Zpracování digitalizovaného obrazu (ZDO) - Klasifikace

Přehled klasifikačních metod

Ing. Zdeněk Krňoul, Ph.D.

Katedra Kybernetiky
Fakulta aplikovaných věd
Západočeská univerzita v Plzni

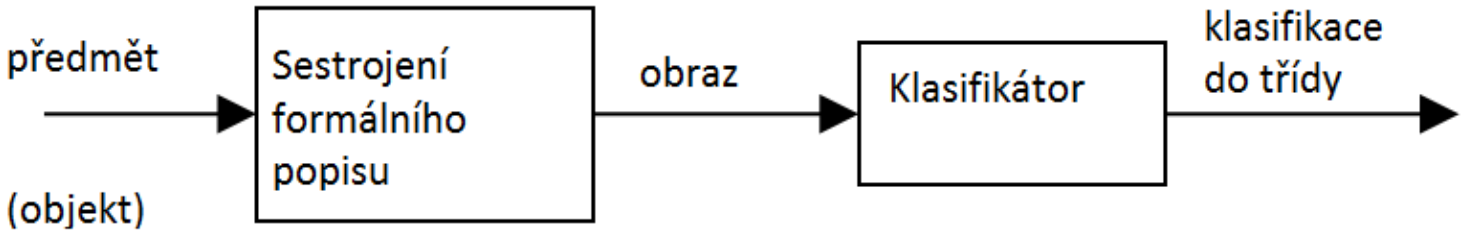
DEPARTMENT OF CYBERNETICS

Obsah:

- klasické přístupy
- DNN (CNN)

Klasifikace

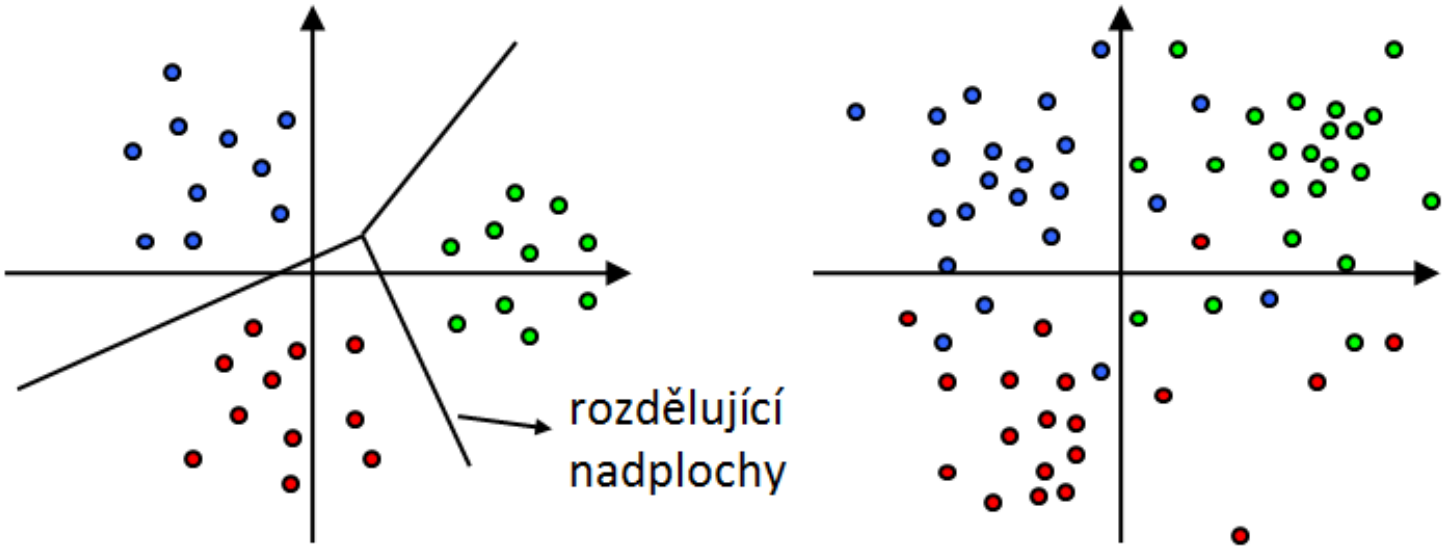
- Rozpoznávání (klasifikace, angl. Pattern recognition) – zařazování předmětů do tříd
- Klasifikátor nerozeznává objekty, nýbrž jejich obrazy (popisy)
- PŘÍZNAKOVÉ ROZPOZNÁVÁNÍ
- STRUKTURÁLNÍ (SYNTAKTICKÉ) ROZPOZNÁVÁNÍ



DEPARTMENT OF CYBERNETICS

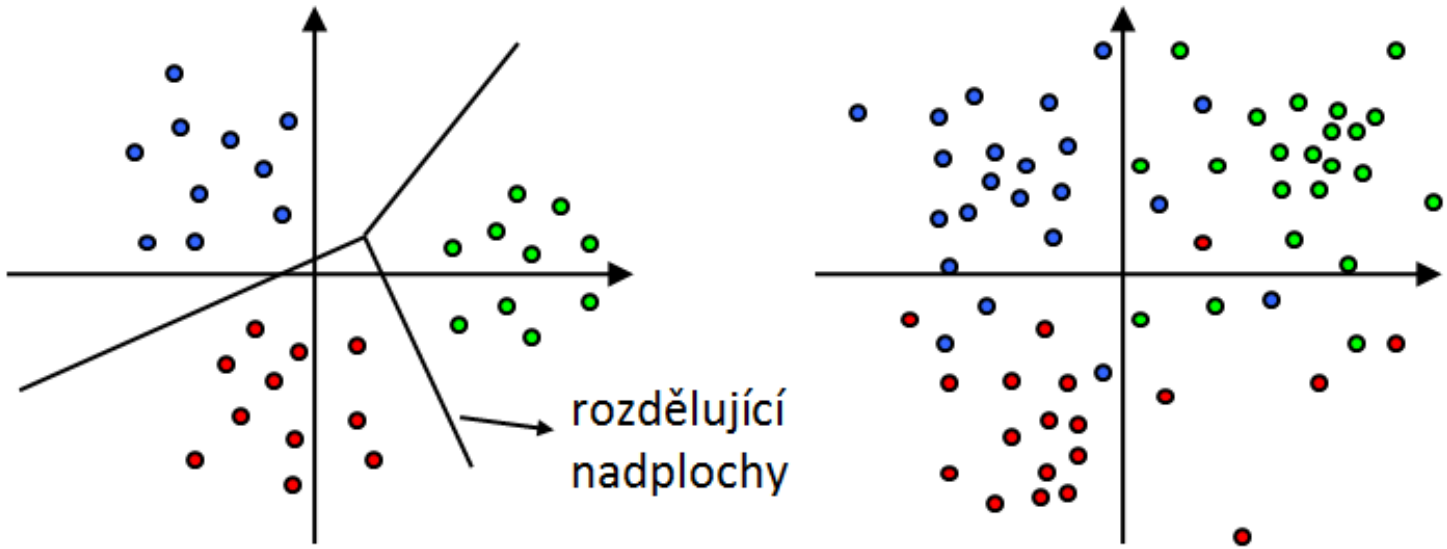
PŘÍZNAKOVÉ ROZPOZNÁVÁNÍ

- obrazy jsou charakterizovány vektorem, jehož souřadnice tvoří hodnoty jednotlivých příznaků.
- množina všech možných obrazů vytváří n-rozměrný obrazový prostor.
- při vhodném výběru příznaků je podobnost předmětů v každé třídě vyjádřena geometrickou blízkostí jejich obrazů



DEPARTMENT OF CYBERNETICS

- pokud lze obrazy jednotlivých tříd (různých) od sebe oddělit rozdělující nadplochou $\rightarrow$ (mluvíme o separabilních množinách obrazů),
- úloha klasifikace je relativně jednoduchá a lze očekávat bezchybnou klasifikaci
- ve valné většině případů však množiny obrazů nejsou stoprocentně separabilní a část předmětů bude vždy chybně klasifikována, viz obrázek vpravo



DEPARTMENT OF CYBERNETICS

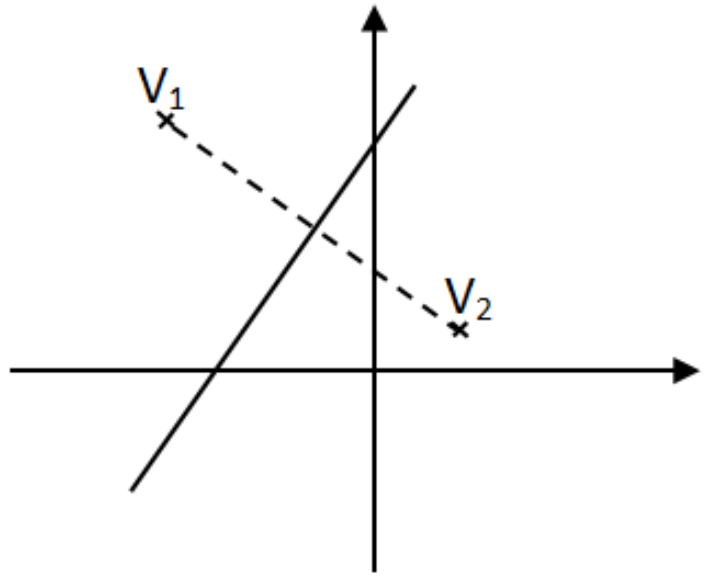
Učení s učitelem - Supervised Learning

- Základem je trénovací množina vzorových obrazů
- $+$ u každého je uvedeno zařazení do správné třídy.
- na základě této trénovací množiny je pak určena reprezentace tříd,
- např. pro metodu **minimální vzdálenosti**

DEPARTMENT OF CYBERNETICS

Metoda minimální vzdálenosti

- metodou minimální vzdálenosti jsou vypočteny centroidy (průměry) vzorových obrazů pro jednotlivé třídy.
- každá třída je reprezentována jedním vzorovým obrazem
- tento obraz lze vypočítat např. průměrem všech vzorových obrazů dané třídy



DEPARTMENT OF CYBERNETICS

Klasifikátor KNN (K-Nearest Neighbours)

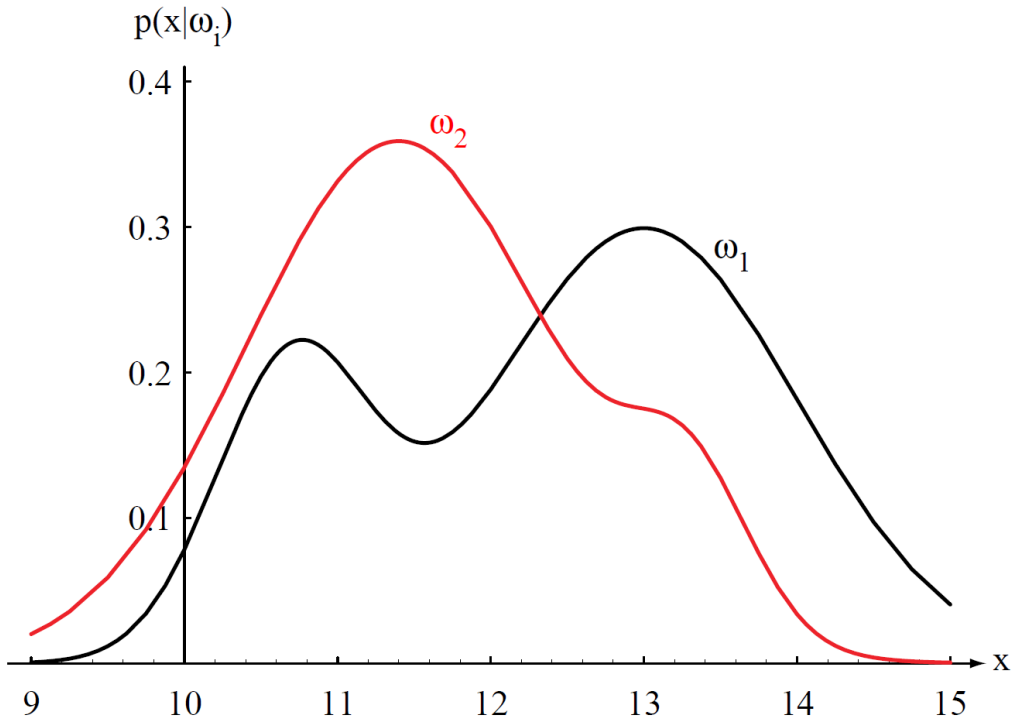
- each image is matched with all images in training data
- top K with minimum distances are selected
- majority class of those top K is predicted as output class of the image
- various distance metrics can be used like
  - L1 distance (sum of absolute distance) $E = \sum_{i=1}^{N} |y_i - f(x_i)|$,
  - L2 distance (sum of squares) $E = \sum_{i=1}^{N} (y_i - f(x_i))^2$,
  - etc.

DEPARTMENT OF CYBERNETICS

Bayesův klasifikátor

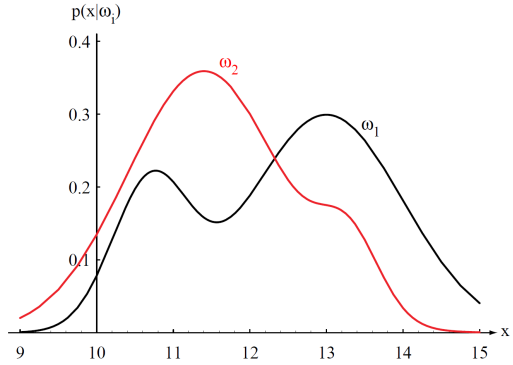
- základní statistický přístup ke klasifikaci založený na Bayesovu teorému (Thomas Bayes 1701–1761)
- patří do množiny tzv. pravděpodobnostních klasifikátorů, tedy vyčíslení kompromisu mezi různými klasifikačními rozhodnutími za využití pravděpodobnosti
- princip je znám od 50-tých a 60-tých let minulého století, kdy byl používán na úlohu vyhledávání informací v textu (information retrieval), kategorizaci textu, vyhodnocení diagnoz v lékařství aj.
- stále je považován za alternativu k více sofistikovanějším klasifikačním metodám, jako je např. SVM (support vector machine)

- ze statistického pohledu Bayesův klasifikátor minimalizuje pravděpodobnost chybné klasifikace,
- princip založen na podmíněné distribuci $P(Y|X)$), kde Y je konečná množina možných tříd a X je vstup (měřené příznaky).
- potom *rozhodnutí* pro nejlepší třídu pro námi naměřené X je dáno $\hat{y} = \underset{y}{\operatorname{argmax}}\ P(Y = y|X)$
- trénování Bayesova klasifikátoru ... často metoda maximální věrohodnosti (maximum likehood, MLE)
- pokud jsou proměnné X nezávislé, pak je potřeba určit celkem **málo parametrů**, např. x se řídí normálním (Gausovským) rozdělením $\rightarrow$ statistický popis jednotlivých tříd je střední hodnota a rozptyl (není potřeba plná kovarianční matice)



- třídy ω_1 a ω_2
- $p(x|\omega_j)$ ... značí podmíněné závislosti vzorku,
- x ... spojitá náhodná veličina, jejíž distribuce závisí na *přirozeném stavu*
- dále zavedeme apriorní pravděpodobnost, která předurčuje příslušnost vzorku k dané přídě ... tedy pro třídu j označíme jako $P(\omega_j)$,

DEPARTMENT OF CYBERNETICS



Trénování:

- pro dvě třídy hledáme $p(x|\omega_1)$ a $p(x|\omega_2)$ pro podmíněnou hustotu pravděpodobnosti třídy ω_1 resp. ω_2
- dále dvě apriorní pravděpodobnost $P(\omega_1)$, která předurčuje, že vzorek patří do třídy ω_1, resp. $P(\omega_2)$ pro třídu ω_2 ... neexistují žádné další třídy, proto $P(\omega_1) + P(\omega_2) = 1$
- rozdíl mezi $p(x|\omega_1)$ a $p(x|\omega_2)$ popisuje rozdíl mezi jednotlivými třídami

Sdružená hustota pravděpodobnosti $p(\omega_j, x)$

- sdružená hustota pravděpodobnosti popisuje, že vzorek patří do třídy ω_j a má hodnotu příznaku x
- může být zapsáno dvěma způsoby:
  $p(\omega_j, x) = P(\omega_j|x)p(x) = p(x|\omega_j)P(\omega_j)$,
- přeuspořádáním tohoto vztahu dostaneme Bayesův vztah:

$$P(\omega_j|x) = \frac{p(x|\omega_j)P(\omega_j)}{p(x)}, \tag{1}$$

  kde $p(x) = \sum_j p(x|\omega_j)P(\omega_j)$
- jde o Bayesův vztah, který může být vyjádřen slovně jako:

$$\text{posterior} = \frac{\text{likelihood} \times \text{prior}}{\text{evidence}} \tag{2}$$

DEPARTMENT OF CYBERNETICS

Učení bez učitele - shluková analýza

- u trénovací množiny není udána informace o příslušnosti obrazů k třídám
- snahou je rozdělit obrazy do k tříd tak, aby byla minimalizována hodnota kritéria optimality
- často globální minimum nelze z výpočetních důvodů nalézt $\rightarrow$ přijatelné lokální minimum
- Metody lze rozdělit na:
  1. hierarchické – vytvářejí shlukovací strom:
     aglomerativní – vycházíme od jednotlivých obrazů a postupně spojujeme menší shluky do větších
     divizní – vycházíme od celé trénovací množiny jako jednoho shluku a postupně dělíme větší shluky na menší
  2. nehierarchické – různé iterační metody – např. MacQueenův algoritmus (k-means)

DEPARTMENT OF CYBERNETICS

STRUKTURÁLNÍ (SYNTAKTICKÉ) METODY

- Syntaktický popis
  - je vhodný tam, kde potřebujeme zachytit strukturu objektů
  - nebo kde pro jejich složitost chceme využít strukturu pro rozpoznávání
- Syntaktický popis objektu je hierarchická struktura jeho elementárních vlastností
- elementární vlastnosti se zde nazývají **primitiva**
- Obraz je pak reprezentován **řetězcem primitiv**
- Množina všech primitiv bývá nazývána **abecedou**
- Množina všech řetězců, pomocí nichž lze charakterizovat obrazy jedné třídy, se nazývá **jazyk popisu**
- Jazyk je generován nějakou **gramatikou**.
- Gramatika je soubor pravidel, pomocí nichž lze ze symbolů abecedy vytvářet řetězce, charakterizující možné tvary objektů.

DEPARTMENT OF CYBERNETICS

STRUKTURÁLNÍ METODY - Trénování:

1. volba primitiv
2. konstrukce gramatiky generující řetězce, reprezentující možné tvary objektu
   - na základě znalosti úlohy, zkušenosti
   - případně z trénovací množiny
   - tento krok je převážně prováděn ručně
   - úloha automatické inference (odvozování) gramatik je jen velmi obtížně řešitelná

STRUKTURÁLNÍ METODY - Rozpoznávání:

1. příslušnost neznámého obrazu do dané třídy testujeme procesem tzv. syntaktické analýzy.
2. snažíme se pomocí gramatiky charakterizující danou třídu vygenerovat neznámý řetězec
3. např. konečné automaty

DNN (CNN)

- jedna z technik machine learning pro zpracování obrazu
- více vrstev - nelineální přístup jak pro extrakci příznaků, tak i jejich zpracování
- obrázek reprezentován jako vektor
- často jako feedforward NN ale i recurrent NN (LSTM language modeling),
- trénování backpropagation algoritmem s gradient descent technikou

Rozdíly oproti standardním přístupům:

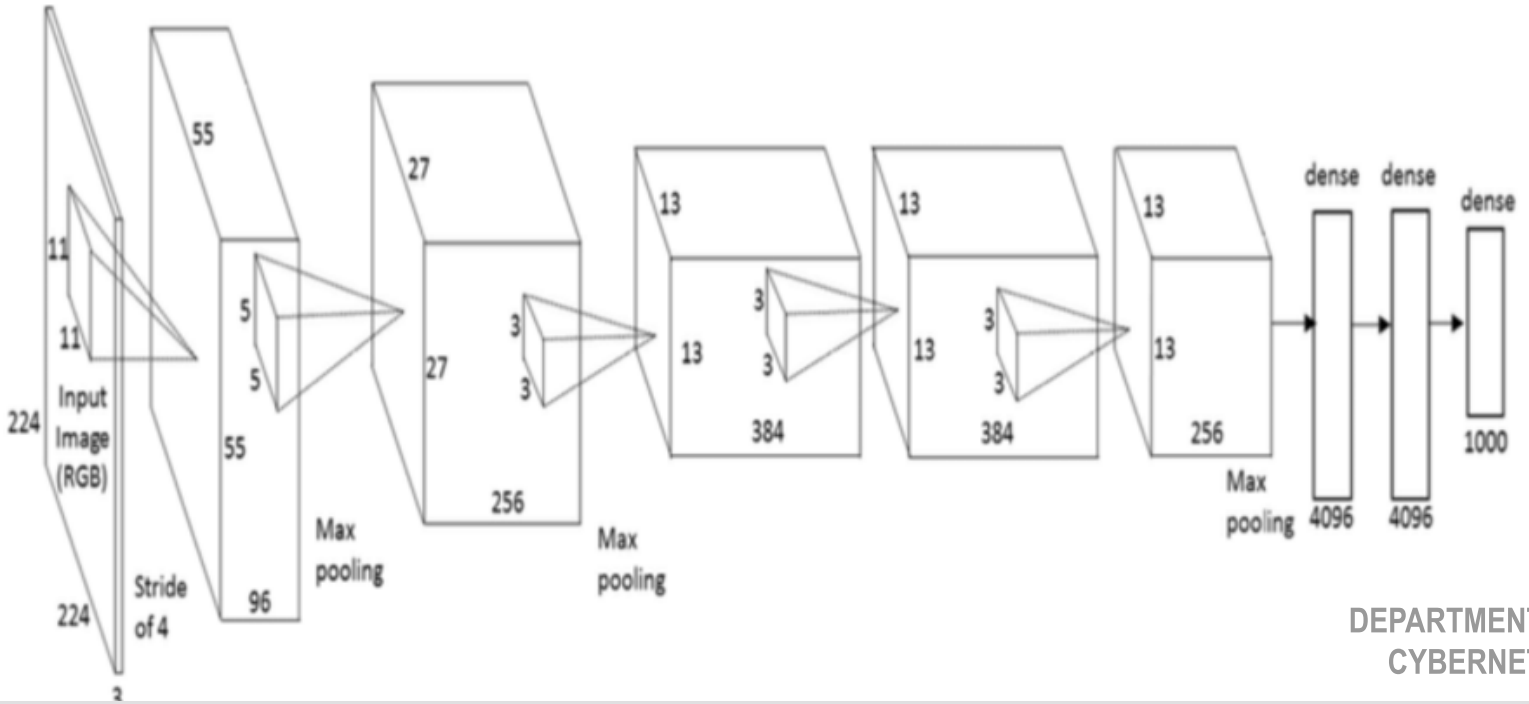
- trénovatelné feature extraktory
- end-to-end systém, tedy včetně klasifikace (regrese)
- feauture extractory ve více úrovních

DEPARTMENT OF CYBERNETICS

Klasická CNN je kombinací následujících vrstev:

1. Convolution Layer
2. Pooling Layer
3. Fully Connected Layer

- první konvoluční vrstvy se starají o detekci hran
- Následující vstrvy kombinují informace do jednodušších tvarů
- následuje modelování menších kousků informace
- předposlední vrtvy (FC) kombinují (vážená kombinace) do finální predikce



Vlastnosti:

- konvoluční jádro (ve více vrstvách)
- maxpooling - maximální hodnota v oblasti (okně) ... dělá pod-vzorkování výběrem pouze významných odezev
- aktivační funkce (často RELu - rychlejší konvergence SGD a implementačně jednodušší)
- Dropout (regularizace přetrénování) - náhodné vymaskování některých neuronů v FC vrstvách
- inicializace vah

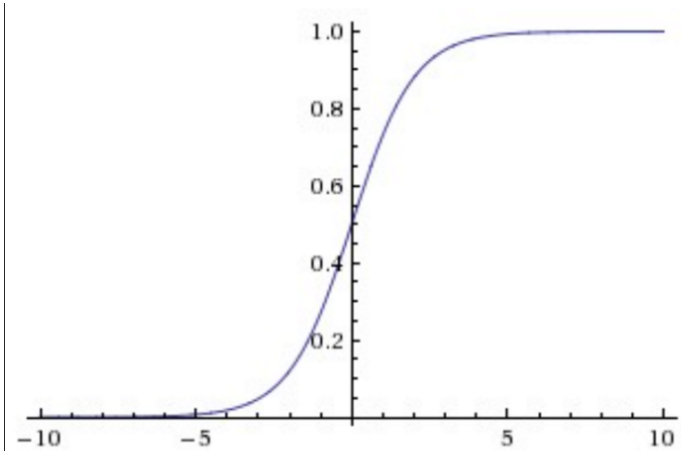
Inicializace vah

- Gaussian Random Variables
  - $\mu = 0$ a $\sigma = 0.01 \ldots 10^{-5}$
- Xavier Initialization
  - incicializace Gaussem pro větší sítě způsobuje, že výstupy neuronů budou blízké nule
  - variance gausovská distribuce závislá na počtu vstupů neuronu (velikosti předchozí vrstvy)
  - doporučuje se $var(w) = \frac{1}{n_{in}}$ (Caffe framework) nebo i $var(w) = \frac{2}{n_{in}+n_{out}}$

DEPARTMENT OF CYBERNETICS

Aktivační funkce

Sigmoid Function $\sigma(x) = \frac{1}{(1+e^{-x})}$

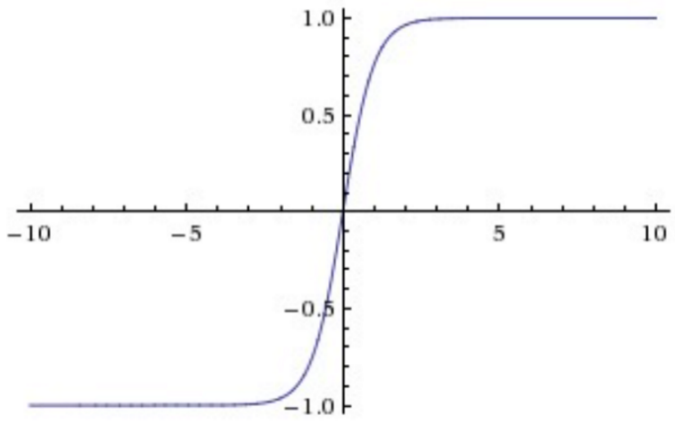


- existují problémy v souvislosti CNN
- saturované neurony potlačují gradient (viz -5 +5) kde gradient je skoro nula
- při zpětné propagaci se gradient na tomto neuronu "ztratí"
- náročný výpočet exp funkce
- výstup $(0, 1)$ není centrován na nulu ... nevýhoda jako vstup do další vrstvy
- není proto využívána pro CNN

DEPARTMENT OF CYBERNETICS

Aktivační funkce

tanh activation - tvar hyperbolické tangent funkce

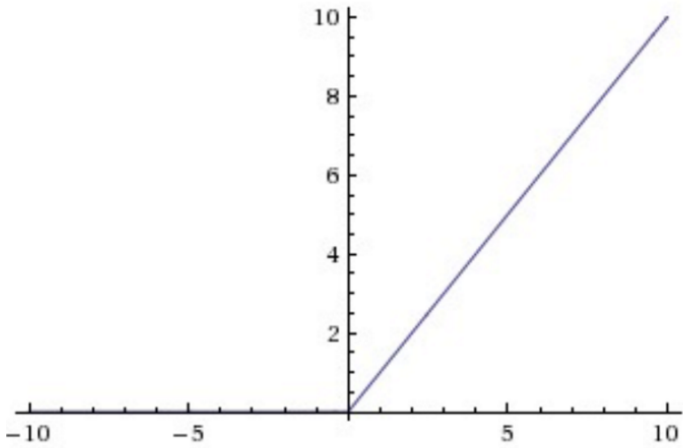


- stále problém ... saturované neurony potlačují gradient (viz -5 +5) kde gradient je skoro nula
- při zpětné propagaci se gradient na tomto neuronu "ztratí"
- náročný výpočet

DEPARTMENT OF CYBERNETICS

Aktivační funkce

ReLU (Rectified Linear Unit) $f(x) = \max(0, x)$

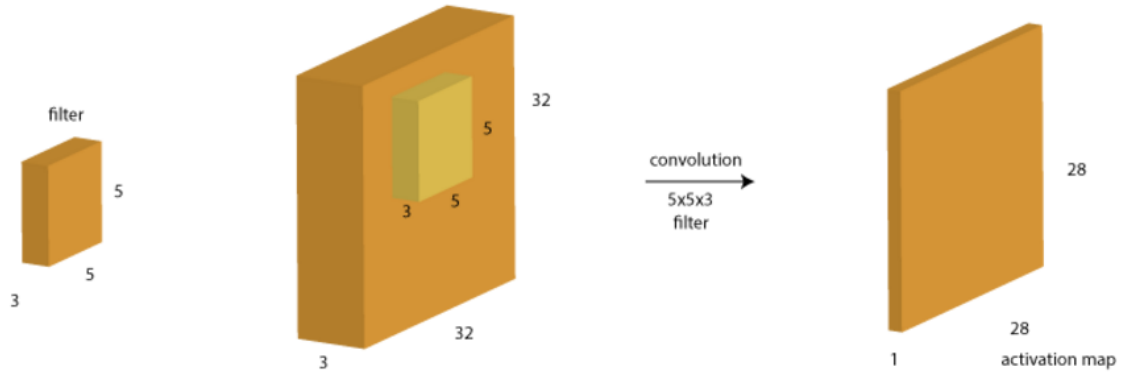


- nejčastěji používaná aktivační funkce pro CNN
- rychlá konvergence trénování
- implementačně jednodušší
- gradient je stále potlačován pro záporné x

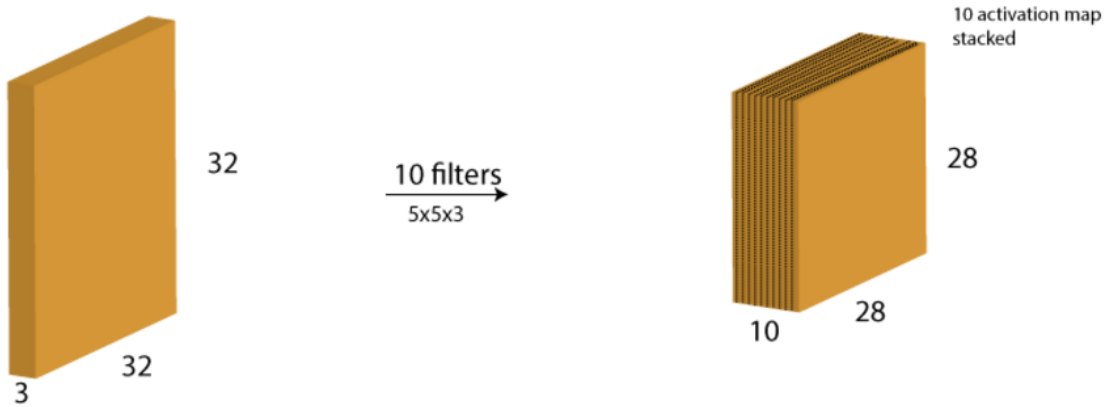
DEPARTMENT OF CYBERNETICS

Konvoluční vrstva - Convolution Layer

Pro vstupní obrázek a daný konvoluční filtr máme výstup jako odezvu (aktivační mapu)



Pro více filtrů (zde na ukázku 10) pak získáváme tvar ve formě 10 odezev:



DEPARTMENT OF CYBERNETICS

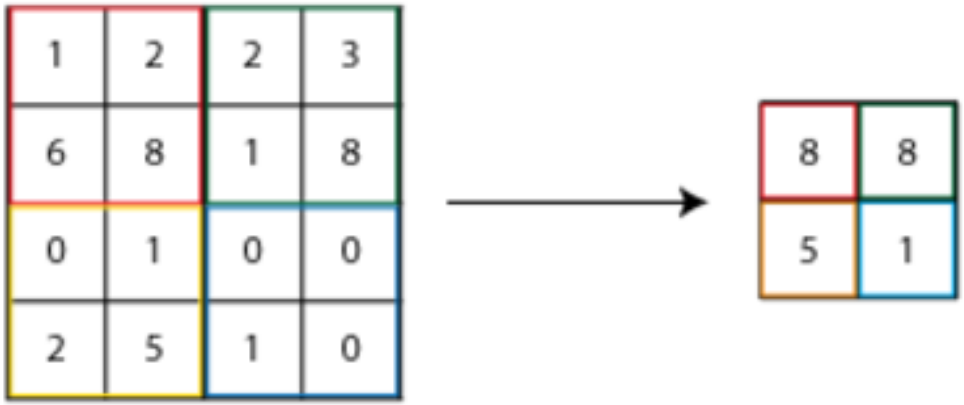
Vlastnosti:

- posun okénka (Stride) je definován pro danou konvoluční vrstu
- hodnota posunu z principu 1 nebo více, závisí také i na zvolené velikosti filtru (okénka)
- ovlivňuje velikost následující vrstvy
- "Zero-padding"vyplnění nulou za okrajem obrázku (výpočet konvoluce)

- počet parametrů $V * K + K$
- V velikost (objem) filtru např. 5x5x3
- K je počet filtrů

Pooling vrstva - Pooling Layer

- Když použijeme zerro padding a stride jedna, pak velikost výstupu konvoluční vrstvy, tedy aktivační mapa(y), mají stejnou velikost
- musí se provést redukce velikosti, např. odezva 4x4 je redukována na 2x2 metodou max-pooling
- možné i jiné techniky např. average pooling (GooGleNet)



DEPARTMENT OF CYBERNETICS

Plně spojená vrstva - Fully Connected Layer

- jsou použity ke konci CNN
- každý neuron trénováním získává danou funkci důležitou pro rozhodnutí
- při trénování aplikován Dropout
- poslední vrstva má velikost rovnou počtu klasifikačních tříd, nebo velikosti regresního vektoru
- FC vrstvy nemusí být v architektuře použity

DEPARTMENT OF CYBERNETICS

DNN as a machine learning model

- in the case of classification or regression we try to:
  - such values of ω_i that will minimize a chosen criterion
  - criteriun usually incorporates information from teacher: t

Classification criteria:

- Binary cross entropy :

$$E_k = -\sum_i t_i \log o_i + (1 - t_i) \log(1 - o_i) \tag{3}$$

- Categorical cross entropy:

$$E_k = -\sum_i t_{k,i} \log o_{k,i} \tag{4}$$

(with soft-max layer)

Regression criteria: Mean squared/absolute error

$$E_k = \sum_i (t_i - o_i)^2 \tag{5}$$

Trénování

- často použit algoritmus Gradient Descent

$$\omega_{n+1} = \omega_n - \gamma_n \nabla E(\omega_n) \tag{6}$$

- kde $E(\omega_n)$ je zvolené kritérium (nelineární funkce) a ω_n parametry v kroku n ,
- γ_n Learning rate ... u GD metod bohužel neznáme velikost kroku optimalizace

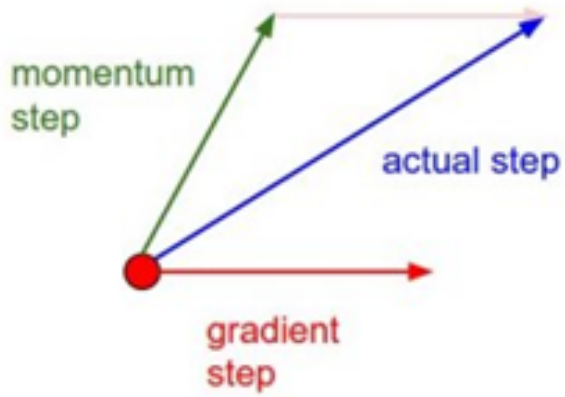
DEPARTMENT OF CYBERNETICS

Realizace:

- omezená velikost paměti
- mini-batch (Stochastic Gradient Descent - SGD metoda)
- náhodně rozdělená data, update postupně na balíčcích
- rychlost konvergence - momentum SGD

$$\Delta\omega_n = \gamma_n \nabla E(\omega_n) + \alpha \Delta\omega_{n-1} \tag{7}$$

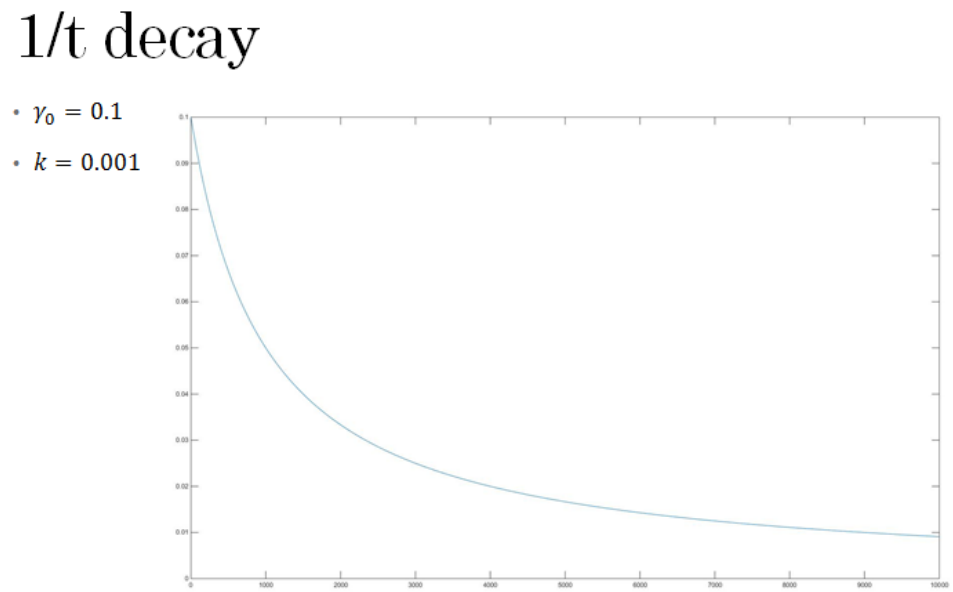
- α momentum $=$ váha jak *cumulative gradient direction* bude při optimalizaci směřován



DEPARTMENT OF
CYBERNETICS

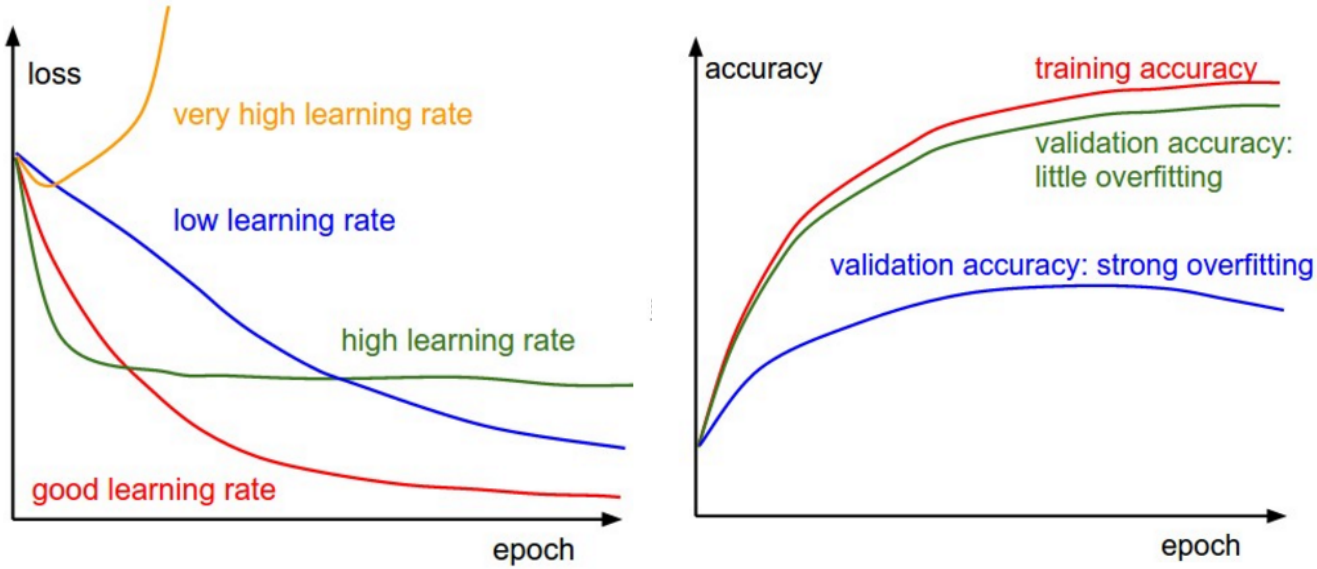
Learning rate decay:

- změna learning rate během trénování
- **Step decay** skokové snižování po určitém počtu trénovacích epoch
- **Exponential decay**
- **1/t decay**



Per-parameter adaptive LR methods:

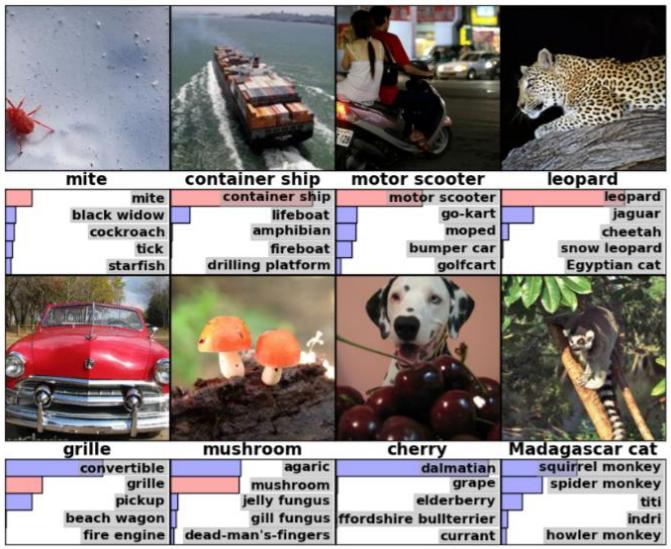
- adaptivně mění Learning rate pro jednotlivé parametry
- **Adagrad**
- **RMSprop**
- **Adam**



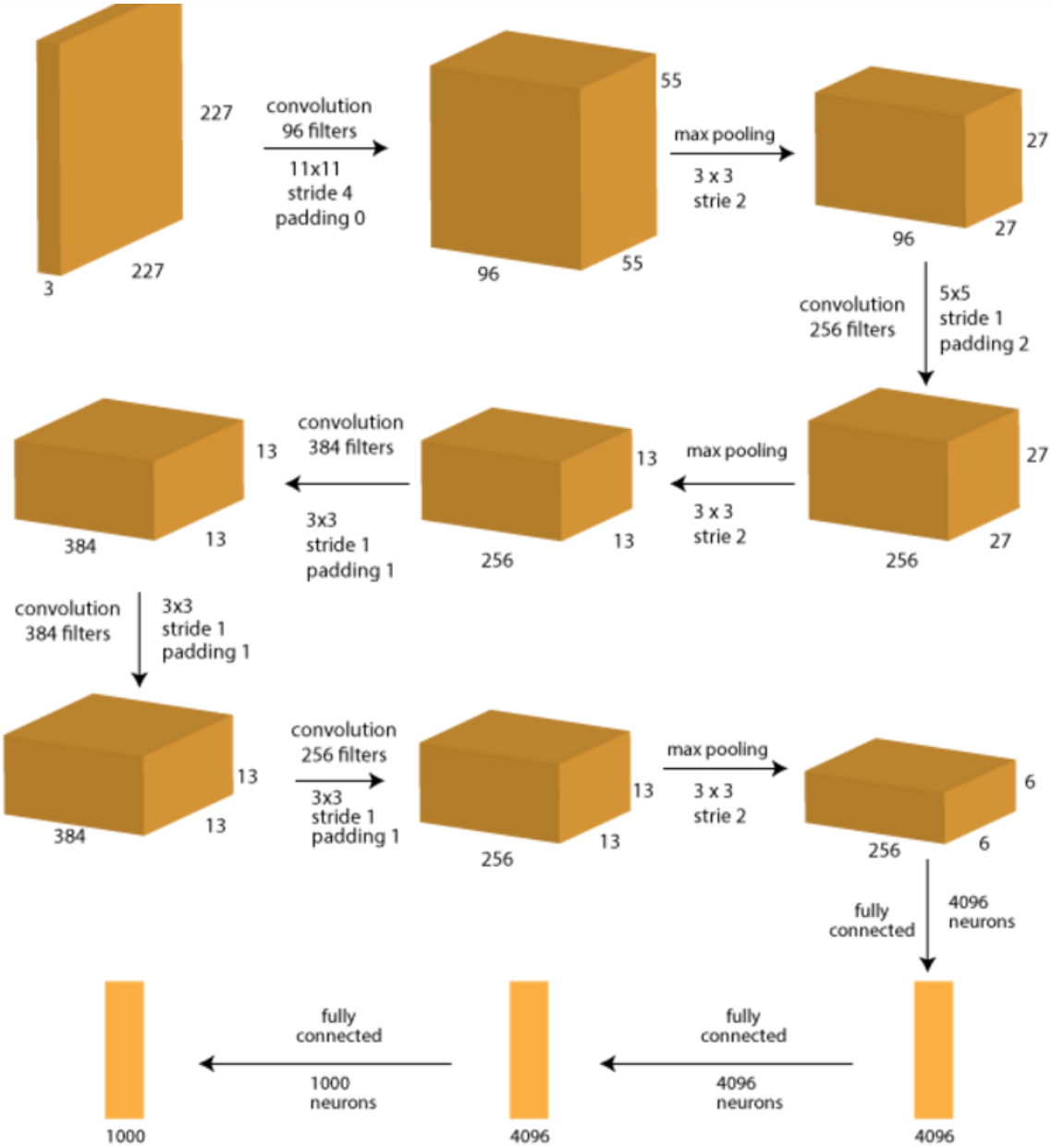
- zde vidíme efekt rozdílného Learning rate
- rozdíl mezi trénovací a validační přesností udává míru přetrénování

AlexNet (2012)

- vítěz IMAGENET Challenge 2012
- 1000 tříd, 1.3 mil. trénovacích obrázků
- nebyl zcela zvládnut problém trénování na GPU
- v dané době získali významně lepších výsledků než jiné přístupy
- označena jako první z "moderních"CNN
- hloubka 11
- možné použít jako inicializace pro podobnou úlohu + dotrénování

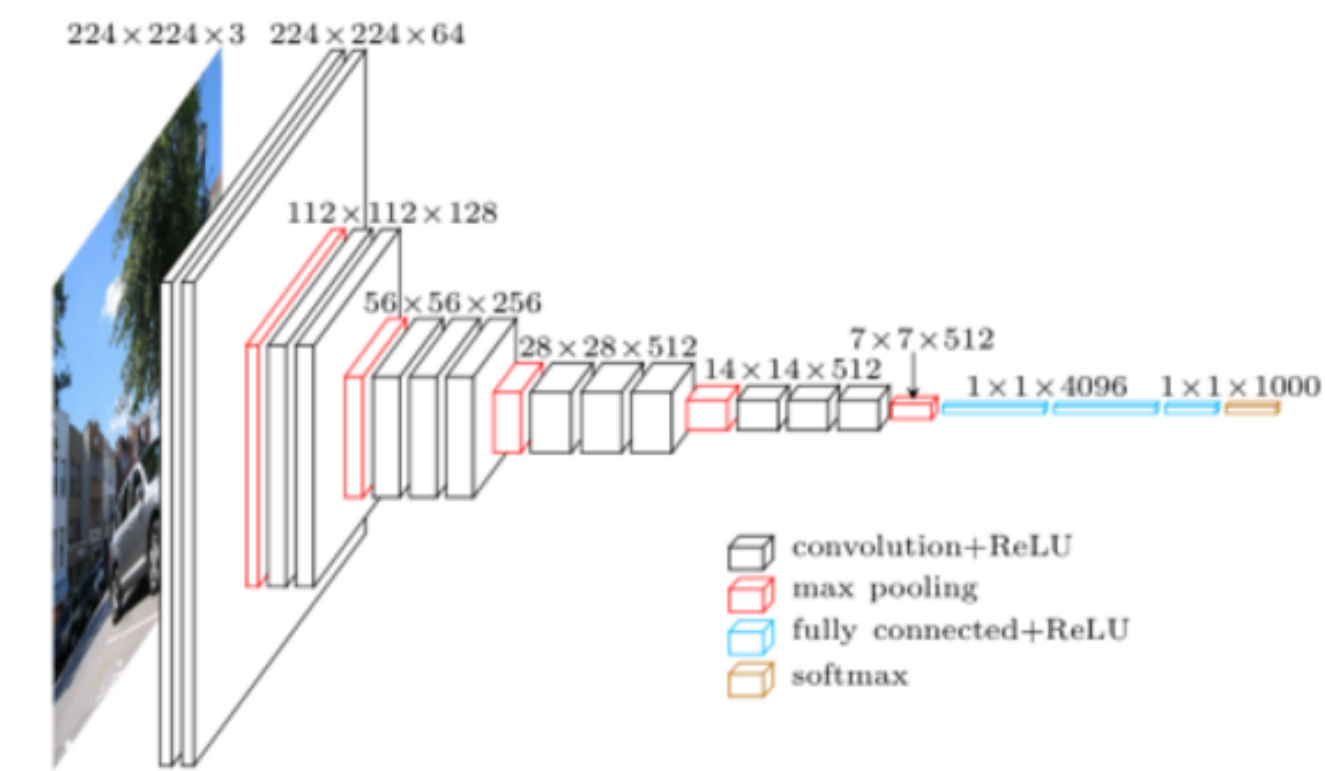


DEPARTMENT OF CYBERNETICS


227
227
3
convolution
96 filters
11x11
stride 4
padding 0
55
55
96
max pooling
3 x 3
strie 2
27
27
96
convolution
256 filters
5x5
stride 1
padding 2
27
27
256
max pooling
3 x 3
strie 2
13
13
256
convolution
384 filters
3x3
stride 1
padding 1
13
13
384
convolution
384 filters
3x3
stride 1
padding 1
13
13
384
convolution
256 filters
3x3
stride 1
padding 1
13
13
256
max pooling
3 x 3
strie 2
6
6
256
fully
connected
4096
neurons
4096
fully
connected
4096
neurons
4096
fully
connected
1000
neurons
1000

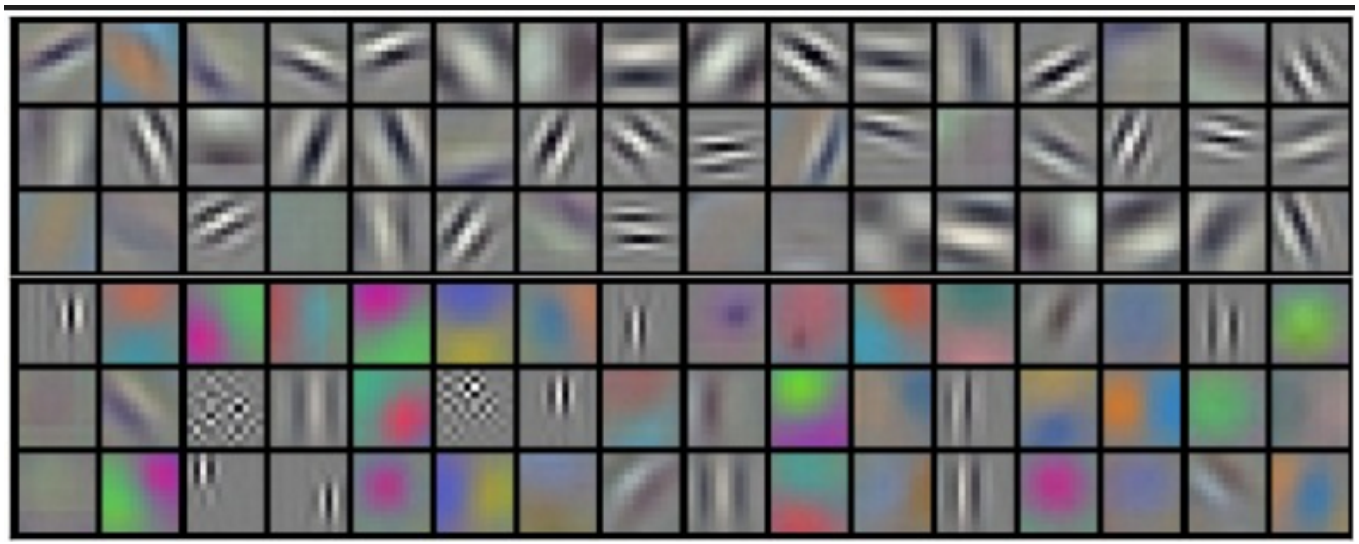
VGGNet (2014)

- malé jen 3x3 kernely, stride 1
- augmentace dat
- hloubka 19



GoogLeNet (2014)

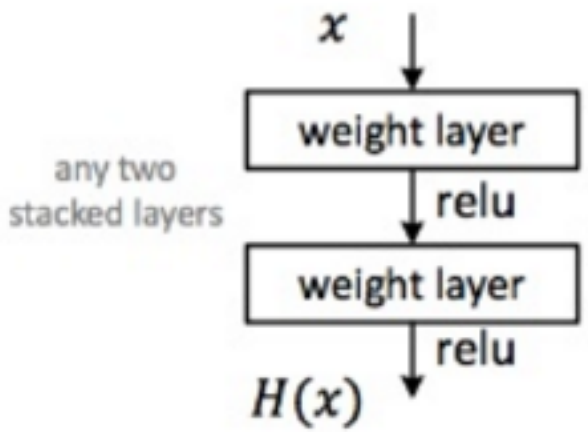
- vítěz IMAGENET Challenge 2014
- jen kernely
- vzniká otázka ztráty gradientu pro hlubší sítě (BackPropagation)
- problémy s FC vrstvami (tendence k přetrénování)
- hloubka 22

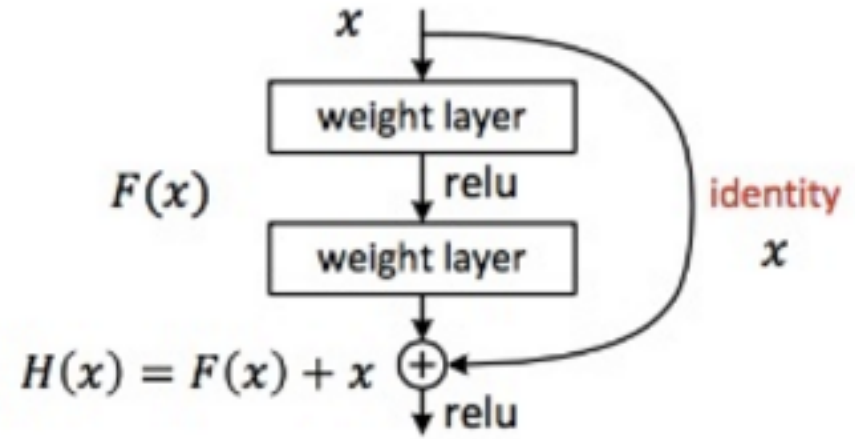


DEPARTMENT OF CYBERNETICS

ResNet (2014)

- vítěz IMAGENET Challenge 2015
- inspirováno VGG, ale hloubka **152**
- aktivační funkce P-RELU (adaptivní RELU - více parametrů)
- vzniká problém ztráty gradientu pro hlubší sítě (BackPropagation) ... cca už od hloubky 30
- vymyšleny a použity tzv. **shortcut connection**

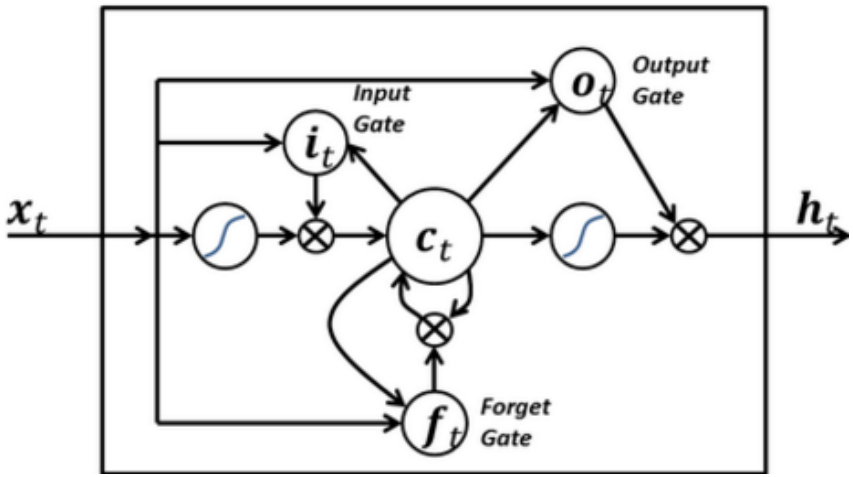




DEPARTMENT OF CYBERNETICS

Long short-term memory (LSTM) 1997

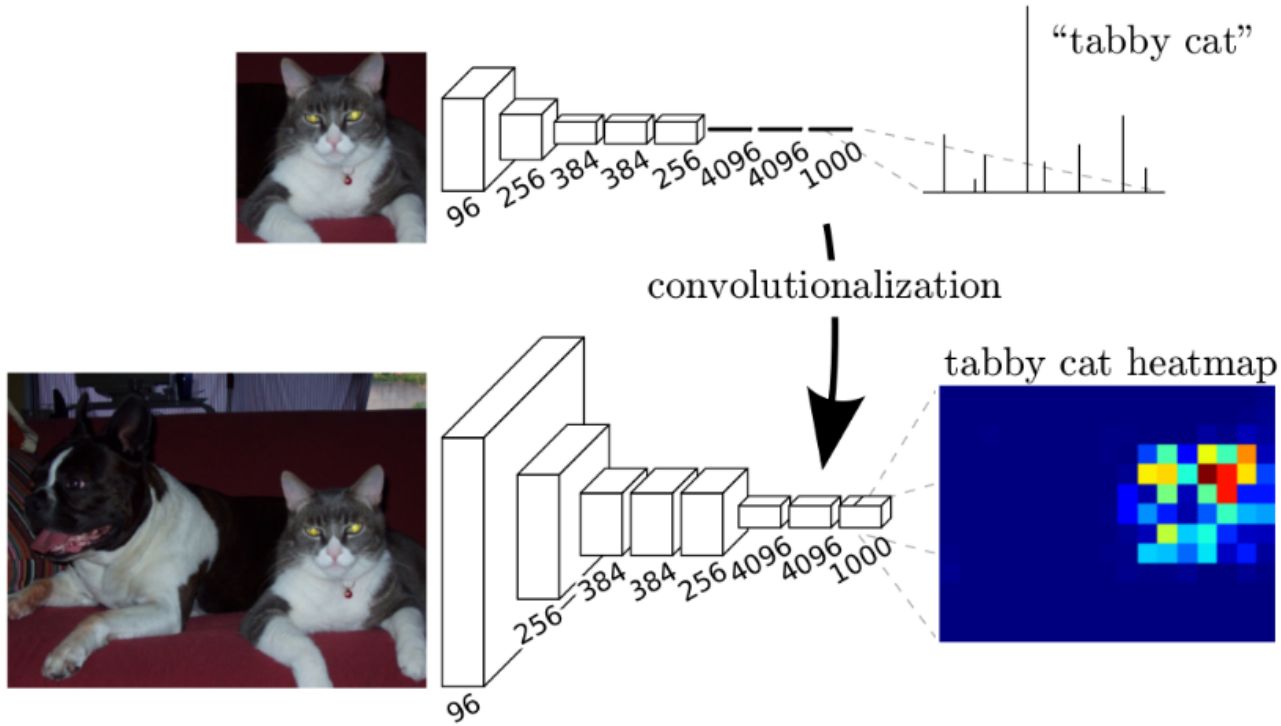
- RNN architektura, jako alternativa k HMM
- v současnosti velký úspěch v aplikacích v ASR nebo AVSR (audio + obraz rtů), Handwriting recognition, Robot control aj.
- řeší problém RNN se ztrátou gradientu s rostoucí dobou mezi důležitými událostmi
- na množině trénovacích sekvencích



DEPARTMENT OF CYBERNETICS

FCN - Fully Convolutional Networks:

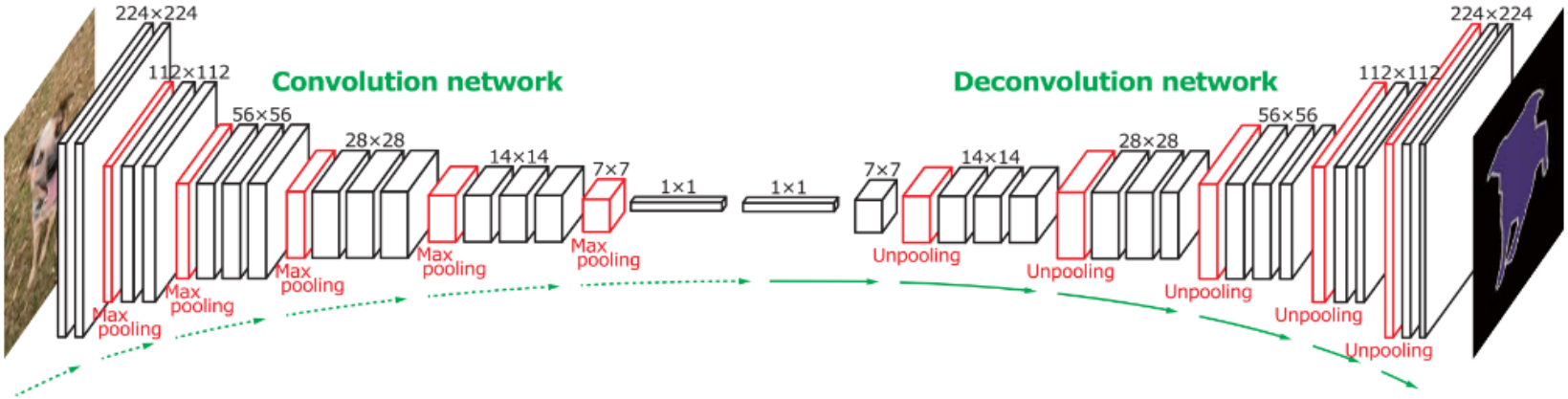
- FCN má pouze konvoluční vrstvy
- nejsou FC vrstvy
- přebírá úlohu segmentace - výstupem je heat mapa
- použitelná na libovolnou velikost vstupního obrázku



DEPARTMENT OF CYBERNETICS

Deconvolution Network:

- místo full connected vrstev jsou použity de-konvoluční a un-pooling vrstvy
- dekonvoluce časo-prostorových feature map
- produkuje per-pixel predikci (labels)
- může např. vycházet AlexNet, VGG aj. architektur

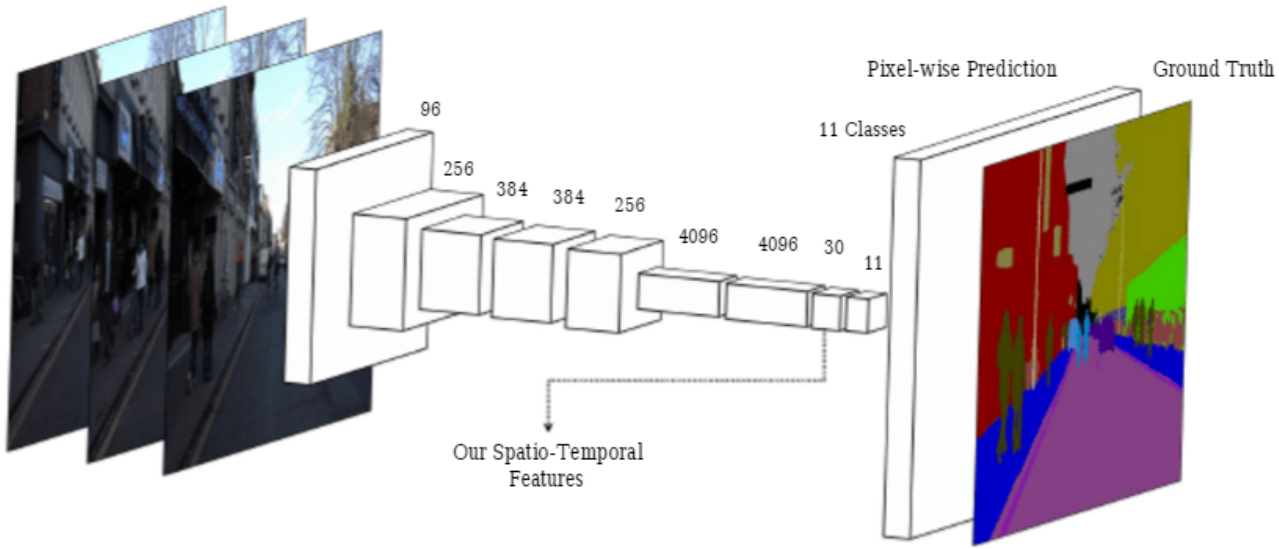


[1] Learning Deconvolution Network for Semantic Segmentation

DEPARTMENT OF CYBERNETICS

Spatio-Temporal FCN for Semantic Video Segmentation:

- vychází z AlexNet architektury
- použity video data LSTM verze rekurze
- dekonvoluce časo-prostorových feature map ... produkuje per-pixel predikci



[2]

[2] STFCN: Spatio-Temporal FCNfor Semantic Video Segmentation

DEPARTMENT OF CYBERNETICS

AARSHAY JAIN, Deep Learning for Computer Vision – Introduction to Convolution Neural Networks