

Rozpoznávání snímků, detekce objektů, moderní trendy

Strojové vidění a zpracování obrazu (BI-SVZ)

Úlohy v oboru počítačového vidění

- Klasifikace obrázků
- Lokalizace objektů
- Detekce objektů
- Sémantická segmentace
- Segmentace instance
- Panoptická segmentace
- Textový popis obrázků
- Stovky dalších...

Nejčastější úlohy v počítačovém vidění

Classification



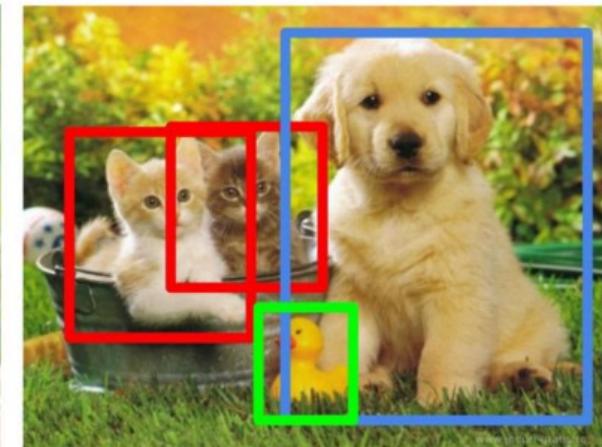
CAT

Classification + Localization



CAT

Object Detection



CAT, DOG, DUCK

Instance Segmentation



CAT, DOG, DUCK

Single object

Multiple objects

Přístupy k úlohám detekce a rozpoznávání

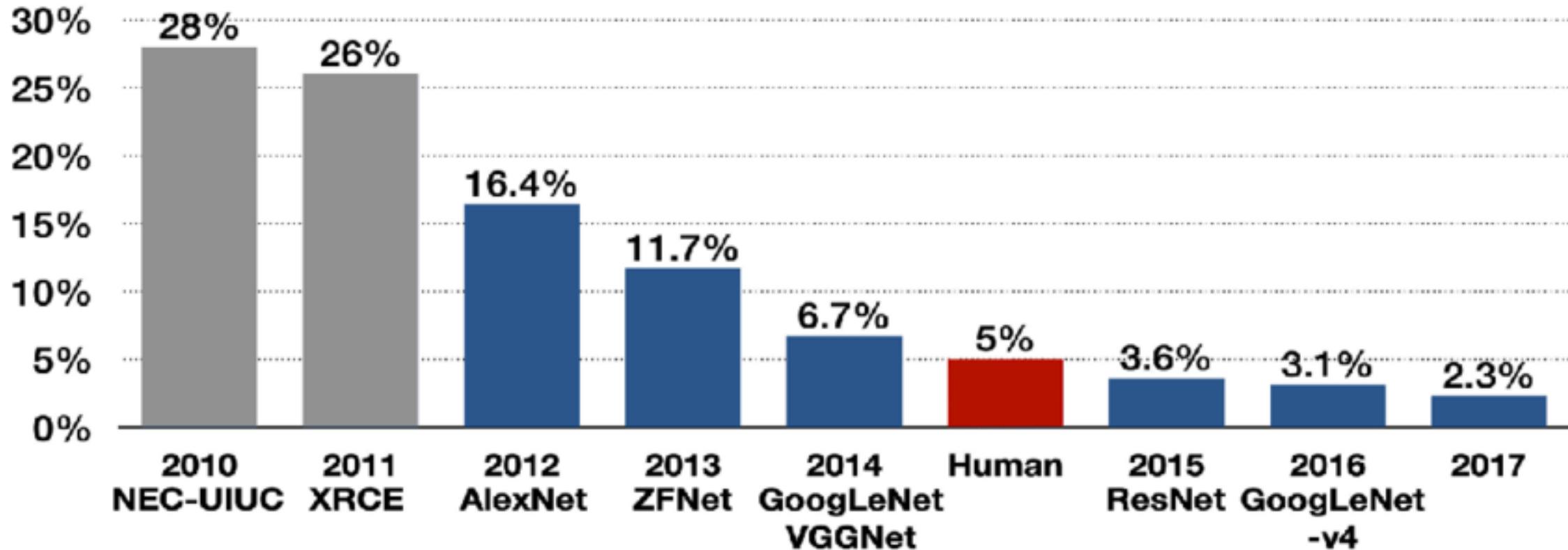
- Tradiční způsoby
 - Barevné rozpoznávání
 - Tvarové rozpoznávání
 - Šablonové rozpoznávání
 - Výpočty příznaků
 - Klasifikace
- Metody založené na hlubokém učení (deep learning)
- Kombinace předchozích způsobů

Historie rozpoznávání snímků a detekce objektů

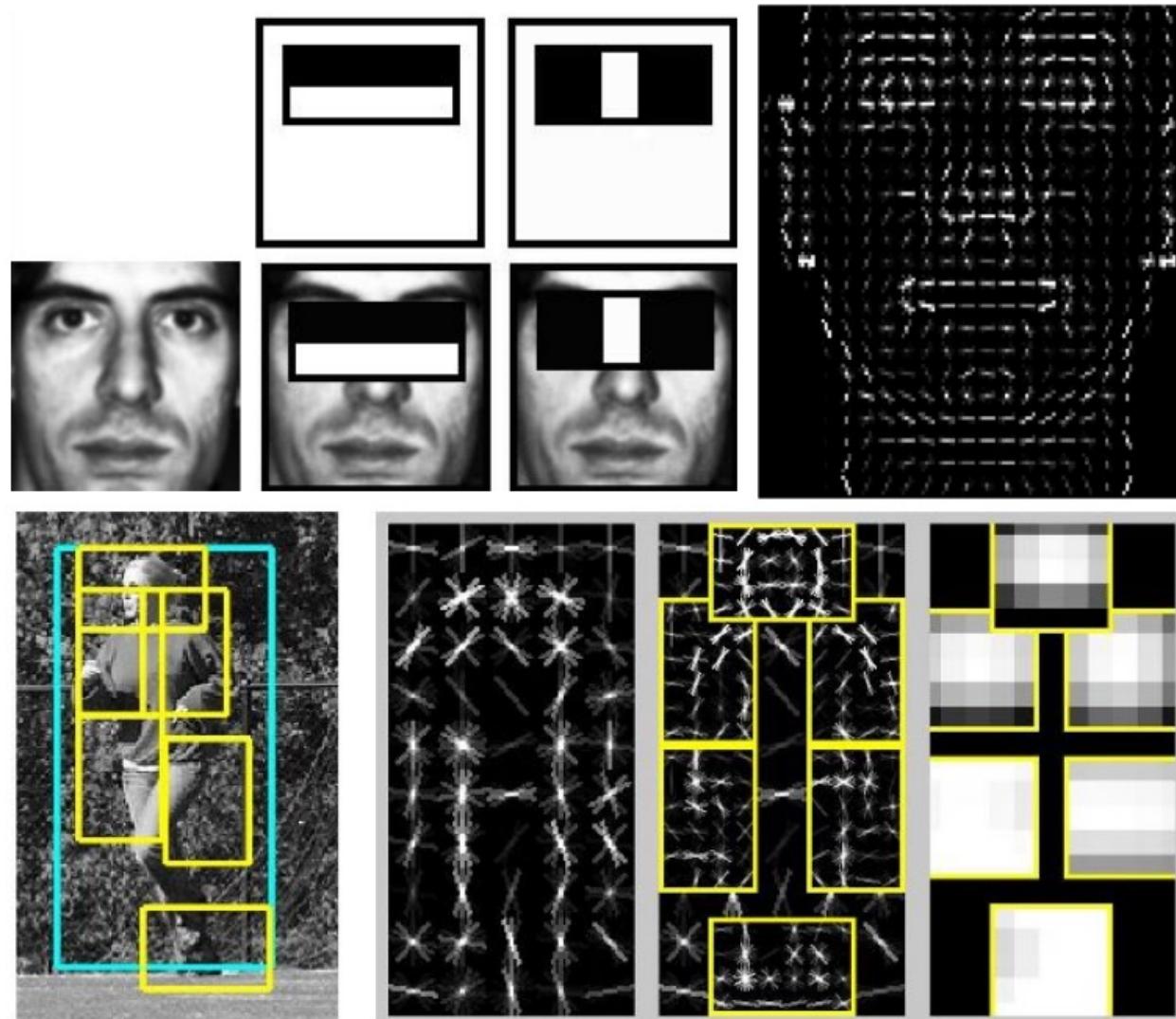
- 2001
 - První real-time algoritmus pro detekci obličejů od Paul Viola a Michael Jones
 - Známý jako Haar Cascades k nalezení v OpenCV
- 2005
 - První použitelný algoritmus pro detekci osob od Navneet Dalal and Bill Triggs.
 - Známý jako deskriptor Histograms of Oriented Gradients (HOG), k nalezení v OpenCV
- 2012
 - Deep learningová síť od autorů Alex Krizhevsky, Ilya Sutskever, a Geoffrey Hinton šokuje svět výhrou v soutěži ImageNet dramatickým zvýšením přesnosti rozpoznávání
- 2015
 - Deep learning je mainstream, algoritmy překonaly přesnost rozpoznávání lidí
 - Přesnost klasifikace snímků převyšuje 95 %

ImageNet-1k benchmark

Top-5 error



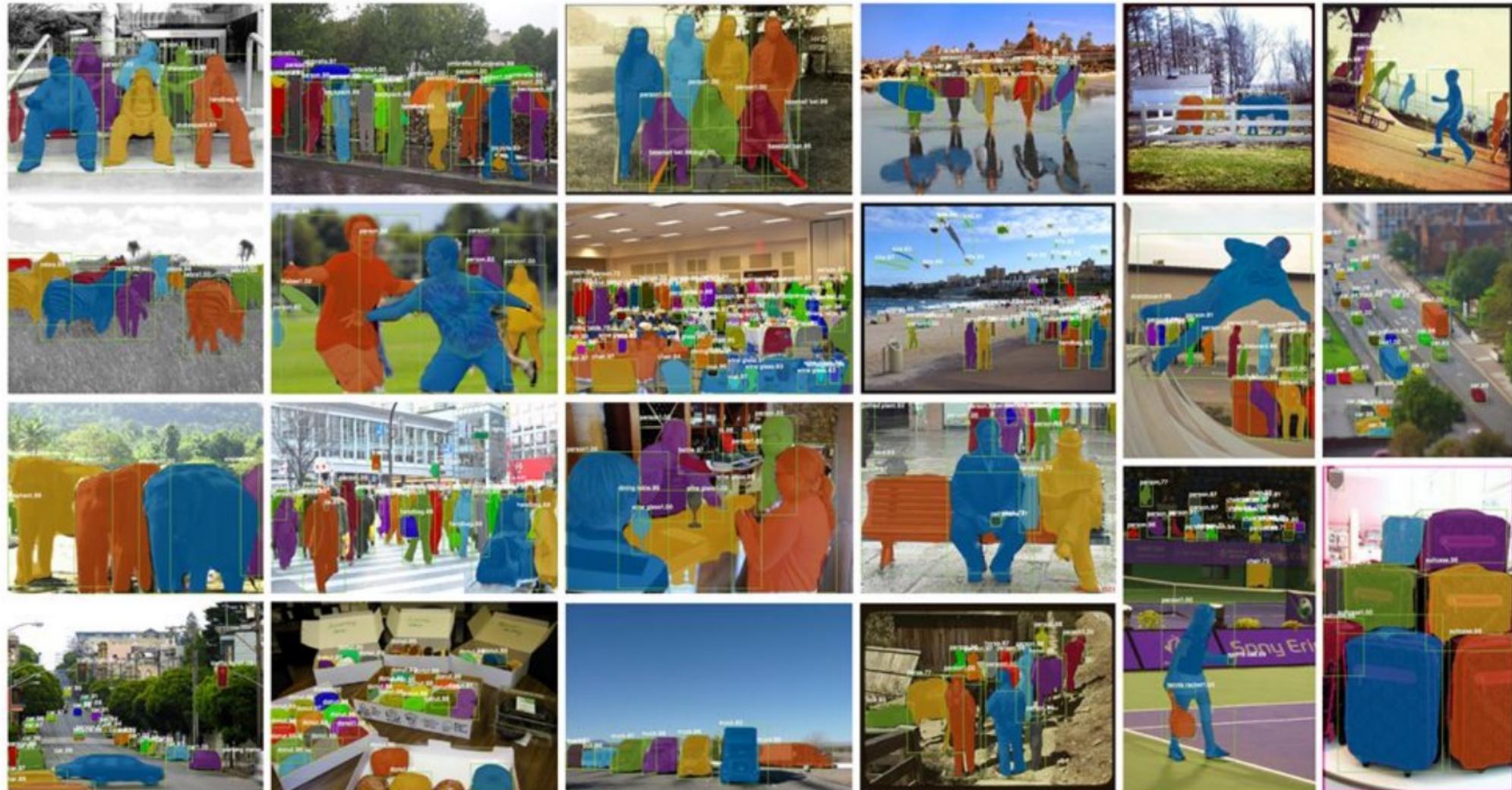
Detekce objektů 2001 – 2007



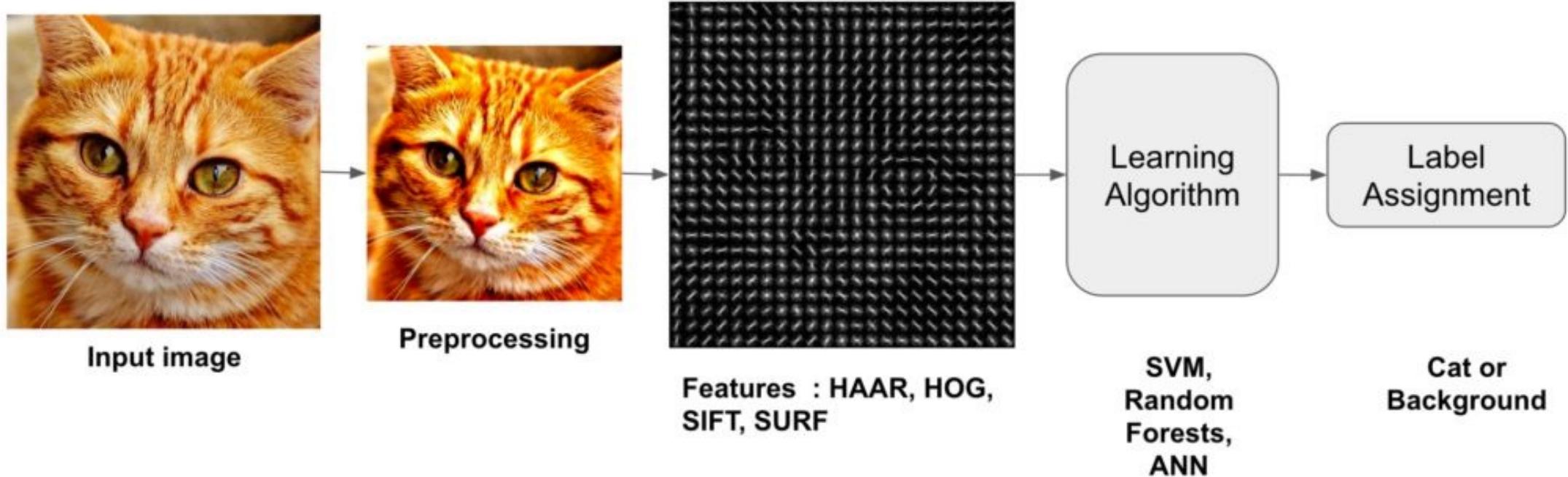
Detekce objektů 2007 – 2012



Detekce objektů nyní



Tradiční techniky počítačového vidění - pipeline



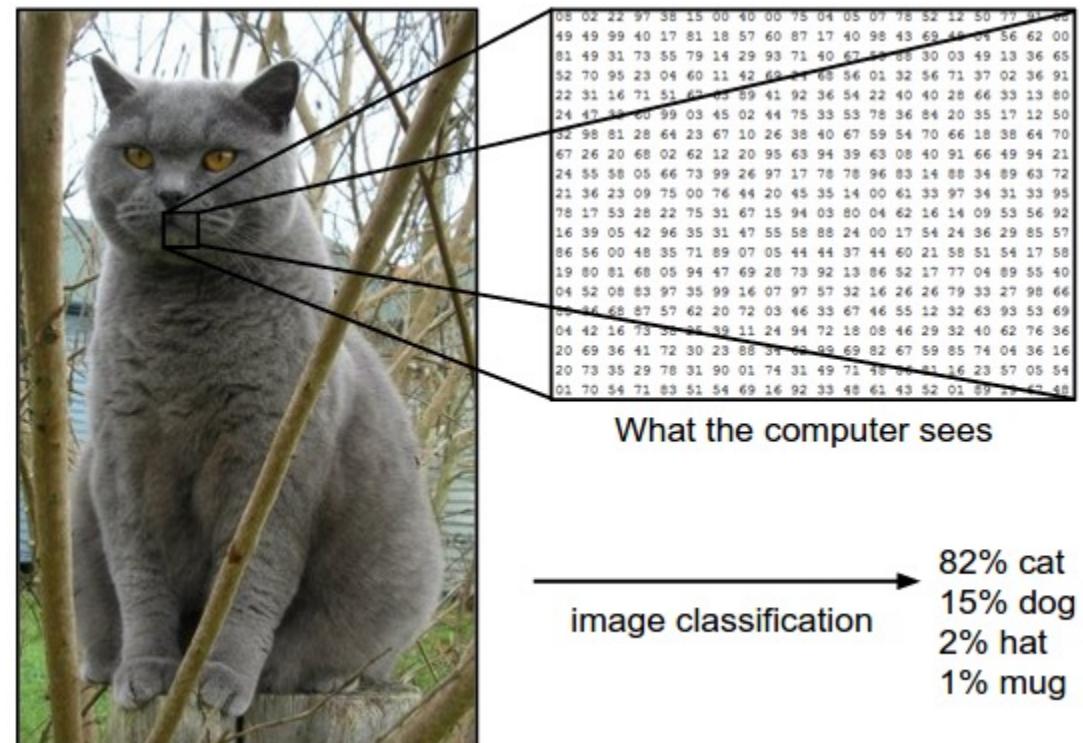
- Většina tradičních způsobů respektuje tuto pipeline, deep learningové algoritmy zpravidla přeskakují část s předzpracováním a extrakcí příznaků
- Představme nyní pipeline na problémů klasifikace

Pipeline rozpoznávání obrazu – předzpracování

- Ve většině případu je vstupní obraz předzpracován, časté operace:
 - Úprava jasu a kontrastu
 - Gamma korekce
 - Ekvalizace histogramu
 - Odečtení průměru obrazu a vydělení standardní odchylkou (normalizace)
 - Převod do různých barevných prostorů (RGB na HSV)
 - Perspektivní transformace
 - Ořez a škálování do předem daných rozměrů (kvůli extrakci příznaků)
- Trik je v tom, že nikdo ve skutečnosti neví, jaké operace předzpracování je potřeba
- Vše vychází z experimentů, kde ověřujeme to, který druh předzpracování poskytuje přesnější výsledky.

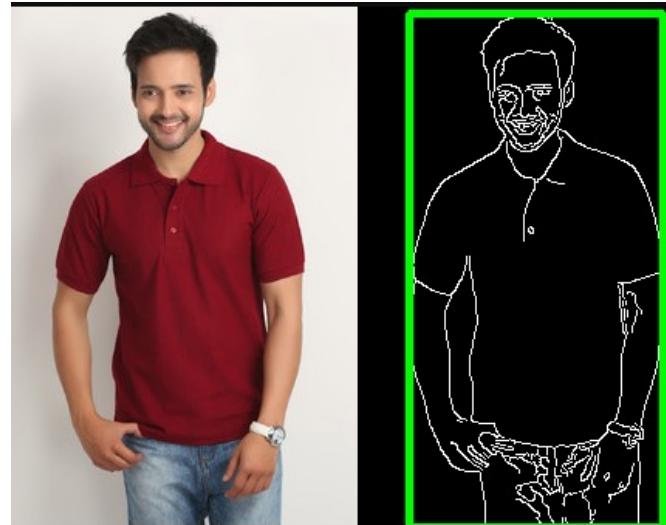
Pipeline rozpoznávání obrazu – extrakce příznaků

- Vstupní předzpracovaný obraz má v sobě příliš mnoho informací, které nejsou pro klasifikaci nutné
- Obraz s šířkou 248, výškou 400, v RGB prostoru obsahuje 297 600 čísel

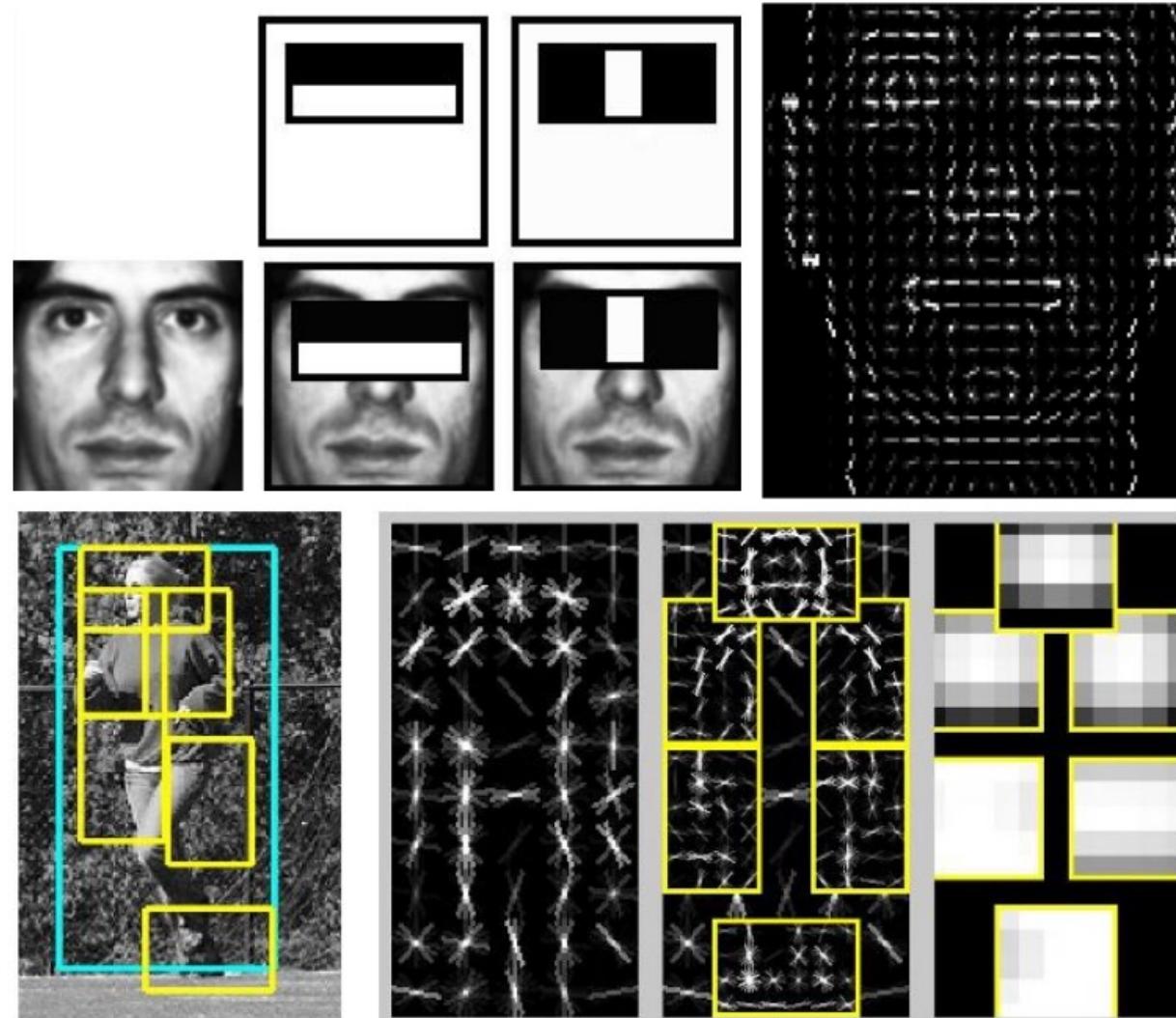


Pipeline rozpoznávání obrazu – extrakce příznaků

- Proto je prvním krokem v klasifikaci obrázků zjednodušení obrazu extrahováním důležitých informací (příznaků) obsažených v obrazu a vynecháním ostatních informací
- Pokud budeme chtít hledat knoflíky na košili v obrazu, zjistíme, že máme v okolí knoflíků významně rozdílné hodnoty RGB pixelů
- Nicméně, pomocí spuštění hranového detektoru můžeme obrázek zjednodušit a stále dokážeme snadno rozpoznat kruhový tvar knoflíků. Tedy zachováváme klíčové informace a zahazujeme nepotřebné (RGB hodnoty)



Pipeline rozpoznávání obrazu – extrakce příznaků



Pipeline rozpoznávání obrazu – extrakce příznaků

- V tradičních přístupech je navrhování extrahovaných příznaků rozhodující pro přesnost algoritmu
- Existují samozřejmě robustnější způsoby k extrakci příznaků, než je samotná hranová detekce, mezi nejznámější patří:
 - [Haar Cascades](#)
 - [Histogram of Oriented Gradients](#)
 - [Harris corners](#)
 - [SIFT](#)
 - [SURF](#)
 - [Local Binary Patterns \(LBP\)](#)
 - Histogramy

Pipeline rozpoznávání obrazu – klasifikátor

- Již víme, jak převést obraz na vektor důležitých příznaků
- Nyní potřebujeme vzít tento vektor, dát ho na vstup klasifikátoru a rozhodnout o jakou třídu objektu se jedná – kočka nebo pes
- Předtím než je klasifikátor použitelný, musíme jej natrénovat na mnoho příkladech
- Je možno použít velké množství klasifikátorů
 - Donesavna velmi používané Support Vector Machines - SVM
 - Více o klasifikátorech viz předměty [BI-ML1](#) a [BI-ML2](#)

Oblíbený challenge – dog vs mop



Problémy tradičních technik klasifikace?

Problémy tradičních technik klasifikace

- Rotace
- Barevnost
- Osvětlení
- Škálování
- Doménový posun v datech (concept drift, domain shift)
- **Obecně nedostatečná generalizace**

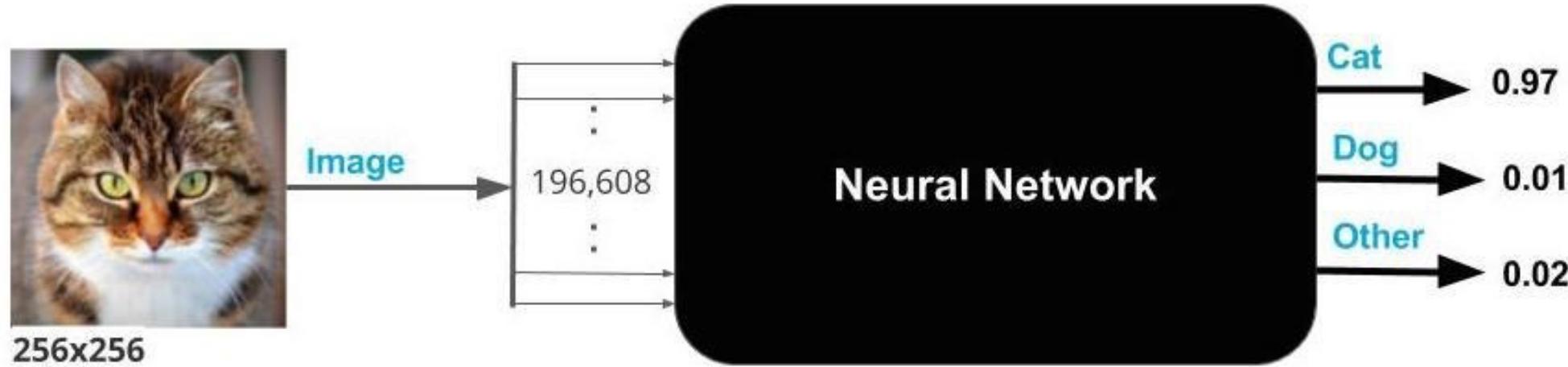
Deep learning, neuronové sítě - pipeline

- Minimální úsilí na předzpracování snímku a využívání black-boxu



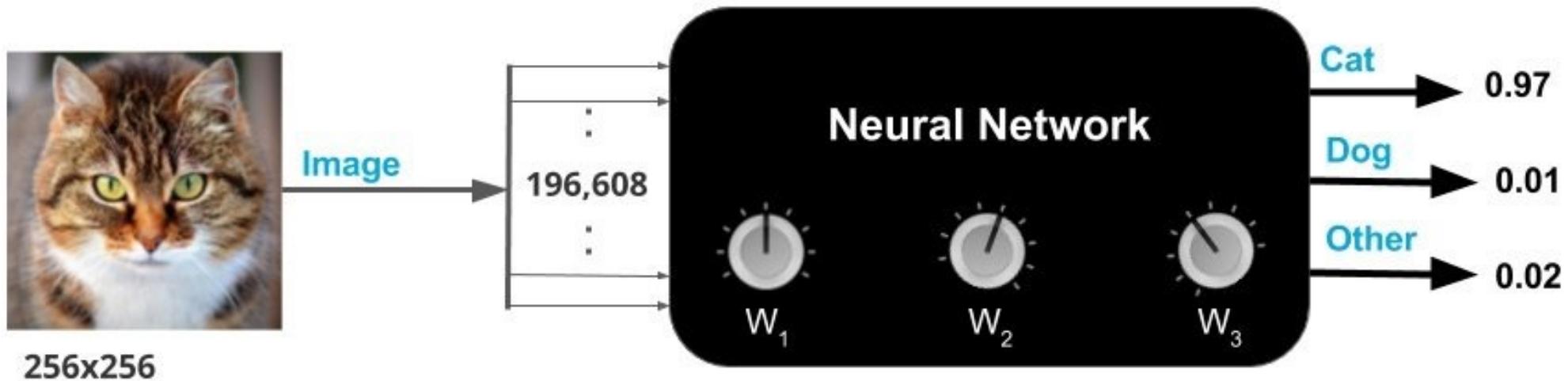
Deep learning, neuronové sítě - pipeline

- Vstupní snímek však musíme převést alespoň do vhodné reprezentace – vektoru fixní délky

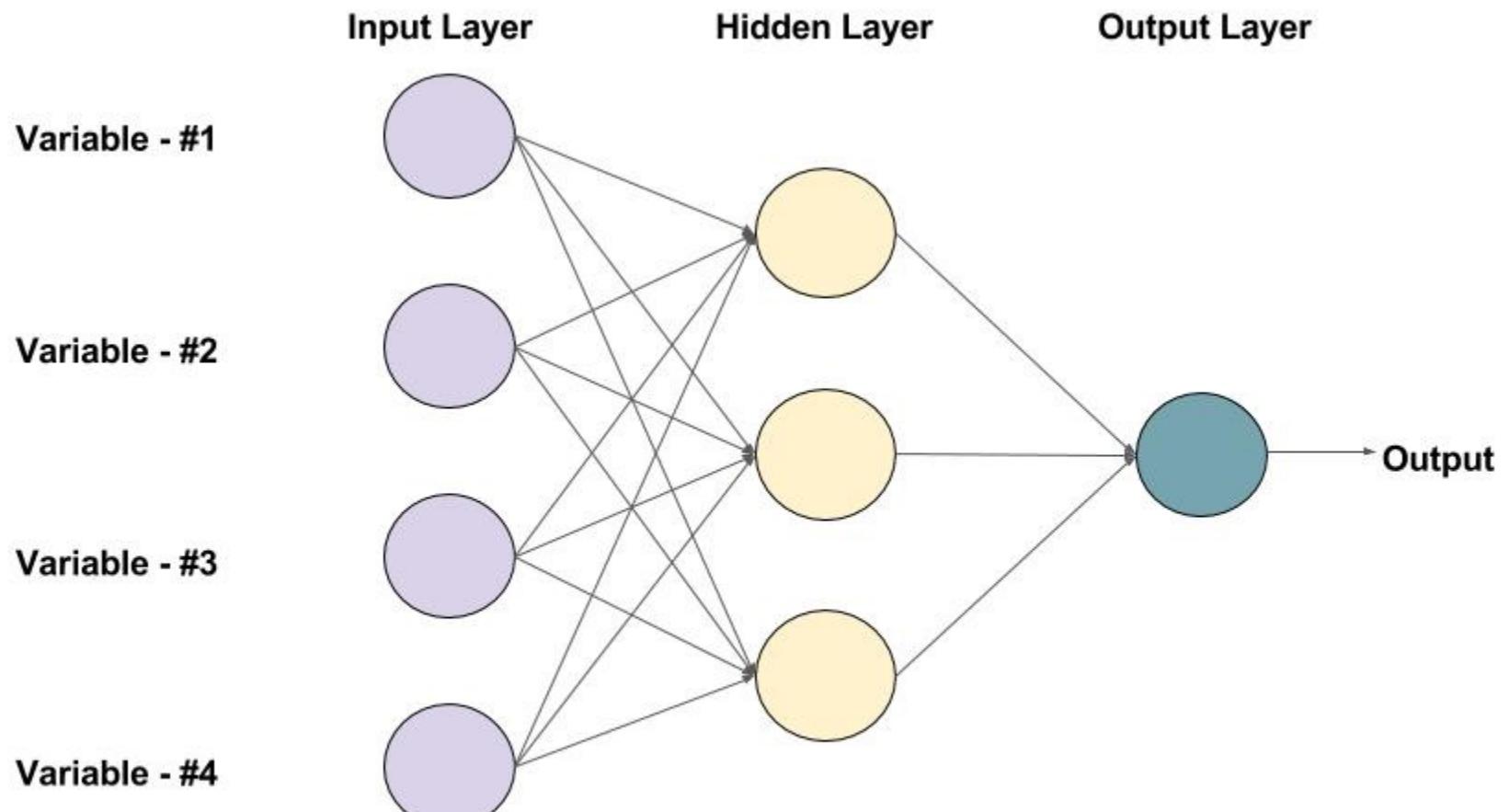


Deep learning, neuronové sítě - pipeline

- Neuronové sítě k predikci využívají miliony vnitřních parametrů (váhy neuronů), které je k dosažení smysluplných výsledků, nutno správně natrénovat



Feed-forward NN

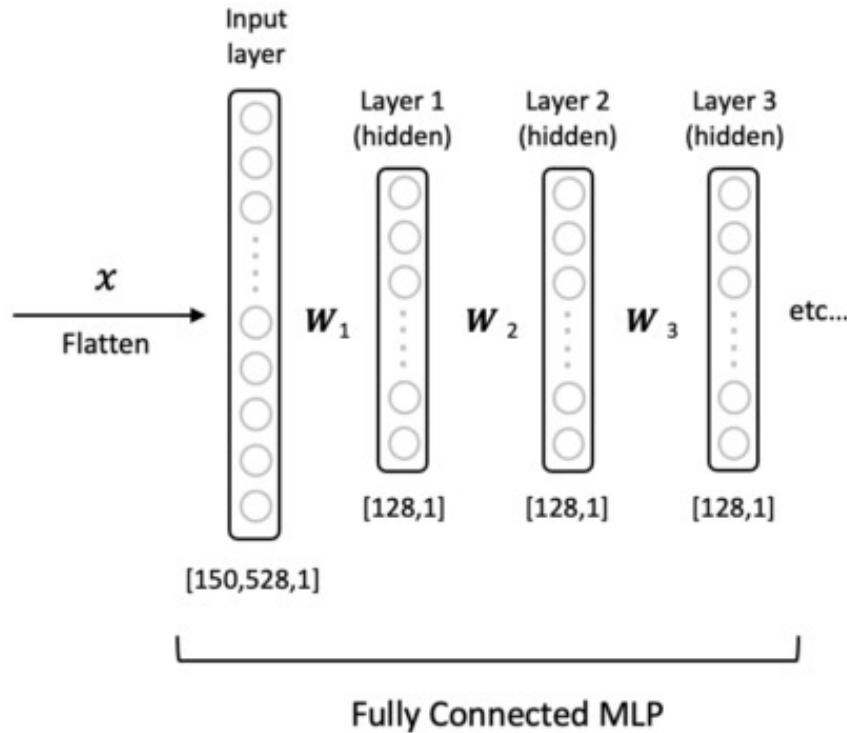


An example of a Feed-forward Neural Network with one hidden layer (with 3 neurons)

Feed-forward NN



Input Image
 $224 \times 224 \times 3$



Number of Trainable Weights

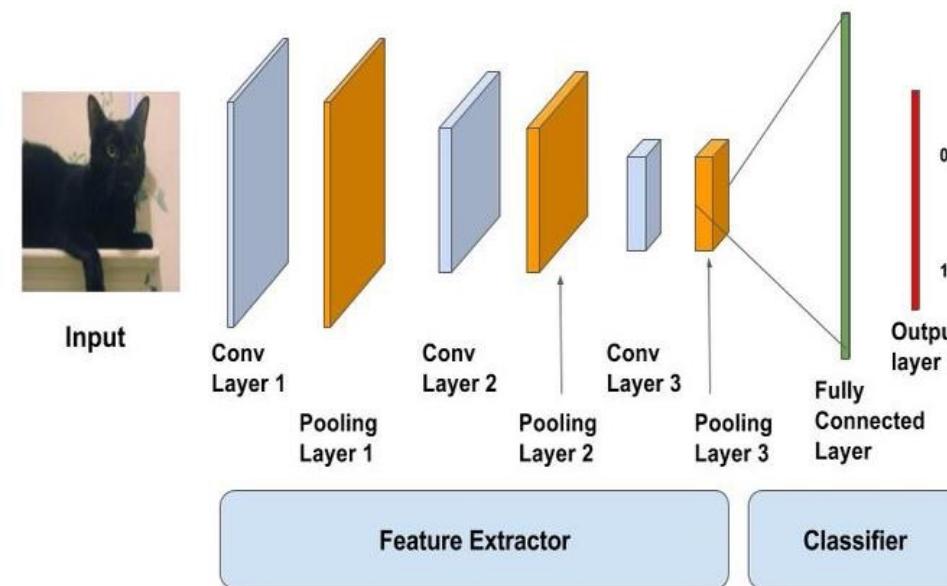
$$W_1 = 150,528 \times 128 = 19,267,584 \text{ Weights}$$

$$W_2 = 19,267,584 \times 128 \sim 2.4 \text{ Billion}$$

$$W_3 = 2.4B \times 128 > 300 \text{ Billion}$$

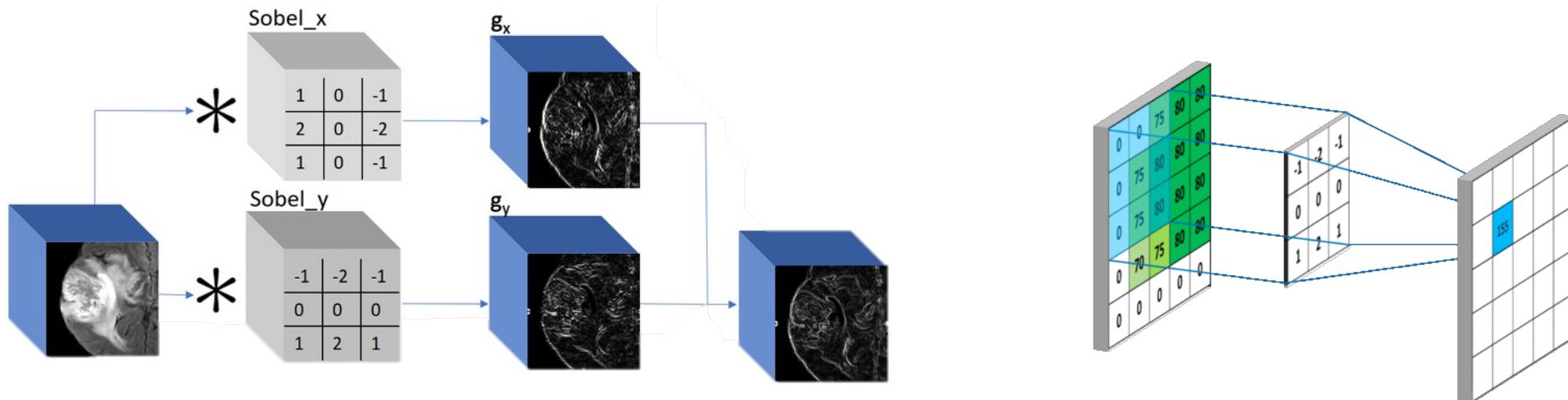
Convolutional neural networks - CNN

- Pro obrázky se často používají **Convolution Neural Networks** místo samotné **Feed-forward NN**
 - CNN se skládá z konvolučních a pooling prstev, pomocí kterých se snažíme získat vhodné atributy.
 - Následně tyto atributy použijeme na vstup Feed-forward NN pro klasifikaci



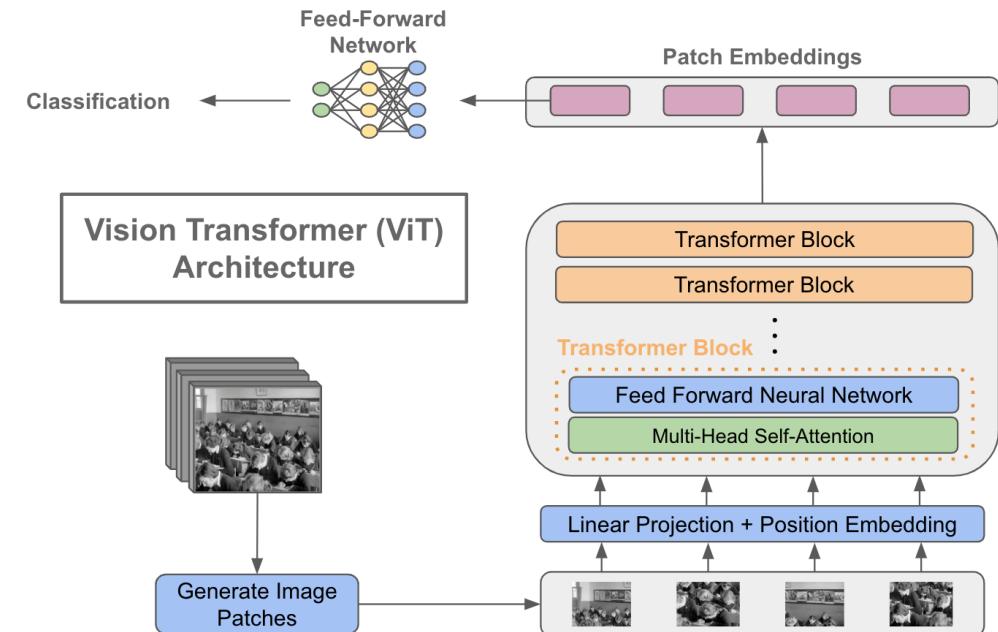
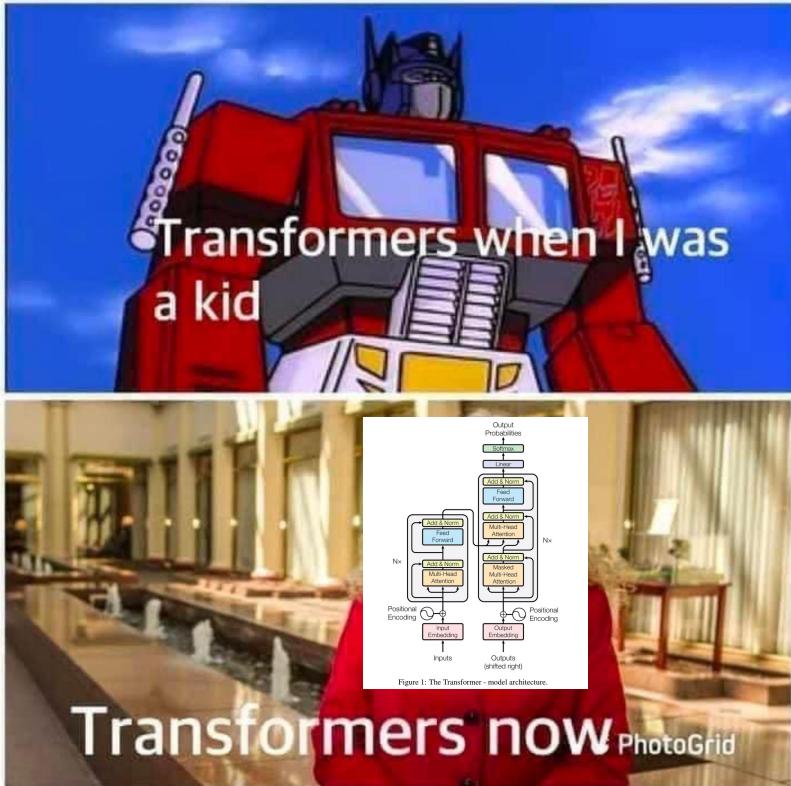
CNN – konvoluce

- Operace konvoluce se u tradičních technik používá pro hranovou detekci, unsharp masking apod.



Transformery

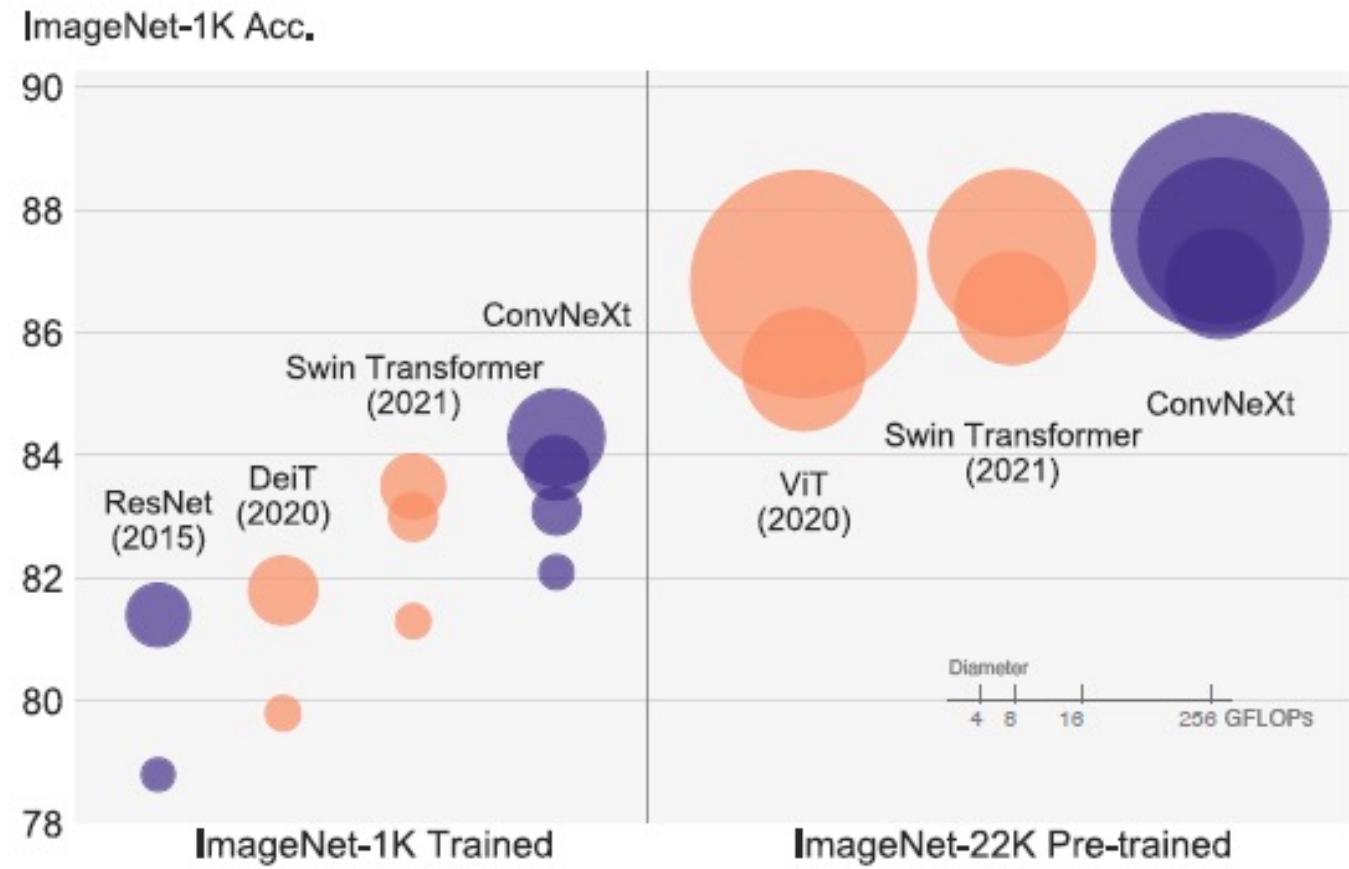
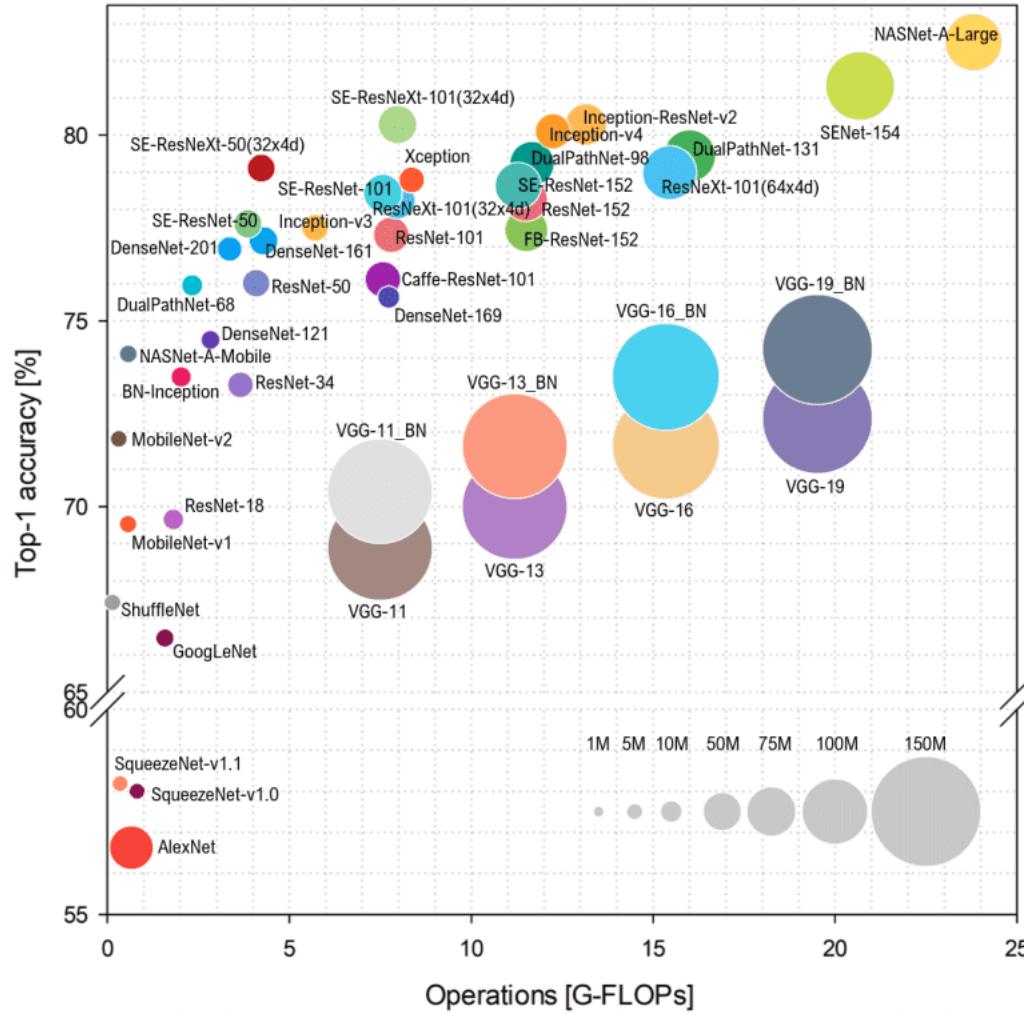
- Attention is all you need, článek z 2017
- Hlavní použití pro Natural Language Processing – NLP
- Podrobný popis architektury viz předmět NI-MVI



Architektury

- AlexNet (2012)
- VGG (2014)
- ResNet (2015)
- ResNeXt (2016)
- SENet (2017)
- EfficientNet (2019)
- **ResNet-RS (2021)**
- **Vision Transformer ViT (2020)**
- **Swin Transformer (2021)**
- **ConvNeXt (2022)**
- Inception (2014)
- MobileNet (2017)
- SE-ResNeXt (2017)
- DenseNet (2017)
- RegNet (2020)
- **EfficientNetV2 (2021)**
- Swin-v2 (2021)
- NFNets (2021)

Architektury

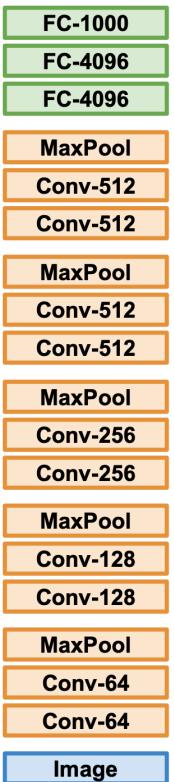


Transfer learning

- Části architektury jež stojí za ziskem atributů (feature extrakce) se přezdívá **backbone**
- Klasifikační části se přezdívá **hlava (head)**
- Praktické použití viz předmět [NI-PDL](#)
- Pro klasifikaci typicky nahradíme **hlavu** za vlastní
- Předtrénovaná **backbone** však může být použita pro úlohy jako
 - Video processing
 - Detekce objektů
 - Sémantická segmentace
 - ...

Transfer learning

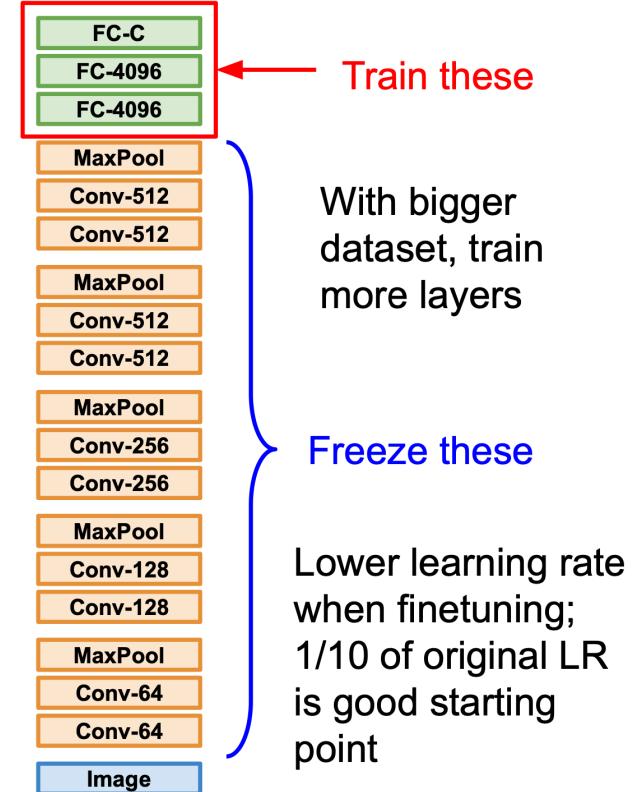
1. Train on Imagenet



2. Small Dataset (C classes)

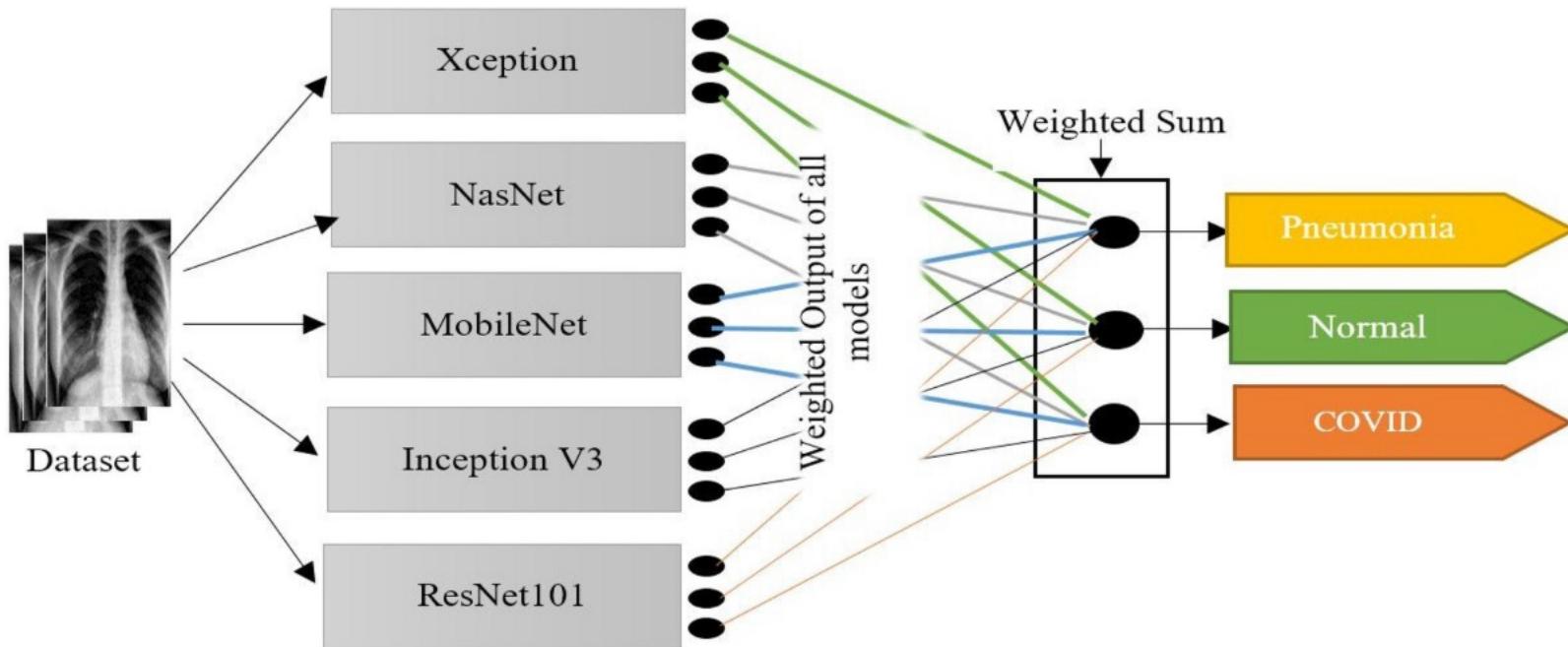


3. Bigger dataset



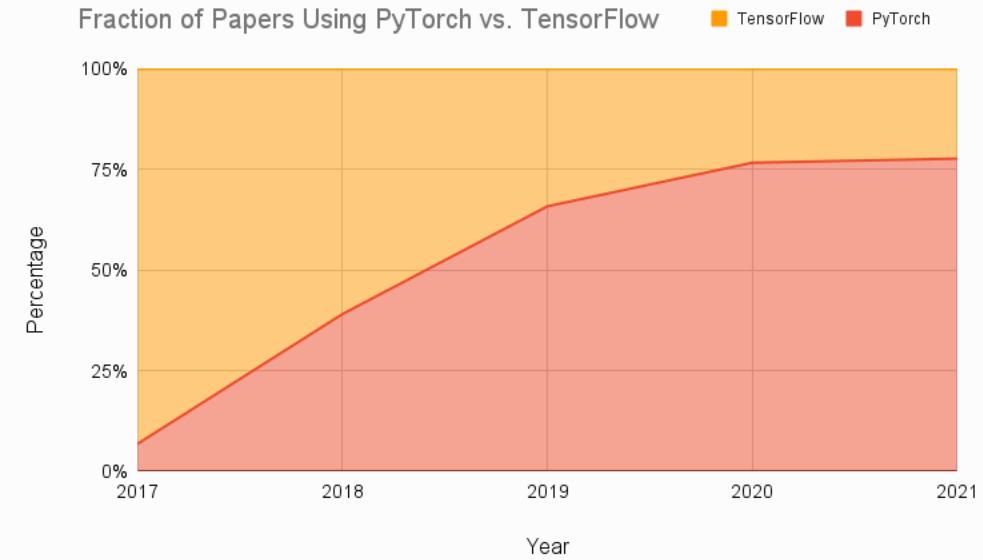
Model ensemble

- Technika spojení výstupu více modelů pro zvýšení robustnosti
- Hodí se kombinovat CNN a Transformer přístupy



Technologie

- Pytorch - Timm
- Tensorflow
- HuggingFace – knihovny a modely



Hugging Face



Problémy neuronových sítí – skryté vzory

- [Video](#) – neviditelný pattern



x
“panda”
57.7% confidence

+ .007 ×



$\text{sign}(\nabla_x J(\theta, x, y))$
“nematode”
8.2% confidence

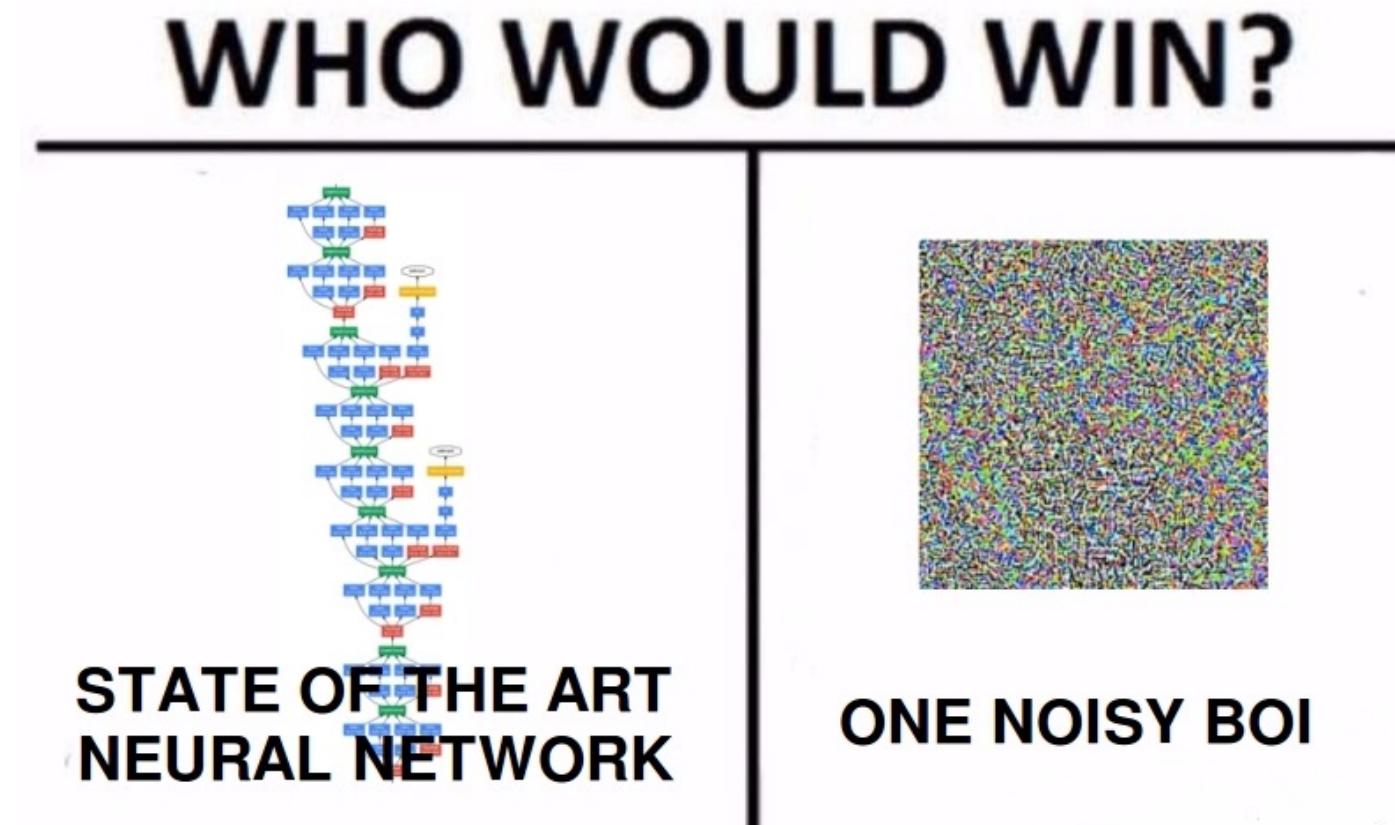
=



$x + \epsilon \text{sign}(\nabla_x J(\theta, x, y))$
“gibbon”
99.3 % confidence

Problémy neuronových sítí – skryté vzory

- [Odkaz na blog post](#)



Problémy neuronových sítí – počítání, lokalizace, rotace



person 0.88

person



reddish orange color 0.78

light brown color 0.78

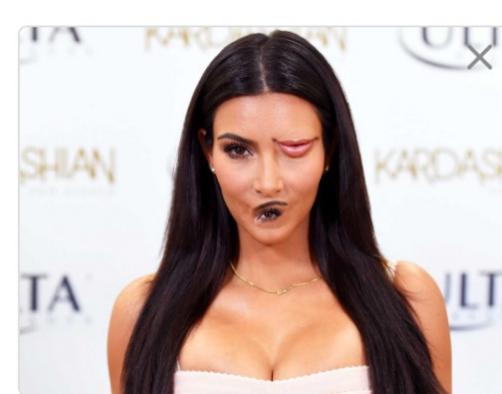
starlet 0.66

entertainer 0.66

female 0.60

woman 0.59

young lady (heroine) 0.59



person 0.90

person



light brown color 0.84

starlet 0.77

entertainer 0.77

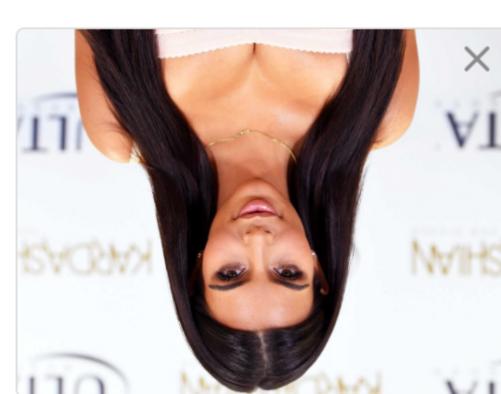
female 0.65

woman 0.64

young lady (heroine) 0.64

reddish orange color 0.64

newsreader 0.50



coal black color 0.79

coal black color



hairpiece (hair) 0.71

dress 0.71

maroon color 0.71

person 0.58

toupee (hairpiece) 0.58

woman 0.56

Earrings 0.55

female 0.50

Datasety

2007

Pascal VOC

2013

ImageNet ILSVRC

2015

MS COCO

- 20 tříd
- 1,464 obrázků pro trénování,
- 1,449 obrázků pro testování a validaci
- Používán jako standard, nyní již pouze k rychlému otestování nového algoritmu

- 1000 tříd
- 1.3M trénovacích obrázků
- Pascal VOC na steroidech

- 80 tříd
- 118K trénovacích obrázků
- Více kategorií v jednotlivých obrázcích. Zaměřený spíše na malé objekty
- Výrazně složitější než ImageNet