



EVROPSKÁ UNIE
Evropské strukturální a investiční fondy
Operační program Výzkum, vývoj a vzdělávání

MŠMT
MINISTERSTVO ŠKOLSTVÍ,
MLÁDEŽE A TĚLOVÝCHOVY



Rozvoj lidských zdrojů TUL pro zvyšování relevance, kvality a přístupu ke vzdělání v podmínkách Průmyslu 4.0

CZ.02.2.69/0.0/0.0/16_015/0002329

Úvod do zpracování obrazů

Mechatronika

Prezentace přednášky č. 10 b

Parametrizace a rozpoznávání obrazu

doc. Ing. Josef Chaloupka, Ph.D.



TECHNICKÁ UNIVERZITA V LIBERCI
www.tul.cz



<http://www.ite.tul.cz>



Příznaky pro detekování objektů

- **Viola-Jonesův detektor** (VJD) obličej využívá příznaky podobné Haarovým báзовým funkcím a klasifikátor založený na **AdaBoost** algoritmu



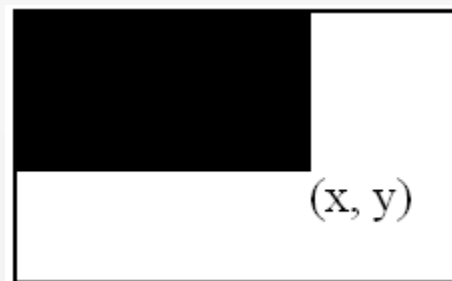
A

B

C

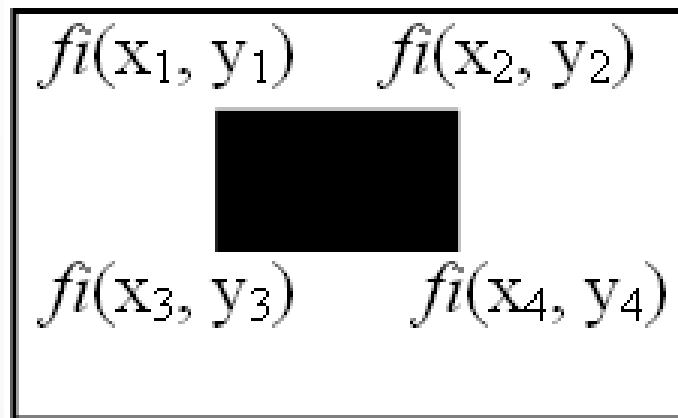
příznak - suma hodnot obrazových bodů nacházející se pod bílou oblastí 2D funkce mínus suma hodnot obrazových bodů pod černou oblastí funkce, zjednodušení:

Integrovaný obraz





Příznaky pro detekování objektů



- suma byla spočítána ze čtyř hodnot na základě využití dvou operací sčítání a jedné operace odčítání: $(f_i(x_1, y_1) + f_i(x_4, y_4)) - (f_i(x_3, y_3) + f_i(x_2, y_2))$
- Pro příznak typu A (obr. 2.8) by poté bylo potřeba použít 6 hodnot z integrálního obrazu, pro příznak B by to bylo 8 a 9 pro příznak typu C

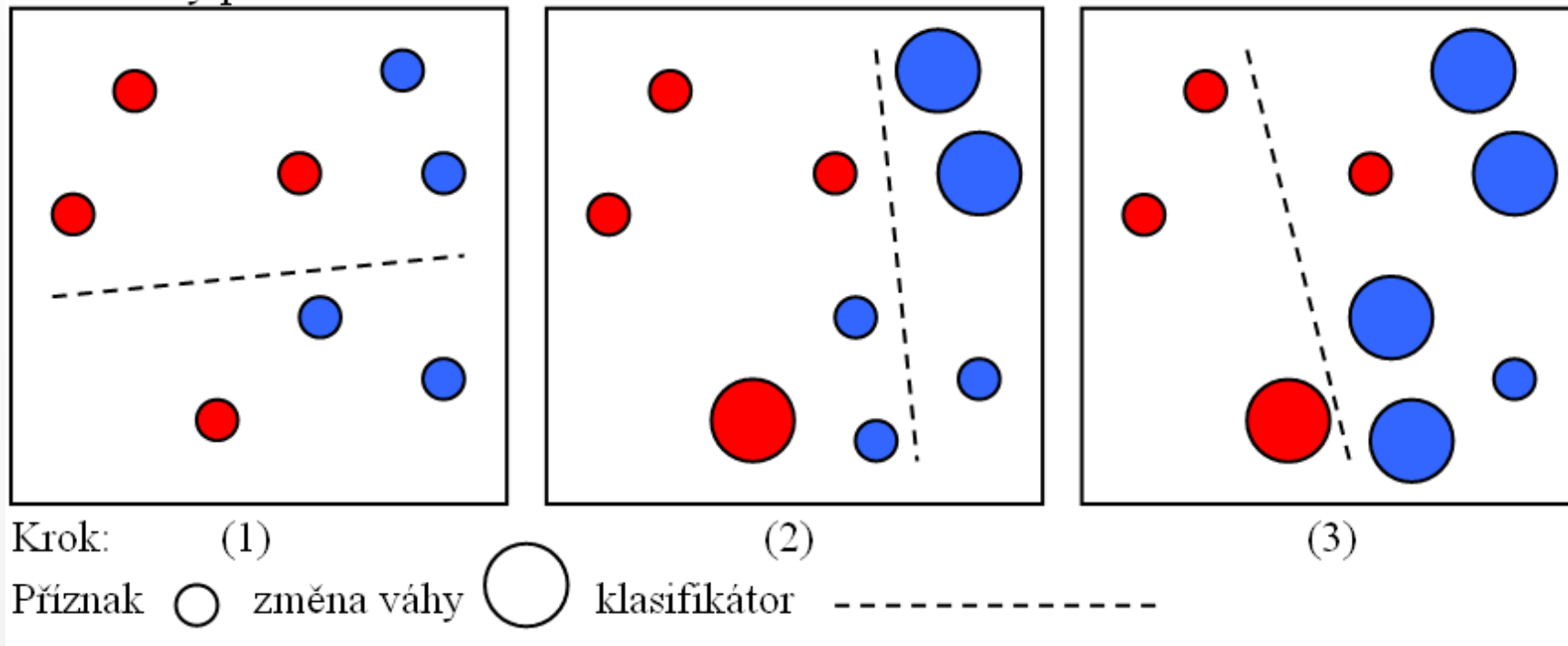




Příznaky pro detekování objektů

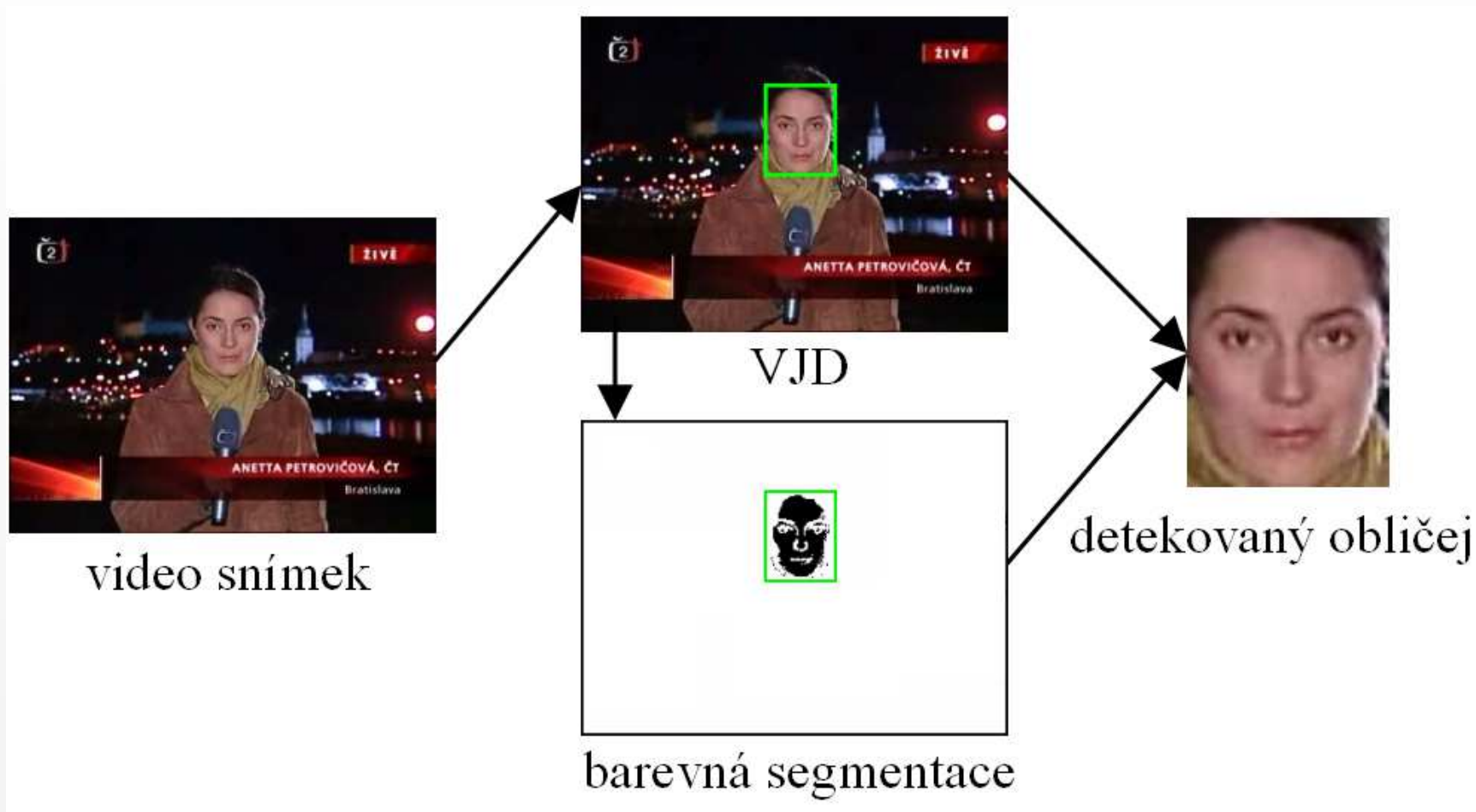
- **Algoritmus AdaBoost**

Příznakový prostor



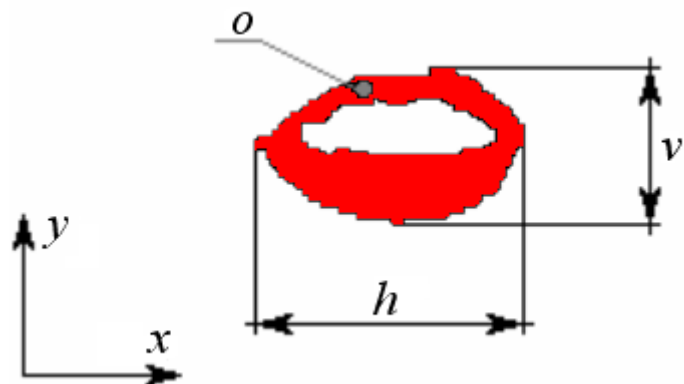


Příznaky pro detekování objektů





Využití dynamiky chování rozpoznávaných objektů



horizontální rozšíření rtů h :

$$h = \max_{y=0..N-1} \sum_{x=0}^{M-1} f(x, y)$$

Vertikální rozšíření rtů v :

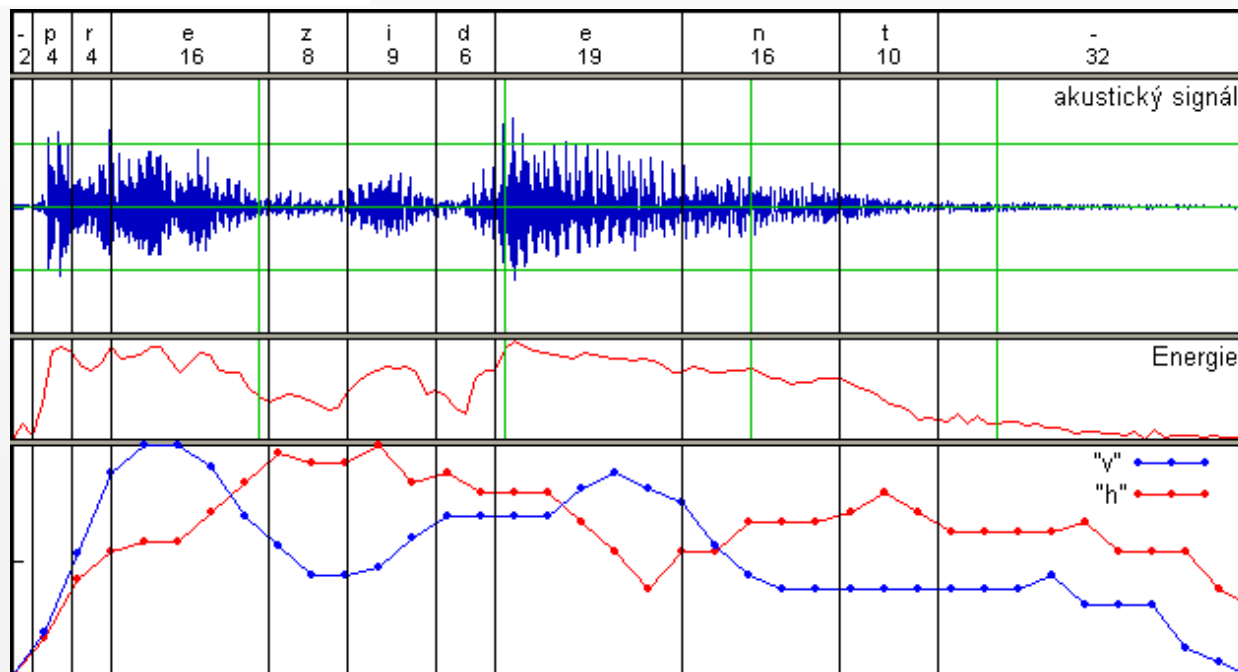
$$v = \max_{x=0..M-1} \sum_{y=0}^{N-1} f(x, y)$$

oblast rtů o :

$$o = \sum_{y=0}^{N-1} \sum_{x=0}^{M-1} f(x, y)$$

zaokrouhlení rtů r :

$$r = \frac{v}{h}$$

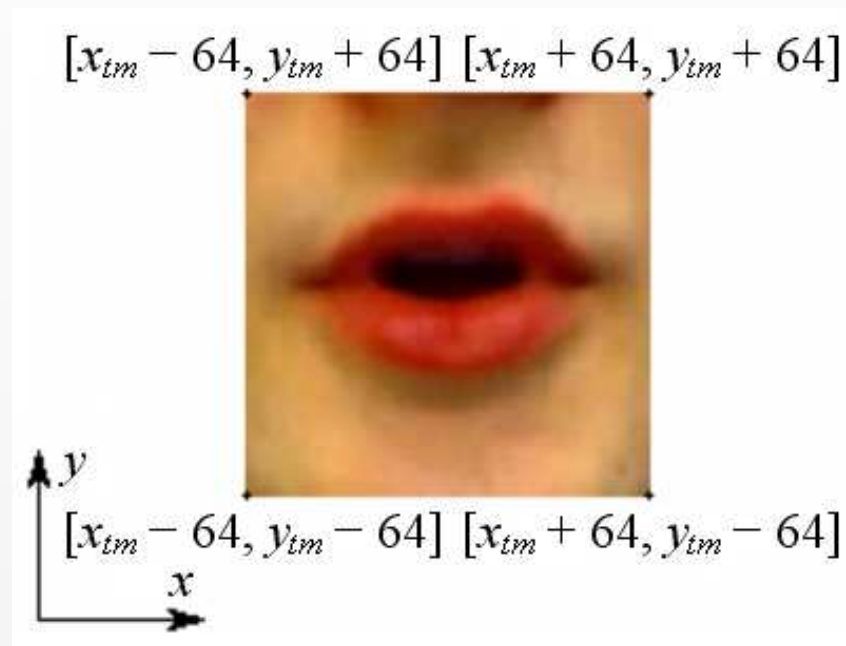




Volba, výběr a redukce příznaků

2D Diskrétní kosinová transformace:

$$F(u,v) = \frac{2c(u)c(v)}{N} \sum_{m=0}^{N-1} \sum_{n=0}^{N-1} f(m,n) \cos\left(\frac{2m+1}{2N}u\pi\right) \cos\left(\frac{2n+1}{2N}v\pi\right) \quad c(k) = \begin{cases} \frac{1}{\sqrt{2}} & \text{pro } k = 0 \\ 1 & \text{pro } k > 0 \end{cases}$$



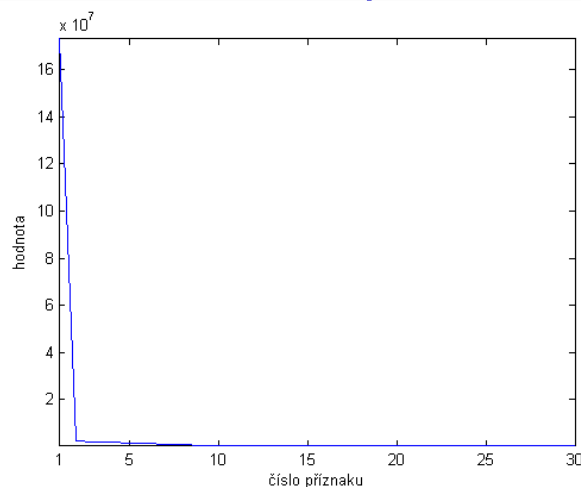
N nejvyšších hodnot energie: $E(u,v) = F(u,v)^2$



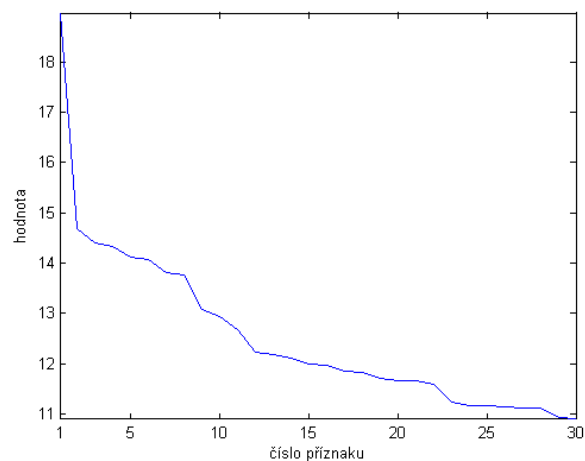


Volba, výběr a redukce příznaků

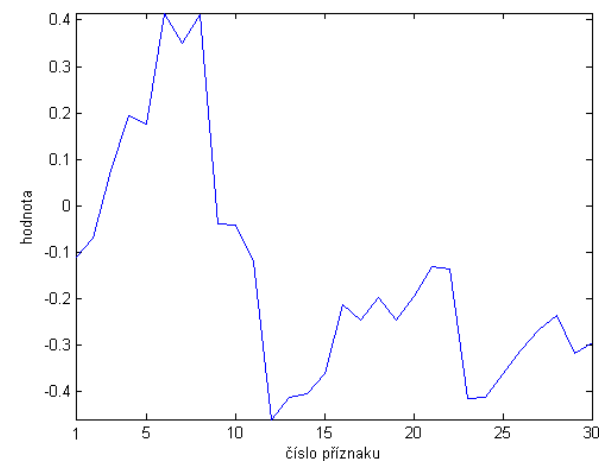
Normalizace příznakového vektoru:



$E(u,v)$



$\log(E(u,v))$



$\log(E(u,v)) - FMS$

Výpočet dynamických příznaků:

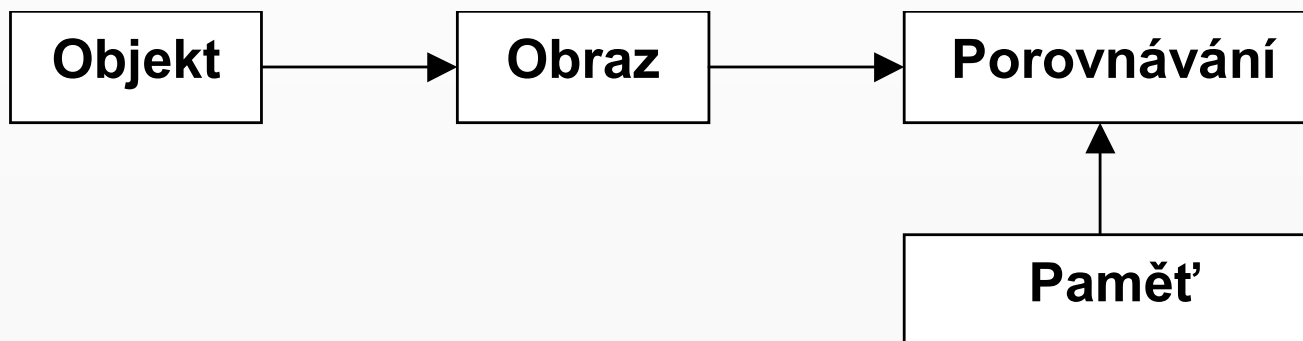
$$\begin{aligned} x'[n] &= x[n] - x[n-1] &<<< \text{rychlost} \\ x''[n] &= x'[n] - x'[n-1] &<<< \text{zrychlení} \end{aligned}$$





Rozpoznávání obrazů

- aby se dalo rozpoznávat, musí dojít k **učení – není rozpoznávání bez učení**



Termíny:

umělá inteligence (artificial intelligence, AI)

- stroj disponující AI vykazuje v dané situaci takové chování, že jej nelze odlišit od člověka

Rozpoznávání obrazů (pattern recognition) PR

teorie rozpoznávání obecných (abstraktních) obrazů

- obrazem může být 1- i více-rozměrných signál, statický či měnící se v čase





Rozpoznávání obrazů



Základní úloha:

- zařadit OBJEKT na základě jeho OBRAZU do jedné ze TŘÍD

TŘÍDA (class) – množina objektů s podobnými vlastnostmi

Příklad: objekty určené k sezení

a) židle, křesla, stoličky, sedla do auta,

b) sedačky s nohami a bez nohou

PODOBNOST (similarity) – vlastnost měřitelná na obrazu objektu umožňující vyjádřit vztah ke každé ze tříd

OBRAZ (pattern) – reprezentace objektu vyjádřená pomocí kvantitativního či kvalitativního popisu, použitelná pro rozpoznávání





Rozpoznávání obrazů

Jak volit obraz objektů?

1. tak aby byla chyba klasifikace co nejmenší (chyba zařazení do špatné třídy),
2. aby byl obraz snadno (levně) získatelný,
3. aby bylo vlastní porovnání obrazů co nejefektivnější

Jsou-li již objekty popsány svými obrazy, jak provádět rozpoznávání?

UČENÍ (training) – systému je předkládán jistý soubor obrazů objektů, u nichž je a priori známo jejich zařazení do daných tříd – tzv. **trénovací množina**

Příklad: učení abecedy

A, a, A, A, **A**, - třída A

GENERALIZACE – schopnost zobecňovat z naučeného

Učení může být sekvenční s průběžným testováním a zpřesňováním

A vs. B - rozdíl ve spodní hraně

TESTOVÁNÍ (testing) – měří se chyba klasifikace na skupině objektů se známým zařazením – **testovací množina**





Rozpoznávání obrazů

Možnosti klasifikace:

1. zařadit vždy do jedné z tříd
2. zařadit do jedné ze tříd s možností odmítnutí zařazení (rejection)
3. možnost zařadit objekt do více tříd

Chyba klasifikace:

- a) objekt je zařazen do nesprávné třídy – classification error
- b) objekt je nesprávně odmítnut – rejection error

Cena chyby (error cost):

Příklad: ovládání stroje

- a) cena špatně vykonaného pohybu - zničení stroje
- b) cena nevykonaného pohybu – zdržení
- c) rozpoznávání falešných bankovek

TRÉNOVÁNÍ, TESTOVÁNÍ, VYUŽÍVÁNÍ (s možností adaptace)

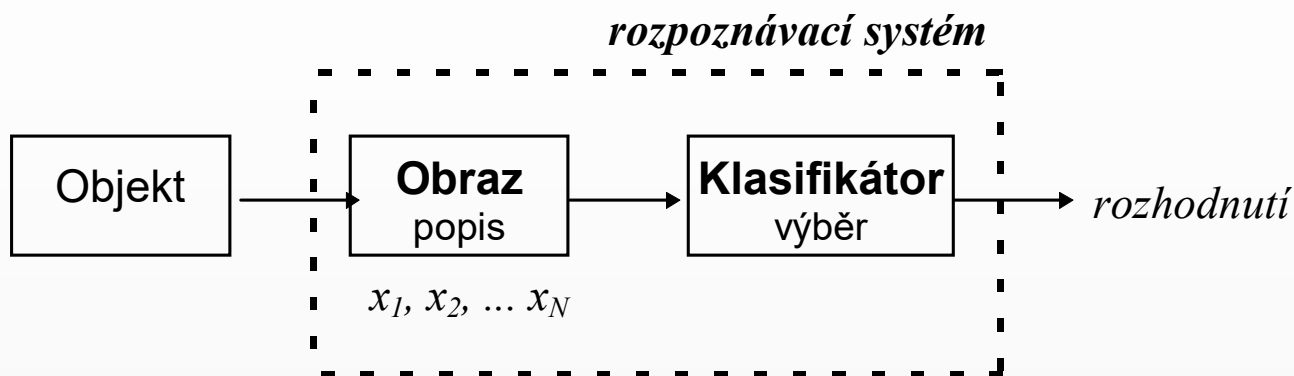
Další aspekty rozpoznávání

neexistují dva identické objekty – rozpoznávání tedy pracuje s daty, jejichž popis má náhodný charakter, **obrazy jsou popisovány pomocí příznaků**

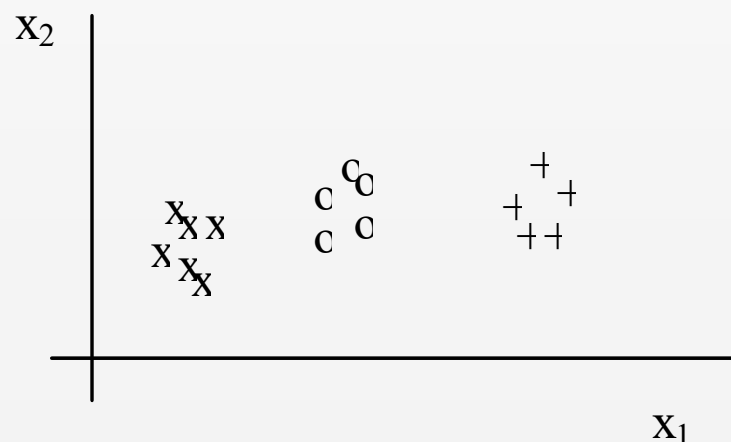




Rozpoznávání obrazů



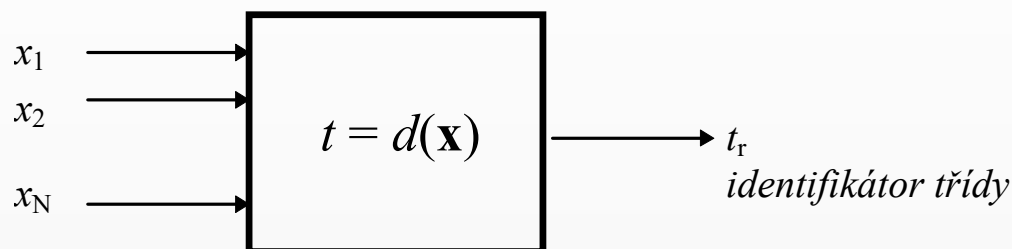
- Obraz reprezentován příznakovým vektorem $x = [x_1, x_2, \dots, x_N]$
- Obrazový prostor: euklid. prostor dimenze N (množina všech možných obrazů)
- Příklad: obraz popsán dvěma příznaky:





Rozpoznávání obrazů

Klasifikátor - systém s N vstupy a jedním výstupem



Úloha klasifikátoru:

Dány třídy T_1, T_2, T_R , obraz popsán přízn. vektorem \mathbf{x} - klasifikátor na základě **pravidla** $t = d(\mathbf{x})$ rozhodne o přiřazení \mathbf{x} do třídy T_r tím, že na výstupu vyše index t_r .

Rozhodovací pravidlo - nejčastější formy:

- a) metoda diskriminačních funkcí
- b) metoda minimální vzdálenosti
- c) metoda maximální pravděpodobnosti





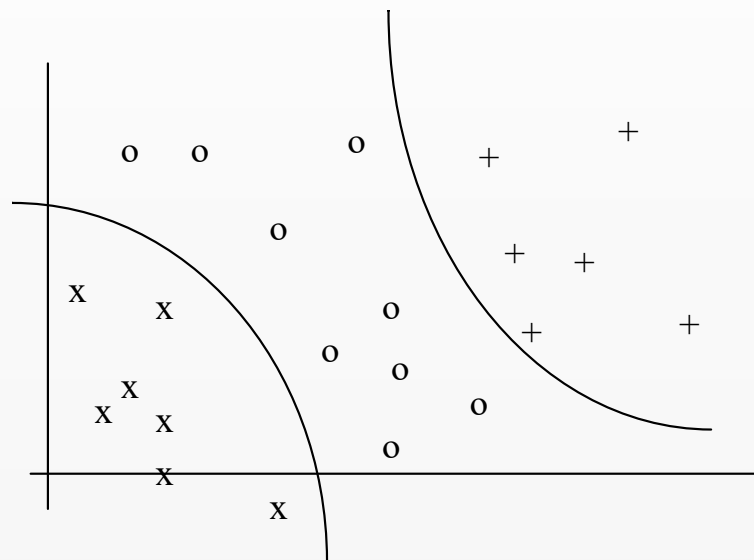
Rozpoznávání obrazů

a) Metoda diskriminačních funkcí

vychází z předpokladu, že obrazový prostor lze rozdělit na disjunktní části pomocí rozdělovacích nadploh:

v E2 rozdělovací křivky

v E3 rozdělovací plochy



Rozdělovací nadplochy lze určit pomocí **diskriminačních funkcí** g_1, \dots, g_R ,

přičemž g_r je vybrána tak, aby pro všechna platilo $g_r(x) > g_s(x) \quad s = 1, \dots, R, \quad s \neq r$

Rozdělovací plocha mezi třídami T_r a T_s je dána rovnicí: $g_r(x) - g_s(x) = 0$





Rozpoznávání obrazů



b) Metoda minimální vzdálenosti

Princip: každá třída je reprezentována svými představiteli (vzorky trénovací množiny)

Rozpoznávání:

Pravidlo nejbližšího souseda - NN (Nearest Neighbour)

neznámý vzorek se zařadí do té třídy, k jejímuž představiteli má nejmenší vzdálenost

Pravidlo k nejbližších sousedů - kNN (k Nearest Neighbours)

neznámý vzorek se zařadí do té třídy, jejíž představitelé jsou nejvíce zastoupeny v uspořádané k-tici nejbližších sousedů

Vzdálenost: nejčastěji Euklidovská

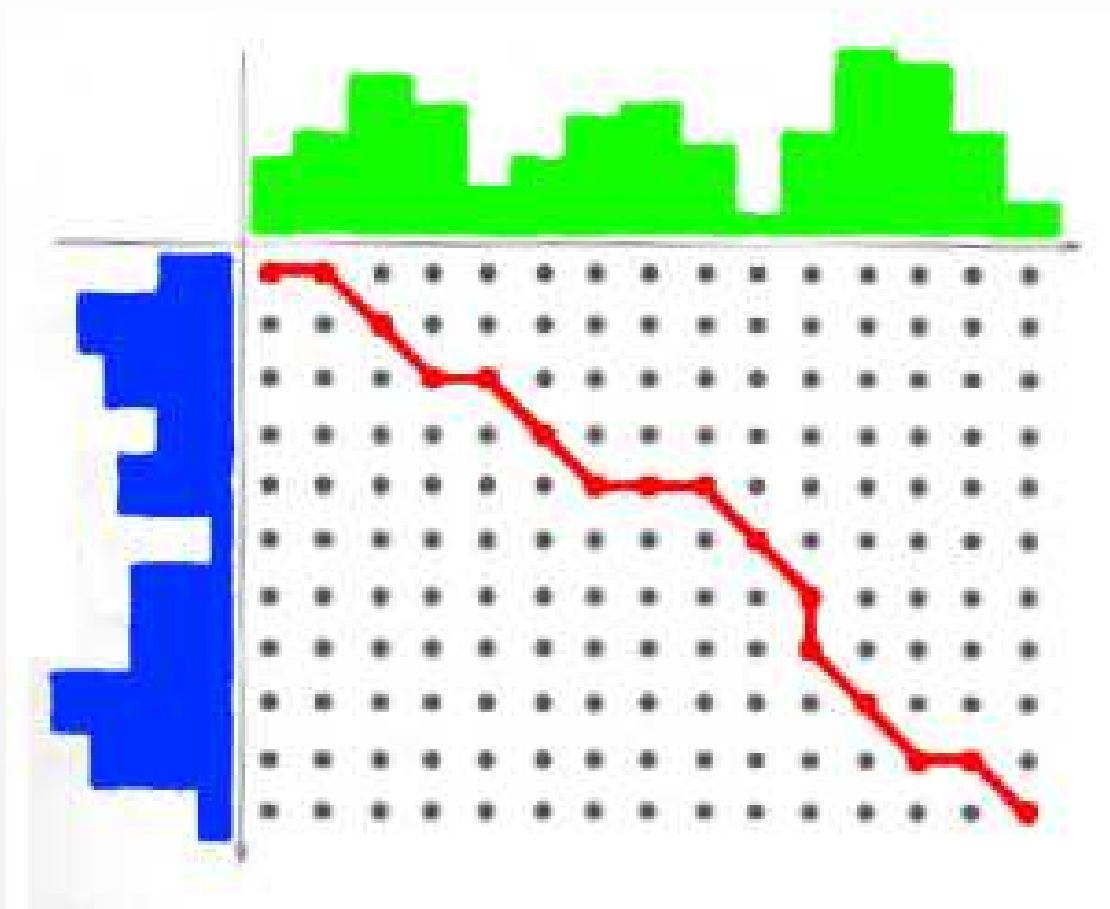
$$d(\mathbf{x}, \mathbf{y}) = \sqrt{\sum_{i=1}^P (x_i - y_i)^2}$$





Rozpoznávání obrazů

Dynamické borcení času: DTW - Dynamic time warping





Rozpoznávání obrazů



b) Metoda minimální vzdálenosti

Reprezentace pomocí etalonů:

Každá třída je reprezentována **etalonem** - vzorkem třídy, který ji nejlépe reprezentuje ve smyslu minimální vzdálenosti.

Etalon je buď skutečným prvkem třídy, nebo může vzniknout výpočtem, např. průměrováním, z příznakových vektorů třídy.

Při klasifikaci se měří vzdálenosti $|\mathbf{x} - \mathbf{e}_r|$ a vybere se T_r , aby $|\mathbf{x} - \mathbf{e}_r| = \min_{s=1,2..R} |\mathbf{x} - \mathbf{e}_s|$





Rozpoznávání obrazů

c) Metoda maximální pravděpodobnosti (též nazývaná metoda minimální chyby)

Princip: každá třída je reprezentována

a) **apriorní** pravděpodobností třídy
pravděpodobností výskytu prvků této třídy $P(T_r)$

musí platit
$$\sum_{r=1}^R P(T_r) = 1$$

b) **podmíněnou hustotou pravděpodobností** $p(\mathbf{x} | T_r)$
udává rozložení pravděpodobnosti vektoru příznaků \mathbf{x} pro třídu T_r

Trénování: pro každou třídu se na trénovací množině určí (odhadnou) výše uvedené pravděpodobnosti

Rozpoznávání: aplikace Bayesova pravidla

$$P(T_r | \mathbf{x}) = \frac{p(\mathbf{x} | T_r)P(T_r)}{p(\mathbf{x})}$$

$P(T_r | \mathbf{x})$... **aposteriorní** pravděpodobnost, že \mathbf{x} patří do třídy T_r

$p(\mathbf{x}) = \sum_{i=1}^R p(\mathbf{x} | T_i)P(T_i)$ absolutní pravd. hustota rozložení vektoru příznaků
(nezávisle na třídě)





Rozpoznávání obrazů

c) Metoda maximální pravděpodobnosti – příklad

V útulku se nachází 70% psů a 30% koček.

20% koček a 10% psů je černých.

Z dálky kamerou snímáme černé zvíře – jaká je pravděpodobnost, že to je pes?

- **apriorní** pravděpodobnost:

třída - psy $P(T_1) = 70\%$, kočky $P(T_2) = 30\%$ $\sum_{r=1}^R P(T_r) = 1$

- **podmíněnou hustotou pravděpodobností**

náhodně vybraný pes je černý $p(\mathbf{x} | T_1) = 10\%$

náhodně vybraná kočka je černá $p(\mathbf{x} | T_2) = 20\%$

náhodně vybrané zvíře je černé $p(\mathbf{x}) = 0,7 \times 0,1 + 0,3 \times 0,2 = 0,13$

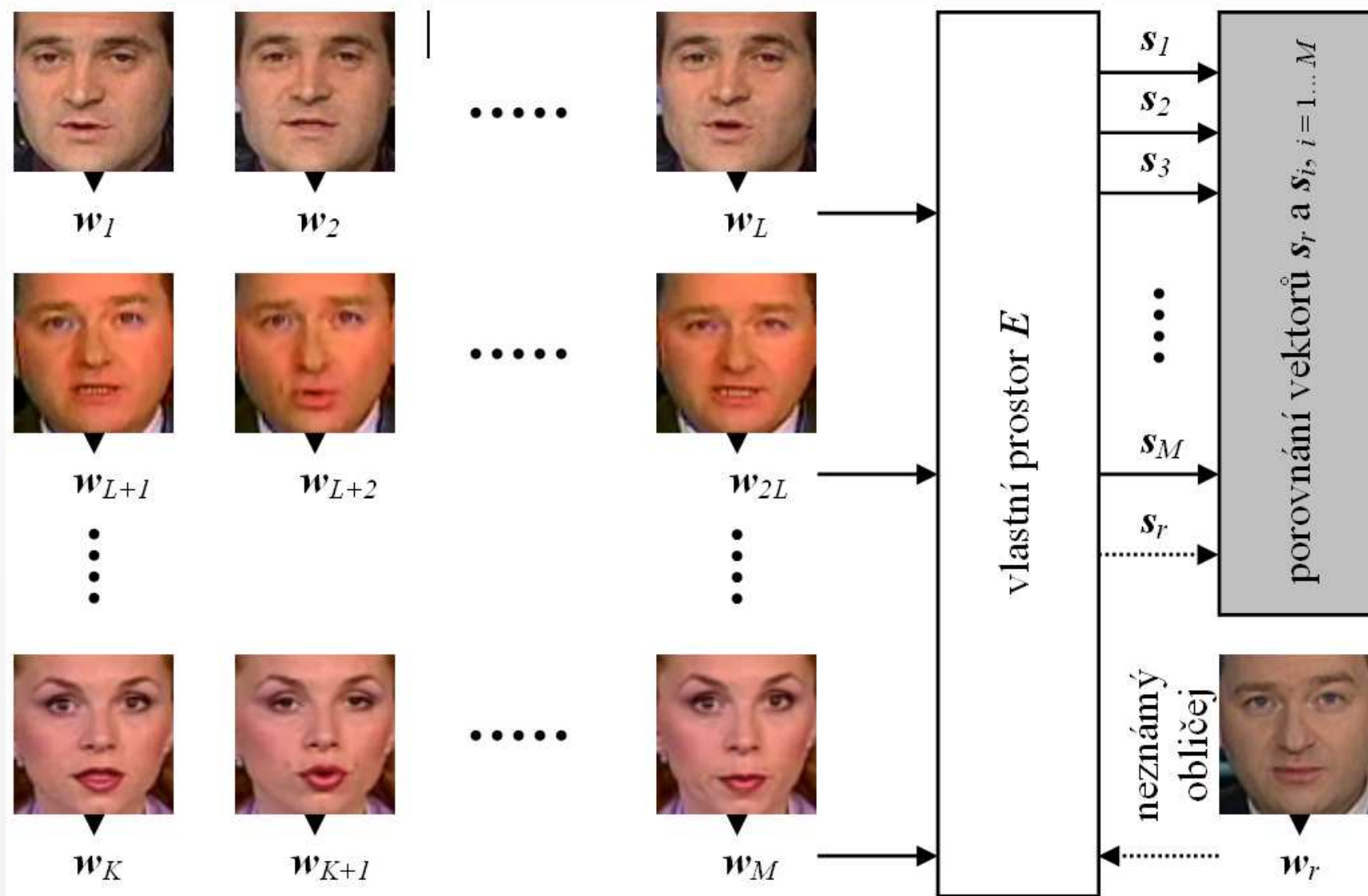
pozorované černé zvíře je pes $p(T_1 | \mathbf{x}) = (0,1 \times 0,7) / 0,13 = 54\%$

pozorované černé zvíře je kočka $p(T_2 | \mathbf{x}) = (0,2 \times 0,3) / 0,13 = 46\%$





Rozpoznávání obrazů - PCA





Rozpoznávání obrazů - PCA



Př. 9 známých obrázků o velikosti 64 x 64 pixelů, 1 neznámý, postup:

Trénovací část

- 1) Obrázky převedeny na šedotónové, z obrázku vytvořen vektor o délce 4096 – seřazení sloupců (nebo řádků) matice obrazu za sebe
- 2) Ze známých obrázků (vektorů) vytvořena matice W_p – velikost 4096 x 9
- 3) Z řádků matice W_p spočítán průměrný vektor w_p – délka 4096
- 4) Vytvoření matice W – od sloupců W_p odečten w_p
- 5) Vytvoření kovarianční matice $C = W^T * W$ – velikost 9 x 9
- 6) Z matice C spočítány vlastní čísla a jím náležející vlastní vektory
- 7) Z vlastních vektorů vytvořena matice E_p – vlastní vektory seřazené podle velikosti (od největšího k nejmenšímu) vlastního čísla – velikost 9 x 9
- 8) Vytvoření matice (vlastní prostor – EigenSpace) $E = W * E_p$ – velikost 4096 x 9
- 9) Projekce známých vektorů do vlastního prostoru $PI = E^T * W$

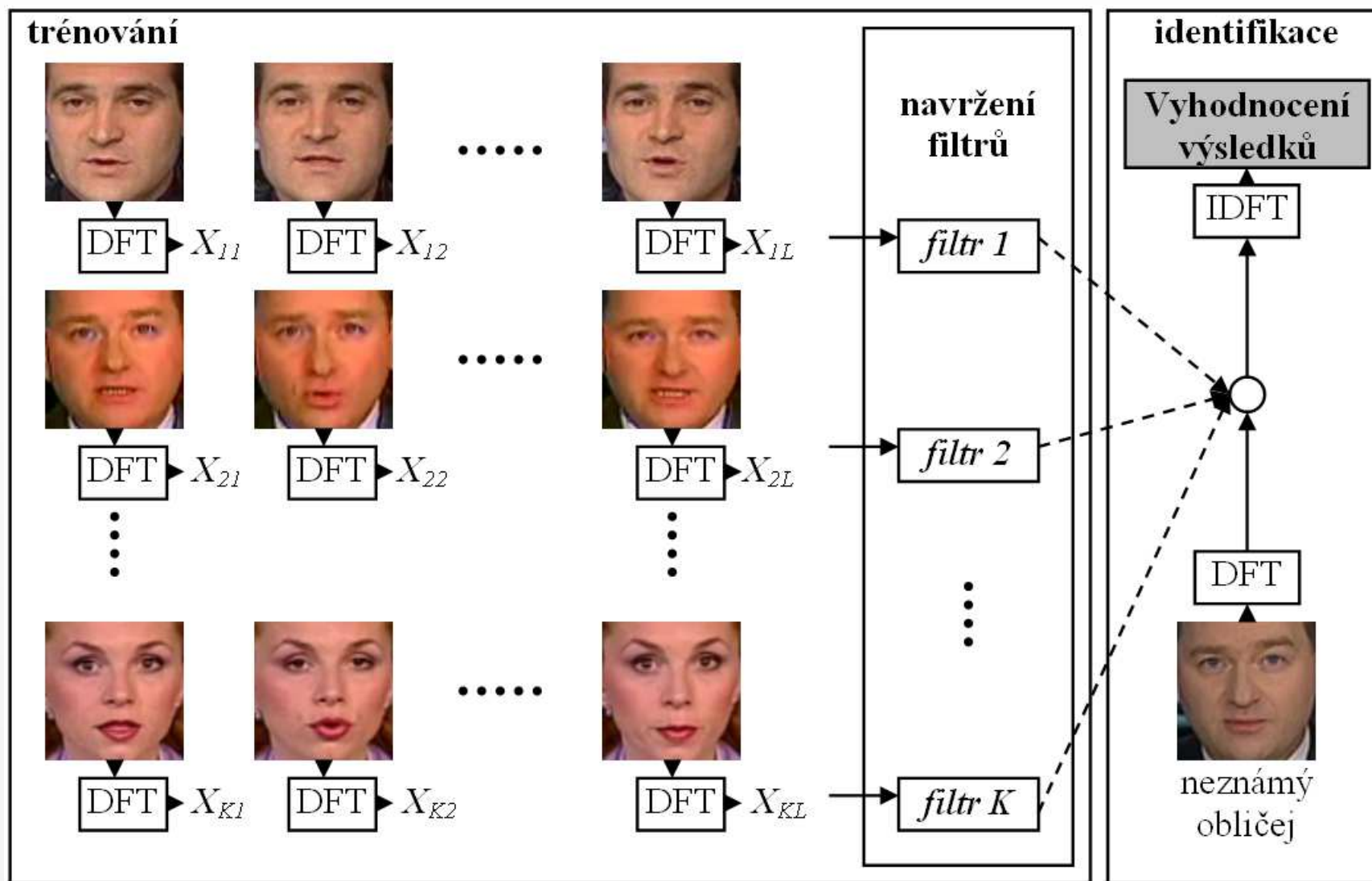
Testovací část

- 1) Převedení neznámého obrázku do stupně šedi a vytvoření vektoru w_p
- 2) Vektor $w_u = w_p - w_p$
- 3) Projekce neznámého vektoru $PT = E^T * w_u$
- 4) Porovnání známých příznakových vektorů $PI(i)$ a neznámého PT – např. dle minimální vzdálenosti





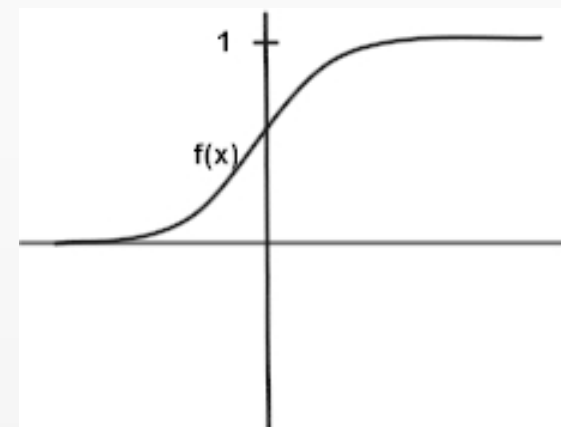
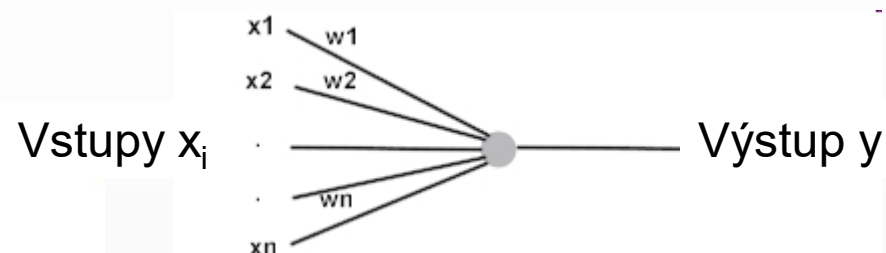
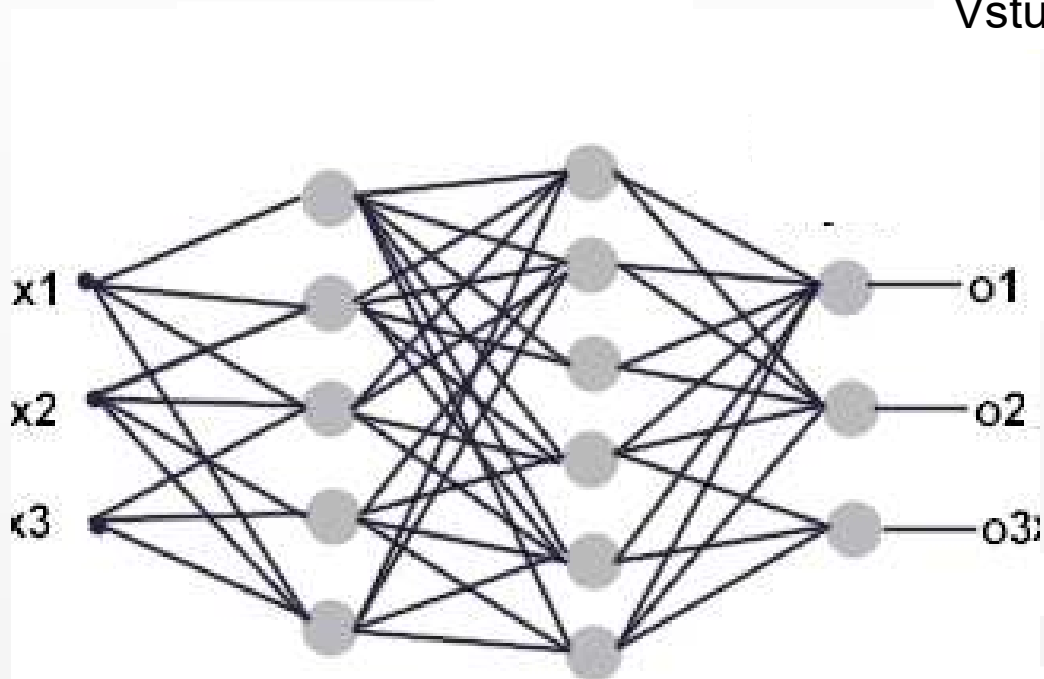
Rozpoznávání obrazů - MACE





Rozpoznávání obrazů - ANN

Umělé neuronové sítě – ANN Artificial Neural Network

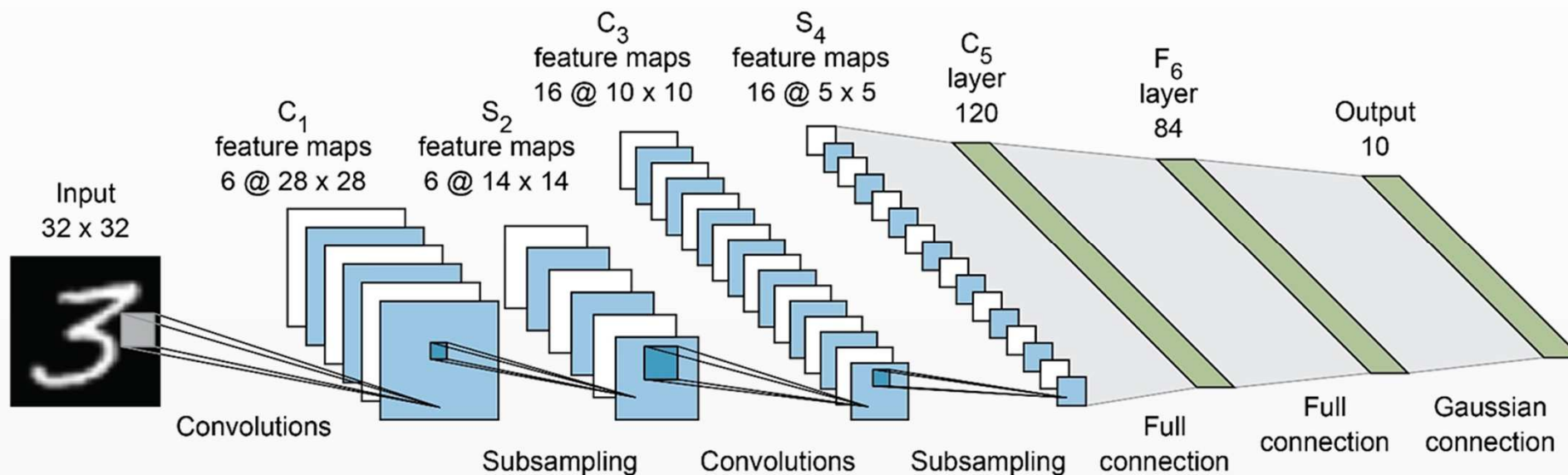


$$y = f\left(\sum_{i=1}^N w_i x_i - T\right)$$

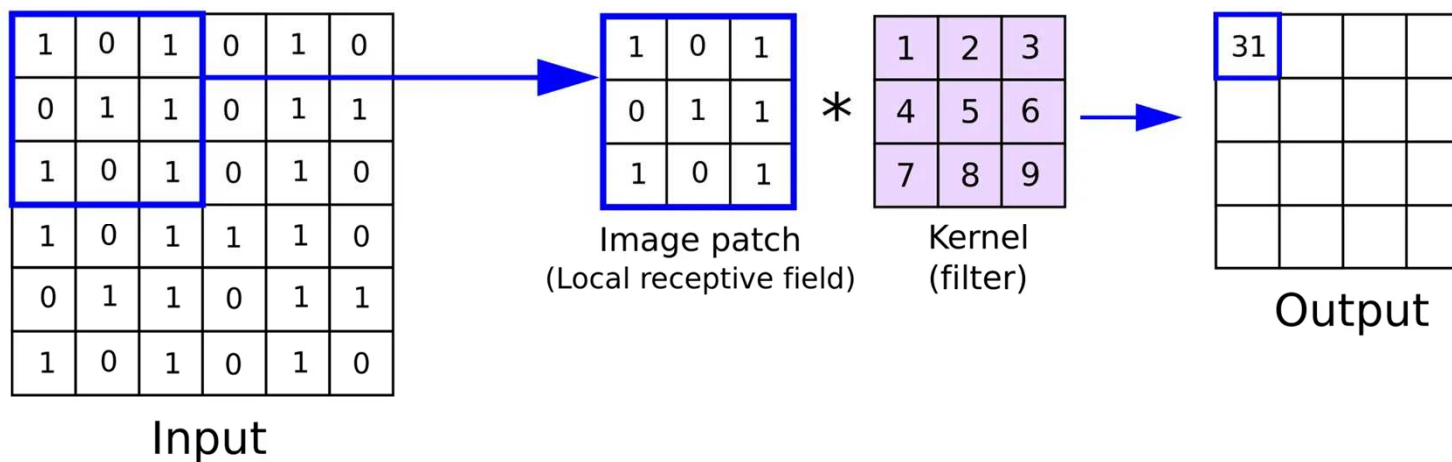




Rozpoznávání obrazů - CNN



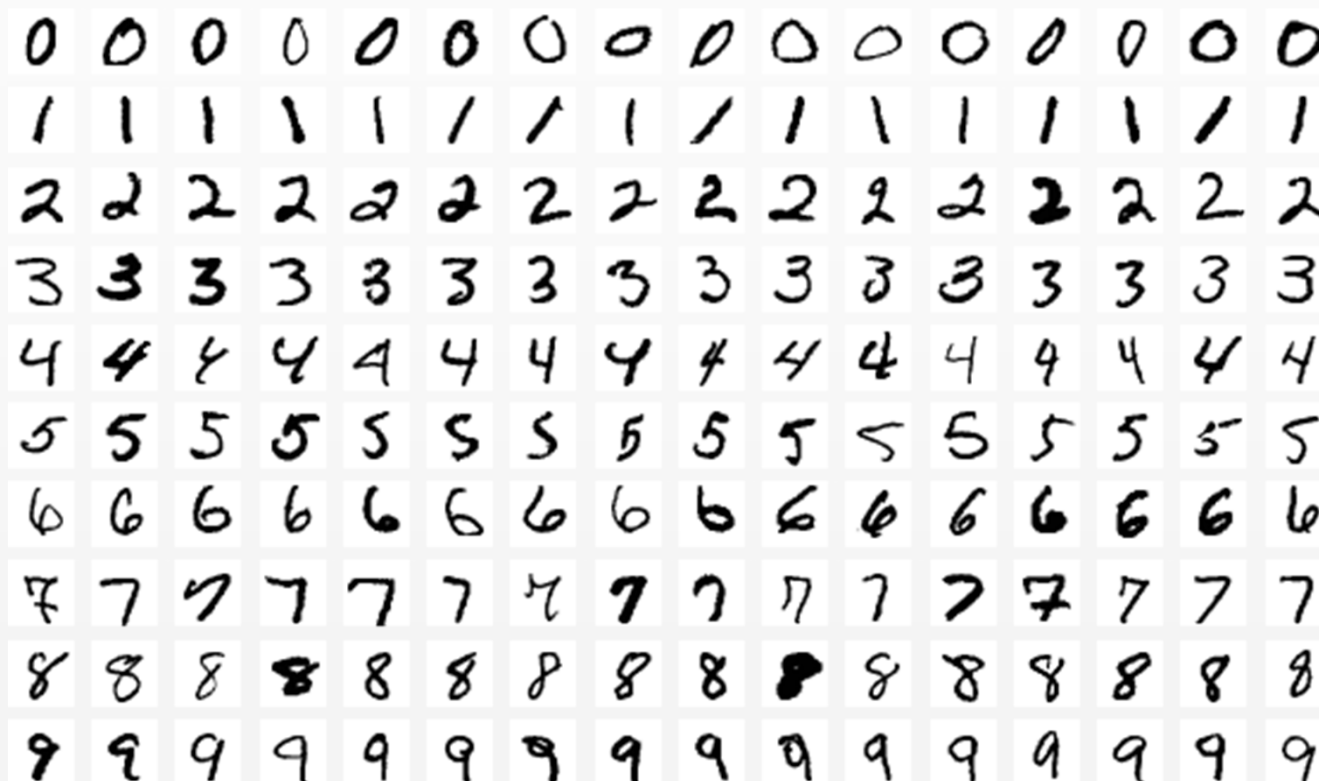
iq.opengenus.org





Rozpoznávání obrazů - CNN

- Modified National Institute of Standards and Technology (MNIST) database (1994)
- 60 000 obrázků pro trénink, 10 000 pro testování





Rozpoznávání obrazů - CNN



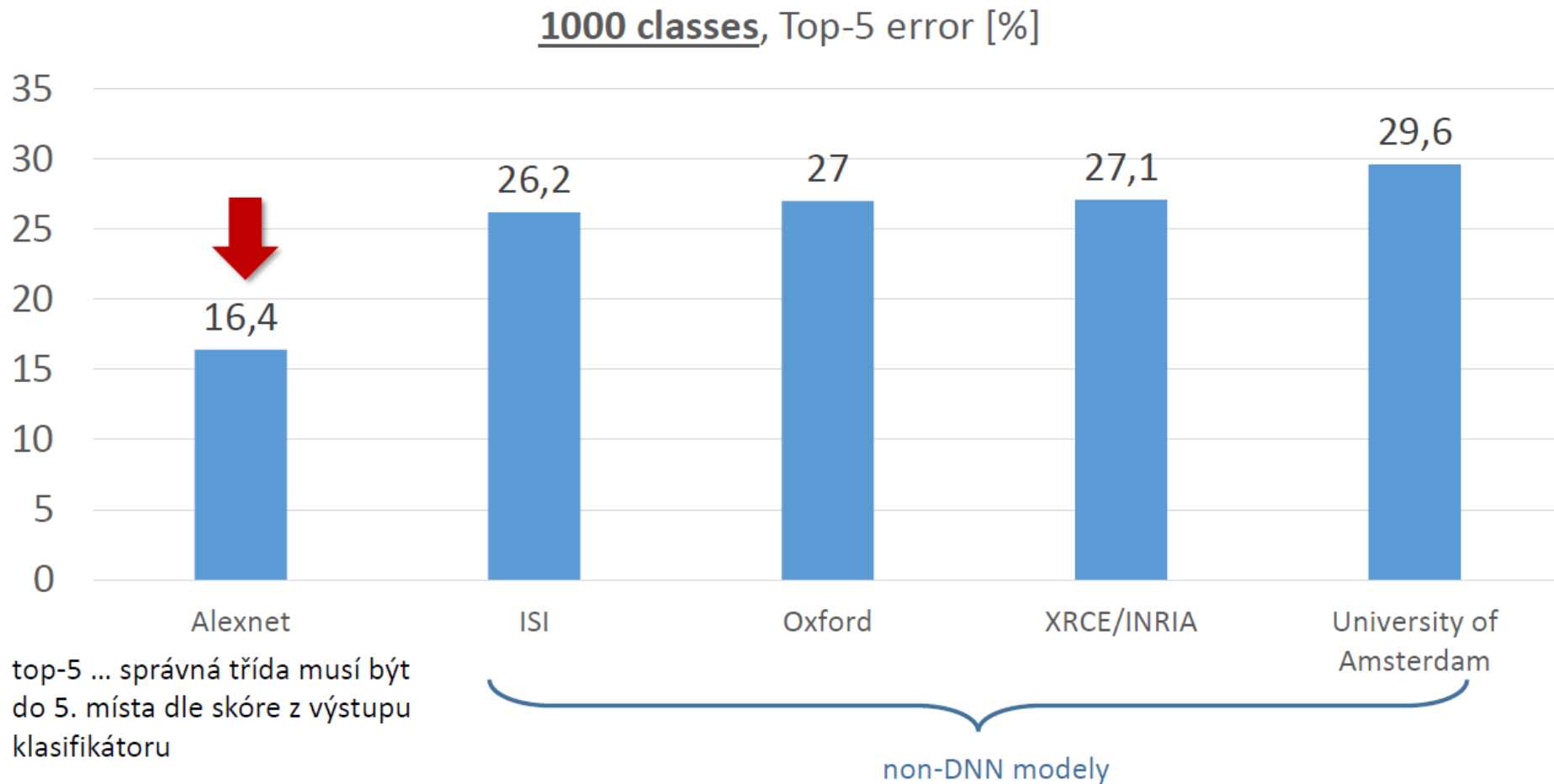
Type	Classifier	Distortion	Preprocessing	Error rate (%)
Linear classifier	Pairwise linear classifier	None	Deskewing	7.6 ^[10]
Decision stream with Extremely randomized trees	Single model (depth > 400 levels)	None	None	2.7 ^[28]
K-Nearest Neighbors	K-NN with rigid transformations	None	None	0.96 ^[29]
K-Nearest Neighbors	K-NN with non-linear deformation (P2DHMDM)	None	Shiftable edges	0.52 ^[30]
Boosted Stumps	Product of stumps on Haar features	None	Haar features	0.87 ^[31]
Non-linear classifier	40 PCA + quadratic classifier	None	None	3.3 ^[10]
Random Forest	Fast Unified Random Forests for Survival, Regression, and Classification (RF-SRC) ^[32]	None	Simple statistical pixel importance	2.8 ^[33]
Support-vector machine (SVM)	Virtual SVM, deg-9 poly, 2-pixel jittered	2002	Deskewing	0.56 ^[34]
Neural network	2-layer 784-800-10	None	None	1.6 ^[35]
Neural network	2-layer 784-800-10	2003	Elastic distortions	0.7 ^[35]
Deep neural network (DNN)	6-layer 784-2500-2000-1500-1000-500-10	2010	Elastic distortions	0.35 ^[36]
Convolutional neural network (CNN)	6-layer 784-40-80-500-1000-2000-10	2016	Expansion of the training data	0.31 ^[37]
Convolutional neural network	6-layer 784-50-100-500-1000-10-10	None	Expansion of the training data	0.27 ^[38]
Convolutional neural network (CNN)	13-layer 64-128(5x)-256(3x)-512-2048-256-256-10	None	None	0.25 ^[22]
Convolutional neural network	Committee of 35 CNNs, 1-20-P-40-P-150-10	Elastic distortions	Width normalizations	0.23 ^[17]
Convolutional neural network	Committee of 5 CNNs, 6-layer 784-50-100-500-1000-10-10	None	Expansion of the training data	0.21 ^{[24][25]}
Random Multimodel Deep Learning (RMDL)	10 NN-10 RNN - 10 CNN	None	None	0.18 ^[27]
Convolutional neural network	Committee of 20 CNNs with Squeeze-and-Excitation Networks ^[39]	None	Data augmentation	0.17 ^[40]
Convolutional neural network	Ensemble of 3 CNNs with varying kernel sizes	2020	Data augmentation consisting of rotation and translation	0.09 ^[41]

https://en.wikipedia.org/wiki/MNIST_database



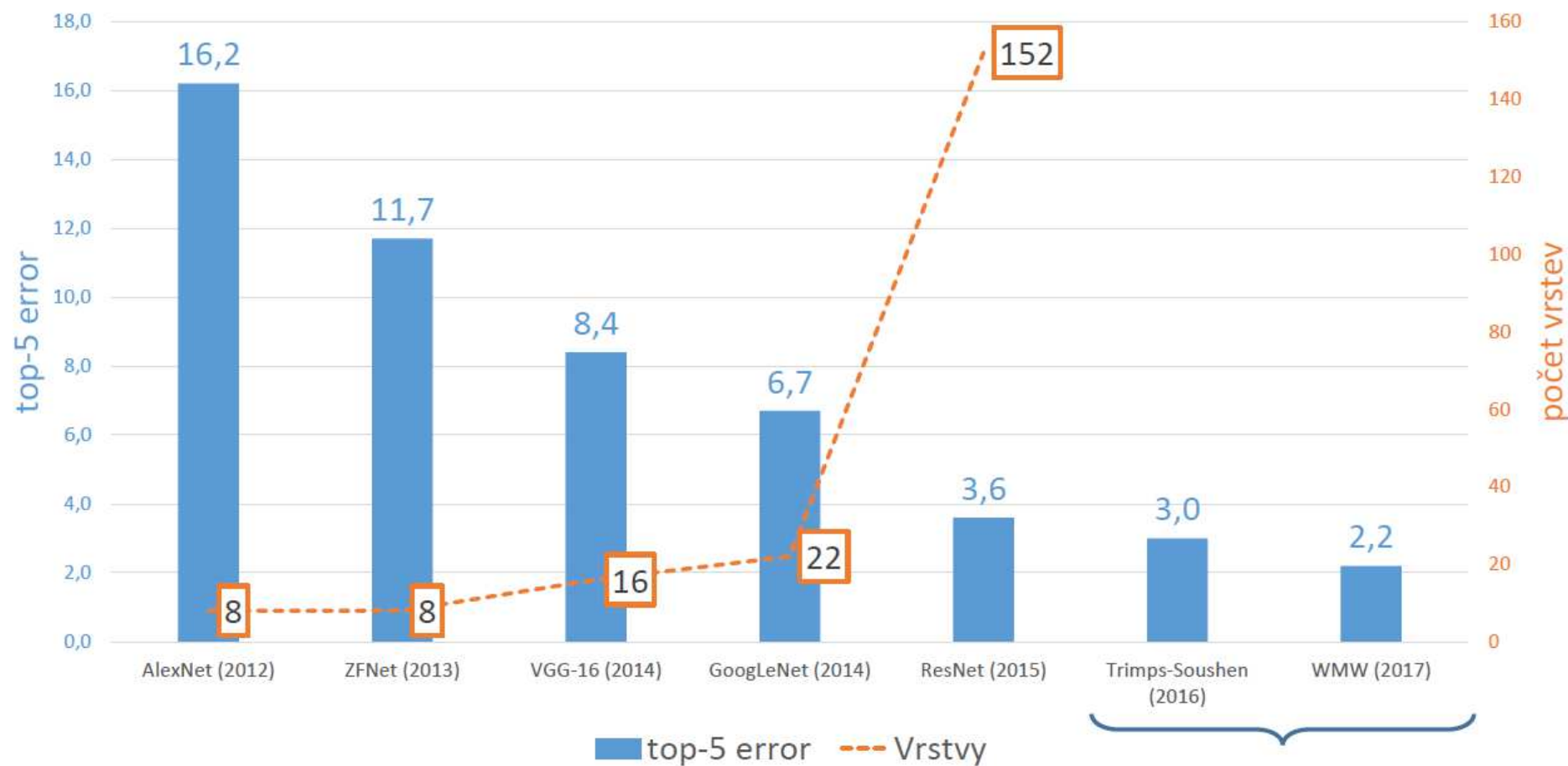


CNN, ImageNet - Alexnet (2012)





CNN, ImageNet

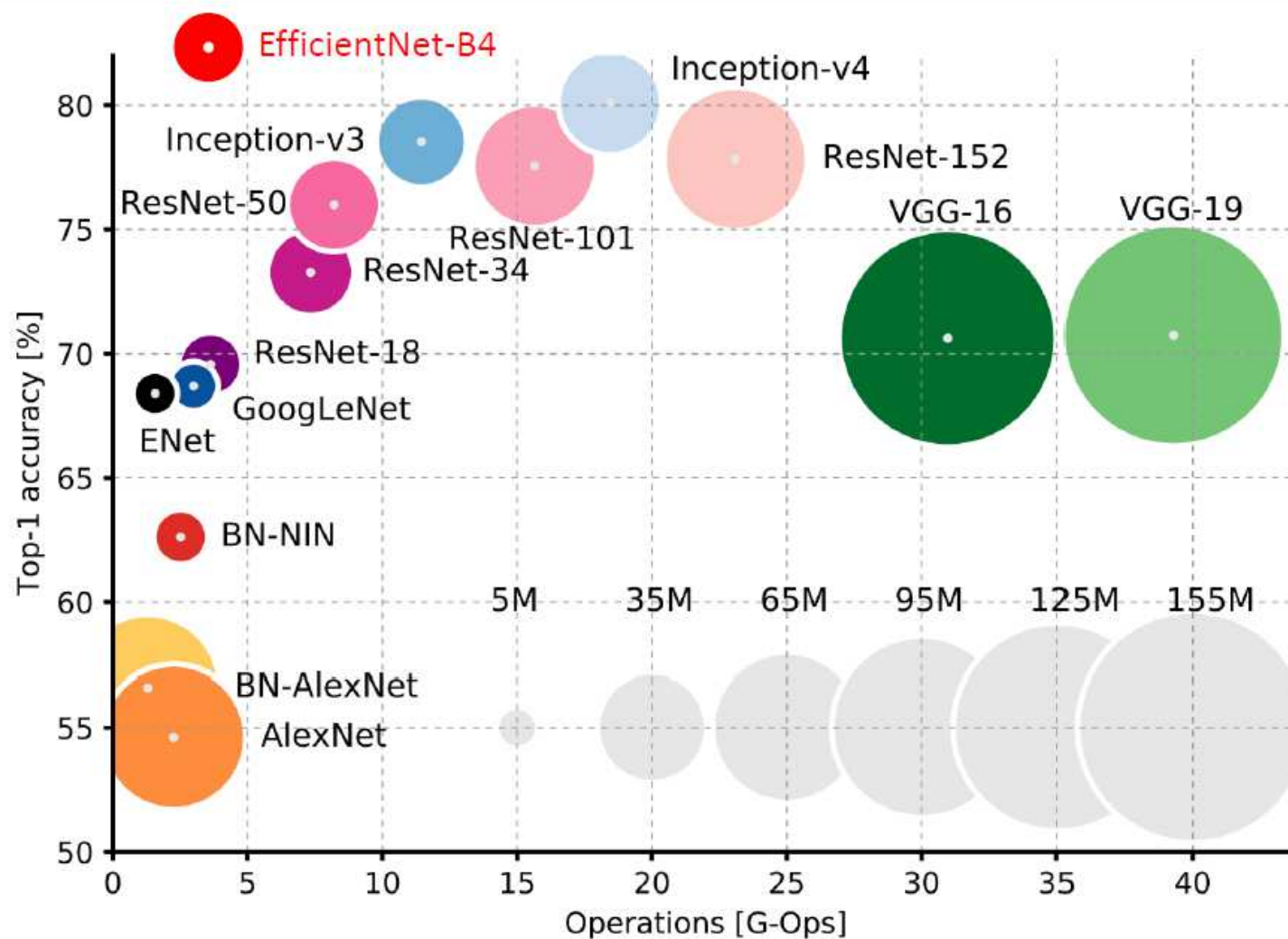


kombinace více modelů (ensemble)





CNN, ImageNet



obrázek: [Canziani et al.: "An Analysis of Deep Neural Network Models for Practical Applications"](https://arxiv.org/abs/1802.02762)

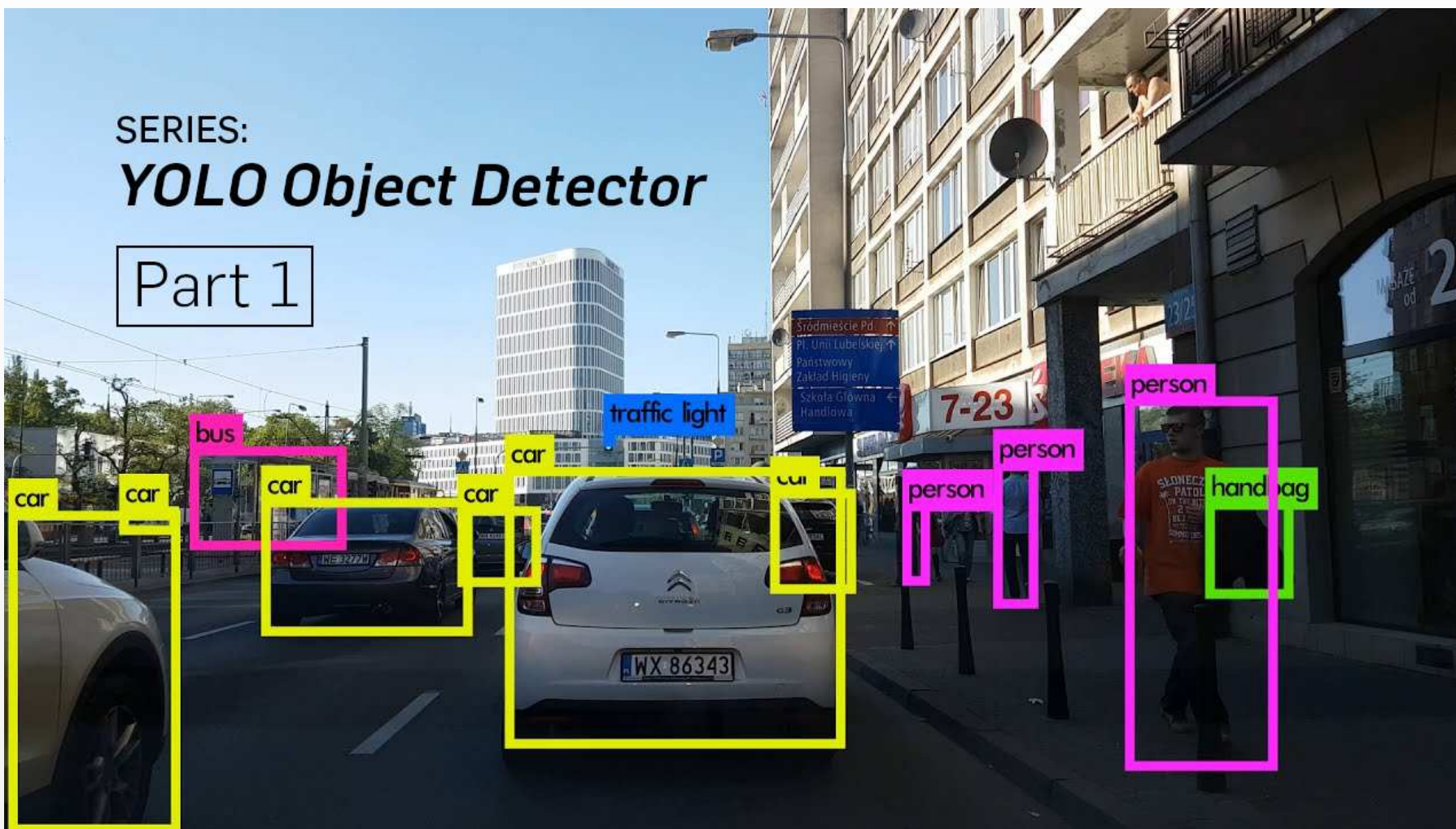




CNN, YOLO

SERIES:
YOLO Object Detector

Part 1



<http://www.ite.tul.cz>

