



**FAKULTA  
INFORMAČNÍCH  
TECHNOLOGIÍ  
ČVUT V PRAZE**

## ZADÁNÍ BAKALÁŘSKÉ PRÁCE

**Název:** Rozpoznávání želv  
**Student:** Kristýna Pátková  
**Vedoucí:** Ing. Radek Richtr, Ph.D.  
**Studijní program:** Informatika  
**Studijní obor:** Znalostní inženýrství  
**Katedra:** Katedra aplikované matematiky  
**Platnost zadání:** Do konce letního semestru 2021/22

### Pokyny pro vypracování

Účelem práce je vytvořit a otestovat prototyp k rozpoznávání želv (např. rodu Testudo Graeca) se zvláštním přihlédnutím k návrhu nového datasetu, pro úlohu určeného.

- 1) Proveďte rešerši problematiky rozpoznávání želv, debatujte použitelné příznaky a druhy želv, zaměřte se především na druh Testudo Graeca.
- 2) Proveďte analýzu stávajících řešení.
- 3) Navrhněte ideální podobu želvího datasetu, zhodnoťte dataset vytvořený v předchozí iteraci projektu.
- 4) Navrhněte a následně implementujte prototyp pro rozpoznávání želv.
- 5) Prototyp otestujte v ÚTIA AVČR na existujícím datasetu.
- 6) Debatujte efektivitu prototypu a výsledky.

### Seznam odborné literatury

Matěj Sedláček, Michal Haindl, D. Formanová: An Automatic Tortoise Specimen Recognition. Progress in Pattern Recognition, Image Analysis, Computer Vision, and Applications: 21st Iberoamerican Congress, CIARP 2016, 52-59. Springer International Publishing, Cham 2017.

Recognition of tortoises (Rozpoznávání želv), Matěj Sedláček, FJFI, 2015,

Majtánová, Adriána. Rozpoznávanie korytnačiek. Bakalárska práca. Praha: České vysoké učení technické v Praze, Fakulta informačních technologií, 2020.

Ing. Karel Klouda, Ph.D.  
vedoucí katedry

doc. RNDr. Ing. Marcel Jiřina, Ph.D.  
děkan

V Praze dne 3. února 2021





**FAKULTA  
INFORMAČNÍCH  
TECHNOLÓGIÍ  
ČVUT V PRAZE**

Bakalářská práce

## Rozpoznávání želv

*Kristýna Pátková*

Katedra aplikované matematiky  
Vedoucí práce: Ing. Radek Richtr, Ph.D.

25. června 2021





---

## Poděkování

Tímto bych chtěla poděkovat Ing. Radkovi Richtrovi, Ph.D. za užitečné rady a trpělivost při vedení této práce a dále také mé rodině a přáteli za jejich obrovskou podporu po celou dobu mého studia.



---

# Prohlášení

Prohlašuji, že jsem předloženou práci vypracovala samostatně a že jsem uvedla veškeré použité informační zdroje v souladu s Metodickým pokynem o dodržování etických principů při přípravě vysokoškolských závěrečných prací.

Beru na vědomí, že se na moji práci vztahují práva a povinnosti vyplývající ze zákona č. 121/2000 Sb., autorského zákona, ve znění pozdějších předpisů. V souladu s ust. § 2373 odst. 2 zákona č. 89/2012 Sb., občanský zákoník, ve znění pozdějších předpisů, tímto uděluji nevýhradní oprávnění (licenci) k užití této mé práce, a to včetně všech počítačových programů, jež jsou její součástí či přílohou a veškeré jejich dokumentace (dále souhrnně jen „Dílo“), a to všem osobám, které si přejí Dílo užít. Tyto osoby jsou oprávněny Dílo užít jakýmkoli způsobem, který nesnižuje hodnotu Díla a za jakýmkoli účelem (včetně užití k výdělečným účelům). Toto oprávnění je časově, teritoriálně i množstevně neomezené.

V Pardubicích dne 25. června 2021

.....

České vysoké učení technické v Praze

Fakulta informačních technologií

© 2021 Kristýna Pátková. Všechna práva vyhrazena.

*Tato práce vznikla jako školní dílo na Českém vysokém učení technickém v Praze, Fakultě informačních technologií. Práce je chráněna právními předpisy a mezinárodními úmluvami o právu autorském a právech souvisejících s právem autorským. K jejímu užití, s výjimkou bezúplatných zákonných licencí a nad rámec oprávnění uvedených v Prohlášení na předchozí straně, je nezbytný souhlas autora.*

### **Odkaz na tuto práci**

Pátková, Kristýna. *Rozpoznávání želv*. Bakalářská práce. Praha: České vysoké učení technické v Praze, Fakulta informačních technologií, 2021.

---

# Abstrakt

Tato bakalářská práce se zabývá rozpoznáváním jednotlivých exemplářů želv druhu *Testudo Graeca* na základě fotografií jejich plastronů. Rozpoznávací systém je vytvořen v programovacím jazyce Python. Pro automatickou detekci plastronů na fotografii je použita metoda Mask R-CNN a pro extrakci příznaků určených ke klasifikaci (multiclass či binární) jsou testovány 3 architektury konvolučních neuronových sítí – VGG16, InceptionV3 a ResNet50. Všechny tyto metody jsou v práci vysvětleny, také je zde popis příbuzných druhů želv z čeledi Testudinidae, jejich značení a použitelné identifikační příznaky na plastronech. Práce též obsahuje analýzu současných řešení problematiky automatického rozpoznávání želv a použitých datasetů, včetně návrhu ideální podoby želvího datasetu.

**Klíčová slova** Rozpoznávání želv, automatické rozpoznávání, extrakce příznaků, *Testudo graeca*, Mask R-CNN, VGG16, InceptionV3, ResNet50, Python

---

# Abstract

This bachelor thesis deals with the recognition of individual tortoises of the *Testudo Graeca* species based on photographs of their plastrons. The recognition system is written in the Python programming language. For automatic detection of plastrons in the photograph, the Mask R-CNN method is used, and for the feature extraction, 3 architectures of convolutional neural networks – VGG16, InceptionV3 and ResNet50, are tested. All these methods are explained in this thesis, there is also a description of related species of tortoises from the family Testudinidae, their marking and suitable identification features on the plastron. Thesis also contains an analysis of the current solutions to the issue of automatic recognition of tortoises and of used datasets, including proposal of the ideal form of a tortoises dataset.

**Keywords** Recognition of tortoises, automatic recognition, feature extraction, *Testudo graeca*, Mask R-CNN, VGG16, InceptionV3, ResNet50, Python

---

# Obsah

<b>Úvod</b>	<b>1</b>
Cíl(e) práce . . . . .	2
<b>1 Problematika rozpoznávání želv</b>	<b>3</b>
1.1 Druhy želv . . . . .	3
1.2 Značení želv . . . . .	4
1.3 Plastron a jeho příznaky . . . . .	5
<b>2 Současný stav řešení problému</b>	<b>9</b>
2.1 L. Tichý a K. Kintrová . . . . .	9
2.2 Matěj Sedláček . . . . .	9
2.3 Adriána Majtánová . . . . .	10
2.4 Zdeněk Svatoň . . . . .	12
<b>3 Datasetsy</b>	<b>13</b>
3.1 Dataset České inspekce životního prostředí . . . . .	13
3.2 Dataset chovatele Petra Petráse . . . . .	14
3.3 Návrh ideální podoby želvího datasetu . . . . .	15
<b>4 Použité metody</b>	<b>17</b>
4.1 Detekce plastronu . . . . .	17
4.1.1 Konvoluční neuronové sítě . . . . .	17
4.1.1.1 Detektory typu R-CNN . . . . .	19
4.1.2 Augmentace dat . . . . .	24
4.1.3 Lineární regrese a Metoda nejmenších čtverců . . . . .	25
4.1.4 RANSAC . . . . .	26
4.1.5 Eukleidovská vzdálenost . . . . .	27
4.1.6 Výpočet úhlu mezi 2 vektory . . . . .	27
4.1.7 Rotace . . . . .	28

4.2	Extrakce příznaků, klasifikace . . . . .	28
4.2.1	Architektury konvolučních neuronových sítí . . . . .	28
4.2.2	Logistická regrese . . . . .	32
4.2.3	Metoda nejbližších sousedů . . . . .	32
4.2.4	Ladění hyperparametrů . . . . .	32
<b>5</b>	<b>Implementace prototypu pro rozpoznávání želv</b>	<b>35</b>
5.1	Použité nástroje . . . . .	35
5.1.1	Python + Jupyter notebook . . . . .	35
5.2	Trénování modelu Mask R-CNN . . . . .	36
5.3	Detekce plastronu na fotografii . . . . .	39
5.4	Extrakce příznaků a jejich klasifikace . . . . .	41
<b>6</b>	<b>Experimenty, diskuze</b>	<b>45</b>
	<b>Závěr</b>	<b>51</b>
	<b>Literatura</b>	<b>53</b>
<b>A</b>	<b>Seznam použitých zkratek</b>	<b>59</b>
<b>B</b>	<b>Obsah přiloženého média</b>	<b>61</b>



---

## Seznam obrázků

1.1	Identifikační fotografie plastronu druhu <i>Testudo graeca</i> s vyznačenými názvy štítků . . . . .	6
1.2	Vývoj černých skvrn na plastronu . . . . .	7
3.1	Nevhodné fotografie plastronů z datasetu Petra Petráse . . . . .	14
4.1	Max pooling . . . . .	18
4.2	Struktura konvoluční neuronové sítě . . . . .	19
4.3	Průběh algoritmu Selective search . . . . .	20
4.4	Architektura R-CNN . . . . .	21
4.5	Architektura Fast R-CNN . . . . .	22
4.6	RPN (Region proposal network) . . . . .	23
4.7	Princip fungování metody RoIAlign . . . . .	24
4.8	Detekce outlierů a inlierů metodou RANSAC . . . . .	27
4.9	Architektura konvoluční neuronové sítě VGG16 . . . . .	29
4.10	Inception modul konvoluční neuronové sítě GoogLeNet . . . . .	30
4.11	Reziduální bloky konvoluční neuronové sítě ResNet . . . . .	31
4.12	Křížová validace . . . . .	33
5.1	Anotovaná fotografie . . . . .	38
5.2	Oblasti detekované pomocí natrénovaného modelu Mask R-CNN . . . . .	40
5.3	Extrahované oblasti plastronu a centrálního švu . . . . .	41



---

## Seznam tabulek

2.1	Porovnání úspěšnosti předchozích řešení . . . . .	12
3.1	Nejčastější vlastnosti fotografií plastronů z datasetu České inspekce životního prostředí . . . . .	14
3.2	Vlastnosti fotografií plastronů z datasetu chovatele Petra Petráše .	15
5.1	Úspěšnost detekce jednotlivých oblastí pomocí modelu Mask R-CNN	39
6.1	Úspěšnost multiclass klasifikace na datasetu ČIŽP . . . . .	46
6.2	Úspěšnost multiclass klasifikace na datasetu Petra Petráše . . . . .	46
6.3	Úspěšnost binární klasifikace na datasetu ČIŽP . . . . .	47
6.4	Úspěšnost binární klasifikace na datasetu Petra Petráše . . . . .	47



---

# Úvod

Želvy žlutohnědé (*Testudo graeca*) jsou v současné době na červeném seznamu ohrožených druhů IUCN<sup>1</sup> označeny jako zranitelné. Především jsou ohrožené kvůli nelegálnímu trhu se zvířaty. V České republice proto tento druh na základě Washingtonské úmluvy podléhá registraci a u každé želvy se eviduje fotografie spodní strany krunýře (plastronu) sloužící k identifikaci jednotlivých exemplářů.

Výsledek práce by měl tuto identifikaci usnadnit. Nyní, když má například kontrolor České inspekce životního prostředí posoudit, zda je želva zaregistrována, musí osobně porovnat krunýř dané želvy se stovkami různě kvalitních fotografií v databázi. Zatím totiž neexistuje žádná využitelná aplikace, která by dokázala na základě fotografií jednotlivé želvy automaticky rozpoznávat.

Z výše uvedeného důvodu jsem se rozhodla pro výběr tématu Rozpoznávání želv, protože právě vytvořením systému pro rozpoznávání želv mohu pomoci ke vzniku takové aplikace, a navíc díky tomu získám nové zkušenosti z oboru.

V práci se tedy zabývám tvorbou systému pro rozpoznávání jednotlivých exemplářů želv druhu *Testudo graeca* na základě fotografií jejich plastronů, naopak cílem mé práce není vytvoření uživatelského prostředí nebo mobilní aplikace.

Nejdříve se věnuji rešerši problematiky rozpoznávání želv, tedy výběru vhodných příznaků, které se s věkem želvy nemění a které jsou zároveň pro každého jedince unikátní a rešerši různých druhů želv se zaměřením na druh *Testudo graeca*.

Dále analyzuji stávající řešení a na jejich základě navrhuji a implementuji svůj vlastní prototyp pro rozpoznávání želv obsahující sadu výše zmíněných příznaků. Takový prototyp musí umět správně lokalizovat želvu na fotografii,

---

<sup>1</sup>Mezinárodní svaz ochrany přírody (anglicky International Union for Conservation of Nature).

vyextrahovat zvolené příznaky a porovnat je s příznaky ostatních želv z fotografií v databázi.

V další části provádím sérii experimentů na existujícím datasetu v ÚTIA AVČR<sup>2</sup> (a na datasetu chovatele želv Petra Petráse) a debatuji efektivitu prototypu v závislosti na použitých příznacích.

Tato práce je součástí širšího projektu věnujícího se rozpoznávání želv druhu *Testudo graeca* ve spolupráci s ÚTIA AVČR a zabývá se stejným tématem jako diplomová práce *Recognition of tortoises (Rozpoznávání želv)* [1] Matěje Sedláčka, studenta Fakulty jaderné a fyzikálně inženýrské ČVUT v Praze, bakalářská práce *Rozpoznávání korytnaček* [2] Adriány Majtánové a diplomová práce *Detekce želvých plastronů pro rozpoznávání* [3] Zdeňka Svatoně z Fakulty informačních technologií ČVUT v Praze.

## Cíl(e) práce

Cílem této bakalářské práce je vytvoření automatického systému pro rozpoznávání jedinců želv druhu *Testudo graeca* z fotografií jejich plastronů.

Prvním cílem je získání přehledu literatury o želvách (zejména druhu *Testudo graeca*), o předzpracování obrazu, extrakci obrazových příznaků a jejich klasifikaci a důkladné prostudování této literatury z důvodu prozkoumání stávajících metod řešících tuto problematiku. Dalším cílem je výběr vhodných příznaků sloužících k identifikaci jednotlivých želv. Navazujícím cílem je posouzení vhodnosti stávajících metod pro jejich využití v praktické části mé práce, případné navržení metod vlastních a následné vytvoření prototypu pro rozpoznávání želv. Posledním cílem je otestování mého prototypu na existujícím datasetu v ÚTIA AVČR (a na datasetu Petra Petráse) a zhodnocení jeho efektivitu.

---

<sup>2</sup>Ústav teorie informace a automatizace Akademie věd České republiky.

---

# Problematika rozpoznávání želv

## 1.1 Druhy želv

Na evropském kontinentu se ve volné přírodě vyskytují pouze 4 druhy suchozemských želv (čeleď Testudinidae). Jedná se především o želvy zelenavé (*Testudo hermanni*), želvy žlutohnědé (*Testudo graeca*), želvy vroubené (*Testudo marginata*) a okrajově také o želvy stepní (*Testudo horsfieldi*), jejichž výskyt do Evropy zasahuje jen malou částí v oblasti jižního Uralu [4].

Všechny tyto druhy želv mají řadu určovacích znaků, podle kterých je možné mezi nimi rozlišovat, pro laika je to ale obtížné i podle těchto popsanych detailů.

Želva stepní má plochý a téměř kruhovitý krunýř, její karapax (horní část krunýře) má většinou pískovou barvu s výraznými černými znaky, na stehnech má shluk různě velkých trnů a také má na předních končetinách pouze čtyři prsty, podle čehož se někdy také označuje jako želva čtyřprstá.

V identifikaci želvy zelenavé nejvíce napomáhá rohovitý trn na konci ocasu, který žádný jiný z těchto druhů želv nemá. Dalším určovacím znakem je poté rozdělený nadocasní štítek, žlutozelené tóny ve zbarvení krunýře nebo absence šupinatých výrůstků na stehnech.

Želva vroubená je největší evropskou želvou, její krunýř je podlouhlý, zadní štítky prodloužené a mírně zahnuté nahoru, naopak spodní část krunýře bývá vklenutá dovnitř.

Želvu žlutohnědou lze jednoznačně určit podle nápadného rohovitého výrůstku na zadních stehnech, kvůli těmto ostruhám se tomuto druhu někdy nesprávně přezdívá želva ostruhatá<sup>3</sup>.

Barva, velikost, váha i způsob života želv se může lišit i mezi jedinci stejného druhu.

---

<sup>3</sup>Tento název je vyhrazen africkému druhu *Geochelone sulcata*.

Například populace želvy žlutohnědé se vyskytuje roztroušeně od jižní Evropy po Střední východ a dokonce i v severní Africe. Kromě lišícího se zabarvení se liší i způsobem zimování – evropské druhy zimují alespoň 2–4 měsíce v roce, ale pro ostatní jedince je zimování nepřípustné, z důvodu teplejšího zimního období v jejich domovině prožívají jen kratší období útlumu a skutečné zimování je pro ně smrtelné.

U želv chovaných v zajetí rychlost růstu ovlivňují hlavně chovné podmínky a strava. Obrovský rozdíl ve velikostech je znatelný u želv, které jsou po celý rok chované uvnitř (bez zimování) a jsou nesprávně krmené a mezi těmi, které mají venkovní výběh, jsou krmeny přírodní stravou a část zimy prospí.

U všech těchto druhů se projevuje také sexuální dimorfismus. Samice bývají větší, mívají kratší ocas než samci a spodní strana krunýře bývá u samců zakřivená dovnitř kvůli reprodukci. Určení pohlaví u těchto druhů do velikosti 10–13 cm je mnohdy však nepřesné.

### 1.2 Značení želv

Kromě želvy stepní patří všechny druhy želv popsané výše mezi ohrožené druhy a vztahuje se na ně tedy úmluva CITES<sup>4</sup>, jejíž cílem je regulace obchodu s exempláři ohrožených druhů získaných z volné přírody, ale i kontrola obchodu s živočichy odchovanými v zajetí nebo člověkem vypěstovanými rostlinami druhů, které jsou v přírodě ohroženy.

Z CITES vyplývá, že každá želva musí mít své doklady. Ty je možné obdržet již vyplněné se želvou při její koupi od chovatele a nebo je nutné je vytvořit při registraci nově vylíhnuté želvy. Doklady se skládají z tzv. bílého papíru – registračního listu a tzv. žlutého papíru – výjimky ze zákazu obchodní činnosti. Součástí obou těchto dokladů je identifikační fotografie exempláře.

#### Identifikační fotografie

Dle metodiky AOPK ČR<sup>5</sup> [5] by měla být identifikační fotografie pořizována za jasného rozptýleného denního světla, být ostrá, barevně věrná a jejím předmětem by měl být plastron (břišní část krunýře). Fotografovaný exemplář by měl být položen na kontrastní podložce s přiloženým dobře čitelným měřítkem a zaujímat cca 80 % této fotografie.

Želvy by měly být fotografovány optimálně těsně před registrací a dále by se jejich identifikační fotografie měla aktualizovat vždy, když se výrazně změní kresba jejich plastronu. Tuto povinnost mají chovatelé až do té doby,

---

<sup>4</sup>Úmluva o mezinárodním obchodu s ohroženými druhy volně žijících živočichů a planě rostoucích rostlin (anglicky Convention on International Trade in Endangered Species of Wild Fauna and Flora), též označována jako Washingtonská úmluva.

<sup>5</sup>Agentura ochrany přírody a krajiny České republiky.



dokud želvy nedorostou konečné velikosti a přestane se jim měnit kresba na plastronu, nebo dokud nejsou označeny mikročipem.

Podle fotografie v dokladech CITES musí být exemplář vždy jednoznačně a nezaměnitelně identifikovatelný.

#### Čipování

Běžný mikročip je tvořen kapslí z bioskla obvykle o rozměrech cca  $2 \times 11$  mm, použití těchto standardních mikročipů však není vhodné pro malé exempláře želv do velikosti plastronu 10 cm a hmotnosti 500 g. V České republice proto není čipování želv v současné době zákonnou povinností a je umožněna alternativní identifikace exemplářů pomocí fotografií plastronů a každý chovatel se může tedy sám rozhodnout, zda nechá své želvy čipovat.

Čipování želv je výhodou pro větší chovatele, protože jim tím odpadá povinnost aktualizace fotografií plastronů u velkého množství exemplářů a je to nejsnadnější identifikační metoda, při které nemůže dojít k záměně želv.

Zavedení čipu s sebou vždy ale nese určité riziko zdravotních komplikací – čip může při nevhodném umístění želvě způsobit krvácení a vést tak až k jejímu úhynu, je finančně nákladnější než pořizování fotografií – kromě samotné ceny čipu a jeho zavedení, si chovatel pro identifikaci jednotlivých exemplářů musí pořídit i vhodné čtecí zařízení čipů a vzhledem k dlouhověkosti želv, otázkou také zůstává, zda bude tato registrační technologie a funkčnost čipu zachována po celý život dané želvy.

Dalším problémem je i možnost zneužití čipů pomocí těžko odhalitelného nelegálního transferu čipu z uhynulého do živého zvířete.

#### Čísla na krunýřích

Dle CITES musí být každý exemplář nezaměnitelně označen, aby nedošlo k záměně dokladů mezi chovanými jedinci. Proto u větších chovatelů můžeme narazit ještě na další typ značení želv – číselný popis na krunýřích. Tento typ krátkodobého značení se používá především u mláďat želv, pro které by byla aplikace čipu vzhledem k malé velikosti příliš nebezpečná.

### 1.3 Plastron a jeho příznaky

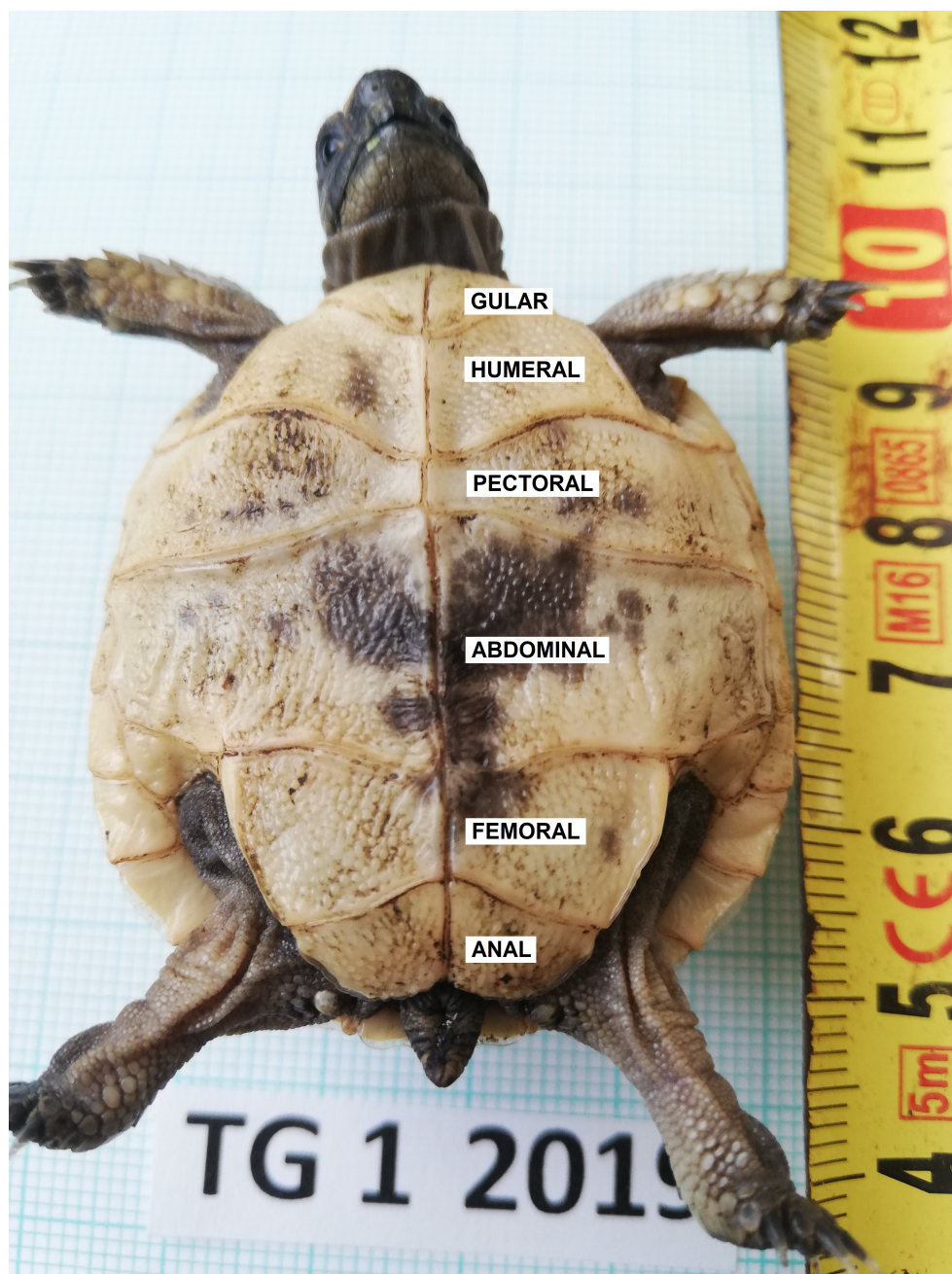
Plastron je spodní část krunýře želvy, je rozdělen do plošek, tzv. štítků, které se uprostřed plastronu připojují podél centrálního švu. Tyto štítky tvoří na plastronu šest symetrických párů a jsou označovány jako chytací (gular), končetinové/pažní (humeral), prsní (pectoral), břišní (abdominal), stehenní (femoral) a anální (anal). Jejich umístění je vyznačeno na fotografii 1.1.

Toto uspořádání štítků je stejné u všech druhů želv zmíněných výše, plastrony se u nich liší především ve zbarvení a celkové velikosti, proto většinu

## 1. PROBLEMATIKA ROZPOZNÁVÁNÍ ŽELV

---

popsaných příznaků lze použít k identifikaci i ostatních druhů a ne pouze druhu *Testudo graeca*, na který se tato práce primárně zaměřuje.



Obrázek 1.1: Ukázka plastronu želvy druhu *Testudo graeca* s vyznačenými názvy štítků, identifikační fotografie z osobního archivu chovatele Petra Petráše

## Plastron druhu *Testudo graeca*

V této sekci jsou uvedeny všechny uvažované identifikační příznaky plastronu želv druhu *Testudo graeca*.

### Černé skvrny

U druhu *Testudo graeca* je na plastronu patrná nepravidelná černá kresba, která je mezi jednotlivými exempláři unikátní. Tato kresba se spolu se želvou vyvíjí. Dochází k tomu, že i zprvu nevýrazné skvrnky se postupem času na jednotlivých štítcích rozrůstají a v dospělosti mají jedinci tohoto druhu plastron často úplně černý. Tento vývoj je zachycen na sérii čtyř fotografií 1.2 prospívající želvy žlutohnědé ve věku 0, 2, 4 a 6 měsíců v článku [6] Lubomíra Tichého.

Z tohoto důvodu není vhodné k identifikaci využít pouze tyto skvrny, tímto způsobem by bylo možné porovnávat pouze fotografie pořízené ve stejném období.

Pokud by se ale porovnávala aktuálně pořízená fotografie s fotografiemi staršími, mohly by skvrny posloužit jako určitý filtr, který by z porovnání odebíral takové fotografie plastronů, na kterých černé skvrny zabírají větší plochu než na plastronu z nejnovejší fotografie. Skvrny totiž časem pouze rostou, ale neztrácejí se.



Obrázek 1.2: Ukázka vývoje černých skvrn na plastronu prospívající želvy žlutohnědé ve věku 0, 2, 4 a 6 měsíců, převzato z [6]

### Marginální švy

Použití naměřených délek štítků marginálních švů (společně s délkami štítků centrálního švu) k identifikaci jedinců želv druhu *Testudo graeca* dosahuje dle článku [7] L. Tichého a K. Kintrové výborných výsledků, pro tuto práci ale není využití marginálních švů možné, protože na fotografiích v dostupných datasetech není tato část plastronu vždy dostatečně viditelná.

### Centrální šev

S růstem krunýře se mění délky štítků přisedajících na centrální šev i úhly, které s ním svírají švy postranní (švy mezi štítky na levé a pravé straně plastronu), tyto příznaky jsou proto pro identifikaci jedinců zachycených v různých časových obdobích nevhodné.

Příznak, který se s věkem želvy nemění, je vzájemný kontakt postranních štítků podél centrálního švu. Postranní švy se na centrálním švu buď vzájemně dotýkají nebo jsou mírně posunuty jedním či druhým směrem. Tyto 3 možnosti dávají na 5 spojení na plastronu celkově až  $3^5 = 243$  různých kombinací, což není pro každou želvu unikátní, a tedy ani samostatně k identifikaci použitelné, ale alespoň k odfiltrování nevhodných jedinců při porovnávání to případně postačující příznak je.

Pro identifikaci použitelné příznaky, které již byly popsány v článku [7] L. Tichého a K. Kintrové a diplomové práci [1] M. Sedláčka, které se považují za neměnné, lišící se pro různé želvy a které jsou viditelné na všech fotografiích plastronů, jsou délky štítků připojujících se na centrální šev (zvláště z levé a pravé strany) a rozdíly ve spojení postranních švů se švem centrálním, vše znormalizované celkovou délkou centrálního švu – z důvodu eliminace změn způsobených růstem plastronu či různým rozlišením fotografií.

---

## Současný stav řešení problému

Jak již bylo zmíněno výše, (nejen) v České republice se u každé registrované želvy druhu *Testudo graeca* eviduje fotografie spodní strany krunýře – plastronu, který je podél centrálního švu uprostřed rozdělen do plošek – štítků, a který slouží pro identifikaci jednotlivých exemplářů.

V minulosti se proto již objevilo několik pokusů o identifikaci želv pomocí fotografií jejich plastronů. Některé tyto pokusy jsou popsány a mezi sebou porovnány v této kapitole.

### 2.1 L. Tichý a K. Kintrová

L. Tichý a K. Kintrová se ve svém článku [7] snažili pomocí morfometrie přijít na spolehlivou neinvazivní metodu pro identifikaci želv, a tím nahradit široce využívané čipování, které ale u některých mladých želv může způsobovat zdravotní problémy. V tomto článku [7] se tedy zaměřili na relativní charakteristiky plastronu, které jsou nezávislé na pohlaví a věku, liší se v populaci a jsou široce dokumentovány u chovaných želv. Vycházeli zde z poznatků J. E. Mosimanna [8]. Ke zjištění míry odlišnosti mezi dvěma jedinci využili kombinaci měření relativních rozdílů v délkách štítků centrálního a marginálních švů normalizovaných celkovou délkou švů na plastronu. Tato měření přinesla vynikající výsledky přesahující dle autorů 99% spolehlivost, byly ale prováděny manuálně.

### 2.2 Matěj Sedláček

O automatický rozpoznávací systém se ve své diplomové práci *Recognition of tortoises (Rozpoznávání želv)* [1] pokusil Matěj Sedláček, který pro identifikaci jednotlivých želv použil délky štítků připojujících se na centrální šev (zvláště z levé a pravé strany) a rozdílly ve spojení postranních švů s centrálním švem, vše znormalizované celkovou délkou centrálního švu. Délky štítků margi-

nálních švů nebylo možné použít, protože ne na všech fotografiích od chovatelů jsou okraje plastronů dobře viditelné.

Tento automatický systém se skládal ze dvou částí – automatické extrakce výše zmíněných příznaků a jejich následné klasifikace. Bylo třeba zjistit, jak je plastron orientován, přesně lokalizovat centrální šev, na něm najít místa připojení jednotlivých štítků, změřit jejich délky, z nich sestavit vektor příznaků a ten porovnat s vektory příznaků ostatních fotografií plastronů v databázi pomocí klasifikace, na jejímž základě bylo možné určit, zda je či není daná želva zaregistrovaná.

Konkrétně tento proces probíhal tak, že se vstupní fotografie nejprve zmenšila takovým způsobem, aby její nejdelší strana měla rozměr 500 pixelů a zároveň aby zůstal zachován poměr jejích stran. Dále byla fotografie rotována okolo svého středu v úhlech  $-15^\circ$ ,  $-10^\circ$ ,  $-5^\circ$ ,  $0^\circ$ ,  $5^\circ$ ,  $10^\circ$ ,  $15^\circ$ ,  $75^\circ$ ,  $80^\circ$ ,  $85^\circ$ ,  $90^\circ$ ,  $95^\circ$ ,  $100^\circ$  a  $105^\circ$ . V každém z těchto úhlů se provedla detekce hran pomocí Cannyho hranového detektoru a ta natočená fotografie, která měla ve své středové oblasti napočítáno největší množství vertikálních hran, tedy taková fotografie, na které se želví plastron nacházel ve vertikální pozici, byla použita pro následující operace.

V dalším kroku se pomocí zobecněné Houghovy transformace využívající nakreslenou šablonu plastronu určila pozice a orientace plastronu na fotografii – jestli je na ní želva natočena hlavou nahoru či dolů, a případně bylo natočení změněno tak, aby hlava mířila směrem vzhůru.

Dále byla pomocí grafového algoritmu nalezena místa připojení štítků na centrální šev. Začátek a konec centrálního švu a s nimi logicky i hrany plastronu u hlavy a ocasu, byly rozeznány opět pomocí zobecněné Houghovy transformace. Po přesném nalezení těchto spojení bylo již možné sestavit vektor příznaků a přistoupit tak ke klasifikaci prováděné algoritmem  $k$ -nejbližších sousedů.

Dosažená celková úspěšnost identifikace želvy tímto automatickým rozpoznávacím systémem v databázi České inspekce životního prostředí obsahující 276 nekvalitních fotografií plastronů s různým nastavením scény byla 43 % (u jiného datasetu obsahujícího kvalitnější fotografie pak 60,7 %), proto tento systém dosud nebyl vhodný pro reálné uplatnění v aplikaci.

### 2.3 Adriána Majtánová

V roce 2020 se řešením problematiky rozpoznávání želv zabývala Adriána Majtánová ve své bakalářské práci *Rozpoznávanie korytnačiek* [2]. Její řešení ale nebylo plně automatické, protože jako vstup jejího rozpoznávacího systému byly využité předem ručně ořezané fotografie želv pouze na oblast plastronu z důvodu eliminace rušivých vlivů pozadí.

Ke klasifikaci přistupovala ze 2 hledisek:

- Za prvé prozkoumala tzv. multiclass klasifikaci, kde na základě vstupní fotografie predikovala unikátní identifikační číslo želvy a tím i třídu, ke které má daná želva na fotografii nejbližší. Tento přístup je vhodný spíše pro chovatele, kteří chtějí rychle rozlišovat jednotlivé exempláře želv mezi velkým množstvím chovaných jedinců. Naopak tento přístup není v praxi vhodný pro zjišťování, zda je či není daná želva zaregistrována.
- Za druhé k této problematice přisupovala jako k binárnímu problému, tedy jako k porovnání dvou fotografií, pro které vypočítala vektory příznaků a následně jejich rozdíl. Dle výsledného rozdílu bylo možné určit, jestli se jedná o stejné jedince a tím pádem i zda je daný jedinec registrovaný a to na základě toho, zda porovnávaná fotografie je součástí databáze legálně chovaných jedinců.

K extrakci příznaků využila a mezi sebou porovнала deskriptory SIFT (Scale Invariant Feature Transform) a ORB (Oriented FAST and Rotated BRIEF) – oba algoritmy společně s konceptem Bag of Visual Words, dále LBP (Local Binary Patterns) s konceptem uniformních vzorů a CNN (konvoluční neuronové sítě) – konkrétně konvoluční neuronovou sítí VGG16. K výpočtu rozdílu mezi dvěma vektory příznaků, získaných ze dvou různých fotografií plastronů, zvolila Eukleidovskou vzdálenost. Výsledné histogramy či vektory příznaků následně použila jako vstup pro vybrané klasifikátory.

Těmito testovanými klasifikačními modely pro predikci identifikačních čísel želv byly SVM (Support Vector Machine), multinomiální naivní Bayesův klasifikátor a logistická regrese a pro porovnání dvou jedinců želv algoritmus  $k$ -nejbližších sousedů.

Součástí této práce [2] je i série experimentů provedených na již existujícím datasetu v ÚTIA AVČR a také na novém datasetu, který se A. Majtánové podařil získat od chovatele želv Petra Petráse, a který obsahuje 51 fotografií plastronů 4 různých dospělých jedinců želv druhu *Testudo graeca* odfočených z různých úhlů, vzdáleností a na různých površích.

Z těchto experimentů vyplynulo, že pro predikci identifikačních čísel želv, se jako nejúspěšnější extraktor příznaků jevila předtrénovaná konvoluční neuronová síť s úspěšností 100 % na datasetu Petra Petráse a s úspěšností 56,67 % na datasetu ÚTIA AVČR. Naopak pro tento typ úlohy se ukázala jako nejméně vhodná metoda LBP. Při porovnání dvou jedinců želv byla v průměru nejúspěšnější opět předtrénovaná konvoluční neuronová síť s maximální dosaženou úspěšností 77,78 %, přičemž této maximální úspěšnosti dokázal dosáhnout i algoritmus SIFT.

## 2.4 Zdeněk Svatoň

Jako poslední se o řešení problematiky rozpoznávání želv, konkrétně o část detekce vlastností želvích krunýřů z fotografie, pokusil v roce 2021 Zdeněk Svatoň. Ve své diplomové práci *Detekce želvých plastrů pro rozpoznávání* [3], kde navázal na Michala Sedláčka, se mu podařilo optimalizovat grafový algoritmus pro detekci spojení štítků na centrální šev a místo zobecněné Houghovy transformace se rozhodl pro lokalizaci plastru na fotografii použít pro větší robustnost neuronovou síť.

Problémem této práce [3] je ale nemožnost zopakování uvedených postupů, protože chybí potřebné detaily, součástí nejsou žádné přílohy a celková pochopitelnost textu je velmi malá.

## Porovnání

V následující tabulce 2.1 jsou pro lepší přehlednost znázorněny úspěšnosti jednotlivých řešení popsanych výše. Kromě řešení L. Tichého a K. Kintrové byla všechna tato řešení vyhodnocována na datasetu fotografií plastrů České inspekce životního prostředí, jehož aktuální podoba je popsána v následující kapitole.

Tabulka 2.1: Porovnání úspěšnosti předchozích řešení

	Detekce		Klasifikace
	způsob	úspěšnost	úspěšnost
L. Tichý a K. Kintrová	manuálně	-	97,5–99,9 %
M. Sedláček	automaticky	63,41 %	43 %
A. Majtánová	manuálně	-	66,67–77,78 %
Z. Svatoň	automaticky	83,94 %	-



---

## Datasety

Výsledný prototyp pro rozpoznávání želv byl testován na datasetech, které jsou podrobně popsány v této kapitole. Kromě těchto použitých datasetů je zde také navržena ideální podoba želvího datasetu, jež může problém automatického rozpoznávání želv usnadnit a přinést lepší výsledky.

### 3.1 Dataset České inspekce životního prostředí

První použitý dataset je pro tuto práci zapůjčen z ÚTIA AVČR. Jedná se o dataset České inspekce životního prostředí (ČIŽP) skládající se z 299 fotografií plastronů želv druhu *Testudo graeca*, z čehož 239 fotografií obsahuje různé jedince a pouze 30 želv je v tomto datasetu vyfocených dvakrát.

Fotografie v tomto datasetu pocházejí z různých časových období, od různých chovatelů želv, jsou pořizovány různými zařízeními<sup>6</sup> a liší se tedy v kvalitě, rozlišení i v nastavení scény. Želvy jsou fotografovány na různých podkladech, v horizontální či vertikální pozici, při různé intenzitě a směru osvětlení, někdy je při snímání použitý blesk, což vytváří na plastronu odlesky nebo stíny, také plastron není vždy dostatečně očištěný a želvy jsou často mírně nakloněné na pravý či levý bok, čímž dochází k tomu, že na fotografii nejsou viditelné marginální švy.

Zařízení a jeho nastavení, které bylo k pořizování fotografií plastronů použito nejčastěji, je znázorněno v tabulce 3.1.

Tento dataset není veřejný, tudíž ani v této práci není možné publikovat příklady fotografií z tohoto datasetu.

---

<sup>6</sup>Fotografie byly pořízeny pomocí fotoaparátů Nikon D50 a COOLPIX S210, Canon PowerShot S3 IS, Olympus C3000Z, Panasonic DMC-LX2 a DMC-FZ20 ale například i pomocí scanneru HP Scanjet G3110.

### 3. DATASETY

---

Tabulka 3.1: Nejčastější vlastnosti fotografií plastronů z datasetu České inspekce životního prostředí

Zařízení	Nikon D50
Rozlišení (px)	3008 × 2000
Ohnisková vzdálenost (mm)	55
Clonové číslo	f/5,6
Expoziční čas (s)	1/60

### 3.2 Dataset chovatele Petra Petráse

Druhý dataset, který byl použit v této práci, je dataset plastronů 4 jedinců želv druhu *Testudo graeca* od chovatele Petra Petráse. Tento dataset se podařilo získat A. Majtánové při tvorbě její práce *Rozpoznávanie korytnačiek* [2].

P. Petrás pořídil v jednom časovém období celkem 248 fotografií svých 4 chovaných jedinců želv druhu *Testudo graeca* z různých úhlů a vzdáleností a na různých površích. Ne všechny fotografie jsou ale pro problém rozpoznávání želv použitelné, viz. příklady fotografií 3.1. Tyto fotografie buď zachycují plastrony z velké dálky a nebo jsou na nich plastrony z části zakryté, špinavé a vůbec se nepodobají identifikačním fotografiím CITES, a tedy ani neodpovídají doporučením uvedeným v metodice AOPK ČR [5].



Obrázek 3.1: Ukázka nevhodných fotografií plastronů z datasetu chovatele Petra Petráse

Z tohoto důvodu bylo A. Majtánovou vybráno pouze 50 fotografií, jež lépe odpovídají doporučením a jsou vhodné pro testování prototypu pro rozpoznávání želv.

Specifikace fotografií z tohoto datasetu je znázorněna v tabulce 3.2.

Tabulka 3.2: Vlastnosti fotografií plastronů z datasetu chovatele Petra Petráše

Zařízení	Huawei ANE-LX1
Rozlišení (px)	3456 × 4608
Ohnisková vzdálenost (mm)	3,81
Clonové číslo	f/2,2
Expoziční čas (s)	1/100

### 3.3 Návrh ideální podoby želvího datasetu

Fotografie v ideálním datasetu by měly dodržovat doporučení uvedená v metodice AOPK ČR [5]. Plastron by na nich měl být tedy umístěn na kontrastní podložce (s přiloženým měřítkem), měl by zabírat co největší část (cca 80 %) fotografie, neměl by být rozmazaný a fotografie by měly být pořizovány za jasného rozptýleného denního světla – bez použití blesku, aby nedocházelo k odleskům, přesvícení či ztrátě kresby a přirozeného zbarvení plastronu.

Dle pozorování provedených při tvorbě této práce, by bylo vhodné k doporučením zmíněným výše doplnit ještě tato:

- Plastron by měl být před samotným fotografováním důkladně očištěn od prachu, hlíny a dalších nečistot, které by mohly zakrýt příznaky důležité pro jeho identifikaci. Vhodné by bylo plastron navlhčit, aby na něm kresba byla zřetelnější, ale to pouze za použití takového osvětlení, které by nezpůsobilo jeho lesknutí.
- Želva by měla být při fotografování vypočtena, aby se nenaklápěla k jedné či druhé straně, a na fotografii byly viditelné i marginální švy, které by mohly dopomoci k lepším výsledkům identifikace [7].
- K usnadnění předzpracování fotografií by dopomohlo, kdyby na všech fotografiích byly želvy umístěny stejným způsobem – například vertikálně s hlavou směřující vzhůru.

Pro potřeby testování automatických rozpoznávacích systémů by takový ideální dataset měl také obsahovat co největší množství fotografií jednoho jedince v různých fázích svého vývoje, aby bylo možné posoudit, zda jsou dané systémy schopné želvy správně rozpoznávat i přes všechny změny projevující se na jejich plastronech.



## Použité metody

V této kapitole jsou popsány všechny metody použité při řešení problematiky rozpoznávání želv. První sekce se zaměřuje na metody potřebné pro detekci plastronu na fotografii a druhá sekce uvádí metody extrakce příznaků a jejich klasifikace.

### 4.1 Detekce plastronu

Tato sekce je zaměřena pouze na metody, které byly v této práci použity pro detekci plastronu na fotografii.

#### 4.1.1 Konvoluční neuronové sítě

Konvoluční neuronové sítě (zkráceně CNN – z anglického Convolutional Neural Network) jsou druhem neuronových sítí, jež jsou využívány zejména v oblasti rozpoznávání obrazových dat a jejich klasifikace.

Svůj název získaly podle toho, že k extrakci příznaků ze vstupního obrazu využívají operaci konvoluce, kterou je možné pro dvojrozměrný obraz vyjádřit následujícím vztahem:

$$g(x, y) = f(x, y) * h(x, y) = \sum_{i=-S/2}^{S/2} \sum_{j=-R/2}^{R/2} f(x-i, y-j) \cdot h(i, j), \quad (4.1)$$

kde  $g(x, y)$  značí výstupní obraz – mapu příznaků,  $f(x, y)$  vstupní obraz a  $h(x, y)$  tzv. konvoluční jádro (nebo také filtrační jádro či pouze filtr) o rozměrech  $R \times S$ .

V závislosti na hodnotách těchto konvolučních jader je možné detekovat různé druhy obrazových příznaků (hrany, křivky...) a konvoluční neuronové sítě si v průběhu trénovacího procesu dokáží samy vhodné hodnoty konvolučních jader určit. Před konvolučním krokem je však nutné určit hyperparametry, které mají efekt na výslednou velikosti mapy příznaků:

#### 4. POUŽITÉ METODY

---

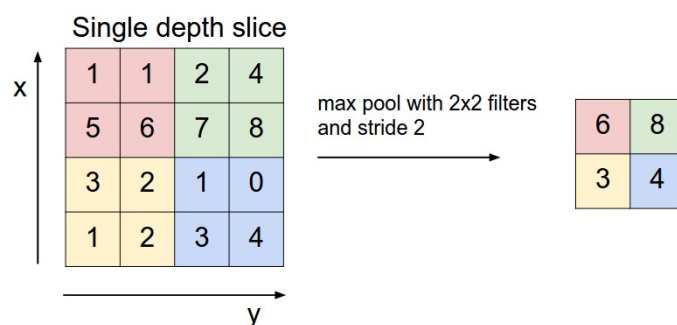
- Počet použitých filtrů udává hloubku mapy příznaků.
- Velikost kroku, značící o kolik pixelů se v každém kroku posouvá matice filtru po matici vstupního obrazu, ovlivňuje počet aplikací filtru – větší velikost kroku vede k méně aplikacím filtru a tedy k menší velikosti výstupního obrazu.
- Nulové odsazení umožňuje použít filtrační matici na všechny prvky matice vstupního obrazu takovým způsobem, že prvkům spadajícím za hranice matice přiřazuje nulové hodnoty<sup>7</sup>. Díky tomu je možné získat výstup srovnatelné velikosti se vstupem.
- Velikost samotného filtru, od čehož se vše zmíněné výše odvíjí.

Konvoluční vrstva CNN je typicky následována aktivační funkcí. Nejčastěji využívanou aktivační funkcí je ReLU (Rectified Linear Unit):

$$f(\xi) = \max(0, \xi) = \begin{cases} \xi & \text{pro } \xi \geq 0 \\ 0 & \text{pro } \xi < 0 \end{cases}, \quad (4.2)$$

kteřá v mapě příznaků nahrazuje všechny negativní hodnoty pixelů hodnotou nula<sup>8</sup> a jejím účelem je zavést do CNN jistý prvek nelinearity.

Dálší vrstvou CNN je tzv. pooling vrstva. Tato vrstva redukuje dimenzionalitu mapy příznaků, ale přitom zachovává nejdůležitější informace – díky ní se snižuje množství parametrů, výpočtů a slouží i jako určitá kontrola proti přeučení sítě. Nejběžněji se využívá typ max pooling s filtrem o rozměrech  $2 \times 2$  a krokem velikosti 2, při němž se z podoblastí mapy příznaků, upravené v předchozím kroku, vybírá nejvyšší hodnota, viz. obrázek 4.1.



Obrázek 4.1: Ukázka max pooling s filtrem o rozměrech  $2 \times 2$  a krokem velikosti 2, převzato z [9]

<sup>7</sup>Pro konvoluci používající nulové odsazení se v angličtině používá označení wide convolution a pro konvoluci, která nulové odsazení nepoužívá, narrow convolution.

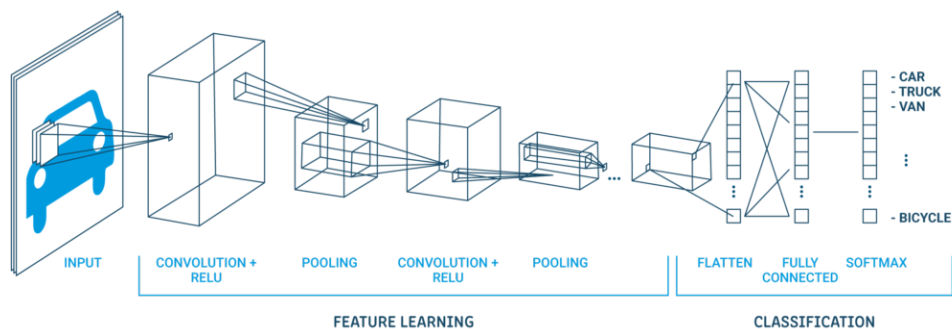
<sup>8</sup>Takto upravená mapa příznaků se v angličtině označuje jako rectified feature map.

Na závěr je výstup z poolingové vrstvy převeden na 1 dimenzionální vektor příznaků a připojen k poslední, plně propojené vrstvě, což je klasický vícevrstvý perceptron<sup>9</sup>, který pro finální klasifikaci (predikci hodnot  $1, \dots, c$ ) využívá jako aktivační funkci softmax:

$$f_i(\boldsymbol{\xi}) = \frac{e^{\xi_i}}{e^{\xi_1} + e^{\xi_2} + \dots + e^{\xi_c}} \quad (4.3)$$

kde  $\boldsymbol{\xi} = (\xi_1, \dots, \xi_c)^T$  je vektor vnitřních potenciálů  $c$  neuronů a  $f_i(\boldsymbol{\xi})$  je aktivační funkce  $i$ -tého neuronu, jejíž výstup je interpretován jako pravděpodobnost příslušnosti ke třídě  $i$ , tj. jako  $\hat{P}(Y = i | \mathbf{X} = \mathbf{x})$  [10].

Konvoluční, ReLU a pooling vrstvy jsou základními stavebními kameny konvolučních neuronových sítí – dohromady všechny jejich sady slouží k extrakci příznaků, které jsou na závěr klasifikovány v plně propojené vrstvě [11]. Kompletní struktura CNN je znázorněna na obrázku 4.2



Obrázek 4.2: Ukázka kompletní struktury konvoluční neuronové sítě, převzato z [12]

#### 4.1.1.1 Detektory typu R-CNN

V této sekci je popsán vývoj modelů strojového učení, určených pro počítačové vidění a konkrétně detekci objektů, založených na konvolučních neuronových sítích, až po model Mask R-CNN, který tato práce využívá.

### R-CNN

Metoda R-CNN (Regions with CNN features) [13] byla představena v roce 2013 týmem Rosse Girshicka a byla navržena pro řešení problému detekce

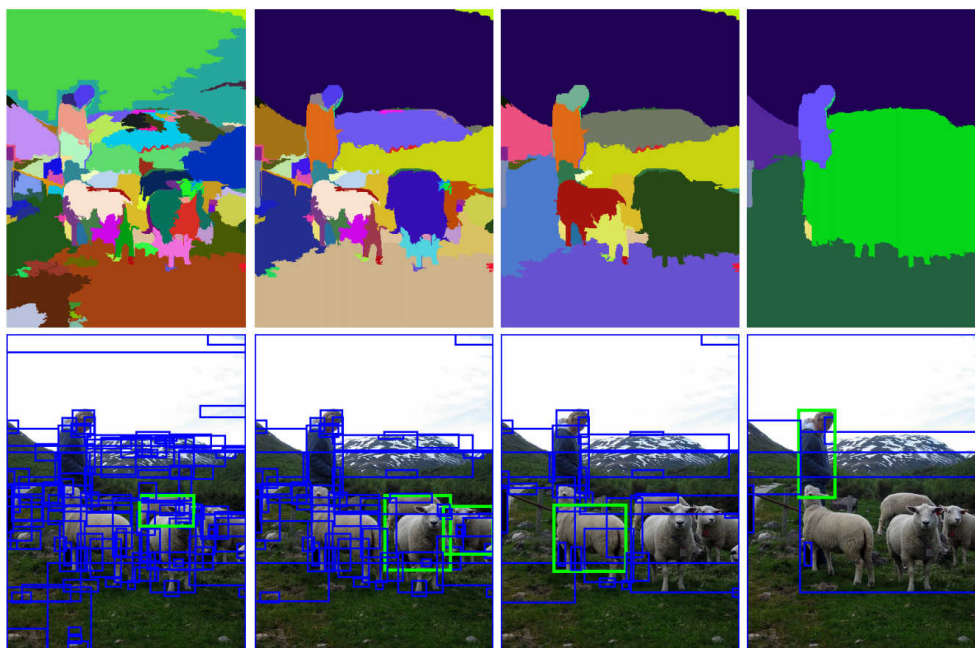
<sup>9</sup>Neuronová síť skládající se z vrstev, které jsou propojené tak, že výstupy neuronů z jedné vrstvy tvoří vstupy neuronů do další vrstvy.

objektů v obraze. Jedná se o detekční systém skládající se ze 3 základních částí.

První část používá algoritmus Selective search pro generování přibližně 2000 oblastí, ve kterých se pravděpodobně nacházejí hledané objekty, nehlédě na jejich kategorie. Tyto oblasti (v angličtině označované jako region proposals) jsou ve druhé části předány konvoluční neuronové síti, která z nich vypočítává vektory příznaků, jež jsou následně předány modelu SVM<sup>10</sup>, aby klasifikoval objekty přítomné v těchto navržených oblastech.

Selective search [14] kombinuje výhody vyhledávání hrubou silou a segmentace – jako při vyhledávání hrubou silou se snaží nalézt všechny možné oblasti výskytu objektů a jako při segmentaci využívá strukturu obrazu ke zjišťování podobností pixelů. Algoritmus postupně shlukuje pixely podobných vlastností (na základě více kritérií – barvy, textury...) a tak rekurzivně generuje oblasti s pravděpodobným výskytem hledaných objektů, viz. obrázek 4.3.

Před předáním vygenerovaných oblastí konvoluční neuronové síti<sup>11</sup>, je nutné ještě tyto oblasti upravit tak, aby nabývaly rozměru  $227 \times 227$  pixelů. Pro každou takto upravenou oblast je pak pomocí CNN vyextrahován 4096 dimenzionální vektor příznaků.



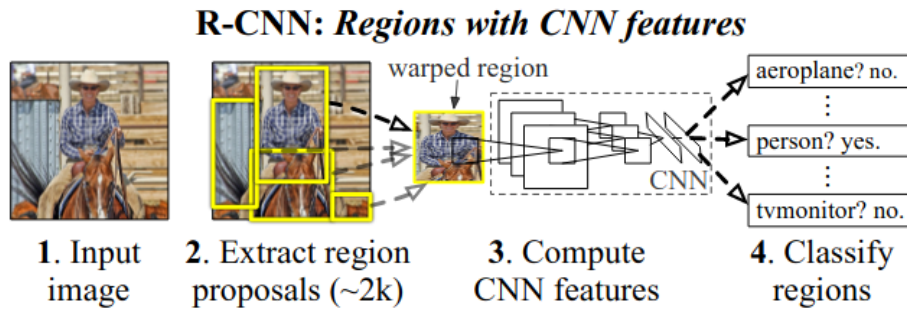
Obrázek 4.3: Ukázka průběhu algoritmu Selective search znázorňující důležitost použití různých měřítek pro nalezení oblastí obsahujících hledané objekty, převzato z [14]

<sup>10</sup>Support Vector Machines neboli Metoda podpůrných vektorů.

<sup>11</sup>V originální práci [13] byla použita upravená architektura AlexNet skládající se z 5 konvolučních vrstev a 2 plně propojených vrstev.



Klasifikátory SVM, natrénované zvláště pro všechny kategorie, následně vyhodnocují, zda se v daných oblastech nějaké objekty nacházejí a do jakých kategorií spadají. V této fázi také probíhá tzv. Bounding-box regrese, jejíž úkolem je zlepšení lokalizace výsledných ohraničujících rámečků detekovaných objektů. Celý proces je znázorněn na obrázku 4.4.



Obrázek 4.4: Ukázka jednotlivých fází systému R-CNN, převzato z [13]

Tento model má však spoustu nedostatků. Algoritmus Selective search nelze trénovat, což může někdy vést ke generování nevhodných oblastí. Všechny vygenerované oblasti procházejí konvoluční neuronovou sítí, což pro každý obrázek znamená cca 2000 průchodů a také každou část R-CNN je třeba trénovat nezávisle (CNN, SVM i Bounding box regresor). Z tohoto všeho vyplývá, že R-CNN model je velmi pomalý při trénování i detekci<sup>12</sup>.

### Fast R-CNN

V roce 2015 Ross Girshick představil novou metodu, pojmenovanou jako Fast R-CNN [15], u které vylepšil proces trénování tím způsobem, že spojil 3 nezávisle trénované části modelu R-CNN do jednoho společně trénovaného celku, čímž se celý proces zrychlil a došlo i ke zlepšení přesnosti detekce.

U metody Fast R-CNN je nejprve opět použit algoritmus Selective search pro generování oblastí s pravděpodobným výskytem objektů, tyto oblasti už ale nejsou CNN předávány samostatně, nýbrž všechny najednou, společně s celým vstupním obrazem.

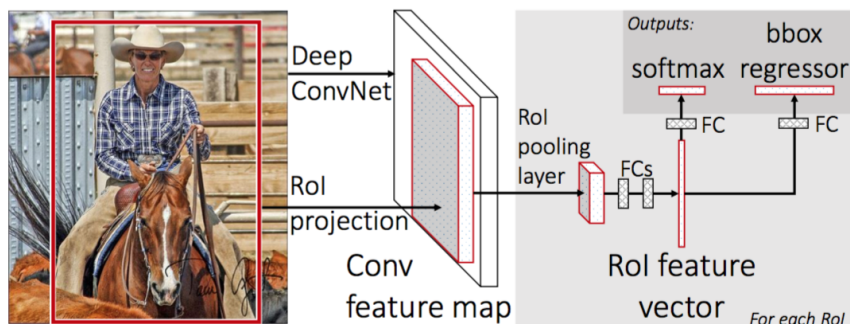
CNN celý tento vstupní obraz zpracovává a vypočítává pro něj mapu příznaků. Pro každou navrženou oblast zájmu je poté pomocí RoI pooling vrstvy<sup>13</sup> vyextrahován z mapy příznaků vektor příznaků fixní délky.

Každý vektor příznaků je následně předán plně propojeným vrstvám, které se na závěr větví do dvou paralelních výstupních vrstev. Jedna větev produkuje pomocí aktivační funkce softmax pravděpodobnosti příslušnosti nalezených

<sup>12</sup>Zpracování jednoho obrázku trvá přibližně 50 s.

<sup>13</sup>Zkratka RoI pochází z anglického Region of Interest.

objektů do daných kategorií a druhá větev vrací čtyři reálná čísla určující pozici ohraničujících rámečků detekovaných objektů. Ukázku celé architektury Fast R-CNN je možné vidět na obrázku 4.5.



Obrázek 4.5: Ukázka architektury Fast R-CNN, převzato z [15]

Problém architektury Fast R-CNN však stále zůstává v generování oblastí zájmu pomocí algoritmu Selective search, který je časově náročný a nelze ho trénovat.

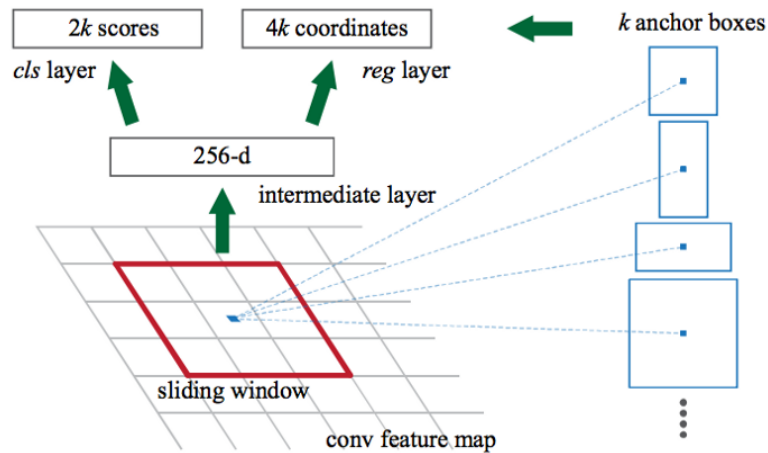
### Faster R-CNN

Dalším nástupcem je detekční systém Faster R-CNN [16] z roku 2016. Tento systém se skládá ze 2 částí – z hluboké plně konvoluční neuronové sítě RPN (Region Proposal Network) určené ke generování oblastí zájmu a z Fast R-CNN detektoru, jenž vygenerované oblasti zpracovává. Obě tyto části sdílí množinu konvolučních vrstev, celý systém je tedy jednou sjednocenou neuro-novou sítí sloužící k detekci objektů.

Pro generování oblastí zájmu RPN využívá malou síť, kterou posouvá po mapě příznaků – výstupu poslední sdílené konvoluční vrstvy. Vstupem této malé sítě je výřez z mapy příznaků o rozměrech  $n \times n$ . Každé takové posuvné okno je mapováno na příznak nižší dimenze, jenž je předáván 2 plně propojeným vrstvám – vrstvě Bounding box regrese a klasifikační vrstvě. V každé pozici tohoto posuvného okna je současně navrženo několik oblastí zájmu – maximálně  $k$ . Z toho plyne, že vrstva Bounding box regrese má  $4k$  výstupů kódujících souřadnice ohraničujících rámečků a klasifikační vrstva  $2k$  výstupních hodnot znázorňujících pravděpodobnosti toho, zda se na dané pozici nachází nějaký objekt.

Těchto  $k$  predikovaných oblastí zájmu je parametrizováno vzhledem k  $k$  referenčním ohraničením – tzv. kotvám<sup>14</sup>. Kotva je v daném posuvném okně vycentrována a přizpůsobena určenému měřítku a poměru stran, viz. 4.6.

<sup>14</sup>V angličtině jsou tyto referenční ohraničení označovány jako anchors.



Obrázek 4.6: RPN – Ukázka malé sítě v jedné své pozici, převzato z [16]

Běžně se používají 3 měřítka ( $128^2$ ,  $256^2$  a  $512^2$  pixelů) a 3 poměry stran ( $1 : 1$ ,  $1 : 2$  a  $2 : 1$ ) poskytující  $k = 9$  kotev v každé pozici. Pro mapu příznaků o rozměrech  $W \times H$  je tedy kotev celkem  $WHk$ .

Metoda Faster R-CNN v rychlosti významně překonává své předchůdce, může být dokonce použita i pro detekci objektů v reálném čase.

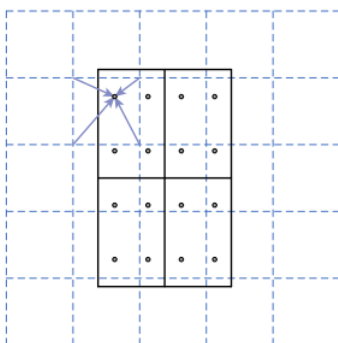
### Mask R-CNN

Metoda Mask R-CNN [17] byla představena v roce 2017 týmem z Facebook AI Research. Jedná se o rozšíření Faster R-CNN o paralelní větev umožňující predikci segmentační masky pro každou instanci nalezeného objektu. Na rozdíl od všech svých předchůdců tedy dokáže řešit problém instanční segmentace objektů na úrovni pixelů.

Pro získání přesně zarovnané binární masky již ale není dostačující použití RoI pooling vrstvy – z důvodu zaokrouhlování. Například, oblast zájmu o rozměrech  $15 \times 15$  pixelů, vybraná ze vstupního obrazu velikosti  $128 \times 128$  pixelů, odpovídající příslušné pozici na mapě příznaků velikosti  $25 \times 25$  pixelů, nabývá rozměrů  $2,93 \times 2,93$  pixelů<sup>15</sup>. Při použití RoIPoolu by byly místo 2,93 pixelů vybrány zaokrouhlené pixely 2, což by způsobilo mírné vychýlení výsledné masky vůči vstupnímu obrazu.

Proto byla tato RoIPool vrstva nahrazena vrstvou RoIAlign, která zaokrouhlování nepoužívá. Přesné hodnoty vstupních příznaků místo toho počítá pomocí bilineární interpolace na 4 pravidelně vzorkovaných bodech v každé buňce oblasti zájmu, ze kterých bere nejvyšší hodnotu (nebo průměr). Tento proces je znázorněn na obrázku 4.7.

<sup>15</sup>Každý pixel ve vstupním obrazu odpovídá  $25/128$  pixelům v mapě příznaků, proto pro  $15$  pixelů je třeba vybrat  $15 \cdot 25/128 \approx 2,93$  pixelů.



Obrázek 4.7: RoIAlign – Mřížka tvořená přerušovanou čarou reprezentuje mapu příznaků, plné čáry oblast zájmu (rozdělenou do  $2 \times 2$  buněk) a tečky 4 vzorkovací body v každé buňce, jejichž hodnoty jsou počítány z mřížky pomocí bilineární interpolace, převzato z [17]

Zavedení vrstvy RoIAlign mělo velký dopad, v modelu Mask R-CNN se zvýšila přesnost masky o relativních 10–50 % a při jejím použití došlo i k vylepšení výsledků modelu Faster R-CNN.

#### 4.1.2 Augmentace dat

Technika augmentace dat se využívá k rozšíření datasetu sloužícího k trénování modelu hlubokého učení. Z již existujících dat se buď generativními adversiálními sítěmi, nebo za použití různých transformací generují nová realistická data, jež zvyšují diverzitu trénovacího datasetu a lépe tak vystihují danou problematiku.

V této práci byly pro generování nových dat využity následující transformace:

- **Horizontální/vertikální překlopení obrazu** – prohození dvou pixelů stejně vzdálených od středové osy obrazu,
- **Rotace obrazu** kolem počátku souřadnicové soustavy o úhel  $\theta$  – aplikace rotační matice  $R = \begin{bmatrix} \cos(\theta) & -\sin(\theta) & 0 \\ \sin(\theta) & \cos(\theta) & 0 \\ 0 & 0 & 1 \end{bmatrix}$ ,
- **Úprava jasu obrazu** – vynásobení hodnoty jasu obrazu zadaným číslem,
- **Úprava kontrastu obrazu** – úprava každého pixelu  $v$  obrazu o zadané číslo  $\alpha$  dle:

$$127 + \alpha(v - 127) . \quad (4.4)$$

### 4.1.3 Lineární regrese a Metoda nejmenších čtverců

Lineární regrese se využívá pro predikci vysvětlované proměnné  $Y$  na základě  $p$  příznaků  $X_1, \dots, X_p$ . V modelu se předpokládá lineární závislost vysvětlované proměnné na hodnotách příznaků. [18]

Hodnota vysvětlované proměnné  $Y$  v bodě  $(x_1, \dots, x_p)^T$  je:

$$Y = \omega_0 + \omega_1 x_1 + \dots + \omega_p x_p + \varepsilon, \quad (4.5)$$

kde  $E\varepsilon = 0$  a koeficient  $\omega_0$  se nazývá intercept a odpovídá očekávané výchozí hodnotě  $Y$  při nulových příznacích.

Po zavedení nového konstantního příznaku  $X_0 = x_0 = 1$  a vektorovém značení  $\mathbf{x} = (x_0, x_1, \dots, x_p)^T$  a  $\boldsymbol{\omega} = (\omega_0, \omega_1, \dots, \omega_p)^T$ , lze tento zápis zkrátit na:

$$Y = \boldsymbol{\omega}^T \mathbf{x} + \varepsilon = \mathbf{x}^T \boldsymbol{\omega} + \varepsilon. \quad (4.6)$$

Model pro trénovací množinu tvořenou  $N$  páry  $(Y_i, \mathbf{x}_i)$  je  $Y_i = \mathbf{x}_i^T \boldsymbol{\omega} + \varepsilon_i$ , při značení  $\mathbf{x}_i = (1, x_{i;1}, \dots, x_{i;p})^T$ , tedy dohromady:

$$\begin{pmatrix} Y_1 \\ \vdots \\ Y_N \end{pmatrix} = \begin{pmatrix} 1 & x_{1;1} & x_{1;2} & \cdots & x_{1;p} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 1 & x_{N;1} & x_{N;2} & \cdots & x_{N;p} \end{pmatrix} \begin{pmatrix} \omega_1 \\ \vdots \\ \omega_p \end{pmatrix} + \begin{pmatrix} \varepsilon_1 \\ \vdots \\ \varepsilon_N \end{pmatrix}. \quad (4.7)$$

Maticově pak:

$$\mathbf{Y} = \mathbf{X}\boldsymbol{\omega} + \boldsymbol{\varepsilon}, \quad (4.8)$$

$$\text{kde } \mathbf{X} = \begin{pmatrix} \mathbf{x}_1^T \\ \vdots \\ \mathbf{x}_N^T \end{pmatrix} = \begin{pmatrix} 1 & x_{1;1} & \cdots & x_{1;p} \\ \vdots & \vdots & \ddots & \vdots \\ 1 & x_{N;1} & \cdots & x_{N;p} \end{pmatrix}, \quad \mathbf{Y} = \begin{pmatrix} Y_1 \\ \vdots \\ Y_N \end{pmatrix} \quad \text{a} \quad \boldsymbol{\varepsilon} = \begin{pmatrix} \varepsilon_1 \\ \vdots \\ \varepsilon_N \end{pmatrix}.$$

Součet chyb přes všechny body trénovací množiny, tj. přes všechny dvojice  $(Y_i, \mathbf{x}_i)$  pro  $i = 1, \dots, N$  pro kvadratickou ztrátovou funkci  $L(Y, \hat{Y}) = (Y - \hat{Y})^2$  aplikovanou na skutečnou hodnotou proměnné  $Y$  a odpovídající predikci  $\hat{Y}$  je:

$$\text{RSS}(\boldsymbol{\omega}) = \sum_{i=1}^N L(Y_i, \mathbf{x}_i^T \boldsymbol{\omega})^2 = \sum_{i=1}^N (Y_i - \mathbf{x}_i^T \boldsymbol{\omega})^2 = \|\mathbf{Y} - \mathbf{X}\boldsymbol{\omega}\|^2 \quad (4.9)$$

a nazývá se reziduální součet čtverců<sup>16</sup>. Minimalizací tohoto výrazu se získá odhad  $\hat{\boldsymbol{\omega}}_{OLS}$  vektoru koeficientů  $\boldsymbol{\omega}$ . Tento postup se nazývá Metoda nejmenších čtverců<sup>17</sup>.

Za předpokladu, že je matice  $\mathbf{X}^T \mathbf{X}$  regulární, existuje jediné řešení minimalizující  $\text{RSS}(\boldsymbol{\omega})$ :

$$\hat{\boldsymbol{\omega}}_{OLS} = (\mathbf{X}^T \mathbf{X})^{-1} \mathbf{X}^T \mathbf{Y}. \quad (4.10)$$

<sup>16</sup>Zkratka RSS pochází z anglického Residual Sum of Squares.

<sup>17</sup>Zkratka OLS pochází z anglického Ordinary Least Squares.

Hodnotu  $Y$  v bodě  $x$  je poté možné predikovat vztahem:

$$\hat{Y} = \hat{\omega}_0 + \hat{\omega}_1 x_1 + \dots + \hat{\omega}_p x_p , \quad (4.11)$$

nebo opět maticově:

$$\hat{Y} = \hat{\omega}_{OLS}^T \mathbf{x} = \mathbf{x}^T \hat{\omega}_{OLS} = \mathbf{x}^T (\mathbf{X}^T \mathbf{X})^{-1} \mathbf{X}^T \mathbf{Y} . \quad (4.12)$$

#### 4.1.4 RANSAC

Iterační metoda RANSAC (z anglického RANdom SAmple Consensus), představená v roce 1981 v článku *Random Sample Consensus: A Paradigm for Model Fitting with Applications to Image Analysis and Automated Cartography* [19], slouží pro odhad parametrů matematického modelu z naměřených dat obsahujících odlehlé hodnoty, tzv. outliery, přičemž tyto odlehlé hodnoty nemají mít na hodnoty odhadů žádný vliv – proto ji lze také interpretovat jako metodu detekce outlierů.

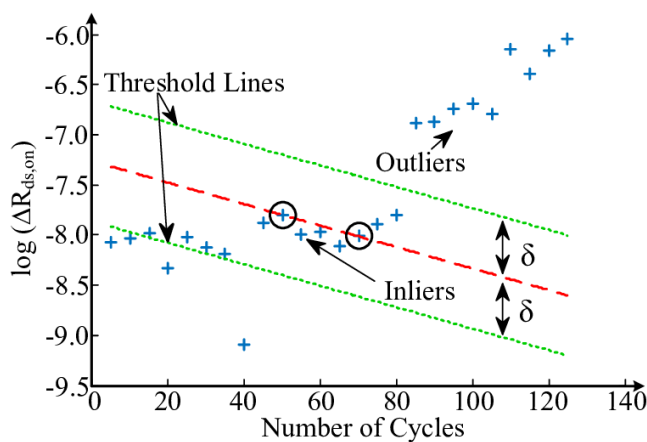
Jedná se o nedeterministický algoritmus, který produkuje správný výsledek s určitou pravděpodobností, přičemž tato pravděpodobnost se zvyšuje s počtem iterací.

V každé iteraci se provádějí tyto kroky:

1. Z naměřených hodnot je vybrána náhodná minimální podmnožina.
2. Model je na této vybrané podmnožině natrénován.
3. Ostatní hodnoty jsou testovány vůči natrénovanému modelu. Na základě thresholdu – prahu  $\delta$  jsou hodnoty klasifikovány jako outliery či inliery – blízké hodnoty. Za inliery jsou považovány takové hodnoty, jejichž chyba je menší než daný threshold  $\delta$ , jinak jsou považovány za outliery, viz. obrázek 4.8.
4. Pokud je celková chyba právě testovaného modelu nižší (počet inlierů je vyšší), než dosud nejlepšího nalezeného modelu, je tento model uložen jako nejlepší.

Průběh algoritmu je ukončen buď po provedení daného maximálního počtu iterací nebo při dosažení nějakého předem definovaného ukončujícího kritéria – např. pokud je určitý počet hodnot označen jako inlier či pokud model dosahuje určitého skóre.

Po skončení jsou hodnoty rozděleny zvlášť na outliery a inliery a použití nejlepšího nalezeného modelu odpovídá použití modelu lineární regrese (popsané v sekci 4.1.3) pouze na část inlierů.



Obrázek 4.8: Ukázka detekce outlierů a inlierů na modelu natrénovaném na podmnožině obsahující 2 hodnoty – jedná se o problém proložení daných hodnot přímkou, převzato z [20]

Počet iterací  $T$  s požadovanou úspěšností modelu  $p$  je možné vypočítat pomocí vzorce:

$$T = \frac{\log(1-p)}{\log(1-(1-e)^s)} , \quad (4.13)$$

kde  $e$  značí poměr počtu outlierů vůči počtu všech hodnot a  $s$  velikost minimální podmnožiny.

#### 4.1.5 Eukleidovská vzdálenost

Eukleidovská vzdálenost nebo také  $L_2$  vzdálenost (Minkowského  $k$ -metrika, kde  $k = 2$ ) mezi dvěma body  $\mathbf{x} = (x_0, x_1, \dots, x_{p-1})$  a  $\mathbf{y} = (y_0, y_1, \dots, y_{p-1})$  z  $\mathbb{R}^p$  je definována jako:

$$\|\mathbf{x} - \mathbf{y}\|_2 = \sqrt{\sum_{i=0}^{p-1} (x_i - y_i)^2} . \quad (4.14)$$

#### 4.1.6 Výpočet úhlu mezi 2 vektory

Úhel  $\theta$  mezi 2 nenulovými vektory  $u$  a  $v$  se počítá jako:

$$\theta = \arccos\left(\frac{u}{\|u\|} \cdot \frac{v}{\|v\|}\right) , \quad (4.15)$$

kde  $\frac{u}{\|u\|}$ ,  $\frac{v}{\|v\|}$  jsou jednotkové vektory vzniklé skalárním dělením vektorů  $u$ ,  $v$  jejich normami  $\|u\|$  a  $\|v\|$ .

### 4.1.7 Rotace

Rotace kolem obecného bodu o úhel  $\theta$  se skládá ze základních transformací – posunutí bodu do počátku, rotace kolem počátku a posunutí zpět, tedy:

$$\begin{aligned} M &= T(x, y) \cdot R \cdot T(-x, -y) \\ &= \begin{bmatrix} \cos(\theta) & -\sin(\theta) & -x \cdot \cos(\theta) + y \cdot \sin(\theta) + x \\ \sin(\theta) & \cos(\theta) & x \cdot \sin(\theta) + y \cdot \cos(\theta) - y \\ 0 & 0 & 1 \end{bmatrix}, \end{aligned} \quad (4.16)$$

kde  $R = \begin{bmatrix} \cos(\theta) & -\sin(\theta) & 0 \\ \sin(\theta) & \cos(\theta) & 0 \\ 0 & 0 & 1 \end{bmatrix}$  značí rotaci kolem počátku souřadnicové soustavy a  $T(x, y) = \begin{bmatrix} 1 & 0 & x \\ 0 & 1 & y \\ 0 & 0 & 1 \end{bmatrix}$  posunutí o  $(x, y)^T$ .

## 4.2 Extrakce příznaků, klasifikace

V této sekci jsou uvedeny metody použité k extrakci příznaků z fotografií a jejich následné klasifikaci.

### 4.2.1 Architektury konvolučních neuronových sítí

Tato sekce popisuje jednotlivé architektury konvolučních neuronových sítí, jež byly v této práci využity jako extraktory příznaků z fotografií.

#### VGGNet

Architektura sítě VGG [21] byla poprvé představena v roce 2014 skupinou Visual Geometry Group z Oxfordské univerzity v soutěži ImageNet Large Scale Visual Recognition Challenge (ILSVRC) 2014, ve které tato architektura získala první místo v kategorii lokalizace objektu v obraze a druhé místo v kategorii klasifikace.

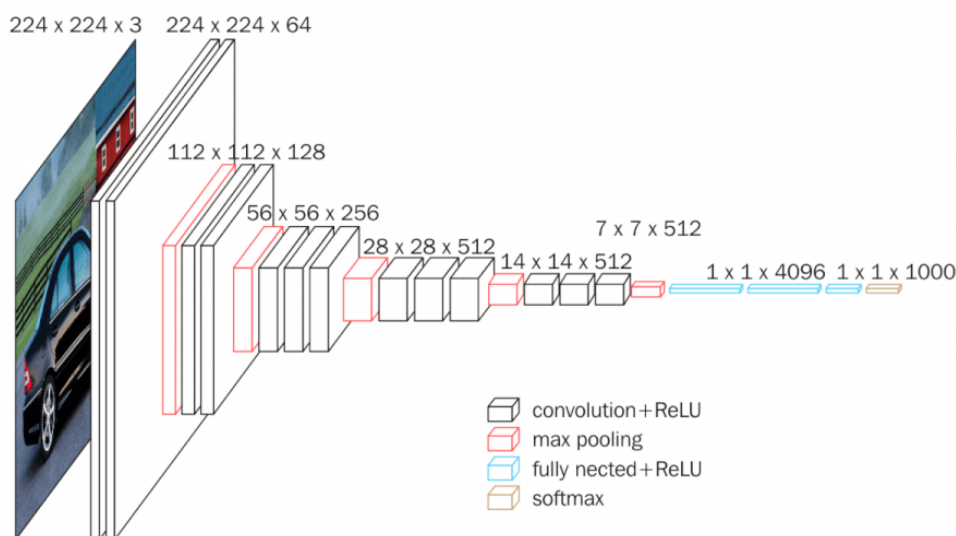
Skupina VGG zkoumala vliv hloubky konvoluční neuronové sítě na její přesnost a experimentovala tak s různými architekturami skládajícími se z 11–19 vrstev (8–16 konvolučních vrstev + 3 plně propojených vrstev). Z těchto architektur je velice známá síť VGG19 (obsahující 19 vrstev) a VGG16 (16 vrstev), jež byla použita i v této práci a proto je zde blíže popsána:

Vstupem této architektury je RGB obraz fixních rozměrů  $224 \times 224$  pixelů, na kterém se provádí jediné předzpracování – odečtení průměrné RGB hodnoty vypočítané na trénovacích datech od každého pixelu. Tento obraz je dále



zpracováván v bloku několika za sebou naskládaných konvolučních vrstev<sup>18</sup>, využívajících velmi malá konvoluční jádra o rozměrech  $3 \times 3$ , a jednou max-pooling vrstvou. Velikost konvolučního kroku i velikost odsazení je nastavena na 1 pixel a max pooling je prováděn pomocí filtru o rozměrech  $2 \times 2$  a krokem velikosti 2. Bloky konvolučních a pooling vrstev jsou následovány třemi plně propojenými vrstvami a zakončeny vrstvou využívající aktivační funkci softmax. Celá tato architektura je znázorněna na obrázku 4.9.

Nelinearita je do architektury vnášena pomocí aktivační funkce ReLU, jež je obsažena ve všech skrytých vrstvách.



Obrázek 4.9: Ukázka architektury konvoluční neuronové sítě VGG16, převzato z [22]

## GoogLeNet/Inception

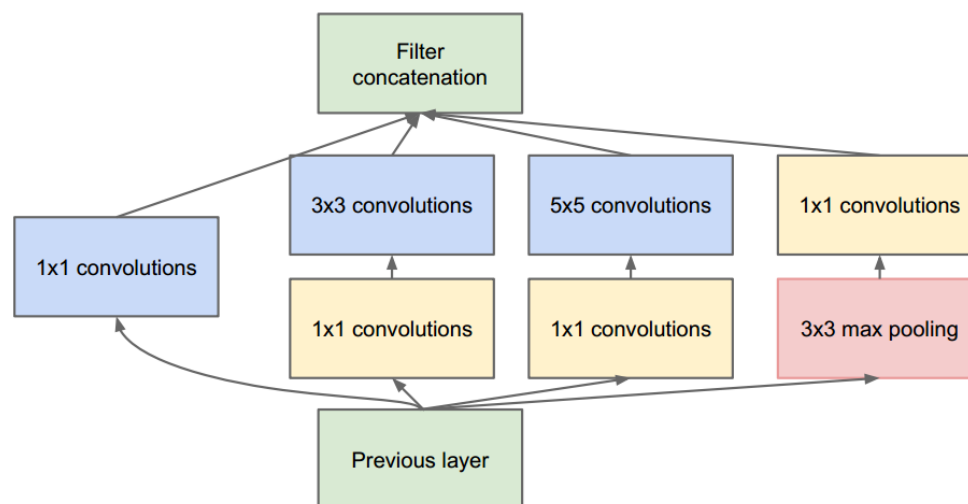
První místo v kategorii klasifikace i kategorii detekce objektů a druhé místo v kategorii lokalizace objektu v obraze získala, na již zmiňované soutěži ILSVRC 2014, architektura GoogLeNet (InceptionV1) [23] společnosti Google. Tato architektura se od ostatních lišila tím, že nešla pouze do hloubky, ale také do šířky – využívala totiž na stejné úrovni konvoluční jádra různých velikostí.

Inception architektura je tedy CNN skládající se z tzv. Inception modulů zřetězených za sebou a z mezi ně příležitostně vložených max-pooling vrstev s krokem 2 pro snížení rozlišení mřížky na polovinu.

<sup>18</sup>Díky naskládání 3 konvolučních vrstev s jádrem velikosti  $3 \times 3$  za sebe lze dosáhnout efektivní oblasti vnímání o velikost  $7 \times 7$ . Výhodou oproti využití jedné konvoluční vrstvy velikosti  $7 \times 7$  je snížení počtu parametrů a začlenění 3 nelineárních ReLU funkcí místo jedné.

#### 4. POUŽITÉ METODY

V Inception modulu na vstupu probíhá paralelně konvoluce s filtry o rozměrech  $1 \times 1$ ,  $3 \times 3$ ,  $5 \times 5$  a max pooling. Pro snížení výpočetní náročnosti je před  $3 \times 3$ ,  $5 \times 5$  konvolucí a po max-poolingu provedena ještě konvoluce  $1 \times 1$ . Výstupy těchto částí jsou následně zřetězeny a poslány do dalšího modulu. Struktura Inception modulu je znázorněna na obrázku 4.10.



Obrázek 4.10: Ukázka Inception modulu, na kterém je založena architektura konvoluční neuronové sítě GoogLeNet, převzato z [23]

Vítězná síť GoogLeNet je 22 vrstev hluboká (27 včetně pooling vrstev)<sup>19</sup>, na konci posledního Inception modulu používá tzv. global average pooling a s problémem mizejícího gradientu (vanishing gradient problem) se při trénování dokáže vypořádat díky 2 pomocným klasifikátorům s funkcí softmax.

Tato práce využívá již vylepšenou verzi sítě GoogLeNet – InceptionV3 [24], jež pro zefektivnění konvolucí ve smyslu výpočetní složitosti a snížení ztráty informací (representational bottleneck) používá mimo jiné metody faktorizace (konvoluce  $5 \times 5$  je faktorizována do dvou konvolucí  $3 \times 3$  či obecně jsou faktorizovány konvoluce  $n \times n$  na konvoluce  $1 \times n$  a  $n \times 1$ ), rozšiřuje Inception modul místo jeho prohlubování a dále i optimalizuje pomocné klasifikátory, čímž snižuje možnost přeučení sítě při jejím trénování.

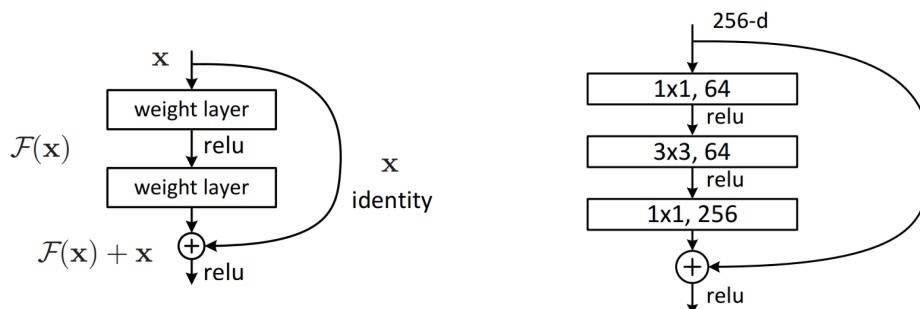
#### ResNet

V roce 2015 byla týmem ze společnosti Microsoft představena reziduální neuronová síť ResNet [25], jež byla s hloubkou až 152 vrstev výrazně hlubší než všechny doposud existující sítě. Tato architektura byla velmi úspěšná – stala se vítězem soutěže ILSVRC 2015 v kategoriích klasifikace, lokalizace i detekce

<sup>19</sup>Celkově ale obsahuje cca 100 nezávislých stavebních bloků.

a také vítězem soutěže MS COCO Image Captioning Challenge 2015 v kategoriích detekce a segmentace.

Architektura ResNet může dosahovat takto velké hloubky a nepotýkat se při tom s problémem mizejícího gradientu díky využití tzv. reziduálních bloků (viz. obrázek 4.11), které umožňují některé vrstvy sítě přeskočit a zachovat tak sílu šířícího se signálu napříč vrstvami.



Obrázek 4.11: **Vlevo:** Obecný popis reziduálního bloku konvoluční neuronové sítě ResNet, kde  $x$  značí vstupní vektor a  $\mathcal{F}(x)$  reziduální funkci, **Vpravo:** „Bottleneck“ reziduální blok, jež je využíván u sítí ResNet majících 50 vrstev a více, převzato z [25]

V těchto reziduálních blocích jsou mezi několika vrstvami vedené zkratky (skip connections), jež provádí „identity mapping“, což znamená, že jejich výstupy jsou jednoduše přidány k výstupům z ostatních, za sebe poskládaných, vrstev bez využití dodatečných parametrů či výpočetní složitosti. Celkový výstup reziduálního bloku je tedy:

$$\mathcal{H}(x) = \mathcal{F}(x) + x , \quad (4.17)$$

kde  $\mathcal{F}(x)$  značí reziduální funkci a  $x$  vstupní vektor reziduálního bloku a zároveň výstup dané zkratky. Blok se nazývá reziduální z toho důvodu, že se při trénování sítě snaží sice naučit skutečný výstup  $\mathcal{H}(x)$ , ale přeskočené vrstvy reziduální sítě se reálně učí jen zbytek (residual)  $\mathcal{F}(x)$ , pro který platí:

$$\mathcal{F}(x) = \mathcal{H}(x) - x . \quad (4.18)$$

U ResNet architektur hlubokých více než 50 vrstev je v reziduálním bloku použita pro snížení výpočetní náročnosti stejná technika jako v Inception modulu (popsaném v předešlé sekci) a to přidání  $1 \times 1$  konvoluce na začátek a konec bloku, čímž lze zredukovat počet parametrů sítě, aniž by to snížilo její výkonnost.

V této práci byla jako extraktor příznaků využita<sup>20</sup> právě jedna z takto hlubokých sítí – ResNet50 obsahující 50 vrstev.

<sup>20</sup>Dále jako backbone Mask R-CNN byla v této práci využita architektura se 101 vrstvami – ResNet101.

### 4.2.2 Logistická regrese

Logistická regrese je navzdory svému názvu metoda určená pro klasifikaci. Stejně jako u lineární regrese (viz. sekce 4.1.3), rozhodnutí modelu logistické regrese se konstruuje pomocí lineární kombinace příznaků. Místo diskrétního problému se u tohoto modelu však řeší problém spojitý a nejedná se tedy o predikci hodnoty vysvětlované proměnné  $Y$ , nýbrž o predikci její pravděpodobnosti.

Model pro odhad této pravděpodobnosti závislé na hodnotách příznaků  $\mathbf{x} = (x_0, x_1, \dots, x_p)$  a koeficientů  $\boldsymbol{\omega} = (\omega_0, \omega_1, \dots, \omega_p)$  má tvar:

$$P(Y = 1 \mid \mathbf{x}, \boldsymbol{\omega}) = \frac{e^{\boldsymbol{\omega}^T \mathbf{x}}}{1 + e^{\boldsymbol{\omega}^T \mathbf{x}}} , \quad (4.19)$$

ve kterém je s dosazenou hodnotou lineárního výrazu využit speciální případ logistické funkce – sigmoida<sup>21</sup>:

$$f(x) = \frac{1}{1 + e^{-x}} = \frac{e^x}{1 + e^x} . \quad (4.20)$$

Učení modelu – výběr hodnot parametrů  $\boldsymbol{\omega}$  na základě trénovacích dat, u kterých je (kromě příznaků) známa i hodnota  $Y$ , je prováděn metodou MLE<sup>22</sup>. [26]

### 4.2.3 Metoda nejbližších sousedů

Algoritmus  $k$ -nejbližších sousedů ( $k$ -NN<sup>23</sup>) je metoda supervizovaného učení (neboli učení s učitelem), při které v podstatě neprobíhá učení modelu, protože trénovací data vlastně sama o sobě naučeným modelem jsou.

Při predikci hodnoty vysvětlované proměnné pro datový bod  $x$  je v trénovacích datech nalezeno  $k$  bodů, jež jsou od  $x$  v nejbližší vzdálenosti – nejběžněji se využívá Eukleidovská vzdálenost (definovaná již v sekci 4.1.5). Predikce je pak v případě klasifikace založena na nejčastější hodnotě vysvětlované proměnné mezi těmito  $k$  body.

### 4.2.4 Ladění hyperparametrů

Hyperparametry jsou přizpůsobitelné parametry modelu, které určují jeho tvar, komplexitu a které umožňují řídit proces trénování modelu. Laděním hyperparametrů se rozumí (za pomoci validačních dat) hledání optimálních hyperparametrů, pro něž model vykazuje nejnižší chybu.

Obvykle se proto data náhodně rozdělují na 3 podmnožiny – trénovací, validační a testovací. Nejčastěji se data dělí v poměru 80 : 20 (nebo také

---

<sup>21</sup>Sigmoida je obvyklou volbou, lze ale využít i jiná funkce, jejíž obor hodnot je podmnožinou intervalu  $[0,1]$ .

<sup>22</sup>Maximum likelihood estimation neboli Metoda maximální věrohodnosti.

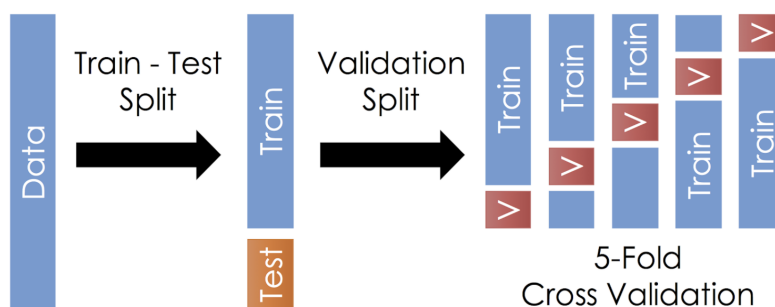
<sup>23</sup>Zkratka  $k$ -NN pochází z anglického  $k$ -nearest neighbors.

75 : 25 či 70 : 30)<sup>24</sup> na trénovací a testovací množinu a následně je takto získaná trénovací množina opět v tomtéž poměru rozdělena na trénovací a validační množinu.

Model je pro různé hodnoty hyperparametrů naučen na trénovací množině a jeho chyba je měřena na množině validační. Optimální hyperparametry jsou zvoleny dle této nejmenší chyby a tedy nejvyšší přesnosti. Poté je kvalita modelu se zvolenými nejlepšími hyperparametry posouzena na testovacích datech, které se doposud na trénování modelu nijak nepodílely a lze je tak použít jako odhad skutečné chyby modelu.

### Křížová validace

Dalším možným přístupem k ladění hyperparametrů modelu je tzv. křížová validace ( $k$ -fold cross-validation). Při ní se data nerozdělují na 3 podmnožiny, ale pouze na 2 – trénovací a testovací. Trénovací množina je pak dále náhodně rozdělena na  $k$  stejně velkých částí, přičemž platí, že  $k$  je rovno alespoň 2 a nejvýše počtu bodů v trénovací množině. Z těchto  $k$  částí je vždy postupně jedna považována za validační a zbylých  $k - 1$  za trénovací, viz. 4.12.



Obrázek 4.12: Ukázka rozdělení dat při křížové validaci, převzato z [27]

Model je tedy pro různé hodnoty hyperparametrů natrénován postupně na všech  $k - 1$  částech tvořících trénovací data a vyhodnocen na zbývající validační části – na validační části je vypočtena chyba modelu  $e$ . Tento proces je opakován  $k$ krát, aby se jednotlivé části prostřídaly a každá z nich jednou sloužila jako validační. Po všech opakováních je vypočítána průměrná cross-validační chyba modelu:

$$\frac{1}{k} \sum_{i=1}^k e_i . \quad (4.21)$$

Na základě nejmenší cross-validační chyby je následně možné vybrat optimální hodnoty hyperparametrů a kvalitu modelu s tímto nastavením na závěr opět posoudit na testovací množině.

<sup>24</sup>Neexistuje žádný obecně optimální poměr velikostí těchto množin, rozdělení je třeba upravit v závislosti na použitém datasetu a jeho velikosti.



---

# Implementace prototypu pro rozpoznávání želv

Tato kapitola se zabývá výběrem nástrojů použitých k implementaci prototypu pro rozpoznávání želv, popisuje způsob trénování modelu Mask R-CNN na fotografiích plastronů z datasetu zapůjčeném z ÚTIA AVČR, detekci plastronu na fotografii a jeho následnou klasifikaci.

## 5.1 Použité nástroje

V této sekci je popsán zvolený programovací jazyk a důležité knihovny a frameworky použité pro tvorbu této práce.

### 5.1.1 Python + Jupyter notebook

Prototyp je vytvořen v programovacím jazyce Python, jenž se v posledních letech drží na prvním místě žebříčku nejpoužívanějších programovacích jazyků pro použití v oblasti strojového učení [28]. K implementaci byl tento jazyk zvolen především z důvodu možnosti využití velkého množství knihoven a frameworků na něm postavených a také pro jeho jednoduchou syntaxi a čitelnost. Pro přehlednější prezentaci kódu byla použita interaktivní webová aplikace Jupyter notebook [29].

#### Knihovny a frameworky

- **imgaug** (0.4.0)<sup>25</sup> je knihovna sloužící k augmentaci dat. Nabízí širokou škálu augmentačních technik a umožňuje je snadno kombinovat a provádět v náhodném pořadí. [30]

---

<sup>25</sup>Čísla v závorkách značí verze jednotlivých knihoven/frameworků použitých v této práci.

- **Keras** (2.3.1) je framework vyšší úrovně abstrakce vystavěný na TensorFlow knihovně a slouží pro práci s (hlubokými) neuronovými sítěmi. [31]
- **NumPy** (1.16.6) je knihovna poskytující infrastrukturu pro práci s vektory, maticemi, vícerozměrnými poli a nabízí také řadu matematických funkcí. [32]
- **OpenCV** (4.5.1) je multiplatformní knihovna využívaná pro počítačové vidění a zpracování obrazu. [33]
- **Pandas** (0.25.3) je knihovna nabízející množství datových struktur sloužících pro analýzu dat a jejich snadnou manipulaci. [34]
- **Scikit-image** (0.16.2) je kolekce algoritmů sloužící pro zpracování obrazu, je vystavěna na matematických a vědeckých knihovnách NumPy, SciPy a matplotlib. [35]
- **Scikit-learn** (0.24.1) je jednoduchá a efektivní knihovna určená pro prediktivní analýzu dat. Je stejně jako Scikit-image vystavěna na matematických a vědeckých knihovnách NumPy, SciPy a matplotlib. [36]
- **TensorFlow** (2.0.0) je knihovna využívající Keras API jako intuitivní nadstavbu pro hluboké učení. [37]
- **matterport/Mask\_RCNN** je repozitář obsahující implementaci modelu Mask R-CNN [17] (s několika menšími odchylkami pro zjednodušení kódu) vystavěný na Python 3, Keras a TensorFlow. Model je založen na architektuře ResNet101 a jako extraktor příznaků používá tzv. Feature Pyramid Network (FPN). Tento repozitář obsahuje různé ukázky modelů s již natrénovanými vahami pro detekci objektů, např. na datasetu MS COCO<sup>26</sup> [38], i návody pro trénování vah na datasetu vlastním, což usnadňuje porozumění této implementace a je díky tomu možné snadno model upravit pro vlastní potřeby. Samotná implementace modelu je pak obsažena v souborech *model.py*, *utils.py* a *config.py*. [39]

### 5.2 Trénování modelu Mask R-CNN

Aby bylo možné využít model Mask R-CNN pro problematiku rozpoznávání želv, bylo nejprve nutné ho natrénovat na příslušných datech. V tomto případě bylo využito 121 fotografií plastronů z datasetu České inspekce životního prostředí<sup>27</sup> zapůjčeného z ÚTIA AVČR, které byly rozděleny do 2 částí – na 100

---

<sup>26</sup>Zkratka MS COCO pochází z anglického Microsoft Common Objects in Context.

<sup>27</sup>K trénování (a validaci) modelu Mask R-CNN byly využity jen ty fotografie, na kterých nebyla v další části práce testována klasifikace, tedy takové fotografie plastronů, jejichž třída byla v datasetu zastoupena pouze jedenkrát.



trénovacích a 21 validačních fotografií. Všechny tyto fotografie bylo následně třeba opatřit anotacemi. K tomu byl použit jednoduchý nástroj VGG Image Annotator (VIA) [40], který běží jako offline aplikace ve webovém prohlížeči a nevyžaduje žádnou instalaci ani nastavení.

Na každé fotografii tak bylo vyznačeno 8 oblastí spadajících do 4 kategorií: plastron, centrální šev, anální štítky<sup>28</sup> a 5 míst připojení postranních švů na centrální šev, viz. obrázek 5.1. Podobné značení bylo využito již v diplomové práci [3] Zdeňka Svatoně.

Takto vytvořené anotace byly (v případě využití nástroje VIA) pro všechny fotografie z dané části (trénovací/validační) zapsány do jednoho JSON<sup>29</sup> souboru, kde je maska každé oblasti uložena ve formě jednotlivých bodů tvořících polygon.

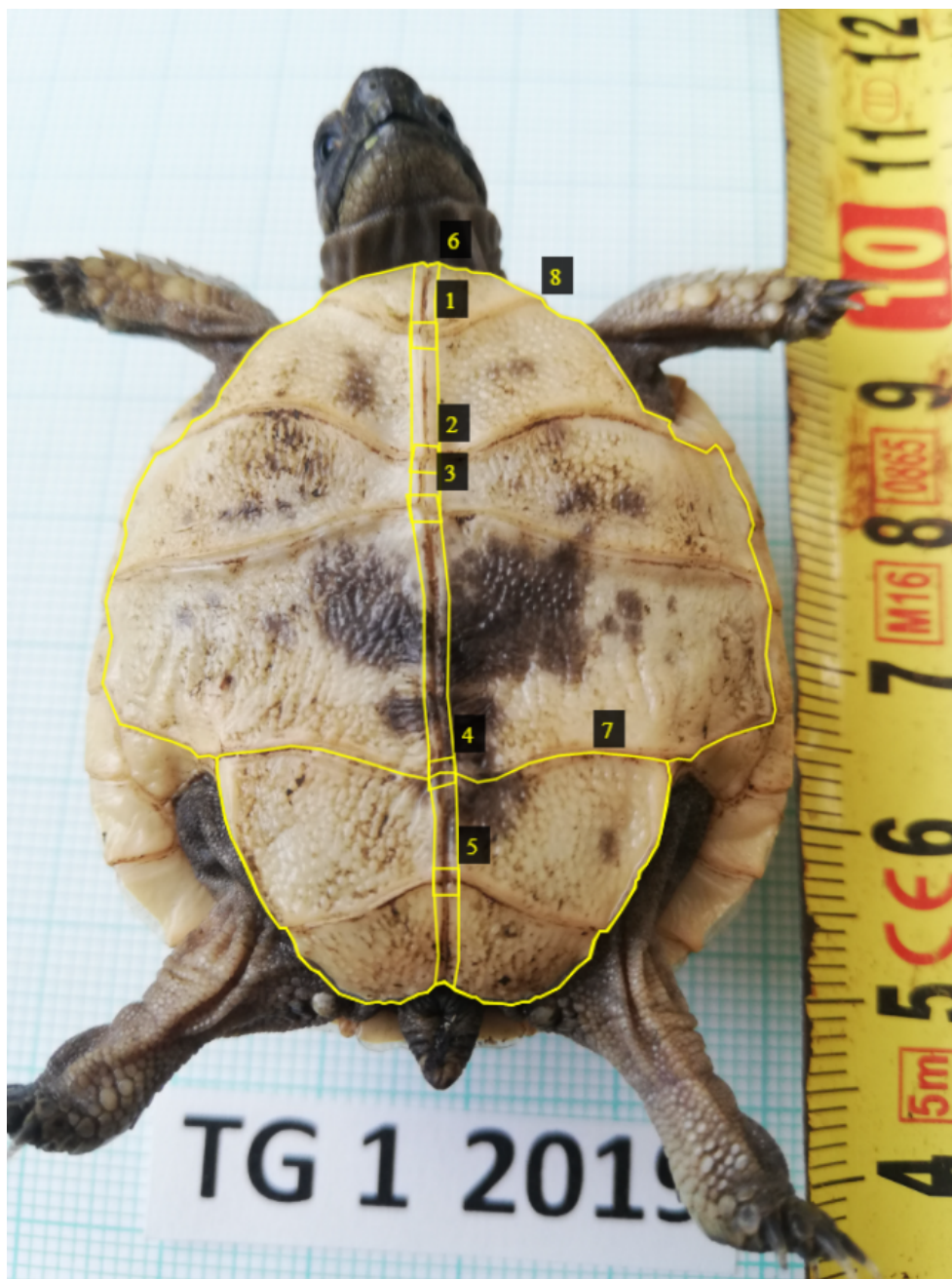
Po pořízení anotací již bylo možné přistoupit k samotnému trénování modelu, které bylo prováděno v prostředí Google Colaboratory, jež poskytuje bezplatný výpočetní výkon. Bylo tak možné spustit kód v Jupyter notebooku ve virtuálním stroji za použití 12GB NVIDIA Tesla K80 GPU běžícího nepřetržitě až 12 hodin a to opět bez nutnosti složitého nastavení.

Trénování modelu bylo uskutečněno metodou tzv. transfer learningu. Model nebyl trénován od nuly, ale jako výchozí bod byly využité již natrénované váhy na datasetu MS COCO [38]. Tento dataset se skládá z 328 tisíc fotografií a obsahuje 80 různých kategorií objektů, proto váhy na něm natrénované umí rozpoznávat spoustu typů příznaků a jsou vhodné jako základ trénování modelu zaměřujícího se na jinou problematiku.

Model byl trénován vícekrát s experimentálně navolenými parametry a také s nebo bez použití augmentace. Nejlepších výsledků dosahoval model natrénovaný na datech augmentovaných pomocí horizontálního a vertikálního překlopení obrazu, rotací a jemné úpravě jasu a kontrastu, viz. příložený notebook *train\_plastron\_model.ipynb*. Nejpodstatnějšími augmentačními technikami byly právě zmíněné rotace a to hlavně z toho důvodu, že plastrony na fotografiích v datasetu ČIŽP jsou orientované převážně směrem vzhůru či směrem doleva a kvůli tomu docházelo k přeučení modelu na toto nastavení a při natočení plastronu jiným směrem nastával problém v záměně análních (anal + femoral) štítků za štítky chytací (gular) + končetinové (humeral) a tím později i k nesprávné detekci plastronu na fotografii. Také takto přeučený model špatně rozpoznával oblast centrálního švu na plastronech, které nebyly na fotografii umístěny čistě v horizontální či vertikální poloze.

<sup>28</sup>Konkrétně se jedná dohromady o štítky stehenní (femoral) a anální (anal). V anotacích pro zjednodušení nazvány souhrnně jako anální.

<sup>29</sup>Zkratka JSON pochází z anglického JavaScript Object Notation, tedy JavaScriptový objektový zápis.



Obrázek 5.1: Ukázka anotací pořízených nástrojem VIA na identifikační fotografii z osobního archivu chovatele Petra Petráše, kde čísla oblastí 1–5 znázorňují místa připojení postranních švů na centrální šev, 6 značí centrální šev, 7 anální štítky a 8 plastron

V tabulce 5.1 je znázorněna vyzorovaná úspěšnost detekce jednotlivých oblastí na fotografiích z datasetu České inspekce životního prostředí a datasetu Petra Petráře pomocí nejlépe natrénovaného modelu Mask R-CNN.

U datasetu P. Petráře je patrná nižší úspěšnost detekce všech 5 míst připojení postranních švů na centrální šev z důvodu velké rozmanitosti nastavení fotografií. Pro finální detekci plastronu popsanou v následující sekci však není nutné znát všech 5 těchto míst připojení, postačující jsou pouze místa 2, což všechny fotografie z datasetu P. Petráře splňují<sup>30</sup>.

Tabulka 5.1: Počet správně detekovaných oblastí plastronů, centrálních švů, análních štítků a všech 5 míst připojení postranních švů na centrální šev v datasetech České inspekce životního prostředí a Petra Petráře a jejich procentuální úspěšnost

dataset	ČIŽP		P. Petráře	
	počet	úspěšnost	počet	úspěšnost
plastron	60/60	100 %	50/50	100 %
centrální šev	59/60	98,33 %	48/50	96 %
anální štítky	59/60	98,33 %	50/50	100 %
5 míst připojení švů	57/60	95 %	23/50	46 %

### 5.3 Detekce plastronu na fotografii

Na fotografii byly nejprve pomocí natrénovaného modelu Mask R-CNN detekovány všechny oblasti, viz. obrázek 5.2.

Středovými body nalezených míst připojení postranních švů na centrální šev byla následně proložena přímka představující linii centrálního švu<sup>31</sup>. V případě správné detekce oblasti centrálního švu k tomu byla použita metoda lineární regrese či metoda RANSAC v případě opačném. Díky metodě RANSAC bylo případně možné odfiltrovat nesprávně detekovaná místa připojení postranních švů na centrální šev a zabránit tak vychýlení směru prokládané přímky.

Dále byl nalezen směrový vektor této přímky na základě pozice oblasti análních štítků<sup>32</sup> a vypočítán orientovaný úhel mezi tímto vektorem a vektorem představujícím osu  $y$ . Dle tohoto úhlu byla poté určena příslušná ro-

<sup>30</sup>V datasetu Petra Petráře bylo detekováno všech 5 míst připojení postranních švů na centrální šev na 23 fotografiích, 4 místa na 20 fotografiích, 3 místa na 6 fotografiích a 2 místa pouze na 1 fotografii z celkového počtu 50 fotografií.

<sup>31</sup>Toto proložení bylo možné uskutečnit pouze za předpokladu, že byla k dispozici alespoň 2 detekovaná místa připojení postranních švů na centrální šev. Pokud jich bylo méně, přímka byla prokládána i mezi středovými body ostatních oblastí.

<sup>32</sup>Pokud oblast análních štítků nebyla detekována, bylo dále předpokládáno, že hlava želvy směřuje vlevo – jedná se o nejčastější nastavení v datasetu ČIŽP.

## 5. IMPLEMENTACE PROTOTYPU PRO ROZPOZNÁVÁNÍ ŽELV



Obrázek 5.2: Ukázka všech oblastí (a jejich pravděpodobností) správně detekovaných pomocí natrénovaného modelu Mask R-CNN na fotografii plastronu z osobního archivu chovatele Petra Petráse – zeleně je zde znázorněna oblast plastronu, žlutě oblast análních štítků, červeně oblast centrálního švu a modře oblasti míst připojení postranních švů na centrální šev



tační matice a fotografie i masky detekovaných oblastí byly pomocí ní otočeny a zvětšeny tak, aby nedocházelo k jejich nechtěnému oříznutí.

Následně byly z otočených masek oblastí přepočítány souřadnice jejich ohraničujících rámečků a pomocí nich byly z dané fotografie vyříznuty oblasti plastronu a centrálního švu. Podoba vyříznutých částí fotografie je znázorněna na obrázku 5.3.



Obrázek 5.3: Ukázka extrahovaných oblastí plastronu a centrálního švu – první obrázek je určen ohraničujícím rámečkem plastronu, druhý pomocí masky plastronu a třetí obrázek, obsahující pouze centrální šev, je vyříznut jako prostřední pětina obrázku prvního

Tímto způsobem byly úspěšně vyříznuty oblasti plastronů (ohraničujících rámečků i masek) a centrálních švů ze všech 60 fotografií z datasetu České inspekce životního prostředí i ze všech 50 fotografií z datasetu Petra Petráse. Na obou těchto datasetech bylo tedy dosaženo 100% úspěšnosti automatické detekce plastronu a všechny vyříznuté části tak mohly být v další části této práce použity ke klasifikaci.

Celý průběh této automatické detekce plastronu na fotografii je popsán (i graficky) a implementován v přiloženém notebooku *predict\_plastron.ipynb*.

## 5.4 Extrakce příznaků a jejich klasifikace

Z bakalářské práce [2] A. Majtánové vyplývá, že nejúspěšnějším extraktorem příznaků z fotografií ručně ořezaných na oblast plastronu se jeví konvoluční neuronová síť VGG16 společně s logistickou regresí jako klasifikátorem v případě tzv. multiclass klasifikace (predikce ID želvy) a s klasifikátorem  $k$ -nejbližších sousedů v případě klasifikace binární (porovnání 2 jedinců želvy),

viz. sekce 2.3. Proto i tato práce byla zaměřena na konvoluční neuronové sítě, konkrétně na populární architektury VGG16, InceptionV3 a ResNet50.

Všechny tyto architektury CNN byly využité bez plně propojených vrstev a s váhami předtrénovanými na datasetu ImageNet [41]. Fotografie z datasetů byly nejprve vhodně předzpracovány dle potřeb jednotlivých architektur CNN<sup>33</sup> a pro následnou klasifikaci se využívaly už pouze vyextrahované vektory příznaků společně s příznakem ID, znázorňujícím příslušnost želvy k dané třídě. Tento příznak je součástí názvu všech fotografií v použitých datasetech, pro jeho získání tedy stačilo nalézt v názvu fotografie číselnou hodnotu<sup>34</sup>.

### Multiclass klasifikace

Pro predikci ID želv byla tedy sestavena tabulka (struktura DataFrame z knihovny Pandas) skládající se ze 3 sloupců obsahujících názvy fotografií, ID želv a vyextrahované vektory příznaků – označené jako *Filename*, *ID* a *Features*, přičemž pro trénování klasifikačního modelu byly využity pouze sloupce *ID* a *Features* a sloupec *Filename* sloužil pouze k případnému dohledání zpracovávané fotografie.

Data v tabulce byla rozdělena na trénovací a testovací množiny a pomocí křížové validace byly nalezeny nejlepší hyperparametry klasifikátoru logistické regrese.

### Binární klasifikace

Pro porovnání 2 jedinců želv a posouzení toho, zda jsou stejní či rozdílní, vznikla další tabulka spojením všech různých dvojic záznamů z tabulky předchozí – určené k multiclass klasifikaci. Tato tabulka tak sestávala z 6 sloupců: *Filename1*, *ID1*, *Features1*, *Filename2*, *ID2*, *Features2*, ke kterým byl přidán další sloupec *Distance* značící Eukleidovskou vzdálenost mezi vektory příznaků ve sloupcích *Features1* a *Features2* a sloupec *Class* s hodnotou 1, pokud se rovnaly hodnoty ve sloupcích *ID1* a *ID2* či s hodnotou 0, pokud byly ID těchto jedinců odlišné.

Následně byla tato tabulka vybalancována tak, aby v ní byl stejný počet záznamů s hodnotami 0 a 1 ve sloupci *Class* – tedy aby se počet dvojic stejných a rozdílných jedinců shodoval.

---

<sup>33</sup>V případě použití architektur VGG16 a ResNet50 byla velikost fotografií přizpůsobena rozměrům  $244 \times 244$  pixelů, obraz byl převeden z barevného modelu RGB do BGR a na každém barevném kanálu bylo provedeno tzv. nulové centrování, tedy odečtení průměrné hodnoty barevného kanálu dle datasetu ImageNet nebo v případě použití architektury InceptionV3 byla velikost fotografií nastavena na  $299 \times 299$  pixelů a hodnoty všech pixelů byly normalizovány do intervalu od -1 do 1.

<sup>34</sup>Například název *Tg001a.jpg* značí, že daná želva druhu *Testudo graeca* na fotografii patří do 1. třídy – její ID je 1. Díky písmenu *a* je dále možné rozlišovat mezi jednotlivými fotografiemi jedinců z této třídy.

Pro trénování klasifikačního modelu  $k$ -nejbližších sousedů byly použity pouze sloupce *Distance* a *Class*, data v nich byla opět rozdělena na trénovací a testovací a ladění hyperparametrů probíhalo pomocí křížové validace.

V případě multiclass i binární klasifikace byla úspěšnost klasifikačního modelu vyhodnocována na testovací množině o  $n$  záznamech dle:

$$accuracy(y, \hat{y}) = \frac{1}{n} \sum_{i=0}^{n-1} 1(\hat{y}_i = y_i) , \quad (5.1)$$

kde  $\hat{y}$  značí predikovanou hodnotu a  $y$  jí odpovídající hodnotu reálnou.

Extrakce příznaků a jejich klasifikace je implementována v příloženém notebooku *plastron\_classification.ipynb* a dosažené výsledky na jednotlivých datasetech jsou uvedeny v následující kapitole.





---

## Experimenty, diskuze

Experimenty probíhaly na 60 fotografiích z datasetu České inspekce životního prostředí a na 50 fotografiích z datasetu chovatele Petra Petráse. K extrakci příznaků a jejich klasifikaci však nebyly použity původní fotografie, ale jejich výřezy získané pomocí automatické detekce v sekci 5.3, tedy vyříznutý ohraničující rámeček plastronu, maska plastronu a centrální šev (podslůžky *plastron\_bbox*, *plastron\_mask* a *central\_seam*).

Jako extraktory příznaků z fotografií byly testované 3 architektury konvolučních neuronových sítí – VGG16, InceptionV3 a ResNet50. V případě multiclass klasifikace byla jako klasifikátor použita metoda logistické regrese a v případě binární klasifikace metoda nejbližších sousedů. Data byla rozdělena na trénovací a testovací množiny v poměrech<sup>35</sup> 70 : 30 a 80 : 20 a hyperparametry klasifikačních modelů byly laděny pomocí křížové validace. Pouze v rámci multiclass klasifikace při použití datasetu ČIŽP, kde pro každou třídu (ID želvy) existují jenom 2 záznamy, byla data rozdělena v poměru 50 : 50 a úspěšnost modelu pro dané hodnoty hyperparametrů byla vyhodnocována na celé trénovací množině.

Na obou datasetech (jejich výřezech) a za použití výše popsaných extraktorů proběhlo cca 50 měření. Tabulky 6.1, 6.2, 6.3 a 6.4 zachycují rozsah hodnot získaných během jednotlivých měření (nejnižší naměřená úspěšnost – nejvyšší naměřená úspěšnost klasifikace) a jejich průměr. Čísla v tabulkách jsou zaokrouhlena na 2 desetinná místa.

---

<sup>35</sup>V tabulkách výsledků jsou uvedeny souhrnně úspěšnosti rozdělení dat ve všech těchto poměrech.

Tabulka 6.1: Úspěšnost multiclass klasifikace na fotografiích z datasetu České inspekce životního prostředí

dataset	VGG16		InceptionV3		ResNet50	
ČÍŽP	rozsah (%)	průměrná úspěšnost (%)	rozsah (%)	průměrná úspěšnost (%)	rozsah (%)	průměrná úspěšnost (%)
plastron_bbox	50,00–66,67	58,37	43,33–63,33	54,58	50,00–66,67	58,75
plastron_mask	66,67–76,67	71,42	43,33–60,00	52,42	56,67–73,33	63,42
central_seam	40,00–70,00	58,08	26,67–40,00	32,92	30,00–50,00	40,33

Tabulka 6.2: Úspěšnost multiclass klasifikace na fotografiích z datasetu Petra Petráše

dataset	VGG16		InceptionV3		ResNet50	
P. Petráše	rozsah (%)	průměrná úspěšnost (%)	rozsah (%)	průměrná úspěšnost (%)	rozsah (%)	průměrná úspěšnost (%)
plastron_bbox	93,33–100	99,17	90,00–100	99,08	80,00–100	96,33
plastron_mask	86,67–100	98,00	93,33–100	99,67	93,33–100	99,67
central_seam	93,33–100	99,50	90,00–100	98,58	86,67–100	99,00

Tabulka 6.3: Úspěšnost binární klasifikace na fotografiích z datasetu České inspekce životního prostředí

dataset ČÍŽP	VGG16		InceptionV3		ResNet50	
	rozsah (%)	průměrná úspěšnost (%)	rozsah (%)	průměrná úspěšnost (%)	rozsah (%)	průměrná úspěšnost (%)
plastron_bbox	61,11–100	82,22	58,33–91,67	72,50	50,00–83,33	71,88
plastron_mask	66,67–100	83,47	50,00–91,67	73,06	50,00–83,33	70,83
central_seam	38,89–83,33	61,32	50,00–83,33	65,90	50,00–88,89	72,64

Tabulka 6.4: Úspěšnost binární klasifikace na fotografiích z datasetu Petra Petráše

dataset P. Petráše	VGG16		InceptionV3		ResNet50	
	rozsah (%)	průměrná úspěšnost (%)	rozsah (%)	průměrná úspěšnost (%)	rozsah (%)	průměrná úspěšnost (%)
plastron_bbox	61,86–73,64	66,97	64,95–79,85	72,39	62,02–74,23	68,72
plastron_mask	56,19–71,32	63,88	60,47–72,09	67,14	68,99–79,90	74,35
central_seam	55,81–65,89	60,30	55,04–66,67	61,29	57,36–70,54	64,47

Pro multiclass klasifikaci je k dispozici 60 fotografií želv z datasetu ČIŽP, z čehož je 30 jedinců zachycených dvakrát, a 50 fotografií z datasetu P. Petráse, ve kterém jsou celkově pouze 4 různí jedinci s následujícím počtem fotografií: 8, 10, 19 a 13. Pro binární klasifikaci tak lze u datasetu ČIŽP sestavit 30 dvojic stejných jedinců a 30 dvojic náhodně zvolených rozdílných jedinců želv, a u datasetu P. Petráse 322 dvojic stejných a 322 dvojic rozdílných jedinců želv.

Naměřené rozsahy úspěšností jsou velmi velké, je to způsobeno právě tím, že ve všech testovaných případech se jedná o poměrně malé datasey, některá data jsou při binární klasifikaci vybírána náhodně, rozdělení dat je testováno v různých poměrech a klasifikátory nemají pevně nastavené hyperparametry, protože jsou vždy laděny vzhledem k použitému datasetu.

Z průměrných hodnot lze vypožorovat, že s multiclass klasifikací na datasetu ČIŽP (tabulka 6.1) si nejlépe poradila nejméně hluboká CNN VGG16 na vyříznutých maskách plastronů s dosaženou průměrnou úspěšností 71,42 %. Naopak nejméně úspěšná byla architektura InceptionV3 na vyříznutých oblastech centrálních švů s průměrnou úspěšností pouhých 32,92 %. Na datasetu P. Petráse při multiclass klasifikaci (tabulka 6.2) všechny extraktory dosahovaly ve většině případů 100% úspěšnosti, je tedy zřejmé, že pokud má klasifikátor větší množství trénovacích dat pro jednu třídu (a ne pouze 1, jako je tomu u datasetu ČIŽP), jsou příznaky vyextrahované pomocí jednotlivých CNN dostatečně robustní a lze je použít pro problematiku určování ID želv. V datasetu P. Petráse jsou však fotografie plastronů zachycených pouze v jediném časovém období, nelze tak jednoznačně dokázat, jestli je tato klasifikace odolná proti vývojovým změnám na plastronech.

V případě klasifikace binární na datasetu ČIŽP (tabulka 6.3) dosahovala nejlepších výsledků opět architektura VGG16 s průměrnou úspěšností 83,47 % na vyříznutých maskách (a s téměř stejně vysokou úspěšností i na ohraničujících rámečkách) plastronů a na datasetu P. Petráse (tabulka 6.4) architektura ResNet50 s průměrnou úspěšností 74,35 %, také na vyříznutých maskách plastronů.

Výsledky naměřené v této práci dosahují vyšších hodnot než výsledky naměřené v práci [2] A. Majtánové na fotografiích ručně ořezaných na oblast plastronu, což jen potvrzuje správnost automatické detekce plastronu provedené v sekci 5.3, díky které jsou plastrony na fotografiích v jednotném nastavení a metody použité při klasifikaci se tak nemusejí potýkat s rozdílnými úhly a umístěním plastronů na fotografiích.

Celkově klasifikace byla nejúspěšnější na vyříznutých maskách plastronů, nejspíše z toho důvodu, že u nich byly eliminovány veškeré rušivé vlivy pozadí, jež mohly být nevhodně interpretovány extraktory příznaků a narušovat tak klasifikaci. Překvapivě nejméně úspěšná byla klasifikace vyříznutých oblastí centrálních švů. Z hlediska identifikačních příznaků, popsanych v sekci 1.3,

---

by tyto vyříznuté části měly obsahovat vše potřebné k úspěšné klasifikaci, ale použité extraktory příznaků – již předtrénované (na datasetu ImageNet) a pro problematiku rozpoznávání želv dále nijak speciálně neuzpůsobené CNN, tyto specifické příznaky však zřejmě zachytit nedokázaly.

Dosažené výsledky měření ukázaly, že využití konvolučních neuronových sítí k extrakci obrazových příznaků v rámci problematiky automatického rozpoznávání želv je možné, ale jak už zde bylo zmíněno dříve, datasety, na kterých byly prováděny experimenty, nejsou dostatečně rozsáhlé na to, aby se dala jednoznačně určit kvalita testovaných architektur CNN a pro použití daných metod v praxi je zajisté nutné další testování v budoucnosti, až bude k dispozici nějaký obsáhlejší dataset, zachycující jedince želv druhu *Testudo graeca* několikrát v průběhu jejich vývoje, aby bylo možné sledovat postupné změny na jejich plastronech a zjišťovat, jak jsou zvolené metody proti těmto změnám odolné.



---

## Závěr

Cílem této bakalářské práce bylo vytvoření automatického systému pro rozpoznávání jedinců želv druhu *Testudo graeca* z fotografií jejich plastronů.

Nejprve byla prozkoumána literatura o želvách z čeledi Testudinidae se zaměřením na druh *Testudo graeca*. V práci byly popsány způsoby značení želv a plastron včetně jeho použitelných identifikačních příznaků.

Dále byla provedena analýza předchozích řešení problému rozpoznávání želv a dostupných datasetů. Také zde byla navržena ideální podoba želvího datasetu – doporučení pro pořizování fotografií plastronů.

Následně byly představeny všechny metody v práci použité pro detekci plastronu i pro extrakci příznaků a jejich klasifikaci, a na základě poznatků z předchozích prací byl v programovací jazyce Python implementován vlastní prototyp pro rozpoznávání želv, který byl na závěr testován na datasetu České inspekce životního prostředí zapůjčeném z ÚTIA AVČR a na datasetu chovatele želv Petra Petráse. Všechny dílčí cíle stanovené v úvodu práce tak byly naplněny.

Automatická detekce plastronu na fotografii byla provedena za pomoci modelu Mask R-CNN a proběhla úspěšně u všech 110 fotografií (z obou zmiňovaných datasetů). U klasifikace byly testovány 2 přístupy – multiclass klasifikace pro predikci ID želv a binární klasifikace pro porovnání 2 jedinců želv a jako extraktory příznaků při nich byly vybrány 3 již předtrénované architektury konvolučních neuronových sítí – VGG16, InceptionV3 a ResNet50.

Z naměřených průměrných úspěšností multiclass klasifikace 71,42 % na datasetu ČIŽP a až 100 % na datasetu P. Petráse a binární klasifikace 83,47 % na datasetu ČIŽP a 74,35 % na datasetu P. Petráse vyplynulo, že využití konvolučních neuronových sítí k extrakci obrazových příznaků v rámci problematiky automatického rozpoznávání želv je možné, ale vyžaduje ještě další testování na rozsáhlejším datasetu zachycujícím vývojové změny na plastronech.





---

## Literatura

- [1] SEDLÁČEK, Matěj. *Recognition of Tortoises (Rozpoznávání želv)*. Praha, 2015. Diplomová práce. České vysoké učení technické v Praze, Fakulta jaderná a fyzikálně inženýrská, Katedra matematiky. Vedoucí práce Michal Haindl.
- [2] MAJTÁNOVÁ, Adriána. *Rozpoznávanie korytnáčiek*. Praha, 2020. Bakalářská práce. České vysoké učení technické v Praze, Fakulta informačních technologií, Katedra aplikované matematiky. Vedoucí práce Radek Richtr.
- [3] SVATONĚ, Zdeněk. *Detekce želvých plastronů pro rozpoznávání*. Praha, 2021. Diplomová práce. České vysoké učení technické v Praze, Fakulta informačních technologií, Katedra aplikované matematiky. Vedoucí práce Michal Haindl.
- [4] TICHÝ, Lubomír. *Evropské suchozemské želvy: jejich chov a odchov* [online]. s. 7–13 [cit. 2021-04-03]. Dostupné z: <https://www.sci.muni.cz/botany/salicornia/EvrSuchZel.pdf>.
- [5] *Jak značit exempláře CITES?: metodika AOPK ČR*. Praha: Agentura ochrany přírody a krajiny České republiky, 2019, 31 s. ISBN 978-80-7620-034-0.
- [6] TICHÝ, Lubomír. *CITES a chov suchozemských želv* [online]. [cit. 2021-03-08]. Dostupné z: [https://www.sci.muni.cz/botany/salicornia/zzelvy\\_clan\\_cites1.htm](https://www.sci.muni.cz/botany/salicornia/zzelvy_clan_cites1.htm).
- [7] TICHÝ, L. a K. KINTROVÁ. Specimen identification from time-series photographs using plastron morphometry in *Testudo graeca iberica*. In: *Journal of Zoology* [online]. 2010, roč. 281, č. 3, s. 210–217 [cit. 2020-03-28]. ISSN 09528369. Dostupné z: doi:10.1111/j.1469-7998.2010.00693.x.

- [8] MOSIMANN, James Emile. *Variation and relative growth in the plastral scutes of the turtle *Kinosternon integrum* Leconte*. Ann Arbor: Museum of Zoology, University of Michigan, 1956.
- [9] KARPATY, Andrej. *Convolutional neural networks (cnns/convnets)*. In: *CS231n Convolutional Neural Networks for Visual Recognition* [online]. 2016 [cit. 2021-04-14]. Dostupné z: <https://cs231n.github.io/convolutional-networks/>.
- [10] KLOUDA, Karel, Juan Pablo Maldonado LOPEZ a Daniel VAŠATA. *Vytěžování znalostí z dat (BI-VZD): Neuronové sítě* [online]. 2020 [cit. 2021-04-15]. Po přihlášení dostupné z: <https://courses.fit.cvut.cz/BI-VZD/@B201/lectures/files/BI-VZD-11-cs-slides.pdf>.
- [11] KARN, Ujjwal. An Intuitive Explanation of Convolutional Neural Networks. In: *The data science blog: machine learning, deep learning, nlp, data science* [online]. 2016 [cit. 2021-04-14]. Dostupné z: <https://ujjwalkarn.me/2016/08/11/intuitive-explanation-convnets/>.
- [12] Deep Convolutional Neural Networks. In: *Run:AI: a Platform for AI Virtualization and Orchestration* [online]. [cit. 2021-04-14]. Dostupné z: <https://www.run.ai/guides/deep-learning-for-computer-vision/deep-convolutional-neural-networks/>.
- [13] GIRSHICK, Ross, Jeff DONAHUE, Trevor DARRELL a Jitendra MALIK. Rich Feature Hierarchies for Accurate Object Detection and Semantic Segmentation. In: *2014 IEEE Conference on Computer Vision and Pattern Recognition* [online]. IEEE, 2014, s. 580–587. ISBN 978-1-4799-5118-5. Dostupné z: doi:10.1109/CVPR.2014.81.
- [14] UIJLINGS, Jasper R. R. et. al. Selective Search for Object Recognition. In: *International Journal of Computer Vision* [online]. 2013, roč. 104, č. 2, s. 154–171 [cit. 2021-04-15]. ISSN 0920-5691. Dostupné z: doi:10.1007/s11263-013-0620-5.
- [15] GIRSHICK, Ross. Fast R-CNN. In: *2015 IEEE International Conference on Computer Vision (ICCV)* [online]. IEEE, 2015, s. 1440–1448 [cit. 2021-04-15]. ISBN 978-1-4673-8391-2. Dostupné z: doi:10.1109/ICCV.2015.169.
- [16] REN, Shaoqing, Kaiming HE, Ross GIRSHICK a Jian SUN. Faster R-CNN: Towards Real-Time Object Detection with Region Proposal Networks. In: *IEEE Transactions on Pattern Analysis and Machine Intelligence* [online]. 2017, roč. 39, č. 6, s. 1137–1149 [cit. 2021-04-15]. ISSN 0162-8828. Dostupné z: doi:10.1109/TPAMI.2016.2577031.
- [17] HE, Kaiming, Georgia GKIOXARI, Piotr DOLLAR a Ross GIRSHICK. Mask R-CNN. In: *2017 IEEE International Conference on Computer*

- Vision (ICCV)* [online]. IEEE, 2017, s. 2980–2988 [cit. 2021-04-13]. ISBN 978-1-5386-1032-9. Dostupné z: doi:10.1109/ICCV.2017.322.
- [18] KLOUDA, Karel, Juan Pablo Maldonado LOPEZ a Daniel VAŠATA. *Vytěžování znalostí z dat (BI-VZD): Lineární regrese* [online]. 2020 [cit. 2021-04-12]. Po přihlášení dostupné z: <https://courses.fit.cvut.cz/BI-VZD/@B201/lectures/files/BI-VZD-07-cs-slides.pdf>.
- [19] FISCHLER, Martin A. a Robert C. BOLLES. Random sample consensus: A Paradigm for Model Fitting with Applications to Image Analysis and Automated Cartography. In: *Communications of the ACM* [online]. New York: Association for Computing Machinery, 1981, roč. 24, č. 6, s. 381–395 [cit. 2021-04-12]. ISSN 0001-0782. Dostupné z: <https://dl.acm.org/doi/10.1145/358669.358692>.
- [20] DUSMEZ, Serkan, Mehrdad HEYDARZADEH, Mehrdad NOURANI a Bilal AKIN. Remaining Useful Lifetime Estimation for Power MOSFETs Under Thermal Stress With RANSAC Outlier Removal. In: *IEEE Transactions on Industrial Informatics* [online]. 2017, roč. 13, č. 3, s. 1271–1279 [cit. 2021-04-12]. ISSN 1551-3203. Dostupné z: doi:10.1109/TII.2017.2665668.
- [21] SIMONYAN, Karen a Andrew ZISSERMAN. Very Deep Convolutional Networks for Large-Scale Image Recognition. In: *3rd International Conference on Learning Representations*. 2015.
- [22] HASSAN, Muneeb ul. VGG16 – Convolutional Network for Classification and Detection. In: *Neurohive – Neural Networks* [online]. 2018 [cit. 2021-6-15]. Dostupné z: <https://neurohive.io/en/popular-networks/vgg16/>.
- [23] SZEGEDY, Christian et al. Going deeper with convolutions. In: *2015 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)* [online]. IEEE, 2015, s. 1–9 [cit. 2021-6-16]. ISBN 978-1-4673-6964-0. Dostupné z: doi:10.1109/CVPR.2015.7298594.
- [24] SZEGEDY, Christian, Vincent VANHOUCKE, Sergey IOFFE, Jon SHLENS a Zbigniew WOJNA. Rethinking the Inception Architecture for Computer Vision. In: *2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)* [online]. IEEE, 2016, s. 2818–2826 [cit. 2021-6-7]. ISBN 978-1-4673-8851-1. Dostupné z: doi:10.1109/CVPR.2016.308.
- [25] HE, Kaiming, Xiangyu ZHANG, Shaoqing REN a Jian SUN. Deep Residual Learning for Image Recognition. In: *2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)* [online]. IEEE,

- 2016, s. 770–778 [cit. 2021-6-7]. ISBN 978-1-4673-8851-1. Dostupné z: doi:10.1109/CVPR.2016.90.
- [26] KLOUDA, Karel, Juan Pablo Maldonado LOPEZ a Daniel VAŠATA. *Vytěžování znalostí z dat (BI-VZD): Logistická regrese* [online]. 2020 [cit. 2021-06-15]. Po přihlášení dostupné z: <https://courses.fit.cvut.cz/BI-VZD/@B201/lectures/files/BI-VZD-09-cs-slides.pdf>.
- [27] MARTINEZ, Hector Ian. Train/Test Split, Cross-Validation, & You. In: *Medium – Where good ideas find you* [online]. 2018 [cit. 2021-6-13]. Dostupné z: <https://medium.com/@hi.martinez/train-test-split-cross-validation-you-b87f662445e1>.
- [28] PIATETSKY, Gregory. Python leads the 11 top Data Science, Machine Learning platforms: Trends and Analysis. In: *KDnuggets* [online]. 2019 [cit. 2021-5-1]. Dostupné z: <https://www.kdnuggets.com/2019/05/poll-top-data-science-machine-learning-platforms.html>.
- [29] KLUYVER, T. et al. Jupyter Notebooks – a publishing format for reproducible computational workflows. In: LOIZIDES, F. a B. SCHMIDT, eds. *Positioning and Power in Academic Publishing: Players, Agents and Agendas*. 2016, s. 87–90.
- [30] JUNG, Alexander B. et al. *imgaug* [online]. Github, 2020 [cit. 2021-5-12]. Dostupné z: <https://github.com/aleju/imgaug>.
- [31] CHOLLET, Francois et al. *Keras* [online]. GitHub, 2015 [cit. 2021-5-1]. Dostupné z: <https://github.com/fchollet/keras>.
- [32] HARRIS, Charles R., K. Jarrod MILLMAN, Stéfan J. VAN DER WALT, et al. Array programming with NumPy. In: *Nature* [online]. 2020, roč. 585, č. 7825, s. 357–362 [cit. 2021-5-1]. ISSN 0028-0836. Dostupné z: doi:10.1038/s41586-020-2649-2.
- [33] BRADSKI, G. The OpenCV Library. In: *Dr. Dobb's Journal of Software Tools*. 2000.
- [34] MCKINNEY, Wes. Data Structures for Statistical Computing in Python. In: WALT, Stéfan van der a Jarrod MILLMAN, eds. *Proceedings of the 9th Python in Science Conference* [online]. 2010, s. 56–61 [cit. 2021-6-17]. Dostupné z: doi:10.25080/Majora-92bf1922-00a.
- [35] WALT, Stéfan van der et al. Scikit-image: image processing in Python. In: *PeerJ* [online]. 2014, roč. 2 [cit. 2021-5-1]. ISSN 2167-8359. Dostupné z: doi:10.7717/peerj.453.
- [36] PEDREGOSA, Fabian et al. Scikit-learn: Machine learning in Python. In: *Journal of machine learning research*. 2011, roč. 12, č. 85, s. 2825–2830.

- 
- [37] ABADI, Martín, et al. Tensorflow: A system for large-scale machine learning. In: *12th USENIX Symposium on Operating Systems Design and Implementation (OSDI'16)*. 2016, s. 265–283.
- [38] LIN, Tsung-Yi et al. Microsoft COCO: Common Objects in Context. In: FLEET, David, Tomas PAJDLA, Bernt SCHIELE a Tinne TUYTELAARS, eds. *Computer Vision – ECCV 2014* [online]. Cham: Springer International Publishing, 2014, s. 740–755 [cit. 2021-5-3]. ISBN 978-3-319-10601-4. Dostupné z: doi:10.1007/978-3-319-10602-1\_48.
- [39] ABDULLA, Waleed. *Mask R-CNN for object detection and instance segmentation on Keras and TensorFlow* [online]. Github, 2017 [cit. 2021-5-1]. Dostupné z: [https://github.com/matterport/Mask\\_RCNN](https://github.com/matterport/Mask_RCNN).
- [40] DUTTA, Abhishek a Andrew ZISSERMAN. The VIA Annotation Software for Images, Audio and Video. In: *Proceedings of the 27th ACM International Conference on Multimedia* [online]. New York: ACM, 2019, s. 2276–2279 [cit. 2021-5-3]. ISBN 9781450368896. Dostupné z: doi:10.1145/3343031.3350535.
- [41] DENG, Jia et al. ImageNet: A large-scale hierarchical image database. In: *2009 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)* [online]. IEEE, 2009, s. 248–255 [cit. 2021-6-7]. ISBN 978-1-4244-3992-8. Dostupné z: doi:10.1109/CVPR.2009.5206848.



---

## Seznam použitých zkratk

- AOPK ČR** Agentura ochrany přírody a krajiny České republiky
- CITES** Convention on International Trade in Endangered Species of Wild Fauna and Flora (Úmluva o mezinárodním obchodu s ohroženými druhy volně žijících živočichů a planě rostoucích rostlin)
- CNN** Convolutional neural network (Konvoluční neuronová síť)
- ČIŽP** Česká inspekce životního prostředí
- FPN** Feature pyramid network
- ILSVRC** ImageNet Large Scale Visual Recognition Challenge
- IUCN** International Union for Conservation of Nature (Mezinárodní svaz ochrany přírody)
- JSON** JavaScript Object Notation (JavaScriptový objektový zápis)
- $k$ -NN**  $k$ -nearest neighbors ( $k$ -nejbližších sousedů)
- LBP** Local Binary Patterns (Lokální binární vzor)
- MLE** Maximum likelihood estimation (Metoda maximální věrohodnosti)
- MS COCO** Microsoft Common Objects in Context
- OLS** Ordinary least squares (Metoda nejmenších čtverců)
- ORB** Oriented FAST and Rotated BRIEF
- RANSAC** Random sample consensus
- ReLU** Rectified linear unit (Oříznutá lineární funkce)

## A. SEZNAM POUŽITÝCH ZKRATEK

---

**ResNet** Residual neural network (Reziduální neuronová síť)

**RoI** Region of interest

**RPN** Region proposal network

**RSS** Residual sum of squares (Reziduální součet čtverců)

**R-CNN** Regions with CNN features

**SIFT** Scale Invariant Feature Transform

**SVM** Support vector machine (Metoda podpůrných vektorů)

**ÚTIA AVČR** Ústav teorie informace a automatizace Akademie věd České republiky.

**VGG** Visual Geometry Group

**VIA** VGG Image Annotator



---

## Obsah přiloženého média

README.md.....	stručný popis obsahu média
requirements.txt.....	seznam použitých Python balíčků
src.....	zdrojové kódy implementace
Mask_RCNN.....	využitá implementace Mask R-CNN
Testudo_graeca	
mask_rcnn_plastron.h5.....	natrénované váhy Mask R-CNN
Petras.....	adresář s datasetem P. Petrás
plastron_bbox.....	adresář s vyříznutými plastrony
plastron_mask.....	adresář s vyříznutými maskami plastronů
central_seam.....	adresář s vyříznutými centrálními švy
text	
thesis.pdf.....	text práce ve formátu PDF
thesis.....	zdrojová forma práce ve formátu L <sup>A</sup> T <sub>E</sub> X