



**VYSOKÉ UČENÍ TECHNICKÉ V BRNĚ**

BRNO UNIVERSITY OF TECHNOLOGY

**FAKULTA INFORMAČNÍCH TECHNOLOGIÍ**

FACULTY OF INFORMATION TECHNOLOGY

**ÚSTAV POČÍTAČOVÉ GRAFIKY A MULTIMÉDIÍ**

DEPARTMENT OF COMPUTER GRAPHICS AND MULTIMEDIA

**GENERÁTOR SYNTETICKÉ DATOVÉ SADY  
PRO DOPRAVNÍ ANALÝZU**

SYNTHETIC DATASET GENERATOR FOR TRAFFIC ANALYSIS

**BAKALÁŘSKÁ PRÁCE**

BACHELOR'S THESIS

**AUTOR PRÁCE**

AUTHOR

**ONDREJ SVOREŇ**

**VEDOUCÍ PRÁCE**

SUPERVISOR

**prof. Ing. ADAM HEROUT, PhD.**

BRNO 2018

**Vysoké učení technické v Brně - Fakulta informačních technologií**

Ústav počítačové grafiky a multimédií

Akademický rok 2017/2018

**Zadání bakalářské práce**

Řešitel: **Svoreň Ondřej**

Obor: Informační technologie

Téma: **Generátor syntetické datové sady pro dopravní analýzu**  
**Synthetic Dataset Generator for Traffic Analysis**

Kategorie: Počítačová grafika

**Pokyny:**

1. Seznamte se s problematikou monitorování dopravy počítačovým viděním; zaměřte se na algoritmy strojového učení a jejich požadavky na datové sady.
2. Po konzultaci s vedoucím identifikujte vhodné datové sady, které by bylo vhodné generovat synteticky; specifikujte požadované vlastnosti.
3. Vytvořte generátor/y řešených datových sad, data generujte.
4. Získejte a nebo vytvořte příklady aplikací strojového učení a experimentujte s jejich chováním na generovaných datových sadách.
5. Iterativně vylepšujte generátor/y tak, aby syntetické datové sady měly co nejlepší vlastnosti pro strojové učení a jeho vyhodnocování.
6. Zhodnoťte dosažené výsledky a navrhněte možnosti pokračování projektu; vytvořte plakátek a krátké video pro prezentování projektu.

**Literatura:**

- Gary Bradski, Adrian Kaehler: Learning OpenCV; Computer Vision with the OpenCV Library, O'Reilly Media, 2008
- Richard Szeliski: Computer Vision: Algorithms and Applications, Springer, 2011
- Jakub Špaňhel et al.: Holistic Recognition of Low Quality License Plates by CNN using Track Annotated Data, IWT4S-AVSS 2017
- Jakub Sochor et al.: BrnoCompSpeed: Review of Traffic Camera Calibration and A Comprehensive Dataset for Monocular Speed Measurement, arXiv:1702.06441

Pro udělení zápočtu za první semestr je požadováno:

- Body 1 a 2, značné rozpracování bodů 3 až 5.

Podrobné závazné pokyny pro vypracování bakalářské práce naleznete na adrese <http://www.fit.vutbr.cz/info/szz/>

Technická zpráva bakalářské práce musí obsahovat formulaci cíle, charakteristiku současného stavu, teoretická a odborná východiska řešených problémů a specifikaci etap (20 až 30% celkového rozsahu technické zprávy).

Student odevzdá v jednom výtisku technickou zprávu a v elektronické podobě zdrojový text technické zprávy, úplnou programovou dokumentaci a zdrojové texty programů. Informace v elektronické podobě budou uloženy na standardním nepřepisovatelném paměťovém médiu (CD-R, DVD-R, apod.), které bude vloženo do písemné zprávy tak, aby nemohlo dojít k jeho ztrátě při běžné manipulaci.

Vedoucí: **Herout Adam, prof. Ing., Ph.D., UPGM FIT VUT**

Datum zadání: 1. listopadu 2017

Datum odevzdání: 16. května 2018

**VYSOKÉ UČENÍ TECHNICKÉ V BRNĚ**

Fakulta informačních technologií  
Ústav počítačové grafiky a multimédií  
L.S.612 66 Brno, Božetěchova 2



---

doc. Dr. Ing. Jan Černocký  
vedoucí ústavu

## Abstrakt

Táto bakalárska práca sa zaoberá vytvorením generátora syntetickej dátovej sady pre dopravnú analýzu. V práci sa venujem problematike analýzy dopravy prostredníctvom počítačového videnia, spôsobu a podmienkam vytvorenia generátora syntetickej dátovej sady, potenciálnemu využitiu výsledkov na strojové učenie a možnostiam ďalšieho vývoja. Zo zastaraných fotografií automobilov Českej republiky, Slovenska, Poľska a Maďarska som vytvoril generátor syntetických poznávacích značiek týchto štátov, ktoré tvoria po grafických úpravách a po spojení s fotografiami automobilov výslednú dátovú sadu určenú na strojové učenie. Samotné riešenie je rozdelené do troch skriptov v jazyku Python s použitím knižnice OpenCV. Výsledná dátová sada slúži systému využívajúcemu strojové učenie na reidentifikáciu evidenčných čísel vozidiel z fotografií zachytených v cestnej premávke.

## Abstract

This bachelor thesis deals with the creation and customization of synthetic dataset generator for traffic analysis. It focuses on traffic analysis by means of computer vision, methods and conditions of creating the generator of synthetic dataset, possible application of achieved results in machine learning and additional development opportunities. Using available automobile photographs from the Czech Republic, Slovakia, Poland and Hungary, a synthetic license plate number generator was created, which, after graphical adjustment and after joining with the vehicle photographs creates the resulting dataset for machine learning. The solution itself is divided into the three scripts written in Python using the OpenCV library. The resulting dataset serves as an input for the machine learning system to re-identify the license plate numbers from photographs captured in the flow of traffic.

## Klíčová slova

analýza dopravy, generátor syntetickej dátovej sady, strojové učenie, reidentifikácia ŠPZ, rozpoznávanie ŠPZ, registračné číslo, registračná značka, syntetická dátová sada, syntetická ŠPZ, OpenCV, sledovanie dopravy radarovými systémami, spracovanie obrazu, sledovanie vozidiel

## Keywords

traffic analysis, synthetic dataset generator, machine learning, reidentification of license numbers, recognition of license numbers, vehicle identification number, license plate, synthetic dataset, synthetic license plate, OpenCV, traffic surveillance, image processing, vehicle tracking

## Citace

SVOREŇ, Ondrej. *Generátor syntetické datové sady pro dopravní analýzu*. Brno, 2018. Bakalářská práce. Vysoké učení technické v Brně, Fakulta informačních technologií. Vedoucí práce prof. Ing. Adam Herout, PhD.

# Generátor syntetické datové sady pro dopravní analýzu

## Prohlášení

Prehlasujem, že som túto bakalársku prácu vypracoval samostatne pod vedením pána prof. Ing. Adama Herouta, Ph.D. Uviedol som všetky literárne pramene a publikácie, z ktorých som čerpal.

.....  
Ondrej Svoreň  
16. května 2018

## Poděkování

Rád by som poďakoval svojmu vedúcemu prof. Ing. Adamovi Heroutovi, Ph.D. za odborné vedenie tejto bakalárskej práce, ktoré pozostávalo z uvedenia do problematiky, objasnenia potrebných postupov a nasmerovania činnosti ako aj z konštruktívnej kritiky, čo v konečnom dôsledku viedlo k úspešnému vypracovaniu tejto práce.

# Obsah

<b>1</b>	<b>Úvod</b>	<b>2</b>
<b>2</b>	<b>Počítačové videnie a strojové učenie v doprave</b>	<b>3</b>
2.1	Praktické využitie počítačového videnia na analýzu dopravy . . . . .	3
2.2	Problémy vznikajúce pri analýze dopravy za pomoci počítačového videnia .	4
2.3	Rozpoznávanie štátnych poznávacích značiek . . . . .	5
2.4	Vizuálne vlastnosti dátovej sady vhodnej na strojové učenie . . . . .	9
2.5	Existujúce riešenia a ich vlastnosti . . . . .	9
<b>3</b>	<b>Pravidlá tvorby poznávacích značiek jednotlivých krajín obsiahnutých v riešení</b>	<b>11</b>
3.1	Pravidlá tvorby ŠPZ Českej republiky . . . . .	11
3.2	Pravidlá tvorby ŠPZ Slovenskej republiky . . . . .	12
3.3	Pravidlá tvorby ŠPZ Poľska . . . . .	13
3.4	Pravidlá tvorby ŠPZ Maďarska . . . . .	17
<b>4</b>	<b>Návrh a implementácia generátora syntetickej dátovej sady</b>	<b>20</b>
4.1	Získavanie dát potrebných v procese vytvárania generátora syntetickej dátovej sady . . . . .	20
4.2	Popis a štruktúra implementácie syntetickej dátovej sady krok za krokom .	22
4.3	Použitie skriptov na generovanie dát pre analýzu dopravy . . . . .	35
4.4	Odhaľovanie chýb a vylepšovanie vlastností dátovej sady . . . . .	37
<b>5</b>	<b>Záver</b>	<b>40</b>
	<b>Literatura</b>	<b>41</b>
<b>A</b>	<b>Obsah CD</b>	<b>43</b>

# Kapitola 1

## Úvod

Táto bakalárska práca rieši problematiku a využitie počítačového videnia a strojového učenia v doprave, konkrétne návrh, implementáciu a použitie generátora dátovej sady pozostávajúcej z fotografií automobilov so synteticky vytvorenými a generovanými štátnymi poznávacími značkami (ďalej ŠPZ) Českej republiky, Slovenska, Poľska a Maďarska určenej na strojové učenie. Súčasťou práce je aj rozbor používaných techník a existujúcich prístupov k problematike počítačového videnia a strojového učenia v doprave. Samostatnú kapitolu podstatnú pre vytvorenie zmieňovanej syntetickej dátovej sady tvoria pravidlá tvorby ŠPZ krajín, ktoré sú predmetom tejto práce. Samotná aplikácia pozostáva z troch modulov, ktoré reprezentujú jednotlivé časti generátora. Práca opisuje celý proces tvorby generátora od návrhu jednotlivých modulov a spôsobu získavania vstupných dát cez popis implementácie a použitia aplikácie až po špecifikáciu výstupných dát a dosiahnutých výsledkov.

Cielom práce je vytvoriť dostatočne veľkú a čo najreálnejšie vyzerajúcu dátovú sadu, ktorá má byť použitá na zdokonalenie rozpoznávacích schopností systému určeného na reidentifikáciu poznávacích značiek automobilov z fotografií zachytených v nevyhovujúcich podmienkach, čo sa odráža na kvalite fotografií a spôsobuje zlú čitateľnosť značiek.

## Kapitola 2

# Počítačové videnie a strojové učenie v doprave

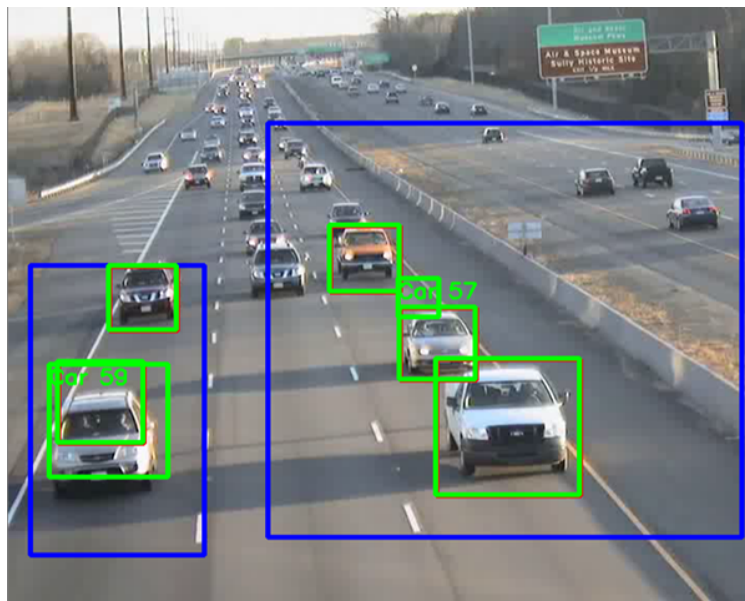
V dnešnej dobe je počítačové videnie používané v mnohých situáciach, kedy je potrebné z určitých obrazových dát získať konkrétnu informáciu, ktorá reprezentuje či už samotný výsledok alebo medzikrok určený na ďalšie spracovanie. Táto činnosť mnohokrát zahŕňa komplexnú a podrobnú analýzu a reprezentáciu získaných informácií, ktoré sú vo väčšine prípadov nedokonalé, skreslené alebo nedostačujúce.

### 2.1 Praktické využitie počítačového videnia na analýzu dopravy

Vhodným príkladom použitia počítačového videnia v praxi sú rôzne druhy problémov týkajúce sa dopravy. Konkrétne môže ísť o použitie zmieňovanej technológie za účelom analýzy dopravnej situácie a s ňou súvisiace získavanie a hromadenie špecifických informácií použitých napríklad na štatistické účely, čo je možné vidieť na obrázku 2.1. Vďaka týmto štatistickým údajom je možné lepšie zvládať management dopravy, vhodne plánovať dianie v cestnej premávke, predvídať určité druhy situácií, ako sú dopravné špičky a podobne.

Zaznamenávanie vlastností automobilov, či už ide o farbu, veľkosť alebo model, ktoré prejdú za určitý čas monitorovaným úsekom pozemných komunikácií, pričom dopravná situácia je po celý čas zaznamenávaná kamerou patrí medzi úlohy, ktoré nachádzajú svoje využitie práve v štatistike. Z takto získaných záberov je možné pomocou počítačového videnia okrem iného zistiť aj rýchlosť vozidiel, trajektóriu vozidiel, ale aj to, či vozidlo disponuje diaľničnou značkou. Všetky tieto informácie je možné získať prostredníctvom systémov určených na sledovanie dopravy, ktoré sú obvykle súčasťou statických alebo dynamických pozorovacích staníc využívaných napríklad dopravnou políciou. Systémy s podobnými vlastnosťami sa používajú aj v prípade autonómnych navigácií, ktoré sa stávajú bežnou súčasťou výbavy dnešných automobilov.

Počítačové videnie má v doprave široké spektrum využitia, vďaka čomu prispieva ku skvalitneniu dopravy a v konečnom dôsledku zvýšeniu bezpečnosti v doprave, o čom sa zmienili vo svojej publikácii aj Huang a Russell [6]. Systém určený na sledovanie vozidiel môže produkovať dáta špecifické pre každé vozidlo (odstup vozidiel, rýchlosť, akcelerácia), ktoré môžu viesť k presnejšiemu modelovaniu premávky a lepšiemu pochopeniu správania vodičov. Tento systém môže zaznamenať určité znaky konkrétneho vozidla, na základe ktorých identifikuje to isté vozidlo v rôznych pozorovacích stanicach.



Obrázek 2.1: Ukážka zaznamenávania a analýzy dopravnej situácie pomocou systému určeného na sledovanie vozidiel a dopravy, konkrétne ide o počítanie vozidiel v dopravnej špičke. Prevzaté z [www.jsexperts.com/img/car\\_counting.png](http://www.jsexperts.com/img/car_counting.png).

## 2.2 Problémy vznikajúce pri analýze dopravy za pomoci počítačového videnia

V nasledujúcej časti sú podrobnejšie popísané problémy vznikajúce pri analýze dopravy pomocou počítačového videnia.

Jednou z najväčších možných komplikácií vznikajúcich pri používaní počítačového videnia je nízka kvalita zaznamenaného obrazu. Nedostatočná kvalita obrazu je spôsobená širokým spektrom nežiadúcich podmienok, čo spôsobuje obmedzené možnosti analýzy obrazových dát. Rovnako obmedzujúca môže byť kvalita zariadenia zaznamenávajúceho obraz a svetelné podmienky, čo v konečnom dôsledku môže spôsobiť vyššie uvedený problém.

V publikácií od Huanga a Russella [6] sa píše aj o kvalite systémov a zariadení určených na sledovanie dopravy. Pre dosiahnutie čo najefektívnejších výsledkov použitím nástrojov na sledovanie dopravy, by mali tieto systémy spracúvajúce obraz a video spĺňať nasledujúce kritériá:

- Automatická segmentácia každého vozidla od pozadia a od ostatných vozidiel tak, že sú rozpoznané všetky vozidlá
- Správne rozpoznanie všetkých typov vozidiel – motocykle, osobné automobily, autobusy, kamióny a pod.
- Fungovanie v rôznych dopravných situáciách – svetelné križovatky, dopravné zápchy, premenlivá rýchlosť vozidiel, jazda v rôznych jazdých pruhoch
- Fungovanie v rôznych svetelných podmienkach a počasí (slnečno, zamračené, súmrak, noc, dážď)
- Fungovanie v reálnom čase



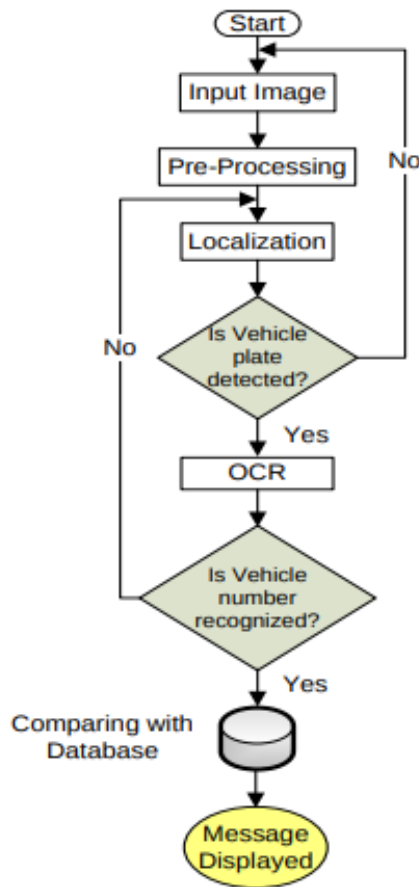


Obrázek 2.2: České ŠPZ pred a po procese reidentifikácie systémom pracujúcim s neurónovými sieťami. Prevzaté z [14].

V problematike dopravy a počítačového videnia sú pre nás dôležité určité rozlišovacie znaky, na základe ktorých sa rozhodujeme, kategorizujeme zozbierané dáta a premieňame ich na štatistické údaje. Môžu to byť – v závislosti na riešenom probléme – napríklad farba alebo tvar karosérie, spomínané diaľničné známky, značky identifikujúce model automobilu alebo štátne poznávacie značky.

### 2.3 Rozpoznávanie štátnych poznávacích značiek

Jednoznačným, medzinárodne uznávaným rozlišovacím znakom vozidiel vo väčšine krajín sú štátne poznávacie značky, ktoré sú vďaka svojej unikátnosti nezameniteľné a ich identifikácia je pri analýze dopravy často kľúčová. Presné zaznamenanie a rozpoznanie poznávacích značiek kamerou je preto veľmi podstatné, avšak do istej miery problematické, predovšetkým pokiaľ sa snažíme zachytiť vozidlo v pohybe alebo iných pre zaznamenávanie obrazu nepriaznivých podmienkach. Kvalita týchto záberov hrá v mnohých prípadoch zásadnú rolu pri riešení problémov vznikajúcich v doprave (ako sú napríklad dopravné priestupky). Nízka kvalita častokrát znemožňuje presne rozpoznať predpis poznávacej značky a týmto spôsobom jednoznačne identifikovať vozidlo. Riešením tohoto problému je spojenie technológie počítačového videnia so strojovým učením, čo vedie k schopnosti reidentifikovať predpisy registračných čísel zachytených v horšej kvalite a previesť ich do čitateľnej podoby, viď obrázok 2.2.

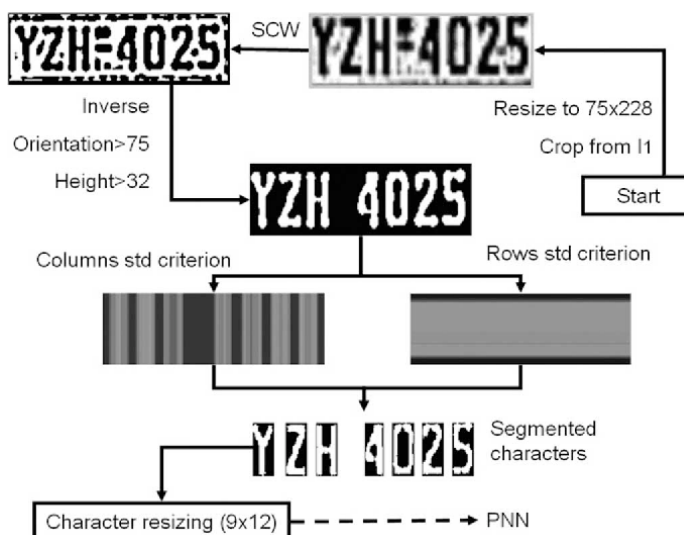


Obrázek 2.3: Graf znázorňuje proces rozpoznávania ŠPZ pozostávajúci z detekcie resp. lokalizácie ŠPZ na fotografii a optického rozpoznania jednotlivých znakov ŠPZ. Prevzaté z [13].

### 2.3.1 Automatické rozpoznávanie štátnych poznávacích značiek

Automatické rozpoznávanie ŠPZ je v publikácií od Saqiba Rasheeda a kolektívu [13] definované ako technika spracovania obrazu používaná na identifikáciu vozidla a jeho majiteľa na základe ŠPZ. Proces rozpoznávania ŠPZ obvykle pozostáva z dvoch častí, konkrétne detekcie resp. lokalizácie ŠPZ na obrázku a rozpoznania jednotlivých znakov ŠPZ, obrázok 2.3. V súčasnosti je automatické rozpoznávanie bežne používané radarovými systémami na sledovanie dopravy a tiež inteligentnými systémami (parkovacie a bezpečnostné systémy), ktoré zabezpečujú a kontrolujú dopravu. Kamerové zariadenia môžu byť stacionárne, prispôbené na snímanie ŠPZ, či už používaným rozlíšením alebo ich orientáciou v priestore. V týchto prípadoch dosahujú rozpoznávacie systémy takmer bezchybné výsledky. V praxi sa ale častejšie používajú pohyblivé systémy, ktorých výsledky nie sú natoľko presné.

Existuje niekoľko prístupov k automatickému rozpoznávaniu poznávacích značiek [17]. Veľké množstvo existujúcich algoritmov používa segmentáciu znakov od pozadia, čo má však citelný dopad na rýchlosť rozpoznávania. Ak je segmentácia nekorektná, rozpoznanie jednotlivých znakov ŠPZ je rovnako nesprávne. Problémy pri segmentácií môžu vzniknúť vplyvom rozmazania, šumu či neprímeraného osvetlenia.



Obrázek 2.4: Graf spracovania registračného čísla algoritmom segmentácie a optického rozpoznávania jednotlivých znakov. Prevzaté z [3].

Efektivita algoritmu na rozpoznávanie ŠPZ určeného pre inteligentné systémy v doprave používajúceho techniku segmentácie obrazu a analýzu spojených súčastí v spojení s neuronovou sieťou určenou na rozpoznávanie znakov od Anagnostopoulou a kolektívu [3] je overená na 1334 obrázkoch automobilov, ktoré boli zachytené z rôznych uhlov a vzdialeností, z ktorých bolo správne rozpoznávaných 1287 obrázkov, čo je 96.5%. Graf na obrázku 2.4 znázorňuje jednotlivé kroky algoritmu smerujúce k segmentácii znakov od pozadia.

Rozpoznávací metóda predstavená v publikácii od Jakuba Španhela a kolektívu [17] ponúka alternatívny prístup, nakoľko nepoužíva segmentáciu znakov a efektívnou nezaostáva za inými dostupnými rozpoznávacími metódami. Výsledky dosahované použitím tejto metódy potvrdzujú, že tento prístup je použiteľný aj pri zhoršených podmienkach rozpoznávania (nízka kvalita obrazu, rozmazanie, šum, neprimerané svetelné podmienky), v prípade ktorých by bolo využitie segmentácie znakov nepoužiteľné. Ukážka výsledkov dosahovaných touto metódou je zobrazená na obrázku 2.5. Tento prístup používa na rozpoznávanie ŠPZ konvolučné neuronové siete. V záujme predísť použitiu segmentácie znakov je sieťou spracovávaný celý RGB obrázok. Sieť je navrhnutá na rozpoznanie až 8 znakov nachádzajúcich sa na značke. V priebehu spracovania prechádza obrázok ôsmimi súvislými vrstvami siete, kde každá vrstva rozpoznáva jeden znak vždy na rovnakej pozícii. Použitím holistickej metódy na rozpoznávanie ŠPZ [17] je možné rozpoznať aj dáta, ktorých kvalita je na nízkej úrovni. Táto metóda je však vyvinutá pre rozpoznávanie ŠPZ, ktoré sú zarovnané voči vodorovnej a zvislej ose fotografie a preto tento prístup nie je priamo použiteľný na rozpoznávanie neupravených fotografií ŠPZ zachytených z rôznych pozorovacích uhlov.

V ďalšej publikácii od vyššie uvedeného autorského kolektívu [16] je ponúknuté riešenie tohoto problému. Riešenie spočíva vo vytvorení konvolučnej neuronovej siete, ktorej účelom je predpovedať súradnice rohov resp. krajné body ŠPZ v nezarovnanom obrázku, na základe ktorých je možné uskutočniť transformáciu fotografie a zarovnať ju tak, aby mohla byť spracovaná spomínanou holisticou metódou. Aj napriek oddelenému tréningu neuronových sietí pre holistické rozpoznávanie a geometrické zarovnávanie ŠPZ pracujú obe riešenia súčasne ako jeden celok, čím sa zabraňuje potenciálnym komplikáciám. Zmenšenie výskytu chýb a zvýšenie stability a rýchlosti rozpoznávania použitím neuronovej siete

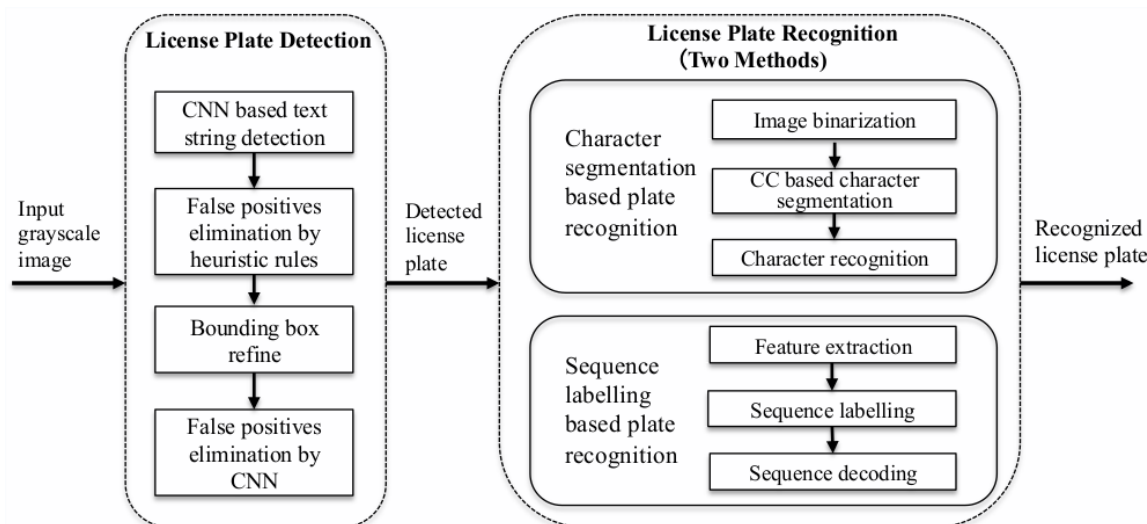


Obrázok 2.5: Ukážka holistickej metódy rozpoznávania ŠPZ. Poznávací znak je odfotený viackrát po sebe, aby sa zvýšila pravdepodobnosť zachytenia čo najlepšieho výsledku. Prevezaté z [17].

zarovnávajúcej fotografie ŠPZ je experimentálne dokázané použitím tejto neurónovej siete na rôzne druhy deformácií vyskytujúcich sa na fotografiách ŠPZ ako napríklad nesymetrické zarovnanie ŠPZ vzhľadom ku fotografiám, rozmazanie, šum a podobne. Vďaka vyššie uvedeným zisteniam sa stáva tento prístup zaujímavým pre inteligentné systémy používané v doprave.

Prístup, ktorý ponúka Li a Shen [10] je zameraný na rozpoznávanie ŠPZ v prirodzených podmienkach, ktoré nie sú nijak upravené pre dosahovanie lepších výsledkov. Ponúkané riešenie problematiky je zamerané na rozpoznávanie ŠPZ ako postupnosti znakov, čím sa vyhýba ich segmentáciám a s ňou spojeným komplikáciám. Táto koncepcia využíva na oddelenie postupnosti znakov od pozadia konvulčné neurónové siete a je zobrazená na obrázku 2.6, ktorý zobrazuje obidva prístupy k rozpoznávaniu znakov.

V ďalšej súvisiacej publikácii [5] je rozpoznávanie ŠPZ rozdelené do troch kategórií v závislosti na druhu aplikácií, ktoré ich používajú, konkrétne ide o rozpoznávanie ŠPZ pri overovaní povolenia prístupu, pri dokazovaní porušenia pravidiel cestnej premávky a pri použití dopravnými hliadkami. Rozdielne sú aj nastavenia parametrov používané na rozpoznávanie v jednotlivých prípadoch. Každá kategória je charakterizovaná parametrami ako je poloha kamery, veľkosť ŠPZ na obraze a svetelné podmienky, pričom rozsah hodnôt, ktoré môžu tieto premenné nadobúdať závisí na konkrétnej kategórii. Aplikácie s väčším rozsahom hodnôt premenných vyžadujú sofistikovanejšie spracovanie a sú výpočtetne náročnejšie ako ostatné, takže by bolo neefektívne používať metódy vyvíjané pre rôzne použitia dopravnými hliadkami aj pri overovaní povolenia prístupu a naopak, metódy používané pri overovaní prístupu nie je možné použiť na riešenie prípadov dopravných hliadok.



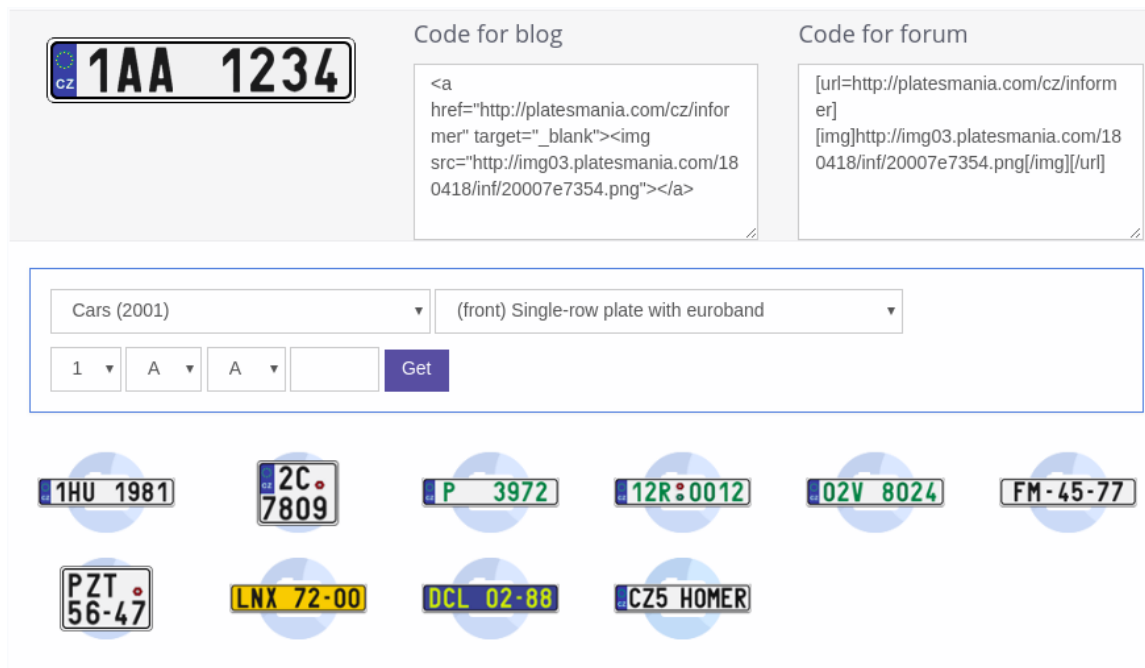
Obrázek 2.6: Koncept jedného z prístupov k detekcii a rozpoznávaniu ŠPZ. Prvá časť (vľavo) zobrazuje detekciu ŠPZ. Časť zobrazujúca rozpoznávanie ŠPZ (vpravo) znázorňuje dve nezávislé metódy: prvá (hore) pracuje na báze segmentácie jednotlivých znakov a druhá (dole) na báze rozpoznávania postupností znakov. Prevzaté z [10].

## 2.4 Vizuálne vlastnosti dátovej sady vhodnej na strojové učenie

V prípade učenia systému za účelom rozpoznávania obrazu vo všeobecnosti platí, že dátová sada určená na strojové učenie by mala spĺňať čo najviac vizuálnych detailov, ktorými disponujú reálne objekty a mala by pozostávať z čo najväčšieho množstva dát spĺňajúcich túto podmienku. Učenie na reálnych dátach prirodzene dosahuje najlepšie výsledky. Napriek tomu je preto v mnohých prípadoch nedostačujúce, nakoľko zozbierať dostatočne veľké množstvo reálnych dát je často nereálne. Za účelom získania dostatočne veľkého množstva dát určených na strojové učenie preto vznikajú generátory syntetických dátových sád, ktorých úlohou je dosiahnuť čo najpresnejší výsledok podobný realite. Dôležité je, aby synteticky vytvorené obrázky neobsahovali elementy, ktoré sa v skutočnosti na objektoch daného typu nevyskytujú alebo vyzerajú odlišne a naopak obsahovali všetky elementy špecifikujúce dané objekty. Riešenie týchto problémov pri implementácii samotného generátora syntetickej dátovej sady pre dopravnú analýzu je podrobnejšie popísané v podkapitole 4.2.3.

## 2.5 Existujúce riešenia a ich vlastnosti

Doména [www.platesmania.com](http://www.platesmania.com) [2] pokrýva veľkú časť poznávacích značiek nie len európskych, ale aj ázijských krajín a prostredníctvom jednoduchého grafického užívateľského rozhrania poskytuje možnosť vygenerovať si veľkú škálu typov registračných značiek od tých najbežnejších až po tie najzriedkavejšie, viď obrázok 2.7, pričom originálny font špecifický pre každú krajinu ostáva zachovaný. Ponúka možnosť vygenerovať si poznávacie značky s vlastným popisom, rovnako je možné si vygenerovať ŠPZ, ktoré sa už nepoužívajú, historické ŠPZ, dočasné ŠPZ, ŠPZ obsahujúce dva rady znakov určené pre motocykle, ŠPZ vyskytujúce sa na diplomatických vozidlách alebo špeciálne ŠPZ určené na komerčné účely.



Obrázek 2.7: Grafické rozhranie domény [www.platesmania.com](http://www.platesmania.com) [2] určenej na generovanie vlastnej syntetickej poznávacej značky rôznych typov.

Na druhej strane však takto vygenerované registračné značky pôsobia synteticky, mnoho z nich sa nevyskytuje v cestnej premávke a ich účelom nie je nahrádzať reálne ŠPZ. Generovanie veľkého množstva takýchto ŠPZ by bolo nepraktické, nakoľko užívateľské rozhranie neponúka možnosť generovať viacero značiek naraz.

Na podobnom princípe funguje generátor syntetických poznávacích značiek<sup>1</sup>, pomocou ktorého je možné vygenerovať si vlastnú ŠPZ patriacu do niektorého zo Spojených štátov amerických. Tieto poznávacie značky majú špecifický vzhľad a sú odlišné od značiek európskych štátov.

Spoločnosť Microsoft dokonca ponúka vlastnú aplikáciu s názvom License Plate Generator<sup>2</sup> určenú pre operačné systémy Windows 10 a Windows 10 Mobile. Princíp a využitie ostáva rovnaké, ako v predchádzajúcich prípadoch. Aplikácia však dokáže generovať len značky patriace štátu Ontário v Kanade.

Generátor syntetických poznávacích značiek umožňujúci okrem výberu vlastného znenia značky aj zmenu farby pozadia a písma<sup>3</sup> dokáže generovať značky obsahujúce aj rôzne symboly, ktoré pravidlá tvorby ŠPZ nedovoľujú. V databáze má pomerne veľké množstvo pozadí ŠPZ, medzi ktorými sa nachádzajú aj reálne nepoužívané tabule. Možnosť generovania je obmedzená na jeden font a väčšina pozadí je dostupná len po zaplatení určitej sumy.

Všetky vyššie uvedené alternatívne riešenia sú len čiastočné a ich použitie pri riešení problematiky rozpoznávania ŠPZ by bolo neefektívne, nakoľko nie sú určené na vytváranie dátových sád pre strojové učenie.

<sup>1</sup><https://www.acme.com/licensemaker/>

<sup>2</sup><https://www.microsoft.com/cs-cz/store/p/license-plate-generator/9p7cg94h8zkh>

<sup>3</sup><https://www.customeuropeanplates.com/>

## Kapitola 3

# Pravidlá tvorby poznávacích značiek jednotlivých krajín obsiahnutých v riešení

Nasledujúca kapitola obsahuje podrobný popis pravidiel, ktorými sa riadi tvorba a pridelovanie ŠPZ v Českej republike, Slovenskej republike, Poľsku a Maďarsku. Popis vychádza zo súčasného aktuálne používaného vzhladu štátnych poznávacích značiek a pravidiel, ktorými sa riadi tvorba ŠPZ v jednotlivých krajinách.

### 3.1 Pravidlá tvorby ŠPZ Českej republiky

V Českej republike [11, 15] obsahujú ŠPZ vydané po máji 2004 modrý prúžok s dvanástimi žltými hviezdami a skratkou štátu, viď obrázok 3.1. Medzi označením kraja a poradovým číslom je miesto vyhradené pre známky technickej (červená) a emisnej (zelená) kontroly. Tieto známky sa nachádzajú len na zadných značkách. Od 1.1.2015 sa používa len známka pre označenie technickej kontroly, táto práca však berie do úvahy aj staršiu konvenciu a syntetická ŠPZ vytvorená na základe tejto práce môže obsahovať v prípade potreby aj známku emisnej kontroly.

Znenie ŠPZ pozostáva zo siedmich znakov. Druhý znak označuje kraj, viď obrázok 3.2 a tabuľku 3.1. Prvý znak a posledné štyri znaky sú číslice určujúce poradie značky v danom kraji. Na tretej pozícii (znak za označením kraja) sa môže vyskytovať alfanumerický znak (z písmen len A, B, C, D, E, I, J, K, L, M, N, P, R, S, T, U, V, X, Y, Z). To však len v prípade, že pre daný kraj nestačia na označenie poradia číslice. Od januára 2016 je



Obrázok 3.1: Poznávacia značka automobilov Českej republiky z Jihomoravského kraja v podobe, akej sa vyskytuje v cestnej premávke dnes (od mája 2004).



Obrázek 3.2: Na obrázku je mapa Českej republiky s označeniami jednotlivých krajov. Prevzaté z <https://upload.wikimedia.org/wikipedia/commons/f/fa/CZ-cleneni.svg>.



Obrázek 3.3: Na obrázku je znázornená štátna poznávacia značka Slovenskej republiky z okresu Trenčín v podobe, akej sa vyskytuje v cestnej premávke dnes (od júna 2006).

možné zažiadať o poznávaciu značku s vlastným textom. Poznávacie značky automobilov s vlastným textom môžu pozostávať z 8 alfanumerických znakov obsahujúcich aspoň jednu číslicu. Značky nemôžu obsahovať vulgárne a hanlivé výrazy, názvy úradov a písmená G, O, Q a W, ktoré sa nemôžu vyskytovať ani na bežnej poznávacej značke, kvôli ich podobnosti so znakmi C, 6, 0, V.

### 3.2 Pravidlá tvorby ŠPZ Slovenskej republiky

Slovenské ŠPZ majú od júna 2006 nový typ tabuliek so symbolom EÚ, skratkou štátu a štátnym znakom Slovenska, ktorý nahradil dovtedajšiu pomlčku. Štandardné evidenčné čísla pozostávajú z piatich znakov, konkrétne z dvoch písmen nasledovaných tromi číslicami a znova dvomi písmenami (napr. BA 123AB). Prvá dvojica označuje skratku okresu, ďalšie trojčíslenie je plynulo rastúcim číslom v intervale 001 až 999 (kombinácia 000 nie je využitá). Zostávajúca dvojica písmen vzrastá podľa abecedy v intervale od A po Z tak, že po vy-



Tabulka 3.1: Tabulka skratiek jednotlivých krajov Českej republiky vyskytujúcich sa na štátnych poznávacích značkách na druhom mieste v poradí.

Označenie	Kraj
A	Hlavné mesto Praha
S	Stredočeský
U	Ústecký
L	Liberecký
K	Karlovarský
H	Královéhradecký
E	Pardubický
P	Plzeňský
C	Jihočeský
J	Vysočina
B	Jihomoravský
M	Olomoucký
T	Moravskoslezský
Z	Zlínský

Tabulka 3.2: Tabulka ukazuje príklad rastúcej postupnosti slovenských ŠPZ.

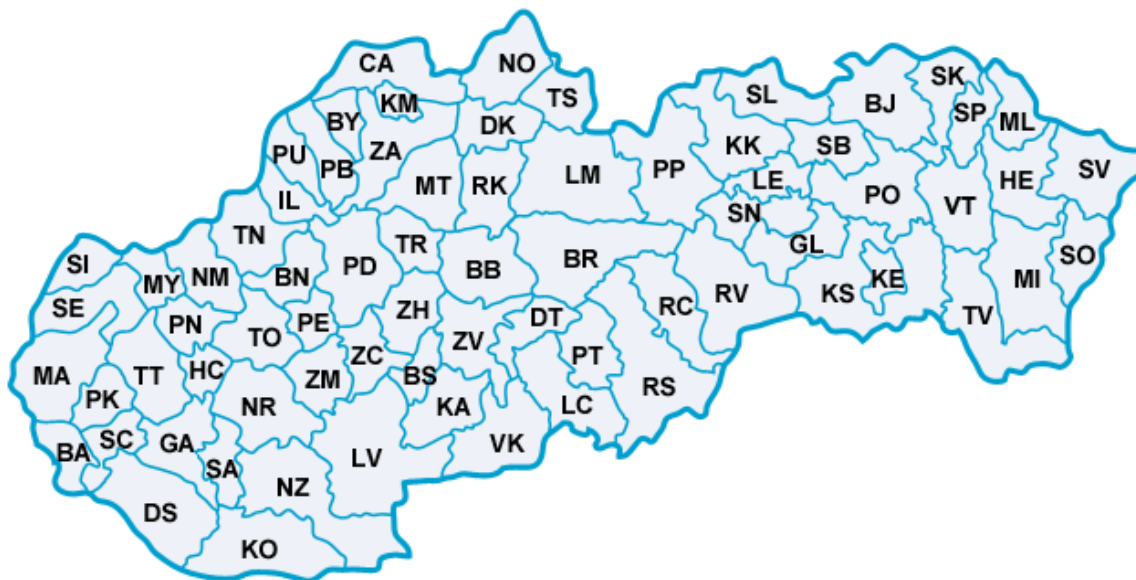
KE 999AA	KE 001AB	KE 002AB
PP 999EZ	PP 001FA	PP 002FA
BA 998ZZ	BA 999ZZ	BL 001AA

čerpaní intervalu trojčíslia vzrastie v intervale posledné písmeno a ako posledné vzrastie po vyčepaní intervalu predposledné písmeno. Príklad takejto postupnosti je v tabuľke 3.2. Táto kombinácia určuje poradie značky v danom okrese, viď obrázok 3.3.

Na rozdiel od štandardného evidenčného čísla sa pri evidenčných číslach s vlastným textom za označením okresu uvádzajú písmená na prvom až piatom mieste (napr. BA AAAAA), písmená na prvom až štvrtom mieste a číslica od jeden po deväť na piatom mieste (napr. BA AAAA1) alebo písmená na prvom až treťom mieste a kombinácia čísel od jeden po deväť na štvrtom a piatom mieste (napr. BA AAA11). Skratky označení okresov sú zobrazené na obrázku 3.4.

### 3.3 Pravidlá tvorby ŠPZ Poľska

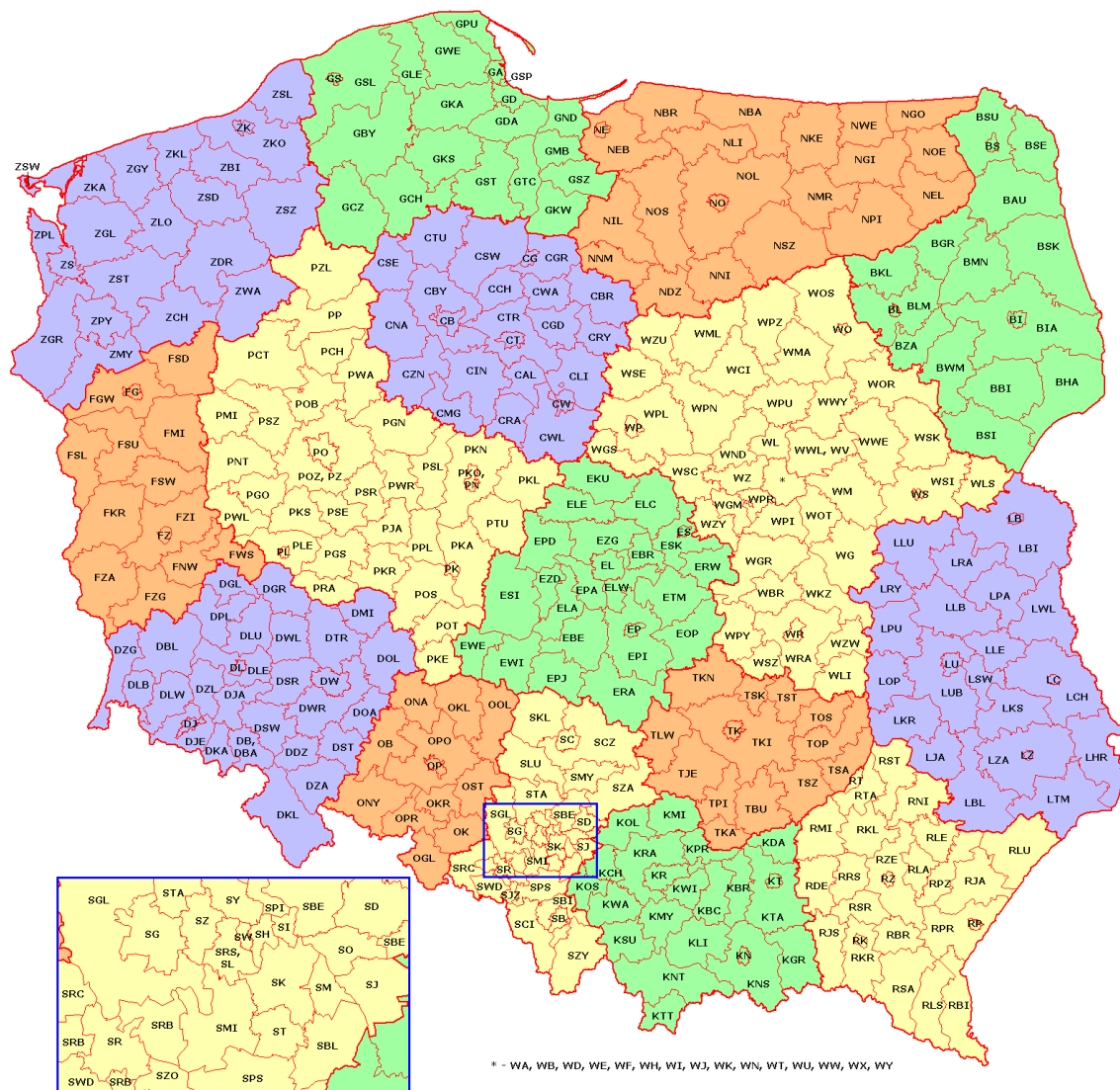
Poľské registračné značky [7, 12] existujú v dnešnej podobe od mája 2006, kedy poľskú vlajku nahradil znak Európskej únie 3.5. Prvé písmeno na značke označuje vojvodstvo resp. provinciu, ďalšie jedno alebo dve písmená okres, obrázok 3.6. V prípade zemských okresov ide o dve písmená, v prípade mestských okresov a štvrtí o písmeno jedno. Nasleduje medzera, v ktorej sa nachádza verifikačná hologramová nálepka a za ňou identifikácia konkrétneho vozidla, ktorá pozostáva zo štyroch alebo piatich znakov v závislosti na type okresu. Identifikácia však nemôže obsahovať písmená B, D, I, O, Z, ktoré sú príliš podobné číslam 8, 0, 1, 0, 2. Celkovo teda značka obsahuje sedem alebo osem znakov a jej rozmery sú 520 x 114 milimetrov, čo je na výšku o 4 milimetre viac ako v predchádzajúcich prípadoch.



Obrázek 3.4: Na obrázku je znázornená mapa Slovenskej republiky so skratkami jednotlivých okresov vyskytujúcich sa na slovenských ŠPZ. Prevzaté z [http://skauting.sk/img/sk\\_mapa.png](http://skauting.sk/img/sk_mapa.png).



Obrázek 3.5: Na obrázku je znázornená štátna poznávacia značka Poľska z Veľkopoľského vojvodstva v podobe, akej sa vyskytuje v cestnej premávke v dnešnej dobe (od mája 2006).



Obrázek 3.6: Na obrázku je mapa Polska so všetkými skratkami jednotlivých vojvodstiev a mestských aj zemských okresov. Prevzaté z <https://upload.wikimedia.org/wikipedia>.

V prípade ŠPZ patriacej do mestského okresu môžu nastať nasledovné identifikačné kombinácie:

- 2 písmená + 5 číslic, napr. AB 12345
- 2 písmená + 4 číslica + 1 písmeno napr. AB 1234C
- 2 písmená + 3 číslice + 2 písmená, napr. AB 123CD
- 2 písmená + 1 číslica + 1 písmeno + 3 číslice napr. AB 1C234
- 2 písmená + 1 číslica + 2 písmená + 2 číslice napr. AB 1CD23

V prípade ŠPZ patriacej do zemského okresu môžu nastať nasledovné identifikačné kombinácie:

- 3 písmená + 1 písmeno + 3 čísla, napr. ABC D123
- 3 písmená + 2 čísla + 2 písmená, napr. ABC 12DE
- 3 písmená + 1 číslo + 1 písmeno + 2 čísla, napr. ABC 1D23
- 3 písmená + 2 čísla + 1 písmeno + 1 číslo, napr. ABC 12D3
- 3 písmená + 1 číslo + 2 písmená + 1 číslo, napr. ABC 1DE2
- 3 písmená + 2 písmená + 2 čísla, napr. ABC DE12
- 3 písmená + 1 písmeno + 2 čísla + 1 písmeno, napr. ABC D12E
- 3 písmená + 1 písmeno + 1 číslo + 2 písmená, napr. ABC D1EF
- 3 písmená + 5 čísel, napr. ABC 12345
- 3 písmená + 4 čísla + 1 písmeno, napr. ABC 1234D
- 3 písmená + 3 čísla + 2 písmená, napr. ABC 123DE

Poľské poznávacie značky teda nemusia vždy obsahovať rovnaký počet znakov. Ich počet závisí na príslušnosti vozidla do zemského či mestského okresu.

- 2 písmená (vojvodstvo + mestský okres) + 5 znakov identifikácie (5 rôznych kombinácií písmen a číslic identifikácie)
- 3 písmená (vojvodstvo + zemský okres) + 4 znaky identifikácie (8 rôznych kombinácií písmen a číslic identifikácie)
- 3 písmená (vojvodstvo + zemský okres) + 5 znakov identifikácie (3 rôzne kombinácie písmen a číslic identifikácie)

Individuálne značky musia obsahovať označenie vojvodstva, tabuľka 3.3 a obrázok 3.7 a jednu nasledujúcu číslicu. Potom nasleduje priestor pre tri až päť ďalších písmen, z ktorým môžu byť maximálne posledné dve nahradené číslicami.

Tabulka 3.3: Označenia poľských vojvodstiev, ktoré sú prvým znakom na poznávacej značke Poľska.

Označenie	Vojvodstvo
B	Podleské
C	Kujavsko-pomoranské
D	Dolnosliezske
E	Lodžské
F	Lubuské
G	Pomoranské
K	Malopoľské
L	Lubelské
N	Varmsko-mazurské
O	Opolské
P	Veľkopoľské
R	Podkarpatské
S	Sliezske
T	Svätokrížske
W	Mazovské
Z	Západopomoranské

Tabulka 3.4: Tabuľka ukazuje príklad rastúcej postupnosti maďarských ŠPZ.

AAA-999	AAB-000	AAB-001
AAZ-999	ABA-000	ABA-001
AZZ-999	BAA-000	BAA-001

### 3.4 Pravidlá tvorby ŠPZ Maďarska

Pravidlá tvorby maďarských ŠPZ [9] tak, ako ich poznáme dnes boli zavedené v roku 1990. Na rozdiel od slovenských, českých alebo poľských ŠPZ nie je z maďarských značiek možné rozoznať miesto registrácie vozidla a pri pridelení značky sa vychádza len z času registrácie a typu vozidla. Značky obsahujú šesť znakov, kde prvé tri sú písmená, za ktorými nasleduje pomlčka a posledné tri sú číslice. Písmená označujú sériu a číslice poradie značky v danej sérii 3.4. Postupnosť sérií pri pridelení poznávacích značiek je znázornená tabuľkou 3.5.

Súčasný systém zahŕňa niekoľko druhov poznávacích značiek. V prvom rade sú to klasické ŠPZ, ktoré sa vyskytujú v dvoch variantách. Jedna varianta obsahuje v ľavej časti značky nad označením národnosti maďarskú vlajku, obrázok 3.8, druhá varianta zase modrý prúžok s dvanástimi hviezdami označujúci príslušnosť do Európskej únie 3.9. Značky taxislužby, nákladných automobilov, motoriek, elektromobilov, automobilov jazdiacich na plyn a pomalých vozidiel sa od klasických automobilov odlišujú farebným prevedením, či dokonca vyhradeným intervalom trojpísmenových skratiek série, tabuľka 3.6.

Vlastné poznávacie značky majú tak, ako v predchádzajúcich prípadoch vlastné pravidlá tvorby. Prvé tri znaky takýchto poznávacích značiek musia byť písmená a posledný znak musí byť číslica, takže do úvahy prichádzajú kombinácie písmen a číslíc AAA-000, AAAA-00, AAAAA-0.



Obrázek 3.7: Na obrázku je mapa Polska s označeniami jednotlivých vojvodstiev. Prevzaté z <https://upload.wikimedia.org/wikipedia>.



Obrázek 3.8: Ukážka maďarskej poznávacej značky staršieho formátu, vydávanej od roku 1990 do roku 2004. Tento typ je stále platný.



Obrázek 3.9: Novšia varianta maďarskej poznávacej značky symbolizujúca príslušnosť do Európskej únie vydávaná od mája 2004.

Tabulka 3.5: Časová následnosť sérií poznávacích značiek Maďarska od roku 1990.

Počiatkové písmeno série	Obdobie vydávania
A, B, C	1990 - 1992
D	1992 - 1993
E	1993 - 1995
F	1995 - 1996
G	1996 - 1998
H	1998 - 2002
I	2002 - 2003
J	2003 - 2005
K	2005 - 2007
L	2007 - 2012
M	2012 - 2015
N	2015 - 2016
P	2016 - súčasnosť

Tabulka 3.6: Tabuľka znázorňuje trojpísmenové intervaly sérií špeciálnych typov vozidiel a farebné prevedenie daných poznávacích značiek.

Séria	Vyhradený typ vozidiel	Farebné prevedenie
EAA-EZZ	vozidlá taxi	čierne písmo na žltom podklade
FAA-FZZ	nákladné automobily	čierne písmo na žltom podklade
UAA-UZZ	motocykle	štandardné
WAA-XZZ	prívesy	štandardné
YAA-YZZ	pomalé vozidlá	červené písmo na bielom pozadí

## Kapitola 4

# Návrh a implementácia generátora syntetickej dátovej sady

V nasledujúcej kapitole je podrobne opísaný prístup k návrhu a implementácií generátora syntetickej dátovej sady použitý v tejto práci. Práca sa zaoberá generovaním dátovej sady pozostávajúcej z fotografií automobilov so syntetickými ŠPZ a kompletne dokumentuje implementáciu generátora. Analogickým postupom je možné vytvoriť v prípade potreby aj generátory poznávacích značiek iných štátov.

V tejto práci je použitá aktuálna podoba ŠPZ všetkých vyššie uvedených krajín tak, ako sa vyskytuje v cestnej premávke dnes. Generovanie ŠPZ sa riadi pravidlami podrobnejšie popísanými v kapitole 3. Jednotlivé poznávacie značky symbolizujú príslušnosť do Európskej únie (12 hviezd v kruhu) a tiež štátnu príslušnosť, tabuľka 4.1.

### 4.1 Získavanie dát potrebných v procese vytvárania generátora syntetickej dátovej sady

Štátna poznávací značka každého z vyššie uvedených štátov pozostáva z pozadia, znakov definujúcich jej znenie a doplnkových elementov symbolizujúcich štátnu príslušnosť, príslušnosť do Európskej únie alebo známok technickej či emisnej kontroly.

#### 4.1.1 Získavanie obrazových dát potrebných na generovanie ŠPZ Českej republiky

V prípade českých ŠPZ som na vytvorenie znakovkej sady použil font, ktorý je možné získať na [www.feudal.cz](http://www.feudal.cz), obrázok 4.2. Každý znak som z textového editora typu Microsoft Word exportoval do formátu PNG v čo najväčšom rozlíšení. V ďalšom kroku som každý znak vektorizoval pomocou grafického editora InkScape a opäť exportoval vo vysokom rozlíšení do formátu PNG. Vektorizácia je proces konverzie rastrového obrázka na obrázok vektorový. Nakoniec som zmenil veľkosť obrázka na požadovanú veľkosť v závislosti na pomere reálnej veľkosti znakov v milimetroch a veľkosti znakov v pixeloch. V prípade českých ŠPZ som zvolil pomer reálnej veľkosti v milimetroch a veľkosti v pixeloch 1:3. Tento pomer je použitý aj pri vytváraní znakovkej sady poznávacích značiek ďalších štátov. Nepotlačené pozadie značky som vytvoril vymazaním znakov z obrázka reálnej ŠPZ a úpravami vlastností farieb tohto obrázka, obrázok 4.7. Okrem pozadia, znakov abecedy a číslíc sa na českých ŠPZ vyskytujú aj obrázky známok technickej a emisnej kontroly. Obrázky známok som



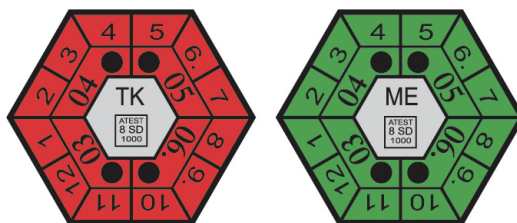
Tabulka 4.1: Tabulka skratiek jednotlivých štátov Európskej únie vyskytujúcich sa na štátnych poznávacích značkách.

štát	skratka
Rakúsko	A
Belgicko	B
Bulharsko	BG
Chorvátsko	HR
Cyprus	CY
Česká republika	CZ
Dánsko	DK
Estónsko	EST
Fínsko	FIN
Francúzsko	F
Nemecko	DE
Gibraltár	GBZ
Grécko	GR
Maďarsko	H
Írsko	IRL
Taliansko	I
Lotyšsko	LV
Luxemburgsko	L
Malta	M
Holandsko	NL
Portugalsko	P
Rumunsko	RO
Slovensko	SK
Slovinsko	SLO
Španielsko	E
Švédsko	S

získal vystrihnutím z obrázka 4.1 pomocou grafického editora, vektorizáciou, exportovaním do formátu PNG v čo najväčšom rozlíšení a zmenšením na požadovanú veľkosť.

#### 4.1.2 Získavanie obrazových dát potrebných na generovanie ŠPZ Slovenskej republiky

V prípade evidenčných čísel Slovenskej republiky som získal všetky používané obrázky znakov z fotografií reálnych ŠPZ. Pre čo najlepší výsledok je potrebné pri tomto prístupe fotografovať v čo najväčšom rozlíšení a z čo najväčšej blízkosti. Fotografie som previedol do čiernobielej reprezentácie nastavením hodnoty threshold v grafickom editore GIMP. Thresholding je najjednoduchšia metóda segmentácie obrazu, ktorá nahradí každý pixel obrázka čiernym, v prípade, ak je jeho intenzita menšia ako vhodne zvolená konštanta, alebo bielym, v prípade ak je jeho intenzita väčšia ako táto konštanta [4]. Čiernobiely reprezentáciu každého znaku (písmena a číslice) som exportoval do formátu PNG. Pri tomto spôsobe získavania obrazových dát je potrebné v grafickom editore odstrániť z fotografií nežiaduce elementy, predovšetkým nečistoty vyskytujúce sa na značkách. Každý znak prevedený do



Obrázek 4.1: Znamky technickej kontroly (vľavo) a emisnej kontroly (vpravo) vyskytujúce sa na zadných poznávacích značkách Českej republiky. Od 1.1.2015 sa známka merania emisií viac nepoužíva. Prevzaté z <http://autoservis-dobris.cz/priprava-a-zajisteni-stk>.

čiernobielej podoby som vektorizoval, uložil vo vysokom rozlíšení vo formáte PNG a zmenšil na požadovanú veľkosť pre čo navyššiu kvalitu, obrázok 4.3 podobne, ako pri znakovkej sade pre české ŠPZ. Na evidenčných číslach Slovenskej republiky sa nachádza aj slovenský znak, ktorý som vektorizoval, uložil vo vysokom rozlíšení a zmenšil na požadovanú veľkosť. Prázdnu značku som vytvoril analogicky, ako v prípade českých ŠPZ.

#### 4.1.3 Získavanie obrazových dát potrebných na generovanie ŠPZ Poľska

Štátne poznávacie značky Poľska sa líšia od slovenských a českých už na prvý pohľad, nakoľko sú znaky náchádzajúce sa na poľských značkách rôznej šírky. Takisto ich počet nie je vždy rovnaký, čo je potrebné zohľadniť pri implementácii. Znakovú sadu, prázdnu značku a hologramovú verifikačnú známku, ktorá je typická pre ŠPZ Poľska som získal z fotografií reálnych značiek analogicky ako v prípade poznávacích značiek Slovenska. V prípade znakov aplikáciou treshold filtra, vektorizáciou a zmenou na požadovanú veľkosť aj v prípadoch ostaných elementov vyskytujúcich sa na značkách. Generovanie poľských značiek si však vyžaduje o niečo väčšiu réžiu vzhľadom na zložitosť pravidiel, ktoré sa pri ich vytváraní používajú. Proces generovanie a implementácia generátora sú podrobnejšie popísané v kapitole 4.2.1.

#### 4.1.4 Získavanie obrazových dát potrebných na generovanie ŠPZ Maďarska

Všetky potrebné elementy tvoriace maďarské poznávacie značky som získal, podobne ako v predchádzajúcich prípadoch, z fotografií reálnych automobilov. Pozadia značiek boli vytvorené vymazaním jednotlivých elementov, ktoré sa na nich nachádzali, zvýraznením farieb a odstránením nečistôt, obrázok 4.4. Špeciálne značky nachádzajúce sa na automobiloch taxi služby, nákladných vozidlách, či automobiloch poháňaných elektrickým prúdom alebo plynom boli vytvorené prefarbením pozadia značky určeného na dosadenie znakov a tak isto zmenou farby pozadia jednotlivých znakov na príslušný odtieň. Všetky znaky som rovnako, ako pri znakoch ostatných štátov upravil treshold filtrom, vektorizoval, exportoval do PNG formátu v čo najväčšom rozlíšení a nakoniec zmenšil na požadovanú veľkosť.

## 4.2 Popis a štruktúra implementácie syntetickej dátovej sady krok za krokom

Nasledujúca podkapitola je venovaná popisu implementácie jednotlivých súčastí generátora syntetickej dátovej sady. Samotná implementácia pozostáva z troch skriptov, pričom výstup

# SPZ 2004 REGULAR

© COPYRIGHT: TYPE DESIGN, LAYOUT & TEXT BY DALIBOR FEUEREISL, PRAGUE 2006

CODEPAGE ISO 1250

WINDOWS | MACINTOSH

CHARACTER SET  
CAPS

ABCDEFGHIJKLMNOPQRSTUVWXYZ

LOWERCASE

ABCDEFGHIJKLMNOPQRSTUVWXYZ

PROPORTIONAL LINING NUMERALS

1234567890

ACCENTED & LANGUAGE CAPS & LOWERCASE

ÁÉÍÓÚÛÝŽŠČŘĎŤŇ

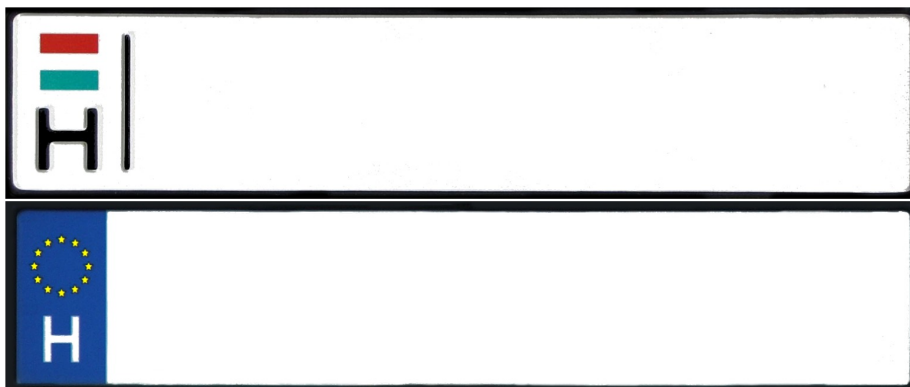
PUNCTUATION, SYMBOLS & MISCELLANEOUS

\$ € CZ FEU DAL 1 CZ . + - . ' © ° ´ ˘

Obrázek 4.2: Na obrázku je znázornený font použitý na generovanie českých poznávacích značiek dostupný na [www.feudal.cz/spz/html/autorske\\_pismo.htm](http://www.feudal.cz/spz/html/autorske_pismo.htm). Rovnaký font sa používa na tvorbu českých ŠPZ od mája 2004.



Obrázek 4.3: Na obrázku vľavo je fotografia jedného zo znakov nachádzajúcich sa na slovenských poznávacích značkách, ktorú je potrebné pre ďalšie použitie upraviť. Výsledný obrázok po grafických úpravách je zobrazený vpravo. Takto upravené znaky sa používajú na vytvorenie syntetických poznávacích značiek.



Obrázek 4.4: Pozadia ŠPZ Maďarska vytvorené grafickými úpravami fotografií. V hornej časti sa nachádza pozadie používané v rokoch 1990 až 2004 obsahujúce maďarskú vlajku. V dolnej časti je zobrazené pozadie používané v súčasnosti.

každého z nich ovplyvňujú vstupné argumenty príkazového riadku. Skripty sú implementované v jazyku Python s použitím knižníc určených na spracovanie obrazu, konkrétne PIL (Python Image Library) a OpenCV [1].

#### 4.2.1 Implementácia skriptu generujúceho syntetické ŠPZ

Prvý z trojice skriptov tvoriacej generátor syntetickej dátovej sady má za úlohu vygenerovať požadované množstvo syntetických poznávacích značiek vopred zvolenej krajiny tak, že obrázok pozadia značky sa prekryje znakmi a elementami, ktoré sa na značke majú nachádzať. V prípade spracovania obrázkov v prvom skripte používam knižnicu PIL (Python Image Library). Súradnice uloženia znakov a iných symbolov na pozadie značky som si vypočítal na základe pomeru 1:3 veľkosti reálnej značky v milimetroch a veľkosti syntetickej značky v pixeloch.

Zdrojový kód je logicky rozložený do funkcií. Generátory poznávacích značiek všetkých štátov zdieľajú jednu implementáciu, ktorá sa líši len v časti implementácie pravidiel tvorby poznávacích značiek jednotlivých štátov. V hlavnej funkcii s názvom `main()` nasleduje po volaní funkcie `argparse()`, ktorá má na starosti rozparovanie vstupných argumentov iterovanie v adresári `source` obsahujúcom obrázky znakov jedného zo štyroch štátov na základe hodnoty argumentu príkazového riadku, ktorý symbolizuje štátnu príslušnosť. Jednotlivé súbory reprezentujúce znaky sa vo funkcií `createImagesArray()` ukladajú do poľa `images` pomocou `images.append()`, kde sa každý súbor uloží ako usporiadaná dvojica `[img, file_name]` obsahujúca informácie o danom súbore potrebné na manipuláciu s obrázkom znaku a reťazec obsahujúci názov súboru, ktorý sa uloží do premennej `file_name` v každej iterácii adresárom `source` pomocou `os.walk(source)`. Nasleduje načítanie ďalších symbolov vyskytujúcich sa na ŠPZ v závislosti na hodnote argumentu príkazového riadku predstavujúceho štátnu príslušnosť.

V prípade generovania slovenských značiek je okrem znakov sady potrebné prázdne pozadie, obrázok 4.5 a štátny znak, obrázok 4.6, ktorý je hneď po načítaní pomocou funkcie z knižnice PIL `Image.open(path_to_image)` prilepený na pozadie značky na pozíciu určenú súradnicami použitím funkcie `background.paste(image, coordinates)`.

Po pridaní všetkých znakov vyskytujúcich sa na slovenských ŠPZ do poľa `images` je z poľa skratiek okresov Slovenska `sk_shortcuts` vo funkcií `addDistrictSK()` náhodne vy-



Obrázek 4.5: Pozadie ŠPZ Slovenskej republiky slúži ako základ pre ďalšie grafické elementy ako znaky písmen a číslíc a štátny znak, ktoré sú do pozadia v priebehu generovania postupne pridávané.



Obrázek 4.6: Štátny znak Slovenskej republiky je prvým grafickým prvkom, ktorý sa v procese generovania prilepí na súradnicami vopred určené miesto na pozadí ŠPZ. Obrázok znaku bol vektorizovaný, exportovaný do formátu PNG vo vysokom rozlíšení a zmenšený na požadovanú veľkosť, vďaka čomu si zachováva kvalitu.

braná niektorá z dvojpísmenových skratiek reprezentujúca určitý okres. Obrázky znakov sú na základe vybranej skratky postupne načítané použitím funkcie `Image.open()` a následne prilepené na pozadie na vopred vypočítané súradnice, ktoré sú argumentami funkcie `addDistrictSK()`. Po prilepení dvojpísmenovej skratky okresu nasleduje volanie funkcie `generate_char(offset, images, nation)`, kde `offset` je usporiadaná dvojica súradníc definujúca budúcu polohu aktuálne generovaného znaku na pozadí a `nation` je skratka národnosti. Táto funkcia je volaná pre každú dvojicu súradníc definujúcu polohu ostatných znakov na poznávacej značke. Pomocou funkcie `random.choice(images)` sa náhodne vyberie jeden súbor reprezentujúci znak zo znakovkej sady daného štátu. Ak je náhodne vybraný znak možné umiestniť na základe pravidiel implementovaných pomocou regulárneho výrazu na pozíciu určenú súradnicami, funkcia vráti názov súboru, ktorý reprezentuje tento znak. V opačnom prípade dochádza k rekurzívnemu volaniu funkcie `generate_char()`, ktoré sa ukončí až keď sa náhodne vyberie znak spĺňajúci podmienky definované regulárnym výrazom. Poznávacie značky Slovenska obsahujú celkovo 7 znakov. Ku každej synteticky vygenerovanej značke sa navyše viaže textový súbor obsahujúci refazec pozostávajúci z náhodne vygenerovanej skratky okresu a ostatných znakov, ktoré v konečnom dôsledku tvoria jej znenie, za ktorým nasleduje v textovom súbore medzera a skratka štátu.

V prípade generovania českých značiek sú okrem pozadia, obrázok 4.7 a špecifickej znakovkej sady potrebné aj známky technickej a emisnej kontroly, pre ktoré je na značke vyhradené miesto. Na generovanie jednotlivých znakov vyskytujúcich sa na českých značkách sa používa rovnaký postup ako v prípade slovenských značiek. Rozdiel je však v pravidlách tvorby 3.1. Pomocou funkcií z knižnice PIL, ktoré slúžia na manipuláciu s obrázkami, pomocou regulárnych výrazov, konkrétne funkcie `re.search(regex, character)`,



Obrázek 4.7: Pozadie ŠPZ České republiky slúži ako základ pre ďalšie grafické elementy ako znaky písmen a číslíc a známky technickej kontroly a merania emisií, ktoré sú na pozadie v priebehu generovania postupne pridávané.

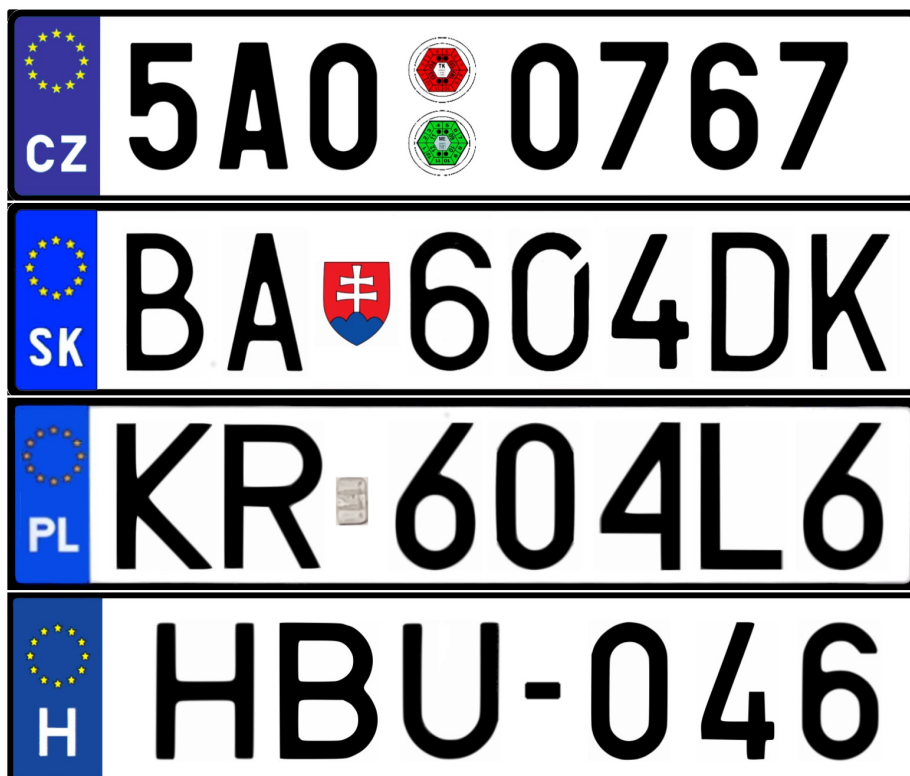


Obrázek 4.8: Pozadie ŠPZ Poľska slúži ako základ pre prilepenie hologramovej verifikačnej nálepky a znakov vygenerovaných na základe pravidiel.

kde `regex` je regularny vyraz, ktorému musí odpovedať vygenerovaný znak uložený v premennej `character` a pomocou rekurzívneho volania funkcie `generate_char()`, pre každú dvojicu súradníc definujúcich polohu jednotlivých znakov na značke je tak vytvorená syntetická ŠPZ obsahujúca 7 znakov, ku ktorej sa vytvorí textový súbor obsahujúci celé náhodne vygenerované identifikačné číslo a skratku štátu.

Implementácia generátora poľských poznávacích značiek si vyžaduje najväčšiu pozornosť vzhľadom na zložitosť pravidiel tvorby 3.3. Ako prvá sa na pozadie značky prilepí hologramová verifikačná nálepka. Podobne, ako pri poznávacích značkách Slovenska sú súčasťou poľských značiek skratky vojvodstiev a okresov, ktoré sú uložené do dvoch polí podľa počtu znakov, odkiaľ sú v procese generovania náhodne vyberané. Skratky môžu pozostávať z dvoch alebo troch znakov, za ktorými nasleduje identifikačný reťazec. Na základe rôznorodosti validných kombinácií znakov, ktoré môžu tvoriť identifikáciu značiek, obsahuje generátor celkovo 16 pravidiel, na základe ktorých sa značky generujú. V prvom rade sa na prázdnu značku, obrázok 4.8 prilepí podľa zvoleného pravidla skratka okresu vo funkcií `addDistrictPL(offset1, offset2, offset3, background, source)`, kde sa podľa počtu predaných offsetov (usporiadaných dvojíc súradníc  $(x,y)$  definujúcich polohu znaku) rozhodne o tom, koľko znakov bude obsahovať okres v danom prípade.

Pozicovanie znakov a hologramovej značky sa deje rovnakým spôsobom ako pri českých a slovenských značkách, teda pomocou vopred vypočítaných súradníc. Súradnice všetkých elementov, ktoré majú byť prítomné na značke závisia od vybraného pravidla, ktoré je v danej iterácii použité (rozdielne súradnice pre 7 a 8 miestne značky). Výber pravidla usporiadania znakov tvoriacich znenie značky sa deje náhodne použitím funkcie `random.randint(min, max)`, čím sa pre každú pozíciu znaku definuje, aký znak je pre danú pozíciu v danom pravidle validný. Nasleduje funkcia `generate_pl_char(character, images, offset, background)`, kde `character` je typ znaku (čísllica, písmeno alebo číslica okrem 0) vo forme reťazca, ktorý sa predáva ďalej do funkcie `return_char(character, images)` vracajúcej požadovaný typ znaku. Vybraný znak sa prilepí na obrázok pozadia pomocou funkcie z knižnice PIL.



Obrázek 4.9: Príklad synteticky vygenerovaných poznávacích značiek Česka, Slovenska, Poľska a Maďarska.

Maďarské poznávacie značky obsahujú celkovo 6 znakov a pravidlá ich tvorby sú podstatne jednoduchšie ako v predchádzajúcich prípadoch. Generovanie a pozicovanie znakov na pozadí sa deje rovnako, ako v predchádzajúcich prípadoch. Prvé tri miesta sú vyplnené náhodne vybranými znakmi písmen a zvyšné tri znakmi číslic pomocou funkcie `generate_char()`. Na pozadí značky sa nachádza ešte znak pomlčky.

Na generovaní poznávacích značiek staršieho dátá, značiek taxislužby, nákladných automobilov alebo automobilov poháňaných elektrickým prúdom dochádza zriedkavejšie (raz za 1500 vygenerovaných značiek). Použité sú pritom upravené znakové sady a pozadia značiek (rozdiel vo farbe pozadia). Rovnako, ako v predchádzajúcich prípadoch vzniká aj textový súbor s popisom značky, skratkou národnosti a v prípade špeciálnych značiek aj s označením TAXI, TRUCK alebo E-CAR.

Výsledok generovania poznávacej značky každého zo spomínaných štátov môžeme vidieť na obrázku 4.9.

#### 4.2.2 Implementácia skriptu umožňujúceho získavanie súradníc poznávacích značiek pomocou grafického rozhrania

V ďalšom kroku si zaobstaráme sadu fotografií automobilov, ktoré budú slúžiť ako základ pre synteticky vygenerované poznávacie značky z predchádzajúceho kroku. Program v každej iterácii fotografiami v adresári načíta jednu fotografiu pomocou funkcie z knižnice OpenCV [1] `cv2.imread(path_to_file)`, kde `path_to_file` je cesta k aktuálne spraco-



Obrázek 4.10: Na obrázku vidíme ukážku grafického rozhrania skriptu, ktorý slúži na naklikanie súradníc reprezentujúcich polohu poznávacej značky vo fotografií. Červené body na obrázku symbolizujú súradnice, ktoré boli vytvorené kliknutím ľavého tlačidla myši. Súradnice sa po ich potvrdení klávesou X uložia do textového súboru.

vávanej fotografií automobilu. Načítaná fotografia je zmenšená do štvrtinovej veľkosti kvôli zobrazeniu na obrazovke pomocou funkcie `cv2.resize()`.

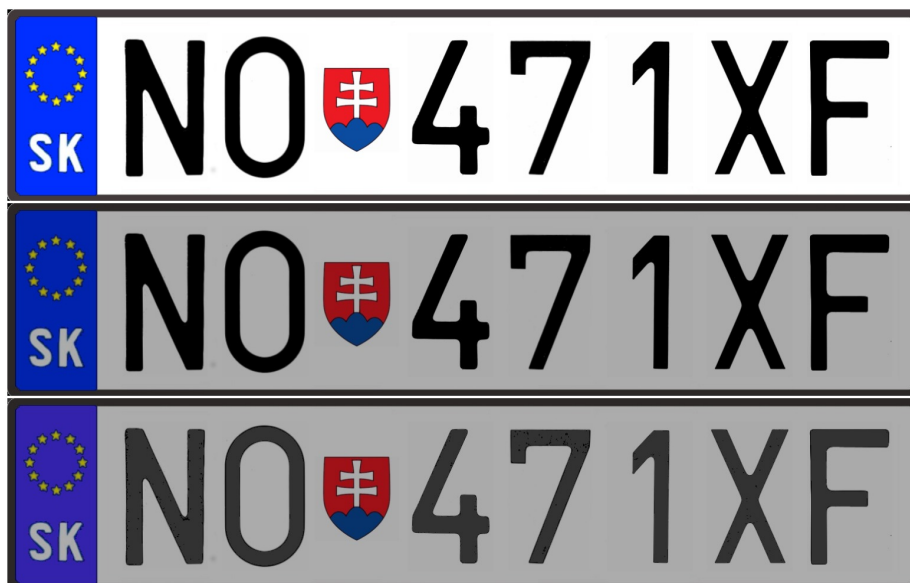
Nasleduje volanie funkcie `get_four_points(image)`, ktorá vracia pole súradníc, ktoré určujú polohu značky na fotografií. Fotografia sa zobrazí použitím funkcie `cv2.imshow()`. Pomocou ďalšej funkcie z knižnice OpenCV `cv2.setMouseCallback()` sa zavolá funkcia reagujúca na vstup zadávaný z klávesnice alebo myši. Pomocou jednoduchého ovládania klávesami a myšou je možné kliknutím ľavého tlačidla myši zvoliť súradnice, ktoré odpovedajú polohe značky na danej fotografií a následne ich uložiť do poľa `data`. Na mieste kliknutia sa na pozadie, ktoré predstavuje aktuálne spracovávaná fotografia vykreslí použitím funkcie `cv2.circle()` červená bodka, obrázok 4.10. Pohyb bodky po pozadí resp. zmenu súradníc a ich zápis do poľa `data` je možné ovládať klávesami W, A, S, D, Z a X. Ovládanie je podrobnejšie opísané v kapitole 4.3.2. Vo výsledku tak získame z každej fotografie štyri usporiadané dvojice reprezentujúce pozíciu značky resp. súradnice jej rohov na danej fotografií. Do textového súboru sú súradnice uložené pomocou funkcií na manipuláciu so súbormi (`open()`, `write()`, `close()`).

Týmto spôsobom je možné získať ku každej fotografií súradnice poznávacej značky, ktoré budú použité ďalším skriptom na vytvorenie transformačnej matice, na základe ktorej sa synteticky vygenerovaná poznávacia značka deformuje a bude pridaná do fotografie.

#### 4.2.3 Implementácia skriptu spájajúceho syntetické ŠPZ a fotografie do jedného celku

Výstupom prvého skriptu je sada synteticky vygenerovaných ŠPZ, výstupom druhého skriptu sú textové súbory obsahujúce súradnice ŠPZ automobilov na jednotlivých fotografiách. Tretí





Obrázek 4.11: Poznávacia značka zobrazená hore je výstupom skriptu generujúceho syntetické ŠPZ `rzgenerator.py`. Po zmene kontrastu a jas v skripte `supertool.py` je značka zobrazená v strede. Dole sa nachádza značka po úprave hodnoty čiernej a bielej farby.

skript má za úlohu spojiť tieto dáta a vytvoriť dátovú sadu fotografií automobilov so syntetickými ŠPZ určenú na strojové učenie.

Hlavná funkcia `main()` začína volaním funkcie `argparse()` slúžiacej na spracovanie argumentov príkazového riadku. Použitie vstupných argumentov je podrobnejšie rozpísané v podkapitole 4.3.3. Potom sa vo funkcií `createOutputDir()` vytvorí výstupný adresár pomocou funkcie `os.makedirs()`. V nasledujúcej funkcií `createTemplates()` sa vytvorí pole fotografií s názvom `templates`, v ktorom sa majú nachádzať fotografie automobilov a textové súbory so súradnicami poznávacích značiek získané pomocou predchádzajúceho skriptu. Cesta k adresáru obsahujúcemu fotografie automobilov môže byť zadaná ako voliteľný argument pri spustení skriptu. Po návrate do funkcie `main()` sa začína cyklus prechádzajúci synteticky vygenerované obrázky poznávacích značiek. V prípade, že sa počas iterovania v adresári so syntetickými ŠPZ narazí na súbor končiaci na `.txt` obsahujúci znenie značky a skratku štátu, je tento textový súbor preskočený. Prítomnosť textového súboru je zisťovaná na základe názvu súboru pomocou funkcie `file_name.endswith('.txt')`. V každej iterácii cyklu sa otvorí vo funkcií `getSourceImage()` jeden obrázok značky pomocou funkcie `cv2.imread()` a uloží sa do premennej `im_src`.

Pomocou funkcie `cv2.cvtColor()` sa mení farebný priestor obrázka synteticky vygenerovanej poznávacej značky (z RGB na HSV a naopak) a následne sa mení kontrast a jas obrázka pre autentickejší vzhľad, obrázok 4.11.

```
# Brightness and contrast changing
im_hsv = cv2.cvtColor(im_src, cv2.COLOR_BGR2HSV)

for x in range(0, len(im_hsv)):
    for y in range(0, len(im_hsv[0])):
        im_hsv[x,y][2] *= 0.675
```

```
im_src = cv2.cvtColor(im_hsv, cv2.COLOR_HSV2BGR)
```

V ďalšom kroku sa obrázok zmenší na základe vhodne zvolených koeficientov a metódy interpolácie, ktoré sú parametrami funkcie `cv2.resize()`, čím sa dosiahne vyhladenie hrán (antialiasing), ktoré by mohli na obrázku značky vzniknúť pri priestorových transformáciach. Vo funkcií `changeBlackAndWhiteValues()` sa pomocou funkcie `cv2.split()`, ktorá získa hodnoty jednotlivých farebných zložiek obrázka a funkcie `cv2.merge()`, ktorá zoskupí zmenené farebné zložky obrázka, zmení intenzita jednotlivých farebných zložiek RGB, obrázok 4.11. Dôvodom tejto zmeny je dosiahnutie autentickjšieho vzhľadu značiek po spojení s fotografiami automobilov. Ďalšie úpravy vzhľadu syntetickej poznávacej značky zahŕňajú aj aplikáciu Gaussovského šumu pomocou funkcie `add_gaussian_noise()`, ktorá vyzerá nasledovne:

```
# Function adds Gaussian noise
def add_gaussian_noise(image_in):
    noise_sigma = 20
    temp_image = np.float64(np.copy(image_in))

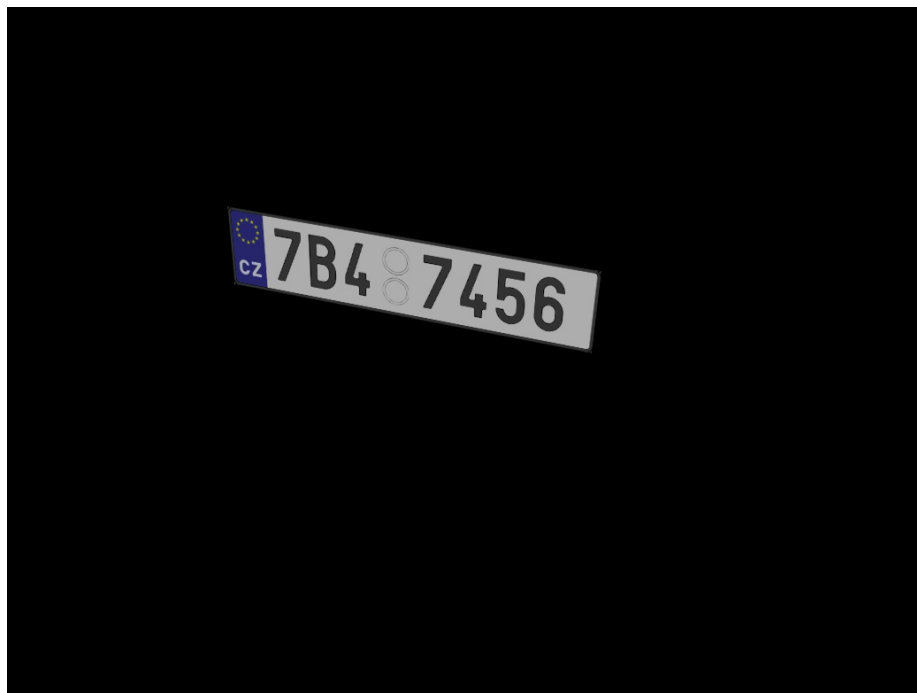
    h = temp_image.shape[0]
    w = temp_image.shape[1]
    noise = np.random.randn(h, w) * noise_sigma
    noisy_image = np.zeros(temp_image.shape, np.float64)

    noisy_image[:, :, 0] = temp_image[:, :, 0] + noise
    noisy_image[:, :, 1] = temp_image[:, :, 1] + noise
    noisy_image[:, :, 2] = temp_image[:, :, 2] + noise

    return noisy_image
```

Takto upravený obrázok syntetickej poznávacej značky je vo funkcií `main()` uložený do premennej s názvom `im_src`. Ďalej je potrebné zistiť súradnice rohov syntetickej značky, ktoré sa neskôr použijú spolu so súradnicami polohy reálnych značiek v jednotlivých fotografiách automobilov na vytvorenie transformačnej matice. Rozmer synteticky vygenerovaných poznávacích značiek je  $1560 \times 330$  px (podľa pomeru veľkosti reálnej ŠPZ v milimetroch a syntetickej ŠPZ v pixeloch 1:3). Uloženie usporiadaných dvojíc súradníc do poľa s názvom `pts_src` sa deje vo funkcií `getSourcePoints()`, kde sa najprv zistí rozmer obrázka pomocou funkcie `size = image.shape`, kde `size` je dvojprvkové pole, pričom na nultom indexe je uložená výška a na prvom indexe šírka obrázka, a potom sa toto pole použije pre získanie súradníc nasledovne:

```
# Getting source image points
pts_src = np.array(
    [
        [0,0],
        [size[1] - 1, 0],
        [size[1] - 1, size[0] - 1],
        [0, size[0] - 1 ]
    ], dtype=int
);
```



Obrázek 4.12: Deformovaná synteticky vytvorená poznávacia značka pripravená na prilepenie do fotografie.

Funkcia vracia pole usporiadaných dvojíc resp. pole súradníc `numpy.array` typu `integer`. Po návrate do funkcie `main` sa z fotografií uložených do poľa s názvom `templates` vyberie pomocou funkcie `random.choice(templates)` v každej iterácii jedna fotografia, ktorá je vo funkcií `getDestinationImage()` uložená do premennej `im_dst` a vrátená do rovnomennej premennej vo funkcií `main()`. Pred návratom do funkcie `main()` je ešte upravená veľkosť tejto fotografie pomocou funkcie `cv2.resize()`. Táto fotografia bude slúžiť ako pozadie pre syntetickú poznávaciu značku vybranú v jednej iterácii.

Súradnice polohy značky na fotografií sú uložené v textovom súbore vo formáte `[[y0, x0], [y1, x1], [y2, x2], [y3, x3]]`. Vo funkcií `getDestinationPoints()` sa daný textový súbor obsahujúci súradnice otvorí pomocou funkcie `open()` a pomocou funkcie `read()` sa súradnice uložia do poľa, ktoré táto funkcia po pretypovaní z horizontálneho do vertikálneho uloženia pomocou `numpy.vstack()` vracia. Týmto spôsobom sú získavané všetky súradnice potrebné pre spojenie zdrojového a cieľového obrázka do jedného celku.

Funkcia `createOutputImage()` dostáva v podobe vstupných argumentov všetky dáta potrebné pre spojenie syntetickej značky s fotografiou, teda samotné obrázky a k nim polia súradníc. Použitím polí súradníc cieľového (fotografia) a zdrojového (syntetická ŠPZ) obrázka ako argumentov funkcie `cv2.findHomography()` získame transformačnú maticu, ktorá použitím vo funkcií `cv2.warpPerspective()` spôsobí požadovanú deformáciu obrázka syntetickej značky tak, aby zodpovedala polohe skutočnej ŠPZ vo vybranej fotografii, obrázok 4.12. Následne sa pomocou funkcie `cv2.fillConvexPoly()` začierni plocha značky na fotografii a tým sa pripraví na spojenie s vhodne deformovaným obrázkom syntetickej značky, obrázok 4.13.

Takto vzniknutý výsledný obrázok 4.14 je pomocou funkcie `cv2.resize()` zmenšený na veľkosť, ktorá je používaná pri strojovom učení, obrázok 4.15. Na obrázok je v prípade po-



Obrázek 4.13: Plocha poznávej značky na fotografií sa začierni pre bezchybné prilepenie syntetickej ŠPZ.



Obrázek 4.14: Obrázok, ktorý vznikol spojením syntetickej poznávej značky a fotografie automobilu na základe nájdenia homografie medzi rovinou syntetickej značky a polohou reálnej značky v rovine.



Obrázek 4.15: Výsledný obrázek zmenšený na velikost vhodnú na strojové učenie.



Obrázek 4.16: Výsledný obrázok po aplikácii šumu imitujúceho pohyb.

treby možné aplikovať ešte rôzne filtre a šumy, napríklad šum imitujúci pohyb 4.16 pomocou funkcie `applyMotionBlur()`, ktorej implementácia môže vyzerá nasledovne [8].

```
# Function applies motion blur to the image
def applyMotionBlur(image):

    size = 17
    # generating the kernel
    kernel_motion_blur = np.zeros((size, size))
    kernel_motion_blur[int((size-1)/2), :] = np.ones(size)
    kernel_motion_blur = kernel_motion_blur / size

    # applying the kernel to the input image
    output = cv2.filter2D(image, -1, kernel_motion_blur)

    return output
```

Výsledok je uložený pomocou funkcie `cv2.imwrite()` na miesto určené jedným zo vstupných argumentov skriptu. V priemere sa 5000 obrázkov vygeneruje za približne 2 hodiny, čo znamená, že pri spustení skriptu na troch počítačoch naraz sa 500 000 obrázkov vygeneruje za približne 66 hodín. Všetky podstatné informácie o vygenerovanom obrázku získané z textových súborov k zdrojovému a cieľovému obrázku sa vo funkcií `createJson()` uložia do súboru vo formáte JSON:

```
data = {
    'lp_text' : info[0],
    'nation' : info[1],
    'point0' : coordinates[0],
    'point1' : coordinates[1],
    'point2' : coordinates[2],
    'point3' : coordinates[3],
    'gaussian_blur' : gauss,
    'motion_blur' : motion,
}
```

Význam jednotlivých premenných:

- `lp_text` – znenie identifikačného čísla ŠPZ
- `nation` – skratka národnosti
- `gaussian_blur`, `motion_blur` – typy použitých filtrov a šumov
- `point0-3` – súradnice poznávacej značky v tvare `[x,y]`

V prípade, že sa jedná o zvláštny typ maďarskej ŠPZ (taxi, nákladné vozidlá alebo vozidlá poháňané elektrickým prúdom), je navyše do atribútu `cartype` objektu `data` uložená hodnota `info[2]` (s hodnotou TAXI, TRUCK alebo E-CAR) v závislosti na type značky. Takto vyzerajúci objekt je uložený do súboru vo formáte JSON pomocou funkcií `json.dumps(data)` a funkcií `open()`, `write()` a `close()`.

## 4.3 Použitie skriptov na generovanie dát pre analýzu do-pravy

Každý z trojice skriptov sa spúšťa pomocou interpretu jazyka Python v termináli ako konzolová aplikácia s argumentami príkazového riadku.

### 4.3.1 Použitie generátora syntetických poznávacích značiek

V prípade potreby je možné si nechať vypísať do konzole nápovedu týmto spôsobom:

```
$ python rzgenerator.py -h
```

Príklad použitia skriptu:

- Spustením skriptu sa vygeneruje 10 000 poznávacích značiek Českej republiky.

```
$ python rzgenerator.py -i 10000 -t cz
```

- Spustením skriptu sa vygeneruje 10 000 poznávacích značiek Slovenskej republiky.

```
$ python rzgenerator.py -i 10000 -t sk
```

- Spustením skriptu sa vygeneruje 10 000 poznávacích značiek Poľska.

```
$ python rzgenerator.py -i 10000 -t pl
```

- Spustením skriptu sa vygeneruje 10 000 poznávacích značiek Maďarska.

```
$ python rzgenerator.py -i 10000 -t h
```

Argumenty príkazového riadku sú 3, pričom 2 sú povinné a majú nasledovný význam:

- `-i` množstvo značiek, ktoré sa má vygenerovať
- `-o` cesta k adresáru, do ktorého sa majú vygenerované obrázky značiek uložiť (tento argument je nepovinný a v prípade, že nie je zadaný sa vygenerované obrázky uložia do adresára RZ/)
- `-t` skratka štátu, ktorého poznávacie značky majú byť vygenerované (sk/cz/pl/h)

### 4.3.2 Použitie skriptu s užívateľským rozhraním pre zaznamenanie súrad-níc polohy poznávacích značiek vo fotografiách

V prípade potreby je možné si nechať vypísať do terminálu nápovedu týmto spôsobom:

```
$ python clicker.py -h
```

Skript očakáva jediný argument, ktorý je voliteľný a v prípade, že nebude zadaný, cesta k adresáru je nastavená na `photo_templates/`.

```
$ python clicker.py -i path/to/source/
```

Po spustení skriptu sa v prípade správneho zadania cesty k fotografiám zobrazí v okne jedna z fotografií. V opačnom prípade program vypíše do konzole chybové hlásenie. Kliknutím ľavého tlačidla myši sa na fotografií zobrazí červený bod. Stlačením kláves W a S sa mení y-ová súradnica a stlačením kláves A a D sa mení x-súradnica, vplyvom čoho sa bod posúva po pozadí fotografie. Konkrétne, po stlačení klávesy W sa y-ová súradnica dekrementuje, čo spôsobí posun bodu nahor, stlačením klávesy S sa y-ová súradnica inkrementuje, čo spôsobí posun bodu nadol, po stlačení klávesy A sa dekrementuje x-ová súradnica, čo spôsobí posun bodu doľava a sa po stlačení klávesy D sa x-ová súradnica inkrementuje, čo bude mať za následok posun bodu vpravo. V prípade stlačenia klávesy Z sa zobrazí obrázok spred posledného kliknutia, takže bod, ktorý bol vykreslený ako posledný sa odstráni. To je však možné len predtým, než sa poloha tohto bodu uloží stlačením klávesy X. Každú polohu červeného bodu reprezentujúceho dvojicu súradníc  $(x, y)$  je potrebné stlačením klávesy X uložiť do poľa súradníc. Klávesou ENTER je potrebné potvrdiť vybrané súradnice a klávesou ESC prejsť na ďalší obrázok. Prehľad významu jednotlivých kláves je v tabuľke 4.2.

Tabuľka 4.2: Tabuľka znázorňuje ovládanie konzolovej aplikácie `clicker.py` určenej na uloženie súradníc poznávacej značky nachádzajúcej sa na fotografií automobilu.

klávesa	efekt vyvolaný stlačením
W	posun nahor, dekrementácia y-ovej súradnice
A	posun vľavo, dekrementácia x-ovej súradnice
S	posun vpravo, inkrementácia x-ovej súradnice
D	posun nadol, inkrementácia y-ovej súradnice
Z	krok späť
X	uloženie usporiadanej dvojice symbolizujúcej pozíciu jednej zo štyroch súradníc
ENTER	potvrdenie vybraných súradníc zobrazených červenými bodmi
ESC	prechod na nasledujúci obrázok

### 4.3.3 Použitie skriptu generujúceho výslednú dátovú sadu určenú na strojové učenie

V prípade potreby je možné si nechať vypísať do terminálu nápovedu týmto spôsobom:

```
$ python supertool.py -h
```

Skript očakáva na vstupe 3 argumenty, pričom sú všetky voliteľné a ich význam je nasledovný:

- `-t` je cesta k adresáru obsahujúcemu fotografie automobilov
- `-r` je cesta k adresáru obsahujúcemu synteticky vygenerované poznávacie značky
- `-o` je cesta k adresáru, do ktorého sa má uložiť výstup skriptu

Ak hodnoty týchto argumentov nie sú zadané na vstupe, cesty sú:

- `-t photo_templates_front2/`
- `-r RZ/`
- `-o final/`



## 4.4 Odhaľovanie chýb a vylepšovanie vlastností dátovej sady

Za účelom odhalenia chýb a zlepšenia vlastností dátovej sady bol vytvorený dotazník, obrázok 4.17 pozostávajúci z fotografií áut s reálnymi a syntetickými ŠPZ pomere 1:2. Užívateľ má pri vyplňovaní rozhodnúť pri každej fotografii, či sa jedná o automobil so skutočnou alebo syntetickou ŠPZ. Na základe zozbieraných dát sú vytvorené štatistiky, vďaka ktorým je možné odhaliť znaky prezrádzajúce, že sa jedná o syntetickú ŠPZ. Dotazník je implementovaný pomocou webových technológií (HTML/CSS/JS/PHP) a umiestnený na fakultnom serveri<sup>1</sup>.

Dotazník bol predložený 50 respondentom. Odpovede boli zozbierané a štatisticky vyhodnotené. Na obrázku 4.18 sú fotografie automobilov s poznávacími značkami, ktoré boli pri každom zobrazení v dotazníku označené za syntetické. Príčinou môžu byť svetelné podmienky, čo spôsobuje rozdiely v intenzite farieb medzi fotografiou a syntetickou ŠPZ, odraz svetla, ktorý sa na syntetickej ŠPZ nevyskytuje a podobne. Ďalšou z príčin môže byť absencia nedokonalostí, ktoré reálne značky majú (napríklad výraznejšie flaky alebo poškodenia). Riešenie taktiež neberie do úvahy tieňe pokrývajúce len niektoré časti poznávacej značky a nedokonalosti môžu vznikáť aj pri nepresnom určení súradníc rohov reálnych značiek na fotografiách.

Na druhej strane, obrázok 4.19 znázorňuje fotografie, ktoré boli označené vo všetkých prípadoch ako reálne. Fotografie sú zachytené z rôznych pozorovacích uhlov, vďaka čomu zaniknú niektoré detaily na značke a značka vyzerá reálne. Pri vytváraní dátovej sady fotografií a získavaní súradníc značiek automobilov na fotografiách je dôležité tak isto presné ohraňovanie značky, vďaka čomu pôsobí obrázok autenticky. Správna miera aditívnych šumov aplikovaných na plochu syntetickej ŠPZ má priaznivý vplyv na vzhľad výsledného obrázka a vytvára realistický dojem.

Osobitnú kategóriu tvoria fotografie, ktoré boli niekoľkokrát označené za reálne aj za syntetické. Na základe zistených skutočností o jednotlivých fotografiách je možné konštatovať, že významnú rolu v procese určovania originality obrázkov vygenerovaných pomocou generátora hrajú vonkajšie vplyvy, ale aj vlastnosti farieb a prídavné efekty v podobe filtrov a šumov aplikovaných na syntetické poznávacie značky.










---

<sup>1</sup><http://www.stud.fit.vutbr.cz/~xsvore01/index.php>

## Registration number originality recognition quiz

There are couple of images of cars with czech, slovak and polish license plate. Please answer "Real" if you think, that the licence plate on the picture is real or "Fake" if you think, that it's fake. All images are required.

Please choose one option:

		
<input type="radio"/> Real <input type="radio"/> Fake	<input type="radio"/> Real <input type="radio"/> Fake	<input type="radio"/> Real <input type="radio"/> Fake
		
<input type="radio"/> Real <input type="radio"/> Fake	<input type="radio"/> Real <input type="radio"/> Fake	<input type="radio"/> Real <input type="radio"/> Fake
		
<input type="radio"/> Real <input type="radio"/> Fake	<input type="radio"/> Real <input type="radio"/> Fake	<input type="radio"/> Real <input type="radio"/> Fake
<input type="button" value="Submit and Next"/>		

Obrázek 4.17: Dotazník originality poznávacích značek vytvorený za účelom odhalenia chýb a vylepšovanie syntetickej dátovej sady.



Obrázek 4.18: Obrázky automobilov so syntetickými poznávacími značkami vyskytujúce sa v dotazníku originality ŠPZ označované vo všetkých prípadoch ako syntetické.



Obrázek 4.19: Obrázky automobilov so syntetickými poznávacími značkami vyskytujúce sa v dotazníku originality ŠPZ označované vo všetkých prípadoch ako reálne.

# Kapitola 5

## Záver

Táto bakalárska práca slúži na generovanie syntetickej dátovej sady, ktorá je určená pre analýzu dopravy pomocou systému využívajúceho počítačové videnie a umelú inteligenciu. Učenie systému s umelou inteligenciou závisí od vlastností vstupnej dátovej sady, ktorá je mu poskytnutá. Monitorovanie dopravy pomocou počítačového videnia má široké spektrum využitia, na základe ktorého sa líšia aj požiadavky na dátové sady používané na strojové učenie.

Výhodou generátora syntetickej dátovej sady je možnosť prispôbenia vlastností generovaných dát pomocou vstupných argumentov jednotlivých skriptov. Generátor dokáže vygenerovať dostatočne veľkú dátovú sadu validných poznávacích značiek Českej republiky, Slovenska, Poľska a Maďarska pomocou skriptu `rzgenerator.py` a obrazových dát potrebných na vytvorenie značiek jednotlivých štátov, ktoré sú uložené v repozitári verzovacieho systému<sup>1</sup>, kde sa nachádza aj 100 fotografií automobilov spolu s textovými súbormi obsahujúcimi súradnice značiek, ktoré boli vytvorené pomocou skriptu `clicker.py` a sú pripravené na ďalšie spracovanie skriptom `supertool.py` vytvárajúcim konečnú dátovú sadu. Každý takto vytvorenej fotografií odpovedá súbor vo formáte JSON, ktorý obsahuje všetky potrebné informácie k danej fotografií potrebné v procese učenia. Rýchlosť generovania závisí od výkonnosti počítača.

Vytvorený generátor je možné ďalej prispôbovať v závislosti na požiadavkách vyššie uvedeného systému a vylepšovať vizuálne vlastnosti generovaných dát podľa potreby, napríklad aplikáciou filtrov spôsobujúcich deformácie a nedokonalosti na syntetických značkách alebo úpravou svetlosti niektorých miest na značke na základe dopadu svetla vo fotografií tak, aby výsledok spĺňal vlastnosti dôležité pre čo najlepšiu efektivitu strojového učenia. K zlepšovaniu vlastností generátora môže dopomôcť aj dotazník originality poznávacích značiek, z ktorého výsledkov je možné štatisticky odpozorovať výrazné nezrovnalosti a extrémny.

V tejto fáze implementácie generátora syntetickej dátovej sady určenej na strojové učenie by bolo vhodné zamerať sa na experimentovanie s vygenerovanými dátami a na základe výsledov učenia vylepšovať schopnosti generátora, napríklad rozšírením sady fotografií automobilov. V prípade potreby je možné sadu skriptov využiť na generovanie poznávacích značiek ďalších krajín, čomu predchádza získanie potrebných obrazových dát analogickými postupmi, ako v prípade štátov podrobne rozobratých v tejto práci.

---

<sup>1</sup><https://git.fit.vutbr.cz/xsvore01/IBP>

# Literatura

- [1] OpenCV library. <https://opencv.org/>.
- [2] Photos of vehicles and license plates. <https://www.platesmania.com/cz/informer>.
- [3] Anagnostopoulos, C. N. E.; Anagnostopoulos, I. E.; Loumos, V.; aj.: A License Plate-Recognition Algorithm for Intelligent Transportation System Applications. *IEEE Transactions on Intelligent Transportation Systems*, ročník 7, č. 3, Sept 2006: s. 377–392, ISSN 1524-9050, doi:10.1109/TITS.2006.880641.
- [4] Gonzalez; C., R.; amd Richard E., W.: *Thresholding. In Digital Image Processing*. Pearson Education, 2002, ISBN 81-7808-629-8.
- [5] Hsu, G. S.; Chen, J. C.; Chung, Y. Z.: Application-Oriented License Plate Recognition. *IEEE Transactions on Vehicular Technology*, ročník 62, č. 2, Feb 2013: s. 552–561, ISSN 0018-9545, doi:10.1109/TVT.2012.2226218.
- [6] Huand, T.; Russell, S.: *Object identification: a Bayesian analysis with application to traffic surveillance*. Elsevier B.V., 1998.
- [7] Jaruga, J.: *Historia polskich tablic rejestracyjnych*. Centrum Obsługi Kancelarii Prezesa Rady Ministrów, 2007, ISBN 9788392561002.  
URL <https://books.google.cz/books?id=SUPjlgEACAAJ>
- [8] Joshi, P.: *OpenCV with Python By Example*. Packt Publishing, 2015, ISBN 9781785283932.  
URL  
<https://www.packtpub.com/application-development/opencv-python-example>
- [9] Lencsés, C.: *Meglepetésbetűvel jön az új rendszám*. [Online; navštíveno 21.06.2016].  
URL <http://www.vezess.hu/magazin/2016/06/21/meglepetesbetuvel-jon-az-uj-rendszam/>
- [10] Li, H.; Shen, C.: Reading Car License Plates Using Deep Convolutional Neural Networks and LSTMs. *CoRR*, ročník abs/1601.05610, 2016, 1601.05610.  
URL <http://arxiv.org/abs/1601.05610>
- [11] Marinov, P.: *České poznávací značky - historie, dnešek, budoucnost*. Kampe, 2007, ISBN 978-80-902955-7-5.
- [12] Parker, N.; Weeks, J.: *Registration Plates of the World*. Europlate, 2004, ISBN 9780950273570.  
URL <https://books.google.cz/books?id=ij4WLwAACAAJ>

- [13] Rasheed, S.; Naeem, A.; Ishaq, O.: Automated Number Plate Recognition Using Hough Lines and Template Matching.
- [14] Svoboda, P.; Hradis, M.; Marsik, L.; aj.: CNN for License Plate Motion Deblurring. *CoRR*, ročník abs/1602.07873, 2016, [1602.07873](https://arxiv.org/abs/1602.07873).  
URL <http://arxiv.org/abs/1602.07873>
- [15] Zelený, F.; Feuereisl, D.: *Poznávací značky v Čechách, na Moravě a ve Slezsku*. SAXI, 2011, ISBN 978-80-904767-2-1.
- [16] Špaňhel, J.; Sochor, J.; Juránek, R.; aj.: CNN for License Plate Motion Deblurring. *CoRR*, ročník abs/1602.07873, 2016, [1602.07873](https://arxiv.org/abs/1602.07873).  
URL <http://arxiv.org/abs/1602.07873>
- [17] Špaňhel, J.; Sochor, J.; Juránek, R.; aj.: Holistic recognition of low quality license plates by CNN using track annotated data. In *2017 14th IEEE International Conference on Advanced Video and Signal Based Surveillance (AVSS)*, Aug 2017, s. 1–6, doi:10.1109/AVSS.2017.8078501.

# Příloha A

## Obsah CD

- `report.pdf` – technická správa
- `latex/` – zdrojové súbory  $\text{\LaTeX}$ u
- `src/` – zdrojové kódy aplikácie
- `temp/` – syntetická dátová sada
- `poster.pdf` – demonštračný plagát
- `video.mp4` – video demonštrujúce fungovanie aplikácie