

18NES1 Neuronové sítě 1 - Konvoluční neuronové sítě

18NES1 - 19. hodina, LS 2024/25

Zuzana Petříčková

24. dubna 2025

Co jsme probírali minule

Konvoluční neuronové sítě

- Klasická bipyramidální architektura konvoluční neuronové sítě + vizualizace filtrů, map příznaků, citlivostních (Saliency) map
- Učení modelu CNN od nuly na úloze rozpoznávání květin
- Techniky pro zlepšení schopnosti CNN zobecňovat
- Přenesené učení (transfer learning) - úvod

Tento týden

Konvoluční neuronové sítě - dokončení

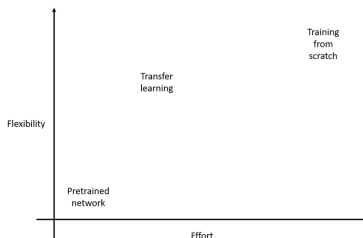
- Přenesené učení (transfer learning), fine-tuning
- Známé architektury konvolučních neuronových sítí
- Aplikace konvolučních neuronových sítí pro řešení různých typů úloh
- Lehký úvod do dalších modelů hlubokého učení
- Ukázky a příklady

Příště

- Návrat k „mělkým“ modelům strojového učení a neuronových sítí: **Učení bez učitele (samoorganizace, unsupervised learning)**

Možnosti vytvoření a učení konvolučních neuronových sítí

- učení od začátku (training from scratch)
- použití již naučeného modelu (pretrained network)
- přenesené učení (transfer learning)
- doučení již naučeného modelu (fine-tuning)



Zdroj :

<https://matlabacademy.mathworks.com/details/deep-learning-onramp/deeplearning>

Konvoluční neuronové sítě a zobecňování

- V porovnání s MLP se CNN učí pomaleji, obzvláště na CPU
- Konvoluční neuronové sítě mívají výraznější problémy s přeučením
- Problémem je, pokud máme k dispozici málo dat (stovky, malé tisíce vzorů)
- Jak zobecňování zlepšit?
 - **standardní regularizace**: early stopping, dropout, normalizace (pro hluboké modely)
 - **data augmentation**: rozšíření dat „za běhu“
 - **transfer learning**

Augmentace dat

- Uměle zvětšuje velikost a rozmanitost trénovací množiny
- Zabraňuje přeučení a posiluje robustnost modelu tím, že ho vystavuje širší škále variací vstupu
- Augmentace probíhá **průběžně**, což šetří paměť a zvyšuje variabilitu
- V Kerasu implementujeme pomocí vrstev jako `RandomFlip`, `RandomRotation`, `RandomZoom` atd.
 - Různé (náhodné) transformace obrazu (rotace, posun, zrcadlení, zešíkmení, měna rozlišení, změna jasu a kontrastu, oříznutí, přidání šumu, blur, kombinace)



Zdroj :

<https://matlabacademy.mathworks.com/details/deep-learning-onramp/deeplearning>

Připomenutí: Ukázka modelu učeného od začátku na malých datech: klasifikace květin

- Zdroj dat: **Oxford 102 Flower Dataset**
paperswithcode.com/dataset/oxford-102-flower
- Obsahuje **8189 obrázků** ve **102 třídách** květin (každá třída cca 40–258 obrázků)
- Velikost datasetu: cca **330 MB** (JPEG)
- Vhodné pro testování trénování modelu CNN od nuly i přeneseného učení

Úlohy:

- 1 klasifikace květin do 3 nejpočetnějších tříd
- 2 klasifikace květin do všech 102 tříd
- 3 na obou datech jsme naučili základní -bipyramidální -CNN model

CNN_from_scratch_flowers_3.ipynb

CNN_from_scratch_flowers_102.ipynb

Připomenutí: Ukázka modelu učeného od začátku na malých datech: klasifikace květin

CNN_from_scratch_flowers_3.ipynb Pozorování

- Oproti MLP se CNN učí poměrně pomalu (je třeba GPU).
- **3 třídy:** Základní model (bez regularizace) se naučil poměrně dobře, i když se mírně přeučil, k dokonalému naučení účinně pomohla kombinace technik: dropout, early stopping a augmentace dat
- **102 tříd:** Model se výrazně přeučil. Regularizační techniky (augmentace dat + dropout + early stopping) výrazně pomohou, ale stále je Accuracy po 50 %.

Další krok

- Co takhle použít již naučený model, popř. přenesené učení (transfer learning)?

Použití předučeného modelu

Co takhle použít již nějaký existující model naučený na velké sadě dat?

Předučené modely:

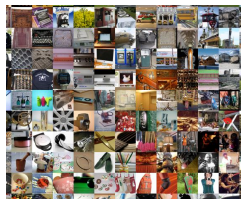
- <https://keras.io/api/applications/>
- oblíbené architektury konvolučních neuronových sítí
 - VGG16
 - MobileNet
 - ResNet
 - ...
- váhy těchto modelů jsou buď náhodně inicializované, nebo jsou sítě předučené, typicky na datové sadě **ImageNet**

pretrained_model.ipynb

- praktická ukázka využití předučeného modelu

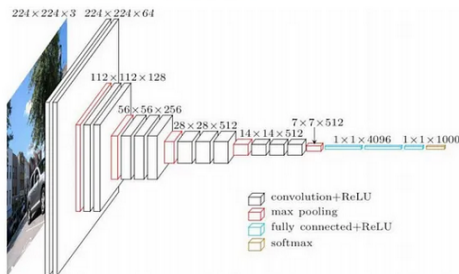
Datová sada ImageNet

- 16 milionů barevných obrázků z 20 tisíc kategorií
- Vznikla v rámci soutěže ImageNet Large-Scale Visual Recognition Challenge (ILSVRC, 2010-2017)
- Tato soutěž odstartovala boom konvolučních neuronových sítí v rozpoznávání obrazu
- ImageNet se stal standardní referenční sadou pro porovnávání modelů (nahradil MNIST)



Příklad předučeného modelu: VGGNet

- Karen Simonyan a Andrew Zisserman, 2014, rodina modelů (např. VGG16, VGG19)
- Klasická bipyramidální architektura, poměrně mělký model (16, resp. 19 vrstev)



Zdroj: <https://medium.com/nerd-for-tech/vgg-16-easiest-explanation-12453b599526>

Použití předučeného modelu

- Obrazová data je třeba převést (přeškálovat /ořezat) na požadované rozměry a typicky do RGB



- Požadovaná velikost obrázků může být u každého modelu jiná (ale zpravidla bývá poměrně malá, kolem 200×200)
- O normalizaci hodnot pixelů se obvykle postará samotný model.

Od předučeného modelu k přenesenému učení (transfer learning)

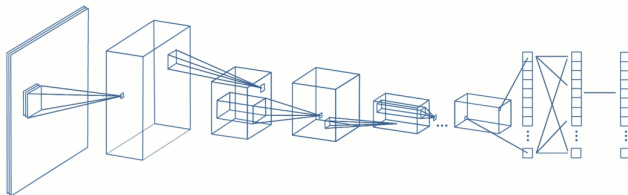
Použití předučeného modelu je skvělý začátek - ale obvykle to takhle jednoduše nepůjde

- ImageNet sice obsahuje 20 000 tříd, ale v případě našich dat i tak klasifikace nemusí být příliš přesná.
- viz. náš příklad: **pretrained_model.ipynb**

Ale jak přesně na to?

Od předučeného modelu k přenesenému učení

- Náš naučený model klasifikuje do 20000 tříd:

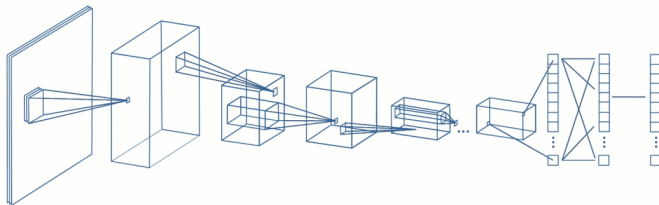


Zdroj :

<https://matlabacademy.mathworks.com/details/deep-learning-onramp/deeplearning>

Od předučného modelu k přenesnému učení

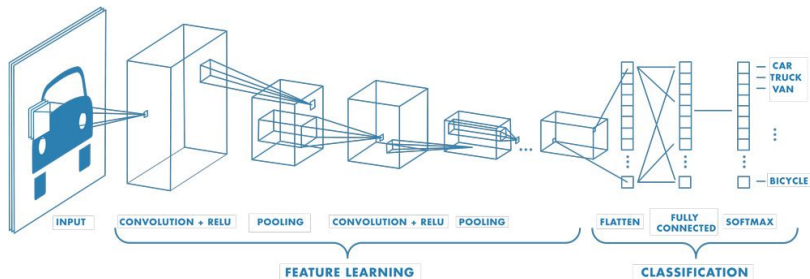
- je třeba klasifikovat do trochu jiných tříd:



Zdroj :

<https://matlabacademy.mathworks.com/details/deep-learning-onramp/deeplearning>

Připomenutí: Architektura konvoluční neuronové sítě



Části konvoluční neuronové sítě

- Konvoluční báze - pro extrakci hierarchických příznaků
- Flattening vrstva - převede data na vektor čísel
- Vrstevnatá neuronová síť (plně propojené vrstvy) pro klasifikaci ... **klasifikační hlava**

Zdroj :

<https://matlabacademy.mathworks.com/details/deep-learning-onramp/deeplearning>

Od předučeného modelu k přenesenému učení (transfer learning)

Jak přesně na to? ... první nápad:

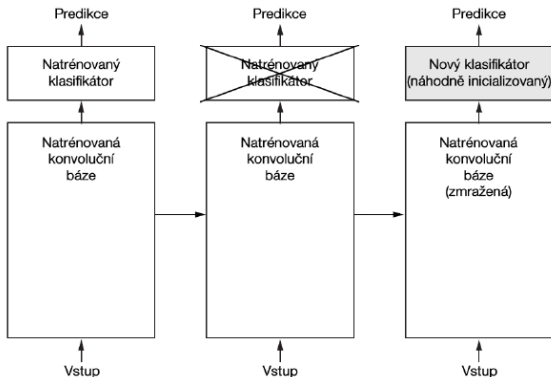
- Vezmeme síť natrénovanou na ImageNetu
- Odstraníme její klasifikační hlavu
- Tu použijeme na extrakci příznaků z dat ... vytvoříme trénovací množinu
- Vytvoříme novou vrstevnatou neuronovou síť a naučíme ji na extrahovaných příznacích

Tento přístup je velmi efektivní, ale často nepraktický

- Statický přístup, nelze jednoduše použít knihovní metody pro augmentaci dat

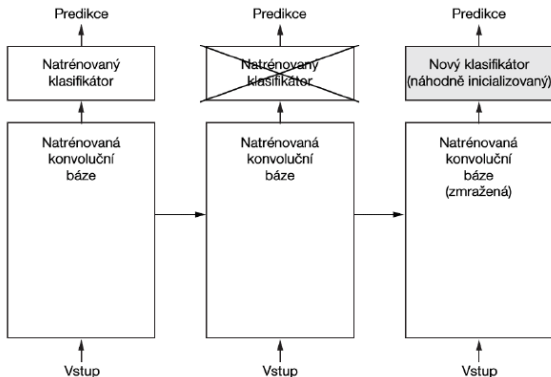
Přenesené učení (transfer learning)

- Vezmeme síť natrénovanou na ImageNetu
- Klasifikační hlavu neuronové sítě (nebo jen její vrchní část) nahradíme novou (náhodně inicializovanou)



Přenesené učení (transfer learning)

- Novou klasifikační hlavu doučíme na nových datech (váhy předchozích vrstev zafixujeme, neboli „zmrazíme“)



Přenesené učení (transfer learning)

Praktické poznámky

- Typicky volíme menší parametr učení než v případě sítě učené od začátku
- Vhodné je použití regularizace (dropout, augmentace dat)

Příklady: Rozpoznávání květin - Pokračování

- `CNN_transfer_learning_flowers_3.ipynb`,
- `CNN_transfer_learning_flowers_102.ipynb`

Tři varianty

- 1 využití předučené konvoluční báze k efektivní extrakci příznaků z dat
- 2 plnohodnotné přenesené učení (spojené s augmentací dat a dropoutem)
- 3 dodatečný fine-tuning

Příklady: Rozpoznávání květin - Pokračování

Pozorování (obecně):

- **Pouze extrakce příznaků:** Výroba trénovací množiny je časově náročné, ale následné učení MLP je velmi rychlé, naučený model má velmi dobrou **Accuracy**.
- **Plnohodnotné přenesené učení** Učení je výrazně pomalejší (opakovaná inference dat skrz konvoluční bázi je časově náročná), ale regularizační techniky (augmentace dat, dropout) vedou ještě k lepší přesnosti

Pozorování (jednotlivé úlohy):

- **3 třídy:** Oba modely se učí rychle v počtu epoch a mají výbornou výslednou přesnost
- **102 tříd:** Výrazné zlepšení oproti učení modelu od začátku, Accuracy nad 80 %.

Přenesené učení a fine-tuning (doladění)

- model naučený s pomocí přeneseného učení můžeme dále vylepšit pomocí fine-tuning-u
 - 1 Nejprve aplikujeme přenesené učení (váhy mimo klasifikační hlavu zafixujeme)
 - 2 Následně „nejhlubší“ část konvoluční báze „rozmrazíme“ a doučíme spolu s novými vrstvami

Praktické poznámky

- Je třeba postupovat opatrně: začneme velmi malým parametrem učení - ideálně začneme s o něco menší hodnotou, na které končil původní model
- Můžeme použít regularizaci

CNN_transfer_learning_flowers_102.ipynb

Přenesené učení - omezení

- přenesené učení se hodí, pokud chce naučit model na podobných datech (obecný model → konkrétnější model)
- nelze použít, pokud by bylo třeba změnit vstupní vrstvy
 - větší počet kanálů (např. rentgenový snímek apod.)
 - rozdílná úroveň abstrakce
 - úplně jiná data
- někdy nezbyvá než učit model od začátku

příklad: fotografie ramene rozšířená o rentgen a MRI



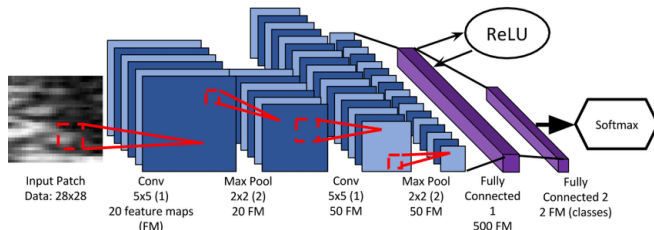
<https://radiopaedia.org/cases/normal-shoulder-mri>

Doplnění: Známé architektury konvolučních neuronových sítí

- nejstarší typy architektur: široké, mělké, často hlubší plně propojená část, odpovídají více základnímu schématu

LeNet 5

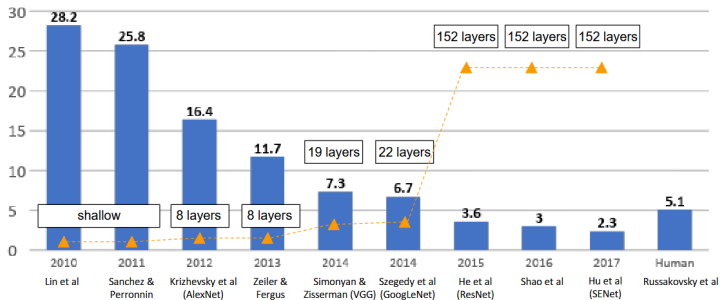
- jedna z původních architektur (Yann LeCun, 1998), poměrně jednoduchá, učená na MNIST



Známé architektury konvolučních neuronových sítí

- ILSVRC soutěž: zásadní soutěž nad datovou sadou ImageNet

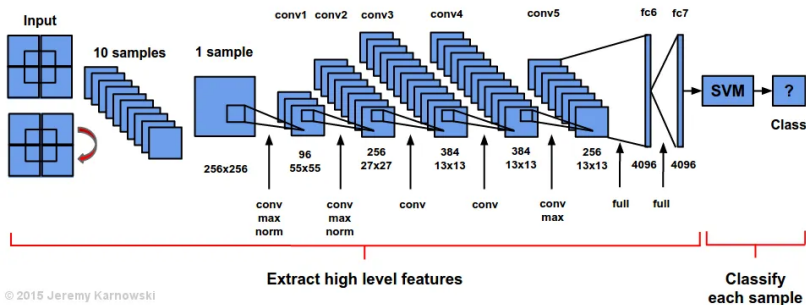
ImageNet Large Scale Visual Recognition Challenge (ILSVRC) winners



Zdroj obrázku: https://cs231n.stanford.edu/slides/2024/lecture_6_part_1.pdf

AlexNet

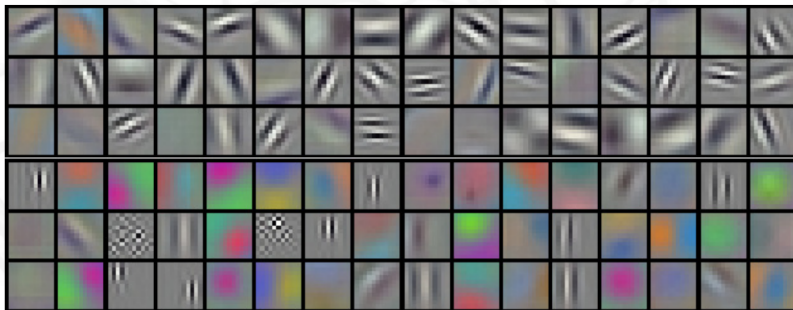
- první CNN-vítěz soutěže ILSVRC (2012) nad ImageNet-úspěšnost 84,7 % v top-5),
- již výrazně složitější (8 vrstev, 61M parametrů, 2 GPU 5-6 dnů)



Zdroj obrázku: <https://medium.com/@jkarnows/alexnet-visualization-35577e5dcd1a>

AlexNet

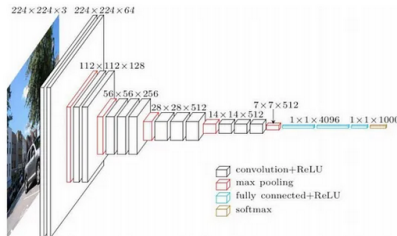
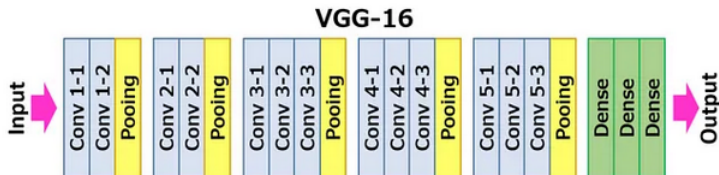
- 96 filtrů $11 \times 11 \times 3$ v první konvoluční vrstvě u AlexNet



Zdroj obrázku: Alex Krizhevsky et al.: "ImageNet Classification with Deep Convolutional Neural Networks", Figure 3

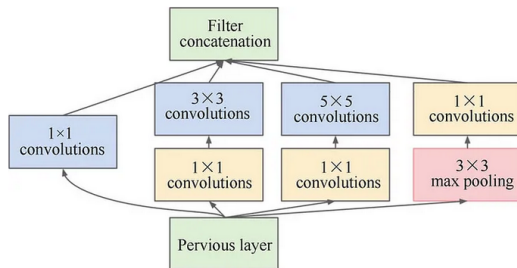
VGGNet

- účastník ILSVCR 2014, rodina modelů (např. VGG16 - 16 vrstev, VGG19 - 19 vrstev) - úspěšnost 92,7 % v top-5
- pyramidová architektura



GoogLeNet (Inception v1), 2014

- vítěz soutěže ILSVRC 2014, 22 vrstev
 - úspěšnost 93,33 % v top-5
- revoluční prvek: Inception bloky -
 - extrakce příznaků v různých škálách,
 - parallelismus, efektivita



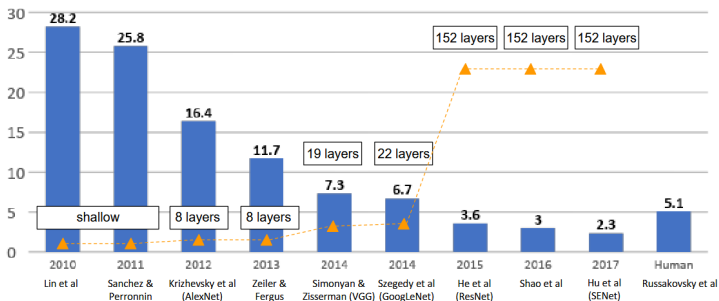
The illustration is from the original paper[1]



Známé architektury konvolučních neuronových sítí

- ILSVRC soutěž po roce 2014: výrazně hlubší a užší architektury (omezení redundance: hloubkově oddělené konvoluce), modularita

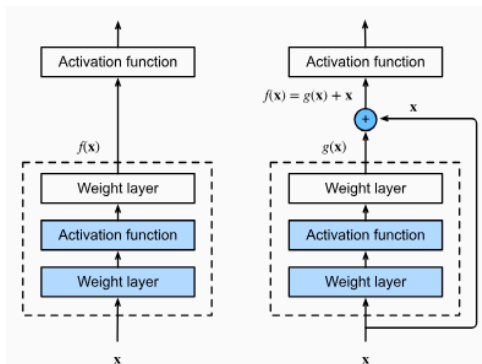
ImageNet Large Scale Visual Recognition Challenge (ILSVRC) winners



Zdroj obrázku: https://cs231n.stanford.edu/slides/2024/lecture_6_part_1.pdf

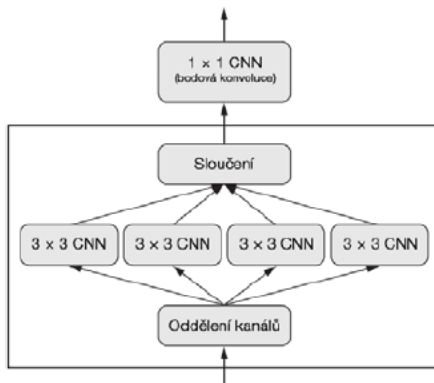
ResNet, 2015

- vítěz soutěže ILSVRC 2015, 152 vrstev - úspěšnost 96,4 % v top-5
- revoluční prvek: skip connections - řeší problém tiché pošty
- residuální bloky:



Hlubkově oddělitelné konvoluce (Xception, 2017)

- další revoluční prvek na cestě ke zvýšení efektivity a zeštíhlení vrstev
- prostorová konvoluce pro každý kanál nezávisle, výstupní kanály se nakonec smíchají pomocí bodové konvoluce



Další populární architektury

- Inception v2, v3 (2015, 2016)
- DenseNet: (Gao Huang, 2016)
- MobileNet (Google, 2017), EfficientNet (2019), SqueezeNet (2016)- další úspora výpočetních prostředků
- NASNet (Neural Architecture Search Network, 2017) - model se naučil vhodnou architekturu sám, genetické algoritmy, reinforcement learning
- ConvNeXt, 2020
- ...

V Kerasu:

- <https://keras.io/api/applications/>

Další populární benchmarkové datové sady

CIFAR-10, CIFAR-100

<https://www.cs.toronto.edu/~kriz/cifar.html>

- CIFAR-10: 60000 barevných obrázků, 10 tříd (auta, psi, lodě,...)
- CIFAR-100: 100 tříd
- malé obrázky, dataset populární pro rychlé testování a porovnání modelů

COCO (Common Objects in Context)

- detekce objektů, segmentace a popis obrázků

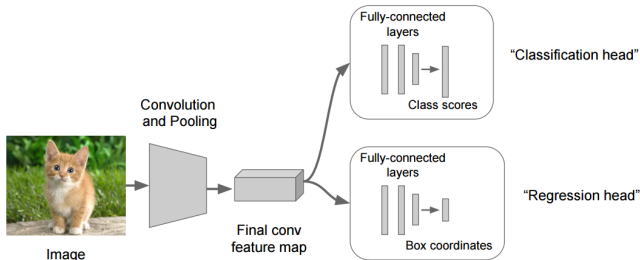
COCO 2020 Panoptic Segmentation Task



<https://cocodataset.org/>

Aplikace konvolučních neuronových sítí

- klasifikační hlavu neuronové sítě můžeme nahradit jinou hlavou a použít naučené příznaky pro řešení jiné úlohy nad stejnými (nebo jinými) daty
 - jiná klasifikační úloha - klasifikační hlava
 - regresní úloha - regresní hlava
 - ...



Aplikace konvolučních neuronových sítí

- klasifikace obrázků - klasifikační hlava
- regresní úloha - regresní hlava

Příklad regrese: predikce sklonu číslic



Zdroj : <https://matlabacademy.mathworks.com>

Aplikace konvolučních neuronových sítí

- segmentace obrazu (sémantická a instanční), detekce objektů:

Other Computer Vision Tasks

Semantic Segmentation



GRASS, CAT,
TREE, SKY

No objects, just pixels

Classification + Localization



CAT

Single Object

Object Detection



DOG, DOG, CAT

Multiple Object

Instance Segmentation



DOG, DOG, CAT

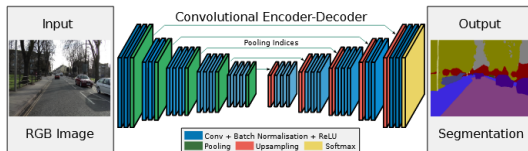
This image is CC0 public domain

Zdroj : https://cs231n.stanford.edu/slides/2017/cs231n_2017_lecture11.pdf

Aplikace konvolučních neuronových sítí

Sémantická segmentace (SegNet, U-net,...)

- enkodér-dekodér architektura
- enkodér: klasická CNN, extrakce příznaků
- latentní prostor: kompaktní reprezentace
- dekodér: obnovuje obrázek do původních rozměrů



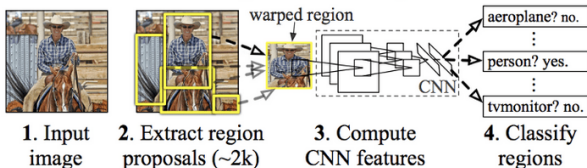
Zdroj : SegNet: A Deep Convolutional Encoder-Decoder Architecture for Image Segmentation <https://arxiv.org/pdf/1511.00561>

Aplikace konvolučních neuronových sítí

Detekce objektů: např. R-CNN (Region-based Convolutional Neural Network)

- 1 obrázek je rozdělen na různé oblasti (ROI)
- 2 každá oblast je zvětšena na stejnou velikost a předložena předučené konvoluční neuronové síti (VGG-16 naučená na ImageNet v původním článku).
- 3 Nakonec jsou použity dvě hlavy: klasifikátor pro klasifikaci každé oblasti (SVM) a regresor pro doladění jeho hranic.

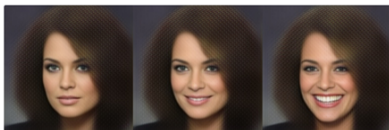
R-CNN: *Regions with CNN features*



Aplikace konvolučních neuronových sítí

Variační autoencodery

- Kódují data jako pravděpodobnostní rozdělení
- Schopné generovat nová data nebo vytvářet interpolace:



F. Chollet: Deep learning v jazyku Python, obr. 5.7

Zdroj : Tom White: Sampling generative networks, <https://arxiv.org/pdf/1609.04468>

Aplikace konvolučních neuronových sítí

Zpracování 2D obrazu

- klasifikace obrázků - klasifikační hlava
- regresní úloha - regresní hlava
- detekce objektů - detekční hlava
- segmentace obrazu - segmentační hlava - architektura encoder-decoder
- restaurování obrazu, stylizace obrazu, generování obrazu - architektura autoencoder
- generování textových popisků k obrázkům, určení póz, rysů obličejů, hodnocení podobnosti obrazů,...

Jiný typ dat

- analýza videa (3D konvoluce) - akční detekce (např. sport. záznam)
- sekvenční data (1D konvoluce) - časové řady, audio data, omezeně i přirozený jazyk)

Výhody a nevýhody konvolučních neuronových sítí

- Ušité na míru datům, která jsou uspořádaná do mřížky
- Invariance vůči posunutí, velikosti, odstínu,...
- Robustní k šumu v datech
- Výpočetně náročné učení, vyžaduje velké množství dat, GPU
- Riziko přeučení, zejména u menších datasetů
- Adversariální příklady – stále nevyřešený problém
(*drobná neviditelná změna obrázku, která způsobí nesprávnou klasifikaci*)
https://www.tensorflow.org/tutorials/generative/adversarial_fgsm

Další modely hlubokého učení

Rekurentní neuronové sítě (RNNs)

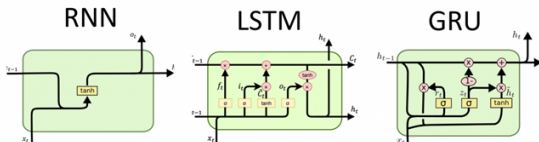
- Analýza a modelování sekvenčních dat (časové řady, jednorozměrné signály, řeč, text, písmo)
- Rozpoznání řeči a písma, strojový překlad

Sítě s dlouhou-krátkodobou pamětí (Long Short-Term Memory Networks, LSTMs)

- Model schopný zachytit i dlouhodobé závislosti v datech

Gated Recurrent Unit Networks (GRU)

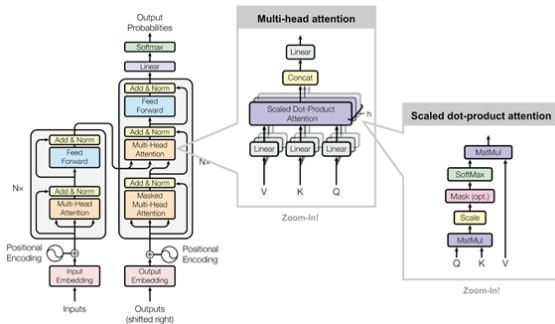
- Zjednodušená verze LSTM



Další modely hlubokého učení

Transformer Networks (Gemini, GPT,...)

- Pro úlohy zpracování přirozeného jazyka (klasifikace textů, překlady, generování textu apod.).
- udržují kontext díky mechanismu self-attention



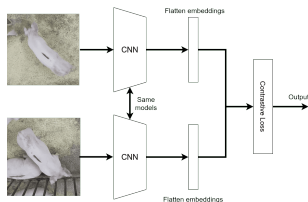
Attention is all you need, A. Vaswani, N. Shazeer, N. Parmar, J. Uszkoreit, L. Jones, A. Gomez, et.al.. Advances in Neural Information Processing Systems , page 5998–6008. (2017)

Další modely hlubokého učení

- často modulární architektura

Siamese Networks

- Pro úlohy rozpoznání obrazu, object-tracking.
- Počítají podobnost mezi dvěma různými vstup
- **Ukázka:** *Animal recognition with Siamese Networks and Mean Embeddings*

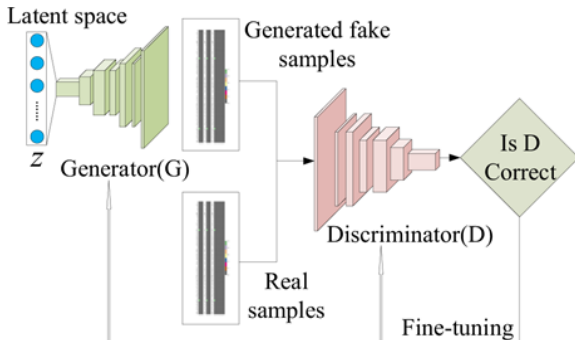


<https://erdem.pl/2021/02/animal-recognition-with-siamese-networks-and-mean-embeddings>

Další modely hlubokého učení

Generative Adversarial Networks (GANs)

- Generování nových dat na základě naučených vzorů.



Dan, Y., Zhao, Y., Li, X. et al. Generative adversarial networks (GAN) based efficient sampling of chemical composition space for inverse design of inorganic materials. npj Comput Mater 6, 84 (2020). <https://doi.org/10.1038/s41524-020-00352-0>

Další modely hlubokého učení

- **Deep Belief Networks (DBNs)** - Generativní model učený bez učitele (unsupervised).
- **Deep Q-Networks (DQNs)** - Pro úlohy zpětnovazebného učení.
- **Capsule Networks** - Pro úlohy rozpoznání obrazu. Modelují hierarchické vztahy mezi částmi objektů



<https://ilkereliozkan.com.tr/publication/sengul-new-2022/sengul-new-2022.pdf>