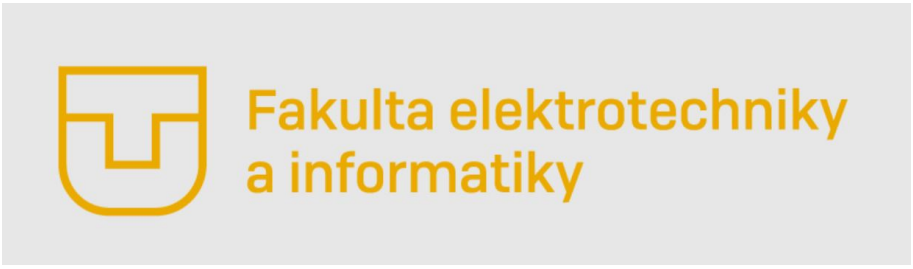


Fakulta Elektrotechniky a Informatiky

Katedra kybernetiky a umelej inteligencie

Predmet : **Základy hlbokého učenia**
kurz 2018 / 2019

Zadanie číslo # 1 (esej) :

Rozpoznávanie objektov

kľúčové slova : rozpoznávanie objektov, R-CNN, YOLO

Spracoval :
Dávid Gajdoš

Obsah

1 Úvod **2**

2 Prehľad state-of-the-art výskumu **3**
2.1 R-CNN . . . 3
2.2 Fast R-CNN . . . 3
2.3 Faster R-CNN . . . 4
2.4 YOLO . . . 5
2.4.1 Algoritmus . . . 5
2.4.2 Topológia . . . 5
2.4.3 Loss funkcia . . . 6

3 Porovnanie jednotlivých metód **6**

4 Záver **8**

1 Úvod

Človek pri pohľade na akýkoľvek obrázok okamžite vie určiť aké objekty sa na ňom nachádzajú a takisto vie určiť ich polohu na obrázku. Ľudský vizuálny systém je veľmi rýchly a presný, čo nám umožňuje vykonávať komplexné úlohy akými je napríklad vedenie vozidla. Rýchle, presné a spoľahlivé algoritmy by umožnili počítačom plniť podobne náročné úlohy bez špecializovaných senzorov, asistovať ľuďom pri plnení úloh a takisto vytvorenie rezponzívnych robotických systémov.
Rozpoznávanie objektov a ich lokalizcia je veľmi dôležitou úlohou pre detailné porozumenie obrázkov. Pri klasifikácii obrázkov sa snažíme čo najpresnejšie zatriediť celý obrázok do určitej triedy. Pri detekcii je úlohou určiť polohu objektov na danej snímke. Obr. 1 znázorňuje detekciu a klasifikáciu jednotlivých objektov.

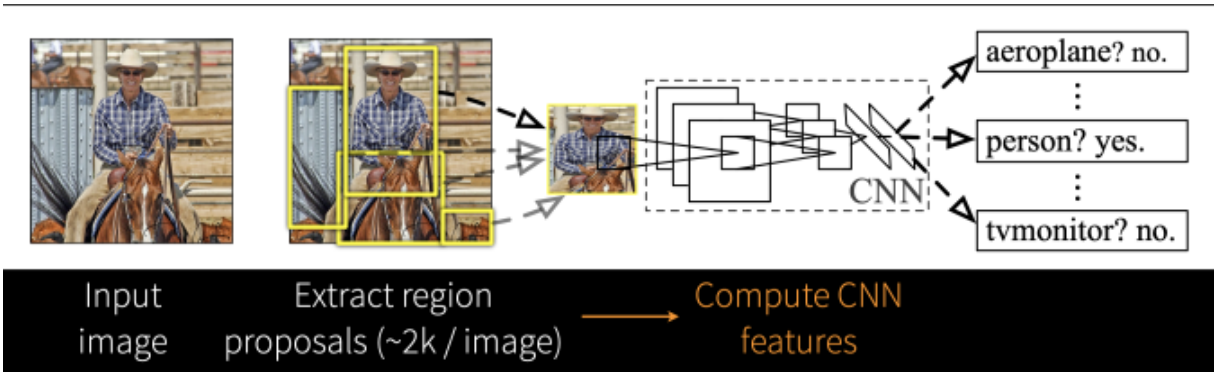
Obr. 1: Detekcia a klasifikácia objektov na obrázku [1]

Tradičné prístupy na detekciu objektov využívajú rôzne prístupy. Jedným z najpoužívanejších je prístup tzv. "posuvného okna". Ten dovoľuje aj objekty s rôznymi rozmermi resp. s rôznymi pomermi strán. Jeho nevýhodou však je to, že je pre neho náročné klasifikovať objekty do viacerých tried. [2, 3]
Vďaka pokrokom v oblasti hlbokého učenia (DL) a tzv. konvolučných neurónových sietí (CNN) však vznikli detektory objektov viacerých tried s akceptovateľnou presnosťou vhodné pre komerčné využitie [4, 5].
Na rozdiel od tradičných prístupov ide o hlboké architektúry, ktoré umožňujú učenie komplexnejších modelov. Sú však náročné na výpočtový výkon.

2 Prehľad state-of-the-art výskumu

2.1 R-CNN

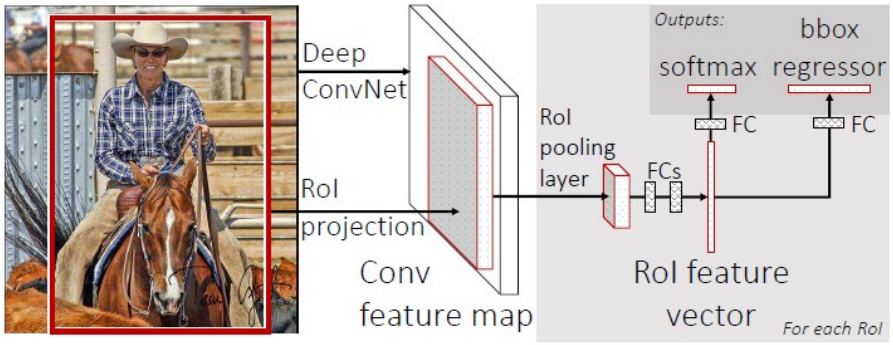
Region-based Convolutional Network method (R-CNN) využíva na detekciu objektov hlbokú konvolučnú neurónovú sieť. Jej výhodou je vysoká presnosť pri detekcii, avšak ide o pomalú a výpočtovo náročnú metódu.
Metóda najprv generuje návrhy regiónov (ROI), ktorých je okolo 2000 pre každý obrázok. Využíva algoritmus selektívneho vyhľadávania. Pre každý návrh sa vytvorí vektor príznakov, ktoré sa následne pomocou konvolučnej siete klasifikujú a pomocou regresie určia ich polohy na obrázku. Architektúru metódy znázorňuje Obr. 2.



Obr. 2: R-CNN flowchart [6]

2.2 Fast R-CNN

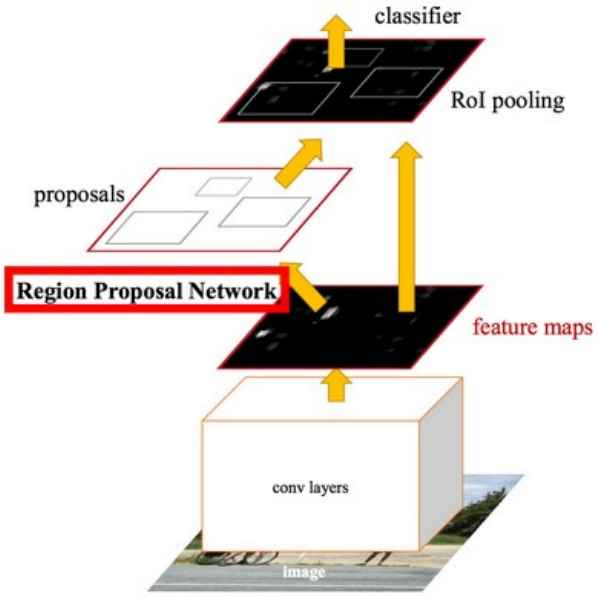
Vstupom do Fast R-CNN je celý obrázok, z ktorého je pomocou konvolučnej siete vytvorený príznakový priestor. Následne sa vyhľadávajú regióny záujmu (ROI), z ktorých sa vytvoria príznakové vektory pevnej veľkosti. Tie sa následne pomocou plne-prepojených vrstiev pomocou softmax klasifikátorov klasifikujú a pomocou regriesie sa určia 4 súradnice pre okraje objektu. Architektúru metódy Fast R-CNN znázorňuje Obr. 3. [7]
Hlavný rozdiel od R-CNN je to, že R-CNN vytvára návrhy oblastí na úrovni pixelov, kým Fast R-CNN vytvára príznakovu mapu jednotlivých oblastí.



Obr. 3: Fast R-CNN architektúra [7]

2.3 Faster R-CNN

Metóda Faster R-CNN nevuyžíva špeciálnu metódu na návrh oblastí záújmu na rozdiel od predošlých metód. Využíva sieť návrhu oblastí, ktorej vstupom je príznakový priestor a výstupom návrhy oblastí záujmu. Tie sú následne vyhodnocované podobne ako v metóde Fast R-CNN a určené ich súradnice. Architektúru metódy popisuje Obr. 4.



Obr. 4: Faster R-CNN architektúra [8]

2.4 YOLO

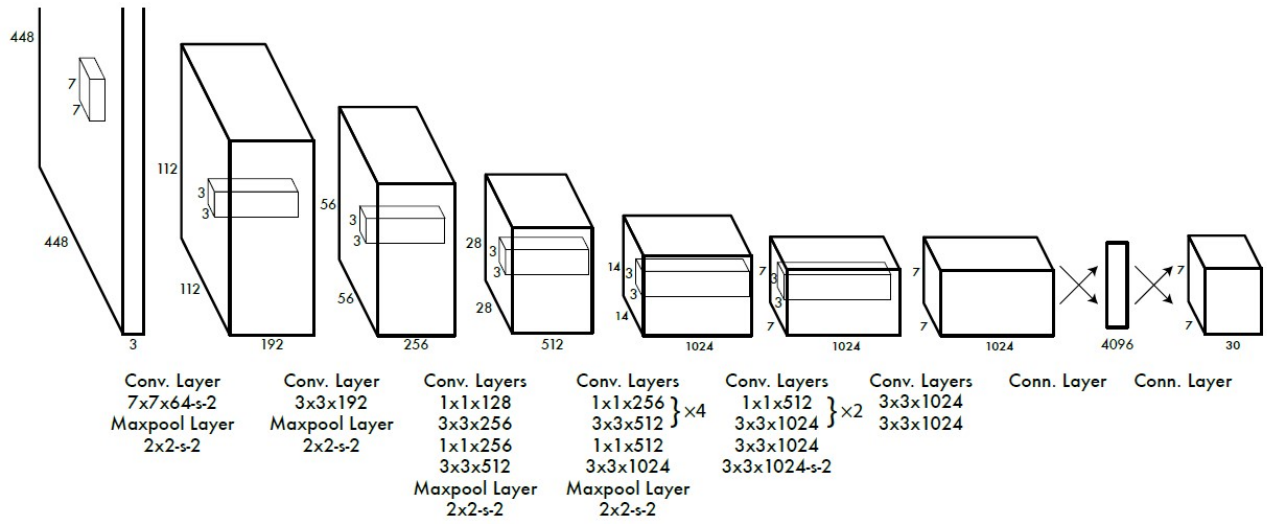
Metóda YOLO (You Only Look Once) využíva jednu konvolučnú sieť súčasne na vyhľadávanie ohraničení objektov a ich klasifikáciu. Na rozdiel od predošlých metód vníma obrázok ako celok a je schopná porozumieť vzájomným kontextom medzi oblasťami. Ide o veľmi rýchlu metódu oproti predošlým, ktorú je možné jednoducho prispôsobiť na nové domény a podáva spoľahlivejšie výsledky na neočakávané vstupy ako predošlé metódy. Jej presnosť je však nižšia hlavne pri malých objektoch. [1]

2.4.1 Algoritmus

Každý obrázok je rozdelený do S x S mriežky. Ak sa stred objektu nachádza v danej bunke mriežky, daná bunka zodpovedá za detekciu objektu. V každej bunke sa môže nachádzať rôzny počet objektov.
Oblasti záujmu majú 5 parametrov: x, y, w, h, a confidence level. Parametre (x,y) určujú súradnice stredov, parametre (w,h) určujú šírku a výšku oblasti a confidence level určí úroveň spoľahlivosti klasifikácie do určitej triedy. [1]

2.4.2 Topológia

Sieť má 24 konvolučných vrstiev, za ktorými sa nachádzajú 2 plne prepojené vrstvy. Alternujúce 1x1 konvolučné vrstvy slúžia na zmenšenie príznakového priestoru. Výstupom je 7x7x30 uzlov predikcie. Topológiu siete popisuje Obr. 5.



Obr. 5: YOLO - topológia siete [1]

2.4.3 Loss funkcia

$\lambda_{coord} \sum_{i=0}^{S^2} \sum_{j=0}^{B} \mathbb{1}obj[(x_i - \hat{x}_i)^2 + (y_i - \hat{y}_i)^2]$
$+ \lambda_{coord} \sum_{i=0}^{S^2} \sum_{j=0}^{B} \mathbb{1}j^{obj}[(\sqrt{w_i} - \sqrt{\hat{w}_i})^2 + (\sqrt{h_i} - \sqrt{\hat{h}_i})^2]$
$+ \sum_{i=0}^{S^2} \sum_{j=0}^{B} \mathbb{1}ij^{obj}(C_i - \hat{C}_i)^2 + \lambda_{noobj} \sum_{i=0}^{S^2} \sum_{j=0}^{B} \mathbb{1}_{ij}^{noobj}(C_i - \hat{C}_i)^2$

x_i, y_i predstavuje polohu stredu oblasti,
w_i, h_i predstavujú šírku a výšku oblasti,
C_i predstavuje interval spoľahlivosti, že daná oblasť obsahuje objekt,
$p_i(c)$ predstavuje klasifkačnú stratu,
$\mathbb{1}ij^{obj}$ sa rovná 1, ak daná oblasť obsahuje objekt,
$\mathbb{1}ij^{noobj}$ sa rovná 1, ak daná oblasť neobsahuje objekt,
$\mathbb{1}^{obj}$ sa rovná 1, ak bola pedikovaná príslušna trieda.

3 Porovnanie jednotlivých metód

Priemerná presnosť a presnosť pre jednotlivé triedy je zobrazená v tabuľke Obr. 7. Stĺpec mAP predstavuje priemernú presnosť jednotlivých metód a v ostatných stĺpcoch sú presnosti klasifikácie pre jednotlivé triedy. Metódy Fast R-CNN a Faster R-CNN dosahujú vysokú presnosť klasifikácie. [9]
Rýchlosť klasifikácie popisuje tabuľka Obr. 6. Stĺpec FPS predstavuje počet spracovaných snímkov za sekundu. Metóda YOLO a jej modifikácia Fast YOLO predstavujú veľmi rýchle metódy, ktoré dosahujú dostatočnú presnosť klasifikácie.[9]
Presnosť a rýchlosť jednotlivých metód bola vyhodnocovaná v rámci Pascal VOC Challenge, kde sa nachádza 20 tried. Dataset obsahuje 11 530 obrázkov s popisom, v ktorých sa nachádza 27 450 regiónov obsahujúcich objekty. Dataset pre rok 2007 obsahuje 9 963 obrázkov a 24 640 regiónov obsahujúcich objekty. Výsledky klasifikácie jednotlivých metód sa porovnávajú s testovacou vzorkou s rozložením trénovacej a testovacej množiny 1:1. [9]

Real-Time Detectors	mAP	FPS
Fast YOLO	52.7	**155**
YOLO	**63.4**	45
Less Than Real-Time		
Fastest DPM	30.4	15
R-CNN Minus R	53.5	6
Fast R-CNN	70.0	0.5
Faster R-CNN VGG-16	73.2	7
Faster R-CNN ZF	62.1	18
YOLO VGG-16	66.4	21

Obr. 6: Porovnanie rýchlosti jednotlivých metód detekcie objektov na PASCAL VOC 2007 Leaderboard. [1]

VOC 2012 test	mAP	aero	bike	bird	boat	bottle	bus	car	cat	chair	cow	table	dog	horse	mbike	person	plant	sheep	sofa	train	tv
MR_CNN_MORE_DATA [11]	**73.9**	**85.5**	**82.9**	**76.6**	**57.8**	**62.7**	**79.4**	77.2	86.6	**55.0**	**79.1**	**62.2**	87.0	**83.4**	**84.7**	78.9	45.3	73.4	65.8	80.3	74.0
HyperNet_VGG	71.4	84.2	78.5	73.6	55.6	53.7	78.7	**79.8**	87.7	49.6	74.9	52.1	86.0	81.7	83.3	**81.8**	**48.6**	**73.5**	59.4	79.9	65.7
HyperNet_SP	71.3	84.1	78.3	73.3	55.5	53.6	78.6	79.6	87.5	49.5	74.9	52.1	85.6	81.6	83.2	81.6	48.4	73.2	59.3	79.7	65.6
Fast R-CNN + YOLO	70.7	83.4	78.5	73.5	55.8	43.4	79.1	73.1	**89.4**	49.4	75.5	57.0	**87.5**	80.9	81.0	74.7	41.8	71.5	68.5	**82.1**	67.2
MR_CNN_S_CNN [11]	70.7	85.0	79.6	71.5	55.3	57.7	76.0	73.9	84.6	50.5	74.3	61.7	85.5	79.9	81.7	76.4	41.0	69.0	61.2	77.7	72.1
Faster R-CNN [27]	70.4	84.9	79.8	74.3	53.9	49.8	77.5	75.9	88.5	45.6	77.1	55.3	86.9	81.7	80.9	79.6	40.1	72.6	60.9	81.2	61.5
DEEP_ENS_COCO	70.1	84.0	79.4	71.6	51.9	51.1	74.1	72.1	88.6	48.3	73.4	57.8	86.1	80.0	80.7	70.4	46.6	69.6	**68.8**	75.9	71.4
NoC [28]	68.8	82.8	79.0	71.6	52.3	53.7	74.1	69.0	84.9	46.9	74.3	53.1	85.0	81.3	79.5	72.2	38.9	72.4	59.5	76.7	68.1
Fast R-CNN [14]	68.4	82.3	78.4	70.8	52.3	38.7	77.8	71.6	89.3	44.2	73.0	55.0	**87.5**	80.5	80.8	72.0	35.1	68.3	65.7	80.4	64.2
UMICH_FGS_STRUCT	66.4	82.9	76.1	64.1	44.6	49.4	70.3	71.2	84.6	42.7	68.6	55.8	82.7	77.1	79.9	68.7	41.4	69.0	60.0	72.0	66.2
NUS_NIN_C2000 [7]	63.8	80.2	73.8	61.9	43.7	43.0	70.3	67.6	80.7	41.9	69.7	51.7	78.2	75.2	76.9	65.1	38.6	68.3	58.0	68.7	63.3
BabyLearning [7]	63.2	78.0	74.2	61.3	45.7	42.7	68.2	66.8	80.2	40.6	70.0	49.8	79.0	74.5	77.9	64.0	35.3	67.9	55.7	68.7	62.6
NUS_NIN	62.4	77.9	73.1	62.6	39.5	43.3	69.1	66.4	78.9	39.1	68.1	50.0	77.2	71.3	76.1	64.7	38.4	66.9	56.2	66.9	62.7
R-CNN VGG BB [13]	62.4	79.6	72.7	61.9	41.2	41.9	65.9	66.4	84.6	38.5	67.2	46.7	82.0	74.8	76.0	65.2	35.6	65.4	54.2	67.4	60.3
R-CNN VGG [13]	59.2	76.8	70.9	56.6	37.5	36.9	62.9	63.6	81.1	35.7	64.3	43.9	80.4	71.6	74.0	60.0	30.8	63.4	52.0	63.5	58.7
YOLO	57.9	77.0	67.2	57.7	38.3	22.7	68.3	55.9	81.4	36.2	60.8	48.5	77.2	72.3	71.3	63.5	28.9	52.2	54.8	73.9	50.8
Feature Edit [32]	56.3	74.6	69.1	54.4	39.1	33.1	65.2	62.7	69.7	30.8	56.0	44.6	70.0	64.4	71.1	60.2	33.3	61.3	46.4	61.7	57.8
R-CNN BB [13]	53.3	71.8	65.8	52.0	34.1	32.6	59.6	60.0	69.8	27.6	52.0	41.7	69.6	61.3	68.3	57.8	29.6	57.8	40.9	59.3	54.1
SDS [16]	50.7	69.7	58.4	48.5	28.3	28.8	61.3	57.5	70.8	24.1	50.7	35.9	64.9	59.1	65.8	57.1	26.0	58.8	38.6	58.9	50.7
R-CNN [13]	49.6	68.1	63.8	46.1	29.4	27.9	56.6	57.0	65.9	26.5	48.7	39.5	66.2	57.3	65.4	53.2	26.2	54.5	38.1	50.6	51.6

Obr. 7: Porovnanie presnosti jednotlivých metód detekcie objektov na PASCAL VOC 2012 Leaderboard. [1]

4 Záver

Cieľom tejto práce bolo predstaviť rôzne metódy detekcie objetov, popísať princíp ich fungovania a porovnať presnosť a rýchlosť jednotlivých metód.
Metódy R-CNN, Fast R-CNN a Faster R-CNN sú veľmi presné, avšak možnosť ich využitia v real-time aplikáciách je obmedzená z dôvodu nízkej rýchlosti a výpočtovej náročnosti.

Literatúra

[1] Joseph Redmon, Santosh Divvala, Ross Girshick, Ali Farhadi *You Only Look Once: Unified, Real-Time Object Detection*, IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2016

[2] Viola, P.A., Jones, M.J. *Robust real-time face detection*, Int. J. Comput. Vis. 57(2), 137–154 (2004)

[3] Dollár, P., Appel, R., Belongie, S.J., Perona, P. *Fast feature pyramids for object detection*, IEEE Trans. Pattern Anal. Mach. Intell. 36(8), 1532–1545 (2014)

[4] Alex Krizhevsky, Ilya Sutskever, Geoff Hinton *Imagenet classification with deep convolutional neural networks*, Advances in Neural Information Processing Systems 25, 2012.

[5] Geoffrey E Hinton and Ruslan R Salakhutdinov *Reducing the dimensionality of data with neural networks*, Science, 313(5786):504–507, 2006.

[6] Ross Girshick Jeff Donahue Trevor Darrell Jitendra Malik *Rich feature hierarchies for accurate object detection and semantic segmentation*, IEEE Conference on Computer Vision and Pattern, 2014

[7] Ross Girshick *Fast R-CNN*, IEEE International Conference on Computer Vision (ICCV), 2015

[8] Sik-Ho Tsang, *Review: Faster R-CNN (Object Detection)*, TowardsDataScience, 2018 Dostupné online na [https://towardsdatascience.com/review-faster-r-cnn-object-detection-f5685cb30202]

[9] Everingham, M. and Van Gool, L. and Williams, C. K. I. and Winn, J. and Zisserman, A., *The PASCAL Visual Object Classes Challenge 2012 (VOC2012) Results*, Dostupné online na [http://www.pascal-network.org/challenges/VOC/voc2012/workshop/index.html"]