

Vizualizace a hodnocení augmentačních technik v počítačovém vidění

Bc. Petr Buček

Diplomová práce
2025



Univerzita Tomáše Bati ve Zlíně
Fakulta aplikované informatiky

Univerzita Tomáše Bati ve Zlíně

Fakulta aplikované informatiky

Ústav informatiky a umělé inteligence

Akademický rok: 2024/2025

ZADÁNÍ DIPLOMOVÉ PRÁCE

(projektu, uměleckého díla, uměleckého výkonu)

Jméno a příjmení: **Bc. Petr Buček**

Osobní číslo: **A23749**

Studijní program: **N0613A140022 Informační technologie**

Specializace: **Softwarové inženýrství**

Forma studia: **Kombinovaná**

Téma práce: **Vizualizace a hodnocení augmentačních technik v počítačovém vidění**

Téma práce anglicky: **Visualization and Evaluation of Augmentation Techniques in Computer Vision**

Zásady pro vypracování

1. Prostudujte existující metody augmentace dat v oblasti počítačového vidění a analyzujte, jaký vliv mají různé augmentace na rozmanitost datového souboru.
2. Navrhněte a implementujte postup pro aplikaci zvolených augmentací na vstupní dataset včetně parametrů pro každou augmentaci.
3. Vytvořte t-SNE vizualizaci dat po aplikaci augmentací a analyzujte rozložení a vzory datových bodů ve 2D prostoru.
4. Zhodnotěte a porovnejte výsledky t-SNE vizualizace s ohledem na interpretaci výsledků a vliv jednotlivých augmentací na diverzitu dat.
5. Navrhněte doporučení pro výběr vhodných augmentací pro konkrétní scénáře v oblasti počítačového vidění a diskutujte o jejich potenciálním využití v praxi.

Forma zpracování diplomové práce: **tištěná/elektronická**

Seznam doporučené literatury:

1. XU, Mingle, et al. A comprehensive survey of image augmentation techniques for deep learning. *Pattern Recognition*, 2023, 109347.
2. AYYADEVARA, V. Kishore; REDDY, Yeshwanth. *Modern Computer Vision with PyTorch: Explore deep learning concepts and implement over 50 real-world image applications*. Packt Publishing Ltd, 2020.
3. FLEET, David, Tomas PAJDLA, Bernt SCHIELE a Tinne TUYTELAARS. *Computer Vision – ECCV 2014*. Berlin: Springer Nature, 2014. ISBN 978-3-319-10601-4.
4. VAN DER MAATEN, Laurens; HINTON, Geoffrey. Visualizing data using t-SNE. *Journal of machine learning research*, 2008, 9.11.
5. VAN DYK, David A.; MENG, Xiao-Li. The art of data augmentation. *Journal of Computational and Graphical Statistics*, 2001, 10.1: 1-50.
6. STOCKMAN, George; SHAPIRO, Linda G. *Computer vision*. Prentice Hall PTR, 2001.

Vedoucí diplomové práce: **Ing. Adam Viktorin, Ph.D.**
Ústav informatiky a umělé inteligence

Datum zadání diplomové práce: **27. října 2024**

Termín odevzdání diplomové práce: **2. června 2025**



L.S.

doc. Ing. Jiří Vojtěšek, Ph.D. v.r.
děkan

prof. Mgr. Roman Jašek, Ph.D., DBA v.r.
ředitel ústavu

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.
- že při tvorbě této práce jsem použil nástroj generativního modelu AI Gemini 2.5 Flash dostupné na <https://gemini.google.com> a ChatGPT o4-mini dostupné na <https://chat.openai.com> za účelem inspirace při návrhu a zpracování práce, jako asistenta při programování a optimalizaci kódu a také pro lepší formulování textu. Po použití těchto nástrojů jsem provedl kontrolu obsahu a přebírám za ně plnou zodpovědnost.

Ve Zlíně, dne

.....

podpis autora

ABSTRAKT

Tato práce se zabývá analýzou vlivu augmentace dat na rozmanitost datových souborů a výkon modelu hlubokého učení v oblasti počítačového vidění. Cílem práce bylo analyzovat, jak různé augmentační metody ovlivňují rozmanitost dat a jejich reprezentaci v latentním prostoru neuronové sítě. Bylo vybráno sedm typů augmentací: rotace, horizontální převrácení, perspektivní transformace, změna barevnosti, převod do stupně šedi, Gaussův šum a náhodné mazání. Tyto metody byly aplikovány na datový soubor CIFAR-10 a jejich dopad byl hodnocen pomocí kombinace kvantitativních metrik. Výsledky ukázaly, že některé augmentace, například horizontální převrácení a Gaussův šum, významně rozšiřují prostor příznaků při zachování nízkého doménového posunu, což z nich činí vhodné nástroje pro zlepšení generalizace a odolnosti modelu vůči variacím v obrazových datech. Perspektivní transformace se ukázala jako nejproblematičtější tím, že výrazně narušila strukturu tříd a zvýšila doménový posun.

Klíčová slova: augmentace dat, počítačové vidění, t-SNE, analýza dat

ABSTRACT

This thesis analyzes the impact of data augmentation on dataset diversity and the performance of deep learning models in the field of computer vision. The aim of the work was to examine how different augmentation methods influence data diversity and their representation in the latent space of a neural network. Seven types of augmentations were selected: rotation, horizontal flipping, perspective transformations, color jitter, grayscale conversion, Gaussian noise, and random erasing. These methods were applied to the CIFAR-10 dataset, and their effects were evaluated using a combination of quantitative metrics. The results showed that certain augmentations, such as horizontal flipping and Gaussian noise, significantly expanded the feature space while maintaining a low domain shift, making them suitable tools for improving model generalization and robustness to variations in image data. Perspective transformations proved to be the most problematic, as they considerably disrupted class structure and increased domain shift.

Keywords: data augmentation, computer vision, t-SNE, data analysis

Rád bych poděkoval vedoucímu své práce Ing. Adamu Viktorinovi Ph.D., za odborné vedení, cenné rady a podporu během celého procesu. Dále děkuji své rodině a blízkým za trpělivost, motivaci a povzbuzení.

OBSAH

ÚVOD	15
I TEORETICKÁ ČÁST	16
1 AUGMENTACE DAT V POČÍTAČOVÉM VIDĚNÍ	17
1.1 PROBLEMATIKA KVALITY A MNOŽSTVÍ DAT V POČÍTAČOVÉM VIDĚNÍ	17
1.2 ÚLOHA AUGMENTACE DAT PŘI ZLEPŠOVÁNÍ VÝKONU MODELŮ HLUBOKÉHO UČENÍ	18
1.3 KLASIFIKACE A CHARAKTERISTIKA AUGMENTAČNÍCH METOD	19
1.3.1 Geometrické transformace (rotace, škálování, ořez, zrcadlení)	19
1.3.2 Intenzivní transformace (změna jasu, kontrastu, aplikace šumu)	20
1.3.3 Pokročilé augmentační metody (Cutout, Mixup, CutMix, GAN-based augmentace)	21
1.3.4 Doménově specifické augmentační techniky	23
2 VIZUALIZACE DOPADU AUGMENTACÍ	24
2.1 PRINCIP METODY T-SNE A JEJÍ MATEMATICKÝ ZÁKLAD	24
2.2 APLIKACE T-SNE V KONTEXTU ANALÝZY AUGMENTOVANÝCH DAT	25
2.3 MOŽNOSTI INTERPRETACE VÝSLEDKŮ T-SNE VIZUALIZACE	26
3 HODNOCENÍ VLIVU AUGMENTACÍ NA DIVERZITU DAT	29
3.1 METRIKY PRO KVANTIFIKACI DIVERZITY A ROZMANITOSTI DAT	29
3.2 ANALÝZA VLIVU JEDNOTLIVÝCH AUGMENTAČNÍCH TECHNIK NA DISTRIBUCI DAT	32
3.2.1 Kvantifikace dopadu augmentací na distribuci dat	33
3.3 KOMPARATIVNÍ HODNOCENÍ EFEKTIVITY RŮZNÝCH PŘÍSTUPŮ K AUGMENTACI	33
II PRAKTICKÁ ČÁST	35
4 IMPLEMENTACE A EXPERIMENTÁLNÍ POSTUP	36
4.1 POPIS DATOVÉHO SOUBORU	36
4.2 VÝBĚR AUGMENTACÍ	37
4.3 IMPLEMENTACE AUGMENTACÍ	39

4.4	VÝBĚR METRIK PRO HODNOCENÍ AUGMENTACÍ	42
4.5	IMPLEMENTACE VYBRANÝCH METRIK	43
4.5.1	Rozptylová diverzita.....	43
4.5.2	Index afinity a diverzity	43
4.5.3	Pokrytí prostoru příznaků	44
4.5.4	Skóre posunu domény	45
4.5.5	Čistota shluků	46
4.5.6	KL Divergence	47
4.5.7	Siluetové skóre	47
4.6	IMPLEMENTACE T-SNE VIZUALIZACE.....	48
4.7	IMPLEMENTACE TRÉNOVACÍHO ALGORITMU	49
5	EXPERIMENTÁLNÍ VÝSLEDKY	50
5.1	ROTACE	50
5.1.1	Interpretace výsledků.....	50
5.1.2	Vizualizace t-SNE.....	51
5.2	HORIZONTÁLNÍ PŘEVRÁCENÍ	53
5.2.1	Interpretace výsledků.....	53
5.2.2	Vizualizace t-SNE.....	54
5.3	PERSPEKTIVNÍ TRANSFORMACE	54
5.3.1	Interpretace výsledků.....	54
5.3.2	Vizualizace t-SNE.....	57
5.4	ZMĚNA BAREVNOSTI	57
5.4.1	Interpretace výsledků.....	57
5.4.2	Vizualizace t-SNE.....	60
5.5	STUPNICE ŠEDI	60
5.5.1	Interpretace výsledků.....	62
5.5.2	Vizualizace t-SNE.....	62
5.6	GAUSSŮV ŠUM.....	63
5.6.1	Interpretace výsledků.....	63
5.6.2	Vizualizace t-SNE.....	65

5.7	NÁHODNÉ MAZÁNÍ	65
5.7.1	Interpretace výsledků.....	67
5.7.2	Vizualizace t-SNE.....	68
5.8	POROVNÁNÍ A CELKOVÉ ZHODNOCEŇÍ.....	68
6	ZÁVĚREČNÉ SHRNUTÍ A DOPORUČENÍ	71
6.1	SHRNUTÍ A DISKUSE KLÍČOVÝCH ZJIŠTĚNÍ.....	71
6.2	DOPORUČENÍ PRO VÝBĚR A VYUŽITÍ AUGMENTACÍ.....	72
6.3	DOPORUČENÍ PRO VÝBĚR AUGMENTACÍ PRO KONKRÉTNÍ SCÉNÁŘE	73
	ZÁVĚR.....	74
	SEZNAM POUŽITÉ LITERATURY	77
	SEZNAM OBRÁZKŮ	80
	SEZNAM TABULEK	81
	SEZNAM VÝPISŮ	82
	SEZNAM POUŽITÝCH SYMBOLŮ A ZKRATEK	83
	SEZNAM PŘÍLOH.....	84

ÚVOD

Počítačové vidění hraje klíčovou roli v mnoha odvětvích, od autonomního řízení přes lékařské zobrazování až po průmyslovou automatizaci. Základem těchto systémů jsou především modely hlubokého učení, které vyžadují obrovské množství trénovacích dat, aby fungovaly správně a spolehlivě v těchto disciplínách. Získání a anotace takto rozsáhlého a rozmanitého datového souboru, je ovšem náročné, zdlouhavé a finančně nákladné. Reálná obrazová data navíc obsahují různé měnící se podmínky, jako jsou změny v osvětlení, úhlů pohledu na objekty, jejich částečné zakrytí nebo šum. Model natrénovaný na omezených a málo rozmanitých datech má jisté riziko přeучení a selhává při klasifikaci dat, které se liší od těch, na kterých byl trénován.

Tyto problémy se snaží řešit právě augmentace dat, kdy se uměle rozšiřuje trénovací soubor tak, že se provádějí různé transformace nad obrazovými daty. Tyto transformace zahrnují například rotaci, změny jasu, kontrastu a barevnosti nebo šum. To pomáhá modelu naučit se klasifikovat objekty v širší škále podmínek bez nutnosti sběru nových dat. Model se tedy stane robustnějším a nabude lepší schopnosti generalizace, přičemž se zároveň sníží riziko jeho přeучení.

Cílem této práce je analyzovat vliv různých augmentačních metod na rozmanitost datového souboru a reprezentaci dat v latentním prostoru modelu. V práci se zaměříme na kvantitativní hodnocení dopadu augmentací pomocí vybraných metrik a vizuální analýzy pomocí t-SNE vizualizace. V práci budou vybrány a popsány implementace augmentací a vyhodnocovacích metrik, interpretovány jejich výstupní hodnoty a na konci bude jejich shrnutí a doporučení výběru augmentací pro konkrétní scénáře.

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 počítačového 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 chybami 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 nepřeberné 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 [2]. Největší přínos spočívá v umělém rozšíření vzorů, změnou domény, nebo zvýšení rozmanitosti bez nutnosti dalšího sběru a anotace dat. Tím augmentace zaštiťuje efektivní způsob navýšení velikosti a rozmanitosti trénovacích dat, což vede k lepšímu výsledku modelu.

Efektivita konvolučních neuronových sítí v oblasti hlubokého učení je úzce spjata s dostupností rozsáhlé a kvalitní datové sady. Velké množství kvalitních dat zajišťuje snížení rizika přeучení modelu. Ovšem realita je taková, že v mnoha aplikačních odvětvích, jako je například analýza medicínských snímků, často nejsou rozsáhlé datové sady dostupné [3]. To může být způsobeno různými faktory:

- Sběr takto specializovaných dat je vysoce nákladný
- V datové sadě se nachází nerovnoměrné zastoupení různých tříd a data jsou tedy nevyvážená
- Anotace dat je vysoce časově náročná
- Existence etických a právních omezení týkající se ochrany osobních údajů

Z těchto důvodů často vzejde neúplná datová sada a jejím klíčovým řešením se stává právě augmentace dat, která dokáže původní datovou sadu obohatit a model může dosahovat výsledků srovnatelných s těmi, které by byly dosaženy na robustních a kvalitních reálných datových sadách.

Jeden ze skelních problémů v oblasti počítačového vidění je přeучení, k němuž dochází když se model přílišně naučí kopírovat trénovací data včetně artefaktů jako je šum nebo nerealistických či neexistujících souvislostí. Augmentace dat tento problém řeší následujícími metodami:

- **Zvýšení variability dat:** Původní vzorky jsou obohaceny různými transformacemi, čímž se zajistí odolnost modelu vůči specifickým variacím, které transformace odrážejí. Model se poté dokáže lépe zaměřit na obecnější charakteristiky v datové sadě místo kopírování specifických detailů dané trénovací sady.

- **Umělé rozšíření:** Augmentační metody dokáží s efektivitou zvýšit počet trénovacích vzorků, to vede k lepší generalizaci modelu.
- **Vyvážení dat:** Augmentace pomáhají zvýšit a tím vyrovnat zastoupení málo reprezentovaných tříd v dané datové sadě.

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 [4].

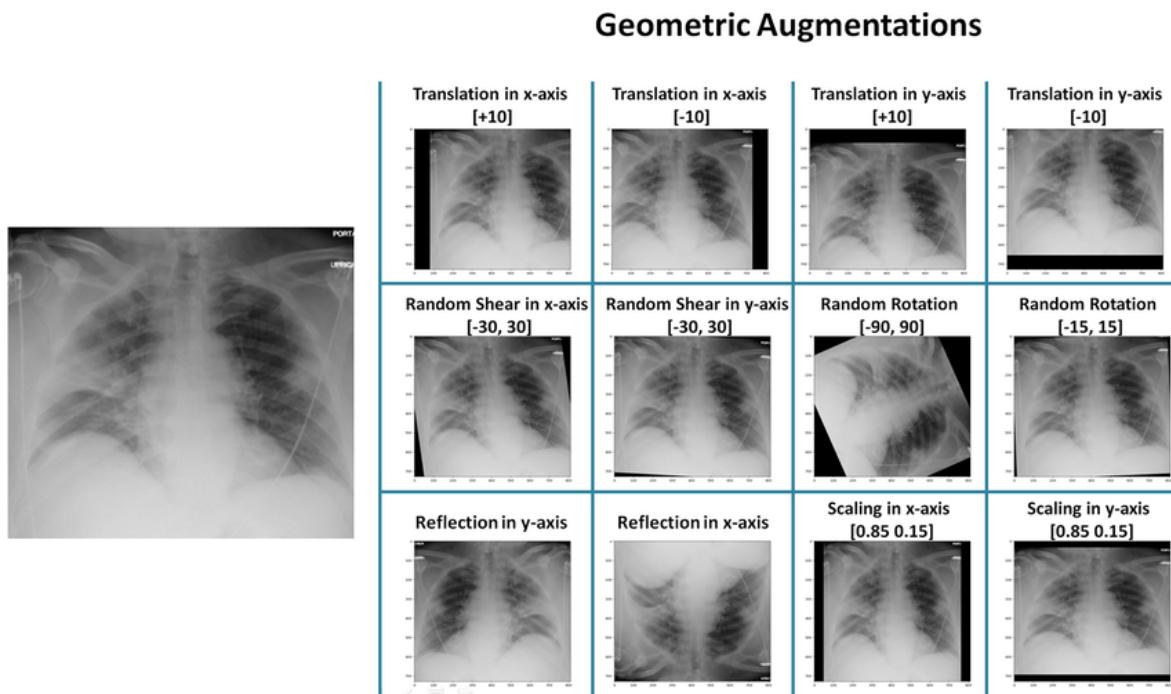
1.3.1 Geometrické transformace (rotace, škálování, ořez, zrcadlení)

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 orientace 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.



Obrázek 1.1 Ukázka geometrických transformací

(zdroj: 5)

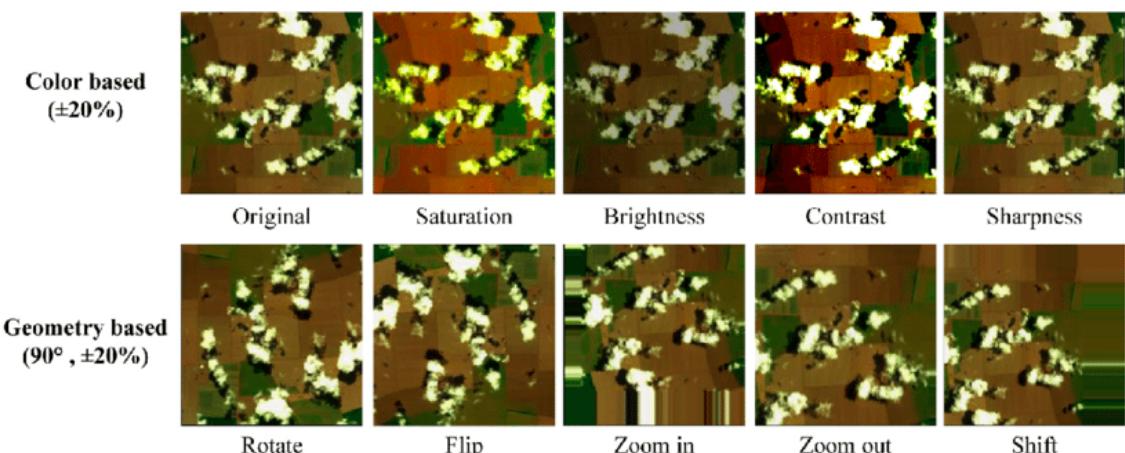
1.3.2 Intenzivní transformace (změna jasu, kontrastu, aplikace šumu)

Tyto transformační techniky jsou stěžejní pro zlepšování odolnosti modelů počítačového videní vůči různým světelným podmínkám, nebo různorodé kvality v obrazových datových sadách. Tato odolnost modelu je důležitá, zejména 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 mo-

del by byl přetrénovaný při specifických světelných podmírkách, například by mohl správně určit objekt pouze při denním světle. [6]

- **Změna kontrastu:** Úprava kontrastu je na datovou sadu aplikována tak, že mapuje hodnoty intenzity obrazu na nový rozsah. Tím se mění rozdíl mezi nejtmavšími a nejsvětlejšími části obrazu. [7] Model natrénovaný na datové sadě augmentované touto metodou je lepší v rozlišování důležitých rysů při různých kontrastních podmírkách.
- **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.



Obrázek 1.2 Ukázka intenzivních transformací

(zdroj: 8)

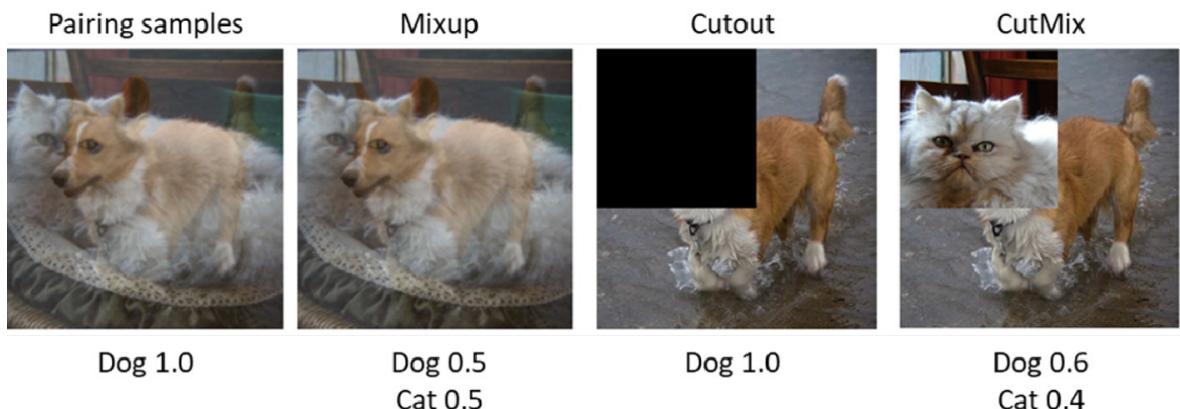
1.3.3 Pokročilé augmentační metody (Cutout, Mixup, CutMix, GAN-based augmentace)

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 této augmentací pramení z faktu, že v reálných obrazových datech zachycené objekty mohou být částečně skryty. Podle zdroje [9] 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 medicín-ských 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 [10].

- **Mixup:** Je metodou, která se snaží o minimalizaci okolního rizika, tedy generuje nové vzorky v blízkosti původních dat, spojením dvou obrazových dat. Tudíž 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 [11].
- **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 proporcionalně 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 o 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 rozpoznat 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 [12]. Ve srovnání s ostatními metodami augmentací dat, tato metoda poskytuje robustnější obohacení datové sady zcela novými vzorky dat, namísto pouhých transformací dat původních.



Obrázek 1.3 Ukázka pokročilých augmentačních metod

(zdroj: 1)

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, zvlášť tak u chorob, které jsou vzácné. Doménově specifické augmentace v medicínském snímkování musí zachovat relevantní informace pro diagnózu [13]. Dále když navštívíme oblast autonomních vozidel, jsou zde 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 představující 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 byli vybíráni 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 nabývají hodnot téměř nulových. [14]

Metoda SNE sice dokáže sestrojit dobrou vizualizaci, ale je těžké optimalizovat nákladovou funkci a vzniká tak problém přeplnění bodů. Další problém nastává z povahy SNE a to, že podmíněné pravděpodobnosti jsou v asymetrii. Tato metoda je tedy rozšířena o novou techniku zvanou "t-Distributed Stochastic Neighbor Embedding", která tyto problémy řeší. Děje se to pomocí zavedením symetrické verze pravděpodobnosti a následným použitím t-distribuce namísto normálního rozdělení.

Symetrická verze SNE metody pracuje se společnými pravděpodobnostmi namísto podmíněných pravděpodobností, tedy místo výpočtu $p_{j|i}$ se používá symetrická pravděpodobnost:

$$p_{ij} = \frac{p_{j|i} + p_{i|j}}{2n}$$

kde n značí celkový počet datových vzorků. Symetrizace zjednodušuje gradient při optimalizaci a algoritmus je stabilnější.

T-distribuce v nízkodimenzionálním prostoru využívá Student-t distribuci na rozdíl od normálního rozdělení pro výpočet podobnosti bodů s jedním stupněm volnosti:

$$q_{ij} = \frac{(1 + \|y_i - y_j\|^2)^{-1}}{\sum_{k \neq l} (1 + \|y_k - y_l\|^2)^{-1}}$$

kde y_i a y_j jsou body v nízkodimenzionálním prostoru. Distribuce Student-t má takovou vlastnost, že vzdálené body ve vysokodimenzionální prostoru jsou reprezentovány jako vzdálené i v nízkodimenzionálním prostoru, přičemž ale nedochází k penalizaci jako u normálního rozdělení.

Perplexita je klíčový parametr algoritmu t-SNE, který určuje efektivní počet nejbližších sousedů pro každý datový bod a je definována jako:

$$\text{Perp}(P_i) = 2^{H(P_i)}, \quad \text{kde } H(P_i) = - \sum_j p_{j|i} \log_2 p_{j|i}$$

kde $H(P_i)$ je Shannonova entropie pravděpodobnostního rozdělení P_i . Provádí se binární hledání hodnoty tak, aby perplexita u každého z bodů odpovídala zadané hodnotě.

2.2 Aplikace t-SNE v kontextu analýzy augmentovaných dat

Vizualizace vysokodimenzionálních dat pomocí t-SNE je účinným nástrojem pro analýzu dopadu augmentačních metod na distribuci dat v kontextu počítačového vidění. Díky této vizualizační metodě jsme schopni porovnat distribuce dat před a po aplikaci augmentací v dvourozměrném prostoru, to nám pomáhá posoudit, jak jednotlivé augmentační metody ovlivňují strukturu dat. Při analýze je také důležité zhodnotit, jestli augmentační metoda správně rozšiřuje datový prostor, nebo zda vytváří vzory nerealistické pro přirozenou variabilitu původních dat. Podle příkladů augmentací z první kapitoly této práce může být struktura ovlivněna například:

- **Geometrické transformace** by měli vytvářet body, které vytvářejí shluky kolem původních neaugmentovaných bodů, přičemž ale jejich distribuci jemně rozšiřují.

- **Intenzivní transformace** by měli ve vizualizaci vytvářet jakési "paprsky", které vychází z původních bodů. Paprsky jsou tvořeny tak, že tento typ transformací ve vizualizaci body lehce rozptyluje do blízkého okolí původních bodů buďto směrem od nebo do středu, či radiálně. Každý z paprsků pak reprezentuje konkrétní transformaci.
- **Pokročilé augmentační metody** mohou ve vizualizaci vytvářet zcela nové shluky.

Standardní implementace t-SNE vizualizace zachovává lokální strukturu dat, to následně může vést ke ztrátě informací o globální struktuře, což se může projevit při hodnocení vlivu augmentací, jež mění i globální vlastnosti dat. Tento problém řeší rozšířený algoritmus globální t-SNE (g-SNE), který uvažuje navíc s globální nákladovou funkcí k původní nákladové funkci t-SNE algoritmu:

$$C_{\text{g-SNE}} = C_{\text{t-SNE}} + \lambda C_{\text{global}}$$

Kde váhový parametr λ kontroluje rovnováhu mezi lokálními shluky a globální distribucí dat. Tato globální nákladová funkce na rozdíl od původní nákladové funkce je citlivá na velké vzdálenosti v datovém prostoru a tedy dokáže lépe zachovat vztahy mezi shluky [1].

Dopad augmentačních metod lze sledovat i v průběhu učení modelu. Extrahované příznaky lze vizualizovat z jakékoli vrstvy sítě. Pokud tak učiníme před a po aplikaci augmentačních metod, získáme poznatky o tom, jak síť interpretuje augmentovaná data v různých fázích učení a to nám dokáže nabídnout lepší porozumění jak augmentace dokáží ovlivnit učení modelu.

Pomocí tohoto vizualizačního nástroje lze identifikovat augmentace, které dokáží pomoci modelu lépe generalizovat, nebo také rozpoznat augmentace, které naopak mohou způsobit přetrénování nebo jiné problémy při učení modelu. Také lze optimalizovat augmentační strategii dle vizuální zpětné vazby jejího dopadu na augmentovaná data.

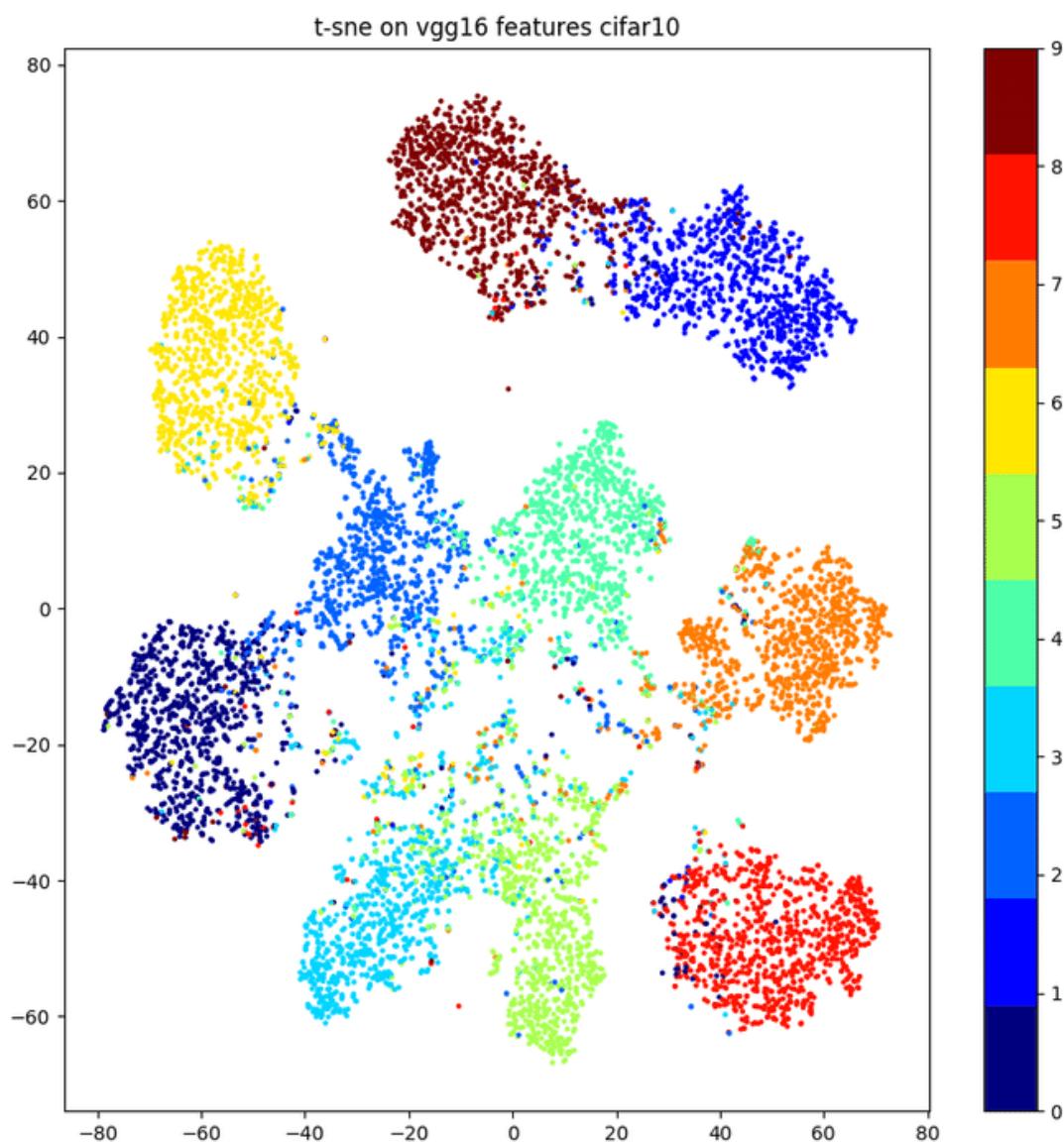
2.3 Možnosti interpretace výsledků t-SNE vizualizace

Při interpretaci t-SNE vizualizace je potřeba abychom měli na paměti několik úskalí této metody. Kvůli povaze t-SNE zachovávání lokálních vzdáleností, nemusí vzdálenosti mezi jednotlivými shluky odpovídat vzdálenostem v původním datovém prostoru. Také jejich velikost a hustota nemusí být reprezentativní k reálné velikosti a hustotě v datovém prostoru. Dále

změna parametrů algoritmu, jako je například parametr perplexity, může ovlivnit výslednou vizualizaci. Stochastická povaha algoritmu znamená, že algoritmus může produkovat mírně odlišné vizualizace pro různé běhy.

Z vizualizace můžeme zkoumat tyto aspekty:

- **Překrývání tříd:** Pokud se po augmentaci třídy více překrývají, může to znamenat, že augmentace vytvářejí datové vzory, které mohou vytvořit zmatek při klasifikaci. To lze ve vizualizaci lehce poznat.
- **Rozšíření datového prostoru:** Že augmentace správně rozšiřují datový prostor ve vizualizaci zjistíme tím, že existující shluky jaksi "přetíkají" nebo se postupně rozšiřují.
- **Zachování identity tříd:** Správně by augmentační metoda měla zachovávat identitu tříd po augmentaci. Ve vizualizaci by měly augmentovaná data zůstávat poblíž původních dat stejné třídy.
- **Nové shluky:** Vznik nových shluků mohou do datového prostoru zanést některé z pokročilých augmentačních metod. Zde je důležité zjistit, zda-li tyto nové shluky smysluplně rozšiřují datový prostor, nebo zda-li vytváří zavádějící a odlehlé hodnoty, které mohou spíše ublížit efektivitě modelu.



Obrázek 2.1 Ukázka t-SNE na datasetu CIFAR-10

(zdroj: 15)

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 rozdeleními není žádný posun. Záporná hodnota pak značí, že augmentovaná data se nacházejí mimo distribuční oblast očekávanou modelem. [16]

D_{train} a D_{val} označují trénovací a validační datové sady, které jsou nezávisle a identicky rozdelené 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 affinity $T[a; m; D_{\text{val}}]$ 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 rozširovat trénovací datovou sadu. Diverzita reflektuje jakousi intuici, že komplexní nebo vícefaktorové augmentační strategie zvyšují počet vzorků v trénovací datové sadě, čímž snižují riziko přeúčení modelu [16].

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:

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

- **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 [17] 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 $f_0(\cdot)$ 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 [18].

Metriky shlukování a homogeneity 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 vhodna.

- **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 ovlivňuje diverzitu obrazových dat. Pochození 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.

Augmentační metody jsou transformacemi nad původními datovými sadami s účelem vytvoření obohaceného a rozmanitějšího trénovacího souboru, to lze kvantifikovat pomocí různých metrik a vizualizovat pomocí algoritmů jako je t-SNE, což nám umožní nahlédnout na změnu distribuce dat, jak se nově augmentovaná data rozprostírají dimenzemi a jak se například mění jejich shlukování oproti původním datům. Augmentační metody také dokáží modifikovat strukturu a rozmanitost původních dat, což může mít vážný dopad na výkon modelu a je potřeba zhodnotit, zda je tento dopad přínosný nebo ne.

- **Geometrické transformace (rotace, škálování, ořez, zrcadlení)** zachovávají sémantický obsah a charakteristiku obrazu, ale mění jeho prostorovou orientaci, tedy rozšiřují datový prostor o další možné variace a pomáhají modelu se obrnit vůči těmto změnám v reálných datech.
- **Intenzivní transformace (změna jasu, kontrastu, šum)** mění vizuální vlastnosti ob-

razu jako takového, to může pomoci zvětšit robustnost modelu vůči změnám v osvětlení a šumu v reálných datech.

- **Pokročilé metody (Cutout, Mixup, GAN-based)** vytváří zcela nové umělé vzory, ty mohou výrazně rozšířit původní distribuci dat mimo jejich původní vzory a shluhy, model poté lépe dokáže generalizovat. Ovšem tyto metody také mohou zavést artefakty do augmentovaných dat, proto je velmi důležité tyto metody správně vyhodnotit.

3.2.1 Kvantifikace dopadu augmentací na distribuci dat

Pro analýzu vlivu augmentačních metod na distribuci dat se dají použít tyto metriky:

- **Statistické metriky rozptylu a hustoty dat:** tyto metriky kvantifikují rozptyl dat v různých dimenzích po augmentaci.
- **Metriky divergence distribucí:** kvantifikují rozdíly mezi distribucí původní a augmentované datové sady.
- **Metriky vzdáleností v latentním prostoru:** tyto metriky odrážejí, jak se augmentační metody projevují v prostoru reprezentací modelu, jde například o průměrnou vzdálenost mezi body.

Z výsledků těchto analýz lze vidět, že různé augmentační metody mají různé dopady na rozmanitost a distribuci dat. Jednoduché geometrické augmentace často vedou k mírnému rozšíření distribuce, naopak metody pokročilé, které dokáží vytvořit zcela nové vzorky dokáží distribuci výrazně rozšířit, ale s rizikem zavedení artefaktů a vzorů neodpovídající těm reálným.

Pro správné nastavení tréninkového procesu modelu je klíčové pochopit, jak augmentace mění distribuci dat. Přílišné augmentace, které ostře rozšířují distribuci dat, mohou vést k učení nereálných vzorů. Naopak nedostatečné augmentace nemusí dostatečně pokrýt variabilitu reálných dat a model se může lehce přeucít. Je tedy důležité vyvážit kvalitu a diverzitu augmentovaných dat i s ohledem na danou úlohu a datovou sadu.

3.3 Komparativní hodnocení efektivity různých přístupů k augmentaci

Komparativní hodnocení efektivity augmentačních metod je klíčové pro výběr optimální strategie pro danou úlohu. Augmentační metody se liší v principech jak rozšiřují distribuci

dat, v dopadu na výkon modelu, robustnosti vůči variabilitě v reálných datech a schopnosti generalizace.

Kritéria hodnocení efektivity augmentací:

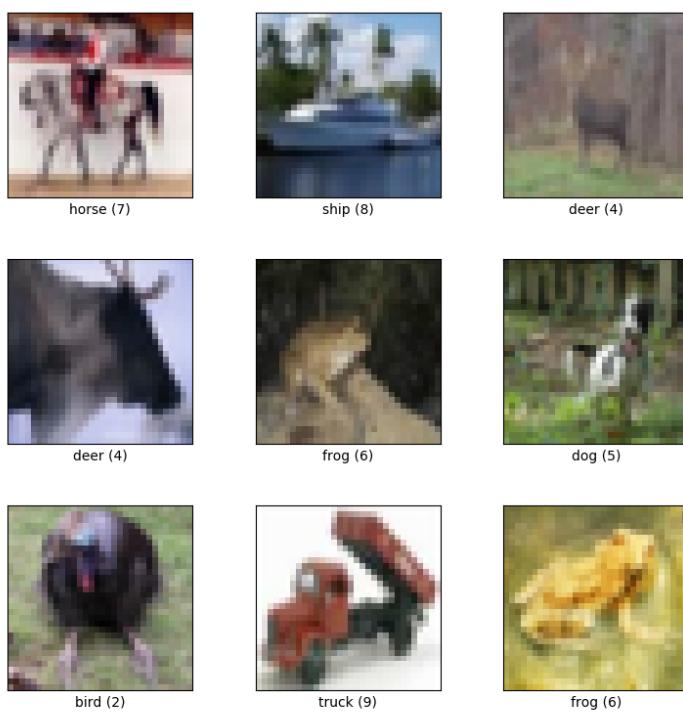
- Robustnost modelu vůči rušivým vlivům jako je například šum, deformace, variace v osvětlení.
- Výpočetní náročnost a složitost na implementaci jednotlivých augmentačních metod.
- Diverzita augmentovaných dat, která ovlivňuje jak se model dokáže naučit variabilitu reálných dat.
- Dopad na konvergenci tréninku, tedy jak rychle a stabilně model dokáže dosáhnout optimálních parametrů.
- Zlepšení přesnosti modelu na validačních a testovacích datových sadách, to dokáže odhalit přínos augmentační metody pro schopnost modelu lépe generalizovat.

II PRAKTICKÁ ČÁST

4 IMPLEMENTACE A EXPERIMENTÁLNÍ POSTUP

4.1 Popis datového souboru

Pro účely této práce byl vybrán datový soubor CIFAR-10. Tento datový soubor je standardizovaný dataset často využívaný v oblasti strojového učení a počítačového vidění pro výzkum. CIFAR-10 obsahuje v celku 60 000 obrazových vzorků s rozlišením 32x32 pixelů. Ty jsou rozděleny do 10 sémantických tříd a každá třída je reprezentována přesně 6000 vzorky. Dále je datový soubor rozdělen na trénovací a testovací sadu v poměru 5:1, tedy 50 000 obrazových vzorků pro trénování a 10 000 vzorků pro testování. Obrazové vzorky jsou uloženy ve formátu trojrozměrných tenzorů odpovídající kanálům rgb tedy 32x32x3, kde každý pixel má hodnotu v rozsahu 0-255 reprezentovaný 8 bity [19].



Obrázek 4.1 Ukázka datasetu CIFAR-10

(zdroj: 20)

Sémantické třídy datového souboru CIFAR-10 jsou následující:

- Letadla

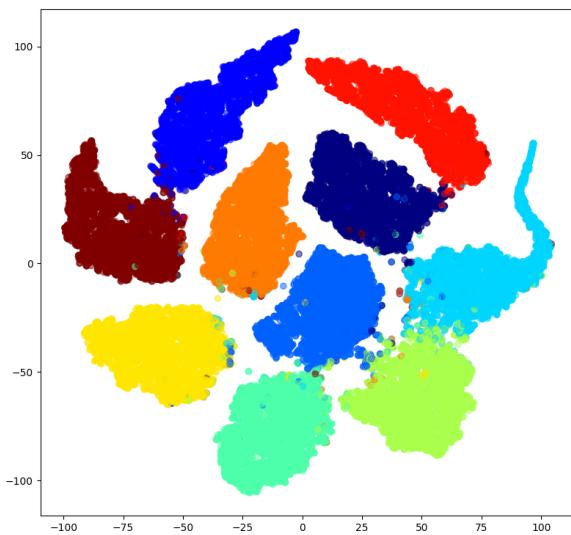
- Automobily
- Ptáci
- Kočky
- Jeleni
- Psi
- Žáby
- Koně
- Lodě
- Nákladní automobily

Výběr datové sady CIFAR-10 jako základu pro experimenty vychází z několika jeho vlastností. Jednak jde o nízké rozlišení obrazových vzorků díky němuž je trénovací proces rychlý a umožňuje rychlejší iteraci při testování. Dále jeho skromnější komplexita co se množství tříd týče poskytuje rozumný testovací prostor pro přehlednou vizualizaci a vyhodnocení dopadu augmentací. Navíc jeho využití v jiných studiích poskytuje přímé srovnání s již existujícími metodami či výzkumy. Graf 4.2 zobrazuje vizualizaci datového souboru CIFAR-10 pomocí t-SNE.

4.2 Výběr augmentací

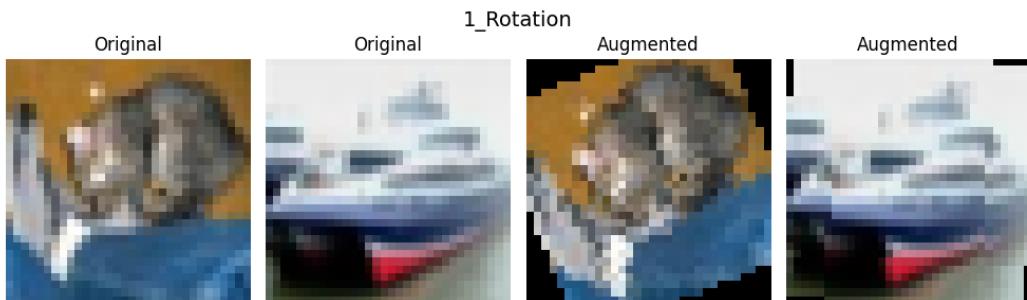
Pro tuto práci byly vybrány augmentační metody, které pokrývají několik kategorií augmentačních technik. Výběr je založen na schopnostech rozšíření distribuce datové sady z hlediska geometrických změn, fotometrických transformací, simulace šumu a kombinace obrazových vzorů. Tabulka 4.1 níže uvádí jednotlivé augmentace, kategorie do které patří, parametry dané augmentace a její očekávaný vliv na data.

- **Geometrické transformace** byly zvoleny pro rozšíření prostorové variability datových vzorků, s účelem pomocí modelu naučit se odchylky orientací a pozic objektů v obrazových datech.



Obrázek 4.2 Vizualizace t-SNE pro CIFAR-10

(zdroj: vlastní)



Obrázek 4.3 Ukázka augmentace: Rotace

(zdroj: vlastní)

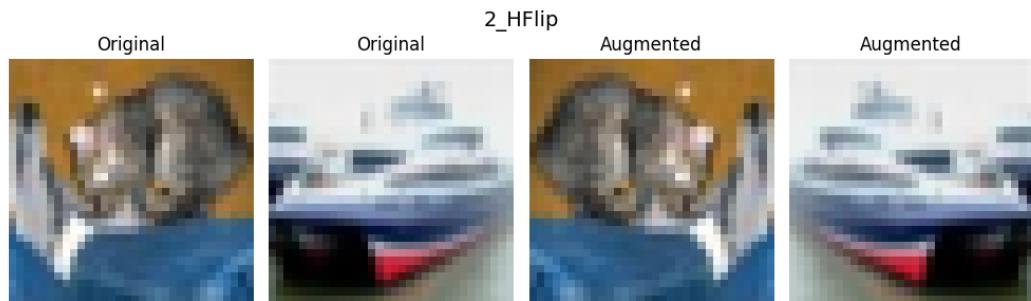
- **Fotometrické augmentace** by měli pomocí modelu lépe rozpoznávat objekty nebo scény při změně osvětlení, barevnosti vstupních dat nebo i při absenci informací o barvě, což odráží reálné scénáře při sběru dat.
- **Augmentace založené na šumu** napodobují šum zachycený na reálných obrazových vzorcích, zvyšující robustnost modelu vůči nekvalitním vstupním datům.

Výběr těchto augmentací umožňuje analýzu jejich vlivu na diverzitu datového souboru CIFAR-10, poskytuje také podklad pro t-SNE vizualizaci prostoru příznaků. Analýzou jednotlivých augmentačních metod poté budeme sledovat, jak se mění rozložení dat v prostoru příznaků. Výpočtem a analýzou různých metrik budeme uvažovat, jak mohou přispět k lepší generalizaci, robustnosti a výkonu modelu.

Tabulka 4.1 Tabulka použitých augmentací

Exp.	Název augmentace	Kategorie	Parametry	Očekávaný vliv na data
1	Rotace	Geometrická	$\pm 30^\circ$	Změna orientace, větší variabilita pozice objektů
2	Horizontální převrácení	Geometrická	$p = 0.5$	Otočení horizontálně, mírná diverzita bez ztráty informace
3	Perspektivní transformace	Geometrická	$zkreslení = 0.5, p = 1.0$	Simulace zkreslení perspektivy, zvýšení variace tvaru a úhlu pohledu
4	Změna barevnosti	Fotometrická	Jas, kontrast, saturace, odstín (0.4)	Změna barevnosti, rozšíření barevné variability
5	Stupnice šedi	Fotometrická	Převod na odstíny šedi	Ztráta barevné informace, testování odolnosti na barevné změny
6	Gaussův šum	Šum	mean = 0, std = 0.1	Přidání zrnění, simulace reálného šumu v obrazech
7	Náhodné mazání	Šum/Okluze	Měřítko = (0.02, 0.33)	Náhodné vymazání částí obrazu, simulace výpadků a překryvů

(zdroj: vlastní)



Obrázek 4.4 Ukázka augmentace: Horizontální převrácení

(zdroj: vlastní)

4.3 Implementace augmentací

Augmentace jsou definovány v Python slovníku, kde klíčem je název augmentace a číslo experimentu, hodnotou je pak samotná augmentace. Většina augmentací jsou transformační operace z knihovny `torchvision.transforms` a u Gaussova šumu se jedná o vlastně implementovanou třídu.

```

1 AUGMENTATIONS = {
2     "1_Rotation": transforms.RandomRotation(30),
3     "2_HFlip": transforms.RandomHorizontalFlip(p=0.5),
4     "3_RandomPerspective": transforms.RandomPerspective(
5         distortion_scale=0.5, p=1.0),
6     "4_ColorJitter": transforms.ColorJitter(brightness=0.4,
7         contrast=0.4, saturation=0.4, hue=0.1),
8     "5_Grayscale": transforms.Grayscale(num_output_channels

```



Obrázek 4.5 Ukázka augmentace: Perspektivní transformace

(zdroj: vlastní)



Obrázek 4.6 Ukázka augmentace: Změna barevnosti

(zdroj: vlastní)

```

7     =3),
8     "6_GaussianNoise": GaussianNoise(std=0.1),
9     "7_RandomErasing": transforms.RandomErasing(),
}
```

Výpis 4.1 Implementace augmentací

Gaussův šum je implementován pomocí třídy `GaussianNoise`, která přijímá dva parametry a funguje jako standardní transformační operace. Šum lze řídit pomocí dvou parametrů: `mean`, který ovládá posun šumu a `std`, který určuje intenzitu šumu. Návratová hodnota je tensor stejného tvaru jako ten vstupní, ovšem s přidanými náhodnými hodnotami s Gaussovým rozdělením se zadaným středem `mean` a směrodatnou odchylkou `std`.

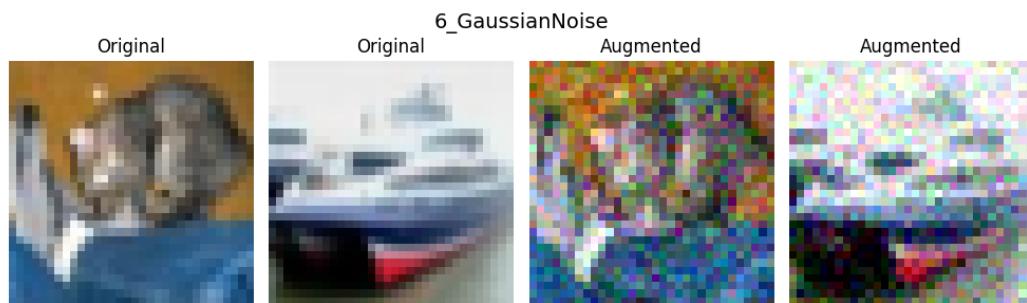
```

1 class GaussianNoise:
2     def __init__(self, mean=0., std=0.1):
3         self.mean = mean
4         self.std = std
5     def __call__(self, tensor):
6         return tensor + torch.randn_like(tensor) * self.std +
               self.mean
```



Obrázek 4.7 Ukázka augmentace: Stupnice šedi

(zdroj: vlastní)



Obrázek 4.8 Ukázka augmentace: Gaussův šum

(zdroj: vlastní)

Výpis 4.2 Implementace Gaussova šumu

Augmentace jsou poté do celé pipeline vkládány pomocí funkce `get_transform(aug_key)`, která následně sestaví sekvenci transformací pro trénovací data. Každý obraz je převeden na tensor a následně normalizován, poté se přidá daná augmentace na začátek seznamu transformací. Funkce navrácí objekt, který představuje sekvenci transformačních operací.

```

1 def get_transform(aug_key=None):
2     base = [
3         transforms.ToTensor(),
4         transforms.Normalize((0.5, 0.5, 0.5), (0.5, 0.5, 0.5))
5     ]
6     if aug_key in AUGMENTATIONS:
7         base.insert(0, AUGMENTATIONS[aug_key])
8
9     return transforms.Compose(base)

```

Výpis 4.3 Funkce `get_transform`



Obrázek 4.9 Ukázka augmentace: Náhodné mazání

(zdroj: vlastní)

4.4 Výběr metrik pro hodnocení augmentací

Pro hodnocení vlivu augmentačních metod na diverzitu dat, pokrytí prostoru příznaků, separaci shluků a posunu domény byly vybrány metriky popsané v tabulce 4.2 níže.

Tabulka 4.2 Tabulka vybraných metrik

Metrika	Rozsah hodnot	Význam nízké hodnoty	Význam vysoké hodnoty
Rozptylová diverzita	≥ 0.0	Nízká variabilita, data jsou homogenní	Vysoká rozmanitost, data pokrývají širší spektrum
Index afinity	0.0 - 1.0	Nízká diverzita nebo přílišná odlišnost od původních dat	Vysoká podobnost k originálním datům, zachování klíčových rysů
Index diverzity	0.0 - 1.0	Nízká variabilita, data jsou homogenní	Vysoká rozmanitost, data pokrývají širší spektrum variací
Pokrytí prostoru příznaků	$-\infty, \infty \%$	Špatné pokrytí prostoru příznaků, omezená variabilita	Dobré pokrytí prostoru příznaků, široká reprezentace dat
Skóre posunu domény	≥ 0.0	Malý posun domény, data jsou podobná originálu	Velký posun, riziko ztráty relevantnosti dat
Čistota shluků	0.0 - 1.0	Špatné shlukování, třídy se překrývají	Dobré shlukování, třídy jsou jasně oddělené
KL divergence	≥ 0.0	Malý rozdíl mezi distribucemi, dobrá věrnost vůči původní distribuci	Velký rozdíl, augmentace mění distribuci výrazně
Siluetové skóre	-1.0 - 0.0	Špatná separace a kompaktnost shluků	Dobrá separace a kompaktnost shluků

(zdroj: vlastní)

Tyto vybrané metriky poskytnou ucelený pohled na kvalitu augmentací z hlediska variability, pokrytí prostoru příznaků, udržení konzistence vůči původní datové sadě, zachování struktury a posunu domény. To umožní kvantitativní hodnocení efektivity jednotlivých augmentačních metod a jejich vhodnost pro konkrétní scénáře.

4.5 Implementace vybraných metrik

Tato podkapitola se zaměří na popis jednotlivých implementací metrik vybraných pro tuto práci.

4.5.1 Rozptylová diverzita

Funkce `variance_diversity` vyhodnotí rozptyl pravděpodobnostních výstupů modelu na datech. Pro každý obrazový vzorek se vypočítá pravděpodobnosti tříd pomocí funkce `nn.functional.softmax`. Tyto pravděpodobnosti jsou uloženy do pole a následně se z nich spočítá průměrná variance přes všechny třídy.

```

1 def variance_diversity(model, dataloader):
2     model.eval()
3     probs = []
4     with torch.no_grad():
5         for images, _ in dataloader:
6             images = images.to(DEVICE)
7             outputs = model(images)
8             prob = nn.functional.softmax(outputs, dim=1)
9             probs.append(prob.cpu().numpy())
10    probs = np.concatenate(probs, axis=0)
11    var_div = np.var(probs, axis=0).mean()
12    return var_div

```

Výpis 4.4 Implementace rozptylové diverzity

4.5.2 Index affinity a diverzity

Funkce `affinity_diversity_index` vypočítává dvě metriky naráz. Metriku affinity, která měří jak moc jsou si augmentovaná data podobná datům původním z hlediska klasifikace a metriku diversity, která měří variabilitu augmentovaných dat. Augmentovaná data se postupně procházejí a počítají se predikce. Jednak se spočítá přesnost predikce vůči skutečnému štítku pro affinity a poté se vypočítá rozdíl průměrných ztrát augmentovaných a původních dat pro index diversity. Funkce navrací obě tyto hodnoty.

```

1 def affinity_diversity_index(base_model, augmented_loader,
2     aug_losses, base_losses):
3     base_model.eval()
4     preds = []
5     labels = []
6     with torch.no_grad():
7         for images, lbls in augmented_loader:
8             images = images.to(DEVICE)
9             outputs = base_model(images)
10            preds.append(outputs.argmax(dim=1).cpu().numpy())
11            labels.append(lbls.numpy())

```

```

11     preds = np.concatenate(preds)
12     labels = np.concatenate(labels)
13     affinity = accuracy_score(labels, preds)
14     diversity = np.mean(aug_losses) - np.mean(base_losses)
15
16     return affinity, diversity

```

Výpis 4.5 Implementace indexů afinity a diverzity

4.5.3 Pokrytí prostoru příznaků

Výpočet této metriky je realizován do dvou fází kdy prvně se použije PCA (principiální analýza komponentů) pro redukci dimenzionality dat a poté se spočítá konvexní obal, který kvantifikuje míru pokrytí prostoru příznaků. Dimenzionalitu zde redukujeme, protože výpočet konvexního obalu pro více jak tři dimenze je nestabilní a výpočetně náročný. Nejprve se zkombinují data původních a augmentovaných příznaků a projdou redukcí pomocí PCA do 2D/3D prostoru. Výpočet konvexního obalu je následně vypočítán pomocí třídy ConvexHull z knihovny `scipy.spatial`. Konvexní obal se vypočítá jak pro model naučený na čistých datech, tak pro model naučený na datech augmentovaných a výstupem je procentuální hodnota, která určuje o kolik konvexní obal augmentovaného modelu se zvětšil nebo zmenšil od toho neu augmentovaného.

```

1 if len(baseline_features.shape) > 2:
2     baseline_features = baseline_features.reshape(
3         baseline_features.shape[0], -1)
4 if len(augmented_features.shape) > 2:
5     augmented_features = augmented_features.reshape(
6         augmented_features.shape[0], -1)
7 combined_features = np.concatenate([baseline_features,
8                                   augmented_features], axis=0)
9 actual_n_components = min(n_components, combined_features
10                           .shape[0] - 1, combined_features.shape[1])
11 if actual_n_components <= 1:
12     print("Warning: Not enough data points or dimensions
13           for meaningful PCA and ConvexHull calculation.
14           Returning 0.")
15     return 0.0
16
17 pca = PCA(n_components=actual_n_components)
18 pca.fit(combined_features)
19 baseline_reduced = pca.transform(baseline_features)
20 augmented_reduced = pca.transform(augmented_features)
21 if baseline_reduced.shape[0] < actual_n_components + 1:
22     print(f"Warning: Not enough points ({baseline_reduced
23           .shape[0]}) for baseline convex hull in {
24           actual_n_components} dimensions. Returning 0.")
25 else:
26     baseline_hull = ConvexHull(baseline_reduced)
27     baseline_volume = baseline_hull.volume

```

```

26         except Exception as e:
27             print(f"Error calculating baseline convex hull: {e}. Returning 0.")
28             baseline_volume = 0.0
29
30     if augmented_reduced.shape[0] < actual_n_components + 1:
31         print(f"Warning: Not enough points ({augmented_reduced.shape[0]}) for augmented convex hull in {actual_n_components} dimensions.
32             Returning 0.")
33     augmented_volume = 0.0
34 else:
35     try:
36         augmented_hull = ConvexHull(augmented_reduced)
37         augmented_volume = augmented_hull.volume
38     except Exception as e:
39         print(f"Error calculating augmented convex hull: {e}. Returning 0.")
40     augmented_volume = 0.0
41
42     if baseline_volume == 0:
43         if augmented_volume > 0:
44             return float('inf')
45         else:
46             return 0.0
47     else:
48         percentage_change = ((augmented_volume - baseline_volume) / baseline_volume) * 100
        return percentage_change

```

Výpis 4.6 Implementace pokrytí prostoru příznaků

4.5.4 Skóre posunu domény

Výpočet této metriky je založen na natrénování jednoduchého klasifikátoru, který je schopen rozlišit původní a augmentovaná data. Výstupem je potom jeho přesnost v rozlišování těchto dat, pokud je klasifikátor schopen lehce rozlišit augmentovaná data od těch originálních, znamená to, že mezi nimi existuje výrazný posun domény, pokud ovšem klasifikátor selhává, je to ukazatel, že doménový posun je mezi nimi malý. Klasifikátor je implementován pomocí jednoduché konvoluční neuronové sítě, která je implementována jako třída `DomainClassifier`, která má za vstup obrazový vzorek o rozměrech 32x32x3, ten převede na 1D vektor o délce 3072. Výstupem jsou pravděpodobnosti pro třídu 0 (originální vzorek) a třídu 1 (augmentovaný vzorek). Trénink začíná tak, že se zkombinují původní a augmentovaná data, přičemž každá z dávek obsahuje polovinu původních obrazových dat a polovinu dat augmentovaných. Model se učí 3 epochy. Výpočet přesnosti zajišťuje funkce `accuracy_score` z knihovny `sklearn.metrics`.

```

1 def train_domain_classifier(original_loader, augmented_loader):
2     class DomainClassifier(nn.Module):
3         def __init__(self):
4             super().__init__()
5             self.net = nn.Sequential(

```

```

6             nn.Flatten(),
7             nn.Linear(3*32*32, 256),
8             nn.ReLU(),
9             nn.Linear(256, 2)
10            )
11        def forward(self, x):
12            return self.net(x)
13
14    model = DomainClassifier().to(DEVICE)
15    criterion = nn.CrossEntropyLoss()
16    optimizer = optim.Adam(model.parameters(), lr=1e-3)
17
18    def combined_loader():
19        for (x1, _), (x2, _) in zip(original_loader,
20                                      augmented_loader):
21            yield x1, torch.zeros(x1.size(0), dtype=torch.
22                                  long)
23            yield x2, torch.ones(x2.size(0), dtype=torch.long)
24
25    model.train()
26    for epoch in range(3):
27        for images, labels in combined_loader():
28            images, labels = images.to(DEVICE), labels.to(
29                DEVICE)
30            optimizer.zero_grad()
31            outputs = model(images)
32            loss = criterion(outputs, labels)
33            loss.backward()
34            optimizer.step()
35
36    model.eval()
37    all_preds = []
38    all_labels = []
39    with torch.no_grad():
40        for images, labels in combined_loader():
41            images, labels = images.to(DEVICE), labels.to(
42                DEVICE)
43            outputs = model(images)
44            preds = outputs.argmax(dim=1)
45            all_preds.append(preds.cpu().numpy())
46            all_labels.append(labels.cpu().numpy())
47    all_preds = np.concatenate(all_preds)
48    all_labels = np.concatenate(all_labels)
49    acc = accuracy_score(all_labels, all_preds)
50
51    return acc

```

Výpis 4.7 Implementace skóre posunu domény

4.5.5 Čistota shluků

Výpočet této metriky se skládá ze tří kroků:

- **Shlukování** pomocí algoritmu K-Means do 10 shluků. Výstupem je, že každý vzorek je přiřazen do jednoho ze shluků.
- **Kontigenční matici** je vytvořena pomocí funkce `confusion_matrix`, která porovnává skutečné štítky vůči těm přiřazeným ke shlukům. Řádky odpovídají skutečným třídám a sloupce shluků.
- **Výpočet purity** se děje tak, že pro každý shluk se najde maximální hodnota, ta

představuje dominantní třídu daného shluku. Tyto hodnoty jsou sečteny napříč všemi shluky a jejich suma je dělena celkovým počtem všech vzorků a tím je získána hodnota purity, která se pohybuje v rozmezí 0.0 až 1.0.

```

1 def cluster_purity(features, true_labels, n_clusters=10):
2     kmeans = KMeans(n_clusters=n_clusters, random_state
3                      =42).fit(features)
4     cm = confusion_matrix(true_labels, kmeans.labels_)
5     purity = np.sum(np.max(cm, axis=0)) / np.sum(cm)
6     return purity

```

Výpis 4.8 Implementace čistoty shluků

4.5.6 KL Divergence

Funkce `kullback_leibler_divergence` kvantifikuje rozdíl mezi dvěma diskrétními pravděpodobnostními rozděleními p a q . Nejprve se obě rozdělení normalizují na pravděpodobnostní rozdělení, kde součet všech prvků v daném rozdělení je roven 1. Poté funkce `distance.rel_entr` vypočítá relativní entropii normalizovaných rozdělení a poté se nad ním provede suma což odpovídá KL divergenci. Ke každému rozdělení se přičítá malá hodnota 1×10^{-10} , která zabraňuje případu, kdyby se dělilo nulou.

```

1 def kullback_leibler_divergence(p, q):
2     p_norm = p / np.sum(p)
3     q_norm = q / np.sum(q)
4     return distance.rel_entr(p_norm + 1e-10, q_norm + 1e-10).
5           sum()

```

Výpis 4.9 Implementace KL divergence

4.5.7 Siluetové skóre

Implementace siluetového skóre probíhá tak, že je nejprve vytvořen model `KMeans` s počtem shluků 10 což odpovídá datovému souboru CIFAR-10, pevně daným náhodným seedem a počtem inicializací na 10, což zajistí větší stabilitu výsledků. Model poté provede shlukování a vrátí přiřazení každého datového bodu ke shluku. Poté se pomocí funkce `silhouette_score` spočítá průměrné siluetové skóre pro všechny datové body.

```

1 def silhouette(features):

```

```

2     kmeans = KMeans(n_clusters=10, random_state=42, n_init
3         =10)
4     cluster_labels = kmeans.fit_predict(features)
5     return silhouette_score(features, cluster_labels)

```

Výpis 4.10 Implementace siluetového skóre

4.6 Implementace t-SNE vizualizace

Vizualizace t-SNE probíhá v několika krocích, nejdříve se zkombinují trénovací (augmentovaná) a validační data do jednoho pole `all_features` a k němu se vytvoří ještě pole `labels`, ve kterém 0 označuje trénovací data a 1 validační data pro jejich rozlišení. Dále proběhne redukce dimenzionality dat pomocí t-SNE do 2D prostoru. Tato funkce poté generuje dva grafy vedle sebe a to jeden nalevo s vizualizací rozložení tříd v augmentovaných datech a napravo překryv původních dat v modré barvě a dat augmentovaných v barvě červené.

```

1 def visualize(train_features, train_labels, val_features,
2     val_labels, name, path):
3
4     fig, axs = plt.subplots(ncols=2, figsize=(16, 8))
5
6     tsne = TSNE(n_components=2, random_state=42)
7     features_2d = tsne.fit_transform(train_features.reshape(
8         train_features.shape[0], -1))
9     axs[0].scatter(features_2d[:, 0], features_2d[:, 1], c=
10        train_labels, cmap="jet", alpha=0.5)
11    axs[0].set_title('t-SNE of augmented dataset features')
12
13    all_features = np.vstack((train_features, val_features))
14    labels = np.array([0] * len(train_features) + [1] * len(
15        val_features))
16
17    tsne = TSNE(n_components=2, random_state=42)
18    features_2d = tsne.fit_transform(all_features.reshape(
19        all_features.shape[0], -1))
20
21    train_2d = features_2d[:len(train_features)]
22    val_2d = features_2d[len(train_features):]
23
24    axs[1].scatter(val_2d[:, 0], val_2d[:, 1], c='blue',
25        alpha=0.7, label='Validation')
26    axs[1].scatter(train_2d[:, 0], train_2d[:, 1], c='red',
27        alpha=0.1, label='Augmented (train)')
28
29    fig.suptitle(f"Augmentations: {name}", fontsize=12)
30    fig.tight_layout()
31    plt.savefig(path)
32    plt.close(fig)
33
34    return features_2d

```

Výpis 4.11 Implementace t-SNE vizualizace

4.7 Implementace trénovacího algoritmu

Funkce `train_model` implementuje standardní tréninkový cyklus neuronové sítě pomocí knihovny PyTorch, kde se pomocí optimalizátoru Adam iterativně upravují váhy modelu za účelem minimalizace cross-entropy ztráty. Data jsou zpracovávána po dávkách a model je trénován po předem stanovený počet epoch. Po každé epoše se vypočítá a zaznamená průměrná ztráta, což umožňuje sledovat průběh učení modelu.

```
1 def train_model(model, train_loader):
2     criterion = nn.CrossEntropyLoss()
3     optimizer = optim.Adam(model.parameters(), lr=LR)
4     model.train()
5     epoch_losses = []
6     for epoch in range(EPOCHS):
7         total_loss = 0
8         for images, labels in train_loader:
9             images, labels = images.to(DEVICE), labels.to(
10                 DEVICE)
11             optimizer.zero_grad()
12             outputs = model(images)
13             loss = criterion(outputs, labels)
14             loss.backward()
15             optimizer.step()
16             total_loss += loss.item()
17             avg_loss = total_loss / len(train_loader)
18             epoch_losses.append(avg_loss)
19             print(f"Epoch [{epoch+1}/{EPOCHS}], Loss: {avg_loss
20                   :.4f}")
21     return epoch_losses
```

Výpis 4.12 Implementace trénovacího algoritmu

5 EXPERIMENTÁLNÍ VÝSLEDKY

Tato kapitola prezentuje výsledky experimentů, zaměřuje se na vyhodnocení dopadu augmentačních metod na datovém souboru CIFAR-10. Výsledky jsou strukturovány dle vybraných metrik a doplněny o t-SNE vizualizaci.

5.1 Rotace

Tato augmentace představuje transformaci původních dat pootočením orientace obrazových vzorků o ± 30 stupňů, s cílem zkoumání jak tato jednoduchá geometrická transformace ovlivní reprezentaci dat a klasifikační schopnosti modelu. Pro tuto augmentaci byly získány následující hodnoty popsané v tabulce 5.1:

Tabulka 5.1 Tabulka hodnot pro rotaci

Metrika	Hodnota
Rozptylová diverzita	0.0691
Index afinity	0.5977
Index diverzity	0.3251
Pokrytí prostoru příznaků	18.25 %
Skóre posunu domény	0.9357
Čistota shluků	0.5376
KL Divergence	0.0007
Siluetové skóre	0.2590

(zdroj: vlastní)

5.1.1 Interpretace výsledků

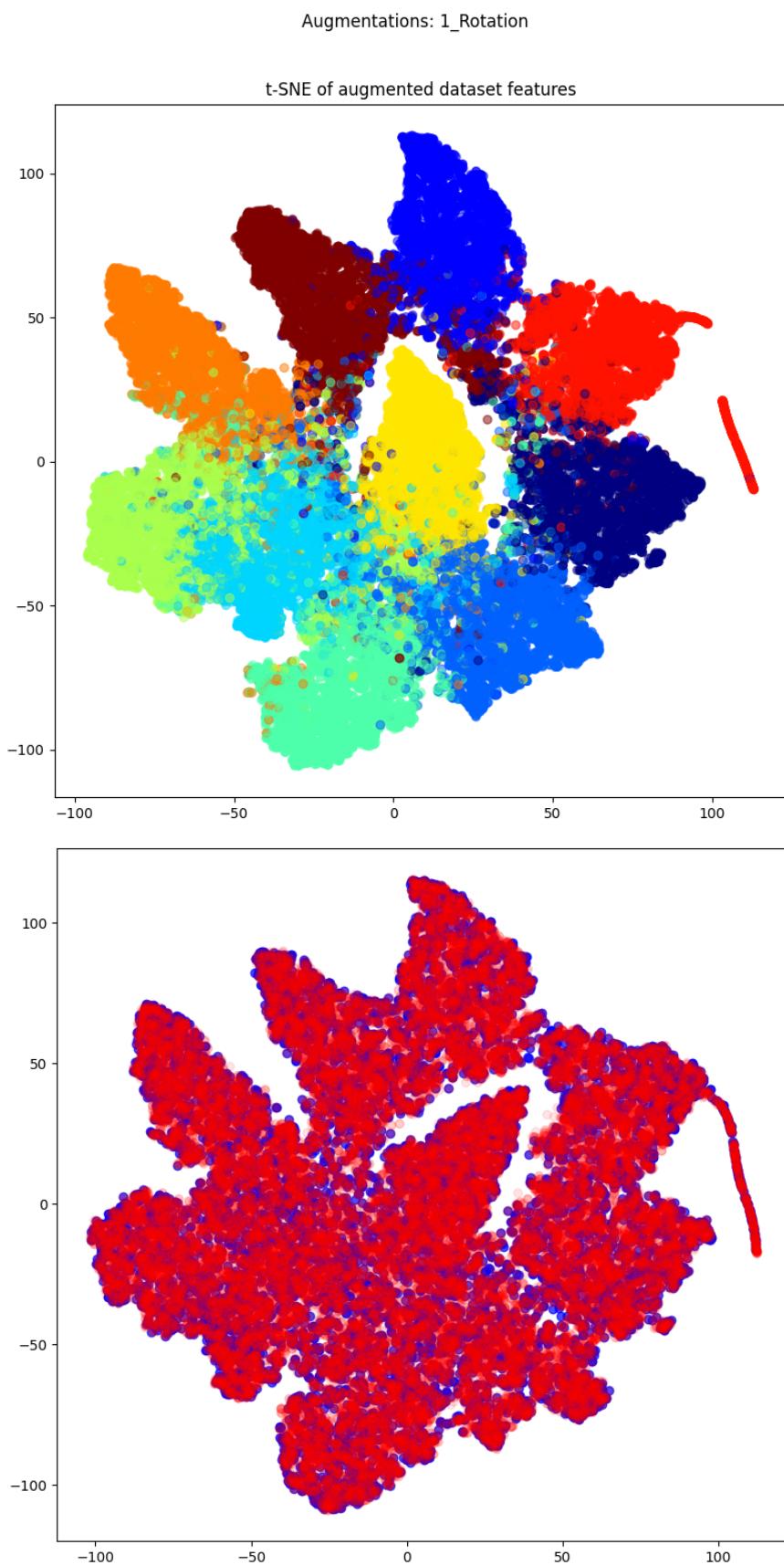
- **Rozptylová diverzita: 0.0691** Tato hodnota naznačuje velice mírné zvýšení rozmanitosti v predikcích modelu po aplikaci rotace, což znamená mírné obohacení datového souboru v rámci schopnosti modelu rozpoznávat objekty.
- **Index afinity: 0.5977** Tato hodnota naznačuje, že predikce na augmentovaných datech se přibližuje původnímu modelu zhruba z 60%, to indikuje, že rotace nezpůsobuje významnou ztrátu ve schopnosti modelu rozpoznávat objekty.
- **Index diverzity: 0.3251** Hodnota této metriky značí, že model má větší chybu na

augmentovaných datech ve srovnání s těmi původními. Rotace tedy vnáší novou variabilitu do datového souboru.

- **Pokrytí prostoru příznaků: 18.25%** Augmentovaná data rozšiřují pokrytí prostoru příznaků o 18.25%, což může vést k lepší generalizační schopnosti modelu, když se setká s jinak natočenými objekty v dosud neviděných obrazových datech. Procento pokrytí je ale mírné a nedochází k výrazné změně.
- **Skóre posunu domény: 0.9357** Vysoká hodnota této metriky naznačuje, že klasifikátor dokáže dobře rozlišit původní data od těch augmentovaných což indikuje výrazný posun domény.
- **Čistota shluků: 0.5376** Tato hodnota značí, že rotace částečně narušuje shlukování vzorků, rotace tedy může částečně vést ke smíšení tříd.
- **KL Divergence: 0.0007** Velmi nízká hodnota Kullback-Leiblerovy divergence naznačuje, že pravděpodobnostní rozdělení augmentovaných a původních obrazových dat jsou si podobná. To indikuje, že rotace nemění významným způsobem statistické vlastnosti dat, model tedy vnímá augmentovaná data jako velice podobná těm původním.
- **Siluetové skóre: 0.2590** Tato hodnota indikuje, že rotace udržuje rozumnou shlukovou strukturu. Shluky jednotlivých tříd jsou částečně překryty, ale jsou natolik oddělené, aby byly rozpoznatelné.

5.1.2 Vizualizace t-SNE

Vizualizace t-SNE pro rotaci 5.1 se skládá ze dvou grafů, přičemž nahoře je vizualizace rozložení příznaků v augmentovaných datech. Zde jde vidět jisté větší překrývání datových bodů ve shlucích napříč třídami, zejména tak ve spodních shlucích. Dole se pak nachází překryv dvou t-SNE vizualizací, kde modrou barvou jsou zvýrazněny datové body původních dat a jsou překryty datovými body augmentované sady, které jsou vyznačeny barvou červenou. Drobná změna některých shluků mezi těmito dvěma grafy vychází ze stochastické podstaty algoritmu t-SNE a jeho zachovávání pouze lokální struktury a ne té globální.



Obrázek 5.1 Vizualizace t-SNE pro rotaci

(zdroj: vlastní)

5.2 Horizontální převrácení

Tato transformace převrací některé obrazové vzorky dle jejich horizontální osy. Děje se podle zadaného parametru zhruba v 50% případů. Cílem je zjistit, jestli tato augmentační metoda dokáže ovlivnit reprezentaci dat a klasifikační schopnosti modelu. Pro tuto augmentaci byly získány následující hodnoty 5.2:

Tabulka 5.2 Tabulka hodnot pro horizontální
převrácení

Metrika	Hodnota
Rozptylová diverzita	0.0797
Index afinity	0.8760
Index diverzity	0.0935
Pokrytí prostoru příznaků	236.53 %
Skóre posunu domény	0.5460
Čistota shluků	0.407
KL Divergence	0.0013
Siluetové skóre	0.2508

(zdroj: vlastní)

5.2.1 Interpretace výsledků

- **Rozptylová diverzita: 0.0797** Poměrně nízká, ale ne zcela zanedbatelná hodnota napovídá, že horizontální převrácení velmi mírně zvyšuje variabilitu výstupů modelu.
- **Index afinity: 0.8760** Vysoká hodnota této metriky značí, že model klasifikuje převrácená obrazová data velmi podobně jako ta původní. To znamená, že horizontální převrácení dobře zachovává sémantický obsah obrazových dat.
- **Index diverzity: 0.0935** Nízká hodnota zde značí, že převrácení obrazu vede k minimálnímu nárůstu chyby modelu ve srovnání s původními obrazovými daty. Model tedy považuje převrácená obrazová data za velmi podobné těm původním.
- **Pokrytí prostoru příznaků: 236.53%** Tato augmentace významně rozšiřuje pokrytí prostoru příznaků a to o více jak dvojnásobek. To značí, že i když se augmentované

vzorky statisticky tolik neliší od původních vzorků, tak ve vícedimenzionálním prostoru mají podstatně větší objem. To je pozitivní pro robustnost modelu, protože se učí zobecňovat větší spektrum variací v datech.

- **Skóre posunu domény: 0.5460** Doménový klasifikátor dokáže rozlišit augmentovaná data od původních s průměrnou přesností, což znamená, že tyto nové vzorky se nijak výrazně neliší od původních. Z tohoto hlediska je tato augmentace bezpečná.
- **Čistota shluků: 0.407** Nižší čistota shluků naznačuje, že augmentace způsobuje větší překryv mezi třídami.
- **KL Divergence: 0.0013** Malý rozdíl mezi distribučními vlastnostmi originálních a augmentovaných dat naznačuje, že horizontální převrácení nikterak zásadně nenarušuje celkovou distribuci. Model považuje takto augmentované vzorky za podobné původním.
- **Siluetové skóre: 0.2508** Tato hodnota naznačuje, že po horizontálním převrácení si data v prostoru udržují docela rozpoznatelnou shlukovou strukturu s jistým překryvem. Sice nejde o dokonalé rozdělení a překryvy existují, ale jsou dostatečně koherentní.

5.2.2 Vizualizace t-SNE

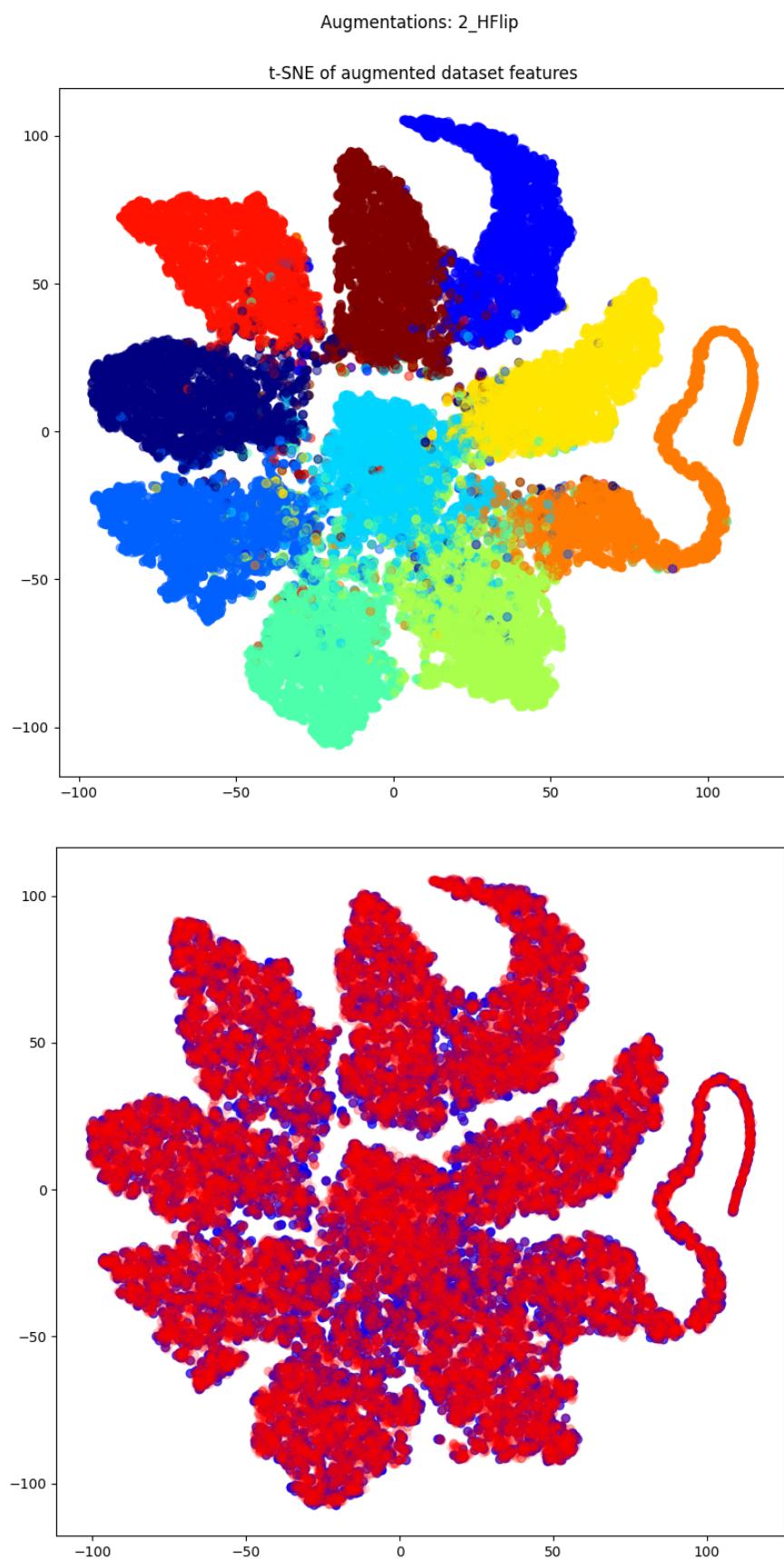
Vizualizace t-SNE pro horizontální převrácení 5.2 se skládá ze dvou grafů. V horním jde opět vidět mírné překrývání některých datových bodů napříč třídami. Nevznikají žádné významné změny mezi shluky, je zde ale mírné překrývání tříd. V překryvu původních a augmentovaných dat lze vidět, že shluky drží stejnou kompozici, ale některé datové body se posunuly.

5.3 Perspektivní transformace

Tato augmentační metoda se snaží simulovat změnu v perspektivitě obrazu, jakoby byl pozorován z jiného úhlu. Implementace augmentace je s parametry transformace maximálního zkreslení 0.5 pro každý prvek. Pro tuto augmentaci byly získány následující hodnoty 5.3:

5.3.1 Interpretace výsledků

- **Rozptylová diverzita: 0.0565** Mezi augmentacemi tato hodnota představuje tu nejnižší, což naznačuje, že perspektivní transformace nepřidává významnou rozmanitost



Obrázek 5.2 Vizualizace t-SNE pro horizontální převrácení

(zdroj: vlastní)

Tabulka 5.3 Tabulka hodnot pro perspektivní transformace

Metrika	Hodnota
Rozptylová diverzita	0.0565
Index affinity	0.3134
Index diverzity	0.5526
Pokrytí prostoru příznaků	31.14 %
Skóre posunu domény	0.9793
Čistota shluků	0.4215
KL Divergence	0.0109
Siluetové skóre	0.1191

(zdroj: vlastní)

do predikcí modelu.

- **Index affinity: 0.3134** Velmi nízká hodnota, která je nejnižší z augmentací, znamená, že model natrénovaný na původních datech má velké potíže s klasifikací perspektivně augmentovaných dat. Tedy tato augmentace významně mění způsob jakým model data vidí.
- **Index diverzity: 0.5526** Index diverzity je u této augmentace nejvyšší, což koreluje s povahou této augmentace, která silně mění rysy v obrazových datech, tudíž zvětne diverzitu v datech. Ovšem s ohledem ostatních metrik, je tohle zvýšení diverzity nestabilní.
- **Pokrytí prostoru příznaků: 31.14%** Tato hodnota značí, že pokrytí prostoru příznaků se zvětšilo zhruba o třetinu, tedy nikterak významně.
- **Skóre posunu domény: 0.9793** Velice vysoké skóre posunu domény naznačuje, že doménový klasifikátor dokáže snadno rozlišit mezi augmentovanými a původními vzorky. To koreluje s povahou augmentace, která značně mění obrazová data a klasifikátor je poměrně účinný tuto změnu odhalit.
- **Čistota shluků: 0.4215** Nižší čistota shluků naznačuje, že tato augmentace částečně narušuje shlukování tříd. Shluky se překrývají a nejsou tak kompaktní.

- **KL Divergence: 0.0109** Nízká hodnota KL divergence značí, že tato augmentace výrazně nemění statistické vlastnosti dat.
- **Siluetové skóre: 0.1191** Tato augmentace má nejnižší siluetové skóre, což koreluje s nižší hodnotou čistotu shluků. Nejvíce tedy narušuje shlukovou strukturu.

5.3.2 Vizualizace t-SNE

Vizualizace t-SNE pro horizontální převrácení 5.3 se skládá ze dvou grafů, přičemž nahoře je vizualizace rozložení příznaků v augmentovaných datech. U této augmentace jde vidět, že shluky tříd se značně překrývají a nejsou skoro vůbec homogenní, tedy jde vidět, že tato augmentace opravdu významně narušuje strukturu shluků tříd.

5.4 Změna barevnosti

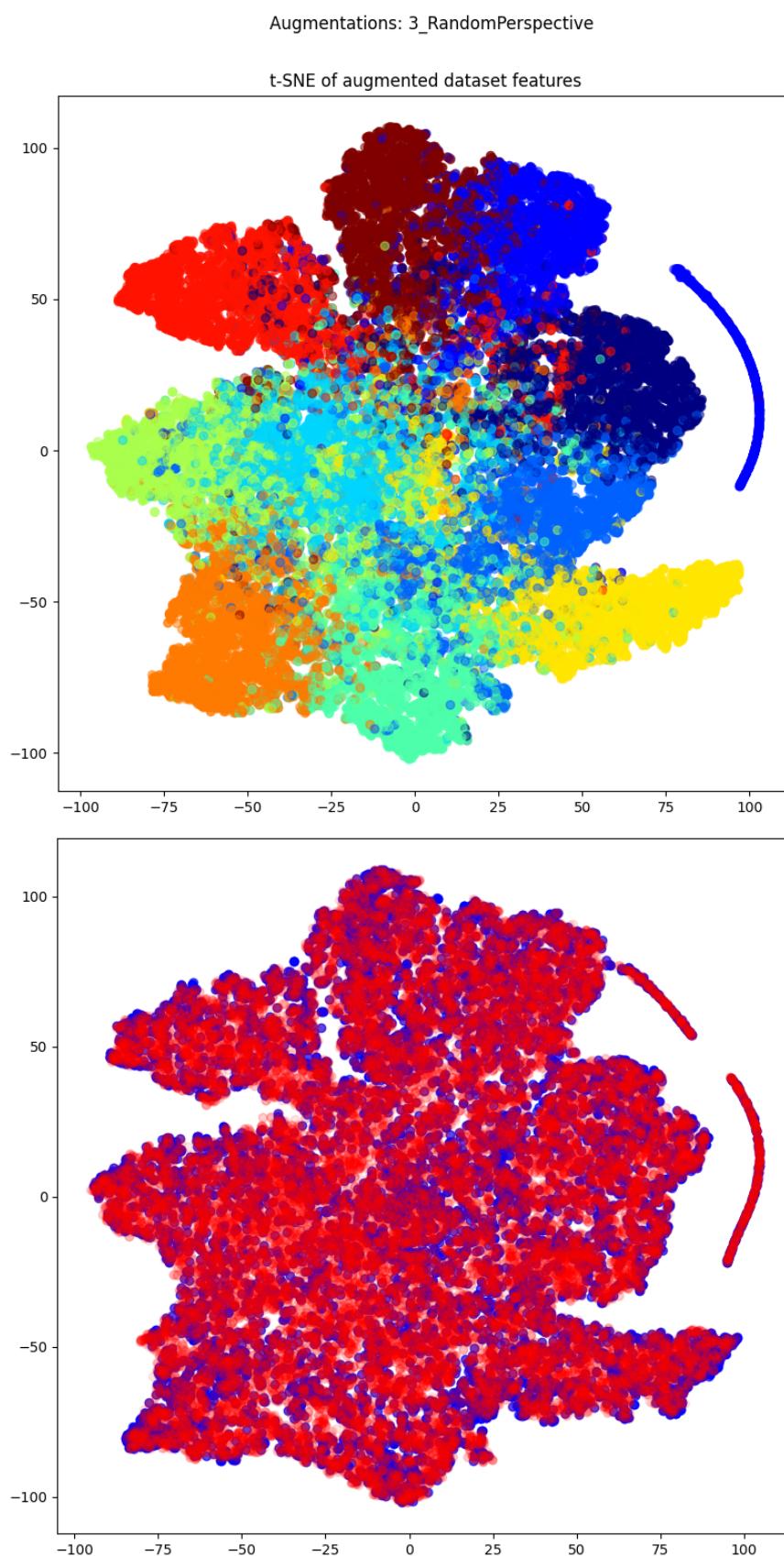
Augmentace změny barevnosti je technikou náhodné úpravy barevnosti obrazových vzorků. Konkrétně tak jasu, kontrastu, saturace a odstínu. Parametry pro tuto augmentaci byly následující:

- Jas: 0.4
- Kontrast: 0.4
- Saturace: 0.4
- Odstín: 0.1

Tím jsou simulovaly různé světelné podmínky a barvení. Pro tuto augmentaci byly získány následující hodnoty popsané v tabulce 5.4:

5.4.1 Interpretace výsledků

- **Rozptylová diverzita: 0.0828** Tato nízká hodnota značí, že tato augmentace velice mírně zvyšuje rozmanitost datového souboru oproti původním datům. To je sice žádoucí, ale jedná se o mírné zlepšení.
- **Index affinity: 0.8799** Vysoká hodnota této metriky indikuje, že model natrénovaný na původních datech zvládne s velkou přesností klasifikovat takto augmentovaná data.



Obrázek 5.3 Vizualizace t-SNE pro perspektivní transformace

(zdroj: vlastní)

Tabulka 5.4 Tabulka hodnot pro změnu
barevnosti

Metrika	Hodnota
Rozptylová diverzita	0.0828
Index afinity	0.8799
Index diverzity	0.0793
Pokrytí prostoru příznaků	24.51 %
Skóre posunu domény	0.6280
Čistota shluků	0.6181
KL Divergence	0.0005
Siluetové skóre	0.2662

(zdroj: vlastní)

Model je tedy schopen generalizovat na nové barevné variace bez významné ztráty přesnosti.

- **Index diverzity: 0.0793** Nízká hodnota zde může značit, že tato augmentace způsobuje minimální nárůst chyby modelu, model tedy považuje nové augmentované vzorky jakoby součást tréninkových dat. To koreluje s nízkou KL divergencí a vysokým afinity indexem.
- **Pokrytí prostoru příznaků: 24.51%** Tato augmentace mírně rozšiřuje pokrytí prostoru příznaků. Shluky tedy nejsou významně upraveny, pouze rozšířeny, což je očekávané, protože tato augmentace je pouze fotometrickou transformací, která nemění sémantický obsah ani geometrii objektů.
- **Skóre posunu domény: 0.6280** Doménový klasifikátor rozlišoval mezi augmentovanými a původními prvky s lehce nadprůměrnou přesností. Barevné změny v obrazových vzorcích jsou tedy stále příbuzné původním obrazovým vzorkům.
- **Čistota shluků: 0.6181** Lehce nadprůměrná hodnota této metriky značí, že struktura shluků zůstala relativně zachována. Augmentace tedy nezpůsobuje výrazné promíchaní tříd.
- **KL Divergence: 0.0005** Velice nízká hodnota této metriky indikuje, že distribuce augmentovaných a původních dat je statisticky téměř identická. Model tedy vnímá

tyto barevné variace v datech jako velmi podobné originálům.

- **Siluetové skóre: 0.2662** Tato hodnota naznačuje, že po této augmentaci je udržena do- cela rozumná shluková struktura v prostoru příznaků. Shluky nejsou zcela kompaktní a jsou částečně překryty, ale jsou dostatečně oddělené.

5.4.2 Vizualizace t-SNE

Vizualizace t-SNE pro změnu barevnosti 5.4 je složena ze dvou grafů. V horním z grafů jde vidět mírnější překrývání tříd než u předchozích augmentací, ale stále se zde vyskytuje, shluky jako takové vypadají být dobře odděleny. Dole se pak nachází překryv dvou t-SNE vizualizací, kde jde vidět že shluky mají velmi podobnou strukturu, přičemž některé datové body jsou po augmentaci posunuty.

5.5 Stupnice šedi

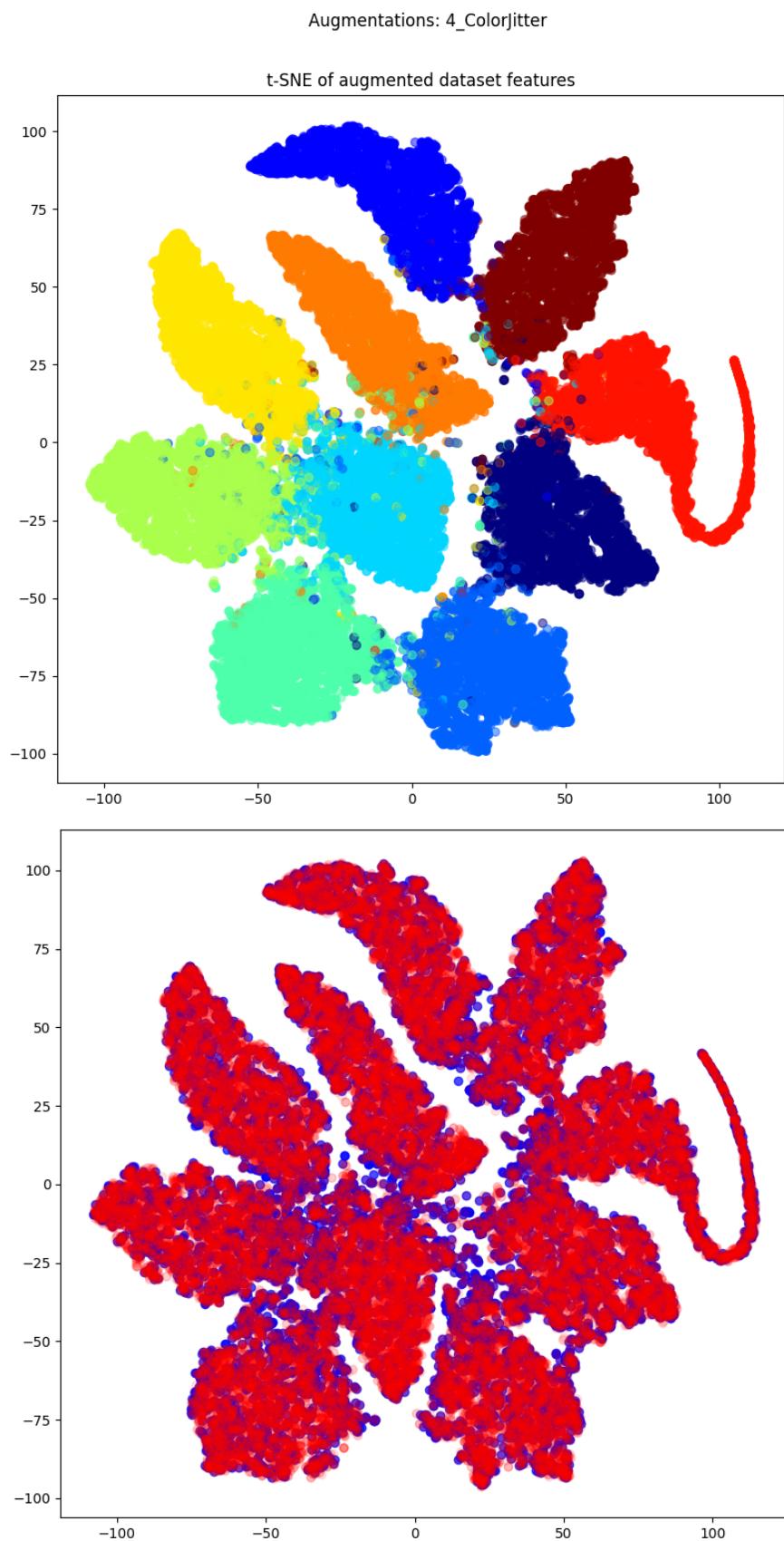
Tato augmentační metoda transformuje barevné obrazové vzorky na černobílé, tady záměrně způsobujeme ztrátu informace o barevnosti obrazu, což nutí model učit se více například z tvarových rysů v obrazových datech.

Pro tuto augmentaci byly získány následující hodnoty popsané v tabulce 5.5:

Tabulka 5.5 Tabulka hodnot pro stupnici šedi

Metrika	Hodnota
Rozptylová diverzita	0.0846
Index afinity	0.8148
Index diverzity	0.0769
Pokrytí prostoru příznaků	-7.99 %
Skóre posunu domény	0.9486
Čistota shluků	0.6133
KL Divergence	0.0012
Siluetové skóre	0.2584

(zdroj: vlastní)



Obrázek 5.4 Vizualizace t-SNE pro změnu barevnosti

(zdroj: vlastní)

5.5.1 Interpretace výsledků

- **Rozptylová diverzita: 0.0846** Tato hodnota je nejvyšší mezi augmentacemi, ale i tak je velmi nízká. Vede tedy jen k mírnému rozšíření rozmanitosti augmentovaných dat.
- **Index afinity: 0.8148** Poměrně vysoký Index afinity naznačuje, že model trénovaný na barevných datech si stále dobře klasifikuje obrazová data ve stupnici šedi. To potvrzuje jeho schopnost rozpoznávat objekty i bez barevné informace.
- **Index diverzity: 0.0769** Tato hodnota značí, že augmentace nevedou k velké variabilitě, to koreluje s nízkou rozptylovou diverzitou.
- **Pokrytí prostoru příznaků: -7.99 %** Tato hodnota značí, že pokrytí prostoru příznaků se po augmentaci lehce zmenšilo. To nejspíše proto, že se ztrácí barevná dimenze, která přispívala k širšímu rozprostření dat.
- **Skóre posunu domény: 0.9486** Velmi vysoká hodnota této metriky indikuje, že doménový klasifikátor s velmi výraznou přesností dokáže odlišit mezi augmentovanými a původními daty. To koreluje s povahou této augmentace, která vynechává barevnou informaci obrazu a vytváří tak zřetelný rozdíl, který klasifikátor dokáže dobře rozlišit.
- **Čistota shluků: 0.6133** Tato hodnota značí, že shluky jednotlivých tříd zůstávají relativně dobře seskupené. Ztráta barevné informace tedy nijak zásadně nenarušuje separabilitu tříd.
- **KL Divergence: 0.0012** Nízká hodnota zde ukazuje, že statické rozdělení augmentovaných a původních dat jsou si velmi podobná. Model i přes ztrátu barevné informace je schopen extrahovat podobné rysy.
- **Siluetové skóre: 0.2584** Jako u předchozích augmentací je tato hodnota v podobném rozsahu, což znamená, že shluky tříd v prostoru mají rozumnou strukturu, ale vyskytují se překryvy shluků některých tříd.

5.5.2 Vizualizace t-SNE

Vizualizace t-SNE pro stupnici šedi 5.5 se skládá ze dvou grafů. V horním grafu s vizualizací rozložení příznaků jde vidět pouze částečné překrývání shluků tříd, ale shluky jako takové

vypadají být dobře odděleny a koherentní, udržují lepší strukturu, než u ostatních augmentací. Dole se pak nachází překryv dvou t-SNE vizualizací, kde jde vidět že shluky mají velmi podobnou strukturu, ovšem u okrajů některých tříd jsou nové body lehce posunuty a můžeme vidět jakoby původní modrá data ty augmentovaná "lemovaly". Opět rozdíl ve tvaru a pozici některých shluků mezi dvěma grafy je vysvětlen stochastickou povahou t-SNE vizualizace.

5.6 Gaussův šum

Tato augmentace přidává do obrazových vzorků náhodný šum s normálním rozdělením. Parametry Gaussova šumu pro tuto augmentaci byly průměr $\text{mean} = 0$ a směrodatná odchylka $\text{std} = 0.1$. Cílem této augmentace je napodobit reálný šum v obrazových datech a zvýšit tak robustnost modelu vůči šumu.

Pro tuto augmentaci byly získány následující hodnoty popsané v tabulce 5.6:

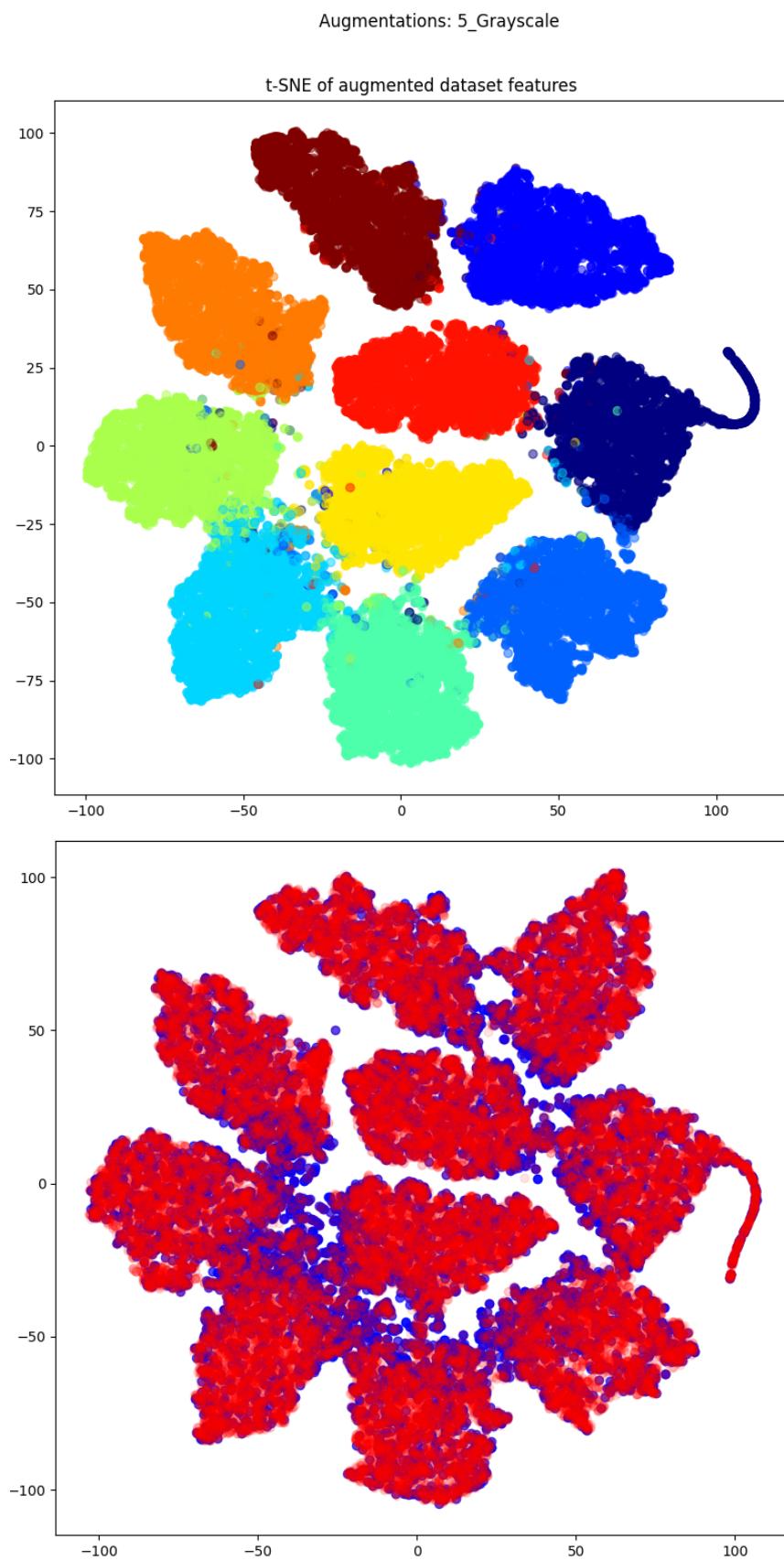
Tabulka 5.6 Tabulka hodnot pro Gaussův šum

Metrika	Hodnota
Rozptylová diverzita	0.0838
Index afinity	0.8709
Index diverzity	0.0603
Pokrytí prostoru příznaků	72.62 %
Skóre posunu domény	0.5490
Čistota shluků	0.5922
KL Divergence	0.0032
Siluetové skóre	0.2505

(zdroj: vlastní)

5.6.1 Interpretace výsledků

- **Rozptylová diverzita: 0.0838** Poměrně nízká hodnota rozptylové diverzity indikuje, že tato augmentace přidává určitou rozmanitost, ale ne zdaleka zásadní.
- **Index afinity: 0.8709** Relativně vysoká hodnota této metriky znamená, že většina augmentovaných vzorků je stále podobná originálním datům, i přes přidaný šum.
- **Index diverzity: 0.0603** Nízká hodnota indexu diverzity naznačuje, že augmentace



Obrázek 5.5 Vizualizace t-SNE pro stupnici šedi

(zdroj: vlastní)

nevnesla dostatečnou diverzitu do datového souboru, to ovšem koreluje s nízkou rozptylovou diverzitou.

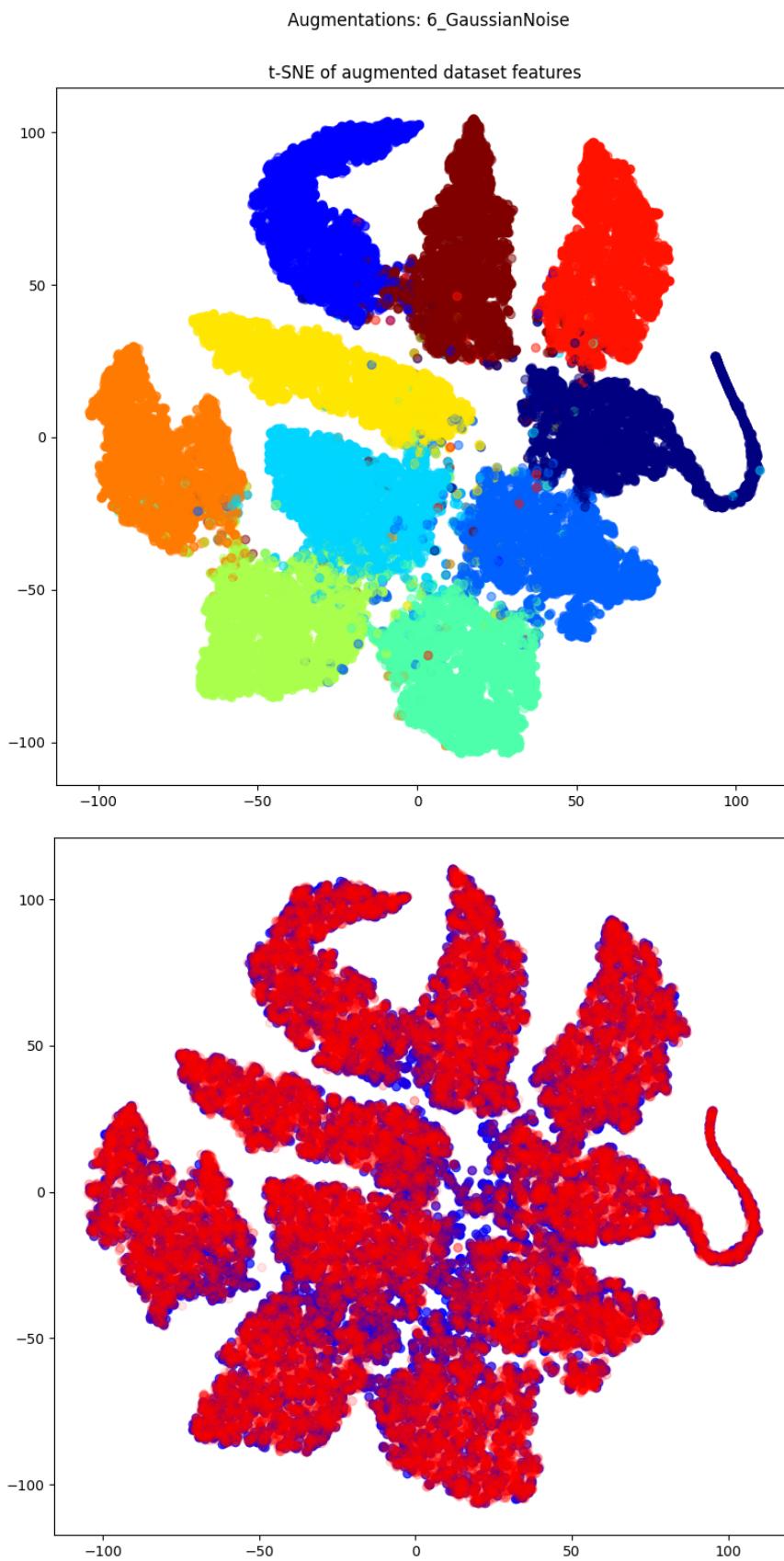
- **Pokrytí prostoru příznaků: 72.62%** Tato hodnota naznačuje poměrně výrazné rozšíření oblasti, kterou nově augmentovaná data zaujímají v prostoru příznaků. To může modelu pomoci zobecňovat širší spektrum variací v datech.
- **Skóre posunu domény: 0.5490** Tato hodnota značí, že doménový klasifikátor dokáže odlišit augmentované a původní data se zhruba 55% účinností. Tedy šum je modelem vnímán spíše jako varianta než jako odlišná doména.
- **Čistota shluků: 0.5922** Tato hodnota indikuje, že i po augmentaci šumem jednotlivé třídy zachovávají poměrně dobrou konzistenci. Šum tedy významně nenarušuje sémantickou separabilitu tříd.
- **KL Divergence: 0.0032** Hodnota této metriky je o něco vyšší než u ostatních augmentací, ale je stále poměrně dost nízká. Distribuce rysů je tedy i po augmentaci šumem velmi podobná té původní.
- **Siluetové skóre: 0.2505** Tato hodnota naznačuje, že shluky jsou rozpoznatelné, i když nejsou dokonale oddělené, což potvrzuje, že šum nemění sémantický obsah obrazových vzorků.

5.6.2 Vizualizace t-SNE

Vizualizace t-SNE pro stupnici šedi 5.6 se skládá ze dvou grafů, přičemž nahoře je vizualizace rozložení příznaků v augmentovaných datech. I u tohoto grafu jde vidět, že dochází k mírnému překrývání tříd, ale shluky jsou relativně dobře odděleny. Dole v grafu, který ukazuje překryv původních a augmentovaných dat jde vidět že shluky drží svou strukturu i po augmentaci, jen některé body se posunuly oproti původním.

5.7 Náhodné mazání

Tato augmentace představuje techniku, při které je vymazána náhodná obdélníková oblast v obrazovém vzorku. Implementace této augmentace je pomocí funkce `transforms.RandomErasing()`, která má standardní parametr rozsahu vymazané oblasti 0.02 až 0.33. Cílem je simulovat zakrytí částí objektů v reálných obrazových datech.



Obrázek 5.6 Vizualizace t-SNE pro Gaussův šum

(zdroj: vlastní)

Pro tuto augmentaci byly získány následující hodnoty popsané v tabulce 5.7:

Tabulka 5.7 Tabulka hodnot pro náhodné mazání

Metrika	Hodnota
Rozptylová diverzita	0.0772
Index afinity	0.8218
Index diverzity	0.2048
Pokrytí prostoru příznaků	-0.43 %
Skóre posunu domény	0.6025
Čistota shluků	0.4772
KL Divergence	0.0003
Siluetové skóre	0.2383

(zdroj: vlastní)

5.7.1 Interpretace výsledků

- **Rozptylová diverzita: 0.0772** Tato hodnota je v souladu se zbytkem augmentací a značí, že došlo k pouze velmi mírnému obohacení rozmanitosti dat.
- **Index afinity: 0.8218** Tato poměrně vysoká hodnota indexu afinity naznačuje, že model trénovaný na úplných datech si v celku dobré poradí s klasifikací obrazových vzorků s náhodně vymazanými částmi.
- **Index diverzity: 0.2048** Hodnota indexu diverzity je poměrně vysoká oproti ostatním augmentacím, i když obecně je nízká. Vyšší hodnota může být vysvětlena tak, že odstranění části obrazu představuje pro model větší výzvu než pouhá změna orientace nebo barevnosti
- **Pokrytí prostoru příznaků: -0.43%** Tato hodnota naznačuje, že náhodné mazání nezvyšuje objem dat v prostoru příznaků, dokonce jej trochu zmenšuje. Tato augmentace tedy prostor příliš nemění.
- **Skóre posunu domény: 0.6025** Tato poměrně vyšší hodnota indikuje, že tato augmentace zavádí jistý doménový posun, ale ne nijak významný.

- **Čistota shluků: 0.4772** Čistota shluků je menší než u jiných augmentací a znamená to, že třídy se více překrývají a jsou méně kompaktní, ale ne výrazně.
- **KL Divergence: 0.0003** Tato velmi nízká hodnota značí, že distribuce rysů původních a náhodně vymazaných obrazových vzorků jsou si statisticky téměř identické
- **Siluetové skóre: 0.2383** Tato hodnota je v podobném rozsahu jako u ostatních augmentací, dochází k překrytí shluků tříd, ale shluky jsou docela rozpoznatelné.

5.7.2 Vizualizace t-SNE

Vizualizace t-SNE pro stupnici šedi 5.7 se skládá ze dvou grafů, přičemž nahoře je vizualizace rozložení příznaků v augmentovaných datech. Zde jde dobře vidět překryv některých tříd, který se zdá být výraznější než u ostatních, což koreluje s nižší hodnotou čistoty shluků. Dole se pak nachází překryv dvou t-SNE vizualizací, kde jde vidět že struktura shluků je zachována. I zde jsou drobné rozdíly mezi dvěma grafy způsobeny stochastickou povahou t-SNE vizualizace.

5.8 Porovnání a celkové zhodnocení

V této podkapitole budou shrnutý a porovnány výsledky všech augmentačních experimentů, s cílem identifikovat, které augmentace nejfektivněji přispívají k lepší generalizaci a robustnosti modelu.

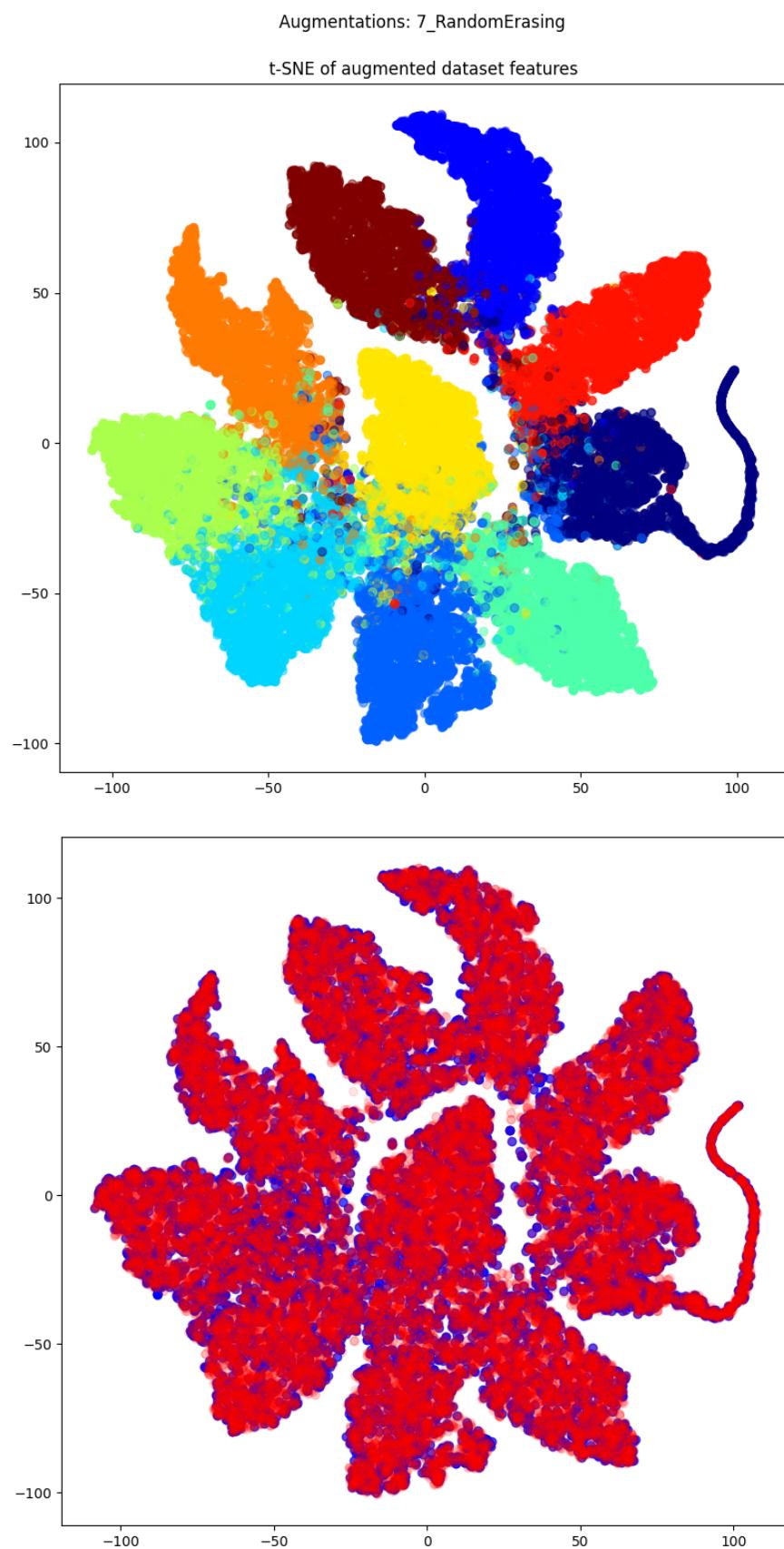
V následující tabulce 5.8 jsou shrnutý hodnoty všech metrik pro každou z augmentací.

Tabulka 5.8 Souhrnná tabulka hodnot metrik pro všechny augmentační metody

Metrika	Rotace	H. převrácení	Perspektiva	Barevnost	St. šedi	Šum	Mazání
Rozptylová div.	0.0691	0.0797	0.0565	0.0828	0.0846	0.0838	0.0772
Index afinity	0.5977	0.8760	0.3134	0.8799	0.8148	0.8709	0.8218
Index diverzity	0.3251	0.0935	0.5526	0.0793	0.0769	0.0603	0.2048
Pokrytí pros. př. [%]	18.25	236.53	31.14	24.51	-7.99	72.62	-0.43
Posun domény	0.9357	0.5460	0.9793	0.6280	0.9486	0.5490	0.6025
Čistota shluků	0.5376	0.4070	0.4215	0.6181	0.6133	0.5922	0.4772
KL Divergence	0.0007	0.0013	0.0109	0.0005	0.0012	0.0032	0.0003
Siluetové skóre	0.2590	0.2508	0.1191	0.2662	0.2584	0.2505	0.2383

(zdroj: vlastní)

Z analýzy jednotlivých augmentací a jejich výstupních metrik vyplývá několik klíčových poznatků. Hodnota KL divergence je u všech augmentací extrémně nízká, v rozmezí 0.0003



Obrázek 5.7 Vizualizace t-SNE pro náhodné mazání

(zdroj: vlastní)

až 0.0109, to indikuje, že žádná z augmentací významně nemění globální statistické rozdělení rysů, modely tedy vnímají augmentované vzorky jako variace těch původních. Siluetové skóre se pohybuje mezi 0.1191 až 0.2662, tyto konzistentní hodnoty naznačují, že všechny augmentace relativně zachovávají strukturu shluků, ale dochází k určitým překryvům a to zejména u perspektivních transformací. Hodnoty pokrytí prostoru příznaků mají mezi sebou největší rozdíly a zde se nachází klíčová zjištění. Augmentace horizontálního převrácení zde vyniká s velkým nárůstem 236.53%, což znamená, že tato augmentace výrazně rozšířila objem latentního prostoru, který model dokázal pokrýt, to je pozitivní pro robustnost modelu. Dále Gaussův šum také výrazně rozšířil prostor příznaků a to o 72.62%, zatímco stupnice šedi a náhodné mazání jej mírně zmenšilo. Všechny augmentace měli poměrně vysoký index affinity, to značí že model trénovaný na původních datech si dobře poradí s klasifikací takto augmentovaných dat. Index diverzity byl ve většině augmentací poměrně nízký, ale u augmentací rotace (0.3251) a náhodného mazání (0.2048) byl vyšší, to znamená, že tyto augmentace vedou k mírnému obohacení rozmanitosti dat, více než u ostatních augmentací. Nejvyšší ale byl u perspektivních transformacích (0.5526), ale s ohledem ostatních metrik je tato augmentace poněkud nestabilní a narušuje strukturu shluků tříd. Skóre posunu domény bylo významné u augmentací rotace (0.9357) a stupnice šedi (0.9486), to značí, že doménový klasifikátor dokáže s vysokou přesností rozlišit, zda je obrazový vzorek původní nebo augmentovaný. U rotace to může být způsobeno změnami v orientaci rysů, u stupnice šedi je to ztrátou barevné informace. Všechny augmentace mají poměrně malou hodnotu rozptylové diverzity, což značí, že žádná z augmentací nevedla k výraznému zvýšení rozmanitosti v predikcích modelu. Dále augmentace změny barevnosti (0.6181) a stupnice šedi (0.6133) mezi augmentacemi udržují nejvyšší čistotu shluků, tedy i po těchto augmentacích zůstávají shluky tříd relativně homogenní.

6 ZÁVĚREČNÉ SHRNUTÍ A DOPORUČENÍ

Tato závěrečná kapitola shrnuje poznatky z provedených experimentů s augmentačními metodami v oblasti počítačového vidění. Na základě analýzy vlivu různých augmentací na rozmanitost datového souboru, t-SNE vizualizací a porovnání metrik budou formulována doporučení pro výběr a použití augmentací v praktických scénářích.

V této práci jsme prozkoumali sedm augmentačních metod: rotaci, horizontální převrácení, perspektivní transformace, změnu barevnosti, stupnici šedi, Gaussův šum a náhodné mazání. Každá z těchto augmentací byla aplikovaná na datový soubor CIFAR-10 a jejich dopady byly kvantifikovány metrikami: rozptylová diverzita, index afinity a diverzity, pokrytí prostoru příznaků, skóre posunu domény, čistota shluků, KL divergence a siluetové skóre. Poté k nim byla vytvořena t-SNE vizualizace.

6.1 Shrnutí a diskuse klíčových zjištění

- **Diverzita a robustnost:** Všechny augmentace měly pozitivní, ale mírný nárůst rozptylové diverzity, což indikuje, že přispívají k rozmanitosti predikcí modelu, ale jen velice mírně. Nejvíce tak v tomto ohledu augmentace stupnice šedi (0.0846) a Gaussův šum (0.0838). Dále byl u všech augmentací, vyjma perspektivních transformací vysoký index afinity, to ukazuje, že model natrénovaný na neupravených datech si poradí s klasifikací augmentovaných vzorků, je tedy robustní vůči těmto typům variací v datech. Naopak rotace a náhodné mazání sice vykázaly nižší index afinity a vyšší index diverzity, což značí, že pro model představují větší výzvu, ale zároveň nutí model učit se odolnosti vůči výraznějším geometrickým změnám a okluzi.
- **Vliv na prostor příznaků:** Zde vynikla augmentace horizontálního převrácení, která pokrytí prostoru příznaků více jak zdvojnásobila (236.5%), to může vést ke zlepšení generalizace modelu. Dále pak Gaussův šum také přispěl k významnějšímu rozšíření pokrytí (72.62%). Naopak stupnice šedi (-7.99%) a náhodné mazání (-0.43%) prostor příznaků zmenšily, což odpovídá k povahám těchto augmentací, kdy jsou ztraceny některé informace z obrazových dat, ale nejedná se o významnou změnu.
- **Doménový posun:** U rotace, stupnice šedi a perspektivních transformací se tato hodnota nacházela velmi vysoko, tedy doménový klasifikátor s velkou přesností dokázal

odlišit augmentovaná data od těch původních. To je v souladu s povahami těchto augmentací, které vytvářejí výrazný rozdíl v augmentovaných datech. U stupnice šedi je to způsobeno ztrátou informace o barvě, u rotace a perspektivních transformací se také mění struktura obrazu a rysy v nich jsou poměrně dost zkresleny.

- **Zachování sémantické struktury:** U všech augmentací byla získána velmi nízká hodnota KL divergence, což naznačuje, že statistické distribuce rysů původních a augmentovaných dat jsou si velmi podobné. Model tedy vnímá augmentovaná data jako varianty původních dat. Konzistentní bylo i siluetové skóre naznačující, že sémantická struktura tříd je v prostoru příznaků částečně zachována.

6.2 Doporučení pro výběr a využití augmentací

Výběr správné augmentace by měl být založen na specifických požadavcích daného scénáře z oblasti počítačového vidění a na charakteru datového souboru jež má obohatovat. Z experimentů provedených v této práci lze shrnout následující doporučení:

- **Zlepšení generalizace a rozšíření latentního prostoru:** Dle experimentů je zde horizontální převrácení dobrou volbou, protože tato augmentace dokázala významně rozšířit pokrytí prostoru příznaků s minimálním doménovým posunem. Hodí se do typu scénářů kde je relevantní horizontální symetrie, jako je například rozpoznávání obličejů lidí. Dále v tomto ohledu pomohl i Gaussův šum, který také významně rozšířil latentní prostor a je vhodný pro zvýšení robustnosti modelu vůči šumu v reálných datech.
- **Zvýšení robustnosti vůči proměnlivým podmínkám:** V tomto ohledu je ideální augmentace změny barevnosti, která se snaží simulovat různé světelné podmínky nebo barevné zkreslení, které se objevuje v reálných datech. Tato augmentace měla vysoký index affinity a poměrně rozumnou čistotu shluků. Dále augmentace stupnice šedi je dobrou volbou, pokud chceme, aby model byl nezávislý na barevné informaci a více se zaměřoval na texturální a tvarové rysy v obrazových datech. Je tedy dobrá volba pro scénáře, kde se pracuje s černobílými obrazy nebo v případech, kde je barva irelevantní.
- **Zvýšení robustnosti vůči geometrickým deformacím a okluzi:** Pokud se očekává,

že objekty v reálných datech konkrétního scénáře budou mít různé orientace, například při fotografování z různých úhlů nebo rotující objekty, nabízí se zde dvě augmentace a to rotace a perspektivní transformace. Ovšem u obou těchto augmentací byl naměřen velký posun domény a perspektivní transformace měli nežádoucí vliv na strukturu shluků tříd, tedy nejspíše představují pro model určitou výzvu. Dále se augmentace náhodného mazání zdá být dobrým nástrojem ve scénářích, kde objekty mohou být částečně zakryty a model se dokáže naučit je rozpoznávat i bez úplných informací.

6.3 Doporučení pro výběr augmentací pro konkrétní scénáře

- **Průmyslová kontrola kvality:** Modely určené pro detekci vad na produkčních lincech mohou těžit z augmentace Gaussova šumu, která dokáže nasimulovat šum ze senzorů nebo kamer. Ještě je zde cenná augmentace náhodného mazání, která dokáže simulovat částečné zakrytí objektů na produkční lince.
- **Lékařské snímky:** Zde je cenná robustnost vůči šumu a změn barevnosti. Byly by zde přínosné augmentace Gaussova šumu a změn barevnosti. Dále dle typu snímků i náhodné mazání nebo mazání specifické části obrazu, jako je například logo výrobce nebo jiné artefakty, které v reálném lékařském snímku mohou existovat.
- **Bezpečnostní systémy:** Systémy, které chrání nebo monitorují nějakou oblast mohou používat počítačové vidění na detekci osob nebo objektů. Zde se nabízí augmentace změny barevnosti pro simulaci reálných světelných podmínek a náhodného mazání pro simulaci okluze. Pro systémy používající černobílé kamery by byla vhodná augmentace stupnice šedi.

Závěrem, augmentace dat představuje dobrý nástroj pro trénink robustnějších modelů s lepší schopností generalizace v oblasti počítačového vidění. Pochopení dopadu jednotlivých augmentačních metod na metriky latentního prostoru a t-SNE vizualizace je důležité pro informovaný výběr a optimalizaci tréninkových strategií modelů počítačového vidění, což může vést ke zlepšení výkonu modelů v reálných aplikacích.

ZÁVĚR

Tato práce se zabývala rolí augmentace dat v oblasti počítačového vidění, jejím cílem bylo prozkoumat, jak různé augmentační metody ovlivní rozmanitost datového souboru ale i generalizační schopnosti a robustnost modelů hlubokého učení. To se dělo na základě implementace a následné analýzy sedmi vybraných augmentačních metod - rotace, horizontálního převrácení, perspektivních transformacích, změny barevnosti, stupnici šedi, Gaussova šumu a náhodného mazání. Tyto augmentace byly aplikovány na datový soubor CIFAR-10 a následně byly zhodnoceny jejich dopady pomocí souboru metrik a t-SNE vizualizace.

Experimentální výsledky ukázaly, že každá z augmentací ovlivnila latentní prostor modelu trochu jiným způsobem. Konzistentní napříč všemi augmentacemi bylo siluetové skóre a velmi nízká KL divergence, což potvrzuje, že základní sémantická struktura a statistické vlastnosti augmentovaných dat je zachována. Model dokáže i přes aplikaci augmentací objekty v nich rozpoznat a tyto nová data vnímá jako varianty těch původních.

Nejvýznamnější zjištění z těchto experimentů byly:

- Z pohledu rozšíření prostoru příznaků nejvíce vynikla augmentace horizontálního převrácení, která jej více jak zdvojnásobila, přičemž si ale udržela poměrně nízký doménový posun, to z ní činí dobrou volbu pro zvýšení schopnosti modelu generalizovat.
- Gaussův šum se také ukázala jako užitečná augmentace, která významně rozšířila pokrytí prostoru příznaků s malým doménovým posunem a model tedy vnímal tyto augmentované data jako varianty těm původním. Tato augmentace je klíčová pro zvýšení robustnosti modelu vůči šumu, který se objevuje v reálných datech.
- Augmentace změny barevnosti, která simuluje různé světelné podmínky a barevné zkreslení, dokázala mírně zvětšit prostor příznaků s vysokou hodnotou indexu afinity. Stupnice šedi pak pomocí ztráty barevné informace nutí model se učit více z rysů objektů. Ta vedla mírnému zmenšení prostoru příznaků a výrazně změně posunu domény.
- Augmentace perspektivních transformací se ukázalo jako pro model nejnáročnější augmentací s nejnižším siluetovým skóre, nejvýraznějším doménovým posunem a nízkou čistotou shluků. V t-SNE vizualizaci této augmentace bylo vidět, že shluky tříd ztrácejí svou strukturu a dochází k jejich významnému překryvu. To značí, že způsobuje

výraznou deformaci dat, se kterou se model hůře vyrovnává.

Tyto poznatky mají uplatnění v nejrůznějších konkrétních scénářích od kontroly v průmyslové výrobě přes lékařské snímky a bezpečnostní systémy, kde model, který dokáže správně klasifikovat objekt navzdory variacím v obrazových datech, jako je například změna světelných podmínek, změna v orientaci, šum a okluze, je klíčový. Tato práce tak poskytla náhled do toho, jak tyto augmentace dat ovlivňují jejich reprezentaci v modelech hlubokého myšlení.

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>. [cit. 2025-02-28].
- [2] 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>. [cit. 2025-04-12].
- [3] SHORTEN, Connor a KHOSHGOFTAAR, Taghi M, 2019. A survey on image data augmentation for deep learning. *Journal of Big Data*. 2019.
- [4] AYYADEVARA, V. Kishore a REDDY, Yeshwanth, 2020. *Modern computer vision with PyTorch*. Packt. ISBN 978-1-83921-347-2.
- [5] ELGENDI, Mohamed; NASIR, Muhammad; TANG, Qunfeng; SMITH, David; GRENIER, John-Paul et al., 2021. The Effectiveness of Image Augmentation in Deep Learning Networks for Detecting COVID-19: A Geometric Transformation Perspective.
- [6] 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>. [cit. 2025-05-02].
- [7] UNIVERSITY, Northwestern, 2025. Analyzing and Enhancing Images (Image Processing Toolbox). Dostupné z: <http://www.ece.northwestern.edu/local-apps/matlabhelp/toolbox/images/enhanc12.html>. [cit. 2025-05-05].
- [8] YU, Junchuan; LI, Yichuan; ZHENG, Xiangxiang; ZHONG, Yufeng a HE, Peng, 2020. An Effective Cloud Detection Method for Gaofen-5 Images via Deep Learning.
- [9] DEVRIES, Terrance a TAYLOR, Graham W., 2017. Improved Regularization of Convolutional Neural Networks with Cutout. Dostupné z: <https://arxiv.org/abs/1708.04552>. [cit. 2025-05-05].

- [10] MILVUS, 2025. How does cutout augmentation work? Dostupné z: <https://blog.milvus.io/ai-quick-reference/how-does-cutout-augmentation-work>. [cit. 2025-05-08].
- [11] CHOI, Juhwan a KIM, YoungBin, 2024. Colorful Cutout: Enhancing Image Data Augmentation with Curriculum Learning. Dostupné z: <https://arxiv.org/abs/2403.20012>. [cit. 2025-05-08].
- [12] GOODFELLOW, Ian J.; POUGET-ABADIE, Jean; MIRZA, Mehdi; XU, Bing; WARDE-FARLEY, David et al., 2014. Generative Adversarial Networks. Dostupné z: <https://arxiv.org/abs/1406.2661>. [cit. 2025-05-12].
- [13] 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>. [cit. 2025-05-12].
- [14] 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>. [cit. 2025-05-12].
- [15] SARKER, Krishanu; YANG, Xiulong; LI, Yang; BELKASIM, Saeid a SHIAO, Ji, 2020. A Unified Plug-and-Play Framework for Effective Data Denoising and Robust Abstention.
- [16] LOPEZ, 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>. [cit. 2025-05-13].
- [17] LIU, Zirui; JIN, Haifeng; WANG, Ting-Hsiang; ZHOU, Kaixiong a HU, Xia, 2021. Di-vAug: Plug-in Automated Data Augmentation with Explicit Diversity Maximization. Dostupné z: <https://arxiv.org/abs/2103.14545>. [cit. 2025-05-14].
- [18] STATLECT, 2025. Domain shift — statlect.com. Dostupné z: <https://www.statlect.com/machine-learning/domain-shift>. [cit. 2025-05-06].

- [19] TENSORFLOW, 2025. cifar10 - TensorFlow Datasets — tensorflow.org. <https://www.tensorflow.org/datasets/catalog/cifar10>. [cit. 2025-05-19].
- [20] KRIZHEVSKY, Alex, 2009. Learning multiple layers of features from tiny images.

SEZNAM OBRÁZKŮ

Obrázek 1.1	Ukázka geometrických transformací	20
Obrázek 1.2	Ukázka intenzivních transformací	21
Obrázek 1.3	Ukázka pokročilých augmentačních metod.....	23
Obrázek 2.1	Ukázka t-SNE na datasetu CIFAR-10.....	28
Obrázek 4.1	Ukázka datasetu CIFAR-10	36
Obrázek 4.2	Vizualizace t-SNE pro CIFAR-10.....	38
Obrázek 4.3	Ukázka augmentace: Rotace	38
Obrázek 4.4	Ukázka augmentace: Horizontální převrácení.....	39
Obrázek 4.5	Ukázka augmentace: Perspektivní transformace.....	40
Obrázek 4.6	Ukázka augmentace: Změna barevnosti	40
Obrázek 4.7	Ukázka augmentace: Stupnice šedi	41
Obrázek 4.8	Ukázka augmentace: Gaussův šum	41
Obrázek 4.9	Ukázka augmentace: Náhodné mazání.....	42
Obrázek 5.1	Vizualizace t-SNE pro rotaci	52
Obrázek 5.2	Vizualizace t-SNE pro horizontální převrácení	55
Obrázek 5.3	Vizualizace t-SNE pro perspektivní transformace.....	58
Obrázek 5.4	Vizualizace t-SNE pro změnu barevnosti	61
Obrázek 5.5	Vizualizace t-SNE pro stupnici šedi	64
Obrázek 5.6	Vizualizace t-SNE pro Gaussův šum	66
Obrázek 5.7	Vizualizace t-SNE pro náhodné mazání	69

SEZNAM TABULEK

Tabulka 4.1	Tabulka použitých augmentací.....	39
Tabulka 4.2	Tabulka vybraných metrik	42
Tabulka 5.1	Tabulka hodnot pro rotaci.....	50
Tabulka 5.2	Tabulka hodnot pro horizontální převrácení	53
Tabulka 5.3	Tabulka hodnot pro perspektivní transformace.....	56
Tabulka 5.4	Tabulka hodnot pro změnu barevnosti	59
Tabulka 5.5	Tabulka hodnot pro stupnici šedi	60
Tabulka 5.6	Tabulka hodnot pro Gaussův šum	63
Tabulka 5.7	Tabulka hodnot pro náhodné mazání	67
Tabulka 5.8	Souhrnná tabulka hodnot metrik pro všechny augmentační metody	68

SEZNAM VÝPISŮ

Výpis 4.1	Implementace augmentací	39
Výpis 4.2	Implementace Gaussova šumu	40
Výpis 4.3	Funkce get_transform	41
Výpis 4.4	Implementace rozptylové diverzity	43
Výpis 4.5	Implementace indexů affinity a diverzity	43
Výpis 4.6	Implementace pokrytí prostoru příznaků	44
Výpis 4.7	Implementace skóre posunu domény	45
Výpis 4.8	Implementace čistoty shluků	47
Výpis 4.9	Implementace KL divergence	47
Výpis 4.10	Implementace siluetového skóre	47
Výpis 4.11	Implementace t-SNE vizualizace	48
Výpis 4.12	Implementace trénovacího algoritmu	49

SEZNAM POUŽITÝCH SYMBOLŮ A ZKRATEK

GAN	Generative Adversarial Network
PCA	Principal Component Analysis
SNE	Stochastic Neighbor Embedding
t-SNE	t-Distributed Stochastic Neighbor Embedding
g-SNE	global t-Distributed Stochastic Neighbor Embedding

SEZNAM PŘÍLOH

PŘÍLOHA P I: Zdrojový kód a výstupy

PŘÍLOHA P II: Práce

PŘÍLOHA P I: ZDROJOVÝ KÓD A VÝSTUPY

```
└── data-augmentation-pipeline.py
└── tsne-non-aug.png
└── experiments
    ├── 1_Rotation
    │   └── log.txt
    │   └── tsne.png
    ├── 2_HFlip
    │   └── log.txt
    │   └── tsne.png
    ├── 3_RandomPerspective
    │   └── log.txt
    │   └── tsne.png
    ├── 4_ColorJitter
    │   └── log.txt
    │   └── tsne.png
    ├── 5_Grayscale
    │   └── log.txt
    │   └── tsne.png
    ├── 6_GaussianNoise
    │   └── log.txt
    │   └── tsne.png
    └── 7_RandomErasing
        └── log.txt
        └── tsne.png
```

,

PŘÍLOHA P II: PRÁCE

fulltext.pdf