



Středoškolská technika 2024

Setkání a prezentace prací středoškolských studentů na ČVUT

Automatické rozpoznávání SPZ

Jiří Klimeš

SPŠE a VOŠ Pardubice
Karla IV. 13, 530 02 Pardubice I

Anotace

V posledních letech pokroky v počítačovém vidění a nositelné technologii usnadnily inovativní aplikace v různých oblastech. Tato práce představuje nový přístup k automatickému rozpoznávání registračních značek v reálném čase pomocí chytrých brýlí integrovaných s malou kamerou. Inovace spočívá ve využití jedné konvoluční neuronové sítě (CNN) pro současnou detekci a optické rozpoznávání znaků (OCR) registračních značek, což eliminuje potřebu více inferenčních kroků a výrazně snižuje dobu zpracování. Využitím jednofázové predikce dosahuje systém vysokorychlostního zpracování, 30 snímků za sekundu (FPS) bez grafického akcelérátoru. Po detekci umožňuje bezproblémovou integraci s databází okamžité získání relevantních informací a poskytuje včasnou zpětnou vazbu prostřednictvím vyskakovacích oznámení zobrazených na chytrých brýlích. Toto optimalizované řešení ukazuje praktické aplikace v řízení dopravy, bezpečnosti a dalších oblastech a představuje potenciál integrovaných systémů pro rozpoznávání registračních značek v reálných scénářích a slibné pokroky v efektivitě a účinnosti.

Klíčová slova: počítačové vidění, nositelná technologie, automatické rozpoznávání registračních značek, ANPR, chytré brýle, konvoluční neuronová síť, CNN, detekce, optické rozpoznávání znaků OCR, jednorázové předpovědi, v reálném čase, doba zpracování, hardware s omezenými prostředky, integrace s databází, vyskakovací oznámení, řízení dopravy, bezpečnost, efektivita, účinnost.

Annotation

In recent years, advancements in computer vision and wearable technology have facilitated innovative applications across various domains. This paper introduces a novel approach to real-time automatic number plate recognition (ANPR) using smart glasses integrated with a small camera. The innovation lies in utilizing a single Convolutional Neural Network (CNN) for simultaneous detection and optical character recognition (OCR) of number plates, eliminating the need for multiple inference steps and significantly reducing processing time. Leveraging single-shot predictions, our system achieves high-speed processing, 30 frames per second (FPS) without a graphical accelerator. Upon detection, seamless integration with a database enables instant retrieval of relevant information, providing timely feedback through toast notifications displayed on the smart glasses. This streamlined solution demonstrates practical applications in traffic management, security, and beyond, showcasing the potential of integrated ANPR systems in real-world scenarios and promising advancements in efficiency and effectiveness.

Keywords: computer vision, wearable technology, automatic number plate recognition ANPR, smart glasses, Convolutional Neural Network, CNN, detection, optical character recognition, OCR, single-shot predictions, real-time, processing time, resource-constrained hardware, database integration, toast notifications, traffic management, security, efficiency, effectiveness

Obsah

1 Úvod	5
2 ANALÝZA	5
2.1 LOKALIZACE SPZ	6
2.2 ČTENÍ SPZ	14
3 NÁVRH PROJEKTU	15
3.1 HARDWARE	15
3.2 SOFTWAREVÉ TECHNOLOGIE A VÝVOJOVÉ NÁSTROJE	16
3.3 VÝBĚR ARCHITEKTURY	17
3.4 INTEGRACE DETEKCE A OCR DO JEDNOHO MODELU	18
3.5 PŘÍPRAVA DAT PRO TRÉNINK	19
3.6 POUŽITÍ TENSORRT	20
3.7 ZLEPŠENÍ PŘESNOSTI DETEKCE	21
4 ZÁVĚR	22
4.1 SHRUTÍ PROJEKTU	22
4.2 VÝZNAM A PŘÍNOSY	23
4.3 BUDOUCÍ ROZVOJ A PLÁNY	24
SEZNAM OBRÁZKŮ	26
SEZNAM POUŽITÉ LITERATURY A ZDROJŮ OBRÁZKŮ	27

1 Úvod

V současné době se technologie automatického rozpoznávání registračních značek vozidel (ANPR) stává stále důležitější pro širokou škálu aplikací, od bezpečnostních systémů po inteligentní parkovací služby. Vývojáři se proto neustále snaží inovovat, aby zlepšili rychlost, přesnost a efektivitu těchto systémů. Jednou z hlavních výzev je schopnost provádět detekci a rozpoznávání SPZ v reálném čase, což je zásadní pro jejich aplikaci v chytrých brýlích a dalších zařízeních.

V rámci tohoto projektu se plánuje představit inovativní postup, který odstraní potřebu samostatného OCR modelu pro rozpoznávání textu na SPZ. Místo toho se upraví architektura YOLOv8 tak, aby byla schopna provádět obě úlohy, detekci SPZ a následné rozpoznávání textu v jednom kroku. Tento přístup by měl výrazně zrychlit celkový proces a zredukovat výpočetní nároky systému, což umožní jeho efektivní běh i na hardwaru s nižším výkonem.

Další klíčovou inovací projektu bude využití Kalmanova filtru pro sledování pohybu SPZ mezi jednotlivými snímky. Tato metoda by měla umožnit systému efektivně sledovat a identifikovat SPZ i v dynamickém prostředí a minimalizovat chyby v detekci způsobené pohybem nebo změnami v osvětlení.

Tento inovativní přístup, který kombinuje upravenou architekturu YOLOv8 s aplikací Kalmanova filtru, by měl znamenat významný pokrok v technologii ANPR. Díky tomu bude možné algoritmus spustit v reálném čase i na hardwaru s omezenými výpočetními zdroji. Představte si, že s chytrými brýlemi XREAL Air za pouhých 8 tisíc korun, webkamerou za 500 korun a grafickým akcelerátorem s výkonem 2 TFLOPS získáte plně funkční řešení. Toto řešení zvládne bez problémů běžet v reálném čase a může být využito například pro řízení vjezdu do parkoviště.

2 Analýza

Problém provádění ANPR na chytrých brýlích lze rozdělit na několik částí. Chytré brýle sloužící jako kamera a obrazovka, kamera vytváří videozáznam a detekce může být prováděna na jednotlivých snímcích. Tento proces můžeme rozdělit na detekci celých SPZ, jejich oříznutí a provádění OCR na vybraných oblastech. Tyto kroky pojmenujeme lokalizace a čtení. Poslední částí je zobrazení uživatelského rozhraní na obrazovce.

2.1 Lokalizace SPZ

Lokalizace SPZ, neboli detekce registračních značek vozidel, je klíčovým krokem v procesu ANPR. Zahrnuje analýzu obrazových dat s cílem identifikovat a ohraničit oblasti obsahující registrační značky na snímcích vozidel. Výsledkem lokalizace jsou tzv. ohraničující rámečky (bbox), což jsou obdélníkové rámečky, které přesně vymezují polohu SPZ na obrázku.

2.1.1 Tradiční počítačové vidění

Detekce hran a obdélníků jsou základní techniky v počítačovém vidění pro identifikaci SPZ, využívající analýzu obrazu pro rozpoznání hran a geometrických tvarů. Tyto metody jsou jednoduché a vyžadují málo výpočetních zdrojů, což je činí vhodnými pro rychlou lokalizaci SPZ za předpokladu dobré viditelnosti a osvětlení.

Hlavní nevýhody zahrnují závislost na světelných podmínkách, což může vést k problémům při detekci v náročném osvětlení, a obtíže s detekcí SPZ pod úhlem kvůli deformaci a perspektivnímu zkreslení.

Optimalizace může zahrnovat vylepšení algoritmů pro lepší odolnost vůči různým světelným podmínkám a zpracování obrazu s perspektivním zkreslením. Navzdory efektivitě v ideálních podmínkách, omezení v náročných situacích naznačují potřebu integrace s pokročilejšími metodami.

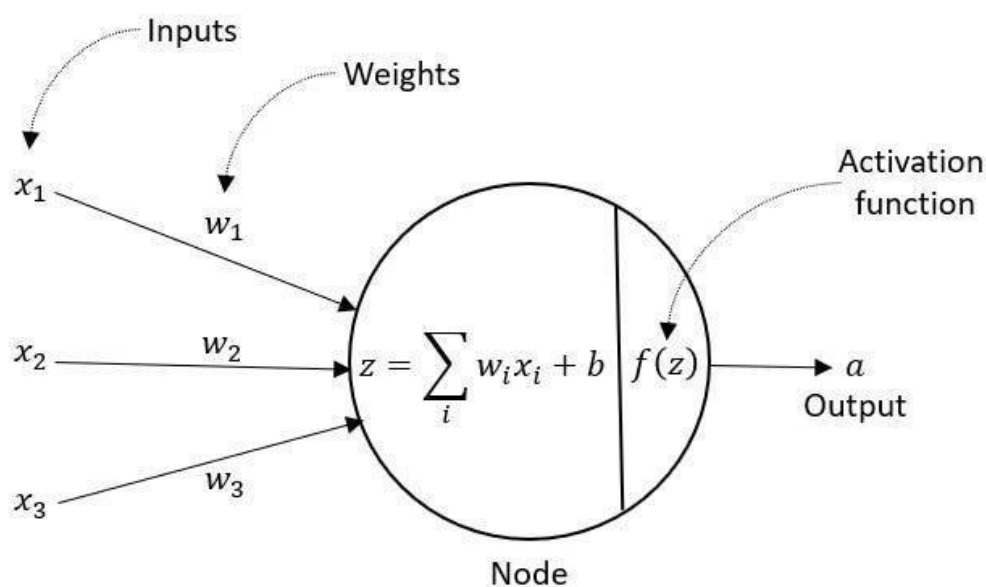
2.1.2 Párování šablon

Použití šablon pro detekci SPZ je metoda založená na porovnávání získaných obrazů s předem definovanou sadou šablon SPZ. Tento přístup je obzvláště účinný pro identifikaci a lokalizaci SPZ, které mají specifické tvary nebo velikosti odpovídající určitému regionu nebo typu vozidla. Jeho hlavní výhody spočívají ve zvýšené přesnosti detekce SPZ, které přesně odpovídají šablonám v knihovně, a v efektivitě omezení falešných pozitivů díky porovnání s přesnými šablonami.

Nicméně, metoda čelí významným výzvám. Jednou z hlavních nevýhod je nutnost udržovat rozsáhlou a aktuální knihovnu šablon, která musí pokrývat širokou škálu možných SPZ. To může být náročné na údržbu a aktualizaci, zejména v kontextu globalizace a neustálého přibývání nových typů SPZ. Dalším významným omezením je špatná výkonnost při částečném zakrytí nebo poškození SPZ. Pokud SPZ není plně viditelná nebo je poškozena tak, že se výrazně liší od šablon v knihovně, může být její správná detekce a rozpoznání obtížné.

2.1.3 Neuronové sítě

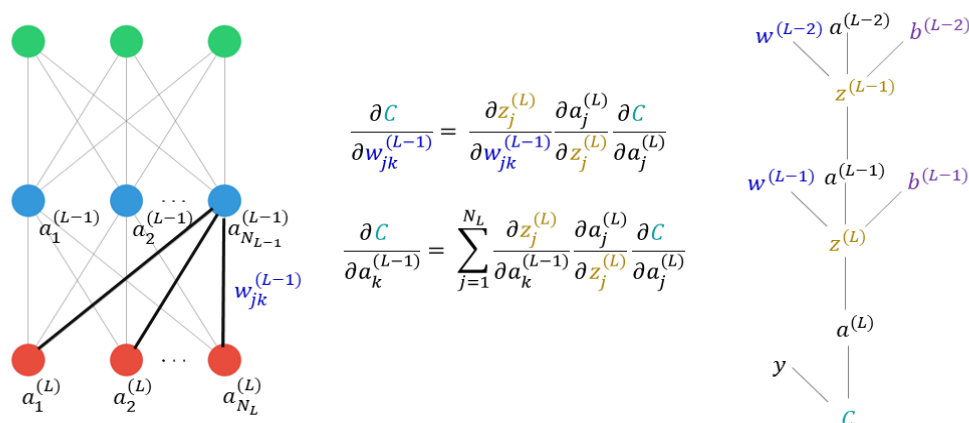
Neuronové sítě představují základní stavební kámen moderního strojového učení a umělé inteligence. Tyto výpočetní modely jsou inspirovány strukturou a funkcí lidského mozku a mají schopnost učit se z dat, identifikovat vzory a provádět složité úlohy, jako je rozpoznávání obrazu, zpracování přirozeného jazyka a mnoho dalších. V posledních letech došlo k významnému pokroku v oblasti neuronových sítí, což umožnilo dosáhnout průlomů v mnoha aplikacích.



Neuronové sítě se skládají z vrstev neuronů, kde každý neuron přijímá vstupy, provádí jednoduchý výpočet a generuje výstup, který je dále předáván do dalších vrstev sítě. Podle počtu vrstev a způsobu jejich propojení můžeme neuronové sítě rozdělit na nehluboké a hluboké. Zatímco nehluboké sítě mají obvykle jen několik vrstev, hluboké neuronové sítě (tzv. deep learning) obsahují mnoho vrstev, což jim umožňuje učit se složitější a abstraktnější reprezentace dat.

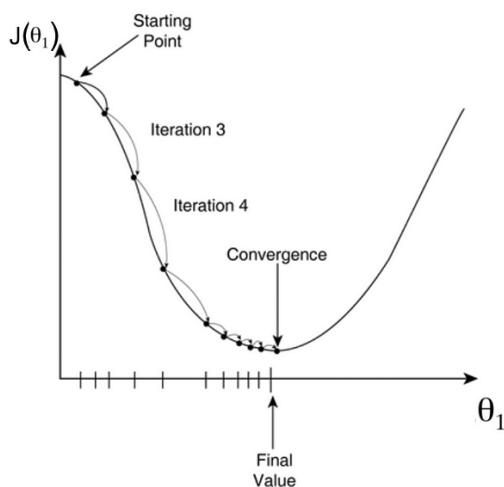
Vývoj a využití neuronových sítí v ANPR systémech přineslo značné výhody oproti tradičním metodám zpracování obrazu, zejména v přesnosti detekce a schopnosti adaptace na různé podmínky. Díky schopnosti neuronových sítí učit se z velkého množství dat a generalizovat naučené znalosti na nové, neviděné případy, se staly klíčovým nástrojem pro vývoj efektivních a spolehlivých ANPR systémů.

2.1.3.1 Učení neuronových sítí



Obrázek 2 – Zpětná propagace

Proces učení hlubokých neuronových sítí je založen na principu učení s učitelem, kde model využívá anotovaná data k naučení se správných odpovědí. Každý vzorek v tréninkové sadě obsahuje vstupní data a odpovídající "správnou" odpověď, což modelu umožňuje se "učit" tím, že porovnává své predikce s těmito správnými odpověďmi a upravuje své interní váhy, aby minimalizoval rozdíly.



Cost Function – "One Half Mean Squared Error":

$$J(\theta_0, \theta_1) = \frac{1}{2m} \sum_{i=1}^m (h_{\theta}(x^{(i)}) - y^{(i)})^2$$

Objective:

$$\min_{\theta_0, \theta_1} J(\theta_0, \theta_1)$$

Derivatives:

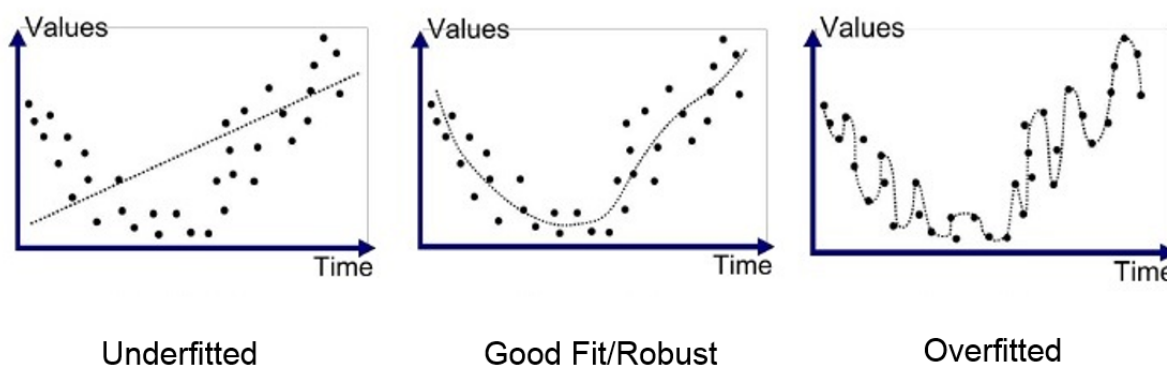
$$\frac{\partial}{\partial \theta_0} J(\theta_0, \theta_1) = \frac{1}{m} \sum_{i=1}^m (h_{\theta}(x^{(i)}) - y^{(i)})$$

$$\frac{\partial}{\partial \theta_1} J(\theta_0, \theta_1) = \frac{1}{m} \sum_{i=1}^m (h_{\theta}(x^{(i)}) - y^{(i)}) \cdot x^{(i)}$$

Základem učení je algoritmus zpětné propagace spolu s metodou gradientního sestupu. Zpětná propagace je proces, při kterém se chyba (rozdíl mezi predikovanou a skutečnou hodnotou) propaguje zpět skrze síť, což umožňuje modelu pochopit, jak jednotlivé neurony ovlivňují celkovou chybu. Gradientní sestup je iterativní optimalizační algoritmus, který postupně upravuje váhy v neuronové síti s cílem minimalizovat chybu mezi predikovanými a skutečnými hodnotami. V průběhu učení algoritmus vypočítává gradient chybové funkce vzhledem k váhám a iterativně upravuje tyto váhy v opačném směru gradientu, což vede k postupnému snižování celkové chyby. Váhy v neuronové síti, které určují sílu spojení mezi neurony, jsou klíčové pro určení výstupu sítě a jejich optimalizace je zásadní pro úspěšné

učení

modelu.



Obrázek 4 - Přeučení

Během učení je klíčové řešit problémy jako přeučení (overfitting), kdy se model příliš přizpůsobí tréninkovým datům a ztrácí schopnost generalizovat na nová data. K řešení tohoto problému se využívají techniky regularizace, jako je dropout, který náhodně "vypíná" některé neurony během tréninku, čímž zabráňuje modelu stát se příliš závislým na konkrétních vzorcích v tréninkových datech. Další technikou je L1/L2 regularizace, která přidává penalizaci k velikosti vah, což pomáhá udržovat model jednodušší a odolnější vůči přeučení.

Optimalizace hyperparametrů, jako je rychlost učení a velikost dávky, je dalším důležitým aspektem. Rychlost učení určuje, jak rychle model aktualizuje své váhy v reakci na chybu, zatímco velikost dávky ovlivňuje, kolik vzorků dat se zpracuje najednou před aktualizací vah. Správné nastavení těchto parametrů je klíčové pro efektivní a úspěšné učení.

Celý proces učení vyžaduje nejen technické znalosti o fungování neuronových sítí, ale také hluboké pochopení teorie strojového učení, aby bylo možné správně nastavit a optimalizovat model pro konkrétní úlohu. Pokroky v algoritmech a technikách trénování, spolu s rostoucí dostupností výpočetních zdrojů, umožňují vývoj stále pokročilejších a efektivnějších modelů, které mohou řešit stále složitější úlohy.

2.1.3.2 Potřeba hlubokých neuronových sítí

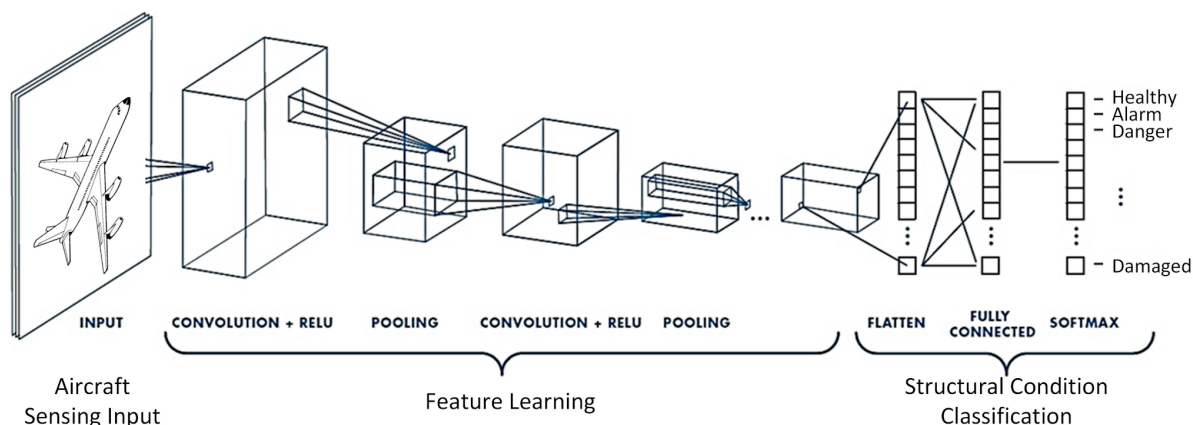
Pro účinné rozpoznávání složitých vzorů, jako je detekce registračních značek vozidel (SPZ), se ukázalo, že tradiční metody a jednoduché neuronové sítě s omezeným počtem vrstev jsou nedostatečné. Tyto přístupy selhávají kvůli své omezené schopnosti abstrakce a adaptace na variabilitu ve vizuálních datech. Na rozdíl od nich, hluboké neuronové sítě, které obsahují mnoho vrstev, poskytují výrazně lepší výsledky.

Hlavním důvodem, proč jedna nebo pouze několik vrstev nestačí, je, že složité úlohy rozpoznávání vyžadují extrakci a kombinaci různých úrovní abstrakce z vizuálních dat. Nízkoúrovňové rysy, jako jsou hrany a textury, které jednoduché sítě mohou identifikovat, jsou základní, ale pro rozpoznání komplexních objektů, jako jsou SPZ, nejsou samy o sobě dostatečné. Hluboké neuronové sítě překonávají tuto omezenost tím, že postupně kombinují tyto základní rysy ve vyšší úrovní abstrakce, jako jsou objekty a scény, což umožňuje efektivní rozpoznávání a klasifikaci i ve složitých a proměnlivých podmínkách.

Každá vrstva v hluboké neuronové síti se specializuje na detekci určitých rysů, přičemž počáteční vrstvy se zaměřují na jednoduché rysy a postupně, jak data procházejí sítí, se

kombinují do složitějších vzorů. Tento hierarchický přístup umožňuje hlubokým neuronovým sítím naučit se rozpoznávat složité vzory s vysokou přesností. Proto jedna nebo jen několik vrstev není schopna poskytnout dostatečnou úroveň abstrakce potřebnou pro účinné rozpoznávání složitých objektů, jako jsou SPZ, v různorodých vizuálních scénářích.

2.1.3.3 Konvoluční neuronové sítě (CNN)



Obrázek 5 - Konvoluční síť

Konvoluční neuronové sítě (CNN) představují významný pokrok v oblasti zpracování obrazu, neboť jsou optimalizovány pro efektivní zachycení prostorových a lokálních vztahů mezi pixely. Klíčovou operací v CNN je konvoluce, která aplikuje filtr na vstupní obraz, umožňující detekci specifických rysů, jako jsou hrany a textury. Filtr, charakterizovaný velikostí, krokem a počtem výstupních kanálů, se pohybuje přes obraz, vytvářející mapu rysů, která zachycuje klíčové informace o struktuře obrazu. Díky sdílení filtrů a hierarchické struktuře, CNN s menším počtem parametrů efektivně zpracovávají obrazy, přecházejí od jednoduchých rysů k rozpoznávání složitějších vzorů. Velikost filtru určuje, kolik pixelů je zpracováno najednou, zatímco stride určuje rychlost pohybu filtru přes obraz. Počet výstupních kanálů určuje, kolik různých typů rysů je síť schopna z obrazu extrahovat, což umožňuje hlubší abstrakci a efektivní rozpoznávání komplexních objektů, jako jsou SPZ, a zajišťuje jejich schopnost generalizace pro rozpoznávání v nových situacích.

Na rozdíl od konvolučních neuronových sítí, hustě propojené sítě, známé také jako plně propojené sítě, tvoří základ raných architektur neuronových sítí, kde každý neuron je propojen s každým neuronem v následující vrstvě. Ačkoliv jsou tyto sítě efektivní pro řešení jednoduchých úloh, při složitějších úlohách, jako je detekce SPZ, narazí na omezení spojená s neefektivním zpracováním obrazových dat a nedostatečnou schopností zachytit prostorové vztahy. Proto se pro úlohy zpracování obrazu a rozpoznávání objektů často upřednostňují CNN, které lépe zvládají zachycení a interpretaci složitých vzorů a struktur v obrazových datech.

2.1.3.4 Výhody použití většího počtu menších filtrů

Použití menších filtrů v konvolučních neuronových sítích (CNN) přináší zlepšení v efektivitě, přesnosti a generalizaci při zpracování obrazových dat. Menší filtry snižují počet výpočetních operací, což umožňuje rychlejší zpracování a snižuje nároky na výpočetní zdroje. To je klíčové pro zpracování velkých obrazových dat a umožňuje efektivní nasazení modelů v reálném čase a na zařízeních s omezenými kapacitami.

Menší filtry jsou také efektivnější v detekci jemných obrazových rysů, jako jsou textury a hrany. To umožňuje CNN lépe abstrahovat a rozpoznávat složité vzory a objekty. Díky tomu jsou sítě s menšími filtry schopné naučit se více generalizovatelných rysů z trénovacích dat, což zlepšuje adaptabilitu modelu na nové, neviděné obrazy.

2.1.3.5 Separabilní konvoluce

Bodová (pointwise) a hloubková (depthwise) konvoluce jsou klíčové techniky v moderních architekturách hlubokých konvolučních neuronových sítí (CNN), zvyšující efektivitu a zpracování obrazových dat. Hloubková konvoluce zpracovává každý kanál obrazu nezávisle, což snižuje počet výpočtů a parametrů. Tím se efektivně extrahují lokální rysy, jako jsou textury a hrany, s minimálními výpočetními nároky.

Bodová konvoluce, používající 1x1 filtry, kombinuje výstupy z hloubkové konvoluce, což umožňuje síti lépe kombinovat informace z různých kanálů a rozpoznávat složitější vzory. Tato metoda také efektivně redukuje dimenzionalitu a zvyšuje hloubku sítě bez výrazného nárůstu parametrů.

Kombinace obou technik, známá jako separabilní konvoluce, umožňuje efektivní a přesnou analýzu obrazových dat s menším počtem parametrů a nižšími výpočetními nároky. Tento přístup, porovnáván s tradičními konvolučními operacemi, může snížit počet potřebných výpočetních operací až o 8krát při zachování srovnatelného výkonu. Tím se stává ideální volbou pro aplikace v reálném čase a na zařízeních s omezenými zdroji, jako jsou mobilní telefony a vestavěné systémy, kde je důležitá vysoká efektivita při omezeném výkonu.

2.1.3.6 Downsampling

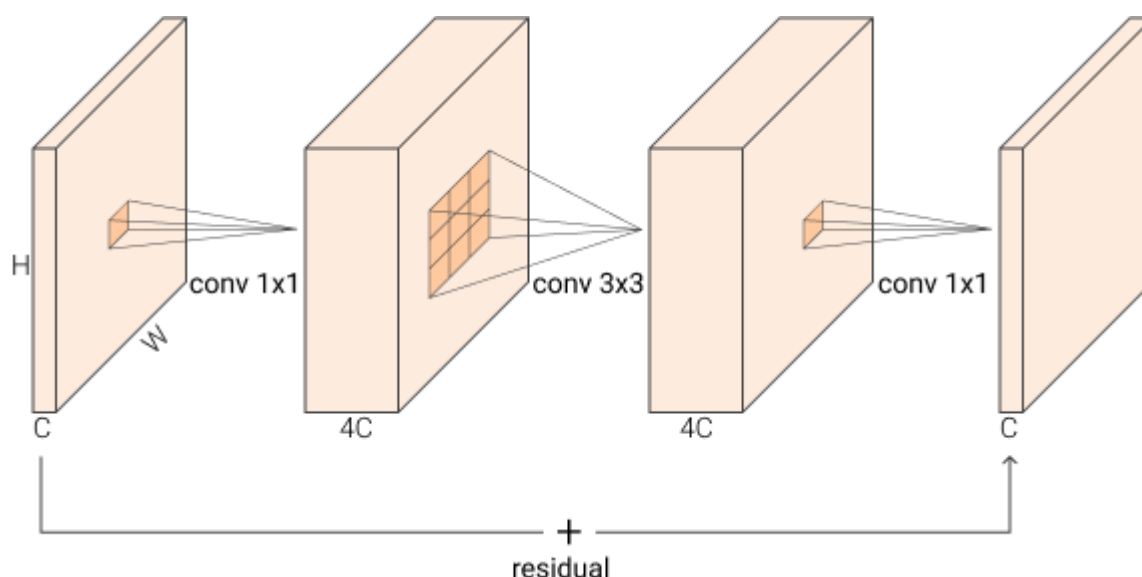
V CNN je důležitým krokem zpracování obrazu downsampling, který umožňuje postupnou abstrakci informací z obrazu. Vstupní obraz obvykle obsahuje 3 kanály (RGB), kde každý pixel vyjadřuje intenzitu barvy v daném bodě. Avšak, jak obraz prochází hlubšími vrstvami sítě, dochází k abstrakci na vyšší úrovni. Místo intenzit barev se hodnoty v kanálech začínou týkat přítomnosti hran, textur nebo dokonce celých objektů. Tento proces zvyšuje počet kanálů, protože síť se učí rozpoznávat a reprezentovat širokou škálu rysů a objektů.

S postupem do větších hloubek sítě se rozlišení obrazu snižuje. Tento jev je logický, jelikož při přechodu na vyšší úroveň abstrakce, kde jedna hodnota může reprezentovat celý objekt, tato hodnota pokrývá větší region ve vstupním obrazu.

Downsampling je realizován pomocí dvou hlavních operací: pooling a konvoluce s větším krokem. Pooling operace, jako jsou max pooling a average pooling, redukuje rozměry obrazu tím, že selektivně zachovávají důležité informace. Max pooling vybírá maximální hodnotu z každé podoblasti mapy rysů, což umožňuje efektivní zachycení dominantních rysů a zároveň zmenšuje velikost dat pro další zpracování. Average pooling vypočítává průměr všech hodnot v podoblasti, což vede k hladší reprezentaci rysů a je vhodné pro zachycení obecného rozložení rysů.

Konvoluce s větším krokem je další metodou pro downsampling, kde konvoluční filtr "skáče" o větší počet pixelů při aplikaci na obraz. Tím se přímo snižuje rozlišení výstupní mapy rysů bez nutnosti samostatné pooling operace. Tento přístup může zlepšit integraci downsamplingu a extrakce rysů, avšak může vést k vzniku artefaktů v obraze, pokud není pečlivě navržen.

2.1.3.7 Bottle neck a inverse bottle neck bloky



Obrázek 6 – Obrácené úzké hrdlo

V architektuře hlubokých CNN se využívají speciálně navržené bloky, známé jako "bottle neck" (úzké hrdlo) a "inverse bottle neck" (obrácené úzké hrdlo), pro efektivní zpracování a redukcí datových rozměrů, přičemž se zachovává či zvyšuje schopnost modelu učit se složité rysy z obrazových dat. Bloky jsou základní stavební jednotky v CNN, které definují specifické operace zpracování dat a jsou klíčové pro zlepšení výkonu sítě a snížení výpočetních nároků.

"Bottle neck" bloky snižují dimenzi vstupních dat před jejich zpracováním v hlubších vrstvách. Typicky zahrnují sérii konvolucí, kde první 1×1 konvoluce redukuje počet kanálů, následuje hlubší konvoluce pro zpracování těchto dat, a končí další 1×1 konvolucí, která obnovuje počet kanálů. Tímto způsobem umožňují efektivní zpracování s menším počtem výpočtů.

"Inverse bottle neck" bloky pracují na stejném principu, ale v opačném pořadí: začínají rozšířením počtu kanálů pomocí 1×1 konvolucí, pokračují hlubší konvolucí pro zpracování rozšířených dat a končí 1×1 konvolucí, která snižuje počet kanálů zpět. Tento přístup umožňuje lepší extrakci a učení složitějších rysů tím, že zpracovává informace v rozšířeném prostoru.

Oba typy bloků využívají koncept "rozšíření a redukce" kanálů, což zvyšuje efektivitu zpracování obrazu. Většina výpočtů se odehrává v prostoru s redukováným nebo rozšířeným počtem kanálů, což snižuje výpočetní nároky a počet parametrů, aniž by byla ohrožena schopnost modelu naučit se reprezentovat složité rysy obrazu.

2.1.3.8 Reziduální spojení

Reziduální spojení, zavedená v architektuře ResNet, jsou zásadním průlomem v hlubokém učení a v architektuře CNN. Umožňují efektivní trénink velmi hlubokých sítí řešením problému mizejícího gradientu. Tato technika zavádí "zkratky", které umožňují signálu obcházet jednu nebo více vrstev, což zajišťuje lepší šíření gradientu během zpětné propagace a usnadňuje učení.

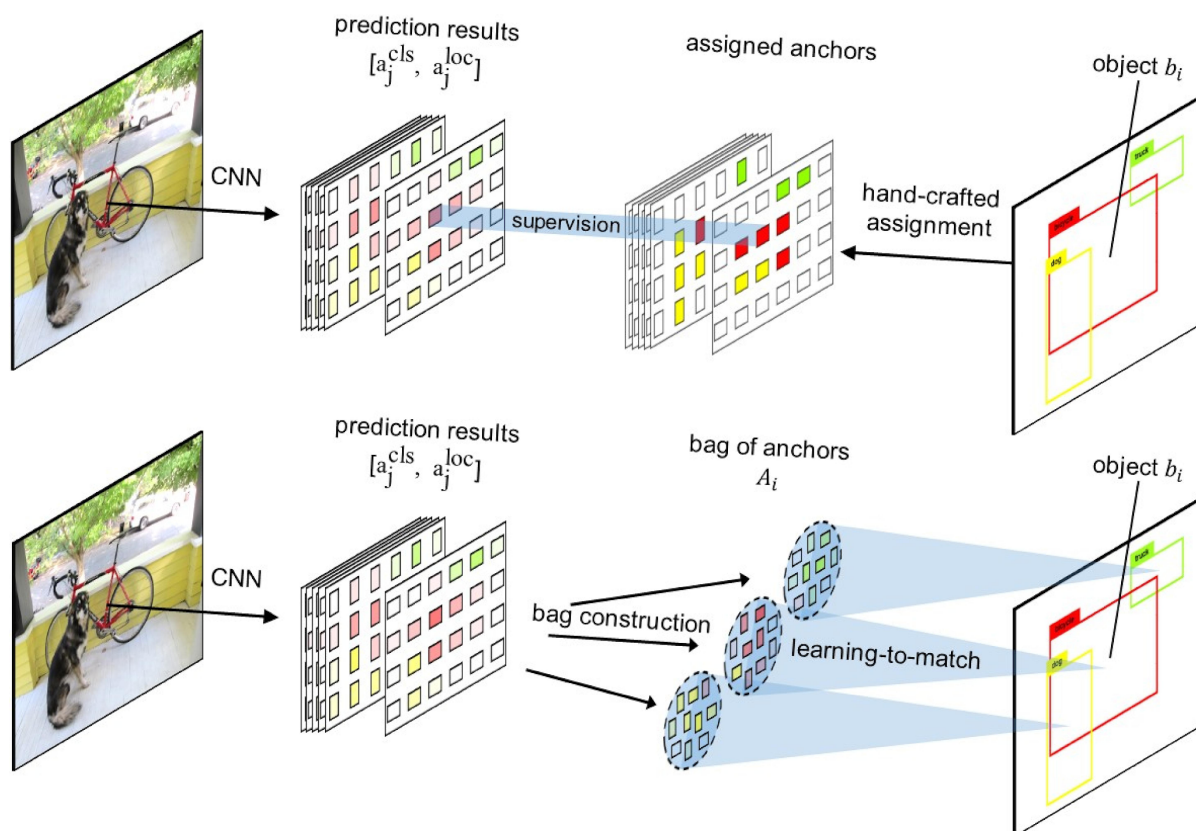
Reziduální spojení umožňují vrstvám učit se pouze změny oproti vstupu, nikoli celou rekonstrukci, což zvyšuje efektivitu učení. Pokud vrstvy identifikují užitečné transformace, další vrstvy mohou tyto transformace vylepšit. Pokud ne, mohou se naučit identitní transformaci, čímž se předejde zhoršení výkonu sítě při přidávání dalších vrstev.

Díky reziduálním spojením mohou sítě dosahovat lepšího výkonu v úlohách jako klasifikace, detekce objektů a segmentace, i při velmi hlubokých architekturách. Zvyšují také modularitu a flexibilitu sítě, což umožňuje experimentování s různými konfiguracemi vrstev.

Reziduální spojení tak představují klíčovou techniku pro překonání výzev trénování hlubokých sítí a otevírají cestu k vývoji pokročilejších modelů pro strojové vidění a umělou inteligenci.

2.1.3.9 Anchor-free Detekce

V oblasti detekce objektů se v poslední době prosazuje nové paradigma zvané anchor-free detekce, které se odchyľuje od tradičních anchor-based metod. Tyto metody se opírají o předem definované kotvicí boxy sloužící jako referenční body pro identifikaci a lokalizaci objektů. Naopak, anchor-free detekce přímo určuje umístění a rozměry objektů bez nutnosti kotvicích boxů, což přináší zlepšení v přesnosti, efektivitě a flexibilitě modelů.



Obrázek 7 - Anchor-free detekce

Tradiční metody vyžadují náročné ladění kotvicích boxů, což je časově náročné a zvyšuje výpočetní nároky. Anchor-free přístup eliminuje tyto potřeby tím, že přímo predikuje klíčové body nebo oblasti zájmu, což umožňuje modelům efektivněji se přizpůsobit různým formám a velikostem objektů a snižuje počet falešně pozitivních detekcí.

Díky odstranění potřeby generování a zpracování kotvicích boxů mohou být modely navrženy kompaktněji, což usnadňuje trénování a nasazení modelů, zejména na zařízeních s omezenými zdroji. Anchor-free metody také využívají pokročilé techniky strojového učení, jako je regrese klíčových bodů a segmentace sémantických map, což zlepšuje lokalizaci a rozpoznávání objektů a umožňuje integraci s dalšími úlohami zpracování obrazu.

2.1.3.10 Single Shot Detekce

Single Shot Detekce (SSD) je revoluční technika v detekci objektů, umožňující CNN identifikovat a lokalizovat objekty v jednom kroku. Na rozdíl od tradičních metod, které vyžadují vícefázové zpracování, SSD integruje detekci a klasifikaci do jediného procesu, což značně zvyšuje rychlost a efektivitu.

SSD využívá více konvolučních vrstev k predikci klasifikačních skóre a polohových údajů pro objekty v různých měřítkách, což umožňuje detekovat objekty různých velikostí a tvarů bez externích návrhů regionů. Tato schopnost zajišťuje vysokou přesnost detekce při zachování rychlého zpracování, což je klíčové pro real-time aplikace a zařízení s omezenými výpočetními zdroji.

Díky své flexibilitě a snadné integraci s různými architekturami CNN, SSD umožňuje výzkumníkům a vývojářům experimentovat s různými konfiguracemi pro optimální výkon v konkrétních aplikacích. I přes některá omezení, jako je obtížná detekce extrémně malých nebo těsně seskupených objektů, neustálý vývoj v oblasti hlubokého učení slibuje další zlepšení výkonu SSD.

2.2 Čtení SPZ

Po úspěšném vyříznutí obrazu SPZ z celkového snímku je nutné z tohoto obrazu extrahovat a interpretovat jednotlivé znaky, aby bylo možné získat čitelný a použitelný textový řetězec odpovídající registrační značce vozidla.

2.2.1 Použití šablon pro OCR

Metoda závisí na knihovně šablon pro specifické fonty, což omezuje její použitelnost na SPZ, které odpovídají těmto předem definovaným fontům. Efektivita této metody klesá, pokud je SPZ snímána z nepříznivých úhlů, což omezuje její univerzálnost a použitelnost v reálných podmínkách.

2.2.2 OCR pomocí neuronových sítí (RNN)

Rekurentní neuronové sítě (RNN) jsou speciálně navrženy pro zpracování sekvenčních dat, což je klíčové pro úlohy, kde je důležitá kontextová informace a vzájemná závislost mezi jednotlivými prvky v sekvenci. V kontextu rozpoznávání SPZ to znamená, že RNN jsou schopné efektivně zpracovat a interpretovat sekvenci znaků, které byly identifikovány na SPZ.

2.2.2.1 Sekvenční zpracování znaků

RNN využívají svou schopnost "pamatovat" předchozí informace v sekvenci, což umožňuje modelu zohlednit kontext při rozhodování o tom, jaký znak je aktuálně zpracováván. Tato

vlastnost je zvláště užitečná pro rozlišení znaků, které mohou být vizuálně podobné, ale jejich význam se liší v závislosti na pozici nebo kontextu v sekvenci.

2.2.3 Kombinace RNN a CNN pro komplexní OCR systémy

Ačkoliv RNN jsou silné v sekvenčním zpracování a interpretaci, pro maximální efektivitu a přesnost jsou často kombinovány s CNN, které slouží k detekci a extrakci znaků z obrazu. CNN jsou schopné lokalizovat a rozpoznat jednotlivé znaky na SPZ, zatímco RNN následně zpracovávají tyto rozpoznané znaky jako sekvenci, což umožňuje přesné rekonstruování celého textu SPZ.

Tato kombinace CNN a RNN v OCR systémech pro rozpoznávání SPZ představuje pokročilý přístup, který využívá silné stránky obou typů neuronových sítí, což vede k výraznému zlepšení v přesnosti a spolehlivosti rozpoznávání SPZ v různých podmínkách.

3 Návrh projektu

S ohledem na potřebu rychlosti zpracování v reálném čase, typicky 30 snímků za sekundu, což znamená méně než 40 ms na jeden snímek, bylo nutné provést důkladné optimalizace a využít experimentální techniky. Správný výběr architektury modelu, příprava a augmentace datové sady, spolu s implementací pokročilých algoritmů pro zpracování obrazu, byly klíčové pro dosažení požadované efektivitu a rychlosti projektu.

3.1 Hardware

3.1.1 Display

Pro zobrazení informací byly vybrány chytré brýle XREAL Air, které nabízejí 60Hz Full HD displej pro každé oko. Díky schopnosti připojení přes USB-C v režimu alternativního módu (DisplayPort) se tyto brýle staly ideální volbou pro snadnou integraci s dalšími zařízeními.

3.1.2 Kamera



Pro získání obrazových dat byla zvolena webkamera "Niceboy Stream", která disponuje clonou 2.2, rozlišením 2Mpx a snímkovou frekvencí 30 FPS. Ačkoli tato kamera nenabízí Full HD rozlišení a může mít nižší snímkovou frekvenci, byla preferována kvůli své dostupnosti na trhu. Hlavní nevýhodou kamery je však pokles výkonu v špatných světelných podmínkách, kdy může snímková frekvence klesnout až na 8 FPS. Aby bylo možné kameru snadno integrovat s brýlemi, provedla se modifikace, včetně výroby nového krytu pomocí 3D tisku, což umožňuje její jednoduché připevnění k rámu brýlí bez potřeby lepidla nebo šroubů.

3.1.3 Výpočetní jednotka

Pro jednoduchost testování a vývoje byl použit notebook s grafickou kartou NVIDIA RTX 3070, který poskytuje dostatečný výkon pro trénink a inferenci neuronových sítí. Nicméně, model byl navržen tak, aby byl schopen běžet v reálném čase na TPU Coral, což umožňuje jeho efektivní nasazení na zařízeních s omezenými výpočetními zdroji.

3.2 Softwarové technologie a vývojové nástroje

Pro vývoj algoritmu pro automatické rozpoznávání registračních značek byl zvolen programovací jazyk Python, který je v oblasti strojového učení a počítačového vidění široce používán díky své flexibilitě a rozsáhlé podpoře knihoven.

3.2.1 Použité knihovny a frameworky

Hlavními knihovnami použitými pro implementaci algoritmu byly TensorFlow a PyTorch. Tyto knihovny poskytují rozsáhlé možnosti pro trénink a nasazení hlubokých neuronových sítí, včetně konvolučních sítí, které jsou klíčové pro úspěšnou detekci a rozpoznávání objektů v obraze. Pro rychlou inferenci byl využit TensorRT od společnosti NVIDIA, což je optimalizační engine pro hluboké učení, který zvyšuje výkon a efektivitu modelů na hardware NVIDIA.

Pro konverzi modelů do optimalizovaných formátů byl použit ONNX (Open Neural Network Exchange), což umožnilo snadnou interoperabilitu mezi různými frameworky.

Pro matematické operace a manipulaci s daty byla použita knihovna NumPy, která je základem pro většinu vědeckých výpočtů v Pythonu. Pro práci s obrazem byla zvolena knihovna OpenCV, která poskytuje širokou škálu nástrojů pro zpracování obrazu a počítačové vidění. Pro zobrazení informací na chytrých brýlích byla použita knihovna Pillow, která umožňuje snadné vytváření a manipulaci s obrazovými daty.

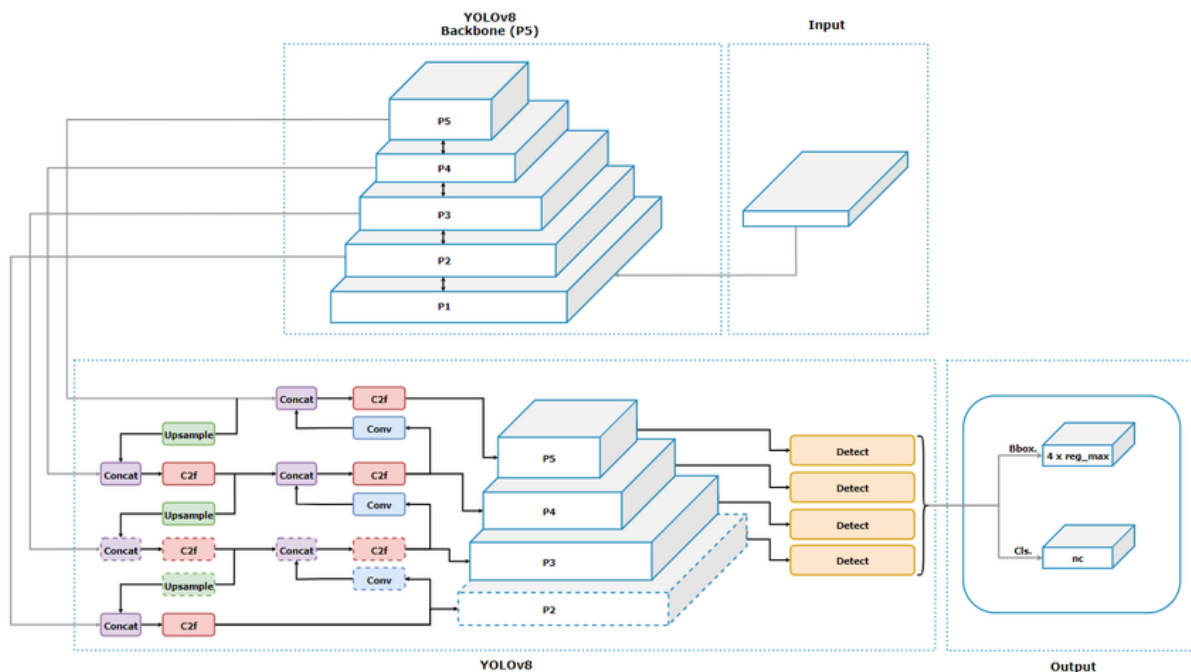
3.2.2 Vývoj webové aplikace

Pro správu databáze SPZ a interakci s uživatelem byla vyvinuta jednoduchá webová aplikace. Backend aplikace byl postaven na frameworku FastAPI, který je navržen pro rychlý vývoj moderních webových API s vysokým výkonem. FastAPI poskytuje podporu pro asynchronní programování, což umožňuje efektivní zpracování požadavků. Jako ASGI server pro spuštění aplikace byl použit Uvicorn, který je kompatibilní s FastAPI a zajišťuje rychlou a spolehlivou komunikaci.

Frontend webové aplikace byl vytvořen s využitím základních technologií HTML, CSS a JavaScript, což umožnilo vytvoření uživatelsky přívětivého rozhraní pro správu a prohlížení dat o SPZ.

3.3 Výběr architektury

Pro účely ANPR byla zvolena architektura YOLOv8, což je nejnovější iterace v rodině modelů YOLO (You Only Look Once), která je známá svou vysokou rychlostí a přesností v detekci objektů v reálném čase, jsou totiž navrženy tak, aby v jednom inferenčním kroku předpovídaly ohraničující rámečky a třídy pro detekované objekty. YOLOv8 přináší řadu inovací a vylepšení, které dále zlepšují výkon a efektivitu modelu, což je klíčové pro aplikace vyžadující rychlou a spolehlivou analýzu obrazu.



Obrázek 9 - Architektura YOLOv8

Jednou z klíčových vlastností YOLOv8, která byla rozhodující pro její výběr, je využití konvolučních bloků C2F (Cross-Stage Partial Bottleneck with 2 convolutions) a SPPF (Spatial Pyramid Pooling-Fast), které umožňují modelu efektivně zpracovávat informace na různých úrovních rozlišení a zachytit tak komplexní rysy objektů v obrazu. Tato schopnost je zásadní pro úspěšnou detekci a rozpoznávání SPZ, které mohou být na snímcích různě orientovány a v různých měřítkách.

Dalším důležitým aspektem architektury YOLOv8 je použití MaxPooling namísto většího kroku v konvolučních vrstvách. Tento přístup pomáhá předcházet potenciálním artefaktům, které mohou vzniknout při použití většího kroku v konvolucích, a zároveň zachovává vysokou úroveň detailů extrahovaných rysů.

3.4 Integrace detekce a OCR do jednoho modelu

Po výběru YOLOv8 pro detekci SPZ bylo zjištěno, že tradiční OCR modely založené na RNN jsou pro real-time aplikace pomalé kvůli nutnosti sekvenčního zpracování dat. Navíc, využití dvou oddělených modelů pro detekci SPZ a následné čtení textu by znamenalo redundantní extrakci rysů, což by celý proces zpomalilo.

Proto bylo rozhodnuto upravit architekturu YOLOv8, aby byla schopna zvládnout obě úlohy najednou. Detekci SPZ a rozpoznávání textu. Tato úprava zahrnovala přidání další vrstvy pro extrakci rysů do CNN, což umožňuje detekci i velmi malých písmen, která by pro standardní architekturu YOLOv8 mohla představovat problém.

Následně byl vyvinut vlastní mechanismus pro zpracování detekcí, který rozděljuje všechny detekce do dvou kategorií: SPZ a znaky. Pro každou detekovanou SPZ tento mechanismus nejprve identifikuje všechny znaky, které se nacházejí v její oblasti. Poté aplikuje specifickou filtraci těchto znaků, kde nejprve odstraní ty ohraničující rámečky znaků, jejichž průsečík s odpovídající SPZ je menší než jedna třetina jejich plochy. Tento krok zajišťuje, že jsou

zachovány pouze znaky, které jsou s vysokou pravděpodobností součástí SPZ, a minimalizuje se tak riziko zahrnutí náhodných nebo nežádoucích detekcí.

Po filtraci znaků následuje další krok, ve kterém jsou detekce řazeny podle jejich konfidenčního skóre a je vybráno top 7 detekcí, což odpovídá běžnému počtu znaků na SPZ. Tento proces zajišťuje, že i v případě, kdy model předpoví více ohraničujících boxů pro tentýž znak budou použity pouze nejspolehlivější detekce.

Nakonec jsou tyto vybrané znaky seřazeny podle jejich x-ových souřadnic, což umožňuje správné uspořádání znaků do konečného predikovaného textu SPZ. Tímto způsobem bylo možné získat OCR informace "zdarma", tedy bez potřeby dalšího samostatného OCR modelu.

3.5 Příprava dat pro trénink

Proces přípravy dat pro trénink modelu pro detekci a rozpoznávání SPZ zahrnoval několik klíčových kroků, které byly nezbytné pro úspěšné naučení modelu OCR funkcionality. Vzhledem k tomu, že reálný dataset obsahoval omezený počet znaků a nebyl dostatečně rozmanitý pro naučení modelu efektivně rozpoznávat a číst znaky, bylo nutné přistoupit k alternativní strategii.

3.5.1 Syntetický dataset pro OCR

Prvním krokem bylo vytvoření syntetického datasetu pro OCR část úlohy. Tento přístup zahrnoval stahování obrazů z internetu, které byly následně filtrovány pomocí segmentačního modelu pro text, aby byly vybrány pouze ty obrazy, které neobsahovaly text a poskytovaly tak čisté pozadí. Na tyto obrazy byly následně renderovány náhodné znaky v široké škále fontů a velikostech, doplněné o různé transformace a rotace. Tímto procesem vznikl rozsáhlý syntetický dataset, který zahrnuje znaky v mnoha kontextech a na rozmanitých pozadích. Tento dataset byl klíčový pro předtrénink modelu a zajištění, že váhy modelu již obsahovaly základní znalosti potřebné pro OCR úlohu ještě před samotným tréninkem lokalizace.

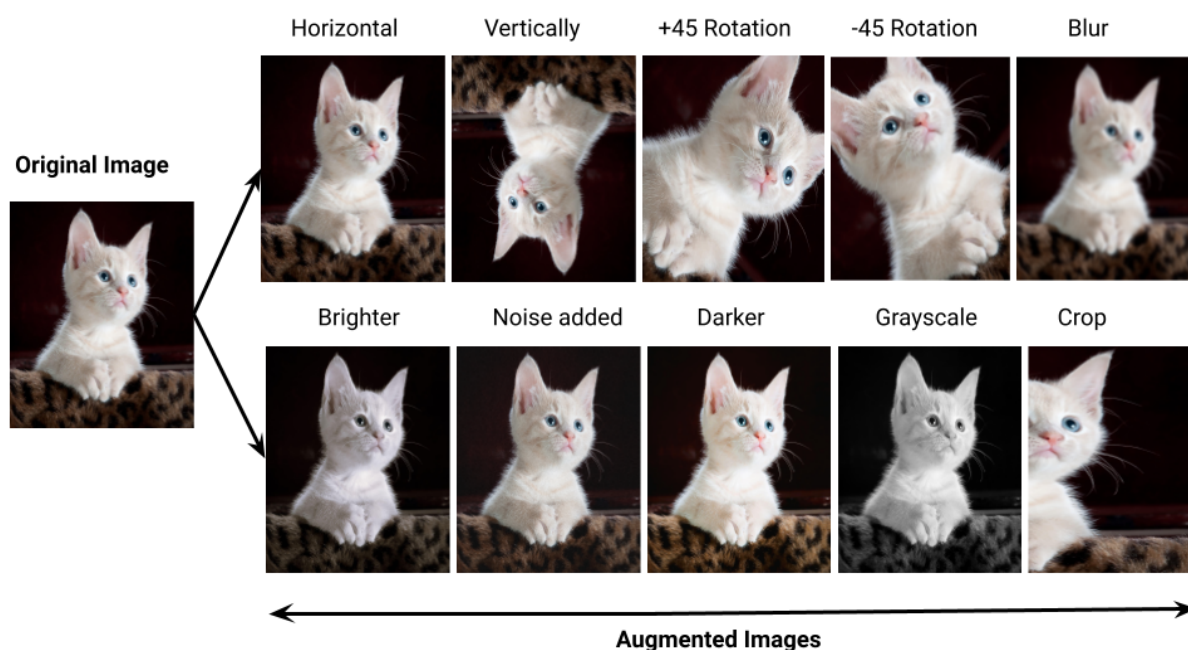
3.5.2 Předtrénink modelu

S využitím syntetického datasetu byl model předtrénován, aby se naučil základní aspekty rozpoznávání a čtení znaků. Tento krok byl nezbytný pro zajištění, že model bude mít solidní základ pro další trénink na reálných datech.

3.5.3 Reálný dataset

Pro detekci samotných SPZ bylo potřeba jemné doladění. Na to byl použit reálný dataset SPZ, který byl profesionálně označen. Označení spočívalo v ručním kreslení ohraničujících boxů okolo SPZ na obrázcích. Tento dataset poskytl důležitou diverzitu reálných obrazů SPZ, která byla nezbytná pro úspěšný trénink modelu na rozpoznávání SPZ v různých reálných podmínkách.

3.5.4 Augmentace a změna hyper parametrů pro fine tuning



Obrázek 10 - Augmentace dat

Pro další zlepšení robustnosti a schopnosti modelu generalizovat byla použita rozsáhlá augmentace dat, včetně translací, deformací, šumu, pixelizace, změn v osvětlení (HSV posuny), rozmazání a dalších technik. Tyto augmentace pomohly modelu naučit se rozpoznávat SPZ i v obtížných vizuálních podmínkách. Současně byla provedena úprava hyperparametrů modelu, především snížením rychlosti učení a změnou optimalizačního algoritmu z ADAM na SGD s momentem, čímž se snížila pravděpodobnost, že model zapomene to, co se naučil při předtrénování na syntetickém OCR datasetu.

3.6 Použití TensorRT

Pro zajištění maximální efektivity a rychlosti inferenčního procesu bylo rozhodnuto využít TensorRT, výkonný inferenční engine od společnosti NVIDIA, který je navržen pro vysokorychlostní inferenci na hardware NVIDIA. Tento krok zahrnoval konverzi trénovaného a jemně doladěného modelu do formátu *.engine*, což umožnilo aplikovat řadu optimalizačních technik nabízených TensorRT. Mezi tyto techniky patří kvantizace, vrstvené fúzování a další optimalizace, které mají za cíl redukovat latenci a zvýšit propustnost modelu při zachování jeho přesnosti.

Díky optimalizaci pro NVIDIA GPU hardware, využití TensorRT vedlo k výraznému zvýšení rychlosti inferenčního procesu, což je zásadní pro aplikace vyžadující rychlou a v reálném čase probíhající analýzu obrazu, jako je detekce a rozpoznávání SPZ. Optimalizace modelu pomocí TensorRT přinesla významné zlepšení inferenční rychlosti, snížení spotřeby paměti a zvýšení celkové efektivity modelu při běhu na cílovém hardwaru.

3.7 Zlepšení přesnosti detekce

3.7.1 Sledování SPZ v čase

Pro zvýšení přesnosti a spolehlivosti detekce SPZ byl implementován mechanismus, který umožňuje sledování a rozpoznávání stejných SPZ v sérii po sobě jdoucích snímků. Tento přístup vychází z předpokladu, že SPZ vozidla se v krátkém časovém úseku výrazně neposune a její poloha mezi dvěma po sobě jdoucími snímky bude relativně konzistentní. Klíčovým aspektem tohoto zlepšení je využití prostorového překryvu ohraničujících rámečků detekovaných SPZ. Pokud se rámečky detekované v aktuálním snímku překrývají s rámečky z předchozího snímku nad určitou prahovou hodnotu, lze je považovat za stejnou SPZ. Tento přístup využívá fakt, že SPZ je typicky umístěna na pevném místě vozidla (například na nárazníku) a v běžných provozních podmínkách se nestává, že by se SPZ různých vozidel významně překrývaly, pokud nejsou vozidla v bezprostřední blízkosti nebo v situaci dopravní nehody.

Navíc, byl do systému integrován algoritmus, který efektivně zvládá situace, kdy se v jednom snímku objeví více SPZ. Díky použití jednoduchého algoritmu z dynamického programování, systém hledá nejlepší možné párování mezi detekcemi SPZ v po sobě jdoucích snímcích. Tento algoritmus umožňuje systému optimalizovat párování detekcí tak, aby bylo dosaženo nejvyšší pravděpodobnosti správné identifikace SPZ.

3.7.2 Aplikace Kalmanova filtru na pohyb SPZ

Při nízké frekvenci snímkování nebo rychlém pohybu vozidel může dojít k situaci, kdy se detekce SPZ ve dvou po sobě jdoucích snímcích nepřekrývají, což by mohlo vést k nemožnosti považovat je za stejné vozidlo. Pro řešení tohoto problému byl využit Kalmanův filtr, algoritmus určený k odhadu stavu lineárního dynamického systému na základě nepřesných měření. Klíčovým aspektem je, že při příchodu nových detekcí SPZ není hodnoceno jejich překrytí s pozicemi, kde se SPZ nacházely dříve, ale s pozicemi, kde by se měly nacházet, pokud by pokračovaly v lineárním pohybu.

Kalmanův filtr funguje ve dvou fázích: predikce a korekce. V predikční fázi vytváří odhad budoucí polohy SPZ na základě aktuálního pohybu. V korekční fázi filtr upravuje tento odhad podle nově získaných dat, tedy podle skutečné detekce SPZ v novém snímku. Během korekční fáze se filtr také učí informace o rychlosti SPZ, což umožňuje zlepšení přesnosti budoucích odhadů. Tento proces umožňuje plynulé sledování pohybu SPZ a zvyšuje přesnost detekce v proměnlivém prostředí.

Využití Kalmanova filtru zvyšuje přesnost systému pro detekci a rozpoznávání SPZ, zejména v situacích s nízkou frekvencí snímkování nebo rychlým pohybem vozidel. Díky predikcím filtru je možné nové detekce efektivně porovnávat s odhady polohy, což snižuje riziko ztráty sledování vozidla v případě, že se detekce mezi snímky nepřekrývají.

3.7.3 Korekce chyb v detekci znaků na SPZ

Díky možnosti přiřadit totožným SPZ identifikátor nezávislý na samotném textu na SPZ, který může být chybně rozpoznán, lze výrazně zlepšit celkovou přesnost a spolehlivost

systému. Tento přístup umožňuje systému sledovat a identifikovat SPZ i v případě, že dojde k chybám v detekci nebo rozpoznávání jednotlivých znaků.

V praxi to znamená, že každé detekované SPZ je přiřazen unikátní identifikátor, který je odvozen z její polohy a pohybu v čase, a ne z přesného textu na SPZ. Tento identifikátor pak slouží jako klíč pro sledování SPZ napříč snímky, což umožňuje systému zachovat kontinuitu identifikace i v případě, že se v některých snímcích znaky na SPZ nerozpoznají správně nebo vůbec.

Pro korekci chyb v detekci znaků na SPZ systém využívá tabulku počtu výskytů znaků na jednotlivých pozicích, která je aktualizována pro každý identifikátor SPZ. Pokud systém zaznamená chybu nebo chybějící znak v některém ze snímků, může se obrátit na tuto tabulku a určit, které znaky jsou nejčastěji detekovány na dané pozici pro danou SPZ. Tímto způsobem může systém "opravit" chybné nebo chybějící detekce tím, že jako reálné znaky SPZ použije ty, které se vyskytují nejčastěji.

Tento mechanismus korekce zajišťuje, že i v případě, kdy detekce znaků selže nebo udělá chybu na několika snímcích, celková identifikace SPZ zůstává přesná a konzistentní. Tento přístup tak přináší další úroveň robustnosti do systému pro detekci a rozpoznávání SPZ, což je zásadní pro jeho spolehlivé použití v reálném světě.

4 Závěr

4.1 Shrnutí projektu

V rámci tohoto projektu byl vyvinut inovativní systém pro automatické rozpoznávání registračních značek vozidel s využitím chytrých brýlí. Hlavním cílem bylo vytvořit efektivní a rychlý nástroj, který by umožnil identifikaci SPZ v reálném čase s minimálními hardwarovými požadavky. Projekt kombinuje pokročilé technologie počítačového vidění a strojového učení, konkrétně konvoluční neuronové sítě pro detekci a rozpoznávání SPZ, a implementuje je do nositelného zařízení.



Obrázek 11 - ANPR Brýle

Vývoj systému zahrnoval návrh a trénink modelu, který byl optimalizován pro rychlou a přesnou detekci SPZ, a jeho integraci s chytrými brýlemi XREAL Air. Tyto brýle slouží jako zobrazovací jednotka a jsou doplněny o externí webkameru "Niceboy Stream" pro získávání obrazových dat. Navzdory omezením kamery, jako je nižší frekvence snímků a slabší výkon za špatných světelných podmínek, bylo dosaženo uspokojivých výsledků díky sofistikovaným algoritmům zpracování obrazu a optimalizaci modelu.

Projekt také zahrnoval vývoj uživatelského rozhraní pro brýle, které vizuálně informuje o statusu detekovaných vozidel, a jednoduchého webového rozhraní pro správu databáze SPZ. Celý systém byl navržen s ohledem na možnost budoucího rozšíření a vylepšení, jak na úrovni hardwaru, tak softwaru.

Celkově tento projekt představuje významný krok vpřed v oblasti ANPR technologií a nositelných zařízení, nabízí nové možnosti pro aplikace v bezpečnosti, řízení dopravy a automatizaci přístupových systémů.

4.2 Význam a přínosy

Projekt automatického rozpoznávání registračních značek pomocí chytrých brýlí přináší řadu významných přínosů a otevírá nové možnosti v oblasti bezpečnosti, monitorování dopravy a automatizace přístupových systémů. Jeho implementace demonstruje praktické využití pokročilých technologií počítačového vidění a strojového učení v kombinaci s nositelnou technologií, což představuje inovativní přístup k problému identifikace vozidel.

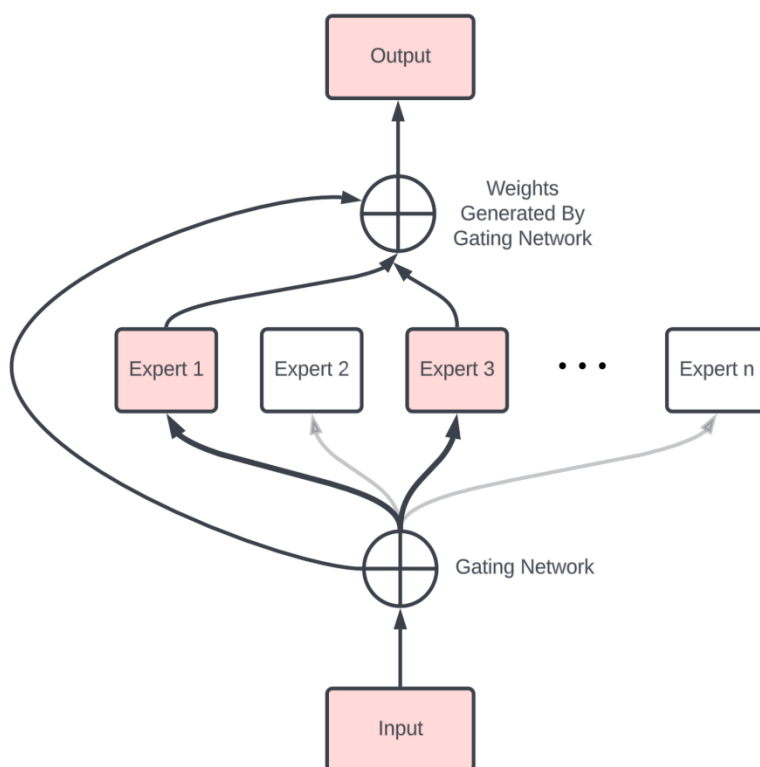
Přínosy pro bezpečnost a monitorování jsou značné. Systém umožňuje rychlou a efektivní identifikaci vozidel, což je klíčové pro zvýšení bezpečnosti na parkovištích, v areálech firem, na hranicích nebo v jiných kontrolovaných oblastech. Umožňuje okamžitou verifikaci

oprávnění vozidla vstoupit do dané oblasti a může sloužit jako nástroj pro monitorování a analýzu dopravních toků.

Automatizace přístupových systémů představuje další významný přínos projektu. Nabízí efektivní řešení pro automatizaci přístupových systémů bez nutnosti instalace drahých stacionárních ANPR kamer. Nosným prvkem je flexibilita a mobilita systému, který může být snadno přenesen a použit v různých lokacích podle aktuální potřeby.

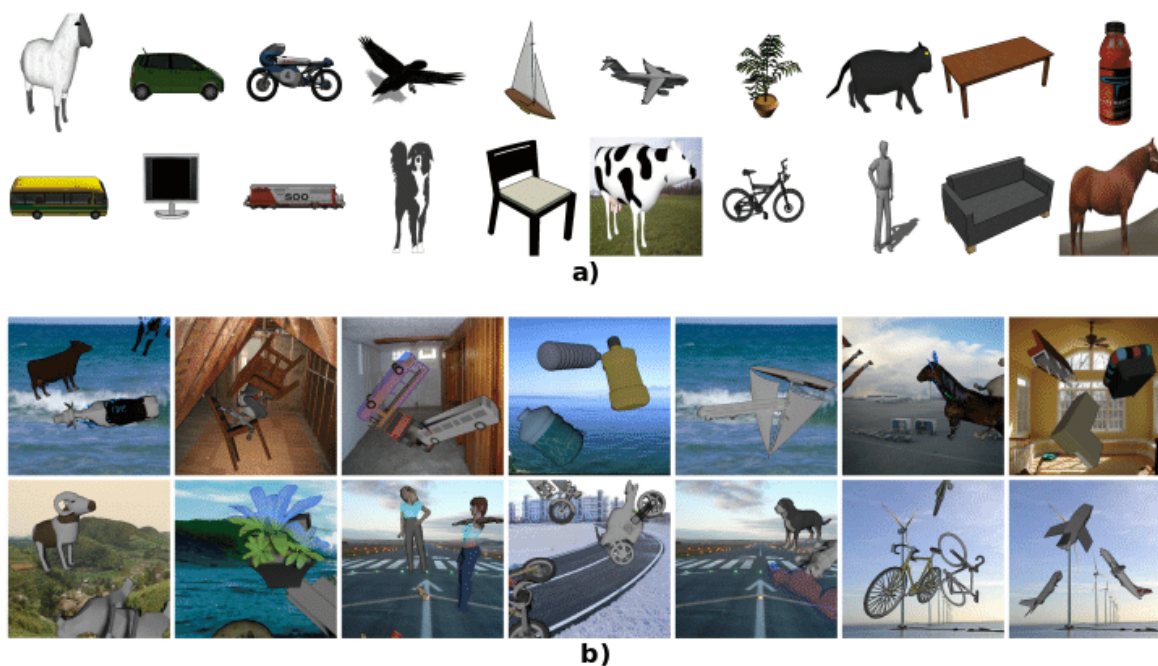
Integrace ANPR systému do chytrých brýlí představuje významný pokrok v oblasti nositelných technologií.

4.3 Budoucí rozvoj a plány



Obrázek 12 - Mixture of Experts

V rámci budoucího rozvoje projektu automatického rozpoznávání registračních značek s využitím chytrých brýlí je kladen důraz na vylepšení a rozšíření stávajících technických a softwarových řešení. Mezi klíčové plány patří začlenění pokročilých metod strojového učení, jako jsou Mixture of Experts (MoE) a attention mechanismy, které zvýší adaptabilitu a přesnost modelu při rozpoznávání SPZ.



Obrázek 13 - Syntetický dataset

Pro zlepšení schopnosti modelu generalizovat a rozpoznávat širokou škálu SPZ, včetně těch s odlišnými fonty a z různých zemí, je plánováno rozšíření trénovacího datasetu. Kromě sběru reálných dat se počítá s využitím 3D renderovaných scén a virtuální kamery pro generování syntetických dat, což umožní modelu trénovat na širokém spektru scénářů.



Obrázek 14 - XReal Air 2 Pro

V oblasti hardwaru je zamýšleno přejít na výkonnější a specializovanější zařízení, včetně možného nasazení malých počítačů jako Raspberry Pi ve spojení s TPU Coral od Google pro zvýšení mobility a energetické efektivity systému. Dále je plánováno přejít na nové brýle s integrovanou kamerou, možná model XREAL Air 2 Pro, což by odstranilo potřebu externí kamery a zjednodušilo celkovou konfiguraci.

Pro řešení problémů s nelineárním pohybem SPZ způsobeným pohybem hlavy uživatele je zvažováno zpětné inženýrství komunikace mezi brýlemi a zařízením. Chytré brýle použité v projektu jsou vybaveny gyroskopem, který by mohl být klíčovým prvkem pro přesnější sledování pohybu SPZ. Nicméně, gyroskop je v současné době využíván pouze proprietárními aplikacemi brýlí a neexistuje pro něj open source kód, který by umožnil jeho přímé využití v tomto projektu. Tento krok by umožnil využití dat z gyroskopu brýlí pro přesnější predikce pohybu SPZ pomocí modifikovaných Kalmanových filtrů, které by lépe zohledňovaly dynamiku pohybu. Zpětné inženýrství komunikace a psaní vlastních ovladačů pro gyroskop by tak otevřelo cestu k výraznému zlepšení schopnosti systému sledovat a rozpoznávat SPZ i v náročných podmínkách, kdy se uživatel rychle pohybuje nebo prudce otáčí hlavou.

Seznam obrázků

Obrázek 1 - Perceptron	10
Obrázek 2 – Zpětná propagace	11
Obrázek 3 - Gradientní sestup	12
Obrázek 4 - Přeučení	12
Obrázek 5 - Konvoluční síť	14
Obrázek 6 – Obrácené úzké hrdlo	17
Obrázek 7 - Anchor-free detekce	19
Obrázek 8 - Kamera Niceboy Stream	22
Obrázek 9 - Architektura YOLOv8	24
Obrázek 10 - Augmentace dat	27
Obrázek 11 - ANPR Brýle	30
Obrázek 12 - Mixture of Experts	32
Obrázek 13 - Syntetický dataset	33
Obrázek 14 - XReal Air 2 Pro	33

Seznam použité literatury a zdrojů obrázků

- [1] LeCun, Y., Boser, B., Denker, J., Henderson, D., Howard, R., Hubbard, W., and Jackel, L. 1989. Backpropagation Applied to Handwritten Zip Code Recognition. *Neural Computation*, 1(4), p.541-551.
- [2] Krizhevsky, A., Sutskever, I., and Hinton, G. 2012. ImageNet Classification with Deep Convolutional Neural Networks. In *Advances in Neural Information Processing Systems*. Curran Associates, Inc..
- [3] Karen Simonyan, and Andrew Zisserman. (2015). Very Deep Convolutional Networks for Large-Scale Image Recognition.
- [4] François Chollet. (2017). Xception: Deep Learning with Depthwise Separable Convolutions.
- [5] Jost Tobias Springenberg, Alexey Dosovitskiy, Thomas Brox, and Martin Riedmiller. (2015). Striving for Simplicity: The All Convolutional Net.
- [6] He, K., Zhang, X., Ren, S., and Sun, J. 2016. Deep Residual Learning for Image Recognition. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*.
- [7] Joseph Redmon, Santosh Divvala, Ross Girshick, and Ali Farhadi. (2016). You Only Look Once: Unified, Real-Time Object Detection.
- [8] Diederik P. Kingma, & Jimmy Ba. (2017). Adam: A Method for Stochastic Optimization.
- [9] Jocher, G., Chaurasia, A., & Qiu, J. (2023). Ultralytics YOLO (Version 8.0.0) [Computer software]. <https://github.com/ultralytics/ultralytics>
- [10] TensorRT. Online. C2024. Dostupné z: developer.nvidia.com/tensorrt [cit. 2024-03-18].
- [11] FastAPI. Online. C2024. Dostupné z: fastapi.tiangolo.com [cit. 2024-03-18].
- [12] Kalman filter. In: *Wikipedia: the free encyclopedia* [online]. San Francisco (CA): Wikimedia Foundation, C2001. Dostupné z: https://en.wikipedia.org/wiki/Kalman_filter [cit. 2024-03-18]
- [13] Tensorflow. Online. C2015. Dostupné z: tensorflow.org [cit. 2024-03-18].
- [14] Medium. Online. In: *Medium.com*. Dostupné z: https://miro.medium.com/v2/resize:fit:640/1*sPg-0hha7o3iNPjY4n-vow.jpeg. [cit. 2024-04-02].
- [15] Toward data science. Online. In: *towardsdatascience.com* Dostupné z: https://miro.medium.com/v2/resize:fit:1200/1*XJ7ioX3mFycK5FwsLqVJ8w.png. [cit. 2024-04-02].

- [16] KD Nuggets. Online. In: www.kdnuggets.com. Dostupné z: https://miro.medium.com/max/1400/1*tQTcGTLZqnI5rp3JYO_4NA.png. [cit. 2024-04-02].
- [17] Medium. Online. In: medium.com. Dostupné z: https://miro.medium.com/v2/resize:fit:1125/1*_7OPgojau8hkiPUiHoGK_w.png. [cit. 2024-04-02].
- [18] MDPI. Online. In: www.mdpi.com. Dostupné z: https://www.mdpi.com/sensors/sensors-19-04933/article_deploy/html/images/sensors-19-04933-g001.png. [cit. 2024-04-02].
- [19] Toward data science. Online. In: towardsdatascience.com. Dostupné z: https://miro.medium.com/v2/resize:fit:571/1*hq62bMEMgvgBgU_f8cW7jbw.png. [cit. 2024-04-02].
- [20] Papers with code. Online. In: paperswithcode.com. Dostupné z: https://production-media.paperswithcode.com/methods/new_3Dpipeline.jpg. [cit. 2024-04-02].
- [21] Research Gate. Online. In: www.researchgate.net. Dostupné z: <https://www.researchgate.net/publication/372207753/figure/fig2/AS:11431281173322965@1688824328575/The-improved-YOLOv8-network-architecture-includes-an-additional-module-for-the-head.ppm>. [cit. 2024-04-02].
- [22] UbiAI. Online. In: ubiai.tools. Dostupné z: https://ubiai.tools/wp-content/uploads/2023/11/0_LNtz0G4cngapDH41.png. [cit. 2024-04-02].
- [23] XReal. Online. In: Xreal.com. Dostupné z: https://us.shop.xreal.com/cdn/shop/files/Air2Pro_2.png?v=1705909511&width=1946. [cit. 2024-04-02].
- [24] DeepGram. Online. In: Deepgram.com. Dostupné z: <https://www.datocms-assets.com/96965/1695407447-image1.png>. [cit. 2024-04-02].
- [25] Research Gate. Online. In: www.researchgate.net. Dostupné z: <https://www.researchgate.net/publication/324762925/figure/fig2/AS:629770475814913@1527160390335/a-3D-models-used-for-synthetic-dataset-b-Synthetic-images-with-multiple-objects-c.png>. [cit. 2024-04-02].
- [26] NiceBoy. Online. In: Niceboy.eu. Dostupné z: <https://niceboy.eu/files/produkt/stream/stream-layout-04.jpg>. [cit. 2024-04-02].