

**Projekt do predmetu MRBT**

**Detekcia áut v obraze pomocou Neurónových sietí**

**Vedúci práce:** Ligocki Adam, Ing.

**Vypracovali:** Hadzima Jaroslav     (173648)   
                Sabó Marek     (171152)   
                Kratochvíla Lukáš     (171242)

**Odovzdané:** 4.5.2018

**Zadanie :**

1. Preskúmajte súčasný state-of-art v oblasti detekcie vozidiel pomocou neuronových sietí.
2. Oživte najmenej dva open source projekty a demonštrujte ich presnosť na vlastnom datasete.
3. Spravte benchmark v ktorom porovnáte výstupný výkon a presnosť jednotlivých sietí.

**Detekcia objektov**

V dnešnej dobe stále narastá počet odlišných oblasti, v ktorých môžeme využiť systémy na detekciu/rozpoznávanie objektov. V oblastiach automatizácie, kde stroj prichádza do kontaktu s človekom sme však konzervatívnejší. Dbáme na bezpečnosť a tým spomaľujeme nasadenie v bežných aplikáciách. Na druhej strane to núti ľudí stále viac a viac inovovať a ošetrovať používané systémy, prípadne vymýšľať nové tak, aby splňovali na ne kladené kritéria. Typickým príkladom je nasadenie systému rozpoznávania vozidiel a chodcov pre aplikácia autonómnych vozidiel, kde môžu výrazne znížiť počet nehôd a smrteľných úrazov.

V minulosti sa na aplikácie zautomatizovania dopravy používalo veľké množstvo iných senzorov (lasery GPS, radar, inteligentné mapy). Tie však samé o sebe mali svoje osobité obmedzenia. Metódy počítačového videnia majú potenciál nahradiť tieto senzory a zároveň zvýšiť bezpečnosť cestujúcich. Metódy detekcia vozidiel sa však stretávajú so svojimi vlastnými ***problémami***:

* *Rozmery objektu* - So snímaním kamerou strácame informáciu o hĺbke. Vzniká projekčná chyba, kde objekty vzdialené majú na snímke opticky menšiu veľkosť a objekty blízko kamery zase veľkú. Na dvoch snímkach teda môže byť ten istý predmet rôzne veľký.
* *Lokalizácia v obraze* – Detekované objekty budeme mať len veľmi vzácne vycentrované na stred obrazu. Preto metódy detekcie vozidiel zvádzajú na použitie kĺzavého okna po celom obraze a porovnávania s modelom vozidla. Zvyšuje to však časovú náročnosť funkcie. Ako si kážeme neskôr v texte, existujú aj inteligentnejšie riešenia.
* *Veľké množstvo typov objektov a ich variácia vo vzhľade* – Vozidiel máme veľké množstvo najrôznejších značiek. Môže sa jednať napríklad o osobné autá, nákladné autá, kabriolety alebo pracovné stroje typu báger. Rovnako môžu byť rôznej farby. Tiež treba myslieť na to, že vozidlo môže byť snímané pod rôznym uhlom.
* *Modelovanie* – Vzniká otázka ako zároveň spojiť tieto problémy do jedného modelu. Obrázok môže byť rôzne veľký, umiestnený kdekoľvek v obrázku, rôzne otočený a s rôznou farbou.

**História vývoja metód používaných pre detekciu vozidiel v obraze**

***Bežný prístup:***

V priebehu rokov sa vyskytlo veľké množstvo metód pre detekciu áut v obraze. Tieto metódy však nevykazovali dostatočné realtime vlastnosti. Porovnávanie s databázou a kĺzavé okná sú príliš výpočetne náročné. Jeden z dvoch metód, ktoré sú stále používané a ktoré zvolili odlišný prístup sú napríklad:

**Viola-Jones framework (2001)** [10.18]

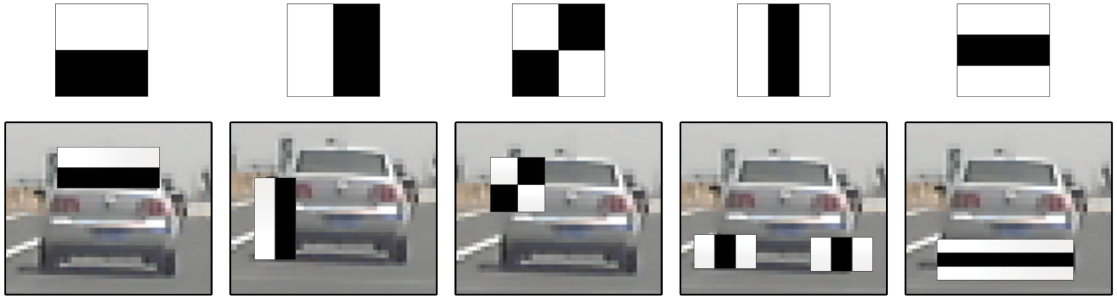
Publikované v článku ‘Robust Real-time Object Detection’  Paulom Michaelom Jonesom. Pracuje na princípe generovania stoviek až tisícok jednoduchých binárnych Háárových klasifikátorov. Využívalo sa to pre detekciu tváre pomocou web-kamery. Samotný princíp je jednoduchý a relatívne rýchly, zvláda aj realtime processing. Klasifikátory využívajú na posudzovanie multiscale kĺzavé okno v kaskáde. Táto metóda bola rýchlo implementovaná do *OpenCV****.*** Veľkou nevýhodou bol fakt, že tvár dokázali rozpoznať iba z rovnakého uhlu, v akom boli naučené. Ak bola tvár naklonená alebo pootočené, algoritmus zlyhával. Pre autá bol problém rovnaký.



Obr.1) Ukážka Háárových obrazcov, použitých na detekciu tváre [10.18]

*Základný princíp pri tom pozostával z nasledujúcich krokov:*

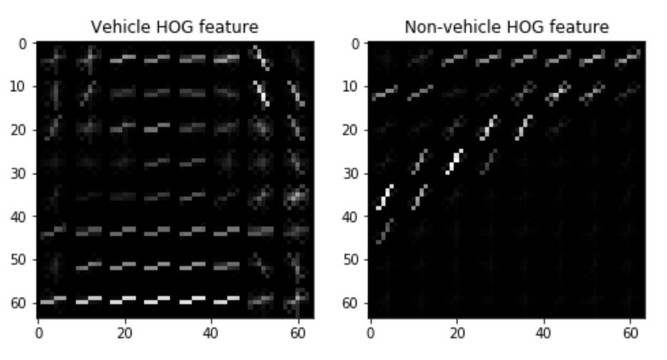
1. Zostrojenie datasetu s tisíckami obrázov tváre pod rovnakým uhlom a natočením.
2. Natvrdo zakódovať pomocou Háárových vlastností vlastnosti tváre. Napríklad, ako je v obrázku vidieť, oči sú tmavšie ako miesto medzi nimi odpovedajúce nosu alebo napríklad tmavá oblasť očí a svetlá obasť pod očami.
3. Natrénujeme SVM klasifikátor na daný dataset.
4. Využijeme natrénovaný klasifikátor na nové obrázky z web-kamery.



Obr.2) Háárove klasifikátory pre detekciu áut [10.19]

**HOG (Histogram of oriented gradients**) **(2005)** [10.17]

Využívajú tzv. a SVM(Support Vector Machine) na klasifikáciu. Rovnako ako *Viola-Jones framework* využíva multiscale kĺzavé okno. Dosahuje väčšiu presnosť ako Viola-Jones, ale je výrazne pomalší. Pôvodne vyvinutý na detekciu chodcov. Vlastnosti sú však rovnako zakódované ‘natvrdo‘.



Obr.3) Ukážka funkcie HOG [10.17]

*Základný princíp:*

1. Pre každý pixel zistíme smer k najtmavšiemu pixelovému susedovi v jeho blízkom okolí.
2. Nakreslíme v tomto mieste šípku smerujúcu k tmavému susedovi. Vytvoríme tak gradient intenzity v obraze. Gradienty ukazujú tok od svetlých oblastí ku tmavým oblastiam.
3. Celý obraz rozdelíme na subatice o veľkosti zvyčajne 16x16 pixelov. Vytvoríme jednu centrálnu šípku podľa vhodného kritéria, napríklad funkciou max().
4. Pre náš model auta sme vykonali rovnaký postup. Porovnávame oboje Gradienty a hľadáme zhodu.

*Nevýhody:*

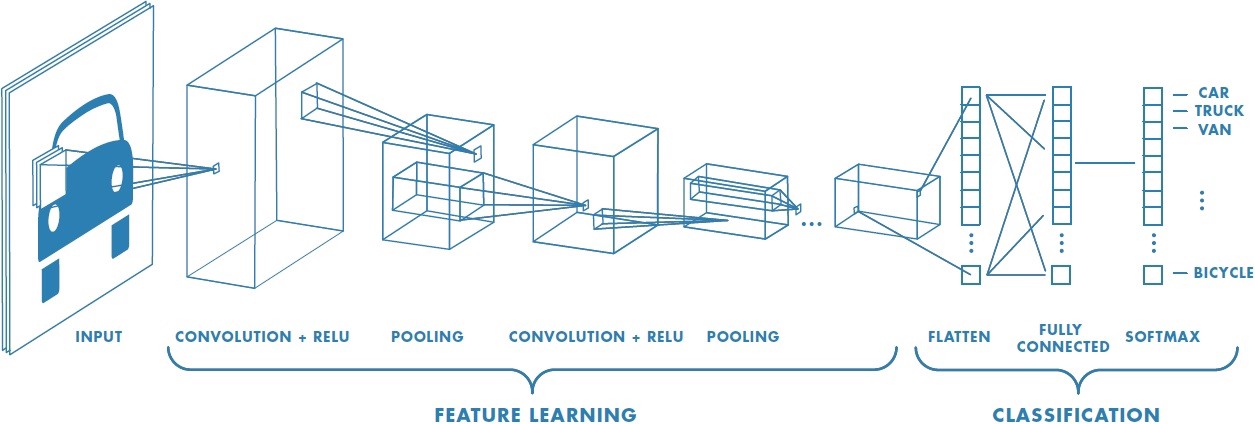
Vlastnosti sú zakódované natvrdo. Pri zmene obrazu alebo šum zlyhávajú. Neposkytujú teda užívateľovi dostatočnú dynamiku. Auto musí byť natočené rovnako, ako bolo modelové auto pri učení. Využívajú kĺzavé okno, takže spomalenie programu.

***Prístup s využitím neurónových sietí:***

Nie je žiadnym prekvapením, že neurónové siete nahradili v efektivite bežne používané postupy ako aj v detekcii objektov tak aj pri ich klasifikácii.

**CNN (2012)** [10.15]

Výsledky výkonnostného testu na Kriszhevskeho CNN ImageNete ukázali možnosť využitia konvolučných neurónových sietí. Používa sa kĺzavé okno pohybujúce sa po celom obraze a prebieha klasifikácia. Vznikne nám tak obrovský počet klasifikovaných objektov, väčšina však má veľmi malé confidence skóre – pravdepodobnosť, že sa nám v obrázku nachádza objekt porovnávaného klasifikátoru. Tento spôsob funguje, ale je pomalý. Je to spôsobené práve veľkým počtom porovnávaní.



Obr.4) Architektúra bežných CNN sietí [10.15]

*Princíp:*

1. Zoberieme obrázok, rozsekáme ho na sériu submatíc obrazu.
2. Na každú submaticu obrazu spustíme CNN klasifikátor.
3. Pre každú class klasifikátora vypočítame confidence skóre, reprezentujúce pravdepodobnosť, že sa v obrázku nachádza objekt daného klasifikátoru.
4. Uchováme si iba značky s dostatočne vysokým confidence skóre.

*Nevýhody:*

Je pomalý. Prebieha tu veľký počet výpočtov a len ťažko použiteľný pre real-time aplikácie.

**OverFeat(2013)**[10.9]

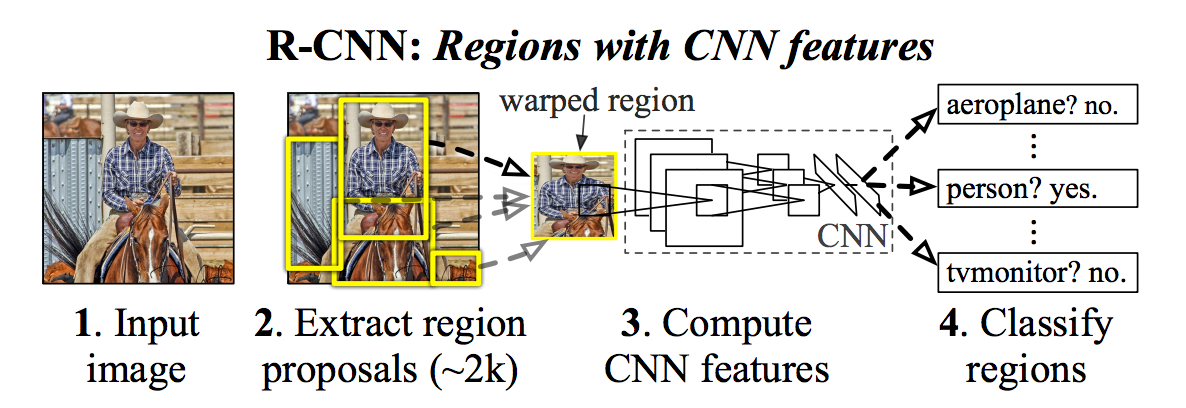
Jeden z prvých pokrokov od začiatku používania CNN.Využíva viacero rôzne veľkých okien kĺzajúcich po obraze. V publikovanej verzii využívali 6 rôzne veľkých okien. Nevýhodou bolo, že výpočet prebiehal iba pre jedno okno súčasne ale na druhej strane bola zväčšená robustnosť kódu.

****

Obr.5) Zjednodušená ukážka funkcie kĺzajúcého okna [10.9]

**R-CNN (2014)** [10.0]

Inteligentnejšia metóda ako bežné CNN. Využíva procesu nazývaného Selektívne hľadanie. Selektívne hľadanie vytvára hraničné oblasti tzv. *Bounding boxes*. Na každý obrázok je pozerané cez okná rôznych veľkostí a  hľadá pixelovo podobné oblasti označujúce regiónové umiestnenie objektu. Následne sa extrahujú vlastnosti v obraze pomocou CNN siete. Na klasifikáciu sa použije SVM.



Obr.6) Ukážka princípu R-CNN [10.0]

*Postup:*

1. Vytvoríme vo vloženom obrázku návrhy na Bounding boxy.
2. Jednotlivé oblasti vybraté pomocou Bounding boxov sú analyzované CNN sieťou a sú získané vlastnosti v obraze.
3. SVM nám ponúkne najlepšiu zhodu s jedným z klasifikátorov.
4. Na bounding boxy použijeme lineárny regresný model na získanie presnejších koordinátov už klasifikovaného objektu.

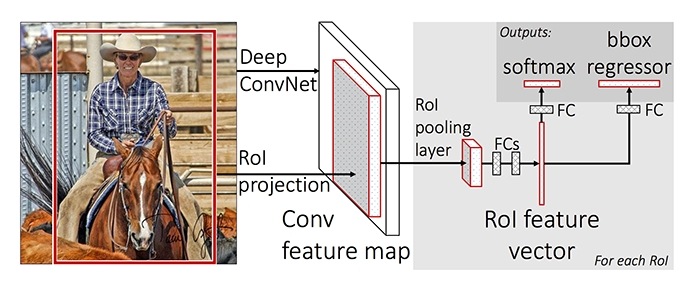
*Nevýhody:*

Pomalé pri tréningu aj pri testovaní. Každý regiónový návrh je posunutý do CNN samostatne. Učenie neurónovej siete na VOC 2007 datasete trvalo 84 hodín. Vlastnosti získané z CNN nie sú updatované na základe dát z SVM a regresie.

***Vylepšenia implementácie R-CNN:***

**Fast R-CNN (2015)** [10.16]

Ponúka viaceré vylepšenia oproti bežnému R-CNN v rýchlosti trénovania a aj testovania. Zároveň je zvýšená presnosť detekcie. Pre dataset z *PASCAL VOC (2012)* bola dosiahnutá 9x väčšia rýchlosť trénovania a 140-213x (podľa implementácie) pre testovanie ako pre bežné R-CNN. R-CNN je teda výrazne pomalšie. Fast využíva vlastnosti z R-CNN ako je napríklad návrhy bounding boxov a selektívne hľadanie. Rozdiel je v tom, že R-CNN extrahuje jednotlivé regióny samostatne a tie posiela do neurónovej sieti a na klasifikovanie pomocou SVM. Fast posiela celý obrázok s extrahovanými vlastnosťami. Fast R-CNN použije CNN na celý obrázok a následne nájde a zlúči regióny záujmu- ***Region of Interest (RoI) Pooling*** na obraz vlastností z CNN*.* Následne sa až posiela na klasifikáciu a regresiu. Tento prístup bol rýchlejší a umožňoval jednoduchšie učenie siete.



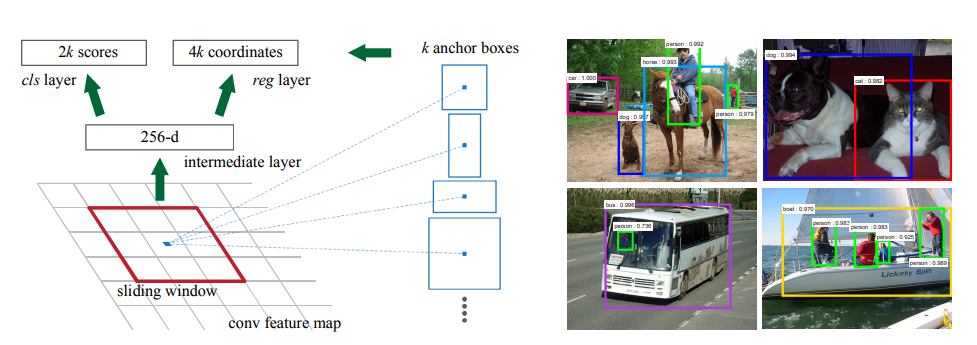
Obr.7) Ukážka funkcie Fast R-CNN [10.16]

Nevýhoda:

Jeho funkcia stále závisela na selektívnom vyhľadávaní alebo inej metóde regiónovo založeného algoritmu. Stalo sa to najslabším a najpomalším miestom tejto metódy.

**Faster R-CNN (2015):**[10.6] [10.7]

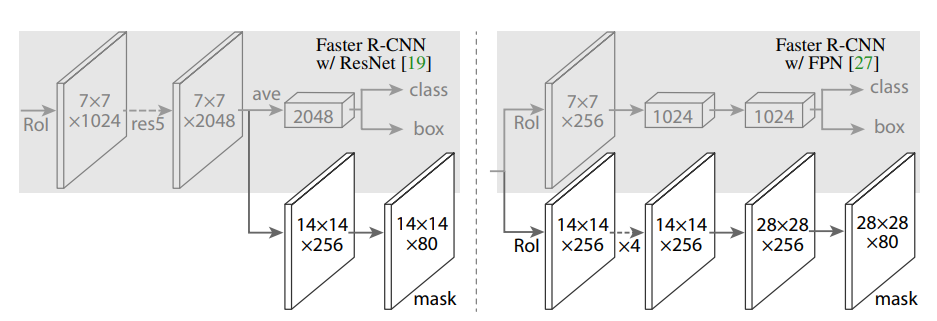
Výrazné zrýchlenie rýchlosti. Pre trénovanie na *VOC 2007* datasete rýchlosť učenia menej ako 1 hodina. Dosahuje až 250x zrýchlenie oproti R-CNN. Za zrýchlenie vďačia hlavne pridaním vrstiev pre odhad regiónov ***Region Proposal Network (RPN)***. Odstránili tak predchádzajúci slabý článok Fast R-CNN a to bolo selektívne vyhľadávanie. Výstupom RPN sú objekty s vlastnými objektovými hodnoteniami. Tie charakterizujú, ako moc sú si istý prítomnosťou objektov v obraze. Tieto objekty sú následne poslané do ďalších vrstiev CNN na hľadanie regiónov záujmu a klasifikáciu. Dosiahnem tak jednoducho kompletne trénovatelný model neurónovej siete.



Obr.8) Ukážka funkcie návrhu regiónov pomocou RPN. Testovanie prebehlo na VOC 2007 datasete. [10.7]

**Mask R-CNN(2017)** [10.8]

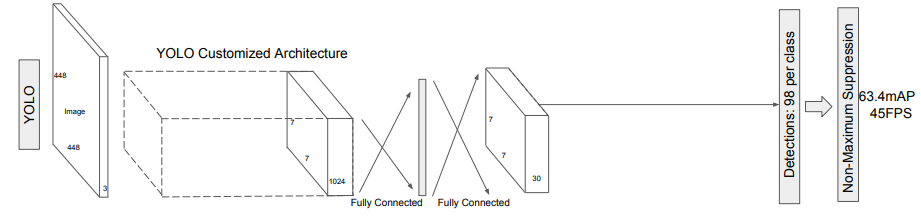
Využíva podobnú dvoj-stupňovú architektúru ako Faster R-CNN. Časť *Region Proposal Network (RPN)* je identická ako v metóde Faster R-CNN. V druhej časti, paralelne s predikciou skupiny Mask R-CNN vyhodnocuje aj binárnu masku pre každý *Region of Interest (RoI).* Maska je naučená pomocou plne prepojenej neurónovej siete. Pomáha to v pixelovom učení a hľadaní pixelovo-objektovej zhody príslušiacej jednému objektu na báze podobnosti. Získali sme tak veľmi dobrú objektovú segmentáciu. Nevýhodou je mierne zmenšenie v rýchlosti ktoré je však zanedbateľné pri zohľadnení prínosu, ktorý prináša.



Obrázok. 9) Ukážka implementácie Mask R-CNN na architektúre Faster R-CNN s využitím ResNetu a FPN [10.8]

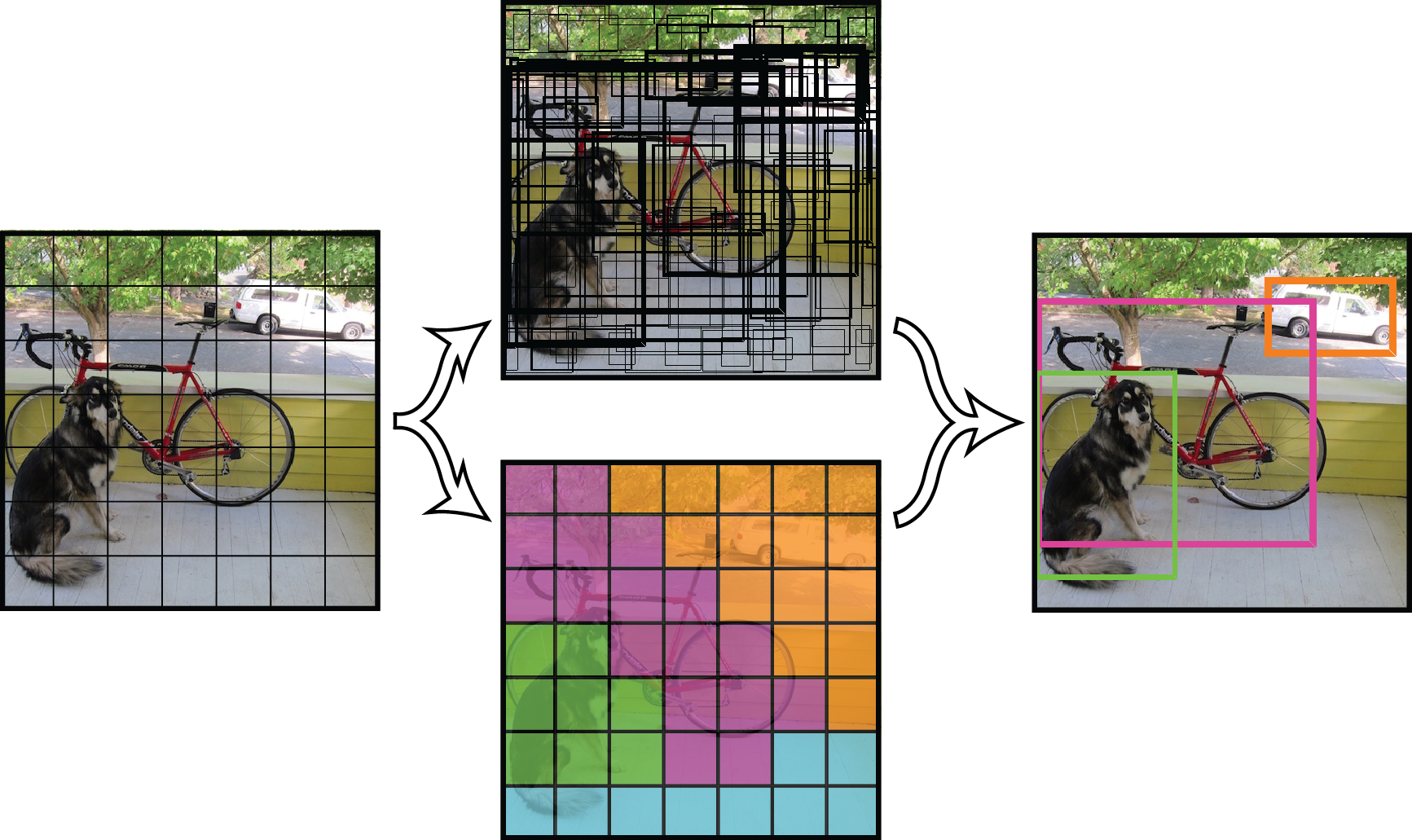
**YOLO - You Only Look Once (2016)** [10.10]

Všetky do teraz spomínané siete využívajú lokalizáciu a klasifikáciu objektov pre detekciu. Používajúce model na časti obrazu s rôznou veľkosťou a pohybujúce sa po obraze. YOLO však na rozdiel od nich využíva iba jednu jedinú sieť na celý obraz.



Obr.10) Architektúra YOLO [10.13]

Táto sieť rozdelí obraz na regióny a predikuje Bounding Boxy a pravdepodobnosti pre každý Bounding Box. V obraze sa nachádza obrovské množstvo detekcií ale väčšina má predvídanú pravdepodobnosť že sa v nich objekt nachádza veľmi nízku. Ak naprahujeme všetky nájdené Bounding Boxy a ponecháme v obraze iba tie, ktoré majú väčšiu pravdepodobnosť ako je dopredu stanovený prah, získame iba pár detekcií výrazne vyššie odpovedajúce skutočnému výskytu objektov.



Obr.11) Ukážka priebehu hľadania Bounding Boxov pre YOLO [10.10]

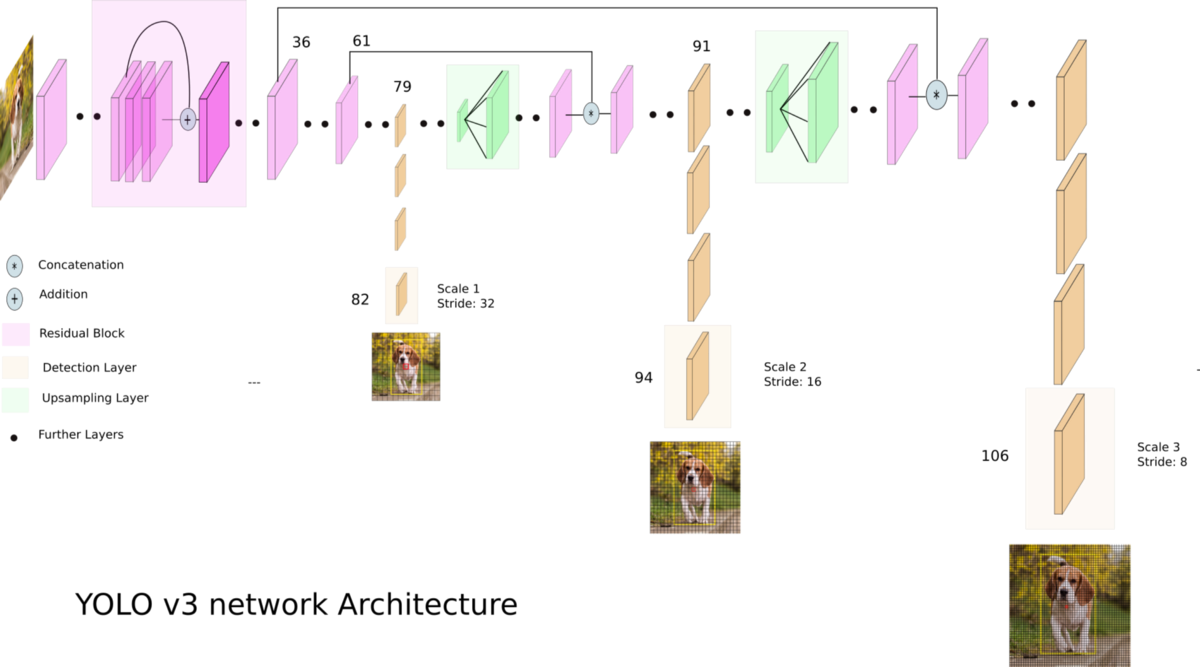
YOLO použije iba jediný priechod cez CNN pre celý obraz. Nepoužíva tisíce priechodov pre všetky oblasti obrazu. Je tým dosiahnutý výrazne rýchlejší chod a po prvý krát kompletne real-time aplikácia. Táto metóda pri tom dosahuje vysoké presnosti. Od vtedy však táto metóda prešla sériou vylepšení.

**YOLO v2 (9000): Better, Faster, Stronger 2017**[10.11]

Využíva zopár zlepšení a trikov ako zlepšiť efektivitu a rýchlosť pôvodného YOLO. Používa architektúru darknetu s 19 vrstvovou sieťou s ďalšími 11 vrstvami pre objektovú detekciu. Na Titan X podľa čerpanej literatúry dosahuje rýchlosť až 45 fps. Jeho veľkou nevýhodou bolo, že strácal kvalitu detekcie malých objektov z dôvodu podvzorkovania vstupného obrazu. Z tohto dôvodu sa začali používať zlúčenia aktuálnych vlastnostných máp a máp z predchádzajúcich vrstiev. Podobne ako OverFeat a SSD používa plne konvolučný model ale stále sa trénuje na celom obrázku. Podobne ako pri Faster R-CNN upravujeme hranice na Bounding Boxoch namiesto priamej počiatočnej predikcie hraníc.

**YOLO v3** **: An Incremental Improvement 2018** [10.12]

YOLO v2 patrilo v svojom čase k najrýchlejším a najpresnejším detektorom objektov v obraze. Netrvalo však dlho a prišlo sa s metódami ako RetinaNet a SSD(popísané nižšie), ktoré ho výkonnostne predbehli. Rýchlostne však stále patril k najrýchlejším. YOLO v3 používa až 53 pôvodných vrstiev Darknetu pre extrahovanie vlastností obrazu a ďalších 53 vrstiev pre detekciu. Prišlo k zhoršeniu rýchlosti za cenu zvýšenia presnosti detekcie. Dosahuje iba 30 fps. Veľkou výhodou je aj zlepšenie detekcie malých objektov vďaka doprednému pričítaniu predchádzajúcich nepodvzorkovaných máp vlastností.

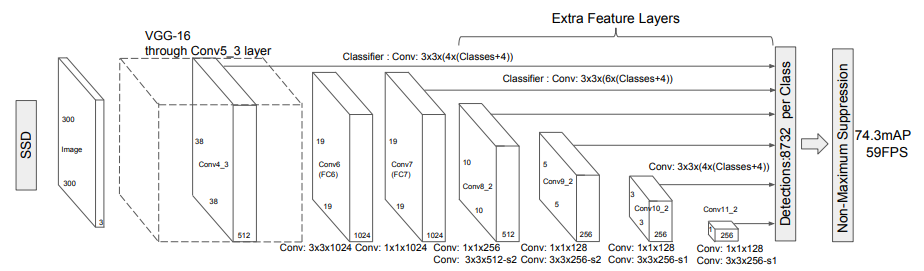


Obr.12) Architektúra YOLO v3 [10.12]

Pre YOLO v3 je detekcia navrhnutá ako aplikáciou 1x1 detekčných ker nelov na mapu vlastností o troch odlišných veľkostiach a v iných častiach siete.

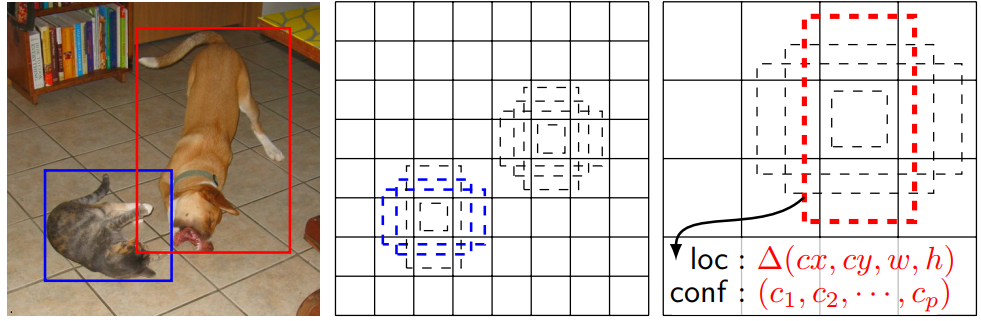
**SSD - Single Shot MultiBox Detector** [10.12]

Predchádzajúce metódy detekcie objektov zdielali jednu rovnakú vlastnosť. Sú rozdelené na dve časti. Prvá časť je určená k odhadom regiónového ohraničenia objektov pomocou Bounding Boxes a druhá časť s kvalitnými klasifikátormi na klasifikáciu objektov v ohraničení.



Obr.13) Architektúra SSD [10.13]

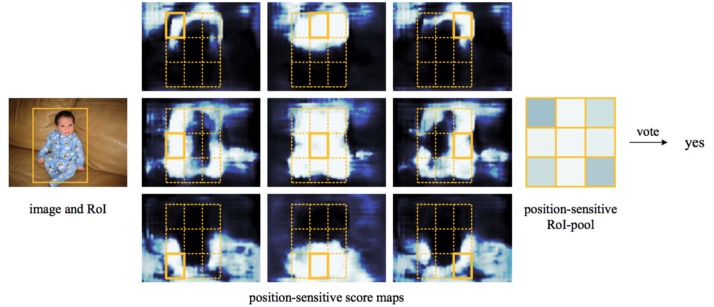
Tieto metódy sú veľmi presné, ale platia za to vysokými výpočetnými nákladmi – malé fps. Z tohto dôvodu nie sú vhodným kandidátom na použitie v embedded zariadeniach. SSD je navrhnuté tak, aby skombinovali tieto dve úlohy do jednej siete. Namiesto siete vytvárajúcej návrhy na Bounding Boxy môžeme mať predefinované boxy, v ktorých hľadáme objekty. Používajúc vlastností obrazu získaných z konvolučných máp z vrstiev, nachádzajúcich sa v druhej polovici siete, použijeme malé konvolučné filtre na predikciu máp vlastností z obrazu na predikciu pravdepodobnosti výskytu objektu v obraze a samotného bounding boxu. Pre nájdenie Bounding Boxov a triedy objektu používame viaceré aktivačné mapy s rôznymi veľkosťami pre predikciu.



Obr. 14) Ukážka nájdenia vlastností pomocou máp vlastností v obraze [10.13]

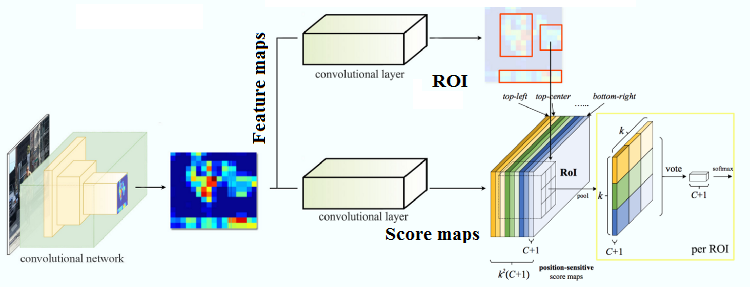
**R-FCN - Region-based Fully Convolutional Networks [10.14]**

R-FCN prináša zlepšenie rýchlosti vďaka zníženiu potrebnej práce potrebnej pre každú oblasť záujmu. Mapa vlastností je posielaná na paralelne vetvy. V jednej prebieha klasický výpočet oblastí záujmov a v druhej sa nachádzajú tzv. Score maps. V týchto skóre mapách sa nachádza zhoda časti obrazu s objektom. Na nasledujúcom obrázku je ukážka princípu skóre máp. Oblasť Bounding boxu je rozčiastkovaná v tomto prípade na 9 subblokov s vlastnými mapami vlastností a v každom sa porovnáva zhoda s časťou objektu, ktorá by sa a nachádzať na rovnakej pozícii, ak by tam objekt bol. V obrázku sa snažíme nájsť v objekte dieťa. Lokalizácia prebehla úspešne.

****

Obr.15) Ukážka princípu Skóre máp [10.14]

Samostatné mapy vlastností pre každý subblok nedáva výrazne precíznu odpoveď. Ak ich ale skombinujeme viacej, v tomto prípade 9, a vieme čo sa má nachádzať v ktorej časti objetu, dokážeme výrazne spresniť detekciu. Architektúra siete je na nasledujúcom obrázku. Využíva podobnú architektúru ako Faster R-CCN ale iba s konvolučnými sieťami.

****

Obr.16) Architektúra R-FCN [10.14]

**Výber vhodného datasetu [10.0]**

Dataset môže zohrávať významnú rolu v procese učenia neurónových sietí. Pri učení sa požaduje dostatočná variácia zisťovaných objektov ale zároveň dostatočná specificita oblasti použitia. Napríklad pri detekcii značiek v Českej Republike bude vhodné učiť NN na značky získané v ČR. Odlišnosti v značke môžu zhoršovať ich detekciu. Na druhej strane napríklad pri príchode zahraničného vozidla musí byť značka detekovaná tiež.

Pri väčšom množstve dát pri učení neurónovej sieti stúpa efektívnosť neurónovej sieti. Nezabúdame pri tom na možnosť preučenia. V praxi sa však ťažko dostávame k väčšiemu objemu obrázkov použiteľných pre učenie úloh v oblasti klasifikácie objektov. Firmy nerady dávajú svoje dáta zadarmo a získavanie obrázkov napríklad pomocou Google môže byť prácne na anotáciu a triedenie. V nasledujúcej tabuľke je zoznam najznámejších bežne používaných voľne dostupných datasetov:

|  |  |  |  |
| --- | --- | --- | --- |
| **Názov datasetu** | **Počet obrázkov** | **Počet tried** | **Posledný update** |
| COCO [10.1] | 200K | 80 | 2017 |
| Pascal Voc [10.2] | 12K | 20 | 2012 |
| Oxford-IIIT Pet [10.3] | 7,5K | 37 | 2012 |
| KITTI Vision [10.4] | 15K | Only cars 3 | 2012 |
| Stanford Car dataset [10.5] | 16K | 196 of cars | 2012 |

Tab 1) Najčastejšie používané datasety pre úlohy klasifikácie objektov

Voľnodostupné porovnania výkonnosti z iných meraní:

Použité neurónové siete:

**SSD: Single Shot MultiBox Object Detector využívajúci PyTorch [5.0]**

**citácie:**

[10.0] https://tryolabs.com/blog/2017/08/30/object-detection-an-overview-in-the-age-of-deep-learning/

[10.1] http://cocodataset.org/#download

[10.2] http://host.robots.ox.ac.uk/pascal/VOC/

[10.3] http://www.robots.ox.ac.uk/~vgg/data/pets/

[10.4] http://www.cvlibs.net/datasets/kitti/eval\_object.php?obj\_benchmark=2d

[10.5] <http://ai.stanford.edu/~jkrause/cars/car_dataset.html>

[10.6] https://web.cs.hacettepe.edu.tr/~aykut/classes/spring2016/bil722/slides/w05-FasterR-CNN.pdf

[10.7] <https://arxiv.org/pdf/1506.01497.pdf>

[10.8] <https://arxiv.org/pdf/1703.06870.pdf>

[10.9] <http://slideplayer.com/slide/10395667/>

[10.10] <https://arxiv.org/pdf/1506.02640.pdf>

[10.11] <https://pjreddie.com/media/files/papers/YOLO9000.pdf>

[10.12] https://arxiv.org/pdf/1512.02325.pdf

[10.12 ]<https://towardsdatascience.com/yolo-v3-object-detection-53fb7d3bfe6b>

[10.13] <https://arxiv.org/pdf/1512.02325.pdf>

[10.14] <https://medium.com/@jonathan_hui/understanding-region-based-fully-convolutional-networks-r-fcn-for-object-detection-828316f07c99>

[10.15] <https://medium.com/@rohanthomas.me/convolutional-networks-for-everyone-1d0699de1a9d>

[10.16] <https://arxiv.org/pdf/1504.08083.pdf>

[10.17]<https://towardsdatascience.com/detecting-vehicles-using-machine-learning-and-computer-vision-e319ee149e10>

[10.18] <https://www.superdatascience.com/opencv-face-detection/>

[10.19]http://journals.plos.org/plosone/article/figure?id=10.1371/journal.pone.0173424.g002

[5.0] https://github.com/amdegroot/ssd.pytorch