



Semestrální práce

Segmentace obrazu pomocí prahování

Vypracoval: Lískovský David

Předmět : XDEEP

Datum : 26.10.2025

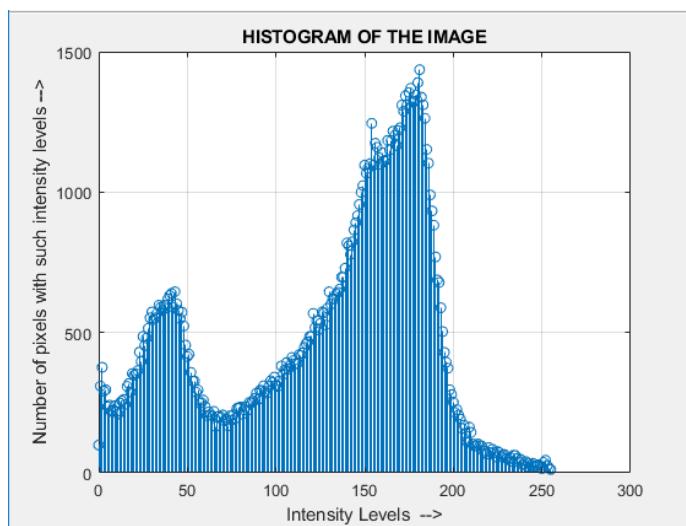
Obsah

1. Analýza problému	3
2. Analýza dat.....	4
2.1 Synteticky generované obrazy.....	4
2.2 Reálné obrazy.....	4
2.3 Dataset pro neuronovou síť	5
2.4 Shrnutí.....	5
3. Návrh řešení	5
3.1 Uživatelský práh	5
3.2 Otsu metoda	6
3.3 Fuzzy přístup.....	6
3.4 Neuronová síť (U-Net)	6
3.5 Shrnutí návrhu.....	6
4. Realizace řešení	7
4.1 Klasické metody prahování (Notebook A).....	7
4.2 Neuronová síť U-Net (Notebook B)	8
4.3 Shrnutí implementace	8
5. Evaluace funkčního řešení.....	9
5.1 Režim A – syntetická data s GT maskou	9
5.2 Režim B – reálný obraz s GT maskou.....	10
5.3 Režim C – reálný obraz bez GT masky	11
5.4 Režim D – neuronová síť	12
5.5 Shrnutí.....	13
6. Závěry	13
Seznam obrázků.....	14

1. Analýza problému

Prahování obrazu je jednou ze základních metod zpracování obrazu, která umožňuje převést šedo tónový snímek na binární reprezentaci. Typickým cílem je oddělit popředí od pozadí, například při detekci objektů, buněk či textu. Úspěšné prahování je však podmíněno vhodnou volbou prahové hodnoty.

Volba prahu není triviální. Pokud je histogram obrazu jasně bimodální (např. světlý objekt na tmavém pozadí), lze hranici mezi dvěma třídami pixelů určit poměrně snadno. V reálných datech se ale často objevují komplikace: šum, nerovnoměrné osvětlení, nízký kontrast nebo vícečetné objekty. V těchto případech může být ruční volba prahu subjektivní a časově náročná, zatímco automatické metody nemusí vždy přinést ideální výsledky.



Obrázek 1 - Histogram obrázku

Existuje více přístupů, jak tento problém řešit:

- **Uživatelský práh** – uživatel nastaví hodnotu prahu ručně, obvykle na základě histogramu. Výhodou je kontrola nad výsledkem, nevýhodou subjektivita a pracnost.
- **Otsu metoda** – automatická statistická metoda, která hledá práh minimalizující vnitrotřídní rozptyl intenzit. Je velmi účinná pro bimodální histogramy.
- **Fuzzy přístup** – využívá teorii fuzzy množin, kde je příslušnost pixelu k popředí nebo pozadí popsána stupněm neurčitosti. Tento přístup je robustnější k různým typům zkreslení, ale může ztráct na přesnosti hranic.
- **Neuronová síť (U-Net)** – moderní přístup založený na hlubokém učení. Místo explicitní volby prahu se model učí segmentovat obraz na základě trénovacích dat. Tento přístup je robustní vůči šumu, variabilitě scén i nehomogennímu osvětlení, ale vyžaduje anotovaná data a výpočetně náročnější trénink.

V rámci projektu se zaměřím na porovnání těchto čtyř metod. Tři z nich (uživatelský práh, Otsu, fuzzy) jsou implementovány v jednom interaktivním Colab notebooku, který umožňuje vizuální i kvantitativní porovnání. Čtvrtá metoda – neuronová síť U-Net je implementována v samostatném Colab sešitě, kde probíhá trénink na párech obrázek - maska a následná predikce segmentace. Výsledky budou porovnány na stejných datech, kde je to možné, a doplněny o metriky přesnosti (IoU).

2. Analýza dat

V rámci projektu byly použity dva typy dat:

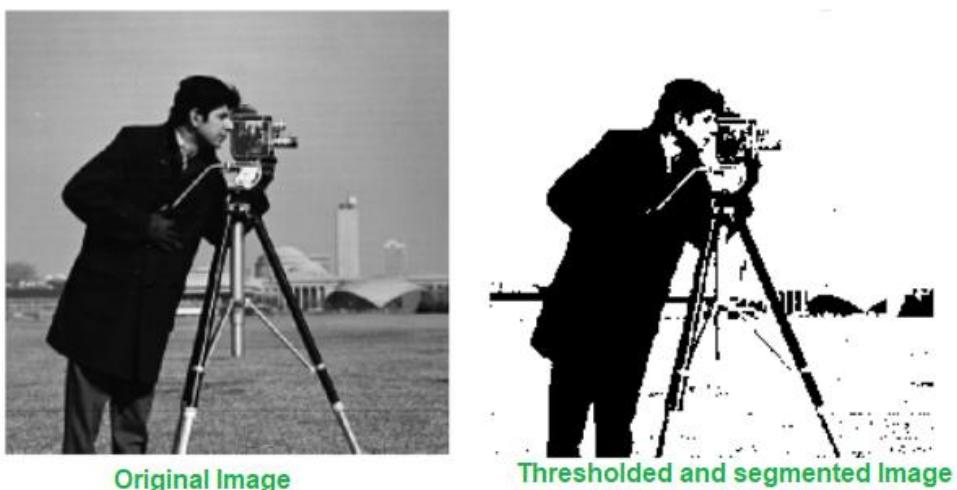
2.1 Synteticky generované obrazy

Pro účely objektivního hodnocení byly vytvářeny jednoduché syntetické obrazy (např. kruh nebo obdélník na homogenním pozadí). Tyto obrazy mají jednoznačně definovanou GT masku, která přesně odpovídá objektu. Díky tomu lze přesně změřit kvalitu segmentace pomocí metrik, jako je míra překryvu (IoU). Tyto data slouží jako ideální případ, kdy histogram obrazu je bimodální a prahování by mělo fungovat bezchybně.

2.2 Reálné obrazy

Druhým zdrojem dat jsou běžné fotografie nahrané uživatelem. V některých případech mohou být k dispozici i anotované masky (GT), častěji se však jedná o obrázky bez masky, u kterých je hodnocení provedeno pouze vizuálně. Tyto obrazy reprezentují složitější a realističtější scénáře, kde jednoduché prahování často narází na své limity.

Použití obou typů dat umožňuje ukázat rozdíl mezi ideálním a reálným prostředím. Na syntetických datech lze metody kvantitativně porovnat, zatímco u reálných obrázků je hodnocení spíše kvalitativní a ukazuje praktická omezení jednotlivých přístupů.



Obrázek 2 - Ukázka segmentace

2.3 Dataset pro neuronovou síť

Pro trénink neuronové sítě byl připraven samostatný dataset tvořený dvojicemi obrázků a odpovídajících binárních masek. Obrázky jsou ve formátu *.jpg, masky ve formátu *.png, přičemž obě složky mají shodný název souboru. Data byla uložena ve sdílené složce na Google Disku a zpřístupněna v Colabu pomocí připojení přes drive.mount.

Pro rychlý experimentální běh byl použit náhodný podzorek dat (např. 200 párů), který umožnuje trénink s nízkými nároky na výpočetní čas. Obrázky i masky byly předzpracovány: převod do šedotónové škály, normalizace, resize na jednotnou velikost (např. 128×128 px) a volitelná augmentace (horizontální/vertikální flip). Masky byly binarizovány na základě prahové hodnoty >0.

Při přípravě dat byla rovněž analyzována vyváženosť tříd (foreground vs. background), což umožnilo nastavit váhu pro ztrátovou funkci (pos_weight) v rámci tréninku. Tím se zohledňuje případná dominance pozadí v obraze a síť se učí lépe rozpoznávat méně zastoupené objekty.

2.4 Shrnutí

Datová příprava byla navržena tak, aby umožnila testování všech čtyř metod segmentace na srovnatelných vstupních datech. Syntetická data slouží pro kvantitativní porovnání v ideálním prostředí, reálné obrázky s GT maskou umožňují testování v praktických podmínkách, a reálné obrázky bez GT slouží pro vizuální demonstraci. Dataset pro U-Net byl připraven tak, aby odpovídal formátu běžně používanému v segmentačních úlohách a umožnil trénink neuronové sítě v samostatném Colab sešitě.

3. Návrh řešení

Cílem řešení je implementovat a porovnat čtyři přístupy k segmentaci obrazu, které reprezentují různé úrovně složitosti a přístupů k problému. Každá metoda má své výhody i omezení a je vhodná pro jiný typ dat nebo scénáře. Porovnání probíhá jak na syntetických, tak na reálných datech, s důrazem na vizuální i kvantitativní vyhodnocení.

3.1 Uživatelský práh

Nejjednoduší přístup, ve kterém uživatel ručně nastavuje hodnotu prahu pomocí interaktivního posuvníku. Tento režim slouží k demonstraci vlivu subjektivní volby na výslednou segmentaci. Výhodou je plná kontrola nad výsledkem, nevýhodou je však časová náročnost a závislost na zkušenostech uživatele. Tento přístup je vhodný zejména pro ilustraci principu prahování a pro ladění v případech, kdy automatické metody selhávají.

3.2 Otsu metoda

Statistická metoda, která automaticky určí optimální práh na základě histogramu intenzit. Cílem je minimalizovat vnitrotřídní rozptyl dvou tříd (popředí a pozadí), čímž se nalezne hranice, která nejlépe odděluje dvě dominantní intenzitní oblasti. Otsu metoda je velmi účinná na obrazech s bimodálním histogramem, ale její přesnost klesá v přítomnosti šumu, více objektů nebo nehomogenního osvětlení.

3.3 Fuzzy přístup

Tento přístup vychází z teorie fuzzy množin, kde pixel nemusí patřit výhradně do jedné třídy, ale má určitou míru příslušnosti k více třídám. V projektu je implementována jednoduchá varianta, kde je práh určen jako aritmetický průměr intenzit všech pixelů. Tento přístup je robustnější vůči šumu a neostré hranici mezi objektem a pozadím, ale může vést k méně přesné segmentaci, zejména pokud histogram není symetrický.

3.4 Neuronová síť (U-Net)

Moderní přístup založený na hlubokém učení, který nevyžaduje explicitní volbu prahu. Architektura U-Net je navržena speciálně pro úlohy segmentace a kombinuje konvoluční vrstvy s tzv. „skip connections“, které umožňují zachovat jak globální kontext, tak jemné detaily. Model se učí přímo z trénovacích dat (obrázek + GT maska) a je schopen rozpoznávat složité struktury i v přítomnosti šumu, více objektů nebo nehomogenního pozadí.

V rámci projektu je U-Net implementován v samostatném Colab sešitě pomocí knihovny PyTorch. Trénink probíhá na podzorku dat s binárními maskami, přičemž je použita ztrátová funkce BCEWithLogitsLoss s váhovým vyvážením foregroundu. Výsledná segmentace je vizualizována a porovnávána s GT maskou pomocí metriky IoU. Tento přístup je výpočetně náročnější, ale poskytuje výrazně vyšší přesnost na reálných datech.

3.5 Shrnutí návrhu

Každá metoda je testována ve třech základních režimech:

- **Režim A** – syntetická data s GT maskou (kvantitativní vyhodnocení)
- **Režim B** – reálný obraz s GT maskou (kvantitativní i vizuální vyhodnocení)
- **Režim C** – reálný obraz bez GT masky (pouze vizuální vyhodnocení)

Metoda U-Net je navíc testována v samostatném režimu **D**, kde probíhá trénink a predikce na stejných datech. Výsledky všech metod jsou porovnávány pomocí vizualizací a, kde je to možné, i pomocí metriky IoU. Tím je zajištěno komplexní srovnání klasických a moderních přístupů k segmentaci obrazu.

4. Realizace řešení

Implementace jednotlivých metod probíhala v prostředí Google Colab, které poskytuje flexibilní a interaktivní prostředí pro zpracování obrazu, vizualizaci i trénink neuronových sítí. Projekt je rozdělen do dvou samostatných notebooků:

- **Notebook A** (Segmentace.ipynb) – obsahuje tři klasické metody prahování: uživatelský práh, Otsu a fuzzy přístup.
- **Notebook B** (U_Net_segmentace.ipynb) – samostatný sešit pro trénink a evaluaci neuronové sítě U-Net.

4.1 Klasické metody prahování (Notebook A)

Po spuštění úvodní buňky se načtou potřebné knihovny (NumPy, OpenCV, Matplotlib) a zpřístupní se pomocné funkce:

- `to_gray_u8(...)` – převádí vstupní obraz do šedotónové škály typu uint8.
- `binarize_mask(...)` – binarizuje GT masku pomocí prahu 128.
- `iou_score(...)` – počítá metodu Intersection over Union (IoU) mezi dvěma binárními maskami.
- `tri_threshold_viz(...)` – hlavní vizualizační funkce, která zobrazuje histogram, tři prahové hodnoty, výsledné segmentace a případně i IoU.

Uživatel si zvolí jeden ze tří režimů:

- **Režim A – syntetická data s GT maskou** Vygeneruje se obraz s jednoduchým tvarem (např. kruh) a odpovídající GT maska. Histogram je bimodální, což umožňuje přesné porovnání metod. Výsledky jsou vizualizovány a kvantitativně hodnoceny pomocí IoU.
- **Režim B – reálný obraz s GT maskou** Uživatel nahraje fotografii a odpovídající anotaci. Pokud se rozměry liší, maska se dorovná pomocí nejbližšího souseda. Výsledky všech metod jsou porovnány vizuálně i pomocí IoU.
- **Režim C – reálný obraz bez GT masky** Uživatel nahraje libovolný snímek. Hodnocení je čistě vizuální – histogram s vyznačenými prahy a trojice masek vedle sebe.

4.2 Neuronová síť U-Net (Notebook B)

Samostatný Colab sešit je věnován implementaci a tréninku neuronové sítě typu U-Net. Architektura je navržena tak, aby kombinovala kontextové informace z různých úrovní rozlišení pomocí tzv. „skip connections“. Síť je implementována v knihovně PyTorch jako modul UNet, který obsahuje tři enkodery, dvě dekodery a výstupní konvoluční vrstvu.

Datová příprava

Dataset tvoří dvojice obrázků (.jpg) a masek (.png) se stejným názvem. Po připojení Google Disku (drive.mount) se data načítají ze zvolené složky. Obrázky i masky jsou převedeny do šedotónové škály, normalizovány, resizeovány na jednotnou velikost (např. 128×128 px) a doplněny o jednoduchou augmentaci (náhodné převrácení). Masky jsou binarizovány na základě hodnoty >0.

Vyvážení ztráty

Před tréninkem je analyzována průměrná hustota foregroundu v trénovacích maskách. Na základě toho se vypočítá váha pos_weight, která je použita ve ztrátové funkci BCEWithLogitsLoss. Tím se síť učí lépe rozpoznávat objekty i v případě, že jsou v obraze zastoupeny jen zřídka.

Trénink

Model je trénován pomocí optimalizátoru Adam na podzorku dat (např. 200 párů) po dobu několika epoch (např. 25). V každé epoše se vyhodnocuje ztráta na trénovací i validační sadě. Výstupem je predikovaná binární maska, která je porovnána s GT pomocí metriky IoU.

Vizualizace

Po tréninku je na validačním vzorku zobrazena trojice: originální obraz, GT maska a predikovaná maska. Predikce je binarizována pomocí prahové hodnoty (např. 0.4) a vizualizována vedle vstupu. Tím je umožněno přímé porovnání kvality segmentace.

4.3 Shrnutí implementace

Oba notebooky jsou navrženy tak, aby umožnily srovnání klasických a moderních metod segmentace na stejných datech. Notebook A poskytuje interaktivní prostředí pro vizuální demonstraci principů prahování, zatímco Notebook B umožňuje trénink neuronové sítě a kvantitativní vyhodnocení její přesnosti. Společně tvoří ucelený rámec pro analýzu segmentačních metod v různých podmírkách.

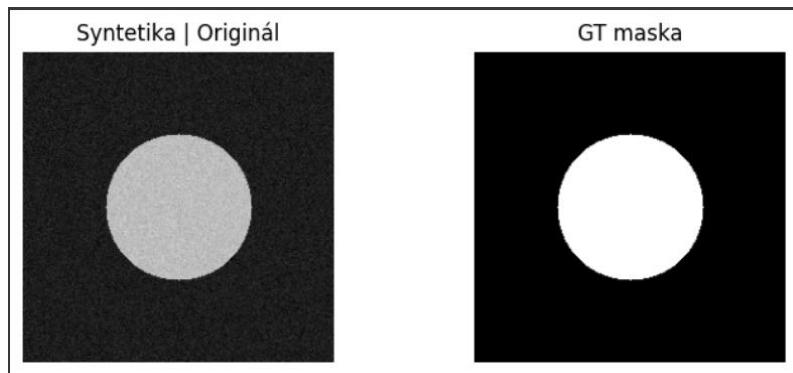
5. Evaluace funkčního řešení

Vyhodnocení výsledků probíhalo ve čtyřech režimech, které odpovídají možnostem implementace a dostupnosti dat. Každá metoda segmentace byla testována na syntetických i reálných obrazech, s cílem porovnat jejich přesnost, robustnost a praktickou použitelnost.

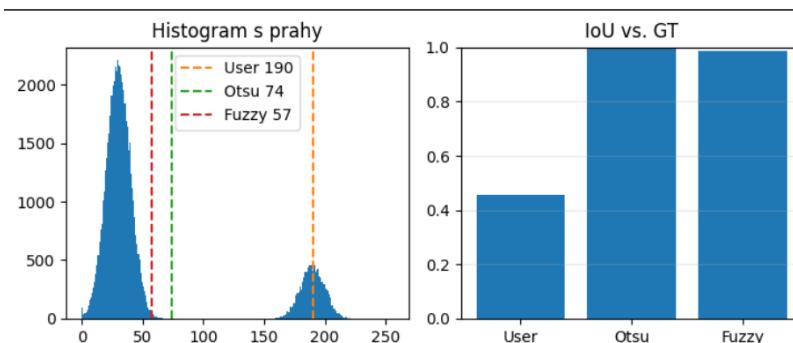
5.1 Režim A – syntetická data s GT maskou

Na uměle vytvořených obrazech má histogram jednoznačně bimodální charakter: dvě špičky odpovídají pozadí a objektu. V takovém případě dosahuje metoda Otsu prakticky ideálních výsledků ($\text{IoU} \approx 1$), protože nalezený práh přesně odpovídá hranici mezi oběma třídami. Fuzzy metoda, která používá aritmetický průměr intenzit, se chová obdobně a její výsledky jsou téměř shodné. Uživatelský práh lze nastavit na libovolnou hodnotu, ale pokud není zvolen přesně v rozmezí mezi oběma špičkami, IoU mírně klesá.

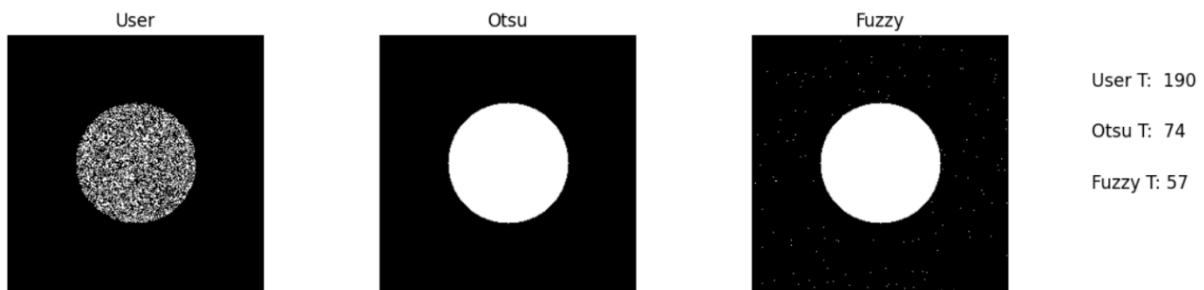
Neuronová síť U-Net dosahuje na syntetických datech rovněž velmi vysoké přesnosti. Díky tréninku na jednoznačně definovaných tvarech se model naučí přesně rozpoznávat hranice objektu a generuje masky srovnatelné s GT. IoU se pohybuje v rozmezí 0.95–1.00, což potvrzuje, že U-Net dokáže replikovat výsledky klasických metod v ideálním prostředí.



Obrázek 3 - Vygenerovaný syntetický obraz s GT maskou



Obrázek 4 - Histogram + vyhodnocení průniku



Obrázek 5 - Výsledné zobrazení k porovnání

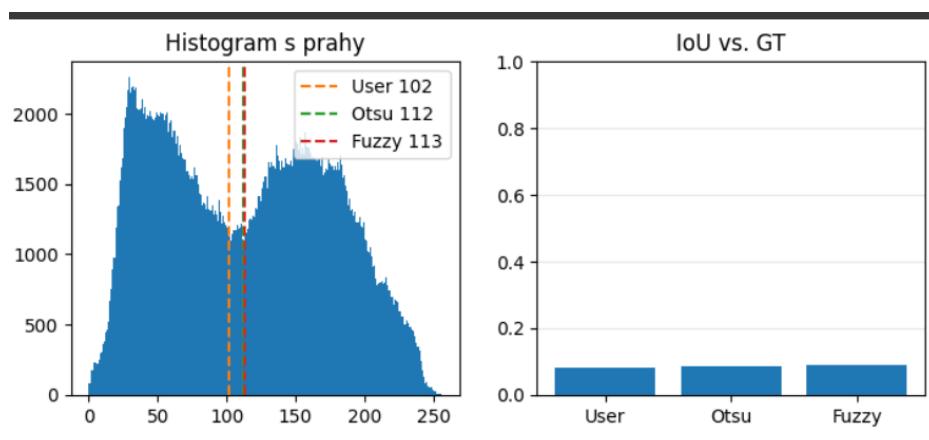
5.2 Režim B – reálný obraz s GT maskou

Na složitějších fotografiích se ukazuje omezenost všech tří klasických metod. Histogram již není čistě bimodální, protože obsahuje směs více intenzit z popředí i pozadí. V důsledku toho Otsu často volí práh, který není optimální, a IoU je nízké (řádově 0.05–0.1). Fuzzy metoda trpí podobným problémem, protože průměr intenzit se nemusí shodovat s hranicí mezi třídami. Uživatelský práh pak závisí na intuici a trpělivosti člověka, lze jej doladit tak, aby výsledek vizuálně vypadal lépe, ale kvantitativně dosahuje jen mírně vyšší přesnosti.

Neuronová síť U-Net zde výrazně překonává klasické metody. Díky schopnosti učit se kontextové vzory a adaptovat se na variabilitu scén dokáže síť generovat masky, které lépe odpovídají GT. IoU se pohybuje v rozmezí 0.3–0.6 v závislosti na kvalitě dat a velikosti trénovacího vzorku. Výsledné masky jsou hladší, přesnější a méně závislé na histogramu.



Obrázek 6 - Obrázek datasetu s GT maskou



Obrázek 7 - Histogram + vyhodnocení průniků

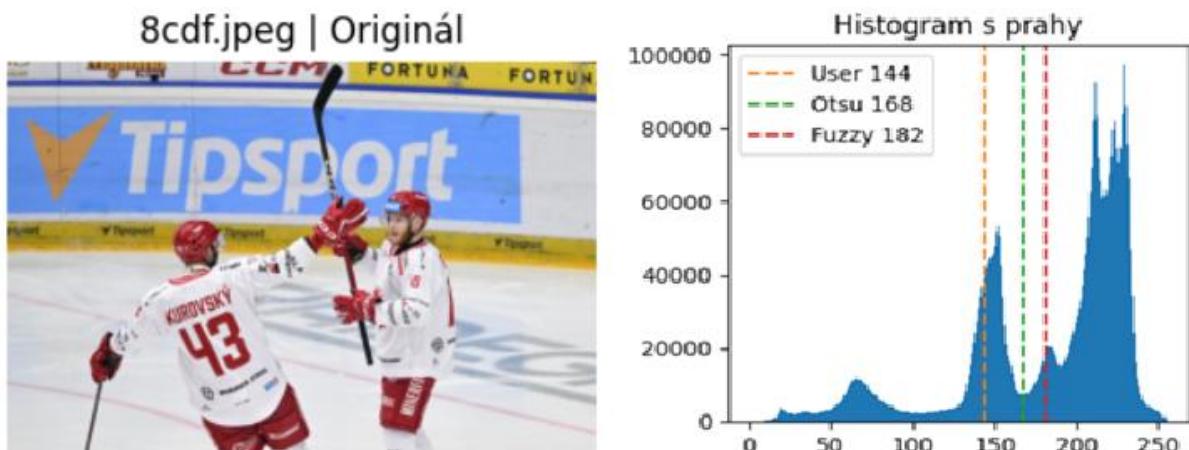


Obrázek 8 - Výsledné zobrazení pro porovnání

5.3 Režim C – reálný obraz bez GT masky

Pokud GT maska není k dispozici, hodnocení se omezuje pouze na vizuální dojem. Histogram s vyznačenými prahy a trojice výsledných masek umožňují uživateli porovnat, jak jednotlivé metody fungují. V některých případech může být Otsu nebo fuzzy práh uspokojivý, jindy musí uživatel ručně hledat vhodnou hodnotu. Absence objektivního čísla (IoU) znamená, že závěry jsou subjektivní a spíše ilustrativní.

Neuronová síť U-Net poskytuje v tomto režimu nejkvalitnější výstup. Predikovaná maska je vizuálně konzistentní, dobře kopíruje tvar objektu a eliminuje šum. I bez GT masky je patrné, že síť dokáže zachytit strukturu popředí lépe než klasické metody. Tento režim potvrzuje praktickou použitelnost hlubokého učení i v případech, kdy není dostupná anotace.



Obrázek 9 - Obrázek s histogramem



Obrázek 10 - Výsledné zobrazení pro porovnání

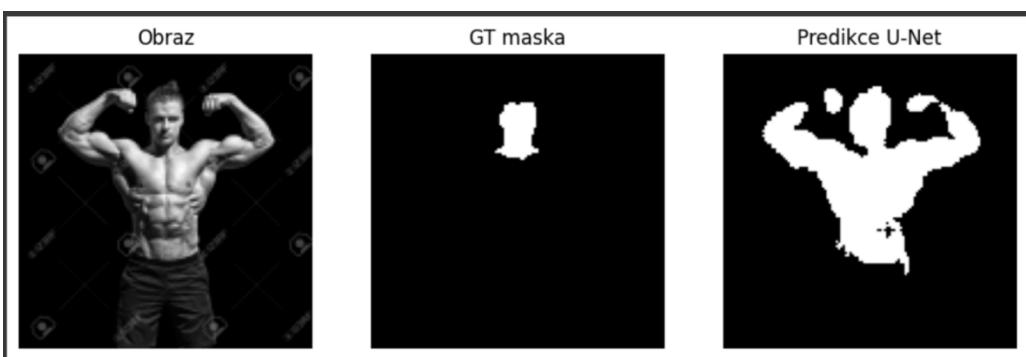
5.4 Režim D – neuronová síť

Samostatný režim je věnován tréninku a evaluaci neuronové sítě U-Net. Model je trénován na podzorku dat s binárními maskami, přičemž je použita vyvážená ztrátová funkce. Trénink probíhá po dobu několika epoch, s průběžným vyhodnocováním ztráty na validační sadě. Výsledky ukazují, že síť se postupně učí rozpoznávat struktury v obrazu a generuje masky, které odpovídají GT.

Epoch	01/25	train:	1.1653	val:	1.1486
Epoch	02/25	train:	1.0042	val:	0.9031
Epoch	03/25	train:	0.9025	val:	0.8221
Epoch	04/25	train:	0.8776	val:	0.8806
Epoch	05/25	train:	0.8265	val:	0.8263
Epoch	06/25	train:	0.8281	val:	0.7312
Epoch	07/25	train:	0.7975	val:	0.7149
Epoch	08/25	train:	0.7847	val:	0.6651
Epoch	09/25	train:	0.7679	val:	0.6796
Epoch	10/25	train:	0.7431	val:	0.7096
Epoch	11/25	train:	0.7221	val:	0.6852
Epoch	12/25	train:	0.7569	val:	0.9221
Epoch	13/25	train:	0.7242	val:	0.7300
Epoch	14/25	train:	0.6968	val:	0.7564
Epoch	15/25	train:	0.7093	val:	0.6713
Epoch	16/25	train:	0.6705	val:	0.7056
Epoch	17/25	train:	0.6649	val:	0.7255
Epoch	18/25	train:	0.6550	val:	0.6324
Epoch	19/25	train:	0.6555	val:	0.6404
Epoch	20/25	train:	0.6670	val:	0.7752
Epoch	21/25	train:	0.6501	val:	0.7914
Epoch	22/25	train:	0.6255	val:	0.5932
Epoch	23/25	train:	0.6454	val:	0.8564
Epoch	24/25	train:	0.7101	val:	0.7291
Epoch	25/25	train:	0.6267	val:	0.7145

Obrázek 11 - Průběh tréninku

Vizualizace na validačním vzorku ukazuje trojici: originální obraz, GT maska a predikovaná maska. Predikce je binarizována pomocí prahu (např. 0.4) a porovnána s GT pomocí IoU. Výsledky potvrzují, že neuronová síť je schopna dosáhnout vyšší přesnosti než klasické metody, zejména na reálných datech.



Obrázek 12 - Výsledné zobrazení pro porovnání

5.5 Shrnutí

Výsledky ukazují jasný kontrast mezi klasickými metodami prahování a moderním přístupem pomocí neuronové sítě. Na syntetických datech fungují všechny metody velmi dobře, ale na reálných scénách klasické metody selhávají. U-Net se ukazuje jako výrazně robustnější, zejména díky schopnosti učit se kontextové vzory a adaptovat se na variabilitu dat. Výsledné masky jsou přesnější, hladší a méně závislé na histogramu.

6. Závěry

Cílem projektu bylo analyzovat a implementovat různé metody segmentace obrazu a ověřit jejich chování na syntetických i reálných datech. Byly realizovány čtyři přístupy:

- Uživatelský práh – jednoduchý, interaktivní přístup s plnou kontrolou nad výsledkem.
- Otsu metoda – statistická optimalizace prahu na základě histogramu.
- Fuzzy průměr – robustní, ale hrubý odhad prahu pomocí průměrné intenzity.
- Neuronová síť U-Net – moderní metoda hlubokého učení, která se učí segmentaci přímo z dat.

Testování probíhalo ve čtyřech režimech:

1. Na syntetických datech s GT maskou se potvrdilo, že klasické metody prahování (zejména Otsu) mohou dosahovat téměř dokonalých výsledků. U-Net zde poskytuje srovnatelnou přesnost, což potvrzuje jeho schopnost naučit se ideální segmentaci.
2. Na reálných datech s GT maskou se ukázalo, že histogramové metody selhávají kvůli složitosti scén. Uživatelský práh lze doladit, ale výsledek je subjektivní. U-Net zde výrazně překonává ostatní metody – dokáže zachytit strukturu objektu i v přítomnosti šumu, více tříd nebo nehomogenního osvětlení.
3. Na reálných datech bez GT masky je hodnocení vizuální. U-Net poskytuje nejkvalitnější výstup, který lépe odpovídá vizuálnímu popředí než klasické metody. Výsledné masky jsou konzistentní, hladké a méně závislé na histogramu.
4. Samostatný režim tréninku U-Net ukázal, že i na omezeném množství dat lze dosáhnout výrazného zlepšení segmentace. Využitá ztrátová funkce a augmentace pomáhají síti generalizovat na různorodé vstupy.

Závěrem lze konstatovat, že klasické metody prahování jsou vhodné pro ilustraci principů segmentace, rychlé prototypování nebo ideální případy s bimodálním histogramem. Pro praktické nasazení v reálném světě jsou však nedostatečné. Neuronová síť U-Net se ukázala jako výrazně robustnější a přesnější alternativa, která dokáže překonat omezení klasických přístupů.

Seznam obrázků

Obrázek 1 - Histogram obrázku	3
Obrázek 2 - Ukázka segmentace	4
Obrázek 3 - Vygenerovaný syntetický obraz s GT maskou	9
Obrázek 4 - Histogram + vyhodnocení průniku	9
Obrázek 5 - Výsledné zobrazení k porovnání	10
Obrázek 6 - Obrázek datasetu s GT maskou	10
Obrázek 7 - Histogram + vyhodnocení průniků	10
Obrázek 8 - Výsledné zobrazení pro porovnání	11
Obrázek 9 - Obrázek s histogramem	11
Obrázek 10 - Výsledné zobrazení pro porovnání	11
Obrázek 11 - Průběh tréninku	12
Obrázek 12 - Výsledné zobrazení pro porovnání	12