

Odhalenie vplyvu jasú obrázkov na presnosť modelu

1st Mariia Paik

Technická univerzita v Košiciach

mariia.paik@student.tuke.sk

Košice, Slovensko

2nd Yaroslav Zhuravskiy

Technická univerzita v Košiciach

yaroslav.zhuravskiy@student.tuke.sk

Košice, Slovensko

3rd Pavlo Smitiuk

Technická univerzita v Košiciach

pavlo.smitiuk@student.tuke.sk

Košice, Slovensko

4th Anatolii Porplytsia

Technická univerzita v Košiciach

anatolii.porplytsia@student.tuke.sk

Košice, Slovensko

5th Iryna Popovych

Technická univerzita v Košiciach

iryna.popovych@student.tuke.sk

Košice, Slovensko

Abstract—Klasifikačné modely obrázkov dosiahli v posledných rokoch významné pokroky, ale ich rozhodovacie procesy sú často neprehľadné. Táto štúdia si kladie za cieľ preskúmať vplyv manipulácie jasú obrázkov patriacich do konkrétnych tried v rámci datasetu na presnosť a významnosť klasifikačného modelu. Postupným prispôbovaním jasú jednej triedy nahor a druhej nadol, pričom tretiu triedu nechávame nezmenenú, skúmame, ako celkový jas príkladov tried ovplyvňuje rozhodovací proces modelu. Okrem toho je počas experimentu sledovaná presnosť modelu, aby sme pozorovali akékoľvek zmeny vo výkonnosti.

Index Terms—explainable ai, image brightness, image classification

I. ÚVOD

V posledných rokoch sme svedkami významného pokroku v oblasti umelej inteligencie (AI) a strojového učenia (ML), ktorý sa prejavuje najmä vo zlepšení schopností týchto systémov v rozpoznávaní obrazov a hlasu. Hlboké neurónové siete, ktoré sú hlavným pilierom mnohých pokročilých modelov, vykazujú vysokú presnosť v riešení úloh, avšak sú charakterizované vysokým stupňom komplexity. Tieto siete disponujú významným počtom váh a parametrov, čo sťažuje ich interpretáciu a porozumenie ich rozhodovacích mechanizmov. V literatúre sa preto často stretávame s termínom "čierna skrinka", ktorý poukazuje na nedostatok transparentnosti a porozumenia fungovania týchto modelov. [1]

V kontexte kritických odvetví, ako je zdravotníctvo, právo a súdnictvo, a doprava, kde rozhodnutia AI môžu mať priame dôsledky na ľudské životy, je nevyhnutné zdôrazniť potrebu prehĺbovať pochopenie fungovania týchto systémov. V tomto zmysle sa významne rozvíja koncept Explainable AI (XAI), ktorý sa zameriava na vývoj techník a metód umožňujúcich interpretovateľnosť a zrozumiteľnosť rozhodnutí vykonaných AI systémami. Cieľom XAI je poskytnúť ľuďom nástroje na pochopenie a analyzovanie dôvodov, ktoré stoja za rozhodnutiami alebo predikciami vydanými umelou inteligenciou, čím sa zvyšuje dôvera v tieto systémy a znižuje riziko nežiaducich dôsledkov ich použitia.

Náš projekt sa zameriava na pochopenie toho, ako zmeny jasú ovplyvňujú schopnosť modelu správne klasifikovať obrázky, pričom využíva dataset stromových orechov ako základ pre experimenty.

Prvá časť v tejto práci je kapitola venujúca sa súvisiacim prácam, kde sme preskúmali podobné experimenty a zistili veľa nových vecí. Tieto súvisiace práce poskytujú kontext pre náš vlastný experiment a umožňujú nám porovnať naše výsledky s existujúcimi metódami a prístupmi k aktívnemu rozpoznávaniu objektov a prispôbovaniu jasú. Ich analýza posilňuje dôveryhodnosť našich vlastných výsledkov a poskytuje priestor na diskusiu o nových objavoch a možných smeroch budúceho výskumu.

Okrem toho, dokumentácia poskytuje pohľad na dáta, ktoré sa používajú v projekte, vrátane dôvodov, prečo bol vybraný konkrétny dataset stromových orechov a ich popisu. Nasleduje detailný popis postupu experimentu, vrátane prípravy dát a tréningu modelu. Experimentálne výsledky sa potom analyzujú z hľadiska zlepšenia presnosti modelu v závislosti na zmenách jasú obrázkov v jednotlivých triedach.

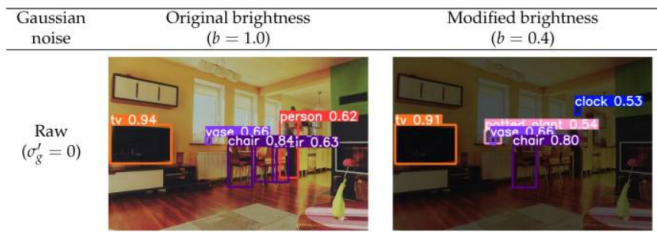
V záverečnej časti textu sa diskutujú zistenia a ich význam pre oblasť strojového učenia a rozpoznávania obrazov. Identifikujú sa dôležité aspekty, ako citlivosť modelu na zmeny jasú a variabilný vplyv manipulácie jasú na rôzne triedy obrázkov. Diskutuje sa aj o potrebe ďalšieho výskumu v tejto oblasti a možných aplikáciách zistení pre rôzne odvetvia, ako je medicína či automatizované vizuálne systémy.

Celkový text poskytuje pohľad na experiment, ktorý rozširuje naše pochopenie vplyvu manipulácie jasú obrázkov na presnosť klasifikačných modelov, a naznačuje dôležité implikácie pre budúci výskum a aplikácie strojového učenia.

II. SÚVISIACE PRÁCE

Pred začatím nášho experimentu sme vykonali systémový prehľad existujúcej literatúry s cieľom hlbšieho pochopenia problematiky a metodológie. Jeden z najpodrobnejšie popísaných článkov, ktorý sme našli, bol článok autorov Rodriguez et al. [2], zameraný na analýzu vplyvu intenzity šumu a jasú na detekciu objektov. Aj keď sme sa v našej práci zamerali výlučne na vplyv jasú, bolo zaujímavé sa dozvedieť viac aj o vplyve šumu. V spomenutom článku boli využívané metódy detekcie objektov, ako napríklad YOLO a Faster R-CNN. Zatiaľ čo YOLO sa vyznačuje rýchlejšou implementáciou a jednoduchosťou, Faster R-CNN často dosahuje presnejšie

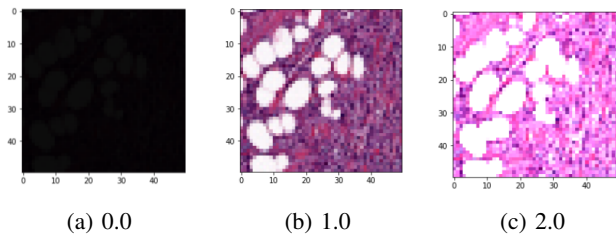
výsledky, obzvlášť v situáciách s výskytom viacerých objektov a ich prekrytím.



Obr. 1: Vplyv klesajúcej intenzity jasu na presnosť detekcie objektov v obraze

Na priloženom Obr. 1 je zrejmé, ako s klesajúcou intenzitou jasu dochádza k postupnému poklesu presnosti detekcie. Autori článku dôrazne poukazujú na skutočnosť, že konzistentné jasové prostredie bez výrazných zmien intenzity osvetlenia umožňuje lepšie vizuálne zaznamenanie ľudí.

V ďalšom článku Kandel et al. [3] sa zaoberali jasom ako technikou augmentácie. Riešili, či nie je táto metóda zbytočná ako spôsob predspracovania údajov. V článku spomínali aj geometrickú augmentáciu, ale nás zaujímal konkrétne jas, takže upustíme od podrobností týkajúcich sa tohto aspektu. Využívali histopatologický dataset, pričom menili jas v rozsahu od 0.0, čo je takmer čierny obraz, až po 2.0 - svetlý, ale ešte nie biely (Obr. 2). Pričom hodnota 1.0 reprezentuje originálny jas. Na dosiahnutie presnejších výsledkov používali štyri konvolučné neurónové siete - ResNet50, DenseNet, InceptionV3 a Xception. Následne boli podobné výsledky pre každú neurónovú sieť - presnosť je pomerne nízka pri hodnotách jasu v rozpätí od 0.0 do 0.25, následne stúpa až k hodnotám 0.75-1.00, avšak potom klesá smerom k hodnotám 1.75-2.00.



Obr. 2: Príklad zmeny jasu obrázkov v článku

III. POPIS DÁT

A. Prečo Tree Nuts dataset

Rozhodli sme sa pre tento dataset z dôvodu, že sa snažíme demonštrovať, ako sa strojové učenie môže naučiť rozlišovať objekty na základe ich tvarov, nie na základe farieb. Rozhodli sme sa pre stromové orechy - konkrétne mandle, vlašské orechy a lieskovce - pretože tieto druhy majú výrazne odlišné tvary, aj keď ich farba môže byť podobná. Tento výber nám umožní preskúmať, ako model reaguje na variácie tvaru a

či dokáže efektívne ignorovať farbu ako faktor rozhodovania. Tým pádom tento dataset ponúka ideálnu príležitosť preskúmať vplyv jasu na rozhodovanie modelu, čo je hlavný cieľ nášho experimentu.

B. Popis datasetu

Tento dataset obsahuje 1163 obrázkov určených na tréning, 50 na testovanie a ďalších 50 na overovanie. Obrázky majú rozlíšenie 224 x 224 pixelov a sú uložené vo formáte JPEG s tromi farebnými kanálmi.

Dataset sa zameriava na klasifikáciu 10 rôznych druhov orechov stromov, ale pre účely tohto projektu sme sa rozhodli využiť iba 3 z nich. Tieto tri druhy orechov zahŕňajú mandle, vlašské orechy a lieskovce. Tento výber bol zvolený z dôvodu rozmanitej formy a podobnej farby týchto orechov.

IV. POSTUP EXPERIMENTU

A. Príprava dát a tréning modelu

Po pripojení knižníc a načítaní datasetu sa začína proces prípravy dát. Využívame knižnicu torchvision.transforms na definovanie súboru transformácií pre tréning modelu. Jedna z týchto transformácií, transform_train, slúži na úpravu vstupných obrázkov. Príkladom takejto úpravy je zmena veľkosti obrázkov na rozmer 224x224 pixelov, aby mali jednotnú veľkosť. Okrem toho sú obrázky náhodne otočené o uhol do 10 stupňov a orezané. Následne sú tieto obrázky transformované na tenzor a normalizované pomocou priemerových hodnôt a štandardných odchýlok pre každý farebný kanál RGB.

Druhá časť transform definuje súbor transformácií pre validáciu a testovanie modelu. Sú takmer rovnaké, s výnimkou absencie augmentácií (náhodných transformácií) a orezovania. To znamená, že na obrázkoch pre validáciu a testovanie nie sú aplikované náhodné zmeny, ako je zrkadlenie alebo otočenie, a ani nie sú orezané.

Používame knižnicu torchvision.datasets.ImageFolder na načítanie dát tréningového, validačného a testovacieho datasetu. Potom sa dáta načítajú do DataLoader, ktorý rozdeľuje dáta do dávok (batchov) veľkosti 32 obrázkov každá. Parameter shuffle=True pre tréningový DataLoader znamená, že dáta budú zamiešané pred každou epochou tréningu, aby sa model učil na rôznych príkladoch v každej epoche. Pre testovací a validačný DataLoader parameter shuffle=False, pretože nie je potrebné zamiešať tieto dáta, ale stále ich chceme rozdeliť do dávok pre výpočtovú efektivitu.

Model konvulčnej neurónovej siete (CNN) prispôbený na klasifikáciu obrázkov. Architektúra CNN pozostáva z konvulčných vrstiev (conv1 a conv2) na extrakciu príznakov, po ktorých nasledujú vrstvy dávkovej normalizácie (bn1 a bn2) na normalizáciu, maximálneho združovania (pool) na redukciu priestorovej dimenzie a vynechávania (dropout) na zabránenie nadmernému prispôbieniu. Plne prepojené vrstvy (fc1 a fc2) sa starajú o klasifikáciu. Počas priameho prechodu vstupné údaje prechádzajú konvulciou, normalizáciou, aktiváciou, poolovaním a sploštením pred prechodom cez plne prepojené vrstvy. Tréning zahŕňa nastavenie načítavania údajov, definovanie stratových funkcií a optimalizátorov

a použitie plánovania miery učenia. Model sa trénuje s využitím vlastných spätných volaní na úpravu miery učenia a Torchbearerovho procesu na zefektívnenie vykonávania. Po skončení trénovania sa vykoná vyhodnotenie modelu a história trénovania vrátane meraní strát a presnosti sa vizualizuje v priebehu epoch, aby bolo možné získať prehľad.

Na funkciu strát sa používa krížová entropia CrossEntropyLoss, ktorá je všeobecne prijímanou stratovou funkciou pre klasifikačné úlohy. Optimalizátor Adam sa používa na aktualizáciu parametrov modelu s prihliadnutím na gradienty zo straty. Dodatočne sa používa parameter weight_decay, ktorý ovplyvňuje rýchlosť učenia a regularizáciu modelu. Zníženie rýchlosti učenia sa vykonáva pomocou objektu scheduler s krokom 10 a koeficientom zníženia 0.1 pre každý krok. To pomáha stabilizovať proces učenia a zvýšiť jeho efektívnosť.

Potom definujeme spätný volanie SchedulerStepCallback, ktoré volá metódu scheduler.step() po dokončení každej epochy trénovania na zníženie rýchlosti učenia. Potom spúšťame proces trénovania.

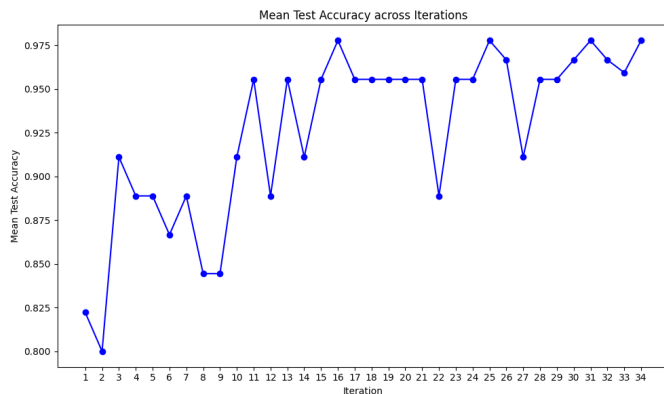
B. Vyhodnotenie výkonu modelu

Na dôkladné preskúmanie vplyvu zmeny jasu na úspešnosť klasifikácie obrázkov sme zvolili postupnú zmenu jasu datasetu s krokom 3%. Týmto spôsobom sme vykonali celkovo 34 iterácií, systematicky testujúc rôzne úrovne jasu a zhromažďujúc dáta na vyhodnotenie vplyvu na klasifikačnú presnosť.

Trieda "Almonds". Na základe týchto obrázkov sme analyzovali vplyv zvýšenia jasu na správnosť klasifikácie daného datasetu.

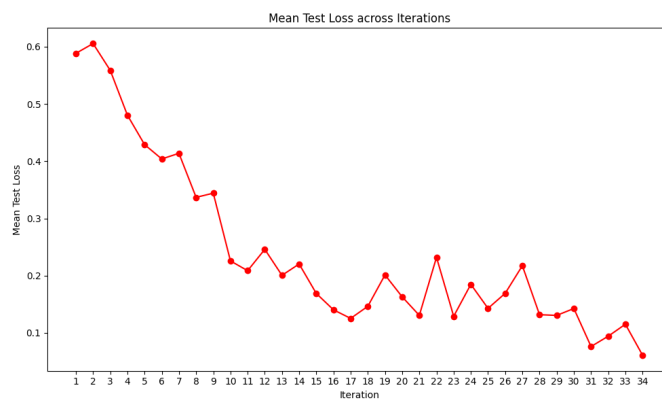
Trieda "Hazelnuts". Tieto dáta nám slúžili na reprezentáciu vplyvu znižovania jasu na presnosť klasifikácie vybraného datasetu.

Trieda "Walnuts". Posledná zvolená sada údajov sa zaoberala výsledkami zachovania pôvodných hodnôt jasu na efektívnosť klasifikácie.



Obr. 3: Priemerná presnosť počas 34 iterácií.

Nakoniec sme súbežne s úpravou jasu obrázkov sledovali zlepšenie presnosti modelu, ako je zrejmé z výsledkov stále sa zvyšujúcich hodnôt presnosti testovania počas iterácií (Obr.



Obr. 4: Priemerná strata počas 34 iterácií

3 a 4). Počas prvej iterácie, kde bol jas nezmenený, bola hodnota priemernej straty 0,58841 s presnosťou 0,82223. Tieto hodnoty sa výrazne líšia od výsledkov v 34. kroku, kde bol jas upravený na 99% - priemerná strata 0,06065 s presnosťou 0,97778.

V. ZHODNOTENIE A ZÁVER

Na základe výsledkov nášho experimentu sme identifikovali niekoľko významných poznatkov. Potvrdili sme hypotézu, že pri určitých zmenách jasu môže model neurónovej siete začať zachytávať jas ako dôležitý aspekt a rozhodovať na základe neho. Toto zistenie má zásadný význam nielen pre náš projekt, ale aj pre širšie pole výskumu v oblasti vysvetliteľnej umelej inteligencie (XAI). Pochopenie fungovania jednotlivých modelov je kritické nielen pre vývojárov pracujúcich v oblasti neurónových sietí, ale aj pre zabezpečenie dôveryhodnosti a spoľahlivosti týchto modelov vo viacerých odvetviach.

Naše zistenia majú potenciálne aplikácie v oblasti detekcie objektov, najmä v prostrediach s variabilným osvetlením. Porozumenie toho, ako sa model správa v závislosti od zmien jasu, môže poskytnúť cenné informácie pre návrh neurónových sietí a iných klasifikačných modelov, ktoré majú byť nasadené v týchto dynamických prostrediach.

Celkovo sme identifikovali dôležitý aspekt fungovania klasifikačných modelov obrázkov a otvorili sme tak dvere pre ďalší výskum v tejto oblasti.

VI. DISKUSIA

Naša štúdia zameraná na vplyv manipulácie jasu obrázkov na presnosť klasifikačného modelu poskytla zaujímavé výsledky. Zistili sme, že model je citlivý na zmeny jasu a že tieto zmeny môžu viesť k zlepšeniu presnosti klasifikácie. Tieto zistenia otvárajú niekoľko smerov pre ďalšie výskumy a vylepšenia projektu.

Jedným možným smerom je ďalšie skúmanie vplyvu rôznych úrovní jasu na presnosť modelu. Môžeme experimentovať s väčším rozsahom úprav jasu a pozorovať, ako sa mení presnosť modelu v závislosti na rozdielnych úrovniach jasu. To by mohlo poskytnúť hlbšie pochopenie toho, ako jas ovplyvňuje rozhodovacie procesy klasifikačného modelu.

Ďalším možným vylepšením projektu je skúmanie iných foriem úprav obrázkov a ich vplyv na výsledky modelu. Okrem manipulácie jasou môžeme skúmať aj iné faktory, ako sú farebné filtre, kontrast alebo rozostrenie, a ich vplyv na presnosť klasifikácie. Tieto alternatívne experimenty by mohli poskytnúť ďalšie informácie o tom, ako rôzne úpravy obrázkov ovplyvňujú rozhodovacie procesy modelu.

Ďalším krokom môže byť rozšírenie nášho datasetu a skúmanie vplyvu manipulácie jasou na presnosť modelu pri použití rôznych typov obrázkov. V našej štúdii sme sa sústredili na klasifikáciu obrázkov rôznych typov orechov, ale podobné experimenty by sa mohli vykonať aj na iných typoch obrázkov, aby sme zistili, či naše zistenia sú všeobecne platné pre rôzne kategórie obrázkov.

Vzhľadom na obmedzenia a obmedzenia nášho súčasného výskumu je dôležité tiež zvážiť ďalšie faktory, ktoré by mohli ovplyvniť výsledky projektu. Medzi tieto faktory patrí napríklad použitá architektúra modelu, veľkosť datasetu a nastavenia trénovania. Skúmanie týchto faktorov by mohlo poskytnúť lepšie pochopenie toho, ako a prečo sa model správa v určitých situáciách.

V závere je naším cieľom neustále zlepšovať a rozširovať naše pochopenie o fungovaní klasifikačných modelov obrázkov. Naša diskusia o budúcich plánoch projektu a možnostiach ich zlepšenia nám poskytuje rámec pre ďalší výskum a experimenty, ktoré by mohli prispieť k lepšiemu pochopeniu tohto zložitého problému.

POUŽITÁ LITERATÚRA

- [1] P. P. Angelov, E. A. Soares, R. Jiang, N. I. Arnold, and P. M. Atkinson, "Explainable artificial intelligence: an analytical review," *WIREs Data Mining and Knowledge Discovery*, vol. 11, no. 5, p. e1424, 2021. [Online]. Available: <https://wires.onlinelibrary.wiley.com/doi/abs/10.1002/widm.1424>
- [2] J. A. Rodríguez-Rodríguez, E. López-Rubio, J. A. Ángel Ruiz, and M. A. Molina-Cabello, "The impact of noise and brightness on object detection methods," *Sensors (Basel)*, vol. 24, no. 3, p. 821, Jan 2024.
- [3] I. Kandel, M. Castelli, L. Manzoni *et al.*, "Brightness as an augmentation technique for image classification," *Emerging Science Journal*, vol. 6, no. 4, pp. 881–892, 2022.