



Pedagogická  
fakulta  
Faculty  
of Education

Jihočeská univerzita  
v Českých Budějovicích  
University of South Bohemia  
in České Budějovice

**Jihočeská univerzita v Českých Budějovicích**

Pedagogická fakulta

Katedra aplikované fyziky a techniky

**Fyzikální aspekty digitálního obrazu a detekce  
objektů aplikací konvolučních neuronových sítí**

Diplomová práce

**Autorka práce:** Bc. Lucie Štindlová

**Vedoucí práce:** doc. RNDr. Petr Bartoš, Ph.D.

**Konzultant práce:** Mgr. Ing. Roman Bumbálek

České Budějovice 2023

## Prohlášení

Prohlašuji, že jsem autorem této kvalifikační práce a že jsem ji vypracovala pouze s použitím pramenů a literatury uvedených v seznamu použitých zdrojů.

V Českých Budějovicích dne .....

.....

*Podpis*

## Abstrakt

Diplomová práce se zabývá problematikou počítačového vidění a porovnáním trénovaných deep learning modelů. V teoretické části je rozpracován podrobný přehled fyzikálních vlastností elektromagnetického záření viditelného spektra a principů záznamu digitálního snímku, včetně jeho vlastností. Následuje rešerše metod zpracování obrazu, kde jsou stručně charakterizovány konvenční metody a poté i konvoluční neuronové sítě. Praktická část je zaměřena na vytvoření vhodného datasetu s anotacemi, jenž je dále aplikován pro trénování vybraných modelů. Z hlediska zvolených vyhodnocovacích metrik, a také na základě experimentálního stanovení vstupních parametrů pro trénování, je analyzována variabilita a přesnost získaných výsledků.

**Klíčová slova:** Elektromagnetické záření, digitální obraz, zpracování obrazu, konvoluce, anotace dat, trénování neuronových sítí

## Abstract

The thesis is focused on computer vision and the comparison of trained deep learning models. In the theoretical part, a detailed overview of the physical properties of the visible electromagnetic spectrum and the principles of the digital image, including it, is processed. This is followed by a review of image processing methods, where conventional methods and then convolutional neural networks are briefly characterized. The practical part is focused on creating a suitable dataset with annotations, which is further applied for training selected models. The variability and accuracy of the obtained results is analyzed from the point of view of the selected evaluation metrics, as well as based on the experimental determination of input parameters for training.

**Keywords:** Electromagnetic spectrum, digital image, image processing, convolution, data annotation, training neural networks

## Poděkování

Ráda bych poděkovala vedoucímu mé diplomové práce, panu doc. RNDr. Petru Bartošovi, Ph.D., za vedení práce a cenné zkušenosti získané během studia. Dále velmi děkuji panu Mgr. Ing. Romanovi Bumbálkovi za podnětné rady, věnovaný čas a ochotu při zpracovávání celé problematiky. Poděkování patří i mému manželovi, rodině a kolegům za jejich trpělivost a ochotu.

# Obsah

Úvod	8
<b>1 Obraz jako elektromagnetické záření</b>	<b>9</b>
1.1 Popis záření . . . . .	9
1.1.1 Fotoelektrický jev . . . . .	10
1.1.2 Comptonův jev . . . . .	12
1.2 Přehled elektromagnetického záření . . . . .	13
1.3 Typy obrazových záznamů . . . . .	14
<b>2 Proces vzniku digitálního obrazu</b>	<b>16</b>
2.1 Vlastnosti digitálního obrazu . . . . .	17
2.2 Snímače . . . . .	18
<b>3 Zpracování obrazu</b>	<b>21</b>
3.1 Konvenční metody zpracování obrazu . . . . .	21
3.1.1 Lineární úprava obrazu . . . . .	21
3.1.2 Nelineární úprava obrazu . . . . .	23
3.1.3 Úprava histogramu . . . . .	23
3.1.4 Prahování . . . . .	24
3.1.5 Hranová detekce . . . . .	24
<b>4 Metody hlubokého učení</b>	<b>26</b>
4.1 Konvoluční neuronové sítě . . . . .	26
4.2 Porovnání konvolučních neuronových sítí . . . . .	26
4.3 Klasifikace . . . . .	27
4.4 Detekce . . . . .	28
4.5 Segmentace . . . . .	29
<b>5 Anotace digitálního obrazu</b>	<b>32</b>
5.1 Anotační nástroje . . . . .	32

5.1.1	CVAT . . . . .	33
5.1.2	V7 . . . . .	33
5.1.3	Labelbox . . . . .	34
5.1.4	LabelMe . . . . .	34
5.1.5	VGG Image Annotator . . . . .	35
<b>6</b>	<b>Augmentace dat</b>	<b>36</b>
6.1	Postup při augmentaci dat . . . . .	36
<b>7</b>	<b>Vybrané vstupy a parametry procesu trénování</b>	<b>38</b>
7.1	Vlastnosti vstupního datasetu . . . . .	38
7.2	Architektura YOLO . . . . .	39
7.3	Vybrané hyperparametry . . . . .	40
7.4	Metriky pro vyhodnocení . . . . .	41
<b>8</b>	<b>Vyhodnocení získaných výsledků</b>	<b>43</b>
8.1	Porovnání YOLOv5n, YOLOv5m a YOLOv5x dle vybraných hyperparametrů . . . . .	43
8.1.1	Confusion matrix . . . . .	43
8.1.2	Precision-Recall Curve . . . . .	45
8.1.3	Čas trénování a vytížení GPU . . . . .	47
8.1.4	F1-Curve . . . . .	47
8.1.5	Výsledky průběhu trénování modelů YOLOv5 . . . . .	49
8.2	Porovnání YOLOv8n, YOLOv8m a YOLOv8x dle vybraných hyperparametrů . . . . .	51
8.2.1	Confusion matrix . . . . .	51
8.2.2	Precision-Recall Curve . . . . .	53
8.2.3	Čas trénování a vytížení GPU . . . . .	53
8.2.4	F1-Curve . . . . .	55
8.2.5	Výsledky průběhu trénování modelů YOLOv8 . . . . .	57
8.3	Vliv hyperparametru Learning rate . . . . .	59

8.3.1	Precision-Recall Curve . . . . .	59
8.3.2	Čas trénování a vytížení GPU . . . . .	61
8.3.3	F1-Curve . . . . .	61
8.3.4	Výsledky průběhu trénování modelu YOLOv8n . . . . .	63
8.4	Vliv hyperparametru Batch size . . . . .	65
8.4.1	Precision-Recall Curve . . . . .	65
8.4.2	Čas trénování a vytížení GPU . . . . .	67
8.4.3	F1-Curve . . . . .	68
8.4.4	Výsledky průběhu trénování modelu YOLOv8n . . . . .	70
8.5	Diskuze . . . . .	72
<b>9</b>	<b>Závěr</b>	<b>74</b>
	<b>Seznam literatury</b>	<b>75</b>

# Úvod

Současná doba je mimo jiné charakterizována neustálým vývojem informačních technologií, které se dostávají do popředí ve stále více odvětvích. Své místo zaujímá i umělá inteligence, a to převážně v technických oborech či ve zdravotnictví. Využít ji ale lze například i v zemědělství. Neuronové sítě se učí na trénovacích datech, která musí být vhodně zpracována. V případě zpracování obrazu existuje několik různých metod zpracování. Důležitým krokem před samotným trénováním modelu neuronové sítě je také anotace obrazu a případná augmentace dat. Po dosažení nejvyšší možné přesnosti mohou neuronové sítě sloužit k rozpoznání řeči nebo ke klasifikaci objektů na snímku.

Cílem diplomové práce je nejprve vytvořit validní vstupní dataset s anotacemi vybrané plodiny a následně vytrénovat a porovnat varianty zvolené architektury detekční neuronové sítě a zároveň ověřit vliv stanovených hyperparametrů na výsledky získané v rámci procesu trénování.



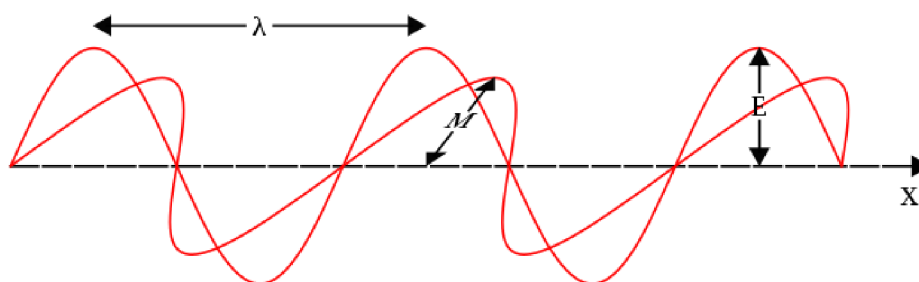
# 1 Obraz jako elektromagnetické záření

## 1.1 Popis záření

Stav objektu je ovlivňován mnoha faktory, mimo jiné i působením silového pole. Můžeme hovořit o gravitačním, magnetickém nebo elektromagnetickém silovém poli. Aby vznikl obraz, využívá se právě elektromagnetického pole. Informace o objektu jsou přenášeny pomocí vln (Pavelka, 1999). Veličiny, které popisují elektromagnetické pole, jsou:

- intenzita elektrického pole  $E$ ,
- magnetická indukce  $B$ ,
- elektrická indukce  $D$ ,
- intenzita magnetického pole  $H$ .

Dle Maxwellovy teorie, která se zabývá vlnovým popisem záření, je elektromagnetické záření považováno za příčné vlnění. Toto vlnění se skládá ze dvou složek, které jsou na sebe kolmé. Jedná se o velikost intenzity elektrického pole a velikost intenzity magnetického pole. Tyto dvě složky (velikosti) se periodicky mění ve směru šíření vlny (Franklin, 2017).



Obrázek 1: Světlo složené ze dvou složek (Vo, 2023)

Rychlost vlnění je ovlivněna prostředím, respektive vlastnostmi tohoto prostředí. Konstanty popisující prostředí, kterým se vlnění šíří, nazýváme permitivita  $\epsilon$  a permeabilita  $\mu$ . Důležitou veličinou, která popisuje již samotné vlnění,

je frekvence  $f$ . Tato veličina udává, kolikrát nastane periodická změna za jednu sekundu. Dráha, kterou vlna urazí za jednu periodu, se nazývá vlnová délka. Platí vztah:

$$\lambda = \frac{c}{f}.$$

S nástupem moderní (kvantové) fyziky, která přinesla mnoho převratů a nových myšlenek, přicházely nové pohledy i na záření. Tzv. kvantový popis záření je založen na faktu, že záření se skládá z částic. Tyto částice jsou nazývány fotony. Zatímco vlnový popis záření vysvětloval řadu jevů z oblasti optiky, kvantový popis objasnil fotoelektrický a Comptonův jev.

Může se zdát, že pravdivá je jen jedna z teorií. V tomto případě jsou ovšem obě teorie správné. Elektromagnetické záření má vlnové i částicové vlastnosti. Klasický (vlnový) popis a kvantový popis záření se tedy doplňují. Ačkoliv foton má vlnové i částicové vlastnosti, není možné ho považovat za částici. Tato představa může být velice obtížná, nicméně se musíme vzdát té myšlenky, že lze přesně určit aktuální polohu fotonu. Pokud by foton byl částicí, nemohlo by docházet k interferenčnímu jevu. Při pokusu na dvojštěrbině by foton musel proletět oběma štěrbinami zároveň, což ovšem nelze. A pokud by proletěl pouze jednou štěrbinou, nebylo by možné pozorovat na stínítku interferenční obrazec. Místo dopadu fotonu na stínítku můžeme tedy určit jen s určitou pravděpodobností. Tedy to, co se vlní, je pravděpodobnost výskytu fotonu v daném místě (Halliday, 2000).

### 1.1.1 Fotoelektrický jev

Ačkoliv je s fotoelektrickým jevem spojován výhradně Albert Einstein, za objevitele je považován německý fyzik Heinrich Hertz. Tento jev pozoroval jako vedlejší produkt při pokusu, ve kterém zkoumal vlnové vlastnosti elektromagnetického záření. Na základě toho začali stavět další fyzikové, a to Wilhelm Hallwachs a Alexandr Stoletov. Ti pracovali s různými vodivými materiály,

které ozařovali. Pozorovali, že dosáhne-li záření určité frekvence (mezní frekvence), začne docházet k uvolňování záporného náboje z vodiče, přičemž je tato frekvence pro každý materiál různá. Ačkoliv Hallwasch a Stoletov několikrát pozorovali a potvrdili fotoelektrický jev, nikdy se neodvážili teoreticky ho vysvětlit. Zlom nastal s experimentem, který provedl Joseph John Thomson. Právě jemu je přikládán objev elektronu. Dva roky po tomto objevu identifikoval elektrony, které unikaly z ozařovaného kovu. Na tento experiment následně navázal Philipp Lenard, který jako první určil kinetickou energii uvolňujících se elektronů (Martinásková, 2014).

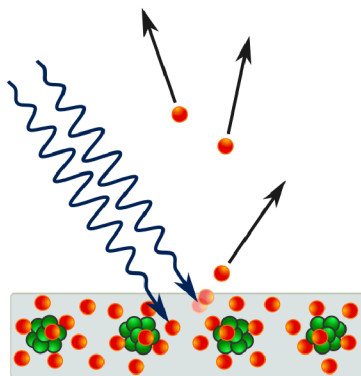
Bylo provedeno mnoho experimentů, ovšem i přesto žádný fyzik nedokázal fotoelektrický jev teoreticky vysvětlit. A to mimo jiné i proto, že by jakékoliv vysvětlení vyvracelo klasickou fyziku. Podle klasické fyziky by totiž s rostoucí intenzitou světla rosla i rychlost uvolňujících se elektronů. Zjistilo se ovšem, že intenzita záření ovlivňuje pouze počet uvolněných elektronů, nikoliv však jejich rychlost. Dalším předpokladem klasické fyziky bylo, že k emisi elektronů bude docházet pouze při určitých (rezonančních) frekvencích, nikoliv však ve spojitém spektru. Zároveň by k emisi nemělo docházet neprodleně po dopadu záření, podle klasické fyziky by měla nastat určitá prodleva mezi dopadem záření a emisí elektronů. Během pozorování fotoelektrického jevu ovšem žádná prodleva zaznamenána nebyla, k emisi dochází okamžitě po dopadu záření (Králová, 2007).

S klíčovou myšlenkou, která dokázala fotoelektrický jev zcela objasnit, přišel Albert Einstein. Dle jeho vysvětlení elektrony absorbují energii záření po kvantech. Matematicky je tento stav popsán následovně:

$$h \cdot f = W_0 + E_K,$$

kde  $h$  je Planckova konstanta,  $f$  je frekvence záření,  $W_0$  je výstupní práce a  $E_K$  je kinetická energie emitujícího elektronu. Tyto kvantové částice nazval světelné kvantum, s pojmem foton se pracovalo až ve 30. letech 20. století. Za vysvětlení fotoelektrického jevu získal Albert Einstein v roce 1921 Nobelovu cenu. V praxi

se s tímto jevem často setkáváme. Hraje klíčovou roli u snímačů digitálních fotoaparátů či kamer, zároveň je využíván i ve fotovoltaiice (Kulhánek, 2020).



Obrázek 2: Vizualizace fotoelektrického jevu (Wikimedia Commons)

### 1.1.2 Comptonův jev

Výsledkem Einsteinova experimentů byla další hypotéza o fotonech. Podle ní předávají fotony částicím nejen energii, ale i hybnost. Tak jako energie, je i hybnost předávána v kvantech. Dojde-li ke srážce fotonu a elektronu, měl by platit zákon zachování energie a hybnosti. Americký fyzik Arthur Holly Compton provedl v roce 1922 experiment, během kterého bylo rentgenové záření rozptýlené na materiálech se slabými elektronovými vazbami. Paprsky byly rozptýleny do různých směrů, Compton po rozptylu měřil vlnovou délku a intenzitu těchto paprsků (Králová, 2007).

Pokud bychom tento experiment popisovali opět z pohledu klasické fyziky, docházelo by po dopadu rentgenového záření k rozkmitání elektronů. Frekvence kmitu elektronů by se pak rovnala frekvenci elektromagnetického záření, které by elektrony vyzařovaly. Rozptýlené záření by bylo charakterizováno jednou vlnovou délkou  $\lambda$ . Compton během experimentu ovšem získal docela jiné výsledky (Králová, 2007).

Compton našel v rozptýleném záření zdvojené spektrální čáry. Jedna vlnová délka odpovídala původní vlnové délce záření, druhá vlnová délka byla dokonce

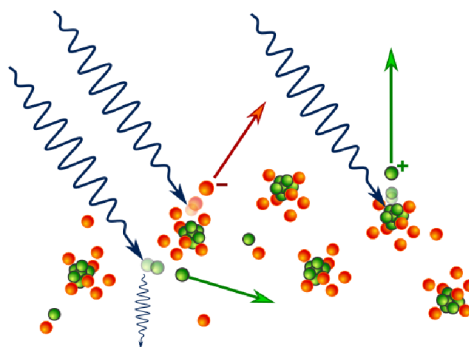
větší. Záření s větší vlnovou délkou vzniklo díky rozptylu původního záření na slabě vázaných elektronech. Existenci druhé vlnové délky určuje vztah:

$$\lambda - \lambda' = \frac{h}{m_0 \cdot c} \cdot (1 - \cos \varphi).$$

Comptonova vlnová délka je vyjádřena vztahem:

$$\frac{h}{m_0 \cdot c}.$$

Ke Comptonovu jevu dochází při každé srážce fotonu a elektronu. Pozorovatelný je jev pouze u záření, jehož fotony mají vyšší hmotnost, což je například i záření rentgenové (Halliday, 2020).

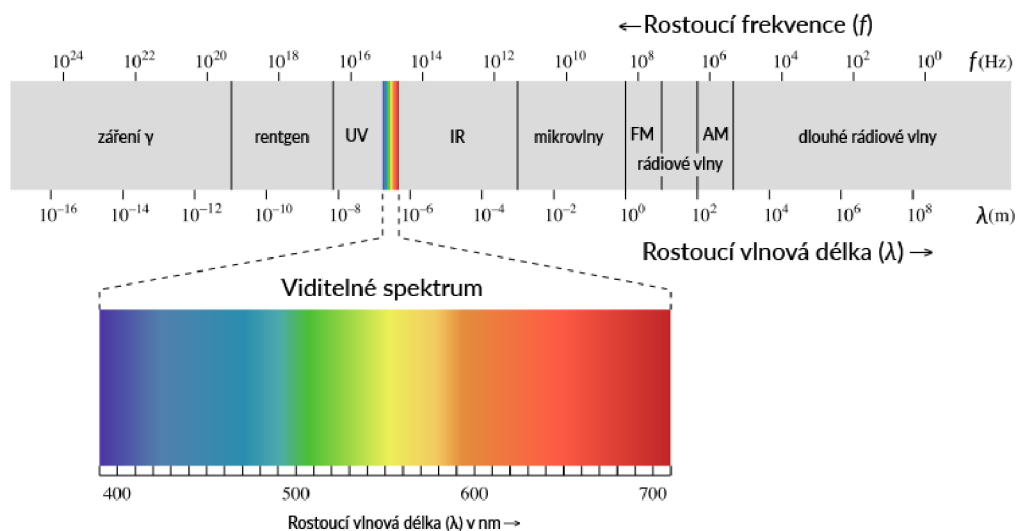


Obrázek 3: Vizualizace Comptonova jevu (Wikimedia Commons)

## 1.2 Přehled elektromagnetického záření

Elektromagnetické záření rozdělujeme do několika skupin v závislosti na energii daného záření. Seřazení těchto druhů záření do řady nazýváme elektromagnetické spektrum. Jen jeho velmi malá část je viditelná lidským okem. Jedná se o světlo. Vlnová délka viditelného světla se pohybuje zhruba od 400 do 700 nm. Mezi více energetická záření (tedy záření s vyšší frekvencí a menší vlnovou délkou) patří UV záření, rentgenové záření a gama záření. Tato záření jsou vysoce energetická a tak nebezpečná pro lidský organismus. I přesto ale

nalézájí využití. Například s rentgenovým zářením se běžně setkáváme v lékařství. Všechna ostatní jsou méně energetická než světlo. Jedná se o infračervené záření, mikrovlnné záření a rádiové vlny (Geoff, 2009).



Obrázek 4: Elektromagnetické spektrum (Vo, 2022)

### 1.3 Typy obrazových záznamů

Aby byl obraz zachycen a zaznamenán, je k tomu zapotřebí technického zařízení, jež je součástí například i klasické kamery. Jedná se o tzv. detektor. Ten je schopen reagovat na energii, kterou má záření dopadající na detektor. Tuto energii definujeme vztahem:

$$E = h \cdot f,$$

kde  $h$  je Planckova konstanta a  $f$  je frekvence záření. Frekvence je nepřímo úměrná vlnové délce záření. To znamená, že čím větší má záření frekvenci (neboli čím více energetické záření je), tím je jeho vlnová délka kratší. Detektory můžeme dělit na základě toho, jak reagují na energii dopadajícího záření. Známe detektory: tepelné, fotonové, integrální, fotochemické. Tepelné detektory převádějí optickou energii na energii tepelnou. Dopadající optické záření

zvyšuje teplotu čidla detektoru. Tento typ detektoru se nejčastěji využívá pro snímání infračerveného záření. Fotonové detektory pracují na principu vnitřního fotoelektrického jevu. Přeměňují dopadající záření na elektrický singál. Integrální detektory po celou dobu expozice kumulují záření, které je vyhodnoceno dodatečně. Příkladem takového detektoru je dozimetr. U fotochemických detektorů pak dochází k přeměně dopadajícího záření na chemickou reakci. Fotografické materiály (například fotografické emulze) tak reagují na dopadající záření díky vznikajícím chemickým reakcím (Vrbová, Jelínková a Gavrilov, 1998).

## 2 Proces vzniku digitálního obrazu

Digitální fotografie vděčí za svůj vznik integrovaným obvodům. Pomocí nich lze obraz zaznamenat jako elektrické signály. Obraz lze definovat jako dvou-rozměrnou funkci  $f(x,y)$ , kde  $x$  a  $y$  jsou rovinné souřadnice. Jsou-li tyto hodnoty diskrétní, nazýváme obraz digitálním obrazem. Digitální obraz se skládá z konečného počtu prvků, přičemž každý tento prvek má určitou hodnotu a lokalizaci. Tyto prvky, základní elementy obrazu, se nazývají pixely (Zmeškal, Buchníček a Sedlák, 2001).

I přesto, že člověk je schopen vnímat pouze malou část elektromagnetického spektra (viditelné světlo), existuje mnoho technologií, které jsou schopné zaznamenávat elektromagnetické spektrum téměř v celém rozsahu (od gama záření po rádiové vlny). Digitalizaci je možné rozdělit do tří úrovní: nízká, střední a vysoká. Nízkoúrovňové procesy zahrnují triviální operace, mezi které patří například redukce šumu či úprava kontrastu. Pro nízkoúrovňovou digitalizaci je charakteristické, že vstupem i výstupem je vždy obraz. Ve střední úrovni digitalizace jsou zastoupeny procesy jako segmentace (rozdělení obrazu na části nebo objekty podle daných podmínek) a klasifikace (rozeznávání jednotlivých objektů). U střední úrovně je vstupem také obraz, ale výstupem jsou již atributy popisující tento obraz. Digitalizace na nejvyšší úrovni už se pak zabývá počítačovým viděním. Na této úrovni je tedy hlavní podstatou porozumění obrazu jako celku z pohledu počítače (Gonzalez a Woods, 1987).

Digitalizace je tedy převod spojitého signálu na diskrétní. Na detektor dopadá světelné záření, které je následně přeměněno na jednotlivé pixely obrazu. Tato přeměna je mimo jiné závislá na intenzitě dopadajícího záření. Tento proces, digitalizace, se provádí pomocí vzorkování obrazu v matici, která má  $M \times N$  bodů. Dále musí dojít ke kvantování jasové funkce do  $K$  intervalů. Čím větších hodnot nabývá  $M$ ,  $N$ ,  $K$ , tím lépe odpovídá digitalizovaný obraz tomu skutečnému (Hlaváč, 2005).

Důležitý údaj, který musíme znát, je interval vzorkování. Jedná se o vzdá-



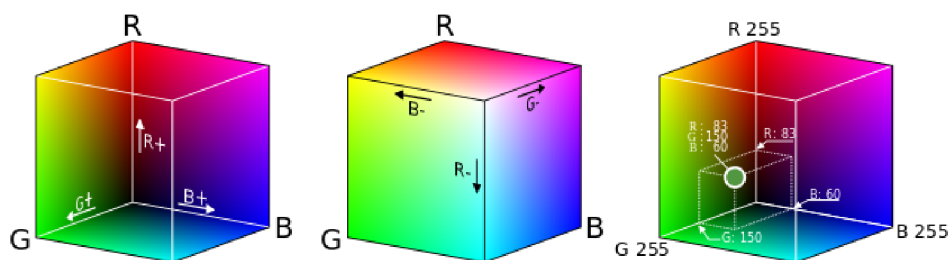
lenost mezi sousedícími vzorkovacími body v obraze. Tato vzdálenost je také udávána jako plošná vzorkovací frekvence. Interval vzorkování je třeba vhodně zvolit, a to tak, aby jeho hodnota byla menší nebo rovna polovině rozměru nejmenšího detailu obrazu. Kromě intervalu vzorkování je důležité vhodně určit i vzorkovací mřížku, která představuje plošné uspořádání bodů obrazu během vzorkování. Zpravidla se pro vzorkování používá pravidelná mřížka, nejčastěji čtvercová. Dále je možné použít mřížku z rovnostranného trojúhelníku nebo šestiúhelníku. Pouze tyto geometrické útvary (čtverec, rovnostranný trojúhelník, pravidelný šestiúhelník) svou sítí zcela pokrývají plochu roviny. Proto jsou při volbě pravidelné mřížky nejlepší možností (Petrou a Bosdogianni, 1999).

## 2.1 Vlastnosti digitálního obrazu

- Jas udává celkovou světlost obrazu. S vyšším jasnem sice dochází k redukci šumu, zároveň ale i k blednutí barev. Jas lze upravit nastavením expozice na fotoaparátu. Jas digitálního obrazu je přímo úměrný intenzitě světla, které dopadá na snímač (Geoff, 2009).
- Rozlišení digitálního obrazu je udáváno pomocí dvou hodnot  $X$  a  $Y$ , kdy  $X$  je počet pixelů na ose horizontální a  $Y$  je počet pixelů na ose vertikální. Po vynásobení těchto dvou hodnot získáme celkový počet pixelů obrazu. Čím vyšší je celkový počet pixelů, tím vyšší je rozlišení. Takový obraz je velmi detailní, zároveň ale paměťově náročnější.
- Kontrast je charakterizován jako podíl nejsvětějšího a nejtmavšího pixelu.
- Barevná hloubka určuje, kolik barevných možností má jednotlivý pixel a zda je průhledný. Jako příklad můžeme uvést 24bitovou hloubku. Takový obraz je často nazýván *true-color*. Celkový počet barev, kterých může pixel nabývat, je  $2^{24}$  - tedy 16 777 216 barev (Wallstrom, 2015).

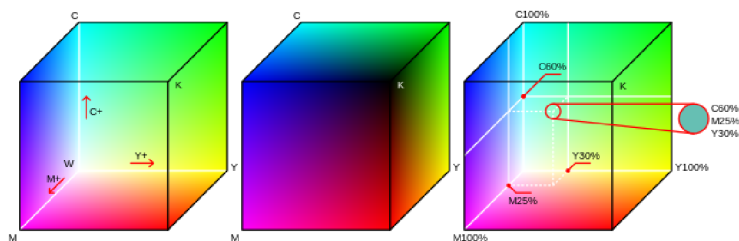
- Barevný model

Model RGB se skládá ze tří barev o určitých vlnových délkách: červená ( $\lambda=700$  nm), zelená ( $\lambda=546$  nm) a modrá ( $\lambda=435,8$  nm). Sloučením všech tří složek vzniká výsledná barva (Dobeš, 2008).



Obrázek 5: Model RGB (Wikimedia Commons)

Výsledná barva podle modelu CMYK je dána rozdílem složek tohoto modelu. Jedná se o barvy: azurová, purpurová, žlutá a černá.

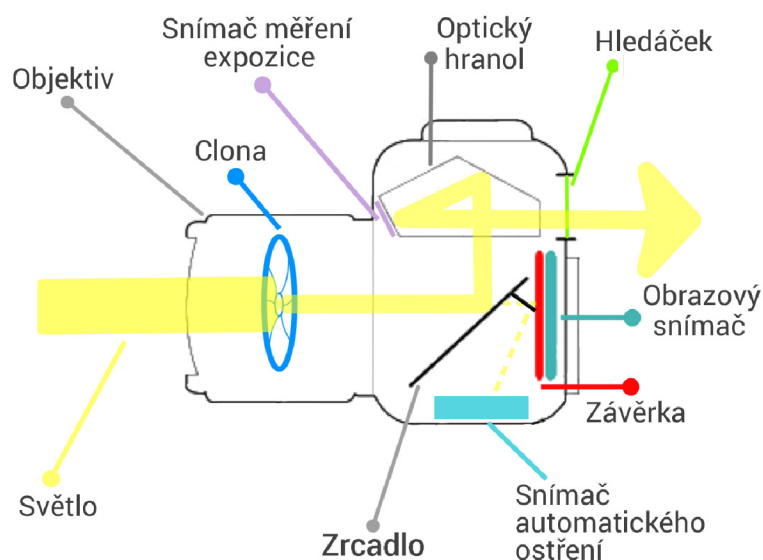


Obrázek 6: Model CMYK (Wikimedia Commons)

## 2.2 Snímače

Při vzniku digitální fotografie putuje světlo od reálného objektu jako první skrze objektiv. Objektiv si lze představit jako soustavu čoček, přes kterou se obraz promítá na snímač. Po zmáčknutí spouště nastává proces, který je nazýván expozice. Závěrka se zvedne a po určitý čas nechá dopadat světlo na snímač. Tento čas lze u dnešních fotoaparátů nastavit podle potřeby. Další hodnotou je clonové číslo (zkráceně nazýváno clona), které udává velikost otvoru v objektivu, skrz který dopadá světlo na snímač. Měnit lze i citlivost

snímače, což je označováno jako ISO. Důležitým parametrem, který charakterizuje objektiv, je ohnisková vzdálenost. Určuje, jak široký a detailní záběr dokáže objektiv zachytit (Minková, 2006).



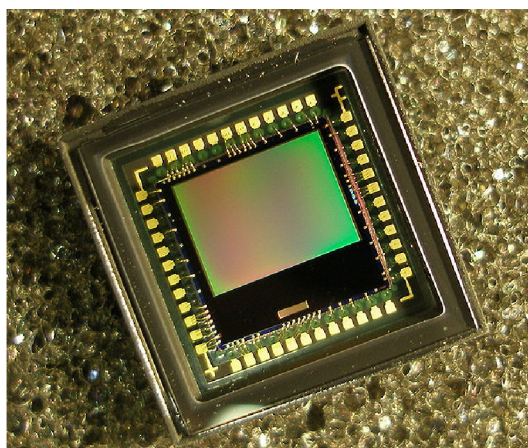
Obrázek 7: Schéma fotoaparátu (Fotolab)

První typ snímače se označuje jako CCD snímač (Charge Coupled Devices). Tento snímač je využíván v kamerách a skenerech už od 80. let 20. století. Kamery zpravidla obsahují tři takové senzory, přičemž přenos elektrického náboje probíhá pomocí paměťového pole. U digitálních fotoaparátů jsou to paměťové registry. Výjimkou nejsou ani systémy s mechanickou závěrku, která slouží k tomu, aby během kompozice světlo nedopadalo na snímač. Modernější snímače tohoto typu umožňují postupné čtení obrazových bodů (pixelů) nebo čtení vymezeného pole senzoru (Zmeškal, Buchníček a Sedlák, 2001). Tento typ senzoru je drahý a na výrobu časově náročný, zároveň ale disponuje mnoha výhodami. Má vysokou světelnou citlivost, dosahuje vyššího rozlišení a vzniklý obraz je kvalitní.

Druhým nejčastěji používaným typem snímače je CMOS (Complementary Metal Oxide Semiconductor). Tento typ je mladší, poprvé byl použit v roce 1996. Přenos náboje je u těchto snímačů realizován přímou adresací pixelů.

Digitalizace obrazu neprobíhá v jeden moment, nýbrž v každém pixelu zvlášť. Snímače CMOS mají menší světelnou citlivost, vzniklý obraz by tedy mohl být při špatném osvětlení tmavý. To vše na úkor výrazně nižší spotřeby elektrické energie. Zároveň i samotná výroba CMOS snímačů není tak časově náročná a vyjde levněji (Helmers a Chellenberg, 2002).

Dynamický vývoj technologií je ovšem zjevný i u snímače typu CMOS. Dnešní snímače tohoto typu mohou samy generovat světlo díky jevu, který se nazývá elektroluminiscence. Během tohoto jevu prochází elektrický proud luminoforem (vhodným materiálem) a dochází k přeměně elektrické energie na světlo. Původní nevýhody byly postupně eliminovány a snímače CMOS se nakonec dostaly do popředí (Xie, 2022).



Obrázek 8: Snímač typu CMOS (Wikipedia)

## 3 Zpracování obrazu

Zpracování obrazu je využíváno pro zdokonalení vlastností digitálního obrazu (například jas, kontrast, rozlišení) či k získání potřebných informací a charakteristik obrazu. Výstupem je tedy obraz s novými parametry. Existuje mnoho metod zpracování obrazu, od jednoduchých (transformace jasu a kontrastu) po ty nejsložitější (klasifikace či segmentace). Specifickou metodou zpracování obrazu je operace založená na konvoluci - lineární úprava obrazu. Těmto metodám je věnována následující část práce. (Hozman, 1996).

### 3.1 Konvenční metody zpracování obrazu

Mezi základní (konvenční) metody zpracování obrazu řadíme prahování, úpravu histogramu, detekci hran nebo také úpravu kontrastu či převod do stupňů šedi.

Způsob zpracování obrazu musí být vhodně zvolen, neboť kombinace některých metod nemusí vést k příznivému výsledku. Některé metody pracují se šumem. Pokud se pokusíme šum odstranit, dojde přitom i k eliminaci ostrých hran, které jsou ovšem pro zpracování obrazu také důležité. Opačnou operací je ostření obrazu, neboli zvýraznění míst s vyšší frekvencí. Vedlejším výsledkem tohoto procesu je zvýšení šumu, čímž se dostáváme zpět na začátek. Proto je nutné volit pro zpracování obrazu vždy tu nejefektivnější metodu a snažit se nalézt kompromisní řešení.

#### 3.1.1 Lineární úprava obrazu

Lineární úpravou rozumíme aplikaci lineárních filtrů, čímž dojde k potřebným změnám vlastností obrazu. Tento způsob operace je založen na konvoluci, což je matematická operace pracující s jednotlivými pixely. Tato operace se skládá z funkce a jádra, kdy jádrem je konvoluční maska, která je přiložena na určité místo v obrazu (postupně se posouvá po celém obraze). Všechny pixely, které přiložená maska překrývá, jsou jednotlivě vynásobeny koeficientem v buňce a následně jsou tyto dvě hodnoty sečteny. Tak vznikne nový pixel. Po aplikaci

této metody bude výsledkem rozostřený obraz. Konvoluce je často využívána při zpracování obrazu (Smith, 1997).

Jádro, neboli kernel, označuje při zpracování obrazu konvoluční matici či masku, která se využívá pro vhodnou úpravu obrazu (rozostření či detekce hran). Jedna z nejnámějších lineárních metod spočívá v aplikaci tzv. Gaussova filtru. Ten se používá k odstranění či redukci šumu, jež je obvykle charakterizován vysokou frekvencí. V tomto případě je aplikována konvoluce s maskou, která vychází z Gaussovy funkce. Výsledkem této metody je rozmazaný snímek (Yun, 2022).



Obrázek 9: Rozostření aplikováním Gaussova filtru

Další metodou pro redukci šumu je průměrování snímku. Tato úprava digitálního obrazu spočívá ve výpočtu aritmetického průměru intenzity pro každý pixel. Tato metoda je založena na předpokladu, že šum na snímku je náhodný. Tyto náhodné výchyly se průměrováním dalších snímku postupně vyrovnají (Gonzalez a Woods, 1992).



Obrázek 10: Rozostření průměrováním

### 3.1.2 Nelineární úprava obrazu

Nejznámějším nelineárním filtrem pro redukci šumu je mediánový filtr. Hodnota každého pixelu je nahrazena mediánem sousedících pixelů. Výhodou této metody je zachování hran na snímku, což umožňuje další možnosti zpracování obrazu (Tay, 2021).



Obrázek 11: Rozostření užitím mediánu

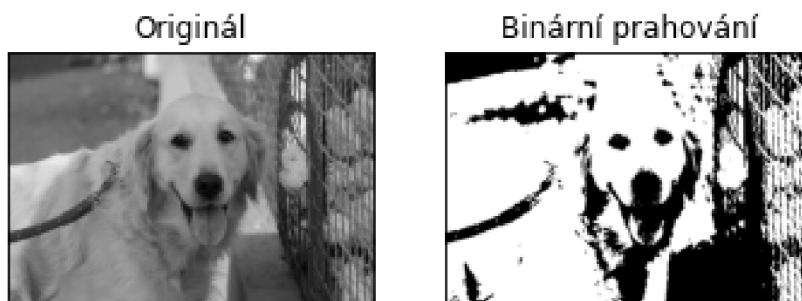
### 3.1.3 Úprava histogramu

Histogram digitálního obrazu s úrovněmi intenzity v rozsahu  $[0, L - 1]$  je diskrétní funkce  $h(r_k) = n_k$ , kde  $r_k$  je  $k$ -tá hodnota intenzity a  $n_k$  je počet pixelů v obraze s intenzitou  $r_k$ . Běžně se v praxi histogram upravuje tak, že se každá z jeho složek vydělí celkovým počtem pixelů v obrázku. Tento počet získáme součinem hodnot  $M$  a  $N$ , kdy  $M$  a  $N$  jsou rozměry obrazu. Práce s histogramem může být využita při dalších operacích jako je komprese nebo segmentace (Gonzalez a Woods, 1992).

Po grafické stránce je histogram znázorněn jako zastoupení intenzit složek barev a jejich odstínů. Častou operací, která je aplikována na histogram, je ekvalizace. Ta je vhodná v případě, kdy je pásmo intenzit úzké a v obraze tak nejsou zřejmé některé prvky. Ekvalizací dojde z rozšíření pásma intenzit, díky čemuž vyniknou i méně viditelné objekty a detaily (Freeman, 2004).

### 3.1.4 Prahování

Prahování je často využívaná metoda pro segmentaci obrazu. Jako první musí být určena prahová hodnota. S touto stanovenou hodnotou jsou porovnávány jednotlivé pixely. Hodnota, která je vyšší než prahová hodnota, je přepsána na 255 (bílá barva) a hodnota, která je nižší než prahová hodnota, je přepsána na 0 (černá barva). Prahovou hodnotu je možné určit například podle histogramu, kterému je věnována kapitola 3.1.3. Prahováním dosáhneme oddělení objektu a pozadí na obraze (Šťastný, 2006).



Obrázek 12: Prahování snímku

### 3.1.5 Hranová detekce

Hrana je místo v obrazu, kde dochází ke skokové změně jasové funkce. Obdobně se vyznačuje i míra šumu, a tak by mohlo dojít k detekci tzv. falešných hran. Pokud to vstupní obraz vzhledem k míře šumu vyžaduje, je vhodné ho před samotným zpracováním upravit tak, aby byl výskyt falešných hran minimalizován. K takové úpravě lze použít například Gaussův filtr, který eliminuje šum. Po této operaci následuje samotná hranová detekce. Pro tu jsou často využívány klasické hranové filtry, jedná se o Robertsův, Prewittové, Sobelův, Robinsonův, Kirschův a Laplaceův hranový detektor. Tato operace spočívá v hledání maximálních prvních derivací s využitím konvolučních masek. Po aplikaci jednoho z detektorů získáme signál obsahující informaci o všech hranách v obraze (Hájovský, 2018).



Originál

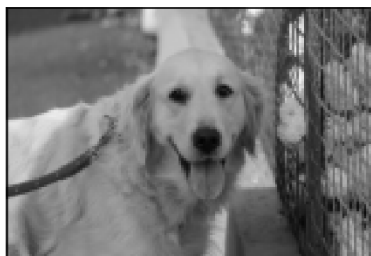


Laplaceův operátor



Obrázek 13: Detekce hran užitím Laplaceova operátoru

Originál



Sobelův operátor



Obrázek 14: Detekce hran užitím Sobelova operátoru

## 4 Metody hlubokého učení

Hluboké učení je typ strojového učení využívající umělou inteligenci. Jedná se o velmi výkonný nástroj v mnoha oblastech, jež je založen na neuronových sítích. Metody hlubokého učení se od konvenčních metod liší v mnoha aspektech. Pro využití metod hlubokého učení je zapotřebí velké množství dat. Trénování konvolučních neuronových sítí vyžaduje předchozí zpracování dat, vyhodnocení nikoliv (Nagpal, 2020). Mezi metody hlubokého učení patří:

- klasifikace,
- detekce objektu,
- segmentace obrazu.

### 4.1 Konvoluční neuronové sítě

Konvoluční neuronové sítě jsou podmnožinou neuronových sítí, inspirovány zrkovou soustavou koček (Dong, 2021). Zpravidla se vyznačují složením z několika typů vrstev (konvoluční a pooling vrstvy). Mezi hlavní znaky konvolučních neuronových sítí tedy patří nízká konektivita (každý neuron reaguje na malé množství hodnot vstupu) a sdílení parametrů (Goodfellow, Bengio a Courville, 2016). Jejich využití zahrnuje zpracování obrazu a rozpoznávání ručně psaných znaků, jako je například poštovní směrovací číslo. V současnosti je hojně využívána pro rozpoznání hlasu či detekci objektu. Mezi nejznámější konvoluční neuronové sítě patří AlexNet, VGGNet a ResNet. Tyto sítě se nejčastěji využívají pro klasifikaci. Liší se výkonostními parametry, způsobem detekce i oblíbeností (Dong, 2021).

### 4.2 Porovnání konvolučních neuronových sítí

Alex Krizhevsky vyvinul v roce 2012 konvoluční neuronovou síť s názvem AlexNet. Tato neuronová síť je složena z 8 vrstev, z nichž 5 je konvolučních,

a obsahuje více než 60 milionů parametrů. Neuronová síť AlexNet se nejčastěji využívá pro klasifikaci, bývá použita i jako páteř (tzv. backbone) při detekci obrazu k extrakci příznaků.

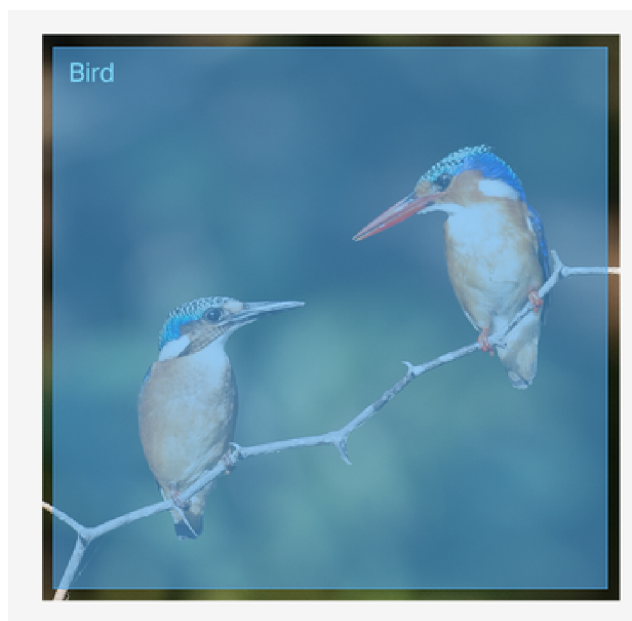
V roce 2014 byla vytvořena neuronová síť pod názvem VGG-16. Síť obsahuje celkem 16 vrstev, z toho 13 konvolučních, a více než 138 milionů parametrů. V porovnání s AlexNet má tedy VGG-16 více vrstev a pracuje s menšími filtry. Neuronová síť VGG je také často určena pro detekci objektů, přičemž je ale výkonější než AlexNet.

Mario Szegedy přišel v roce 2014 s neuronovou sítí, která je od AlexNet a VGG odlišná. Síť je založena na víceúrovňové konvoluci, která mimo jiné slouží i ke snížení rozměrů vstupního obrazu. Síť je tvořena 22 vrstvami a 5 miliony parametrů.

Konvoluční sítě se s přibýváním vrstev stále prohlubují. Kaiming He v roce 2015 vyvinul neuronovou síť ResNets, jejíž nejhlubší varianty mají stovky vrstev. Princip těchto neuronových sítí je založen na možnosti přeskokování vrstev (většinou dvou nebo tří). Díky možnosti přeskokování je k trénování dat zapotřebí méně vrstev, což tuto hlubokou neuronovou síť zjednodušuje (Amjoud, 2019).

### 4.3 Klasifikace

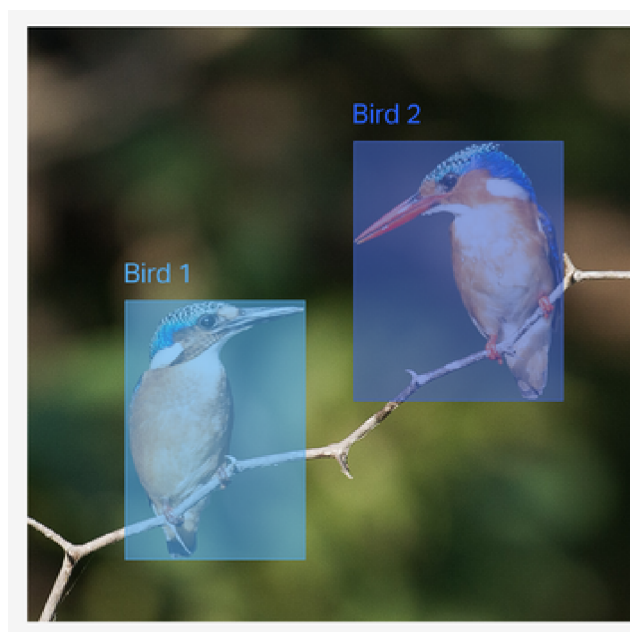
Při klasifikaci je vstupem obraz jako celek, cílem klasifikace je pak rozčlenit obraz do určitých kategorií. Všechny objekty jsou do kategorií rozdělovány na základě svých vizuálních vlastností. Výstupem klasifikace je popis jednotlivých kategorií objektů, které se na snímku nacházejí (Brownlee, 2019). S vývojem konvolučních neuronových sítí se výrazně zvýšila efektivnost této metody, klasifikaci lze aplikovat na rozsáhlé množství dat, zvyšuje se i její přesnost. Čím více anotovaných dat do sítě vložíme, tím přesnější a spolehlivější trénovaná síť bude (Khanday a Sofi, 2021).



Obrázek 15: Klasifikace objektů (V7's Blog)

#### 4.4 Detekce

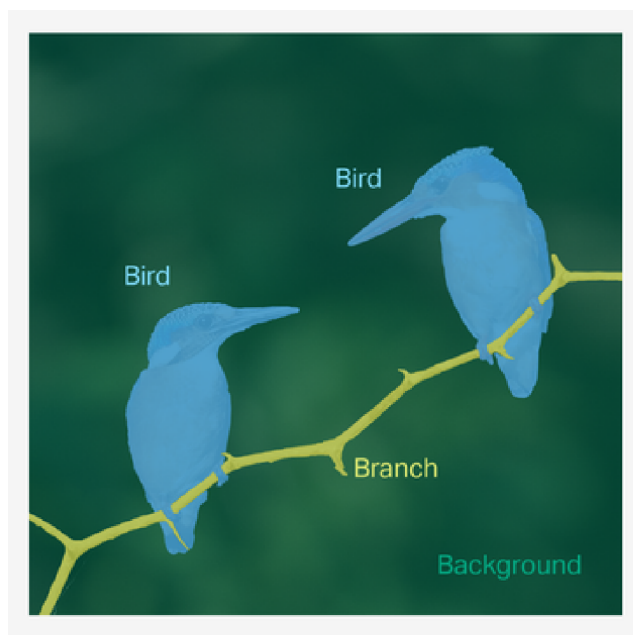
Detekce již zajišťuje určitou lokalizaci daného objektu, a to za použití čtyřúhelníků, které tyto prvky označují. Cílem detekce je lokalizovat a klasifikovat objekty vybraných kategorií na snímku. Komplikace mohou nastat ve chvíli, kdy síť pracuje s velkým množstvím kategorií. Přesnost této metody může negativně ovlivnit i nasvícení pozadí objektů nebo poměry stran. Neuronové sítě využívající tuto metodu jsou velmi výkonné, zároveň však již poměrně náročné na výpočetní výkon. Velkou výzvou do budoucna je v této oblasti detekce 3D objektů. Samotná detekce objektů má v posledních letech široké spektrum využití, například pro rozpoznávání obličejů nebo pro vývoj autonomních vozidel (Khanday a Sofi, 2021).



Obrázek 16: Detekce objektů (V7's Blog)

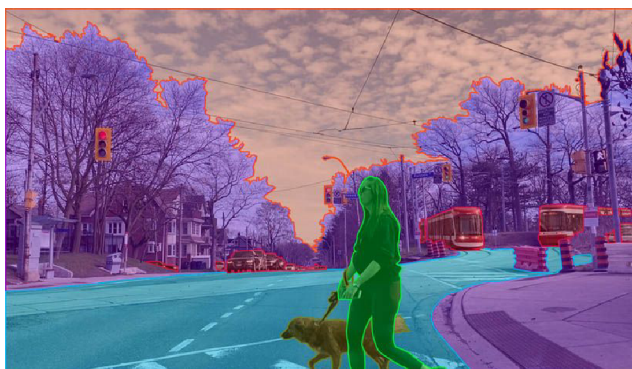
## 4.5 Segmentace

Segmentace je nejsložitější ze zmíněných metod hlubokého učení. Pracuje s obrazem na úrovni pixelů, které seskupuje podle kategorie, do které patří. Segmentace postupně pomáhá zdokonalovat algoritmy v mnoha problematikách, například zkoumání satelitních snímků. Dosažení co nejpřesnějších výsledků často komplikuje členité a nasvícené pozadí objektů. Pro přesnou segmentaci je také třeba dodat co největší množství anotovaných dat. Pokud samotná anotace obsahuje nepřesnosti, nelze čekat přesné výsledky segmentace. Náročným úkolem pro neuronovou síť během segmentace je mimo jiné rozlišení objektů, které se vyznačují podobnými vlastnostmi. Všechny tyto aspekty je nutné zahrnout do závěrečného vyhodnocování výsledků (Dong, 2021). Existují tři základní druhy segmentace: sémantická segmentace, segmentace instancí a panoptická segmentace.



Obrázek 17: Segmentace objektů (V7's Blog)

Sémantická segmentace je proces, jehož cílem je přidělit každému pixelu snímku konkrétní označení třídy. Tak dojde k hlubšímu rozdělení snímku do jednotlivých částí. K užití sémantické segmentace založené na hlubokém učení je potřeba velké množství anotovaných dat (Liu, 2022).



Obrázek 18: Sémantická segmentace (Keymark)

Další způsob segmentace spočívá v nalezení objektů na snímku. Tato metoda hlubokého učení zaznamenala v poslední době výrazný pokrok. Od sémantické segmentace se liší tím, že dokáže odlišit různé prvky stejné kategorie.

Tento typ segmentace tedy může poskytnout podrobnější informace o snímku. Kombinací těchto dvou metod vzniká panoptická segmentace (Gu, 2022).



Obrázek 19: Segmentace instancí (Keymark)

## 5 Anotace digitálního obrazu

Anotace digitálního obrazu je proces, většinou prováděný člověkem (aplikací již natrénované konvoluční neuronové sítě), při kterém jsou označovány konkrétní objekty digitálního obrazu. V rámci procesu anotování vznikají doplňková data, která nesou informace o poloze. Na začátku anotace jsou určeny a pojmenovány objekty, které bude anotátor označovat. Samotný proces anotace se může lišit v závislosti na požadavcích zadavatele nebo výběrem anotačního nástroje (Jiu a Sahbi, 2022).

K potížím může dojít při rozhodování, jak zacházet s objekty v případě, že nejsou v obraze celé nebo jsou zakryty jiným objektem. Už před započítím anotace je tedy nutné rozhodnout se, jak v takových případech budeme postupovat.

Anotaci lze současně chápat jako přípravu obrazů (fotografií) pro trénování neuronové sítě, která by v ideálním případě měla následně zvládnout anotovaný objekt rozpoznat. V současnosti jsou veřejně dostupné rozsáhlé datasey obsahující velké množství dat jako je Open Image Dataset V6 od společnosti Google, který je tvořen 9 miliony obrázky.

### 5.1 Anotační nástroje

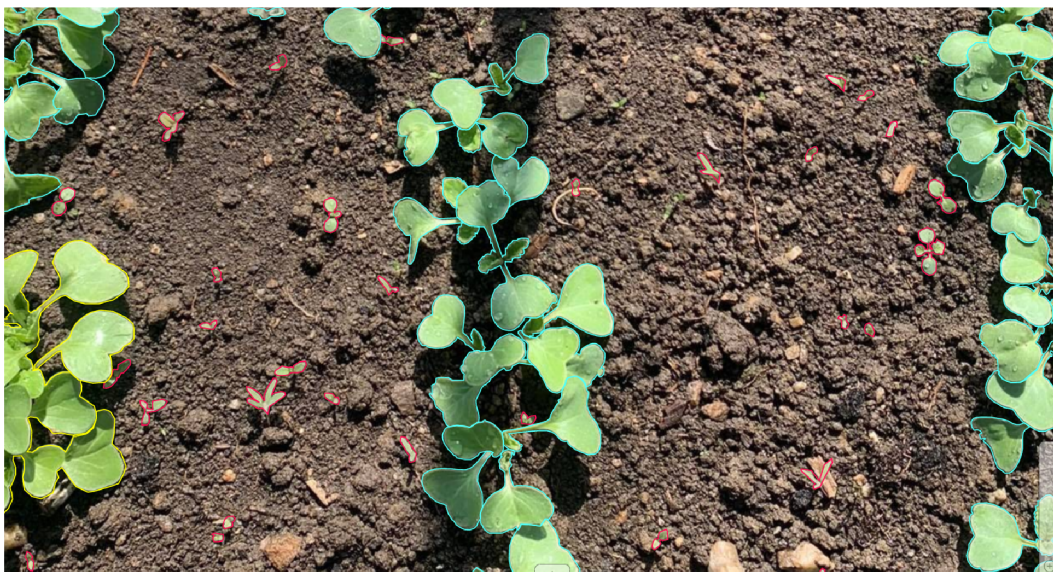
Anotační nástroje mohou být rozděleny do skupiny na základě několika vlastností, například dle typu licence na placené, freeware pro nekomerční využití či open-source, které jsou bezplatné pro libovolné použití. Dále lze nástroje kategorizovat na webové, které jsou k dispozici online, a desktopové, jenž je nutné nainstalovat. Bezplatné anotační nástroje jsou zpravidla dostačující na základní úlohy. Mnoho z nich podporuje i anotaci videa. Placené nástroje mají často nadstandartní funkce jako zpracování 3D dat nebo textových dokumentů. Výhodou online anotačních nástrojů je snadný přístup z jakéhokoliv zařízení. Práci si lze uložit do projektů. Jako nevýhodu pak lze považovat závislost na internetovém připojení. Offline anotační nástroje tento problém



řešit nemusí, ovšem přenos rozpracovaného projektu je pak zdatelně složitější. Několik příkladů anotačních nástrojů je vyjmenováno a popsáno v této části práce.

### 5.1.1 CVAT

CVAT, neboli Computer Vision Annotation Tool, je open source nástroj pro anotaci obrazu i videa. Tento nástroj je bezplatný a umožňuje anotaci obrazu pro základní metody hlubokého učení, mezi které patří detekce objektu, klasifikace a segmentace. Před samotnou anotací musí uživatel definovat třídy, do kterých budou objekty snímku přiřazovány. Poté objekt označí (rámečkem nebo mnohoúhelníkem). Tento software vyvinutý společností Intel je distribuován pod licenci MIT.



Obrázek 20: Anotace objektu v prostředí CVAT

### 5.1.2 V7

Anotační nástroj s názvem V7 je výkonný software, který disponuje mnoha možnostmi anotace. Umožňuje anotaci snímků, videí, ale také i 3D dat nebo mikroskopických snímků. Bývá využíván i v oblasti lékařství pro anotaci dat

označovaných jako DICOM. Do této skupiny dat patří snímky pořízené lékařskými snímacími metodami jako je CT či ultrazvuk. Software V7 nabízí 14denní zkušební verzi, poté následuje zpoplatnění.

### 5.1.3 Labelbox

Velmi oblíbeným anotačním nástrojem je LabelBox. Byl vyvinut v roce 2018 a rychle si získal popularitu. Nabízí klasickou anotaci snímku pomocí rámečků a mnohoúhelníků, zároveň ale umožňuje i anotaci textových dokumentů. Rozhraní pro anotaci je zjednodušené, snadno se v něm tak zorientuje i začátečník. Běží přes webový prohlížeč a nabízí tři možnosti užívání. První z nich zahrnuje možnost anotace 10 000 snímků zdarma. Druhá varianta je zpoplatněna a je určena pro větší společnosti nebo pro anotaci lékařských dat. Poslední forma softwaru již nabízí speciální prostředí pro anotaci a je doporučována firmám, které se plně orientují na trénování několika neuronových sítí.

### 5.1.4 LabelMe

Tento software slouží pro anotaci obrázků. Jedná se také o open source software, který je bezplatný. Jde o výsledek projektu společnosti MIT's Computer Science and Artificial Intelligence Laboratory. LabelMe se vyznačuje tím, že je navržen tak, aby dokázal rozpoznat třídy objektů a umožňuje vytváření dalších nových tříd. Mezi výhody tohoto nástroje bezpochyby patří poskytování dat (obrázků), na které se nevztahují autorská práva, což umožňuje vytváření anotací, které jsou pak snadno dostupné veřejnosti. Pro označení určitého objektu musí uživatel myší vytvořit mnohoúhelník, který co nejvíce odpovídá tvaru objektu a následně ho zařadit do vhodné třídy. U těchto veřejných dat může jiný uživatel udělat změnu (například pokud usoudí, že anotace není provedena správně). Volně přístupná data se tak neustále mění na základě toho, jak s nimi pracují uživatelé.

### **5.1.5 VGG Image Annotator**

VGG Image Annotator, zkráceně VIA, je jednoduchý software umožňující anotaci snímků i videí. Je to online software běžící ve webovém prohlížeči, který nevyžaduje instalaci. Nástroj je vyvinut společností Visual Geometry Group (proto zkratka VGG) a jedná se o open source. Může tak být využit pro soukromé i komerční účely.

## 6 Augmentace dat

Modely strojového učení vyžadují velké množství kvalitně anotovaných dat. Tento aspekt zásadně ovlivní přesnost neuronových sítí. Kromě kvality je důležitá i rozmanitost dat, což si lze představit jako záběry z různých úhlů a vzdáleností nebo rozdílná umístění objektů. Anotace je ovšem časově náročná. Tuto nevýhodu je možné vyřešit augmentací dat. Jejím cílem je zvýšit objem dat pro trénování neuronové sítě. Augmentace dat je široce používána pro metody hlubokého učení sloužící ke zpracování obrazu, jako je detekce a klasifikace objektů, či sémantická segmentace. Generováním nových a různorodých instancí pro trénovací datové sady augmentace zlepšuje výkon a výsledky hlubokého učení (Xu et. al., 2023).

### 6.1 Postup při augmentaci dat

Při augmentaci dat jsou aplikovány konvenční metody zpracování obrazu za účelem úprav, které lze rozčlenit do několika hlavních kategorií (Šonka et al., 2018):

- geometrické transformace obrazu,
- transformace jasu obrazových bodů,
- metody lokálního předzpracování obrazu.

Po zpracování dat je nutné, aby došlo ke kontrole dat a systém obdržel zpětnou vazbu. Pokud jsou data v pořádku, může je uživatel využít pro trénování neuronové sítě.

Existuje mnoho způsobů, jak provádět augmentaci dat. Jako transformaci obrazu můžeme chápat náhodné oříznutí či oříznutí na střed, otáčení, převrácení nebo změnu velikosti obrazu. Z hlediska barev lze měnit jas, kontrast či sytost obrazu nebo obraz převést do stupňů šedi.



Obrázek 21: Ukázka augmentace

Při augmentaci můžeme rozlišovat data rozšířená a syntetická. Syntetická data jsou generována uměle bez využití skutečných obrázků. Rozšířená data vznikla určitou transformací skutečných obrázků (například rotace či převrácení). Tak dojde k rozšíření datasetu pro trénování neuronové sítě.

V následujících bodech jsou uvedené hlavní důvody, proč je augmentace výhodná.

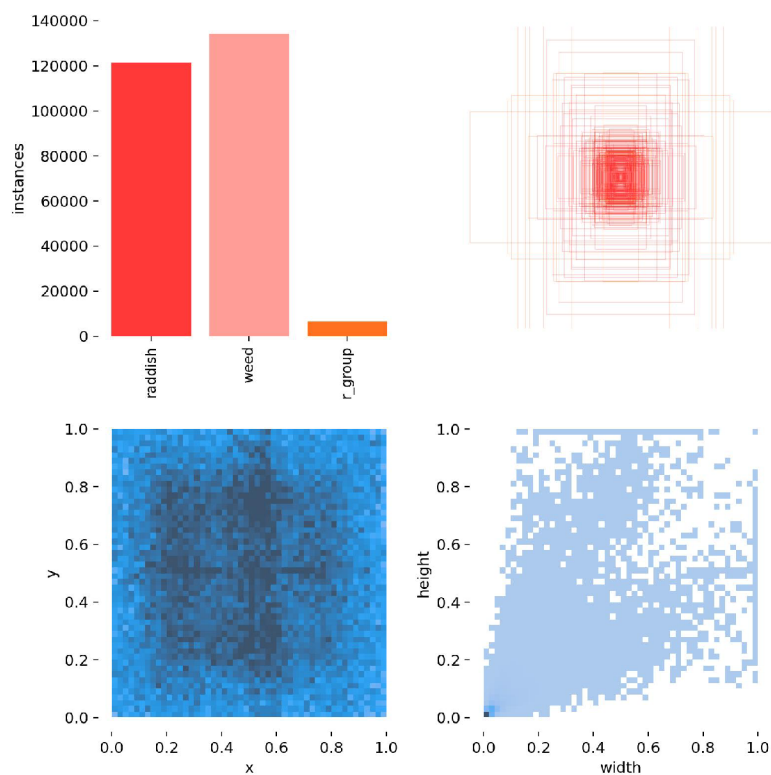
- Augmentace může zvýšit přesnost inferencí neuronové sítě. Aplikuje se u metod hlubokého učení, mezi které patří detekce objektů, klasifikace či sémantická segmentace.
- Snižuje časovou náročnost související se sběrem dat. Vytváření datasetů je obecně náročné, a to nejen časově. Techniky augmentace (či rozšíření dat) mohou tuto časovou náročnost snížit.

Tento proces je hojně využíván v mnoha důležitých oblastech. Jednou z ní je zdravotnictví. Je ovšem nutné dbát na výběr vhodného způsobu augmentace. U lékařských snímků je vhodné vyhnout se transformačním metodám. Dalším odvětvím, ve kterém se setkáme s augmentací dat, je vývoj autonomních vozidel. S pomocí augmentace získá síť dostatečné množství dat pro trénování v simulovaném prostředí (Shah, 2020).

## 7 Vybrané vstupy a parametry procesu trénování

### 7.1 Vlastnosti vstupního datasetu

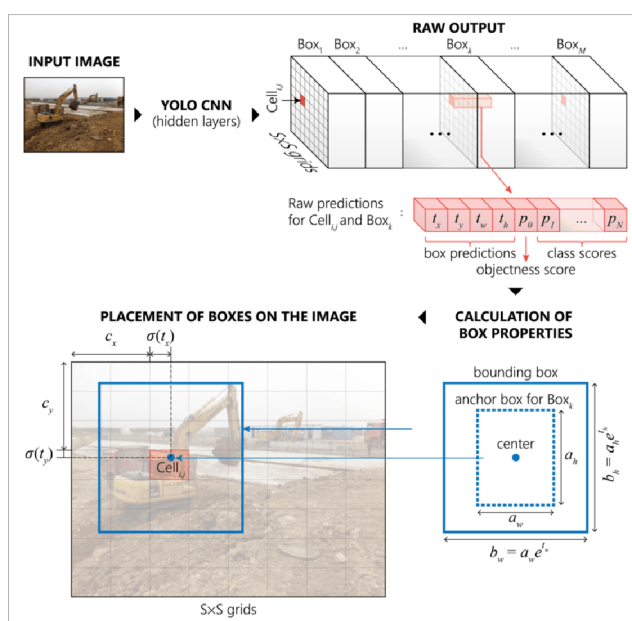
Vstupní data byla zpracována pomocí softwaru CVAT, následně byla augmentována (výřez na šířku, zmenšení, rotace o 90°, 180°, 270°, změna perspektivy, rozostření, úprava jasu a kontrastu, přidání šumu, posun barevné škály). Vzniklo 15 138 trénovacích obrázků, které obsahují 121 303 labelů ředkviček, 134 139 labelů plevelů a 6850 labelů označujících skupinu ředkviček. Evaluačních snímků vzniklo 2000. Sítě byly trénovány na GPU serveru, který obsahoval 3× GPU NVIDIA A30 s 24 GB VRAM a 16 GB RAM.



Obrázek 22: Znázornění vlastností anotací vytvořeného vstupního datasetu

## 7.2 Architektura YOLO

YOLO (You only look once) je jednou z nejpopulárnějších modelových architektur detekční neuronové sítě. Algoritmus má za úkol lokalizovat objekt ohraničením a následně ho přiřadit do třídy, a to s určitou pravděpodobností predikce. U nalezeného objektu je název třídy, do které ho síť přiřadila, ponechán. Každý rámeček ohraničující objekt je popsán pomocí tří parametrů - výška, šířka a střed rámečku (Siddell, 2022).



Obrázek 23: Schéma YOLO (Nath a Behzadan, 2020)

Výhodou této architektury je rychlost, kterou prochází snímky. Tento aspekt umožňuje užití YOLO v reálném čase, čehož může být využito například v autonomních automobilech. Od prvního vydání systému YOLO, které se připisuje k roku 2016, vzniklo mnoho dalších verzí. První tři verze představil Joseph Redmon v letech 2016 - 2018. Tyto verze jsou označovány jako YOLOv1, YOLOv2 a YOLOv3. Každá nová verze se od té předchozí lišila především ve vyšší rychlosti zpracování snímků. Zlepšovala se také přesnost. Redmon ve vývoji nadále nepokračoval, i přesto dál vznikaly nové verze. YOLOv4 využívá nové metody BOF (Bag of Freebies). Jedná se o soubor metod či tréninkových po-

stupů, jejichž aplikace vede ke zlepšení přesnosti modelu (Shah, 2020). Model YOLOv5 byl vydán krátce po zveřejnění čtvrté verze. Je to první z modelů napsaný v rámci PyTorch (jedná se o framework pro trénování neuronových sítí). V porovnání s předchozí verzí je jednodušší pro použití z uživatelského hlediska (Solawetz, 2020). Další verze, čili YOLOv6 a YOLOv7 předčily své předchůdce nejen vysokou rychlostí, ale i nižšími nároky na hardware. Nejnovější z modelů, YOLOv8, je vylepšena novými funkcemi, které zvyšují rychlost, přesnost a zároveň umožňují využití pro detekci objektu, segmentaci obrazu i klasifikaci obrazu (Boesch, 2023).

### 7.3 Vybrané hyperparametry

Hyperparametry jsou charakteristickou proměnnou ovlivňující trénování dané sítě. Jsou zadány před začátkem trénování. Pro optimální výsledky je nutné hyperparametry vhodně nastavit. Tato část kapitoly se zabývá několika takovými hyperparametry.

- Learning rate představuje velikost kroku, kterým dochází k průchodům ztrátové funkce optimalizačního algoritmu, přičemž cílem je nalézt globální optimum nastavení vah. Zjednodušeně řečeno se jedná o rychlost, kterou se model učí. Příliš vysoký Learning rate zapříčiní vynechání lokálních minim. Pokud by naopak byl příliš nízký, bylo by trénování časově velmi náročné. K nalezení optima je třeba více trénovacích epoch, avšak může také dojít k nežádoucímu uvíznutí v lokálním minimu (Murphy, 2012).
- Batch size udává počet vzorků z trénovací sady dat, na kterých model provádí trénování v jednom čase. Menší hodnota tohoto parametru může i nemusí být výhodná. Nízká hodnota Batch size znamená nízké nároky na paměť. Po dokončení trénování sady vzorků může uživatel parametry aktualizovat. Nevýhodou tohoto postupu je ovšem zhoršení přesnosti (Brownle, 2022).



- **Epocha.** Počet epoch definuje, kolikrát model pracuje s celým souborem dat. Tím dochází k minimalizaci chybového vyhodnocení (Brownle, 2022).
- **Optimizéry** jsou metody ladící metody neuronové sítě během jejího trénování. Tyto algoritmy mohou měnit váhy nebo rychlost učení takovým způsobem, aby výsledky byly co nejpřesnější (Doshi, 2019).

## 7.4 Metriky pro vyhodnocení

Metriky chápeme jako funkce pro posouzení výkonu modelu neuronové sítě. Tyto a další podobné výstupy trénovaného modelu jsou popsány metrikami.

- **Confusion matrix** představuje souhrn výsledků predikce. Každý sloupec matice představuje predikované instance, každý řádek matice představuje skutečné instance. Při vyhodnocení výsledků dělíme predikce na pozitivní (P) a negativní (N), skutečnost je pak popsána stavy True (T) a False (F). Sloučíme-li predikci a pravdivost, získáme (Mishra, 2021):
  - **True Positive (TP).** P představuje pozitivní predikci, T potvrzuje, že je predikce pravdivá.
  - **True Negative (TN).** N zamená negativní predikci, T potvrzuje pravdivost predikce.
  - **False Positive (FP).** P vyjadřuje pozitivní predikci, F ovšem znamená, že předpověď pravdivá není a jedná se o chybné vyhodnocení.
  - **False Negative (FN).** N ukazuje, že predikce je negativní, F vyjadřuje chybnou predikci.
- **Precision (Přesnost)** je definována jako podíl správných pozitivních výsledků a všech pozitivních výsledků (Mishra, 2018).

$$P = \frac{TP}{TP + FP}$$

- Recall vyjadřuje podíl správných pozitivních výsledků a součtu všech správně pozitivních a falešně negativních výsledků (Mishra, 2018).

$$R = \frac{TP}{TP + FN}$$

- Intersection over Union (IoU) vyhodnocuje míru překrytí pravdy a predikce. Hodnota funkce se pohybuje mezi 0 a 1, kdy 0 znamená žádné překrytí a 1 dokonalé překrytí. Tato funkce je získávána z prahových hodnot (Koech, 2020).

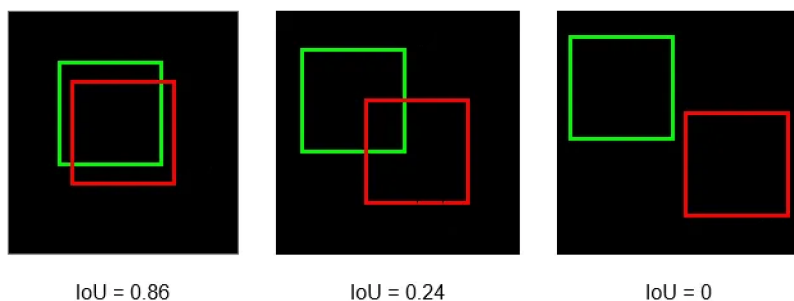
$$IoU = \frac{overlap}{union}$$

- Mean Average Precision (mAP) či Average Precision (AP) vyjadřuje průměrnou přesnost modelu. V mnoha pracích je AP uváděn jako průměr pro každou třídu zvlášť a mAP pro průměr z více tříd. Aktuálně je však častěji užíván jeden obecný termín AP (Maxwell, Warner a Guillén, 2021).

$$AP = \frac{1}{N} \sum_{i=1}^N AP_i$$

- F1 vyjadřuje vážený průměr přesnosti.

$$F1 = 2 \cdot \frac{Precision \cdot Recall}{Precision + Recall}$$



Obrázek 24: Různé hodnoty metriky IoU (Koech, 2020)

## 8 Vyhodnocení získaných výsledků

Několik vybraných modelů architektury YOLO bylo trénováno na vytvořeném datasetu a následně porovnáno z hlediska vyhodnocovacích metrik. Zkoumán je také vliv změny dvou zvolených hyperparametrů (*Learning rate* a *Batch size*).

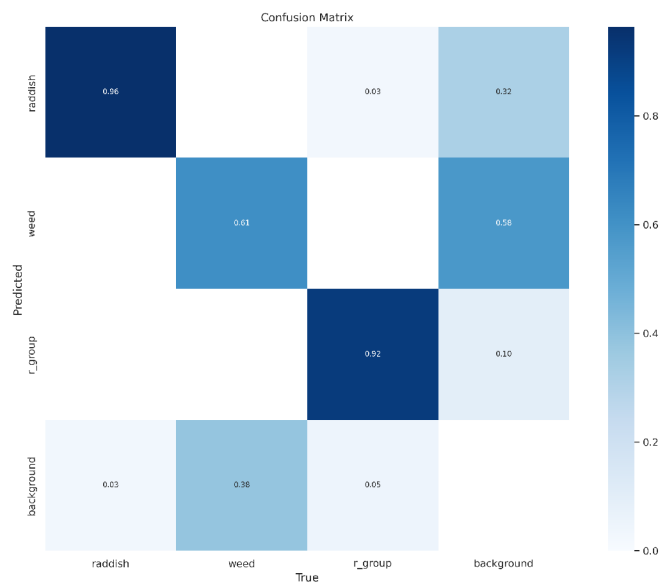
### 8.1 Porovnání YOLOv5n, YOLOv5m a YOLOv5x dle vybraných hyperparametrů

Jako první jsou porovnávány výsledky tří modelů - YOLOv5n, YOLOv5m a YOLOv5x, tedy nejmenšího, středního a největšího modelu dle počtu parametrů (Rath, 2022). Velikost vybraných hyperparametrů je v tomto případě konstantní. Hodnota Learning rate je rovna 0,01, velikost Batch size se rovná 6.

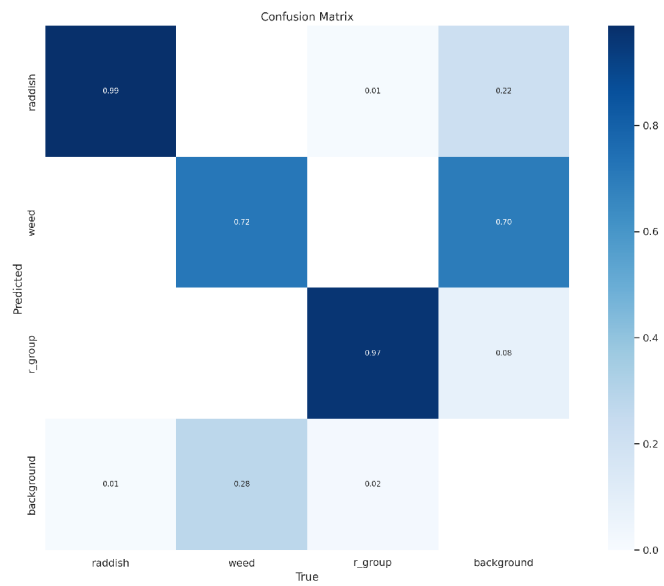
#### 8.1.1 Confusion matrix

Výsledky jsou vyhodnoceny na základě metriky zvané Confusion matrix. V matici existují tři třídy instancí, které sít rozeznává. V řádku se nachází skutečný název instance a ve sloupci předpovězená instance. Hodnoty v matici představují, kolik procent ze všech vzorků dané třídy předpověděl model správně. Modely YOLOv5m a YOLOv5x předpověděly 99 % ředkviček do správné třídy (*raddish*). Model YOLOv5n byl méně přesný, správně predikováno bylo 96 % vzorků třídy *raddish*. Mezi další třídy patří plevel a skupina ředviček (*weed* a *r\_group*).

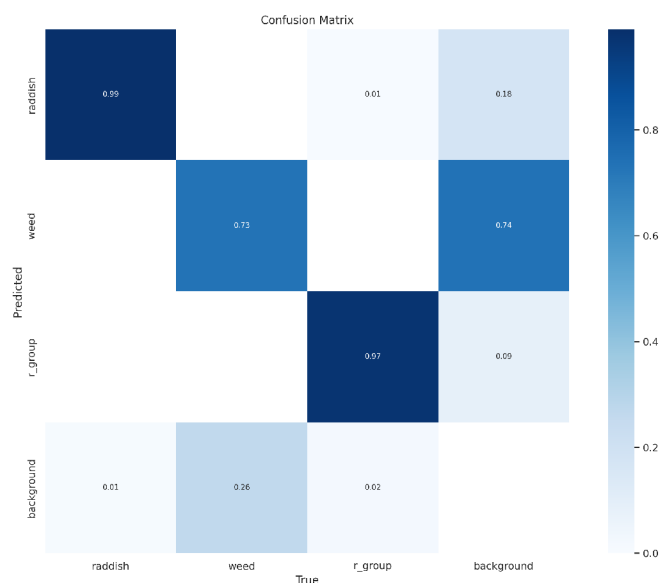
Background vyjadřuje falešně pozitivní výsledky, tedy kolik procent vzorků bylo jako instance dané třídy vyhodnoceno chybně. Dle matice byl z těchto tří modelů nejpřesnější YOLOv5x, neboť zařadil nejvyšší počet vzorků do správných tříd.



Obrázek 25: Confusion matrix - YOLOv5n



Obrázek 26: Confusion matrix - YOLOv5m

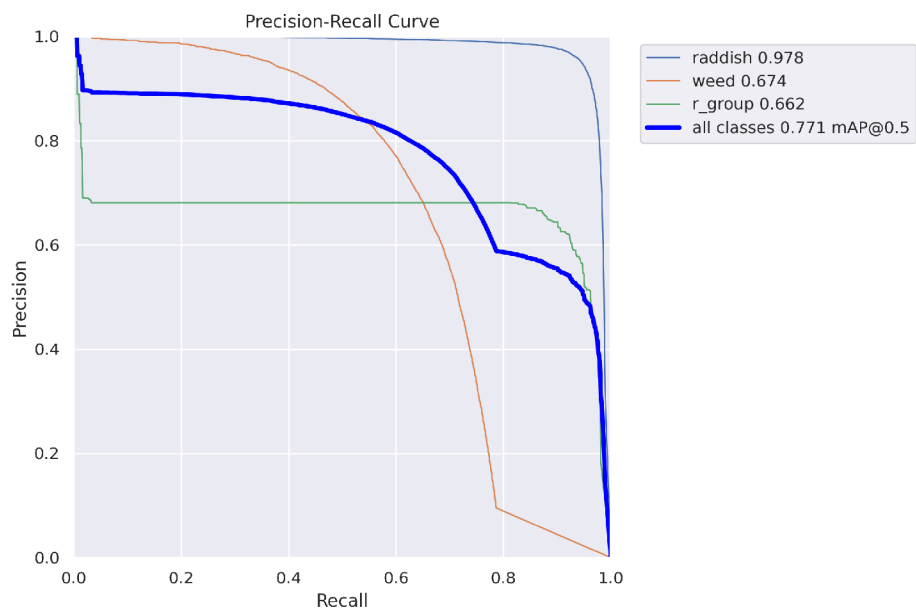


Obrázek 27: Confusion matrix - YOLOv5x

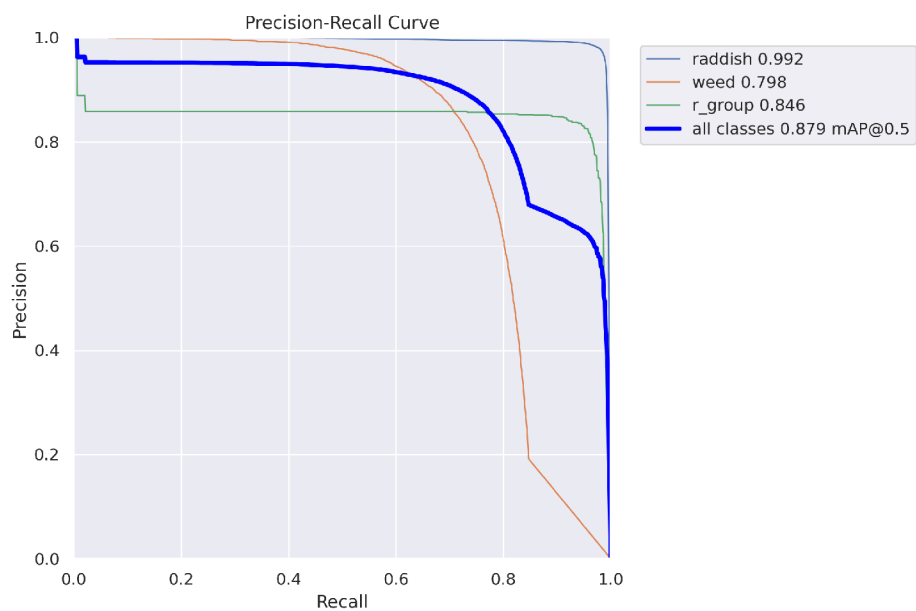
### 8.1.2 Precision-Recall Curve

Tato křivka vzniká sloučením dvou metrik, které jsou popsány v kapitole 7.2. Jedná se o *Precision* (přesnost) a *Recall* (senzitivita vybavení). Hodnoty *Precision* jsou naneseny na ose  $y$ , *Recall* pak na ose  $x$ . Čím větší je plocha pod zvolenou křivkou, tím přesnější je predikce vzorků dané třídy.

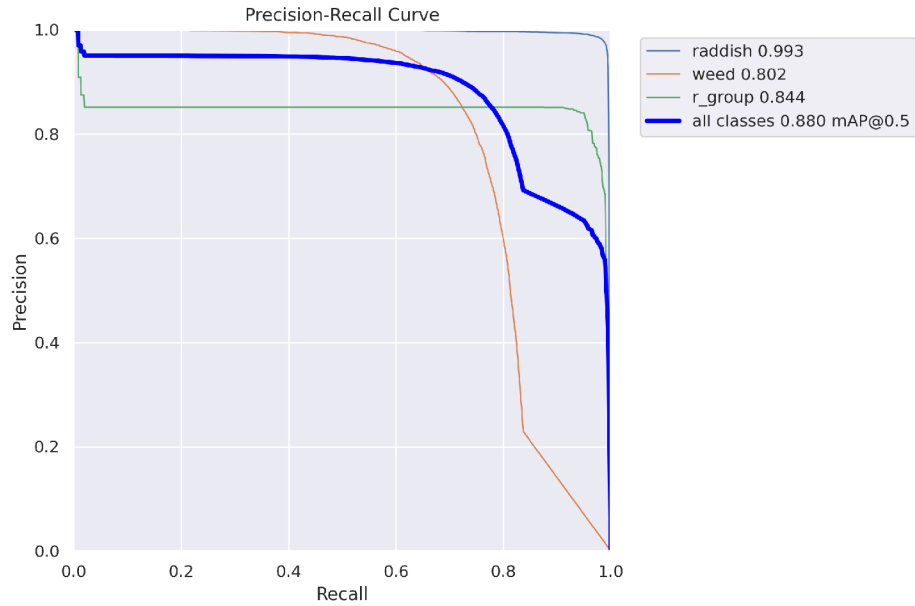
Hodnoty pro jednotlivé třídy instancí jsou uvedeny vedle grafů, a to včetně celkové průměrné přesnosti (plocha pod křivkou) při hodnotě  $IoU = 0,5$ . Vysoká hodnota přesnosti v praxi znamená nízkou četnost falešně pozitivních výsledků. Vysoká hodnota senzitivity vybavení značí nízkou frekvenci výskytu falešně negativních výsledků. Nejvyšší průměrná přesnost je u modelu YOLOv5x. Nejméně příznivých výsledků dosáhl dle této metriky, stejně jako u konfuzní matice, model YOLOv5n.



Obrázek 28: Precision-Recall Curve - YOLOv5n



Obrázek 29: Precision-Recall Curve - YOLOv5m



Obrázek 30: Precision-Recall Curve - YOLOv5x

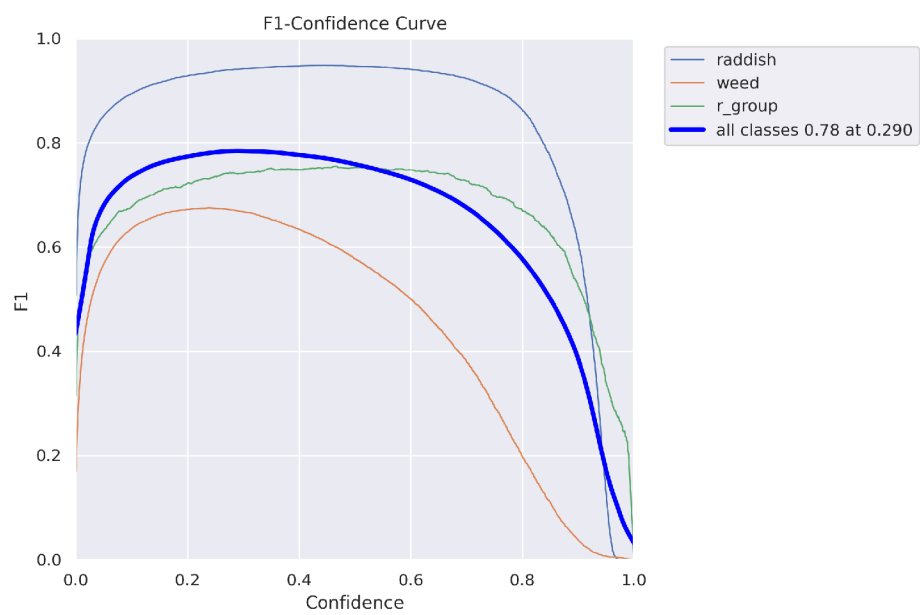
### 8.1.3 Čas trénování a vytížení GPU

Model	Průměrný čas epochy [s]	Průměrné vytížení GPU [GB]
YOLOv5n	234,41	3,43
YOLOv5m	288,69	5,74
YOLOv5x	1035,58	10,75

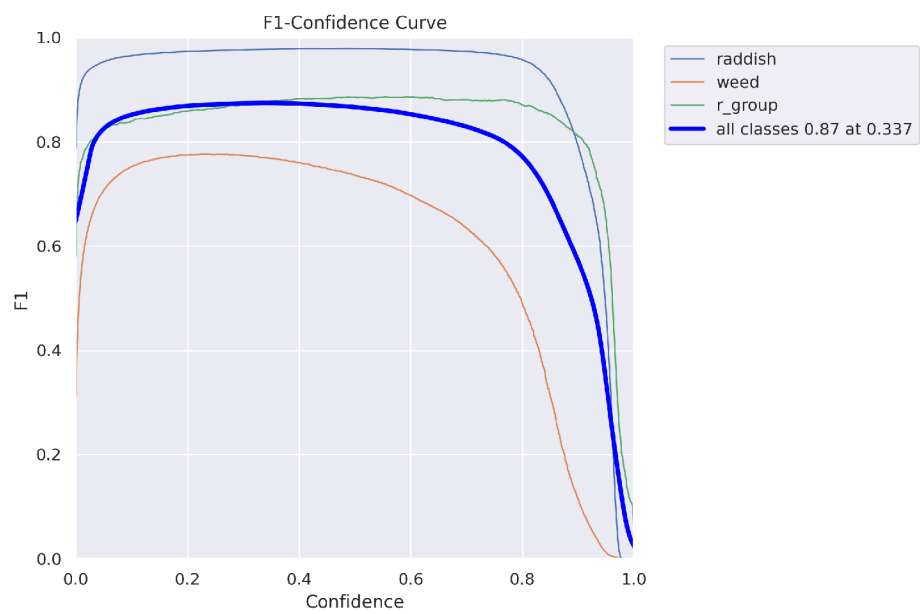
Tabulka 1: Průměrný čas epochy a vytížení GPU při trénování modelů YOLOv5

### 8.1.4 F1-Curve

F1 představuje vážený harmonický průměr přesnosti (Precision) a senzitivity vybavení (Recall), přičemž obě tyto metriky mají stejnou důležitost. Křivka F1 posuzuje prediktivní schopnosti modelu (Yan a Zhang, 2021). Hodnota skóre F1 může nabývat hodnot od 0 do 1. Nejvyšší hodnota F1 byla získána také u největšího modelu YOLOv5x. Hodnota spolehlivosti se zde rovná 0,376.

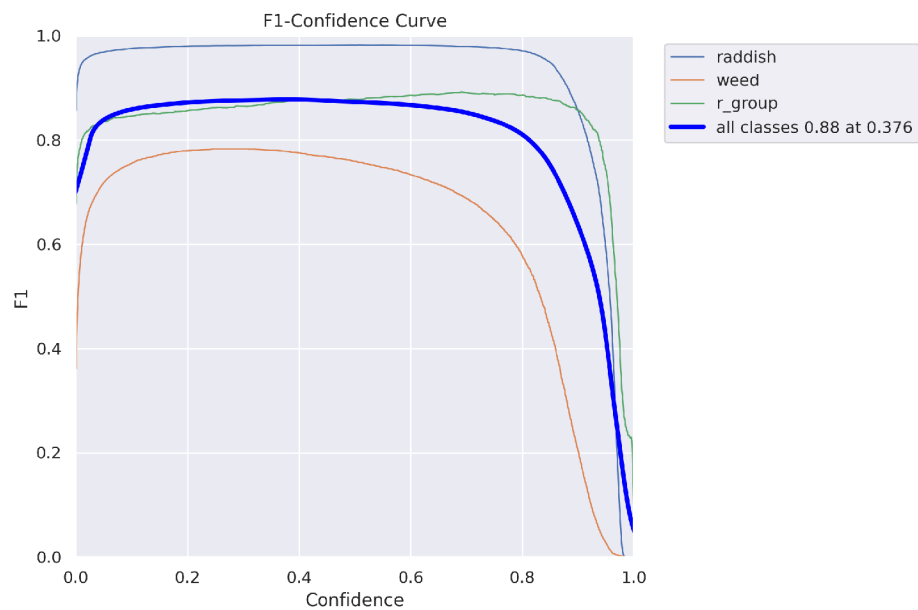


Obrázek 31: F1 Curve - YOLOv5n



Obrázek 32: F1 Curve - YOLOv5m

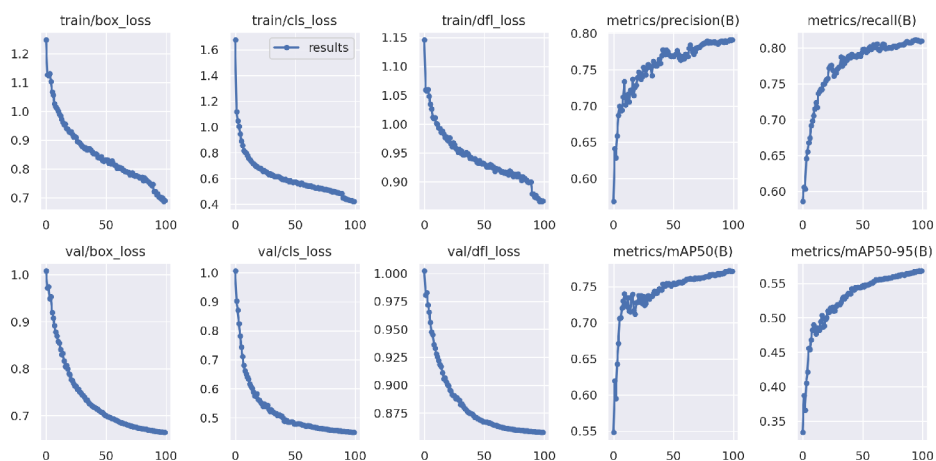




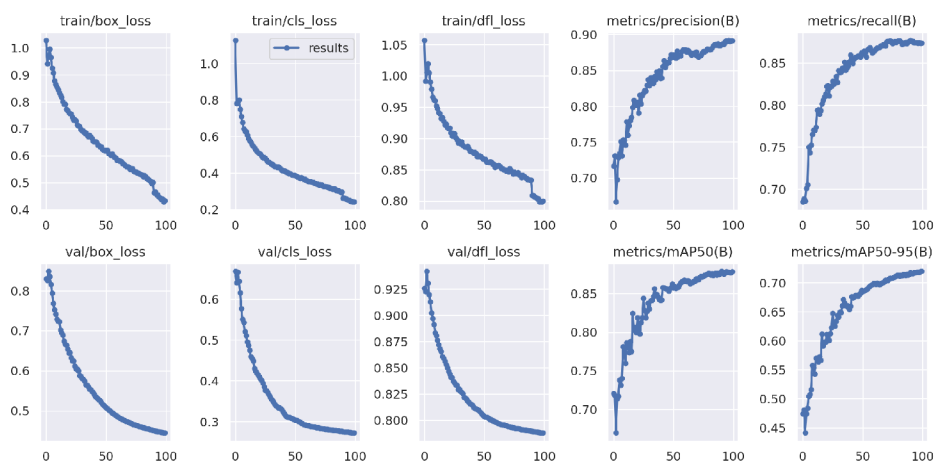
Obrázek 33: F1 Curve - YOLOv5x

### 8.1.5 Výsledky průběhu trénování modelů YOLOv5

Pro určení nejspolehlivějších výsledků je třeba zaměřit se nejen na zmíněné metriky, ale i na sledování všech typů ztrát. V celkových výsledcích trénování jsou tři typy ztrát: *Box Loss*, *Classification Loss* a *Distributional Focal Loss*.

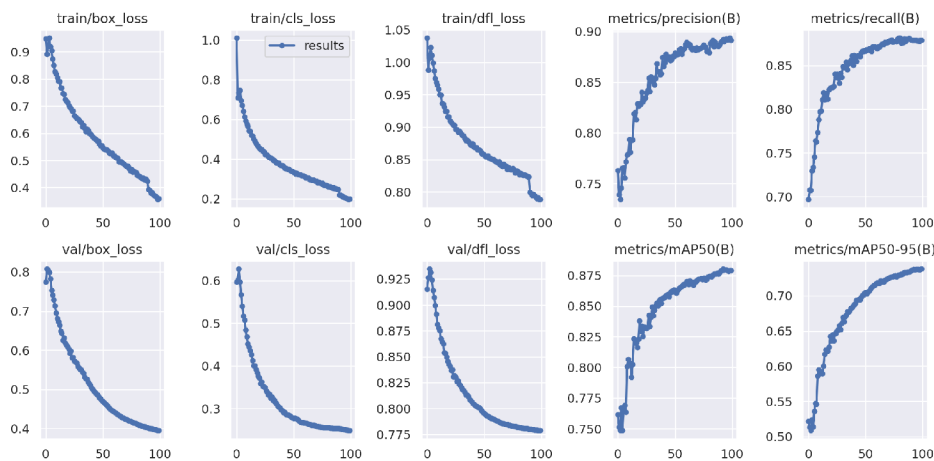


Obrázek 34: Průběh trénování - YOLOv5n



Obrázek 35: Průběh trénování - YOLOv5m

*Box Loss* představuje, jak dobře dokáže algoritmus lokalizovat střed objektu a jak ohraničující rámeček pokrývá objekt. *Classification Loss* popisuje správnost predikce, *Distributional Focal Loss* optimalizuje rozložení rámečku. V prvním řádku jsou ztráty trénování, neboli *Training Loss*. Posuzují chybu modelu na trénovací množině dat. Na druhém řádku se nacházejí ztráty validace (*Validation Loss*). Sada dat je vyhrazena pro ověření výkonu modelu.



Obrázek 36: Průběh trénování - YOLOv5x



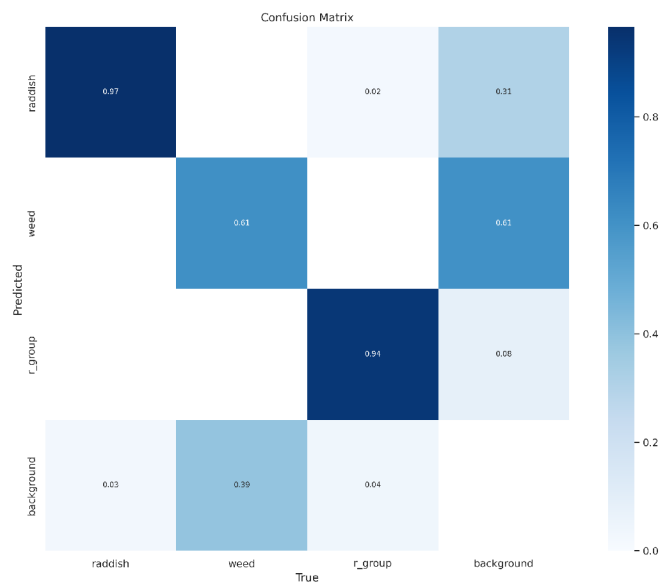
Obrázek 37: Trénování modelu YOLOv5x

## 8.2 Porovnání YOLOv8n, YOLOv8m a YOLOv8x dle vybraných hyperparametrů

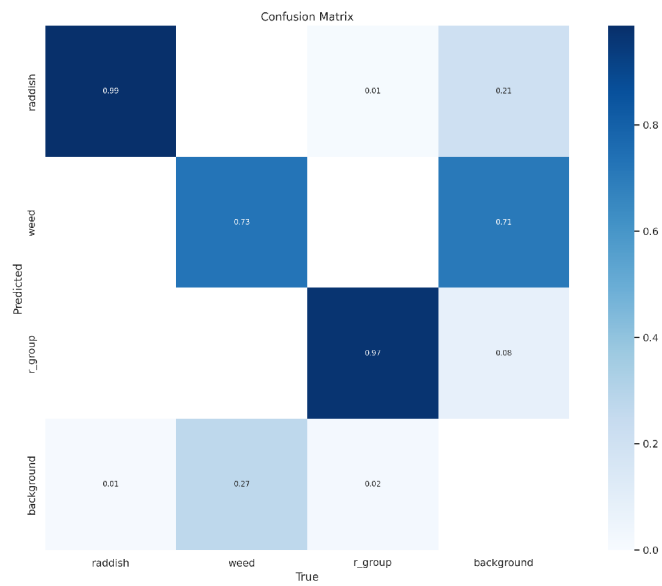
Trénování proběhlo i na novějším modelu architektury, opět na jeho třech verzích - YOLOv8n, YOLOv8m a YOLOv8x. Hodnoty hyperparametrů jsou stejné jako u předchozí verze.

### 8.2.1 Confusion matrix

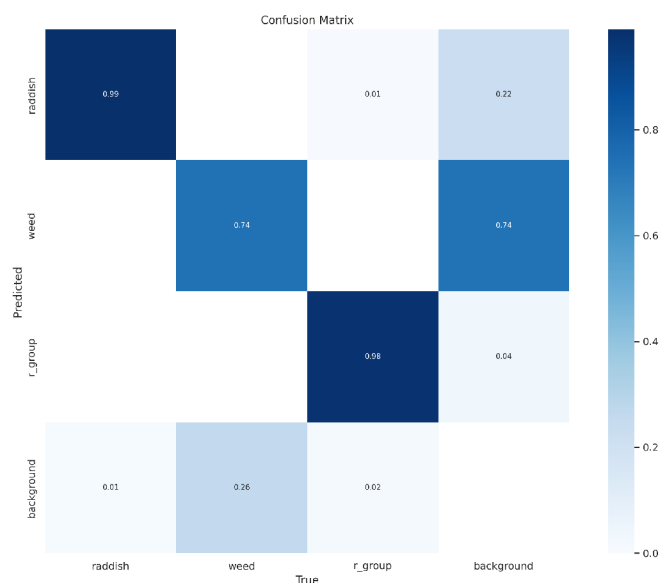
Z porovnání jednotlivých matic modelů vyplývá, že s nejnižší přesností rozeznával instance model YOLOv8n, a to především u třídy plevel (*weed*). Pouze 61 % vzorků této třídy předpověděl správně, zbylých 39 % vzorků jsou falešně pozitivní výsledky. To znamená, že model klasifikoval tento podíl vzorků jako plevel, ovšem chybně. Zbylé dva modely byly přesnější, dle matice má největší přesnost předpovědí model YOLOv8x.



Obrázek 38: Confusion matrix - YOLOv8n



Obrázek 39: Confusion matrix - YOLOv8m



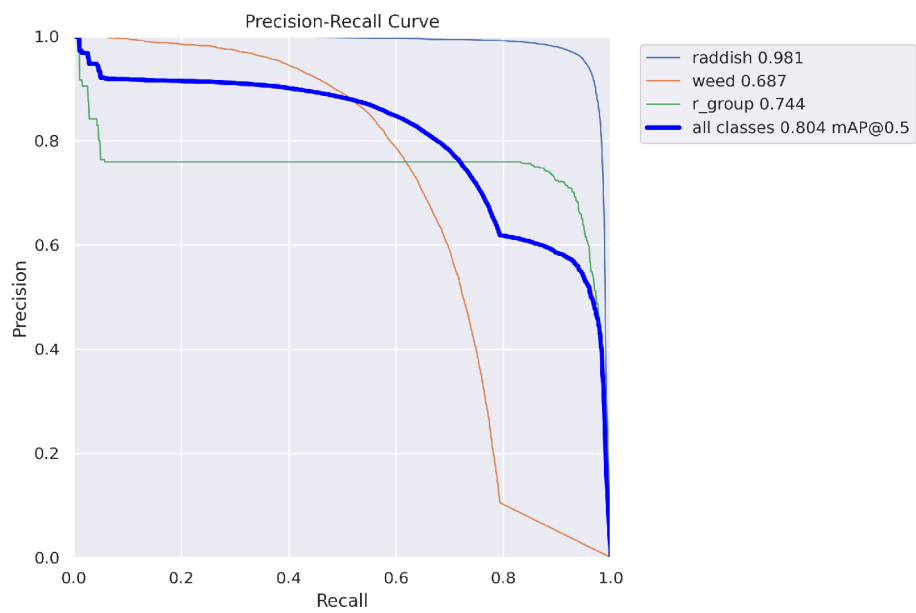
Obrázek 40: Confusion matrix - YOLOv8x

### 8.2.2 Precision-Recall Curve

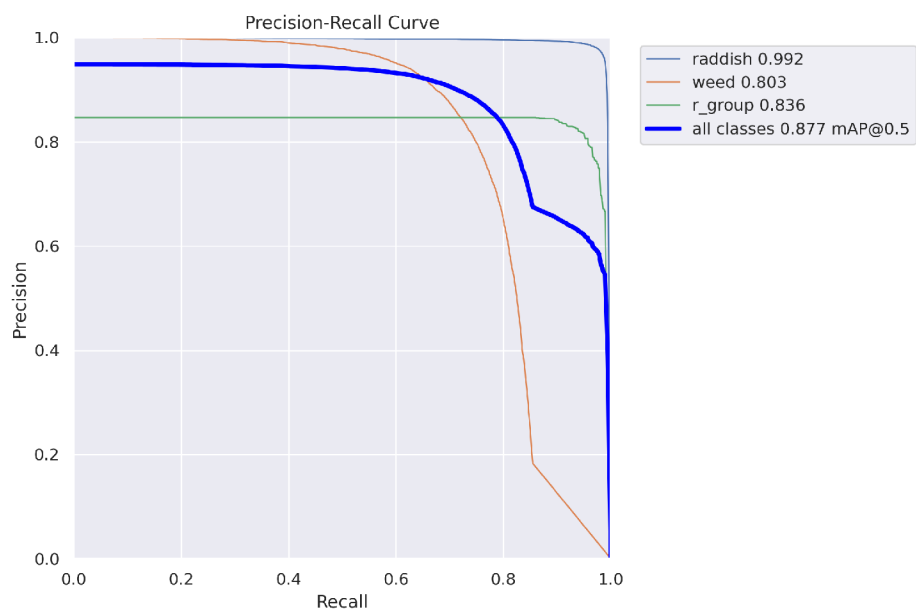
U této vyhodnocovací metriky o přesnosti modelu vypovídá plocha pod křivkou každé třídy, a také plocha pod křivkou průměrné přesnosti modelu. S největší průměrnou přesností vyhodnocoval podle této metriky model YOLOv8x, kdy hodnota průměrné přesnosti je rovna 0,934. Nejmenší z modelů, YOLOv8n, má nejnižší průměrnou přesnost ze tří sledovaných modelů, a to 0,804. Středně velký model, YOLOv8m, má průměrnou přesnost 0,877. Ze všech tří křivek je patrné, že největší potíže při rozeznávání působily instance třídy *weed*.

### 8.2.3 Čas trénování a vytížení GPU

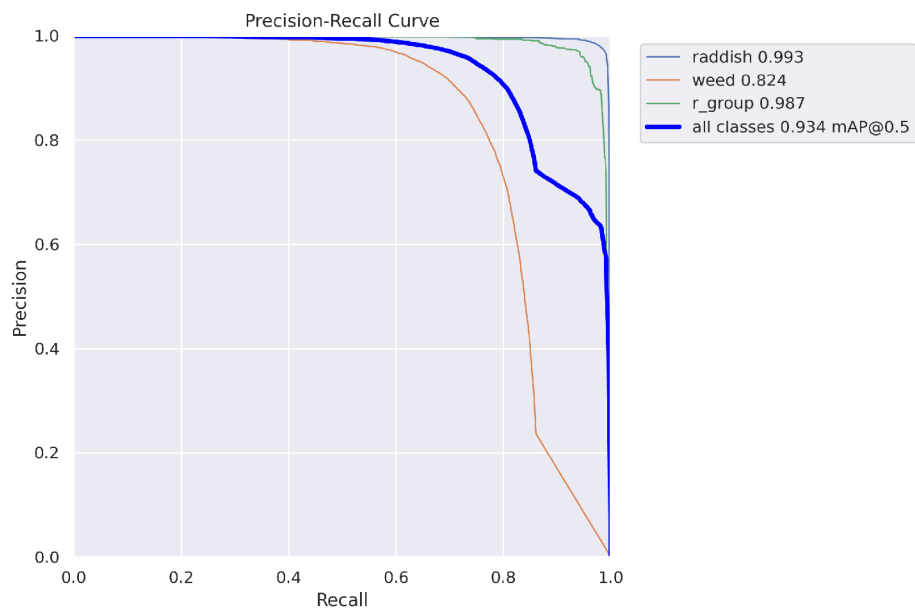
Z hlediska časové náročnosti dopadl nejlépe model YOLOv8m (viz Tabulka 2). Průměrné vytížení GPU se zvyšuje s velikostí modelu. Pokud by záleželo pouze na čase trénování a vytížení GPU, jeví se střední model YOLOv8m jako optimální variantou.



Obrázek 41: Precision-Recall Curve - YOLOv8n



Obrázek 42: Precision-Recall Curve - YOLOv8m



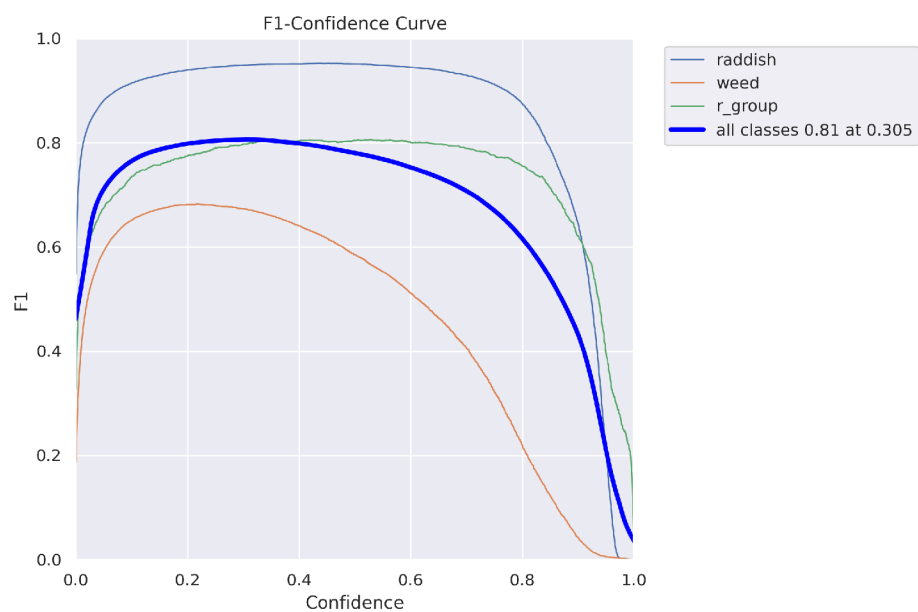
Obrázek 43: Precision-Recall Curve - YOLOv8x

Model	Průměrný čas epochy [s]	Průměrné vytížení GPU [GB]
YOLOv8n	352,43	3,66
YOLOv8m	297,21	5,79
YOLOv8x	580,47	9,11

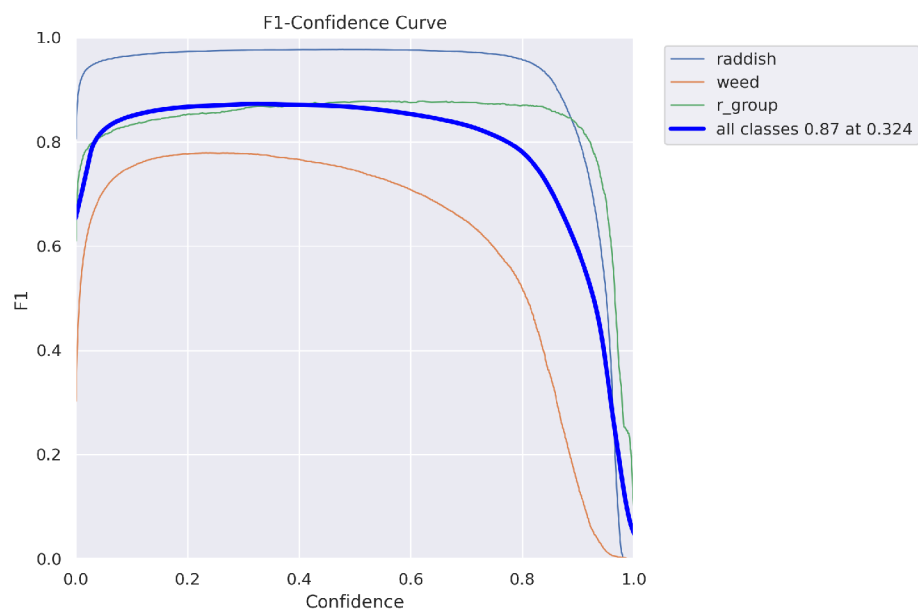
Tabulka 2: Průměrný čas epochy a vytížení GPU při trénování modelů YOLOv8

### 8.2.4 F1-Curve

Hodnota F1 je důležitým ukazatelem, který vyjadřuje rovnováhu mezi přesností (*Precision*) a senzitivitou vybavení (*Recall*). Nejvyšší hodnota F1 je dosaženo ve druhém případě, tedy u modelu YOLOv8m, kdy se tato hodnota rovná 0,324.

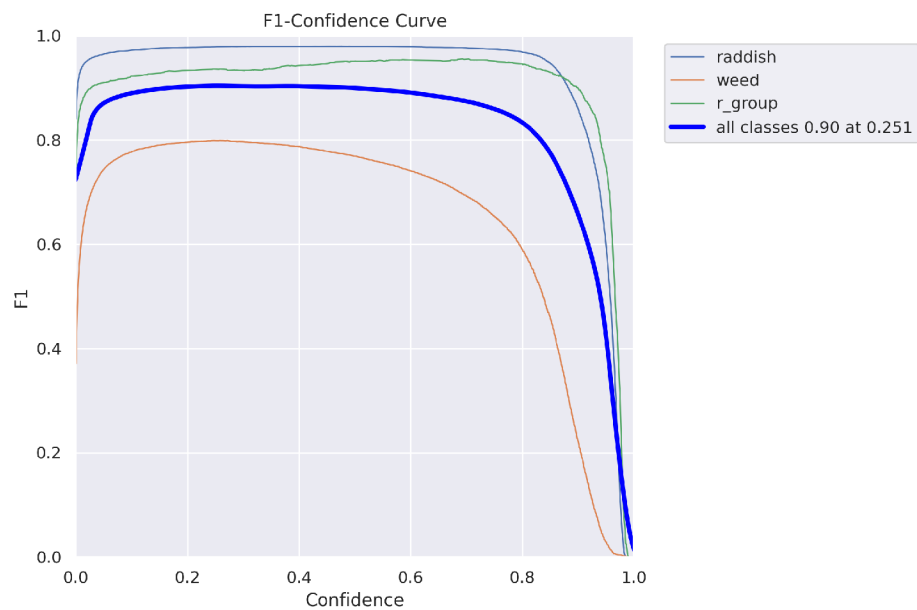


Obrázek 44: F1 Curve - YOLOv8n



Obrázek 45: F1 Curve - YOLOv8m

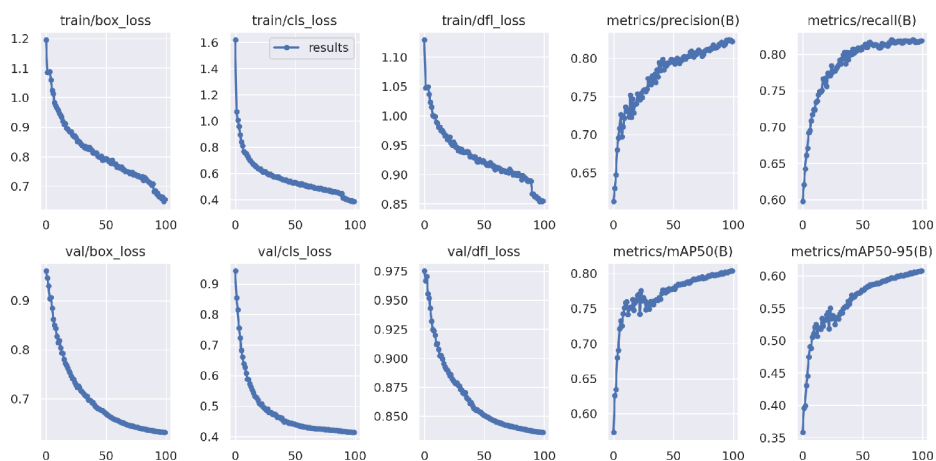




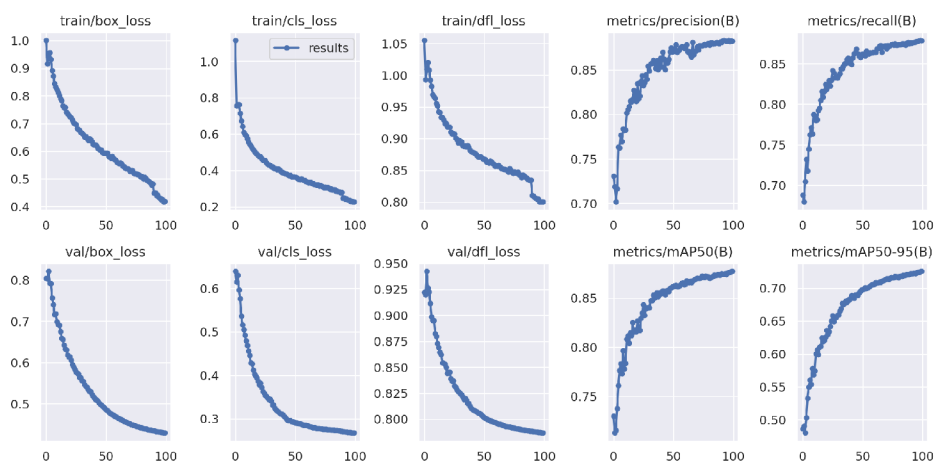
Obrázek 46: F1 Curve - YOLOv8x

### 8.2.5 Výsledky průběhu trénování modelů YOLOv8

Osa  $x$  vyjadřuje trénovací epochu, na ose  $y$  jsou naneseny hodnoty dané metriky. U nejmenšího modelu se ztráty snižovaly s časem trénování. Průměrná přesnost modelu se začala ustálovat při trénování epochy 80.

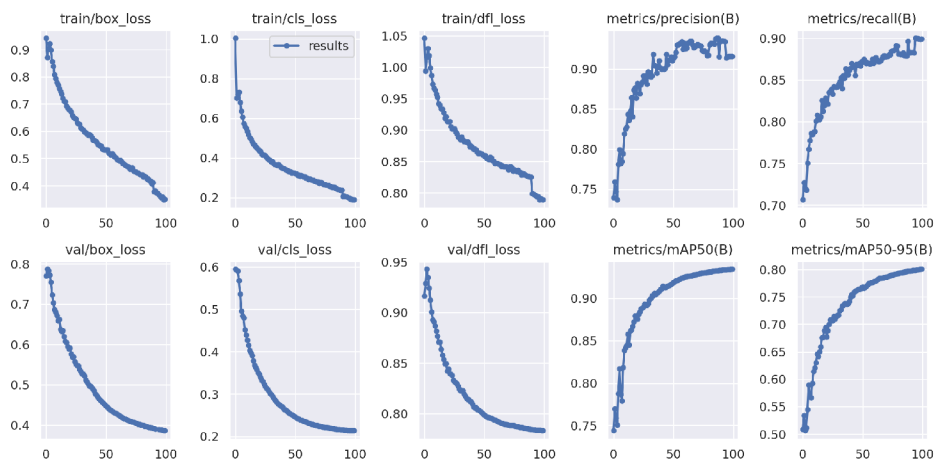


Obrázek 47: Průběh trénování - YOLOv8n



Obrázek 48: Průběh trénování - YOLOv8m

U dalšího z modelů (YOLOv8m) byly ztráty už na počátku nižší než u předchozího modelu, průměrná přesnost modelu se ustálila při 70. epoše. Počáteční hodnoty ztrát u YOLOv8x byly opět nižší. Metriky *Precision* a *Recall* nabývaly při trénování epochy 50 vysokých hodnot a s každou další epochou trénování se stabilizovaly.



Obrázek 49: Průběh trénování - YOLOv8x



Obrázek 50: Trénování modelu YOLOv8x

### 8.3 Vliv hyperparametru Learning rate

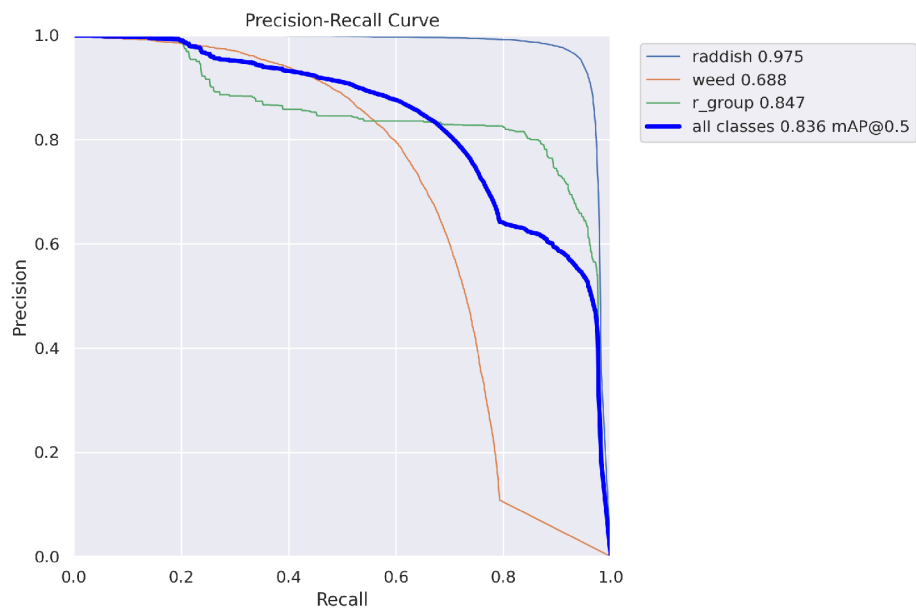
Ve druhém případě trénování probíhá na modelu YOLOv8n. Porovnány jsou výsledky jednoho modelu, avšak při třech různých hodnotách hyperparametru *Learning rate*. Cílem je posoudit, jak změna této hodnoty ovlivňuje výsledky trénování. Jednotlivé hodnoty zvolených hyperparametrů jsou následující:

1. *Learning rate* = 0,1; *Batch size* = 36,
2. *Learning rate* = 0,01; *Batch size* = 36,
3. *Learning rate* = 0,001; *Batch size* = 36.

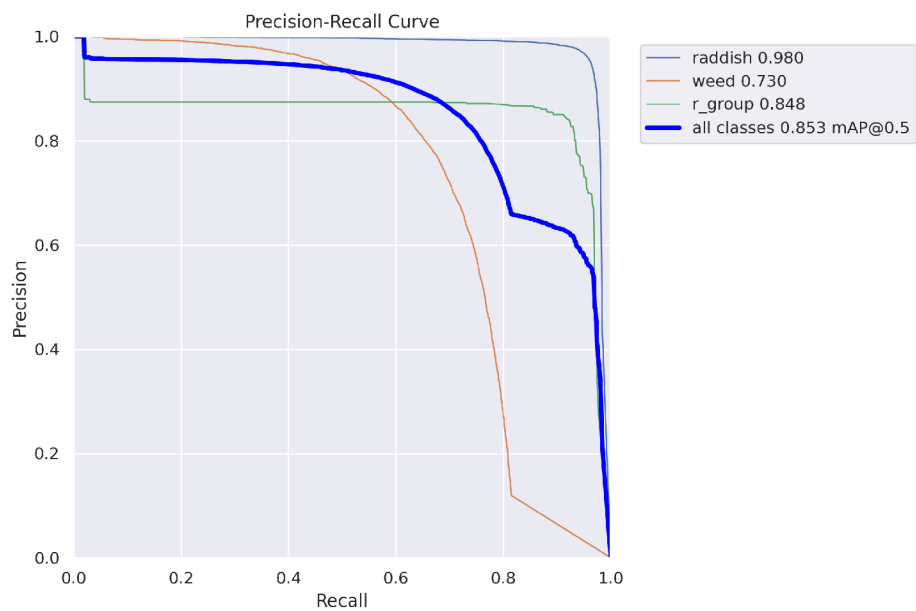
#### 8.3.1 Precision-Recall Curve

Je-li senzitivita vybavení vysoká, ale přesnost nízká, vrací model mnoho výsledků, ovšem často nesprávných. Naopak při vysoké přesnosti a nízké senzivitě vybavení model ve většině případů vrací výsledky správné, avšak v malém množství. Proto je ideální najít kompromis mezi přesností a senzitivitou (Powers, 2011).

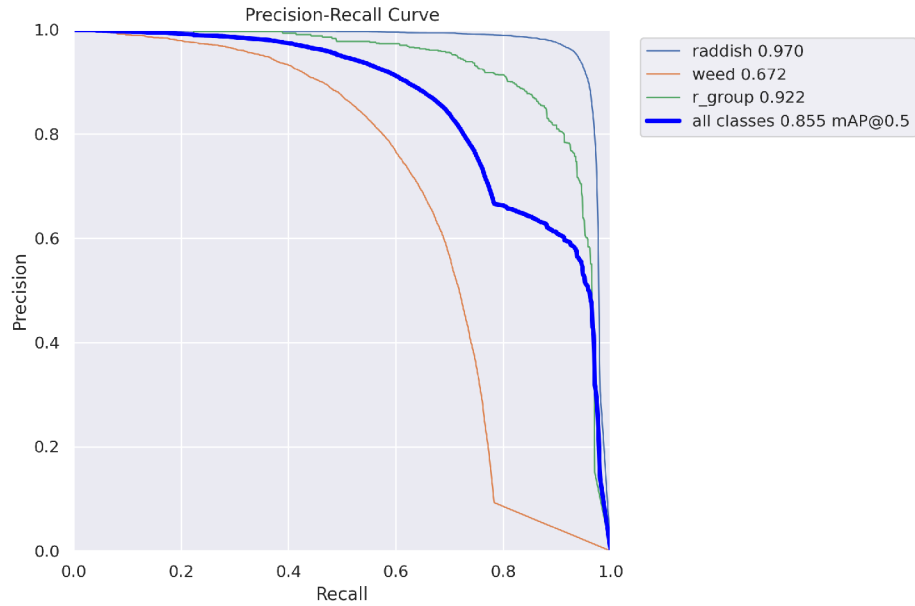
Vyšší průměrnou přesnost má model, kde *Learning rate* = 0,001. Výsledek modelu, kde *Learning rate* = 0,1, výrazně ovlivňuje přesnost třídy *r\_raddish*. Ta je u prvního trénování zřetelně nižší a ovlivňuje tak výsledky modelu.



Obrázek 51: Precision-Recall Curve - *Learning rate = 0,1; Batch size = 36*



Obrázek 52: Precision-Recall Curve - *Learning rate = 0,01; Batch size = 36*



Obrázek 53: Precision-Recall Curve -  $Learning\ rate = 0,001$ ;  $Batch\ size = 36$

### 8.3.2 Čas trénování a vytížení GPU

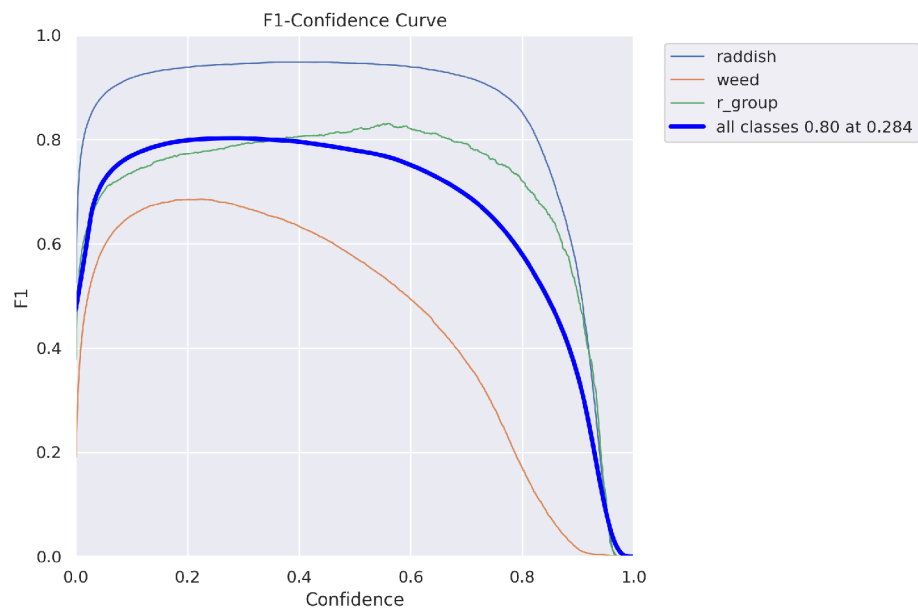
Jak ukazuje Tabulka 3, čas trénování modelů se zásadně neliší. Průměrné vytížení GPU je dokonce ve dvou případech stejné. Optimální nastavení parametru není dle těchto metrik jednoznačné. Je nezbytné zaměřit se na další ukazatele.

Hodnota parametru	Průměrný čas epochy [s]	Průměrné vytížení GPU [GB]
Learning rate = 0,1	113,76	14,97
Learning rate = 0,01	115,26	15,91
Learning rate = 0,001	113,49	14,97

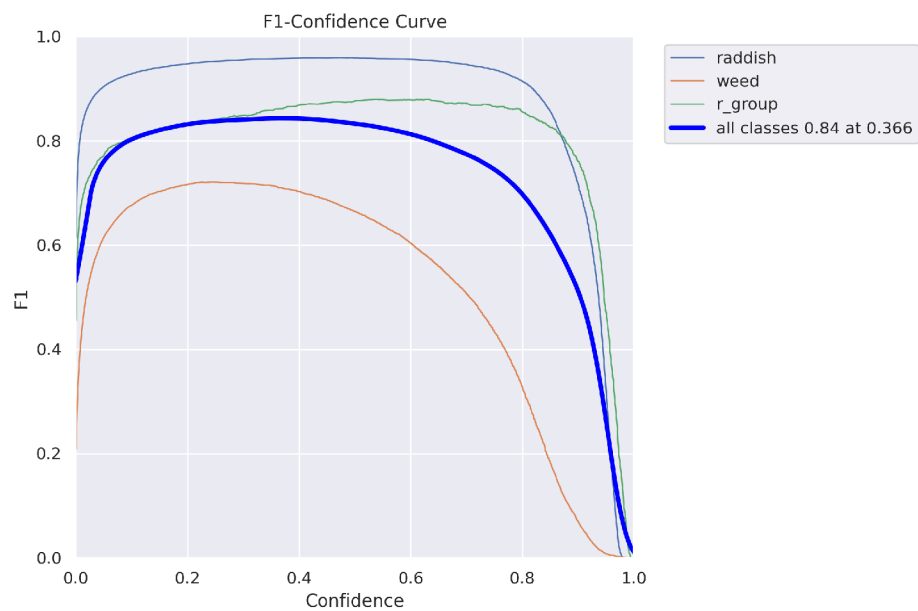
Tabulka 3: Průměrný čas epochy a vytížení GPU při rozdílných hodnotách hyperparametru Learning rate

### 8.3.3 F1-Curve

F1 vyjadřuje rovnováhu mezi přesností (*Precision*) a senzitivitou vybavení (*Recall*), přičemž obě tyto metriky mají stejnou důležitost.



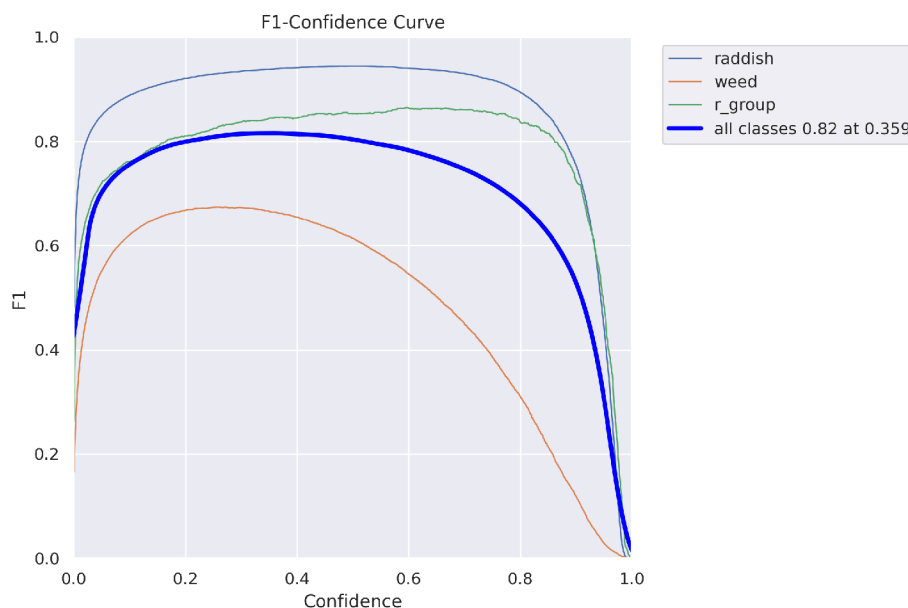
Obrázek 54: F1 Curve; *Learning rate = 0,1; Batch size = 36*



Obrázek 55: F1 Curve; *Learning rate = 0,01; Batch size = 36*

Graf říká, že pokud je zvýšena spolehlivost (*Confidence*), zvýší se i přesnost (*Precision*). Na úkor toho se však sníží senzitivita vybavení (*Recall*) (Lebiedzinski, 2021).

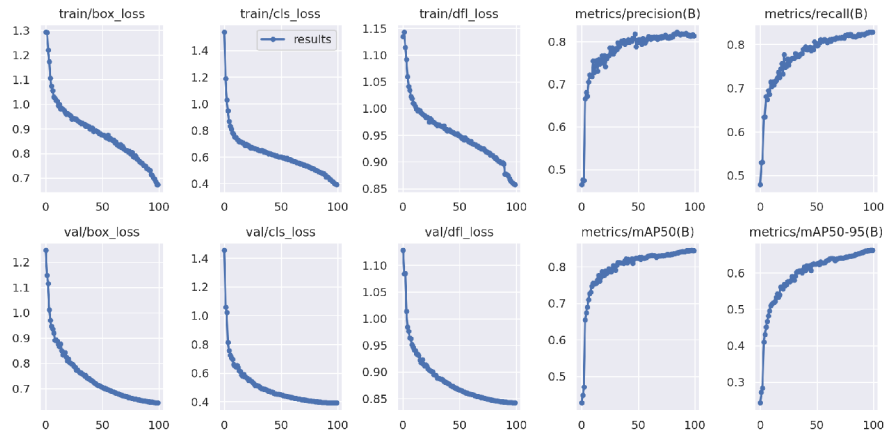
Nejpříznivější výsledek na základě hodnoty skóre F1 je pozorovatelný během druhého trénování, kdy *Learning rate* = 0,01.



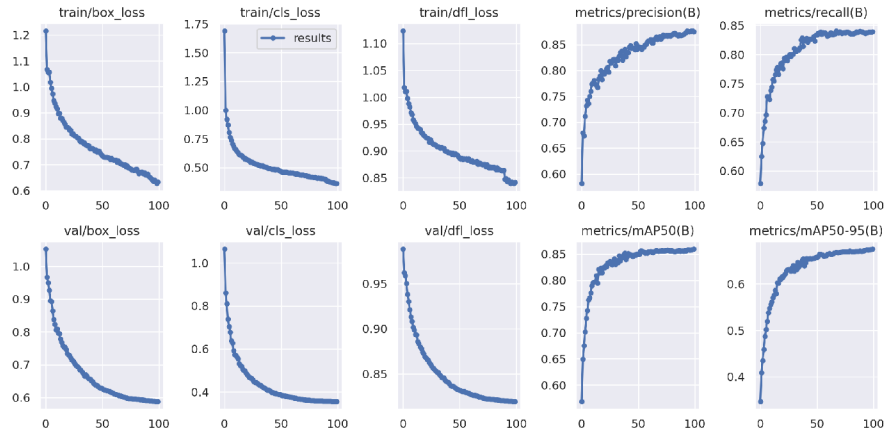
Obrázek 56: F1 Curve; *Learning rate* = 0,001; *Batch size* = 36

### 8.3.4 Výsledky průběhu trénování modelu YOLOv8n

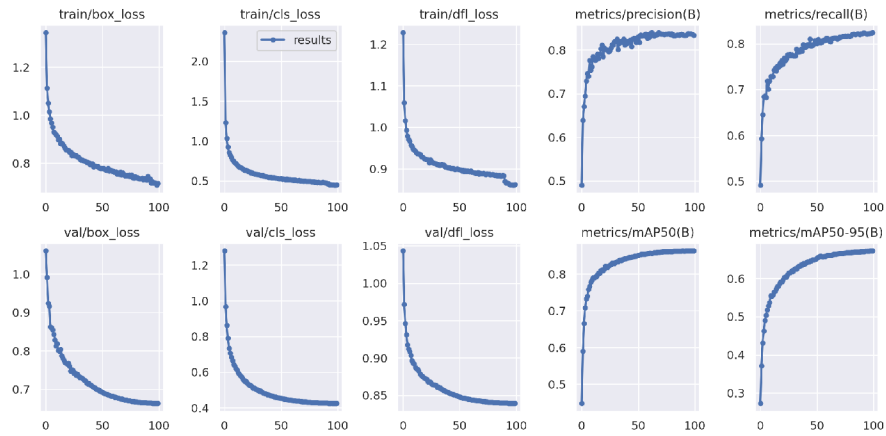
Shrnutí průběhu trénování je založeno nejen na vyhodnocovacích metrikách, ale i na sledování ztrát. Největší ztráty jsou zpočátku patrné u prvního trénování, kdy *Learning rate* = 0,1. Tyto ztráty byly ovšem postupně eliminovány. Další z důležitých ukazatelů, průměrná přesnost, je nejstabilnější u posledního trénování, čili kdy *Learning rate* = 0,001. Zároveň jsou v tomto případě zaznamenány nejnižší ztráty.



Obrázek 57: Průběh trénování;  $Learning\ rate = 0,1$ ;  $Batch\ size = 36$



Obrázek 58: Průběh trénování;  $Learning\ rate = 0,01$ ;  $Batch\ size = 36$



Obrázek 59: Průběh trénování;  $Learning\ rate = 0,001$ ;  $Batch\ size = 36$



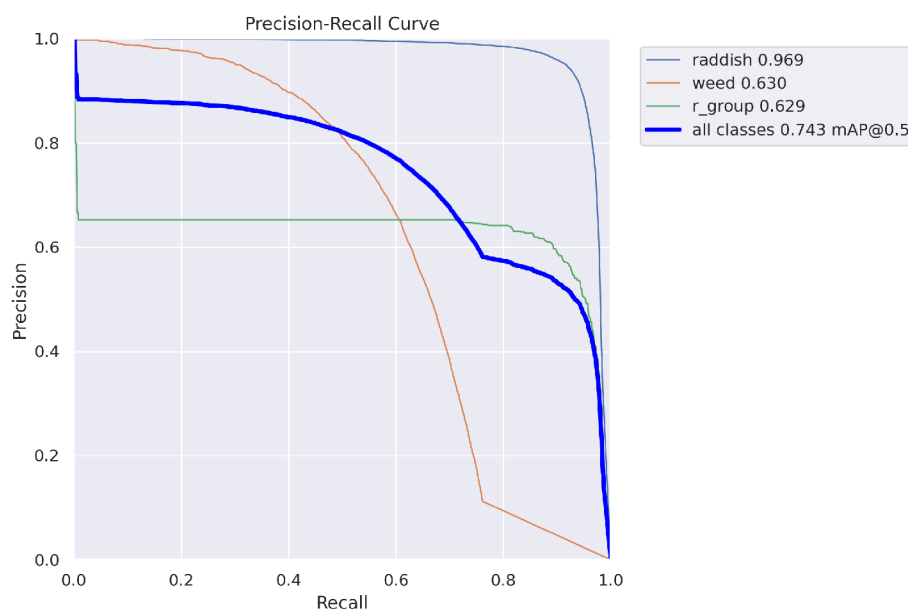
## 8.4 Vliv hyperparametru Batch size

Další z hyperparametrů, jehož vliv na výsledky trénování je v práci zkoumán, se nazývá Batch size. Porovnány jsou výsledky čtyř trénování, kdy hodnoty hyperparametrů jsou následující:

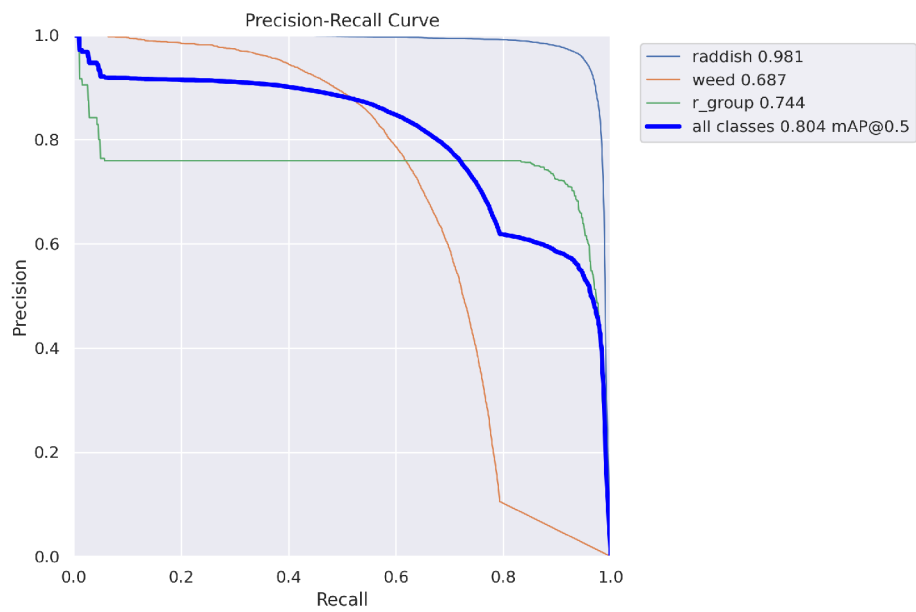
- *Learning rate* = 0,01; *Batch size* = 3,
- *Learning rate* = 0,01; *Batch size* = 6,
- *Learning rate* = 0,01; *Batch size* = 12,
- *Learning rate* = 0,01; *Batch size* = 36.

### 8.4.1 Precision-Recall Curve

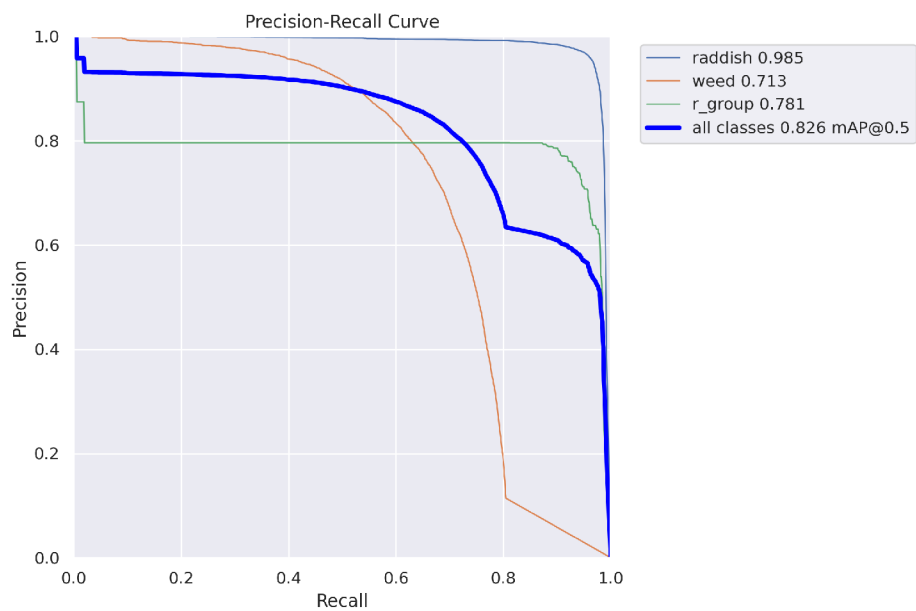
Výsledky jsou srovnány na základě uvedené metriky. Nejméně spolehlivé výsledky měl model opět v prvním případě (*Batch size* = 3). Průměrná přesnost modelu byla ze všech nejnižší.



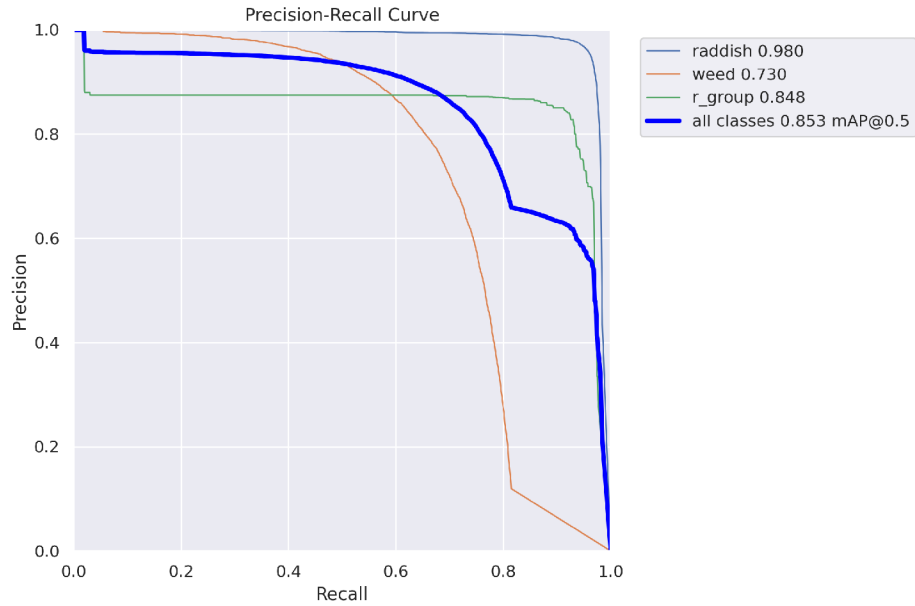
Obrázek 60: Precision-Recall Curve - *Learning rate* = 0,01; *Batch size* = 3



Obrázek 61: Precision-Recall Curve -  $Learning\ rate = 0,01$ ;  $Batch\ size = 6$



Obrázek 62: Precision-Recall Curve -  $Learning\ rate = 0,01$ ;  $Batch\ size = 12$



Obrázek 63: Precision-Recall Curve -  $Learning\ rate = 0,01$ ;  $Batch\ size = 36$

Chybné predikce jsou patrné především u vzorků plevelu a skupiny ředkviček (*weed* a *r\_raddish*). Celková průměrná přesnost se zde zřetelně liší od ostatních. Největší průměrná přesnost je dle této metriky v posledním případě trénování, tedy kdy  $Batch\ size = 36$ .

#### 8.4.2 Čas trénování a vytížení GPU

Je zřejmé, že průměrný čas jedné epochy je nepřímo úměrný parametru  $Batch\ size$ . Vytížení GPU se ovšem zvyšuje úměrně s tímto parametrem.

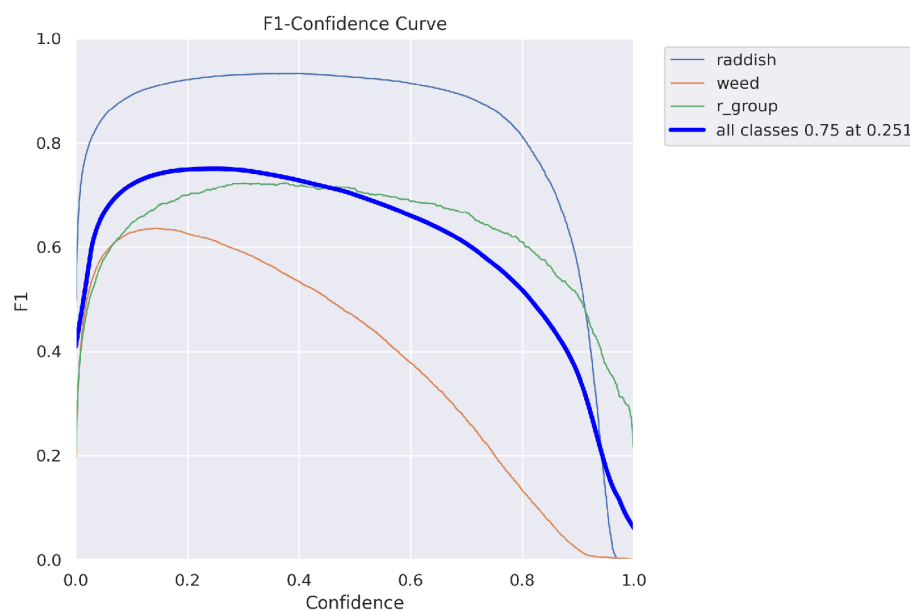
Hodnota	Průměrný čas epochy [s]	Průměrné vytížení GPU [GB]
Batch-size = 3	762,87	2,47
Batch-size = 6	354,26	3,66
Batch-size = 12	287,15	8,89
Batch-size = 36	115,26	15,91

Tabulka 4: Průměrný čas epochy a vytížení GPU při rozdílných hodnotách hyperparametru  $Batch\ size$

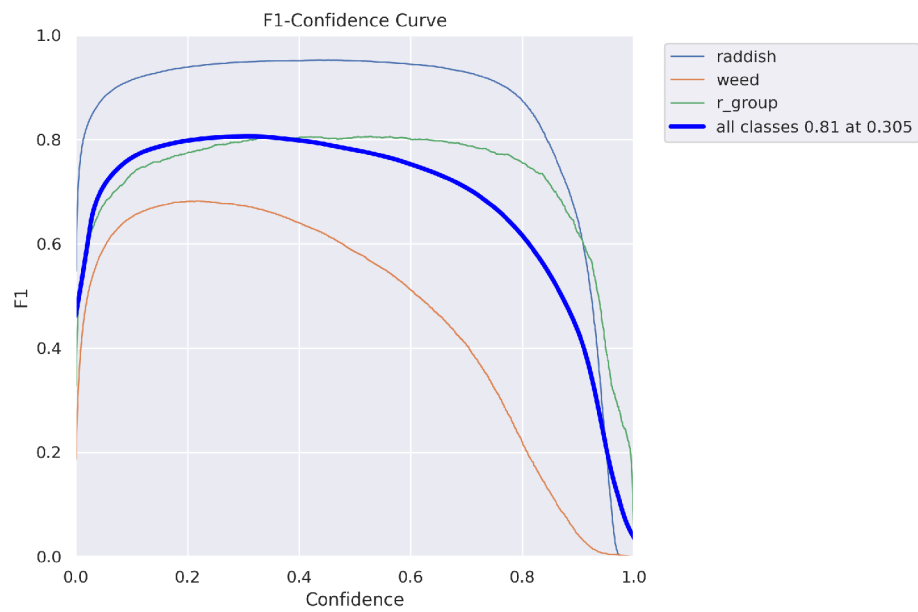
### 8.4.3 F1-Curve

Skóre F1 ukazuje, že spolehlivost modelu je přímo úměrná hodnotě hyperparametru *Batch size*. V případě prvního nastavení hyperparametrů (*Batch size* = 3) je skóre F1 v porovnání v ostatními případy nízké. Proto lze usoudit, že nízká hodnota parametru *Batch size* není optimální.

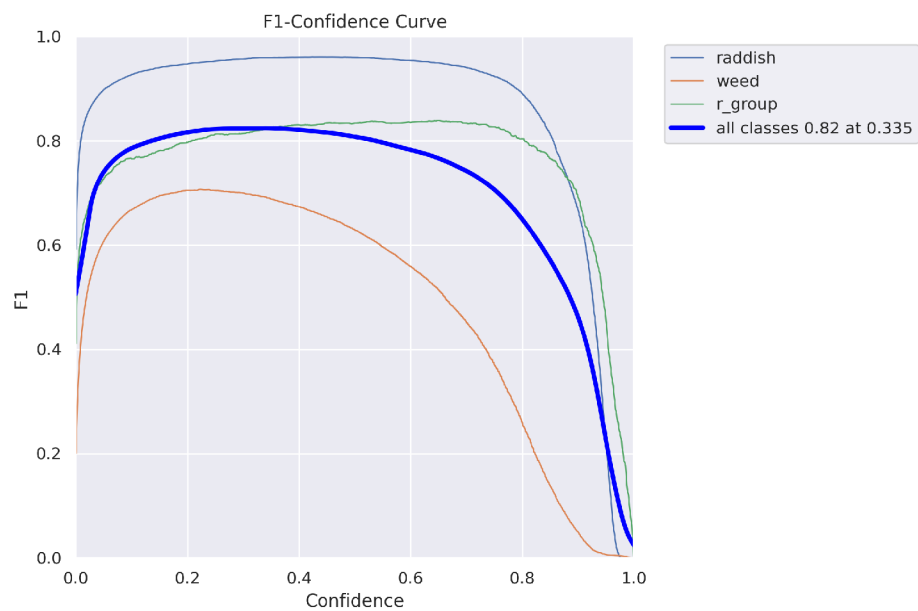
Nejpříznivějších výsledků dosahuje soudě dle této metriky model s nejvyšší hodnotou *Batch size*. Zohlednit ovšem musíme i další metriky, dle kterých jsou výsledky modelů analyzovány. Vysoká hodnota tohoto hyperparametru znamená také vysoké vytížení GPU (viz Tabulka 4).



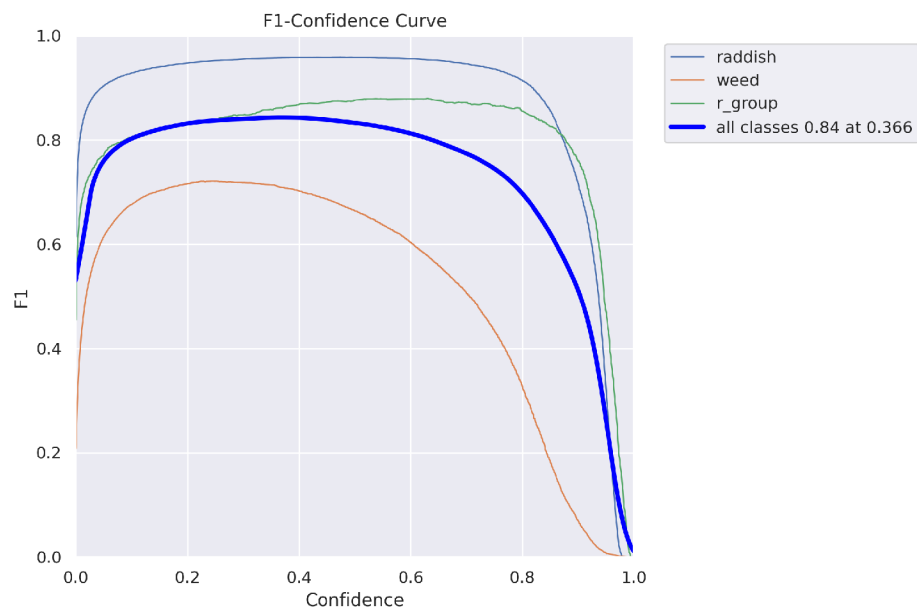
Obrázek 64: F1 Curve; *Learning rate* = 0,01; *Batch size* = 3



Obrázek 65: F1 Curve; *Learning rate = 0,01; Batch size = 6*



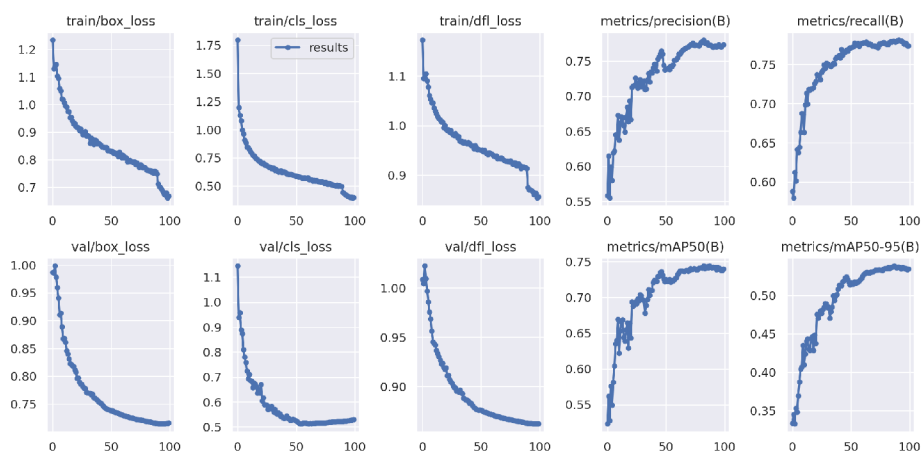
Obrázek 66: F1 Curve; *Learning rate = 0,01; Batch size = 12*



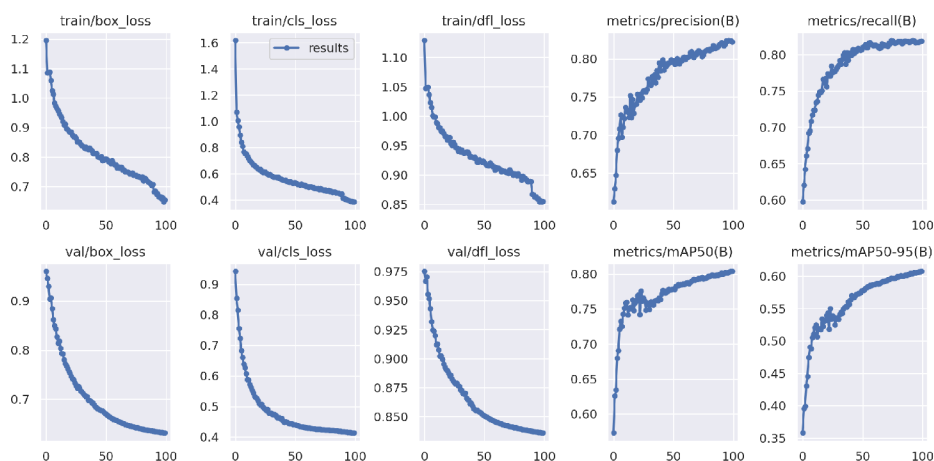
Obrázek 67: F1 Curve;  $Learning\ rate = 0,01$ ;  $Batch\ size = 36$

#### 8.4.4 Výsledky průběhu trénování modelu YOLOv8n

Příznivé výsledky při nejvyšší hodnotě  $Batch\ size$  jsou pozorovatelné i v následujících grafech. V prvním případě, kdy  $Batch\ size = 3$ , průměrná přesnost modelu zpočátku ztelně kolísá a během trénování posledních epoch se dokonce snižuje.

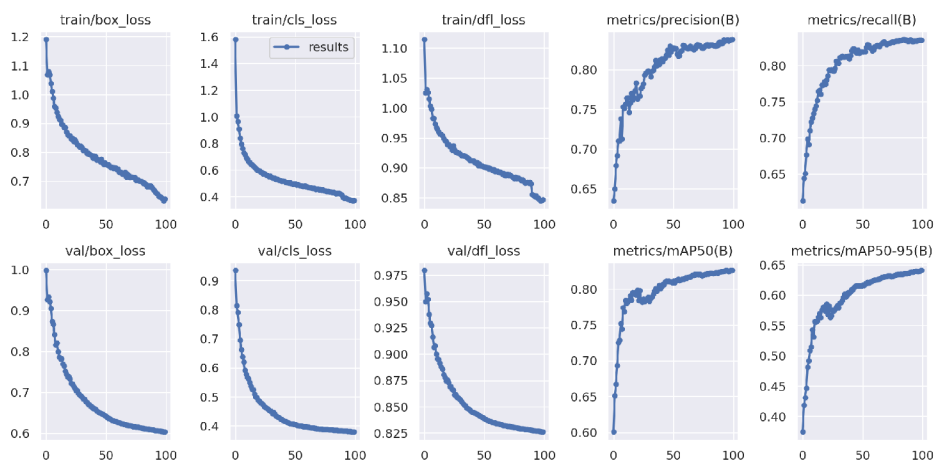


Obrázek 68: Průběh trénování;  $Learning\ rate = 0,01$ ;  $Batch\ size = 3$

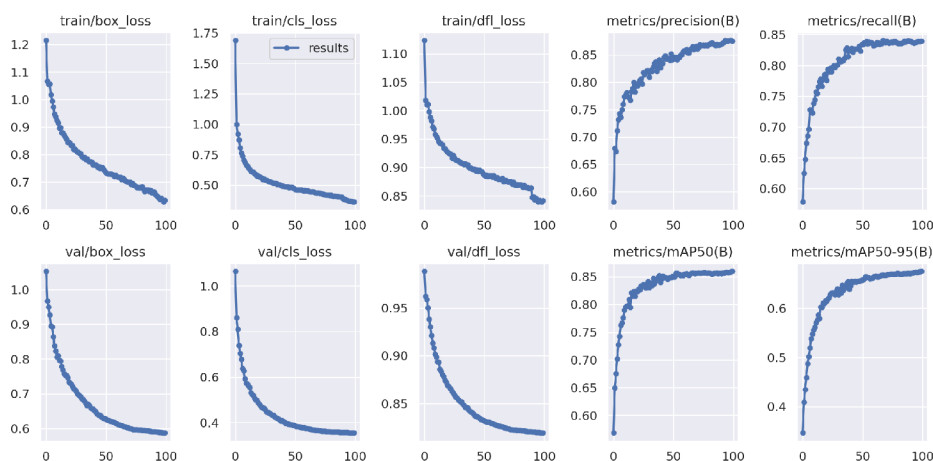


Obrázek 69: Průběh trénování;  $Learning\ rate = 0,01$ ;  $Batch\ size = 6$

Během třetího trénování při hodnotě  $Batch\ size = 12$  je zřejmé postupné ustálení průměrné přesnosti modelu. Počáteční hodnoty ztrát nejsou změnou  $Batch\ size$  výrazně ovlivněny. Model s nejvyšší hodnotou  $Batch\ size$  prokazuje vysokou a stabilní průměrnou přesnost i na základě výsledků průběhu trénování.



Obrázek 70: Průběh trénování;  $Learning\ rate = 0,01$ ;  $Batch\ size = 12$



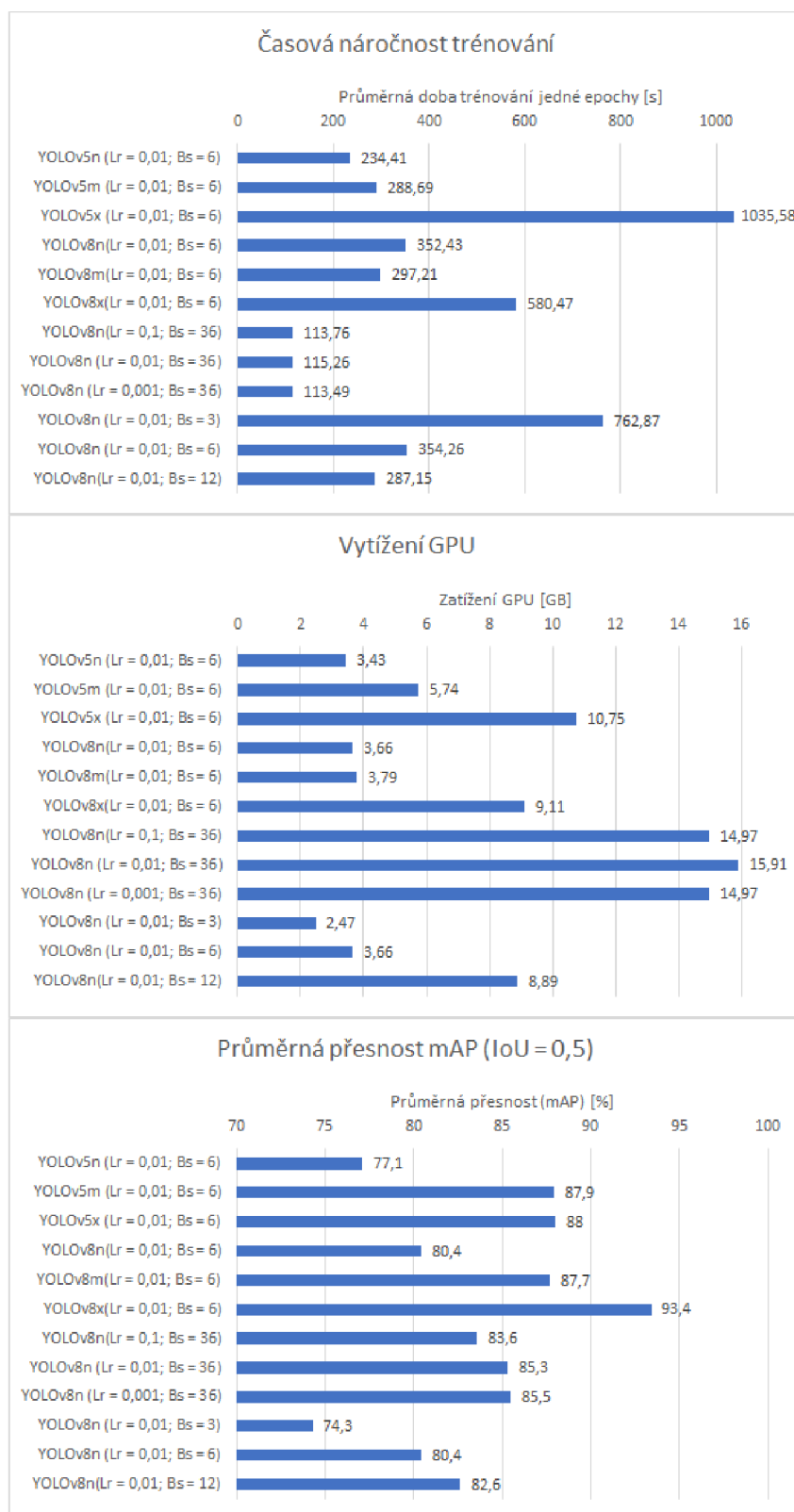
Obrázek 71: Průběh trénování;  $Learning\ rate = 0,01$ ;  $Batch\ size = 36$

## 8.5 Diskuze

V práci jsou porovnány výsledky tří různých verzí dvou modelů architektury YOLO (YOLOv5 a YOLOv8). Na základě vyhodnocení výsledků podle zvolených metrik lze jako nejpřesnější model z prvních tří sledovaných uvést největší z nich, a to YOLOv5x. To s sebou ovšem nese vyšší časovou náročnost a větší vytížení GPU. V případě druhého porovnávání to byl model YOLOv8x. V závěru vedle sebe lze postavit i tyto dva modely. Největší rozdíly jsou patrné u průměrné přesnosti, která je u YOLOv8x vyšší. Tuto verzi je tedy možné na základě získaných výsledků označit jako model s největší přesností vyhodnocení ze šesti trénovaných verzí. Dále je patrné, že trénování na novějším modelu probíhalo rychleji a s nižším vytížením GPU, což je bezpochyby výhodnější.

Po sledování vlivu hyperparametrů lze uvést, že čas trénování ani vytížení GPU nejsou výrazně ovlivněné změnou  $Learning\ rate$ . Velikost  $Batch\ size$  ovlivňuje čas trénování, a to tak, že se zvyšující se hodnotou tohoto hyperparametru doba trénování klesá. Zatížení GPU se zvyšuje. Pro ideální nastavení parametrů je vhodné hledat kompromis na základě těchto atributů, a také zohlednit přesnost modelu. Z toho mimo jiné vyplývá fakt, že nízká hodnota  $Batch\ size$  není vhodná. Při hodnotě  $Batch\ size = 3$  lze pozorovat malé vytížení GPU, nicméně doba trénování je vysoká a přesnost je nízká.





Obrázek 72: Grafy pro porovnání všech modelů dle uvedených aspektů

## 9 Závěr

V případě porovnání různých verzí modelů neuronové sítě vyhodocoval s nejnižší spolehlivostí nejmenší a nejrychlejší z modelů. Je zřejmé, že vhodné nastavení parametrů ovlivní samotné trénování sítě i následnou spolehlivost modelu během detekce či klasifikace. Dalším z faktorů, který zásadně ovlivňuje celkový výstup neuronové sítě, je zpracování dat. V oblasti anotace může dojít k nesprávnému postupu z pozice uživatele, který data zpracovává. Model se učí na základě dat, která dostane. Chybně anotovaná nebo nevhodně zpracovaná data mohou snížit celkovou přesnost modelu.

Neuronové sítě se jeví jako vhodné pro detekci plodin. Problémové je rozpoznání plevele, což může být způsobeno různorodostí instancí této třídy a jejich malou velikostí.

## Seznam literatury

AGARWAL, Saloni, Rami HALLAC, Rashika MISHRA, Chao LI, Ovidiu DAESCU a Alex KANE. *Image Based Detection of Craniofacial Abnormalities using Feature Extraction by Classical Convolutional Neural Network*. 2018. 10.1109/ICCABS.2018.8541948.

AMJOUR BENALI, Ayoub a Mustapha AMROUCH. *Convolutional Neural Networks Backbones for Object Detection*. Marrakesh, Morocco, 2020. PMID: PMC7340949.

*Best practices for successful image annotation* [online]. [cit. 2022-11-29]. Dostupné z: <https://labelbox.com/image-annotation-overview>

BOESCH, Gaudenz. *YOLOv7: The Most Powerful Object Detection Algorithm (2023 Guide)* [online]. Dostupné z: <https://viso.ai/deep-learning/yolov7-guide/>

BROWNLEE, Jason. *A Gentle Introduction to Object Recognition With Deep Learning* [online]. 2019 [cit. 2023-02-20]. Dostupné z: <https://machinelearningmastery.com/object-recognition-with-deep-learning/>

BROWNLEE, Jason. *Difference Between a Batch and an Epoch in a Neural Network* [online]. 2022 [cit. 2023-02-20]. Dostupné z: <https://machinelearningmastery.com/difference-between-a-batch-and-an-epoch/>

*CMOS image sensor*. In: [wikipedia.org](https://en.wikipedia.org/wiki/Active-pixel_sensor#/media/File:Matrixw.jpg) [online]. 18. března 2009 [cit. 2022-05-02]. Dostupné z: [https://en.wikipedia.org/wiki/Active-pixel\\_sensor#/media/File:Matrixw.jpg](https://en.wikipedia.org/wiki/Active-pixel_sensor#/media/File:Matrixw.jpg)

DOBEŠ, Michal. *Zpracování obrazu a algoritmy v C*. Praha: BEN - technická literatura, 2008. ISBN 978-80-7300-233-6.

DONG, Shi, Ping WANG a Khusnhood ABBAS. *A survey on deep learning and its applications*. School of Computer Science and Technology, Zhokou Normal University, China, 2021.

DOSHI, Sanket. *Various Optimization Algorithms For Training Neural Network* [online]. 2019 [cit. 2023-03-10].

Dostupné z: <https://towardsdatascience.com/optimizers-for-training-neural-network-59450d71caf6>

FRANK, Horst. *CMYK*. In: *commons.wikimedia.org* [online]. 22. února 2008 [cit. 2022-04-30]. Dostupné z: [https://commons.wikimedia.org/wiki/File:CMYK\\_farbwuerfel.svg](https://commons.wikimedia.org/wiki/File:CMYK_farbwuerfel.svg)

FRANKLIN, Jerrold. *Classical electromagnetism*. Second edition. Mineola, New York: Dover Publications, 2017. ISBN 978-0-486-81371-4.

FREEMAN, Michael. *Průvodce světem digitální fotografie*. Praha: Svojtka, 2004. ISBN 80-7237-962-3.

GEOFF, Dougherty. *Digital Image Processing for Medical Applications*. Cambridge: Cambridge University Press, 2009. ISBN 978-0-521-86085-7.

GONZALEZ, Rafael C. a Richard E. WOODS. *Digital image processing*. Reading, Mass.: Addison-Wesley, 1992. ISBN 0201508036.

GOODFELLOW, Ian, Yoshua BENGIO, and Aaron COURVILLE. *Deep learning: The MIT Press*, 2016. ISBN: 0262035618.

GU, Wenchao, Shuang BAI a Lingxing KONG. *A review on 2D instance segmentation based on deep neural networks*. Image and Vision Computing, Volume 120, 2022.

HÁJOVSKÝ, Radovan, Radka PUSTKOVÁ a František KUTÁLEK. *Zpracování obrazu v měřicí a řídicí technice: učební text : studijní materiály pro studijní obor Měřicí a řídicí technika, Elektronika Fakulty elektrotechniky a informatiky*. Ostrava: Vysoká škola báňská - Technická univerzita, 2012. ISBN 978-80-248-2596-0.

HALLIDAY, David, Robert RESNICK a Jearl WALKER, DUB, Petr, ed. *Fyzika*. 2., přeprac. vyd. Přeložil Miroslav ČERNÝ. Brno: VUTIUM, c2013. Překlady vysokoškolských učebnic. ISBN 978-80-214-4123-1.

HELMERS Heinz a Markus SCHELLENBERG. *CMOS vs. CCD sensors in speckle interferometry*. Carl von Ossietzky Universit at Oldenbur, 2002.

HLAVÁČ, Václav a Miloš SSEDLÁČEK. *Zpracování signálů a obrazů*. Vyd. 2., přeprac. Praha: Vydavatelství ČVUT, 2005. ISBN 80-01-03110-1.

HOZMAN, J., BERNAS, M., KLÍMA, M. a DVOŘÁK, P. *Zpracování obrazové informace*. (Skriptum). České vysoké učení technické v Praze. Vydavatelství ČVUT: Praha, 1996.

*Instance vs. Semantic Segmentation: What Are the Key Differences?* [online]. 2021 [cit. 2023-04-01]. Dostupné z: <https://keymakr.com/blog/instance-vs-semantic-segmentation/>

*Jak funguje fotoaparát: čas – clona – ISO* [online]. 2020 [cit. 2022-06-12]. Dostupné z: <https://www.fotolab.cz/blog/jak-funguje-fotoaparát/>

JIU Mingyuan Jiu a Hichem SAHBI. *Context-aware deep kernel networks for image annotation*. Neurocomputing, 2022.

KHANDAY, Nadeem Yousuf a Shabir Ahmad SOFI. *Taxonomy, state-of-the-art, challenges and applications od visual understanding: A review*. National Institute of Technology, Srinagar, India, 2021.

KOECH, Kiprono Elijah. *Object Detection Metrics With Worked Example* [online]. 2020 [cit. 2023-3-30] Dostupné z: <https://towardsdatascience.com/object-detection-metrics-with-worked-example-216f173ed31e>

KRÁLOVÁ, Magda. *Comptonův jev* [online]. [cit. 2022-05-10]. Dostupné z : <http://edu.techmania.cz/cs/encyklopedie/fyzika/kvanta/comptonuv-jev>

KRÁLOVÁ, Magda. *Fotoelektrický jev* [online]. [cit. 2022-05-10]. Dostupné z: <http://edu.techmania.cz/cs/encyklopedie/fyzika/kvanta/fotoelektricky-jev>

KULHÁNEK, Petr. *Vybrané kapitoly z teoretické fyziky*. Vydání: druhé, aktualizované a změněné. Praha: AGA, 2020. ISBN 978-80-906638-2-4.

*Light-matter interaction diagrams*. In: *commons.wikimedia.org* [online]. 19. června 2020 [cit. 2022-04-25]. Dostupné z: [https://commons.wikimedia.org/wiki/File:Light-matter\\_interaction\\_-\\_schematic.svg](https://commons.wikimedia.org/wiki/File:Light-matter_interaction_-_schematic.svg)

LIU Haiqiang, Meibao YAO, Xueming XIAO a Hutao CUI. *A hybrid attention semantic segmentation network for unstructured terrain on Mars*. *Acta Astronautica*, 2023.

MARTINÁSKOVÁ, Hana. *Vývoj představ o světle a možnosti jeho využití v gymnaziálním kurzu fyziky* [online]. Brno, 2007 [cit. 2022-06-18]. Dostupné z: <https://is.muni.cz/th/jwps8/>. Diplomová práce. Masarykova univerzita, Přírodovědecká fakulta. Vedoucí práce Aleš LACINA.

MAXWELL, Aaron E., Timothy A. WARNER a Luis Andrés GUILLÉN. *Accuracy Assessment in Convolutional Neural Network-Based Deep Learning Remote Sensing Studies—Part 1: Literature Review* [online]. [cit. 2023-01-25]. ISSN 2072-4292. Dostupné z: [doi:10.3390/rs13132450](https://doi.org/10.3390/rs13132450)

MINKOVÁ, Markéta. *Digitální fotografie pro pokročilé*. Čestlice: Rebo, 2006. ISBN 80-7234-506-0.

MISHRA, Aditya. *Metrics to Evaluate your Machine Learning Algorithm* [online]. 2018 [cit. 2023-03-15]. Dostupné z: <https://towardsdatascience.com/metrics-to-evaluate-your-machine-learning-algorithm-f10ba6e38234>

MURPHY, Kevin P. *Machine Learning: A Probabilistic Perspective*. Cambridge: MIT Press, 2012. ISBN 978-0-262-01802-9.

NAGPAL, Shaveta, Munish KUMAR, MARUTHI, Rohit AYYAGARI. *A Survey of Deep Learning and Its Applications: A New Paradigm to Machine Learning*. Archives of Computational Methods in Engineering, 2020.

NATH, Nipun D. a Amir H. Behzadan. *Deep Convolutional Networks for Construction Object Detection Under Different Visual Conditions*. Frontiers in Built Environment, 2020.

NOVOTNÝ, Karel. *Elektromagnetické pole a vlny : teorie elektromagnetického pole II.*. Praha: Vydavatelství ČVUT, 2001. 168 s. ISBN 80-01-02429-6.

PAVELKA, Karel. *Zpracování obrazových záznamů DPZ*. Praha: Vydavatelství ČVUT, 1999. ISBN 80-01-02031-2.

PETROU Maria a BOSNOGIANNI P. *Image Processing – The Fundamentals*. John Wiley Sons, New York, 1999. ISBN 0-471-99883-4.

*Photoelectric effect in a solid: ultraviolet light ejects electrons from a crystal*. In: *commons.wikimedia.org* [online]. 28. července 2020 [cit. 2022-04-25]. Dostupné z: [https://commons.wikimedia.org/wiki/File:Photoelectric\\_effect\\_in\\_a\\_solid\\_-\\_diagram.svg](https://commons.wikimedia.org/wiki/File:Photoelectric_effect_in_a_solid_-_diagram.svg)

POWERS, David. *Evaluation: From Precision, Recall and F-Measure to ROC, Informedness, Markedness Correlation*. Journal of Machine Learning Technologies, 2008.

RATH, Sovit. *YOLOv5 – Custom Object Detection Training* [online]. 2022 [cit. 2023-04-01]. Dostupné z: <https://learnopencv.com/custom-object-detection-training-using-yolov5/>

*RGB color cube*. In: *commons.wikimedia.org* [online]. 6. ledna 2015 [cit. 2022-04-30]. Dostupné z: [https://commons.wikimedia.org/wiki/File:RGB\\_color\\_cube.svg](https://commons.wikimedia.org/wiki/File:RGB_color_cube.svg)

SHAH, Deval. *The Essential Guide to Data Augmentation in Deep Learning* [online]. 2023 [cit. 2023-03-30]. Dostupné z: <https://www.v7labs.com/blog/data-augmentation-guide>

SHAH, Deval. *YOLOv4 — Version 1: Bag of Freebies. An Introductory Guide on the Fundamentals and Algorithmic Flows of YOLOv4 Object Detector* [online]. 2020 [cit. 2023-03-29]. Dostupné z: <https://medium.com/visionwizard/yolov4-bag-of-freebies-dc126623fc2d>

SIDDEL, Kathleen. *What YOLO Model Architecture Is Best for You?* [online]. 2022 [cit. 2023-03-15]. Dostupné z: <https://alwaysai.co/blog/best-yolo-model-for-you>

SMITH, W. Steven. *Digital signal processing: scientist and engineer's guide*. California: California Technical Publishing, 1997. ISBN 978-0966017632.

SOLAWETZ, Jacob. *How to Train A Custom Object Detection Model with YOLO v5* [online]. 2020 [cit. 2023-03-29]. Dostupné z: <https://towardsdatascience.com/how-to-train-a-custom-object-detection-model-with-yolo-v5-917e9ce13208>

*Světlo: Elektromagnetické vlnění, spektrum a fotony* [online]. [cit. 2022-05-27]. Dostupné z: <https://cs.khanacademy.org/science/obecna-chemie/>

SVOBODA, Emanuel, et al. *Přehled středoškolské fyziky*. 4. vydání. Praha : Prometheus, 2008. ISBN 9788071963073.



ŠŤASTNÝ, Jiří. *Netradiční metody a algoritmy pro rozpoznávání objektů technologické scény: Nontraditional methods and algorithms for object re-cognition of technological scene: zkrácená verze habilitační práce*. Brno: VUTIUM, 2006. ISBN 80-214-3117-2.

TAY, David B. *Sensor network data denoising via recursive graph median filters*. Signal Processing, 2021.

VO, Kevin et al. *Electromagnetic Radiation* [online]. [cit. 2022-06-02]. Dostupné z: <https://chem.libretexts.org>

VRBOVÁ, Miroslava, Helena JELÍNKOVÁ a Petr GAVRILOV. *Úvod do laserové techniky*. Dot. 1. vyd. Praha: České vysoké učení technické, 1998. ISBN 80-01-01108-9.

V7's Blog - Computer vision [online]. [cit. 2022-06-28]. Dostupné z: <https://www.v7labs.com/blog>

WALLSTROM, Conny, *What Color Depth You Should Use And Why It Matters* [online]. 2015 [cit. 2022-06-20]. Dostupné z: <http://www.diyphotography.net/8-bit-vs-16-bit-color-depth-use-matters/>

XIE, Shuang. BJT Induced Dark Current in CMOS Image Sensors. School of Microelectronics, Shandong University, China, 2022.

XU, Mingle, Sook YOON, Alvaro FUENTES a Dong Sun PARK. *A Comprehensive Survey of Image Augmentation Techniques for Deep Learning*. Pattern Recognition, 2023.

YAN, Kun a Zhang ZHIUA. *Automated Asphalt Highway Pavement Crack Detection Based on Deformable Single Shot Multi-Box Detector Under a Complex Environment*. IEEE Access. PP. 1-1. 10.1109/ACCESS.2021.3125703.

YUN, Sehyun, Renato ZANETTI a Brandon A. JONES. *Kernel-based ensemble gaussian mixture filtering for orbit determination with sparse data*. Advances in Space Research, 2022.

ZMEŠKAL, Oldřich, Martin NEŽÁDAL a Ondrej SEDLAK. *The usage of fractal analysis for evaluating the quality of print*. 2001.