



VYSOKÉ UČENÍ TECHNICKÉ V BRNĚ
BRNO UNIVERSITY OF TECHNOLOGY

FAKULTA INFORMAČNÍCH TECHNOLOGIÍ
FACULTY OF INFORMATION TECHNOLOGY

ÚSTAV INTELIGENTNÍCH SYSTÉMŮ
DEPARTMENT OF INTELLIGENT SYSTEMS

ANALÝZA KONVOLUČNÍCH NEURONOVÝCH SÍTÍ PRO DETEKCI A KLASIFIKACI POŠKOZENÍ OTISKU PRSTU

ANALYSIS OF CONVOLUTIONAL NEURAL NETWORKS FOR DETECTION AND CLASSIFICATION
OF DAMAGES IN FINGERPRINT IMAGES

DIPLOMOVÁ PRÁCE
MASTER'S THESIS

AUTOR PRÁCE
AUTHOR

Bc. KATEŘINA FOŘTOVÁ

VEDOUCÍ PRÁCE
SUPERVISOR

Ing. ONDŘEJ KANICH, Ph.D.

BRNO 2022

Zadání diplomové práce



25006

Studentka: **Fořtová Kateřina, Bc.**
Program: Informační technologie a umělá inteligence
Specializace: Bioinformatika a biocomputing
Název: **Analýza konvolučních neuronových sítí pro detekci a klasifikaci poškození otisku prstu**
Analysis of Convolutional Neural Networks for Detection and Classification of Damages in Fingerprint Images
Kategorie: Zpracování obrazu
Zadání:

1. Prostudujte literaturu týkající se rozpoznávání podle otisků prstů s důrazem na poškozené či nekvalitní snímky z důvodu onemocnění na prstu či dalších poškození. Seznamte se s možnostmi detekce a klasifikace s využitím neuronových sítí.
2. Navrhněte několik metod založených na konvolučních neuronových sítí pro detekci a klasifikaci minimálně tří tříd poškození otisků prstů (např. atopický ekzém, bradavice, psoriáza).
3. Implementujte navržené metody z předchozího bodu. Experimentujte s nastavením parametrů zvolených architektur neuronových sítí.
4. Analyzujte přesnost metod implementovaných v předchozím bodě na vhodné datové sadě. Porovnejte jejich výkonnost.
5. Dosažené výsledky shrňte a diskutujte. Uveďte možná vylepšení a rozšíření vašeho řešení.

Literatura:

- Maltoni, D., Maio, D., Jain, A.K. and Prabhakar, S.: *Handbook of Fingerprint Recognition*. Springer, 2009, p. 512. ISBN 978-1-8488-2254-2.
- Drahanský, M.: *Hand-Based Biometrics: Methods and technology*, IET 2018, p. 430, ISBN 978-1-78561-224-4.
- Barotová, Š.: *Detector of Skin Diseases by Fingerprint Technology*, 2017. Bachelor's thesis. FIT BUT in Brno, Brno.

Při obhajobě semestrální části projektu je požadováno:

- Body 1 a 2.

Podrobné závazné pokyny pro vypracování práce viz <https://www.fit.vut.cz/study/theses/>

Vedoucí práce: **Kanich Ondřej, Ing., Ph.D.**

Vedoucí ústavu: Hanáček Petr, doc. Dr. Ing.

Datum zadání: 1. listopadu 2021

Datum odevzdání: 18. května 2022

Datum schválení: 3. listopadu 2021

Abstrakt

Cílem této diplomové práce je analyzovat přístupy detekce a klasifikace s využitím konvolučních neuronových sítí na problému poškození otisků prstů. První část práce se zabývá studiem literatury týkající se biometrie a zpracování otisků prstů s důrazem na možná onemocnění, které mohou oblast prstu postihnout. Následně se práce zaměřuje na rozpoznávání s využitím neuronových sítí. Diplomová práce popisuje architektury konvolučních neuronových sítí a modelů pro detekci objektů až po nejnovější přístupy. Je navrhnut několik metod pro detekci a klasifikaci onemocnění otisků prstů využívající moderní architektury, rozličné typy páteřních sítí a metod pro detekci. Pro experimenty je zvoleno osm modelů založených na čtyřech různých přístupech pro detekci a klasifikaci. Následně je každý model několikrát natrénován s využitím úprav konfiguračních parametrů. Modely jsou na základě různých metrik posouzeny a porovnány z hlediska využití páteřní sítě i zvolené metody pro detekci. Při testu správně detekované a klasifikované plochy na reálných otiscích bylo dosaženo nejlepšího výsledku 76,875 %. Nejvíce problematickým onemocněním pro detekci a klasifikaci byl atopický ekzém, jehož příznaky se mohou projevovat mnoha způsoby.

Abstract

The aim of this Master's thesis is to analyze detection and classification approaches using convolutional neural networks on the problem of fingerprint damage. The first part of the thesis deals with the study of literature related to biometrics and fingerprint processing with emphasis on possible diseases that may affect the fingertip area. Subsequently, the thesis focuses on neural network-based recognition. The thesis describes the architectures of convolutional neural networks and object detection approaches up to the latest research. Several detection methods for detection and classification of skin diseases affecting fingertip are proposed using modern architectures, different types of backbone networks and detection methods. Eight models based on four different detection and classification approaches are chosen for the experiments. Subsequently, each model is trained several times using configuration parameter adjustments. The models are assessed on the basis of various metrics and compared in terms of the use of the backbone network and the chosen method for detection. The best result of 76.875 % was achieved in the test of correctly detected and classified area on real fingerprint images. The most problematic disease for detection and classification was atopic eczema, whose symptoms can manifest in many ways.

Klíčová slova

onemocnění kůže, otisky prstů, konvoluční neuronové sítě, detekce, klasifikace

Keywords

skin diseases, fingerprints, convolutional neural networks, detection, classification

Citace

FOŘTOVÁ, Kateřina. *Analýza konvolučních neuronových sítí pro detekci a klasifikaci poškození otisku prstu*. Brno, 2022. Diplomová práce. Vysoké učení technické v Brně, Fakulta informačních technologií. Vedoucí práce Ing. Ondřej Kanich, Ph.D.

Analýza konvolučních neuronových sítí pro detekci a klasifikaci poškození otisku prstu

Prohlášení

Prohlašuji, že jsem tuto diplomovou práci vypracovala samostatně pod vedením Ing. Ondřeje Kanicha, Ph.D.. Uvedla jsem všechny literární prameny, publikace a další zdroje, ze kterých jsem čerpala.

.....
Kateřina Fořtová
18. května 2022

Obsah

1	Úvod	2
2	Rozpoznávání podle otisků prstů	3
2.1	Biometrie	3
2.2	Struktura kůže	4
2.3	Nemoci kůže ovlivňující otisky prstů	6
2.4	Zpracování otisku prstů	10
2.5	Typy senzorů	13
2.6	Generování syntetických otisků prstů	16
3	Detekce a klasifikace pomocí neuronových sítí	18
3.1	Umělý neuron a perceptron	18
3.2	Vícevrstvý perceptron	19
3.3	Aktivační funkce	20
3.4	Loss funkce	21
3.5	Trénování neuronové sítě	22
3.6	Konvoluční neuronové sítě	24
3.7	Existující architektury neuronových sítí pro zpracování obrazu	25
3.8	Přístupy pro detekci objektů	28
4	Návrh řešení a implementace	33
4.1	Databáze otisků prstů	33
4.2	Anotace dat	34
4.3	Předzpracování otisků prstů	35
4.4	Použité metody pro detekci a klasifikaci	37
4.5	Výstup detekce a klasifikace onemocnění otisků prstů	39
4.6	Vyhodnocovací metriky	39
5	Experimenty	42
5.1	Parametry modelu pro experimentování	42
5.2	Experimenty s jednotlivými přístupy pro detekci objektů	43
5.3	Zhodnocení experimentů	58
6	Závěr	62
	Literatura	63
	A Ukázka predikcí onemocnění	68

Kapitola 1

Úvod

V současném světě je stále více rozšířeným užití biometrie jako součásti bezpečnostních mechanismů. Jednou z nejrozšířenějších biometrických charakteristik je otisk prstu. Pro snímání by měly být zaručeny ideální podmínky, avšak jedná se o nesplnitelný úkol v běžném nasazení. Snímky získané ze senzoru mohou být nekvalitní, mohou obsahovat nečistoty, které dále komplikují proces rozpoznávání. Vysoké procento populace trpí kožními onemocněními. Mnohá široce rozšířená onemocnění, například atopický ekzém nebo bradavice, postihují i oblast bříška prstu. U některých z onemocnění (např. bradavice) je jejich vliv na rozpoznávání nižší, protože papilární linie jsou zachovány v nepostiženém místě otisku prstu, u tolik rozšířeného atopického ekzému však jsou ve vážnějších podobách papilární linie těžko rozpoznatelné. Potenciální detektor a klasifikátor těchto onemocnění by mohl odhalit možná požadované uživatele využívající biometrický systém.

Problematika detekce a klasifikace pro rozličné typy objektů je stále jedním z předních témat současnosti. Techniky, které kdysi spíše využívaly ruční extrakci rysů, byly nahrazeny konvolučními neuronovými sítěmi, které dosahují vyšší přesnosti, ale i výpočetní náročnosti. Moderní detektory typů Faster R-CNN (Region Based Convolutional Neural Network) nebo SSD (Single Shot Detector) jsou využívané v reálných aplikacích a při rozpoznávání rozličných typů objektů. Cílem této práce je využít několika typů těchto moderních detektorů na úloze detekce a klasifikace různých typů onemocnění otisku prstu. Přitom je kladen důraz na experimenty s dostupnými architekturami a snahou nalézt co neuniverzálnější detektor, který by zvládl rozlišit co nejvíce typů onemocnění – pro experimenty jsou vybrány onemocnění atopický ekzém, psoriáza, dyshidróza a bradavice.

V kapitole 2 je popsáno biometrické rozpoznávání se zaměřením na otisky prstu. Cílem této kapitoly je popsat základní termíny oboru biometrie, strukturu kůže, některé ze základních typů senzorů a metod určené pro snímání a následné zpracování otisku prstu. Kapitola poukazuje i na různá onemocnění, které mohou snímání otisku prstu ovlivnit. Nakonec se kapitola věnuje problematice generování syntetických otisků prstů, které slouží pro rozšíření trénovacího datasetu. Kapitola 3 se zabývá detekcí a klasifikací s využitím neuronových sítí. Návrh řešení problému a jeho implementace je popsána v kapitole 4. Tato kapitola popisuje tvorbu datasetu, problematiku anotace bounding boxů pro onemocnění a předzpracování datasetu. Následně jsou popsány vybrané modely užité pro trénování a metriky, které budou složit pro vyhodnocení přístupů při experimentech uvedených v kapitole 5. V kapitole 6 jsou shrnutý závěrem výsledky diplomové práce a možné budoucí rozšíření.

Kapitola 2

Rozpoznávání podle otisků prstů

Lidské otisky prstů jsou jednou z nejznámějších a nejužívanějších biometrických charakteristik. V roce 1823 bylo vytvořeno Janem Evangelistou Purkyně první klasifikační schéma otisků prstů, které rozdělovalo otisky prstů do devíti tříd na základě konfigurace hřebenů papilárních linií. V roce 1863 bylo ve Velké Británii akceptováno, že žádní dva lidé nemají totožné otisky prstu [36]. Následující kapitola se zabývá stručným popisem oboru biometrie a jejími základními pojmy. Hlavním z cílů této kapitoly je však představit onemocnění, které mohou plochu otisku prstu postihnout a popsat základní přístupy pro zpracování a rozpoznávání otisků prstů.

2.1 Biometrie

Cílem biometrie (biometrického rozpoznávání) je použití anatomických a behaviorálních charakteristik, které se nazývají biometrické rysy, k automatickému rozpoznávání jedinců. Biometrie se postupem času stala efektivním řešením lidské identifikace reprezentující tělesnou identitu jednotlivce [36]. Představuje jednu z technologií, které mohou umožnit naší společnosti bezpečnější a poskytnout vyšší uživatelské pohodlí [36]. Biometrický systém však může být napadnutelný a biometrická identita nemůže být v případě prozrazení anulována [14].

2.1.1 Identita, identifikace, verifikace a autentizace

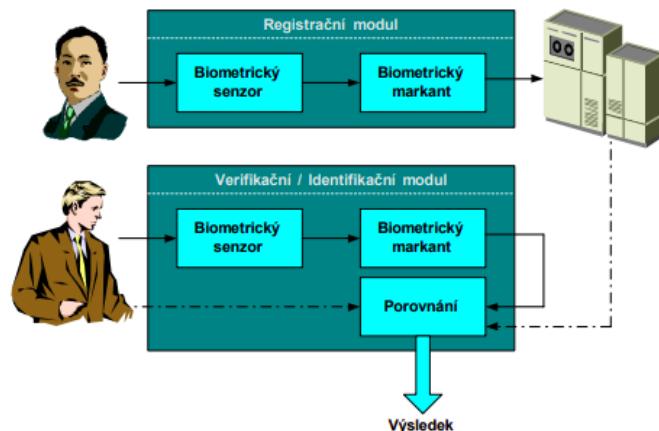
Identita, identifikace, verifikace a autentizace jsou základními pojmy, které biometrie využívá. Rozpoznání je založeno na jednoznačné identitě jedince. Identita je jednoznačnou charakteristikou každého z nás. Rozlišujeme fyzickou a elektronickou identitu. Fyzická identita je pouze jediná a je definována vzhledem a chováním. Elektronických identit můžeme mít ovšem vytvořených více (např. více účtů na sociálních sítích) [14].

Identifikace je situací, kdy daná osoba zadá svoji biometrickou vlastnost systému, ale nesdílí mu svoji identitu. Daný systém pak má za úkol rozpoznat identitu uživatele – porovnává se vzorek zadaný na vstupu s celou databází. Systém poté naleze nebo nenaleze danou identitu. Identifikaci můžeme chápat jako porovnání 1:N [14]. Při verifikaci je sdělena systému elektronická identita uživatele a dojde k ověření fyzické identity. V databázi je vyhledán daný záznam uživatele, který obsahuje biometrická data. Pokud je záznam nenalezen, dojde k zamítnutí přístupu uživatele, jinak dochází k porovnání dat – výsledkem je potvrzení či nepotvrzení identity. Proces verifikace je také jinak nazýván porovnání 1:1 [14]. Při autentizaci má systém za úkol potvrdit autentičnost (hodnověrnost) osoby

žádající o přístup. Autentizace může probíhat jak při identifikaci, tak při autentizaci. Rozhodnutí je obvykle prováděno na základě nastavené hodnoty prahu [14].

2.1.2 Biometrické systémy a vlastnosti

Biometrický systém, znázorněn na obrázku 2.1, je složen z registračního a verifikačního / identifikačního modulu. Součástí obou uvedených modulů je biometrický senzor a biometrický markant – extrahované rysy z biometrické informace na vstupu. Hlavním rozdílem obou modulů je princip práce s databází a biometrickým markantem. Při práci s registračním modulem je biometrický markant uložen do databáze. Verifikační / identifikační modul naopak neukládá biometrický markant do databáze, ale místo toho načítá data z databáze pro porovnání biometrického markantu na vstupu s údaji databáze. Při tomto porovnání obdržíme výsledek na základě nalezení případné shody a případného operačního módu – verifikaci nebo identifikaci [14].



Obrázek 2.1: Schéma biometrického systému [14]

Biometrické charakteristiky můžeme dělit na anatomické a dynamické vlastnosti. U anatomických vlastností je jeden pevný rys jednou biometrickou vlastností. Do anatomických vlastností spadá např. otisk prstu, obličej, duhovka a sítnice oka, dlaň nebo geometrie ruky. Dynamické vlastnosti jsou spjaty s nějakou akcí uživatele, do této kategorie spadá např. chůze, hlas, gestikulace obličeje či dynamické vlastnosti podpisu [14].

2.2 Struktura kůže

Kůže je největším lidským orgánem, který funguje jako počáteční obrana proti patogenům, ultrafialovému (UV) záření, chemikáliím nebo zranění. Také slouží jako regulátor teploty. Plocha kůže tvoří asi $1,5 - 1,8 \text{ m}^2$ a má hmotnost zhruba 4,5 kg [17]. Lidská kůže je složena ze tří vrstev:

- epidermis (pokožka),
- corium (dermis – škára),
- tela subcutanea (subcutis — podkožní vazivo) [17].

Pokožka (epidermis) je složena z několika vrstev plochých buněk. Buňky na povrchu postupně prochází procesem odumírání, rohovatění a odlupování, kdy jsou nahrazeny buňkami z hlubší epidermální vrstvy. Zvláštní vazivové buňky v hlubších vrstvách obsahují kožní pigment melanin. Barva lidské kůže pak závisí na množství melaninu, jeho hloubce uložení a prokrvení kůže. Melanin má za funkci pohlcovat UV složku slunečního záření, která by jinak dokázala poškodit buňky hlubších vrstev [17].

Škára (corium) je složena z vazivových buněk, elastických vláken a tukových buněk. Elastická vlákna mají za úkol zajišťovat pružnost, roztažitelnost, pevnost a štěpitelnost kůže. Ve škáře jsou uloženy dva typy kožních žláz – mazové a potní žlázy. Také se zde vyskytují krevní, mízní cévy, vlasové kořeny a nervy. Ve výběžcích proti epidermální vrstvě se vyskytují nervová zakončení – receptory umožňující vnímání bolesti, tepla nebo chladu [17].

Podkožní vazivo (subcutis) je tvořené sítí kolagenních, elastických vláken a vazivovými buňkami. Slouží jako potenciální tuková tkáň, která je schopna ukládat v buňkách velké množství tukových kapének [17].

Nejdůležitějšími funkcemi kůže jsou:

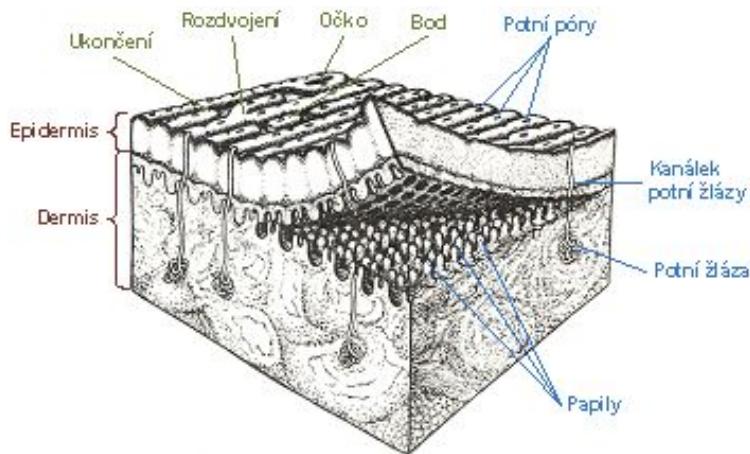
- Ochrana těla – Kůže zabraňuje vnikání škodlivých látek do vnitřního prostředí organismu a je odolnou vůči mechanickému poškození. Pigment melanin chrání proti UV záření [17].
- Smyslová funkce kůže – V kůži se vyskytuje velké množství receptorů sloužící k vnímání mechanických, tepelných a bolestivých počitků. Specializované receptory slouží k vnímání tepla, chladu a hmatu [17].
- Udržování tělesné teploty – Zrohovatělá vrstva buněk díky své špatné tepelné vodivosti chrání organismus před většími tepelnými ztrátami. Podkožní vazivo tvoří poměrně silnou vrstvu u každého jedince [17].
- Skladovací funkce kůže – Množství tuku v podkožním vazivu tvoří energetickou zásobárnu organismu. Probíhá zde uskladnění vitamínů rozpustných v tucích (A, D, E, K) [17].
- Vylučovací funkce kůže – Vyloučené sekrety mazových a potních žláz se uplatňují při ochraně kůže i celého organismu [17].

2.2.1 Papilární linie

Na prstech rukou a nohou jsou přítomné vyvýšené kožní útvary, které se nazývají papilární linie. Základ papilárních linií je formovaný na hranici kožních vrstev epidermis a corium, kde dochází k vzájemné interakci spodní epidermální vrstvy a horní vrstvy coria. Po určitém zvlnění epidermální vrstvy se na bříškách prstů objeví papilární linie [15].

Struktura kůže a papilárních linií je znázorněna na obrázku 2.2, na obrázku 2.3 je pak zobrazen nasnímaný otisk prstu tvořený charakteristickými papilárními liniemi. Papilární linie slouží k určení identity jedince, tvoří dermatoglyfy, které jsou pro jedince charakteristické. Přesný důvod vzniku papilárních linií však není známý – uvažuje se o hypotéze, že to jsou např. výsledky napětí a tlaků v kůži během embryonálního vývoje. Uspořádání papilárních linií je z 90 % podmíněno geneticky, z 10 % je pak závislé na vnějších podmínkách [10]. Ani jednovaječná dvojčata nemají shodné otisky prstů – shodují se sice v základních vztazích, ale odlišují se v drobných detailech rozhodujícími pro identifikaci osob [10]. Papilární linie se vyvíjejí během nitroděložního vývoje a po celý život jedince vykazují stabilitu, kdy

se s přibývajícím věkem mění rozměry plošek prstů, dlaní či chodidel, avšak struktura zůstává stejná [15].



Obrázek 2.2: Struktura kůže [15]



Obrázek 2.3: Příklad nasnímaného otisku prstu [16]

2.3 Nemoci kůže ovlivňující otisky prstů

Nemoci kůže představují důležitý, ale často opomíjený faktor ovlivňující snímání otisků prstů a jejich následnou analýzu. Uvádí se, že zhruba 20 – 25 % pacientů trpí některou z kožních chorob [13]. Tato kapitola uvádí výčet onemocnění, které mohou postihnout oblast prstů, a způsob, jakým se projevují na zachyceném snímku na biometrickém senzoru.

2.3.1 Atopický ekzém

Atopický ekzém (viz obrázek 2.4), je jedním z druhů ekzému – nejčastějšího chronického zátnělivého onemocnění. Původ onemocnění zahrnují například genetické faktory, defekt kožní bariéry a vliv prostředí. Předpokládá se, že atopický ekzém je spjatý s příbuznými onemocněními jako astma, potravinová alergie nebo alergická rýma. Pokud jeden z rodičů dítěte trpí atopickým ekzémem, je více jak poloviční pravděpodobnost, že bude symptomy postihnut i jeho potomek. Při postižení obou rodičů je tato šance dokonce 80 % [31]. Atopický ekzém se projevuje svěděním, snížením hydratace a zvýšením suchosti pokožky. Postižené místo

je citlivé k alergenům a dráždivým látkám [31]. Při léčbě je nutné vyhýbat se spouštěcím faktorům a zamezit vysušování pokožky.

Při snímání se onemocnění atopickým ekzémem projevuje více způsoby. Příklad snímku atopického ekzému zachyceného biometrickým senzorem je znázorněn na obrázku 2.5. U pacientů s mírnějším průběhem se nemoc projevuje úzkými bílými liniemi přecházející přes papilární linie. Papilární linie zůstávají přerušené však pouze na několika místech a obvykle jsou stále uspokojivě čitelné. Ovšem otisky prstů pacientů s vážnějším průběhem nemoci jsou velmi poškozené, papilární linie jsou skoro nečitelné a vedle bílých linií se často vyskytují i tmavé nepravidelné skvrny.



Obrázek 2.4: Snímek pacienta trpící atopickým ekzémem [4]



Obrázek 2.5: Atopický ekzém zachycený biometrickým senzorem [4]

2.3.2 Bradavice

Bradavice (verruca vulgaris) (viz obrázek 2.6) jsou běžnými novotvary způsobené lidskými papilomaviry (HPV). Tyto viry, kterých existuje více jak 100 různých typů, dokáží vstoupit do kůže prostřednictvím malých řezných ran a způsobit zvýšený růst buněk. Kůže se tedy na daném místě postupně stává silnější – je vytvořena bradavice. Bradavice se mohou vyskytovat jednotlivě nebo ve skupinách na různých částech lidského těla. Zbavení se bradavic může být náročné zejména pro lidi se slabým imunitním systémem. U lidí se zdravým imunitním systémem obvykle bradavice zmizí postupem času [27]. Bradavice se na snímcích projevují jako bílé skvrny kulatého tvaru, občas s černými tečkami uvnitř (viz obrázek 2.7). Poškození není tolik závažné, jako např. u atopického ekzému a psoriázy, a papilární linie zbytku otisku prstu jsou nepoškozené.



Obrázek 2.6: Snímek pacienta trpící bradavici [13]



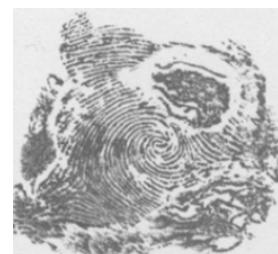
Obrázek 2.7: Bradavice zachycená biometrickým senzorem [4]

2.3.3 Psoriáza

Psoriáza (lupénka) (viz obrázek 2.8) se vyskytuje zhruba u 1 – 3 % lidské populace. Původ onemocnění je neznámý, avšak je uvažováno o dědičných predispozicích. Nejčastějším projevem tohoto chronického onemocnění je odlupování pokožky. Výskyt onemocnění na dlaních a prstech se vyznačuje červenými plaky. Při zhoršeném průběhu se můžou objevit bolestivé trhliny a krvácení. Psoriáza stejně jako bradavice nemá žádnou dostupnou léčbu. Obvykle se nemoc postupně zhoršuje a zlepšuje [37, 13]. Onemocnění je často na snímcích rozpoznatelné pomocí tmavé skvrny nepravidelného tvaru, která je ohraničena bílými liniemi (viz obrázek 2.9). Papilární linie jsou obvykle čitelné pouze na menší části snímku.



Obrázek 2.8: Snímek pacienta trpící psoriázou [13]



Obrázek 2.9: Psoriáza zachycená biometrickým senzorem [4]

2.3.4 Dyshidróza

Dyshidróza (viz obrázek 2.10) je jednou z nejčastějších kožních poruch. Původ onemocnění je neznámý, ale uvažuje se například o důvodu styku s kontaktním alergenem. Onemocnění je charakteristické výskytem drobných svědivých puchýřků na dlaních a stranách prstu. Kůže může být vlhká a zarudlá. Postupem času tyto puchýřky prasknou a objevují se zarudlé šupinaté skvrny [13]. Příklad zachycení onemocnění dyshidrózy biometrickým senzorem je zobrazen na obrázku 2.11. Při snímcích na senzoru se onemocnění projevuje více způsoby – většinou se jedná o bílé čáry a nepravidelné bílé skvrny. Při vážnějším poškození nejsou dostatečně viditelné papilární linie.



Obrázek 2.10: Snímek pacienta trpící dyshidrózou [13]



Obrázek 2.11: Dyshidróza zachycená biometrickým senzorem [4]

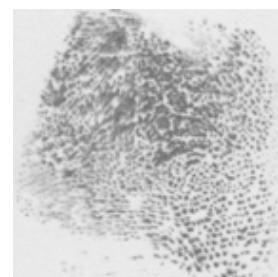
2.3.5 Acrodermatitis continua

Acrodermatitis continua (viz obrázek 2.12) je chronickým onemocněním, jedním ze zvláštních druhů psoriázy. Projevuje se výskytem pustul – dutin vyplněných hnisek na špičkách

prstů, většinou vyskytující se častěji na prstech dlaní než chodidel. Toto onemocnění nemá žádný standardizovaný postup léčby a jeho původ není známý, uvažuje se však o např. infekčním nebo genetickém původu. Při vážném průběhu onemocnění může onemocnění poškodit oblast nehtu [50]. Snímky na senzoru jsou typické nerozpoznatelnými papilárními liniemi, celý otisk prstu pokrývají tmavé menší skvrny různých tvarů (viz obrázek 2.13).



Obrázek 2.12: Snímek pacienta trpící Acrodermatitis continua [50]



Obrázek 2.13: Acrodermatitis continua zachycená biometrickým senzorem [4]

2.3.6 Raynaudova nemoc

Raynaudova nemoc (viz obrázek 2.14) je charakteristická křečovitými stahy drobných tepen a tepének z důsledku chladu a stresu. Dochází k poruše krevního zásobení postižené oblasti a postižená periferní oblast (např. prsty) pak postupně mění barvu na bílou, modrou a červenou. Odhad prevalence – počtu jedinců trpící nemocí vzhledem k počtu jedinců v dané populaci – odhadem činí 5 až 20 %. Nemoc se častěji vyskytuje u žen a problémy jsou obvykle dlouhodobé. Jedním z důvodů postižení tímto onemocněním je např. onemocnění pojivové tkáně nebo hematologické příčiny [13]. Toto onemocnění se projevuje pouze změnou zbarvení prstů a snímky ze senzoru se proto jeví jako nepoškozené (viz obrázek 2.15).



Obrázek 2.14: Snímek pacienta trpící Raynaudovou nemocí [38]



Obrázek 2.15: Raynaudova nemoc zachycená biometrickým senzorem [4]

2.3.7 Systémová sklerodermie

Systémová sklerodermie (viz obrázek 2.16) je chronickým autoimunitním onemocněním. Pro více než 90 % pacientů je typické onemocnění Raynaudovou nemocí a postižené jsou více ženy [13]. Původ nemoci je neznámý. Onemocnění se projevuje otoky prstů, malými jizvičkami na špičce prstu nebo sklerodaktylií. Při sklerodaktylii se v tkáních ruky ukládá

vazivo, čímž je narušená pohyblivost a vzniká charakteristická ztuhlá ruka. Nemoc je neléčitelná, ke tlumení projevů se mohou využít imunosupresivní léky tlumící imunitní reakce v těle – kortikoidy. Kožní postižení není pouze jediným projevem systémové sklerodermie – může dojít i k plicnímu postižení nebo postižení oběhového systému [51].



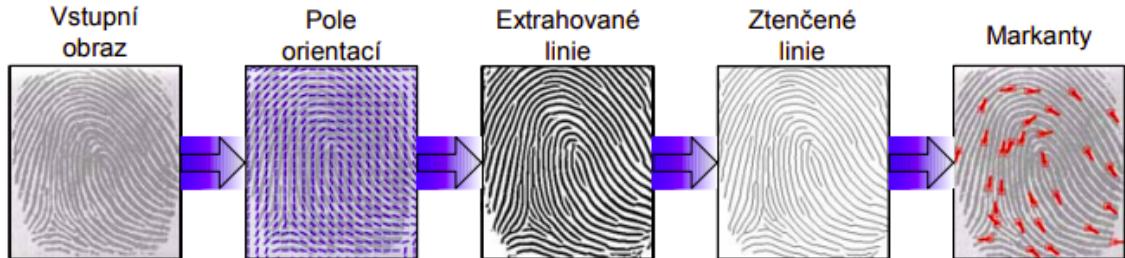
Obrázek 2.16: Snímek pacienta trpící systémovou sklerodermií [13]



Obrázek 2.17: Systémová sklerodermie zachycená biometrickým senzorem [4]

2.4 Zpracování otisku prstů

Nasnímané otisky prstu jsou následně zpracovány za cílem nalezení vhodných markantů, které mohou být využity pro následnou detekci a rozpoznávání. Nejběžnější způsob zpracování otisku prstu je znázorněn na obrázku 2.18.



Obrázek 2.18: Schéma zpracování otisku prstu [14]

Na začátku samotného zpracování otisku prstu je načten **vstupní obraz**. Snímek může být získaný z různých druhů senzorů (např. optické, kapacitní nebo bezdotykové) – vznikne tak digitální otisk prstu. Vstupní obraz obvykle obsahuje vysokou hodnotu šumu. Je nutné dbát na kontrolu živosti prstu nebo na vlivy jako je znečištění a poranění [14].

Následně probíhá výpočet **pole orientací**. Principem tohoto algoritmu je spočítání směru papilární linie pro každý pixel obrazu, kdy se využívá šedotónových hodnot. Následně je vypočtena lokální orientace celého bloku pixelů. Vznikne orientované pole, které může být namapováno na původní otisk prstu [14].

Při **extrahování linií** se používají různé úpravy vstupního obrazu. Patří k nim např. úprava histogramu, se kterou se spjata kontrola kvality obrazu. 2D Gaborovu funkce je jedním z možných způsobů filtrování. Pro samotnou detekci papilárních linií je možné použít schéma RAT nebo detekci papilárních linií dle Honga [14].

Detekce papilárních linií dle Honga vychází z předpokladu, že papilární linie probíhají paralelně vůči sobě a dosahují maximálně šedotónové úrovně uprostřed samotné linie – jedná se o černé (tmavé) body. Otisk prstu je násobený dvěma maskami h_t a h_b , tyto masky mají navzájem posunutou fázi o 180 stupňů. Po aplikaci algoritmu je nutné ve výsledném obrazu provést kontrolu a úpravu porušení papilárních linií a zrušení zlomů v průběhu papilárních linií [16].

Detekce papilárních linií pomocí schématu RAT rozděluje obrázek na bloky o velikosti 8×8 pixelů. V této oblasti je spočtená průměrná úroveň šedé a levá část 8×4 pixelů se nastaví na tuto získanou hodnotu. Poté je posunuto okno o 4 body doprava. Pokud je dosažený pravý okraj, tak se posouvá okno o 8 bodů dolů a opět se začíná zleva [14].

Pro extrakci markantů je nutné následně **ztenčit papilární linie** na šířku 1 pixelu. V tomto případu je nutné, aby algoritmus ztenčování byl vše směrový, kdy v žádném směru nesmí ubývat papilární linie. Jednou z vhodných metod je Metoda Emyroglu, která používá dva typy bodů – RMP (Ridge Meeting Point) a RCP (Ridge Continuity Point) [14].

Obvykle se detekují dva základní **typy markantů** – ukončení papilární linie a vidlička. Typů markantů existuje mnohem více, avšak jsou kombinací těchto dvou základních typů. Ke každému markantu je pak uložena jeho pozice X, pozice Y, typ markantu (ukončení nebo vidlička) a gradient – směr pokračování papilární linie [14].

Metod pro porovnání otisků prstů existuje celá řada. Metoda založená na markantech, která je nejčastěji se vyskytující metodou, užívá generování globálního překryvu a následného hledání lokálního překryvu markantů. Také se zavádí oblast tolerance při hledání lokálního překryvu – porovnání. Dalšími způsoby porovnání otisků prstů je metoda založená na korelace, která využívá 2D korelace mezi vstupem a šablonou, a metoda porovnání založená na rozpoznání vzorů, která využívá neuronových sítí a je použitelnou i při malé ploše otisků (např. senzory na mobilních telefonech) [16].

2.4.1 Daktyloskopie

Daktyloskopie je jednou z nejstarších disciplín kriminalistické techniky, které se zabývají identifikací osob [43]. Platí následující daktyloskopické zákony:

- Papilární linie jsou pro každého z jedinců unikátní [14].
- Vzor, který tvoří papilární linie, zůstává po celý život jedince relativně neměnným [14].
- Dorůstáním kůže na povrchu prstů jsou papilární linie obnovovány. Papilární linie nemohou být pozmeněny nebo odstraněny, pokud není poškozena dermální vrstva kůže. Pokud však k poškození dojde, pak již na tomto místě nedojde k obnově papilárních linií [14].
- Konfigurační typy se mohou individuálně měnit, ovšem jedná se pouze o malé změny, které leží v tolerančních limitech a umožňují systematickou klasifikaci [14].

2.4.2 Typy otisků prstů, druhy markantů a třídy otisků prstů

Otisky prstů jsou děleny na základě způsobu pořízení následovně (viz obrázek 2.19):

- válený otisk (barvený, rolovaný),
- píchaný otisk (živý),

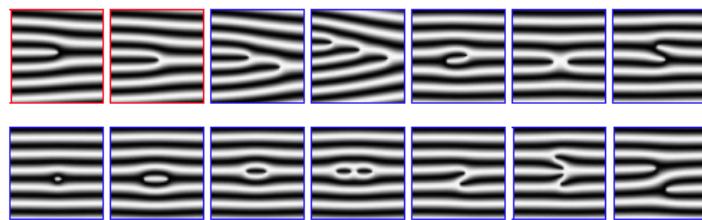
- latentní otisk (skrytý) [14].

Válené otisky jsou získávány způsobem rolování prstu po celé délce, kdy je snahou zachytit všechny detaity hřebenů papilárních linií. Píchané otisky vznikají přiložením bříška prstu k plochému povrchu – rolování není prováděno. Píchané otisky prstů pokrývají menší plochu než válené otisky prstů a typicky nejsou zkresleny právě procesem rolování. Oba tyto typy jsou získávány pomocí senzoru nebo off-line způsobem snímání – např. nanesením inkoustu na bříško prstu. Latentní otisky prstu jsou obvykle získávány z objektů, kterých se daná osoba dotkla – typickým prostředím pro práci s tímto typem otisku prstů je kriminalistika [28]. Ve srovnání s válenými a píchanými otisky, které mají obvykle dobrou kvalitu a obsahují mnoho informace, jsou latentní otisky těžší na správné získání – obvykle se používají speciální fyzikální a chemické metody (např. daktyloskopický prášek, kyanoakrylát nebo ninhydrin) [43].



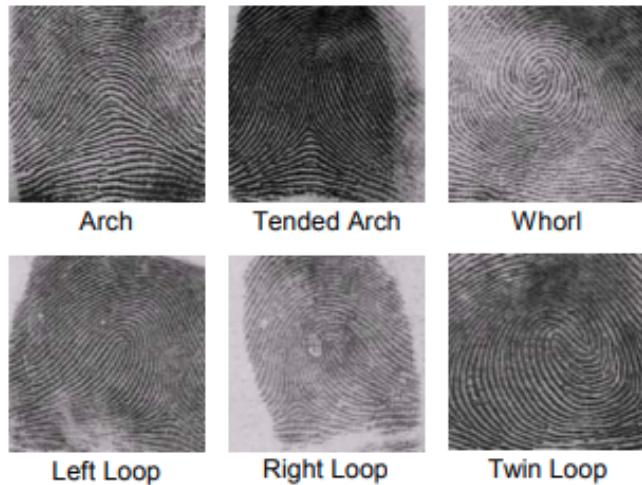
Obrázek 2.19: Válený, píchaný a latentní otisk [14]

Druhy markantů slouží pro základní rozpoznávání otisků prstů. Existuje mnoho typů markantů (viz obrázek 2.20), avšak obvykle se u přístupových systémů pracuje pouze s typem markantu ukončení nebo vidličky: ukončení (*Line Ending*), jednoduchá vidlička (*Simple Bifurcation*), dvojitá vidlička (*Double Bifurcation*), trojitá vidlička (*Triple Bifurcation*), hák (*Hook*), křížení (*Crossing*), boční kontakt (*Side Contact*), bod (*Point*), interval (*Interval*), jednoduchá smyčka (*Single Whorl*), dvojitá smyčka (*Double Whorl*), jednoduchý most (*Single Bridge*), dvojitý most (*Twin Bridge*), průsečná linie (*Through Line*) [14].



Obrázek 2.20: Typy markantů ve vyjmenovaném pořadí výše [14]

Otisky prstů mohou být také rozřazeny do následujících tříd na základě charakteristického průběhu papilárních linií (viz obrázek 2.21): oblouk (*Arch*), klenutý oblouk (*Tended Arch*), spirála / závit (*Whorl*), levá smyčka (*Left Loop*), pravá smyčka (*Right Loop*), dvojitá smyčka (*Twin Loop*) [14]. V daktyloskopických (kriminalistických) systémech se obvykle používají všechny třídy otisků prstů s výjimkou dvojité smyčky [14].



Obrázek 2.21: Typy tříd otisků prstů [14]

2.4.3 Měření kvality otisků prstů

Verifikace otisku prstu je často ovlivněna kvalitou snímku otisku prstu poskytnutým senzorem. Dokonce nejlepší verifikační systémy mají mnohdy problémy se zašuměnými a nekvalitními snímky. V mnoha studiích bylo zjištěno, že změna typu senzoru může mít vliv na rozpoznávání – každý z nich poskytuje snímek jiné kvality. Vlhkost / suchost otisku prstu, šum, nečistoty jak na prstu, tak na senzoru nebo onemocnění mohou mít důležitý vliv na biometrický systém. Existuje mnoho různých způsobů pro měření kvality otisků prstu – vedle způsobů zaměřených na analýzu šířky a kontinuity papilárních linií se může využít i např. analýza na základě jasnosti papilárních linií [3].

Měření na základě šířky papilárních linií:

- úroveň jistoty orientace – měření koncentrace energie okolo dominantního směru hřebenů s využitím gradientu intenzity,
- směrodatná odchylka výstupu Gaborova filtru – snímek je filtrován Gaborovým filtrem, pro otisk prstu s kvalitními oblastmi bude jeden z mnoha výstupů filtru vyšší jak ostatní, směrodatná odchylka m výstupu filtru je použita pro ohodnocení kvality [3].

Měření na základě propojení (kontinuity) papilárních linií:

- kvalita lokální orientace – průměrný absolutní rozdíl úhlu papilární linie s okolními bloky obrazu,
- kontinuita pole orientací – u kvalitních snímků předpokládá plynulý průběh hřebenů a údolí v lokálně konstantním směru, analyzuje náhlé změny směru mezi sousedními bloky [3].

2.5 Typy senzorů

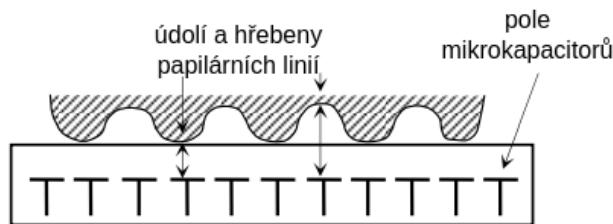
První techniky pro získávání otisků prstů byly typu off-line. Prsty byly očerněny inkoustem a přitisknutý nebo rolovány po papírové kartě. Tato karta byla dále naskenována a

převedena tak do digitální podoby. Tato technika je stále užívanou v kriminalistice pod názvem daktoloskopická karta. Vedle této metody se však v přístupových systémech užívají techniky pro živé naskenování otisku prstu pomocí biometrického senzoru – tento způsob je mnohem více efektivním, levným a uživatelsky přívětivým. Může být tedy nasazen v široké oblasti aplikací [36]. Existuje mnoho typů senzorů pro snímání otisků prstů, tato kapitola popisuje pouze dosud nejužívanější – kapacitní a optický senzor.

2.5.1 Kapacitní senzor

Jedná se o jeden z nejužívanějších typů senzorů. Schéma kapacitního senzoru je znázorněno na obrázku 2.22. Kapacitní senzor je tvořen 2D polem mikrokapacitorů, které jsou umístěny na čipu. Po přiložení prstu se mezi povrchem prstu a každou z křemíkových destiček vytvoří malý elektrický náboj, jehož hodnota závisí na vzdálenosti mezi povrchem prstu a kapacitními destičkami. Hřebeny a údolí papilárních linií mají jinou hodnotu elektrického náboje a odlišnou výslednou intenzitu ve finálním snímku. Každý výrobce kapacitních senzorů má obvykle vlastní nastavenou hodnotu, jak rozlišit hřebeny a údolí papilárních linií, kterou je obvykle obtížné upravit. Kapacitní senzory nelze zneužít přiložením fotografie otisku prstu – senzor měří dané vzdálenosti a nasnímán může být pouze 3D povrch [36].

Povrchovou vrstvu kapacitního senzoru je třeba chránit před chemickými látkami, které jsou přítomné v potu (např. sodík). Povrchová vrstva senzoru nesmí být však příliš silná, protože pak je snížena rozlišovací schopnost senzoru mezi hřebeny a údolími papilárních linií. Povrchová vrstva nesmí být ani příliš tenká, protože pak není odolná vůči mechanickému poškození [36].

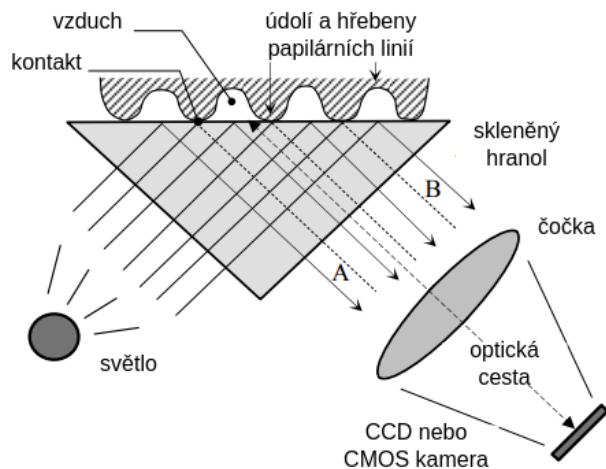


Obrázek 2.22: Schéma kapacitního senzoru [36]

2.5.2 Optický senzor

Optický senzor FTIR (Frustrated Total Internal Reflection) je nejstarší a dosud nejužívanější způsob pro získávání snímků otisků prstů. Schéma optického senzoru FTIR je znázorněno na obrázku 2.23. Prst je přiložen na povrch hranolu, vrcholy papilárních linií se povrchu přímo dotýkají a hřebeny papilárních linií zůstávají od povrchu vzdálené. Levá část hranolu je obvykle nasvícena např. LED (Light-Emitting Diode) diodami. Toto světlo, které vstupuje do hranolu, je odráženo údolími a náhodně rozptylováno hřebeny papilárních linií. Hřebeny se pak díky neschopnosti odrážet světlo jeví na obrazu jako tmavé oblasti ve srovnání s bílými oblastmi údolí, které světlo odrážejí. Odražené světlo od povrchu hranolu dopadá přeš čočku na CCD (Charge-Coupled Device) nebo CMOS (Complementary Metal–Oxide–Semiconductor) kamery, která nasnímá výsledný obraz. Stejně jako kapacitní senzory snímá FTIR senzor 3D povrch, a proto nemůže být podvrhnut např. vytisknutou fotografií [36].

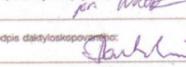
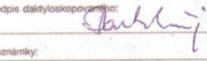
Pokud je prst velmi suchý, nemusí docházet k rovnoramennému kontaktu s povrchem senzoru FTIR. Někteří výrobci FTIR senzorů proto používají typicky silikonovou vrstvu na povrchu hranolu, která zlepšuje optický kontakt kůže s hranolem. Často se z důvodu ceny používá na výrobu hranolu a čoček plast místo skla a CMOS kamery místo nákladnějších CCD kamer [36].



Obrázek 2.23: Schéma FTIR senzoru [36]

2.5.3 Daktyloskopická karta

Daktyloskopická karta je off-line způsobem snímání, která se stále užívá v kriminalistice. Obecně obsahuje osobní údaje o daktyloskopované osobě, otisky prstů a další záznamy týkající se daktyloskopování. Příklad daktyloskopické karty Policie ČR je zobrazen na obrázku 2.24.

 KARTA			
Příjmení: <input type="text"/> Datum narodení: <input type="text"/> R. &: <input type="text"/> Jméno: <input type="text"/> Podpis příjmení: <input type="text"/> Misto narodení: <input type="text"/> Národnost: <input checked="" type="checkbox"/> česká Umravený ročník: <input type="text"/> KOTEHLÍKY Pohlaví: muž <input type="checkbox"/> ženec <input type="checkbox"/>			
Trvalý pobyt: <input type="text"/> Výška v cm: <input type="text"/> 181 Číslo, datum a místo vydání dokladu totožnosti (IČ, pas, aj.): AA 222222 18.10.1989, Kosojedy			
Daktyloskopován dne: 1.1.2001 Kde: 5 PČR OŘ Jihlava Pro: krádež Ev. číslo foto: 567/2001			
Podpis daktyloskopujícího:  Podpis daktyloskopovaného:  Poznámky: <input type="text"/>			
Levá ruka (kontrolní otisky čtyř prstů) Kontrolní otisky palců Pravá ruka (kontrolní otisky čtyř prstů) <table border="1" style="width: 100%;"> <tr> <td>Levý</td> <td>Pravý</td> </tr> </table>		Levý	Pravý
Levý	Pravý		
<small>1) Uvedete dleborň daktyloskopovaného číslo. 2) Viz katalogový přehled směru teleskopu. 3) Uzaví se blauflátnou, pouze se předem zjistí náročnost na číslování.</small>			

Obrázek 2.24: Příklad daktyloskopické karty Policie ČR [43]

Daktyloskopická karta obsahuje následující otisky prstů:

- válené otisky pro každý prst (každý umístěn a označen ve zvláštním poli),
- píchaný otisk pro levý a pravý palec (každý umístěn a označen ve zvláštním poli),
- píchané otisky ostatních prstů (otisky umístěny ve dvou polích podle levé nebo pravé ruky) [43].

Při daktyloskopování je dbáno na to, aby kresba papilárních linií na získaných otiscích byla plynulá, dobře čitelná a nepřerušovaná. Také se snažíme o zachycení co největší plochy dané otiskované části prstu [43].

2.6 Generování syntetických otisků prstů

Generování syntetických otisků prstů je inverzní biometrickou úlohou. Syntetické otisky mohou být vytvořeny z náhodných bodů nebo ISO šablony a vygenerovaný otisk je obvykle dokonalým. Mohou být užity pro hodnocení výkonu nebo učení modelů neuronových sítí. Výhodou je, že nepodléhají ochraně osobních údajů, a proto mohou být snadněji zakomponovány do součástí vědeckých článků a prezentací výsledků [16].

Syntetické otisky prstů jsou nedílnou součástí pro učení modelů konvolučních neuronových sítí, zvláště při malé velikosti datasetu reálných otisků prstů získaných od pacientů trpících onemocněním. Model je však potřeba otestovat i na reálných otiscích, protože neuronová síť může mít tendenci se přetrénovat na pouze vygenerovaných syntetických otiscích prstů.

Jelikož syntetický otisk prstu je obvykle dokonalým, při generování je kladen důraz na generování poškození nebo šumu pro větší podobnost s reálným otiskem prstu. Na obrázku 2.25 je zobrazen příklad syntetického otisku prstu generovaným generátorem SFinGe, na obrázcích 2.26, 2.27, 2.28 a 2.29 se pak nachází příklad generování různých onemocnění do syntetických otisků prstů – bradavic, dyshidrózy, atopického ekzému a psoriázy.



Obrázek 2.25: Syntetický otisk prstu vygenerovaný generátorem SFinGe

Obrázek 2.26: Generované onemocnění bradavice do syntetického otisku prstu

Obrázek 2.27: Generované onemocnění dyshidrózy do syntetického otisku prstu



Obrázek 2.28: Generované onemocnění atopického ekzému do syntetického otisku prstu



Obrázek 2.29: Generované onemocnění psoriázy do syntetického otisku prstu

Kapitola 3

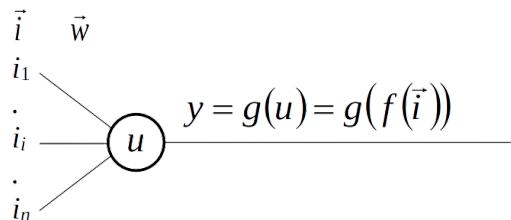
Detekce a klasifikace pomocí neuronových sítí

Neuronová síť je abstraktním modelem, který simuluje chování lidského mozku. Od roku 1943, kdy bylo poprvé popsáno chování neuronů a implementovaná jednoduchá neuronová síť využívající elektrické obvody, prošla problematika velikým vývojem [18]. Následující kapitola popisuje základní myšlenku funkce perceptronu, způsob trénování a učení neuronové sítě. Následně se práce zaměřuje speciálně na chování konvolučních neuronových sítí. Vedle obecného principu je popsán obsáhlý výčet dnes užívaných architektur – od jedné z prvních VGG-16 až po architektury z minulých let, jako je EfficientNet vyvinutá společností Google roku 2019. Následně jsou představeny modely pro detekci objektů – opět jsou kapitoly zaměřeny na popis začátků tohoto problému, až po přístupy ze současných let.

3.1 Umělý neuron a perceptron

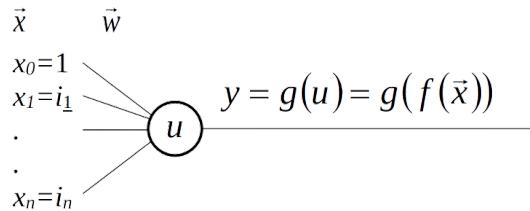
Pro umělý neuron se stal inspirací biologický neuron, který je základem nervové soustavy živých organismů. Lidský mozek obsahuje přibližně 10^{11} neuronů – 10^{14} synapsí, což je druh spojení dvou neuronů sloužící k předávání vznaků. Tento vysoký počet synapsí představuje paměťové schopnosti člověka a proces učení spočívá ve způsobu nastavování synaptických vaz jednotlivých neuronů [61].

Obecné schéma umělého neuronu je znázorněno na obrázku 3.1. Pro umělý neuron se používají na rozdíl od biologického neuronu pouze statické výstupní signály. Proměnná u představuje vnitřní potenciál, f bázovou funkci, g aktivační funkci, vektor i_1, i_2, \dots, i_n představuje vstup neuronu a vektor w_1, w_2, \dots, w_n je vektorem vah [61].



Obrázek 3.1: Schéma umělého neuronu [61]

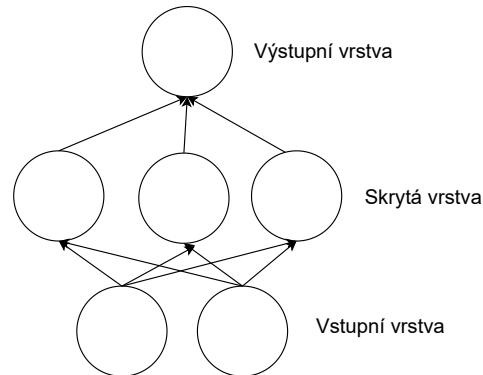
Nejznámější umělý neuron se nazývá perceptron. Schéma perceptronu je znázorněno na obrázku 3.2. Vstupní vektor x_0, x_1, \dots, x_n má fixně nastavenou nultou složku na hodnotu 1 z důvodu zjednodušení vztahů pro výpočet odezvy a pro učení. Váha w_0 je pak nastavena na hodnotu záporného prahu. Vnitřní stav má následující rovnici: $w_0x_0 + \sum_{i=1}^n w_i x_i$, kde x_0 představuje speciální stav bias a w_0 jeho váhu. Na vnitřní reprezentaci je aplikována aktivační funkce a získaný výstup je předán další vrstvě. Při zpětné propagaci jsou pak upraveny váhy neuronové sítě [61].



Obrázek 3.2: Schéma perceptronu [61]

3.2 Vícevrstvý perceptron

Vícevrstvý perceptron (Multilayer Perceptron – MLP) je klasickým typem neuronové sítě. Obecné schéma sítě MLP je zobrazeno na obrázku 3.3. Data jsou přivedena na vstup vstupní vrstvě, následně síť obsahuje několik skrytých vrstev, která zajišťuje různé úrovně abstrakce. Výstupní vrstva poté provádí výslednou predikci. Všechny vrstvy jsou plně propojené – každý perceptron z vrstvy je propojen s každým perceptronem následující vrstvy [6].



Obrázek 3.3: Schéma sítě MLP

Sítě MLP jsou vhodnými pro využití např. pro klasifikaci na základě atributů při tabulkových datasetech [6]. Jejich využití na problémy, jako je zpracování obrazu, je však velmi nevýhodné. MLP sítě obsahují velké množství parametrů a nejsou invariantní vůči posunutí. Pokud by byl obraz převeden na 1D vektor a zpracován MLP sítí, tak je ztracena prostorová informace [12].

3.3 Aktivační funkce

Aktivační funkce definuje, jak se suma vstupů bude transformovat na výstupy v průběhu průchodu neuronovou sítí. Aktivační funkce jsou obvykle diferencovatelné a různé aktivační funkce mohou být užity v různých částech neuronové sítě. Volba aktivační funkce má vliv na výkon a konečné výsledky zvoleného modelu [9].

Dobrá aktivační funkce by měla mít následující vlastnosti:

- Monotónní funkce – Aktivační funkce by měla být buď zcela nerostoucí, nebo neklesající. Pokud aktivační funkce monotónní není, pak zvýšení váhy neuronu může mít menší vliv na minimalizaci hodnoty chyby – průměrné hodnoty loss (viz sekce 3.4) přes celý trénovací dataset [1].
- Diferencovatelnost – U aktivačních funkcí je naším cílem, aby byly diferencovatelné, protože chceme spočítat změnu chyby vzhledem k vahám v daném okamžiku gradientního sestupu (viz sekce 3.5) [1].
- Rychlá konvergence – Aktivační funkce by měla rychle konvergovat k požadovaným hodnotám. Pokud je naším cílem získat např. dva různé výstupy o hodnotách 0 a 1, pak by měla aktivační funkce rychle konvergovat k těmto dvěma hodnotám [1].

Skryté vrstvy obvykle užívají stejnou aktivační funkci, která se obvykle liší od aktivační funkce zvolené pro výstupní vrstvu. Aktivační funkce pro výstupní vrstvu obvykle závisí na typu úlohy, kterou má neuronová síť realizovat (např. úloha binární klasifikace, klasifikace do více tříd) [9].

3.3.1 Aktivační funkce skryté vrstvy

Pro aktivační funkci skryté vrstvy se užívá diferencovatelná nelineární aktivační funkce. Model se pak dokáže naučit složitější funkce než model užívající lineární aktivační funkci. Řešené problémy reálného světa jsou totiž nelineárními [9].

Mezi známé aktivační funkce skryté vrstvy neuronové sítě patří:

- ReLU – ReLU (Rectified Linear Unit) je nejužívanější aktivační funkci. V kladné části nabývá derivace hodnoty 1, v záporné části je pak derivace nulová (tedy konstantní). Funkce a její derivace jsou monotónní [48]. Funkci ReLU lze matematicky vyjádřit jako:

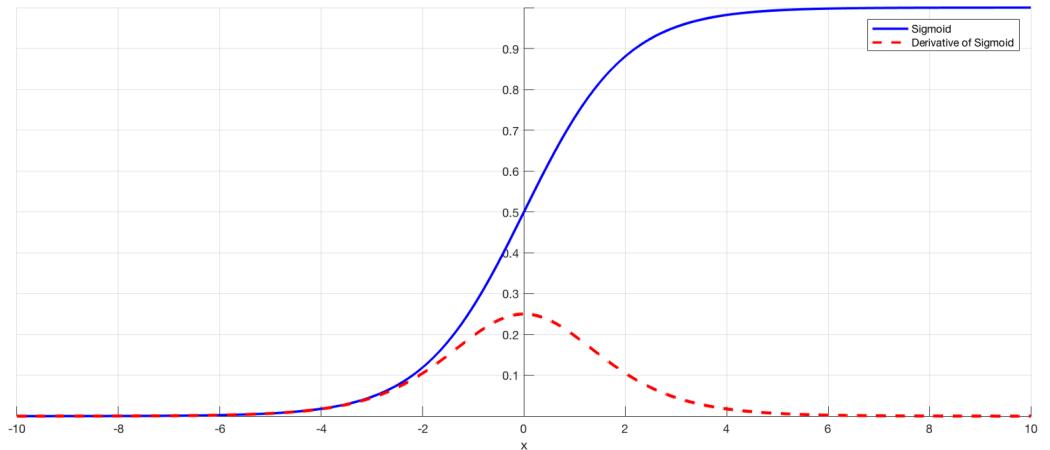
$$R(z) = \max(0, z) \quad (3.1)$$

- Sigmoida – Výstupem sigmoidy jsou hodnoty z intervalu (0, 1). Čím je výstup vyšší, tím bude hodnota blíže hodnotě 1, čím je výstup nižší, tím více bude hodnota blízká hodnotě 0. Funkce je monotónní, avšak derivace funkce monotónní není [48]. Funkce je vhodná pro predikci pravděpodobnosti díky právě jejímu definovanému oboru hodnot. S pojmem sigmoidy však souvisí problém mizejících gradientů. Jedná se o problém, kdy přidání vrstev využívající tuto aktivační funkci má za následek, že gradient loss funkce dosáhne k nule. Trénování modelu se tedy stává obtížným, což může vést k celkové nepřesnosti celé sítě. Vysoké změny vstupu funkce vedou k malým změnám na výstupu – hodnota derivace se blíží k nule. Tento problém je specifický pro sítě o více vrstvách [59]. Graf sigmoidy a její derivace je znázorněný na obrázku 3.4. Sigmoida je definována následujícím předpisem:

$$\sigma(z) = \frac{1}{1 + e^{-z}} \quad (3.2)$$

- Tanh – Rozsah výstupů je v intervalu $(-1, 1)$. Pro záporné vstupy je výsledek funkce negativním. Funkce je monotónní, derivace monotónní není [48]. Tato aktivační funkce je definována následovně:

$$\sigma(z) = \frac{e^z - e^{-z}}{e^z + e^{-z}} \quad (3.3)$$



Obrázek 3.4: Sigmoid funkce a její derivace [59]

3.3.2 Aktivační funkce výstupní vrstvy

Aktivační funkce výstupní vrstvy je závislá na řešeném problému. Obvykle je její výběr spjatý i s výběrem dané loss funkce (viz sekce 3.4), které jsou také charakteristické pro různé zvolené úlohy.

Aktivačními funkcemi užitými pro výstupní vrstvy neuronové sítě jsou:

- Linearita – Lineární aktivační funkce nemění váhovanou sumu vstupu a přímo vrací danou hodnotu. Běžně se užívá při regresi – predikci numerické hodnoty [9].
- Sigmoida – Sigmoida je vhodnou aktivační funkcí např. při binární klasifikaci [9].
- Softmax – Výstupem softmax je vektor hodnot, jejichž součet je jednotkový – jedná se o pravděpodobnost příslušnosti výstupu do dané třídy. Je vhodnou použít při klasifikaci do více než dvou tříd [9].

3.4 Loss funkce

S problematikou loss funkce souvisí pojem objektivní funkce – funkce, která má za úkol vyhodnocovat kandidátní řešení (např. množinu vah) a úkolem modelu je maximalizovat nebo minimalizovat právě tuto objektivní funkci. U neuronových sítí se typicky snažíme o minimalizaci chyby – objektivní funkce je zde právě nazývána loss funkcí [8]. Stejně jako u aktivační funkce výstupní vrstvy výběr dané loss funkce závisí na vybrané úloze, kterou řešíme:

- MAE (Mean Absolute Error) – Jedná se o nejjednodušší, avšak robustní loss funkci užitou pro regresi. MAE je ideální i při operaci s odlehlymi hodnotami. Výpočet je počítán jako průměr druhých mocnin mezi predikovanými a skutečnými hodnotami [8].
- Cross-entropy loss – Tato loss funkce je užívanou při problematice binární klasifikace a klasifikace do více tříd. Při binární cross-entropy loss je predikovaná hodnota porovnána se skutečnou třídou (hodnoty 0 nebo 1) a vypočítané skóre penalizuje pravděpodobnost na základě vzdálenosti od očekávané hodnoty. Penalizace je logaritmická, kdy pro malé vzdálenosti od skutečné hodnoty nabývá malých hodnot a naopak velkých hodnot pro velké rozdíly. Binární cross-entropy loss je vypočítaná jako průměr aplikace cross-entropy přes všechny vzorky. Pro aplikaci cross-entropy loss při klasifikaci do více tříd se používá one-hot kódování pro dané třídy. Skóre penalizace je vypočítané přes každou hodnotu vektoru získaným one-hot kódováním a zprůměrováno přes všechny vzorky [8].

3.5 Trénování neuronové sítě

Cílem algoritmu učení je minimalizace objektivní funkce, což je suma výsledků loss funkce přes všechny prvky datasetu. Snahou je nalézt takové parametry, které minimalizují hodnotu objektivní funkce. Gradienty jsou derivacemi objektové funkce s konkrétními parametry a počítají se pro konkrétní parciální derivaci – vektor reálných čísel. Minimalizuje se objektivní funkce $J(\theta)$ parametrizovaná parametry modelu θ pomocí aktualizace parametrů v opačném směru gradientu objektivní funkce $\nabla_{\theta}J(\theta)$. Parametr η udává learning rate – velikost kroku pro dosažení (lokálního) minima [47]. Při zpětné propagaci se aktualizují váhy neuronové sítě. Při trénování mohou být využity i různé regularizační techniky určené pro snížení komplexity modelu a problému potenciálního přetrénování – neschopnosti modelu generalizovat na testovacím datasetu.

3.5.1 Batch gradient descent

Batch gradient descent počítá gradient objektivní funkce s využitím celé datové sady. Pro vypočítání jedné aktualizace parametrů je tedy zapotřebí vypočítání gradientů pro celý dataset:

$$\theta = \theta - \eta \nabla_{\theta}J(\theta) \quad (3.4)$$

Tento výpočet však může být velmi pomalým a neřešitelným pro datasety, které se svým rozsahem nevezdou do paměti. Tato metoda nám také neumožňuje online aktualizaci modelu – např. při rozšíření datasetu v průběhu trénování [47].

3.5.2 Stochastic gradient descent

Stochastic gradient descent počítá aktualizaci parametrů pro každý prvek z trénovací sady $x^{(i)}$, který spadá do třídy $y^{(i)}$:

$$\theta = \theta - \eta \nabla_{\theta}J(\theta; x^{(i)}; y^{(i)}) \quad (3.5)$$

Tato metoda řeší problém provádění redundantních výpočtů pro rozsáhlé datasety, který je typický pro přístup batch gradient descent. Stochastic gradient descent je také mnohem

rychlejší, umožňuje online aktualizaci modelu a vyhledávání nových možných lepších potenciálních lokálních minim. Při snižování parametru learning rate vykazuje stochastic gradient descent stejně konvergenční chování jako batch gradient descent a konverguje k lokálnímu nebo globálnímu minimu v závislosti na typu úlohy [47].

3.5.3 Mini-batch gradient descent

Mini-batch gradient descent provádí aktualizaci pro každý mini-batch n trénovacích vzorků:

$$\theta = \theta - \eta \nabla_{\theta} J(\theta; x^{(i:i+n)}; y^{(i:i+n)}) \quad (3.6)$$

Snižování počtu aktualizací může vést ke stabilnější konvergenci a při této metodě může být využito maticových optimalizací typických pro současné knihovny strojového učení – výpočet mini-batch gradient descent je pak efektivním. Velikost mini-batch se pohybuje v širokém rozmezí (např. mezi 50 a 256) a liší se na základě řešené úlohy i výpočetního prostředí [47].

3.5.4 Regularizační techniky

Regularizace je množinou různých technik užitých za cílem snížení složitosti neuronové sítě a zabránění přetrénování – schopnosti modelu naučit se velmi dobře rozpoznávat trénovací data, avšak následné neschopnosti generalizovat na testovacím datasetu [42].

Techniky pro regularizaci jsou následující:

- L2 regularizace – L2 regularizace je nejužívanějším typem regularizace. Výpočet loss funkce je zde rozšířen o regularizační část Ω , která se dá vyjádřit jako:

$$\Omega(W) = \|W\|_2^2 = \sum_i \sum_j w_{ij}^2 \quad (3.7)$$

Jedná se tedy o sumu přes druhé mocniny všech vah w matice vah W . Regularizační loss je poté spočtena jako suma loss funkce $\mathcal{L}(W)$ a regularizace, která je váhována s využitím parametru α (míra regularizace):

$$\hat{\mathcal{L}}(W) = \frac{\alpha}{2} \Omega(W) + \mathcal{L}(W) \quad (3.8)$$

Za podpory L2 regularizace jsou hodnoty vah snižovány nepřímo k hodnotě nula. Váhy při každé aktualizaci jsou postupně nižší nezávisle na gradientu loss funkce [42].

- L1 regularizace – Regularizační část Ω je spočtena jako suma absolutních hodnot parametrů vah w v matici vah W .

$$\Omega(W) = \|W\| = \sum_i \sum_j |w_{ij}| \quad (3.9)$$

Regularizace je opět přičtena k loss funkci $\mathcal{L}(W)$ s využitím parametru α :

$$\hat{\mathcal{L}}(W) = \alpha \Omega(W) + \mathcal{L}(W) \quad (3.10)$$

Za podpory L1 regularizace jsou hodnoty vah snižovány přímo k hodnotě nula. Menší váhy obecně snižují vliv neuronů skryté vrstvy, tyto neurony se stávají zanedbatelnými a celková složitost neuronové sítě je snížena [42].

Je nutné brát ohled na zvolení správného parametru míry regularizace α . Pokud je hodnota α příliš vysoká, pak se model stává málo komplexním s rizikem tzv. *underfitting* – neschopnosti modelu se dostatečně naučit na trénovacích datech. Pokud je hodnota α příliš malá, pak bude model příliš komplexním a může docházet k tzv. *overfitting* – přetrénování [42].

- Dropout – Dropout může být využitý při trénování sítě, kdy s nastavenou pravděpodobností P je daný neuron při tréninku nevyužitým. Při např. zvoleném $P = 0,5$ se stane přibližně polovina neuronů neaktivní a nebudou tedy brány v potaz jako součást neuronové sítě – model se tedy stává jednodušším a potenciálně méně náchylným pro přetrénování [42].

3.6 Konvoluční neuronové sítě

Konvoluční neuronové sítě jsou speciálním typem neuronových sítí, které jsou v dnešní době vysokým způsobem užívané při zpracování obrazu a řeči. Jejich název je odvozen od matematické operace konvoluce, která je v jejich architektuře použita. Jelikož konvoluční neuronové sítě explicitně předpokládají, že vstupy jsou např. ve formě obrázků nebo zvuku, tak můžeme zakódovat určité vlastnosti do jejich architektury [12]. Výrazně se snižuje počet parametrů modelu oproti tzv. sítím MLP, které byly popsány v sekci 3.2. Obecně se konvoluční neuronová síť skládá z:

- vstupní obraz,
- konvoluční vrstvy,
- nelineární aktivační funkce,
- pooling vrstvy,
- plně propojené vrstvy [7].

Konvoluční neuronová síť dokáže zohlednit místní konektivitu – konvoluční filtr je postupně posouván po celém vstupu (v našem případě obrázku) podle určité velikosti a kroku. Dochází tedy k nalezení a porovnávání vzorů bez ohledu na to, kde se dané vzory v obrazu nachází. Váhy jsou sdílené a trénování je obecně efektivnější. Každý uzel není propojen s každým dalším uzlem [12].

Na každý pixel obrazu je obecně aplikováno více konvolučních filtrov. První vrstvy konvoluční neuronové sítě obsahují jednoduché filtry reagující pouze na hrany, filtry v jedné z posledních vrstev nám umožňují získat z obrazu detailnější informace (např. filtry reagující na barvy). Aplikací každého filtrov je získána aktivační mapa [12].

Další nedílnou součástí konvolučních neuronových sítí je pooling vrstva. Její parametry jsou nižší, jak velikost získané aktivační mapy konvoluční vrstvou – např. velikost 2×2 a parametr stride (krok posunu konvolučního filtrov) 2 pixely. Při tomto nastavení pooling vrstvy je aktivační mapa dvakrát zmenšena – každá dimenze je poloviční, počet pixelů v aktivační mapě je pak zmenšen čtyřikrát – např. při vstupu aktivační mapy o velikosti 6×6 (36 pixelů) je po aplikaci pooling vrstvy nová velikost aktivační mapy 3×3 (9 pixelů) [7].

Pooling vrstvy jsou neučitelnou částí konvolučních neuronových sítí. Nejznámější operací spojenou s pooling vrstvou je max pooling umožňující nalezení hodnot lokálních maxim. Pooling vrstva efektivně snižuje prostorové dimenze vstupu a pracuje s myšlenkou, že relativní umístění prvku (vlastnosti) vzhledem k ostatním prvkům (vlastnostem) je důležitější jak jeho samotné přesné umístění [7].

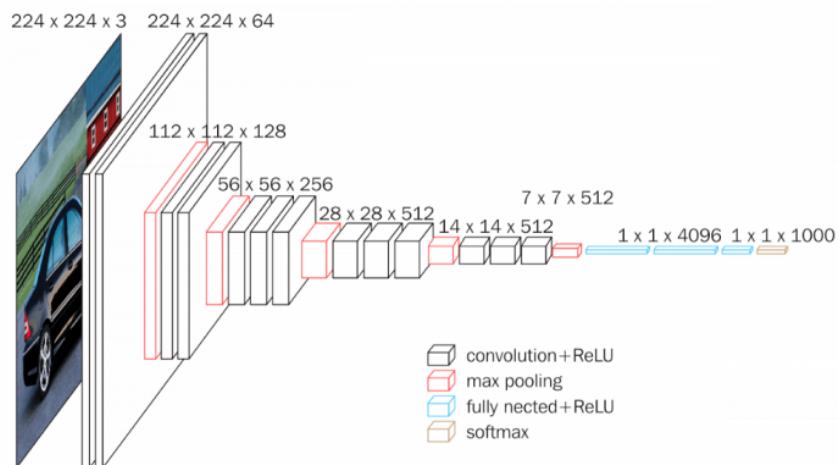
Obraz je poté po dostatečném zpracování s využitím konvolučních vrstev převeden na 1D vektor a přiveden na vstup plně propojených vrstev. Tyto plně propojené vrstvy pak tvoří výstupní část celého modelu a je určen výsledek na základě dané úlohy.

3.7 Existující architektury neuronových sítí pro zpracování obrazu

Páteřní konvoluční neuronová síť je základní částí každého detekčního modelu. V průběhu zpracovávání vstupu konvoluční neuronovou sítí jsou extrahovány rysy různých úrovní – v prvních vrstvách sítě se získávají rysy nízké úrovně, které mají vysoké rozlišení, avšak malý počet informací. Naopak na konci jsou získány rysy s nízkým rozlišením, ale vysokou hodnotou informace. Běžné principy detekce objektů jako Faster R-CNN nebo SSD užívají jako páteřní síť např. ResNet-50 nebo MobileNet. Jeden z nejnovějších přístupů EfficientDet užívá však speciální EfficientNet síť.

3.7.1 VGG-16

VGG-16 je jednoduchou a široce užívanou konvoluční neuronovou sítí představenou roku 2014. Není však příliš výpočetně optimalizovanou. Její architektura je zobrazena na obrázku 3.5. Vstupem sítě je barevný obrázek o rozměrech 224×224 . Všechny konvoluční vrstvy jsou výpočetně stejně složité a z rozlišení se složitost přesouvá do parametru počtu kanálů – při zmenšení rozlišení se dvakrát zvýší počet kanálů. Velikost jádra konvoluční vrstvy je 3×3 . Síť má v hlubších vrstvách více nelinearit a nižší počet parametrů. Vstup první plně propojené vrstvy má dimenze $7 \times 7 \times 512$. Po plně propojených vrstvách následuje predikce s využitím aktivační funkce softmax např. do 1 000 tříd pro dataset ILSVRC [44].



Obrázek 3.5: Architektura VGG16 [44]

3.7.2 ResNet-50

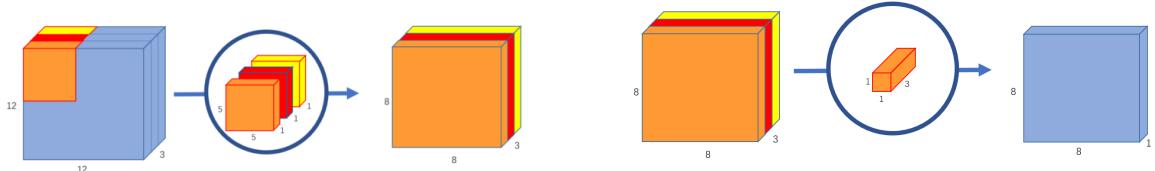
ResNet-50 je variantou modelu ResNet. Je složen z 48 konvolučních vrstev, jedné max pool vrstvy a jedné average pool vrstvy – celkově tedy obsahuje 50 vrstev. Tento model využívá reziduálních připojení - připojení vstupu ResNet-50 modelu na výstup, kdy je spočítána suma právě tohoto vstupu (identity) a zpracovaného vstupu. Výhodou tohoto dodatečného připojení identity na výstup je, že modelu nejsou přidány žádné nové dodatečné parametry. Tento princip řeší problém tendence vyšší chybovosti u modelů s vyšším počtem vrstev. Také je vyřešený problém mizejících gradientů. Dalšími známými variantami ResNet jsou např. ResNet-101 (obsahující 101 vrstev) nebo ResNet-152 (obsahující 152 vrstev) [29].

3.7.3 MobileNet

Architektura MobileNet byla poprvé představena roku 2017. Svojí myšlenkou se vyhýbá trendu tvoření komplikovanějších sítí o více vrstvách pro dosažení vyšší přesnosti. Místo toho se autoři zaměřili na výpočetně limitované platformy – na tvorbu vhodné sítě pro mobilní a vestavěné aplikace, robotiku, nebo samořídící auta. MobileNet využívá tzv. hloubkově oddělitelné konvoluce (Depthwise Separable Convolutions) [25].

Princip základních kroků hloubkově oddělitelné konvoluce je zobrazen na obrázcích 3.6 a 3.7. Při hloubkově oddělitelné konvoluci je v první části provedena operace konvoluce bez změny hloubky – každé z jader je aplikováno pouze na 1 odpovídající kanál vstupu (nikoliv na všechny jako u běžné operace konvoluce). Následně probíhá bodová konvoluce – rozměry velikosti jádra pro bodovou konvoluci jsou 1×1 [58].

Výhodou tohoto přístupu oproti klasické operaci konvoluce je rapidní snížení počtu aplikace operace násobení – např. při velikosti batche obrázků o velikosti 256 s rozměry jednoho barevného snímku $12 \times 12 \times 3$ pixely se jedná o snížení operací násobení z hodnoty 1 228 800 na 52 952. Síť je pak schopná zpracovat více dat za nižší čas [58].

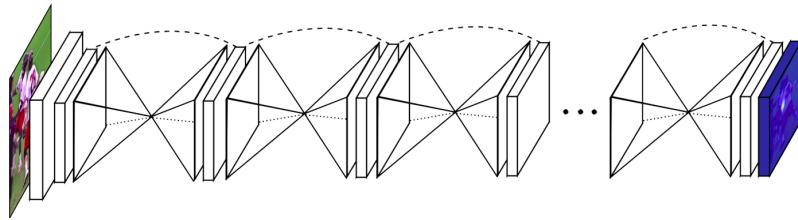


Obrázek 3.6: Hloubkově oddělitelná konvoluce - použití tří konvolučních jader [58]

Obrázek 3.7: Bodová konvoluce - transformace 3-kanálového obrazu na 1-kanálový [58]

3.7.4 Hourglass

Hourglass je typem konvoluční neuronové sítě představené v roce 2016. Poprvé byla využita na problematiku predikce pozice lidského těla. Hlavním úkolem této architektury je zachytit informaci přes různá rozlišení vstupního obrazu. Cílem této symetrické topologie (viz obrázek 3.8) je převést obraz na velmi nízké rozlišení pomocí pooling vrstev, následně probíhá zvětšení rozlišení. Při malém rozlišení síť dokáže efektivně pracovat s velkým kontextem, avšak se ztrácí informace o pozici. Při zvyšování rozlišení se naopak ztrácí hodně kontextu. Proto síť kombinuje rysy přes více rozlišení [39].



Obrázek 3.8: Symetrická topologie Hourglass [39]

3.7.5 EfficientNet

Pokud chceme dosáhnout vyšší přesnosti modelu, je běžným způsobem škálovat model za pomoci využití větší páteřní sítě (škálování do hloubky) nebo zvětšením rozlišení vstupního obrázku (škálování rozlišení). Přestože tyto metody individuálního škálování pouze na základě jednoho z faktorů zvyšují přesnost modelu, obvykle vyžadují zdlouhavé ruční ladění a u sítí obsahujících mnoho vrstev se může vyskytnout problém saturace a mizejících gradientů. Například ResNet dokáže být škálován z verze ResNet-18 na verzi ResNet-200 zvýšením počtu vrstev, řeší problém mizejících gradientů na základě funkce identity, avšak se jedná o rozsáhlou síť s mnoha parametry a dlouhým trénovacím časem. Cílem společnosti Google bylo navrhnout síť o méně parametrech, která dokáže být nasazená např. na mobilních zařízeních nebo méně výkonných strojích a je méně náročnou na výpočetní zdroje a trénovací čas. V roce 2019 byla poprvé prezentována EfficientNet se základním principem efektivního složeného škálování modelu [54, 53].

Můžeme provést následující typy škálování:

- škálování rozlišení – pokud má obraz vyšší rozlišení, můžeme z něj získat komplexní rysy a detailnější textury,
- škálování do hloubky – pokud chceme zpracovávat více komplexní rysy (obraz s vyšším rozlišením), tak síť musí obsahovat více vrstev,
- škálování do šířky – pro zpracování komplexnějších rysů a detailnějších struktur v obrazu síť potřebuje více kanálů (více získaných map rysů) [54, 53].

EfficientNet umožňuje provádět všechny tři formy škálování zároveň pomocí heuristiky, což se dle experimentů projevilo na zlepšení celkového výkonu neuronové sítě ve srovnání s individuálním škálováním vždy pouze na základě jednoho z faktorů užívaným v minulosti. Bylo zavedeno několik verzí modelu B0 – B7, kdy B0 je základním modelem, od kterého jsou všechny další modely odvozeny. Velikost vstupního obrazu pro model B0 je 224×224 pixelů. Architektura B0 byla navrhnuta s využitím NAS (Neural Architecture Search) s využitím optimalizace přesnosti a efektivity (FLOPS) [54, 53].

NAS je přístupem, který slouží k nalezení co nejlepší architektury neuronové sítě pro daný problém. Jedná se o automatizaci procesu ručního ladění modelu a snahu o nalezení složitějších architektur na základě učení. Ve vyhledávacím prostoru je testováno velké množství možných architektur a je vybrán model maximalizující fitness funkci – model odpovídající nejlépe řešení daného problému [20].

Při škálování je pracováno s koeficienty škálování $\alpha = 1,2$ (škálování do hloubky), $\beta = 1,1$ (škálování do šířky) a $\gamma = 1,15$ (škálování rozlišení). Pokud se tedy rozlišení obrazu zvýší o 15 %, pak se počet vrstev sítě má zvýšit o 20 % a počet kanálů o 10 % [54, 53].

3.8 Přístupy pro detekci objektů

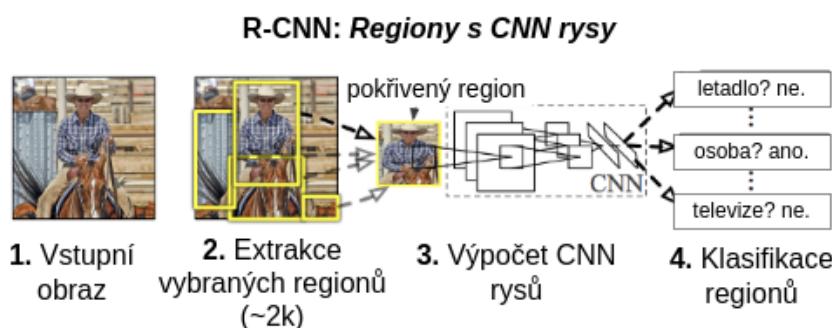
Problematika detekce objektů je jednou ze základních úloh strojového učení. Úkolem je vykreslit predikovaný bounding box okolo daného objektu. Naivní přístup k detekci by mohl být založen na pohyblivém okně přes dané pixely obrazu, kdy každé podokno je ohodnocené na základě predikce modelu skóre pro danou třídu. Podokna s vysokým skórem by pak byly částmi výsledného bounding boxu. Tento přístup je však výpočetně náročný. Proto se užívá jiných moderních přístupů.

3.8.1 R-CNN

R-CNN (Region Based Convolutional Neural Network) byla poprvé představena v roce 2014 [22]. Tento přístup kombinuje extrakci rysů z vybraných oblastí obrazu a konvoluční neuronovou síť. Je vybráno pouze např. 2 000 oblastí na základě algoritmu selektivního výběru. Pro predikci není tedy využit velký počet oblastí, jako by tomu mohlo být u naivního algoritmu využívajícího posuvného okna, ale pouze zmenšený počet oblastí. Algoritmus selektivního výběru užívá greedy algoritmus (výběr lokálního optima v každé fázi) pro rekurzivní spojování podobných oblastí do regionů větší velikosti [19].

Princip R-CNN modelu je znázorněn na obrázku 3.9. Regiony jsou poté konvertovány na rozměry na základě modelu konvoluční neuronové sítě, která extrahuje vektory rysů o 4 096 dimenzích. Extrahované vektory rysů jsou klasifikovány pomocí klasifikátoru SVM (Support Vector Machines) do daných tříd. Algoritmus vedle predikce dané třídy objektu také predikuje čtyři hodnoty offsetu pro získání vyšší přesnosti polohy bounding boxu [19].

Nevýhodami daného přístupu je stálá vysoká doba trénování a celková nepoužitelnost nasazení v reálném čase. Regionů je stálá vysoký počet, což vyžaduje tisíce zpětných průchodu modelem pro detekci objektů [19].

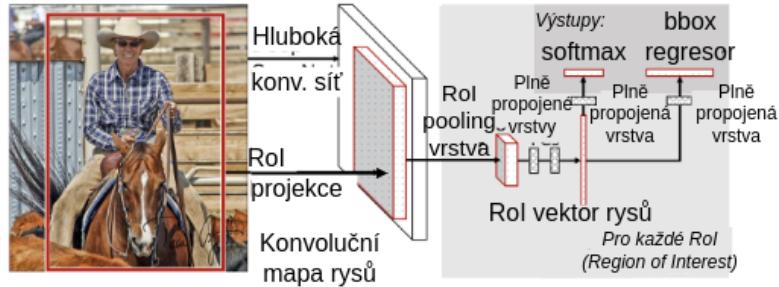


Obrázek 3.9: Princip R-CNN [19]

3.8.2 Fast R-CNN

Ross Girshick, jeden z předních autorů přístupu R-CNN, v roce 2015 vyřešil některé z nevýhod předešlého přístupu [21]. V R-CNN byl vysoký počet regionů postupně zpracováván konvoluční neuronovou sítí, v Fast R-CNN je na vstup modelu přiveden celý vstupní obraz. Je získána konvoluční mapa rysů, ze které identifikujeme vhodné regiony, které jsou s využitím ROI pooling vrstvy převedeny na pevnou délku. Tyto vektory mohou být klasifikovány pomocí MLP sítě a softmax aktivační funkce. Vedle predikované třídy získáváme i hodnoty offsetu pro daný bounding box [19].

Fast R-CNN je znatelně rychlejší jak R-CNN, protože z daného obrazu nemusí být pokaždé vybíráno několik tisíc regionů, které jsou následně zpracovávány modelem. Na obrázku 3.10 je znázorněn její princip. Operace konvoluce je aplikována pouze jednou pro každý obraz, ze které je poté zpracována mapa rysů. Algoritmus selektivního výběru je však stále časově náročným procesem, který je fixně implementovaný a není schopný procesu učení a tedy i možného zlepšení přesnosti při hledání vhodných regionů v obrazu [19].

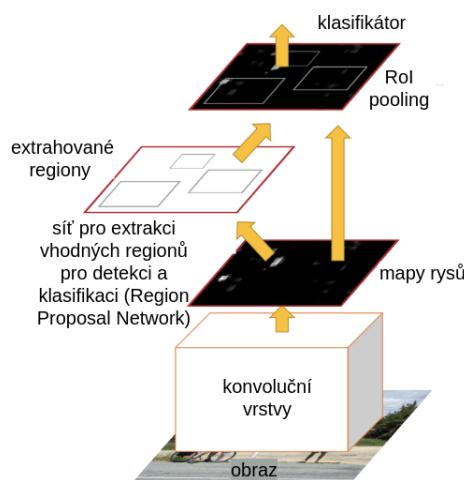


Obrázek 3.10: Princip Fast R-CNN [19]

3.8.3 Faster R-CNN

Faster R-CNN je metodou představenou v roce 2016, která řeší hlavně nevýhody R-CNN a Fast R-CNN z hlediska využívání algoritmu selektivního výběru [45]. Jelikož se jedná o algoritmus, který ovlivňuje výkon neuronové sítě, bylo využito myšlenky zrušení tohoto algoritmu. Princip modelu Faster R-CNN je zobrazen na obrázku 3.11. Celý obraz je opět převeden na vstup konvoluční neuronové sítě, kdy je vygenerována mapa rysů. Je ovšem využito i druhé oddělené sítě, která predikuje vhodné oblasti pro následnou detekci. Tyto oblasti jsou po změně dimenzí při průchodu ROI pooling vrstvou využity pro následnou klasifikaci obrazu v dané oblasti a je opět predikována i odchylka pro bounding boxy [19].

Faster R-CNN je díky vylepšením nedostatků přístupů R-CNN a Fast R-CNN znatelně rychlejší a je tedy vhodnou pro nasazení v reálných aplikacích [19].

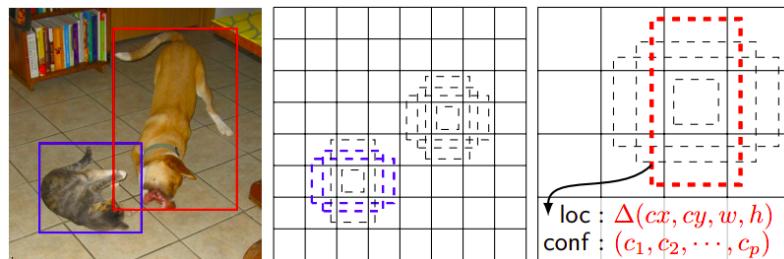


Obrázek 3.11: Princip Faster R-CNN [19]

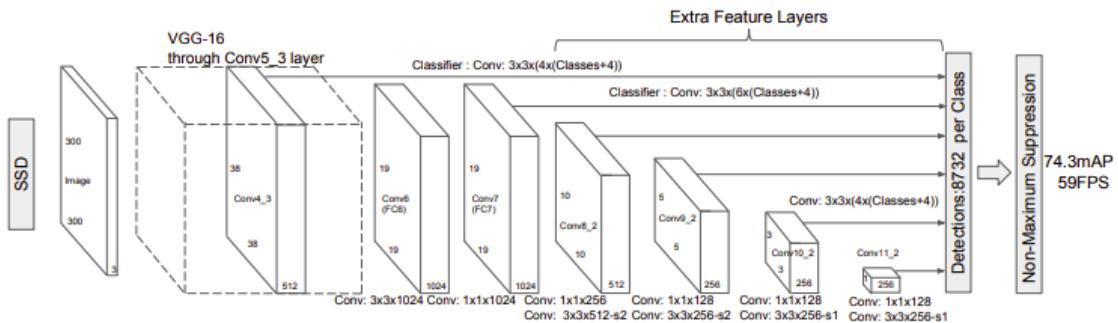
3.8.4 SSD

SSD (Single Shot Detector) je přístupem pro detekci objektů představeným roku 2016, který je založen na odlišném přístupu než Faster R-CNN. Principem Faster R-CNN je zpočátku hledat vhodné regiony a poté u těchto vybraných regionů provést detekci a klasifikaci. Fáze hledání vhodných regionů je u metody SSD zrušena. Místo toho se v několika konvolučních vrstvách o různém rozlišení provádí predikce bounding boxů a třídy pro každý z daných objektů (viz schéma architektury 3.13). Pro každý objekt je při trénování predikováno 8 732 bounding boxů – díky tomuto vysokému množství získáme větší pokrytí dané lokace v obrazu a tak možnost nalezení bounding boxu, který má nejvyšší překryv se skutečným anotovaným bounding boxem objektu (viz obrázek 3.12). SSD následně vybírá právě na základě skóre překryvu bounding boxy se skóre vyšším jak 50 %. Následně jsou eliminovány duplicitní bounding boxy a je vybráno pouze 100 bounding boxů s nejvyšším skóre jistoty pro každý obrázek [34, 62].

Síť díky predikcím polohy bounding boxů a příslušnosti do třídy v různých měřítkách dokáže predikovat objekty rozličných velikostí. SSD je přístupem, který zvláště díky více vrstvám pro predikci v různých měřítkách objektu dosahuje dobré přesnosti za snížení trénovacího času a může být tedy nasazeným v reálných aplikacích. Také dosahuje dobré přesnosti i na datasetech o nižším rozlišení [34, 62].



Obrázek 3.12: Princip získání většího počtu bounding boxů pro daný objekt [34]



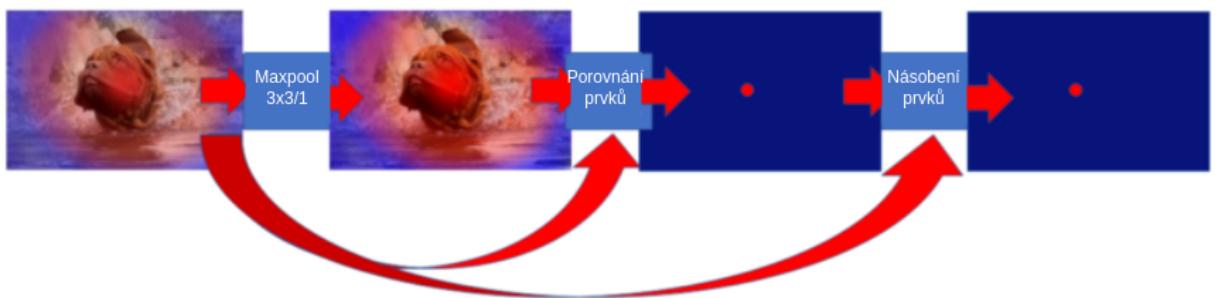
Obrázek 3.13: Architektura modelu SSD [34]

3.8.5 CenterNet

Mnoho přístupů k detekci (např. SSD) využívá predikování tisíců bounding boxů pro jeden objekt a vybírá bounding boxy s vysokou hodnotou překryvu. Tento přístup má následující nevýhody – složitost je úměrná druhé mocnině počtu predikcí a model je nucen

vypočítávat předpovědi pro všechny bounding boxy, což je operace náročná na čas i výpočetní prostředky. CenterNet využívá jiný způsob predikce – pro každý objekt je vytvořena tzv. *center confidence heatmap*, kdy ve středu objektu nabývá hodnota distribuce funkce svého maxima (a tedy nejvyšší jistoty). Vedle generování této mapy jsou také jako u jiných detektorů predikovány hodnoty pro rozměry bounding boxů a offsety související se středem bounding boxů predikovanými mapou jistoty. CenterNet predikuje *center confidence heatmap* pro každou třídu [2].

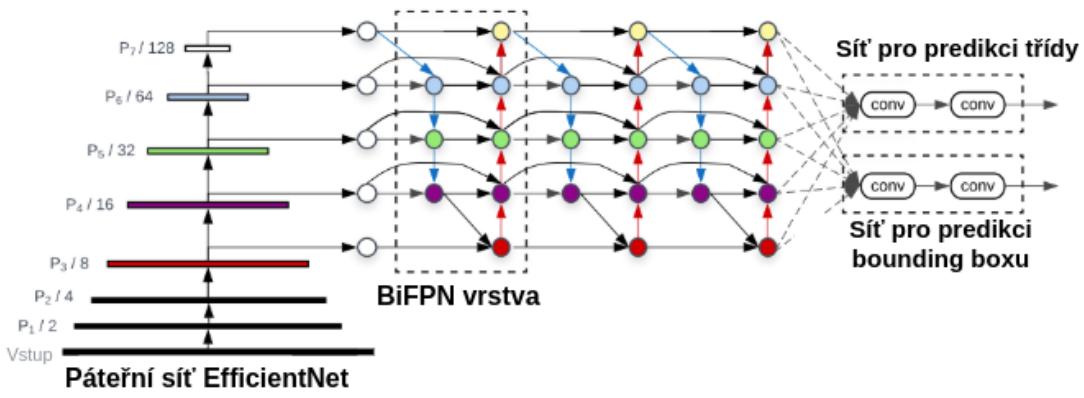
Právě užití *center confidence heatmap* nám umožňuje okamžitě odstranit irrelevantní predikce bez nutnosti predikování celého bounding boxu, princip analýzy je zobrazen na obrázku 3.14. Na mapu je aplikován max pooling a okolo samotného centra je vytvořen malý plochý region. Následně se boolovsky porovnávají jednotlivé prvky mezi vstupem a získaným výstupem po aplikaci max pooling – je vytvořena nová tepelná mapa, kdy jsou lokální maxima nastavena na hodnotu 1, zbytek je nastaven na hodnotu 0. Tato mapa se vynášobí s původním vstupem a výsledné skóre jistoty je vyprahováno. Tímto způsobem se dále pracuje jen s bounding boxy, které odpovídají lokálním maximům [2].



Obrázek 3.14: Princip analýzy *center confidence heatmap* [2]

3.8.6 EfficientDet

EfficientDet je modelem vyvinutým společností Google v roce 2020. Díky svému nízkému počtu parametrů se řadí k sítím s nižšími nároky na výpočetní zdroje a vyšší rychlostí trénování. Pro páteřní síť není užita např. MobileNet nebo ResNet-50 jako u předchozích metod, ale opět společností Google vyvinutá síť EfficientNet. EfficientDet neužívá FPN (Feature Pyramid Network) jako běžné algoritmy pro detekci. Při užití FPN jsou všechny rysy různých rozlišení posuzovány stejným způsobem a je jim tedy přikládána totožná důležitost. FPN je nahrazeno BiFPN (Weighted Bi-directional Feature Pyramid Network), který umožňuje analýzu a získání pouze důležitých rysů. Architektura modelu EfficientDet je znázorněna na obrázku 3.15. Každá z komponent sítě – páteřní síť EfficientNet, BiFPN a síť pro detekci bounding boxu a predikci třídy objektu – má jeden ze složených faktorů škálování a je prováděno složené škálování rozlišení, škálování do hloubky a škálování do šířky pro každou z těchto částí [55, 56].



Obrázek 3.15: Architektura EfficientDet [55]

Verze EfficientNet jsou značeny v rozsahu B0 – B7 a verze EfficientDet označením D0 – D7. Každá z verzí EfficientDet má také přiřazenou hodnotu ϕ , která figuruje ve výpočtu dalších charakteristik celého modelu [55, 56].

Kapitola 4

Návrh řešení a implementace

Tato kapitola se zabývá návrhem a implementací řešení. V první řadě jsou popsány dvě nejdůležitější části práce – dataset otisků prstů a použité modely. Následně je kladen důraz na předzpracování vstupního otisku prstu ještě na přivedení na vstup samotné konvoluční neuronové síti. Všechna onemocnění také musí být anotována bounding boxy s danými třídami.

4.1 Databáze otisků prstů

Databáze onemocnění otisků prstů získaných od pacientů je součástí stále pokračujícího výzkumu ve výzkumné skupině STRaDe na Fakultě informačních technologií v Brně. Ve své práci se zabývám v současné době pouze následujícími onemocněními – atopickým ekzémem, bradavicemi, psoriázou a dyshidrózou. Kompletní databáze však obsahuje další onemocnění, z nichž mnohá byla popsaná v kapitole 2.3. Databáze však není příliš rozsáhlou co se týče jednotlivých kategorií, a proto je nutné dataset rozšířit i o syntetické otisky prstů vygenerované generátorem.

Pro generování syntetických otisků bylo využito prostředí generátoru SFinGe, syntetickým otiskům je následně vygenerováno onemocnění generátorem popsaným v bakalářské práci Generování onemocnění kůže do syntetických otisků prstů z SFinGe [52], v bakalářské práci Detekce a klasifikace poškození otisku prstu s využitím neuronových sítí [57] byl generátor následně upraven a přidána funkce automatického anotování bounding boxů onemocnění v JSON formátu. S využitím tohoto generátoru byla vygenerována onemocnění atopického ekzému, dyshidrózy a bradavic. Pro generování onemocnění psoriázy byl využit generátor popsaný v bakalářské práci Generování projevu kožního onemocnění do syntetických otisků prstů z generátoru Anguli [32].

Tabulka 4.1 udává počty zastoupení jednotlivých kategorií snímků celkového datasetu užitého v této diplomové práci. Další databází živých otisků prstů je databáze americké organizace NIST – Národní institut standardů a technologie. Jedná se konkrétně o NIST Special Database 4, která obsahuje 4 000 snímků získaných převážně z daktyloskopických karet s neznámými běžnými poškozeními. Z této databáze byly ručně vybrány snímky otisků prstů, které nejsou poškozené. Tato část datasetu otisků prstů byla doplněna ještě vygenerovanými syntetickými snímky z generátoru SFinGe, které nebyly poškozeny žádným onemocněním.

Onemocnění	Reálné otisky	Syntetické otisky
Atopický ekzém	286	411
Bradavice	42	225
Psoriáza	98	271
Dyshidróza	96	200
Bez poškození	1131	400

Tabulka 4.1: Celkové počty snímků datasetu

Celkový dataset je tvořen 3 160 snímků, reálné otisky tvoří 52,3 % a syntetické otisky 47,7 %. Toto zastoupení je tedy celkově přibližně rovnoměrné a je tedy kladen větší důraz na trénování modelu na snímcích reálných pacientů, kdy onemocnění se často může projevovat více komplexně než způsobem implementovaným v generátoru onemocnění do syntetických otisků prstů.

Dále je dataset rozdělen na trénovací, validační a testovací. Na konkrétním testovacím datasetu budou dále vyhodnoceny všechny experimenty. Jelikož je pro trénování vybráno více různých modelů založených na konvolučních neuronových sítích, dataset byl rozdělen na tyto tři podmnožiny pouze fixně před prvotním trénováním, všechny další modely jsou pak natrénované na stejných snímcích, aby byly co nejlépe porovnatelné. Snímků každé kategorie jsou rozděleny v jiném poměru. Pro reálné otisky prstů s onemocněním jsou snímků děleny v poměru 60 : 20 : 20 hlavně z důvodu menšího počtu snímků a většího počtu vyhodnocených snímků na testovacím datasetu v rámci experimentů. Syntetické otisky a snímků z databáze NIST jsou pak děleny v poměru 80 : 10 : 10.

Onemocnění	Trénovací sada		Validační sada		Testovací sada	
	Reálné	Syntetické	Reálné	Syntetické	Reálné	Syntetické
Atopický ekzém	171	328	57	41	58	42
Bradavice	25	180	8	22	9	23
Psoriáza	58	216	20	27	20	28
Dyshidróza	57	160	19	20	20	20
Bez poškození	904	320	113	40	114	40

Tabulka 4.2: Počty snímků datasetu pro trénovací, validační a testovací podmnožiny

4.2 Anotace dat

Každý snímek je anotován v XML formátu. Pro anotaci reálných otisků prstů s onemocněními byl užit program labelImg dostupný pod MIT licencí¹. Níže je uveden jeden z příkladů anotace snímků s onemocněním atopického ekzému:

```
<annotation>
    <folder>atopic_png</folder>
    <filename>atopic_eczema_FP_00008.png</filename>
    <path>D:\images\atopic_png\atopic_eczema_FP_00008.png</path>
    <source>
        <database>Unknown</database>
    </source>
```

¹<https://github.com/tzutalin/labelImg>

```

<size>
    <width>527</width>
    <height>597</height>
    <depth>1</depth>
</size>
<segmented>0</segmented>
<object>
    <name>atopic</name>
    <pose>Unspecified</pose>
    <truncated>0</truncated>
    <difficult>0</difficult>
    <bndbox>
        <xmin>13</xmin>
        <ymin>297</ymin>
        <xmax>159</xmax>
        <ymax>332</ymax>
    </bndbox>
</object>
</annotation>

```

Snímky získané generátorem dyshidrózy a bradavic obsahují anotaci bounding boxů onemocnění v JSON formátu: [{"Type": "wart", "p1": [126, 232], "p2": [198, 304]}]

Pro potřeby práce byly tyto anotace v JSON formátu pro generování onemocnění dyshidrózy a bradavic převedeny do XML formátu. XML formát mimo jiné obsahuje důležité informace ohledně pozice jednotlivých bounding boxů (`<xmin>`, `<ymin>`, `<xmax>` a `<ymax>`) a názvu dané třídy onemocnění (`<name>`).

4.3 Předzpracování otisků prstů

Všechny otisky prstu byly před samotným začátkem trénování konvoluční neuronové sítě předzpracovány. Šedotónové snímky jsou normalizovány (viz obrázek 4.2). Jedná se o metodu, která má za účel snížit rozptyl šedotónových hodnot podél hřebenů a údolí a usnadnit tak další zpracování [24]. Papilární linie jsou obecně po aplikaci normalizace více viditelné a lépe rozeznatelné.



Obrázek 4.1: Snímek atopického ekzému



Obrázek 4.2: Normalizovaný obraz

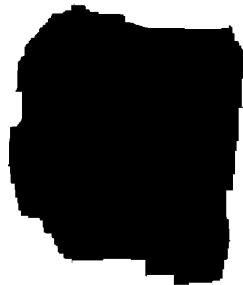
Následně je obraz prahován pomocí metody Otsu (viz obrázek 4.3). Rozdílem této metody na rozdíl od globálního prahování je princip výběru prahu. Při globálním prahování je

tento práh konstantní pro všechny snímky (obvykle hodnoty 127), Otsu metoda tento práh zjišťuje speciálně pro každý snímek, konkrétně podle histogramu. Je tedy nalezen optimální práh díky principu minimalizace váženého vnitrotřídního rozptylu [40].

Obraz je segmentován s využitím morfologických operací. Ze snímku je tedy vybrán pouze otisk prstu a pozadí zůstává bílé. Mnoho snímků databáze pochází z daktyloskopických karet, které jsou naskenované – pozadí je tedy různých variant, jeho podoba není potřebná při další analýze, a proto je tímto způsobem odstraněno. Morfologické operace jsou operacemi, které mohou být provedeny na binárním obrazu složeném z pixelů dvou možných barev (obvykle černobílé snímky). Každá morfologická operace využívá dva vstupy – originální obraz a jádro. Základními operacemi jsou eroze a dilatace. Eroze slouží k odstranění a ztenčení hranic objektu. Jádro je aplikováno na pixely obrazu stejným principem jako 2D konvoluce – pixel v původním obrázku (hodnoty 0 nebo 1) bude považován za 1 pouze v případě, pokud všechny pixely pod jádrem jsou hodnoty 1, jinak je nastaven na hodnotu 0. Dilatace je opakem eroze, kdy je pixel obrazu nastaven na hodnotu 1, pokud aspoň jeden pixel pod jádrem je hodnoty 1. Linie objektu v popředí se stávají silnějšími [41].



Obrázek 4.3: Otsu prahování



Obrázek 4.4: Získaná maska

Pro předzpracování otisků prstů bylo využito jádro o velikosti 51×51 a morfologická operace morfologického otevření – metoda eroze následována dilatací. Morfologické otevření je vhodným přístupem pro odstranění šumu v obrazu, v našem případě se bude jednat o vyplnění světlých oblastí mezi papilárními liniemi, tak aby vznikla maska určující pozici otisku prstu (viz obrázek 4.4). Původní snímek a získaná maska jsou sečteny – výsledkem je předzpracovaný otisku prstu s bílým pozadím (viz obrázek 4.5).



Obrázek 4.5: Výsledný předzpracovaný snímek

4.4 Použité metody pro detekci a klasifikaci

Cílem této práce je analyzovat více různých přístupů pro detekci a klasifikaci onemocnění otisků prstů – bude experimentováno s mnoha typy detekčních modelů popsaných v kapitole 3.8. Konkrétně byly vybrány následující modely, bude brán ohled na experimentování pro různé přístupy k detekci i varianty páteřních sítí:

- SSD MobileNet V2 FPNLite 320×320 ,
- SSD ResNet50 V1 FPN 640×640 (RetinaNet50),
- Faster R-CNN ResNet50 V1 640×640 ,
- Faster R-CNN ResNet101 V1 640×640 ,
- EfficientDet D0 512×512 ,
- EfficientDet D1 640×640 ,
- CenterNet HourGlass104 512×512 ,
- CenterNet Resnet101 V1 FPN 512×512 .

4.4.1 TensorFlow a TensorFlow Model Garden

TensorFlow je open-source knihovnou zaměřující se na strojové učení. Umožňuje efektivně provádět nízkoúrovňové tenzorové operace na CPU, GPU a TPU (Tensor Processing Unit). Je schopna škálovat výpočty s využitím mnoha zařízení (např. clustery se stovkami GPU). Keras je vysokoúrovňové API knihovny TensorFlow - klade důraz na uživatelskou přívětivost a nezbytné abstrakce – umožňuje snadným způsobem stavět modely strojového učení pro rozličné úlohy [30].

TensorFlow Model Garden je knihovnou dostupnou pod licencí Apache 2.0. Obsahuje implementace state-of-the-art řešení modelů a poskytuje mimo jiné podporu pro detekci a klasifikaci objektů – TensorFlow 2 Detection Model Zoo²[60]. Všechny modely byly před-trénovány na COCO 2017 datasetu³ – jednom z nejužívanějších datasetů strojového učení. Tento dataset obsahuje více jak 200 000 anotovaných snímků řazených do 80 objektových kategorií, každý objekt je poté zařazen také do subkategorií – např. objekty kategorií *bicycle*, *car*, *motorcycle*, *airplane*, *bus*, *train*, *truck* nebo *boat* jsou zařazeny do superkategorií *vehicle* [49].

4.4.2 Načítání a konfigurace modelu

Každý z detekčních modelů z knihovny TensorFlow Model Garden obsahuje konfigurační soubor s názvem `pipeline.config`. Tento soubor má mnoho parametrů vhodných pro další experimentování. Pro každé trénování je nutné správně specifikovat hlavně následující parametry:

- `num_classes` – počet tříd (v této práci hodnoty 4 pro každé z onemocnění – atopický ekzém, bradavice, dyshidróza, psoriáza),

²https://github.com/tensorflow/models/blob/master/research/object_detection/g3doc/tf2_detection_zoo.md

³<https://cocodataset.org>

- `fine_tune_checkpoint` – cesta k původnímu předtrénovanému modelu na COCO datasetu nebo cesta k již dříve natrénovaném modelu na zadáné úloze,
- `num_steps` – počet trénovacích kroků,
- `fine_tune_checkpoint_type` – parametr nastavený na hodnotu "detection" při načítání původního předtrénovaného checkpointu nebo "full" při pokračování trénování z checkpointu získaného z trénování na stejně úloze,
- `label_map_path` – cesta k `pbtxt` souboru zaznamenávající užité názvy tříd (stejné jako v `xml` anotacích) a jejich číselné indexy, v této práci má soubor následující formát:

```

item {
    name:'atopic'
    id:1
}
item {
    name:'verruca'
    id:2
}
item {
    name:'dysh'
    id:3
}
item {
    name:'psor'
    id:4
}

```

- `train_input_reader` – `input_path` – cesta k trénovacímu datasetu ve formátu TF (TensorFlow) record,
- `eval_input_reader` – `input_path` – cesta k validačnímu datasetu ve formátu TF record.

Pro trénovací a validační dataset je třeba vytvořit tzv. TF records, které obsahují vstupní obrázky reprezentované v bytech, anotované pozice bounding boxů s třídami a další dodatečné informace. Pro konverzi datasetu s XML anotacemi do TF records byl využit Python skript autora Nicholase Renotte⁴. Trénovací dataset je dostupný v souboru `train.record` a validační dataset je uložen v souboru `val.record`. TF records pro trénování všech typů modelů jsou totožné, stejně jako testovací dataset pro vyhodnocení experimentů.

4.4.3 Trénování modelu a následná predikce

Pro trénování bylo využito prostředí Google Colab Pro s GPU T4 a P100. Při následné predikci je vybráno v každém obrazu 100 bounding boxů s nejvyšším skóre jistoty, uživateli je poté vykresleno 20 bounding boxů s aspoň 30 % mírou jistoty pro klasifikaci. Současně jsou do textového souboru zapsány i informace o jednotlivých nejjistějších predikcích. Následně

⁴https://github.com/nicknochnack/GenerateTFRecord/blob/main/generate_tfrecord.py

je třeba model vyhodnotit na testovacím datasetu na základě např. překryvu anotovaných a predikovaných bounding boxů a správnosti klasifikované třídy onemocnění nebo pomocí dalších technik (viz sekce 4.5).

4.5 Výstup detekce a klasifikace onemocnění otisků prstů

Výstupem detekce a klasifikace onemocnění otisků prstů je konzolová aplikace umožňující načtení, konfiguraci a trénování vybraného modelu ať už od počátku, nebo od uvedeného checkpointu. Snímky z testovacího datasetu jsou následně predikovány natrénovaným modelem a jsou vykresleny bounding boxy s klasifikací a klasifikačním skóre jistoty pro dané onemocnění (viz příklady v příloze A). Výsledky predikce jsou zapsány do textového souboru, následně je na testovací sadě ohodnocena i přesnost modelu, výsledky evaluace pro celý testovací dataset jsou zapsány do csv souboru. Při trénování modelu je možné vizualizovat metriky v prostředí TensorBoard. Uživatel může také před začátkem trénování měnit různé parametry modelu v konfiguračním souboru `pipeline.config`. V kapitole 5 jsou uvedeny parametry, se kterými bylo experimentováno.

Při predikci se do souboru `predictions.txt` zapisují informace o predikcích se skóre jistoty aspoň 30 %:

```
FILE NAME: atopic_eczema_FP_00183.png
DETECTIONS:
    CLASS: atopic eczema
    DETECTION SCORE: 98.44 %
    CLASS: atopic eczema
    DETECTION SCORE: 89.35 %
    CLASS: atopic eczema
    DETECTION SCORE: 80.77 %
    CLASS: atopic eczema
    DETECTION SCORE: 78.55 %
    CLASS: atopic eczema
    DETECTION SCORE: 76.88 %
    CLASS: atopic eczema
    DETECTION SCORE: 67.33 %
    CLASS: atopic eczema
    DETECTION SCORE: 62.71 %
```

4.6 Vyhodnocovací metriky

Úspěšnost natrénovaného modelu je vyhodnocena na testovacím datasetu. Pro evaluaci bylo vybráno několik metrik, které jsou počítány pro celý testovací dataset i jeho různé podkategorie. Jedná se o podkategorie jednotlivých onemocnění i typu reálného nebo syntetického otisku prstu.

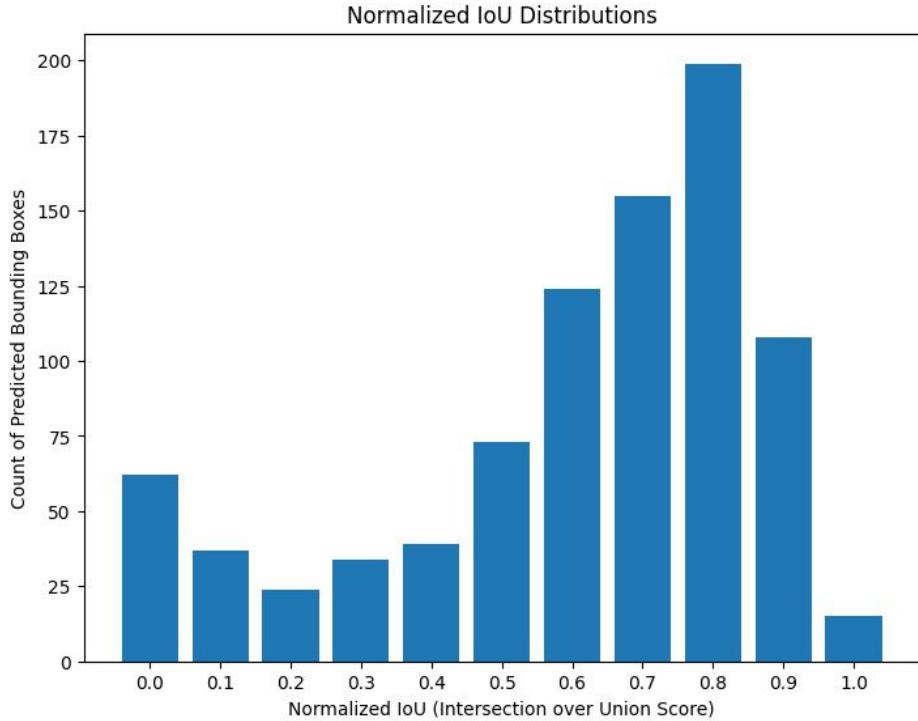
Užité evaluační metriky jsou následující:

- **Test správně detekované plochy** – Tato metrika udává procentuální velikost detekované plochy vzhledem ke všem anotovaným bounding boxům. Pokud by byla tedy hodnota metriky 100 %, pak by všechny detekované bounding boxy byly i správně predikovány modelem. Principem implementace tohoto testu je vypočítání průniku všech

detekovaných a anotovaných pixelů bounding boxů. Protože otisky prstu bez poškození nemají ve svém anotačním souboru žádnou zapsanou anotaci, neobsahují totiž bounding box s jakýmkoliv onemocněním, při predikci jakéhokoliv bounding boxu onemocnění natrénovaným modelem je výsledek tohoto testu 0 %, naopak pokud není predikován žádný bounding box, pak je výsledek 100 %.

- **Test správně detekované a klasifikované plochy** – Tento test je podobný testu správně detekované plochy. Rozdílem je, že se vedle správné detekce počítá i s tím, zda je daný bounding box správně klasifikovaný do třídy onemocnění – atopického ekzému, dyshidrózy, psoriázy nebo bradavic.
- **Test nedetekované plochy** – Jedná se o procentuální velikost plochy anotovaných bounding boxů, která při evaluaci nebyla natrénovaným modelem detekována.
- **Test detekované plochy navíc** – Tento test udává procentuální zastoupení pixelů predikovaných bounding boxů, které byly detekovány navíc vzhledem ke všem predikovaným bounding boxům. Pokud by tedy např. hodnota této metriky byla 30 %, pak 30 % plochy všech predikovaných bounding boxů nebylo anotováno. Implementace využívá rozdílu pixelů predikovaných bounding boxů a pixelů anotovaných bounding boxů.
- **Test detekované plochy navíc se správnou klasifikací** – Tento test je podobný testu detekované plochy navíc, avšak počítá pouze s predikovanou plochou, která byla sice detekovaná navíc, avšak správně klasifikovaná do třídy onemocnění daného otisku prstu.
- **Test průměrného skóre detekce pro správně detekovanou plochu** – Natrénovaný model predikuje bounding boxy s nastaveným skóre detekce aspoň 30 %. Tato hodnota udává míru jistoty natrénovaného modelu o zařazení bounding boxu právě do dané třídy. Po tomto prahovém vyfiltrování je získáno skóre detekce pro každý správně detekovaný pixel. Daný pixel může však být součástí více predikovaných bounding boxů s různým skóre detekce, v tomto případě je počítáno s nejvyšším možným skórem detekce ze všech možných výskytů. Následně jsou všechny správně detekované pixely a jejich detekční skóre sečteny a je vypočten průměr.
- **Test průměrného skóre detekce pro správně detekovanou a klasifikovanou plochu** – Tato metrika je obdobná jako metrika průměrného skóre detekce pro správně detekovanou plochu. Pouze jsou opět předem vyfiltrovány nejen správně detekované pixely, ale i správně klasifikované do třídy onemocnění.
- **Normalizovaná IoU (Intersection over Union)** – IoU je užívanou metrikou, která počítá procentuální překryv predikovaných a anotovaných bounding boxů. Pro tento výpočet byl využit následující postup – Pro každou predikci je spočítána metrika IoU a je uložena nejvyšší hodnota této metriky s jakýmkoliv anotovaným bounding boxem. Následně je vykreslen histogram pro celý testovací dataset s normalizovanými počty [23]. Příklad tohoto výstupu je uveden na obrázku 4.6. Histogram by měl ideálně nabývat největšího počtu bounding boxů s aspoň 50 % – 60 % překryvem. Z příkladu histogramu na obrázku 4.6 je zřejmé, že nejčetněji je predikovaný bounding box překryt s některým z anotovaných bounding boxů z přibližně 80 %. Otisky bez poškození mají normalizovanou hodnotu IoU vždy 0, protože neobsahují žádné

anotované bounding boxy, se kterými by mohlo být počítáno překrytí. Pro každou z kategorií datasetu je následně počítána nejčetnější hodnota této metriky.



Obrázek 4.6: Histogram normalizované IoU

Každý obraz testovacího datasetu je evaluován pomocí těchto metrik a na konci běhu skriptu se uživateli zobrazí průměrné hodnoty všech metrik pro celý testovací dataset. Metriky jsou zobrazeny pro následující kategorie – atopický ekzém (reálné a syntetické otisky), atopický ekzém (reálné otisky), atopický ekzém (syntetické otisky), bradavice (reálné a syntetické otisky), bradavice (reálné otisky), bradavice (syntetické otisky), dyshidróza (reálné a syntetické otisky), dyshidróza (reálné otisky), dyshidróza (syntetické otisky), psoriáza (reálné a syntetické otisky), psoriáza (reálné otisky), psoriáza (syntetické otisky), zdravé otisky bez poškození (reálné a syntetické otisky), zdravé otisky bez poškození (reálné otisky), zdravé otisky bez poškození (syntetické otisky), všechny snímky (reálné a syntetické otisky), všechny snímky (reálné otisky), všechny snímky (syntetické otisky). Tyto metriky jsou vyplány uživateli do terminálu a následně také zapsány do csv souboru.

Kapitola 5

Experimenty

Následující kapitola shrnuje experimenty s osmi vybranými modely zmíněnými v sekci 4.4. Experimenty jsou rozděleny do sekcí na základě daného přístupu pro detekci objektů – SSD, Faster R-CNN, EfficientDet a CenterNet. Kapitola dále popisuje parametry z konfiguračního souboru každého z modelů `pipeline.config`, u kterých byly jejich hodnoty upravovány během experimentů. Nakonec jsou experimenty porovnány a jsou vybrány nejvhodnější modely pro zadanou úlohu.

5.1 Parametry modelu pro experimentování

Každý z modelů byl trénován minimálně dvakrát – poprvé s původními parametry uvedenými v konfiguračním souboru (maximálně došlo ke změně počtu trénovacích kroků). Následně byly u těchto prvních trénování zjištěny evaluační metriky modelu. Poté byly stejné modely trénovány podruhé, s upravenými parametry za cílem zlepšení výsledků v porovnání s prvním trénováním. Tyto modely byly obvykle trénovány vyšší počet trénovacích kroků, avšak muselo být dbáno na problém možného přetrénování modelu hlavně z důvodu datasetu menší velikosti. Změna hodnot upravených parametrů vycházela často z některých příbuzných vědeckých článků, často bylo snahou zachování poměrově podobných velikostí parametrů (asoň co se týče jejich řádů) jako u původních nastavení v konfiguračním souboru.

Během experimentů byly obvykle upravovány následující atributy:

- `l2_regularizer - weight` – L2 regularizace (viz sekce 3.5.4) je příkladem jedné z regularizací neuronových sítí, které mají za cíl snížit složitost modelu a předejít přetrénování. Jedná se o jednu z nejdůležitějších regularizačních technik, avšak často neoptimalizovanou a nastavenou na obecnou hodnotu např. $5e-5$ nebo $1e-5$. Pro experimentování modelů s upravenými parametry byla využita hodnota $3e-4$, uvedena jako jedna z potenciálně vhodných hodnot užitá pro experimenty s využitím CIFAR10 datasetu v článku [46].
- `learning_rate - learning_rate_base` – Všechny modely s výjimkou CenterNet Hourglass využívají postupné snižování learning rate s využitím cosinového sestupu. Learning rate η_t pro každý trénovací krok i je definována následovně [35]:

$$\eta_t = \eta_{min}^i + \frac{1}{2}(\eta_{max}^i - \eta_{min}^i)(1 + \cos(\frac{T_{cur}}{T_i}\pi)) \quad (5.1)$$

Parametry η_{min}^i a η_{max}^i udávají rozsah learning rate, T_{cur} je aktuální počet již provedených trénovacích kroků. Pokud $T_{cur} = T_i$, pak výstup cosinové funkce bude nabývat hodnoty -1 a zároveň $\eta_t = \eta_{min}^i$ [35]. Konfigurační parametr `learning_rate_base` udává právě maximální hodnotu learning rate cosinového sestupu. Při úpravě parametrů modelu při využití cosinového sestupu byla zvolena hodnota $0,03$ nebo $0,003$, se kterými bylo experimentováno v článku [33].

- `learning_rate - total_steps` – Udává celkový počet kroků, po který je cosinový sestup i s využitím warmup fáze prováděn. Při trénování této práce je tato hodnota vždy totožná s hodnotou `num_steps` [11].
- `learning_rate - warmup_learning_rate` – Hodnota veličiny `learning_rate_base` musí být větší nebo rovna hodnotě `warmup_learning_rate`. Tzv. warmup fáze trvá několik prvních trénovacích kroků a využívá se z důvodu problematiky vysoké chybovosti na počátku trénování. Hodnota learning rate je postupně lineárně zvyšována z hodnoty `warmup_learning_rate` do hodnoty `learning_rate_base` [11]. Při experimentech s úpravou konfiguračních parametrů byla tato hodnota volena $0,003$ nebo $0,0003$, kdy byla hodnota `warmup_learning_rate` o jeden řád nižší vzhledem k zvolené `learning_rate_base`, také vzhledem k pravidlu při jejich vzájemném porovnání.
- `learning_rate - warmup_steps` – Tato hodnota udává počet trénovacích kroků warmup fáze. Při úpravě hodnoty byl obvykle zvyšován počet kroků warmup fáze společně se zvyšováním celkového počtu trénovacích kroků.
- `num_steps` – Tento parametr udává celkový počet trénovacích kroků. Při trénování s upravenými parametry byl počet zvyšován za účelem potenciálního snižování chybové funkce a zlepšení přesnosti natrénovaného modelu.

Speciálně pro modely Faster R-CNN byly upraveny v rámci experimentů následující parametry:

- `first_stage_max_proposals, max_total_detections` – Tyto parametry udávají počty zjištěných vhodných regionů pro detekci a klasifikaci a následně počet finálních detekcí. V původním konfiguračním souboru mají tyto atributy hodnotu 300 . Během experimentů bylo podobně jako v článku [26] využito snížení počtu potenciálně vhodných oblastí pro detekci a klasifikaci pro možné snížení výpočetní složitosti. Experimentálně byly tyto dvě metriky nastaveny na hodnotu 250 .
- `use_dropout, dropout_keep_probability` – Tyto parametry slouží pro nastavení regularizační techniky dropout popsané v sekci 3.5.4. Parametr `use_dropout` udává, zda bude během trénování dropout použitý, `dropout_keep_probability` je pak nastavena na pravděpodobnost, se kterou mají být neurony zachovány, během experimentů je tato hodnota nastavena na $0,5$ – tato hodnota byla potenciálně označena za vhodnou na základě studie v článku [5], který obsahuje příklady užití dropout v několika vědeckých pojednáních.

5.2 Experimenty s jednotlivými přístupy pro detekci objektů

Následující sekce je rozdělena na základě vybraného modelu z hlediska vybraného přístupu pro detekci. Pro každý vybraný model jsou pak uvedeny parametry konfiguračního souboru,

které byly upravovány. Následně jsou vyobrazeny tabulky některých z evaluačních metrik – test správně detekované a klasifikované plochy pro jednotlivá onemocnění, otisky prstu bez poškození a celkové průměrné výsledky pro celý testovací dataset, pouze syntetické otisky prstu a pouze reálné otisky prstu, test skóre detekce správně detekované a klasifikované plochy pro jednotlivá onemocnění, test detekované plochy navíc, nejčetnější hodnota metriky IoU pro jednotlivá onemocnění a doba trénování.

Sloupce jednotlivých onemocnění udávají vždy jak reálné, tak syntetické otisky prstu z dané kategorie. Každý natrénovaný model obsahuje `csv` soubor, který obsahuje detailní výsledky a další z metrik i pro kategorie pouze reálných nebo syntetických otisků prstů postižených daným onemocněním. Tento soubor obsahuje také další dílčí metriky (např. test správně detekované plochy, která nebene v potaz správný výsledek klasifikace atd.).

5.2.1 Experimenty s modely SSD

Pro experimenty byly využity dva zástupci detekčního přístupu SSD – SSD MobileNet o velikosti vstupního obrazu 320×320 pixelů a SSD ResNet-50 pracující se vstupem o velikosti 640×640 pixelů. Nastavení parametrů pro SSD MobileNet je uvedeno v tabulce 5.1. I díky skutečnosti, že ze všech vybraných modelů pracuje SSD MobileNet se vstupem o nejnižším rozlišení a trénovací doba je nízká, byl při druhém trénování zdvojnásoben počet trénovacích kroků.

ID	num_steps	lr_base	warmup_lr	warmup_steps	l2_weight
1	50 000	0,08	0,026666	1 000	4e-05
2	100 000	0,03	0,003	2 500	3e-04

Tabulka 5.1: Změny parametrů pro SSD MobileNet

Tabulky 5.2 a 5.3 obsahují průměrné hodnoty metriky správně detekované a klasifikované plochy pro otisky prstů s poškozením, otisky prstů bez poškození a celkové shrnující statistiky pro celý testovací dataset. Při druhém trénování po větší hodnotu trénovacích kroků a úpravě parametrů byla zlepšena tato metrika pro všechna onemocnění s výjimkou atopického ekzému. Model se tedy tato onemocnění učí klasifikovat a detektovat stále přesněji. Avšak při delším trénování SSD MobileNet začíná nesprávně predikovat možná onemocnění otiskům prstů bez poškození. Při analýze výsledků predikcí je pak zjištěno, že v této situaci nejčastěji nesprávně model predikuje onemocnění atopického ekzému.

ID	Atopický ekzém	Bradavice	Dyshidróza	Psoriáza	Bez poškození
1	52,469	84,605	75,999	69,925	79,221
2	51,427	86,219	78,173	75,198	66,234

Tabulka 5.2: Test správně detekované a klasifikované plochy pro SSD MobileNet

ID	Celkově (reálné, syntetické)	Celkově (reálné)	Celkově (syntetické)
1	70,991	68,226	74,984
2	66,412	60,861	74,43

Tabulka 5.3: Test správně detekované a klasifikované plochy pro SSD MobileNet

Tabulka 5.4 zobrazuje průměrné skóre detekce pro správně klasifikovanou a detekovanou plochu. V této situaci by měla teoreticky se zvyšujícím se počtem trénovacích kroků růst

i jistota modelu predikovat dané bounding boxy do tříd onemocnění. Při trénování SSD MobileNet je tato teorie potvrzena a míra jistoty je pro všechna onemocnění při druhém trénování vyšší.

ID	Atopický ekzém	Bradavice	Dyshidróza	Psoriáza
1	49,998	84,249	74,176	68,924
2	54,011	86,281	77,048	73,426

Tabulka 5.4: Test skóre detekce správně detekované a klasifikované plochy pro SSD MobileNet

Při druhém trénování došlo k zvýšení detekce neanotované plochy pro všechna onemocnění kromě dyshidrózy (viz tabulka 5.5). Model při druhém trénování s upravenými parametry má větší tendenci detektovat plochu navíc zejména pro onemocnění psoriázy. Naopak pro bradavice je obecně predikována nejmenší neanotovaná plocha – zhruba pouze 8 % celkové predikované plochy není zároveň anotované.

ID	Atopický ekzém	Bradavice	Dyshidróza	Psoriáza
1	24,093	8,105	19,216	16,673
2	27,002	8,954	14,698	20,489

Tabulka 5.5: Test detekované plochy navíc pro SSD MobileNet

Tabulka 5.6 obsahuje nejčetnější maximální překryvy predikovaného bounding boxu a anotovaných bounding boxů. Obecně největší překryvy mají bradavice – přibližně 90 %. Při delším učení modelu by teoreticky mělo dojít ke zvýšení hodnoty překryvu, kdy predikované a anotované bounding boxy mají stálé vyšší hodnotu překrytí a probíhá přesnější lokalizace. Avšak při trénování s upravenými parametry došlo ke zvýšení nejčetnějšího překryvu pouze pro dyshidrózu.

ID	Atopický ekzém	Bradavice	Dyshidróza	Psoriáza
1	0,7	0,9	0,7	0,6
2	0,7	0,9	0,8	0,6

Tabulka 5.6: Nejčetnější hodnota IoU pro SSD MobileNet

Jak je z tabulky 5.7 patrné, trénování s původními parametry nastavenými v konfiguračním souboru je velmi rychlé a v prostředí Google Colab Pro trvalo pouze přibližně hodinu. Při dvojnásobném navýšení počtu trénovacích kroků se pak potřebná doba pro trénování zdvojnásobila. Delší trénování s upravenými parametry přineslo výhody hlavně pro metriku průměrného skóre detekce, při dalších testech však už nedošlo k výraznému zlepšení.

ID	Trénovací doba
1	1 h 17 min
2	2 h 34 min

Tabulka 5.7: Trénovací doba pro SSD MobileNet

Dalším modelem využívajícím pro detekci objektů přístupu SSD byl model s páteřní sítí ResNet-50 a obrazem o vstupních rozměrech 640×640 pixelů. Parametry byly nastaveny

obdobným způsobem jako u předchozího modelu, avšak počet trénovacích kroků byl navýšen pouze o 10 000 (viz tabulka 5.8).

ID	num_steps	lr_base	warmup_lr	warmup_steps	l2_weight
3	25 000	0,04	0,013333	2 000	0,0004
4	35 000	0,03	0,003	2 500	3e-04

Tabulka 5.8: Změny parametrů pro SSD ResNet-50

Při testu správně detekované a klasifikované plochy došlo bohužel při zvýšení počtu trénovacích kroků a úpravě parametrů ke zhoršení dané metriky (viz tabulky 5.9 a 5.10). Největší zhoršení je patrné pro onemocnění psoriázy a otisky prstů bez poškození. Ve srovnání s SSD MobileNet došlo k mírnému zlepšení pro bradavice a atopický ekzém, avšak průměrná hodnota pro reálné testovací otisky prstů se snížila až na pouze přibližně 55 %.

ID	Atopický ekzém	Bradavice	Dyshidróza	Psoriáza	Bez poškození
3	49,775	90,25	81,215	75,319	76,623
4	45,041	89,302	76,822	67,219	66,234

Tabulka 5.9: Test správně detekované a klasifikované plochy pro SSD ResNet-50

ID	Celkově (reálné, syntetické)	Celkově (reálné)	Celkově (syntetické)
3	70,934	67,123	76,44
4	63,8	55,604	75,639

Tabulka 5.10: Test správně detekované a klasifikované plochy pro SSD ResNet-50

Při dané konfiguraci parametrů s páteřní sítí ResNet-50 došlo k nepotvrzení teorie zvyšování hodnot skóre detekce při větším počtu trénovacích kroků. Největším problém se při této metrice jeví onemocnění psoriázy, kdy při druhém trénování je predikováno modelem pouze s jistotou okolo 39 % (viz tabulka 5.11). V tomto testu bezpodmínečně také více uspěl model SSD MobileNet, u kterého se s úpravou parametrů a trénovacích kroků zvýšilo průměrné skóre detekce pro všechny kategorie.

ID	Atopický ekzém	Bradavice	Dyshidróza	Psoriáza
3	42.031	87.176	60.819	55.139
4	35,278	84,797	50,631	39,491

Tabulka 5.11: Test skóre detekce správně detekované a klasifikované plochy pro SSD ResNet-50

Tabulka 5.12 znázorňuje procento plochy predikovaných bounding boxů, která byla predikována navíc. Zde při druhém trénování došlo k zlepšení metriky pro bradavice a dyshidrózu. Pokud ale výsledky srovnáme s předchozím modelem využívajícího stejného přístupu pro detekci objektů, ale jiné páteřní sítě a menší velikosti vstupního obrazu, pak jsou výsledky této metriky horší.

ID	Atopický ekzém	Bradavice	Dyshidróza	Psoriáza
3	29,352	13,952	23,542	23,835
4	31,655	12,771	23,526	29,449

Tabulka 5.12: Test detekované plochy navíc pro SSD ResNet-50

V průběhu druhého trénování došlo k zvýšení nejčetnějšího překryvu anotací a predikcí u atopického ekzému přibližně o 10 %. Největším problémem však zůstalo onemocnění psoriázy, kdy se nejčetnější překryv zhoršil asi o 30 % (viz tabulka 5.13).

ID	Atopický ekzém	Bradavice	Dyshidróza	Psoriáza
3	0,7	0,9	0,6	0,7
4	0,8	0,9	0,6	0,4

Tabulka 5.13: Nejčetnější hodnota IoU pro SSD ResNet-50

Navýšení trénovacích kroků o 10 000 zvýšil dobu potřebnou pro trénování modelu SSD ResNet-50 o přibližně 1 hodinu (viz tabulka 5.14). Bohužel u tohoto modelu nebyly pokusy o jeho vylepšování příliš úspěšné a pro danou úlohu je vhodnější model SSD MobileNet. Kombinace páteřní sítě MobileNet, která se vyhýbá trendu tvoření komplikovanějších sítí o více vrstvách (jako např. právě ResNet-50) a vstupního obrazu o malých rozměrech, přináší výhodu rychlého trénování. Důvodem lepšího chování prvního z modelů může být i problém nepříliš velkého trénovacího datasetu.

ID	Trénovací doba
3	2 h 10 min
4	3 h

Tabulka 5.14: Trénovací doba pro SSD ResNet-50

5.2.2 Experimenty s modely Faster R-CNN

Jako zástupci modelů Faster R-CNN byly využity modely využívající páteřní sítě ResNet-50 a ResNet-101. Oba tyto modely pracují na vstupním obrazu o rozměrech 640×640 pixelů. Tabulka 5.15 znázorňuje nastavení parametrů pro Faster R-CNN ResNet-50. V druhém a třetím trénování byly využity dvě regularizační techniky – nastavení dropout a L2 regulizace. Počet trénovacích kroků byl navýšen pouze o 5 000.

ID	steps	lr_base	warmup_lr	warmup_steps	l2_weight	dropout	proposals
5	25 000	0,04	0,013333	2 000	0	false	300
6	30 000	0,03	0,003	2 000	0	0,5	250
7	30 000	0,03	0,003	2 000	3e-04	false	250

Tabulka 5.15: Změny parametrů pro Faster R-CNN ResNet-50

Z tabulek 5.16 a 5.17 vyplývá, že navýšení trénovacích kroků a úprava parametrů už příliš nevylepšila metriku správně detekované a klasifikované plochy. Pouze pro bradavice se výsledek při třetím trénování s využitím L2 regularizace vylepšil. Je patrné, že při delším trénování se snižuje tato metrika pro onemocnění psoriázy.

ID	Atopický ekzém	Bradavice	Dyshidróza	Psoriáza	Bez poškození
5	64,397	89,891	79,165	71,577	91,558
6	65,093	88,775	76,846	64,424	88,312
7	64,04	93,944	79,521	64,191	88,312

Tabulka 5.16: Test správně detekované a klasifikované plochy pro Faster R-CNN ResNet-50

ID	Celkově (reálné, syntetické)	Celkově (reálné)	Celkově (syntetické)
5	80,263	76,875	85,158
6	77,851	73,971	83,456
7	78,268	73,808	84,71

Tabulka 5.17: Test správně detekované a klasifikované plochy pro Faster R-CNN ResNet-50

Test skóre detekce správně detekované a klasifikované plochy pro model Faster R-CNN ResNet-50 nabývá pro všechna onemocnění aspoň 85 % (viz tabulka 5.18). Model tedy s velmi vysokou jistotou dokáže zařadit daný bounding box do správné třídy onemocnění. Tato metrika je ve druhém nebo třetím trénování vylepšena pro všechna onemocnění s výjimkou psoriázy.

ID	Atopický ekzém	Bradavice	Dyshidróza	Psoriáza
5	90,053	93,443	89,496	85,608
6	92,372	95,09	86,388	85,361
7	90,439	98,551	89,96	85,062

Tabulka 5.18: Test skóre detekce správně detekované a klasifikované plochy pro Faster R-CNN ResNet-50

Tabulka 5.19 znázorňuje výsledky detekované plochy navíc pro Faster R-CNN ResNet-50. V průběhu druhého a třetího trénování po větší počet trénovacích kroků došlo k snížení této metriky pouze pro psoriázu a při třetím trénování s využitím L2 regularizace pro atopický ekzém.

ID	Atopický ekzém	Bradavice	Dyshidróza	Psoriáza
5	26,643	10,454	18,155	23,176
6	26,944	11,177	21,172	21,479
7	25,215	11,021	20,835	21,702

Tabulka 5.19: Test detekované plochy navíc pro Faster R-CNN ResNet-50

Nejčetnější hodnoty metriky IoU pro onemocnění jsou uvedeny v tabulce 5.20. Pro atopický ekzém a bradavice jsou výsledky pro všechna trénování stejné, pro psoriázu došlo k zvýšení maximálního průniku predikovaného a některého z anotovaných bounding boxů na 80 %.

ID	Atopický ekzém	Bradavice	Dyshidróza	Psoriáza
5	0,7	0,9	0,7	0,7
6	0,7	0,9	0,6	0,8
7	0,7	0,9	0,8	0,8

Tabulka 5.20: Nejčetnější hodnota IoU pro Faster R-CNN ResNet-50

Při využití některé z technik regularizace se trénovací čas i přes zvýšení počtu trénovacích kroků snížil o přibližně hodinu (viz tabulka 5.21). Problémem je však navýšení trénovacích kroků, které by potenciálně měly výsledky vylepšit. Pokusy o vylepšení trénovacích výsledků pro tento model ve srovnání s původními nastaveními v příslušném konfiguračním souboru nevykazovaly příliš dobré výsledky.

ID	Trénovací doba
5	3 h 28 min
6	2 h 33 min
7	2 h 33 min

Tabulka 5.21: Trénovací doba pro Faster R-CNN ResNet-50

Při jednom z úplně prvních experimentů s úpravami parametrů konfiguračního souboru byl model trénován s využitím L2 regularizace a navýšením maximálního počtu regionů vhodných pro další detekci a klasifikaci na hodnotu 400. Následně byl model trénován po dvojnásobný počet trénovacích kroků – 50 000. Při takovém navýšení se však přetrénování projevilo už opravdu markantně – metrika správně detekované a klasifikované plochy pro reálné snímky postižené psoriázou nabývala hodnoty pouze přibližně 36 %. Při analýze výsledků testovacího datasetu pak bylo zjištěno, že snímky reálných pacientů postižené tímto onemocněním jsou často klasifikovány nesprávně do kategorie atopického ekzému. Obecně při dalších experimentech byly pak trénovací kroky zvyšovány mnohem obezřetněji, kvůli možnému přetrénování i z důvodu datasetu menší velikosti.

Páteřní síť druhého modelu využívající Faster R-CNN jako přístup pro detekci objektů je hlubší konvoluční neuronová síť ResNet-101. Jak je z tabulky 5.22 patrné, parametry byly během experimentů upravovány stejným způsobem jako u předchozí Faster R-CNN ResNet-50.

ID	steps	lr_base	warmup_lr	warmup_steps	l2_weight	dropout	proposals
8	25 000	0,04	0,013333	2 000	0	false	300
9	30 000	0,03	0,003	2 000	0	0,5	250
10	30 000	0,03	0,003	2 000	3e-04	false	250

Tabulka 5.22: Změny parametrů pro Faster R-CNN ResNet-101

Následující tabulky 5.23 a 5.24 shrnují výsledky pro metriku správně detekované a klasifikované plochy. Při třetím trénování s využitím L2 regularizace došlo ke zlepšení výsledků pro onemocnění bradavic a psoriázy. Celkové výsledky pro syntetické otisky prstu byly také hlavně oproti prvnímu trénování vylepšeny. Výsledky pro reálný dataset však zůstaly přibližně stejné. Ve srovnání s Faster R-CNN ResNet-50 došlo ke zlepšení výsledků pro bradavice, avšak ostatní výsledky jsou srovnatelné, mnohdy i mírně horší.

ID	Atopický ekzém	Bradavice	Dyshidróza	Psoriáza	Bez poškození
8	64,837	90,629	77,669	68,7	87,013
9	64,403	90,613	73,94	65,881	86,364
10	64,78	93,433	73,812	73,62	81,818

Tabulka 5.23: Test správně detekované a klasifikované plochy pro Faster R-CNN ResNet-101

ID	Celkově (reálné, syntetické)	Celkově (reálné)	Celkově (syntetické)
8	78,043	73,241	73,241
9	76,898	72,616	83,082
10	76,348	70,175	85,264

Tabulka 5.24: Test správně detekované a klasifikované plochy pro Faster R-CNN ResNet-101

Z tabulky 5.25 je zřejmé, že výsledek skóre detekce je vylepšen během trénování s L2 regularizací hlavně pro bradavice – model je schopen třídu bradavic predikovat skoro se 100 % jistotou. Ve srovnání s Faster R-CNN ResNet-50 došlo ke zvýšení minimální hodnoty skóre detekce – z 85 % na 88 %. Pro nemoci dyshidrózy a psoriázy dosáhl model využívající hlubší páteřní síťě lepších výsledků.

ID	Atopický ekzém	Bradavice	Dyshidróza	Psoriáza
8	91,783	96,409	90,456	88,448
9	93,091	95,53	89,294	89,941
10	92,91	99,97	91,855	90,273

Tabulka 5.25: Test skóre detekce správně detekované a klasifikované plochy pro Faster R-CNN ResNet-101

V průběhu trénování s upravenými parametry docházelo ke snížení metriky detekované plochy navíc pro psoriázu a v případě L2 regularizace i pro dyshidrózu (viz tabulka 5.26). V porovnání s předchozím modelem došlo ke snížení této metriky pro všechna onemocnění s výjimkou dyshidrózy.

ID	Atopický ekzém	Bradavice	Dyshidróza	Psoriáza
8	24,457	6,145	22,2	19,432
9	24,351	9,836	24,307	18,452
10	23,69	7,433	17,869	17,083

Tabulka 5.26: Test detekované plochy navíc pro Faster R-CNN ResNet-101

Z tabulky 5.27 je patrné, že pro všechna tři trénování nejčetnější hodnoty IoU pro jednotlivá onemocnění zůstaly zachovány. Při L2 regularizaci pouze došlo ke snížení hodnoty maximálního překryvu bounding boxů pro psoriázu. Faster R-CNN ResNet-50 dosáhl v tomto testu horších výsledků pro atopický ekzém, kdy v případě užití hlubší páteřní sítě došlo při všech třech experimentech k zvýšení nejčetnější hodnoty překryvu.

ID	Atopický ekzém	Bradavice	Dyshidróza	Psoriáza
8	0,8	0,9	0,6	0,8
9	0,8	0,9	0,6	0,8
10	0,8	0,9	0,6	0,7

Tabulka 5.27: Nejčetnější hodnota IoU pro Faster R-CNN ResNet-101

Tabulka 5.28 znázorňuje doby trénování pro každý experiment. Zde, na rozdíl od předchozího modelu, došlo k navýšení trénovacího času při využití technik regularizace. Tato skutečnost však může být i způsobena možným odlišným vytížením výpočetních zdrojů při trénování v prostředí Google Colab Pro.

ID	Trénovací doba
8	2 h 51 min
9	3 h 25 min
10	3 h 41 min

Tabulka 5.28: Trénovací doba pro Faster R-CNN ResNet-101

Při porovnání obou modelů dosáhl model Faster R-CNN ResNet-101 lepších výsledků pro test skóre detekce a test detekované plochy navíc. Jako lepší regularizační technikou se pak ukázala být spíše L2 regularizace, avšak vliv regularizačních technik výsledky obecně příliš nevylepsil. I když se jedná o velmi užívané nástroje, detekční přístup Faster R-CNN byl při trénování na zvolené úloze náhylný k přetrénování a obecně se ukázal jako nejvíce komplikovaným přístupem pro detekci objektů pro další vylepšení. Důvodem tohoto problému může být zejména menší velikost trénovacího datasetu.

5.2.3 Experimenty s modely EfficientDet

Pro experimenty s přístupem pro detekci objektů vyvinutým společností Google byly použity dva modely – EfficientDet D0 pracující se vstupním obrazem o rozměrech 512×512 pixelů a EfficientDet D1 se vstupním obrazem o velikosti 640×640 pixelů. Model EfficientDet D0 byl trénován 50 000 trénovacích kroků stejně jako model SSD MobileNet, následně byl počet trénovacích kroků navýšen a byla upravena i nastavení learning rate a váhy L2 regulátoru (viz tabulka 5.29).

ID	num_steps	lr_base	warmup_lr	warmup_steps	l2_weight
11	50 000	0,08	0,001	2 500	4e-05
12	60 000	0,03	0,003	4 000	3e-04

Tabulka 5.29: Změny parametrů pro EfficientDet D0

Tabulky 5.30 a 5.31 udávají průměrné procento správně detekované a klasifikované plochy. Při druhém trénování byla navýšena tato metrika zejména pro reálné otisky prstu. Tato skutečnost je určitě výhodou, protože model by měl být na reálných otiscích co nejvíce použitelný. Nevýhodou učení s využitím hlavně syntetických otisků prstů může být totiž problém naučení modelu na pouze několika vygenerovaných příznacích onemocnění, výskyt stejného onemocnění na reálném otisku prstu může být však mnohem více komplexní.

ID	Atopický ekzém	Bradavice	Dyshidróza	Psoriáza	Bez poškození
11	59,255	95,736	78,59	78,231	71,429
12	60,261	94,763	78,068	80,028	72,727

Tabulka 5.30: Test správně detekované a klasifikované plochy pro EfficientDet D0

ID	Celkově (reálné, syntetické)	Celkově (reálné)	Celkově (syntetické)
11	71,893	67,633	78,045
12	72,788	70,448	76,168

Tabulka 5.31: Test správně detekované a klasifikované plochy pro EfficientDet D0

Při testu průměrného skóre detekce pro správně detekovanou a klasifikovanou plochu (viz tabulka 5.32) došlo při druhém trénování k navýšení detekčního skóre zvláště pro atopický ekzém, dyshidrózu a psoriázu. Model je trénován delší počet trénovacích kroků, jeho jistota pro klasifikaci bounding boxu do daného onemocnění je tedy zvýšena.

ID	Atopický ekzém	Bradavice	Dyshidróza	Psoriáza
11	50,689	98,132	74,282	73,642
12	56,816	97,954	79,567	79,118

Tabulka 5.32: Test skóre detekce správně detekované a klasifikované plochy pro EfficientDet D0

Test detekované plochy navíc je počítán procentuálně vzhledem ke všem predikovaným bounding boxům. Z evaluace této metriky pro EfficientDet D0 (viz tabulka 5.33) vyplývá, že druhé delší trénování modelu s upravenými parametry mělo za následek snížení této metriky hlavně pro onemocnění bradavic.

ID	Atopický ekzém	Bradavice	Dyshidróza	Psoriáza
11	27,376	11,107	16,117	18,092
12	26,876	8,422	15,614	19,277

Tabulka 5.33: Test detekované plochy navíc pro EfficientDet D0

Při analýze nejčetnější hodnoty IoU byl zvýšen maximální překryv anotovaného a jakéhokoliv predikovaného bounding boxu zejména pro psoriázu (viz tabulka 5.34). Opět je zde viditelné vylepšování metrik zejména pro toto onemocnění při druhém trénování.

ID	Atopický ekzém	Bradavice	Dyshidróza	Psoriáza
11	0,7	0,9	0,8	0,6
12	0,7	0,8	0,7	0,8

Tabulka 5.34: Nejčetnější hodnota IoU pro EfficientDet D0

Následující tabulka 5.35 znázorňuje trénovací doby pro EfficientDet D0. Při upravení parametrů a navýšení počtu trénovacích kroků byla trénovací doba zvýšena pouze zhruba o 30 minut. Celkově se jedná o jeden z modelů, který může být díky vstupnímu obrazu o velikosti 512×512 pixelů trénován více trénovacích kroků po kratší trénovací době.

ID	Trénovací doba
11	3 h 51 min
12	4 h 23 min

Tabulka 5.35: Trénovací doba pro EfficientDet D0

Model EfficientDet D1 využívá větší vstupní velikost obrazu 640×640 pixelů. Při nastavování parametru trénovacích kroků bylo voleno nižšího počtu než u modelu EfficientDet D0, protože trénování modelu je pomalejším. Následně byly parametry při druhém trénování upraveny obdobným způsobem jako u EfficientDet D0 – opět je trénování navýšeno o 10 000 trénovacích kroků, parametry `lr_base`, `warmup_lr` a `l2_weight` byly nastaveny na stejnou hodnotu (viz tabulka 5.36).

ID	num_steps	lr_base	warmup_lr	warmup_steps	l2_weight
13	25 000	0,08	0,001	2 500	4e-05
14	35 000	0,03	0,003	2 500	3e-04

Tabulka 5.36: Změny parametrů pro EfficientDet D1

Následující tabulky 5.37 a 5.38 udávají test správně detekované a klasifikované plochy pro EfficientDet D1. Při druhém trénování modelu s upravenými parametry došlo hlavně ke zvýšení této metriky u atopického ekzému a otisků prstů bez poškození. Celkově došlo k navýšení výsledku tohoto testu u celého testovacího datasetu. Při porovnání s výsledky trénování modelu EfficientDet D0 dosáhl EfficientDet D1 při této metrice lepších výsledků zejména pro otisky prstu bez poškození, průměrný výsledek pro celý testovací dataset byl také vylepšen, opět s přihlédnutím k reálným otiskům prstů.

ID	Atopický ekzém	Bradavice	Dyshidróza	Psoriáza	Bez poškození
13	54,9	94,125	77,428	80,021	72,727
14	61,297	94,606	77,315	80,922	79,87

Tabulka 5.37: Test správně detekované a klasifikované plochy pro EfficientDet D1

ID	Celkově (reálné, syntetické)	Celkově (reálné)	Celkově (syntetické)
13	71,23	68,106	75,744
14	76,027	75,329	77,034

Tabulka 5.38: Test správně detekované a klasifikované plochy pro EfficientDet D1

Při všech onemocnění bylo při druhém trénování o větším počtu trénovacích kroků navýšeno skóre detekce (viz tabulka 5.39). Průměrný výsledek této metriky pro bradavice dosáhl více než 99,5 %. Ve srovnání s výsledky této metriky s modelem EfficientDet D0 je navýšeno skóre detekce při druhém trénování o vyšší hodnoty, model EfficientDet D1 dokázal predikovat všechna onemocnění s vyšší jistotou.

ID	Atopický ekzém	Bradavice	Dyshidróza	Psoriáza
13	43,945	97,446	53,724	58,394
14	54,139	99,588	69,808	76,371

Tabulka 5.39: Test skóre detekce správně detekované a klasifikované plochy pro EfficientDet D1

Tabulka 5.40 udává výsledky průměrné detekované plochy navíc pro EfficientDet D1. Výsledky této metriky byly při druhém trénování sníženy u všech onemocnění, model je tedy schopen méně detektovat neanotovanou plochu, současně je i větší anotovaná plocha detekovaná a klasifikována do správné třídy onemocnění. Ve srovnání s modelem EfficientDet D0 je nejvíce znatelně snížena tato metrika pro bradavice.

ID	Atopický ekzém	Bradavice	Dyshidróza	Psoriáza
13	27,219	5,764	18,451	20,957
14	25,425	6,625	15,536	18,629

Tabulka 5.40: Test detekované plochy navíc pro EfficientDet D1

Nejčetnější hodnota metriky IoU (viz tabulka 5.41) se při obou trénováních až na výsledek pro dyshidrózu nezměnila. EfficientDet D0 dosáhlo v tomto testu lepších výsledků, hlavně pro překryv bounding boxů pro dyshidrózu, výsledky pro atopický ekzém však pro oba modely a všechna trénování zůstaly stejné.

ID	Atopický ekzém	Bradavice	Dyshidróza	Psoriáza
13	0,7	0,9	0,6	0,8
14	0,7	0,9	0,5	0,8

Tabulka 5.41: Nejčetnější hodnota IoU pro EfficientDet D1

Trénovací doba při úpravě parametrů a navýšení trénovacích kroků byla navýšena o více než 1 hodinu 30 minut (viz tabulka 5.42). Upravený model byl trénován o hodinu déle než upravený model EfficientDet D0. Avšak při přihlédnutí k celkově lepším výsledkům modelu, hlavně co se týče testu správně detekované a klasifikované plochy, skóre detekce a skóre detekované plochy navíc je využití modelu EfficientDet D1 na zvolené úloze lepší než EfficientDet D0.

ID	Trénovací doba
13	4 h 4 min
14	5 h 42 min

Tabulka 5.42: Trénovací doba pro EfficientDet D1

5.2.4 Experimenty s modely CenterNet

Při experimentech byly zvoleny dva modely využívající CenterNet jako přístup pro detekci objektů – CenterNet Hourglass a CenterNet ResNet-101. Oba přístupy využívají vstupní obraz o velikosti 512×512 pixelů stejně jako EfficientDet D0. V tabulce 5.43 jsou znázorněny upravené parametry pro první a druhé trénování pro CenterNet Hourglass. Tento model

jako jediný využívá optimalizátor bez využití cosinového sestupu, ale s manuálně nastavenými hodnotami změn learning rate během trénování – tedy např. při trénovacím kroku 16 000 je learning rate snížena z 0,001 na 0,0001. Druhé trénování modelu bylo navýšeno o 5 000 trénovacích kroků.

ID	num_steps	intitial_lr	step1	step1_lr	step2	step2_lr
15	25 000	0,001	16 000	0,0001	21 000	1e-05
16	30 000	0,003	19 000	0,0003	26 000	3e-05

Tabulka 5.43: Změny parametrů pro CenterNet Hourglass

Test průměrné správně detekované a klasifikované plochy je znázorněn v tabulkách 5.44 a 5.45. Při druhém trénování byla vylepšena tato metrika pro otisky prstu bez poškození, hodnota pro celý testovací dataset se také zvýšila s přihlédnutím hlavně k reálným otiskům prstů.

ID	Atopický ekzém	Bradavice	Dyshidróza	Psoriáza	Bez poškození
15	69,491	96,284	71,122	73,086	70,13
16	68,628	95,043	66,351	72,331	81,818

Tabulka 5.44: Test správně detekované a klasifikované plochy pro CenterNet Hourglass

ID	Celkově (reálné, syntetické)	Celkově (reálné)	Celkově (syntetické)
15	72,682	66,736	81,271
16	76,551	72,019	83,098

Tabulka 5.45: Test správně detekované a klasifikované plochy pro CenterNet Hourglass

Tabulka 5.46 obsahuje průměrné skóre detekce pro správně detekovanou a klasifikovanou plochu. Toto skóre je při druhém trénování navýšeno pouze u atopického ekzému.

ID	Atopický ekzém	Bradavice	Dyshidróza	Psoriáza
15	55,896	88,138	59,707	65,637
16	57,064	87,257	53,777	64,404

Tabulka 5.46: Test skóre detekce správně detekované a klasifikované plochy pro CenterNet Hourglass

Tabulka detekované plochy navíc 5.47 ukazuje, že největší procentuální snížení detekované plochy navíc při trénování s upravenými parametry a navýšením trénovacích kroků zaznamenaly bradavice. Dalším onemocněním, kterému se při druhém trénování snížila tato metrika, byla psoriáza.

ID	Atopický ekzém	Bradavice	Dyshidróza	Psoriáza
15	23,647	8,363	12,897	16,064
16	25,923	5,242	13,812	15,023

Tabulka 5.47: Test detekované plochy navíc pro CenterNet Hourglass

Nejvyšší možné překrytí predikováno bounding boxu s anotovaným bounding boxem se v obou trénováních modelu CenterNet Hourglass nezměnilo pro žádné z onemocnění

(viz tabulka 5.48). Výsledek pro psoriázu byl nejnižší z daných onemocnění, avšak ostatní onemocnění dosáhly nejčetnějšího překryvu 0,8. Znamená to tedy, že nejvíce predikovaných bounding boxů se překrývá s některou z anotací aspoň z osmdesáti procent.

ID	Atopický ekzém	Bradavice	Dyshidróza	Psoriáza
15	0,8	0,9	0,8	0,7
16	0,8	0,9	0,8	0,7

Tabulka 5.48: Nejčetnější hodnota IoU pro CenterNet Hourglass

I když CenterNet Hourglass pracuje se vstupními snímky o velikosti 512×512 pixelů jako EfficientDet D0, trénování tohoto modelu je mnohem pomalejší. Při úpravě parametrů a navýšení trénovacích kroků o 5 000 byl trénovací čas navýšen přibližně o hodinu. Celkově se tedy tento model trénoval při druhém trénování skoro 7 hodin (viz tabulka 5.49).

ID	Trénovací doba
15	5 h 45 min
16	6 h 53 min

Tabulka 5.49: Trénovací doba pro CenterNet Hourglass

Model CenterNet ResNet-101 byl trénován v první iteraci 50 000 trénovacích kroků – jedná se o stejnou hodnotu jako prvotní trénování modelů SSD MobileNet a EfficientDet D0. Na rozdíl od modelu CenterNet Hourglass je opět využito cosinového sestupu. V druhé iteraci byly trénovací kroky opět navýšeny o 10 000 stejně jako u EfficientDet D0. Nastavení změn parametrů je znázorněno v tabulce 5.50.

ID	num_steps	lr_base	warmup_lr	warmup_steps
17	50 000	0,001	0,00025	5 000
18	60 000	0,003	0,0003	4 000

Tabulka 5.50: Změny parametrů pro CenterNet ResNet-101

Správně detekovaná a klasifikovaná plocha byla při druhém trénování navýšena hlavně pro onemocnění bradavic a psoriázy (viz tabulky 5.51 a 5.52). Výsledek pro celý testovací dataset byl navýšen, opět s větším přihlédnutím k reálným otiskům prstů. Ve srovnání s CenterNet Hourglass jsou výsledky horší pro atopický ekzém a psoriázu, avšak lepší pro otisky prstů bez poškození. Model CenterNet ResNet-101 tedy více otiskům prstů bez poškození správně nepredikoval žádný bounding box s onemocněním. Souhrnné výsledky této metriky pro celý testovací dataset jsou pak mírně lepší, opět hlavně díky zlepšení této metriky pro reálné otisky prstů.

ID	Atopický ekzém	Bradavice	Dyshidróza	Psoriáza	Bez poškození
17	65,019	95,482	71,698	65,127	85,065
18	65,541	96,186	69,87	69,652	85,065

Tabulka 5.51: Test správně detekované a klasifikované plochy pro CenterNet ResNet-101

ID	Celkově (reálné, syntetické)	Celkově (reálné)	Celkově (syntetické)
17	76,608	73,682	80,834
18	77,193	74,624	80,903

Tabulka 5.52: Test správně detekované a klasifikované plochy pro CenterNet ResNet-101

Test skóre detekce pro správně detekovanou a klasifikovanou plochu je znázorněn v tabulce 5.53. Při druhém trénování došlo i přes navýšení počtu trénovacích kroků ke snížení jistoty modelu pro klasifikaci bounding boxu zejména do tříd onemocnění atopického ekzému a bradavice. Mírně byla však zvýšena jistota predikce modelu do kategorie dyshidrózy. Ve srovnání s CenterNet Hourglass došlo k navýšení skóre detekce pro atopický ekzém, dyshidrózu a psoriázu.

ID	Atopický ekzém	Bradavice	Dyshidráza	Psoriáza
17	66,8	88,77	64,08	74,604
18	62,642	85,385	66,556	74,396

Tabulka 5.53: Test skóre detekce správně detekované a klasifikované plochy pro CenterNet ResNet-101

Při druhém trénování byla snížena plocha detekována navíc hlavně u bradavic a dyshidrózy (viz tabulka 5.54). Výsledky pro tuto metriku pro CenterNet ResNet-101 byly celkově úspěšnější než pro CenterNet Hourglass. Došlo ke snížení obsahu detekované plochy hlavně u onemocnění psoriázy a atopického ekzému.

ID	Atopický ekzém	Bradavice	Dyshidráza	Psoriáza
17	20,895	6,029	12,899	13,552
18	22,628	5,742	10,052	13,922

Tabulka 5.54: Test detekované plochy navíc pro CenterNet ResNet-101

Následující tabulka 5.55 udává nejčetnější normalizované hodnoty IoU pro jednotlivá onemocnění. Při trénování s upravenými parametry byla tato metrika zlepšena pro dyshidrózu a psoriázu – predikované bounding boxy pro toto onemocnění mají tedy s nějakou ze skutečných anotací aspoň překryv 70 %. Výsledky této metriky pro předchozí model CenterNet Hourglass byly lepší zejména pro onemocnění dyshidrózy, kdy byl maximální překryv aspoň 80 %.

ID	Atopický ekzém	Bradavice	Dyshidráza	Psoriáza
17	0,8	0,9	0,6	0,6
18	0,8	0,9	0,7	0,7

Tabulka 5.55: Nejčetnější hodnota IoU pro CenterNet ResNet-101

Trénovací doba pro model CenterNet ResNet-101 je znázorněna v tabulce 5.56. Doba trénování byla vyšší než pro CenterNet Hourglass, při navýšení trénovacích kroků o hodnotu 10 000 byl trénovací čas navýšen skoro na 6 hodin. Výsledky, kterých oba modely s detekčním přístupem CenterNet dosáhly na zkoumané úloze, jsou spíše srovnatelné, největší zlepšení pro CenterNet ResNet-101 bylo pro metriku skóre detekce.

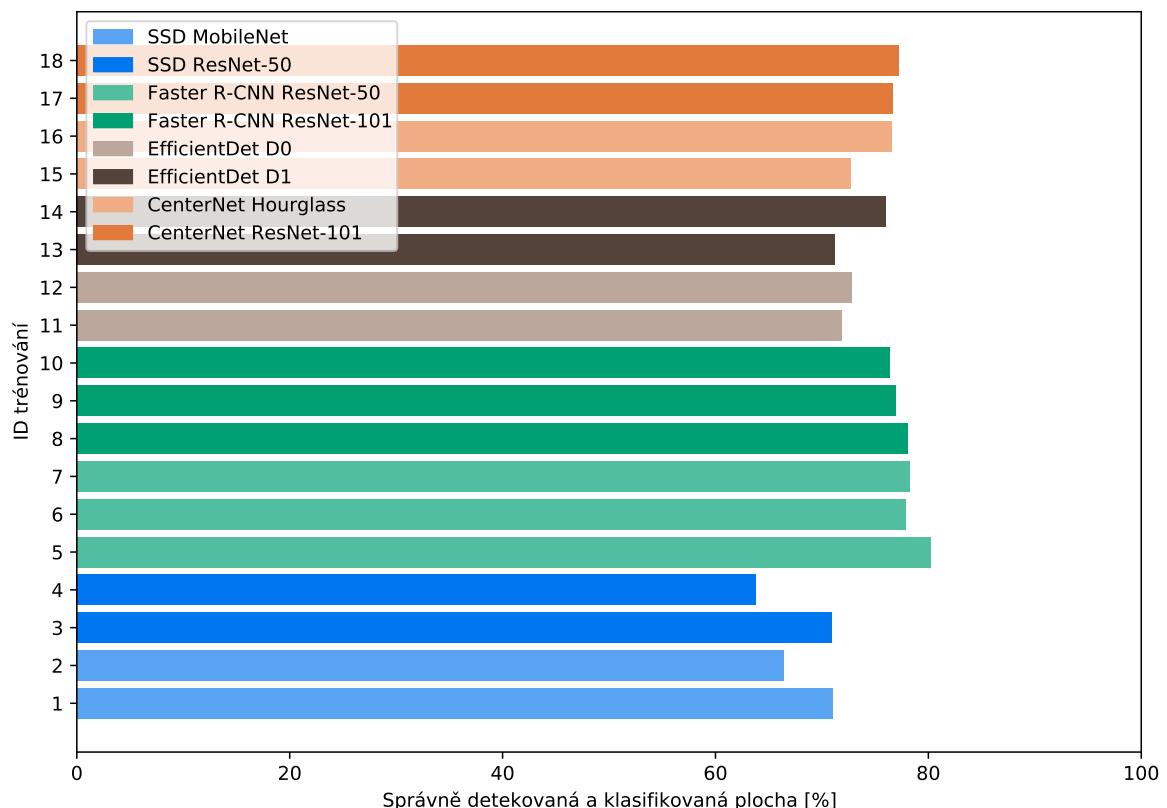
ID	Trénovací doba
17	6 h 9 min
18	7 h 7 min

Tabulka 5.56: Trénovací doba pro CenterNet ResNet-101

5.3 Zhodnocení experimentů

V předchozí sekci 5.2 byly detailně popsány experimenty s osmi modely patřící do čtyřech kategorií na základě zvoleného přístupu pro detekci objektů – SSD, Faster R-CNN, EfficientDet a CenterNet. U každého z modelů byly provedeny minimálně dvě trénování s úpravami parametrů za účelem vylepšení. V mnoha případech při zvýšení trénovacích kroků a změně parametrů k určitému vylepšení došlo, avšak u některých z modelů přínosy nebyly příliš úspěšné – např. model SSD ResNet-50 nebo Faster R-CNN modely, kdy nastával problém s přetrenováním.

Nejdůležitější metrikou, která dané modely popisuje, je test správně detekované a klasifikované plochy. Potenciální model pro detekci a klasifikaci onemocnění by měl s co největší přesností předpovědět možné onemocnění daného pacienta. Tento model potom může být součástí biometrického systému. Graf 5.1 znázorňuje výsledky testu správně detekované a klasifikované plochy pro celý testovací dataset každého z trénování.



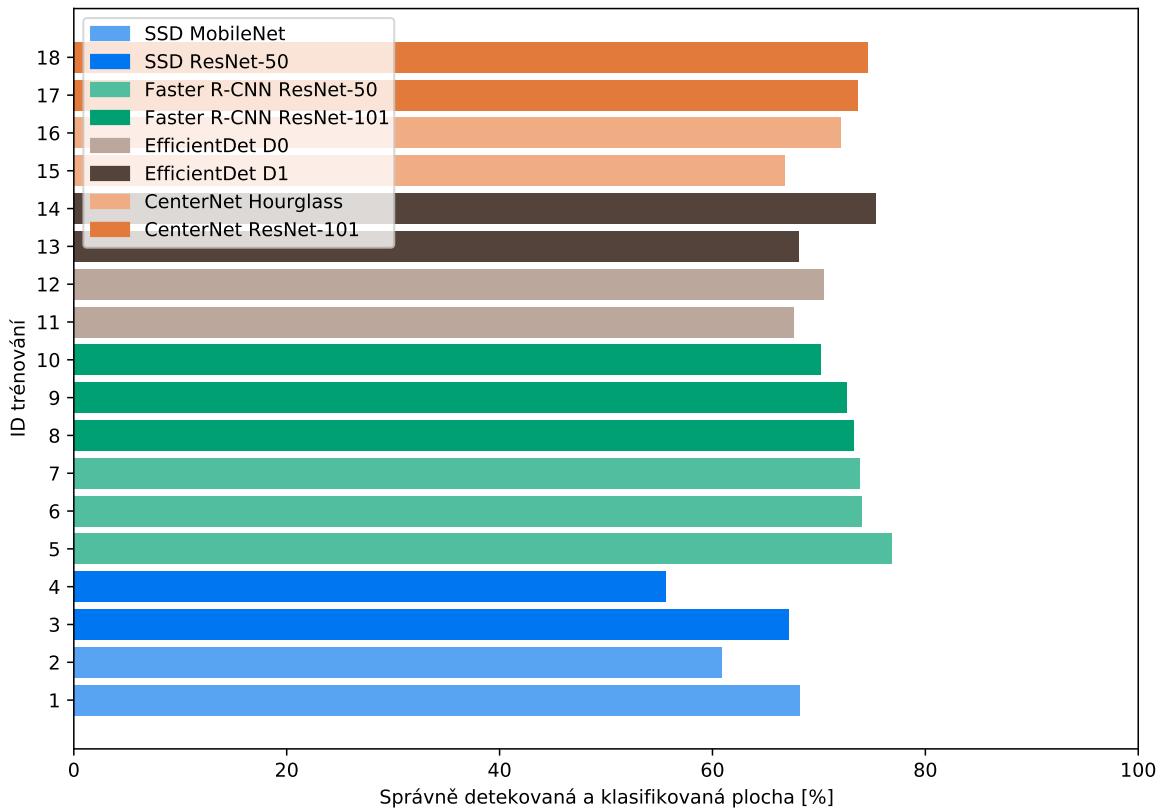
Obrázek 5.1: Test správně detekované a klasifikované plochy pro testovací dataset

Pro každý z 8 modelů bylo vybráno ID trénování, které dosáhlo nejlepších výsledků právě pro tuto metriku na celém testovacím datasetu (viz tabulka 5.57). Všechny tyto modely jsou posléze seřazeny od nejlepšího po nejhorší.

Pořadí	Název modelu	ID trénování	Celkově (reálné, syntetické)
1.	Faster R-CNN ResNet-50	5	80,263
2.	Faster R-CNN ResNet-101	8	78,043
3.	CenterNet ResNet-101	18	77,193
4.	CenterNet Hourglass	16	76,551
5.	EfficientDet D1	14	76,027
6.	EfficientDet D0	12	72,788
7.	SSD MobileNet	1	70,991
8.	SSD ResNet-50	3	70,934

Tabulka 5.57: Test správně detekované a klasifikované plochy pro testovací dataset – srovnání modelů

Graf 5.2 a tabulka 5.58 udávají srovnání modelů opět podle metriky správně detekované a klasifikované plochy, avšak s využitím pouze podmnožiny testovacího datasetu – reálných snímků pacientů. Toto srovnání je důležité hlavně z hlediska analýzy pouze otisků prstů reálných uživatelů bez syntetických otisků a tedy potenciálně spolehlivější pohled na problematiku chování natrénovaného modelu na komplexních reálných otiscích.



Obrázek 5.2: Test správně detekované a klasifikované plochy pro reálné otisky testovacího datasetu

Pořadí	Název modelu	ID trénování	Celkově (reálné)
1.	Faster R-CNN ResNet-50	5	76,875
2.	EfficientDet D1	14	75,329
3.	CenterNet ResNet-101	18	74,624
4.	Faster R-CNN ResNet-101	8	73,241
5.	CenterNet Hourglass	16	72,019
6.	EfficientDet D0	12	70,448
7.	SSD MobileNet	1	68,226
8.	SSD ResNet-50	3	67,123

Tabulka 5.58: Test správně detekované a klasifikované plochy pro reálné otisky testovacího datasetu – srovnání modelů

I přes problémy s vylepšováním trénování pro model Faster R-CNN ResNet-50 dosáhlo jeho první trénování s původními konfiguračními parametry nejlepších výsledků pro celý testovací dataset i jeho podmnožinu obsahující pouze otisky prstů reálných pacientů. I přes delší dobu trénování bych se však spíše přikláněla k využití modelů EfficientDet D1 nebo CenterNet ResNet-101. Jedná se o modernější způsoby, u kterých se zvýšením trénovacích kroků a úpravou parametrů podařilo vylepšit jejich vlastnosti.

Každé z onemocnění – atopický ekzém, bradavice, dyshidróza a psoriáza – mělo odlišný vliv na různé metriky. Bylo provedeno celkem 18 trénování (dva pro každý z osmi modelů, kromě modelů Faster R-CNN, kde byla provedena tři trénování). Přes všech těchto 18 trénování jsou spočítány průměrné hodnoty pro každou z kategorií pro zvolenou metriku.

Tabulka 5.59 udává průměrnou hodnotu testu správně detekované a klasifikované plochy pro jednotlivé kategorie s ohledem na všechna 18 trénování. Nejlepších výsledků v této metrice dosahují bradavice, kdy je správně detekováno a klasifikováno průměrně 91 % celkové predikované plochy. Bradavice patří tradičně k snazším onemocněním otisků prstů pro rozpoznávání hlavně díky svému pravidelnému kulovitému tvaru. Naopak nejtěžším onemocněním pro správnou detekci a rozpoznání je atopický ekzém – toto onemocnění se projevuje nejen bílými čarami přes papilární linie, ale i nepravidelnými tmavými místy. Celkově se jedná o onemocnění zabírající velikou plochu otisku prstu, kdy natrénovaný model může mít potíže komplexní projevy atopického ekzému rozpoznat.

Atopický ekzém	Bradavice	Dyshidróza	Psoriáza	Bez poškození
58,713	91,207	77,469	72,518	79,174

Tabulka 5.59: Test správně detekované a klasifikované plochy

Natrénovaný model dokáže predikovaný bounding box zařadit do onemocnění bradavic s průměrnou přesností přes 92 %. Naopak největší míru nejistoty opět vykazuje klasifikace do onemocnění atopického ekzému (viz tabulka 5.60).

Atopický ekzém	Bradavice	Dyshidróza	Psoriáza
65,553	92,454	73,424	73,791

Tabulka 5.60: Test skóre detekce správně detekované a klasifikované plochy

Pouze průměrně 9 % plochy predikovaného bounding boxu je neanotovaného pro onemocnění bradavic. Pro plošná onemocnění otisku prstu – atopický ekzém, dyshidrózu a

psoriázu – je tato detekovaná plocha navíc vyšší a u atopického ekzému je průměrně dosaženo až $\frac{1}{4}$ celkové predikované plochy (viz tabulka 5.61).

Atopický ekzém	Bradavice	Dyshidróza	Psoriáza
25,744	8,73	17,828	19,294

Tabulka 5.61: Test detekované plochy navíc

Pro všech 18 trénování byly zjištěny nejčetnější hodnoty IoU pro každé z onemocnění. Následně je na rozdíl od předchozích metrik spočítán místo průměru modus – nejčetnější normalizovaná hodnota IoU přes všechna 18 trénování. Výsledky jsou zobrazeny v tabulce 5.62. Nejčastěji k nejmenšímu maximálnímu překryvu predikovaného a anotovaného bounding boxu dochází u dyshidrózy. Naopak překryv pro bradavice je nejčetněji okolo 90 %.

Atopický ekzém	Bradavice	Dyshidróza	Psoriáza
0,7	0,9	0,6	0,8

Tabulka 5.62: Nejčetnější hodnota IoU

Kapitola 6

Závěr

Cílem diplomové práce bylo analyzovat současné přístupy pro detekci a klasifikaci objektů založených na konvolučních neuronových sítích. Tyto metody byly následně využity na úloze detekce a klasifikace onemocnění otisků prstů – pro experimenty byly vybrány onemocnění atopický ekzém, bradavice, dyshidróza a psoriáza.

Hlavním problémem datasetu onemocnění reálných pacientů je jeho malá velikost a absence anotací onemocnění. Tento dataset onemocnění výzkumné skupiny STRaDe byl ručně přetříděn a anotován. Následně bylo využito syntetických otisků prstů a poškozování těchto otisků s využitím generátorů jednotlivých onemocnění.

Pro experimenty bylo využito 8 modelů z knihovny TensorFlow Model Garden založených na čtyřech přístupech pro detekci objektů – SSD, Faster R-CNN, EfficientDet a CenterNet. Každý model byl trénován minimálně dvakrát – poprvé bez úprav příslušného konfiguračního souboru a posléze s úpravami konfiguračních parametrů za účelem zlepšení metrik pro evaluaci. Úspěšnost modelu na testovacím datasetu byla hodnocena např. z hlediska správně detekované a klasifikované plochy, skóre detekce pro správně detekovanou a klasifikovanou plochu, nejčetnejší hodnoty překryvu predikovaného s anotovaným bounding boxem a dalších.

Během delšího trénování a úprav parametrů byly vylepšeny hlavně modely EfficientDet a CenterNet. Nejúspěšnějšími modely z hlediska správně detekované a klasifikované plochy na snímcích reálných pacientů byly modely Faster R-CNN ResNet-50 a EfficientDet D1. Atopický ekzém byl nejvíce komplexním a komplikovaným onemocněním při trénování daných modelů.

Při testu správně detekované a klasifikované plochy pro reálné otisky byla dosažena nejlepší úspěšnost 76,875 %. Atopický ekzém (reálné a syntetické otisky) byl detekován a klasifikován s průměrnou přesností 58,713 %. Výsledky pro onemocnění bradavic byly průměrně 91,207 %. Při otiscích prstů postižených dyshidrózou nebo psoriázou byla hodnota této metriky 77,469 % a 72,518 %. Další z důležitých metrik bylo průměrné skóre detekce. Nejnižšího výsledku dosahoval atopický ekzém s 65,553 %, nejlepšího bradavice s 92,454 %. Detekční skóre pro dyshidrózu a psoriázu bylo velmi podobné – 73,424 % a 73,791 %.

Potenciální rozšíření práce spočívá zejména v rozšíření datasetu onemocnění reálných pacientů. I když je část datasetu této práce složena ze syntetických otisků prstů, při studiu datasetu reálných pacientů je zřejmé, že se plošná onemocnění atopického ekzému, dyshidrózy a psoriázy mnohdy projevují více způsoby. Při rozšiřování datasetu by mělo být dbáno na co nejlepší možnou kvalitu při snímání otisků prstů s onemocněním.

Literatura

- [1] AFTERACADEMY. *Mastering Activation Functions in Neural Networks* [online]. AfterAcademy, prosinec 2019 [cit. 2022-01-20]. Dostupné z: <https://afteracademy.com/blog/mastering-activation-functions-in-neural-networks>.
- [2] ALMOG, U. *CenterNet, Explained* [online]. Towards Data Science, 2021 [cit. 2021-12-22]. Dostupné z: <https://towardsdatascience.com/centernet-explained-a7386f368962>.
- [3] ALONSO-FERNANDEZ, F., ROLI, F., MARCIALIS, G., FIERREZ, J., ORTEGA-GARCIA, J. et al. Performance of fingerprint quality measures depending on sensor technology. *Journal of Electronic Imaging*. 2008, sv. 17, č. 1, s. 1–11. ISSN 1017-9909.
- [4] BAROTOVÁ, S. *Detektor kožních onemocnění u technologie otisků prstů*. Brno, CZ, 2017. Bakalářská práce. Vysoké učení technické v Brně, Fakulta informačních technologií. Dostupné z: <https://www.fit.vut.cz/study/thesis/20057/>.
- [5] BROWNLEE, J. *A Gentle Introduction to Dropout for Regularizing Deep Neural Networks* [online]. Machine Learning Mastery, prosinec 2018 [cit. 2022-05-04]. Dostupné z: <https://machinelearningmastery.com/dropout-for-regularizing-deep-neural-networks/>.
- [6] BROWNLEE, J. *When to Use MLP, CNN, and RNN Neural Networks* [online]. Machine Learning Mastery, červenec 2018 [cit. 2022-01-19]. Dostupné z: <https://machinelearningmastery.com/when-to-use-mlp-cnn-and-rnn-neural-networks/>.
- [7] BROWNLEE, J. *A Gentle Introduction to Pooling Layers for Convolutional Neural Networks* [online]. Towards Data Science, 2019 [cit. 2021-12-23]. Dostupné z: <https://machinelearningmastery.com/pooling-layers-for-convolutional-neural-networks/>.
- [8] BROWNLEE, J. *Loss and Loss Functions for Training Deep Learning Neural Networks* [online]. Machine Learning Mastery, leden 2019 [cit. 2021-12-09]. Dostupné z: <https://machinelearningmastery.com/loss-and-loss-functions-for-training-deep-learning-neural-networks/>.
- [9] BROWNLEE, J. *How to Choose an Activation Function for Deep Learning* [online]. Machine Learning Mastery, leden 2021 [cit. 2021-12-01]. Dostupné z: <https://machinelearningmastery.com/choose-an-activation-function-for-deep-learning/>.
- [10] BRZOBOHATÁ, K. *Dermatoglyfika* [online]. 2017 [cit. 2021-12-09]. Dostupné z: <https://is.muni.cz/el/sci/jaro2017/Bi5120c/Dermatoglyfika.pdf>.

- [11] CORREA, S. *Cosine Learning rate decay* [online]. Medium, září 2019 [cit. 2022-03-29]. Dostupné z: <https://scorrea92.medium.com/cosine-learning-rate-decay-e8b50aa455b>.
- [12] DINESH. *CNN vs MLP for Image Classification* [online]. Medium, 2019 [cit. 2021-12-23]. Dostupné z: <https://medium.com/analytics-vidhya/cnn-convolutional-neural-network-8d0a292b4498>.
- [13] DRAHANSKY, M., DOLEZEL, M., URBANEK, J., BREZINOVA, E. a KIM, T. Influence of Skin Diseases on Fingerprint Recognition. *BioMed Research International*. 2012, sv. 2012, s. 144–157. ISSN 2578-8892.
- [14] DRAHANSKÝ, M. *Biometrické systémy - Studijní opora* [online]. 2006 [cit. 2021-12-07]. Dostupné z: https://www.fit.vutbr.cz/study/courses/BIO/private/BIO_Studijni_opora.pdf.
- [15] DRAHANSKÝ, M., BŘEZINOVÁ, E., ORSÁG, F. a HEJTMÁNKOVÁ, D. Dermatologické faktory ovlivňující snímání otisků prstů. *Kriminalistika*. 2010, sv. 42, č. 3. ISSN 1210-9150.
- [16] DRAHANSKÝ, M. a KANICH, O. *Rozpoznávání podle otisků prstů* [online]. 2020 [cit. 2021-12-07]. Dostupné z: https://www.fit.vutbr.cz/study/courses/BIO/private/04_BIO_Prednaska_OK_2020.pdf.
- [17] DYLEVSKÝ, I. a JEŽEK, P. *Kožní ústrojí - Anatomie kůže* [online]. [cit. 2021-12-10]. Dostupné z: <https://vos.palestra.cz/skripta/anatomie/12a1.htm>.
- [18] FOOTE, K. D. *A Brief History of Neural Networks* [online]. Dataversity, listopad 2021 [cit. 2021-12-25]. Dostupné z: <https://www.dataversity.net/a-brief-history-of-neural-networks/>.
- [19] GANDHI, R. *R-CNN, Fast R-CNN, Faster R-CNN, YOLO — Object Detection Algorithms* [online]. Towards Data Science, září 2018 [cit. 2021-12-03]. Dostupné z: <https://towardsdatascience.com/r-cnn-fast-r-cnn-faster-r-cnn-yolo-object-detection-algorithms-36d53571365e>.
- [20] GEY, T. *What is Neural Architecture Search? And Why Should You Care?* [online]. Towards Data Science, září 2021 [cit. 2021-12-15]. Dostupné z: <https://towardsdatascience.com/what-is-neural-architecture-search-and-why-should-you-care-1e22393de461>.
- [21] GIRSHICK, R. Fast R-CNN. *Computing Research Repository*. 2015, sv. 1504. ISSN 2331-8422.
- [22] GIRSHICK, R. B., DONAHUE, J., DARRELL, T. a MALIK, J. Rich feature hierarchies for accurate object detection and semantic segmentation. *Computing Research Repository*. 2013, sv. 1311. ISSN 2331-8422.
- [23] HOFESMANN, E. *IoU a better detection evaluation metric* [online]. Towards Data Science, srpen 2020 [cit. 2022-03-16]. Dostupné z: <https://towardsdatascience.com/iou-a-better-detection-evaluation-metric-45a511185be1>.

- [24] HONG, L., WAN, Y. a JAIN, A. Fingerprint Image Enhancement: Algorithm and Performance Evaluation. *IEEE Transactions on Pattern Analysis and Machine Intelligence*. 1998, sv. 20, č. 8, s. 777–789. ISSN 0182-8828.
- [25] HOWARD, A. G., ZHU, M., CHEN, B., KALENICHENKO, D., WANG, W. et al. MobileNets: Efficient Convolutional Neural Networks for Mobile Vision Applications. *Computing Research Repository*. 2017, sv. 1704. ISSN 2331-8422.
- [26] HUANG, J., RATHOD, V., SUN, C., ZHU, M., KORATTIKARA, A. et al. Speed/accuracy trade-offs for modern convolutional object detectors. In: *2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*. Los Alamitos, CA, USA: IEEE Computer Society, červenec 2017, s. 3296–3297. ISBN 1063-6919.
- [27] INFORMEDHEALTH.ORG. *Warts: Overview* [online]. Cologne, Germany: Institute for Quality and Efficiency in Health Care (IQWiG), listopad 2019 [cit. 2021-11-24]. Dostupné z: <https://www.ncbi.nlm.nih.gov/books/NBK279586/>.
- [28] JAIN, A. a FENG, J. Latent Fingerprint Matching. *IEEE transactions on pattern analysis and machine intelligence*. 2011, sv. 33, s. 88–100. ISSN 1939-3539.
- [29] KAUSHIK, A. *Understanding ResNet50 architecture* [online]. OpenGenusIQ, březen 2020 [cit. 2021-12-14]. Dostupné z: <https://iq.opengenus.org/resnet50-architecture/>.
- [30] KERAS. *About Keras* [online]. Keras, 2021 [cit. 2021-12-25]. Dostupné z: <https://keras.io/about/>.
- [31] KOLB, L. a FERRER BRUKER., S. J. *Atopic Dermatitis* [online]. StatPearls Publishing, srpen 2021 [cit. 2021-11-24]. Dostupné z: <https://www.ncbi.nlm.nih.gov/books/NBK448071/>.
- [32] KOŠTÁK, D. *Generování projevu kožního onemocnění do syntetických otisků prstů z generátoru Anguli*. Brno, CZ, 2018. Bakalářská práce. Vysoké učení technické v Brně, Fakulta informačních technologií. Dostupné z: <https://www.fit.vut.cz/study/thesis/20890/>.
- [33] LIU, L., JIANG, H., HE, P., CHEN, W., LIU, X. et al. On the Variance of the Adaptive Learning Rate and Beyond. In: *8th International Conference on Learning Representations, ICLR 2020, Addis Ababa, Ethiopia, April 26-30, 2020*. Addis Ababa, Ethiopia: OpenReview.net, Květen 2020.
- [34] LIU, W., ANGUELOV, D., ERHAN, D., SZEGEDY, C., REED, S. et al. SSD: Single Shot MultiBox Detector. *Lecture Notes in Computer Science*. 2016, s. 21–37. ISSN 1611-3349.
- [35] LOSHCHILOV, I. a HUTTER, F. SGDR: Stochastic Gradient Descent with Restarts. In: *5th International Conference on Learning Representations, ICLR 2017, Toulon, France, April 24-26, 2017, Conference Track Proceedings*. Toulon, France: OpenReview.net, Duben 2017.
- [36] MALTONI, D., MAIO, D., JAIN, A. K. a PRABHAKAR, S. *Handbook of Fingerprint Recognition*. 2. vyd. Springer, 2009. ISBN 978-1-84882-253-5.

- [37] NAIR, P. A. a BADRI, T. *Psoriasis* [online]. StatPearls Publishing, srpen 2021 [cit. 2021-11-24]. Dostupné z: <https://www.ncbi.nlm.nih.gov/books/NBK448194/>.
- [38] NATIONAL HEALTH SERVICE. *Raynaud's* [online]. NHS, říjen 2020 [cit. 2021-12-21]. Dostupné z: <https://www.nhs.uk/conditions/raynauds/>.
- [39] NEWELL, A., YANG, K. a DENG, J. Stacked Hourglass Networks for Human Pose Estimation. *Computing Research Repository*. 2016, sv. 1603. ISSN 2331-8422.
- [40] OPENCV. *Image Thresholding* [online]. OpenCV, 2021 [cit. 2021-12-22]. Dostupné z: https://docs.opencv.org/3.4/d7/d4d/tutorial_py_thresholding.html.
- [41] OPENCV. *Morphological Transformations* [online]. OpenCV, 2021 [cit. 2021-12-22]. Dostupné z: https://docs.opencv.org/3.4/d7/d4d/tutorial_py_thresholding.html.
- [42] OPPERMANN, A. *Regularization in Deep Learning — L1, L2, and Dropout* [online]. Towards Data Science, únor 2020 [cit. 2022-03-28]. Dostupné z: <https://towardsdatascience.com/regularization-in-deep-learning-l1-l2-and-dropout-377e75acc036>.
- [43] PERNICKÝ, R. *Pokročilé rozpoznávání podle otisků prstů* [online]. 2020 [cit. 2021-12-21]. Dostupné z: https://www.fit.vutbr.cz/study/courses/BIO/private/08_BIO_Prednaska_RP_2020.pdf.
- [44] PERUMANOOR, T. J. *What is VGG16? — Introduction to VGG16* [online]. Great Learning, září 2021 [cit. 2021-12-13]. Dostupné z: <https://medium.com/@mygreatlearning/what-is-vgg16-introduction-to-vgg16-f2d63849f615>.
- [45] REN, S., HE, K., GIRSHICK, R. B. a SUN, J. Faster R-CNN: Towards Real-Time Object Detection with Region Proposal Networks. *Computing Research Repository*. 2015, sv. 1506. ISSN 2331-8422.
- [46] RIDNIK, T. *Revealing the Dark Magic Behind Deep Learning* [online]. Alibaba Cloud, červenec 2019 [cit. 2022-03-28]. Dostupné z: https://www.alibabacloud.com/blog/revealing-the-dark-magic-behind-deep-learning_595111.
- [47] RUDER, S. *An overview of gradient descent optimization algorithms*. 2017 [cit. 2021-12-20]. Dostupné z: <https://arxiv.org/abs/1609.04747>.
- [48] SHARMA, S. *Activation Functions in Neural Networks* [online]. Towards Data Science, září 2017 [cit. 2021-12-01]. Dostupné z: <https://towardsdatascience.com/activation-functions-neural-networks-1cbd9f8d91d6>.
- [49] SIMALANGO, M. F. *What Object Categories / Labels Are In COCO Dataset?* [online]. Amikelive, 2018 [cit. 2021-12-22]. Dostupné z: <https://tech.amikelive.com/node-718/what-object-categories-labels-are-in-coco-dataset/>.
- [50] SMITH, M. P., LY, K., THIBODEAUX, Q., BHUTANI, T., LIAO, W. et al. Acrodermatitis continua of Hallopeau: clinical perspectives. *Psoriasis: Targets and Therapy*. 2019, sv. 9, s. 65–72. ISSN 2230-326X.
- [51] SOUKUP, T. a VELETA, T. Systémová sklerodermie v roce 2017. *Vnitřní lékařství*. 2018, č. 2, s. 146–154. ISSN 1801–7592.

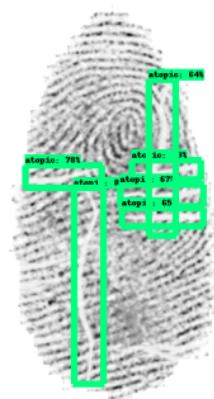
- [52] SVORADOVÁ, V. *Generování onemocnění kůže do syntetických otisků prstů z SFinGe*. Brno, CZ, 2019. Bakalářská práce. Vysoké učení technické v Brně, Fakulta informačních technologií. Dostupné z: <https://www.fit.vut.cz/study/thesis/21893/>.
- [53] TAN, M. a LE, Q. V. *EfficientNet: Improving Accuracy and Efficiency through AutoML and Model Scaling* [online]. Google AI Blog, květen 2019 [cit. 2021-12-14]. Dostupné z: <https://ai.googleblog.com/2019/05/efficientnet-improving-accuracy-and.html>.
- [54] TAN, M. a LE, Q. V. EfficientNet: Rethinking Model Scaling for Convolutional Neural Networks. In: *Proceedings of the 36th International Conference on Machine Learning*. Long Beach, California, USA: PLMR, červenec 2019, s. 6105–6114. ISBN 978-1-5108-8698-8.
- [55] TAN, M., PANG, R. a LE, Q. V. EfficientDet: Scalable and Efficient Object Detection. In: *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*. Seattle, WA, USA: IEEE, červenec 2020, s. 10778–10787. ISBN 978-1-7281-7168-5.
- [56] TAN, M. a YU, A. *EfficientDet: Towards Scalable and Efficient Object Detection* [online]. Google AI Blog, duben 2020 [cit. 2021-12-13]. Dostupné z: <https://ai.googleblog.com/2020/04/efficientdet-towards-scalable-and.html>.
- [57] ŠALKO, M. *Detekce a klasifikace poškození otisku prstu s využitím neuronových sítí*. Brno, CZ, 2020. Bakalářská práce. Vysoké učení technické v Brně, Fakulta informačních technologií. Dostupné z: <https://www.fit.vut.cz/study/thesis/22567/>.
- [58] WANG, C.-F. *A Basic Introduction to Separable Convolutions* [online]. Towards Data Science, srpen 2018 [cit. 2021-12-25]. Dostupné z: <https://towardsdatascience.com/a-basic-introduction-to-separable-convolutions-b99ec3102728>.
- [59] WANG, C.-F. *The Vanishing Gradient Problem* [online]. Towards Data Science, leden 2019 [cit. 2021-12-14]. Dostupné z: <https://towardsdatascience.com/the-vanishing-gradient-problem-69bf08b15484>.
- [60] YU, H., CHEN, C., DU, X., LI, Y., RASHWAN, A. et al. *TensorFlow Model Garden* [online]. 2020 [cit. 2021-12-22]. Dostupné z: <https://github.com/tensorflow/models>.
- [61] ZBOŘIL, F. a ZBOŘIL, F. V. *Přednáška 9 - Rozpoznávání* [online]. 2021 [cit. 2021-12-21]. Dostupné z: https://www.fit.vutbr.cz/study/courses/IZU/private/1819izu_9.pdf.
- [62] ZEREN, M. T., AYTULUN, S. K. a KIRELLI, Y. Comparison of SSD and Faster R-CNN Algorithms to Detect the Airports with Data Set Which Obtained From Unmanned Aerial Vehicles and Satellite Images. *European Journal of Science and Technology*. 2020, č. 19, s. 643–658. ISSN 2148-2683.

Příloha A

Ukázka predikcí onemocnění



Obrázek A.1: Atopický ekzém – reálný otisk



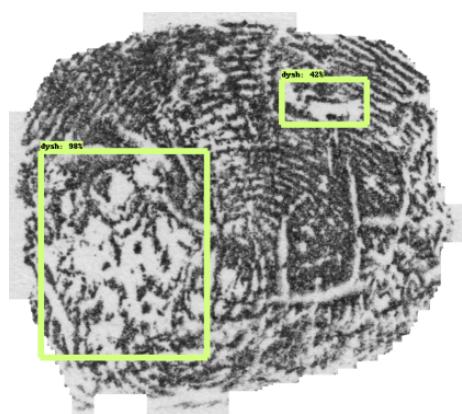
Obrázek A.2: Atopický ekzém – syntetický otisk



Obrázek A.3: Bradavice – reálný otisk



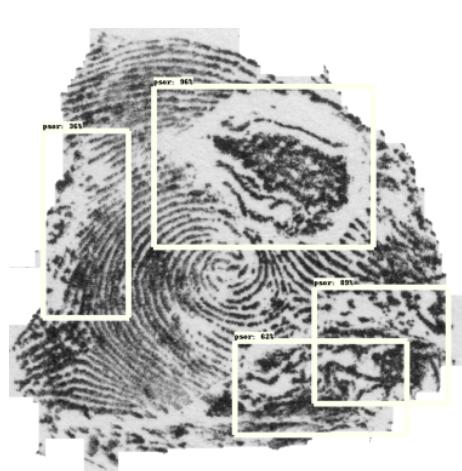
Obrázek A.4: Bradavice – syntetický otisk



Obrázek A.5: Dyshidróza – reálný otisk



Obrázek A.6: Dyshidróza – syntetický otisk



Obrázek A.7: Psoriáza – reálný otisk



Obrázek A.8: Psoriáza – syntetický otisk