

Vizualizace a hodnocení augmentačních technik v počítačovém vidění

Bc. Petr Buček



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

PROHLÁŠENÍ AUTORA DIPLOMOVÉ PRÁCE

Beru na vědomí, že

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

Prohlašuji, že

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

Ve Zlíně, dne

.....
podpis autora

ABSTRAKT

Text abstraktu v jazyce práce.

Klíčová slova: klíčové slovo, klíčové slovo

ABSTRACT

Text abstraktu ve světovém jazyce (angličtině).

Keywords: keyword, keyword

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

OBSAH

ÚVOD	11
I TEORETICKÁ ČÁST	12
1 AUGMENTACE DAT V POČÍTAČOVÉM VIDĚNÍ	13
1.1 PROBLEMATIKA KVALITY A MNOŽSTVÍ DAT V POČÍTAČOVÉM VIDĚNÍ	13
1.2 ÚLOHA AUGMENTACE DAT PŘI ZLEPŠOVÁNÍ VÝKONU MODELŮ HLUBO- KÉHO UČENÍ	14
1.3 KLASIFIKACE A CHARAKTERISTIKA AUGMENTAČNÍCH METOD	14
1.3.1 Geometrické transformace (rotace, škálování, ořez, zrcadlení)	14
1.3.2 Intenzivní transformace (změna jasu, kontrastu, aplikace šumu)	15
1.3.3 Pokročilé augmentační metody (Cutout, Mixup, CutMix, GAN-based augmentace)	15
1.3.4 Doménově specifické augmentační techniky	15
2 VIZUALIZACE DOPADU AUGMENTACÍ	16
2.1 PRINCIP METODY T-SNE A JEJÍ MATEMATICKÝ ZÁKLAD	16
2.2 APLIKACE T-SNE V KONTEXTU ANALÝZY AUGMENTOVANÝCH DAT	16
2.3 MOŽNOSTI INTERPRETACE VÝSLEDKŮ T-SNE VIZUALIZACE	16
3 HODNOCENÍ Vlivu AUGMENTACÍ NA DIVERZITU DAT	17
3.1 METRIKY PRO KVANTIFIKACI DIVERZITY A ROZMANITOSTI DAT	17
3.2 ANALÝZA Vlivu JEDNOTLIVÝCH AUGMENTAČNÍCH TECHNIK NA DIS- TRIBUCI DAT	17
3.3 KOMPARATIVNÍ HODNOCENÍ EFEKTIVITY RŮZNÝCH PŘÍSTUPŮ K AUG- MENTACI	17
II PRAKTICKÁ ČÁST	18
ZÁVĚR	19
SEZNAM POUŽITÉ LITERATURY	21
SEZNAM OBRÁZKŮ	22
SEZNAM TABULEK	23

SEZNAM POUŽITÝCH SYMBOLŮ A ZKRATEK	24
SEZNAM PŘÍLOH.....	25

ÚVOD

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

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

I TEORETICKÁ ČÁST

1 AUGMENTACE DAT V POČÍTAČOVÉM VIDĚNÍ

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

1.1 Problematika kvality a množství dat v počítačovém vidění

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

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

1.2 Úloha augmentace dat při zlepšování výkonu modelů hlubokého učení

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

1.3 Klasifikace a charakteristika augmentačních metod

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

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

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

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

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

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

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

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

1.3.4 Doménově specifické augmentační techniky

2 VIZUALIZACE DOPADU AUGMENTACÍ

- 2.1 Princip metody t-SNE a její matematický základ**
- 2.2 Aplikace t-SNE v kontextu analýzy augmentovaných dat**
- 2.3 Možnosti interpretace výsledků t-SNE vizualizace**

3 HODNOCENÍ VLIVU AUGMENTACÍ NA DIVERZITU DAT

Tato kapitola se zabývá hodnocením vlivu augmentačních technik na diverzitu dat pro strojové učení. Zahrnuje metriky pro kvantifikaci augmentací a porovnává efektivitu jednotlivých přístupů.

3.1 Metriky pro kvantifikaci diverzity a rozmanitosti dat

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

3.2 Analýza vlivu jednotlivých augmentačních technik na distribuci dat

Metriky posunu distribuce:

- **Afinita:** Tato metrika určuje míru jak se posouvá distribuce augmentovaných dat od distribuce čistých dat, na kterých se model naučil. To nám poskytuje pohled na to jak se augmentovaná data odchyľují od modelu naučeného na čistých datech.
- **KL divergence :** Celým názvem Kullbackova-Leiblerova divergence, kvantifikuje míru rozdílnosti mezi dvěma pravděpodobnostními rozděleními. V kontextu augmentace dat pro strojové učení se používá na kontrolu změny rozdělení dat po augmentaci. Matematicky se značí jako $D_{KL}(P||Q)$ a nabývá hodnot $D_{KL}(P||Q) \geq 0$, 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:

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

3.3 Komparativní hodnocení efektivity různých přístupů k augmentaci

II PRAKTICKÁ ČÁST

ZÁVĚR

Text ...

SEZNAM POUŽITÉ LITERATURY

- [1] XU, Mingle; YOON, Sook; FUENTES, Alvaro a PARK, Dong Sun, 2023. A Comprehensive Survey of Image Augmentation Techniques for Deep Learning. *Pattern Recognition*. 2023. 137, s. 109347. ISSN 0031-3203. Dostupné z: <https://www.sciencedirect.com/science/article/pii/S0031320323000481>.
- [2] AYYADEVARA, V. Kishore a REDDY, Yeshwanth, 2020. *Modern computer vision with PyTorch*. Packt. ISBN 978-1-83921-347-2.
- [3] Fleet, Pajdla, Schiele,, Tuytelaars]computer-vision1 FLEET, David; PAJDLA, Tomáš; SCHIELE, Bernt a TUYTELAARS, Tinne, [2014]. *Computer vision - ECCV 2014*. Springer. Dostupné z: <https://doi.org/10.1007/978-3-319-10602-1>.
- [4] 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>.
- [5] 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>.
- [6] SHAPIRO, Linda G. a STOCKMAN, George C., c2001. *Computer vision*. Prentice Hall. ISBN 0-13-030796-3.
- [7] LOPES, Raphael Gontijo; SMULLIN, Sylvia J.; CUBUK, Ekin D. a DYER, Ethan, 2020. Affinity and Diversity: Quantifying Mechanisms of Data Augmentation. *CoRR*. 2020. abs/2002.08973. Dostupné z: <https://arxiv.org/abs/2002.08973>.
- [8] SHORTEN, Connor a KHOSHGOFTAAR, Taghi M, 2019. A survey on image data augmentation for deep learning. *Journal of Big Data*. 2019. 6, č. 1, s. 1–48.

SEZNAM OBRÁZKŮ

SEZNAM TABULEK

SEZNAM POUŽITÝCH SYMBOLŮ A ZKRATEK

CPU	Central Processing Unit
PTFE	Polytetrafluoroethylene
VNA	Vector Network Analyser

SEZNAM PŘÍLOH

PŘÍLOHA P I: První příloha

PŘÍLOHA P II: Druhá příloha

PŘÍLOHA P I: PRVNÍ PŘÍLOHA

Obsah přílohy

,

PŘÍLOHA P II: DRUHÁ PŘÍLOHA

Obsah přílohy