

VYSOKÉ UČENÍ TECHNICKÉ V BRNĚ

Fakulta elektrotechniky  
a komunikačních technologií

DIPLOMOVÁ PRÁCE

Brno, 2021

Bc. Barbora Matoušková



# VYSOKÉ UČENÍ TECHNICKÉ V BRNĚ

BRNO UNIVERSITY OF TECHNOLOGY

## FAKULTA ELEKTROTECHNIKY A KOMUNIKAČNÍCH TECHNOLOGIÍ

FACULTY OF ELECTRICAL ENGINEERING AND COMMUNICATION

## ÚSTAV BIOMEDICÍNSKÉHO INŽENÝRSTVÍ

DEPARTMENT OF BIOMEDICAL ENGINEERING

## DETEKCE CIZÍCH OBJEKTŮ V RENTGENOVÝCH SNÍMCÍCH HRUDNÍKU S VYUŽITÍM METOD STROJOVÉHO UČENÍ

DETECTION OF FOREIGN OBJECTS IN X-RAY CHEST IMAGES USING MACHINE LEARNING METHODS

### DIPLOMOVÁ PRÁCE

MASTER'S THESIS

### AUTOR PRÁCE

AUTHOR

Bc. Barbora Matoušková

### VEDOUCÍ PRÁCE

SUPERVISOR

Ing. Jiří Chmelík, Ph.D.

BRNO 2021



# Diplomová práce

magisterský navazující studijní program **Biomedicínské inženýrství a bioinformatika**

Ústav biomedicínského inženýrství

**Studentka:** Bc. Barbora Matoušková

**ID:** 195716

**Ročník:** 2

**Akademický rok:** 2020/21

## NÁZEV TÉMATU:

### **Detekce cizích objektů v rentgenových snímcích hrudníku s využitím metod strojového učení**

#### **POKyny PRO VYPRACOVÁNÍ:**

1) Seznamte se s problematikou detekce cizích objektů v medicínských obrazových datech a možnostmi jejich automatické detekce. 2) Vypracujte literární rešerši zabývající se metodami strojového učení a jejich využitím pro detekci objektů v medicínských obrazech. 3) Na základě provedené rešerše navrhnete algoritmus pro detekci cizích objektů ve snímcích hrudníku pořízených planárním rentgenem. 4) Ve vhodně zvoleném programovacím jazyce implementujte vybranou metodu strojového učení společně s implementací vhodně zvoleného předzpracování dat. 5) Navržený postup otestujte na dostupné databázi a optimalizujte z hlediska maximalizace úspěšnosti detekce. 6) Statisticky zhodnoťte funkčnost navržené metody a výsledky objektivně porovnejte s jinými existujícími přístupy. Dosažené výsledky vhodně diskutujte.

#### **DOPORUČENÁ LITERATURA:**

[1] DHILLON, Anamika a Gyanendra K. VERMA. Convolutional neural network: a review of models, methodologies and applications to object detection. Progress in Artificial Intelligence. 2020, 9(2), 85-112. DOI: 10.1007/s13748-019-00203-0. ISSN 2192-6352

[2] DESHPANDE, Hrishikesh, Tim HARDER, Axel SAALBACH, Abhivyakti SAWARKAR a Thomas BUELOW. Detection Of Foreign Objects In Chest Radiographs Using Deep Learning. In: 2020 IEEE 17th International Symposium on Biomedical Imaging Workshops (ISBI Workshops). IEEE, 2020, 2020, s. 1-4. DOI: 10.1109/ISBIWorkshops50223.2020.9153350. ISBN 978-1-7281-7401-3.

**Termín zadání:** 8.2.2021

**Termín odevzdání:** 21.5.2021

**Vedoucí práce:** Ing. Jiří Chmelík, Ph.D.

**prof. Ing. Ivo Provazník, Ph.D.**  
předseda rady studijního programu

#### **UPOZORNĚNÍ:**

Autor diplomové práce nesmí při vytváření diplomové práce porušit autorská práva třetích osob, zejména nesmí zasahovat nedovoleným způsobem do cizích autorských práv osobnostních a musí si být plně vědom následků porušení ustanovení § 11 a následujících autorského zákona č. 121/2000 Sb., včetně možných trestněprávních důsledků vyplývajících z ustanovení části druhé, hlavy VI. díl 4 Trestního zákoníku č.40/2009 Sb.

## ABSTRAKT

Cizí objekty v RTG snímcích hrudníku způsobují komplikace během automatického zpracování snímku. Abychom zabránily chybám, které vznikají právě kvůli těmto cizím objektům, je třeba je nejprve automaticky vyhledat a z následné analýzy je vynechat. Jedná se především o knoflíky, šperky, implantáty, dráty či trubičky. Zároveň nalezení pacemakerů a jiných voperovaných zařízení může pomoci při automatickém zpracování. Cílem této práce bylo navrhnout metodu pro detekci cizích objektů v RTG snímcích hrudníku. Pro tento úkol byla zvolena metoda Faster R-CNN s předtrénovanou sítí ResNet50 pro extrakci příznaků, která byla natrénována na 4 000 snímcích a následně otestována na 1 000 snímcích z veřejně dostupné databáze. Po nalezení optimálních učících parametrů se podařilo natrénovat síť, která dosahuje 75% přesnosti, 77% senzitivity a 76% F1 skóre. Určitá část chyby je ovšem tvořena nejednotnými anotacemi objektů v datech, kdy ne všechny anotované cizí objekty se nachází v oblasti plic, jak je udáno v popisu.

## KLÍČOVÁ SLOVA

detekce objektů v obrazech, Faster R-CNN, RTG, plíce, strojové učení, hluboké učení, RPN

## ABSTRACT

Foreign objects in Chest X-ray (CXR) cause complications during automatic image processing. To prevent errors caused by these foreign objects, it is necessary to automatically find them and omit them in the analysis. These are mainly buttons, jewellery, implants, wires and tubes. At the same time, finding pacemakers and other placed devices can help with automatic processing. The aim of this work was to design a method for the detection of foreign objects in CXR. For this task, Faster R-CNN method with a pre-trained ResNet50 network for feature extraction was chosen which was trained on 4 000 images and lately tested on 1 000 images from a publicly available database. After finding the optimal learning parameters, it was managed to train the network, which achieves 75% precision, 77% recall and 76% F1 score. However, a certain part of the error is formed by non-uniform annotations of objects in the data because not all annotated foreign objects are located in the lung area, as stated in the description.

## KEYWORDS

object detection in images, Faster R-CNN, X-ray, lungs, machine learning, deep learning, RPN

MATOUŠKOVÁ, Barbora. *Detekce cizích objektů v rentgenových snímcích hrudníku s využitím metod strojového učení*. Brno, 2021, 66 s. Diplomová práce. Vysoké učení technické v Brně, Fakulta elektrotechniky a komunikačních technologií, Ústav biomedicínského inženýrství. Vedoucí práce: Ing. Jiří Chmelík, Ph.D.



## PROHLÁŠENÍ

Prohlašuji, že svou diplomovou práci na téma „Detekce cizích objektů v rentgenových snímcích hrudníku s využitím metod strojového učení“ jsem vypracovala samostatně pod vedením vedoucího diplomové práce a s použitím odborné literatury a dalších informačních zdrojů, které jsou všechny citovány v práci a uvedeny v seznamu literatury na konci práce.

Jako autorka uvedené diplomové práce dále prohlašuji, že v souvislosti s vytvořením této diplomové práce jsem neporušila autorská práva třetích osob, zejména jsem nezasáhla nedovoleným způsobem do cizích autorských práv osobnostních a/nebo majetkových a jsem si plně vědoma následků porušení ustanovení § 11 a následujících autorského zákona č. 121/2000 Sb., o právu autorském, o právech souvisejících s právem autorským a o změně některých zákonů (autorský zákon), ve znění pozdějších předpisů, včetně možných trestněprávních důsledků vyplývajících z ustanovení části druhé, hlavy VI. díl 4 Trestního zákoníku č. 40/2009 Sb.

Brno .....

.....

podpis autorky

## PODĚKOVÁNÍ

Ráda bych poděkovala vedoucímu své diplomové práce panu Ing. Jiřímu Chmelíkovi, Ph.D. za jeho rady, předání cenných zkušeností a veškeré připomínky, které mi pomohly dobrat se ke správnému řešení problémů. Také chci poděkovat za ochotu a jeho čas, který si na mě vždy udělal i přes jeho pracovní vytížení. Dále bych ráda poděkovala panu Ing. Ondřeji Pechovi, Ph.D., který mi poskytnul hardwarové vybavení, bez kterého bych se jen stěží dobrala takovýchto výsledků. Velice si vážím času, který mi věnoval před zprovozněním celé práce, a hlavně jeho zájmu. Také bych ráda poděkovala svému příteli, který ochotně odpovídal na moje dotazy, podporoval mě při potížích, ale také si vyslechl mé nápady a diskutoval nad nimi. A v neposlední řadě děkuji své rodině, která mě vždy podporovala ve studiu a nebylo tomu jinak ani nyní. Z celého srdce jim děkuji za jejich zájem, péči a hlavně trpělivost.

# Obsah

Úvod	9
<b>1 Teoretická část studentské práce</b>	<b>10</b>
1.1 Základní princip rentgenu	10
1.2 Anatomie a diagnostické zobrazování plic	11
1.3 Metody hlubokého učení používané pro detekci objektů v obrazech	11
1.4 Dostupné metody pro detekci cizích objektů v rentgenových snímcích hrudníku	13
1.4.1 Dostupné metody pro vyhledávání kruhových cizích objektů	13
1.4.2 Dostupné metody pro vyhledávání tvarově nespecifických cizích objektů	14
1.5 Regionální konvoluční neuronové sítě (R-CNN)	16
1.5.1 Architektura Faster R-CNN	16
1.5.2 Trénování Faster R-CNN	19
<b>2 Výsledky studentské práce</b>	<b>20</b>
2.1 Použitá data	20
2.1.1 Úprava dat	24
2.2 Výběr a návrh metody	27
2.2.1 Výběr předtrénované neuronové sítě jako extraktoru příznaků	27
2.2.2 Stanovení počtu a velikostí anchorů	29
2.2.3 Trénování Faster R-CNN	30
2.2.4 Vyhodnocení úspěšnosti detektorů	34
2.2.5 Nalezení nejlepšího natrénovaného detektoru	38
2.2.6 Možné zlepšení kvality detektoru	49
<b>Závěr</b>	<b>51</b>
<b>Literatura</b>	<b>52</b>
<b>Seznam symbolů, veličin a zkratk</b>	<b>60</b>
<b>A Tabulka výsledků</b>	<b>62</b>

# Seznam obrázků

1.1	Základní části RTG přístroje . . . . .	10
1.2	Zobrazení plic na RTG hrudníku . . . . .	12
1.3	Architektura Faster R-CNN . . . . .	17
2.1	Rentgenový snímek s anotovanými cizími objekty . . . . .	21
2.2	Ukázka variability dostupných obrazů . . . . .	22
2.3	Ukázka anotací cizích objektů . . . . .	23
2.4	Příklady dvou RGB vstupních obrázků . . . . .	25
2.5	Ukázka odebraných obrázků z trénovací skupiny . . . . .	27
2.6	Parametry předtrénovaných neuronových sítí . . . . .	28
2.7	K-means pro všechny velikosti anchorů v trénovací skupině . . . . .	29
2.8	Příklad špatného vyhodnocení úspěšnosti originální funkcí . . . . .	35
2.9	Rozdíl mezi metrikou IoU a Min pro překryv bounding boxů . . . . .	36
2.10	Zobrazení vyhodnocení TP, FP a FN . . . . .	37
2.11	Boxplot přesnosti, senzitivity a F1 skóre pro detektor s nejlepším F1 skóre . . . . .	39
2.12	Zobrazení snímků, na nichž jsou detekovány cizí objekty pomocí detektoru s nejvyšším F1 skóre. . . . .	41
2.13	Nejednotné anotování objektů mimo oblast plic. . . . .	42
2.14	Boxplot přesnosti, senzitivity a F1 skóre pro detektor s nejvyšším počtem TP výsledků . . . . .	44
2.15	Zobrazení snímků, na nichž jsou detekovány cizí objekty pomocí detektoru s nejvyšším počtem TP výsledků. . . . .	45
2.16	Boxplot přesnosti, senzitivity a F1 skóre pro detektor s nejnižším počtem FP výsledků . . . . .	47
2.17	Zobrazení snímků, na nichž jsou detekovány cizí objekty pomocí detektoru s nejnižším počtem FP výsledků. . . . .	48

# Seznam tabulek

2.1	Rozdělení dat do trénovací, validační a testovací množiny . . . . .	21
2.2	Průměrné IoU mezi definovanými anchory a všemi anotovanými objekty	31
2.3	Parametry trénování a vyhodnocení úspěšnosti detekce pro detektor s nejlepším F1 skóre. . . . .	39
2.4	Vyhodnocení úspěšnosti detekce pro detektor s nejlepším F1 skóre po úpravě anotací. . . . .	42
2.5	Parametry trénování a vyhodnocení úspěšnosti detekce pro detektor s nejvyšším počtem TP výsledků. . . . .	43
2.6	Parametry trénování a vyhodnocení úspěšnosti detekce pro detektor s nejnižším počtem FP výsledků. . . . .	47
A.1	Použité parametry pro trénování detektorů, práh pro detekci a výsledky	63

# Úvod

Detekce cizích objektů v RTG snímcích hrudníku je důležitá pro automatické zpracování RTG snímků, kdy se jedná především o kovové předměty umístěné na pacientově oblečení, např. knoflíky, šperky, zipy či dráty. Některé tyto předměty totiž mohou při automatickém zpracování snímku vyvolat falešně pozitivní výsledek onemocnění, což může způsobit značné komplikace. Z tohoto důvodu je třeba do postupu automatického zpracování snímků zařadit právě detekci cizích objektů a popřípadě nalezené oblasti cizích objektů z následné analýzy vynechat. Detekce cizích objektů může mít ovšem i lékařsky významnou hodnotu, kdy detekovaným objektem mohou být např. pacemakery, chirurgické svorky a další voperovaná zařízení.

Tato práce je členěna do dvou částí, kdy celá práce začíná teoretickým úvodem k danému tématu. Jsou zde krátce probrány základní principy RTG a anatomie plic. Dále jsou rozebrány metody hlubokého učení pro detekci objektů v obrazech následované rešerší dostupných metod pro detekci cizích objektů v RTG snímcích hrudníku. Teoretická část je ukončena podrobným rozebráním Faster R-CNN metody, která je následně využita v praktické části této práce.

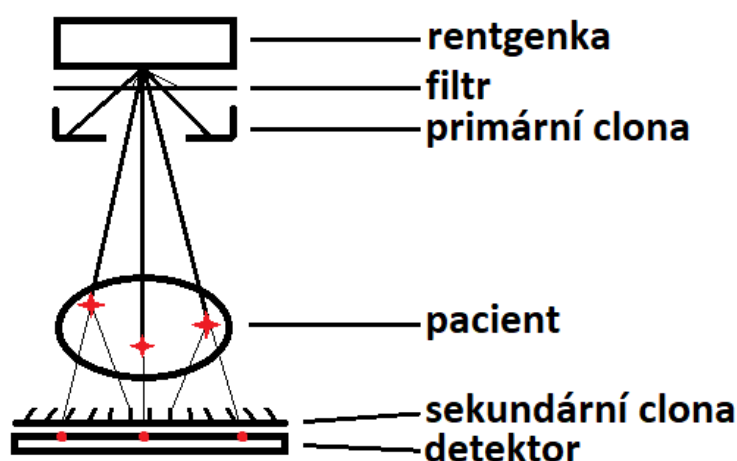
Druhá část této diplomové práce se zabývá praktickou aplikací detekce cizích objektů v RTG snímcích hrudníku. Nejprve jsou v úvodu popsána dostupná data a jejich úprava pro využití ve Faster R-CNN síti. Následuje popis samotné metody, která byla aplikována. Je zde popsán výběr předtrénované sítě jako extraktoru příznaků, dále výběr počtu a velikosti anchorů a jsou podrobněji popsány měněné parametry učení při trénování detektorů. Následuje vyhodnocení úspěšnosti natrénovaných detektorů a vybrané jsou poté rozebrány podrobněji. Na závěr je určen nejlepší detektor pro aplikaci detekce cizích objektů v RTG snímcích hrudníku a jsou naznačeny možnosti, kterými by se úspěšnost detektoru dala vylepšit.

# 1 Teoretická část studentské práce

## 1.1 Základní princip rentgenu

Rentgen (RTG) je diagnostická vyšetřovací metoda využívající ionizující rentgenové záření, tedy elektromagnetické vlny s vlnovou délkou  $10^{-8}$ – $10^{-12}$  m. Vyzářené RTG záření prochází přes pacienta, kde se v závislosti na absorpčním koeficientu tkáně částečně utlumí, a zbytek prochází až na detektor, kde se vytvoří rentgenový obraz. Výsledná intenzita v obraze tedy vzniká díky rozdílným absorpcím RTG záření ve tkáních, které jsou závislé hlavně na tloušťce a hustotě tkáně. Měkké tkáně, které mají menší hustotu, a tedy i nižší absorpci RTG záření, se ve výsledném obraze zobrazují tmavší a naopak hustší kosti, které mají naopak větší absorpci RTG záření, se zobrazují světlejší. Při vytváření stacionárních RTG snímků mluvíme o skiagrafii, pokud RTG obraz pozorujeme dynamicky a kontinuálně, mluvíme o skioskopii. [1, 2]

Rentgenový přístroj se skládá z několika základních částí, viz Obr. 1.1. Zdrojem RTG záření je rentgenka, klasická dioda, z jejíž žhavené katody jsou emitovány elektrony urychlené vysokým napětím směrem k wolframové anodě. Po dopadu elektronů na anodu se velká část jejich energie přemění na teplo a asi 1 % na RTG záření. Tento rentgenový svazek dále pokračuje přes hliníkové či měděné filtry, které mají za úkol odstranit měkké záření způsobující zbytečnou radiační zátěž pro pacienta bez diagnostického významu. Svazek je před dopadem na pacienta dále geometricky vymezen primární clonou. Sekundární clona umístěná za pacientem má za úkol odstranit nežádoucí rozptýlené záření zatímco svazek RTG záření procházející v původním směru zachovat. Využití této clony, která se skládá z olověných lamel umístěných ve směru původního svazku, zlepšuje kontrast v rentgenovém obraze. Na konci přístroje je umístěn detektor, který detekuje RTG záření. [1, 2]



Obr. 1.1: Základní části RTG přístroje

## 1.2 Anatomie a diagnostické zobrazování plic

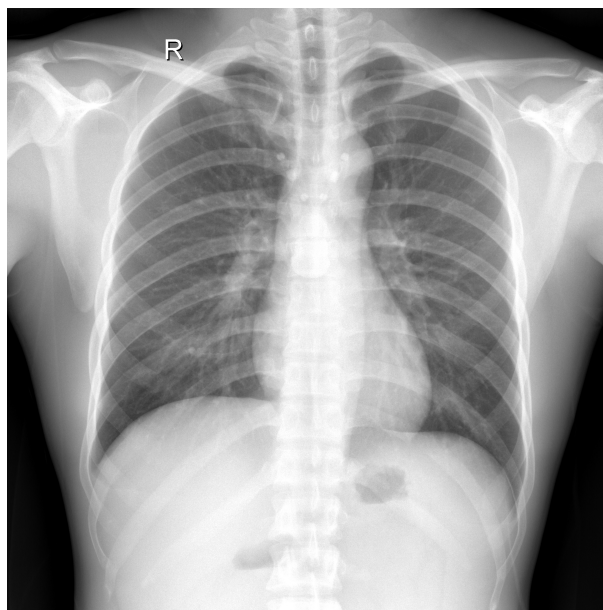
Plíce jsou párový orgán uložený v hrudní dutině, který slouží k zevnímu dýchání. Plíce mají narůžovělou barvu s černým mramorováním a přibližně kuželovitý tvar, kdy jejich baze je orientovaná směrem dolů a přiléhá k bránici, naopak plicní hrot je orientován směrem nahoru. Na mezihrudní ploše plíce orientované směrem do středohrudí se nachází plicní stopka, která umožňuje průchod průduškám, cévám a nervům. Plíce pokrývá blána zvaná poplicnice (*lat. pleura viscelaris*), ta přechází v oblasti plicní stopky na blánu zvanou pohrudnice (*lat. pleura parietalis*), která pokrývá vnitřní stěnu hrudníku. Mezi oběma blanami se nachází pohrudniční dutina (*lat. cavitas pleuralis*) vyplněná malým množstvím vazké tekutiny, která slouží ke snížení tření mezi pohrudnicí a poplicnicí, a umožňuje tak volný pohyb plic. V pohrudniční dutině je nižší tlak než atmosférický, což umožňuje roztažení plic při nádechu. Pravá plíce je větší, skládá se ze tří laloků, levá plíce má laloky pouze dva. Laloky se dále dělí na segmenty, ty na lalůčky. Do každého lalůčku vstupuje průdušinka, která se dělí na chodbičky končící plicními váčky vyklenujícími se v plicní sklípky, ve kterých probíhá samotná výměna plynů. [3, 4, 5]

Pro zobrazení plic se využívá hlavně metod RTG, CT (výpočetní tomografie, *angl. Computed Tomography*) a MRI (magnetická rezonance, *angl. Magnetic Resonance Imaging*), kdy každá z těchto metod má své výhody a nevýhody. Pro počáteční diagnostiku a obecný přehled se využívá RTG hrudníku, kdy se jedná zároveň o nej-používanější metodu pro zobrazování plic, viz. Obr. 1.2. Celkem 25 % z celkového množství RTG snímků jsou právě snímky hrudníku. Velikou výhodou této metody je její dostupnost, rychlost, nízká cena, ale i nízká radiační zátěž pro pacienta. Naopak nevýhodou je omezená rozlišovací schopnost a fakt, že se jedná o sumační snímání, tedy nemáme přehled o prostorovém rozložení. Snímek se provádí nejčastěji ve stoje, a to hned dvakrát, předozadní a zadopřední, kvůli lepší diagnostice. Další výhodou této metody ovšem je, že se dají snímat i pacienti uvázaní k lůžku vleže pomocí mobilních RTG přístrojů. Mezi nemoci, které se dají diagnostikovat pomocí RTG patří pneumonie, rakovina plic, chronická obstruktivní nemoc plic, pneumotorax, plicní výpotek, plicní noduly, infiltráty v plicní tkáni, občas fibrotizující projevy v plicích či kardiomegálie. [6, 7]

## 1.3 Metody hlubokého učení používané pro detekci objektů v obrazech

Detekce objektů spočívá nejen v samotné klasifikaci objektu, ale také v nalezení jeho polohy. Počátky detekce objektů se prováděly nejčastěji metodou posouvání





Obr. 1.2: Zobrazení plic na RTG hrudníku

okna po obraze a hledání objektu v tomto místě. Pokud bychom ale chtěli posouvat po obraze postupně okna všech velikostí a poměrů stran, bylo by to velice výpočetně a časově náročné a navíc by nám vzniklo velké množství redundantních oken. Proto se přistoupilo k anchorům, tedy oknům s předem definovanou velikostí a poměrem stran. V těchto oknech byly dále extrahovány ručně vybrané příznaky pro zařazení do jednotlivých skupin, kdy tato klasifikace byla prováděna metodou strojového učení, např. SVM (*Support Vector Machines*) [8], AdaBoost (*Adaptive Boosting*) [9] či DPM (*Deformable Parts Model*) [10]. Jelikož byl ovšem tento model neefektivní, náročný a příznaky musely být určeny ručně a byly mělké, začalo se následně využívat metod hlubokého učení, které jsou více robustní a extrahují komplexní příznaky. [11, 12]

V současné době se metody hlubokého učení pro detekci objektů dělí na 2 základní skupiny. Tou první jsou *region proposal-based* metody, kam spadá R-CNN (*Region Based Convolutional Neural Networks*) [13], Fast R-CNN [14], Faster R-CNN [15], Mask R-CNN [16], SPP-net (*Spatial Pyramid Pooling Net*) [17], R-FCN (*Region based Fully Convolutional Networks*) [18] a FPN (*Feature Pyramid Networks*) [19]. Tyto metody provádí nejprve skenování celého obrazu, kdy v něm vyhledají zájmové oblasti, na které se následně síť v dalším kroku zaměřuje. Skládají se tedy z několika částí, a to z extrakce příznaků pomocí konvoluční neuronové sítě (CNN), vyhledávání oblastí zájmu a klasifikace a regrese *bounding boxů*, kdy jsou všechny tyto části obvykle trénovány samostatně. Druhou skupinou jsou *classification-based* metody, kam spadá AttentionNet [20], G-CNN (*Grid Convolutional Neural Network*) [21], SSD (*Single Shot Detector*) [22], DSSD (*Deconvolutional Single Shot Detector*) [23],

YOLO (*You Look Only Once*) [24] a DSOD (*Deeply Supervised Object Detector*) [25]. Tyto metody detekují objekty pouze v jednom kroku, kdy z hodnot pixelů v obraze rovnou provádí regresi a klasifikaci *bounding* boxů, což může ušetřit výpočetní čas. Obě tyto skupiny spojuje používání anchorů, které byly představeny ve Faster R-CNN [15]. V současné době tyto metody pokročily tak, že je možné je využívat v aplikacích fungujících v reálném čase. [11, 12]

## 1.4 Dostupné metody pro detekci cizích objektů v rentgenových snímcích hrudníku

Dostupné metody pro detekci cizích objektů v rentgenových snímcích hrudníku byly pro účely této práce rozděleny podle tvaru vyhledávaného objektu na metody vyhledávající buď tvarově nespécifické objekty, nebo předem předdefinované tvary (nejčastěji kruhové objekty). Metody by se ovšem daly také rozdělit na metody hledající objekty pouze v oblasti plic, k čemuž často využívají i předešlou segmentaci plic, a na metody hledající objekty v celém rentgenovém snímku.

### 1.4.1 Dostupné metody pro vyhledávání kruhových cizích objektů

Článek [26] vyhledává kulaté objekty pouze v oblasti plic, které byly předem segmentovány pomocí anatomických atlasů s nerigidní registrací [27]. Na začátku jsou obrazy normalizovány a upraveny pro zvýraznění cizích objektů. Na takto upravené obrazy je poté aplikována hranová detekce, kdy bylo využito pro porovnání čtyř různých hranových operátorů, a to konkrétně Cannyho, Prewittové, Sobelova a Robertsova operátoru. Následně byly využity morfologické operace, nejprve dilatace lineárním elementem, po které následovalo vyplnění děr, a poté eroze strukturním elementem čtvercového tvaru. Na takto upravené obrazy byla aplikována CHT (*Circular Hough Transformation*) a byly nalezeny cizí objekty. Algoritmus byl testován na 50 snímcích s celkem 32 knoflíky. Nejlepšího výsledku dosáhl algoritmus s použitím hranového detektoru Prewittové, a to 100% přesnosti (*angl. precision*) a 100% senzitivity. Stejní autoři následně provedli podrobnější testování v článku [28]. Byl využit stejný algoritmus pro segmentaci plic a následné vyhledávání kruhových cizích objektů v oblasti plic. Na testování bylo použito 400 RTG snímků hrudníku a byla na ně aplikována 10násobná křížová validace. Nejlepšího výsledku opět dosáhl algoritmus s Prewittové hranovým detektorem, a to 96% přesnosti (*angl. precision*), 90% senzitivity a F1 skóre 92 %.

Další dostupná metoda, která se zaměřuje na vyhledávání knoflíků v oblasti plic, je popsána v článku [29]. Metoda se skládá ze čtyř hlavních kroků. Prvním je normalizace intenzity snímků, případně potřebná inverze rentgenových snímků. Následuje úprava snímků s nízkým kontrastem, které se vyhledávají podle prahové hodnoty střední hodnoty a rozptylu intenzit pixelů. Třetím krokem je segmentace plic s využitím algoritmu použitého i v předchozích metodách, tedy pomocí anatomických atlasů s nerigidní registrací [27]. Posledním krokem je nalezení knoflíků v oblastech plic, což bylo provedeno pomocí dvou algoritmů, a to opět CHT (*Circular Hough Transformation*) a Viola-Jones algoritmu. Pro CHT dosáhly výsledky 83% přesnosti (*angl. precision*) a 85% senzitivity, oproti tomu Viola-Jones algoritmus dosáhl 86% přesnosti (*angl. precision*) a 94% senzitivity, jeho nevýhodou je ale nutnost trénování příznaků. Testování bylo provedeno na vybraných 278 snímcích s 505 knoflíky z datasetu *Dataset National Library of Medicine (NLM)* [30].

V dalším článku [31] je popsána detekce kruhových cizích objektů v celém RTG snímku pomocí Faster R-CNN. Pro extrakci příznaků je na vstup umístěna předtrénovaná CNN, následně RPN určí konkrétní zájmové oblasti, ROI (*Region of Interest*) pooling vrstva nalezne požadované příznaky v prostoru a R-CNN nalezne konkrétní objekty. Byly použity dvě konvoluční vrstvy, ReLU (*Rectified Linear Unit*) a Pooling vrstvy, pět FC (*Fully Connected*) vrstev, opět ReLU následovaná FC a na konci Softmax vrstva řadící do tříd. Pro optimalizaci byl využit SGDM (*Stochastic Gradient Descent with Momentum*) algoritmus. Po natrénování na 400 rentgenových snímcích s 325 objekty v plicích bylo dosaženo 97% přesnosti (*angl. precision*), 90% senzitivity a F1 skóre 93 %.

### 1.4.2 Dostupné metody pro vyhledávání tvarově nespecifických cizích objektů

V článku [32] autoři vyhledávali cizí objekty s minimální velikostí 120 pixelů pouze v oblasti plic. Nejprve byla provedena segmentace plic pomocí algoritmu k-nejbližších sousedů, který vyhledává kontury plic, kdy jsou následně vyhledané kontury uzavřeny a dvě největší jsou určeny jako plíce. V dalším kroku byly extrahovány příznaky pro následnou detekci objektů. Jednalo se celkem o 61 příznaků, kdy některé byly získány pomocí Gaussovských filtrů 0., 1. a 2. řádu o různých velikostech, dále se jednalo o upravená vlastní čísla Hessiánu a  $x$  a  $y$  souřadnici. Pro detekci bylo využito několik metod, konkrétně se jednalo o LDA (*Linear Discriminant Analysis*), klasifikaci nejbližších průměrů, algoritmus k-nejbližších sousedů, GentleBoost [33] a SVM. Na konec byla provedena segmentace a odstranění objektů. Metoda byla natrénována na celkem 59 887 pixelech a otestována na 108 052 pixelech, které byly snímány na pacientech se zvýšeným rizikem tuberkulózy. Nejlepších výsledků dosáhl

algoritmus k-nejbližších sousedů s  $k = 19$ , který detekoval 95,6 % objektů a dosáhl 65,0% senzitivity a 99,9% specificity.

Ve článku [34] je popsána metoda zabývající se detekcí cizích objektů v rentgenových snímcích a jejich klasifikací do šesti skupin, konkrétně trubičky a dráty, pacemakery, implantáty, malé cizí objekty, šperky a knoflíky. Byla využita předtřénovaná konvoluční neuronová síť DenseNet [35] a ResNet50 [36] (nejlepších výsledků dosáhl DenseNet161) na datech ImageNet [37] a pro lepší výsledky ještě na *NIH Database ChestX-ray14* [38] a dále byla použita 5násobná křížová validace na 4 704 trénovacích a 1 357 testovacích snímcích z MIMIC CXR [39] datasetu. Data byla normalizována, byla provedena augmentace a jako učící algoritmus byl využit SGD (*Stochastic Gradient Descent*). Nejprve byla síť nastavena pouze pro binární klasifikaci, zda ve snímku je či není cizí objekt, a následně byla síť přepracována na *one-versus-all multilabel* klasifikaci, kdy síť zařazovala nalezené objekty do výše zmiňovaných skupin. Tyto sítě dosáhly AUC (*Area under the ROC Curve*) 0,984 pro binární klasifikaci a 0,969 pro *multilabel* klasifikaci na testovacích datech.

Cizí objekty v rentgenových snímcích pořízených kamerou smartphonu jsou detekovány ve článku [40]. Bylo využito celkem 10 507 snímků z databáze cheXphoto [41], která obsahuje snímky s plicními patologiemi. Detekce se zaměřovala především na objekty v plicích, ale některé falešně pozitivní objekty byly nalezeny i mimo ně. Rentgenové snímky byly normalizovány a detekce objektů probíhala pomocí YOLOv4-based [42] modelu využívající *Generalized Intersection over Union* kritériální funkci, která hledá obdélníkové ohraničující rámečky. Výsledek byl dále vložen do segmentačního modelu U-Net [43] a nalezené objekty byly odstraněny. Detekce objektů dosáhla 76% přesnosti (*angl. precision*), 70% senzitivity a F1 skóre 73 %.

Patologické snímky byly také využity ve článku [44], který popisuje vytvoření 16 různých předtřénovaných sítí, kdy 14 z nich klasifikuje snímky na jednotlivé patologie a další dvě určují, zda je ve snímku detekován cizí objekt a zda je tento objekt medicínská pomůcka. Všechny výstupy sítí pak vstupují do poslední sítě XGBoost (*eXtreme Gradient Boosting*) [45], která klasifikuje snímek jako patologický či fyziologický. Vstupní data byla normalizována a byla na nich provedena augmentace. Obě sítě pro detekci cizích objektů byly Inception-v3 [46] s využitím optimalizačního algoritmu ADAM (*Adaptive Moment Estimation*) [47]. Výsledky dosáhly 96,9% přesnosti (*angl. precision*) a 94,9% senzitivity pro klasifikaci výskytu cizího objektu a 95,7% přesnosti (*angl. precision*) a 92,8% senzitivity pro klasifikaci na medicínské či jiné objekty.

## 1.5 Regionální konvoluční neuronové sítě (R-CNN)

Regionální konvoluční neuronová síť (R-CNN) slouží ke klasifikaci výřezů (regionů) v rámci daného snímku. První byl v roce 2014 představen koncept R-CNN v článku [13], který využívá metodu *Selective Search* pro nalezení oblastí zájmu. V roce 2015 bylo uvedeno Fast R-CNN v článku [14], kde byla představena technika *ROI Pooling* pro extrakci příznaků, která udělala model přibližně  $25\times$  rychlejší, ale stále využívala pro určení zájmových oblastí *Selective Search*. Na konci roku 2015 pak bylo uvedeno Faster R-CNN v článku [15], které již využívá RPN (*Region Proposal Network*) pro určení zájmových oblastí, čímž se celý algoritmus ještě přibližně  $10\times$  urychlil. Faster R-CNN je navíc jako první plně diferencovatelný. [48, 49]

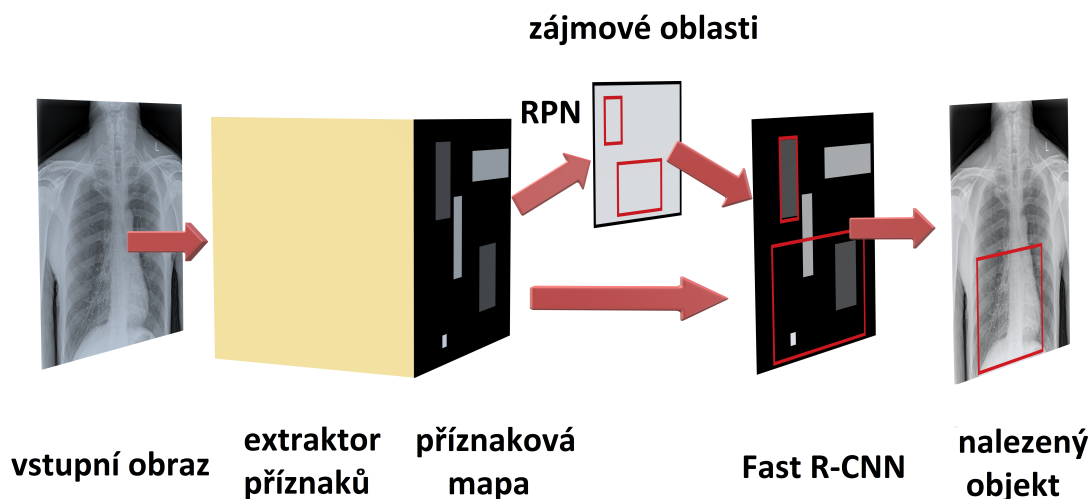
### 1.5.1 Architektura Faster R-CNN

Detektor Faster R-CNN se skládá z několika částí, viz. Obr. 1.3. Jako první je vstupní obraz vložen na začátek předtrénované sítě, která na výstupu vytvoří příznakovou mapu, jedná se tedy o extraktor příznaků pro vstupní obraz složený z konvolučních vrstev. V původním článku byly použity předtrénované sítě VGG-16 [50] a ZF-net [51], dají se ale využít i jiné architektury, kdy se v současné době nejvíce využívá síť ResNet [48]. Následují další dva moduly. Tím první je RPN, který pomocí příznakové mapy obrazu najde definované množství oblastí zájmu, ve kterých se nachází objekty. Druhým modulem je Fast R-CNN, který na začátku z příznakové mapy obrazu a nalezených oblastí zájmu nejprve pomocí *RoI Pooling* nalezne nové příznakové mapy pro oblasti. Dalším krokem je pak pomocí těchto nových map klasifikovat jednotlivé oblasti do tříd a vylepšit pozice ohraničených oblastí, což se provádí pomocí R-CNN. Zjednodušeně se tedy dá říct, že RPN říká Fast R-CNN, kam se má podívat. [15, 48, 49]

#### Region Proposal Networks

RPN (*Region Proposal Networks*) je část Faster R-CNN, která vytváří oblasti zájmu v obraze. Tyto oblasti zájmu mohou být různě veliké a mít různý poměr stran, ale vždy mají obdélníkový tvar. Nejlepším řešením určení těchto oblastí zájmu je definovat je pomocí  $x$  a  $y$  souřadnice středu a dále pomocí šířky a výšky oblasti. [15, 48]

Pro účely Faster R-CNN se využívá anchorů, jedná se o sadu přesně definovaných rámečků s různými velikostmi a poměry stran, které jsou definovány pro každý bod příznakové mapy obrazu. Je nutné si ovšem uvědomit, že extraktor příznaků je složen z konvolučních a *pooling* vrstev, tedy velikost příznakové mapy je oproti vstupnímu obrazu menší. Když tedy definujeme anchory pro každý bod příznakové



Obr. 1.3: Architektura Faster R-CNN

mapy, anchors označují oblasti v originálním vstupním obraze. V původním článku byly anchors vytvořeny pomocí kombinace tří velikostí a tří poměrů stran, čímž vzniklo celkem  $k=9$  anchorů. Pokud tedy máme příznakovou mapu o velikosti  $W \times H$  a  $k$  anchorů, vznikne nám postupným posouváním pozice celkem  $W \times H \times k$  anchorů. [15, 48, 52]

Úkolem RPN je na základě příznakové mapy obrazu vytvořit množinu obdélníkových oblastí zájmu. Na začátku se provádí konvoluce mezi příznakovou mapou a konvoluční vrstvou s  $3 \times 3$  kernelem a 512 kanály. Takto vytvořené příznaky dále vstupují paralelně do dvou konvolučních vrstev s kernelem  $1 \times 1$ . První  $1 \times 1$  konvoluční vrstva má  $2 \times k$  kanálů a určuje pro každý bod příznakové mapy pravděpodobnost, zda anchors v daném bodě obsahují či neobsahují objekt. Jedná se tedy o klasifikátor. Druhá  $1 \times 1$  konvoluční vrstva má  $4 \times k$  kanálů a určuje pro každý bod příznakové mapy čtyři koeficienty umístění pro každý anchor, tedy  $x$  a  $y$  souřadnici středu, výšku a šířku. Jedná se o regresor. [15, 48, 52]

Dále RPN během fáze trénování vytvoří dva druhy predikce, a to binární klasifikaci a vylepšení pozice anchoru. Nejprve RPN tedy nalezne anchors, které se buď překrývají s anotovaným objektem o více než IoU (*Intersection over Union*, viz. vzorec 2.1) 0,7, nebo mají pro daný anotovaný objekt největší IoU, a označí je jako pozitivní. Dále nalezne anchors, které mají IoU s jakýmkoliv anotovaným objektem menší než 0,3 a zařadí je do skupiny negativní. Ostatní anchors se nepoužívají. Z takto vytvořené binární klasifikace se vytvoří mini-batch z anchorů v jednom obraze, který obsahuje 128 pozitivních a 128 negativních anchorů, a aplikuje se na

něj binární vzájemná entropie, tedy vypočte se chyba klasifikace. Dále se vezmou jen pozitivní anchory a vypočte se *Smooth L1* chyba pro regresi, čímž se zároveň posunou anchory směrem k pozici anotovaných objektů.[15, 48, 49, 52]

Během fáze testování je po  $1 \times 1$  konvolučních vrstvách využit postprocessing. Nejprve se regresní koeficienty aplikují do anchorů pro zjištění přesné pozice rámečku zájmové oblasti. Poté se aplikuje NMS (*non-maximum suppression*), který zajistí, aby se nalezené oblasti zájmu nepřekrývaly. Algoritmus funguje iterativně, vybere oblast s nejvyšším skóre a odstraní všechny oblasti, které se překrývají více než je daný IoU práh. Takto algoritmus postupuje, dokud se nevyskytuje žádná oblast s menším skóre. [48, 49, 52]

### Fast R-CNN detektor

Po aplikaci RPN přichází na řadu Fast R-CNN. Na její vstup je dán výstup z extraktoru příznaků, tedy příznaková mapa, která vstupuje i do RPN, jsou mezi nimi tedy sdílené váhy. Dále jsou na její vstup dány i pozice zájmových oblastí vypočtených pomocí RPN. Úkolem Fast R-CNN je vzít zájmové oblasti a určit jejich klasifikaci k jedné ze tříd či k pozadí a následně upravit souřadnice zájmové oblasti. [48, 52]

Prvním krokem je *RoI pooling* vrstva. Na její vstup tedy vstupují příznaková mapa obrazu a zájmové oblasti v obraze. *RoI pooling* vrstva vezme oblasti zájmu a jim odpovídající část příznakové mapy, kterou pomocí např. bilineární interpolace převzorkuje na velikost  $14 \times 14 \times 512$ . Dále následuje *max pooling* vrstva s kernelem  $2 \times 2$ , která příznakovou mapu zmenší na velikost  $7 \times 7 \times 512$  pro každou zájmovou oblast, což je potřebná velikost na vstup následujícího kroku, tedy R-CNN. [15, 48, 52]

Dalším krokem je samotné R-CNN. Pro každou oblast zájmu je její příznaková mapa nejprve zploštěna a následně jsou aplikovány dvě FC vrstvy s ReLU aktivací. Následují opět dvě paralelní větve podobně jako u RPN, tedy klasifikátor a regresor. U klasifikátoru je použita plně propojená síť s  $N+1$  neurony, kde  $N$  je počet tříd a jedna navíc je pro pozadí, a následně *soft-max* vrstva pro získání pravděpodobností zařazení pro každou třídu. U regresoru je použita FC vrstva s  $4 \times N$  neurony, tedy pro každou třídu je určena úprava velikosti a umístění zájmové oblasti zvlášť. [48, 52]

Během trénování se provádí podobné úkony jako během trénování RPN, ale musíme počítat s různými třídami. Nejprve vytvoříme skupinu objektů s třídami ze zájmových oblastí, které mají IoU s některým anotovaným objektem větší než 0,5. Zájmové oblasti s IoU mezi 0,1 a 0,5 s některým z objektů jsou zařazeny do skupiny pozadí. Mini-batch se vytvoří z 25 % vzorky s udanou třídou a ze 75 % ze vzorků pozadí. Pro takovýto mini-batch se vypočte tentokrát *multiclass* vzájemná entropie

jako chyba klasifikátoru a následně pouze z 25 % zařazených vzorků k třídám se vypočte *Smooth L1* chyba pro chybu regresoru. [48]

Při fázi testování opět provedeme podobný postup jako při trénování RPN. Nejprve rozhodneme podle největší pravděpodobnosti, ke které třídě oblast přiřadíme, a následně vyřadíme všechny oblasti, které byly při přiřazování klasifikovány jako pozadí, a aplikujeme *class-based* NMS. V prvním kroku tedy rozřadíme oblasti podle tříd, seřadíme je podle skóre a poté aplikujeme samotné NMS na každou třídu zvlášť. [48]

## 1.5.2 Trénování Faster R-CNN

Trénování Faster R-CNN probíhá ve čtyřech krocích:

1. V prvním kroku je nezávisle trénováno RPN podle popisu v kapitole 1.5.1. Extraktor příznaků má na začátku nastavené váhy podle předtrénované sítě na ImageNet datasetu, následně jsou upraveny pro úkol vyhledávání zájmových oblastí.
2. V druhém kroku je provedeno trénování Fast R-CNN, které je popsáno v kapitole 1.5.1. Váhy RPN jsou zachovány a výsledné zájmové oblasti vypočtené pomocí RPN jsou používány k trénování Fast R-CNN. Váhy extraktoru jsou jako v předchozím kroku inicializovány podle předtrénované sítě na ImageNet datasetu, tedy v této chvíli ještě nedochází ke sdílení vah z extraktoru příznaků mezi RPN a Fast R-CNN.
3. Ve třetím kroku dochází k dotrénování RPN, který má na počátku stejné váhy jako na konci prvního kroku. Váhy Fast R-CNN jsou zachovány. Zároveň dochází k zachování vah u extraktoru příznaků z konce kroku dva, v tuto chvíli tedy již RPN a Fast R-CNN sdílí váhy konvolučních vrstev u extraktoru příznaků.
4. V posledním kroku dochází k dotrénování Fast R-CNN. Inicializace vah je stejná jako na konci druhého kroku, naopak váhy RPN i extraktoru příznaků jsou zachovány konstantní.

[15, 48]



## 2 Výsledky studentské práce

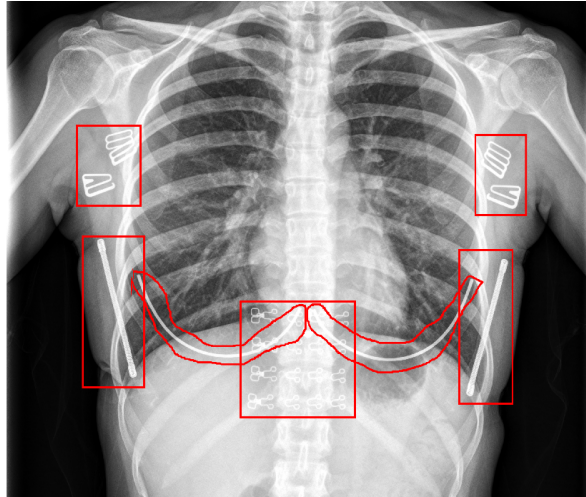
V této kapitole bude probráno konkrétní řešení detekce cizích objektů v rentgenových snímcích hrudníku. Na základě prostudovaných článků s dostupnými metodami pro detekci těchto cizích objektů v rentgenových snímcích hrudníku bylo rozhodnuto, že tato diplomová práce bude využívat metodu regionálních konvolučních sítí, konkrétně tedy metodu Faster R-CNN. V této kapitole budou nejprve popsána dostupná data a jejich úprava pro využití pro tuto konkrétní implementaci. Následně budou rozebrány vyzkoušené parametry pro trénování sítě Faster R-CNN a úspěšnost detekce pro jednotlivé natrénované sítě. V závěru budou vybrané detektory podrobněji prozkoumány a budou probrána možná zlepšení.

Pro účely této diplomové práce byl jako programové prostředí vybrán MATLAB R2020a. Z důvodu vysoké výpočetní náročnosti bylo nutné výpočty provádět na počítači s výkonnou GPU. Pro tyto účely byl používán počítač se specifikacemi: 2× osmijádrový (16 vláken) procesor Intel Xeon E5-2650 se základní frekvencí 2,0 GHz, 128 GB operační paměti, typ disku SSD, grafická karta Tesla K40c s 12 GB video-paměti typu GDDR5, operační systém Windows 10 Pro Education.

### 2.1 Použitá data

Použitá data byla stažena na stránce soutěže *object-CXR* [53]. Data obsahují celkem 10 000 rentgenových snímků hrudníku, kdy polovina z nich obsahuje jeden či více cizích objektů v oblasti plic a druhá polovina žádný cizí objekt v této oblasti neobsahuje. Tyto objekty byly anotovány 12 zkušenými radiology. Anotované objekty se dělí do tří skupin podle jejich tvaru, a to na obdélníkové, eliptické a polygonální. U obdélníkových a eliptických tvarů jsou anotace určeny hodnotami  $x$  a  $y$  souřadnicemi pro levý horní a pravý spodní roh (*typ*,  $x1$ ,  $y1$ ,  $x2$ ,  $y2$ ). U polygonálního tvaru jsou postupně zapsány dvojice  $x$  a  $y$  souřadnic pro každý bod (*typ*,  $x1$ ,  $y1$ , ...,  $xn$ ,  $yn$ ). Na začátku každého vektoru bodů je také zaznamenán typ tvaru, tedy 0 pro obdélník, 1 pro elipsu a 2 pro polygon. Příklad snímku s anotovanými cizími objekty je zobrazen na Obr. 2.1.

Data jsou dále rozdělena do tří skupin. První skupina trénovacích dat obsahuje 8 000 rentgenových snímků, validační skupina pak 1 000 snímků a testovací skupina pak zbývajících 1 000 snímků. V každé z těchto skupin polovina snímků obsahuje cizí objekty a polovina je neobsahuje. Rozdělení dat do jednotlivých skupin je vidět v Tab. 2.1. Pro účely této diplomové práce budou využita data ze skupiny trénovací pro trénování sítě Faster R-CNN a následně budou využita data ze skupiny testovací pro otestování kvality detektoru. Validací data nebylo možné v programovacím prostředí pro validaci během trénování využít. Bylo uvažováno, zda tato



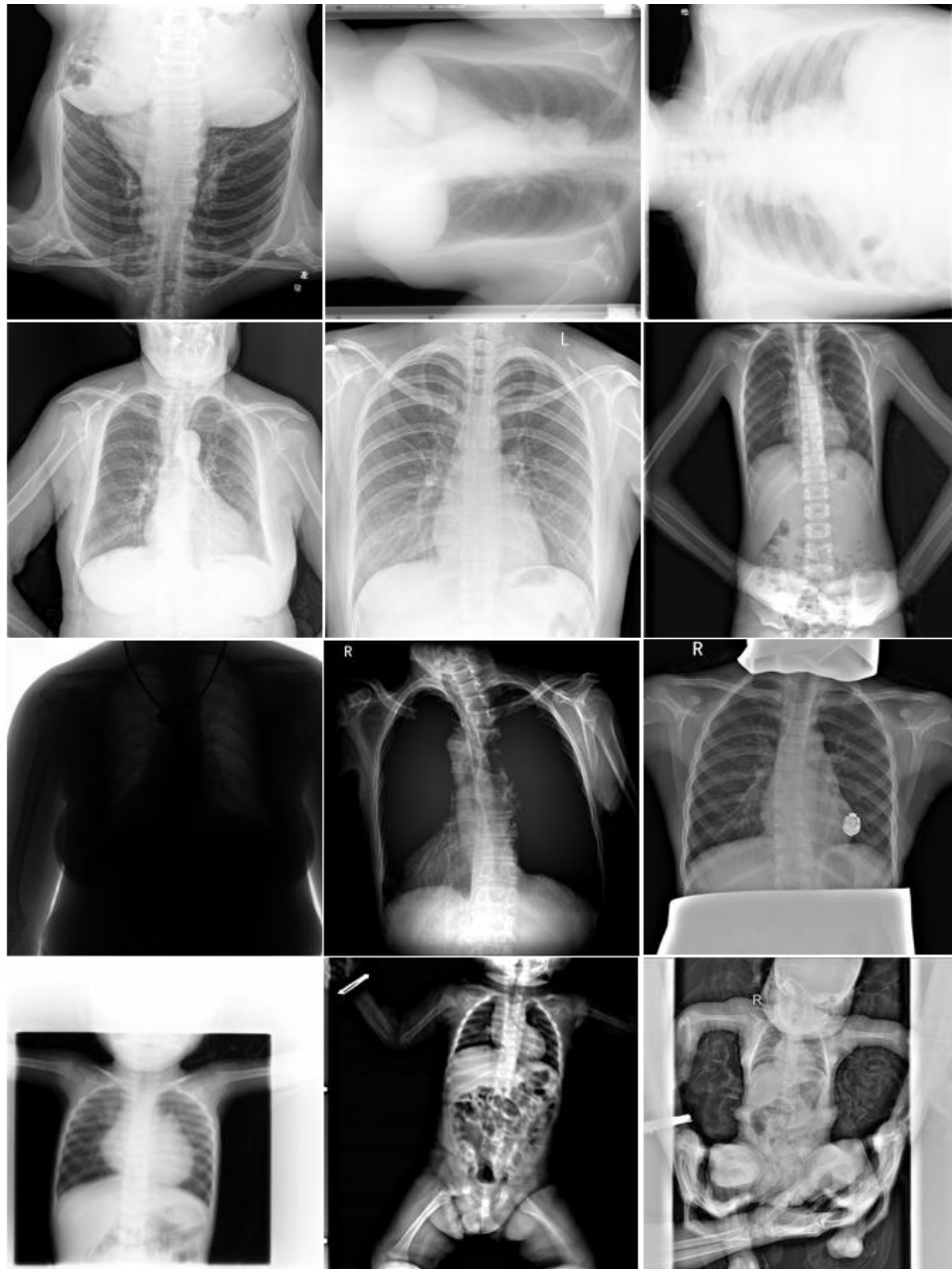
Obr. 2.1: Rentgenový snímek s anotovanými cizími objekty

validační data nezařadit do jedné ze skupin trénovacích či testovacích dat, nakonec bylo ovšem rozhodnuto, že validační skupina dat nebude využita, aby byly výsledky porovnatelné s výsledky v zadané soutěži *object-CXR* [53], ze které pochází použitá data v této diplomové práci.

Tab. 2.1: Rozdělení dat do trénovací, validační a testovací množiny

	S cizími objekty	Bez cizích objektů
Trénovací	4 000	4 000
Validační	500	500
Testovací	500	500

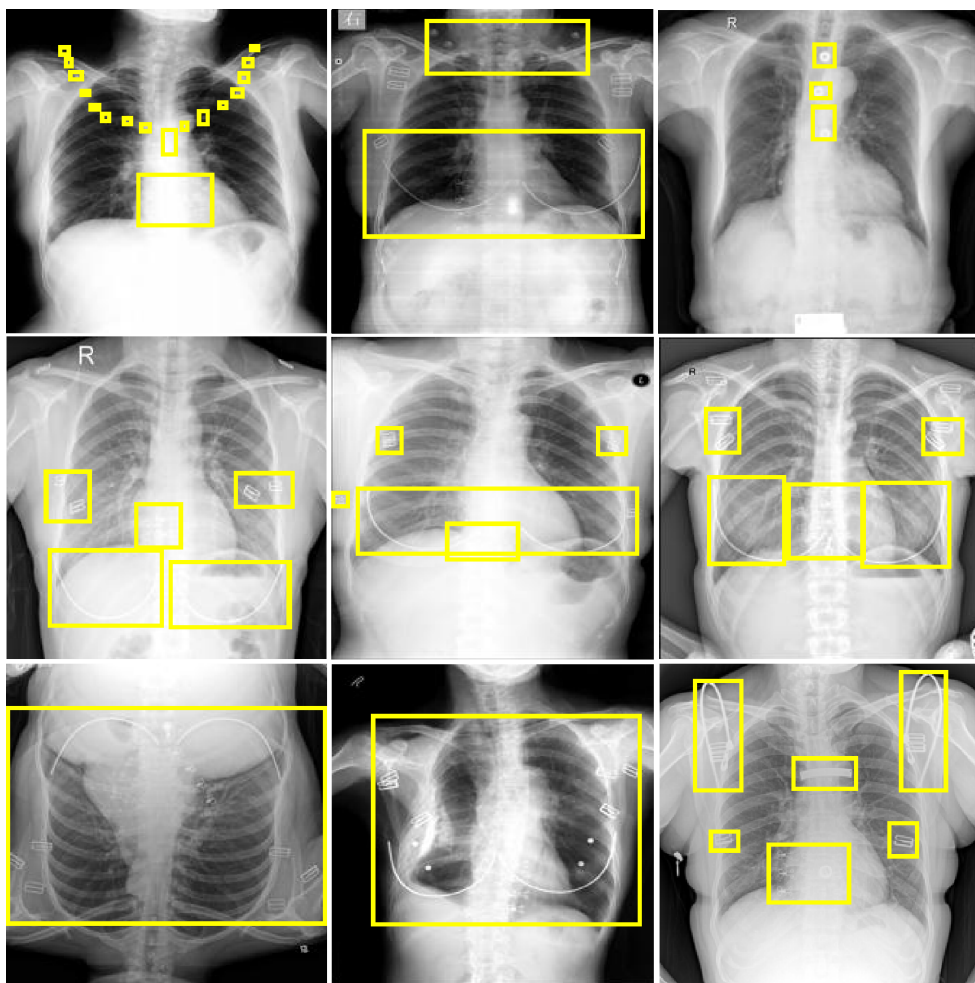
Dostupná data projevují vysokou variabilitu, na Obr. 2.2 lze vidět několik ukávek po převzorkování snímků na stejnou velikost  $224 \times 224$  pixelů. RTG snímky jsou otočeny různými směry, některé jsou podexponované a některé naopak přexponované, oblast snímání je různorodá, někdy jsou snímány pouze plíce a jejich těsné okolí, na některých snímcích byla snímána větší část těla, ba dokonce u dětí celé tělo. Některé snímky jsou negativy nebo se na nich vyskytují různé radiologické ochranné pomůcky. Škála pacientů je také široká, snímkováni byli muži i ženy, ale dokonce i děti různého věku, na některých snímcích jsou vidět i rodiče přidržující dětské pacienty. Z důvodu této vysoké variability a relativně velkého množství dostupných dat nebude v této diplomové práci prováděna augmentace dat. Velké rozdíly jsou také ve velikosti jednotlivých snímků, kdy nejmenší snímek má velikost  $248 \times 300$  pixelů a naopak největší  $4\,096 \times 4\,932$  pixelů.



Obr. 2.2: Ukázka variability dostupných obrazů: obrázky jsou různě natočené, je snímána různá oblast těla, kontrast v obrazech není vždy ideální a pacienti jsou vybráni z různých skupin

Velká variabilita se také projevuje v anotacích cizích objektů, což lze pozorovat na Obr. 2.3. Velice podobné objekty jsou anotovány různými způsoby, na Obr. 2.3 jsou např. šperky na krku na jednom snímku anotovány jako jeden velký objekt, kdežto na jiném obrázku jsou označeny jako mnoho menších objektů. V popisu dat je také napsáno, že anotované objekty se vyskytují pouze v oblasti plic, ale z ukázek je zřejmé, že někdy jsou anotované objekty i mimo oblast plic. Na některých

obrázcích se dá vypořádat, že pokud se vyskytuje více objektů ve stejné oblasti, je někdy anotován objekt pouze jeden, ale v některých případech je anotováno více překrývajících se objektů. Asi největší rozdíl v anotovaných cizích objektech je vidět na snímcích dole, kdy byly všechny objekty ze snímku označeny pouze jako jeden velký objekt, který pokrývá téměř celý RTG snímek. Po převzorkování obrazů na stejnou velikost  $224 \times 224$  pixelů, které je následně popsáno v kapitole 2.1.1, je velikost nejmenšího anotovaného objektu  $4 \times 4$  pixely a naopak největšího  $223 \times 222$  pixelů. Toto nejednotné anotování cizích objektů by mohlo během trénování neuronových sítí působit značné problémy, pokud by tedy úspěšnost natrénované sítě nebyla dostatečná, měla by se provést podrobnější kontrola anotací cizích objektů, sjednotit pravidla pro jejich vytváření a upravit nevyhovující anotace.



Obr. 2.3: Ukázka anotací cizích objektů: Podobné objekty jsou anotovány jako celek nebo více menších objektů, překrývající se objekty jsou anotovány jedním či každý svým rámečkem, v některých případech je vytvořen pouze jeden anotovaný objekt pro všechny objekty, který pokrývá téměř celý obraz

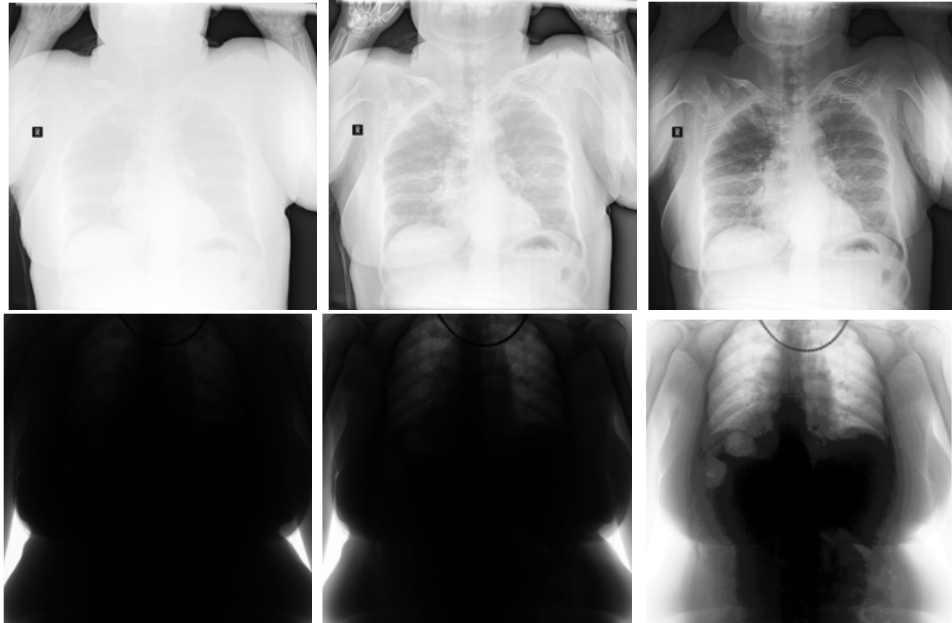
### 2.1.1 Úprava dat

Anotovaná data bylo nejprve potřeba vložit do tabulky, která se vyžaduje na vstup funkcí. Tabulka obsahuje dva a více sloupců, kdy v prvním sloupci obsahuje cestu k obrázku včetně názvu. Druhý sloupec obsahuje anotované objekty v níže popsaném tvaru  $(x, y, šířka, výška)$ . Pokud je v obraze více objektů, budou zapsány v řádcích pod sebou a následně umístěny opět do druhého sloupce tabulky. Třetí sloupec obsahuje dobrovolně typ objektu, čehož se využívá při trénování detektoru, který odlišuje několik druhů objektů. V této práci byla vytvořena tabulka o dvou sloupcích, kdy první obsahuje cestu k obrázku a druhý anotované objekty v obrázku ve správném tvaru. Vytvoření této matice, úprava anotací objektu a vytvoření vstupních dat, které je popsáno v následujících kapitolách, se provádí se skriptech *GTtrain.m* pro trénovací data a *GTtest.m* pro testovací data.

#### Úprava obrazů

Jelikož dostupná data jsou šedotónová, ale na vstup předtrénovaných CNN je možné dodat RGB obrazy, bylo této skutečnosti využito pro vytvoření vstupních dat. Do první vrstvy matice byl tedy vložen originální šedotónový snímek a do dalších dvou vrstev byly vloženy snímky s různou úpravou kontrastu. Jelikož jsou některé snímky přexponované a některé naopak podexponované, a tedy cizí objekty jsou na nich špatně zřetelné, jako nejlepší řešení se zdálo využít takovou úpravu obrazů, která by pro oba dva případy zlepšila kontrast a zvýraznila cizí objekty. První úpravou obrazu je aplikace CLAHE (*Contrast Limited Adaptive Histogram Equalization*) [54] s *clip* limitem 0,01 (při možném rozmezí 0–1) a s rozdělením obrazu na  $8 \times 8$  regionů. Takto upravený obraz je následně umístěn do druhé vrstvy RGB obrazu. Do třetí vrstvy je umístěn obraz po globální ekvalizaci histogramu s 64 stupni šedi. Jednotlivé vrstvy výsledného RGB obrazu lze vidět na Obr. 2.4.

Takto připravené obrazy byly dále upraveny na správnou velikost dle vstupní vrstvy předtrénované CNN. Jak již bylo řečeno v kapitole 2.1, použitá data mají rozmanitou velikost, a to od  $248 \times 300$  až do  $4\,096 \times 4\,932$  pixelů. U takto velkých obrazů se dá postupovat následujícími způsoby, zaprvé se dají všechny zmenšit na shodnou velikost, kdy ovšem hrozí ztráta informace. Druhým způsobem je takto velké obrazy rozdělit na několik menších o požadované velikosti, kdy se tyto jednotlivé části budou používat pro trénování sítě zvlášť jako samostatné obrazy. Z důvodu snahy vytvořit co nejjednodušší model a ten v případě malé úspěšnosti dále rozšiřovat, byl do této diplomové práce vybrán první způsob zpracování dat, tedy převzorkovat všechny obrazy na stejnou velikost. Jelikož je v této diplomové práci využito sítě ResNet50 jako extraktoru příznaků, vstupní obrazy tedy musely být převzorkovány na velikost  $224 \times 224 \times 3$  dle požadavků této sítě. Úprava velikosti obrazu



Obr. 2.4: Příklady dvou RGB vstupních obrázků. Vlevo první vrstva – originální obraz, uprostřed druhá vrstva – obrázek po úpravě pomocí CLAHE, vpravo třetí vrstva - obrázek po ekvalizaci histogramu.

byla provedena pomocí bikubické interpolace, tedy pomocí váhovaného průměru  $4 \times 4$  okolí. Při použití této metody může ovšem někdy dojít k tomu, že výsledná hodnota pixelu bude mimo původní rozsah hodnot, je tedy nutné následně přepočíst výsledné hodnoty do původního rozmezí. Posledním krokem při úpravě obrazů v této diplomové práci je převedení obrazů z původního rozmezí 0–255 do rozmezí 0–1, což je požadováno na vstup předtrénovaných sítí. Zároveň je se zmenšením obrazů také potřeba přepočítat umístění anotovaných objektů tak, aby odpovídaly umístění objektů v nově zmenšených obrazech. Tento postup je popsán v následující kapitole 2.1.1.

### Úprava anotací objektů

Před použitím anotací objektů bylo nejprve třeba úprav pro správné fungování. Původní data jsou anotována v kartézském souřadném systému s počátkem v levém horním rohu, což je systém i většiny programovacích prostředí. Jak již bylo řečeno v kapitole 2.1, stažená data obsahují anotace obdélníkového, eliptického a polygonálního tvaru. Při trénování vybraného detektoru jsou ovšem využívány vymežovací rámečky objektů pouze obdélníkového tvaru, kdy je udána  $x$  a  $y$  souřadnice horního levého rohu a následně šířka a výška rámečku, tedy  $(x, y, \text{šířka}, \text{výška})$  [55]. Z tohoto důvodu bylo potřeba vytvořit rámečky nové.

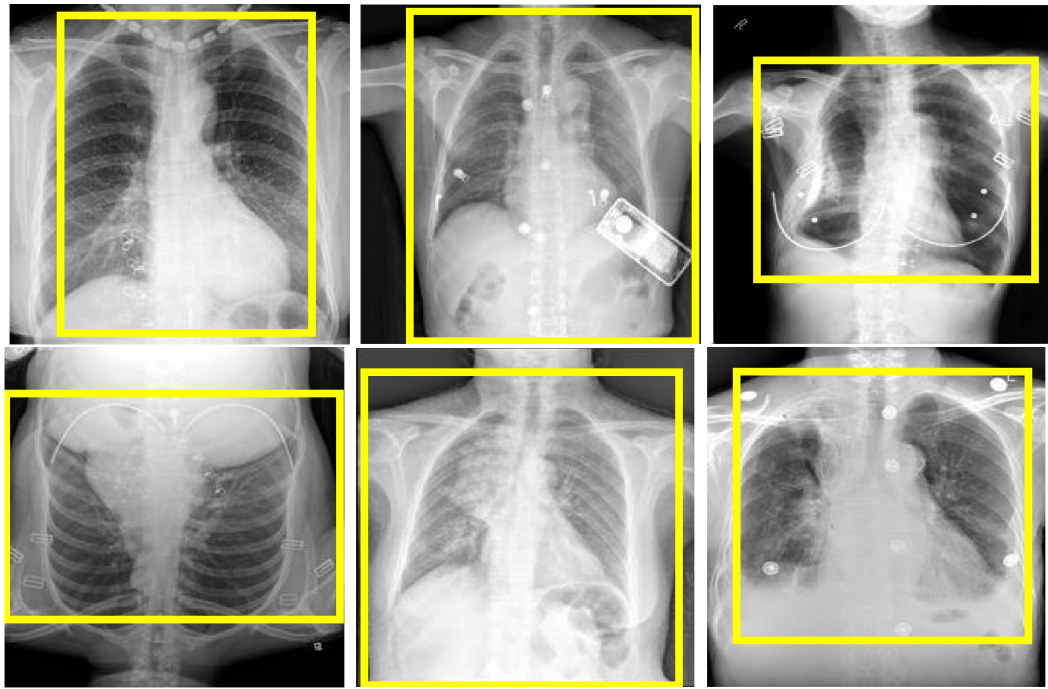
U objektů s tvarem obdélníku či elipsy byly vždy udány souřadnice pro levý horní

a pravý dolní roh, tedy se jednalo pouze o náhradu souřadnic pravého dolního rohu šířkou a výškou rámečku, která byla získána u obou souřadnic pouze odečtením souřadnic levého horního rohu od pravého dolního. U objektů s tvarem polygonu jsou v původních datech postupně uloženy  $x$  a  $y$  souřadnice jednotlivých bodů, proto bylo nejprve třeba uložit zvlášť do vektorů  $x$  a  $y$  souřadnice jednotlivých bodů a následně nalézt minimum a maximum pro obě souřadnice. Z těchto bodů se vytvořila souřadnice horního levého rohu (minimum souřadnic) a dopočítaly se opět šířka a výška rámečku obdobně jako u rámečků obdélníkového a eliptického tvaru. U některých anotovaných objektů byly udané body mimo obrázek, tedy bylo potřeba ohlídat, zda se všechny body vyskytují v rámci obrázku, popřípadě souřadnice posunout na okraj obrázku.

### **Odstranění nevhodně anotovaných dat**

Jak bylo již v kapitole 2.1 ukázáno, v anotacích objektů se projevuje poměrně velká variabilita, kdy jsou některé stejné předměty označovány jinak či jsou skupiny blízkých objektů označovány občas jako jeden objekt, jindy zase jako několik menších objektů. Tato situace pravděpodobně nastala z toho důvodu, že data označovalo 12 různých radiologů, kteří si nestanovili pravidla pro označování cizích objektů. Během diplomové práce bylo tedy uvažováno, že by se tato pravidla pro označování cizích objektů stanovila a některé anotace v snímcích by byly upraveny. Jelikož by se ovšem jednalo o kontrolu 4 500 snímků (4 000 trénovacích snímků s označenými objekty a 500 testovacích snímků s označenými objekty, viz. Tab. 2.1), bylo z této možnosti upuštěno a síť byla trénována na stažených původních datech.

V trénovacích datech byly ovšem prohlédnuty velké anotované objekty, zda neoznačují téměř celý snímek, což by pravděpodobně pro trénování sítě nebylo přínosné. Nejprve byly tedy zobrazeny všechny snímky, kde byla velikost anotovaného objektu  $100 \times 100$  pixelů, následně byly stejně zobrazeny snímky s anotovanými objekty o velikostech  $100 \times 150$  pixelů a  $150 \times 150$  pixelů. Jako nejlepší hranice pro nevhodně anotované objekty byla stanovena velikost  $100 \times 150$  pixelů. Jak lze vidět ve skriptu *spatne\_annotace.m*, tak byly zobrazeny právě takové snímky z trénovací skupiny s anotovaným objektem, na kterých se vyskytuje právě jeden cizí objekt s velikostí větší než  $100 \times 150$  pixelů. Jednalo se celkem o 40 snímků, které byly následně postupně zkontrolovány a byly mezi nimi nalezeny nevhodně anotované snímky. Odstraněno z trénovací skupiny tak bylo celkem 14 snímků, kdy některé z nich lze vidět na Obr. 2.5. Jednalo se především o snímky, které obsahují více oddělených cizích objektů označených jedním velkým anotovaným objektem, který zároveň překrývá téměř celý snímek, tedy oblast někdy zasahuje i mimo oblast plic, kde již cizí objekty dle pokynů zadání nemají být označeny.



Obr. 2.5: Ukázka odebraných obrázků z trénovací skupiny, na kterých označují anotované objekty několik předmětů najednou, a označují tak téměř celý obraz.

## 2.2 Výběr a návrh metody

Na základě prostudovaných metod strojového učení pro detekci objektů v obrazech, viz. kapitola 1.3, byla pro tuto diplomovou práci zvolena metoda Faster R-CNN pro vyhledání cizích objektů v RTG snímcích hrudníku. Tato metoda byla také použita v článku [31], kdy autoři vyhledávali pouze kruhové objekty v oblasti plic. Metoda dosáhla pro tento účel 97% přesnosti (*angl. precision*), 90% senzitivity a F1 skóre 93%. V této diplomové práci se na dostupných datech bude trénovat tento model pro účel vyhledávání tvarově nespécifických cizích objektů v RTG snímcích hrudníku v oblasti plic. Před použitím této metody je potřeba obrazy a anotované objekty nejprve upravit dle požadavků sítě, viz. kapitola 2.1.1. Příklad implementace metody ve vybraném programovacím prostředí MATLAB, kterým byla inspirována i tato diplomová práce, je možné vidět na stránce [56].

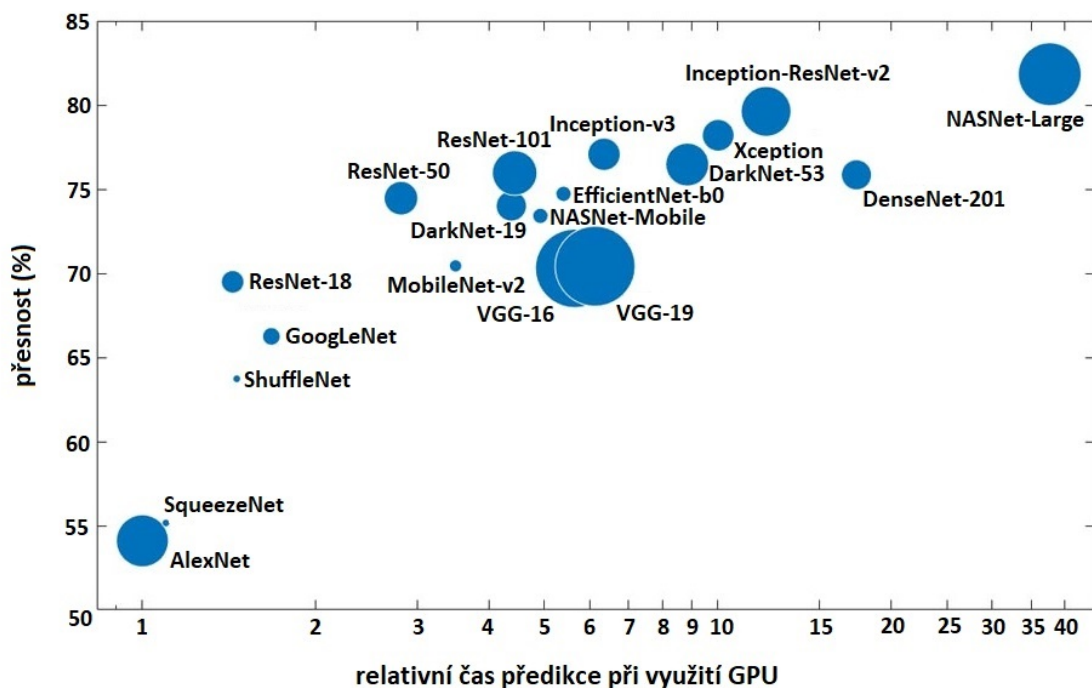
### 2.2.1 Výběr předtrénované neuronové sítě jako extraktoru příznaků

Jak již bylo popsáno v kapitole 1.5.1, vstupní obrazy při trénování Faster R-CNN nejprve přichází na vstup předtrénované neuronové sítě, která má funkci extraktoru příznaků, kdy tyto příznaky pak vstupují do dalších modulů. Předtrénované neuro-



nové sítě se využívají, jelikož jejich trénování je méně výpočetně i časově náročné a zároveň je k jejich přetrénování potřeba menší množství trénovacích obrazů. Tento alternativní přístup, kdy neučíme síť od nuly, ale využijeme předtrénované sítě, se nazývá *transfer learning*. Využívá skutečnosti, že extrakce jednoduchých příznaků na začátku neuronové sítě je shodná pro všechny úlohy. Extrakce příznaků se tedy liší až v posledních vrstvách neuronových sítí, které se právě při *transfer learningu* přeučují podle konkrétní úlohy. [57]

Kvalita extrakce příznaků je ovšem závislá na výběru předtrénované sítě. Při jejím výběru je nutné udělat kompromis mezi přesností výsledků, velikostí sítě a relativní dobou predikce. Příklady některých předtrénovaných sítí s těmito parametry jsou ukázány na Obr. 2.6. Jako jednodušší modely se často využívají sítě AlexNet [58], VGG16 [50], VGG19 [50] či GoogLeNet [59], pro složitější modely pak sítě ResNet-50 [36], Inception-v3 [46] či Densenet-201 [35]. Pro úlohu detekce objektů, což je i tato diplomová práce, se nejčastěji využívá síť ResNet-50, DarkNet-19 [60] či DarkNet-53 [60]. Při porovnání těchto sítí bylo rozhodnuto, že se pro účely této práce bude využívat předtrénovaná síť ResNet-50, jelikož má oproti zbývajícím sítím kratší relativní dobu detekce, ale srovnatelnou přesnost i velikost. [61]

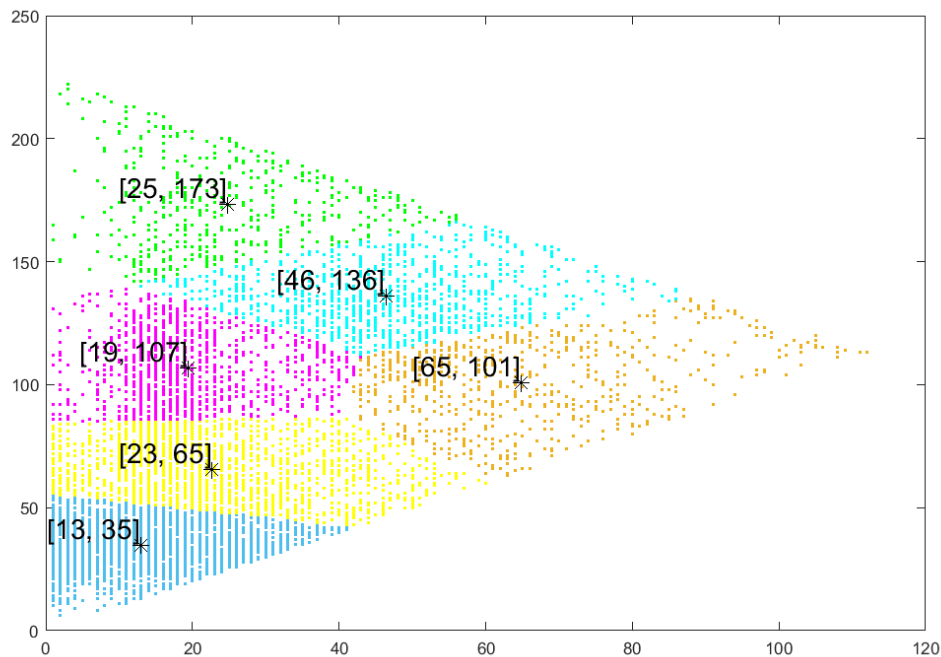


Obr. 2.6: Parametry předtrénovaných neuronových sítí využívaných pro *transfer learning*. V grafu lze sledovat úspěšnost sítí a rychlost predikce sítí, zároveň velikost značky ukazuje na velikost sítě samotné. Převzato a upraveno z: [61]

## 2.2.2 Stanovení počtu a velikostí anchorů

Před samotným trénováním Faster R-CNN je nejprve potřeba stanovit si počet anchorů a jejich velikosti. Stanovení nejlepších velikostí anchorů je vidět ve skriptu *anchors.m*.

Nejprve byla načtena matice anotovaných dat a byly z ní odstraněny obrazy, které neobsahují žádný cizí objekt, a dále obrazy, které byly odstraněny z trénovací skupiny kvůli nevhodné anotaci dat, viz. kapitola 2.1.1. Ve zbylých trénovacích datech pak byly nalezeny všechny velikosti anchorů označující cizí objekty v RTG snímcích. Velikosti anchorů pak byly upraveny tak, aby na první souřadnici byl vždy menší rozměr než na druhé souřadnici. Na tuto množinu pak byl aplikován algoritmus k-means pro rozdělení do tolika skupin, kolik velikostí anchorů bylo během trénování sítě třeba využít. Výsledek pro množství šesti anchorů lze vidět na Obr. 2.7. Lze si ovšem všimnout, že anchorů pravděpodobně nelze vhodně rozdělit do skupin pomocí k-means. Jelikož nemáme k dispozici ani žádný parametr, kterým bychom mohli určit vhodnost těchto velikostí anchorů, byl využit jiný přístup.



Obr. 2.7: K-means pro všechny velikosti anchorů v trénovací skupině

Dále byla tedy využita předpřipravená funkce, která určuje vhodné velikosti anchorů na základě průměrného IoU mezi anchorů a označenými objekty. Do funkce je potřeba zadat počet anchorů, se kterými chceme pracovat, a dále velikosti označených objektů. Výstupem funkce pak jsou konkrétní velikosti anchorů a průměrné IoU

mezi těmito anchory a anotovanými objekty. Jelikož jsou ovšem výstupem funkce pokaždé jiné velikosti anchorů, bylo vždy potřeba pro daný počet anchorů spustit funkci několikrát a vybrat nejlepší výsledek. Jednotlivé výsledky funkce pro různé počty anchorů jsou vidět v Tab. 2.2. Při trénování sítě budou tedy vždy pro konkrétní počet anchorů použity velikosti anchorů s největším průměrným IoU k anotovaným objektům.

Na základě metriky IoU mezi stanovenými velikostmi anchorů a anotacemi v trénovacích datech lze také stanovit minimální potřebný počet anchorů. Dle doporučení [62] je vhodné využít minimálně takový počet anchorů, který má IoU s anotovanými objekty alespoň 65%. Pro tato konkrétní trénovací data je tedy vhodné využít alespoň pět velikostí anchorů. Jak se ale později během trénování sítě ukázalo, přidávání počtu anchorů je velice výpočetně náročné, tedy větší počet než sedm velikostí anchorů nebylo vhodné využít, jelikož se tím omezilo nastavování ostatních parametrů sítě.

### 2.2.3 Trénování Faster R-CNN

Trénování sítě Faster R-CNN je provedeno ve skriptu *trenovanisite.m*. Nejprve je potřeba načíst předpřipravenou matici obsahující názvy obrázků a jim odpovídající anotované objekty, kdy tato matice obsahuje 8 000 snímků ze skupiny trénovací množiny. Také by bylo vhodné zároveň při trénování využít i validační data pro validaci sítě během učení, to ovšem v tomto programovacím prostředí není možné. Před samotným počátkem trénování sítě je zobrazen jeden obrázek i se svými anotovanými objekty, aby byla před trénováním zkontrolována správnost umístění anotací po jejich úpravách a úpravách obrázku (např. po převzorkování atd.).

Dále je potřeba definovat počet a velikosti anchorů, které byly vždy zvoleny na základě největšího IoU s anotovanými objekty, viz. Tab. 2.2. Následuje načtení předtrénované konvoluční neuronové sítě a jí odpovídající vrstvy pro extrakci příznaků, v tomto případě se jedná o předtrénovanou síť ResNet50 a vrstvu ReLU, na jejímž výstupu je podvzorkovaná příznaková mapa. Podle zvolené předtrénované sítě je také nutné definovat velikost vstupních obrázků, která je v tomto případě  $224 \times 224 \times 3$ . Z těchto vstupních parametrů je potřeba vytvořit vrstvy Faster R-CNN sítě, která má za úkol klasifikovat do jedné třídy.

Posledním krokem před samotným trénováním pak je nastavení trénovacích parametrů, např. počtu epoch či počátečního kroku učení. Popis nastavovaných parametrů bude popsán v následující kapitole 2.2.3. Trénování je pak provedeno buď na nově vytvořené Faster R-CNN síti, nebo v případě dotrénování již natrénované sítě se nahrazuje prázdná síť již částečně naučenou sítí.

Tab. 2.2: Průměrné IoU mezi definovanými anchorů a všemi anotovanými objekty. Průměrné IoU bylo zkoumáno pro různý počet anchorů a pro různé velikosti anchorů v rámci daného počtu. V řádku jsou zapsány velikosti anchorů a jim odpovídající IoU s anotovanými objekty (nejlepší výsledek ve spodním řádku označen tučně).

počet anchorů = 4

Velikost anchorů	Průměrné IoU
[4 5], [25 25], [61 48], [81 133]	43,8 %
[4 5], [25 26], [62 50], [70 132]	44,3 %
[22 22], [53 41], [44 84], [89 137]	47,5 %
[23 21], [42 46], [99 80], [57 145]	60,7 %
[20 20], [37 43], [73 49], [72 137]	61,8 %
[19 19], [34 39], [66 49], [73 135]	<b>62,6 %</b>

počet anchorů = 5

Velikost anchorů	Průměrné IoU
[5 4], [19 19], [40 36], [67 70], [71 154]	43,8 %
[5 4], [21 20], [44 39], [67 80], [73 162]	51,9 %
[5 4], [20 20], [41 36], [63 73], [75 158]	53,3 %
[5 4], [22 15], [33 32], [66 84], [78 166]	61,9 %
[19 19], [36 35], [71 52], [48 108], [95 156]	65,2 %
[18 18], [34 34], [65 47], [60 93], [78 169]	<b>66,1 %</b>

počet anchorů = 6

Velikost anchorů	Průměrné IoU
[5 4], [20 19], [39 38], [56 71], [128 54], [47 153]	53,1 %
[5 4], [19 19], [34 34], [50 67], [91 38], [80 150]	53,8 %
[5 4], [19 19], [38 34], [49 74], [109 49], [77 154]	54,8 %
[16 18], [33 25], [42 43], [49 93], [104 54], [84 157]	66,7 %
[17 18], [33 28], [46 47], [50 104], [107 55], [88 161]	67,2 %
[18 19], [39 28], [41 49], [89 67], [43 110], [85 164]	<b>67,5 %</b>

počet anchorů = 7

Velikost anchorů	Průměrné IoU
[5 4], [19 18], [33 33], [52 47], [50 103], [118 57], [84 163]	57,7 %
[5 4], [19 18], [33 30], [46 44], [49 90], [111 50], [82 156]	58,3 %
[5 4], [18 18], [33 28], [45 45], [49 99], [108 55], [84 160]	58,5 %
[16 17], [32 26], [42 41], [46 81], [94 56], [131 133], [67 170]	67,8 %
[17 18], [31 30], [47 51], [59 32], [47 95], [109 63], [82 162]	67,9 %
[17 16], [25 26], [26 29], [41 49], [96 62], [47 102], [84 161]	<b>68,3 %</b>

## Parametry trénování

Během diplomové práce byly hledány optimální parametry pro natrénování Faster R-CNN pro účely detekce cizích objektů v RTG snímcích hrudníku. Již z počátku se ukázalo, že natrénování takovéto sítě je velmi časově a výpočetně náročné, a je tedy nutné výpočty provádět na výkonném počítači za pomoci GPU. Velikost paměti GPU pak silně ovlivňuje možnosti nastavení některých parametrů. Výpočty během této diplomové práce byly prováděny na GPU o velikosti 12 GB, kdy tato paměť byla ve většině případů plně využita. Aby bylo trénování vůbec možné začít, muselo být tedy nejprve nastaveno několik parametrů na správnou hodnotu.

Prvním parametrem, který výrazně ovlivnil výpočetní náročnost při trénování sítě, je parametr metody trénování. Bylo zjištěno, že pro snížení výpočetních nároků je třeba využít klasické trénování ve čtyřech krocích místo trénování celé sítě najednou. Jelikož se nepodařilo síť začít trénovat ani v jednom případě metodou trénování celé sítě najednou, nedá se posoudit, zda nemá tato metoda oproti metodě trénování pomocí čtyř kroků nějakou výhodu, např. zda-li není trénování rychlejší. Dalším parametrem, který ovlivnil výpočetní náročnost, je velikost *minibatche*. Je jasné, že velikost *minibatche* nesmí být jedna, jelikož by během trénování sítě nebyla využita minibatch normalizace. Během této diplomové práce nebylo zkoumáno, jakou minimální velikost *minibatche* je potřeba použít, aby bylo trénování sítě úspěšné, ale velikost *minibatche* se pohyboval v rozmezí 4–10, což se zdá být dostatečná velikost. Posledním parametrem, který ovlivňuje výpočetní náročnost při trénování sítě, je počet generovaných oblastí zájmu v obrázku. Během nastavování parametrů se tak vždy volil kompromis mezi právě počtem oblastí zájmu v obrázku a mezi velikostí *minibatche*.

Při zkoumání, které všechny parametry ovlivňují právě výpočetní náročnost, byly zkoumány i další parametry. Nejprve bylo vyzkoušeno změnit optimalizační algoritmus z algoritmu ADAM na SGDM, ale nebyl zjištěn rozdíl, jelikož se nepodařilo žádný jiný parametr úspěšně zvýšit. Dále bylo vyzkoušeno použít jednodušší předtrénovanou síť pro extrakci příznaků, kdy místo sítě ResNet50 byl vyzkoušen AlexNet a následně i GoogLeNet. Při využití jednodušších sítí bylo sice možné parametr počtu oblastí zájmu trochu zvýšit, ale vzhledem k výrazně nižší úspěšnosti těchto sítí bylo rozhodnuto, že bude využita síť ResNet50 i přes možnost nastavení generování menšího počtu oblastí zájmu v obraze.

Při různém nastavování parametrů bylo také zjištěno, že některé parametry ovlivňují časovou náročnost. Nejvýraznějším parametrem je dle předpokladu použití GPU, kdy při využití CPU se nepodařilo z vysokých časových nároků síť ani dotrénovat. Výrazným parametrem ovlivňující dobu trénování je také dle očekávání počet epoch trénování. Dalším parametrem je velikost *minibatche*, kdy při zvětšování

tohoto parametru se čas lehce zkracuje. Poslední parametr, u kterého byla patrná změna časových nároků při trénování, je nastavení rozmezí IoU mezi anotovanými objekty a oblastmi zájmu pro zařazení do pozitivní a negativní skupiny. Nejvíce patrný tento účinek byl v případě, kdy se minimum IoU pro zařazení do negativní skupiny změnilo z 0,05 na 0,00, čímž se čas trénování sítě prodloužil přibližně na dva a půl násobek.

Všechny nastavované parametry v této diplomové práci jsou popsány zde:

- **solverName** - Použitý optimalizační algoritmus, na výběr je SGDM, RMSPROP či ADAM. V této diplomové práci je využit vždy algoritmus ADAM.
- **MaxEpochs** - Maximální počet epoch při učení sítě, tedy kolikrát bude každý obrázek z množiny trénovacích dat využit pro přepočtení parametrů během trénování sítě.
- **MiniBatchSize** - Velikost *minibatche*, tedy počet dat z trénovací množiny dat, podle kterých jsou upraveny parametry sítě během jedné iterace.
- **InitialLearnRate** - Počáteční krok učení, který musí být vhodně zvolen. Při malém počátečním kroku bude učení probíhat dlouho, při velkém počátečním kroku nemusí být nalezeno optimum.
- **L2Regularization** - Nastavení regularizačního faktoru pro L2 regularizaci, která zabraňuje přeučení sítě.
- **GradientDecayFactor** - Nastavení konstanty  $\beta_1$  pro adaptivní učící koeficient při využití algoritmu ADAM.
- **SquaredGradientDecayFactor** - Nastavení konstanty  $\beta_2$  pro adaptivní moment při využití algoritmu ADAM.
- **NegativeOverlapRange** - Rozmezí IoU mezi skutečným objektem a nalezeným objektem pomocí Faster R-CNN, kdy se nalezený objekt řadí do skupiny negativních vzorků.
- **PositiveOverlapRange** - Rozmezí IoU mezi skutečným objektem a nalezeným objektem pomocí Faster R-CNN, kdy se nalezený objekt řadí do skupiny pozitivních vzorků.
- **NumRegionsToSample** - Počet náhodně generovaných oblastí zájmu v jednom obrázku, u kterých poté síť určuje pravděpodobnost, že se jedná o hledaný objekt.
- **TrainingMethod** - Nastavení trénovací metody, kdy se síť může trénovat klasicky čtyři kroky postupně nebo celá nejednou. Z důvodu nižších výpočetních nároků na trénování je v této diplomové práci použita vždy metoda postupného trénování ve čtyřech krocích.
- **ExecutionEnvironmet** - Jednotka, na které budou probíhat výpočty. Z důvodu vysoké časové a výpočetní náročnosti při trénování Faster R-CNN bude v této diplomové práci vždy využíváno GPU.

Během diplomové práce byly tyto parametry nastavované v různé kombinaci tak, aby se později dalo určit jejich optimální nastavení. Tato nastavení parametrů i s vyhodnocením jejich úspěšnosti, které bude následně popsáno v kapitole 2.2.4, lze nalézt v příloze v Tab. A.1.

## 2.2.4 Vyhodnocení úspěšnosti detektorů

Vyhodnocování kvality detektorů se během této diplomové práce provádělo na 1 000 snímcích hrudníku ze skupiny testovacích dat, kdy na 500 snímcích se vyskytoval alespoň jeden cizí objekt a na zbylých 500 snímcích se cizí objekt nevyskytuje. Vyhodnocování úspěšnosti bude tedy prováděno na těchto dvou skupinách zvlášť, kdy celková kvalita detektoru se poté bude hodnotit podle výsledků obou skupin. Vyhodnocení kvality jednotlivých detektorů lze vidět ve skriptech *vysledky\_detekce\_prahy.m*.

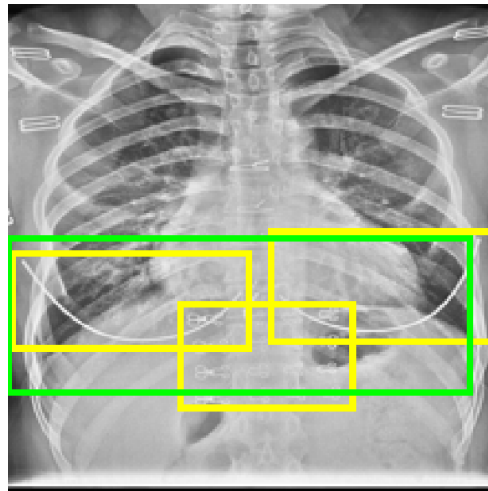
Prvním krokem při vyhodnocení kvality natrénovaného detektoru je detekovat na snímcích cizí objekty. K tomuto je třeba nejprve určit práh detekce, tedy jakou minimální pravděpodobnost, že se jedná o cizí objekt, musí detektor pro oblast zájmu určit, aby byla ve výsledcích označena jako cizí objekt. V originální funkci je práh nastaven na 0,5, bylo ovšem zkoumáno, zda by změna prahu na vyšší hodnotu neodstranila některé falešně pozitivní (FP) výsledky, které by mohly mít menší pravděpodobnost zařazení do skupiny cizího objektu než objekty samotné. Jelikož bylo ovšem testování prahů časově náročné, byly vybrány pouze tři úrovně prahů, které byly testovány, a to prahy 0,5, 0,55 a 0,6. Při určování nejlepšího prahu bylo bráno v potaz, že při úkolu detekce cizích objektů v RTG snímcích hrudníku je důležitější, aby detektor našel všechny objekty nacházející se v snímku, než aby bylo méně falešně pozitivních výsledků. Důvod je ten, že pro lékaře by mělo být v praxi lepší podívat se na nalezené objekty a v případě, že by se nejednalo o cizí objekt, pouze jeden z označených objektů odstranit, než prohlížet celý snímek a hledat neoznačené cizí objekty. S ohledem na tuto skutečnost byla pak hledána vhodná hodnota prahu, kdy se hodnota prahu zvyšovala pouze v případě, kdy se se zvýšením prahu odstranilo alespoň 10krát více falešně pozitivních výsledků než se přidalo falešně negativních výsledků. Pokud se tedy se zvýšením prahu např. přidalo 20 falešně negativních (FN) výsledků a odstranilo se 50 falešně pozitivních výsledků, pak se práh nezvyšoval. Pokud by se ale přidalo 20 falešně negativních výsledků a 203 falešně pozitivních výsledků by se odstranilo, pak by se práh zvýšil. Tento postup byl proveden nejprve pro zvýšení prahu z hodnoty 0,5 na hodnotu 0,55 a poté dále z hodnoty 0,55 na hodnotu 0,6.

Pro určení správného prahu detekce bylo tedy nejprve potřeba pro všechny hodnoty prahu vypočítat hodnoty FP a FN, které se určují na základě anotovaných objektů a detekovaných objektů. Po detekci pomocí jednoho z prahů byl před sa-

motným výpočtem správných a falešných detekcí přidán ještě krok nalezení nejpravděpodobnějšího boxu. Na všech obrázcích postupně jsou tedy vypočteny překryvy všech detekovaných objektů pomocí metriky IoU, jejíž výpočet se provádí pomocí vzorce 2.1. Pokud jsou překryvy pro některé detekované objekty větší než 0,1, pak se z překrývajících se boxů zachová pouze ten s největší pravděpodobností zařazení do skupiny. Takto malý práh byl zvolen z důvodu, že na snímcích se cizí objekty většinou nepřekrývají nebo jsou v případě překrytí označeny jako jeden objekt, tedy není nutné zachovávat např. z poloviny se překrývající objekty jako v jiných aplikacích.

$$IoU = \frac{\text{plocha}(A \cap B)}{\text{plocha}(A \cup B)} \quad (2.1)$$

Po kroku nalezení nejpravděpodobnějších boxů se může již vyhodnocovat úspěšnost detektoru. Jelikož bylo zjištěno, že oficiální vyhodnocovací funkce nefungují v některých případech správně, byl pro vyhodnocení TP, FP a FN výsledků vytvořen vlastní přístup. Při využití originálních funkcí docházelo k selhání např. pro případy, kdy na obrázku je více menších objektů blízko sebe, které jsou v originálních datech označeny jako jeden větší objekt, ale detektor je určí jako každý objekt zvlášť, nebo naopak. Příklad lze vidět na Obr. 2.8. V originálních funkcích se totiž umí přiřazovat pouze jeden anotovaný objekt k jednomu TP boxu. Pokud se tedy jako na obrázku tři originální anotované objekty označí jako jeden objekt, výsledek originální funkce bude TP = 1, FP = 0 a FN = 2, tedy z výsledků vychází, že se dva anotované objekty neoznačily i přes jejich nalezení, tedy správnou detekci všech cizích objektů v obraze.



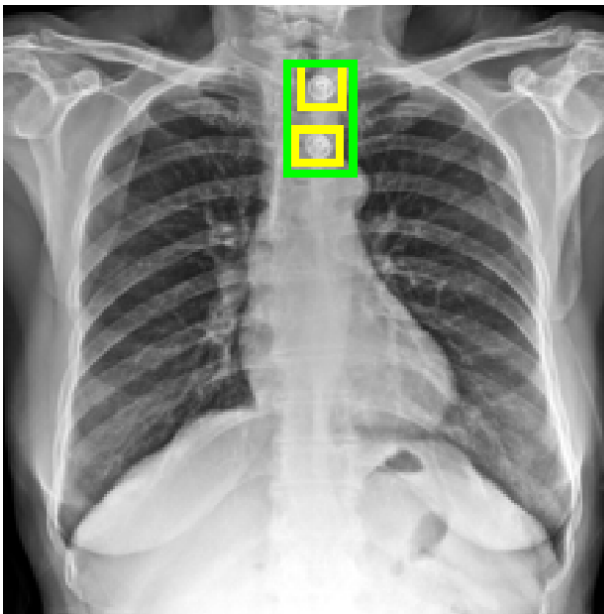
Obr. 2.8: Příklad špatného vyhodnocení úspěšnosti originální funkcí. Žlutě jsou vyznačeny originální anotované objekty, zeleně nalezené objekty. V tomto případě by originální funkce vyhodnotila TP = 1, FP = 0, FN = 2, správně by se mělo vyhodnotit TP = 3, FP = 0, FN = 0.



Vyhodnocení TP, FP a FN výsledků proběhlo následujícím způsobem. Nejprve se pro kombinaci všech anotovaných objektů vytvořila matice překryvů se všemi nalezenými objekty v každém obraze. Metrika překryvů není tentokrát klasické IoU, jedná se o úpravu, kdy ve jmenovateli vzorce není plocha sjednocených boxů, ale pouze plocha menšího z boxů, viz. vzorec 2.2.

$$Min = \frac{plocha(A \cap B)}{\min(plocha(A), plocha(B))} \quad (2.2)$$

Tato metrika byla využita především z důvodu, že detektor často určoval více menších anotovaných objektů jako jeden velký. Pokud tento případ ovšem nastane, metrika IoU velmi rychle klesá, kdežto metrika Min zůstává velmi blízká hodnotě 1. Příklad je vidět na Obr. 2.9, kdy jsou dva objekty detekovány jako jeden a metriky IoU a Min se od sebe velice liší. Z důvodu odlišných již samotných anotací objektů, kdy v datech se občas vyskytuje anotace několika menších cizích objektů do jednoho velkého objektu, budou i při vyhodnocování brány tyto detekce jako správné. Z tohoto důvodu bylo tedy třeba využít právě metriku Min, která zachová překryv podobných detekcí blízký 1, jelikož při využití metriky IoU by byly podobné případy detekce vyhodnoceny jako špatné.



**IoU = [0,27 0,22]**

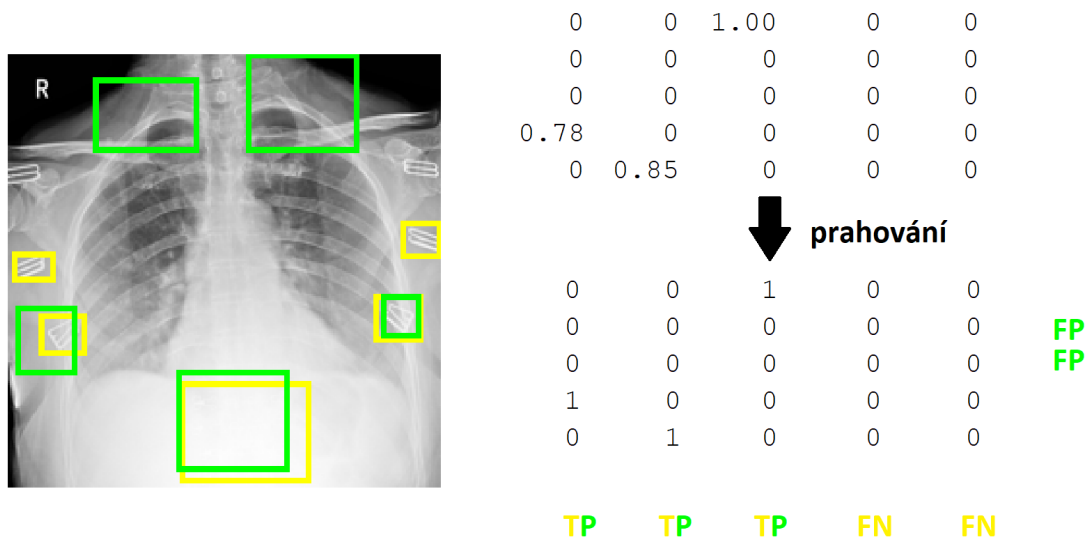
**vs.**

**Min = [1 1]**

Obr. 2.9: Rozdíl mezi metrikou IoU a Min pro překryv bounding boxů. Z důvodu, že podobné detekce budou brány jako správné, se pro vyhodnocení správnosti detekovaného objektu využívá metrika Min.

Po sestavení matice překryvů je tato matice naprahována hodnotou 0,3, což vede k tomu, že pokud se některý detekovaný objekt překrývá s anotovaným objektem

alespoň na 30 % za pomoci metriky Min, poté bude určen jako správně detekovaný objekt. Tato hodnota byla zvolena heuristicky a s její změnou by pochopitelně docházelo ke změně vyhodnocení správně detekovaných objektů. Za pomoci této matice poté dále probíhá vyhodnocení detektoru. Pokud nemá tato matice žádný sloupec, pak na obraze nebyl anotovaný žádný objekt a tedy počet řádků této matice odpovídá počtu FP výsledků. Tato hodnota se zapisuje do speciální proměnné, která bude hodnotit pouze průměrný počet FP výsledků na obrázek v snímcích, kde se žádný cizí objekt nevyskytuje. Dále se kontroluje, zda má matice nějaký řádek. Pokud matice nemá žádný řádek, ale má určitý počet sloupců, znamená to, že na snímku cizí objekty jsou, ale detektor je nenašel. V tomto případě se uloží počet sloupců do hodnoty FN, hodnoty TP a FP budou obě nulové. V případě, že matice má minimálně jeden řádek i sloupec, znamená to, že na snímku se vyskytují cizí objekty a detektor zároveň detekoval některé oblasti. Nejprve se zjišťuje v kolika sloupcích se vyskytují pouze nuly, čímž zjistíme počet FN výsledků (tyto sloupce odpovídají takovým anotovaným objektům, které nemají ani s jedním nalezeným objektem překryv alespoň 30 %). Dále se poté zjišťuje, v kolika řádcích jsou pouze samé nuly, čímž se určí počet FP výsledků (jsou to takové nalezené objekty, které nemají se žádným anotovaným objektem překryv alespoň 30 %). Počet TP se poté počítá jako počet sloupců této matice ponížený o počet FN výsledků. Příklad detekovaných objektů, anotovaných objektů a matice jejich překryvů je vidět na Obr. 2.10.



Obr. 2.10: Zobrazení vyhodnocení TP, FP a FN. Zeleně jsou zobrazeny detekované objekty, žlutě anotované objekty. Sloupce v matici překryvu naznačují anotované objekty, řádky detekované objekty.

Z vektorů TP, FP, FN a FN pro obrazy bez anotovaného objektu se vyhodnocuje kvalita detektoru. Jako první se počítá celkový počet TP, FP a FN výsledků ve snímcích s anotovanými objekty. Dále se počítá přesnost (odpovídající angl. *precision* či *recall*), senzitivita (odpovídající také pozitivní prediktivní hodnotě, angl. *sensitivity*, PPV (positive predictive value)) a F1 skóre v každém obrázku. Ve snímcích, kde byly anotovány objekty, ale žádné objekty nebyly detekovány se hodnota přesnosti nezapočítává, stejně tak F1 skóre ve snímcích, kde je hodnota přesnosti i senzitivity rovna 0. Následně je vypočítána průměrná hodnota přesnosti, senzitivity i F1 skóre přes všechny snímky. V poslední řadě je vypočítán průměrný počet FP výsledků na obrázek jak pro snímky s anotovanými cizími objekty, tak pro snímky bez anotovaných cizích objektů.

Pro lepší přehled pak byla vytvořena tabulka se všemi hodnotícími parametry pro všechny prahy detekce zároveň. V tuto chvíli se také najde ideální práh pro detekci, tedy zjistí se rozdíly hodnot FP a FN mezi jednotlivými prahy detekce a podle pravidla popsaného na začátku této kapitoly se vyhodnotí ideální hodnota prahu. Všechny hodnotící parametry včetně určeného ideálního prahu detekce pro jednotlivé detektory jsou zobrazeny v příloze v Tab. A.1. Podrobnější vyhodnocení některých z detektorů bude poté rozebráno v kapitole 2.2.5.

## 2.2.5 Nalezení nejlepšího natrénovaného detektoru

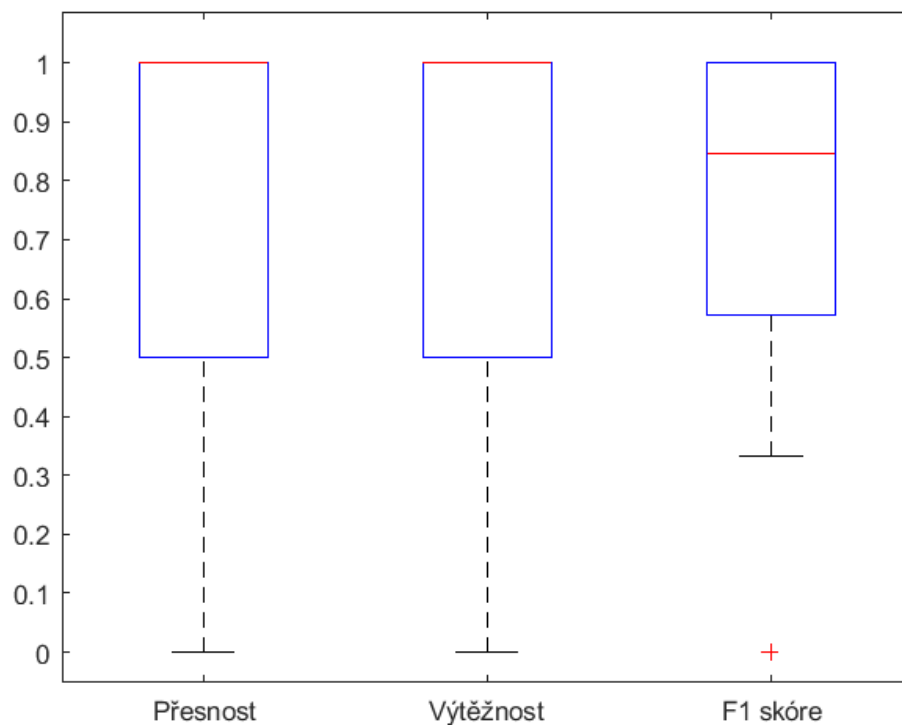
Cílem této kapitoly je podrobněji prozkoumat některé z natrénovaných detektorů a stanovit, který z nich by byl pro praktické využití nejvhodnější. Původním cílem kapitoly bylo také výsledky tohoto detektoru porovnat s výsledky, kterých bylo dosaženo v rámci zadané soutěže, ze které pochází veškerá data v této diplomové práci. Bohužel se ale tyto výsledky nadají dohledat, jelikož webová stránka soutěže, na které byly umístěny i výsledky, byla zrušena a nebyl nalezen žádný článek popisující metodu provedenou v rámci této soutěže.

### Detektor s nejvyšším F1 skóre

Detektor, který dosáhl mezi všemi natrénovanými detektory nejvyššího F1 skóre, dosáhl zároveň i nejvyšší přesnosti. Jelikož při určování optimálního prahu detekce bylo z výpočetních a časových nároků hledání omezeno na hodnoty 0,5–0,6, bylo před samotným zhodnocením detektoru zkontrolováno, zda by nebylo vhodnější práh detekce snížit pro dosažení lepších výsledků. To se ovšem nepotvrdilo, a práh detekce tohoto detektoru tedy zůstal na hodnotě 0,5. Parametry detektoru a podrobnější vyhodnocení úspěšnosti lze pak vidět v Tab. 2.3. Konkrétně tedy tento detektor dosáhl přesnosti 72 %, senzitivity 76 % a F1 skóre 74 %. Přesnosti a senzitivity na jednotlivých obrázcích jsou pak zobrazeny v boxplotu na Obr. 2.11.

Tab. 2.3: Parametry trénování a vyhodnocení úspěšnosti detekce pro detektor s nejlepším F1 skóre.

Parametry trénování		Vyhodnocení detekce	
MaxEpochs	5	TP	729
MiniBatchSize	7	FP	225
InitialLearnRate	10e-5	FN	222
L2Regularization	10e-7	Průměrné FP s objekty	0,45
GradiendDecayFactor	0,95	Přesnost průměr	0,72
SquaredGradiendDecay	0,9	Přesnost sm. odchylka	0,38
NegativeOverlapRange	[0 0, 3]	Senzitivita průměr	0,76
PositiveOverlapRange	[0, 3 1]	Senzitivita sm. odchylka	0,38
NumRegionsToSample	400	F1 skóre průměr	0,74
Počet anchorů	7	F1 skóre sm.odchylka	0,36
Práh detekce	0,5	Průměrné FP bez objektů	0,78
		Časová náročnost	3,3 s



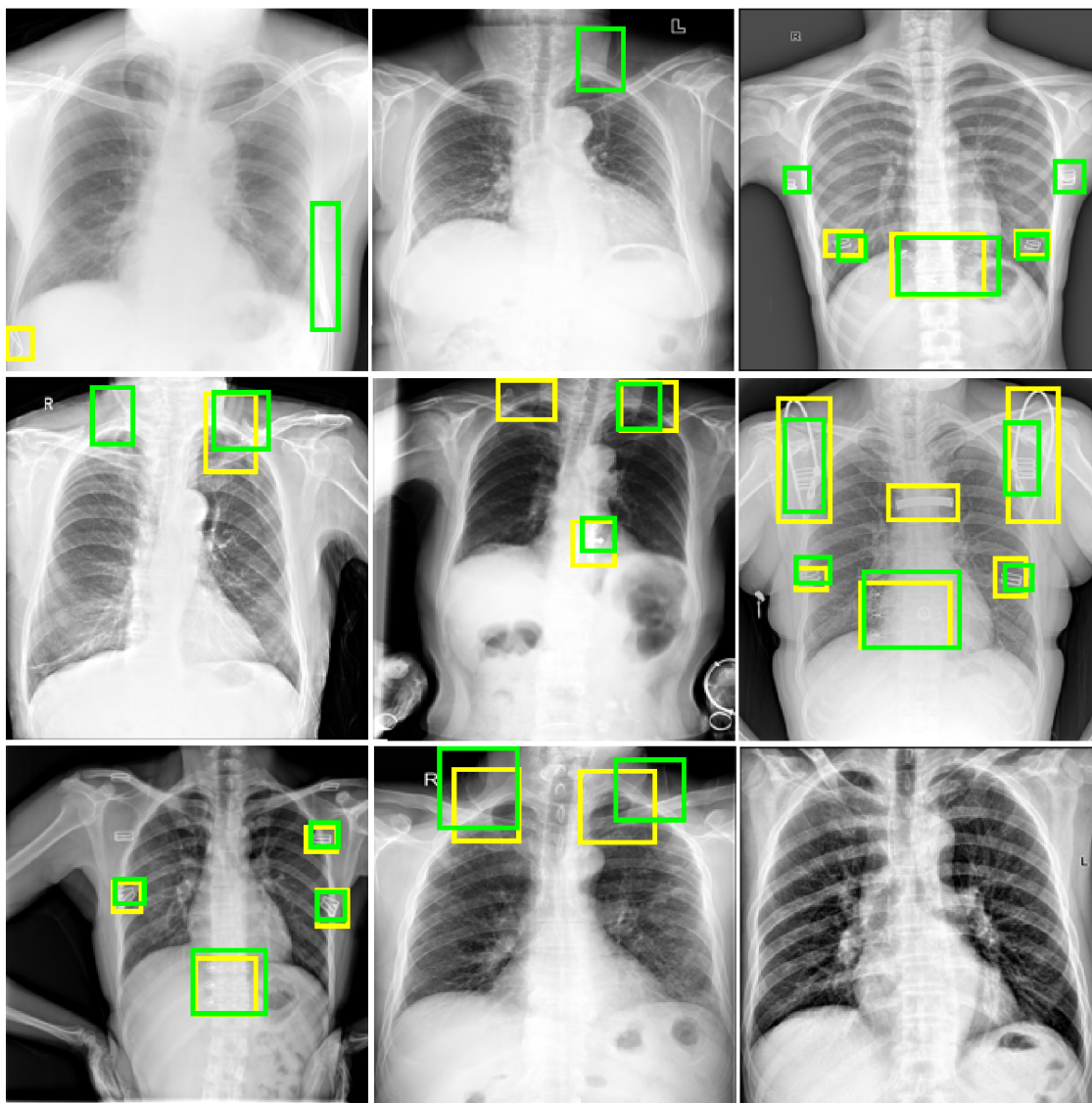
Obr. 2.11: Boxplot přesnosti, senzitivity a F1 skóre pro detektor s nejlepším F1 skóre. Červená čára označuje medián, spodní hranice boxu značí první kvartil, horní hranice třetí kvartil, vousy označují krajní hodnotu bez započítání odlehlých hodnot, které jsou označeny zvlášť červeným křížkem (o odlehlou hodnotu se jedná v případě, kdy její vzdálenost od kraje boxu je alespoň  $1,5\times$  větší než je výška boxu).

Jak lze vidět, tak spodní kvartil všech tří parametrů je větší než 0,5 a medián je pro přesnost a senzitivitu roven 1, pro F1 skóre je roven 0,85. Dá se tedy konstatovat, že minimálně v polovině detekcích byla přesnost či senzitivita rovna 1, ve 228 případech, tedy téměř v polovině všech detekcí, byly obě tyto hodnoty rovny 1, tedy se jednalo o bezchybnou detekci cizích objektů. Ve většině případů pak byly tyto hodnoty větší než 0,5, pouze na 106 snímcích klesla hodnota přesnosti pod 0,5 a hodnota senzitivity klesla pod tuto hodnotu na 97 snímcích. Co se týče snímků bez výskytu cizích objektů, průměrný počet FP detekcí na snímek se rovná 0,78, u snímků s výskytem cizích objektů je pak tato hodnota o něco nižší, a to 0,45. Mělo by se jednat tedy o poměrně spolehlivý detektor, který na snímcích ve většině případů zachytí cizí objekt a zároveň nebude označovat mnoho oblastí, ve kterých se cizí objekt nevyskytuje, a je tedy potřeba je z detekcí odstranit.

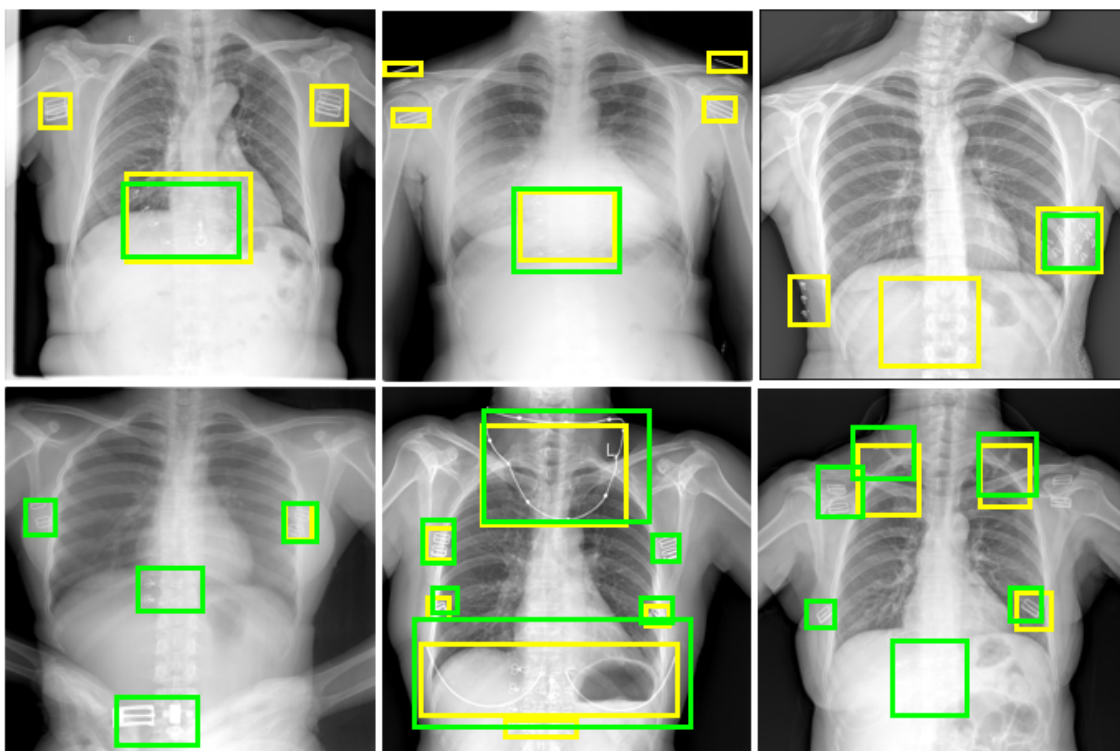
Na Obr. 2.12 jsou vidět příklady některých snímků. Žlutě jsou označeny anotované objekty, zeleně jsou pak označeny objekty detekované. Na spodních snímcích lze vidět případy, kdy detekce cizích objektů byla plně úspěšná. V prostředním řádku pak lze vidět ukázky snímků, na kterých se jeden objekt buď nedetekoval, nebo se naopak detekovala jedna oblast navíc. V horních obrázcích pak můžeme pozorovat příklad horších detekcí, kdy se nedetekuje žádný objekt správně, případně se detekuje více oblastí, ve kterých není anotovaný cizí objekt.

Při pohledu na pravý horní snímek na Obr. 2.12 si lze všimnout, že detektor někdy označuje oblasti, ve kterých se cizí objekt v okolí plic skutečně nachází, ovšem objekt není anotován. Z tohoto důvodu byly podrobněji prohlédnuty snímky, aby bylo zjištěno, zda nedochází k takovýmto detekcím často. Pokud by taková situace nastala, větší úspěšnosti detekce by se mohlo dosáhnout tak, že by byly segmentovány plíce na každém snímku a detekované objekty by byly odstraněny v případě, kdy nezasahují do oblasti plic. Po podrobnějším prohlédnutí snímků se ovšem ukázalo, že samotné anotované objekty se také v některých případech nachází mimo oblast plic, a to v některých případech i poměrně výrazně (viz. Obr. 2.13 nahoře). Na stejném obrázku dole si pak můžeme všimnout, že detektor vyhledává v některých případech objekty, které jsou podobně umístěny jako na jiných snímcích objekty anotované, ale na současném snímku anotovány nejsou.

Z důvodu těchto rozdílných anotací byly postupně projity všechny snímky a bylo zjištěno, že detektor označil celkem 56 oblastí, ve kterých není anotovaný cizí objekt, ale cizí objekt se zde přesto nachází. Tyto detekované objekty pak byly přeřazeny ze skupiny FP do TP a bylo provedeno nové vyhodnocení, které lze vidět v Tab. 2.4. Pro využití detektoru v praxi by bylo třeba udělat více specifické požadavky, v jaké oblasti se mají cizí objekty hledat, a detektor natrénovat buď na nově anotovaných datech, případně by bylo vhodné provést např. odstranění detekovaných oblastí, které nezasahují do oblasti plic.



Obr. 2.12: Zobrazení snímků, na nichž jsou detekovány cizí objekty pomocí detektoru s nejvyšším F1 skóre. Žlutě jsou zobrazeny anotované objekty, zeleně pak detekované objekty.



Obr. 2.13: Nejednotné anotování objektů mimo oblast plic. Úspěšnost detekce může snižovat fakt, že některé objekty mimo oblast plic jsou detekovány, některé ne. Stejně tak jsou některé objekty mimo oblast detekovány, a jiné ne. Žlutě jsou vyznačeny anotované objekty, zeleně pak objekty detekované.

Tab. 2.4: Vyhodnocení úspěšnosti detekce pro detektor s nejlepším F1 skóre po úpravě anotací. Detekované objekty, které nebyly anotované, ale přesto se jednalo o cizí objekt v blízkosti plic, byly vyřazeny ze skupiny FP a zařazeny do skupiny TP.

Vyhodnocení detekce po úpravě	
TP	785
FP	169
FN	222
Průměrné FP s objekty	0,45
Přesnost průměr	0,75
Přesnost sm. odchylka	0,38
Senzitivita průměr	0,77
Senzitivita sm. odchylka	0,38
F1 skóre průměr	0,76
F1 skóre sm. odchylka	0,32
Průměrné FP bez objektů	0,78

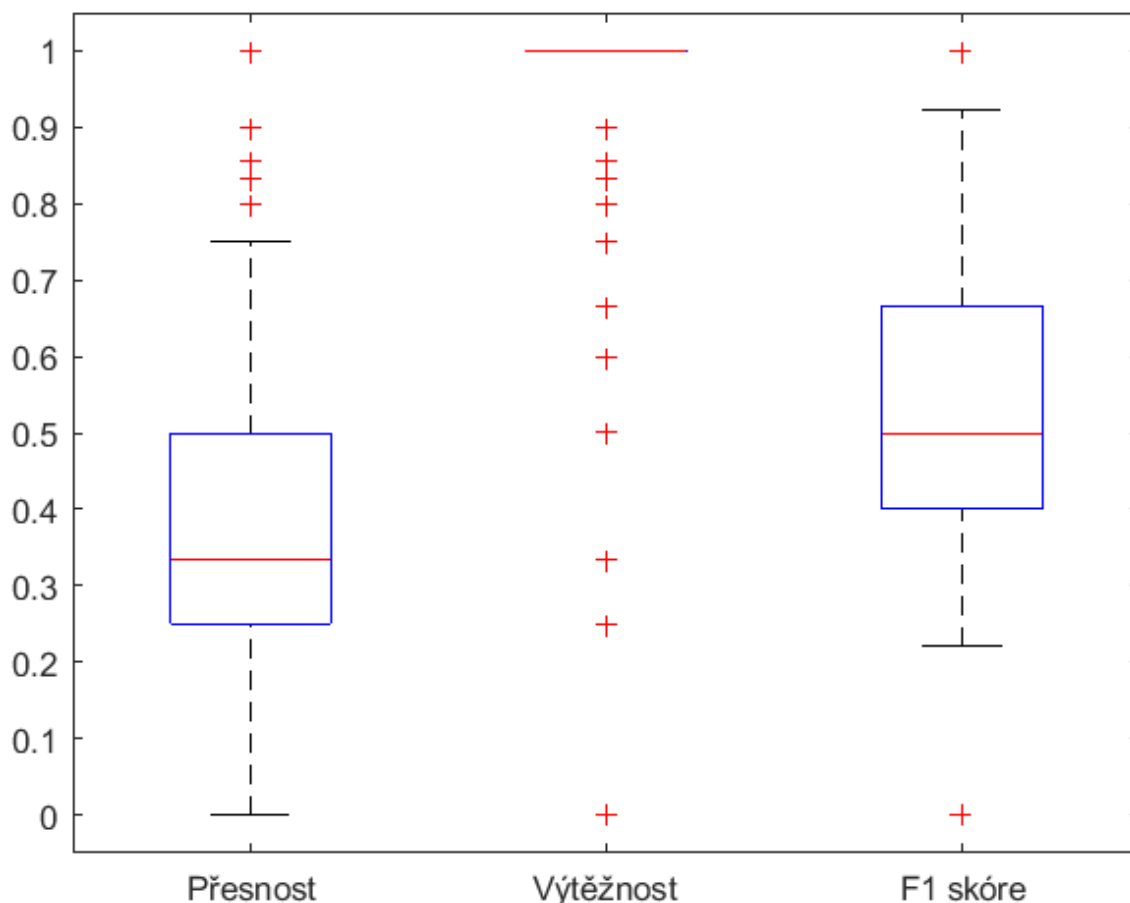
## Detektor s nejvyšším počtem TP výsledků

Detektor s nejvyšším počtem TP výsledků je samozřejmě i detektor s nejmenším počtem FN výsledků, zároveň má tento detektor ale i nejvyšší senzitivitu, a to 91 %. Přesnost tohoto detektoru je ovšem pouze 41 % a F1 skóre 56 %. Podrobnější výsledky jsou vidět v tabulce 2.5 a a na boxplotech pro přesnost, senzitivitu a F1 skóre na jednotlivých snímcích na Obr. 2.14. Jak si lze tedy všimnout, na většině snímků je sice senzitivita 1, ale přesnost rovna 1 vyšla pouze na 33 snímcích. Přesnost i senzitivita rovna jedné vyšla na 28 snímcích, na nich tedy byla detekce úplně správná. Bylo zkontrolováno, zda by další zvýšení prahu, který je v tomto případě již na hodnotě 0,6, nepomohlo k úspěšnější detekci, ovšem podle pravidla určeného v kapitole 2.2.4 již není vhodné práh zvyšovat. Podle těchto výsledků se tedy zdá, že detektor sice nachází téměř všechny cizí objekty ve snímcích, zároveň ale nachází velké množství oblastí, ve kterých se anotované objekty nevyskytují. Při hledání optimálního detektoru je sice snaha najít detektor, který vyhledává co nejvíce cizích objektů i za cenu mírně vyššího množství FP výsledků, v tomto případě se již ale jedná o opravdu velké množství falešných detekcí, které by pravděpodobně způsobilo špatnou orientaci ve snímcích.

Tab. 2.5: Parametry trénování a vyhodnocení úspěšnosti detekce pro detektor s nejvyšším počtem TP výsledků.

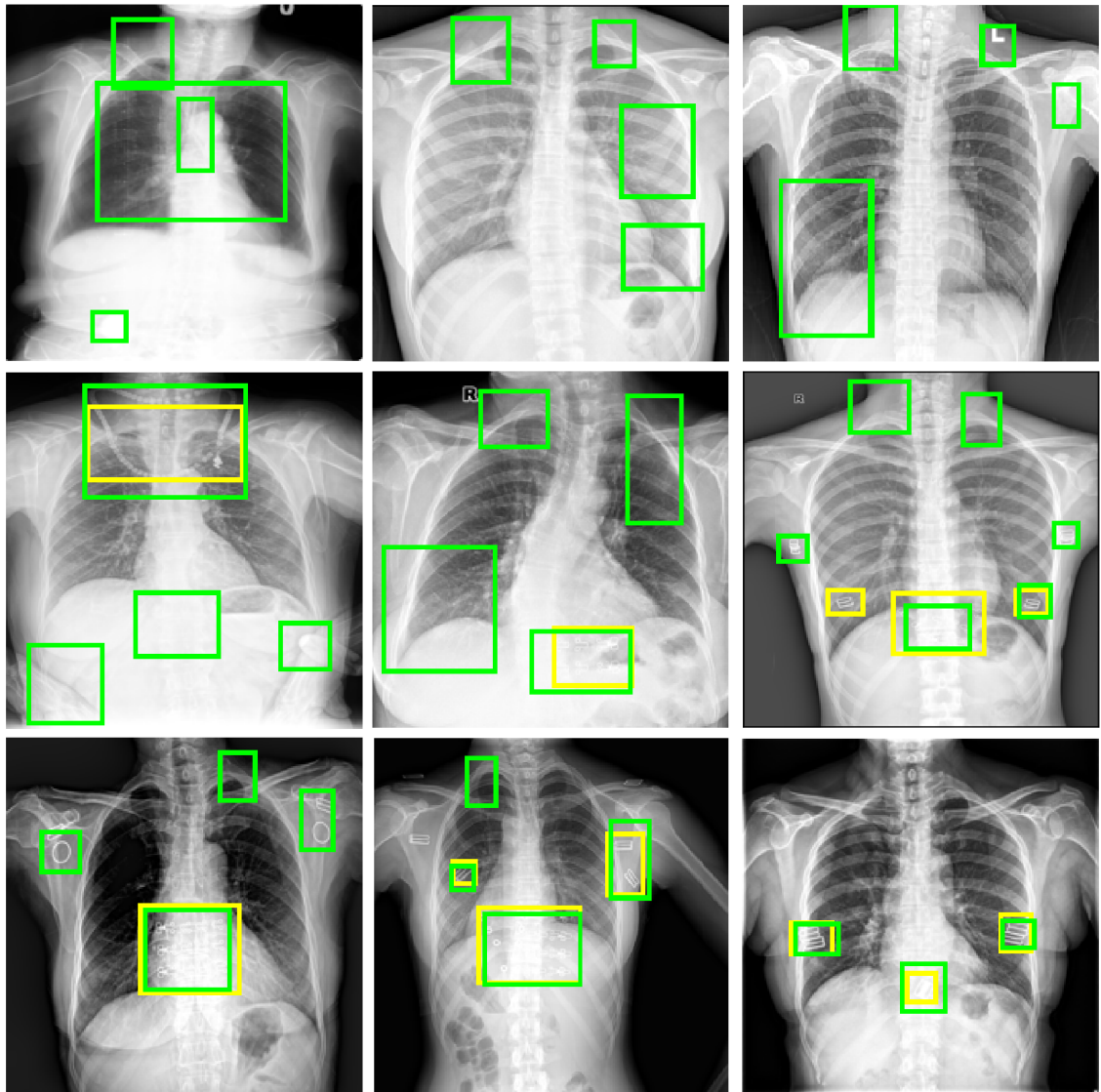
Parametry trénování		Vyhodnocení detekce	
MaxEpochs	9	TP	851
MiniBatchSize	7	FP	1308
InitialLearnRate	10e-5	FN	100
L2Regularization	10e-7	Průměrné FP s objekty	2,62
GradiendDecayFactor	0,95	Přesnost průměr	0,41
SquaredGradiendDecay	0,9	Přesnost sm. odchylka	0,24
NegativeOverlapRange	[0, 05 0, 3]	Senzitivita průměr	0,91
PositiveOverlapRange	[0, 3 1]	Senzitivita sm. odchylka	0,25
NumRegionsToSample	400	F1 skóre průměr	0,56
Počet anchorů	7	F1 skóre sm. odchylka	0,23
Práh detekce	0,6	Průměrné FP bez objektů	3,88
		Časová náročnost	3,4 s





Obr. 2.14: Boxplot přesnosti, senzitivity a F1 skóre pro detektor s nejvyšším počtem TP výsledků. Červená čára označuje medián, spodní hranice boxu značí první kvartil, horní hranice třetí kvartil, vousy označují krajní hodnotu bez započítání odlehlých hodnot, které jsou označeny zvláště červeným křížkem (o odlehlou hodnotu se jedná v případě, kdy její vzdálenost od kraje boxu je alespoň  $1,5\times$  větší než je výška boxu).

Na Obr. 2.15 jsou vidět příklady detekce cizích objektů pomocí tohoto detektoru. Dole jsou vidět lepší detekce, kdy byla detekce úplně správná, tedy našly se všechny anotované objekty a zároveň se nevyhledaly žádné oblasti bez anotovaného objektu, případně nalezených oblastí bez anotovaného objektu je málo nebo se v detekovaných oblastech vyskytují objekty, které nejsou pouze anotovány. Na snímcích níže jsou ovšem zobrazeny snímky, které jsou pro tento detektor běžnější, tedy anotovaný objekt je ve valné většině případů nalezen, ovšem kromě něj se detekují další oblasti, které neobsahují cizí objekt.



Obr. 2.15: Zobrazení snímků, na nichž jsou detekovány cizí objekty pomocí detektoru s nejvyšším počtem TP výsledků. Žlutě jsou zobrazeny anotované objekty, zeleně pak detekované objekty.

Již na první pohled jsou tyto snímky nepřehledné a práce s nimi by pravděpodobně nepřinesla v praxi ulehčení. Pochopitelně by se některé detekované oblasti bez cizích objektů daly automaticky odstranit tím, že by se ve snímcích provedla automatická segmentace plic a oblasti nezasahující do této oblasti plic by byly odstraněny. Také můžeme pozorovat, že některé detekované oblasti označují cizí objekt, který není pouze anotován, tím pádem bychom je jako v ostatních případech vyhodnocení mohli přesunout ze skupiny falešných detekcí do detekcí správných. Ze zobrazených snímků je ovšem zřejmé, že ani tato dvě vylepšení by pravděpodobně

nedosáhla uspokojivého výsledku, jelikož snímky jsou plné detekovaných oblastí bez cizích objektů, které do oblasti plic zasahují.

### **Detektor s nejnižším počtem FP výsledků**

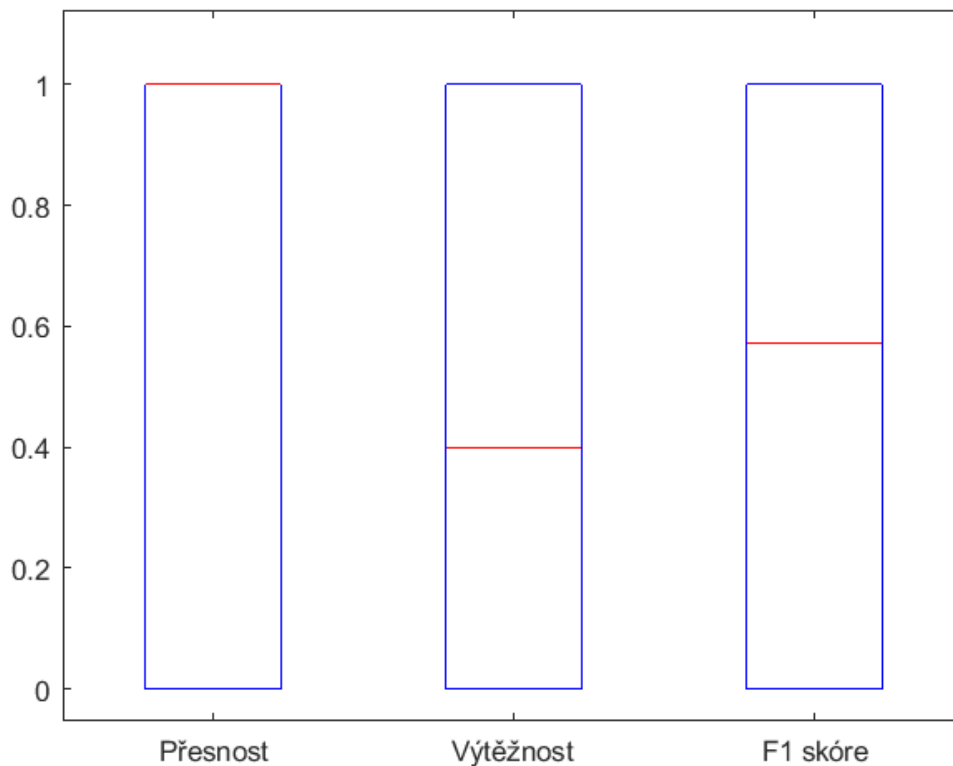
Detektor, který má nejnižší počet FP výsledků ze všech natrénovaných detektorů, má zároveň nejnižší průměrný počet FP výsledků na snímek, a to jak u snímků s výskytem cizích objektů, kde se průměrně vyskytuje 0,09 falešných detekcí na snímek, tak u snímků bez nich, kde se jich průměrně vyskytuje 0,08. Před podrobnějším zhodnocením bylo opět zvažováno, zda není vhodné zmenšit práh detekce, jelikož byl práh zkoumán pouze v rozmezí 0,5–0,6. Bylo ovšem rozhodnuto, že při snížení prahu by pravděpodobně vzhledem k velmi malému množství FP výsledků došlo hlavně ke snížení počtu TP výsledků, a proto bude práh ponechán na hodnotě 0,5. Další parametry použité při trénování a hodnotící parametry detekce jsou zobrazeny v Tab. 2.6, na Obr. 2.16 jsou pak zobrazeny boxploty pro přesnost, senzitivitu a F1 skóre na jednotlivých snímcích. Jak si lze všimnout, detektor sice obsahuje velice malé množství FP výsledků, dokázal ovšem zachytit pouze 424 z 951 cizích objektů, což není ani polovina. Vzhledem k tomu, že pro účely úkolu detekce cizích objektů v RTG snímcích hrudníku je důležitější detekovat cizí objekty i za cenu mírně vyššího množství FP výsledků, které se dají následně odebrat, je tento detektor pro využití v praxi nevhodný.

Na Obr. 2.17 jsou zobrazeny výsledky detekce pomocí tohoto detektoru. Žlutě jsou zobrazeny anotované objekty, zeleně pak objekty detekované. Dole jsou vidět snímky, na kterých je detekce úplně správná. V případě tohoto detektoru se jedná především o snímky, na kterých se nevyskytoval žádný cizí objekt. V horních dvou řádcích jsou pak vidět případy méně úspěšné detekce. Jak je vidět i na obrázcích, detekce selhává především v tom, že nedetekuje všechny cizí objekty ve snímcích. Výjimečný případ je vidět vlevo nahoře, kde jsou naopak detekovány tři oblasti, na ve kterých není anotován žádný cizí objekt.

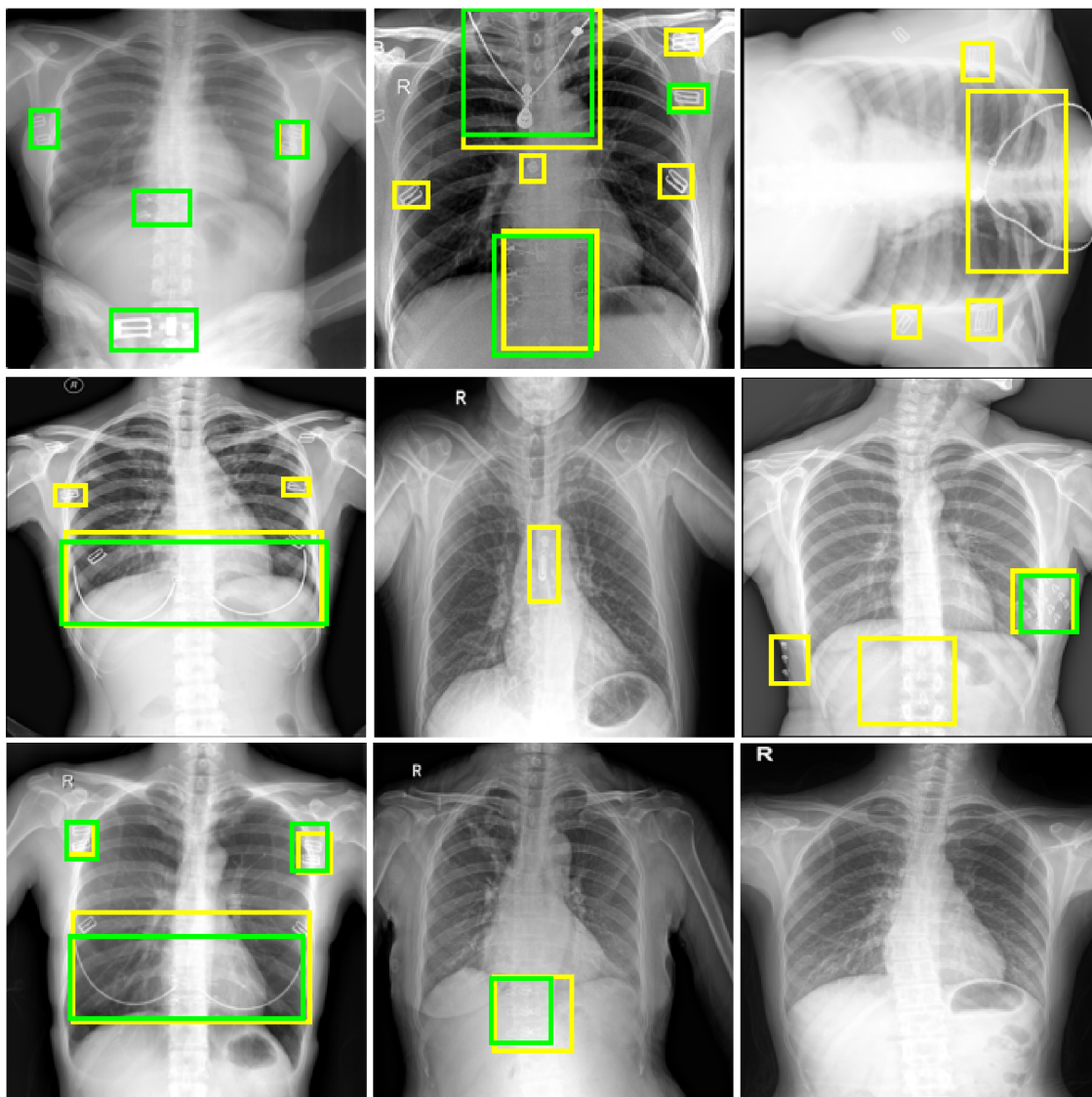
Jak lze vidět i na Obr. 2.17 vlevo nahoře, FP výsledky vznikají opět někdy i tím, že se označí oblasti, ve kterých cizí objekt je, ale není anotován. Bylo tedy provedeno podrobnější prohledání snímků a bylo zjištěno, že tento případ nastal celkem v 18 případech. Tyto detekované objekty byly tedy odstraněny z FP výsledků a byly přesunuty do TP výsledků a bylo provedeno nové vyhodnocení, kdy přesnost dosáhla 59 %, senzitivita 46 % a F1 skóre 51 %. I přes malé zlepšení výsledků ovšem detektor nenašel většinu cizích objektů ve snímcích, a proto je i tak nevhodný pro využití v praxi.

Tab. 2.6: Parametry trénování a vyhodnocení úspěšnosti detekce pro detektor s nej-  
nižším počtem FP výsledků.

Parametry trénování		Vyhodnocení detekce	
MaxEpochs	7	TP	424
MiniBatchSize	7	FP	44
InitialLearnRate	10e-5	FN	527
L2Regularization	10e-7	Průměrné FP s objekty	0,09
GradiendDecayFactor	0,95	Přesnost průměr	0,57
SquaredGradiendDecay	0,9	Přesnost sm. odchylka	0,49
NegativeOverlapRange	[0 0, 3]	Senzitivita průměr	0,45
PositiveOverlapRange	[0, 6 1]	Senzitivita sm. odchylka	0,43
NumRegionsToSample	400	F1 skóre průměr	0,50
Počet anchorů	7	F1 skóre sm. odchylka	0,44
Práh detekce	0,5	Průměrné FP bez objektů	0,08
		Časová náročnost	3,3 s



Obr. 2.16: Boxplot přesnosti, senzitivity a F1 skóre pro detektor s nejnižším počtem  
FP výsledků. Červená čára označuje medián, spodní hranice boxu značí první kvartil,  
horní hranice třetí kvartil, vousy označují krajní hodnotu bez započítání odlehlých  
hodnot, které jsou označeny zvlášť červeným křížkem (o odlehlou hodnotu se jedná v  
případě, kdy její vzdálenost od kraje boxu je alespoň  $1,5\times$  větší než je výška boxu).



Obr. 2.17: Zobrazení snímků, na nichž jsou detekovány cizí objekty pomocí detektoru s nejnižším počtem FP výsledků. Žlutě jsou zobrazeny anotované objekty, zeleně pak detekované objekty.

## Vyhodnocení nevhodnějšího detektoru

Po podrobnějším vyhodnocení některých natrénovaných detektorů se ukázalo, že detektor s nejnižším počtem FN výsledků není nevhodnější, jelikož obsahuje velké množství falešně pozitivních výsledků. I přes to, že se zdá být pro účel detekce cizích objektů v RTG snímcích hrudníků důležitější, aby byly nalezeny všechny cizí objekty i za cenu mírně vyššího počtu falešně detekovaných oblastí, v tomto případě je falešných oblastí nalezeno takové množství, že se snímky stávají nepřehlednými, a tedy nejsou pro využití v praxi vhodné. Také se ukázalo, že detektor s nejnižším počtem FP výsledků označuje sice velmi nízký počet oblastí, ve kterých se objekt nevyskytuje, ale zároveň vyhledá necelou polovinu cizích anotovaných objektů, tedy by bylo třeba snímky vždy znovu ručně prohledat. Tento detektor je pro využití v praxi tedy nevhodný. Naopak detektor s nejvyšším F1 skóre se ukázal být jako nevhodnější z výše zmiňovaných. Tento detektor nachází poměrně velké množství cizích anotovaných objektů a zároveň neoznačuje příliš mnoho oblastí, ve kterých se cizí objekt nevyskytuje.

Pokud vyjdeme tedy z kvality tohoto detektoru, kdy jeho přesnost dosáhla 75 %, senzitivita 77 % a F1 skóre 76 %, pak se můžeme podívat do tabulky A.1 a porovnat tyto hodnoty i s ostatními natrénovanými detektory. Jelikož požadujeme detektor, který bude hledat alespoň stejné množství cizích objektů jako detektor s nejvyšším F1 skóre, budeme v tabulce kontrolot pouze ty detektory, které mají celkový počet TP výsledků minimálně jako zmiňovaný. Takovýchto detektorů, které mají celkový počet TP výsledků vyšší než 729, je pouze sedm. Zároveň ale požadujeme, aby se počet FP nezvýšil příliš mnoho, např. můžeme stanovit maximální hodnotu na dvojnásobek zmiňovaného detektoru. Jelikož má detektor s nejvyšším F1 skóre celkem 225 FP detekcí, tato hranice je určena hodnotou 550. Pokud se podíváme na zmiňovaných sedm detektorů, které mají počet TP výsledků vyšší, zjistíme, že z nich má nejnižší počet FP detekcí detektor s 932 takovýchto detekcí, tzn. téměř čtyřnásobek oproti detektoru s nejvyšším F1 skóre. Po tomto zkontrolování tabulky se dá říct, že nejlepším natrénovaným detektorem je tedy opravdu detektor s nejvyšším F1 skóre. Tento detektor spolu s připravenými skripty lze stáhnout na webové stránce <https://github.com/b-matou/DP/tree/main>.

## 2.2.6 Možné zlepšení kvality detektoru

Nejlépe natrénovaný detektor v této diplomové práci dosáhl 75% přesnosti, 77% výtežnosti a 76% F1 skóre. Před využitím v praxi by bylo ovšem potřeba detekci cizích objektů v RTG snímcích hrudníku pomocí tohoto detektoru ještě vylepšit. V první řadě detektor ne vždy detekuje všechny cizí objekty ve snímku. Tento problém je zásadnější, jelikož pokud bude detektor vynechávat některé cizí objekty, bude potřeba

celý snímek podrobně prohlížet a vyhledávat tyto objekty. Druhým problémem je, že detektor někdy naopak označuje oblasti, ve kterých se cizí objekt nevyskytuje. Tento problém není tolik zásadní, pro využití v praxi by bylo pochopitelně vhodné tento parametr také zlepšit, ale odstranění některých detekovaných oblastí je méně náročné než prohledávat znovu celý snímek.

Pro zlepšení kvality detektoru by bylo pravděpodobně nejprve třeba přesně specifikovat oblast, kde se cizí objekty mají vyhledávat, a upravit podle toho anotace ve snímcích. V těchto datech bylo uvedeno, že se jedná o objekty pouze v oblasti plic, ale bylo na několika obrázcích ukázáno, že anotované objekty se v mnoha případech vyskytovaly mimo oblast plic, v jiných případech nebyly označeny objekty ani na okraji plic. Pokud by se určila jasná pravidla a snímky dle toho byly anotovány, mohlo by to výsledek detektoru poměrně zlepšit. V případě, že by se hledaly objekty pouze v oblasti plic, mohla by se využít i automatická segmentace plic. Prvním způsobem by bylo vysegmentované plíce vložit do jedné vrstvy obrazu, který přichází na vstup Faster R-CNN, druhou možností je detekované objekty na základě vysegmentovaných plic případně odebrat po průchodu sítí.

Dále by výsledek detekce mohlo zlepšit ještě větší množství trénovacích dat, případně by bylo možné trénovací data augmentovat. V případě, že by byl k dispozici pro trénování ještě výkonější počítač, mohly by se zkusit změnit některé parametry trénování, které jsou náročné na paměť. Mohlo by se jednat například o parametr velikosti minibatche či počtu náhodně generovaných oblastí zájmu v jednom obraze. Otázkou ovšem je, zda by změna těchto parametrů pomohla ke zlepšení kvality detektoru.

## Závěr

Tato diplomová práce se zabývá otázkou detekce cizích objektů v RTG snímcích hrudníku. Vyhledat tyto cizí objekty během automatického zpracování snímků je důležité kvůli zabránění falešně pozitivního výsledku některého z onemocnění. Vyhledání lékařsky voperovaných zařízení může mít také významnou roli. Tato diplomová práce si proto kladla za cíl natrénovat vhodný model pro vyhledávání těchto cizích objektů v RTG snímcích hrudníku.

Tato práce nejprve rozebírá téma z teoretického pohledu, kdy byly především provedeny rešerše dostupných metod pro vyhledávání cizích objektů v RTG snímcích hrudníku a následně je podrobněji popsán model Faster R-CNN pro detekci objektů v obrazech. Následně byl tento model prakticky aplikován na dostupná data.

Za účelem detekce cizích objektů v RTG snímcích hrudníku byl tedy natrénován Faster R-CNN model s předtrénovanou sítí ResNet50 jako extraktorem příznaků a s různými nastaveními učebních parametrů. Před samotným trénováním sítě bylo také potřeba provést úpravu snímků a jejich anotací, vybrat vhodný extraktor příznaků a velikosti anchorů. Během diplomové práce se nakonec podařilo natrénovat detektor, který na dostupných datech dosáhl 75% přesnosti, 77% senzitivity a 76% F1 skóre.

Při snaze vylepšit ještě více úspěšnost tohoto detektoru by bylo pravděpodobně třeba upravit anotace snímků, jelikož není jasně určená oblast, ve které se mají objekty detekovat. Dále by bylo pravděpodobně třeba využít k trénování ještě větší množství snímků a před uvedením do praxe pochopitelně detektor otestovat opět na větší databázi.



# Literatura

- [1] SEIDL, Zdeněk. *Radiologie pro studium i praxi*. 1. Praha: Grada, 2012. ISBN 978-80-247-4108-6.
- [2] MALÍKOVÁ, Hana. *Základy radiologie a zobrazovacích metod*. 1. Praha: Univerzita Karlova, nakladatelství Karolinum, 2019. ISBN 978-80-246-4036-5.
- [3] HOLIBKOVÁ, Alžběta a Stanislav LAICHMAN. *Přehled anatomie člověka*. 5. vyd. Olomouc: Univerzita Palackého v Olomouci, 2010. ISBN 978-80-244-2615-0.
- [4] MERKUNOVÁ, Alena a Miroslav OREL. *Anatomie a fyziologie člověka pro humanitní obory*. Praha: Grada, 2008. Psyché (Grada). ISBN 978-802-4715-216
- [5] FIALA, Pavel, Jiří VALENTA a Lada EBERLOVÁ. *Stručná anatomie člověka*. Praha: Univerzita Karlova v Praze, nakladatelství Karolinum, 2015. ISBN 978-802-4626-932.
- [6] *Breathe Easy: How radiology helps to find and fight lung diseases*. 1. Vienna: European Society of Radiology (ESR), 2013. ISBN 978-3950438802.
- [7] WIELPÜTZ, Mark O., Claus P. HEUSSEL, Felix J. F. HERTH a Hans-Ulrich KAUCZOR. Radiological Diagnosis in Lung Disease. *Deutsches Aerzteblatt Online* [online]. 2014, **111**(11), 181-187 [cit. 2020-12-14]. ISSN 1866-0452. Dostupné z: doi:10.3238/arztebl.2014.0181
- [8] CORTES, Corinna a Vladimir VAPNIK. Support-vector networks. *Machine Learning* [online]. 1995, **20**(3), 273-297 [cit. 2021-5-20]. ISSN 0885-6125. Dostupné z: doi:10.1007/BF00994018
- [9] FREUND, Yoav a SCHAPIRE, Robert. Experiments with a New Boosting. *International Conference on Machine Learning*, 1996, 148-156 Dostupné z: <<http://citeseerx.ist.psu.edu/viewdoc/summary?doi=10.1.1.51.6252>>
- [10] FELZENSZWALB, P F, R B GIRSHICK, D MCALLESTER a D RAMANAN. Object Detection with Discriminatively Trained Part-Based Models. *IEEE Transactions on Pattern Analysis and Machine Intelligence* [online]. 2010, **32**(9), 1627-1645 [cit. 2021-5-20]. ISSN 0162-8828. Dostupné z: doi:10.1109/TPAMI.2009.167
- [11] DHILLON, Anamika a Gyanendra K. VERMA. *Convolutional neural network: a review of models, methodologies and applications to object detection*. Progress

- in Artificial Intelligence [online].2020, **9**(2), 85-112 [cit. 2020-11-04]. ISSN 2192-6352. Dostupné z: doi:10.1007/s13748-019-00203-0
- [12] ZHAO, Zhong-Qiu, Peng ZHENG, Shou-Tao XU a Xindong WU. Object Detection With Deep Learning: A Review. *IEEE Transactions on Neural Networks and Learning Systems* [online]. 2019, **30**(11), 3212-3232 [cit. 2020-12-28]. ISSN 2162-237X. Dostupné z: doi:10.1109/TNNLS.2018.2876865
- [13] GIRSHICK, Ross, Jeff DONAHUE, Trevor DARRELL a Jitendra MALIK. Rich Feature Hierarchies for Accurate Object Detection and Semantic Segmentation. *2014 IEEE Conference on Computer Vision and Pattern Recognition* [online]. IEEE, 2014, 580-587 [cit. 2020-12-22]. ISBN 978-1-4799-5118-5. Dostupné z: doi:10.1109/CVPR.2014.81
- [14] GIRSHICK, Ross. Fast R-CNN. *2015 IEEE International Conference on Computer Vision (ICCV)* [online]. IEEE, 2015, 1440-1448 [cit. 2020-12-22]. ISBN 978-1-4673-8391-2. Dostupné z: doi:10.1109/ICCV.2015.169
- [15] REN, Shaoqing, Kaiming HE, Ross GIRSHICK a Jian SUN. Faster R-CNN: Towards Real-Time Object Detection with Region Proposal Networks. *IEEE Transactions on Pattern Analysis and Machine Intelligence* [online]. 2017, **39**(6), 1137-1149 [cit. 2020-12-22]. ISSN 0162-8828. Dostupné z: doi:10.1109/TPAMI.2016.2577031
- [16] HE, Kaiming, Georgia GKIOXARI, Piotr DOLLAR a Ross GIRSHICK. Mask R-CNN. *IEEE Transactions on Pattern Analysis and Machine Intelligence* [online]. 2020, **42**(2), 386-397 [cit. 2021-5-20]. ISSN 0162-8828. Dostupné z: doi:10.1109/TPAMI.2016.2577031
- [17] HE, Kaiming, Xiangyu ZHANG, Shaoqing REN a Jian SUN. Spatial Pyramid Pooling in Deep Convolutional Networks for Visual Recognition. *IEEE Transactions on Pattern Analysis and Machine Intelligence* [online]. 2015, **37**(9), 1904-1916 [cit. 2021-5-20]. ISSN 0162-8828. Dostupné z: doi:10.1109/TPAMI.2015.2389824
- [18] DAI, Jifeng, Yi LI, Kaiming HE a Jian SUN. R-FCN: Object Detection via Region-based Fully Convolutional Networks. *Advances in Neural Information Processing Systems 2016*, **29**, [cit. 2021-5-20]. Dostupné z: <<https://arxiv.org/abs/1605.06409>>.
- [19] LIN, Tsung-Yi, Piotr DOLLAR, Ross GIRSHICK, Kaiming HE, Bharath HARIHARAN a Serge BELONGIE. Feature Pyramid Networks for Object Detection. 2017 IEEE Conference on Computer Vision and Pattern Recognition

- (CVPR) [online]. IEEE, 2017, 2017, 936-944 [cit. 2021-5-20]. ISBN 978-1-5386-0457-1. Dostupné z: doi:10.1109/CVPR.2017.106
- [20] YOO, Donggeun, Sunggyun PARK, Joon-Young LEE, Anthony S. PAEK a In So KWEON. AttentionNet: Aggregating Weak Directions for Accurate Object Detection. *2015 IEEE International Conference on Computer Vision (ICCV)* [online]. IEEE, 2015, 2015, 2659-2667 [cit. 2021-5-20]. ISBN 978-1-4673-8391-2. Dostupné z: doi:10.1109/ICCV.2015.305
- [21] NAJIBI, Mahyar, Mohammad RASTEGARI a Larry S. DAVIS. G-CNN: An Iterative Grid Based Object Detector. *2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)* [online]. IEEE, 2016, 2016, 2369-2377 [cit. 2021-5-20]. ISBN 978-1-4673-8851-1. Dostupné z: doi:10.1109/CVPR.2016.260
- [22] LIU, Wei, Dragomir ANGUELOV, Dumitru ERHAN, Christian SZEGEDY, Scott REED, Cheng-Yang FU a Alexander C. BERG. SSD: Single Shot MultiBox Detector. *Computer Vision – ECCV 2016* [online]. Cham: Springer International Publishing, 2016, 2016-09-17, 21-37 [cit. 2021-5-20]. Lecture Notes in Computer Science. ISBN 978-3-319-46447-3. Dostupné z: doi:10.1007/978-3-319-46448-0\_2
- [23] FU, Cheng-Yang, Wei LIU, Ananth RANGA, Ambrish TYAGI a Alexander C. BERG. DSSD: Deconvolutional Single Shot Detector [online]. 2017 [cit. 2020-10-31]. Dostupné z: <<https://arxiv.org/abs/1701.06659>>.
- [24] REDMON, Joseph, Santosh DIVVALA, Ross GIRSHICK a Ali FARHADI. You Only Look Once: Unified, Real-Time Object Detection. *2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)* [online]. IEEE, 2016, 2016, 779-788 [cit. 2021-5-20]. ISBN 978-1-4673-8851-1. Dostupné z: doi:10.1109/CVPR.2016.91
- [25] SHEN, Zhiqiang, Zhuang LIU, Jianguo LI, Yu-Gang JIANG, Yurong CHEN a Xiangyang XUE. DSOD: Learning Deeply Supervised Object Detectors from Scratch. *2017 IEEE International Conference on Computer Vision (ICCV)* [online]. IEEE, 2017, 2017, 1937-1945 [cit. 2021-5-20]. ISBN 978-1-5386-1032-9. Dostupné z: doi:10.1109/ICCV.2017.212
- [26] ZOHORA, Fatema Tuz a K. C. SANTOSH. *Circular Foreign Object Detection in Chest X-ray Images*. Recent Trends in Image Processing and Pattern Recognition [online]. Singapore: Springer Singapore, 2017, 2017-04-29, **709**, 391-401

- [cit. 2020-10-31]. Communications in Computer and Information Science. ISBN 978-981-10-4858-6. Dostupné z: doi:10.1007/978-981-10-4859-3\_35
- [27] CANDEMIR, Sema, Stefan JAEGER, Kannappan PALANIAPPAN, et al. *Lung Segmentation in Chest Radiographs Using Anatomical Atlases With Nonrigid Registration*. IEEE Transactions on Medical Imaging [online]. 2014, **33**(2), 577-590 [cit. 2020-11-27]. ISSN 0278-0062. Dostupné z: doi:10.1109/TMI.2013.2290491
- [28] ZOHORA, Fatema Tuz a K.C. SANTOSH. *Foreign Circular Element Detection in Chest X-Rays for Effective Automated Pulmonary Abnormality Screening*. International Journal of Computer Vision and Image Processing [online]. 2017, **7**(2), 36-49 [cit. 2020-11-01]. ISSN 2155-6997. Dostupné z: doi:10.4018/IJCVIP.2017040103
- [29] XUE, Zhiyun, Sema CANDEMIR, Sameer ANTANI, L. Rodney LONG, Stefan JAEGER, Dina DEMNER-FUSHMAN a George R. THOMA. *Foreign object detection in chest X-rays*. 2015 IEEE International Conference on Bioinformatics and Biomedicine (BIBM) [online]. IEEE, 2015, 2015, 956-961 [cit. 2020-10-31]. ISBN 978-1-4673-6799-8. Dostupné z: doi:10.1109/BIBM.2015.7359812
- [30] Chest X-rays Dataset (Indiana University) [cit. 2020-10-31]. Dostupné z: <<https://www.kaggle.com/raddar/chest-xrays-indiana-university>>.
- [31] SANTOSH, K.C., Mrinal K DHAR, Ramina RAJBHANDARI a Amul NEUPANE. *Deep Neural Network for Foreign Object Detection in Chest X-Rays*. 2020 IEEE 33rd International Symposium on Computer-Based Medical Systems (CBMS) [online]. IEEE, 2020, 2020, 538-541 [cit. 2020-10-31]. ISBN 978-1-7281-9429-5. Dostupné z: doi:10.1109/CBMS49503.2020.00107
- [32] HOGEWEG, Laurens, Clara I. SÁNCHEZ, Jaime MELENDEZ, Pragnya MADUSKAR, Alistair STORY, Andrew HAYWARD a Bram VAN GINNEKEN. *Foreign object detection and removal to improve automated analysis of chest radiographs*. Medical Physics [online]. 2013, **40**(7) [cit. 2020-10-31]. ISSN 00942405. Dostupné z: doi:10.1118/1.4805104
- [33] FRIEDMAN, Jerome, Trevor HASTIE a Robert TIBSHIRANI. Additive logistic regression: a statistical view of boosting (With discussion and a rejoinder by the authors). *The Annals of Statistics* [online]. 2000, **28**(2) [cit. 2021-5-20]. ISSN 0090-5364. Dostupné z: doi:10.1214/aos/1016218223
- [34] DESHPANDE, Hrishikesh, Tim HARDER, Axel SAALBACH, Abhivyakti SAWARKAR a Thomas BUELOW. *Detection Of Foreign Objects In Chest*

- Radiographs Using Deep Learning*. 2020 IEEE 17th International Symposium on Biomedical Imaging Workshops (ISBI Workshops) [online]. IEEE, 2020, 2020, s. 1-4 [cit. 2020-10-31]. ISBN 978-1-7281-7401-3. Dostupné z: doi:10.1109/ISBIWorkshops50223.2020.9153350
- [35] HUANG, Gao, Zhuang LIU, Laurens VAN DER MAATEN a Kilian Q. WEINBERGER. Densely Connected Convolutional Networks. *2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)* [online]. IEEE, 2017, 2017, 2261-2269 [cit. 2021-5-20]. ISBN 978-1-5386-0457-1. Dostupné z: doi:10.1109/CVPR.2017.243
- [36] HE, Kaiming, Xiangyu ZHANG, Shaoqing REN a Jian SUN. Deep Residual Learning for Image Recognition. *2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)* [online]. IEEE, 2016, 2016, 770-778 [cit. 2021-5-20]. ISBN 978-1-4673-8851-1. Dostupné z: doi:10.1109/CVPR.2016.90
- [37] DENG, Jia, Wei DONG, Richard SOCHER, Li-Jia LI, KAI LI a LI FEI-FEI. ImageNet: A large-scale hierarchical image database. *2009 IEEE Conference on Computer Vision and Pattern Recognition* [online]. IEEE, 2009, 2009, 248-255 [cit. 2021-5-20]. ISBN 978-1-4244-3992-8. Dostupné z: doi:10.1109/CVPR.2009.5206848
- [38] WANG, Xiaosong, Yifan PENG, Le LU, Zhiyong LU, Mohammadhadi BAGHERI a Ronald M. SUMMERS. ChestX-Ray8: Hospital-Scale Chest X-Ray Database and Benchmarks on Weakly-Supervised Classification and Localization of Common Thorax Diseases. *2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)* [online]. IEEE, 2017, 2017, 3462-3471 [cit. 2021-5-20]. ISBN 978-1-5386-0457-1. Dostupné z: doi:10.1109/CVPR.2017.369
- [39] JOHNSON, Alistair, Matt LUNGREN, Yifan PENG, Lu ZHIYONG, Roger MARK, Seth BERKOWITZ a Steven HORNG. MIMIC-CXR-JPG - chest radiographs with structured labels (version 2.0.0). *PhysioNet*. 2019 [cit. 2021-5-20]. Dostupné z: <<https://doi.org/10.13026/8360-t248>>
- [40] LE, Hieu X., Phuong D. NGUYEN, Thang H. NGUYEN a Khanh N.Q. LE. *A novel approach to remove foreign objects from chest X-ray images* [online]. 2020 [cit. 2020-10-31]. Dostupné z: <<https://arxiv.org/abs/2008.06828>>
- [41] PHILLIPS, Nick A., Pranav RAJPURKAR, Mark SABINI, Rayan KRISHNAN, Sharon ZHOU, Anuj PAREEK, Nguyet Minh PHU, Chris WANG, Mudit JAIN, Nguyen Duong DU, Steven QH TRUONG, Andrew Y.

- NGa Matthew P. LUNGREN. CheXphoto: 10,000+ Photos and Transformations of Chest X-rays for Benchmarking Deep Learning Robustness. *Proceedings of the Machine Learning for Health NeurIPS Workshop* [online]. 2020 [cit. 2020-10-31]. Dostupné z: <<https://arxiv.org/abs/2007.06199>>
- [42] BOKCHKOVSKIY, Alexey, Chien-Yao WANG a Hong-Yuan Mark LIAO. *YOLOv4: Optimal Speed and Accuracy of Object Detection* [online]. 2020 [cit. 2020-10-31]. Dostupné z: <<https://arxiv.org/abs/2004.10934>>
- [43] RONNEBERGER, Olaf, Philipp FISCHER a Thomas BROX. U-Net: Convolutional Networks for Biomedical Image Segmentation. *Medical Image Computing and Computer-Assisted Intervention – MICCAI 2015* [online]. Cham: Springer International Publishing, 2015, 2015-11-18, 234-241 [cit. 2021-5-20]. Lecture Notes in Computer Science. ISBN 978-3-319-24573-7. Dostupné z: doi:10.1007/978-3-319-24574-4\_28
- [44] *Advanced neural network solution for detection of lung pathology and foreign body on chest plain radiographs*. *Imaging in Medicine* [online]. 2019, 2019, **11**(6), 57-66 [cit. 2020-10-31]. ISSN 1755-5191. Dostupné z: <<https://www.openaccessjournals.com/articles/advanced-neural-network-solution-for-detection-of-lung-pathology-and-foreign-body-on-chest-plain-radiographs-13104.html>>
- [45] CHEN, Tianqi a Carlos GUESTRIN. XGBoost. *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining* [online]. New York, NY, USA: ACM, 2016, 2016-08-13, 785-794 [cit. 2021-5-20]. ISBN 9781450342322. Dostupné z: doi:10.1145/2939672.2939785
- [46] SZEGEDY, Christian, Vincent VANHOUCKE, Sergey IOFFE, Jon SHLENS a Zbigniew WOJNA. Rethinking the Inception Architecture for Computer Vision. *2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)* [online]. IEEE, 2016, 2016, 2818-2826 [cit. 2021-5-20]. ISBN 978-1-4673-8851-1. Dostupné z: doi:10.1109/CVPR.2016.308
- [47] KINGMA, Diederik P. a Jimmy Ba JIMMY BA. *Adam: A Method for Stochastic Optimization* [online]. 2014 [cit. 2021-5-20]. Dostupné z: <<https://arxiv.org/pdf/1412.6980.pdf>>
- [48] RAY, Javier. Faster R-CNN: Down the rabbit hole of modern object detection. *Tryolabs* [online]. 2018 [cit. 2020-12-22]. Dostupné z: <<https://tryolabs.com/blog/2018/01/18/faster-r-cnn-down-the-rabbit-hole-of-modern-object-detection/>>

- [49] Faster R-CNN. *Alegion* [online]. Meduim, 2019 [cit. 2020-12-22]. Dostupné z: <https://www.alegion.com/faster-r-cnn>
- [50] SIMONYAN, Karen, Andrew ZISSERMAN, Kaiming HE a Jian SUN. *Very Deep Convolutional Networks for Large-Scale Image Recognition* [online]. 2014 [cit. 2021-5-20]. Dostupné z: <https://arxiv.org/pdf/1409.1556v6.pdf>
- [51] ZEILER, Matthew D. a Rob FERGUS. Visualizing and Understanding Convolutional Networks. *Computer Vision – ECCV 2014* [online]. Cham: Springer International Publishing, 2014, 2014, 818-833 [cit. 2021-5-20]. Lecture Notes in Computer Science. ISBN 978-3-319-10589-5. Dostupné z: doi:10.1007/978-3-319-10590-1\_53
- [52] ANANTH, Shilpa. Faster R-CNN for object detection. *Towards Data Science* [online]. Meduim, 2019 [cit. 2020-12-22]. Dostupné z: <https://towardsdatascience.com/faster-r-cnn-for-object-detection-a-technical-summary-474c5b857b46>
- [53] *Object-CXR: Automatic detection of foreign objects on chest X-rays* [online]. GitHub [cit. 2020-11-17]. Dostupné z: <https://jfhealthcare.github.io/object-CXR/>
- [54] ZUIDERVELD, Karel. Contrast limited adaptive histogram equalization. *Graphics gems IV* [online]. Academic Press Professional, 1994, 474-485 [cit. 2021-5-20]. Dostupné z: <https://dl.acm.org/doi/10.5555/180895.180940>
- [55] *TrainFasterRCNNObjectDetector*. *MathWorks: Help Center* [online]. 2020 [cit. 2020-11-26]. Dostupné z: <https://www.mathworks.com/help/vision/ref/trainfasterrcnnobjectdetector.html>
- [56] *Object Detection Using Faster R-CNN Deep Learning*. *MathWorks: Help Center* [online]. 2020 [cit. 2020-1-2]. Dostupné z: <https://www.mathworks.com/help/vision/ug/object-detection-using-faster-r-cnn-deep-learning.html>
- [57] Deep Learning v prostředí Matlab. *Automa* [online]. 2017, **23**(5), 12-14 [cit. 2021-5-9]. ISSN 1210-9592. Dostupné z: <https://automa.cz/page-flip/casopis/automa/2017/05/index.html#page/14>
- [58] KRIZHEVSKY, Alex, Ilya SUTSKEVER a Geoffrey E. HINTON. ImageNet classification with deep convolutional neural networks. *Communications of the ACM* [online]. 2017, **60**(6), 84-90 [cit. 2021-5-20]. ISSN 0001-0782. Dostupné z: doi:10.1145/3065386

- [59] SZEGEDY, Christian, WEI LIU, YANGQING JIA, et al. Going deeper with convolutions. *2015 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)* [online]. IEEE, 2015, 2015, 1-9 [cit. 2021-5-21]. ISBN 978-1-4673-6964-0. Dostupné z: doi:10.1109/CVPR.2015.7298594
- [60] REDMONT, Joseph. Darknet: Open Source Neural Networks in C [online]. 2013–2016 [cit. 2021-5-20]. Dostupné z: <<http://pjreddie.com/darknet/>>
- [61] *Pretrained Deep Learning Models. MathWorks: Deep Learning* [online]. 2020 [cit. 2020-5-9]. Dostupné z: <<https://www.mathworks.com/solutions/deep-learning/models.html>>
- [62] *estimateAnchorBoxes. MathWorks: Help Center* [online]. 2020 [cit. 2020-5-6]. Dostupné z: <[https://www.mathworks.com/help/vision/ref/estimateanchorboxes.html?s\\_tid=doc\\_ta](https://www.mathworks.com/help/vision/ref/estimateanchorboxes.html?s_tid=doc_ta)>



## Seznam symbolů, veličin a zkratek

<b>ADAM</b>	<i>angl. Adaptive Moment Estimation</i>
<b>AUC</b>	<i>angl. Area Under the Curve</i>
<b>CHT</b>	kruhová Houghova transformace, <i>angl. Circular Hough Transformation</i>
<b>CLAHE</b>	adaptivní ekvalizace histogramu s omezeným kontrastem, <i>angl. Contrast Limited Adaptive Histogram Equalization</i>
<b>CNN</b>	konvoluční neuronová síť, <i>angl. Convolution Neural Network</i>
<b>CT</b>	výpočetní tomografie, <i>angl. Computed Tomography</i>
<b>CXR</b>	rentgenový snímek hrudníku, <i>angl. Chest X-Ray</i>
<b>DPM</b>	<i>angl. Deformable Part Models</i>
<b>DSOD</b>	<i>angl. Deeply Supervised Object Detector</i>
<b>DSSD</b>	<i>angl. Deconvolutional Single Shot Detector</i>
<b>FC</b>	plně propojená, <i>angl. Fully Connected</i>
<b>FN</b>	falešně negativní, <i>angl. False Negative</i>
<b>FP</b>	falešně pozitivní, <i>angl. False Positive</i>
<b>FPN</b>	<i>angl. Feature Pyramid Networks</i>
<b>G-CNN</b>	<i>angl. Grid Convolutional Neural Network</i>
<b>IoU</b>	<i>angl. Intersection over Union</i>
<b>LDA</b>	lineární diskriminační analýza, <i>angl. Linear Discriminant Analysis</i>
<b>MRI</b>	magnetická rezonance, <i>angl. Magnetic Resonance Imaging</i>
<b>ReLU</b>	zkorigovaná lineární jednotka, <i>angl. Rectified Linear Units</i>
<b>RMSPROP</b>	<i>angl. Root Mean Square Propagation</i>
<b>ROI</b>	zájmová oblast, <i>angl. Region of Interest</i>
<b>RPN</b>	<i>angl. Region Proposal Network</i>
<b>RTG</b>	rentgenový, rentgen

<b>R-CNN</b>	<i>angl. Region Based Convolutional Neural Networks)</i>
<b>R-FCN</b>	<i>angl. Region-based Fully Convolutional Networks</i>
<b>SGD</b>	stochastický gradientní sestup, <i>angl. Stochastic Gradient Descend</i>
<b>SGDM</b>	stochastický gradientní sestup s momentem, <i>angl. Stochastic Gradient Descend with Momentum</i>
<b>SSD</b>	<i>angl. Single Shot Detector</i>
<b>SSP-net</b>	<i>angl. Spatial Pyramid Pooling</i>
<b>SVM</b>	metoda podpůrných vektorů, <i>angl. Support Vector Machine</i>
<b>TP</b>	správně pozitivní, <i>angl. True Positive</i>
<b>XGBoost</b>	<i>angl. eXtreme Gradient Boosting</i>
<b>YOLO</b>	<i>angl. You Only Look Once</i>

## **A Tabulka výsledků**

Tab. A.1: Použité parametry pro trénování detektorů, práh pro detekci a výsledky. Ve výsledcích je v každém sloupci tučně označena nejlepší hodnota pro daný parametr.

Parametry trénování Faster R-CNN											Výsledky pro obrazy s objekty							bez objektů
MaxEpochs	MiniBatchSize	InitialLearnRate	L2Regularization	GradientDecayFactor	SquaredGradientDecay	NegativeOverlapRange	PositiveOverlapRange	NumRegionsToSample	Počet anchorů	Práh detekce	TP	FP	FN	Průměrné FP	Přesnost	Senzitivita	F1 skóre	Průměrné FP
5	7	10e-5	10e-7	0,95	0,9	[0 0, 3]	[0, 3 1]	400	7	0,5	729	225	222	0,45	<b>0,72</b>	0,76	<b>0,74</b>	0,78
7	7	10e-5	10e-7	0,95	0,9	[0 0, 3]	[0, 3 1]	400	7	0,5	702	210	249	0,42	0,70	0,73	0,72	0,61
9	7	10e-5	10e-7	0,95	0,9	[0 0, 3]	[0, 3 1]	400	7	0,5	685	178	266	0,35	0,70	0,72	0,71	0,53
10	4	10e-5	10e-7	0,9	0,99	[0 0, 3]	[0, 6 1]	300	4	0,5	595	137	356	0,27	0,65	0,62	0,64	0,39
9	7	10e-5	10e-7	0,95	0,9	[0, 05 0, 3]	[0, 6 1]	400	7	0,5	653	390	298	0,78	0,59	0,70	0,64	1,03
5	7	10e-5	10e-7	0,95	0,9	[0, 05 0, 3]	[0, 3 1]	400	7	0,5	813	932	138	1,86	0,50	0,86	0,63	2,88
7	8	10e-5	10e-7	0,9	0,99	[0, 05 0, 5]	[0, 5 1]	500	7	0,5	651	508	300	1,02	0,56	0,69	0,62	1,44
7	7	10e-5	10e-7	0,95	0,9	[0, 05 0, 3]	[0, 3 1]	400	7	0,5	832	1018	119	2,04	0,48	0,89	0,62	2,97
20	4	10e-5	10e-7	0,9	0,99	[0 0, 3]	[0, 6 1]	300	4	0,5	557	108	394	0,22	0,64	0,59	0,61	0,28
11	7	10e-5	10e-7	0,95	0,9	[0 0, 3]	[0, 6 1]	400	7	0,5	532	90	419	0,18	0,64	0,56	0,60	0,23
30	4	10e-5	10e-7	0,9	0,99	[0 0, 3]	[0, 6 1]	300	4	0,5	549	125	402	0,25	0,63	0,57	0,60	0,30

Parametry trénování Faster R-CNN											Výsledky pro obrazy s objekty						bez objektů	
MaxEpochs	MiniBatchSize	InitialLearnRate	L2Regularization	GradientDecayFactor	SquaredGradientDecay	NegativeOverlapRange	PositiveOverlapRange	NumRegionsToSample	Počet anchorů	Práh detekce	TP	FP	FN	Průměrné FP	Přesnost	Senzitivita	F1 skóre	Průměrné FP
13	7	10e-5	10e-7	0,95	0,9	[0 0, 3]	[0, 6 1]	400	7	0,5	507	68	444	0,14	0,65	0,54	0,59	0,18
5	7	10e-5	10e-7	0,95	0,9	[0, 05 0, 3]	[0, 6 1]	400	7	0,5	600	362	351	0,72	0,56	0,64	0,59	0,90
5	8	10e-5	10e-7	0,9	0,99	[0, 05 0, 5]	[0, 5 1]	600	7	0,5	549	358	402	0,72	0,57	0,61	0,59	1,02
10	8	5·10e-6	10e-7	0,9	0,99	[0, 05 0, 5]	[0, 5 1]	400	7	0,5	679	683	272	1,37	0,50	0,71	0,59	1,58
12	7	5·10e-6	10e-7	0,9	0,99	[0, 05 0, 5]	[0, 5 1]	400	7	0,5	636	550	315	1,10	0,53	0,67	0,59	1,62
5	7	5·10e-6	10e-7	0,95	0,9	[0, 05 0, 5]	[0, 5 1]	400	7	0,5	570	305	381	0,61	0,56	0,60	0,58	0,81
7	8	5·10e-6	10e-7	0,9	0,99	[0, 05 0, 5]	[0, 5 1]	400	7	0,5	609	484	342	0,97	0,53	0,65	0,58	1,24
16	7	5·10e-6	10e-7	0,9	0,99	[0, 05 0, 5]	[0, 5 1]	400	7	0,5	625	525	326	1,05	0,51	0,66	0,58	1,45
15	9	10e-5	10e-7	0,9	0,99	[0, 05 0, 3]	[0, 3 1]	200	7	0,5	796	1117	155	2,23	0,44	0,84	0,58	3,06
5	7	10e-5	10e-7	0,95	0,9	[0 0, 3]	[0, 6 1]	400	7	0,5	499	73	452	0,15	0,61	0,52	0,57	0,17
5	7	5·10e-6	10e-7	0,95	0,85	[0, 05 0, 5]	[0, 5 1]	400	7	0,5	521	197	430	0,39	0,59	0,56	0,57	0,48
7	7	10e-5	10e-7	0,95	0,9	[0, 05 0, 3]	[0, 6 1]	400	7	0,5	538	204	413	0,41	0,57	0,57	0,57	0,50
15	10	10e-5	10e-7	0,9	0,99	[0, 05 0, 5]	[0, 5 1]	200	7	0,5	550	366	401	0,73	0,56	0,59	0,57	0,90
5	7	5·10e-6	10e-7	0,7	0,99	[0, 05 0, 5]	[0, 5 1]	400	7	0,5	546	380	405	0,76	0,55	0,59	0,57	1,05

Parametry trénování Faster R-CNN											Výsledky pro obrazy s objekty						bez objektů	
MaxEpochs	MiniBatchSize	InitialLearnRate	L2Regularization	GradientDecayFactor	SquaredGradientDecay	NegativeOverlapRange	PositiveOverlapRange	NumRegionsToSample	Počet anchorů	Práh detekce	TP	FP	FN	Průměrné FP	Přesnost	Senzitivita	F1 skóre	Průměrné FP
14	7	5·10e-6	10e-7	0,9	0,99	[0,05 0,5]	[0,5 1]	400	7	0,5	575	418	376	0,84	0,53	0,61	0,57	1,18
5	7	5·10e-5	10e-7	0,9	0,99	[0,05 0,4]	[0,4 1]	400	7	0,5	588	471	363	0,94	0,53	0,63	0,57	1,26
7	8	5·10e-5	10e-7	0,9	0,99	[0,05 0,5]	[0,5 1]	400	7	0,5	564	457	387	0,91	0,52	0,62	0,57	1,33
9	7	10e-5	10e-7	0,95	0,9	[0 0,3]	[0,6 1]	400	7	0,5	488	72	463	0,14	0,61	0,52	0,56	0,19
5	7	5·10e-6	10e-7	0,95	0,8	[0,05 0,5]	[0,5 1]	400	7	0,5	547	299	404	0,60	0,55	0,58	0,56	0,75
5	7	10e-6	10e-6	0,95	0,9	[0,05 0,3]	[0,3 1]	400	7	0,55	760	1130	191	2,26	0,43	0,80	0,56	3,45
9	7	10e-5	10e-7	0,95	0,9	[0,05 0,3]	[0,3 1]	400	7	0,6	<b>851</b>	1308	<b>100</b>	2,62	0,41	<b>0,91</b>	0,56	3,88
5	7	5·10e-6	10e-7	0,8	0,99	[0,05 0,5]	[0,5 1]	400	7	0,5	505	299	446	0,60	0,55	0,55	0,55	0,77
5	7	5·10e-6	10e-7	0,95	0,99	[0,05 0,5]	[0,5 1]	400	7	0,5	551	377	400	0,75	0,52	0,58	0,55	0,95
7	7	5·10e-6	10e-7	0,9	0,99	[0,05 0,4]	[0,4 1]	400	7	0,5	616	659	335	1,32	0,47	0,67	0,55	1,77
10	7	5·10e-6	10e-7	0,9	0,99	[0,05 0,5]	[0,5 1]	400	7	0,5	636	732	315	1,46	0,46	0,67	0,55	1,94
30	6	10e-5	10e-7	0,9	0,99	[0,05 0,5]	[0,5 1]	200	7	0,5	614	705	337	1,41	0,47	0,65	0,54	1,69
20	6	10e-5	10e-7	0,9	0,99	[0,05 0,5]	[0,5 1]	200	4	0,5	560	556	391	1,11	0,48	0,60	0,53	1,55
5	7	5·10e-6	10e-7	0,95	0,95	[0,05 0,5]	[0,5 1]	400	7	0,5	455	157	496	0,31	0,55	0,49	0,52	0,39

Parametry trénování Faster R-CNN											Výsledky pro obrazy s objekty						bez objektů	
MaxEpochs	MiniBatchSize	InitialLearnRate	L2Regularization	GradientDecayFactor	SquaredGradientDecay	NegativeOverlapRange	PositiveOverlapRange	NumRegionsToSample	Počet anchorů	Práh detekce	TP	FP	FN	Průměrné FP	Přesnost	Senzitivita	F1 skóre	Průměrné FP
5	5	10e-6	10e-7	0,9	0,99	[0,05 0,3]	[0,5 1]	300	4	0,55	750	1362	201	2,72	0,38	0,81	0,52	3,29
7	7	10e-5	10e-7	0,95	0,9	[0 0,3]	[0,6 1]	400	7	0,5	424	<b>44</b>	527	<b>0,09</b>	0,57	0,45	0,50	<b>0,08</b>
5	7	5·10e-6	10e-7	0,9	0,99	[0,05 0,5]	[0,5 1]	400	7	0,5	441	182	510	0,36	0,53	0,47	0,50	0,40
5	7	10e-5	10e-7	0,9	0,99	[0,05 0,3]	[0,5 1]	300	4	0,6	815	1661	136	3,32	0,35	0,87	0,50	4,65
10	6	10e-6	10e-10	0,9	0,99	[0,05 0,6]	[0,6 1]	200	4	0,5	436	948	515	1,90	0,34	0,49	0,40	2,61