

Vizualizace a hodnocení augmentačních technik v počítačovém vidění

Bc. Petr Buček



*** zadání práce (strana 1) / thesis assignment (page 1) ***

PROHLÁŠENÍ AUTORA DIPLOMOVÉ PRÁCE

Beru na vědomí, že

- odevzdáním diplomové práce souhlasím se zveřejněním své práce podle zákona č. 111/1998 Sb., v platném znění bez ohledu na výsledek obhajoby;
- diplomová práce bude uložena v elektronické podobě v univerzitním informačním systému a bude dostupná k nahlédnutí;
- jedno vyhotovení diplomové práce v listinné podobě bude ponecháno Univerzitě Tomáše Bati ve Zlíně k uložení;
- na moji diplomovou práci se plně vztahuje zákon č. 121/2000 Sb. o právu autorském, o právech souvisejících s právem autorským a o změně některých zákonů (autorský zákon) ve znění pozdějších právních předpisů, zejm. § 35 odst. 3;
- podle § 60 odst. 1 autorského zákona má Univerzita Tomáše Bati ve Zlíně právo na uzavření licenční smlouvy o užití školního díla v rozsahu § 12 odst. 4 autorského zákona;
- podle § 60 odst. 2 a 3 mohu užít své dílo – diplomovou práci – nebo poskytnout licenci k jejímu využití jen s předchozím písemným souhlasem Univerzity Tomáše Bati ve Zlíně, která je oprávněna v takovém případě ode mne požadovat přiměřený příspěvek na úhradu nákladů, které byly Univerzitou Tomáše Bati ve Zlíně na vytvoření díla vynaloženy (až do jejich skutečné výše);
- pokud bylo k vypracování diplomové práce využito softwaru poskytnutého Univerzitou Tomáše Bati ve Zlíně nebo jinými subjekty pouze ke studijním a výzkumným účelům (tj. k nekomerčnímu využití), nelze výsledky diplomové práce využít ke komerčním účelům;
- pokud je výstupem diplomové práce jakýkoliv softwarový produkt, považují se za součást práce rovněž i zdrojové kódy, popř. soubory, ze kterých se projekt skládá; neodevzdání této součásti může být důvodem k neobhájení práce.

Prohlašuji, že

- jsem na diplomové práci pracoval(a) samostatně a použitou literaturu jsem řádně citoval(a); v případě publikace výsledků budu uveden(a) jako spoluautor;
- odevzdaná verze diplomové práce a verze elektronická nahraná do IS/STAG jsou obsahově totožné.

Ve Zlíně, dne

.....
podpis autora

ABSTRAKT

Text abstraktu v jazyce práce.

Klíčová slova: klíčové slovo, klíčové slovo

ABSTRACT

Text abstraktu ve světovém jazyce (angličtině).

Keywords: keyword, keyword

Zde je místo pro případné poděkování, popř. motto, úryvky knih atp.

OBSAH

ÚVOD	11
I TEORETICKÁ ČÁST	12
1 AUGMENTACE DAT V POČÍTAČOVÉM VIDĚNÍ	13
1.1 PROBLEMATIKA KVALITY A MNOŽSTVÍ DAT V POČÍTAČOVÉM VIDĚNÍ	13
1.2 ÚLOHA AUGMENTACE DAT PŘI ZLEPŠOVÁNÍ VÝKONU MODELŮ HLUBO- KÉHO UČENÍ	14
1.3 KLASIFIKACE A CHARAKTERISTIKA AUGMENTAČNÍCH METOD	14
1.3.1 Geometrické transformace (rotace, škálování, ořez, zrcadlení)	14
1.3.2 Intenzivní transformace (změna jasu, kontrastu, aplikace šumu)	15
1.3.3 Pokročilé augmentační metody (Cutout, Mixup, CutMix, GAN-based augmentace)	16
1.3.4 Doménově specifické augmentační techniky	17
2 VIZUALIZACE DOPADU AUGMENTACÍ	18
2.1 PRINCIP METODY T-SNE A JEJÍ MATEMATICKÝ ZÁKLAD	18
2.2 APLIKACE T-SNE V KONTEXTU ANALÝZY AUGMENTOVANÝCH DAT	18
2.3 MOŽNOSTI INTERPRETACE VÝSLEDKŮ T-SNE VIZUALIZACE	18
3 HODNOCENÍ VLIVU AUGMENTACÍ NA DIVERZITU DAT	19
3.1 METRIKY PRO KVANTIFIKACI DIVERZITY A ROZMANITOSTI DAT	19
3.2 ANALÝZA VLIVU JEDNOTLIVÝCH AUGMENTAČNÍCH TECHNIK NA DIS- TRIBUCI DAT	22
3.3 KOMPARATIVNÍ HODNOCENÍ EFEKTIVITY RŮZNÝCH PŘÍSTUPŮ K AUG- MENTACI	22
II PRAKTICKÁ ČÁST	23
ZÁVĚR	24
SEZNAM POUŽITÉ LITERATURY	25
SEZNAM OBRÁZKŮ	27
SEZNAM TABULEK	28

SEZNAM POUŽITÝCH SYMBOLŮ A ZKRATEK	29
SEZNAM PŘÍLOH.....	30

ÚVOD

První řádek prvního odstavce v kapitole či podkapitole se neodsazuje, ostatní ano. Vertikální odsazení mezi odstavci je typické pro anglickou sazbu; czech babel toto respektuje, netřeba do textu přidávat jakékoliv explicitní formátování, následuje ukázka sazby tohoto a následujících odstavců). Je vhodné znovu připomenout, že autor práce by si měl pohlídat, čím končí řádky. Určitě by to neměly být jednohláskové spojky ani předložky. Pro takové případy zde máme tildu (~), kterou L^AT_EX vyhodnocuje jako nezlomitelnou mezeru. Příklad použití, ze kterého se můžete poučit, je uveden např. právě na konci tohoto řádku. Konec řádku zde a další řádek hned za ním, aneb podívejte se do zdrojového kódu za spojku „a“.

Šablona je nastavena na oboustranný tisk. Za tímto účelem obsahuje na svém počátku místy prázdnou stranu, aby např. zadání začínalo na liché stránce (vpravo).

I TEORETICKÁ ČÁST

1 AUGMENTACE DAT V POČÍTAČOVÉM VIDĚNÍ

Velké pokroky v posledních letech v oblastech strojového učení a počítačového vidění nám odhalily nové překážky na vyřešení. Zejména efektivita počítačové vidění závisí na třech složkách: schopný hardware, kvalitní algoritmy a početné a rozmanité datové sady. Zatímco jsme byly svědky toho jak efektivita a výkon hardwaru roste a algoritmy se stávají sofistikovanější, tak sběr kvalitních dat je stále stejně drahý a náročný proces. Tento problém se snaží řešit právě augmentace dat. Augmentace dat zaštiťuje soubor metod, které upravují nebo rozšiřují trénovací datové sady a tím zajišťují robustnost a přesnost modelů. [1] Tato kapitola se bude zabývat přehledem augmentačních technik v oblasti počítačového vidění. V následujících sekcích jsou analyzovány jednotlivé aspekty augmentací datových sad počínaje problematikou kvality datových sad, přes samotnou úlohou augmentací ve zlepšování modelů až po jejich praktické aplikace.

1.1 Problematika kvality a množství dat v počítačovém vidění

V disciplíně počítačového vidění je kvalita a velký objem rozmanitých trénovacích dat stěžejním faktorem, který výrazně ovlivňuje úspěšnost modelů. Ovšem reálné datové sady jsou ve skutečnosti nekvalitní, neúplné a často nevyvážené a vedou tedy k nepřesnosti a preferenci určitých tříd nad ostatními.

Kvalitou dat rozumíme několik aspektů, kde prvním z nich je rozmanitost zachycené scény, osvětlení, pozic a dalších různých proměnných, které odráží dynamiku reálného světa. Pokud je model trénován na omezeném spektru této rozmanitosti, tak je jeho efektivita v neobvyklých situacích častokrát nedostačující. To je způsobeno dynamickou povahou reálného světa, kdy se světelné podmínky, úhly pohledu, pozice předmětů, rotace kamery a další aspekty v čase mění. V ideálním případě by měla trénovací datová sada všechny tyto skutečnosti odrážet, ovšem v realitě je to v některých případech nemožné. Dalším důležitým aspektem je přesnost anotací popisů scény, jelikož s nepřesnými nebo chybnými vstupy je supervizované strojové učení předem odsouzeno k nezdaru. Další je aspekt rovnoměrnosti jednotlivých tříd datové sady, zajišťující, že model nebude držet preferenci k určitým třídám. Posledním aspektem je pak samotná velikost trénovacích dat. Hluboké neuronové sítě vyžadují nepřeborné množství dat pro efektivní natrénování s velkou přesností.

1.2 Úloha augmentace dat při zlepšování výkonu modelů hlubokého učení

Jak již bylo nastíněno augmentace dat nabízí řešení pro mnoho výzev spojených s množstvím a kvalitou trénovacích datových sad. Největší přínos spočívá v umělém rozšíření vzorů, změnou domény, nebo zvýšení rozmanitosti bez nutnosti dalšího sběru a anotace dat. Tím augmentace zaštiťují efektivní způsob navýšení velikosti a rozmanitosti trénovacích dat, což vede k lepšímu výsledku modelu.

1.3 Klasifikace a charakteristika augmentačních metod

Klasifikovat augmentační metody lze různými způsoby, nejčastěji je lze klasifikovat podle typu transformací prováděnými nad daty. V této podkapitole jsou popsány hlavní kategorie augmentačních metod. Dále jsou zmíněny jejich charakteristiky, omezení a praktické využití v oblasti trénování počítačového vidění.

1.3.1 Geometrické transformace (rotace, škálování, ořez, zrcadlení)

Tyto transformace znázorňují nejjednodušší možné metody augmentace dat. Modifikují prostorové uspořádání obrazu, ale přitom zachovávají jeho obsah. Mezi nejpoužívanější techniky patří:

- **Rotace:** Otáčení obrazu o daný úhel, čímž zvyšujeme odolnost modelu vůči změně orientaci objektů v prostoru.
- **Škálování:** Změna velikosti obrazu v dané ose o daný násobek, který pomáhá modelu rozpoznávat různě vzdálené objekty, nebo různé velikosti stejných objektů.
- **Ořez:** Ořezání dané části obrazu, tedy zobrazení specifické části obrazu, simuluje částečně viditelné objekty.
- **Zrcadlení:** Horizontální nebo vertikální převrácení obrazu, obecně zvyšuje rozmanitost dat.

Tyto metody jsou snadné na implementaci, jsou pro lidi intuitivní a výpočetně nenáročné. Intuitivnost vychází z faktu, že tyto transformace odpovídají reálným změnám, ke kterým dochází při pořizování obrazu, jako jsou různé úhly pohledu, vzdálenost a orientace kamery a pozice objektů v prostoru. Tyto transformace zachovávají obsah obrazu, tedy transformovaný

obraz stále patří do stejné třídy jako ten původní. Ovšem některé transformace mohou vést ke ztrátě informací jako je to například u ořezu. Omezení takových augmentačních metod pramení z faktu, že nepomohou modelu porozumět změnám negeometrické povahy, jako jsou například změny osvětlení.

1.3.2 Intenzivní transformace (změna jasu, kontrastu, aplikace šumu)

Tyto transformační techniky jsou stěžejní pro zlepšování odolnosti modelů počítačového vidění vůči různým světelným podmínkám, nebo různorodé kvality v obrazových datových sadách. Tato odolnost modelu je důležitá zejména tak, protože reálná obrazová data jsou přímo ovlivněna světelnými podmínkami, ve kterých byly zachyceny, nebo také kvalitou zařízení jež je zachytilo. Tyto transformační techniky jsou často jednoduše implementovatelné, přičemž ale nabízí zvýšení přesnosti modelu v reálných podmínkách s horší kvalitou vstupních dat.

- **Změna jasu:** Úprava jasu je na datovou sadu aplikována tak, že se modifikuje intenzita pixelů v obraze, tedy nezmění se jeho geometrické vlastnosti. Tato úprava simuluje reálné odchylky ve světelných podmínkách a modely, které jsou trénované na takových datových sadách jsou lepší v rozpoznávání objektů, které například byly zachyceny v jinou denní dobu, než data původní. Bez takových augmentací by se mohlo stát, že model by byl přetrénovaný při specifických světelných podmínkách, například by mohl správně určit objekt pouze při denním světle. [2]
- **Změna kontrastu:** Úprava kontrastu je na datovou sadu aplikována tak, že mapuje hodnoty intenzity obrazu na nový rozsah. Tím se mění rozdíl mezi nejtmašími a nejsvětějšími částmi obrazu. [3] Model natrénovaný na datové sadě augmentované touto metodou je lepší v rozlišování důležitých rysů při různých kontrastních podmínkách.
- **Aplikace šumu:** Tato augmentační technika záměrně zanáší do obrazových dat různé typy šumu, jako například Gaussův šum, Poissonův šum, nebo tzv. salt-and-pepper šum. Cílem je obrnit daný model vůči drobným nepřesnostem a poruchám v obrazových datech. Šum se totiž velmi často vyskytuje v reálných obrazových datech, většinou vlivem špatných světelných podmínek, nebo kvalitou zachytávajícího zařízení.

1.3.3 Pokročilé augmentační metody (Cutout, Mixup, CutMix, GAN-based augmentace)

Pokročilé augmentační metody jsou robustnější a sofistikovanější metody augmentace dat. Tyto metody vznikly s cílem zlepšit generalizaci modelů a snížit riziko přetrénování.

- **Cutout:** Tato metoda náhodně zakrývá části obrazových dat. Potřeba těchto augmentací pramení z faktu, že v reálných obrazových datech zachycené objekty mohou být částečně skryty. Podle zdroje [4] model naučený na takto augmentovaných datech bere více v potaz kontext obrazu při rozhodování. Tato metoda může pomoci proti přílišné fixaci modelu na nepodstatné prvky. Například model, který je trénovaný na medicínských snímcích se může lehce přetrénovat na artefakty zanechané samotným zařízením, jako je například značka skeneru někde na snímku, ale tato augmentační metoda mu pomůže se zaměřit na důležitý obsah snímku, tím že tyto artefakty zakryje [5].
- **Mixup:** Je metodou, která se snaží o minimalizaci okolního rizika, tedy generuje nové vzorky v blízkosti původních dat, spojením dvou obrazových dat. Tedy nevytváří nové varianty původních dat, ale vytváří zcela nové trénovací data kombinací existujících dat. Tyto nová data jsou vytvořena například pomocí lineární interpolací dvou obrazových dat a jejich odpovídajících štítků. Tato metoda může vést k lepší generalizaci a robustnosti modelu [6].
- **CutMix:** Kombinuje obě dříve zmíněné augmentační metody v jednu. Jedná se tedy o kombinaci metody Cutout a Mixup. Princip této metody spočívá v tom, že části obrazových dat jsou vyřezávány a poté jsou zase vkládány mezi trénovací obrazová data, přičemž štítky jsou také míchány proporcionálně k ploše a poloze vyřezaných oblastí.
- **GAN-based augmentace:** Představuje augmentační metodu, která umí vytvářet zcela nová realistická data na základě distribuce trénovacích dat. Toho dosahují pomocí generativních adversiálních sítí (GAN). GAN je kombinací dvou neuronových sítí určené ke generování nových dat, která se snaží co nejvíce napodobit vlastnosti trénovacích vzorků. Jedná se dvě neuronové sítě: generátor a diskriminátor, které spolu soupeří v tzv. adversiálním tréninku. Generátor se snaží vygenerovat výstupní nové vzorky dat, které budou co nejvíce podobné původním datům z trénovací množiny. Diskriminátor se poté snaží rozeznat nově generovaná data od těch původních trénovacích. Tím

je zajištěno střídavé zlepšování těchto dvou neuronových sítí a vznikne tak generativní model, který dokáže s velkou přesností napodobit data z trénovací množiny [7]. Ve srovnání s ostatními metodami augmentací dat, tato metoda poskytuje robustnější obohacení datové sady zcela novými vzorky dat, namísto pouhých transformací dat původních.

1.3.4 Doménově specifické augmentační techniky

Doménově specifické augmentační techniky představují soubor pokročilých metod obohacování datových sad, které jsou uzpůsobeny konkrétním charakteristikám dané aplikační oblasti počítačového vidění. Tyto metody jsou využívány zejména v kontextu vysoce specializovaných typů úloh. Například v oblasti zdravotnictví jsou tyto metody často využívány zejména proto, že reálné medicínské snímky jsou často v počtu omezené, zvláště tak u chorob, které jsou vzácné. Doménově specifické augmentace v medicínském snímkování musí zachovat relevantní informace pro diagnózu [8]. Dále když navštívíme oblast autonomních vozidel, zde jsou doménově specifické augmentace klíčové pro zajištění dobrého výkonu modelu v různých podmínkách. Aby bylo zajištěno, že model dokáže správně fungovat za různých světelných podmínek a za různého počasí, jsou použity specializované augmentační metody, které musí umět tyto podmínky nasimulovat a napasovat na původní data. Tyto metody lze charakterizovat těmito klíčovými vlastnostmi:

- Zachovávají důležité informace specifické pro danou doménu
- Využívají znalostí struktury a sémantiky dat v dané doméně
- Simulují realistické variace specifické pro danou doménu

Tímto způsobem doménově specifické augmentační techniky nabízejí mocný nástroj na efektivní tvorbu nových dat, které mohou být obtížné na získání a dokáží natrénovat modely, které excelují na specifických typech úloh.

2 VIZUALIZACE DOPADU AUGMENTACÍ

2.1 Princip metody t-SNE a její matematický základ

Vizualizace dat je klíčový prvek k pochopení a interpretaci jak augmentace upravují původní datovou sadu. Tato kapitola se zaměří na metodu t-SNE (t-Distributed Stochastic Neighbor Embedding), která slouží k vizualizaci vysokodimenzionálních dat ve 2D nebo 3D prostoru, při zachování shluků ve vysokých dimenzích. Tím lze získat jakousi intuici, jak jsou data uspořádána ve vyšších dimenzích.

Základ této metody vychází z metody SNE, celým názvem Stochastic Neighbor Embedding, která převádí vysokodimenzionální euklidovské vzdálenosti datových bodů do podmíněných pravděpodobností představujících jejich podobnost. Podmíněná pravděpodobnost má tento matematický vzorec:

$$p_{j|i} = \frac{\exp\left(-\frac{\|x_i - x_j\|^2}{2\sigma_i^2}\right)}{\sum_{k \neq i} \exp\left(-\frac{\|x_i - x_k\|^2}{2\sigma_i^2}\right)}$$

Podmíněná pravděpodobnost $p_{j|i}$ přímo odpovídá podobnosti bodu x_i a bodu x_j . σ_i značí Gaussovo rozdělení, které je vycentrováno nad bodem x_i . Ve své podstatě $p_{j|i}$ určuje, zda by si bod x_i vybral bod x_j za svého souseda, pokud by sousedé byly vybírání podle proporce pravděpodobností hustoty Gaussova rozdělení σ_i . Pro datové body, které jsou blízko sebe nabývá $p_{j|i}$ relativně vysokých hodnot, zatímco body, které jsou od sebe daleko nabývá hodnot téměř nulových. [9]

Metoda SNE

2.2 Aplikace t-SNE v kontextu analýzy augmentovaných dat

2.3 Možnosti interpretace výsledků t-SNE vizualizace

3 HODNOCENÍ VLIVU AUGMENTACÍ NA DIVERZITU DAT

Tato kapitola se zabývá hodnocením vlivu augmentačních technik na diverzitu dat pro strojové učení a počítačové vidění. Zahrnuje metriky pro kvantifikaci augmentací a popisuje jednotlivé metody.

3.1 Metriky pro kvantifikaci diverzity a rozmanitosti dat

Metriky posunu distribuce:

- **Afinita:** Tato metrika určuje míru posunutí distribuce augmentovaných dat od distribuce dat čistých, na kterých se model naučil. To nám poskytuje pohled na to jak se augmentovaná data odchylují od modelu naučeného na čistých datech. Afinita je definována jako rozdíl validačních přesností modelu naučeného na datech čistých a modelu naučeného na datech augmentovaných. Pokud afinita nabývá nulové hodnoty, mezi rozděleními není žádný posun. Záporná hodnota pak značí, že augmentovaná data se nacházejí mimo distribuční oblast očekávanou modelem. [10]

D_{train} a D_{val} označují trénovací a validační datové sady, které jsou nezávislé a identicky rozdělené dle stejné distribuční funkce. Dále a je stochastická augmentační strategie, přičemž D'_{val} je augmentovaná verze validační datové sady u které byla strategie a aplikována na každý vzorek z datové sady právě jednou:

$$D'_{\text{val}} = \{(a(x), y) \mid (x, y) \in D_{\text{val}}\}$$

Model m je natrénovaný na trénovací datové sadě D_{train} . Dále $A(m, D)$ značí přesnost modelu m na datové sadě D . Samotná metrika afinity $T[a; m; D_{\text{val}}]$ je potom definována jako rozdíl přesnosti modelu na čistých a augmentovaných datech:

$$T[a; m; D_{\text{val}}] = A(m, D'_{\text{val}}) - A(m, D_{\text{val}})$$

- **Diverzita:** Je metrikou, která kvantifikuje augmentaci z hlediska jak dokáže rozšiřovat trénovací datovou sadu. Diverzita reflektuje jakousi intuici, že komplexní nebo vícefaktorové augmentační strategie zvyšují počet vzorků v trénovací datové sadě, čímž snižují riziko přeučení modelu [10].

D'_{train} je trénovací datová sada, na kterou byla aplikována stochastická augmentační strategie a . Dále L_{train} značí trénovací ztrátovou funkci modelu m . Diverzita $D[a; m; D_{\text{train}}]$ je definována jako očekávaná hodnota trénovací ztráty:

$$D[a; m; D_{\text{train}}] := E_{D'_{\text{train}}} [L_{\text{train}}]$$

- **KL divergence** : Celým názvem Kullbackova-Leiblerova divergence, kvantifikuje míru rozdílnosti mezi dvěma pravděpodobnostními rozděleními. V kontextu augmentace dat pro počítačové vidění se používá na kontrolu změny rozdělení dat po augmentaci. Matematicky se značí jako $D_{\text{KL}}(P||Q)$ a nabývá hodnot $D_{\text{KL}}(P||Q) \geq 0$, přičemž když se KL divergence rovná nule, jsou obě rozdělení shodná.

- **P** je skutečné rozdělení (např. původní datová sada)
- **Q** je předpokládané rozdělení (např. augmentovaná datová sada)

Matematický vzorec:

$$D_{\text{KL}}(P||Q) = \sum_x P(x) \log \left(\frac{P(x)}{Q(x)} \right)$$

Metriky diverzity a pokrytí prostoru příznaků:

- **Rozptylová diverzita**: Anglicky "Variance diversity" je metrikou kvantifikující diverzitu augmentovaných dat. Dle zdroje [11] bylo prokázáno experimentem, že existuje korelace mezi touto metrikou a lepší generalizací modelu. Nízká hodnota rozptylové diverzity znamená, že augmentovaná data jsou hodně podobná původním datům, tedy mají nízkou diverzitu. Pokud je hodnota vysoká, znamená to, že data jsou dostatečně různorodá a mají potenciál vylepšit generalizaci modelu.

Pro daný vstup x_i je definována jako průměrný rozptyl embeddingů jednotlivých augmentací $a_j(x_i)$ vzhledem k jejich centroidu (průměrná reprezentace všech augmentací v prostoru příznaků):

$$VD(x_i) = \frac{1}{k-1} \sum_{j=1}^k |f_{\theta}(a_j(x_i)) - \mu_i|_2^2$$

Kde $f_{\theta}(\cdot)$ je model trénovaný na neaugmentovaných datech extrahující příznaky a μ_i je centroid.

- **Pokrytí prostoru příznaků:** Tato metrika, kvantifikuje jak dobře nová augmentovaná data zaplňují prostor příznaků. Pokrytí prostoru příznaků lze měřit různými způsoby. Jednou z variant je výpočet objemu konvexního obalu redukovaného prostoru příznaků, například po PCA transformaci, tím získáme informaci o tom, jak velkou část prostoru augmentovaná data pokrývají. Čím je větší objem, tím se zvětšuje i diverzita a pokrytí prostoru příznaků.

X je množina všech bodů v prostoru příznaků po redukci dimenze pomocí PCA algoritmu. Konvexní obal $\text{Conv}(X)$ je poté množina všech konvexních kombinací těchto bodů:

$$\text{Conv}(X) = \left\{ \sum_{i=1}^n \lambda_i x_i \mid \lambda_i \geq 0, \sum_{i=1}^n \lambda_i = 1 \right\}$$

Samotný objem konvexního obalu V je následovně:

$$V = \text{Vol}_d(\text{Conv}(X))$$

Kde Vol_d znázorňuje objem o dimenzi d , typicky pro dvě dimenze je to plocha a pro tři se jedná o objem.

- **Skóre posunu domény:** V oblasti počítačové vidění výkon modelu může dramaticky klesat, při střetnutí s reálnými daty, která jsou statisticky odlišná od dat, na kterých byl natrénován. Tento problém je znám jako problém posunu domény. Tato metrika kvantifikuje míru změny mezi původní a cílovou doménou. Může být měřena jako statistická vzdálenost mezi distribucemi původních a augmentovaných dat. Velmi vysoký posun může vést k špatné reprezentaci augmentovaných dat vůči těm původním, zatímco příliš malý posun může značit, že augmentace je neefektivní [12].

Metriky shlukování a homogenity dat

- **Čistota shluku:** Anglicky "cluster purity" je metrikou hodnotící, jak moc jsou shluky dat homogenní po aplikaci augmentačních metod. V oblasti augmentačních metod je totiž velice důležité, aby augmentovaná data náležely do stejné sémantické kategorie jako jejich původní neaugmentované verze. Díky této metrice je možné identifikovat případy, kdy augmentace vytváří vzorky překračující hranice samotných tříd a narušují

tak původní strukturu dat. Vysoká hodnota této metriky může značit, že augmentační metody zachovávají sémantickou integritu dat, zatímco malé hodnoty mohou naznačovat, že augmentace není vhodná.

- **Siluetové skóre:** Tato metrika analyzuje shluky augmentovaných dat a hodnotí jak dobře jsou nově augmentované vzorky seskupeny. Tato metrika je počítána pro každý vzorek a jedná se zejména o její kohezi a separaci. Koheze určuje míru podobnosti vzorku vůči svému vlastnímu shluku. Separace poté uvádí míru rozdílnosti vzorku vůči ostatním shlukům. Tím nám tato metrika může naznačit, jestli augmentační metody zachovávají nebo naopak narušují původní struktury v datech. Toto skóre může nabývat hodnot od -1 do 1, přičemž čím vyšší hodnota značí, že shluky jsou dobře definované, to může pomoci modelu lépe rozlišovat samotné třídy.

3.2 Analýza vlivu jednotlivých augmentačních technik na distribuci dat

Tato podkapitola se zaměří na to, jak augmentace ovlivňují diverzitu obrazových dat. Pochoopení těchto vlivů je klíčové pro správné vybrání vhodné augmentační strategie pro konkrétní scénáře v oblasti počítačového vidění, což je i dalším bodem této práce. Chápání těchto metrik je také důležité pro správnou interpretaci jejich dopadů na trénovací proces a následný výkon modelu počítačového vidění.

3.3 Komparativní hodnocení efektivity různých přístupů k augmentaci

II PRAKTICKÁ ČÁST

ZÁVĚR

Text ...

SEZNAM POUŽITÉ LITERATURY

- [1] XU, Mingle; YOON, Sook; FUENTES, Alvaro a PARK, Dong Sun, 2023. A Comprehensive Survey of Image Augmentation Techniques for Deep Learning. *Pattern Recognition*. 2023. 137, s. 109347. ISSN 0031-3203. Dostupné z: <https://www.sciencedirect.com/science/article/pii/S0031320323000481>.
- [2] MILUVS, 2025. What is the impact of brightness adjustment in data augmentation? Dostupné z: <https://milvus.io/ai-quick-reference/what-is-the-impact-of-brightness-adjustment-in-data-augmentation>.
- [3] UNIVERSITY, Northwestern, 2025. Analyzing and Enhancing Images (Image Processing Toolbox). Dostupné z: <http://www.ece.northwestern.edu/local-apps/matlabhelp/toolbox/images/enhanc12.html>.
- [4] DEVRIES, Terrance a TAYLOR, Graham W., 2017. Improved Regularization of Convolutional Neural Networks with Cutout. Dostupné z: <https://arxiv.org/abs/1708.04552>.
- [5] MILVUS, 2025. How does cutout augmentation work? Dostupné z: <https://blog.milvus.io/ai-quick-reference/how-does-cutout-augmentation-work>. [Accessed 08-05-2025].
- [6] CHOI, Juhwan a KIM, YoungBin, 2024. Colorful Cutout: Enhancing Image Data Augmentation with Curriculum Learning. Dostupné z: <https://arxiv.org/abs/2403.20012>.
- [7] GOODFELLOW, Ian J.; POUGET-ABADIE, Jean; MIRZA, Mehdi; XU, Bing; WARDEFARLEY, David et al., 2014. Generative Adversarial Networks. Dostupné z: <https://arxiv.org/abs/1406.2661>.
- [8] GARCEA, Fabio; SERRA, Alessio; LAMBERTI, Fabrizio a MORRA, Lia, 2023. Data augmentation for medical imaging: A systematic literature review. *Computers in Biology and Medicine*. 2023. 152, s. 106391. ISSN 0010-4825. Dostupné z: <https://www.sciencedirect.com/science/article/pii/S001048252201099X>.

- [9] MAATEN, Laurens a HINTON, Geoffrey E., 2008. Visualizing Data using t-SNE. *Journal of Machine Learning Research*. 2008. 9, s. 2579–2605. Dostupné z: <https://api.semanticscholar.org/CorpusID:5855042>.
- [10] LOPES, Raphael Gontijo; SMULLIN, Sylvia J.; CUBUK, Ekin D. a DYER, Ethan, 2020. Affinity and Diversity: Quantifying Mechanisms of Data Augmentation. *CoRR*. 2020. abs/2002.08973. Dostupné z: <https://arxiv.org/abs/2002.08973>.
- [11] LIU, Zirui; JIN, Haifeng; WANG, Ting-Hsiang; ZHOU, Kaixiong a HU, Xia, 2021. DivAug: Plug-in Automated Data Augmentation with Explicit Diversity Maximization. Dostupné z: <https://arxiv.org/abs/2103.14545>.
- [12] STATLECT, 2025. Domain shift — statlect.com. Dostupné z: <https://www.statlect.com/machine-learning/domain-shift>. [Accessed 06-05-2025].
- [13] AYYADEVARA, V. Kishore a REDDY, Yeshwanth, 2020. *Modern computer vision with PyTorch*. Packt. ISBN 978-1-83921-347-2.
- [14] Fleet, Pajdla, Schiele,, Tuytelaars]computer-vision1 FLEET, David; PAJDLA, Tomáš; SCHIELE, Bernt a TUYTELAARS, Tinne, [2014]. *Computer vision - ECCV 2014*. Springer. Dostupné z: <https://doi.org/10.1007/978-3-319-10602-1>.
- [15] DYK, David A. a MENG, Xiao-Li, 2001. The Art of Data Augmentation. *Journal of Computational and Graphical Statistics*. 2001. 10, č. 1, s. 1–50. ISSN 10618600. Dostupné z: <http://www.jstor.org/stable/1391021>.
- [16] SHAPIRO, Linda G. a STOCKMAN, George C., c2001. *Computer vision*. Prentice Hall. ISBN 0-13-030796-3.
- [17] SHORTEN, Connor a KHOSHGOFTAAR, Taghi M, 2019. A survey on image data augmentation for deep learning. *Journal of Big Data*. 2019. 6, č. 1, s. 1–48.

SEZNAM OBRÁZKŮ

SEZNAM TABULEK

SEZNAM POUŽITÝCH SYMBOLŮ A ZKRATEK

GAN Generative Adversarial Network

SEZNAM PŘÍLOH

PŘÍLOHA P I: První příloha

PŘÍLOHA P II: Druhá příloha

PŘÍLOHA P I: PRVNÍ PŘÍLOHA

Obsah přílohy

PŘÍLOHA P II: DRUHÁ PŘÍLOHA

Obsah přílohy