



Zadání bakalářské práce

Název:	Snímání a následná detekce a klasifikace vad skleněných tyčí
Student:	Matěj Latka
Vedoucí:	Ing. Jakub Novák
Studijní program:	Informatika
Obor / specializace:	Znalostní inženýrství
Katedra:	Katedra aplikované matematiky
Platnost zadání:	do konce letního semestru 2022/2023

Pokyny pro vypracování

Cílem práce je detekce a klasifikace povrchových vad skleněných tyčí.

Skleněná tyč se používá jako polotovar například pro frézování čoček. Při výrobě se však mohou vyskytnout vady typu škrábanec, bublina, různé nerovnosti, atd. Práce se zaměřuje na metody detekce a klasifikace vad s použitím metod zpracování obrazu. Práce předpokládá využití pokročilých optických prvků a osvětlení (např. řádkové kamery, UV osvětlení, apod.).

Úkoly:

- 1) Proveďte rešerši v oblasti metod snímání a detekce vad podobných materiálů.
- 2) Sestavte měřicí kamerový systém pro snímání materiálu.
- 3) Vytvořte metodiku měření včetně specifikace vad materiálu.
- 4) Navrhněte algoritmy vyhodnocení vad pomocí metod zpracování obrazu.
- 5) Otestujte výsledky navržených algoritmů a zhodnoťte úspěšnost.
- 6) Vizualizujte výsledky algoritmu na obrazových datech.



**FAKULTA
INFORMAČNÍCH
TECHNOLÓGIÍ
ČVUT V PRAZE**

Bakalářská práce

Snímání a následná detekce a klasifikace vad skleněných tyčí

Matěj Latka

Katedra aplikované matematiky

Vedoucí práce: Ing. Jakub Novák

13. května 2021

Poděkování

Předně děkuji Ing. Jakubu Novákovi za jeho čas, trpělivost, cenné rady a nekonečnou ochotu k diskuzím na téma práce. Dále bych rád poděkoval Bc. Alexandru Bazkovi za hardwarovou podporu a konstrukci speciálního osvětlení, Bc. Petru Kasalickému za pomoc při práci s fakultními servery a všem pracovníkům ImproLabu za poskytnutí potřebného vybavení. Díky patří i mým přátelům z CHS, rodině a příteli za motivaci a podporu během studia.

Prohlášení

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

Beru na vědomí, že se na moji práci vztahují práva a povinnosti vyplývající ze zákona č. 121/2000 Sb., autorského zákona, ve znění pozdějších předpisů, zejména skutečnost, že České vysoké učení technické v Praze má právo na uzavření licenční smlouvy o užití této práce jako školního díla podle § 60 odst. 1 citovaného zákona.

V Praze dne 13. května 2021

.....

České vysoké učení technické v Praze

Fakulta informačních technologií

© 2021 Matěj Latka. Všechna práva vyhrazena.

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

Odkaz na tuto práci

Latka, Matěj. *Snímání a následná detekce a klasifikace vad skleněných tyčí*. Bakalářská práce. Praha: České vysoké učení technické v Praze, Fakulta informačních technologií, 2021.

Abstrakt

Práce se zabývá automatizovanou detekcí a klasifikací vad na skleněných tyčích. Jsou analyzována předchozí řešení podobných problémů a popsán vlastní postup, v rámci kterého byly navrženy čtyři snímací soustavy využívající pokročilých optických prvků a osvětlení. Dále byly upraveny a rozšířeny analyzované metody detekce a klasifikace vad. Druhá z nich, využívající model založený na moderní architektuře Faster R-CNN, detekuje správně na jednom z datasetů 83 % vad a správně klasifikuje 77 % z nich.

Klíčová slova kamerový systém, zpracování obrazu, snímání skla, kontury, tvarové deskriptory, strojové učení, detekce objektů, neuronová síť, Mask R-CNN

Abstract

This thesis focuses on the automated detection and classification of defects on glass rods. Previous solutions of similar problems are analysed and a new approach is proposed. Four camera systems using advanced optical components and lighting are designed. Also, analysed defect detection and classification methods are modified and extended. One of them using a model based on the modern Faster R-CNN architecture is able to detect 83 % of defects and to correctly classify 77 % of them.

Keywords camera system, image processing, glass capturing, contours, geometric features, machine learning, object detection, neural network, Mask R-CNN

Obsah

Úvod	1
1 Rešerše	3
1.1 Optické skleněné tyče	3
1.2 Vady optických skleněných tyčí	4
1.2.1 Bubliny	4
1.2.2 Kamínky	4
1.2.3 Škrábance	5
1.2.4 Smirková čára	5
1.2.5 Nerovnost okraje	5
1.3 Systémy pro snímání a detekci vad obdobných materiálů	7
1.3.1 Detekce a klasifikace defektů pomocí tvarových deskriptorů	7
1.3.2 Detekce vad displejů mobilních telefonů	9
1.3.3 Detekce vad skleněných panelů	9
2 Teoretické zázemí	13
2.1 Osvětlení ve strojovém vidění	13
2.1.1 Výběr osvětlení v závislosti na vlastnostech snímaného objektu	13
2.1.2 Osvětlení mimo viditelné spektrum	14
2.2 Metody zpracování obrazu	16
2.2.1 Segmentace	16
2.2.2 Mediánový filtr	16
2.2.3 High-boost filtr	17
2.2.4 Cannyho detektor	18
2.2.5 Základní morfologické operace	18
2.3 Neuronové sítě	19
2.3.1 Jednovrstvý perceptron	20

2.3.2	Vícevrstvé neuronové sítě	20
2.4	R-CNN metody	22
2.4.1	R-CNN	22
2.4.2	Fast R-CNN	23
2.4.3	Faster R-CNN	25
2.4.4	Mask R-CNN	27
3	Analýza	29
3.1	Metodika měření	29
3.2	Snímací soustavy	30
3.2.1	Monochromatická kamera, modré osvětlení	30
3.2.2	Barevná kamera, modré osvětlení	32
3.2.3	UV kamera, UV osvětlení	33
3.2.4	UV kamera, přizpůsobené UV osvětlení	35
4	Praktická část	37
4.1	Klasifikace tvarových deskriptorů	37
4.1.1	Předzpracování	37
4.1.2	Datasey	38
4.1.3	Klasifikace	39
4.2	Mask R-CNN	39
4.2.1	Datasey	40
4.2.2	Trénování	41
4.2.3	Validace a testování	42
5	Výsledky	45
5.1	Klasifikace tvarových deskriptorů	45
5.2	Mask R-CNN	45
6	Diskuze	49
	Závěr	51
	Literatura	53
	A Seznam použitých zkratk	59
	B Obsah příloženého paměťového média	61

Seznam obrázků

1.1	Řez optické tyče.	3
1.2	Čočky do světlometů v automobilech.	4
1.3	Drobné defekty.	5
1.4	Škrábanec.	6
1.5	Smirková čára.	6
1.6	Nerovnost okraje.	6
1.7	Schéma algoritmu pro detekci a klasifikaci defektů pomocí tvarových deskriptorů.	7
1.8	Algoritmus předzpracování snímků.	8
1.9	Schéma algoritmu detekce vad displejů mobilních telefonů.	10
1.10	Bounding boxy a masky škrábanců.	11
2.1	Porovnání bankovek nasvícených denním světlem a IR zářením.	15
2.2	Bezpečnostní prvky řidičského průkazu pod UV osvětlením.	16
2.3	Jednoduché prahování.	17
2.4	Použití mediánového filtru na extrémně zašuměném obrázku.	17
2.5	Použití high-boost filtru.	18
2.6	Použití Cannyho detektoru.	19
2.7	Aplikace základních morfologických operací.	20
3.1	Schémata typů osvětlení.	30
3.2	Snímací soustava s monochromatickou kamerou a modrým osvětlením.	32
3.3	Snímací soustava s UV kamerou a UV osvětlením.	34
3.4	Snímací soustava s UV kamerou a přizpůsobeným UV osvětlením.	35
4.1	Schéma algoritmu předzpracování pro získání kontur.	38
4.2	Předzpracování snímků pro extrakci kontur.	39
4.3	Hodnoty metrik AP, AP50 a AP75 během trénování na datasetu 1.	43
4.4	Hodnoty metrik AP, AP50 a AP75 během trénování na datasetu 2.	44
4.5	Hodnoty metrik AP, AP50 a AP75 během trénování na datasetu 4.	44

5.1	Detekce drobných defektů.	46
5.2	Detekce smirkové čáry.	47
5.3	Detekce nerovnosti okraje.	48

Seznam tabulek

1.1	Složení optických skleněných tyčí.	5
2.1	Pásma IR záření.	14
2.2	Pásma UV záření.	15
3.1	Parametry kamery Basler acA2500-60um.	31
3.2	Parametry objektivu Kowa LM35SC.	31
3.3	Parametry kamery Basler a2A2590-60ucPRO.	33
3.4	Parametry kamery jAi CM-140 GE-UV.	33
3.5	Parametry objektivu FL-BC2528-VGUV - F2.8/25mm.	33
4.1	Parametry datasetů pro trénování klasifikátoru tvarových deskriptorů.	40
4.2	Hodnoty hyperparametrů vícevrstvé neuronové sítě pro klasifikaci tvarových deskriptorů.	41
4.3	Parametry datasetů pro trénování Mask R-CNN.	42
4.4	Hodnoty hyperparametrů Mask R-CNN.	43
5.1	Výsledky Mask R-CNN.	47

Úvod

Nedílnou součástí průmyslové výroby je kontrola kvality, která by měla odhalit veškeré vady tak, aby výrobek splňoval požadavky odběratele a předepsané normy. Pracovníci musí zkontrolovat velké množství výrobků a nacházet vady, které jsou pro laika mnohdy takřka nepostřehnutelné. Tato činnost je náročná, vyžaduje absolutní soustředěnost, ale je také únavná, monotónní a repetitivní. To vede k častým chybám, v jejichž důsledku je výstup kontrolního procesu subjektivní. Aby se předešlo důsledkům selhání lidského faktoru, je vhodné proces objektivizovat a automatizovat.

Práce se zabývá metodami snímání a detekce vad lesklých materiálů, konkrétně skleněných tyčí. Tyto materiály je možno snímat v různých barevných spektrech, mohou být vhodným způsobem nasvětleny tak, aby byly vady lépe viditelné a bylo možné je snadněji odhalit a vyhodnotit. To umožní zefektivnit výrobní proces a zvýšit kvalitu výrobku.

Cílem práce je navrhnout metody pro snímání, automatizovanou detekci a klasifikaci vad skleněných tyčí. Pro získání optimálních snímků s dobře patrnými vadami jsou využity pokročilé optické prvky a metody strojového vidění, snímky jsou dále vyhodnocovány a klasifikovány modely strojového učení.

Rešerše

V kapitole jsou představeny optické skleněné tyče a existující řešení snímání a detekce vad lesklých materiálů.

1.1 Optické skleněné tyče

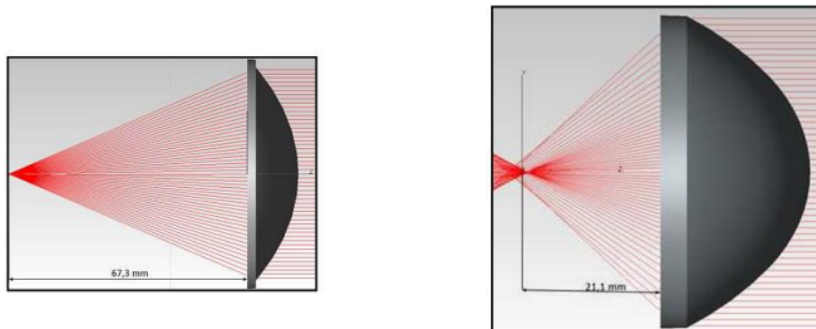
Optické skleněné tyče se používají především k výrobě čoček světlometů v automobilech. Podle technické specifikace [1] jsou 1120 ± 10 mm dlouhé s přibližně čtyřúhelníkovým řezem, jehož výška se v závislosti na konkrétním modelu pohybuje v rozmezí 22–46 mm a šířka v intervalu 28–48 mm. Tyč má jednu plochou **smirkovou stranu**, dvě ploché boční strany a jednu vypouklou spodní stranu (Obrázek 1.1).

Světlometry v automobilech mají různé tvary a optické vlastnosti, kterým se tvar a provedení čoček musí přizpůsobit. Optické tyče jsou proto nařezávány a jednotlivé kusy pak přetavovány do příslušného tvaru. Přetavování probíhá tak, že smirková strana tyče zůstává chladnější a její tvar se nemění, upravuje se pouze tvar spodní vypouklé strany a bočních stran.



Obrázek 1.1: Řez optické tyče.

Na Obrázku 1.2 jsou nákresy dvou různých čoček [2], které vznikly přetavením optické tyče. Jejich rovná strana, na kterou dopadá světlo ze zdroje, odpovídá smirkové straně optické tyče.



Obrázek 1.2: Čočky do světlometů v automobilech.

1.2 Vady optických skleněných tyčí

Součástky používané v automobilovém průmyslu musí splňovat vysoké standardy kvality. Pro optické tyče jsou proto stanoveny normy určující maximální přípustný počet defektů. Nejčastější defekty jsou popsány v následujících podsekcích.

1.2.1 Bubliny

Aby průchod světla tyčí, resp. čočkou nebyl narušen, je třeba, aby bylo sklo homogenní. Jedním z nežádoucích elementů jsou bubliny. „*Přítomnost bublin ve skle používaném v průmyslu je považována za zásadní problém. Při výrobě skla mají bubliny různý fyzikální nebo chemický původ. Fyzikálními příčinami jsou např. velikost částic surovin pro výrobu skla, teplota tavení, délka tavicího procesu, vysoká teplota, odlévání taveniny, chemickou příčinou může být míchání surových materiálů, uvolňování vlhkosti, uvolňování plynů, např. oxidu uhličitého, oxidu uhelnatého, kyslíku nebo dusíku.*“ [3]

Při vhodném nasvícení je bublina viditelná jako drobný světlý bod v tmavém poli (Obrázek 1.3a).

1.2.2 Kamínky

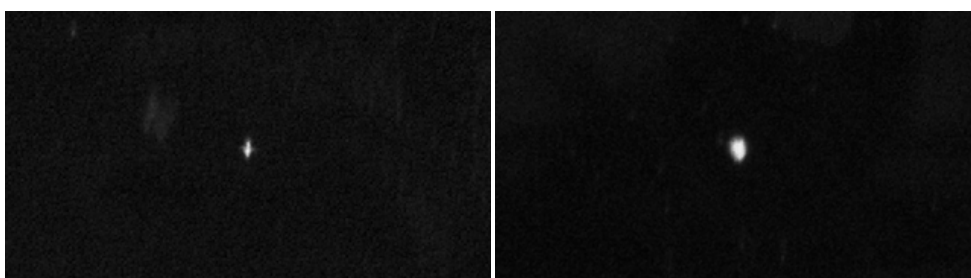
Jak uvádí [1], optické tyče se vyrábí z několika sloučenin, jejichž seznam a poměr je uveden v Tabulce 1.1.

Kromě nich se však v tyčích mohou objevit i cizorodé příměsi, zejména úlomky postupně opotřebovávaných žáruvzdorných vyzdívek sklářských pecí. Úlomky v podobě malých kamínků pak výrazně mění směr světla procházející čočkou.

Tabulka 1.1: Složení optických skleněných tyčí.

Systematický vzorec	Sumární vzorec	Poměr suroviny ve výrobku [%]
Oxid křemičitý	SiO ₂	68
Oxid sodný	Na ₂ O	11
Oxid draselný	K ₂ O	6
Oxid vápenatý	CaO	6
Oxid barnatý	BaO	4,5
Oxid zinečnatý	ZnO	3

Při vhodném nasvícení je kamínek viditelný jako světlý bod v tmavém poli (Obrázek 1.3b), oproti bublině je ale poněkud jasnější a zpravidla větší.



(a) Bublina.

(b) Kamínek.

Obrázek 1.3: Drobné defekty.

1.2.3 Škrábance

Nejčastějšími defekty skleněných tyčí jsou škrábance a oděrky (Obrázek 1.4). Vznikají během výroby v situaci, kdy je tavenina špatně nalitá na dopravník a tuhnoucí hmota pak do dopravníku naráží. K poškození může dojít také při následné manipulaci s hotovým výrobkem. Zásadním problémem jsou škrábance na smirkové straně, jelikož zde zůstávají i po přetavení tyčí na čočky a mají tedy negativní vliv na jejich optické vlastnosti.

1.2.4 Smirková čára

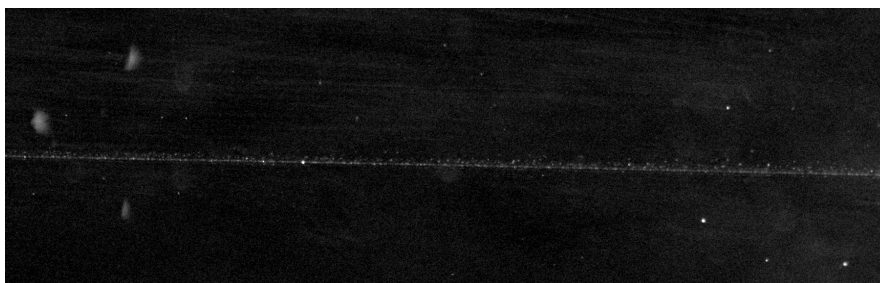
Vada, viditelná na Obrázku 1.5, nemá na rozdíl od jiných vad mechanické či obecně fyzikální příčiny. Jedná se totiž o drobné vysrážené krystaly amoniaku používaného v čistících a leštících prostředcích na sklo. Smirková čára je bez vhodného nasvícení prakticky neviditelná.

1.2.5 Nerovnost okraje

Při nařezávání skleněné hmoty a následné manipulaci s ní často dochází k výrazným deformacím na okrajích (Obrázek 1.6). Tyto části tyčí jsou k výrobě

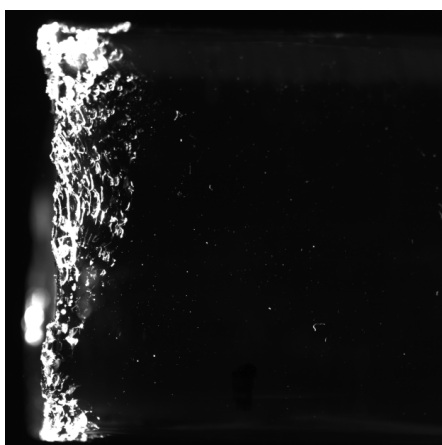


Obrázek 1.4: Škrábanec.



Obrázek 1.5: Smirková čára.

optických čoček zcela nevhodné, před dalším zpracováním je třeba defekty identifikovat a začistit, popřípadě kus tyče úplně odříznout.



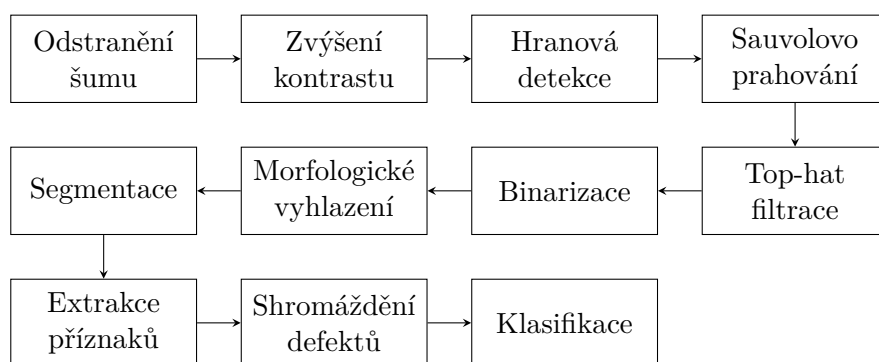
Obrázek 1.6: Nerovnost okraje.

1.3 Systémy pro snímání a detekci vad obdobných materiálů

Sekce popisuje systémy sloužící k detekci a klasifikaci vad na předmětech ze skla nebo obdobných materiálů, např. kovu. Pozornost je věnována zejména použitému druhu kamery a osvětlení, jakož i detekční a klasifikační metodě.

1.3.1 Detekce a klasifikace defektů pomocí tvarových deskriptorů

Hassan a kolektiv v roce 2012 vytvořili systém pro detekci defektů svarů na rentgenových snímcích [4]. Základní schéma algoritmu je popsáno na Obrázku 1.7.



Obrázek 1.7: Schéma algoritmu pro detekci a klasifikaci defektů pomocí tvarových deskriptorů.

Zdrojem dat byly digitalizované rentgenové snímky svarů. Nejprve proběhlo předzpracování. Pro odstranění šumu byl použit **mediánový filtr** (podsekce 2.2.2), pro zvýšení kontrastu high-boost filtr (podsekce 2.2.3), který dobře zvýraznil nepravidelnosti v obrazu. Pomocí **Cannyho detektoru** (podsekce 2.2.4) byly zvýrazněny jednotlivé svary, následně bylo provedeno **Sauvolovo prahování** [5], které určilo, zda je defekt tmavší nebo světlejší než svar v pozadí. Toto prahování je vhodné pro neuniformní pozadí, jelikož je dobře schopno najít lokální práh podle vztahu

$$T = M \left[1 + K \left(\frac{s}{R} - 1 \right) \right] \quad (1.1)$$

počítaném na výřezu o velikosti $N \times N$. M je průměr, s směrodatná odchylka a R maximální možná směrodatná odchylka šedotónových hodnot. Tento krok odlišil dva typy defektů od ostatních. Pokud byly defekty světlejší než svary v pozadí, byl použit originální obraz se zvýrazněnými svary.

Klíčovou částí algoritmu jsou **morfologické operace** (podsekce 2.2.5). Na snímky se zvýrazněným svarem byl aplikován top-hat filtr, tedy odečtení ob-

razu po provedení operace **otevření** od originálního obrázku. Výsledkem byl binární segmentovaný obraz (podsekce 2.2.1) s konturami defektů. Postup algoritmu předzpracování je demonstrován na Obrázku 1.8.



Obrázek 1.8: Algoritmus předzpracování snímků. Vlevo je originální snímek, uprostřed snímek po aplikaci Sauvolova prahování, vpravo segmentovaný snímek.

Pro všechny kontury byly vypočteny následující tvarové deskriptory ($a_{i,j}$ označuje $i., j.$ pixel kontury, p, q, u jsou koeficienty obecné rovnice konturu obepínající elipsy $px^2 + qxy + ry^2 + sx + ty + u = 0$, $a_{c_{i,j}}$ označuje $i., j.$ pixel konvexní obálky):

- plocha

$$A = \sum_{i=0}^n \sum_{j=0}^n a_{i,j} \text{ pro } a_{i,j} = 1, \quad (1.2)$$

- nejdelší průměr

$$L = \sum_{i=0}^n \sum_{j=0}^n p \cdot (a_{i,j} - u), \quad (1.3)$$

- nejkratší průměr

$$e = \sum_{i=0}^n \sum_{j=0}^n q \cdot (a_{i,j} - u), \quad (1.4)$$

- obvod (počet pixelů tvořících hranu kontury),
- plocha konvexní obálky

$$A_c = \sum_{i=0}^n \sum_{j=0}^n a_{c_{i,j}} \text{ pro } a_{c_{i,j}} = 1, \quad (1.5)$$

- celistvost (poměr plochy kontury a její konvexní obálky)

$$S = \frac{A}{A_c}, \quad (1.6)$$

- excentricita

$$E = \sqrt{\frac{1 - \left(\frac{e}{L}\right)^2}{\left(\frac{L}{2}\right)^2}}, \quad (1.7)$$

- orientace (úhel sevřený osou x a hlavní osou obepínající elipsy, tedy nejdelším průměrem).

Pět těchto příznaků, konkrétně plocha konvexní obálky, nejdelší průměr, nejkratší průměr, celistvost a excentricita, bylo použito pro klasifikaci pomocí vícevrstvé neuronové sítě (podsekcce 2.3.2) se dvěma skrytými vrstvami, každá o pěti neuronech. Síť na vstupu přijímá vektory o pěti složkách a vrací vektory o dvou složkách.

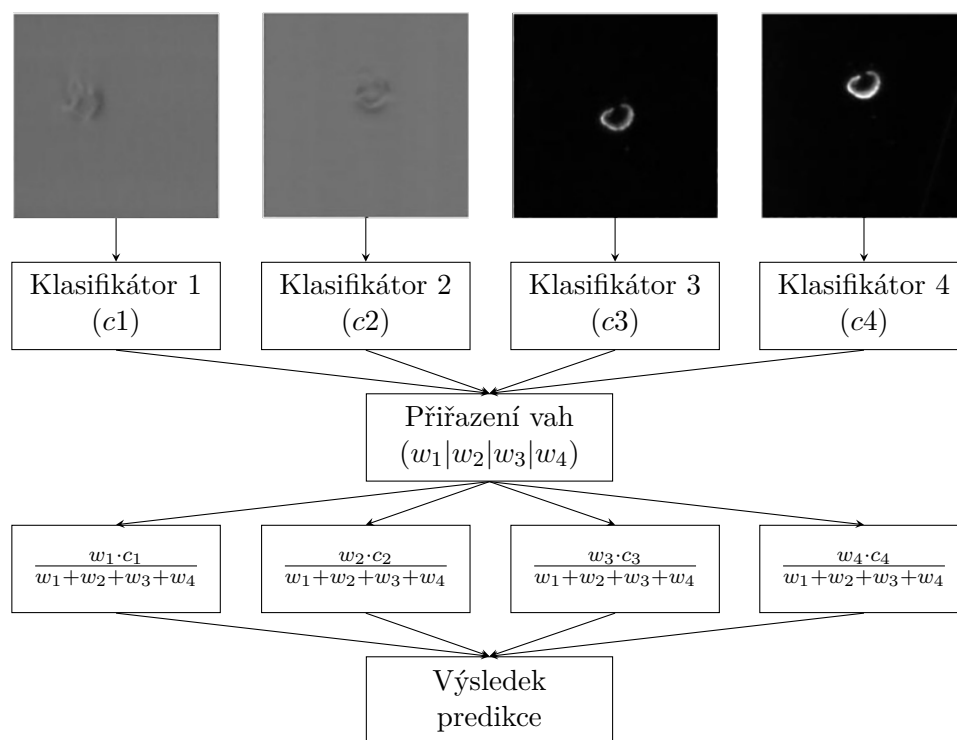
1.3.2 Detekce vad displejů mobilních telefonů

Park a kolektiv v roce 2019 navrhli systém pro detekci a klasifikaci vad skel používaných pro výrobu displejů mobilních telefonů [6], například důlků, škrábanců nebo odštěpků. Autoři pracovali s horizontálními a vertikálními snímky pořízenými řádkovými kamerami ve světlém a temném poli. Na těchto snímcích provedli předzpracování spočívající v odfiltrování šumu pomocí konvoluce s použitím kernelu o velikosti 3×3 . Obrázky získané řádkovou kamerou však měly vysoké rozlišení, detekční algoritmus by tedy potřeboval velmi výkonný hardware a mnoho času, a narušoval tak tempo výroby. Proto byly ze snímků extrahovány oblasti o velikosti 160×160 pixelů podezřelé z výskytu vady. Nad těmito snímky byla provedena **data augmentation** [7], kdy byly snímky náhodně rotovány, měněna jejich velikost, kontrast a podobně. Takto bylo vygenerováno velké množství dalších dat vhodných pro trénování klasifikačních modelů, zároveň se srovnaly počty snímků bez vad a s vadami. Větší množství dat také snížilo pravděpodobnost přeučení modelu, tedy jeho přílišnou fixaci na nereprezentativní vzorek dat. Pro každý ze čtyř typů snímků byla natrénována konvoluční neuronová síť (CNN) založená na síti DarkNet-19 [8]. Finální rozhodnutí, zda se jedná o vadu, popř. jakou, bylo váženým průměrem výsledků jednotlivých sítí. Váhy byly optimalizovány během trénování. Schéma algoritmu je popsáno na Obrázku 1.9.

1.3.3 Detekce vad skleněných panelů

Pan a kolektiv se v roce 2020 zabývali detekcí drobných škrábanců na skleněných panelech používaných na fasádách budov [9]. Větší výskyt těchto vad na malém prostoru by v extrémním případě mohl vést k rozlomení panelu a jeho následnému pádu. Navrhli snímací soustavu s mikroskopickou kamerou, která se systematicky pohybuje postupně nad celým panelem a pořizuje snímky, z nichž každý pokrývá plochu o rozloze $0,01 \text{ m}^2$. Pro nasvícení bylo použito koaxiální osvětlení, které pomocí polopropustného zrcadla odráží světlo do optické dráhy snímače, čímž zabraňuje tvorbě odlesků, které by jinak znemožnily rozpoznávání defektů. Snímky byly anotovány aproximací vad polygonem, zároveň byla provedena segmentace, kdy hodnota pixelů uvnitř polygonu byla stanovena na 1, hodnota pixelů mimo polygon na 0.

Takto anotované snímky byly postoupeny klasifikačnímu algoritmu postavenému na modelu Mask R-CNN (podsekcce 2.4.4), který byl předtrénován

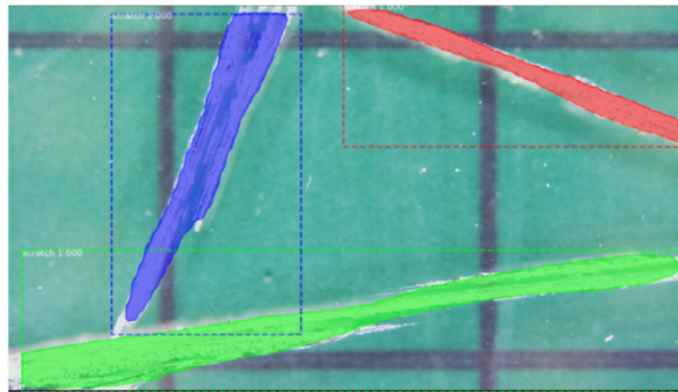


Obrázek 1.9: Schéma algoritmu detekce vad displejů mobilních telefonů.

na COCO datasetu [10]. Následně byl model během 200 epoch dotrénován na vlastním datasetu o 1032 snímcích. Testování proběhlo na datasetu o 72 snímcích.

Úspěšnost modelu byla vyhodnocována pomocí metriky mAP (mean average precision) [11], která byla počítána pro různé prahy IoU (Intersection over Union) [12]. Predikce byla považována za správnou, pokud IoU predikce a skutečného objektu převýšila příslušnou hodnotu prahu. Nejlepší mAP bounding boxů (obdélníků ohraničujících oblast) dosáhl algoritmus pro IoU rovnou 0,5. I na poměrně výrazném pozadí byly škrábance detekovány dostatečně přesně, algoritmus se rovněž nenechal zmást šumem, jak je patrné na Obrázku 1.10.

Výsledky Mask R-CNN modelu byly též porovnány s novějším modelem pro segmentaci masky detekovaných objektů YOLACT [13]. Mask R-CNN dosahoval o 7 % vyšší přesnosti, segmentovaná maska též lépe odpovídala zejména méně kompaktním škrábcům, nicméně YOLACT byl téměř čtyřikrát rychlejší. Volbu modelu je možno přizpůsobit v závislosti na tom, zda je dominantním požadavkem na systém rychlost, nebo přesnost.



Obrázek 1.10: Bounding boxy a masky škrábanců.

Teoretické zázemí

V kapitole jsou teoreticky rozebrány metody a algoritmy použité v analyzovaných řešeních i v praktické části.

2.1 Osvětlení ve strojovém vidění

Strojové vidění nachází uplatnění v rozličných oborech. Jsou snímány různorodé objekty, kterým je třeba přizpůsobit snímací soustavu. Aby mohly být objekty kvalitně rozpoznány, musí být na snímku dostatečně výrazné [14]. Jejich kontrast je tedy třeba maximalizovat, naopak kontrast pozadí je třeba minimalizovat tak, aby nijak neovlivňoval oblast zájmu. Parametry metod zpracování obrazu, kterými získané snímky následně procházejí, je nutno vyladit tak, aby byly metody dostatečně robustní. Proto je nutné, aby snímky stejné scény pořízené v různých časech byly co nejpodobnější. Světelné podmínky při snímání by tedy měly být stálé.

2.1.1 Výběr osvětlení v závislosti na vlastnostech snímaného objektu

Při výběru osvětlení je nutno brát v potaz zejména tvar objektu, jeho barvu, povrch a velikost.

Nejprve je zpravidla analyzován tvar a povrch. Lze použít **směrové osvětlení**, které je vhodné zejména pro materiál s matným povrchem. Na lesklém povrchu by vytvářelo odlesky, které skryjí detaily. Směrové osvětlení též zvýrazňuje prostorové uspořádání, čehož lze využít např. při inspekci reliéfů. Alternativou je **difuzní osvětlení**, které vyzařované světlo rozptýlí a tím odlesky částečně eliminuje. Nevýhodou je zploštění scény, mizí informace o hloubce a prostorovém uspořádání. Odlesky lze odstranit snímáním v **temném poli**, kdy je objekt osvětlen pod malým úhlem, zpravidla menším než 45° . Takto

dopadající světlo se pak na rovném povrchu neodráží směrem do objektivu a netvoří odlesky. Jiná situace však nastává u povrchových defektů, kdy se na jejich stěnách světlo láme a směřuje do všech stran, mj. i do objektivu. Tyto oblasti pak jasně svítí v jinak temném okolí a jsou snadno detekovatelné metodami zpracování obrazu.

Ke zvýraznění **barvy** lze přistoupit dvěma způsoby. Pokud je nutné určitou barvu zesvětlit, je vhodné na ni svítit světlem stejné barvy. Naopak pokud je třeba docílit ztmavení, použije se světlo **komplementární barvy** [15], jelikož každá barva svůj komplementární protějšek pohlcuje. Tyto efekty jsou dobře pozorovatelné na černobílých snímcích pořízených monochromatickou kamerou.

2.1.2 Osvětlení mimo viditelné spektrum

Další možnosti snímání skýtá osvětlení mimo viditelnou část elektromagnetického spektra, konkrétně infračervené (IR) a ultrafialové (UV) záření.

Infračervené záření [16] je elektromagnetické záření s vlnovou délkou 780–1 000 000 nm, tedy vyšší než viditelná část spektra. Dělí se na několik pásem uvedených v Tabulce 2.1.

Tabulka 2.1: Pásmo IR záření.

Název pásma	Vlnová délka [nm]
Blízké	780–1 400
Krátké	1 400–3 000
Střední	3 000–8 000
Dlouhé	8 000–15 000
Vzdálené	15 000–1 000 000

Ve strojovém vidění se zpravidla používá blízké a krátké infračervené záření. Při vlnových délkách v rozmezí 780–1 000 nm lze stále používat monochromatické kamery se standardními čipy, nicméně jejich citlivost se s rostoucí vlnovou délkou záření postupně snižuje a při vyšších hodnotách je nutno kamery chladit. Pro práci ve středním pásmu se využívají termokamery se speciálními objektivy, jejichž čočky již nejsou vyrobeny ze skla, nýbrž ze speciálních materiálů, především germania.

Vyšší vlnová délka infračerveného záření znamená, že jednotlivé vlny méně interagují se zkoumaným materiálem. To snižuje množství odlesků a umožňuje světlu proniknout dovnitř, díky čemuž lze zkoumat vnitřní strukturu objektu, viz Obrázek 2.1. Infračervené osvětlení je možno používat například pro inspekci bezpečnostních prvků na bankovkách nebo lidskému oku nevi-



(a) Denní světlo.



(b) IR osvětlení.

Obrázek 2.1: Porovnání bankovek nasvícených denním světlem a IR zářením.

ditelných defektů na objektech maskovaných barevným povrchem, který je infračerveným zářením zneviditelněn.

Dalším běžně používaným zářením je **ultrafialové záření** [17] s vlnovou délkou 10–400 nm, tedy nižší než viditelná část spektra. Dělí se na několik pásem uvedených v Tabulce 2.2.

Tabulka 2.2: Pásma UV záření.

Název pásma	Vlnová délka [nm]
UVA	315–400
UVB	280–315
UVC	< 280

UV lampy používané ve strojovém vidění emitují záření o vlnové délce 365 nm. To je však obtížně zachytitelné běžnými kamerami, jejichž optické prvky vyrobené ze standardního skla toto záření ve velké míře absorbují. Místo nich se proto používají speciální monochromatické kamery, jejichž optická soustava je vyrobena z čistého křemenného skla, které UV záření propouští.

UV záření má díky své nižší vlnové délce tendenci se intenzivně odrážet od povrchů, díky čemuž je vhodné k inspekci i velmi malých povrchových vad. Dále je vhodné např. pro inspekci bezpečnostních prvků dokladů (Obrázek 2.2), zvýraznění stehů, lepidel a dalších látek.

Kvalitu snímků je možné zvýšit použitím vhodných **filtrů**, které propouštějí pouze světlo určité vlnové délky. Například UV filtr může zvýšit kontrast fluorescentního objektu nebo osvětlení povrchové vady oproti zbytku objektu, popř. odstranit nežádoucí odlesky.



Obrázek 2.2: Bezpečnostní prvky řidičského průkazu pod UV osvětlením.

2.2 Metody zpracování obrazu

Metody zpracování obrazu provádějí obrazové operace a transformace za účelem zvýraznění preferovaných prvků pro jejich další zkoumání, či naopak odstranění rušivých elementů. Upravený obraz může být dále postoupen např. algoritmům strojového učení.

2.2.1 Segmentace

Segmentace je jednou ze základních metod zpracování obrazu. Slouží k rozdělení obrazu na segmenty se stejnými vlastnostmi a významem, například k oddělení objektu zájmu od pozadí. K metodám segmentace patří:

- **jednoduché prahování:** pixely jsou porovnány s hodnotou prahu, pokud je hodnota pixelu větší, je výslednému pixelu přiřazena 1, jinak 0,
- **dvojitě prahování:** pixely jsou porovnány s hodnotami spodního a horního prahu, pokud je hodnota pixelu větší nebo rovna spodnímu prahu a zároveň menší nebo rovna hornímu prahu, je výslednému pixelu přiřazena 1, jinak 0,
- **adaptivní prahování:** pixely jsou porovnány s hodnotou prahu, který je počítán dynamicky jako vážený součet pixelů na $n \times n$ okolí příslušného pixelu. Váhy mohou být definovány např. Gaussovým jádrem [18]

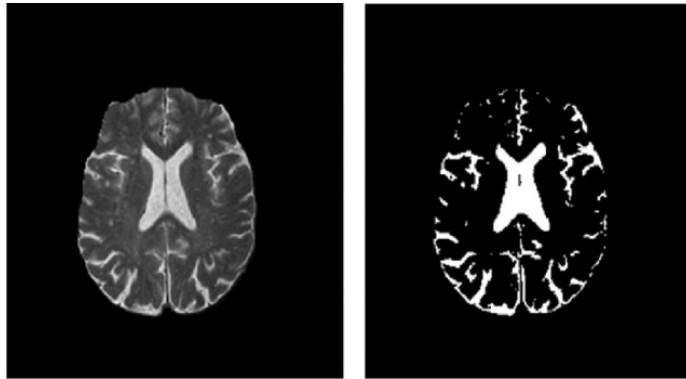
$$G_i = \alpha \cdot e^{-\frac{(i - \frac{n-1}{2})^2}{2 \cdot \sigma^2}},$$

kde $i \in \{0, \dots, n-1\}$ a α je koeficient normalizující součet koeficientů G_i na 1.

Výstupem všech metod segmentace je binární obraz (Obrázek 2.3).

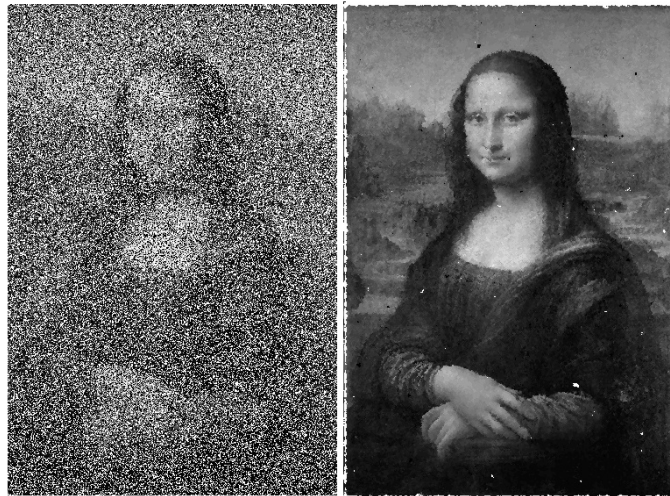
2.2.2 Mediánový filtr

Mediánový filtr je nelineární filtr sloužící především k odstranění šumu z obrazu. Funguje tak, že nahrazuje hodnotu pixelu mediánem hodnot sousedních



Obrázek 2.3: Jednoduché prahování. [19]

pixelů ve výřezu o velikosti $n \times n$. Je výpočetně náročný, neboť pro každý pixel je třeba seřadit hodnoty, ze kterých je vybírán medián. Je však také robustní, poradí si i s extrémními hodnotami, např. šumem „pepř a sůl“ (Obrázek 2.4).



Obrázek 2.4: Použití mediánového filtru na extrémně zašuměném obrázku. [20]

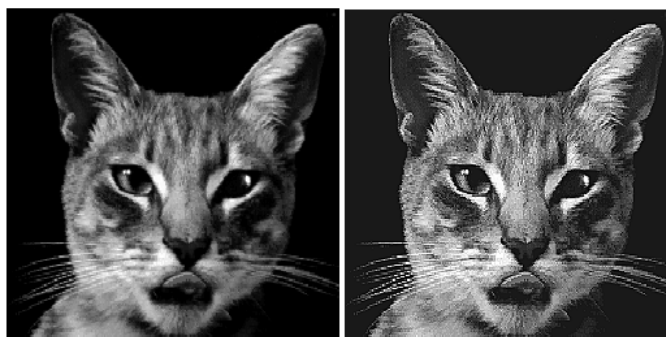
2.2.3 High-boost filtr

High-boost filtr zvýrazňuje světlé detaily a hrany (Obrázek 2.5), zároveň neznehodnocuje tmavší pozadí. Výpočet je proveden konvolučním aplikováním

high-boost kernelu W_{hb} .

$$\begin{aligned} W_{\text{hb}} = W_{\text{ap}} + c \cdot W_{\text{hp}} &= \begin{bmatrix} 0 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 0 \end{bmatrix} + c \begin{bmatrix} 0 & -1 & 0 \\ -1 & 4 & -1 \\ 0 & -1 & 0 \end{bmatrix} \\ &= \begin{bmatrix} 0 & -c & 0 \\ -c & 4c + 1 & -c \\ 0 & -c & 0 \end{bmatrix}, \end{aligned} \quad (2.1)$$

kde W_{ap} je kernel identity a W_{hp} high-pass filtr, tedy filtr propouštějící pouze vyšší hodnoty pixelů. Příklad použití high-boost filtru je na Obrázku 2.5.



Obrázek 2.5: Použití high-boost filtru. [21]

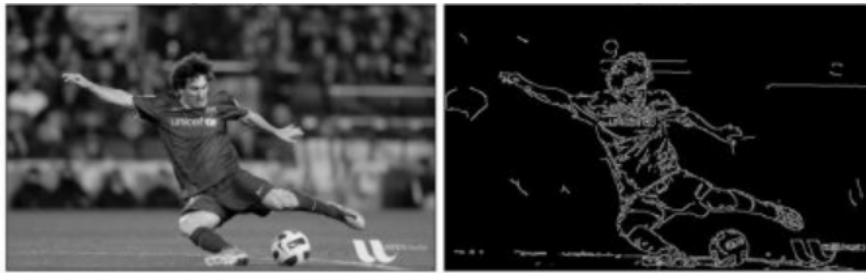
2.2.4 Cannyho detektor

Cannyho detektor slouží k detekci hran (Obrázek 2.6). Pracuje v několika krocích [22]:

1. Vyhlazení obrazu, např. aplikací jednodimenzionálního Gaussova filtru g .
2. Výpočet gradientu (detekce hran), např. pomocí Sobelova filtru nebo jednodimenzionálního derivovaného Gaussova filtru dg .
3. Výpočet konvoluce g a dg , čímž je získán filtr gdg pro simultánní vyhlazení obrazu a detekci hran.
4. Aplikace filtru gdg na původní obraz x , čímž je získán obraz x_1 s detekovanými hranami ve vodorovném směru.
5. Aplikace filtru gdg^T na původní obraz x , čímž je získán obraz x_2 s detekovanými hranami ve svislém směru.
6. Získání výsledného snímku $x_e = \sqrt{x_1^2 + x_2^2}$.

2.2.5 Základní morfologické operace

Metody matematické morfologie pracují s geometrickými útvary v obrazu. Vstupem je binární nebo šedotónový snímek a strukturální element, kterým je



Obrázek 2.6: Použití Cannyho detektoru. [23]

obraz při transformaci procházen. Základní morfologické operace jsou eroze a dilatace, které se dále kombinují v operacích otevření a uzavření.

Při **erozi** (zn. \ominus) je pixelu v binárním obrazu přiřazena hodnota 1, jestliže jsou hodnoty všech pixelů označených strukturním elementem rovny 1. V případě šedotónového obrazu je pixelu přiřazena nejmenší hodnota z hodnot pixelů překrytých strukturním elementem. Při aplikování eroze se plocha objektu zmenšuje.

Při **dilataci** (zn. \oplus) je pixelu v binárním obrazu přiřazena hodnota 1, jestliže je hodnota alespoň jednoho z pixelů označených strukturním elementem rovna 1. V případě šedotónového obrazu je pixelu přiřazena největší hodnota z hodnot pixelů překrytých strukturním elementem. Při aplikování dilatace se plocha objektu zvětšuje.

Při **otevření** (zn. \circ) je aplikována nejprve eroze a poté dilatace. Je vhodná pro odstranění šumu, kdy je erozí eliminován drobný šum a následnou dilatací jsou větší objekty téměř identicky zrekonstruovány do původního tvaru. Otevření využívá top-hat transformace, která od originálního obrazu odečítá jeho verzi po aplikaci otevření.

Při **uzavření** (zn. \bullet) je aplikována nejprve dilatace a poté eroze. Je vhodná pro odstranění drobných „děr“ v objektech, které jsou eliminovány dilatací, následná eroze téměř identicky zrekonstruuje okraje zacelených objektů do původní podoby. Uzavření využívá black-hat transformace, která od obrazu po aplikaci uzavření odečítá jeho originální verzi.

2.3 Neuronové sítě

Neuronová síť je výpočetní model vzdáleně inspirovaný strukturou vzájemně propojených neuronů v mozku [25]. Používá se pro klasifikaci i regresi, složitější architektury dále pro extrakci příznaků, kompresi dat nebo analýzu textu.



Obrázek 2.7: Aplikace základních morfologických operací: vlevo originální obrázek, uprostřed snímek po aplikaci eroze, vpravo snímek po aplikaci dilatace. [24]

2.3.1 Jednovrstvý perceptron

Nejjednodušším modelem neuronové sítě určeným ke klasifikaci je **jednovrstvý perceptron** [26], který na vstupu přijímá vektor příznaků x_1, \dots, x_n pronásobených vahami w_1, \dots, w_n a intercept w_0 . Z těchto hodnot je vypočten vnitřní potenciál ξ

$$\xi = w_0 + \sum_{i=1}^n w_i x_i = w_0 + w^T x = w_0 + [w_1 \ \dots \ w_n] \begin{bmatrix} x_1 \\ \vdots \\ x_n \end{bmatrix}, \quad (2.2)$$

na nějž je poté aplikována **aktivační funkce** f , v případě perceptronu skoková funkce

$$f(\xi) = f(w_0 + w^T x) = \begin{cases} 1 & \text{když } \xi \geq 0, \\ 0 & \text{když } \xi < 0. \end{cases} \quad (2.3)$$

Neuron je aktivován, pokud jeho výstup $f(\xi) = 1$. Tento výpočet se nazývá **dopředný chod**.

Na základě hodnot získaných v dopředném chodu je určena chyba predikce, na jejímž základě je provedena inkrementální aktualizace vah nazvaná **zpětný chod**

$$\begin{aligned} \text{error} &= Y - \hat{Y} \\ w_i &\leftarrow w_i + \text{error} \cdot x_i, \quad i \in \{1, \dots, n\} \\ w_0 &\leftarrow w_0 + \text{error} \end{aligned} \quad (2.4)$$

kde $\hat{Y} = f(w_0 + w^T x)$ je predikce v bodě x a Y je skutečná hodnota.

2.3.2 Vícevrstvé neuronové sítě

Samotný perceptron je však schopen klasifikovat pouze lineárně separabilní data. Pro separaci složitějších dat je třeba použít více neuronů v podobě **vícevrstvého perceptronu**. Tato neuronová síť se skládá z vrstev, kdy výstupy neuronů jedné vrstvy tvoří vstupy neuronů druhé vrstvy. Všechny vrstvy kromě vstupní a výstupní se nazývají **skryté**.

Lze dokázat [27], že síť s jedinou vrstvou dokáže s libovolnou přesností aproximovat jakoukoliv spojitou funkci s kompaktním nosičem v \mathbb{R}^n . Pro to by ale bylo potřeba velké množství neuronů, jejichž nadměrné množství v jedné vrstvě snižuje výkon. Jsou tedy preferovány hlubší sítě s více skrytými vrstvami.

Na hlubší síti však nelze aplikovat zpětný chod (rovnice 2.4). Proto se používá **algoritmus zpětného šíření chyby** [28], který vyžaduje, aby byla neuronová síť skoro všude diferencovatelná, čehož lze docílit volbou vhodné aktivační funkce namísto dříve použité skokové funkce (rovnice 2.3). Použit lze **oříznutou lineární funkci – RELU** (rectified linear unit)

$$f(\xi) = \max(0, \xi) = \begin{cases} \xi & \text{pro } \xi \geq 0, \\ 0 & \text{pro } \xi < 0 \end{cases} \quad (2.5)$$

nebo **hyperbolický tangens**

$$f(\xi) = \tanh(\xi) = \frac{e^\xi - e^{-\xi}}{e^\xi + e^{-\xi}}. \quad (2.6)$$

Ve výstupní vrstvě je třeba hodnoty získané z předchozích vrstev převést na hodnoty vhodné k predikci. Pro regresní úlohu se výstupní vrstva skládá z jediného neuronu s **identitou** $f(\xi) = \xi$ jako aktivační funkcí. Pro binární klasifikaci se použije jeden výstupní neuron, jehož aktivační funkce je **logistická sigmoida**

$$f(\xi) = \frac{1}{1 + e^{-\xi}}, \quad (2.7)$$

jejíž hodnota je interpretována jako pravděpodobnost příslušnosti ke třídě 1, tedy jako $\hat{P}(Y = 1|X = x)$. Predikce v bodě x je potom $\hat{Y} = 1$, pokud $\hat{P}(Y = 1|X = x) > 0,5$, jinak $\hat{Y} = 0$. Při klasifikaci do c tříd se výstupní vrstva skládá z c neuronů s normovanou aktivační funkcí **softmax**

$$f_i(\xi) = \frac{e^{\xi_i}}{e^{\xi_1} + \dots + e^{\xi_c}}, \quad (2.8)$$

kde $\xi = (\xi_1, \dots, \xi_c)^T$ je vektor vnitřních potenciálů c neuronů a $f_i(\xi)$ je aktivační funkce i . neuronu, jejíž výstup je interpretován jako pravděpodobnost příslušnosti k i . třídě $\hat{P}(Y = i|X = x)$. Predikce v bodě x je pak $\hat{Y} = \arg \max_{i \in \{1, \dots, c\}} \hat{P}(Y = i|X = x)$.

Při učení je minimalizována ztrátová funkce L , která měří, jak dobře model predikuje konkrétní hodnotu. V případě klasifikace do c tříd je používána **kategorická relativní entropie**

$$L(Y, \hat{p}) = - \sum_{j=1}^c \mathbb{1}_{Y=j} \log \hat{p}_j = - \log \hat{p}_Y, \quad (2.9)$$

kde $\hat{p}_i = \hat{P}(Y = i|X = x)$, $\hat{p} = (\hat{p}_1, \dots, \hat{p}_c)^T$ a $\mathbb{1}_{Y=j} = 1$, když $Y = j$, jinak $\mathbb{1}_{Y=j} = 0$.

Trénování neuronové sítě s náhodnými počátečními váhami $w = (w_1, \dots, w_m)^T$ na N trénovacích datových bodech $(Y_1, x_1), \dots, (Y_N, x_N)$ probíhá takto:

1. Výpočet průměrné chyby predikce na trénovací množině

$$J(w) = \frac{1}{N} \sum_{i=1}^N L(Y_i, g(x_i; w)). \quad (2.10)$$

2. Výpočet gradientu

$$\nabla_w J = \left(\frac{\partial J}{\partial w_1}, \dots, \frac{\partial J}{\partial w_m} \right)^T. \quad (2.11)$$

3. Přepočítání vah

$$w \leftarrow w - \alpha \nabla_w J, \quad (2.12)$$

kde α je učicí parametr.

Gradientní sestup mívá v prostoru parametrů tam, kde je gradient nulový. Pro jednovrstvou síť se jedná o globální minimum, v případě vícevrstvé sítě to může být i lokální minimum nebo sedlové body, kterých však není mnoho a jejich hodnoty jsou často blízké globálnímu minimu.

2.4 R-CNN metody

V roce 2013 byla navržena metoda detekce objektů v obrazu nazvaná Regiony s příznaky konvolučních neuronových sítí (R-CNN) [29], která dosáhla o 30 % vyšší mAP než do té doby používané metody. Metoda je však poměrně pomalá, detekce neovládne provádět v reálném čase, proto byla v roce 2015 uvedena vylepšená metoda Fast R-CNN [30] a v roce 2017 aktuální metoda Faster R-CNN [31].

2.4.1 R-CNN

Tento systém pro detekci objektů se skládá ze tří částí: extrakce návrhů regionů, extrakce příznaků a jejich klasifikace.

Návrhy regionů jsou generovány algoritmem selective search [32]. Pomocí grafové segmentace [33] jsou vytvořeny prvotní regiony, které jsou dále iterativně slučovány tak, že se na základě barvy a textury spočítají podobnosti sousedních regionů a dva nejpodobnější jsou spojeny dohromady. Výsledkem je přibližně 2000 návrhů regionů.

Z každého regionu je extrahován 4096dimenzionální vektor příznaků, které jsou vypočteny z RGB obrázku o velikosti 227×227 pixelů pomocí CNN s pěti konvolučními a dvěma plně propojenými vrstvami [34]. Před průchodem konvoluční sítí jsou regiony z každé strany zvětšeny o 16 pixelů a následně transformovány na velikost 227×227 pixelů.

Vektory příznaků jsou klasifikovány množinou SVM (support-vector machines) [35], kdy je pro každou třídu přítomen jeden klasifikátor. Následně je nad všemi regiony opatřenými pravděpodobnostmi příslušnosti do jednotlivých tříd provedena non-maximální suprese, kdy jsou vyřazeny regiony, jejichž IoU s regionem s vyšší pravděpodobností příslušnosti je vyšší než práh, jehož hodnota je upravována v průběhu učení.

CNN použitá pro extrakci příznaků z obrázků byla předtrénována na podmnožině ImageNet [36] datasetu použité na ImageNet Large Scale Visual Recognition Challenge (ILSVRC) [37] v roce 2012. Pro obrázky byly dostupné pouze anotace na úrovni jednotlivých snímků (tedy nikoliv bounding boxy). Pro adaptaci na detekci objektů z transformovaných regionů byla síť na těchto regionech dále trénována pomocí stochastického gradientního sestupu (SGD). Vrstva, která se dříve starala o klasifikaci do 1000 tříd datasetu ImageNet, byla nahrazena vrstvou pro klasifikaci do $N + 1$ tříd, kdy jedna třída navíc reprezentuje pozadí. Aby navržený region odpovídal nějaké třídě, musí být jeho IoU s ground-truth bounding boxem (tedy boxem odpovídajícím anotaci definující skutečnou polohu objektu) odpovídající třídy větší než 0,5. V každé iteraci SGD je vybráno 32 bounding boxů odpovídajících některé ze tříd a 96 bounding boxů odpovídajících pozadí tak, aby byl zkonstruován mini-batch (množina trénovacích dat pro jednu iteraci učení) o 128 snímcích.

Po extrakci příznaků a přiřazení třídy jednotlivým bounding boxům je pro každou třídu natrénován jeden lineární SVM klasifikátor. Jelikož jsou trénovací data příliš velká na to, aby se vešla do paměti, je použit hard-negative mining [38]. Metoda redukuje množinu bounding boxů odpovídajících pozadí, která je zpravidla příliš velká a disproporční vzhledem k množině bounding boxů ostatních tříd, pouze na ty, u kterých je pravděpodobné, že budou nesprávně klasifikovány jako pozitivní pro některou ze tříd.

Pro odstranění špatně umístěných predikcí používá R-CNN lineární regresní model pro predikci nového bounding boxu na základě příznaků získaných CNN.

2.4.2 Fast R-CNN

Ačkoliv R-CNN dosahuje poměrně vysoké přesnosti, její použití je velmi výpočetně náročné, jelikož pro každý obrázek na vstupu je potřeba 2000 průchodů konvoluční neuronovou sítí. Např. trénování na 5000 snímcích zabere

2,5 dne výpočetního času na GPU. Model nemůže být použit ani pro real-time systémy, jelikož predikce pro testovací snímek trvá i na GPU až 47 sekund.

Na rozdíl od R-CNN konvoluční síť nejprve zpracuje celý snímek najednou a vrátí mapu příznaků. Následně z ní pro každý navržený region pooling vrstva extrahuje vektor příznaků fixní délky (konkrétně 49). Ten je dále předán plně propojeným vrstvám, které se posléze dělí do dvou paralelních výstupních vrstev. První vrstva pomocí funkce softmax spočítá pravděpodobnosti příslušnosti do N tříd a doplňkové třídy reprezentující pozadí, druhá vrstva pro každou z N tříd vrátí množinu čtyř čísel kódující pozici bounding boxu pro příslušnou třídu.

Fast R-CNN využívá konvoluční síť předtrénovanou na datasetu ImageNet. Při inicializaci síť prochází třemi transformacemi: poslední max pooling vrstva je nahrazena pooling vrstvou extrahující matici příznaků 7×7 , poslední plně propojená vrstva a softmax vrstva jsou nahrazeny dvěma paralelními výstupními vrstvami (viz výše). Nakonec je síť transformována tak, aby na vstupu přijímala obrázky a případně i návrhy regionů.

Pro zrychlení trénování je využito sdílení příznaků. SGD mini-batche jsou sestaveny hierarchicky, kdy je nejprve vybráno N snímků a z každého snímku pak $\frac{R}{N}$ regionů. Regiony ze stejného obrázku pak během dopředného a zpětného chodu sdílejí výpočetní výkon a paměť. Čím je N menší, tím je trénování pomalejší.

Trénování Fast R-CNN je dále zefektivněno tím, že fine-tuning (ladění parametrů modelu pro dosažení maximální přesnosti) softmax klasifikátoru a regresoru bounding boxů je prováděno v jedné fázi na rozdíl od R-CNN, která softmax klasifikátor, SVM a bounding box regresory optimalizuje samostatně.

Každému regionu vrácenému druhou paralelní výstupní vrstvou je přiřazena ground-truth třída u a ground-truth cíl pro bounding box regresi v . Pro společné trénování klasifikace modelu a regrese bounding boxů je na každý region použita víceúčelová ztrátová funkce

$$L(p, u, t^u, v) = L_{\text{cls}}(p, u) + \lambda[u \geq 1]L_{\text{loc}}(t^u, v), \quad (2.13)$$

kde $L_{\text{cls}}(p, u) = -\log p_u$ je zlogaritmovaná ztráta pro ground-truth třídu u . Ztráta L_{loc} je definována nad čtveřicí ground-truth regresních cílů $v = (v_x, v_y, v_w, v_h)$ pro třídu u a čtveřicí souřadnic predikovaného bounding boxu $t^u = (t_x^u, t_y^u, t_w^u, t_h^u)$ pro stejnou třídu. t_x a t_y reprezentují souřadnice levého horního rohu, t_w šířku a t_h výšku. Iversonova závorka $[u \geq 1]$ je rovna 1, pokud $u \geq 1$, jinak je rovna 0. Pro regiony odpovídající pozadí neexistují odpovídající ground-truth bounding boxy, proto je L_{loc} ignorována.

Pro regresi bounding boxů je využita ztrátová funkce

$$L_{\text{loc}}(t^u, v) = \sum_{i \in \{x, y, w, h\}} \text{smooth}_{L_1}(t_i^u - v_i), \quad (2.14)$$

kde

$$\text{smooth}_{L_1}(x) = \begin{cases} 0,5x^2 & \text{pokud } |x| < 1, \\ |x| - 0,5 & \text{jinak,} \end{cases} \quad (2.15)$$

je robustní L_1 ztrátová funkce, která je méně ovlivnitelná odlehlými hodnotami než L_2 funkce použitá v R-CNN.

Každý SGD mini-batch je zkonstruován ze dvou náhodně vybraných snímků. Jsou používány mini-batche o velikosti $R = 128$, tedy z každého snímku bylo extrahováno 64 regionů. Z nich je vybráno 25 % těch, jejichž IoU s ground-truth bounding boxem je alespoň 0,5. Ostatní regiony jsou vybrány z návrhů, jejichž maximální IoU s ground-truth bounding boxem leží v intervalu $(0,1; 0,5)$. Tyto regiony odpovídají pozadí. Během trénování je použita data augmentation, kdy jsou obrázky horizontálně převráceny s pravděpodobností 0,5.

Pro detekci model na vstupu předpokládá snímek a typicky 2000 navržených regionů. Pro každý region r model po skončení dopředného chodu vrátí rozdělení aposteriorních pravděpodobností příslušnosti do jednotlivých tříd a množinu predikovaných offsetů bounding boxů relativních k r (jeden pro každou z K tříd). Každému regionu je za použití odhadnuté pravděpodobnosti

$$\Pr(\text{class} = k | r) \triangleq p_k \quad (2.16)$$

přiřazena jistota detekce. Nakonec je nezávisle pro každou třídu provedena non-maximální suprese obdobně jako u R-CNN.

2.4.3 Faster R-CNN

Ačkoliv je Fast R-CNN násobně rychlejší díky sdílení příznaků generovaných CNN, pro generování návrhů regionů využívá obdobně jako R-CNN výpočetně náročný algoritmus selective search. Bez nutnosti generování regionů je Fast R-CNN schopna provést predikci za 0,32 s, s generováním toto trvá až 2,3 s. Faster R-CNN proto přichází s hlubokou konvoluční sítí pro návrh regionů (Region Proposal Network, RPN), která sdílí konvoluční vrstvy s detekční sítí použité ve Fast R-CNN. Čas potřebný pro výpočet návrhů regionů v jednom snímku je tak snížen až na 10 ms. Po započítání všech metod je Faster R-CNN schopna při použití GPU provádět detekce až v 5 snímcích za sekundu, zároveň dosahuje vysoké přesnosti, což ji činí vhodnou pro některé real-time systémy.

RPN přijímá na vstupu snímek libovolné velikosti a vrací množinu obdélníkových návrhů regionů reprezentovaných vektory (x, y, w, h) , každý se dvěma

ohodnoceními určujícími, s jakou pravděpodobností region náleží do některé z klasifikačních tříd, resp. s jakou pravděpodobností se jedná o pozadí. Pro generování návrhů regionů je použita malá síť přejíždějící přes příznaky dodané poslední sdílenou konvoluční vrstvou. Každý $n \times n$ (zpravidla 3×3) výřez příznaků je transformován do ménědimenzionálního (zpravidla 256 nebo 512) vektoru, který je předán dvěma plně propojeným vrstvám – vrstvě pro regresi bounding boxů a vrstvě pro jejich klasifikaci.

Pro každý výřez je predikováno maximálně k návrhů regionů, které jsou relativní k referenčním bounding boxům, tzv. kotvám. Střed každé kotvy odpovídá středu příslušného výřezu. Kotva je definována velikostí (obvykle 128^2 , 256^2 a 512^2 pixelů) a poměrem stran (obvykle 1 : 1, 1 : 2, 2 : 1), což dává 9 kotev pro každý výřez.

Pro trénování RPN je každé kotvě přiřazeno binární označení, pozitivní, pokud je kotva třídou, a negativní, pokud kotva odpovídá pozadí. Jako pozitivní jsou označeny kotvy mající nejvyšší IoU s ground-truth boxem nebo kotvy, jejichž IoU s jakýmkoliv ground-truth boxem je vyšší než 0,7. Negativní označení je kotvě přiřazeno tehdy, pokud je její IoU se všemi ground-truth boxy nižší než 0,3. Neoznačené kotvy na další trénování nemají vliv. Ztrátová funkce pro celý systém je definována jako

$$L(\{p_i\}, \{t_i\}) = \frac{1}{N_{\text{cls}}} \sum_i L_{\text{cls}}(p_i, p_i^*) + \lambda \frac{1}{N_{\text{reg}}} \sum_i p_i^* L_{\text{reg}}(t_i, t_i^*), \quad (2.17)$$

kde i je index kotvy v mini-batchi, p_i predikovaná pravděpodobnost, že i . kotva je objektem, p_i^* je ground-truth označení, zda se jedná o třídu, nebo pozadí, t_i je vektor definující polohu predikovaného bounding boxu a t_i^* vektor definující polohu ground-truth bounding boxu odpovídajícímu pozitivně označené kotvě. Klasifikační ztráta L_{cls} je zlogaritmovaná ztráta na dvou třídách (objekt, nebo pozadí). Pro regresní ztrátu se využívá $L_{\text{reg}}(t_i, t_i^*) = R(t_i - t_i^*)$, kde R je L_1 ztrátová funkce používaná ve Fast R-CNN (rovnice 2.15). $p_i^* L_{\text{reg}}$ znamená, že regresní ztráta je aktivována pouze pro pozitivní kotvy ($p_i^* = 1$). Výstupem klasifikační, resp. regresní vrstvy jsou množiny $\{p_i\}$, resp. $\{t_i\}$. Základní implementace Faster R-CNN nastavuje balanční parametr $\lambda = 10$.

RPN je trénována pomocí algoritmu zpětného šíření chyby a SGD. Každý mini-batch je tvořen jediným snímkem, který obsahuje množinu pozitivních i negativních kotev, ze kterých je náhodně vybráno 256 kotev tak, aby poměr pozitivních a negativních byl 1 : 1. Pokud je pozitivních kotev méně, je výběr doplněn negativními kotvami.

Celý model byl trénován ve čtyřech etapách:

1. Trénování RPN inicializované modelem předtrénovaným na datasetu ImageNet a následně laděné pro úlohu predikování regionů.

2. Trénování samostatné detekční sítě použité ve Fast R-CNN, která kromě snímků přijímá regiony navržené již natrénovanou RPN. Pro inicializaci byl rovněž použit model předtrénovaný na datasetu ImageNet. V této fázi sítě ještě nesdílejí konvoluční vrstvy.
3. Pro inicializaci trénování RPN je využita detekční síť. Sdílené konvoluční vrstvy jsou zafixovány a dále trénovány pouze vrstvy unikátní pro RPN.
4. Trénování vrstev unikátních pro Fast R-CNN (stále se zafixovanými sdílenými konvolučními vrstvami).

V této chvíli RPN i Fast R-CNN sdílí stejné konvoluční vrstvy a tvoří ucelený systém.

2.4.4 Mask R-CNN

V roce 2017 byl představen model Mask R-CNN [39], který rozšiřuje Faster R-CNN o možnost segmentace instancí objektů. Ke dvěma paralelním výstupním větvím Faster R-CNN, totiž softmax klasifikační vrstvě a regresoru bounding boxů, přidává větev třetí pro simultánní predikci segmentačních masek jednotlivých objektů.

Na rozdíl od systémů, jejichž klasifikace závisí na predikcích masek, např. DeepMask [40], Mask R-CNN predikuje masku a třídu nezávisle. Je definována víceúčelová ztrátová funkce

$$L = L_{\text{cls}} + L_{\text{box}} + L_{\text{mask}}. \quad (2.18)$$

Klasifikační ztráta L_{cls} a ztrátová funkce bounding boxů L_{box} odpovídají obdobným ztrátovým funkcím Fast R-CNN. Vrstva zodpovědná za predikci masky vrací pro každý region Km^2 dimenzionální vektor kódující K binárních masek o velikosti $m \times m$, jednu pro každou třídu. Na všechny pixely jednotlivých masek je postupně aplikována sigmoida. Dále je pro regiony asociované s nějakou ground-truth třídou k definována binární relativní entropie L_{mask} . Definice umožňuje síti generovat masky nezávisle pro každou třídu, predikce masky a třídy je oddělená.

Maska kopíruje kontury objektu. Na rozdíl od názvů tříd nebo offsetů bounding boxů je maska přirozeně zpracovatelná pixel po pixelu pomocí konvolucí. To ale vyžaduje, aby příznaky regionů byly správně zarovnány tak, aby co nejlépe odpovídaly skutečné poloze objektu v prostoru, což si vyžádalo vývoj nové vrstvy RoIAlign.

RoIPool vrstva je ve Fast R-CNN využívána k extrakci malých (např. 7×7) map příznaků z každého regionu. Reálné souřadnice regionů kvantizuje do diskretních map příznaků. To však způsobuje nepřesné zarovnání extrahovaných příznaků oproti regionům. To nemusí mít vliv na klasifikaci, která je

na tyto malé změny dostatečně robustní, ale má negativní dopad na predikci masek. **RoIAlign** odstraňuje kvantizaci a správně zarovnává extrahované příznaky se vstupem. Místo kvantizace používá bilineární interpolaci [41].

Region je považován za pozitivní, pokud je jeho IoU s nějakým ground-truth bounding boxem alespoň 0,5. Ztrátová funkce L_{mask} je definována pouze pro pozitivní regiony. Cílovým bodem masky je průnik regionu a jeho ground-truth masky.

Každý mini-batch obsahuje 2 snímky na jedno GPU a každý obrázek má N vybraných regionů v poměru 1 : 3 ve prospěch negativních regionů. N se mění v závislosti na použité backbone síti.

Během testování je z každého snímku extrahováno 300 nebo 1000 regionů (rovněž v závislosti na použité backbone síti). Na regionech je proveden návrh bounding boxů a non-maximální suprese. Následně jsou ze 100 nejlépe ohodnocených boxů získány masky. Je možné získat K masek pro každý region, ale využívána je pouze každá k . maska, kdy k je třída predikovaná klasifikační částí modelu. Maska velikosti $m \times m$ je následně transformována na velikost regionu a binarizována prahem 0,5.

Vzhledem k tomu, že masky jsou počítány pouze pro 100 detekčních boxů, je Mask R-CNN pouze asi o 20 % výpočetně náročnější než samotná Faster R-CNN.

Analýza

Kapitola popisuje povahu snímaného objektu, jeho vady a možnosti, jak je zvýraznit. Dále popisuje čtyři navržené snímací soustavy.

3.1 Metodika měření

Sklo je pro snímání velmi náročným materiálem. Má lesklý povrch, který je náchylný k tvorbě odlesků, jež znemožňují bližší zkoumání objektu. Skleněné tyče jsou vyrobeny z čirého průhledného skla, nejsou tedy nijak kontrastní vůči pozadí.

Tyč je poměrně dlouhá, nasnímat ji celou najednou je nad možnosti dostupného laboratorního vybavení. Její tloušťka převyšuje hloubku ostroty, které je s průmyslovými objektivy možné dosáhnout. Je tedy vhodné tyč rozdělit na sektory a každý nasnímat ze všech čtyř stran tak, aby byl v dostatečném rozlišení zabrán celý povrch a postupně zaostřena celá tloušťka tyče.

Větší škrábance a nerovnosti okrajů tyčí jsou dobře viditelné, menší a užší škrábance již hůř. Jen málo patrné jsou bubliny a kamínky, které navíc většinou nelze od sebe rozeznat pouhým okem. Pouze školené oko je pak v běžných podmínkách schopno rozeznat smirkovou čáru. Je tedy nutné použít vhodné osvětlení tak, aby byly vady lépe viditelné.

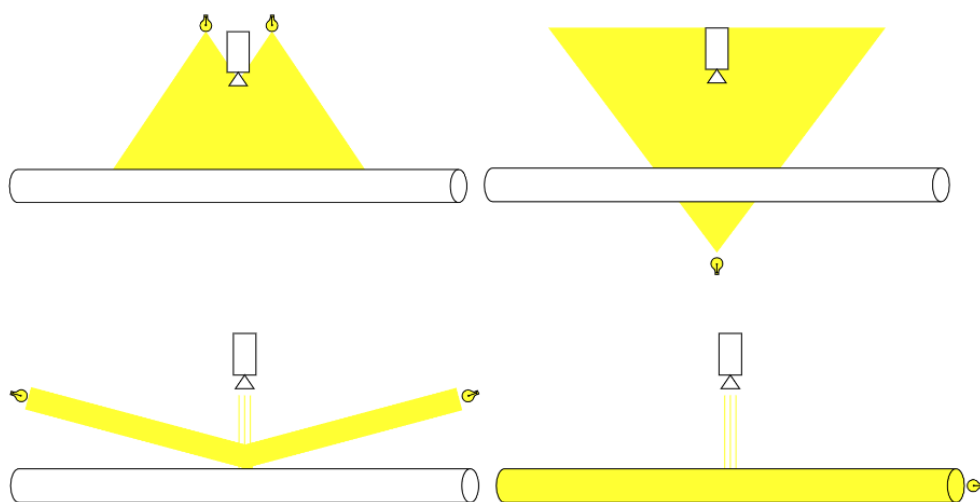
V úvahu přichází několik typů osvětlení (Obrázek 3.1):

- **přímé osvětlení ve světlém poli:** nezvýrazní kontrast vad oproti okolí, vytvoří mnoho odlesků, které vady naopak skryjí,
- **spodní osvětlení ve světlém poli:** netvoří odlesky viditelné kamerou, ale je nevhodné pro nasvícení příliš průsvitných vad (bubliny, smirková čára), které pak nejsou schopny dostatečně blokovat světlo a tvořit tak

dobře viditelné tmavé body v jinak velmi světlé scéně,

- **boční osvětlení v tmavém poli:** při osvětlení pod dostatečně malým úhlem již odražené paprsky nemíří přímo do objektivu, netvoří se tak odlesky. Povrchové vady jsou sice lépe viditelné, vady uvnitř tyčí však nikoliv,
- **boční osvětlení v tmavém poli s prosvícením tyče:** paprsky směřují vnitřkem skleněné tyče, pokud narazí na vadu, zlomí se a odrazí do jiných směrů, mj. i do objektivu. Jsou tak dobře viditelné vady, které kontrastují s tmavým okolím.

Jediný typ osvětlení, který zajišťuje viditelnost všech typů vad, absenci odlesků a maximální kontrast, je boční osvětlení v tmavém poli s prosvícením tyče.



Obrázek 3.1: Schémata typů osvětlení. Nahoře zleva přímé osvětlení ve světlém poli, spodní osvětlení ve světlém poli. Dole zleva boční osvětlení v tmavém poli, boční osvětlení v tmavém poli s prosvícením tyče.

3.2 Snímací soustavy

Byly navrženy čtyři snímací soustavy kombinující různé kamery, objektivy a vlnové délky osvětlení, přičemž se vždy jednalo o boční osvětlení v tmavém poli s prosvícením tyče.

3.2.1 Monochromatická kamera, modré osvětlení

Vhodným druhem kamer jsou monochromatické kamery, které nepotřebují **Bayerovu masku** [42], a jsou tudíž citlivější než kamery barevné.

Pro snímání byla použita plošná monochromatická kamera Basler acA2500-60um (parametry viz Tabulka 3.1) a objektiv Kowa LM35SC (parametry viz Tabulka 3.2).

Tabulka 3.1: Parametry kamery Basler acA2500-60um.

Parametr	Hodnota
Typ senzoru	CMOS
Velikost senzoru	12,4 mm × 9,8 mm
Pracovní spektrum	Viditelné
Rozlišení (výška × šířka)	2590 px × 2048 px
Rozlišení	5 MP
Velikost pixelu	4,8 μm × 4,8 μm

Tabulka 3.2: Parametry objektivu Kowa LM35SC.

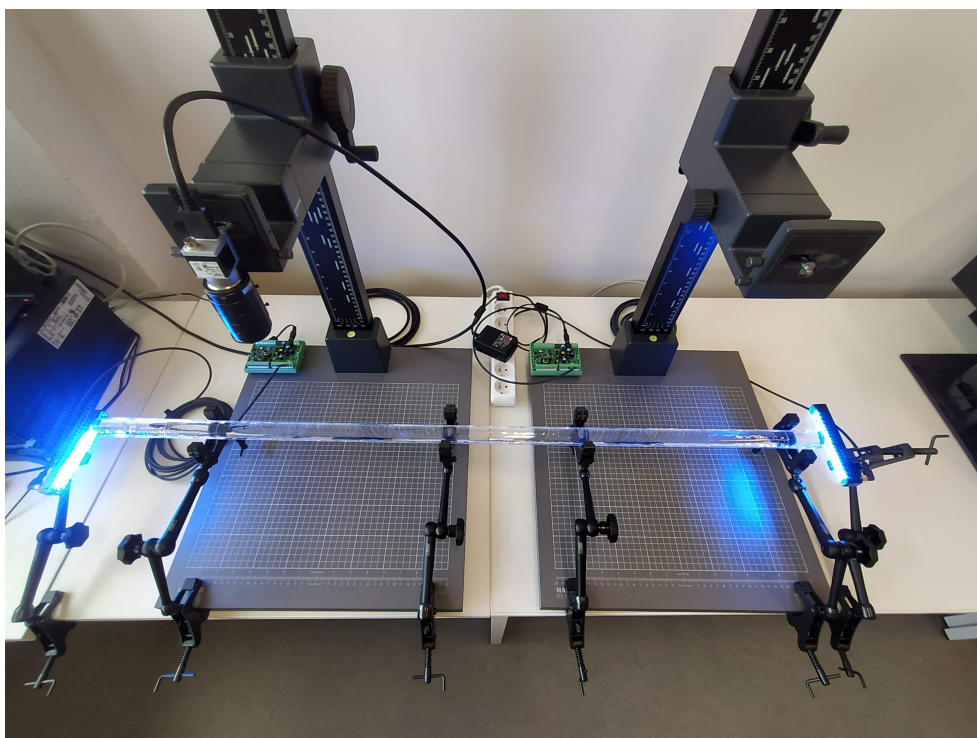
Parametr	Hodnota
Typ	S pevným ohniskem (fixed focal)
Ohnisková vzdálenost	35 mm
Světelnost	F2,0–16
Rozlišení	6 MP
Minimální pracovní vzdálenost	0,2 m

Osvětlení zajišťovala dvě modrá světla umístěná u jednoho nebo obou zcela svislých boků tyče tak, aby byl minimalizován vliv lomu světla. Tyč byla efektivně zevnitř „prosvětlena“, byly zvýrazněny povrchové i vnitřní vady a nebyly tvořeny odlesky. Použitá svítidla jsou běžně dostupná v červené, zelené a modré variantě. Modrá byla zvolena proto, že má z těchto tří barev nejnižší vlnovou délku a její vlastnosti se nejvíce blíží ultrafialovému záření, které zvláště dobře zvýrazňuje detaily (podsekce 2.1.2). Pro eliminaci denního světla, které snižovalo kontrast nasvícených vad oproti zbytku skleněné hmoty a tvořilo odlesky, byla soustava překryta černým dyftýnem, který účinně zabraňuje průchodu světelných paprsků [43]. Popsaná snímací soustava je na Obrázku 3.2.

Některé vady, zejména bubliny a kamínky, jsou velmi malé, bylo tedy nutno je co nejlépe nasvítit a zařídit, aby na senzor dopadalo největší možné množství světla. Z toho důvodu byla clona otevřena na maximum ($F = 2$). To ale mělo negativní vliv na hloubku ostrosti. Pokud byl objektiv zaostřen na povrch tyče, byly viditelné pouze kamínky nebo bubliny v blízkosti povrchu, ostatní drobné defekty se díky rozostření zcela ztratily. Z tohoto důvodu byly tyče snímány ze čtyř stran tak, aby byly pokryty všechny povrchové vady a postupně též celá tloušťka tyče.

3. ANALÝZA

Na snímcích pořízených touto soustavou jsou patrné škrábance, bubliny, kamínky a nerovnosti okrajů. Smirková čára byla zachycena pouze na některých místech, většina výskytů této vady nebyla odhalena. I přes relativně malou hloubku ostrosti jsou patrné škrábance nacházející se mimo zaostřenou oblast, zpravidla na straně vzdálenější od objektivu. Dále jsou místy viditelné odlesky, které se utvořily na bočních stranách tyče působením světla odraženého většími škrábanci. Je nutné, aby klasifikační algoritmus dokázal neostré útvary odlišit od zaostřených defektů.



Obrázek 3.2: Snímací soustava s monochromatickou kamerou a modrým osvětlením.

3.2.2 Barevná kamera, modré osvětlení

Navzdory výhodám monochromatické kamery popsané v podsekcí 3.2.1 byla vyzkoušena i zvláště citlivá plošná barevná kamera Basler a2A2590-60ucPRO, (parametry viz Tabulka 3.3). Citlivost je dána zejména malou velikostí pixelu. Dále byl použit objektiv LM35SC (parametry viz Tabulka 3.2).

Barevný výstup kamery byl pomocí programu Pylon Viewer [44] převeden na výstup monochromatický. Osvětlení bylo opět zajištěno modrými světly.

Díky své citlivosti byla kamera schopna zachytit všechny typy vad včetně

Tabulka 3.3: Parametry kamery Basler a2A2590-60ucPRO.

Parametr	Hodnota
Typ senzoru	CMOS
Velikost senzoru	5,2 mm × 3,9 mm
Pracovní spektrum	Viditelné
Rozlišení (výška × šířka)	2592 px × 1944 px
Rozlišení	5 MP
Velikost pixelu	2 μm × 2 μm

smirkové čáry. Vzhledem k hloubce ostrosti byla tyč také snímána ze čtyř stran. Díky schopnosti zachytit smirkovou čáru je tato kamera v kombinaci s osvětlením ve viditelném spektru považována za nejvhodnější.

3.2.3 UV kamera, UV osvětlení

Snímací soustava se od předchozích liší zejména typem použitého osvětlení, kterému byl přizpůsoben i výběr ostatních komponent. Nasvícení zajišťovalo UV svítidlo Effilux emitující záření o vlnové délce 365 nm. Dále byla použita monochromatická UV kamera jAi CM-140 GE-UV (parametry viz Tabulka 3.4) a objektiv FL-BC2528-VGUV - F2.8/25mm (parametry viz Tabulka 3.5) s optickou soustavou z křemenného skla.

Tabulka 3.4: Parametry kamery jAi CM-140 GE-UV.

Parametr	Hodnota
Typ senzoru	CCD
Velikost senzoru	6,5 mm × 4,8 mm
Pracovní spektrum	Viditelné + UV
Rozlišení (výška × šířka)	1040 px × 1392 px
Rozlišení	1,4 MP
Velikost pixelu	4,65 μm × 4,65 μm

Tabulka 3.5: Parametry objektivu FL-BC2528-VGUV - F2.8/25mm.

Parametr	Hodnota
Typ	S pevným ohniskem (fixed focal)
Ohnisková vzdálenost	25 mm
Světelnost	F2,8–16
Rozlišení (výška × šířka)	640 px × 480 px
Minimální pracovní vzdálenost	0,23 m

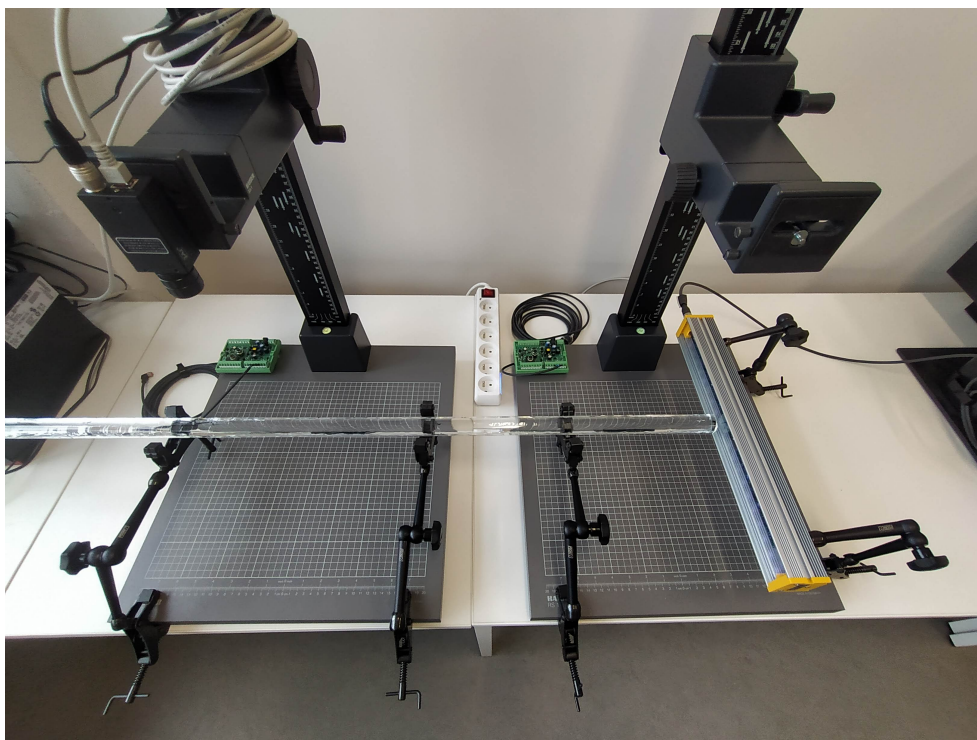
Nakonec byl použit UV filtr Midopt 365-25.4 propouštějící pouze záření o vlnové délce 335–400 nm. Filtr je zvláště vhodný pro použití v kombinaci s os-

3. ANALÝZA

větlením o vlnové délce 365 nm.

Snímací soustava se ukázala jako dosti problematická. Jak je patrné z Obrázku 3.3, je využita pouze malá část UV svítidla, pouze 2–3 diody. Svítidlo není možné přistavit k dlouhému boku tyče, jelikož je třeba s ní v průběhu snímání otáčet, aby mohla být zachycena ze všech čtyř stran. Z průřezu tyče na Obrázku 1.1 je patrné, že takto by se v některých případech svítilo skrz vypouklý bok, což není žádoucí, když v důsledku Snellova zákona [45] by bylo světlo přeměřováno a nasvícení skleněné hmoty, resp. defektů by pak nebylo rovnoměrné.

Nedostatečné osvětlení zneviditelnilo většinu bublin a kamínků, dokonce i mnohé úzké škrábance, které jinak byly zdaleka nejsnadněji snímatelymi vadami. Málo světla v záběru tedy vedlo k maximálnímu otevření clony, které však jen snížilo už dříve nedostatečnou hloubku ostrosti. Již počáteční experimenty prokázaly nevhodnost snímací soustavy, proto nebylo pořízeno větší množství dat.



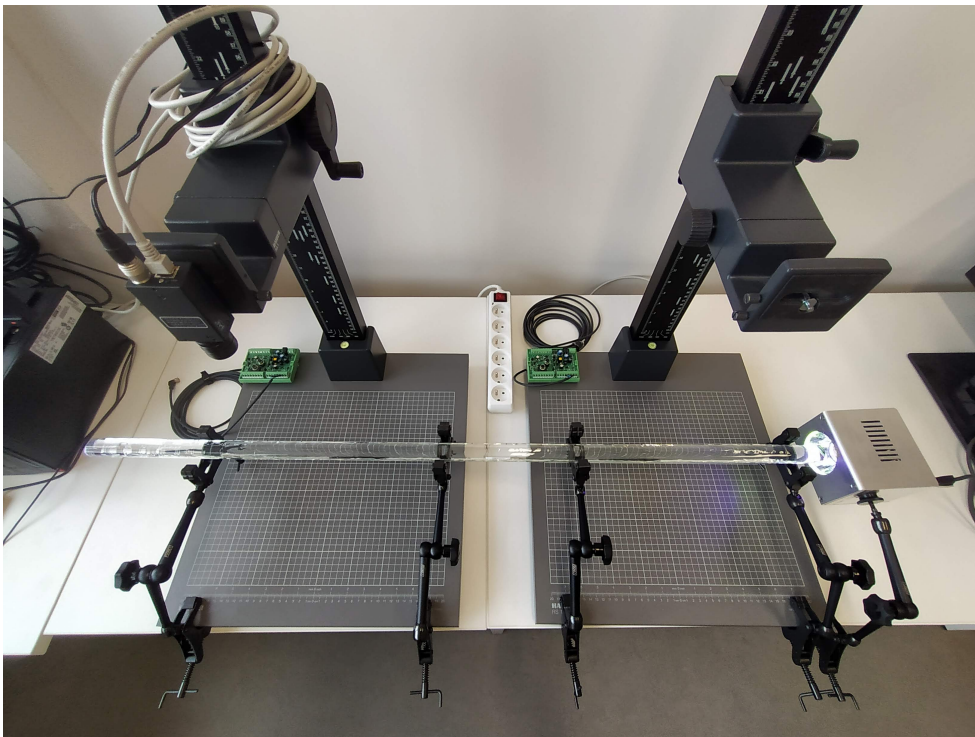
Obrázek 3.3: Snímací soustava s UV kamerou a UV osvětlením.

3.2.4 UV kamera, přizpůsobené UV osvětlení

Zásadním požadavkem na soustavu bylo dostatečně silné osvětlení. Proto bylo sestrojeno speciální bodové UV svítidlo tvarově přizpůsobené profilu snímaných tyčí. Jeho ústřední částí je silná dioda LZP-D0UB00-00U4 o výkonu 20 W, která produkuje záření vlnové délky 390 nm. Byla použita stejná kamera jAi CM-140 GE-UV (Tabulka 3.4), objektiv FL-BC2528-VGUV - F2.8/25mm (Tabulka 3.5) i filtr Midopt 365-25.4. Soustava je zachycena na Obrázku 3.4.

Vyšší výkon svítidla a jeho efektivní využití měly pozitivní dopad na kvalitu snímků a viditelnost vad všech typů a velikostí. Díky dostatečné světelnosti snímků mohla být přivřena clona, což vedlo ke zvýšení hloubky ostrosti. Experimentálně bylo zjištěno, že optimální hodnota clony je $F = 5,6$.

Na snímcích pořízených touto soustavou jsou patrné všechny zkoumané vady, tj. škrábance, bubliny, kamínky, smirková čára i nerovnosti okrajů.



Obrázek 3.4: Snímací soustava s UV kamerou a přizpůsobeným UV osvětlením.

Praktická část

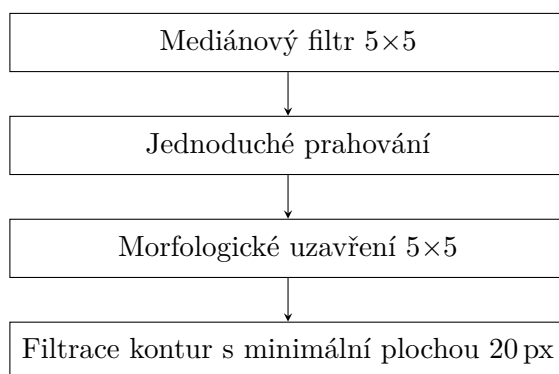
Systémy popsané v sekci 1.3 prováděly inspekci vad výhradně na tenkých skleněných deskách. Pro jejich snímání mohou být použity objektivy s nízkou hloubkou ostrosti, všechny vady je tedy možno dobře zaostřit na jediném snímku pořízeném shora. Optické tyče jsou mnohem tlustší a při relativně nízké pracovní vzdálenosti není vše viditelné najednou. Tyče je tedy nutno snímat z více stran, některé vady jsou viditelné z různých úhlů a při různém zaostření, což klade vysoké nároky na klasifikační algoritmus.

4.1 Klasifikace tvarových deskriptorů

Byl implementován algoritmus provádějící předzpracování obrazu, extrakci kontur, výpočet jejich tvarových charakteristik a finální klasifikaci.

4.1.1 Předzpracování

Bylo nutné najít dostatečně robustní algoritmus předzpracování, který by byl schopen ze všech snímků extrahovat defekty v podobě kontur. Nejprve byl na snímky aplikován mediánový filtr (podsekce 2.2.2) s kernelem o velikosti 5×5 , který ze snímků odstranil drobný šum. Následně bylo použito jednoduché prahování (podsekce 2.2.1), které efektivně segmentovalo jasně svítící bubliny, kamínky, škrábance a nerovnosti okrajů. Bylo též schopno dobře eliminovat poněkud tmavší rozostřené vady, které byly zaostřeny při snímání z jiných stran, a také některé odlesky na bocích tyčí způsobené lomem světla na větších škrábancích. Nakonec bylo provedeno morfologické uzavření (podsekce 2.2.5) s kernelem o velikosti 5×5 , které spojilo dohromady některé blízké kontury a zároveň nenarušilo jejich tvar. Pro další výpočet deskriptorů byly použity kontury o velikosti odpovídající alespoň nejmenším bublinám, tedy 20 pixelů. Schéma algoritmu je na Obrázku 4.1.



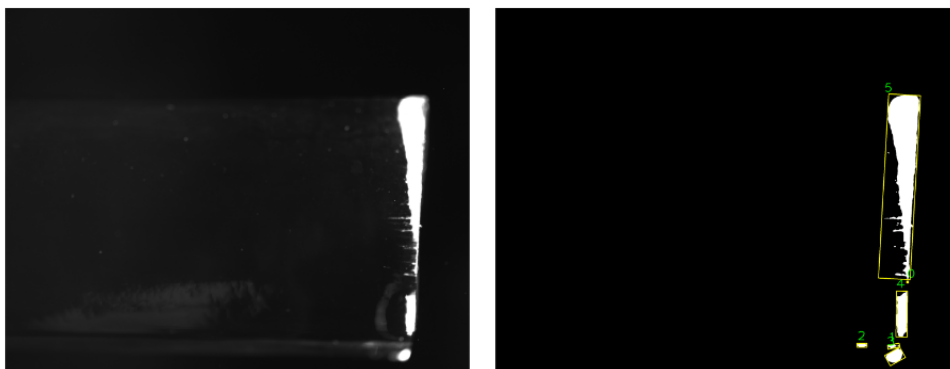
Obrázek 4.1: Schéma algoritmu předzpracování pro získání kontur.

4.1.2 Datasets

Byly pořízeny celkem tři datasety, a to pomocí snímací soustavy s monochromatickou kamerou a modrým osvětlením (podsekce 3.2.1, dále „dataset 1“), soustavy s barevnou kamerou a modrým osvětlením (podsekce 3.2.2, dále „dataset 2“) a soustavy s UV kamerou a specializovaným UV osvětlením (podsekce 3.2.4, dále „dataset 4“). Dataset 3 nebyl pořízen, neboť snímací soustava popsaná v podsekci 3.2.3 neumožnila pořízení dostatečně kvalitních snímků. Snímky byly anotovány ručně s použitím vlastní implementace, která po provedení předzpracování spočítala tvarové deskriptory kontur a kontury vykreslila do obrázků tak, aby bylo možno prostým pohledem určit, o který typ vady se jedná (Obrázek 4.2). Neidentifikovatelné vady byly označeny příznakem „ostatní“. Datasety byly rozděleny na část trénovací, validační a testovací v poměru 64 : 16 : 20. To odpovídá postupu, kdy byla data nejprve rozdělena v poměru 80 : 20 na trénovací a testovací a následně bylo 20 % trénovacích dat použito pro validaci. Podrobnější informace o datasetech jsou uvedeny v Tabulce 4.1.

Vyšší počet škrábanců a místy i nerovností okrajů je dán zejména jejich nekompaktností, která se pak při segmentaci projevila rozdělením vady do více kontur. Na více než polovině snímků též nebyly detekovány žádné kontury, což bylo dáno jednak přítomností snímků bez vad a také vlastnostmi algoritmu předzpracování, který ignoroval rozostřené nedostatečně svítící vady, které byly zaostřeny na snímcích pořízených z jiné strany.

Oproti tvarovým příznakům popsaných v podsekci 1.3.1 byla navíc vypočtena **konvexita** jako podíl obvodu kontury a její konvexní obálky. Nebyl počítán úhel sevřený osou x a hlavní osou elipsy. Pro klasifikaci byly použity všechny vypočítané příznaky. Hodnoty některých z nich, např. celistvost, excentricita nebo konvexita se pohybovaly mezi 0 a 1, jiné, např. plocha, dosahovaly mnohem vyšších hodnot. Aby nebyl klasifikační model příliš ovlivněn hodnotami



Obrázek 4.2: Předzpracování snímků pro extrakci kontur. Vlevo originální snímek, vpravo snímek po předzpracování s vykreslenými rotovanými bounding boxy kontur (žlutě) a jejich číselným označením (zeleně). Kontury 0 a 1 jsou ostatní objekty, kontura 2 je škrábanec, kontury 3, 4 a 5 jsou nerovnosti okraje.

s vyšším rozptylem, byly všechny příznaky **normalizovány** do intervalu $\langle 0, 1 \rangle$, tedy pro každý příznak byla v trénovacích datech nalezena minimální a maximální hodnota \min_x a \max_x a jednotlivé hodnoty byly upraveny podle rovnice

$$x_i \leftarrow \frac{x_i - \min_x}{\max_x - \min_x}. \quad (4.1)$$

4.1.3 Klasifikace

Pro klasifikaci byla použita vícevrstvá neuronová síť (podsekcce 2.3.2), konkrétně její implementace z knihovny Scikit-learn [46, 47]. Laděnými hyperparametry byly: počet neuronů ve skrytých vrstvách, aktivační funkce, optimalizátor vah, regularizační parametr, učicí parametr, počáteční hodnota učicího parametru a maximální počet iterací. Pro opakovatelnost výsledků byl zafixována počáteční hodnota náhodného generátoru iniciálních vah. Hyperparametry byly optimalizovány metodou prohledávání mřížky, kdy byly zkoušeny všechny kombinace hodnot z definovaných množin. Nejlepší přesnosti na validačních datech dosáhly sítě s hyperparametry uvedenými v Tabulce 4.2.

4.2 Mask R-CNN

Alternativou k algoritmu využívajícího tradiční techniky zpracování obrazu jsou metody R-CNN popsané v sekci 2.4. Nejmodernější metodou je Faster R-CNN (podsekcce 2.4.3), která navíc umožňuje provádět predikce v reálném čase. Není zde využito tradičních technik zpracování obrazu. Extrakce příznaků a návrh regionů jsou zajištěny konvolučními vrstvami, klasifikace pak

4. PRAKTICKÁ ČÁST

Tabulka 4.1: Parametry datasetů pro trénování klasifikátoru tvarových deskriptorů.

Parametr	Dataset 1	Dataset 2	Dataset 4
Trénovací data			
Celkový počet snímků	148	121	142
Počet snímků s anotacemi	47	42	49
Škrábance	172	200	122
Bublíny	24	16	2
Kamínky	13	4	4
Nerovnosti okrajů	42	89	27
Ostatní objekty	1	5	0
Validační data			
Celkový počet snímků	37	31	36
Počet snímků s anotacemi	18	8	12
Škrábance	59	52	51
Bublíny	9	2	0
Kamínky	4	1	0
Nerovnosti okrajů	16	18	14
Ostatní objekty	3	2	0
Testovací data			
Celkový počet snímků	47	39	45
Počet snímků s anotacemi	16	11	14
Škrábance	119	80	58
Bublíny	4	2	1
Kamínky	1	3	0
Nerovnosti okrajů	41	15	13
Ostatní objekty	10	1	0

provádějí plně propojené vrstvy. Vzhledem ke značné výpočetní náročnosti bylo trénování spouštěno v prostředí Google Colab.

4.2.1 Datasety

Model Mask R-CNN byl trénován a vyhodnocován na datasetech použitých i pro klasifikaci tvarových deskriptorů (sekce 4.1). Všechny snímky byly ručně anotovány pomocí programu VGG Image Annotator [49]. Vady byly aproximovány polygony a kruhy. Pro účely trénování, validace a testování byly načteny pouze snímky s alespoň jednou anotovanou vadou. Tím nebylo dotčeno vyvážené zastoupení pozadí – vady jsou vesměs velmi malé, většina regionů navrhovaných RPN tedy žádné neobsahovala. Podrobnější informace o datasetech a anotacích jsou uvedeny v Tabulce 4.3.

Tabulka 4.2: Hodnoty hyperparametrů vícevrstvé neuronové sítě pro klasifikaci tvarových deskriptorů.

Hyperparametr	Dataset 1	Dataset 2	Dataset 4
Počet neuronů ve skrytých vrstvách	5, 6	7, 7	5
Aktivační funkce	sigmoida	RELU	sigmoida
Optimalizátor vah	Adam [48]	Adam	Adam
Regularizační parametr	0,01	0,0001	0,0001
Učící parametr	konstantní	konstantní	konstantní
Počáteční hodnota učicího parametru	0,1	0,001	0,1
Maximální počet iterací	1000	1000	1000

Na všechny snímky trénovacího datasetu byla dále náhodně aplikována jedna z následujících technik data augmentation: převrácení snímku horizontálně, převrácení snímku vertikálně, rotace o náhodný úhel z intervalu $\langle -45^\circ, 45^\circ \rangle$. Spolu se snímky byly augmentací upraveny i příslušné masky. Při rotaci, kdy dochází ke změně rozměru snímku, byly pixely mimo snímek vyplněny nulami.

4.2.2 Trénování

Použitá implementace Mask R-CNN [50] nabízí základní framework pro detekci, klasifikaci a segmentaci instancí objektů, který je možné upravit na míru pořízeným datům. Jako základní síť pro extrakci příznaků byla použita architektura ResNet-101. Framework byl rozšířen o vlastní implementaci načítání snímku a jejich anotací, generování masek a vyhodnocování nad testovacím datasetem. Zároveň byly upraveny následující parametry modelu:

- IMAGES_PER_GPU: počet snímků, ze kterých jsou generovány regiony v jedné iteraci učení,
- STEPS_PER_EPOCH: počet úprav gradientu během jedné epochy trénování,
- VALIDATION_STEPS: počet validačních kroků,
- IMAGE_MIN_DIM: minimální délka kratší strany snímku,
- IMAGE_MAX_DIM: maximální délka delší strany snímku,
- RPN_ANCHOR_SCALES: 5 různých velikostí čtvercových kotev RPN v pixelech,
- TRAIN_ROIS_PER_IMAGE: počet návrhů regionů vybraných pro klasifikaci,
- MEAN_PIXEL: průměrná hodnota pixelů ve snímku, která je od hodnot snímku odečítána během extrakce příznaků.

Tabulka 4.3: Parametry datasetů pro trénování Mask R-CNN.

Parametr	Dataset 1	Dataset 2	Dataset 4
Trénovací data			
Celkový počet snímků	148	121	142
Počet snímků s anotacemi	81	71	113
Škrábance	79	57	157
Bubliny	33	30	119
Kamínky	16	19	32
Smirkové čáry	4	5	7
Nerovnosti okrajů	22	15	16
Validační data			
Celkový počet snímků	37	31	36
Počet snímků s anotacemi	24	19	29
Škrábance	26	10	36
Bubliny	7	15	22
Kamínky	4	2	11
Smirkové čáry	1	1	2
Nerovnosti okrajů	8	5	6
Testovací data			
Celkový počet snímků	47	39	45
Počet snímků s anotacemi	29	19	37
Škrábance	21	7	53
Bubliny	12	8	40
Kamínky	5	3	6
Smirkové čáry	2	3	1
Nerovnosti okrajů	9	7	6

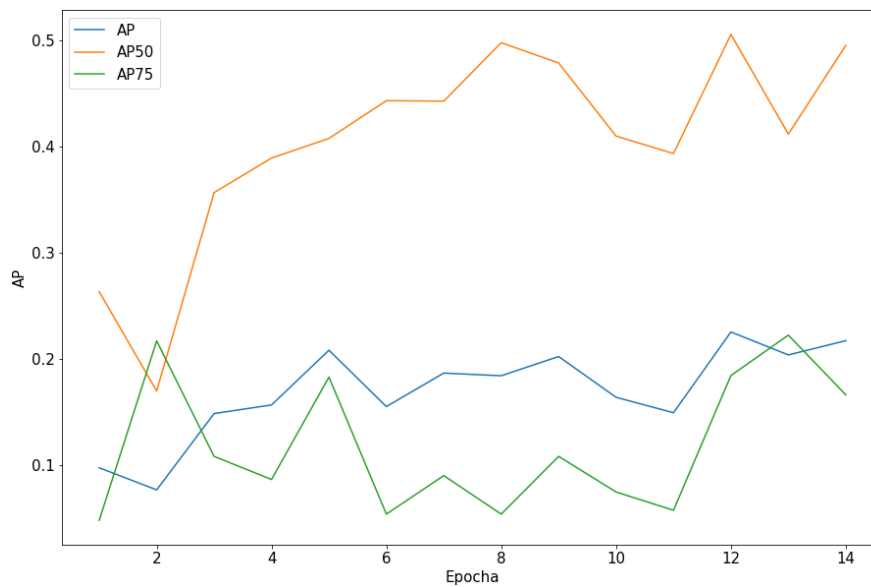
4.2.3 Validace a testování

K hodnocení modelu byly použity metriky AP, AP50 a AP75 podle COCO metodiky [10], kdy AP je průměr hodnot AP_{IoU} na hodnotách prahu IoU 0,5–0,95 s krokem 0,05, AP50 je AP s IoU prahem 0,5 a AP75 je AP s IoU prahem 0,75. Modely nad datasety 1 a 2 byly trénovány na 15 epoch, model nad datasetem 4 byl trénován na 30 epoch. Každá epocha obsahovala 200 iterací. Nejlepších výsledků bylo dosaženo s hyperparametry uvedenými v Tabulce 4.4.

Hodnoty metrik vypočtené na validačních datech po každé epoše trénování pro jednotlivé datasety postupně zobrazeny na obrázcích 4.3, 4.4 a 4.5.

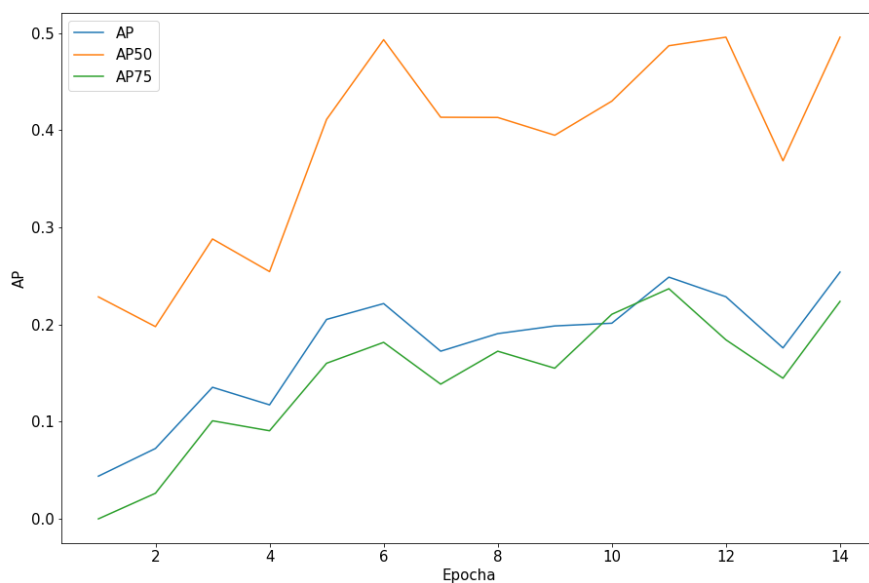
Tabulka 4.4: Hodnoty hyperparametrů Mask R-CNN.

Hyperparametr	Hodnota
Počet snímků v mini-batchi	1
Počet regionů na snímek	128
Učící parametr	0,001
Velikosti RPN kotev	16, 32, 64, 128, 256 px
Práh non-maximální suprese návrhů regionů	0,7
Práh jistoty detekcí	0,7
Práh non-maximální suprese detekcí	0,3
Penalizace vah ve ztrátové funkci	0,0001

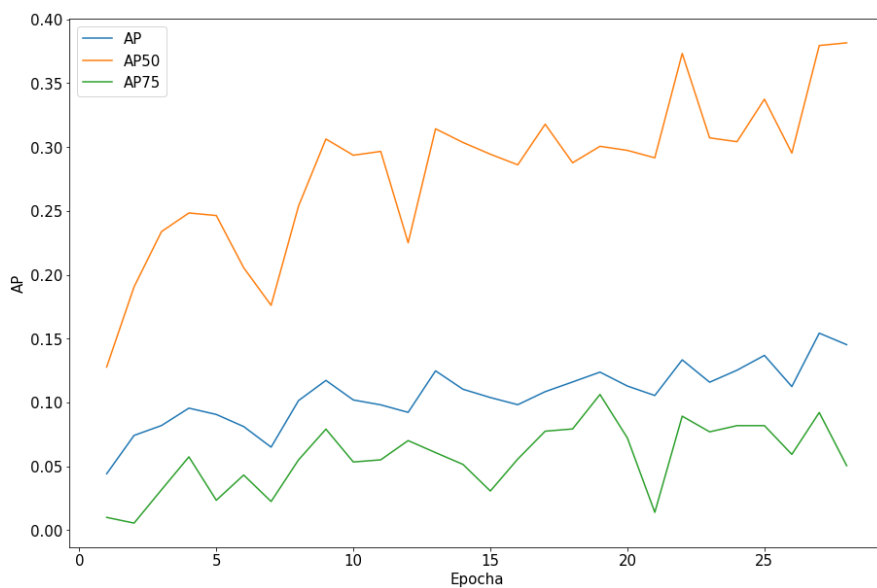


Obrázek 4.3: Hodnoty metrik AP, AP50 a AP75 během trénování na datasetu 1. Model dosahuje nejlepších výsledků po 8. epoše, následně se již hodnoty metriky AP50 drží na podobné úrovni.

4. PRAKTICKÁ ČÁST



Obrázek 4.4: Hodnoty metrik AP, AP50 a AP75 během trénování na datasetu 2. Nejlepších výsledků bylo dosaženo po 6., 11. a 14. epoše. Hodnoty AP a AP75 kopírují vývoj hodnot AP50 přesněji než na Obrázku 4.3.



Obrázek 4.5: Hodnoty metrik AP, AP50 a AP75 během trénování na datasetu 4. Model se zlepšuje pomaleji, nejlepších výsledků bylo dosaženo po 21. a 28. epoše.

Výsledky

Kapitola popisuje výsledky dosažené modelem pro klasifikaci tvarových deskriptorů a modelem Mask R-CNN.

5.1 Klasifikace tvarových deskriptorů

Algoritmus předzpracování popsaný v podsekcí 4.1.1 nebyl schopen extrahovat smirkovou čáru, která se jako jediná vada vyskytuje ve snímku spíše jako shluk bodů než jako ucelený útvar. Bylo experimentováno s různými technikami zvýšení kontrastu a adaptivního prahování (Otsuova metoda [51], Niblackovo [52] a Sauvolovo [5] prahování) tak, aby výsledný algoritmus fungoval univerzálně na všechny druhy vad. Ukázalo se však, že předzpracování vhodné pro snímky se smirkovou čarou není vhodné pro ostatní snímky a naopak. Metody adaptivního prahování měly tendenci segmentovat téměř vše, co bylo světlejší než černé pozadí a ani zvětšení kernelu nevedlo k uspokojivým výsledkům. Vzhledem k převažujícímu zastoupení ostatních druhů vad bylo od segmentování smirkové čáry upuštěno.

Klasifikační modely dosahovaly velmi vysoké přesnosti (95–99 %) na konturách škrábanců, nerovnosti okrajů byly na datasetech 1 a 4 klasifikovány s přesností okolo 50 %, na datasetu 2 dokonce s přesností 100 %. Zcela však selhaly při klasifikaci bublin a kamínků, kdy všechny takto označené kontury vyhodnotily jako škrábance. Nezařazené kontury označené jako „ostatní“ byly klasifikovány většinou jako škrábance, místy jako nerovnost okraje.

5.2 Mask R-CNN

Algoritmus založený na Mask R-CNN ve většině případů dokázal vady správně detekovat a klasifikovat. Místy byla vada detekována správně, ale byla jí přiřazena špatná třída. V drtivé většině případů se jednalo o záměnu bub-

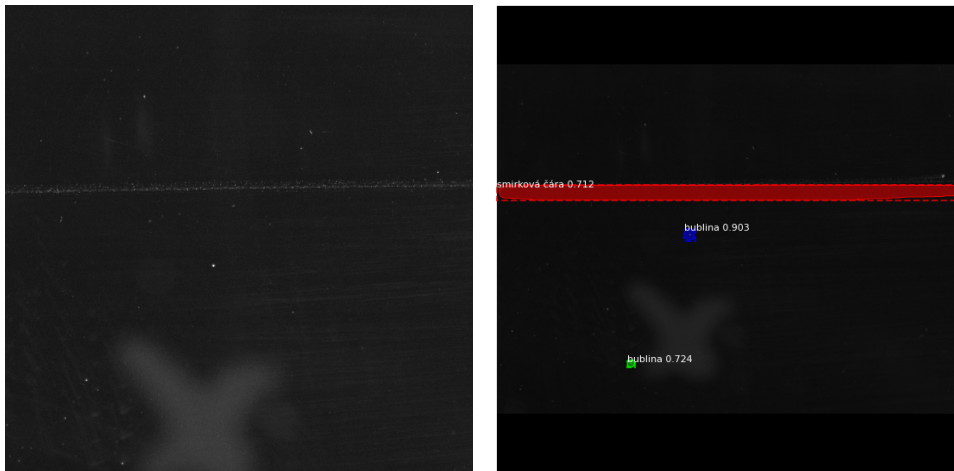
5. VÝSLEDKY

liny za kamínek a naopak. Z detekcí na testovacím datasetu totiž vyplývá, že algoritmus se naučil detekovat drobné kulaté ostré útvary jako bubliny, větší, ne nutně kulaté a ne nutně zcela ostré útvary jako kamínky, ostatní, zpravidla nikoliv kulaté malé útvary jako škrábance (Obrázek 5.1). Pokud byl kamínek v záběru nevýrazný, byl zpravidla detekován jako bublina, naopak větší bubliny byly detekovány jako kamínky. Schopnosti detekce smirkových čar (Obrázek 5.2) a nerovností okrajů se lišily podle datasetu, na kterém byl model trénován. Místy byl za vadu označen neanotovaný objekt, zpravidla se jednalo o drobné bubliny nebo škrábance. Po přezkoumání snímku a tyče v odpovídající oblasti však bylo ve většině případů zjištěno, že detekce byla správná. Výsledky modelu na jednotlivých datasetech jsou rozepsány v Tabulce 5.1.



Obrázek 5.1: Nahoře originální snímek, dole detekce drobných defektů se zobrazeny jistotami modelu, že se skutečně jedná o objekt příslušné třídy.

Mask R-CNN trénovaný na datasetu 1 měl problémy s detekcí smirkové čáry, jednou se v jejím případě dopustil i falešné detekce. Je to dáno zejména ne-



Obrázek 5.2: Vlevo originální snímek, vpravo detekce smirkové čáry.

Tabulka 5.1: Výsledky Mask R-CNN.

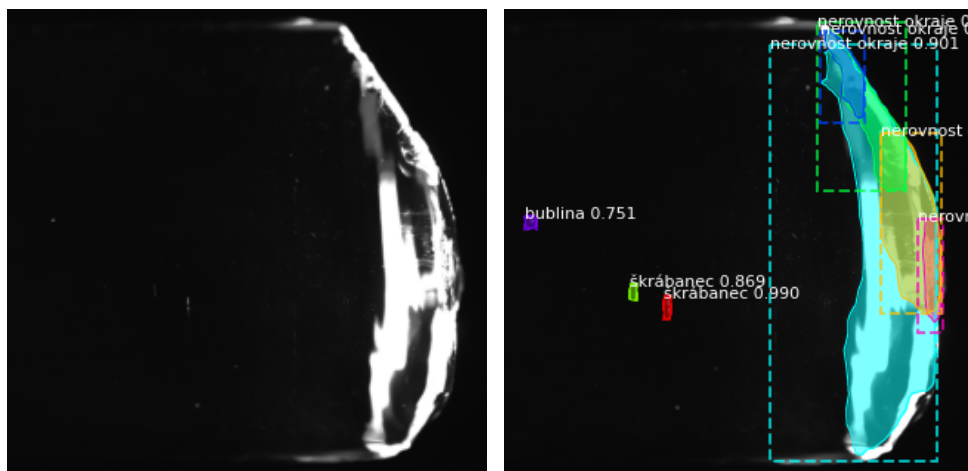
Kategorie	Zastoupení detekcí modelu nad datasetem	Zastoupení detekcí modelu nad datasetem	Zastoupení detekcí modelu nad datasetem
	1	2	4
Pravá detekce, správná třída	71 %	65 %	77 %
Pravá detekce, nesprávná třída	13 %	8 %	6 %
Falešná detekce	6 %	8 %	7 %
Nedetkováno	10 %	19 %	10 %

dostatečnou viditelností vady na snímcích a jejím málo častým výskytem. Nerovnosti okrajů byly detekovány až na výjimky správně, stávalo se však, že větší oblast jedné nerovnosti okraje byla detekována jako více menších nerovností (Obrázek 5.3), to však v praxi není problém, jelikož požadavek na detekci nerovnosti okraje je binární.

Díky lepší viditelnosti smirkové čáry v datasetu 2 byl model schopen tuto vadu většinou detekovat správně. Zlepšila se také detekce nerovností okrajů, predikované regiony již byly větší, kompaktnější a lépe odpovídaly anotovanému tvaru. Zároveň je z Tabulky 5.1 patrné, že došlo ke zlepšení přesnosti v přiřazení třídy jednotlivým vadám, zároveň se ale zmenšil podíl správně klasifikovaných vad a narostl objem těch, které model nebyl schopen zachytit.

Drtivá většina falešných detekcí modelu trénovaném na datasetu 4 byly smirkové čáry v místech odlesků na bocích tyče, což potvrzuje problém s nedostatečným zastoupením těchto vad v datasetech. Zároveň došlo k jedné ra-

5. VÝSLEDKY



Obrázek 5.3: Detekce nerovnosti okraje. Vlevo původní snímek, vpravo snímek s detekcemi. Celá nerovnost okraje je rozdělena do více částí, zvláště jsou detekovány části nerovnosti s odlišnou texturou. Snímek obsahuje i jednu falešnou detekci škrábance (drobná zelená maska vlevo).

ritní detekci, kdy byla detekována nerovnost okraje na boku tyče, kde se však skutečně vyskytovaly povrchové nerovnosti v podobě vroubkování.

Diskuze

Snímací soustavy bylo třeba navrhnout tak, aby byly vidět i ty nejmenší vady. To vedlo mj. k maximálnímu otevření clony, což ale snižovalo hloubku ostrosti a ztěžovalo pořízení kvalitních snímků. Optimální vyvážení těchto faktorů zajistila soustava s UV kamerou a specializovaným UV osvětlením, která poskytla dostatečně silné osvětlení vhodné vlnové délky, díky čemuž bylo možno přivřít clonu až na úroveň tzv. „sweetspotu“ – hodnoty clony, kdy je optimálně vyvážená ostrost a množství světla dopadajícího na snímač. Ještě kvalitnější snímky by mohla přinést jiná metodika snímání, např. použití makroobjektivu s vyšší ohniskovou vzdáleností, který by mohl snímat menší části povrchu tyčí a dosahovat tak lepšího rozlišení a přesnosti.

Aby byla zachycena celá tloušťka tyče, byla tyč snímána postupně ze čtyř stran. Alternativou k tomuto přístupu by mohlo být snímání pouze ze dvou stran. Nejprve by tyč byla snímána z jedné strany, a postupně posouvána hloubka ostrosti tak, aby byla tyč postupně proostřena v celé tloušťce. Následně by byla tyč otočena o 90° a byly by pořízeny snímky zaostřené pouze na nyní vodorovné stěny tyče, které byly během první fáze snímání svislé, a drobné povrchové vady na nich byly obtížně detekovatelné. Tento přístup by zároveň eliminoval opakované snímání vad z různých stran a umožnil by tak vytváření přesnějších statistik výskytů vad.

Algoritmus předzpracování pro extrakci kontur vad nebyl vždy schopen odstranit všechny šum a nečistoty. Pro efektivnější vyčištění snímku by bylo vhodné použít mediánový filtr s větším kernelem, nicméně ten už kromě šumu odstraňoval i drobné bubliny a kamínky. Po segmentaci a extrakci kontur nebylo možné určit přesný počet detekovaných vad, jelikož některé škrábance a nerovnosti okrajů byly během předzpracování rozděleny na více samostatných kontur.

Model pro klasifikaci tvarových deskriptorů dosahoval přijatelných výsledků

pouze pro škrábance a nerovnosti okrajů, bubliny a kamínky nebyl schopen klasifikovat vůbec, smirkovou čáru vzhledem k její jasové podobnosti s odlesky na stěnách tyčí a nečistotami ani nedokázal extrahovat ze snímků. Neschopnost klasifikace bublin a kamínků může být dána jednak zaměnitelností tvarových charakteristik vad navzájem, jednak zaměnitelností s malými škrábanci. Zásadní vliv mělo také jejich nedostatečné zastoupení v datasetech. Zlepšení by s největší pravděpodobností přinesly větší datasety s rovnoměrněji zastoupenými vadami, rozšíření množiny extrahovaných příznaků, případně zcela jiný typ příznaků, které by lépe rozlišily drobné vady. Trendem je extrakce příznaků pomocí konvolučních neuronových sítí.

Algoritmus založený na modelu Mask R-CNN dosahoval uspokojivých výsledků i na relativně malých datasetech s nerovnoměrným zastoupením vad. Jen málokdy se dopouštěl falešné detekce, zpravidla se jednalo o odlesk, který byl podobný nerovnosti okraje, škrábanci nebo smirkové čáře. Ojedinele docházelo k detekci smirkové čáry v oblasti zcela prosté vad, to však bylo důsledkem jejího málo častého výskytu. Model také často rozděloval větší nerovnosti okrajů do několika menších detekcí, což však není problém, jelikož v praxi je část tyče s nerovností okraje nepoužitelná bez ohledu na to, kolikrát se zde vada vyskytuje.

Prostor je také pro další ladění početných hyperparametrů modelu. Mask R-CNN je komplexní model, který je vhodné trénovat na velkých datasetech, pomohlo by tedy také pořízení dalších dostatečně rozmanitých snímků.

Při manipulaci se sklem na jeho povrchu ulpívají různé nečistoty, např. tukové částičky z dlaní nebo rukavic. Tyto nečistoty pak při nasvícení mají tvar odpovídající drobným bublinám nebo škrábancům, což může mást klasifikační model a způsobovat falešné detekce. Bylo by tedy vhodné, aby před spuštěním detekčního algoritmu v průmyslovém prostředí nebylo s tyčí nijak ručně manipulováno.

Závěr

Práce se zabývala problematikou snímání skleněných materiálů, detekcí a klasifikací povrchových vad skleněných tyčí. Byly sestaveny čtyři snímací soustavy, na kterých byly analyzovány různé optické prvky a osvětlení. Byly pořízeny tři datasety, na kterých byly testovány dva algoritmy pro detekci a klasifikaci vad vycházející z analyzovaných řešení.

Algoritmus využívající tradičních metod zpracování obrazu byl schopen ze snímků extrahovat čtyři z pěti typů vad, souběžně použitý model strojového učení pak byl schopen s dostatečnou přesností klasifikovat tři typy. Algoritmus využívající pokročilých metod strojového učení byl na vhodném datasetu schopen detekovat 83 % vad a správně klasifikovat 77 % vad. Jeho limitací byla poměrně malá velikost datasetů a nedostatečné zastoupení některých druhů vad.

Stanovené cíle práce byly splněny, byly sestaveny čtyři snímací soustavy s využitím pokročilých optických prvků a osvětlení a navržen algoritmus pro detekci a klasifikaci vad. Navržené řešení má potenciál uplatnění v průmyslové výrobě optických skleněných tyčí.

Literatura

- [1] Glass Rods Liba 2000+ [online]. Preciosa-Ornela, a. s. [Citováno 2021-03-19]. Dostupné z: <https://www.preciosa-ornela.com/content/files/downloads/glass-rods-liba-2000-2018.pdf>
- [2] Krutský, J.: EcoGlass Lens Catalogue [online]. EcoGlass, a. s. 2020: s. 3, 5, [Citováno 2021-03-19]. Dostupné z: https://www.ecoglass-optic.com/storage/app/media/download/ECG_lens_catalogue.pdf
- [3] Naresh, V.: Homogeneity: bubble formation in the manufacturing of glass [online]. 2015, překlad vlastní. [Citováno 2021-03-20]. Dostupné z: https://www.researchgate.net/publication/273762653_Homogeneity_bubble_formation_in_the_manufacturing_of_glass
- [4] Hassan, J.; Awan, A. M.; Jalil, A.: Welding Defect Detection and Classification Using Geometric Features [online]. In *2012 10th International Conference on Frontiers of Information Technology*, 2012, s. 139–144, doi:10.1109/FIT.2012.33, [Citováno 2021-04-20]. Dostupné z: <https://doi.org/10.1109/FIT.2012.33>
- [5] Sauvola, J.; Pietikäinen, M.: Adaptive document image binarization [online]. *Pattern Recognition*, ročník 33, č. 2, 2000: s. 225–236, ISSN 0031-3203, doi:10.1016/S0031-3203(99)00055-2, [Citováno 2021-04-20]. Dostupné z: [https://doi.org/10.1016/S0031-3203\(99\)00055-2](https://doi.org/10.1016/S0031-3203(99)00055-2)
- [6] Park, J.; Riaz, H.; Kim, H.; aj.: Advanced cover glass defect detection and classification based on multi-DNN model [online]. *Manufacturing Letters*, ročník 23, 2020: s. 53–61, ISSN 2213-8463, doi:10.1016/j.mfglet.2019.12.006, [Citováno 2021-03-23]. Dostupné z: <https://doi.org/10.1016/j.mfglet.2019.12.006>
- [7] Shorten, C.; Khoshgoftaar, T. M.: A survey on Image Data Augmentation for Deep Learning [online]. *Journal of Big Data*, ročník 6, 2019: s. 60–107,

- ISSN 2196-1115, doi:10.1186/s40537-019-0197-0, [Citováno 2021-03-23]. Dostupné z: <https://doi.org/10.1186/s40537-019-0197-0>
- [8] Redmon, J.; Farhadi, A.: YOLO9000: Better, Faster, Stronger [online]. *CoRR*, ročník abs/1612.08242, 2016, [Citováno 2021-03-23]. Dostupné z: <http://arxiv.org/abs/1612.08242>
- [9] Pan, Z.; Yang, J.; er Wang, X.; aj.: Image-based surface scratch detection on architectural glass panels using deep learning approach [online]. *Construction and Building Materials*, ročník 282, 2021, ISSN 0950-0618, [Citováno 2021-03-24]. Dostupné z: <https://www.sciencedirect.com/science/article/pii/S0950061821004773>
- [10] Lin, T.-Y.; Maire, M.; Belongie, S.; aj.: Microsoft COCO: Common Objects in Context [online]. In *Computer Vision – ECCV 2014*, editace D. Fleet; T. Pajdla; B. Schiele; T. Tuytelaars, Cham: Springer International Publishing, 2014, ISBN 978-3-319-10602-1, s. 740–755, [Citováno 2021-03-28]. Dostupné z: https://link.springer.com/chapter/10.1007/978-3-319-10602-1_48
- [11] Hui, J.: mAP (mean Average Precision) for Object Detection [online]. duben 2019, [Citováno 2021-03-28]. Dostupné z: <https://jonathan-hui.medium.com/map-mean-average-precision-for-object-detection-45c121a31173>
- [12] Rosebrock, A.: Intersection over Union (IoU) for object detection [online]. listopad 2016, [Citováno 2021-03-28]. Dostupné z: <https://www.pyimagesearch.com/2016/11/07/intersection-over-union-iou-for-object-detection/>
- [13] Bolya, D.; Zhou, C.; Xiao, F.; aj.: YOLACT: Real-Time Instance Segmentation [online]. In *Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV)*, říjen 2019, [Citováno 2021-03-28]. Dostupné z: https://openaccess.thecvf.com/content_ICCV_2019/html/Bolya_YOLACT_Real-Time_Instance_Segmentation_ICCV_2019_paper.html
- [14] Jiřina, M.; Novák, J.; Brchl, L.: Kamerový systém a zpracování obrazu [online]. říjen 2018, [Citováno 2021-03-29]. Dostupné z: <https://github.com/ImprolabFIT/BI-SVZ-coursework/blob/master/lectures/files/bi-svz-03-kamerovy-system-a-zpracovani-obrazu.pdf>
- [15] MacAdam, D. L.: Photometric Relationships Between Complementary Colors* [online]. *Journal of the Optical Society of America*, ročník 28, č. 4, duben 1938: s. 103–111, doi:10.1364/JOSA.28.000103, [Citováno 2021-03-29]. Dostupné z: <http://doi.org/10.1364/JOSA.28.000103>

-
- [16] Fermum, L.: IR illumination [online]. [Citováno 2021-03-29]. Dostupné z: <https://www.vision-doctor.com/en/ir-illumination.html>
- [17] Fermum, L.: UV illumination [online]. [Citováno 2021-03-29]. Dostupné z: <https://www.vision-doctor.com/en/uv-illumination.html>
- [18] Isteez: Opencv: Image Filtering [online]. , č. 4.5.2-dev, duben 2021, [Citováno 2021-04-21]. Dostupné z: https://docs.opencv.org/master/d4/d86/group__imgproc__filter.html#gac05a120c1ae92a6060dd0db190a61afa
- [19] Thiruvenkadam, K.; Padmanaban, S.; Karuppanagounder, S.: Survey of using GPU CUDA programming model in medical image analysis [online]. *Informatics in Medicine Unlocked*, ročník 9, srpen 2017, doi: 10.1016/j.imu.2017.08.001, [Citováno 2021-04-21]. Dostupné z: <https://doi.org/10.1016/j.imu.2017.08.001>
- [20] Veyselöglu, B.: OpenCV - Examples [software]. červenec 2019, [Citováno 2021-04-21]. Dostupné z: <https://github.com/bveyselöglu/OpenCV-Examples>
- [21] HMC: High-boost filtering [online]. [Citováno 2021-04-21]. Dostupné z: <http://fourier.eng.hmc.edu/e161/lectures/gradient/node2.html>
- [22] Jiřina, M.; Novák, J.; Brchl, L.: Segmentace obrazu - hranové [online]. květen 2020, [Citováno 2021-04-21]. Dostupné z: <https://github.com/ImprolabFIT/BI-SVZ-coursework/blob/master/lectures/files/bi-svz-08-segmentace-obrazu-hranove.pdf>
- [23] Isteez: Opencv: Canny Edge detection [online]. , č. 4.5.2-dev, duben 2021, [Citováno 2021-04-21]. Dostupné z: https://docs.opencv.org/master/da/d22/tutorial_py_canny.html
- [24] Isteez: Opencv: Eroding and Dilating [online]. , č. 4.5.2-dev, duben 2021, [Citováno 2021-04-21]. Dostupné z: https://docs.opencv.org/master/db/df6/tutorial_erosion_dilatation.html
- [25] McCulloch, W. S.; Pitts, W.: A logical calculus of the ideas immanent in nervous activity [online]. *The bulletin of mathematical biophysics*, ročník 5, prosinec 1943: s. 115–133, ISSN 1522-9602, doi:10.1007/BF02478259, [Citováno 2021-04-29]. Dostupné z: <https://doi.org/10.1007/BF02478259>
- [26] Rosenblatt, F.: The perceptron: a probabilistic model for information storage and organization in the brain. [online]. *Psychological review*, ročník 65, č. 6, 1958: str. 386, doi:10.1037/h0042519, [Citováno 2021-04-29]. Dostupné z: <https://doi.org/10.1037/h0042519>

- [27] Cybenko, G.: Approximation by superpositions of a sigmoidal function [online]. *Mathematics of control, signals and systems*, ročník 2, č. 4, 1989: s. 303–314, doi:10.1007/BF02551274, [Citováno 2021-04-29]. Dostupné z: <https://doi.org/10.1007/BF02551274>
- [28] Rumelhart, D. E.; Hinton, G. E.; Williams, R. J.: Learning representations by back-propagating errors [online]. *Nature*, ročník 323, č. 6088, 1986: s. 533–536, doi:10.1038/323533a0, [Citováno 2021-04-29]. Dostupné z: <https://doi.org/10.1038/323533a0>
- [29] Girshick, R. B.; Donahue, J.; Darrell, T.; aj.: Rich feature hierarchies for accurate object detection and semantic segmentation [online]. *CoRR*, ročník abs/1311.2524, 2013, [Citováno 2021-04-14]. Dostupné z: <http://arxiv.org/abs/1311.2524>
- [30] Girshick, R. B.: Fast R-CNN [online]. *CoRR*, ročník abs/1504.08083, 2015, [Citováno 2021-04-14]. Dostupné z: <http://arxiv.org/abs/1504.08083>
- [31] Ren, S.; He, K.; Girshick, R. B.; aj.: Faster R-CNN: Towards Real-Time Object Detection with Region Proposal Networks [online]. *CoRR*, ročník abs/1506.01497, 2015, [Citováno 2021-04-14]. Dostupné z: <http://arxiv.org/abs/1506.01497>
- [32] Uijlings, J.; Sande, K.; Gevers, T.; aj.: Selective Search for Object Recognition [online]. *International Journal of Computer Vision*, ročník 104, 09 2013: s. 154–171, doi:10.1007/s11263-013-0620-5, [Citováno 2021-04-14]. Dostupné z: <https://doi.org/10.1007/s11263-013-0620-5>
- [33] Felzenszwalb, P. F.; Huttenlocher, D. P.: Efficient Graph-Based Image Segmentation [online]. *International Journal of Computer Vision*, ročník 59, září 2004: s. 167–181, ISSN 1573-1405, doi:10.1023/B:VISI.0000022288.19776.77, [Citováno 2021-04-14]. Dostupné z: <https://doi.org/10.1023/B:VISI.0000022288.19776.77>
- [34] Jia, Y.; Shelhamer, E.; Donahue, J.; aj.: Caffe: Convolutional Architecture for Fast Feature Embedding [online]. *CoRR*, ročník abs/1408.5093, 2014, [Citováno 2021-04-14]. Dostupné z: <http://arxiv.org/abs/1408.5093>
- [35] Cortes, C.; Vapnik, V.: Support-vector networks [online]. *Machine Learning*, ročník 20, září 1995: s. 273–297, ISSN 1573-0565, doi:10.1007/BF00994018, [Citováno 2021-04-14]. Dostupné z: <https://doi.org/10.1007/BF00994018>
- [36] Deng, J.; Dong, W.; Socher, R.; aj.: ImageNet: A large-scale hierarchical image database [online]. In *2009 IEEE Conference on Computer Vision and Pattern Recognition*, 2009, s. 248–255, doi:10.1109/

- CVPR.2009.5206848, [Citováno 2021-04-14]. Dostupné z: <https://doi.org/10.1109/CVPR.2009.5206848>
- [37] Russakovsky, O.; Deng, J.; Su, H.; aj.: ImageNet Large Scale Visual Recognition Challenge [online]. *International Journal of Computer Vision (IJCV)*, ročník 115, č. 3, 2015: s. 211–252, doi:10.1007/s11263-015-0816-y, [Citováno 2021-04-30]. Dostupné z: <https://doi.org/10.1007/s11263-015-0816-y>
- [38] Canévet, O.; Fleuret, F.: Efficient sample mining for object detection [online]. In *Asian Conference on Machine Learning*, PMLR, 2015, s. 48–63, [Citováno 2021-05-08]. Dostupné z: <http://proceedings.mlr.press/v39/canevet14a.html>
- [39] He, K.; Gkioxari, G.; Dollár, P.; aj.: Mask R-CNN [online]. In *Proceedings of the IEEE International Conference on Computer Vision (ICCV)*, říjen 2017, [Citováno 2021-03-28]. Dostupné z: https://openaccess.thecvf.com/content_iccv_2017/html/He_Mask_R-CNN_ICCV_2017_paper.html
- [40] Pinheiro, P. H. O.; Collobert, R.; Dollár, P.: Learning to Segment Object Candidates [online]. *CoRR*, ročník abs/1506.06204, 2015, [Citováno 2021-04-16], 1506.06204. Dostupné z: <http://arxiv.org/abs/1506.06204>
- [41] Jaderberg, M.; Simonyan, K.; Zisserman, A.; aj.: Spatial Transformer Networks [online]. *CoRR*, ročník abs/1506.02025, 2015, [Citováno 2021-04-16], 1506.02025. Dostupné z: <http://arxiv.org/abs/1506.02025>
- [42] Bayer, B. E.: Color imaging array [online]. 1975, [Citováno 2021-03-30]. Dostupné z: <https://patents.google.com/patent/US3971065A/en>
- [43] Wikipedie: Dyftýn — Wikipedie: Otevřená encyklopedie [online]. 2020, [Citováno 2021-03-30]. Dostupné z: <https://cs.wikipedia.org/w/index.php?title=Dyft%C3%BDn&oldid=18626450>
- [44] AG, B.: Pylon Viewer [software]. [Citováno 2021-03-30]. Dostupné z: <https://www.baslerweb.com/en/products/software/basler-ylon-camera-software-suite/pylon-viewer/>
- [45] Weik, M. H.: *Snell's law [online]*. Boston: Springer US, 2001, ISBN 978-1-4020-0613-5, str. 1607, doi:10.1007/1-4020-0613-6_17633, [Citováno 2021-05-01]. Dostupné z: https://doi.org/10.1007/1-4020-0613-6_17633
- [46] Pedregosa, F.; Varoquaux, G.; Gramfort, A.; aj.: Scikit-learn: Machine Learning in Python [online]. *Journal of Machine Learning Research*, ročník 12, 2011: s. 2825–2830, [Citováno 2021-05-01]. Dostupné z: <https://www.jmlr.org/papers/volume12/pedregosa11a/pedregosa11a.pdf>

- [47] Buitinck, L.; Louppe, G.; Blondel, M.; aj.: API design for machine learning software: experiences from the scikit-learn project [online]. In *ECML PKDD Workshop: Languages for Data Mining and Machine Learning*, 2013, s. 108–122, [Citováno 2021-05-01]. Dostupné z: <https://arxiv.org/abs/1309.0238>
- [48] Kingma, D.; Ba, J.: Adam: A Method for Stochastic Optimization [online]. *International Conference on Learning Representations*, prosinec 2014, [Citováno 2021-05-03]. Dostupné z: <https://arxiv.org/abs/1412.6980>
- [49] Dutta, A.; Gupta, A.; Zisserman, A.: VGG image annotator (VIA) [software]. 2016, [Citováno 2021-04-19]. Dostupné z: <https://gitlab.com/vgg/via>
- [50] Abdulla, W.: Mask R-CNN for object detection and instance segmentation on Keras and TensorFlow [software]. 2017, [Citováno 2021-04-19]. Dostupné z: https://github.com/matterport/Mask_RCNN
- [51] Otsu, N.: A Threshold Selection Method from Gray-Level Histograms [online]. *IEEE Transactions on Systems, Man, and Cybernetics*, ročník 9, č. 1, 1979: s. 62–66, doi:10.1109/TSMC.1979.4310076, [Citováno 2021-04-29]. Dostupné z: <https://doi.org/10.1109/TSMC.1979.4310076>
- [52] Niblack, W.: *An Introduction to Digital Image Processing*. DNK: Strandberg Publishing Company, 1985, ISBN 8787200554, [Citováno 2021-04-29].

Seznam použitých zkratek

CNN	konvoluční neuronová síť
FPN	Feature Pyramid Network
GPU	graphics processing unit
ILSVRC	ImageNet Large Scale Visual Recognition Challenge
IoU	Intersection over Union
IR	infrared
mAP	mean average precision
RELU	rectified linear unit
RPN	Region Proposal Network
R-CNN	regiony s příznaky konvolučních neuronových sítí
SGD	stochastický gradientní sestup
SVM	support-vector machines
UV	ultraviolet

Obsah přiloženého paměťového média

	readme.txt	stručný popis obsahu média
	geometric-features	implementace klasifikace tvarových deskriptorů
	mask-r-cnn	implementace detekce a klasifikace pomocí Mask R-CNN
	latex	zdrojové kódy práce ve formátu \LaTeX
	text	složka s textem práce
	thesis.pdf	text práce