

Text detection and recognition in natural scene images - a real challenge.

DÁVID GAJDOŠ, TOMÁŠ JUŠČÍK, RICHARD KAČUR
PETER BUGATA, DÁVID HUDÁK



Machine Learning Prague 2020

MARCH 20 - 22

1 000+

ATTENDEES

45

SPEAKERS

10

WORKSHOPS

Code 20% off - mlmu20

VSL Software, a. s.

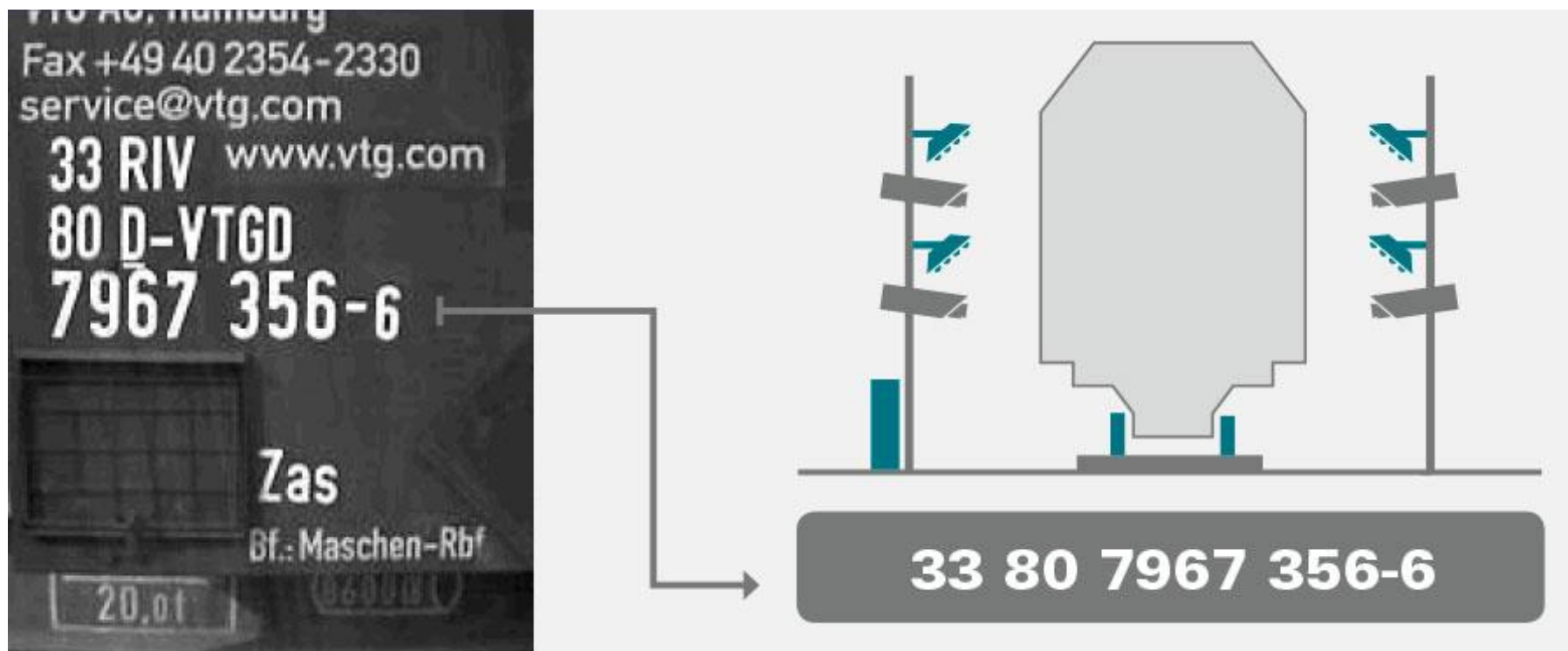
- Návrh, vývoj a údržba informačných systémov
- V minulom roku spoločnosť oslávila 25. výročie vzniku
- Väčšina informačných systémov spracováva veľké objemy údajov uložených v relačných databázach, budovanie dátových skladov
- Oddelenie analýzy dát a strojového učenia
- Spolupráca s PF UPJŠ, projekty v oblasti zdravotného poistenia
- Mladý tím, zapojenie študentov
- Letná stáž 2019, výsledky práce troch študentov z TUKE

DEFINÍCIA ZADANIA

Rozpoznanie 12-ciferného
čísła vozňa číslovaného
podľa UIC



PRÍKLAD BUSINESS RIEŠENIA



MULTIRAIL® IDentify

ÚLOHA

- Ľahká alebo ťažká?
- Natural scene text recognition – všeobecne zložitá úloha
- Rôzne typy písma, rôzne šablóny
- Odlesky, tiene, rozmazanie
- Neexistuje dostatočne veľký verejný dataset vhodný pre našu úlohu

STREET VIEW HOUSE NUMBERS (SVHN)

- Dáta zo skutočného sveta
- Viac ako 600 tisíc obrázkov čísel
- Získaný z Google Street View images
- Častý referenčný dataset pre vývoj a testovanie algoritmov



BUSINESS POHL'AD NA ÚLOHU

- Ian Goodfellow - čísla domov (SVHN)
2013/14 - 95.64% coverage, 98% accuracy
- Požadovaná podobná alebo lepšia presnosť ako človek
- Accuracy vs Coverage
- Kritický pohľad na chyby (problémy v prevádzke)
- Performance

Ukazovateľ	Leto	Cieľ
Coverage	95%	95.0%
Accuracy	98%	99.5%
Performance	-	<200 ms (5-16 fps)

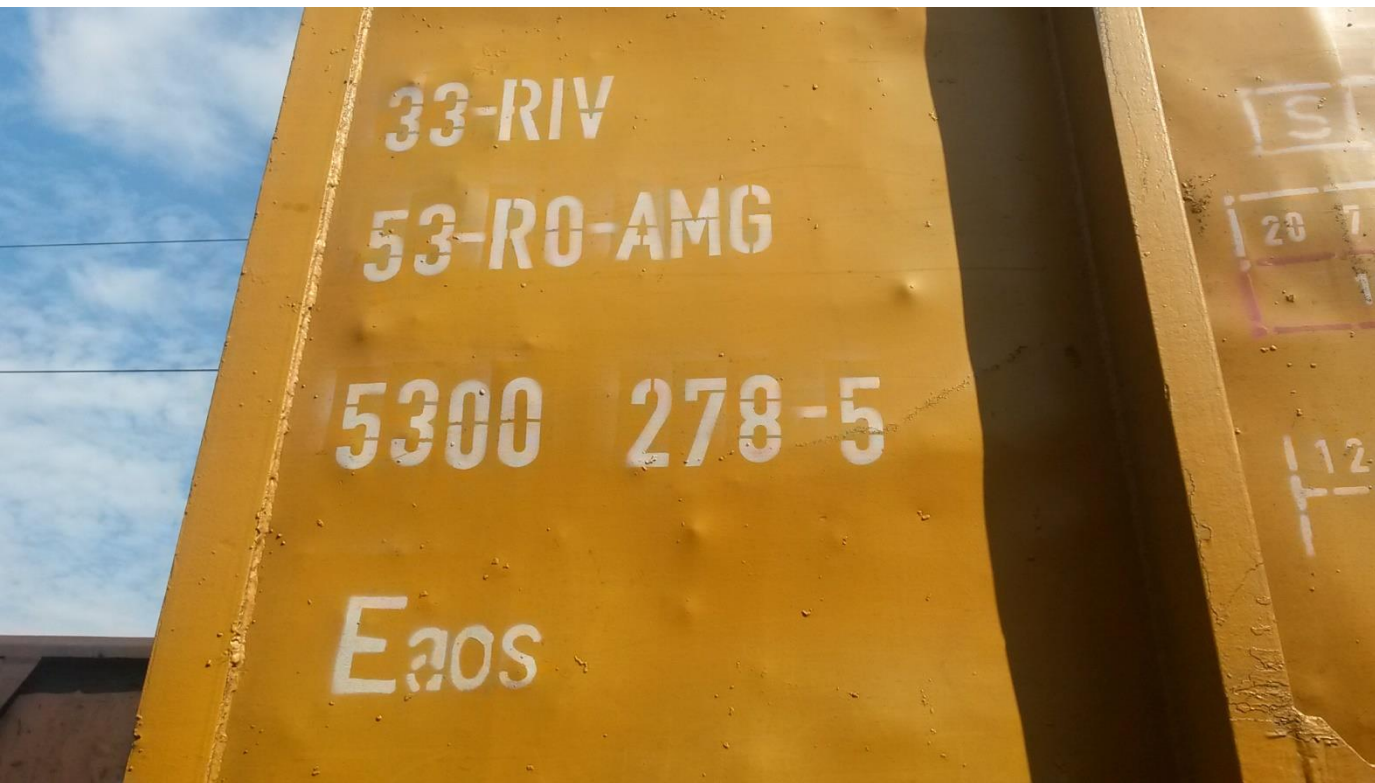
ZHOTOVENIE VLASTNÉHO DATASETU

- Fotenie vozňov 02.07.2019 v Čiernej nad Tisou
- 1503 fotografií rôznych vozňov
- 4 rôzne mobilné telefóny s rôznymi rozlíšeniami
- Súbory pomenované na základe čísla vozňa
- Validácia (ručne, algoritmom na overenie pomocou kontrolnej číslice)
- Odstránenie nevhodných snímok (neúplná fotografia, nečitateľné číslo a pod.) - asi 5 snímok

PRÍKLADY PRE OBRÁZKY Z DATASETU



PRÍKLADY PRE OBRÁZKY Z DATASETU



PRÍKLADY PRE OBRÁZKY Z DATASETU

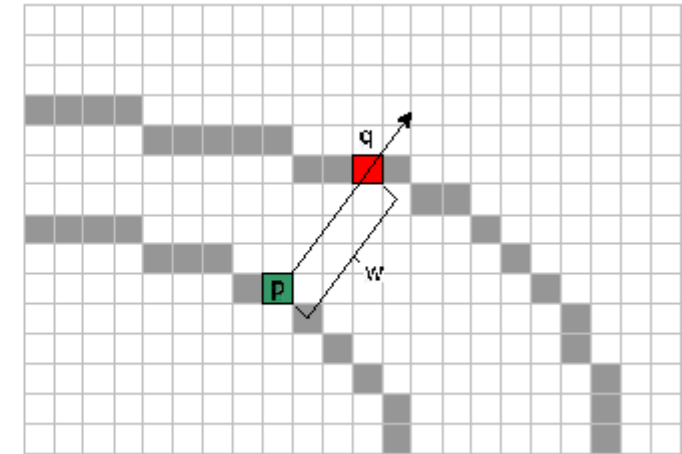
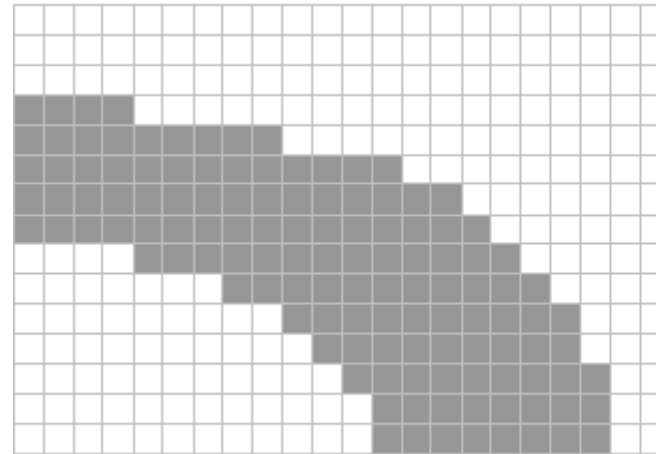


PRÍKLADY PRE OBRÁZKY Z DATASETU



STROKE WIDTH TRANSFORM (SWT)

- Hľadanie písmen na základe šírky ťahu
- Grayscale – podobná intenzita
- Získanie súvislých komponentov
- Vhodné pre rôzne typy, veľkosti a orientácie textov
- Zvláda aj “prírodnú scénu” (natural scene) - množstvo rušivých prvkov, šum
- Pri praktickom využití problém už s detekciou hrán pri zlom osvetlení (Canny detektor) a pri dierach vznikajúcich v písmenách



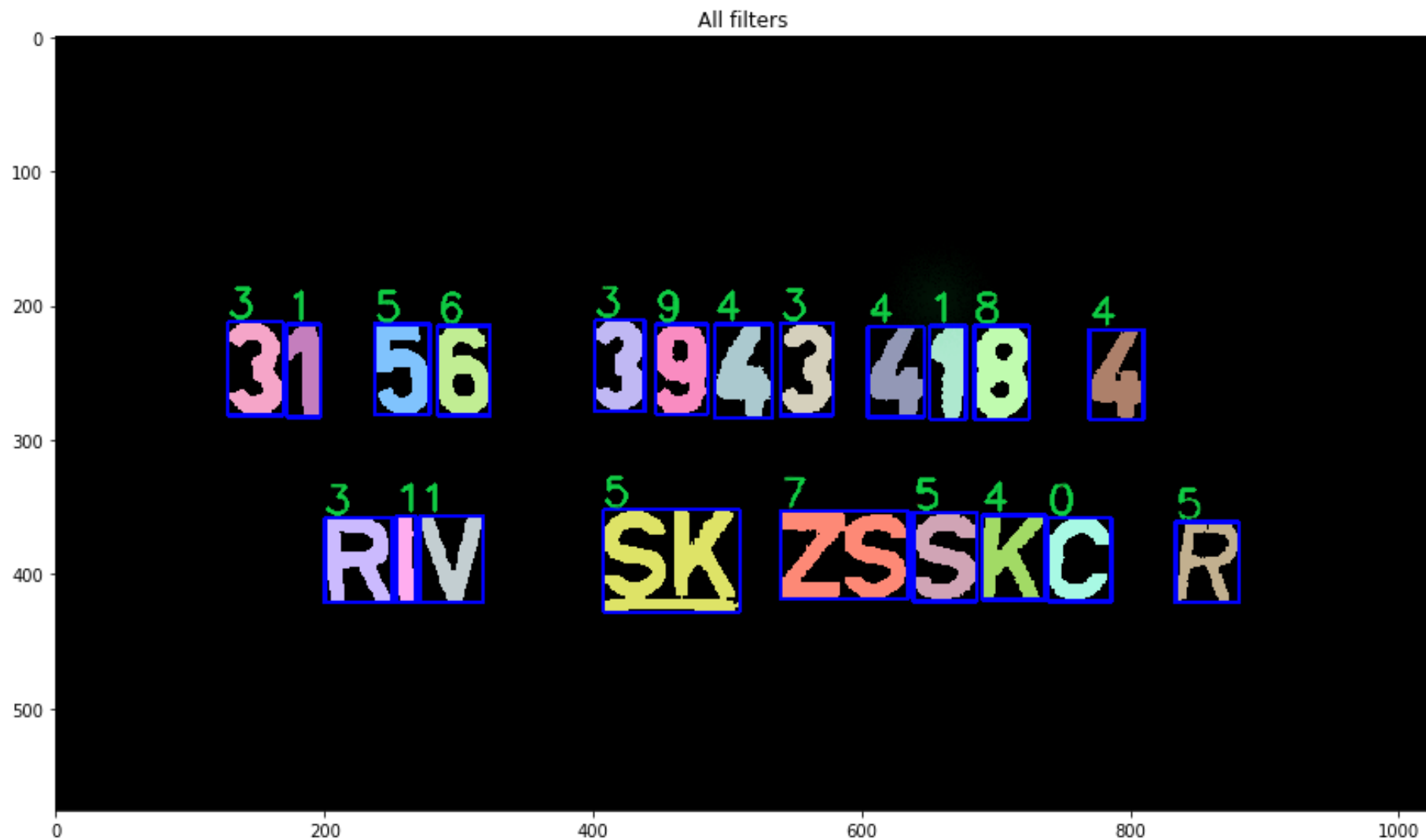
MAXIMALLY STABLE EXTREMAL REGIONS (MSER)

- Hľadanie susediacich oblastí s podobným jasom, ohraničených kontrastným pozadím
- Rôzne úrovne prahovania pre oddelenie popredia a pozadia
- Výsledné regióny sú tie, ktorých tvary sa príliš nemenia pri rôznych prahoch



MSER + CNN (SVHN)

- Prvý funkčný prototyp – kompletný workflow
- Metóda MSER pre získanie čísel (súčasť OpenCV)
- Filtrovanie a spájanie regiónov tvoriacich jedno číslo
- Klasifikácia čísel pomocou konvolučnej neurónovej siete natrénovanej na [Street View House Numbers \(SVHN\) datasete](#)
- Nedostatočná presnosť, množstvo problémov a chybovosť (tiene, medzery v jednom písmene, čísla príliš blízko seba, protisvetlo a odlesky a rozpadávanie čísel)



Classification



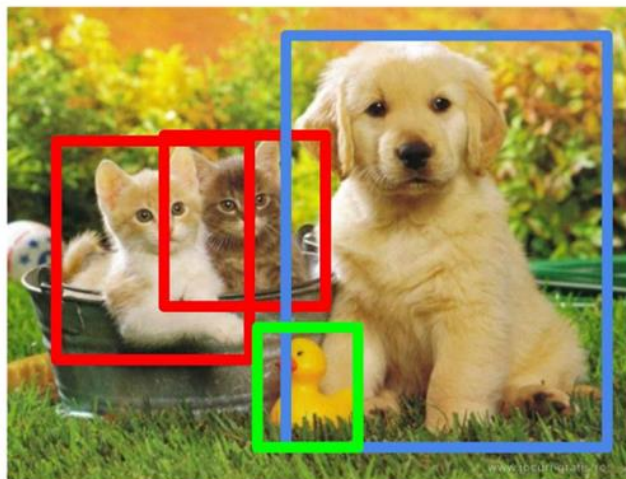
CAT

Classification + Localization



CAT

Object Detection



CAT, DOG, DUCK

Instance Segmentation



CAT, DOG, DUCK

Single object

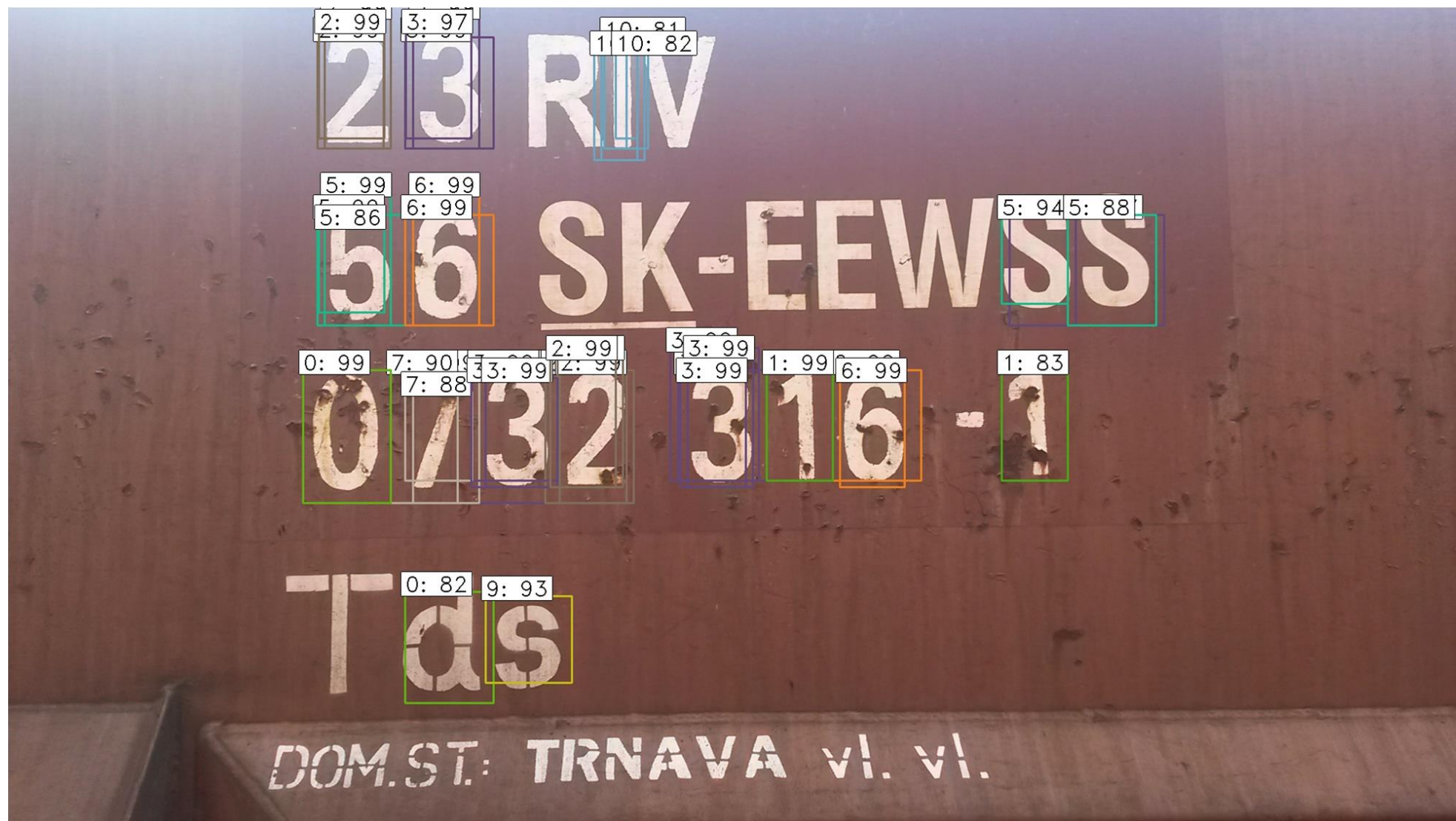
Multiple objects

FASTER R-CNN

- Tretia generácia z rodiny objektových detektorov (po R-CNN a Fast R-CNN)
- Dve siete spojené spoločnou chrbticou - backbone (jedna sieť pre návrh regiónov a extrakciu črt, druhá na klasifikáciu objektu a regresné spresnenie jeho polohy)
- Spoločný tréning oboch sietí (striedavá aktualizácia váh v spoločnej časti)

VYUŽITIE KONVOLUČNEJ SIETE

- Labelovanie 150 + 6 obrázkov pre tréning (10 číslíc + „I“) – stratifikovaný výber obrázkov do tréningovej množiny
- Faster R-CNN dvojstupňový detektor, nastavenie vlastných parametrov (veľkosti objektov a ich pomery strán – nastavenie hyperparametrov na základe štatistiky z trénovacej množiny)
- Data augmentation (umelé rozšírenie trénovacej množiny)
- Vstup – obrázok (3 kanály - R,G,B), výstup – súradnice bounding boxov a predikovaná trieda pre každý objekt
- Tréning na grafickej karte Nvidia GTX 1080, niektoré trénované dlhšie ako 24 hodín (niekedy aj 36 hodín)



SPRACOVANIE VÝSTUPU

- Parsovanie čísel na základe ich polohy na obrázku
- Overenie výsledku pomocou kontrolnej číslice, porovnanie s názvom súboru
- Rekurzívne spracovávanie ak sme nenašli všetky čísla alebo neseďí kontrolná číslica
- Vynechanie najmenej pravdepodobnej číslice, prvej alebo poslednej číslice, kontrola pomocou slovníka, priblíženie na čísla,...

VÝSLEDKY – NAJLEPŠIE (FASTER R-CNN / VGG16 BACKBONE)

Trénovacia množina

Celkový počet	156
Rozpoznalo	154
Chyby	0
Coverage	98.72%
Accuracy	100%

Testovacia množina

Celkový počet	1343
Rozpoznalo	1275
Chyby	3
Coverage	94.937%
Accuracy	99.765%

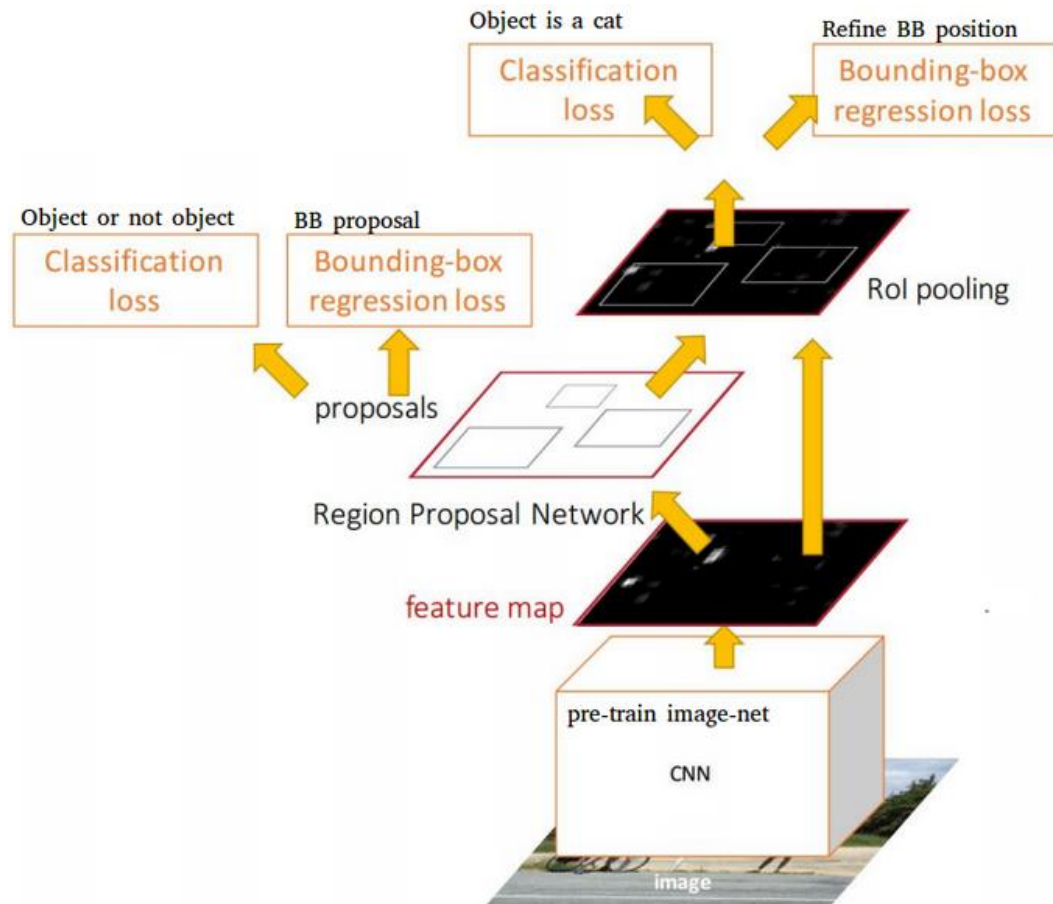
PRÍNOSY LETNEJ STÁŽE

- Práca v tíme
- Kooperácia s firmou
- Práca na vzdialenom serveri prostredníctvom VPN a terminálu
- Rozšírenie znalostí v obore Computer Vision a Deep Learning
- Zlepšenie v jazyku Python

ĎAKUJEME ZA VAŠU POZORNOST



APPENDIX: Faster R-CNN - architektúra



- ✓ potlačenie prekrývajúcich sa objektov
- ✓ spresnenie pozície objektov
- ✓ multiclass klasifikácia s pridanou triedou
- ✓ výber virtuálnych atribútov z poslednej konv. vrstvy

- ✓ regióny s najväčšou pravdepodobnosťou (napr. 300)
- ✓ malé minibatch-e, napr. len po 1 obrázku
- ✓ výber pozitívnych a negatívnych kotviacich boxov
- ✓ viac ako 16 000 kotviacich boxov
- ✓ rôzne rozmery, napr. 800 x 600

APPENDIX: Faster R-CNN – návrh regiónov

- Dvojica neurónových sietí so „spoločnou chrbticou“:
 - Region Proposal Network (RPN)
 - Detector
- RPN má štruktúrovaný výstup – pre každý kotviaci box rieši binárnu klasifikačnú úlohu a regresnú úlohu (štyri súradnice pre upresnenie záujmového regiónu)
- RPN vie spracovať obrázky rôznych rozmerov, typicky 800 x 600 – v takom prípade obvykle viac ako 16 000 kotviacich boxov
- V minibatch-i obvykle malý počet obrázkov, napr. len jeden
- Pri tréningu výber negatívnych prípadov – účelová funkcia sa aplikuje len na vybrané negatívne kotviace boxy.
- Pri predikcii sa vyberá určený počet (napr. 300) regiónov s najvyššou pravdepodobnosťou

APPENDIX: Faster R-CNN – detekcia objektov

- Detector má štruktúrovaný výstup – rieši multiclass klasifikáciu (s pridanou triedou ‘background’) a regresnú úlohu pre spresnenie pozície objektu
- Vysoká rýchlosť predikcie sa dosahuje tým, že Detector dostáva už predspracované regióny (virtuálne premenné), netreba aplikovať konvolučné vrstvy
- Pri tréningu sa využije výstup RPN siete s vhodným pomerom pozitívnych a negatívnych regiónov
- Dvojicu prepojených sietí je možné trénovať postupne alebo „súčasne“, my sme použili súčasný tréning, keď sme striedavo aktualizovali váhy v spoločnej časti na základe gradientu v jednej a druhej sieti

Special thanks for support to :

KOŠICE
IT VALLEY

T · · Systems ·

sas



Data Analytics Meetings

Machine Learning Prague 2020

MARCH 20 - 22

1 000+

ATTENDEES

45

SPEAKERS

10

WORKSHOPS

Code 20% off - mlmu20