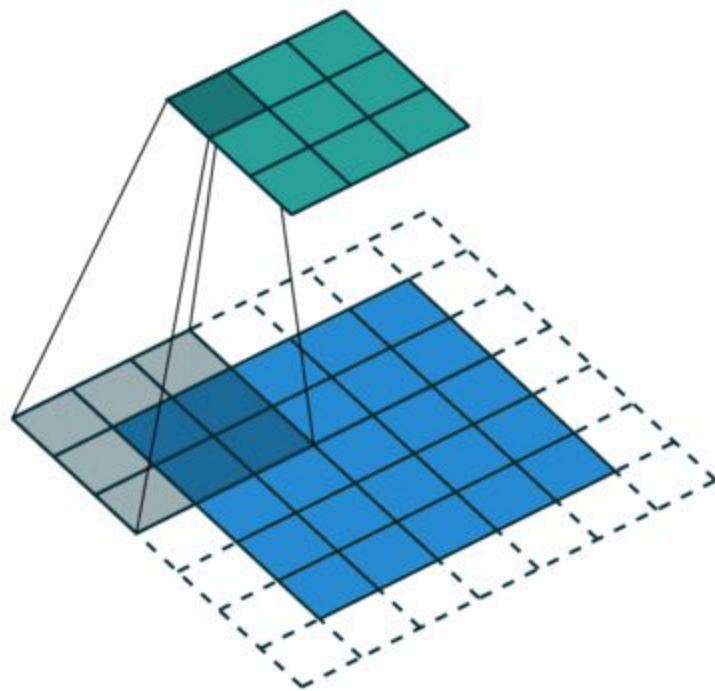
Krok



Y stride = 2



X stride = 2



Výpočty rozměrů

W ... Vstupní výška / délka

O ... Výstupní výška / délka

K ... Velikost filtru (jádro)

P ... Padding

S ... Krok

Výstupní rozměry:

$$O = ((W - K + 2P) / S) + 1$$

Jak velká výplň je zapotřebí pro zachování rozměrů?

$$P = (K - 1) / 2$$

Sdružování/spojování (Pooling)

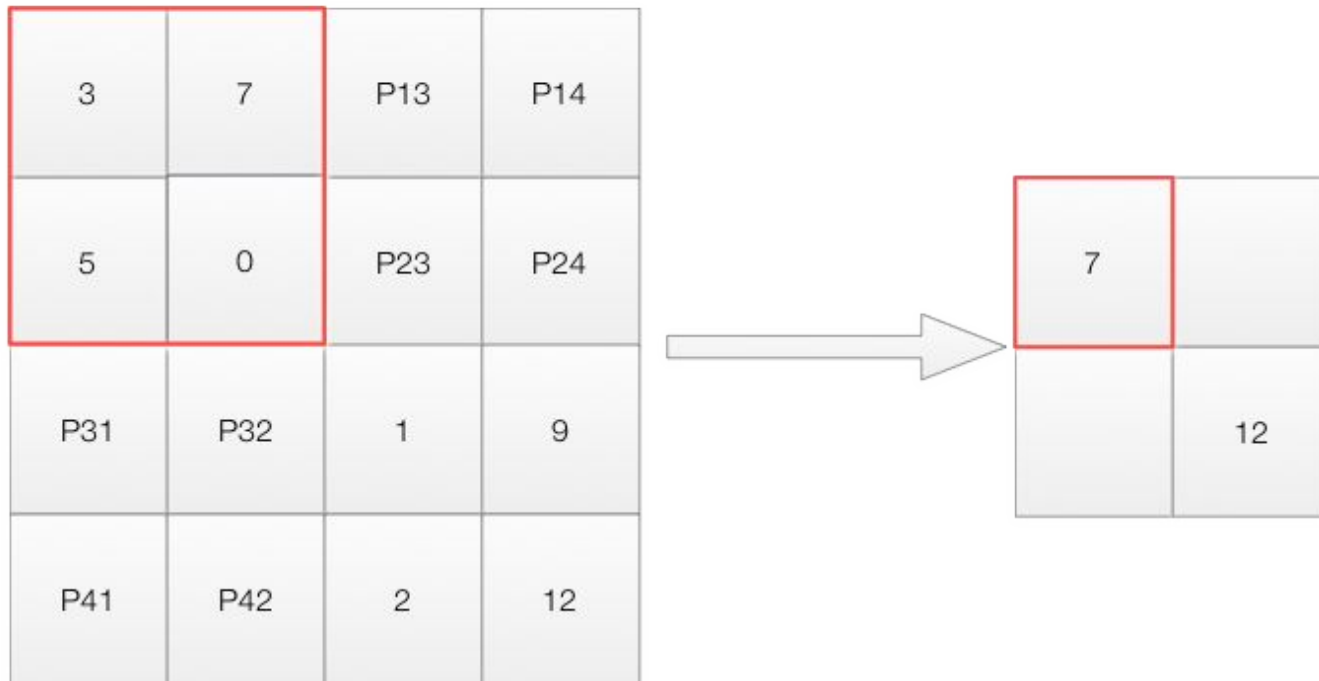
Sdružování

- Sdružující vrstvy zmenšují prostorové rozměry (šířku a výšku) vstupního objemu, čímž snižují počet parametrů a výpočetní složitost sítě.

Sdružování

1. Sdružování nejvyšší hodnotou (**Max Pooling**): okno (obvykle 2x2) se posouvá po vstupní mapě prvků a na každé pozici zachovává maximální hodnotu v rámci okna. Tato operace účinně **snižuje počet vzorků vstupní mapy, přičemž zachovává nejdůležitější prvky a vyřazuje ty méně relevantní.**
2. Sdružování průměrnou hodnotou (Average Pooling): vypočítá průměr všech hodnot v rámci okna. To může být někdy užitečné, **pokud chcete dosáhnout plynulejšího efektu převzorkování.**

Sdružování nejvyšší hodnotou (MaxPooling)



Dávková normalizace (Batch Normalization)

Dávková normalizace

- Dávková normalizace: zlepšení stability učení a urychlení konvergence.
- Zahrnuje normalizaci vrstvy v mini-dávce, obvykle před použitím aktivační funkce.
- Normalizace se provádí odečtením průměru dávky a vydělením směrodatnou odchylkou dávky, poté se normalizované aktivace škálují a posouvají pomocí naučitelných parametrů (gama a beta).

Batch Normalization

$$y = \gamma \frac{x - \mu}{\sqrt{\sigma^2 + \epsilon}} + \beta$$

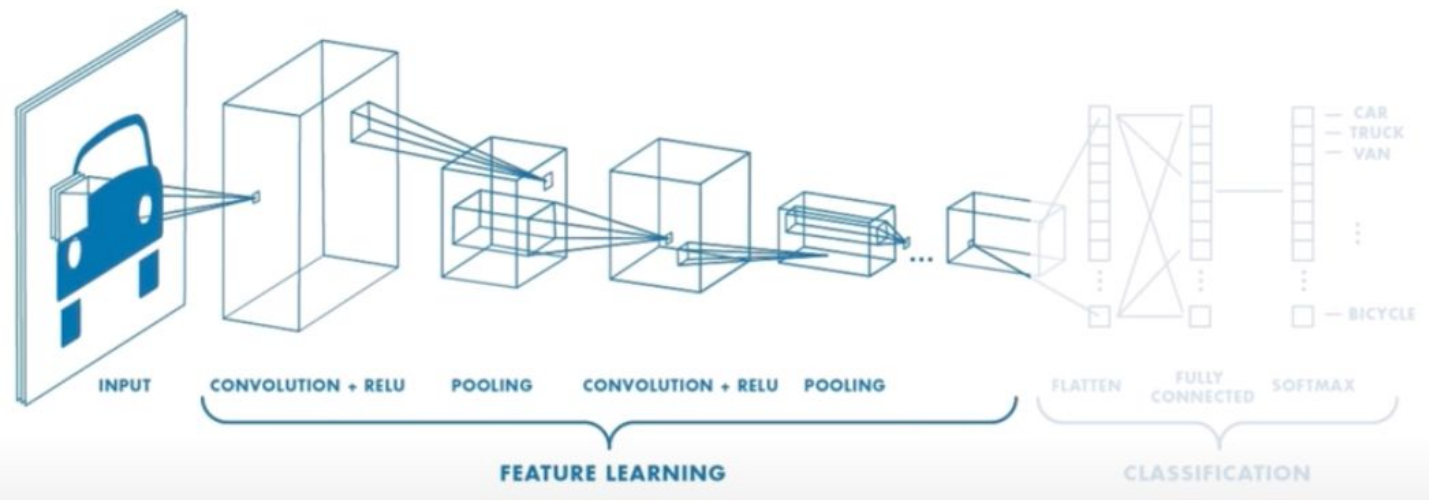
Kde:

- μ je střední hodnota dávky
- σ je směrodatná odchylka dávky
- γ je měřítko, které škáluje normalizovaný výstup (učitelný parametr)
- β je posun, který posouvá normalizovaný výstup (učitelný parametr)
- ϵ je malá konstanta (obvykle 10^{-5}) přidaná do jmenovatele kvůli numerické stabilitě
- x je vstup
- y je normalizovaný výstup

CNN architektury

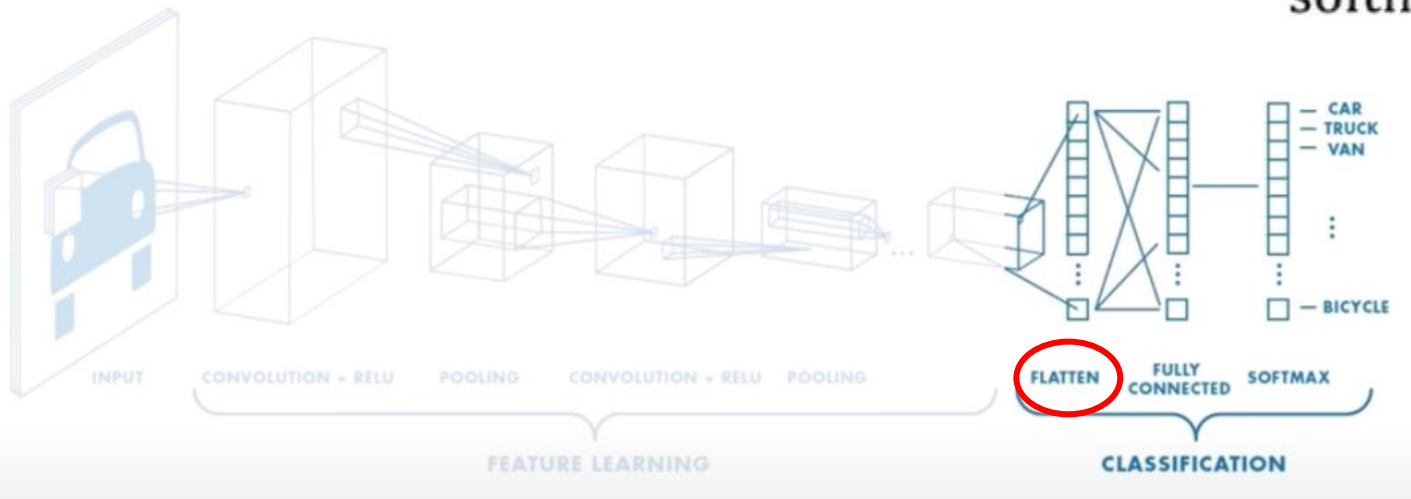
Klasifikační CNN

CNN klasifikační architektura



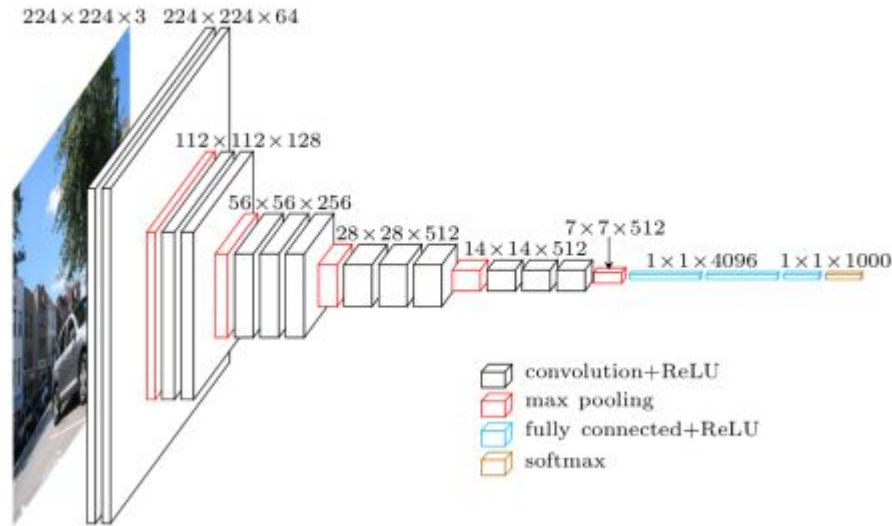
CNN klasifikační architektura

$$\text{softmax}(y_i) = \frac{e^{y_i}}{\sum_j e^{y_j}}$$



- Plně propojená vrstva využívá zvolené funkce pro klasifikaci
- Výstup: Pravděpodobnost zařazení obrazu do určité třídy

VGG (Visual Geometry Group)



- Dosáhl nejlepších výsledků v ILSVRC (ImageNet Large Scale Visual Recognition Challenge) 2014
- <https://www.cs.toronto.edu/~frossard/post/vgg16/>

Segmentační CNN

Segmentační CNN

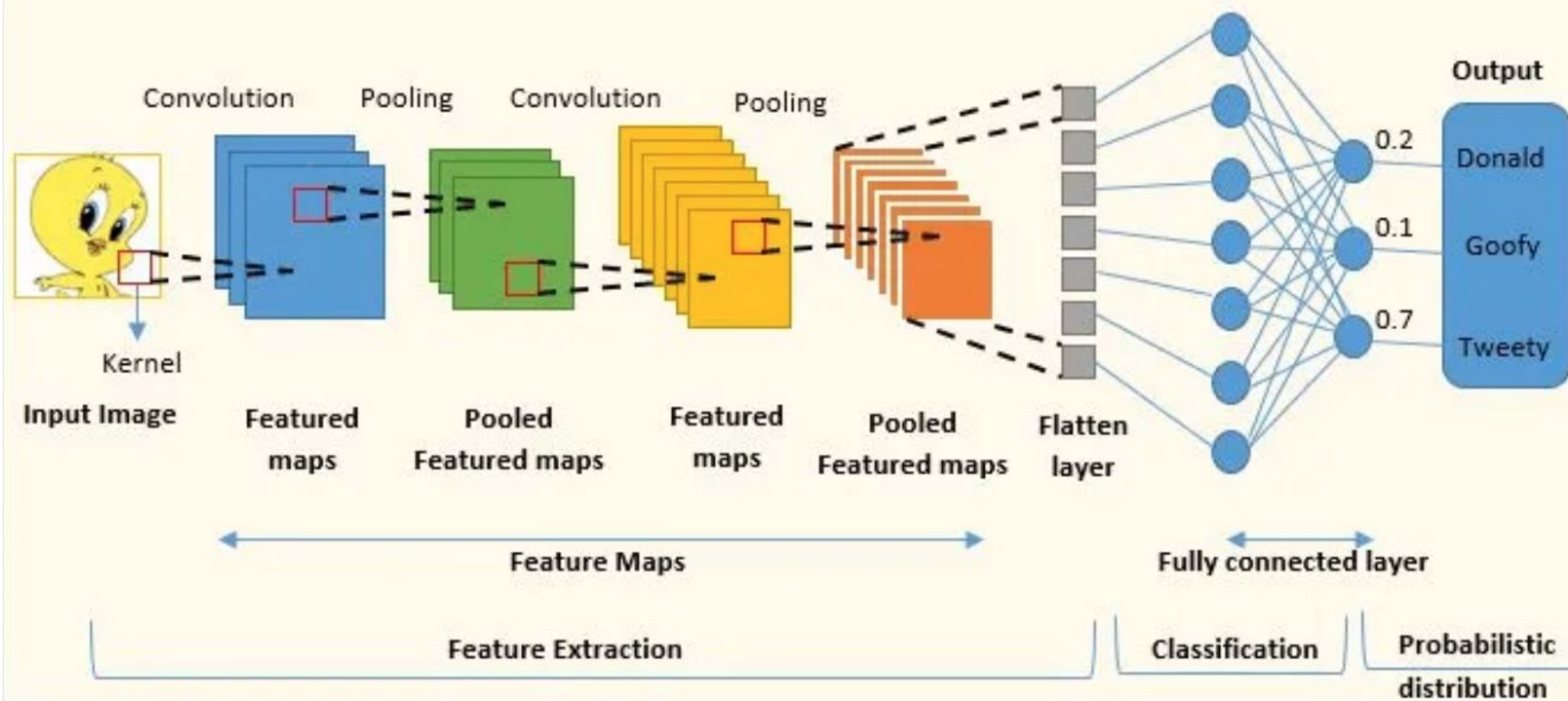
- Segmentace: proces rozdělení obrazu na více segmentů nebo oblastí na základě určitých charakteristik, jako je radiometrie (barva), textura nebo hranice.
- Úlohy segmentace zahrnují přiřazení kódu každému pixelu, čímž se vytvoří segmentační mapa vstupního obrazu v pixelech.

Segmentační CNN

Komponenty CNN:

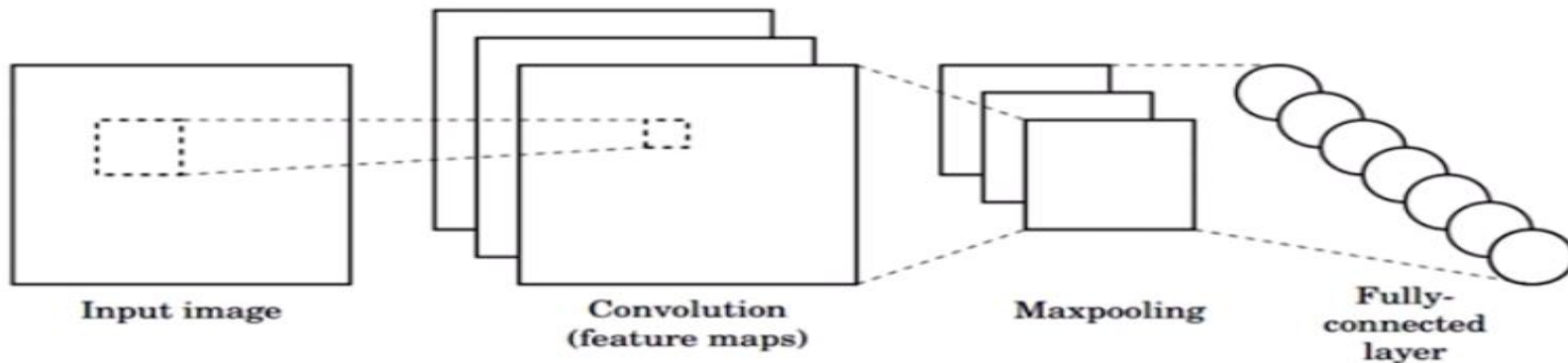
- Kódovač/kodér (encoder): skládá se z několika konvolučních vrstev, po nichž následují operace sdružování
 - extrahuje ze vstupního obrazu hierarchické příznaky, přičemž postupně snižuje prostorové rozměry a zároveň zvyšuje počet map příznaků
- Dekodér (decoder): skládá se z operací inverzních sdružování (unpooling), po nichž následují konvoluční vrstvy
 - Přebírá příznaky extrahované kodérem a převzorkovává je na původní vstupní rozlišení, přičemž snižuje počet map příznaků.
 - cíl -> vygenerovat segmentační mapu v pixelech, která odpovídá rozměrům vstupního obrazu.

A Typical Convolutional Neural Network (CNN)



FCN

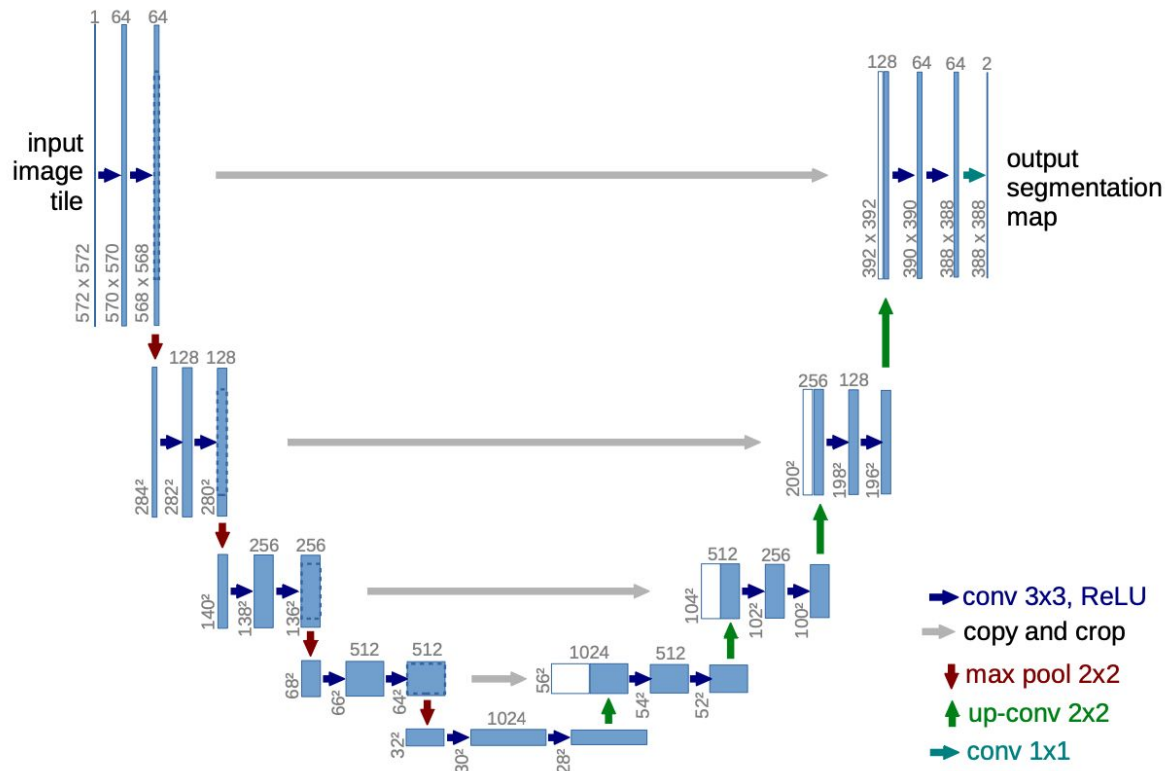
1. Konvoluce: použití filtrů pro generování map prvků
2. Nelinearita: ReLU
3. Sdružování: zmenšování vzorků (zmenšování velikosti)



U-Net

- **Kodér (kontrakční cesta):** několik konvolučních vrstev, po nichž následuje maximálně operace sdružování. Operace max pooling snižují vzorkování map prvků, což pomáhá zachytit stále abstraktnější prvky.
- **Dekodér (expanzivní cesta):** Dekódovací část sítě U-Net se skládá z vrstev pro převzorkování, po nichž následují konvoluční vrstvy. Přebírá rysy extrahované kodérem a převzorkovává je na původní vstupní rozlišení, přičemž snižuje počet map rysů.
- **Přeskokové spojení (*skip connection*):** Přeskokové spoje umožňují síti zachovat prostorové informace a zachytit místní i globální kontext během procesu segmentace. Přímým propojením kodéru a dekodéru ve více rozlišeních síť U-Net usnadňuje tok informací napříč různými úrovněmi abstrakce, což pomáhá zlepšit přesnost segmentace.
- **Závěrečná vrstva:** Na konci sítě se používá závěrečná konvoluční vrstva s aktivační funkcí softmax, která pro každý pixel vypočítá rozdělení pravděpodobnosti v různých třídách.

U-Net Ronneberger et al. 2015



SegNet

- **Kodér:** několik konvolučních vrstev, po nichž následují operace sdružování nejvyšší hodnotou. Tyto vrstvy extrahují ze vstupního obrazu hierarchické rysy, přičemž postupně snižují prostorové rozměry a zároveň zvyšují počet map rysů. Úkolem kodéru je zachytit ze vstupního obrazu jak nízkoúrovňové, tak vysokoúrovňové rysy.
- **Dekodér:** vrstvy s převzorkováním, po nichž následují konvoluční vrstvy. Přebírá rysy extrahované kodérem a převzorkuje je na původní vstupní rozlišení při současném snížení počtu map rysů. Unpooling SegNetu je založen na indexech operace maximálního poolingů provedené v kodéru. Tyto indexy jsou uloženy během operace poolingů a později jsou použity pro unpooling v dekodéru, což umožňuje síti zachovat prostorovou informaci.
- Vrstva **Softmax**: slouží k výpočtu rozdělení pravděpodobnosti nad různými třídami pro každý pixel.

SegNet: Architektura kodéru a dekodéru

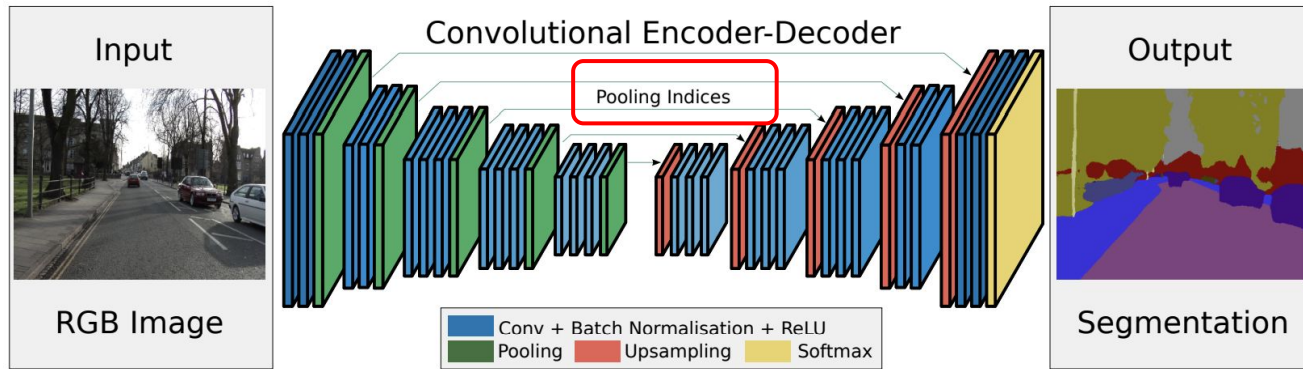


Fig. 2. An illustration of the SegNet architecture. There are no fully connected layers and hence it is only convolutional. A decoder upsamples its input using the transferred pool indices from its encoder to produce a sparse feature map(s). It then performs convolution with a trainable filter bank to densify the feature map. The final decoder output feature maps are fed to a soft-max classifier for pixel-wise classification.

Badrinarayanan et al. 2016: SegNet: A Deep Convolutional Encoder-Decoder Architecture for Image Segmentation, Computer Vision and Pattern Recognition, <https://arxiv.org/abs/1511.00561v3>

CNN

- + prostorový kontext v obraze
- + lepší než algoritmy ML pro analýzu obrazu
- výpočetně náročné
- vyžaduje *a priori* tréninková data plně prostorová (bez mezer)

Otázky?