УНИВЕРЗИТЕТ У БАЊОЈ ЛУЦИ
ЕЛЕКТРОТЕХНИЧКИ ФАКУЛТЕТ

ДЕТЕКЦИЈА УЗОРАКА КОЈИ ОДСТУПАЈУ ОД РАСПОДЈЕЛЕ У КЛАСИФИКАЦИЈИ СЛИКА ДОБИЈЕНИХ ДАЉИНСКОМ ДЕТЕКЦИЈОМ

МАСТЕР РАД

Ментор:
Проф. др Владимир Рисојевић

Кандидат:
Дајана Димитрић

Бања Лука, 2024.

UNIVERSITY OF BANJA LUKA
FACULTY OF ELECTRICAL ENGINEERING

OUT-OF-DISTRIBUTION DETECTION IN REMOTE SENSING SCENE CLASSIFICATION

MASTER THESIS

Mentor:
Prof. dr Vladimir Risojević

Candidate:
Dajana Dimitrić

Banja Luka, 2024.

Информације о ментору и мастер раду

Ментор: проф. др Владимир Рисојевић, Електротехнички факултет, Универзитет у Бањој Луци

Наслов мастер тезе: Детекција узорака који одступају од расподјеле у класификацији слика добијених даљинском детекцијом

Резиме: Класификација слика добијених даљинском детекцијом представља поступак сврставања слика у једну од унапријед дефинисаних категорија, најчешће на основу њеног семантичког значења. Карактеристичан примјер је класификација слика према типу покривача и начину кориштења земљишта (енг. *land cover/land use*). Најбољи резултати се добијају помоћу метода заснованих на дубоком учењу, при чему се посебно треба истаћи значај неуронских мрежа. У пракси је устаљено да се неуронске мреже обучавају на предефинисаном броју класа, очекујући да ће сви тренинг и тестни подаци припадати истој расподјели. Међутим, неизбјежно је да модел у тестној фази буде изложен узорцима који одступају од расподјеле тренинг података. С тим у вези, намеће се питање: Како препознати узорак који не припада ни једној од класа виђених током фазе обучавања? У овом раду је теоријски описано и извршено поређење више метода за детекцију узорака који одступају од расподјеле, те је испитан утицај величине тренинг скупа на перформансе метода. Осим тога, извршено је поређење резултата при употреби различитих архитектура основног класификатора: конволуционих неуронских мрежа и трансформатора за рачунарски вид. На крају је дат приједлог метода за детекцију који при означавању тестних узорака, осим тренинг података, има могућност да користи и ограничен број претходно прикупљених података који одступају од расподјеле тренинг скупа.

Кључне ријечи: класификација слика, даљинска детекција, дубоко учење, неуронске мреже, узорци који одступају од расподјеле

Научна област: Инжењерство и технологија

Научно поље: Електротехника, електроника и информационе технологије

Класификациона ознака: Т 121

Information about mentor and master thesis

Supervisor: Dr Vladimir Risojević, professor, Faculty of Electrical Engineering, University of Banja Luka

Title of master thesis: Out-of-distribution detection in remote sensing scene classification

Abstract: Remote sensing image classification is the process of categorizing images into one of the predefined categories, commonly based on its semantics meaning. A characteristic example is land cover/land use classification. Methods based on machine learning give the best results, where the importance of the neural networks should be emphasized. Neural networks are tipically trained on predefined number of classes, expecting that all trening and test data will belong to the same distribution. But, it is more tnan likely that the model in the test phase will be exposed to samples coming from a different distribution compared to the training data (out-of-distribution samples). About that, there is a question: How to recognize the sample that doesn't belong to any of the classes seen during the training phase? In this paper, several methods for out-of-distribution detection are theoretically described and compared, and the training dataset size influence on the performance of methodes is examined. In addition, results comparison was made when using different architectures of classifier: convolutional neural network and vision transformers. At the end, it is given proposal for detection method which in the process of labeling test samples, except training data, is able to use limited number of offline collected out-of-distribution samples.

Keywords: image classification, remote sensing, deep learning, neural networks, out-of-distribution samples

Scientific area: Engineering and technology

Scientific field: Electrical engineering, electronics and information engineering

Classification code: T 121

Кћерки Софији
и супругу Душку.

Садржај

Листа слика

Листа табела

Листа скраћеница

1. Увод

Појам даљинске детекције (енг. *remote sensing* - RS) [1] се, у општем случају, односи на аквизицију података о одређеном објекту или појави, без остварења физичког контакта. Међутим, у пракси је поменути поступак еквивалентан прикупљању информација о Земљиној површини помоћу сензора постављених изнад ње. Специјалним камерама се добијају снимци Земљине површине и чувају у формату слике, а у зависности од тога да ли је камера постављена у ваздуху или у свемиру, овакви снимци се могу подијелити на *аеро* и *сателитске* снимке. Додатно, даљинска детекција се може подијелити на *активну* и *пасивну*, према типу сензора који се користи. Активни сензори шаљу сопствени електромагнетни сигнал и врше његово мјерење након што се рефлектује о Земљину површину, док пасивни сензори мјере рефлектовану сунчеву свјетлост.

Снимци добијени даљинском детекцијом налазе примјену у разним областима, попут: урбаног планирања, предвиђања и детекције природних катастрофа, прогнозе времена, детекције потенцијалних клизишта, креирања мапа, анализе искористивости земљишта итд. Захваљујући напретку технологије али и све већим захтјевима области у којима се они примјењују, број снимака добијених даљинском детекцијом константно и драстично расте, а паралелно поменутом расте и потреба за све интелигентнијим тумачењем истих [2]. Из тог разлога је неопходно да се константно унапрјеђују већ постојећи и откривају нови методи у областима класификације, сегментације и детекције снимака добијених даљинском детекцијом, а поменуто уједно представља и основни мотив за настанак овог рада.

1.1. Дефиниција проблема

Да би снимци добијени даљинском детекцијом били искористиви, углавном је потребно да се што правилније протумачи њихов садржај и да се што тачније класификују. Са порастом просторне резолуције проблем класификације се преносио са нивоа пиксела, преко нивоа објекта, до нивоа сцене [2]. У посљедње вријеме, у области класификације слика убједљиво најбоље резултате дају методи засновани на дубоком учењу (енг. *deep learning*). Посебно треба истаћи значај појаве конволуционих неуронских мрежа (енг. *convolutional neural networks* - CNN) [3], које налазе своју примјену у многим доменима у којима се ради анализа слика. Неки често кориштени модели, а чије су архитектуре конволуционог типа, су AlexNet [4], VGG [5], ResNet [6], Inception [7], MobileNet [8], EfficientNet [9] итд. Недавно је показано да модел са другачијим типом архитектуре – трансформатор за рачунарски вид (енг. *Vision Transformer* - ViT, у наставку текста трансформатор) [10] може дати још боље резултате. Иако поменута нова архитектура захтијева већи број података и више времена за обучавање, техника трансфера параметара (енг. *transfer learning*) омогућава да се умјесто уобичајених конволуционих модела све чешће користе трансформатори. Трансфер параметара подразумијева употребу већ тренираних модела на неком великом скупу података. Тај претренирани модел може бити, без икаквих модификација, кориштен као екстрактор обиљежја слика. Додатно, уколико је потребно, може да се изврши фино подешавање (енг. *fine tuning*) његових параметара на неком мањем скупу података из домена у којем ће бити рјешаван одређени проблем класификације, детекције или сегментације слика.

Подразумијевано и устаљено у пракси је да се неуронске мреже тренирају на предефинисаном броју класа, очекујући да ће сви тренинг и тестни подаци припадати истој

расподјели (енг. *in-distribution* - ID). Међутим, у пракси то није случај и неизбјежно је да модел буде изложен тестним узорцима који одступају од расподјеле тренинг података (енг. *out-of-distribution* - OOD). Дакле, иако мрежа остварује велику тачност класификације слика у домену ID података, намеће се питање: Како препознати узорак који не припада ни једној од класа виђених током фазе обучавања?

Упркос заједничком мотиву препознавања OOD узорака у тестној фази, литература из области класификације и детекције слика разликује неколико проблема [11]:

- детекција аномалија (енг. *anomaly detection* - AD) - циљ је детекција било којих података који одступају од унапријед дефинисаног „нормалног“, при чему се нормалним сматрају искључиво подаци из класа виђених током фазе тренирања класификатора.
- детекција новина (енг. *novelty detection* - ND) - за разлику од претходног случаја, подаци који одступају од „нормалног“ се не сматрају лошим, већ представљају нову класу и посматрају се као ресурси за будуће побољшање перформанси класификатора.
- детекција изузетака (енг. *outlier detection* - OD) - циљ је откривање узорака који се значајно разликују од осталих у посматраном тестном скупу.
- детекција података ван расподјеле (енг. *out-of-distribution detection* - OODD) - односи се на детекцију узорака у тестној фази који не припадају расподјели података виђених у фази обучавања. Као и у случају детекције аномалија, новина и изузетака, обично се посматра као проблем бинарне класификације, што значи да сваки улазни податак може бити означен само као ID или OOD.
- препознавање у отвореном скупу података (енг. *open set recognition* - OSR) - за разлику од претходно наведених проблема, ова дисциплина представља проблем класификације у више класа. Дакле, класификатор треба да што тачније врши класификацију података из познате расподјеле, али и да детектује податке из непознате расподјеле.

У оквиру овог рада фокус је на проблему детекције података ван расподјеле. Конкретно, везано за домен снимака добијених даљинском детекцијом, OOD узорци могу да се јаве усљед нових, мрежи непознатих, класа у тестном скупу или усљед разлика у географским подручјима и услова под којим се снимање врши.

Проблем препознавања узорака из непознате расподјеле је у литератури први пут споменут 2017. године и од тада се, паралелно у свим претходно наведеним дисциплинама, за детекцију почињу примјењивати различити методи засновани на класификацији (енг. *classification based*), естимацији расподјеле тренинг података (енг. *density based*), мјерењу удаљености у простору обиљежја (енг. *distance based*), те на реконструкцији слике (енг. *reconstruction based*).

Сви претходни методи обично подразумијевају да се врши фино подешавање параметара основног класификатора на неком скупу слика из домена у којем се рјешава проблем детекције. Потом се класификатор са подешеним параметрима користи у сврху издвајања обиљежја корисних за OOD детекцију. Међутим, недавно је истраживање показало да се завидни резултати могу остварити и без примјене финог подешавања параметара мреже, али под условом да је модел довољно добар у погледу издвајања обиљежја слика [12]. Конкретно у оквиру поменутог истраживања, кориштене су двије трансформаторске архитектуре.

У литератури се најчешће могу срести методи који користе класификатор обучен на познатој расподјели, те у том случају детектор узорака који одступају од расподјеле нема никакву информацију о изгледу података ван те расподјеле, а који се потенцијално могу јавити у тестној фази. Међутим, намеће се идеја да би познавање неких узорака ван расподјеле олакшало задатак детекције, те се у циљу побољшања перформанси предлаже

нови приступ проблему - коришћење скупа података који моделује OOD податке (енг. *outlier exposure* - OE) [13-19]. Нови подаци могу бити искоришћени за модификацију параметара само неколико посљедњих слојева мреже или за фино подешавање параметара цијелог модела. Такође, прикупљање узорака ван расподјеле може да се ради и у току тестне фазе, при чему би се вршила модификација система класификатора и детектора кад год се на улазу појави одређен број узорака који припада новој класи (енг. *class incremental learning*) [18]. У том случају, сви OOD узорци који се детектују у току тестне фазе могу бити сматрани једном новом класом [20], али нека истраживања показују да се бољи резултати добијају уколико се и прикупљени OOD узорци групишу у више класа [15]. У случају да се OOD узорак бира прије тестне фазе, прилично је тежак задатак унапријед претпоставити природу узорака ван расподјеле који ће се тек у будућности јавити на улазу система. У циљу да се „покрије“ што већи дио простора обиљежја у којем би се потенцијално OOD узорци могли наћи, понекад се генеришу виртуелни подаци, комбиновањем слика из тренинг скупа и OOD слика узетих из неког постојећег скупа [18]. Понекад се нови, виртуелни узорци, добијају само интерполацијом вектора обиљежја тренинг узорка и узорака ван расподјеле [14], [21].

1.2. Организација рада

Теоријска основа класификације слика је дата у другом поглављу. Описани су најзначајнији методи, а нагласак је на оним најновијим и уједно најуспјешнијим: класификацији слика базираној на конволуционим неуронским мрежама и трансформаторима за рачунарски вид.

У оквиру трећег поглавља су описане технике за детекцију узорака који одступају од расподјеле. Од највећег интереса за овај рад су методи базирани на мјерењу удаљености, па је неколико њих обрађено детаљније. Додатно, описан је нови приступ проблему који укључује излагање детектора узорцима који одступају од расподјеле тренинг скупа. У складу са тим предложена је модификација једног метода базираног на мјерењу удаљености, на тај начин да се омогући коришћење узорака ван расподјеле који се имају на располагању.

У четвртом поглављу су описани коришћени скупови слика и архитектуре модела које су служиле као основни класификатори у оквиру овог рада.

Експериментални резултати су дати у петом поглављу. У оквиру ове цјелине је извршено поређење успјеха метода на различитим скуповима и анализиран је утицај архитектуре класификатора и величине тренинг скупа на перформансе метода. На крају је тестиран нови предложени метод за детекцију узорака који одступају од расподјеле тренинг скупа, у случају када се на располагању има коначан скуп OOD узорака. Дати су резултати у случају различитих скупова из којих се бирају OOD узорци и различитих тестних скупова, те је и у овом случају посматран утицај величине тренинг скупа.

Шесто поглавље представља закључак у којем су рекапитулирани резултати, а иза њега слиједи попис литературе коришћене у току израде овог рада.

1.3. Допринос рада

У раду су тестирани постојећи методи за детекцију узорака који одступају од расподјеле у домену снимака добијених даљинском детекцијом, при чему је акценат стављен на оне методе који су базирани на мјерењу удаљености у простору обиљежја. Испитано је на који начин архитектура основног класификатора који се користи за издвајање обиљежја слика утиче на перформансе посматраних метода и на тај начин је још једном упоређена релативно нова архитектура - трансформатор за рачунарски вид са већ устаљеном у пракси - конволуционом мрежом.

На основу резултата добијених примјеном различитих метода базираних на мјерењу удаљености предложен је метод погодан за примјену на слике добијене даљинском детекцијом. Како су методи који почивају на мјерењу удаљености у простору обиљежја осјетљиви на број тренинг узорака, намеће се питање: Да ли поредак резултата који дају посматрани методи остаје исти и ако се величина тренинг скупа смањи? Мотив долази из праксе, гдје постоји низ ситуација у којима због разних ограничења није могуће обезбједити велики број тренинг узорака. Експерименти показују да су неки методи више робусни у односу на друге, те да у незанемаривој мјери од величине тренинг скупа зависи који метод је најпогоднији за примјену.

Тестирањем метода на различитим скуповима је изведен закључак о утицају домена из којег потичу OOD слике на перформансе детектора. Испитана је разлика у перформансама метода уколико тестни OOD узорци долазе из домена слика добијених даљинском детекцијом или уколико долазе из неких других домена.

На крају рада, детектор је изложен коначном скупу узорака који одступају од расподјеле тренинг скупа и предложена је модификација једног метода базираног на мјерењу удаљености, те је на тај начин детектору омогућено да у обзир узме и поменуте OOD узорке. Скуп се бира прије фазе тестирања и дат је приједлог колекције слика из које се бирају OOD узорци, колико њих, те по којем принципу и којим редом треба да се бирају. Предложени метод не захтијева модификацију основног класификатора, што представља велику предност у погледу уштеде времена, нарочито кад се узме у обзир чињеница да су модели све сложенији и да би поступак финог подешавања параметара могао трајати све дуже. Метод је тестиран под разним околностима и вршено је поређење добијених резултата са разултатима оствареним у ситуацији у којој детектор није изложен ниједном OOD податку. Осим тога, усљед специфичног избора и подјеле базне колекције слика на ID и OOD дио, омогућено је поређење са литературом [18]. На овај начин се процјењује успјех новопредложеног метода и закључује у каквом се то он положају налази у односу на раније објављиване методе.

2. Класификација слика

Како би слика могла бити извор корисних информација, неопходно је да се правилно интерпретира и тако, између осталог, настаје потреба и за класификацијом. Класификација представља поступак означавања слике једном од унапријед дефинисаних категорија, углавном на основу њеног семантичког значења. Дакле, сматра се да су категорије дисјунктне, те да није могуће означити слику ознакама двије или више категорија. Скуп правила, тј. алгоритам на основу којег се означавају слике назива се *класификатор*. Уколико је скуп предефинисаних класа $C = \{c_1, c_2, \dots c_k\}$, модел класификатора је функција $f: X \to C$, која дати узорак $x \in X$ сврстава у класу $y = f(x) \in C$ [22].

Уколико се слике дијеле у двије категорије, класификација се назива бинарном, док се у случају подјеле слика у више од двије категорије ради о класификацији у више класа. Конкретно у домену слика добијених даљинском детекцијом, примјер бинарне класификације је подјела снимака Земљине површине на урбана и рурална подручја, док би се подјела водених површина на поток, ријеку, језеро, бару, море и океан сматрала класификацијом у више класа.

У оквиру проблема класиифкације слика у више класа, паралелно порасту просторне резолуције слика развиле су се три гране:

1. класификација на нивоу пиксела,
2. класификација на нивоу објекта,
3. класификација на нивоу сцене.

На почетку је просторна резолуција била изразито ниска, при чему је стварна величина коју је представљао један пиксел била реда величине објекта од интереса. Стога, поступак означавања сваког пиксела семантичком класом је у том тренутку давао довољно добре резултате. Класификација на нивоу пиксела је и даље активна дисциплина и налази примјену у мултиспектралној и хиперспектралној анализи слика добијених даљинском детекцијом. Са побољшањем просторне резолуције један пиксел је изгубио семантичко значење и још 2001. године су аутори рада [12] констатовали да класификација на нивоу објекта даје боље резултате. Ова дисциплина се своди на препознавање објеката на слици и упркос каснијем порасту просторне резолуције се развија и даље и налази своју примјену у пракси. Зависно од примјене, неријетко је потребно слику сагледати у цјелини и означити је семантичком категоријом у складу са њеним глобалним садржајем. У том случају се каже да се класификација ради на нивоу сцене.

Као и у случају мултимедијалног садржаја уопште, тако се и класификацији слика може приступити на три начина [22]:

1. ручна класификација,
2. класификација кориштењем правила,
3. класфикација кориштењем машинског учења.

У првом и другом приступу експерти из одређене области врше класификацију слика или формирају правила на основу којих се та класификација врши, респективно. Што тачнија класификација подразумијева потребу да се што вјеродостојније свакој слици припишу обиљежја која је описују. На почетку су за то били заслужни експерти и неки предложени методи за описивање слика су [12]: трансформација обиљежја непромјењиве размјере (енг.

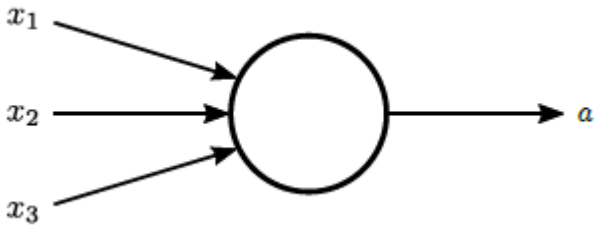
Scale-Invariant Feature Transformation - SIFT) [23], дескриптори текстуре (енг. *Texture Descriptors* - TD) [24-26], хистограми боје (eng. *Color Histogram* - CH) [27], хистограм оријентисаних градијената (*Histogram of Oriented Gradients* - HOG) [28], холистички приступ просторног омотача [29], вектор локално агрегираних дескриптора (енг. *Vector of Locally Agregated Descriptors* - VLADs) [30], скуп визуелних ријечи (енг. *Bag-of-Visual-Words* - BoVW) [31], итд. Након што експерти формирају обиљежја, наведени методи углавном користе неки обучени класификатор за доношење одлуке. Временом, резултати које даје комбинација „ручно израђених“ дескриптора и обучених класификатора су ушли у засићење и показало се да су информације које се добијају „ручно израђеним“ дескрипторима на релативно ниском нивоу [2]. Осим тога, кључан је људски фактор и неизбјежни су проблеми који се односе на трајање класификације, цијену, скалабилност итд. Такође, број слика и потреба за димензионалности вектора обиљежја расту, те постаје све теже и скоро немогуће „ручно“ формирати обиљежја, која би у комбинацији са обученим класификатором дала завидне резултате. Очито, јавила се потреба за формирањем интелигентног алгоритма за рјешење проблема. Све наведено представља мотивацију за употребу техника машинског учења у класификацији слика, гдје више не би постојала потреба за ручним издвајањем обиљежја. Сходно машинском учењу, прибјегава се употреби статистичких, итеративних метода, помоћу којих се параметри класификатора уче од великог броја означених слика. Као доминантна грана машинског учења истиче се дубоко учење (енг. *deep learning*) [32], које углавном подразумјева да се учење репрезентације слика ради помоћу дубоких *неуронских мрежа*, тј. неуронских мрежа са релативно много слојева и параметара.

2.1. Неуронске мреже

Неуронска мрежа представља специфичну архитектуру модела која омогућава да се функције класификације извршавају по узору на људски мозак. Основна градивна јединица неуронске мреже је *неурон*, елемент са више улаза и једним излазом. На Слици 2.1 је дат шематски приказ неурона са три улаза. Ако се на улаз неурона доведу сигнали $x_1, x_2, \dots, x_n$, на његовом излазу се јавља сигнал $a = g(z)$, гдје је $g(\cdot)$ *активациона функција* неурона, а z афина фунција улазних сигнала:

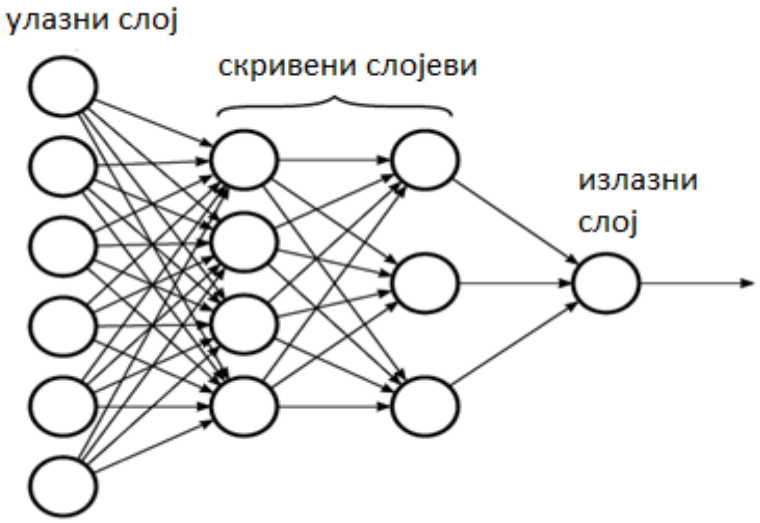
$$z = \omega_1 x_1 + \omega_2 x_2 + \cdots + \omega_N x_N + b. \tag{2.1}$$

Величине $\omega_1, \omega_2, \dots, \omega_N$ представљају тежине неурона, b офсет, док N представља број улазних сигнала. Као активационе функције могу да се користе одскочна, логистичка, хиперболички тангенс, исправљачка функција, итд.



Слика 2.1: Неурон са три узлаза и једним излазом [22].

Могућности које пружа један неурон су недовољне, те је за постизање довољно моћног модела потребно међусобно повезати велики број неурона. На тај начин се креира сложена архитектура кроз коју се сигнал преноси и на одговарајући начин трансформише у облик који је погодан за доношење одлуке у оквиру проблема који се рјешава. У општем случају, неурони могу бити повезани на произвољан начин. Међутим, ради лакшег моделовања и анализе мреже углавном је ријеч о неуронским мрежама у којима су неурони груписани у *слојеве*, при чему не постоје везе између неурона у истом слоју, везе које прескачу слојеве или се враћају уназад. Дакле, за потребе анализе се сматра да један неурон може бити повезан само са неуроном из сљедећег слоја, упркос томе што се везе које прескачу слојеве користе у модерним *конволуционим мрежама*, а о којима ће бити ријеч у наставку овог рада. Примјер вишеслојне неуронске мреже је дат на Слици 2.2.



Слика 2.2: Вишеслојна неуронска мрежа [22].

Први слој неуронске мреже се назива *улазним*, посљедњи *излазним*, док су сви остали унутрашњи или *скривени* слојеви мреже. Неурони из улазног слоја не обрађују сигнал, имају један улаз и један излаз, те је излазни сигнал једнак улазном. Улазни сигнали у неурон скривеног слоја су излази из неурона из претходног слоја. Излазни сигнал неурона из скривеног слоја се води на неуроне у идућем слоју. Сигнали на излазу неурона из излазног слоја се називају излазним сигналима мреже.

У складу са изразом (2.1), примјена афине функције на улазне сигнале у k-ти неурон l-тог слоја даје:

$$z_k^{(l)} = \sum_{j=0}^{n^{(l-1)}} \omega_{kj}^{(l-1)} a_j^{(l-1)}, \tag{2.2}$$

при чему $\omega_{kj}^{(l-1)}$ представља тежину везе од j-тог неурона $(l-1)$-ог слоја ка k-том неурону l-тог слоја, а $a_j^{(l-1)}$ је излаз из j-тог неурона $(l-1)$-ог слоја. Тежина са индексом 0 у изразу (2.2), $\omega_{k0}^{(l-1)}$, представља офсет. Број $n^{(l-1)}$ је број неурона у $(l-1)$-ом слоју. Ако мрежа има L слојева, а посљедњи слој $n^{(L)}$ неурона и ако $\boldsymbol{w}_k^{(L-1)}$ представља вектор тежина улаза k-тог

неурона посљедњег слоја, а $\boldsymbol{a}^{(L-1)}$ вектор излаза свих неурона претпосљедњег слоја, за вектор излаза неурона посљедњег слоја, прије примјене активационе функције, може се писати:

$$\boldsymbol{z}_k^{(L)} = (\boldsymbol{w}_k^{(L-1)})^T \boldsymbol{a}^{(L-1)},\ k = 1, \ldots\ldots, n^{(L)}. \tag{2.3}$$

Ако је x N-димензионални податак, улазни слој мреже има N неурона. Када је ријеч о класификацији узорака, број неурона у излазном слоју мреже је број класа у које се узорци класификују. У том случају се на излазне сигнале из (2.3) примјењује *софтмакс активациона функција*:

$$a_k^{(L)} = g\left(z_1^{(L)}, \ldots, z_{n^{(L)}}^{(L)}\right) = \frac{e^{z_k^{(L)}}}{\sum_{n=1}^{n^{(L)}} e^{z_i^{(L)}}},\ k = 1, \ldots, n^{(L)}. \tag{2.4}$$

Вриједност израза (2.4) представља естимирану вјероватноћу да узорак припада k-тој класи, а узорак се класификује у ону класу за коју је естимирана вјероватноћа максимална.

Вриједности вјероватноћа на излазу мреже зависе од тежинских коефицијената, чије се вриједности одређују у *фази обучавања* мреже. Оптимизација тежина мреже представља начин да се класификатор прилагоди конкретном задатку. Ако је ријеч о *надгледаном обучавању* [33] и задатку класиифкације, што је и подразумјевано у оквиру овог рада, потребно је прикупити скуп узорака за које је позната ознака класе којој припадају. Такав скуп узорака се дијели на три дисјунктна дијела: *тренинг*, *тестни* и *валидациони* скуп. У фази обучавања мреже је потребно одредити параметре модела тако да он тачно класификује што већи број узорака из тренинг скупа. Иако се обучавање мреже своди на њено прилагођавање тренинг подацима, веома је битно да модел оствари добру *генерализацију*, тј. да буде у стању да добро класификује и узорке са којима се није сусрео у фази обучавања. За добру генерализацију је потребно ускладити сложеност модела са величином тренинг скупа. Употреба модела са великим бројем параматера и обучавање на малом тренинг скупу најчешће узрокује лошу генерализацију, јер се у тој ситуацији модел прилагоди обиљежјима која карактеришу само мали број тренинг узорака. Описана појава представља претјерано прилагођење модела (енг. *overfitting*). Лошу генерализацију може узроковати и премало прилагођен модел (енг. *underfitting*), тј. када се на релативно великом броју тренинг података обучава недовољно сложена мрежа. Таква мрежа помоћу малог броја параметара нема могућност да научи све битне карактеристике тренинг скупа и даје лоше резултате класификације не само на тестном, него и на тренинг скупу.

Осим параметара неуронске мреже, потребно је одредити и број скривених слојева и број неурона у скривеним слојевима, тј. хиперпараметре мреже који дефинишу степен сложености модела. Како од поменутих хиперпараметара зависи архитектура мреже, а самим тим и број параметара који се требају одредити, број скривених слојева и неурона у скривеним слојевима се дефинише прије фазе обучавања. Да би се изабрале оптималне вриједности хиперпараметара, потребно је поновити обучавање за различите вриједности и изабрати оне које дају најбоље резултате. Међутим, да би се избјегло претјерано прилагођење хиперпараметара тренинг или тестним подацима, перформансе класификатора се за различите хиперпараметре испитују на валидационом скупу.

Да би се испитала успјешност процеса обучавања мреже, неопходно је провјерити ефикасност модела на тестном скупу. У ту сврху се најчешће одређује *тачност класификације*, као однос броја тачно класификаваних и укупног броја тестних узорака.

Одређивању параметара мреже у фази обучавања може се приступити на различите начине. Полази се од тренинг скупа који се има на располагању, а који је дат у облику (x_i, y_i),

$i = 1, 2, \ldots, n$, где $x_i \in R^N$ вектор обиљежја i-тог тренинг узорка, а y_i ознака класе којој узорак припада. Уколико се на излазу мреже примјењује софтмакс активациона функција, мрежа се обучава минимизацијом *категоричке кросентропије*:

$$J_0(\boldsymbol{W}) = -\frac{1}{n}\sum_{i=1}^{n}\sum_{k=1}^{n^{(L)}} t_{i,k} \ln a_k^{(L)}, \tag{2.5}$$

гдје $\boldsymbol{W}$ представља вектор свих тежина у мрежи, n број тренинг узорака, а $n^{(L)}$ број неурона у излазном слоју. Вриједност $t_{i,k}$ узима вриједност 1 уколико i-ти тренинг узорак има ознаку класе k-те класе, тј. уколико вриједи да је $y_i = c_k$. За $y_i \neq c_k$, $t_{i,k}$ има вриједност 0. Према (2.4), $a_k^{(L)}$ је процијењена вјероватноћа да i-ти узорак припада k-тој класи. Како би се спријечила појава претјераног прилагођења модела, функцији цијене се додаје и регуларизациони члан $R(\boldsymbol{W})$. Хиперпараметром λ се одређује у којој мјери регуларизација има утицај на функцију цијене. Циљ обучавања мреже је да се одреде све тежине тако да функција цијене из (2.6) има минималну вриједност.

$$J(\boldsymbol{W}) = J_0(\boldsymbol{W}) + \lambda R(\boldsymbol{W}). \tag{2.6}$$

Минимизација функције цијене се врши методом *градијентног спуста*, и то алгоритмом *пропагације грешке уназад* [22].

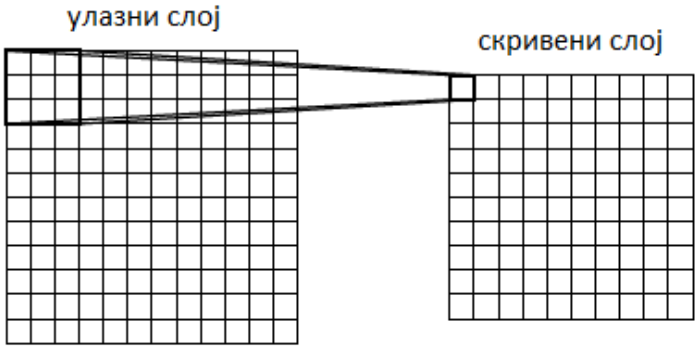
Обученом неуронском мрежом се, идући од улаза ка излазу, из узорка издвајају вектори са информацијама све кориснијим за класификацију. Такви вектори се називају *векторима обиљежја*, чије дужине N одговарају броју неурона у скривеном слоју мреже, помоћу којег се врши издвајање карактеристика. Простор димензионалности која одговара дужини вектора обиљежја N се назива *простором обиљежја*, у којем репрезентација узорка издвојена неуронском мрежом представља једну тачку дефинисану са N координата.

2.1.1. Конволуционе неуронске мреже

У пракси се показало да је обучавање потпуно повезаних неуронских мрежа (мрежа које се састоје од слојева у којима је сваки неурон повезан са свим неуронима из претходног слоја), због великог броја параметара, веома захтјевно. Примјера ради, уколико се обучавање врши на сликама величине 64×64 пиксела, које су притом у RGB колор простору, улазни слој неуронске мреже би имао $64 \times 64 \times 3 = 12.288$ неурона. Ако би сваки неурон из првог скривеног слоја био повезан са свим неуронима из улазног слоја, у том случају би постојало 12.288 различитих тежина које се вежу за само један неурон из скривеног слоја. Када се узме у обзир да овај број треба помножити са бројем неурона из првог скривеног слоја, те поступак поновити крећући се ка излазу мреже, број параметара драстично расте и може се закључити да оптимизација вриједности свих тежина у оваквој мрежи представља прилично захтјеван задатак. Наравно, ситуација се додатно погоршава како резолуција кориштених слика расте. Овако велики број параметара које треба оптимизовати у фази обучавања обично води ка претјераном прилагођењу модела тренинг подацима. Поменути проблеми су били мотив да се размисли о укидању неких веза између неурона и да се формира нова топлогија мреже.

Чињеница је да су везе између просторно блиских пиксела јаче и израженије него оне између удаљених пиксела. Као посљедица тога јавља се идеја да неурон из скривеног слоја буде повезан само са неуронима који одговарају просторно блиским пиксела и да се на тај начин смањи број тежина мреже. Регион на слици сачињен од просторно блиских пиксела

који представљају побудне сигнале неурона у скривеном слоју је *рецептивно поље* тог неурона , а илустрација је дата на Слици 2.3.



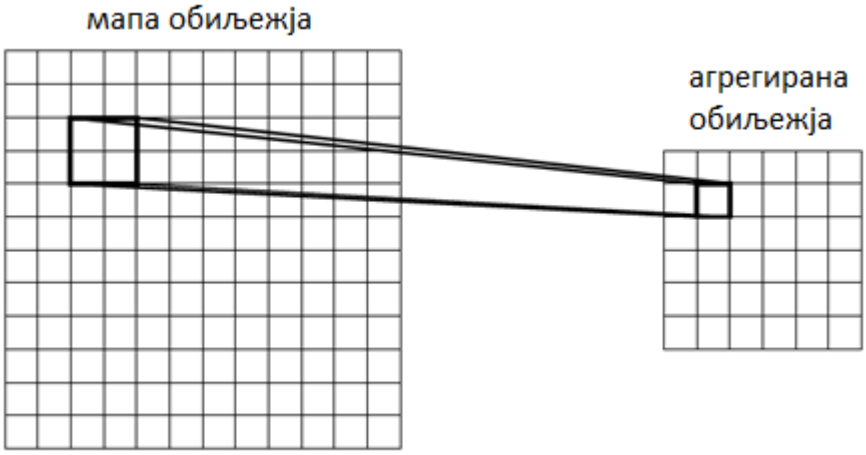
Слика 2.3: Рецептивно поље неурона из скривеног слоја [22].

У циљу поједностављења модела, тежине које одговарају пикселима једног рецептивног поља, без обзира на положај на слици, могу бити једнаке. Скуп тежина једног рецептивног поља представља филтар који служи за издвајање обиљежја из слика. Филтар се транслира по слици и сваком његовом положају одговара један скривени неурон. На тај начин се врши конволуција слике са филтром, што резултује *мапом обиљежја*. За постизање добрих резултата класификације потребно је из слика издвајати обиљежја помоћу више филтара, а у том случају се добијени скривени слој састоји из више мапа обиљежја. Након израчунавања конволуције, на добијени сигнал се може примјенити и одговарајућа нелинеарна активациона функција. Како се обиљежја из слика издвајају примјеном конволуције, скривени слој добијен на описани начин се назива *конволуционим*, а мрежа *конволуционом мрежом*. Дводимензионална конволуција слике f и филтра (конволуционог кернела) h, величине $(2a+1) \times (2b+1)$, се рачуна према:

$$g(i,j) = f(i,j) * h(i,j) = \sum_{k=-a}^{a} \sum_{l=-b}^{b} h(k,l) f(i-k, j-l). \tag{2.7}$$

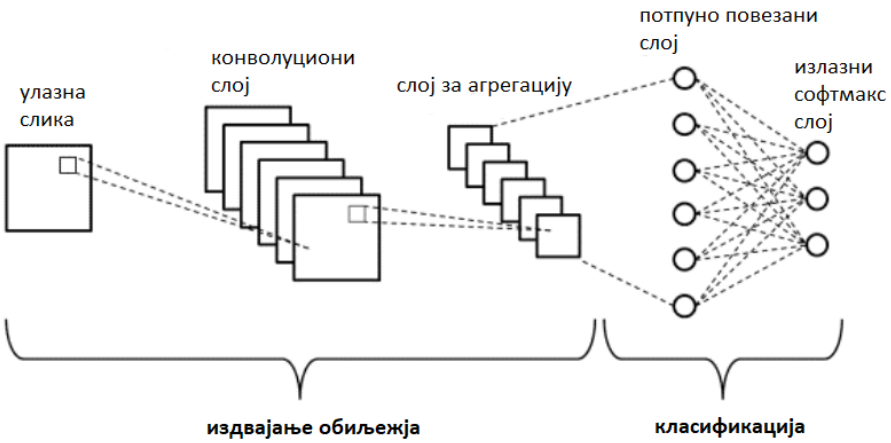
Добијене мапе обиљежја могу представљати улаз у идући конволуциони слој. У случају броја мапа обиљежја већег од 1, користи се тродимензионални конволуциони кернел. Побудни сигнали за нови скривени неурон су онда елементи из рецептивних поља сваке мапе обиљежја.

Како би се количина података додатно смањила, може се извршити пододмјеравање добијених мапа обиљежја. Пододмјеравање се врши рачунањем максимума или средње вриједности елемената из локалног рецептивног поља и слој добијен на овај начин се зове *слојем за агрегацију* (енг. *pooling layer*). При томе, рецептивно поље клизи по мапи обиљежја, али у овом случају не постоји преклапање. Илустративни приказ формирања слоја за агрегацију је дат на Слици 2.4.



Слика 2.4: Слој за агрегацију [22].

Низањем конволуционих слојева и слојева за агрегацију формира се мрежа, а проласком слике кроз низ слојева добија се репрезентација која садржи информације корисне за класификацију. На самом крају мреже се најчешће налази неколико потпуно повезаних слојева, након којих слиједи софтмакс слој којим се процјењују вјероватноће да слика припада појединим класама (енг. *Maximum Softmax Probability* - MSP). Шематски приказ примјера једне конволуционе мреже је дат на Слици 2.5.



Слика 2.5: Шематски приказ архитектуре конволуционе неуронске мреже [22].

2.1.2. Трансформатори за рачунарски вид

Прилично давно је показано да се задаци класификације, детекције или сегментације не могу успјешно рјешавати употребом само сирових података. Из тог разлога су, између осталог, и предлагане сложене архитектуре неуронских мрежа које сирове податке трансформишу у низ информација, корисних да се поменути задаци што успјешније ријеше. Међутим, за разумијевање улазних података је потребно да се они сагледају у цјелини и да се уоче везе између појединих обиљежја. На примјер, у процесу превођења текста са једног језика на други је потребно држати се већ дефинисаног концепта, па је веома битно

анализирати цијеле реченице, а не преводити ријеч по ријеч. У складу са тим је пожељно испитати и везе између ријечи, па закључити у којој мјери је за очување концепта реченице важно да уз једну посматрану ријеч дође и друга. Слична логика се може примјенити и на слике, гдје је битно да се уочи веза између просторно помјерених пиксела. За означавање слике ознаком неке класе или за препознавање облика може бити кључно да се детектује јака веза између више региона пиксела својствена за ту класу. Примјера ради, за препознавање пса или мачке на слици неопходно је да се уоче очи, уши и реп. Уколико се не уоче све три наведене ствари, велика је вјероватноћа да на слици није приказана ни једна од наведених животиња.

У литератури је овакав приступ познат као механизам пажње (енг. *attention mechanism*), којем се раније није придавао посебан значај, иако је био интегрисан у структуру рекурентних и конволуционих мрежа [34]. Међутим, у домену обраде природног језика, 2017. године аутори рада [10] први пут предлажу потпуно нови тип архитектуре који је заснован искључиво на поменутом механизму. Овакав модел, под називом трансформатор (енг. *transformer*), посједује слојеве чија је примарна улога да бројчано опишу везе између појединих ријечи. Резултат *функције пажње* је заправо мјера повезаности између ријечи, те се рачуна по принципу упит-кључ-вриједност. За једну ријеч која се сматра упитом, кључеви и вриједности могу бити све остале и за сваку комбинацију се рачуна скаларни производ вектора репрезентација упита, кључа и вриједности. Коначни вектори представљају нове векторе обиљежја упита, али који након описане процедуре имају више информација о контексту улазних података. Ради лакше анализе и једноставнијег математичког записа сви упити, кључеви и вриједности се могу смјестити у по једну матрицу: Q, K и V. Прије примјене функције пажње вектори обиљежја се могу трансформисати пропуштањем кроз потпуно повезани слој, што је еквивалентно множењу Q, K и V матрица матрицама тежина W^Q, W^K и W^V, чији се елементи уче у процесу обучавања.

$$Q' = Q \cdot W^Q, \tag{2.8}$$

$$K' = K \cdot W^K, \tag{2.9}$$

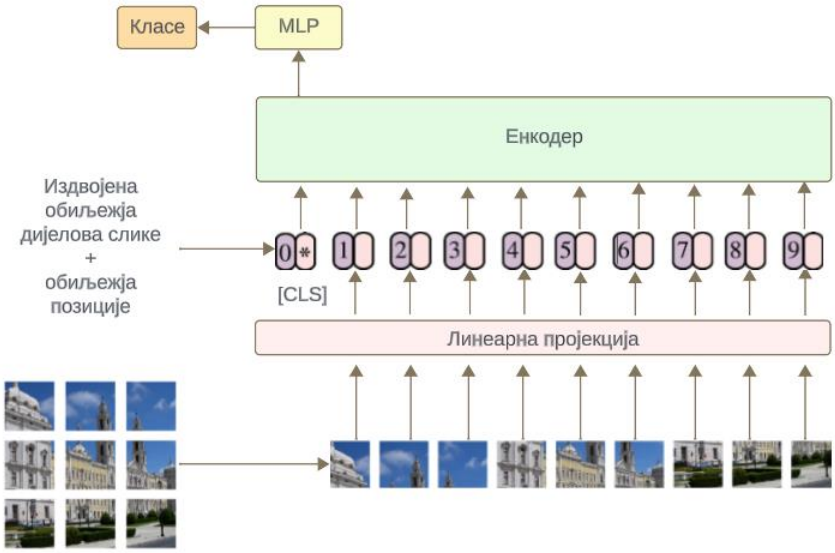
$$V' = V \cdot W^V. \tag{2.10}$$

На описани начин се реализује линеарна пројекција вектора обиљежја упита, кључева и вриједности. Ако је d_k димензија упита и кључа, резултат примјене функције пажње се може записати као:

$$f(Q', K', V') = softmax(\frac{Q' K'^T}{\sqrt{d_k}})V'. \tag{2.11}$$

Ако упите, кључеве и вриједности представљају блокови једне слике или токени једне реченице, ријеч је о механизму самопажње (енг. *self-atenttion mechanism*). Резултат функције пажње говори о томе који токени су највећег приоритета за задатак класификације, те колико су изражене везе између појединих токена.

Инспирисани успјехом у пољу обраде природног језика, аутори у [10] предлажу употребу нове архитектуре и у анализи слика. Иако трансформатори за рачунарски вид почивају на истој идеји као и они кориштени за обраду природног језика, ипак је њихова архитектура модификована и прилагођена раду са сликама. На Слици 2.6 је дат шематски приказ архитектуре трансформатора за рачунарски вид.



Слика 2.6: Шематски приказ архитектуре трансформатора за рачунарски вид.

Највећа разлика у архитектури трансформатора за рачунарски вид у односу на првобитни се односи на одсуство декодера, који је у обради природног језика служио у сврху генерисања нових секвенци ријечи. На примјер, помоћу декодера се генеришу преводи реченица са једног језика на други, док у класификацији слика не постоји потреба тог типа. Дио архитектуре који се задржава и који је кључан за класификацију је *енкодер.*

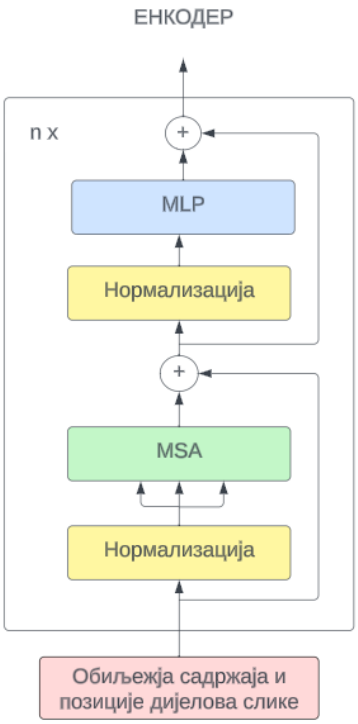
На улазу у трансформатор је потребно слику најприје издијелити на мање квадратне блокове (енг. *patches*). Дакле, слика димензије (H, W) и броја канала C се трансформише у низ N мањих квадратних региона слике димензије (P, P) и броја канала C. У зависности од тога да ли је један квадратни прозор у односу на претходни помјерен за мањи број пиксела од P, између сусједних региона слике постоји или не постоји преклапање. Ако *корак* подјеле слике има вриједност мању од P, преклапање постоји. У супротном, нема преклапања и слика се дијели на број квадратних дијелова који се добија помоћу израза (2.12).

$$N = \mathbb{Z}\{HW/P^2\}. \tag{2.12}$$

Након подјеле слике, сваки од добијених квадратних дијелова је потребно „развући“ у једнодимензионални вектор. На тај начин се добија секвенца од N токена, тј. N вектора пиксела оригиналне слике. Потом се врши линеарна пројекција N једнодимензионалних вектора, што се реализује потпуно повезаним слојем, а математички је представљена изразима (2.8-2.10). У циљу поједностављења модела и смањења броја параметара, на сваки од N једнодомензионалних вектора се примјењује исти вектор тежина. Такође, у оквиру овог корака се може извршити редукција димензионалности вектора који одговарају појединим блоковима на слици. На излазу потпуно повезаног слоја се сваком одређеном вектору обиљежја понаособ додаје и вектор позиције одговарајућег дијела слике. Вектор позиције је исте димензионалности као и вектор обиљежја добијен линеарном пројекцијом, те резултантни вектори обиљежја носе информацију и о садржају појединог региона, али и о његовој позицији на оригиналној слици. Прикупљање информација о позицијама блокова на којима се уочавају поједина обиљежја је једна од битнијих предности трансформатора за

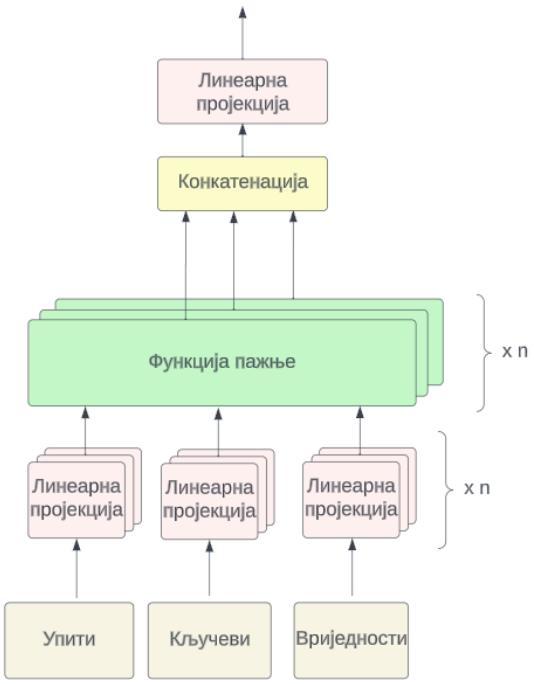
рачунарски вид у односу на раније коришћене архитектуре, а између осталог и конволуционе неуронске мреже.

Вектори обиљежја са информацијама о садржају и позицији се воде на улаз енкодера, чија је структура приказана дијаграмом на Слици 2.7. Енкодер чини низ наизмјенично везаних вишеструких слојева у којима се имплементира механизам самопажње и вишеслојних перцептрона (енг. *Multilayer Perceptron* - MLP). Уобичајено је да се више пута, паралелно, уради линеарна пројекција обиљежја упита, кључева и вриједности, употребом тежина које се уче у процесу обучавања. На тако трансформисане упите, кључеве и вриједности се паралелно примјењује функција пажње, на примјер према (2.8), па се каже да се механизам самопажње реализује у више глава (енг. *multiheaded self-attention* - MSA). На овај начин се моделу омогућава да сагледа информације са бројних гледишта, тако што улазе пројектује у различите просторе репрезентације и у сваком од њих примјени посебан механизам пажње. Добијени резултати се конкатенирају и представљају нову репрезентацију слике. Нова обиљежја се воде на улаз MLP блока, који у овом случају представља релативно малу неуронску мрежу са нелинеарном активационом функцијом у једном од слојева. Прије сваког уласка у MSA или MLP блок врши се нормализација.



Слика 2.7: Структурни блок дијаграм енкодера трасформатора за рачунарски вид.

Са Слике 2.6 се може уочити да се на самом улазу у енкодер, осим вектора обиљежја који одговарају блоковима оригиналне слике, налази још један вектор. У литератури се поменути вектор наводи као [CLS] токен (токен за класификацију, енг. *classify token*) [10]. Вриједности елемената додатног вектора се бирају насумично, док се тежине којима се он трансформише уче у фази обучавања, при чему на излазу енкодера овај вектор представља репрезентацију полазне слике у цјелини. Трансформисани [CLS] токен се са излаза енкодера води у посљедњи MLP блок, у оквиру којег се и врши класификација.



Слика 2.8: Шематски прприказ реализације MSA блока.

Број понављања процедуре са Слике 2.7 унутар енкодера и број паралелних извршавања поступка са Слике 2.8, а самим тим и број параметара модела, зависи од сложености задатка и броја тренинг узорака. У пракси се углавном користе стандардне архитектуре различите сложености са фиксним бројем скривених слојева, као што је предложено у [10], а дато у Табели 2.1.

Табела 2.1: Детаљи различитих архитектура трансформатора за рачунарски вид.

Архитектура модела	**Број скривених слојева**	**Димензионалност скривеног слоја**	**Димензионалност MLP слоја**	**Број „глава“**	**Број параметара**
ViT-Base	12	768	3.072	12	86 мил.
ViT-Large	24	1.024	4.096	16	307 мил.
ViT-Huge	32	1.280	5.120	16	632 мил.

Број скривених слојева се односи на то колико пута се унутар трансформатора понавља поступак са Слике 2.7. Са друге стране, број „глава“ је број паралелних израчунавања функције пажње у оквиру MSA блока. Димензионалност слоја се односи на број неурона у њему, а самим тим и на дужину вектора обиљежја на излазу тог слоја.

У пракси је устаљена нотација за означавање модела из које се лако закључује величина модела и величина дијелова на које се оригинална слика дијели. На примјер, ознака

модела ViT-L/16 се односи на другу по сложености варијанту трансформатора из Табеле 2.1 и на подјелу оригиналне слике на квадратне дијелове величине 16×16 пиксела.

Резултати у [10] показују да се трансфоматорима за рачунарски вид могу остварити бољи резултати него конволуционим мрежама, али уколико су обучени на довољно великом броју података. У пракси су у сврху обучавања трансформатора за рачунарски вид устаљена три скупа слика различитих величина:

1. Imagenet-1k [35] са 1,3 милиона слика подиељених у 1.000 класа,
2. Imagenet-21k [36] са 14 милиона слика подијељених у 21.000 класа,
3. JFT [37, 38] са 300 милиона слика подијељених у 18.000 класа.

У оквиру [10] је показана надмоћ конволуционе архитектуре над трансформаторском уколико су обје претриниране на Imagenet-1k колекцији, док је у случају њиховог обучавања на JFT колекцији изведен супротан закључак. Стога, при избору претренираног трансформаторског модела треба водити рачуна о величини скупа на којем је вршено обучавање. Након избора претренираног модела, уобичајено је да се уради фино подешавање параметара на неком мањем скупу од интереса.

3. Детекција слика које одступају од расподјеле

У пракси се често, ради одржавања поузданости и сигурности система, намеће потреба за детекцијом узорака који одступају од расподјеле тренинг података. На примјер, пожељно је да систем за аутономну вожњу преда контролу човјеку уколико се примјете необичне сцене или се уоче објекти који се разликју од оних виђених у фази обучавања.

Заједничка особина већине метода за детекцију узорака који одступају од расподјеле је да се обиљежја узорака издвајају помоћу основног класификатора, претходно обученог на тренинг, тј. ID подацима. Разликују се по начину на који се формира критеријум за OOD детекцију, те се према томе методе могу сврстати у неколико типова, већ наведених у Уводу.

Тип метода заснованих на класификацији је први који се примјењивао у OOD детекцији. Ако се за детекцију користе информације које даје класификатор, најчешће се при одлучивању којој расподјели припада узорак користи податак о максималној софтмакс вјероватноћи (MSP), добијеној на излазу мреже [39]. Уобичајено је да тачно класификовани узорци имају веће вриједности максималне софтмакс вјероватноће на излазу мреже, него нетачно класификовани. Стога, у случају да је MSP вриједност већа од неког усвојеног прага, тестни узорак се означава као ID, а у супротном као OOD. Као посљедица једноставности овог метода, често се у литератури он сматра полазним и референтним, па се перформансе нових предложених метода пореде са његовим перформансама.

Нешто захтјевнији, али углавном и успјешнији приступ проблему се односи на процјену расподјеле вјероватноћа података из тренинг скупа. Рецимо, може се поћи од претпоставке да тренинг подаци припадају нормалној, Гаусовој расподјели, а потом израчунати њене параметре - центроид и матрицу коваријанси. За сваки тестни узорак се онда може израчунати Махаланобисова удаљеност од центроида (енг. *Mahalanobis distance* - MD), која представља мјеру одступања узорка од расподјеле [40]. Мала Махаланобисова удаљеност значи већу вјероватноћу да узорак припада естимираној расподјели, тј. да ће бити означен као ID узорак. Као посљедица уочавања нешто слабијих резултата које даје овај метод на OOD скуповима који нису много удаљени од тренинг скупа, објављена је његова модификација под називом релативна Махаланобисова удаљеност (енг. *relative Mahalanobis distance* - RMD) [41]. Наиме, након записивања Махаланобисових удаљености у другачијем облику [42] настаје приједлог о естимацији Гаусове расподјеле на тренинг подацима, али не узимајући у обзир припадност класама. Махаланобисова удаљеност узорка од центроида класно-независне расподјеле се одузима од Махаланобисових удаљености узорка од центроида примарно естимиране класно-зависне расподјеле, те се добијена разлика користи за означавање тестног узорка. Показује се да поређење поменуте релативне Махаланобисове удаљености са усвојеним прагом у многим примјенама даје боље резултате.

Идеја од које полазе методи базирани на реконструкцији је да примјена енкодер-декодер система на тестне узорке резултује различитих исходима за ID и OOD узорке [43]. С обзиром на то да се модели за рекострукцију тренирају само на тренинг скупу, очекивано је да реконструкција OOD узорка не буде добра. Стога, лоше перформансе реконструкционог модела на неком тестном узорку упућују на припадност тог узорка OOD скупу. Умјесто реконструкције цијеле слике, у [44] је предложена употреба само насумично одабраног дијела слике, чији квалитет реконструкције даје информацију о томе да ли је оригинална слика OOD узорак.

Методи засновани на мјерењу удаљености су од највећег интереса за овај рад, те ће неки од њих бити детаљније обрађени у наставку текста.

3.1. Методи засновани на мјерењу удаљености

Логика од које се полази у свим методима базираним на мјерењу удаљености [40-42, 45-50] је да се очекује да су тестне ID слике, у простору обиљежја, ближе скупу тренинг слика него тестне OOD слике. За одређивање удаљености двије слике у простору обиљежја могу да се користе различите метрике: косинусна сличност, Еуклидова удаљеност, Менхетн метрика, Минковски метрика итд. Мјера за детекцију OOD узорака се често формира примјеном основних математичких операција на вриједности удаљености два или више парова слика.

3.1.1. Алгоритам k-најближих сусједа

Поступак у којем се неозначени узорак класификује на основу ознака класа неколико блиских означених узорака се први пут спомиње још 1951. године [51], а значајно проширење алгоритма је остварено у раду [52]. Нешто касније, предложени метод постаје познат као алгоритам *k*-најближих сусједа (енг. *k-Nearest Neighbors* - KNN). Правило по којем овај алгоритам ради се односи на сврставање тестног узорка у ону класу којој припада већина означених узорака из скупа њему k најближих сусједа, водећи се претпоставком да су два узорка из исте класе ближе позиционирана у простору обиљежја од оних који припадају различитим класама. Са појавом неуронских мрежа отварају се нове могућности у погледу репрезентације узорака, тј. њиховог представљања вектором обиљежја. Паралелно томе, долази до побољшања перформанси и KNN алгоритма, јер вектори обиљежја издвојени неуронским мрежама носе информације значајније за класификацију од вектора са „сировим“ информацијама. Конкретно за проблем класификације слика, вектор обиљежја слике добијен непосредно прије софтмакс излазног слоја има далеко више информација о семантици слике него вектор вриједности пиксела те слике. У складу са тим, јасно је да је у простору обиљежја креираном помоћу неуронске мреже још израженија близина вектора обиљежја слика из исте класе, него када су обиљежја била само вриједности пиксела.

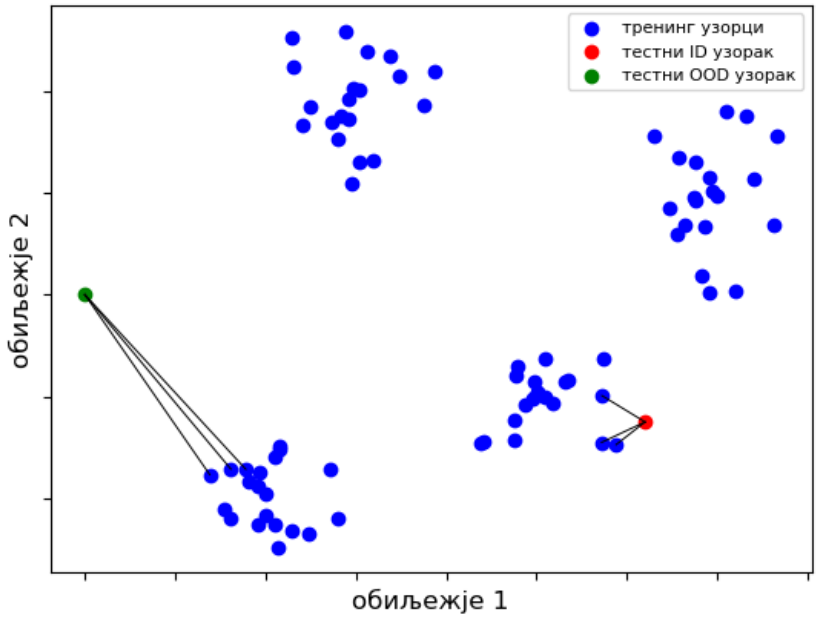
Упркос једноставности KNN алгоритма, требало је дуго времена да се, након популаризације неуронских мрежа, јави идеја за његовим прилагођењем и тестирањем на проблему класификације и детекције слика. Тек 2022. године се у [48] и [49] помиње овако једноставан приступ проблему детекције узорака који одступају од расподјеле. Наиме, као ID/OOD мјера може бити искориштена средња вриједност удаљености тестног узорка од k најближих тренинг узорака. У том случају, ако $\boldsymbol{z}^*$ представља вектор обиљежја тестног узорка, а $\boldsymbol{z}_i$, $i = 1, \dots k$, k најближих сусједа тестном узорку и θ усвојени праг, тестни узорак $\boldsymbol{z}^*$ ће бити детектован као OOD уколико је задовољена неједнакост (3.1).

Као мјера удаљености између два узорка се најчешће користи косинусна удаљеност $d(\boldsymbol{z}^*, \boldsymbol{z}_i)$, која се добија према изразу (3.2), одузимањем косинусне сличности од 1. Угао α представља угао између вектора $\boldsymbol{z}^*$ и $\boldsymbol{z}_i$.

$$-\frac{1}{k} \cdot \sum_{i=1}^{k} d(\boldsymbol{z}^*, \boldsymbol{z}_i) \leq \theta, \tag{3.1}$$

$$d(\boldsymbol{z}^*, \boldsymbol{z}_i) = 1 - \cos\alpha = 1 - \frac{\boldsymbol{z}^* \boldsymbol{z}_i^T}{\|\boldsymbol{z}^*\| \|\boldsymbol{z}_i\|}. \tag{3.2}$$

Број најближих сусједа k се може сматрати хиперпараметром, а праг θ не зависи од OOD података. Бира се тако да 95% тестних ID узорака буде тачно означено, што се



Слика 3.1: Графички приказ детекције OOD узорака помоћу KNN алгоритма у случају $k = 3$

подразумијева и у оквиру осталих метода. На Слици 3.1 је, осим тренинг узорака, приказана и позиција једног ID тестног и једног OOD тестног узорка, као и илустрација принципа по којем се детекција врши.

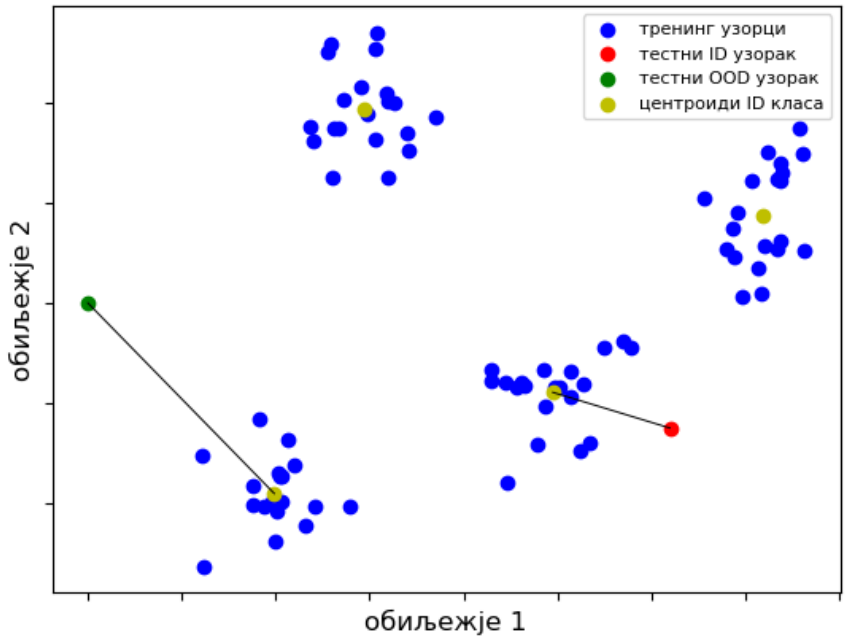
3.1.2. Удаљеност од најближег центроида

Може се поћи од високо вјероватне претпоставке да ће се тестни ID узорак наћи у близини центроида одговарајуће класе, одређеног позицијама тренинг узорака те исте класе. Дакле, још прије тестне фазе је могуће одредити центроиде свих класа коришћених у фази обучавања. Непосредно након појаве тестног узорка, потребно је израчунати удаљености од свих центроида и уочити минималну вриједност међу њима. Одлука да ли је тестни узорак ID или OOD се доноси на основу поређења вриједности удаљености тог узорка од најближег центроида са усвојеним прагом. Узорак се означава као OOD ако је задовољена неједнакост:

$$-\min_{1\leq k\leq n_c}\left\{d(\boldsymbol{z}^*, \boldsymbol{z}_{c,k})\right\} \leq \theta. \tag{3.3}$$

Са $\boldsymbol{z}^*$ је означен вектор обиљежја тестног узорка, са $\boldsymbol{z}_{c,k}$ вектор обиљежја центроида k-те класе, n_c представља број класа и θ усвојени праг.

Метод детекције заснован на одређивању удаљености од најближег центроида (енг. *Nearest Class Mean* - NCM) се може модификовати на начин да се памте тестни OOD узорци који формирају нове класе. На тај начин је могуће реализовати инкрементално побољшање перформанси детектора, тако што се посматрају и вриједности удаљености тестних узорака од центроида нових класа [45, 46]. У овом раду ће се у оквиру овог метода користити само центроиди добијени на основу тренинг слика.



Слика 3.2: Графички приказ детекције OOD слика помоћу NCM детектора

3.1.3. Однос удаљености

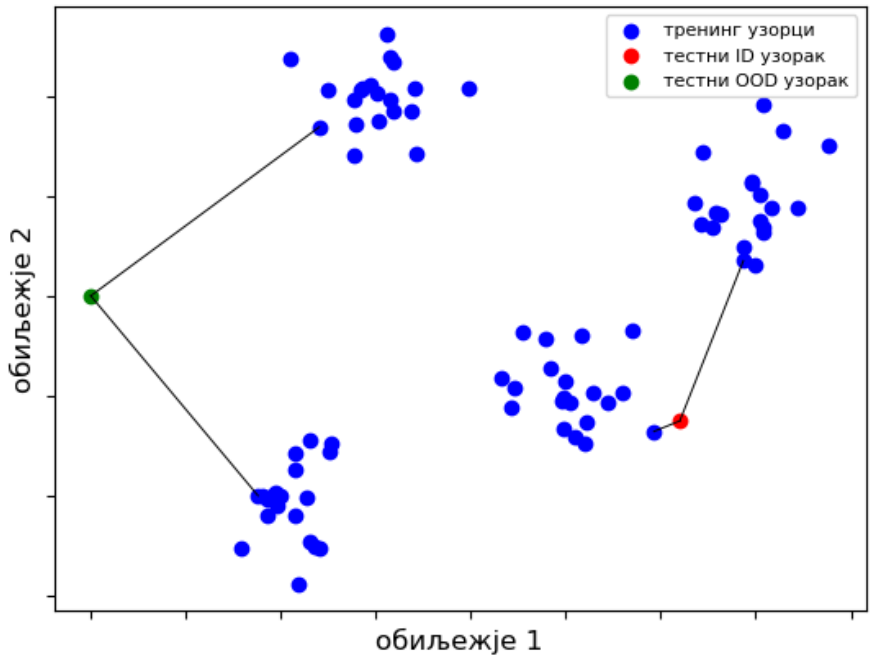
Иако је метод заснован на рачунању односа удаљености тестног узорка од два најближа тренинг узорка из двије различите класе (енг. *Nearest Neighbor Distance Ratio* – NNDR) предложен у сврху рјешавања OSR проблема [53], добијени однос се лако може протумачити на начин да се користи за OOD детекцију. Дакле, за тестни узорак $\boldsymbol{z}^*$ је потребно одредити најближи тренинг узорак $\boldsymbol{z}_1$ и идући најближи тренинг узорак $\boldsymbol{z}_2$, тако да је $g(\boldsymbol{z}_1) \neq g(\boldsymbol{z}_2)$, где $g(\boldsymbol{z}) \in C = \{c_1, c_2, \dots, c_k\}$ представља индекс класе којој припада узорак $\boldsymbol{z}$. Потребно је израчунати однос удаљености:

$$R = d(\boldsymbol{z}^*, \boldsymbol{z}_1)/d(\boldsymbol{z}^*, \boldsymbol{z}_2), \tag{3.4}$$

гдје $d(\boldsymbol{z}, \boldsymbol{z}')$, на примјер, може бити Еуклидова или косинусна удаљеност. Одлука о сврставању тестног узорка у ID или OOD скуп се доноси на основу израза:

$$z^* \in \begin{cases} ID, & \text{ако} - R > T, \\ OOD, & \text{ако} - R \leq T, \end{cases} \tag{3.5}$$

у којем је T усвојени праг из опсега (-1, 0).



Слика 3.3: Графички приказ детекције OOD узорака помоћу NNDR детектора

3.1.4. Махаланобисова удаљеност

У раду [40] је предложен метод за детекцију узорака ван расподјеле тренинг података, који се заснива на претпоставци да узорци једне класе припадају нормалној мултиваријантној Гаусовој расподјели. За процјену параметара Гаусове расподјеле $N(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ користе се d-димензионални вектори обиљежја тренинг узорака, при чему d одговара броју неурона у слоју неуронске мреже помоћу којег се издвајају обиљежја. Ситуација се додатно поједностављује увођењем претпоставке да Гаусове расподјеле свих ID класа дијеле исту матрицу коваријанси. У складу са свим наведеним, d-димензиони центроид $\boldsymbol{\mu}_c$ класе c и $d \times d$ матрица коваријанси $\boldsymbol{\Sigma}$ се рачунају на сљедећи начин:

$$\boldsymbol{\mu}_c = \frac{1}{N_c} \cdot \sum_{i:y_i=c} \boldsymbol{z}_i, \tag{3.6}$$

$$\boldsymbol{\Sigma} = \frac{1}{\mathrm{N}} \sum_c \sum_{i:y_i=c} (\boldsymbol{z}_i - \boldsymbol{\mu}_c)(\boldsymbol{z}_i - \boldsymbol{\mu}_c)^T, \tag{3.7}$$

гдје N_c представља број ID узорака који припадају класи c, N број свих узорака, y_i ознаку класе и $\boldsymbol{z_i}$ вектор обиљежја i-тог узорка. Вриједност корисна за означавање тестног узорка ID или OOD класом је квадрат *Махаланобисове удаљености*, која представља одступање узорка од Гаусове расподјеле и добија се на начин:

$$MD_c^2(\boldsymbol{z}_i) = (\boldsymbol{z}_i - \boldsymbol{\mu}_c)^T \boldsymbol{\Sigma}^{-1} (\boldsymbol{z}_i - \boldsymbol{\mu}_c). \tag{3.8}$$

Потребно је израчунати Махаланобисову удаљеност тестног узорка од Гаусове расподјеле сваке класе, а вриједност кључна за означавање узорка је негативна вриједност

минималног елемента добијеног скупа, која се пореди са усвојеним прагом. Ако је неједнакост у (3.9) испуњена, узорак довољно одступа од процијењених Гаусових расподјела ID класа да би се означио као OOD.

$$-\min_{c} MD_c(\boldsymbol{z}_i) \leq \theta. \tag{3.9}$$

3.1.5. Релативна Махаланобисова удаљеност

Записивање Махаланобисове удаљености вектора обиљежја тестног узорка $\boldsymbol{z}$ од мултиваријабилне Гаусове расподјеле $N(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ преко суме Махаланобисових удаљености везаних за једну димензију простора обиљежја указује на недостатак претходно описаног метода [41]. Наиме, ако $\boldsymbol{v}_d$ и λ_d редом представљају d-ти сопствени вектор и d-ту сопствену вриједност матрице коваријанси $\boldsymbol{\Sigma}$, може се израчунати пројекција $(\boldsymbol{z} - \boldsymbol{\mu})$ на $\boldsymbol{v}_d$ као $l_d = \boldsymbol{v}_d^T(\boldsymbol{z} - \boldsymbol{\mu})$. У том случају се l_d^2/λ_d може посматрати као Махаланобисова удаљеност пројекције вектора $(\boldsymbol{z} - \boldsymbol{\mu})$ од једнодимензионе Гаусове расподјеле $N(0, \lambda_d)$:

$$MD^2(\boldsymbol{z}) = (\boldsymbol{z} - \boldsymbol{\mu})^T \boldsymbol{\Sigma}^{-1}(\boldsymbol{z} - \boldsymbol{\mu}) = \sum_{d=1}^{D} l_d^2/\lambda_d. \tag{3.10}$$

Експерименти на CIFAR-100 ID скупу и CIFAR-10 OOD скупу су показали да OOD узорци имају значајно веће Махаланобисове удаљености од ID узорака по 120 димензија са највећим сопственим вриједностима, док су удаљености по осталим димензијама приближно једнаке за ID и OOD узорке [41]. То значи да уважавање Махаланобисових удаљености по осталим димензијама (димензијама које одговарају мањим сопственим вриједностима) у изразу (3.10) само отежава разликовање ID узорака од OOD. Предложени начин за минимизацију утицаја Махаланобисових удаљености по мање значајним димензијама укључује естимацију мултиваријабилне, класно-независне, Гаусове расподјеле тренинг узорака $N(\boldsymbol{\mu}_0, \boldsymbol{\Sigma}_0)$:

$$\boldsymbol{\mu}_0 = \frac{1}{N}\sum_{i=1}^{N} \boldsymbol{z}_i, \tag{3.11}$$

$$\boldsymbol{\Sigma}_0 = \frac{1}{N}\sum_{i=1}^{N} (\boldsymbol{z}_i - \boldsymbol{\mu}_0)(\boldsymbol{z}_i - \boldsymbol{\mu}_0)^T. \tag{3.12}$$

Махаланобисова удаљеност тестног узорка $\boldsymbol{z}_i$ од класно-независне Гаусове расподјеле $N(\boldsymbol{\mu}_0, \boldsymbol{\Sigma}_0)$ је:

$$MD_0(\boldsymbol{z}_i) = (\boldsymbol{z}_i - \boldsymbol{\mu}_0)^T \boldsymbol{\Sigma}_0^{-1}(\boldsymbol{z}_i - \boldsymbol{\mu}_0). \tag{3.13}$$

Разлика Махаланобисове удаљености узорка од Гаусове расподјеле класе c и процијењене класно-независне Гаусове расподјеле представља *релативну Махаланобисову удаљеност*:

$$RMD_c(\boldsymbol{z}_i) = MD_c(\boldsymbol{z}_i) - MD_0(\boldsymbol{z}_i). \tag{3.14}$$

На крају, узорак се означава као OOD уколико је испуњена неједнакост (3.15), у којој θ представља усвојени праг:

$$-\min_{c} RMD_c(\boldsymbol{z}_i) \leq \theta. \tag{3.15}$$

Како се за одређивање Махаланобисове и релативне Махаланобисове удаљености користи инверз коваријансне матрице, лако се закључује да она мора бити *инвертибилна* (тј. *регуларна*). За разлику од *синуларне* матрице, регуларна матрица има ненулту детерминанту. Најчешћи узрок сингуларности, а самим тим и неинвертибилности процијењене коваријансне матрице је мали број узорака на основу којих се врши процјена, у односу на димензионалност простора обиљежја. Стога, у ситуацијама у којима се располаже малим бројем тренинг узорака матрица коваријанси је често сингуларна, те је неопходно извршити њену *регуларизацију*. Поступак регуларизације је наведен у оквиру израза (3.16), у којем $\boldsymbol{\Sigma}$ представља синуларну матрицу коваријанси, $\boldsymbol{\Sigma}_R$ регуларизовану матрицу коваријанси, $\boldsymbol{I}$ јединичну матрицу, а λ коефицијент регуларизације релативно мале вриједности (нпр. 0,01 или 0,001):

$$\boldsymbol{\Sigma}_R = \boldsymbol{\Sigma} + \lambda \cdot \boldsymbol{I}. \tag{3.16}$$

Дакле, проблем неинвертибилности коваријансне матрице се једноставно може ријешити додавањем мале вриједности λ дијагоналним елементима.

3.2. Излагање детектора узорцима који одступају од расподјеле

Употреба узорака који одступају од ID расподјеле показала се као ефектан начин за побољшање способности детектора да препозна OOD узорке. На тај начин детектор може да научи битне карактеристике узорака који одступају од расподјеле тренинг података, тј. оних узорака који се потенцијално могу јавити на улазу детектора. У наставку текста ће се узорци који одступају од ID расподјеле, а који су изабрани у сврху моделовања тестних OOD узорака, називати OE узорцима. Мотивација за овакав назив потиче од *outlier exposure* [13-19] метода, чији је циљ побољшање перформанси детекције помоћу изабраних узорака који одступају од расподјеле тренинг скупа.

Посебну пажњу треба посветити бирању узорака који одступају од расподјеле, јер од тога зависи колико ће побољшање донијети цијели OE поступак. У складу са тим, поставља се питање о препоручљивој сличности скупа изабраних OE узорака са ID и OOD тестним скупом. У [13] добијени резултати показују да за остварење побољшања перформанси детектора сличност између OE и тестног OOD скупа није кључна. Са друге стране, битно је да се скуп OE налази довољно близу тестног ID скупа у простору обиљежја, јер се на тај начин моделују OOD узорци најизазовнији за детекцију. Међутим, уколико се оствари превелика сличност између OE и тестног ID скупа, детектор у процесу тренирања неће научити нове информације о OOD подацима. Из тог разлога је битно наћи компромис, који се у већ поменутом раду остварује коришћењем како дањих, тако и ближих OE узорака ID узорцима, при чему се ближим, тј. сличнијим, придаје нешто већи значај.

Понекад избор OE узорака представља изразито тежак задатак, нарочито у ситуацијама у којима ID класе припадају истој надкласи. Конкретан примјер би била класификација слика птица у N врста, при чему би у тестној фази чак и слике које приказују неке друге врсте биле означене као OOD. У тим ситуацијама се показало да се моделовање

OOD скупа побољшава креирањем виртуелних OE узорака, који се генеришу примјеном MixUp [21], CutMix [54], ManifoldMixup [55] или сличних операција на парове тренинг слика и реалних OE узорака. Осим тога, у раду [19] је показано да се OE скуп може добити полазећи само од ID података. Такав приступ је предложен у пољу обраде природног језика, гдје се из текста који припада тренинг скупу елиминише токен који у највећој мјери указује на ID расподјелу (метод заснован на формирању псеудо узорака који одступају од расподјеле, енг. *pseudo outliers*). Поступак се понавља док се не добије репрезентација текста која довољно одступа од ID расподјеле. Иако овај метод још увијек није примјењиван на проблем детекције слика које одступају од расподјеле, вриједи у оквиру будућег рада размишљати и у том правцу, водећи се истом логиком.

У оквиру рада [56] добијени резултати на карактеристичном примјеру показују у којој мјери комбинација доброг основног класификатора и добро изабраних OE узорака резултује побољшањем AUROC (енг. *Area Under the Receiver Operating Characteristic*) вриједности. Наиме, уколико се обиљежја издвајају помоћу трансформатора са фино подешеним параметрима на ID тренинг скупу CIFAR100, детекција OOD слика из CIFAR10 се врши са AUROC од 96%. Међутим, са само једним OE узорком по OOD класи се остварује AUROC вриједност од 98,7%, док је та вриједност у случају употребе десет узорака по OOD класи чак 99,46%.

Иако се завидни резултати постижу и употребом OE скупова дисјунктиних са тестним OOD, у оквиру [18] се иде корак напријед и колекција OE се попуњава OOD узорцима који се у тестној фази јаве на улазу детектора. Предност овог метода је могућност да детектор учи директно од тестних узорака, док је основна мана потреба за честим подешавањем параметара сложених модела, што захтијева утрошак значајног времена, али и располагање техничким ресурсима. Из тог разлога се прибјегава потрази за методом који отвара могућности да се OE узорци прикупљају прије тестне фазе, те који не захтјева измјену параметара основног класификатора.

3.2.1. Модификација метода детекције заснованог на рачунању односа удаљености

Ка циљу да се отвори могућност за одабир репрезентативних узорака који одступају од расподјеле тренинг скупа уз тежњу да се бирање врши прије тестне фазе и да се избјегне потреба за модификацијом основног класификатора, јавља се идеја за надоградњом једног од раније описаних метода за детекцију. За почетак, нека су сви тренинг узорци смјештени у ID домен D_{ID}, а одабрани узорци који одступају од расподјеле тренинг података се налазе у OOD домену D_{OOD}. Подразумијевано је да све тачке које у ID домену представљају тренинг узорке, усљед малих удаљености између узорака из исте класе, у простору обиљежја формирају онолико група колико има ID класа. Питање о броју класа у које се дијеле изабрани узорци који одступају од расподјеле, а који су у OOD домену, остаје отворено. Неки приједлози за груписање OE узорака су:

- $k - means$ кластеризација,
- Означавање OE узорака ознаком која указује на најближу ID класу, при чему се удаљеност може рачунати до најближег центроида класе, помоћу алгоритма k најближих сусједа и слично,
- Означавање OE узорака ознаком која указује на ID класу за коју тестни узорак има највећу MSP вриједност.

Као и у оквиру метода за детекцију OOD узорака базираном на рачунању односа удаљености тестног од два тренинг узорка из различитих класа, на идентичан начин можемо да размишљамо и уколико посматрамо OOD домен. Дакле, уколико изабрани OE узорци достојно моделују тестни OOD скуп, очекивано је да се њима у простору обиљежја ближе налазе OOD, него ID тестни узорци. У складу са очекиваним можемо дефинисати три

карактеристична односа удаљености, чијом комбинацијом може да се добије ефектан критеријум за сврставање тестног узорка у ID или OOD класу:

- r_{ID}- однос удаљености тестног од два најближа тренинг узорка која припадају различитим ID класама (одјељак 3.1.3),
- r_{OOD}- однос удаљености тестног од два најближа OE узорка која припадају различитим OOD класама и
- $r_{ID,OOD}$ - однос удаљености тестног узорка од најближег тренинг и од најближег OE узорка.

Треба имати у виду да би тестни ID узорци требали да имају значајно мање односе удаљености r_{ID} и $r_{ID,OOD}$, а значајно већи однос r_{OOD} у поређењу са тестним OOD узорцима, па се у складу са тим предлажу два критеријума за OOD детекцију. Дакле, тестни узорак се означава као OOD уколико је испуњена неједнакост (3.17) или неједнакост (3.18).

$$-\frac{r_{ID}}{r_{OOD}} \leq \theta, \tag{3.17}$$

$$-\frac{r_{ID} \cdot r_{ID,OOD}}{r_{OOD}} \leq \theta. \tag{3.18}$$

У оквиру експерименталног дијела у овом раду тестирана су два метода заснована на комбинацији односа удаљености из израза (3.17) и (3.18) и на основу добијених резултата је дата предност једном од њих.

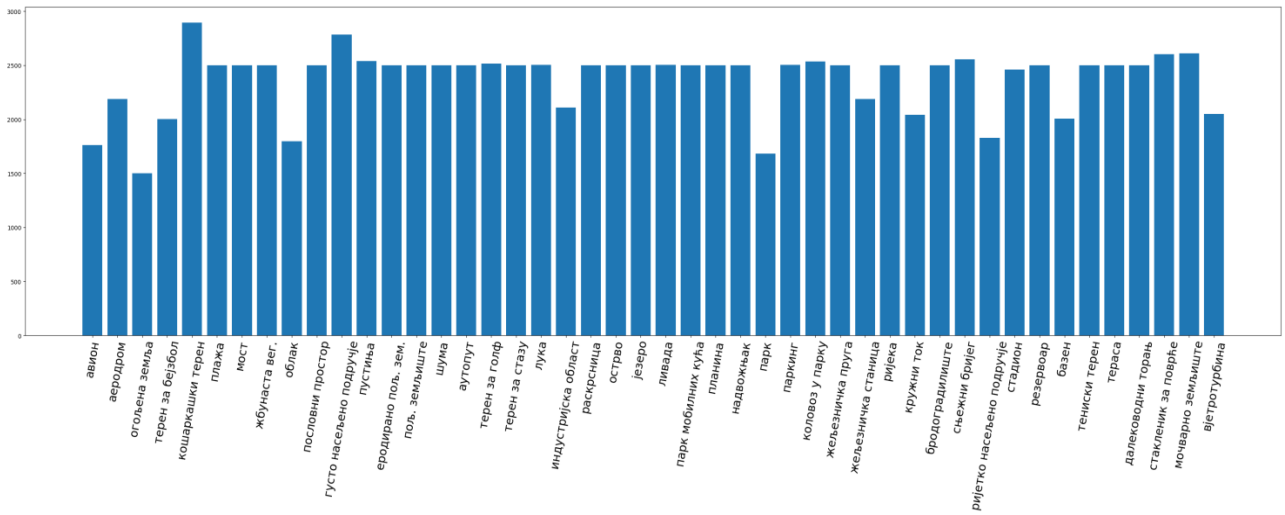
4. Материјал и методологија

У оквиру практичног дијела експерименти су вршени користећи шест колекција слика добијених даљинском детекцијом: MLRSNet [57], NWPU-RESISC45 [58], PatternNet [59], AID [60], UC Merced Land Use [61], MillionAID [62] и три колекције слика из других домена: Food5K [63], ImageNet100 [64] и Imagenette [65]. Све кориштене слике су RGB, а за издвајање обиљежја су кориштени сљедећи модели: ResNet50 [66], ViT-base-patch16-224 [67] и ViT-large-patch16-224 [68].

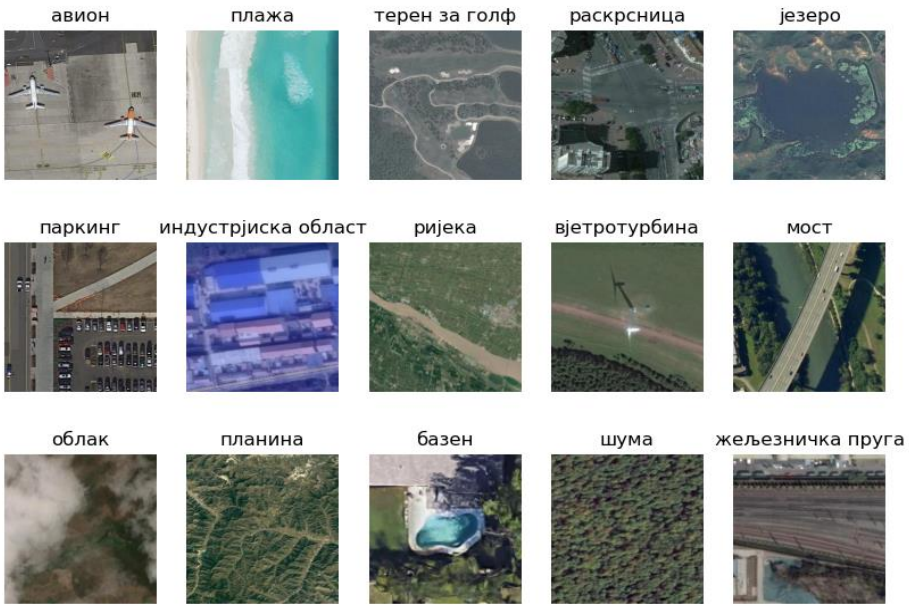
4.1. Колекција слика MLRSNet

MLRSNet скуп слика добијених даљинском детекцијом је представљен у оквиру [57]. У наведеном раду је демонстрирана употреба предложеног скупа података на задацима препознавања и класификације слика, те су разни методи из ових области тестирани на новој MLRSNet колекцији слика.

MLRSNet скуп података се састоји од 109.161 снимака Земљине површине, подијељених у 46 категорија: *авион, аеродром, голо земљиште, терен за бејзбол, кошаркашки терен, плажа, мост, жбунаста вегетација, облак, пословни простор, густо насељено подручје, пустиња, еродирано пољопривредно земљиште, пољопривредно земљиште, шума, аутопут, терен за голф, атлетски стадион, лука, индустријска област, раскрсница, острво, језеро, ливада, парк мобилних кућа, планина, надвожњак, парк, паркинг, коловоз у парку, жељезничка пруга, жељезничка станица, ријека, кружни ток, бродоградилиште, сњежни бријег, ријетко насељено подручје, стадион, резервоар, базен, тениски терен, тераса, далеководни торањ, стакленик за површе, мочварно земљиште и вјетротурбина*. Расподјела слика по класама није униформна, а број узорака по класама варира између 1.500 и 3.000. Расподјела слика по појединим категоријама је графички приказана на Слици 4.1.



Слика 4.1: Расподјела слика по класама за MLRSNet колекцију слика.



Слика 4.2: Примјери слика из колекције MLRSNet.

На Слици 4.2 приказан је по један примјер слике из петнаест насумично одабраних класа из MLRSNet колекције слика. Све слике из ове колекције су фиксне величине од 256×256 пиксела, док је просторна резолуција пиксела промјењива и креће се приближно између 10 m и 0,1 m.

4.2. Колекција слика NWPU-RESISC45

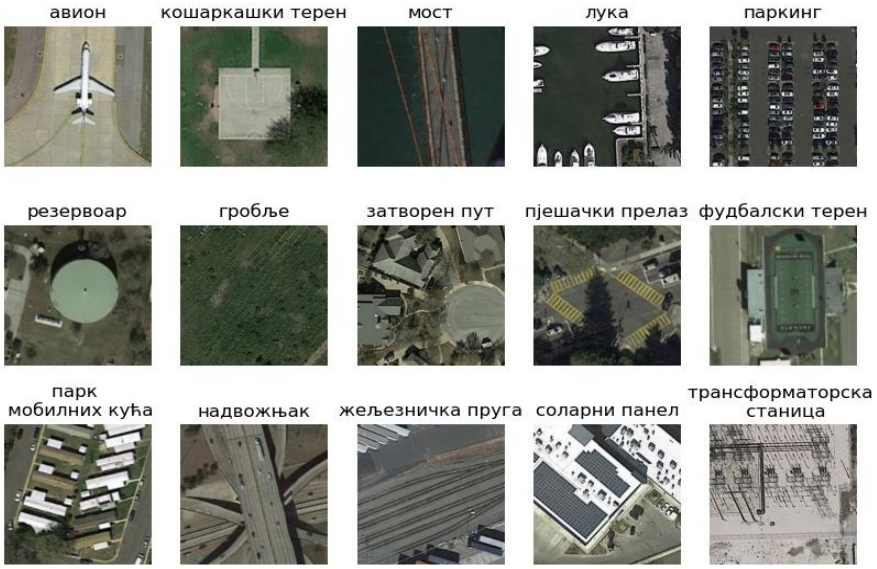
Скуп слика NWPU-RESISC45 [58] се састоји од 31.500 снимака добијених даљинском детекцијом, покривајући на тај начин више од 100 земаља широм свијета. Слике су униформно подијељене у 45 категорија: *авион, аеродром, терен за бејзбол, кошаркашки терен, плажа, мост, жбунаста вегетација, црква, кружно пољопривредно земљиште, облак, пословно подручје, густо насељено подручје, пустиња, шума, аутопут, терен за голф, атлетски стадион, лука, индустријска област, раскрсница, острво, језеро, ливада, средње насељено подручје, парк мобилних кућа, планина, надвожњак, палата, паркинг, жељезничка пруга, жељезничка станица, правоугаоно пољопривредно земљиште, ријека, кружни ток, писта, морски лед, брод, сњежни бријег, ријетко насељено подручје, стадион, резервоар, теniски терен, тераса, термоелектрана и мочварно земљиште.* У свакој од наведених класа се налази 700 слика фиксне величине од 256×256 пиксела. Просторна резолуција једног пиксела слике је између 30 m и 0,2 m, осим за слике из класа *острво, језеро, планина и сњежни бријег*, у којима се могу наћи слике нешто ниже резолуције.

Примјери слика из петнаест насумично одабраних класа из колекције NWPU-RESISC45 су дати на Слици 4.3.

Слика 4.3: Примјери слика из колекције NWPU-RESISC45.

4.3. Колекција слика PatternNet

Скуп слика PatternNet се први пут спомиње у раду [59], чији назив је инспирисан пројектом TerraPattern [69], јавно доступним алатом за откривање облика у неозначеној колекцији сателитских снимака. Скуп је подијељен у 38 категорија: *авион, терен за бејзбол, кошаркашки терен, плажа, мост, гробље, жбунаста вегетација, фарма божићних јелки, затворен пут, приморска вила, пјешачки прелаз, густо насељено подручје, трајектни терминал, фудбалски терен, шума, аутопут, терен за голф, лука, раскрсница, парк мобилних кућа, старачки дом, поље нафтног гаса, нафтна бушотина, паркинг, паркинг простор, жељезничка пруга, ријека, писта, означена писта, бродоградилиште, соларни панел, ријетко насељено подручје, резервоар, базен за пливање, тениски терен, трансформаторска станица и постројења за пречишћавање отпадних вода.*



Слика 4.4: Примјери слика из колекције PatternNet.

Свака класа броји 800 слика фиксне величине од 256×256 пиксела. За разлику од MLRSNet и NWPU-RESISC45 колекција слика, слике из PatternNet колекције карактерише значајно већа просторна резолуција. Највиша уочена просторна резолуција по пикселу је 0,062 m, док је најнижа 4,693 m. Неки примјери слика из PatternNet скупа су дати на Слици 4.4.

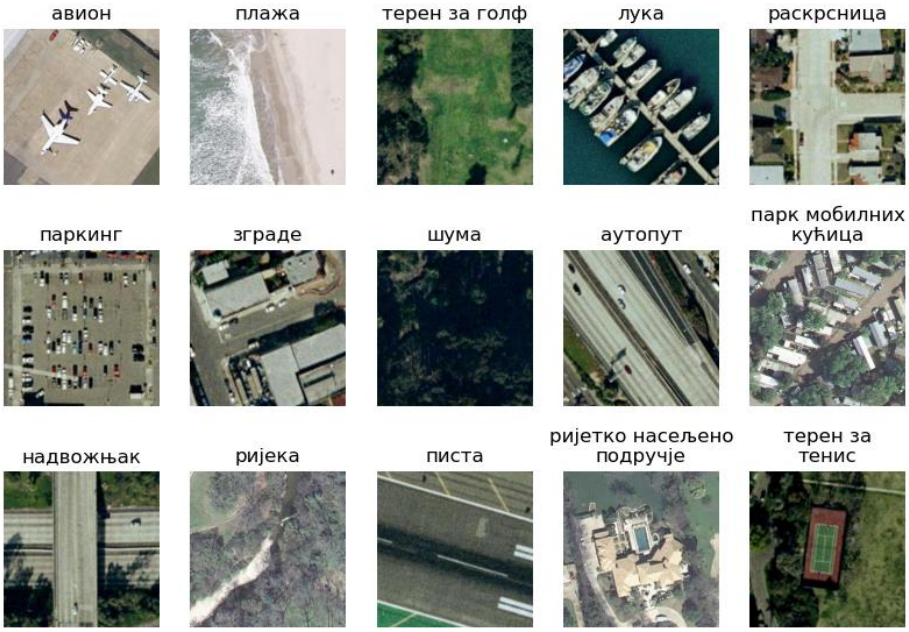
4.4. Колекција слика AID

Слике AID колекције [60] су подијељене у 30 класа: *аеродром, огољено земљиште, терен за бејзбол, плажа, мост, центар, црква, пословно подручје, густо насељено подручје, пустиња, пољопривредно земљиште, шума, индустријско подручје, ливада, средње густо насељено подручје, планина, парк, паркинг, игралиште, бара, лука, жељезничка станица, одмаралиште, ријека, школа, ријетко насељено подручје, квадрат, стадион, резервоари и вијадукт.* Расподјела слика није униформна, а број слика по класи варира између 220 и 420. Величина једне слике је 600×600 пиксела, а просторна резолуција се креће између 0,5 m и 8 m по пикселу. Примјери слика из AID су дати на Слици 4.5.

Слика 4.5: Примјери слика из клекције AID.

4.5. Колекција слика UC Merced Land Use

UC Merced Land Use колекција слика, чији су примјери дати на Слици 4.6, се састоји од 21.000 униформно расподијељених слика у 21 класу: *пољопривредно земљиште, авион, терен за бејзбол, плажа, зграде, жбунаста вегетација, густо насељено подручје, шума, аутопут, терен за голф, лука, раскрсница, средње насељено подручје, парк мобилних кућица, надвожњак, паркинг, ријека, писта, ријетко насељена подручја, резервоари за складиштење и терен за тенис.* Све слике из колекције су величине 256×256 пиксела, просторне резолуције од 0,3 m.

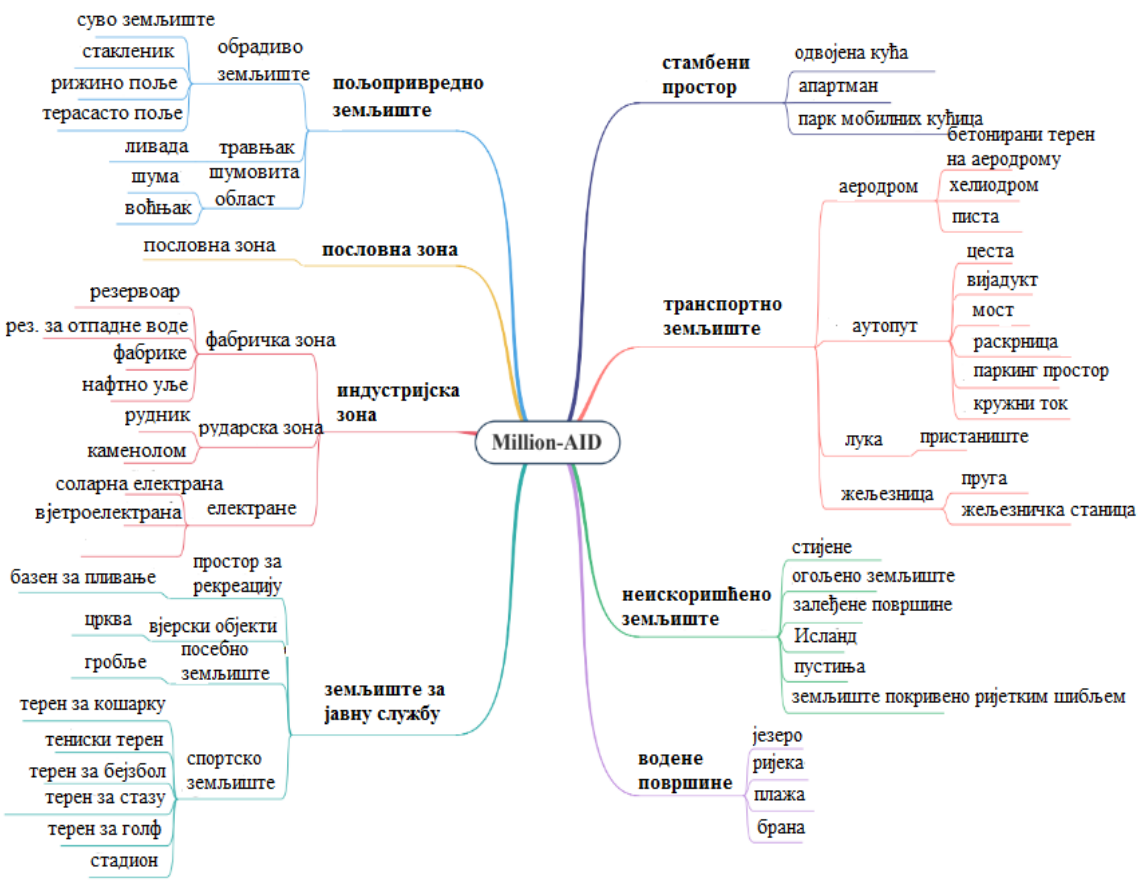


Слика 4.6: Примјери слика из колекције UC Merced Land Use.

4.6. Колекција слика MillionAID

Како би алгоритми за рад са сликама могли напредовати и пратити константно повећање обима и резолуције слика добијених даљинском детекцијом, потребно је имати задовољавајуће колекције слика за обучавање модела и тестирање алгоритама. На путу ка креирању новог скупа слика добијених даљинском детекцијом треба се водити принципима разноликости, богатства и скалабилности [62]. Разноликост подразумјева да у колекцији постоје слике на којима исти објекти показују различите карактеристике. На тај начин се отвара могућност да модел препозна што више могућих особина неког објекта и да се оствари генерализација. Богатство колекција се остварује прикупљањем слика под различитим условима, другачијим сензорима, из разних подручја, са различитим освјетљењем, просторном резолуцијом, у разним годишњим добима и слично. Скалабилност представља могућност за накнадно проширење конструисаног скупа, стога је потребно оставити простора за додавање нових семантичких класа по потреби.

Поштујући претходно наведена правила, 2021. године је конструисан до тада најопширнији скуп слика добијених даљинском детекцијом – Million Aerial Image Dataset (Million AID). Као што сам назив каже, новопредложени скуп се састоји од милион слика организованих у 51 класу, које се додатно могу сврстати у 28 родитељских класа, а оне у 8 главних категорија: *пољопривредно земљиште, пословно подручје, индустријско подручје, земљиште јавне службе, стамбено земљиште, транспортно земљиште, неупотребљиво земљиште и водене површине*. Хијерархија Million AID колекције је приказана на Слици 4.7, која представља модификовану верзију слике из [62].



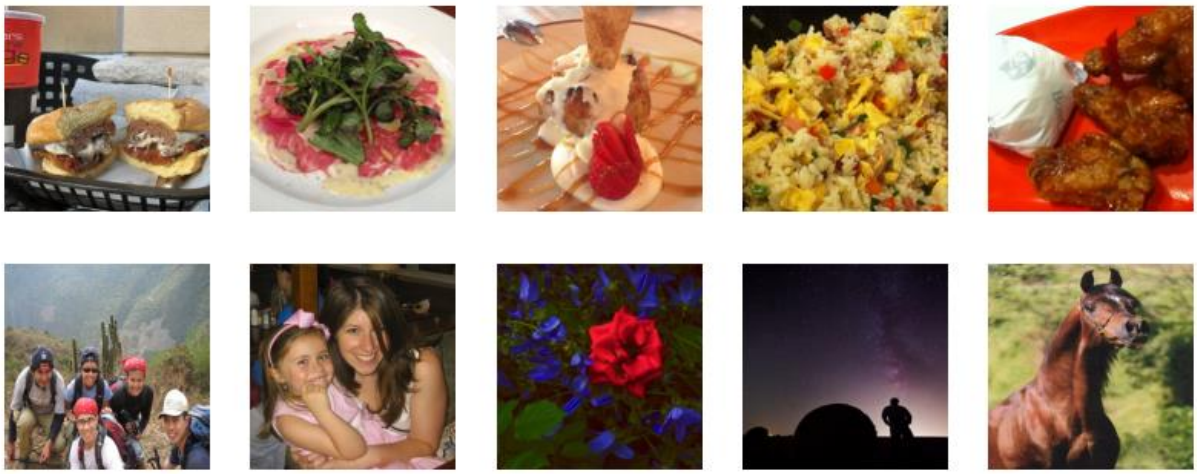
Слика 4.7: Хијерархија слика из Million AID колекције.

Слика 4.8: Примјери слика из колекције Million AID.

Расподјела слика је неуниформна и њихов број по једној класи варира између 2.000 и 45.000. Величина слика је 512×512 пиксела, а просторна резолуција се креће између 0,5 m и 153 m.

4.7. Колекција слика Food5K

Колекција Food5K је кориштена у сврху храна/није храна бинарне класификације у оквиру рада [63]. Састоји се од 2.500 слика хране и 2.500 слика неког другог семантичог значења, при чему је сваки тај скуп додатно подијељен на тренинг од 1.500, валидациони од 500 и тестни подскуп од 500 слика. На Слици 4.9 су приказани примјери слика из Food5K колекције, по пет из сваке од двије категорије којој припадају.



Слика 4.9: Примјери слика хране (први ред) и слика другачијег значења (други ред) из колекције Food5K.

4.8. Колекција слика ImageNet100

ImageNet представља пројекат у оквиру којег је прикупљена велика база слика креирана у сврху развоја алгоритама за препознавање објеката. Колекција ImageNet100 се састоји од 100 насумично одабраних класа из скупа ImageNet1k [64]. Овакав подскуп је додатно подијељен на тренинг и валидациони скуп, који редом садрже 1.300 и 50 слика по класи. За потребе овог рада је искоришћен само валидациони подскуп, дакле укупно 5.000 слика. Примјери слика из ImageNet100 колекције су дати на Слици 4.10.

4.9. Колекција слика Imagenette

И Imagenette [65] је подскуп ImageNet колекције, при чему се он састоји од 10 класа, мање изазовних за класификацију: *лињак, енглески спрингер, касетофон, моторна тестера, црква, хорна, камион за отпад, бензинска пумпа, лоптица за голф* и *падобран*. За потребе тестирања OOD детектора у раду користи се само валидациони дио, тј. 3.925 слика. По један примјер из сваке класе Imagenette колекције слика је дат на Слици 4.11.

Слика 4.10: Примјери слика из ImageNet100 колекције.

Слика 4.11: Примјери слика из Imagenette колекције.

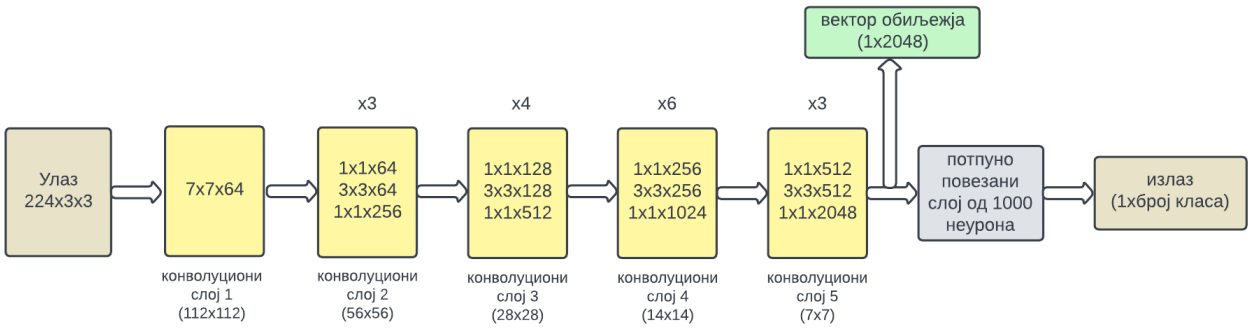
4.10. Конволуциона неуронска мрежа ResNet50

Додавање веза које прескачу слојеве у конволуционим неуронским мрежама резултује настанком нове, ResNet архитектуре [66]. Први предложени модел нове архитектуре је 34-слојни ResNet34 са 3,6 милијарди *floating point* операција у секунди (енг. *floating point operations per second* - FLOPs), што представља само 18% броја параметара VGG-19 модела, до тад најпопуларније VGG архитектуре [70]. У оквиру [66] је показано да 34-слојни ResNet34 и 152-слојни ResNet152 на ImageNet скупу слика дају за око 1% и 4% бољи *top-5 error* резултат, респективно, у односу на VGG архитектуру. Мјера *top-5 error* представља процентуални удио броја тестних узорака за које се ниједна од највећих пет добијених софтмакс вјероватноћа не односи на вјероватноћу оне класе којој узорак припада. Из ових података није тешко извести закључак о предности новопредложене ResNet архитектуре, у смислу временских и техничких ресурса потребних за обучавање модела, али и перформанси класификације.

ResNet50 архитектура садржи сљедеће елементе:

- Конволуциони слој са 64 кернела димензије 7×7, са кораком 2,
- Слој за агрегацију, кернела димензије 3×3 и корака 2,
- По 3 конволуциона слоја са 64 кернела димензија 1×1, са 64 кернела димензије 3×3 и са 256 кернела димензије 1×1,
- По 4 конволуциона слоја са 128 кернела димензије 1×1, са 128 кернела димензије 3×3, са 512 кернела 1×1,
- По 6 конволуционих слојева са 256 кернела димензије 1×1, 256 кернела димензије 3×3 и 1.024 кернела димензије 1×1,
- По 3 конволуциона слоја са 512 кернела димензије 1×1, 512 кернела димензије 3×3 и 2.048 кернела димензије 1×1,
- Слој за агреацију праћен потпуно повезаним слојем од 1.000 неурона, након чега слиједи примјена софтмакс активационе функције.

Шематски приказ архитектуре конволуционог ResNet50 модела дат је на Слици 4.12.



Слика 4.12: Шематски приказ архитектуре ResNet50 модела.

Обиљежја слика се издвајају помоћу слоја који претходи потпуно повезаном слоју непосредно прије примјене софтмакс активационе функције, те је димензионалност вектора обиљежја 2.048.

4.11. Трансформатори за рачунарски вид

Конкретне трансформаторске архитектуре које су у раду коришћене за издвајање обиљежја су ViT-B/16 и ViT-L/16, чије су карактеристике још раније наведене у оквиру Табеле 2.1. Модел ViT-B/16 садржи 12 скривених слојева и број паралелних извршавања MSA блока је 12, а ViT-L/16 садржи 24 скривена слоја и број паралелних извршавања MSA блока је 16. Димензионалност скривеног слоја је 768 у првом случају, а у другом је 1.024. Обје архитектуре улазну слику дијеле на квадратне блокове величине 16x16, а претрениране су на ImageNet21k скупу слика величине 224×224 пиксела. Према наведеним подацима, обје трансформаторске архитектуре дијеле слику на 196 квадратних блокова. Узевши у обзир и додатни [CLS] токен, на излазу сваког скривеног слоја се одређује 197 вектора дужине 768 или 1.024, у зависности од коришћене архитектуре. Вектор који се добија полазећи од [CLS] токена представља репрезентацију цијеле слике, док се осталих 196 односе на по један квадратни блок. У оба случаја се као вектор обиљежја користи један вектор добијен на излазу посљедњег скривеног слоја, и то онај који представља трансформисани [CLS] токен.

4.12. Методологија тестирања

Базни скуп и скуп од којег се полази у експерименталном дијелу је MLRSNet. Основни класификатор је обучаван на дијелу MLRSNet скупа података, те је за почетак било неопходно подијелити га на ID и OOD дио. Додатно је ID дио подијељен на тренинг, валидациони и тестни скуп. Ради могућности пређења са литературом, подјела је извршена на начин описан у [18], како слиједи. Скуп је најприје подијељен на два дисјунктна подскупа, при чему су у први сврстане све класе са прилично јасним значењем, док су у други подскуп сврстане оне класе чије је значење двосмислено. Подјела је заснована на чистој интуицији и дефинитивно није једини начин, али је сасвим разуман постуак који служи у сврху симулације задатка детекције OOD узорака. Класе са двосмисленим значењем се сматрају OOD класама, при чему не постоји оправдан разлог да ће њихова детекција бити тежак задатак, те се из тог разлога тај дио скупа означава као MLRSNet-Holdout. Први подскуп, подскуп у којем су слике са јасним значењем, је додатно подијељен на два дијела. Унутар њега су уочени парови класа са јаким међусобним везама [18], те је по једна класа из сваког таквог пара сматрана OOD класом. На тај начин је формиран MLRSNet-Hard скуп, OOD скуп за који се очекује да ће детектору представљати нешто озбиљнији изазов. Скуп класа са јасним значењем, умањен за по једну елиминисану класу из сваког претходно поменутог пара, представља ID дио. Коначно, формирана 3 подскупа укључују сљедеће класе:

- MLRSNet-ID: *авион, аеродром, терен за бејзбол, кошаркашки терен, плажа, мост, облак, терен за голф, лука, раскрсница, острво, језеро, планина, паркинг, кружни ток, бродоградилиште, сњежни бријег, стадион, резервоар, тераса, далеководни торањ, стакленик за повртње;*
- MLRSNet-Hard: *жбунаста вегетација, пустиња, атлетски стадион, базен, тениски терен, вјетротурбина;*
- MLRSNet-Holdout: *голо земљиште, пословни простор, густо насељено подручје, еродирано пољопривредно земљиште, пољопривредно земљиште, шума, аутопут, индустријска област, ливада, парк мобилних кућа, надвожњак, парк, коловоз у парку, жељезничка пруга, жељезничка станица, ријека, ријетко насељено подручје, мочварно земљиште.*

Из MLRSNet-ID скупа је 42.254 слика одвојено за тренинг, 5.276 слика за валидациони и 5.293 слике за тестни скуп. MLRSNet-Hard броји 14.094, а MLRSNet-Holdout 42.244 слика. Тренинг скуп је искориштен за фино подешавање параметара претходно обучених модела наведених у секцији 4.2. Тако модификовани модели су коришћени за издвајање високодимензионалних вектора обиљежја слика.

Осим на MLRSNet скупу, посматрани методи су тестирани и на другим скуповима слика добијеним даљинском детекцијом (NWPU-RESISC45, PatternNet) и скуповима слика из других домена (Food5K, Imagenette). Јасно је да скупови података који нису из домена даљинске детекције у цјелини представљају OOD узорке, док је скупове слика добијених даљинском детекцијом непосредно прије тестирања неопходно дијелити на ID и OOD дио, у складу са већ утврђеном подјелом MLRSNet скупа. Класе из појединих RS скупова чије слике одступају од расподјеле тренинг података и које су у експерименталном дијелу сматране OOD подацима су:

- NWPU-RESISC45: *жбунаста вегетација, црква, кружно пољопривредно земљиште, пословна зона, густо насељено подручје, пустиња, шума, аутопут, терен за стазу, индустријска зона, ливада, средње насељено подручје, парк мобилних кућица, надвожњак, палата, жељезничка пруга, жељезничка станица, правоугаоно пољопривредно земљиште, ријека, писта, морски лед, брод, ријетко насељено подручје, тениски терен, термоелектрана и мочварно земљиште.*
- PatternNet: *гробље, жбунаста вегетација, фарма божићних јелки, затворен пут, приморска вила, пјешачки прелаз, густо насељено подручје, трајектни терминал, фудбалски терен, шума, аутопут, парк мобилних кућица, старачки дом, нафтно поље, извор нафте, надвожњак, паркинг простор, жељезница, ријека, писта, означена писта, соларсни панел, ријетко насељено подручје, базен за пливање, тениски терен, трансформаторска станица и постројења за пречишћавање отпадних вода.*
- UC Merced Land Use: *пољопривредно земљиште, зграде, жбунаста вегетација, густо насељено подручје, шума, аутопут, средње насељено подручје, парк мобилних кућа, надвожњак, ријека, писта, ријетко насељено подручје и тениски терен.*
- AID: *голо земљиште, центар, црква, пословна зона, густо насељено подручје, пустиња, пољопривредно земљиште, шума, индустријска зона, ливада, средње насељено подручје, парк, жељезничка страница, одмаралиште, ријека, школа, ријетко насељено подручје,квадрат и вијадукт.*
- MillionAID: *апартман, голо земљиште, гробље, црква, пословна зона, брана, пустиња, одвојена кућа, суво поље, шума, стаклена башта, терен за стазу, хелиодром, Исланд, ливада, рудник, парк мобилних кућа, нафтно поље, воћњак, рижино поље, пристаниште, каменолом, жељезница, ријека, пут, стијене, писта, соларна електрана, земљиште пкривено ријетким шибљем, подцентрала, базен за пливање, тениски терен, жељезничка станица, вијадукт, резервоар за складиштење отпадних вода, вјетротурбина и фабрике.*

Величине тестних скупова су дате у описима поменутих колекција слика, у оквиру четвртог поглавља. Да би се испитала зависност успјеха метода од броја тренинг узорака, сви експерименти су урађени и полазећи од неколико мањих, случајно издвојених, подскупова их полазног тренинг скупа. За стицање независности резултата од конкретног избора таквог подскупа, фино подешавање и тестирање је вршено више пута са различитим, случајно изабраним, подскуповима полазног тренинг скупа. У том случају се, за неку релативно малу величину тренинг скупа и конкретан метод, резултат наводи помоћу средње вриједности и стандардне девијације.

На успјех метода детекције OOD узорака директно утиче начин на који се издвајају вектори обиљежја из узорака. Из тог разлога, у експерименталном дијелу су се у улози основног класификатора нашле три различите архитектуре, једна конволуциона и двије

трансформаторске. За сваку изабрани архитектуру класификатора, поређења ради, испитане су перформансе свих метода наведених у одјељку 3.1.

На крају се детектор излаже подацима који одступају од расподјеле ID скупа. У ту сврху је искоришћено неколико различитих скупова, како би се стекао утисак о осјетљивости метода детекције на природу колекције слика из којег се узимају OOD подаци. Детектор је тестиран на различитим OOD скуповима и додатно је испитан и утицај величине тренинг скупа. Када је модел обучаван на неком мањем дијелу тренинг скупа, обавезно је експеримент поновљен више пута из већ раније објашњеног разлога. Пажња је посвећена и начину на који се узимају узорци који оступају од расподјеле, те је у оквиру сваког експеримента испитано више политика бирања узорака.

За оцјену перформанси детектора коришћене су двије метрике: (1) FPR (*false positive rate* кад је *true positive rate* 95%) и (2) AUROC. За рачунање ове двије вриједности неопходно је познавати четири мјере: (1) *true positive* (TP) - број ID узорака класификованих као ID, (2) *true negative* (TN) – број OOD узорака класификованих као OOD, (3) *false positive* (FP) – број OOD узорака класификованих као ID, (4) *false negative* (FN) – број ID узорака класификованих као OOD. Вриједност FPR заправо представља грешку коју детектор прави на OOD скупу података, FPR = FP / (FP + TN). Резултат зависи од изабраног прага, а FPR се рачуна за онај праг када TPR = TP / (TP + FN) износи 95%, тј. кад је 95% ID узорака тачно класификовано. Са друге стране, ROC крива (енг. *receiver operating characteristic curve*) приказује парове TPR и FPR вриједности при различитим вриједностима прага. Пожељније је имати што мању FPR и што већу AUROC вриједност, јер је то добар показатељ успјешности и ефективности детектора.

За мјерење удаљености у простору обиљежја између два узорка су, у оквиру овог рада, коришћене двије метрике: косинусна удаљеност и Еуклидова удаљеност. Ако су $\boldsymbol{z_i}$ и $\boldsymbol{z_j}$ репрезентације два узорка у N-димензионом простору обиљежја, косинусна и Еуклидова удаљеност се рачунају према изразима (4.1-4.2).

$$d_{cosine}(\boldsymbol{z_i}, \boldsymbol{z_j}) = 1 - \frac{\boldsymbol{z_i} \cdot \boldsymbol{z_j}}{\|\boldsymbol{z_i}\| \cdot \|\boldsymbol{z_j}\|} = 1 - \frac{\sum_{k=1}^{N} z_{i,k} \cdot z_{j,k}}{\sqrt{\sum_{k=1}^{N} z_{i,k}^2} \cdot \sqrt{\sum_{k=1}^{N} z_{j,k}^2}}, \tag{4.1}$$

$$d_{euclidean}(\boldsymbol{z_i}, \boldsymbol{z_j}) = \|\boldsymbol{z_i} - \boldsymbol{z_j}\| = \sqrt{\sum_{k=1}^{N} (z_{i,k} - z_{j,k})^2}. \tag{4.2}$$

У оквиру експерименталног дијела су коришћене имплементације косинусне и Еуклидове удаљености из *metrics.pairwise* подмодула библиотеке *sklearn* програмског језика *Python*. У свим тестираним методима базираним на мјерењу удаљености је подразумијевано коришћена косинусна метрика, али су неки од њих имплементирани и рачунањем Еуклидове удаљености, у циљу да се испита зависност успјеха метода од коришћене метрике.

У експерименталном дијелу рада су, за детекцију узорака који одступају од расподјеле, тестирани сљедећи методи:

- MSP - највећа софтмакс вјероватноћа
- KNN - алгоритам *k*–најближих сусједа
- NCM - удаљеност од најближег центроида
- NNDR - однос удаљености од два најближа сусједа

- MD - Махаланобисова удаљеност
- RMD - релативна Махаланобисова удаљеност

5. Експериментални резултати и анализа

Сва три кориштена екстрактора обиљежја (ResNet50, ViT-base-patch16-224, ViT-large-patch16-224) су претрениране неуронске мреже, а додатно је извршено фино подешавање параметара на тренинг дијелу MLRSNet ID скупа.

Конволуциони модел ResNet50 је претрениран на ImageNet скупу слика. Посљедњи слој од 1000 неурона је замијењен слојем од 22 неурона, што одговара броју ID класа. Фино подешавање параметара је извршено на 50 епоха, при чему је кориштен Adam оптимизатор. Корак обучавања је промјенљив, а кориштен је ReduceLROnPlateau планер са подразумијеваним параметрима: величина која се прати је вриједност функције цијене на валидационом скупу; 0,1 је фактор за смањење корака обучавања; 10 је број епоха без побољшања, након којих слиједи смањење корака обучавања; 0 је доња граница за корак обучавања. У циљу да се спријечи претјерано прилагођење модела тренинг скупу, додају се сљедеће аугментације: ротација до 90 степени; зум до 0,15; помјерај по висини и ширини до 0,2; смицање до 0,15 и окретање око хоризонталне и вертикалне осе. Све слике се из RGB колор простора преводе у BGR колор простор, те се средње вриједности свих колор канала своде на 0.

Трансформатори ViT-base-patch16-224 и ViT-large-patch16-224 су претренирани на ImageNet-21k, а извршено је и фино подешавање параметара на ImageNet2012 скупу слика. У овом случају се посљедњи слој од 1000 неурона мијења слојем од 22 неурона, те се тако модел припрема за поступак финог подешавања параметара на MLRSNet ID тренинг скупу. Величина слика се трансформише у 224×224 пиксела и врши се нормализација по RGB каналима са средњом вриједношћу 0 и стандардном девијациојом од 1. На улаз трансформатора се доводи [CLS] токен, помоћу којег се одређује репрезентација цијеле слике. Као вектор обиљежја се користи излаз из посљедњег слоја везаног за [CLS] токен, који у случају трансформатора једноставније архитектуре има димензионалност 768, док је димензионалност у случају сложенијег трансформатора 1.024. Ради превенције појаве превеликог прилагођења модела тренинг скупу, додају се аугментације: окретање око хоризонталне и окретање око вертикалне осе, али се врши и модификација функције губитака помоћу смањења тежина са фактором 0,01. Фино подешавање параметара оба трансформаторска модела се врши на 10 епоха. Распоред мијењања корака обучавања је интегрисан са оптимизатором, при чему се користи техника смањења корака обучавања тако да се његова вриједност смањује почевши од вриједности 0,00003 са фактором 0,01.

5.1. Поређење метода заснованих на мјерењу удаљености

У Табели 5.1 су дате AUROC и FPR вриједности добијене тестирањем метода детекције на четири OOD скупа слика добијена даљинском детекцијом (MLRSNet-Hard, MLRSNet-Holdout, OOD дио NWPU-RESISC45 скупа, OOD дио PatternNet скупа) и два OOD скупа слика који не припадају поменутом домену (Food5K, Imagenette). За потребе експериментисања на NWPU-RESISC45 и PatternNet скуповима и ID и OOD слике долазе из поменутих колекција. У свим осталим тестним сценаријима као ID дио је искоришћен скуп MLRSNet ID. Обиљежја се издвајају помоћу ResNet50 модела, при чему је извршено фино подешавање параметара на тренинг дијелу MLRSNet ID скупа. Тачност класификације на тестном MLRSNet ID скупу је 99,32%.

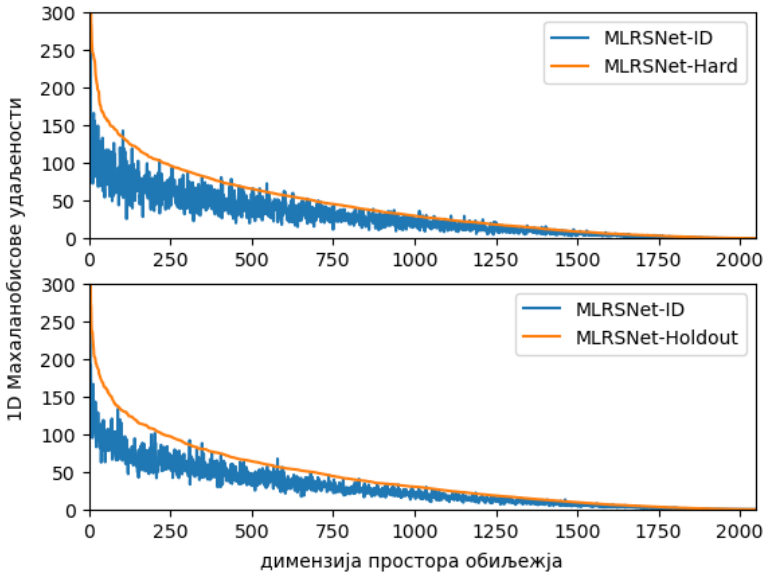
Табела 5.1: Перформансе разматраних метода детекције узорака ван расподјеле на различитим тестним скуповима (нотација = AUROC / FPR при 95% TPR). Тачност класификације на тестном MLRSNet ID скупу је 99,32%.

	MLRSNet-Hard	**MLRSNet-Holdout**	**NWPU-RESISC45**	**PatternNet**	**Food5K**	**Imagenette**
MSP	90,03 / 28,58	96,32 / 11,82	89,42 / 56,55	**88,02** / 53,88	98,31 / 4,96	97,83 / 7,21
KNN	94,01 / 23,51	97,24 / 14,81	91,26 / **49,78**	85,83 / 67,51	**99,81** / **0,44**	99,50 / 1,86
NCM	89,03 / 53,20	95,60 / 28,35	87,85 / 61,57	87,92 / **50,94**	97,08 / 19,76	95,23 / 38,06
NNDR	94,00 / 21,65	96,99 / 12,20	91,52 / 58,20	84,30 / 66,08	99,14 / 1,24	98,59 / 4,33
MD	78,93 / 64,33	78,84 / 73,17	70,84 / 78,96	62,90 / 82,22	98,38 / 6,52	96,23 / 23,44
RMD	**95,62** / **19,93**	**97,63** / **9,86**	**91,59** / 52,73	85,89 / 58,17	99,74 / **0,44**	**99,55** / **1,37**

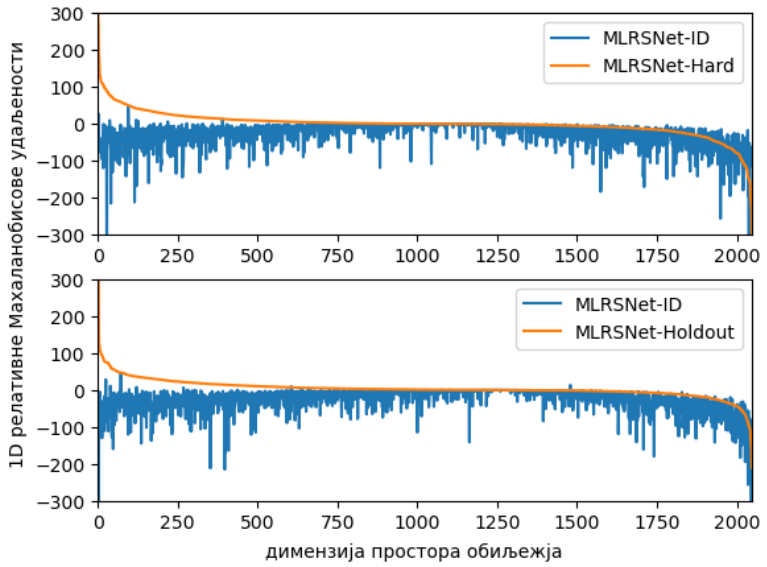
Из Табеле 5.1 се може примјетити да RMD метод даје најбољи резултат на већини испитаних сценарија. Изузетак су експерименти на PatternNet и Food5K скуповима, при чему треба нагласити да AUROC вриједност добијена RMD методом на Food5K заостаје за занемаривих 0,07% у односу на најбољи резултат, добијен KNN методом.

Добијени резултати показују да метод базиран на мјерењу Махаланобисових удаљености на свим испитаним скуповима резултује најмањом AUROC и највећом FPR вриједности, тј. да има најлошије перформансе. График са Слике 5.1 објашњава добијено, јер се може примјетити да се за већински број димензија простора обиљежја компоненте Махаланобисових удаљености OOD узорака не разликују значајно од компоненти Махаланобисових удаљеностти ID узорака. Проблем се рјешава процјеном класно независне распоједјеле ID скупа и одузимањем тако добијених Махаланобисових удаљености од првобитних. На тај начин се изводи RMD метод, за који је већ речено да има најбоље перформансе на већини посматраних скупова. Побољшање се може објаснити поређењем графика са слика 5.1 и 5.2, јер се усљед рачунања релативних Махаланобисових удаљености остварује значајнија разлика између првих 500 једнодимензионих компонената, што доприноси успјешнијем разликовању OOD од ID узорака.

Додатно је уочено да, у односу на резултате који су добијени на MLRSNet-Hard и MLRSNet-Holdout скуповима, скоро сви методи дају лошији резултат на другим скуповима слика добијеним даљинском детекцијом, а бољи резултат на скуповима који нису из тог домена. Наиме, чињеница је да детекција OOD узорака који нису из домена снимака добијених даљинском детекцијом, усљед велике удаљености OOD од ID скупа у простору обиљежја, детектору представља лакши задатак, што објашњава тако добре резултате добијене на Food5K и Imagenette скуповима. Са друге стране, када је ријеч о скуповима слика добијених даљинском детекцијом, удаљеност OOD од ID скупа у простору обиљежја није толико велика. Усљед тога, али и одступања расподјеле скупова NWPU-RESISC45 и PatternNet од расподјеле MLRSNet скупа, детектору је теже да у том случају разликује OOD узорке од ID узорака. Осим тога, основни класификатор је у фази обучавања био изложен узорцима из MLRSNet скупа, па се може рећи да је он као екстрактор обиљежја „наклоњен“ сликама које у тестној фази долазе из исте расподјеле.



Слика 5.1: Компоненте Махаланобисове удаљености.



Слика 5.2: Једнодимензионе компоненте релативне Махаланобисове удаљености.

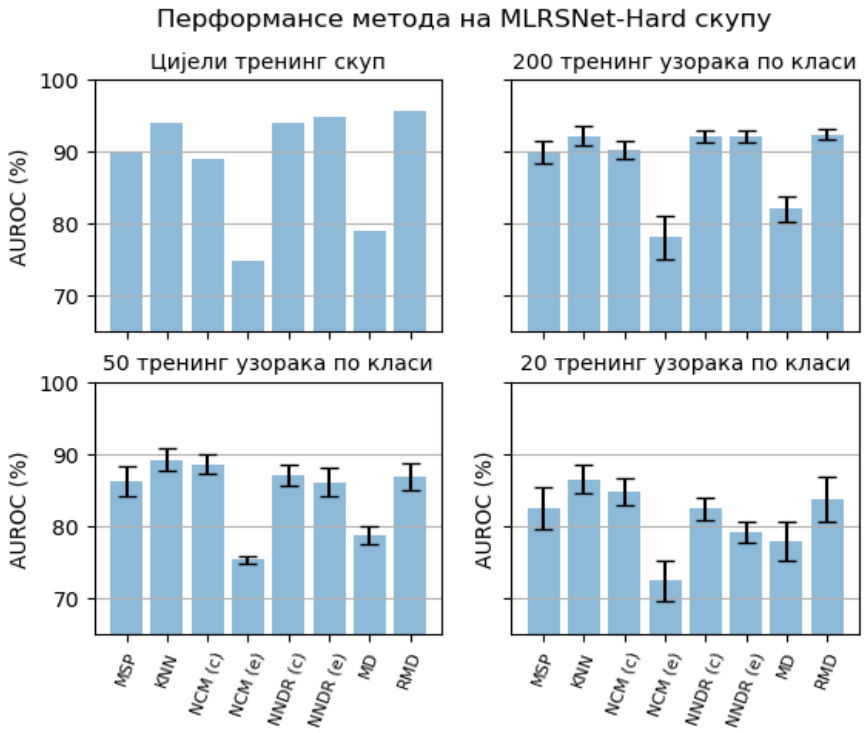
5.2. Зависност перформанси метода детекције од броја тренинг узорака

За потребе испитивања перформанси раличитих метода на MLRSNet-Hard и MLRSNet-Holdout OOD скуповима као основни класификатор је кориштен ResNet50, на којем је извршено фино подешавање параметара помоћу релативно великог тренинг скупа, тачније дијела скупа MLRSNet ID који броји 42.254 слика. Међутим, за неке специфичне примјене у пракси се може десити да не постоји доступна велика колекција слика која би могла бити кориштена у ту сврху, те би било неопходно фино подешавање мреже извршити на знатно мањем тренинг скупу. Усљед недостатка већ сакупљеног, достојног скупа слика, било би потребно креирати нову колекцију за потребе финог подешаваља параметара мреже. Као посљедица разних ограничења (временских, техничких, финансијских...), новокреирани тренинг скуп би потенцијално могао имати мали број узорака.

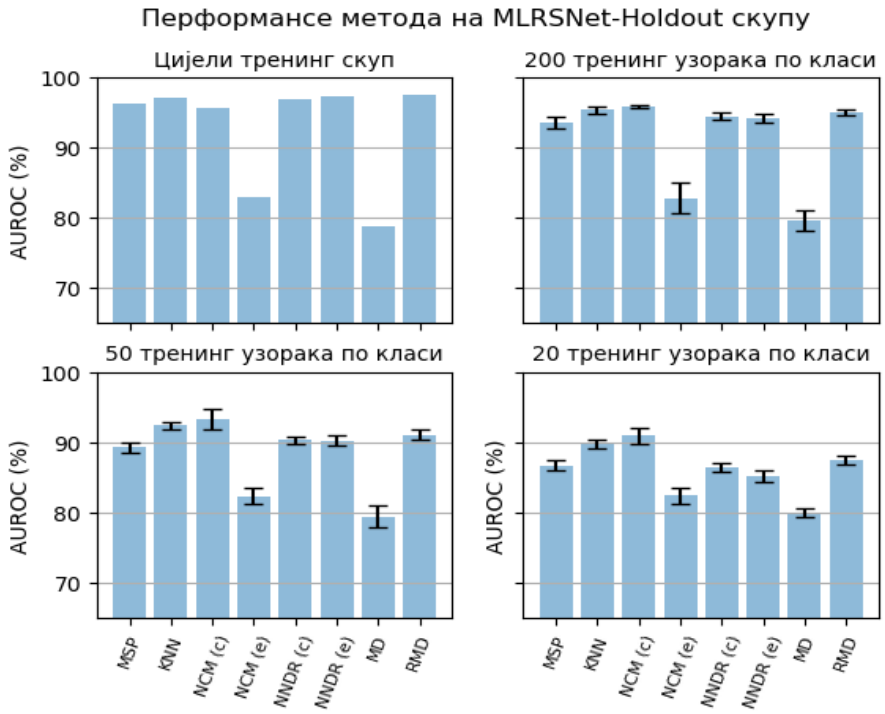
Могућност да се у пракси на располагању има мали тренинг скуп слика представља главни мотив за испитивање перформанси метода и у случају мањих тренинг скупова. Циљ је да се испита како смањење тренинг скупа утиче на перформансе свих метода, те да ли при различитим величинама тренинг скупа исти метод даје најбољи резултат или се поредак остварених резултата мијења са промјеном броја тренинг слика.

Перформансе свих метода су испитане и у случају тренинг скупова који садрже 200, 50 и 20 слика по класи, и то случајно одабраних из полазног тренинг скупа. Како би се остварила независност резултата од специфичног избора мањег тренинг скупа из цијелог полазног, експерименти су поновљени пет пута, за различите насумично одабране тренинг скупове. Резултати су из тог разлога, за поменуте бројеве тренинг узорака, записани преко средњих вриједности и стандардних девијација.

Добијени графици са слика 5.3 и 5.4 показују да, сем MD метода, перформансе свих метода слабе са смањењем броја тренинг слика. Из вриједности из Табеле 5.2 се види да је пад перформанси метода за OOD детекцију у директној корелацији са тачношћу класификације на MLRSNet ID скупу. Методи NCM и NNDR су тестирани користећи двије метрике: косинусну сличност (NCM (c) и NNDR (c) на графицима) и Еуклидову удаљеност (NCM (e) и NNDR (e) на графицима). Ако се пореди резултат добијен за тренинг скуп од 20 узорака по класи и за полазни тренинг скуп, у случају да се користи Еуклидова метрика, нарочито велико смањење AUROC вриједности се може примјетити на NNDR методу, које на MLRSNet-Hard скупу достиже чак 15%. Ова појава се објашњава чињеницом да је за овај метод веома значајно да ID тестни узорак има веома близак ID тренинг узорак. Што је тренинг узорака мање, већа је и шанса да ће тестни ID узорак имати већи однос удаљености између њега и два најближа сусједа из различитих класа, те да ће бити погрешно детектован као OOD. Стога, очекивано је да перформансе овог метода падну нагло са смањењем броја тренинг узорака.



Слика 5.3: Утицај величине тренинг скупа на перформансе метода на MLRSNet-Hard OOD скупу.



Слика 5.4: Утицај величине тренинг скупа на перформансе метода на MLRSNet-Holdout OOD скупу.

Табела 5.2: Тачност класификације ResNet50 класификатора на MLRSNet ID тестном скупу у зависности од величине тренинг скупа.

Број тренинг узорака	**Тачност класификације на MLRSNet ID тестном скупу (%)**
42.254	99,32
4.400 (200 узорака по класи)	97,77
1.100 (50 узорака по класи)	95,28
440 (20 узорака по класи)	89,95

Утицај величине тренинг скупа на перформансе MD метода није значајан, што је изненађујуће кад се узме у обзир да у случају мањих тренинг скупова није могуће прецизно проценити матрицу коваријанси, него је неопходно извршити њену регуларизацију. Очито је већ објашњени узрок лошег резултата чак и у случају употребе пуног тренинг скупа „поништио“ поменути проблем. Да процијењена матрица коваријанси није довољно добра када се има мало тренинг узорака, најбоље показује поређење резултата које дају NCM метод са Еуклидовом метриком и MD метод. Дакле, за мали број тренинг узорака, детектор на MLRSNet-Holdout ради боље када за удаљеност од центроида користи јединичну матрицу, умјесто процијењене коваријансне. Такође, посљедице регуларизације матрице се виде и на RMD методу, чији се разултати значајно кваре са падом броја тренинг узорака.

У случају најмањег испитаног тренинг скупа, од 20 слика по класи, најбољи резултат дају методи KNN и NCM, и то користећи косинусну сличност као метрику. На основу тога се може закључити да су поменута два метода најробуснија, те би били препоручљиви у ситуацијама у којима се нема на располагању велики тренинг скуп. Наиме, како су за NCM метод кључне и једино битне позиције центроида класа, јасно је да овај метод није осјетљив на број тренинг узорака. Често се центроиди не помјере много и кад се из пуног тренинг скупа случајним избором издвоји нпр. 200 узорака по класи. Слична ситуација је и са KNN методом, који такође нема велику осјетљивост на број узорака.

5.3. Утицај архитектуре основног класификатора на перформансе детектора

За потребе демонстрације утицаја архитектуре основног класификатора на перформансе метода, сви методи су тестирани на OOD скуповима MLRSNet-Hard и MLRSNet-Holdout, користећи три различита модела за издвајање обиљежја. Као основни класификатори су искориштени један конволуциони модел ResNet50 и два трансформатора за рачунарски вид: ViT-base-patch16-224 и ViT-large-patch16-224. На сликама 5.5-5.6 су дате AUROC вриједности добијене на MLRSNet-Hard и MLRSNet-Holdout OOD скуповима.

У Табели 5.3 су наведене тачности класификатора на MLRSNet ID тестном скупу. У поређењу са перформансама ResNet50 основног класификатора, нешто већа тачност класификације се постиже употребом сложенијег трансформатора. Са друге стране, кориштени трансформатор једноставније архитектуре, са мањим бројем слојева и параметара, даје за скоро 6% мању тачност класификације на MLRSNet ID тестном скупу.

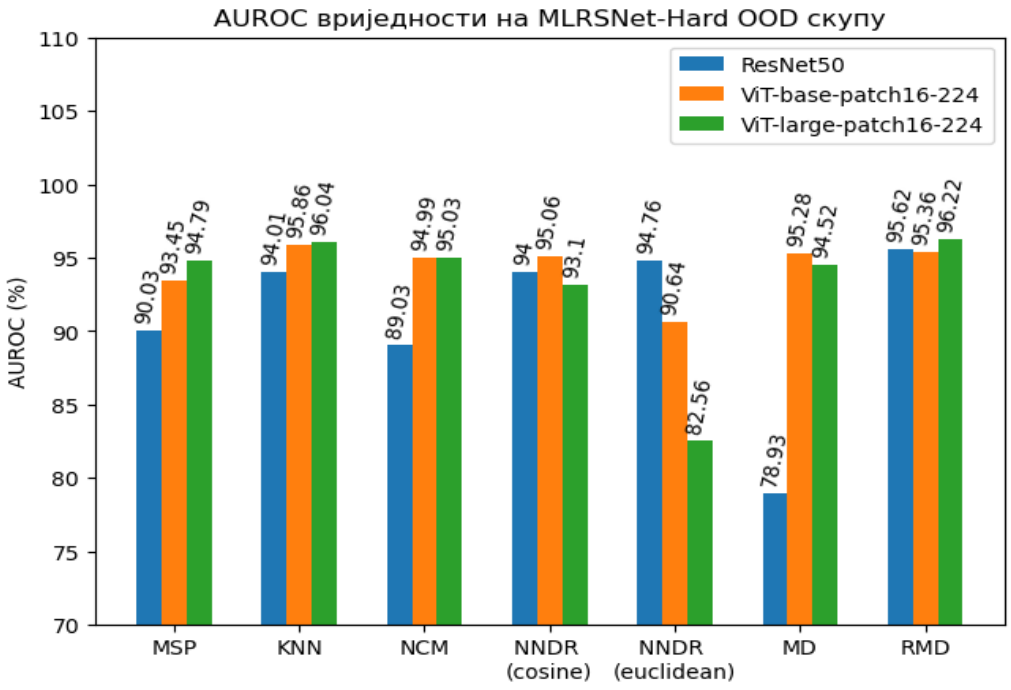
Са приложених графика се може примјетити да употреба трансформатора као екстрактора обиљежја у већини извршених експеримената доноси побољшање у односу на случај када се обиљежја издвајају конволуционом мрежом. Дакле, употреба модела ViT-base-patch16-224, упркос незавидној тачности класификације на MLRSNet ID тестном скупу, на већини испитаних метода за OOD детекцију резултује бољим перформансама у односу на ResNet50. Побољшање перформанси употребом трансформатора се види на методима KNN, RMD и NNDR у случају косинусне сличности као метрике, али не прелази 2% и самим тим побољшање није изражено као на методима MSP и NCM, гдје пораст AUROC прелази и 5%. Међутим, побољшање које доноси замјена конволуцуоне архитектуре класификатора трансформаторском најбоље се види на MD методу, јер AUROC вриједности и на MLRSNet-Hard и MLRSNet-Holdout скуповима порасту за приближно 20%. Узрок томе се види са графика на Слици 5.7, који приказују компоненте Махаланобисове удаљености скупова MLRSNet-ID, MLRSNet-Hard и MLRSNet-Holdout при издвајању обиљежја ResNet50 и ViT-B/16 архитектурама. Како се значајно веће разлике у једнодимензионим Махаланобисовим удаљеностима ID и OOD скупова уочавају уколико се као екстрактор користи трансформатор, може се извести закључак да је трансформаторска архитектура погоднија за примјену MD метода.

Ипак, ако се у оквиру NNDR метода користи Еуклидова удаљеност као метрика, најбољи резултат се добија кад се за издвајање обиљежја користи ResNet50. На основу уоченог се може извести закључак да начин на који трансформаторски модел издваја обиљежја није погодан за наведену специфичну комбинацију метрике и метода.

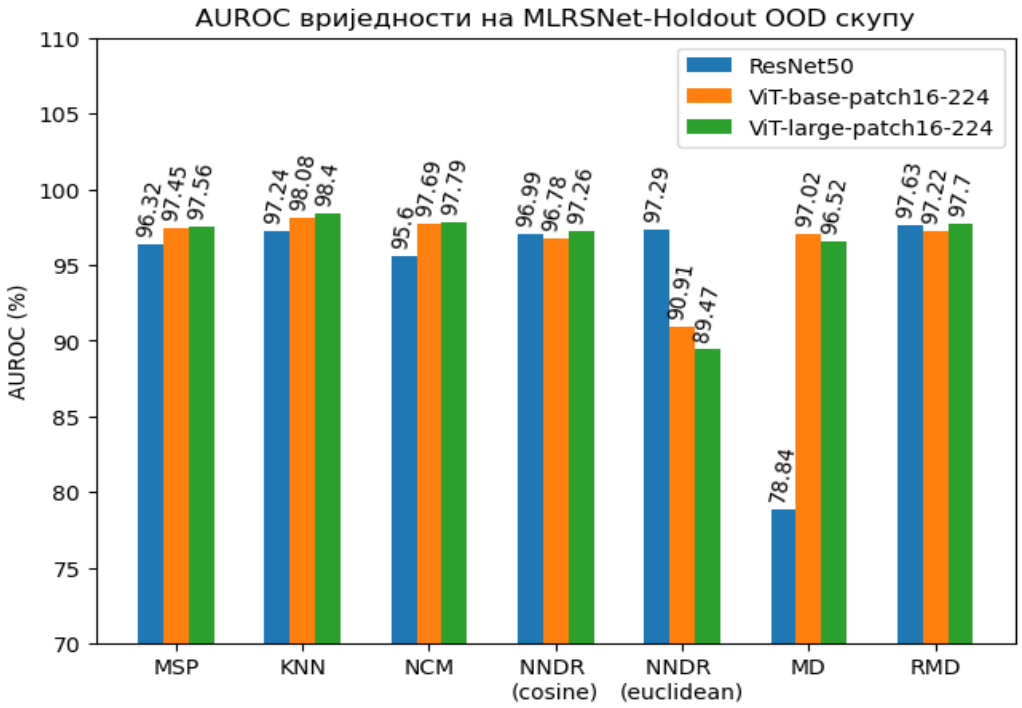
Табела 5.3: Тачности класификације свих посматраних модела основног класификатора на MLRSNet ID тестном скупу.

Модел основног класификатора	**Тачност класификације на MLRSNet ID тестном скупу (%)**
ResNet50	99,32
ViT-base-patch16-224	93,50
ViT-large-patch16-224	99,58

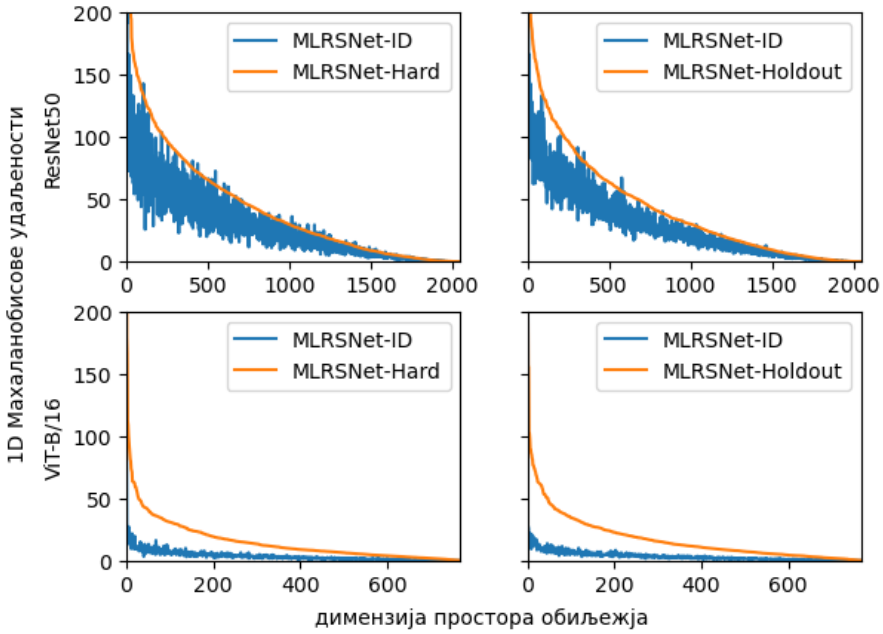
Додатно, усљед минималних разлика у резултатима који су добијени у случају употребе два различита трансформатора, може се закључити да сложенија архитектура не доноси значајну разлику. Ако се користи једноставнији трансформатор, MD и NNDR са Еуклидовом метриком дају боље перформансе на оба посматрана скупа, а NNDR са косинусном метриком на MLRSNet-Hard скупу. Наведено показује да сложенија архитектура не значи нужно и бољи резултат. Такође, није уочена ни директна корелација између тачности класификације двије трансформаторске архитектуре и перформанси OOD детектора који поменуте архитектуре користе као екстрактор обиљежја. Ипак, у већини тестних сценарија сложенији трансформатор доноси побољшање, али је максимални уочени инкремент само 1,34%.



Слика 5.5: Утицај архитектуре класификатора на перформансе метода на MLRSNet-Hard скупу.



Слика 5.6: Утицај архитектуре класификатора на перформансе метода на MLRSNet-Holdout скупу.



Слика 5.7: Једнодимензионе Махаланобисове удаљености у случају издвајања обиљежја помоћу ResNet50 и ViT/B-16 архитектуре.

5.4. Излагање детектора узорцима који одступају од расподјеле

У циљу побољшања резултата детекције, детектор ће бити изложен узорцима који одступају од расподјеле тренинг скупа, те ће они бити искориштени у оквиру метода предложеног у одјељку 3.2. Узорци ван расподјеле који се бирају треба да на што бољи начин моделују OOD узорке који се могу јавити у тестној фази. Притом, треба имати на уму да се у тренутку бирања не располаже било каквим информацијама о тестним узорцима, сем једном - да су слике које долазе на улаз детектора углавном из домена слика добијених даљинском детекцијом. Стога, узорке ван расподјеле би било погодно бирати из неког великог и разноликог скупа слика из домена даљинске детекције, при чему би претходно требало елиминисати оне слике које припадају ID класама. У овом раду се у највећем броју експеримената узорци који одступају од расподјеле бирају из подскупа MillionAID колекције, иако се додатно, у сврху испитивања утицаја скупа узорци бирају и из других колекција слика добијених даљинском детекцијом. Поменути подскуп MillionAID је формиран насумичним бирањем слика из комплетне MillionAID колекције [71]. Састоји се од 10.000 слика, неуниформно подијељених у 51 класу. Након елиминације класа које припадају ID расподјели, скуп садржи 6.997 слика, неуниформно подијељених у 37 класа.

5.4.1. Политике бирања

Ради могућности поређења са литературом, али и испитивања осјетљивости предложеног метода на број изабраних узорака ван расподјеле, одређују се перформансе детектора при разлитим бројевима изабраних узорака. Експерименти ће бити вршени када се на располагању има 88, 132, 264, 506, 1.012, 2.002 и 4.004 узорака који одступају од расподјеле. Сви наведени бројеви су цјелобројни умножак броја 22, тј. броја ID класа, као што је урађено у раду [18]. Намеће се питање: На који начин је оптимално бирати изорке који одступају од расподјеле? До одговора се долази водећи се сличном логиком као и у [18], али додатно наклоњеној природи предложеног метода у овом раду:

- Потребно је да изабрани узорци ван расподјеле „покрију“ што већи дио простора обиљежја, како би била већа шанса да ће се наћи у близини што више тестних OOD узорака.
- Узорке са великим односом удаљености до два најближа сусједа из различитих класа тренинг скупа (израз (3.3), у наставку текста само однос удаљености) је свакако лако детектовати, те би узорци који имају мали поменути однос били бољи модели тестних OOD узорака који представљају већи изазов детектору, тј. оних који су у простору обиљежја ближе ID узорцима.
- Изабрани узорак са малим односом удаљености се потенцијално може наћи у близини тестног ID узорка и на тај начин узроковати да детектор ID узорак погрешно означи као OOD. Мање је вјероватна могућност да ће се описани сценарији догодити уколико се бирају узорци са великим односом.

Ако се погледају претходно наведене ставке, може се примјетити да се друга и трећа противе једна другој, те да у избору политике бирања није једноставно испоштовати обје. Стога, у овом раду се предлаже неколико политика бирања, а из експерименталних резултата ће се закључити која од њих је најповољнија.

Комбиновање описаног у раду [18] са претходно наведеним ставкама резултује сљедећим политикама бирања, названим енглеским терминима у складу са онима у [18]:

- *different*: Тежи се ка томе да изабрани узорци буду што равномјерније распоређени по простору обиљежја, при чему се позиција узорка процјењује на основу његовог положаја у односу на положаје тренинг узорака. Дакле, циљ је да се око сваке ID класе нађе приближно једнак број изабраних узорака

који одступају од расподјеле. При томе, закључак да је неком узорку најближа једна ID класа може се извести користећи добијене вјероватноће на излазу класификатора (MSP), метод *k*-најближих сусједа (KNN) и слично.

- *hard*: Као што сам назив каже, из колекције слика се бирају узорци који су најтежи за детекцију са становишта NNDR матода, тј. они са најмањим односом удаљености.
- *easy*: Контрадикторно претходној политици, бирају се узорци најлакши за детекцију, тј. они са највећим односом удаљености.

Међутим, за специфичан избор неке колекције слика из које се бирају узорци који одступају од расподјеле, може се десити да постоји класа такве природе да је у простору обиљежја веома близу неке ID класе или веома далеко од свих ID класа. У том случају би већина слика из тих класа имала изразито мале или велике односе удаљености респективно. Примјене *hard* и *easy* политика би резултовале тим да већина изабраних узорака ван расподјеле буде из исте класе, што представља веома неповољну ситуацију у погледу моделовања тестних OOD узорака. Како би се заобишли потенцијални проблеми, у овом раду се испитују додатно и још двије хибридне политике бирања:

- *hard + different*: Осим што се бирају узорци са малим односима удаљености, води се рачуна о томе да узорци буду равномјерно распоређени у простору обиљежја.
- *easy + different*: Аналогно претходном, бирају се узорци са највећим односима удаљености, али тако да они буду што равномјерније концентрисани око тренинг узорака. У наставку је дат дио програмског кода који се односи на бирање узорака који одступају од расподјеле баш овом политиком бирања.

У раду је додатно испитана још једна политика, која подразумјева да се узорци бирају насумично. У наставку рада ће ова политика бирања бити звана *random* политиком.

Из приложеног листинга се види да се у библиотеку узорака који одступају од расподјеле додају редом слике из одабране колекције слика. Након што се библиотека попуни, провјерава се у близини које ID класе се налази највише изабраних узорака. На мјесто једног узорка из околине најпопуларније класе, и то оног са најмањим односом удаљености, додаје се сљедећа слика уколико је њен однос удаљености већи од поменутог. Поступак се понавља све док се не прође кроз цијелу колекцију слика из које се бирају узорци ван расподјеле. Слично поступку показаном у оквиру листинга 5.1 имплементирају се и остале политике бирања узорака који одступају од расподјеле.

У оквиру листинга 5.1 се индекс узорку најближе ID класе одређује KNN методом, како се ради и у свим осталим експериментима у овом раду. Поређење KNN и MSP критеријума за означавање узорака ван расподјеле је дато у Табели 5.4. На MLRSNet-Holdout подскупу оба начина дају приближно једнак резултат, при чему се у оба случаја за 4.004 изабраних OE узорака добија побољшање од приближно 1% у односу на 88 изабраних OE узорака. Међутим, разлике су израженије на MLRSNet-Hard подскупу. При употреби MSP начина означавања, за 4.004 изабраних OE узорака се, у односу на 88 изабраних OE узорака, уочава побољшање перформанси од 3%, док је побољшање перформанси при KNN приближно 1%. Међутим, при мањим бројевима изабраних OE узорака се KNN начином означавања добија већа AUROC вриједност за чак 4% и то је главни разлог због којег се KNN критеријуму даје предност.

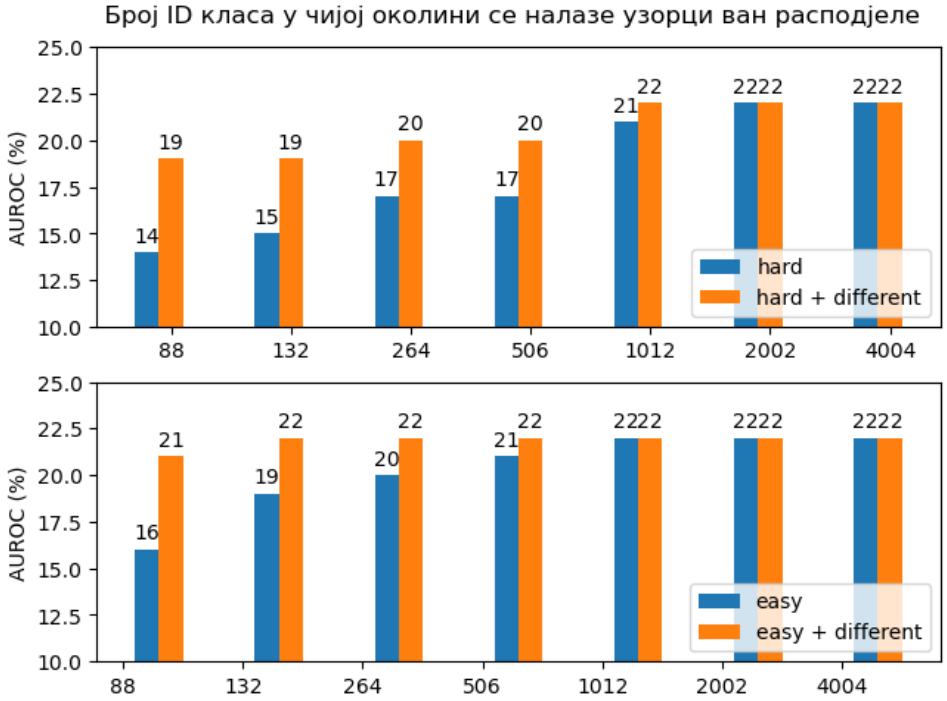
Листинг 5.1: Дио псеудокода за додавање узорка који одступа од расподјеле у OE колекцију *easy* + *different* политиком.

```
define update_OE_lib(curr_lib, lib_size, OE_sample, NNDR_score, KNN_pred):

    # Улази:
    #   curr_lib: тренутна колекција одабраних OE узорака
    #   lib_size: величина колекције OE узорака
    #   OE_sample: нови OE узорак
    #   NNDR_score: однос удаљености OE узорка у ID домену
    #   KNN_pred: предикција коју даје KNN класификатор за OE узорак

    # Излази:
    #   ажурирана колекција OE узорака

    # Ако OE колекција није попуњена, нови OE узорак се одмах додаје:

    if len(curr_lib) < lib_size:
        curr_lib.append((OE_sample, NNDR_score, KNN_pred))
        return curr_lib

    # У супротном, један OE узорак из колекције, а који припада
    # најзаступљенијој класи и има најмању NNDR_score вриједност
    # се мијења новим

    most_popular_class_index = find_most_popular_class(curr_lib)
    hardest_index, hardest_NNDR_score = …
    find_hardest_sample_from_most_popular_class(curr_lib,
                                        most_popular_class_index)

    # Ако је NNDR_score новог узорка већи од претходно пронађеног, треба
    # извршити замјену:

    if NNDR_score > hardest_NNDR_score:
        curr_lib[hardest_index] = (OE_sample, NNDR_score, KNN_pred)

    return curr_lib
```

Додатно, потребно је дати одговор на сљедеће питање: Да ли је повољније да одабрани узорци имају ознаке које одговарају најближим ID класама или их је потребно означити на неки други начин, рецимо употребом *k-means* кластеризације? Испитано је понашање детектора и у једном и другом случају, при чему је вршена *k-means* кластеризација узорака у 5, 10, 22 и 37 кластера. Паралелно утицају кластеризације, испитује се и ефекат који прави комбинација *different* политике са *hard* и *easy* политикама. Резултати су дати у табелама 5.5-5.9. У табелама 5.5-5.9 се колона означена са „/“ односи на случај у којем се *k-means* кластеризација не ради. Међутим, узорци су свакако груписани у 22 категорије према томе у околини које ID класе се налазе, и то KNN критеријумом који се претходно показао као бољи. Из Табеле 5.5 се види да је утицај *k-means* кластеризације на резултате минималан, како на

тестном MLRSNet-Hard скупу, тако и на MLRSNet-Holdout. Слично се показује и у случају *easy* и *easy* + *different* политика на оба OOD подскупа из MLRSNet скупа, што се види из табела 5.7 и 5.9. Примјеном *k-menas* кластеризације се остварују минимална побољшања у перформансама, која за већину посматраних бројева бираних OE, не прелази ни 0,1%. Са друге стране, *k-means* кластеризација доноси побољшање уколико се користе *hard* и *hard* + *different* политике, нарочито на MLRSNet-Holdout OOD скупу. У том случају се, из Табеле 5.8, при кластеризацији у 22 кластера, могу уочити побољшања перформанси детектора и за 1%, у односу на случај без кластеризације. Из табела 5.5-5.9 се види да је разлика у резултатима при кластеризацији узорака у 22 и 37 кластера минимална. Осим тога, број 37 представља број класа у колекцији слика из које се бирају узорци ван расподјеле и тај податак се често нема на располагању. Узевши наведено у обзир, у свим осталим експериментима се подразумјева да се ради *k-means* кластеризација са бројем кластера који одговара броју ID класа, што је у овом специфичном случају 22.

Према резултатима из табела 5.6 и 5.8, закључује се да уважавање *different* политике уз *hard* доноси значајно побољшање на оба посматрана OOD скупа. За мање бројеве бираних OE узорака, из обје табеле се може примјетити побољшање AUROC вриједности и преко 1%. Међутим, употреба *different* са *easy* даје значајан позитиван ефекат само на MLRSNet-Hard скупу, док је на MLRSNet-Holdout скупу разлика занемарљиво мала. Значај *different* компоненте у политици бирања је нарочито примјетан при мањим бројевима узорака који одступају од расподјеле, јер се без примјене *different* изабрани узорци распореде око малог броја ID класа. Када се бира велики број узорака који одступају од расподјеле, чак и у случају само *hard* и *easy* узорци се нађу у околини свих или скоро свих класа из познате расподјеле. Слика 5.8 приказује број ID класа у чијој околини постоје изабрани узорци. Разматрани су случајеви *hard* и *hard* + *different* политика, као и *easy* и *easy* + *different* политика при различитом броју бираних узорака који одступају од расподјеле.



Слика 5.8: Број ID класа у околини којих се налазе изабрани узорци који одступају од расподјеле.

У случају примјене *hard* политике, из табела 5.6 и 5.8 се види да вриједности AUROC углавном расту са порастом броја изабраних OE узорака. За изабраних 4.004 OE узорака се и на MLRSNet-Hard и MLRSNet-Holdout види пораст AUROC вриједности и за 3%, у односу на случај са изабраних 88 узорака. За *hard + different* је поменуто побољшање отприлике 1%. Побољшање перформанси детектора са повећањем броја OE узорака у случају *easy* политике постоји само на MLRSNet-Hard скупу, што се види из Табеле 5.7. Што се *easy + different* политике тиче, из табела 5.7 и 5.9 се види да повећање колекције OE узорака не доноси побољшања перформанси метода на MLRSNet-Hard, као ни на MLRSNet-Holdout скупу. И са графика са Слике 5.17 су видљиви закључци изведени на основу резултата из табела 5.6-5.9.

Табела 5.4 : Утицај критеријума за означавање узорака који одступају од расподјеле у случају *different* политике бирања.

Тестни OOD скуп		MLRSNet-Hard		MLRSNet-Holdout	
Критеријум за означавање узорака који одступају од расподјеле		MSP	KNN	MSP	KNN
Број изабраних узорака који одступају од расподјеле	88	93,11	95,91	96,83	96,71
	132	92,39	96,49	96,38	97,03
	264	93,38	96,27	96,75	96,99
	506	93,40	96,57	96,75	97,33
	1012	96,51	96,46	97,57	97,18
	2002	94,21	96,44	97,86	97,42
	4004	96,54	96,49	97,62	97,50

Табела 5.5 : Утицај *k-means* кластеризације на перформансе метода у случају примјене *different* политике бирања узорака.

Тестни OOD скуп		MLRSNet-Hard				MLRSNet-Holdout			
Број кластера		/	10	22	37	/	10	22	37
Број изабраних узорака који одступају од расподјеле	88	95,91	96,13	95,97	96,05	96,71	96,65	96,70	96,79
	132	96,49	96,42	96,37	96,35	97,03	96,97	97,00	97,02
	264	96,27	96,43	96,33	96,20	96,99	97,05	97,13	97,10
	506	96,57	96,77	96,64	96,62	97,33	97,29	97,39	97,41
	1012	96,46	96,74	96,46	96,63	97,18	97,33	97,40	97,38
	2002	96,44	96,77	96,74	96,64	97,42	97,38	97,56	97,54
	4004	96,49	96,71	96,74	96,90	97,50	97,58	97,69	97,75

Табела 5.6 : AUROC вриједности добијене на тестном OOD скупу MLRSNet-Hard, употребом *hard* и *hard + different* политика бирања, без и са *k-means* кластеризацијом.

Политика бирања		*hard*					*hard + different*				
Број кластера		/	5	10	22	37	/	5	10	22	37
Број изабраних узорака који одступају од расподјеле	88	93,69	93,35	94,38	94,24	94,74	94,94	94,78	95,10	95,53	95,51
	132	94,67	94,92	94,83	95,28	95,46	95,17	95,16	95,46	95,85	95,58
	264	94,56	94,62	94,71	95,52	95,59	95,57	95,61	95,70	96,30	96,09
	506	94,89	95,09	95,07	96,07	95,93	95,70	95,74	95,89	95,93	96,31
	1012	95,21	95,49	95,57	96,09	96,12	95,43	95,51	95,79	96,02	96,20
	2002	96,05	96,34	96,48	96,67	96,70	96,20	96,38	96,42	96,74	96,74
	4004	96,34	96,44	96,65	96,75	96,76	96,29	96,59	96,61	96,82	96,81

Табела 5.7 : AUROC вриједности добијене на тестном OOD скупу MLRSNet-Hard, употребом *easy* и *easy + different* политика бирања, без и са *k-means* кластеризацијом.

Политика бирања		*easy*					*easy + different*				
Број кластера		/	5	10	22	37	/	5	10	22	37
Број изабраних узорака који одступају од расподјеле	88	94,47	94,45	94,49	94,65	94,64	96,25	96,16	96,28	96,23	96,20
	132	94,60	94,56	94,67	94,67	94,74	96,46	96,48	96,54	96,55	96,54
	264	95,21	95,04	95,31	95,27	95,29	95,97	96,26	96,24	96,20	96,01
	506	94,92	94,80	94,90	94,94	94,93	96,02	95,91	96,02	96,06	95,92
	1012	95,38	95,27	95,45	95,32	95,36	97,14	97,15	97,09	96,94	97,02
	2002	95,56	95,37	95,66	95,58	95,50	96,90	96,94	96,92	96,93	96,94
	4004	95,63	95,53	95,87	95,80	95,61	96,37	96,60	96,82	96,52	96,69

Табела 5.8 : AUROC вриједности добијене на тестном OOD скупу MLRSNet-Holdout, употребом *hard* и *hard + different* политика бирања, без и са *k-means* кластеризацијом.

Политика бирања		*hard*					*hard + different*				
Број кластера		/	5	10	22	37	/	5	10	22	37
Број изабраних узорака који одступају од расподјеле	88	94,66	94,47	95,82	95,69	96,33	96,25	96,22	96,43	96,80	96,90
	132	95,57	95,28	95,58	96,44	96,77	96,53	96,55	96,84	97,16	97,22
	264	95,68	95,56	95,83	96,66	96,87	96,32	96,52	96,39	96,91	96,86
	506	96,20	96,30	96,29	97,12	97,19	96,55	96,57	96,91	96,76	97,22
	1012	96,57	96,77	96,70	97,28	97,44	96,76	96,80	97,00	97,21	97,43
	2002	97,26	97,47	97,36	97,66	97,74	97,18	97,26	97,32	97,64	97,70
	4004	97,53	97,68	97,53	97,79	97,83	97,51	97,70	97,59	97,82	97,80

Табела 5.9 : AUROC вриједности добијене на тестном OOD скупу MLRSNet-Holdout, употребом *easy* и *easy + different* политика бирања, без и са *k-means* кластеризацијом.

Политика бирања		*easy*					*easy + different*				
Број кластера		/	5	10	22	37	/	5	10	22	37
Број изабраних узорака који одступају од расподјеле	88	98,03	98,00	97,96	98,04	98,04	97,66	97,68	97,63	97,70	97,64
	132	97,97	97,93	97,94	97,93	97,95	97,86	97,87	97,92	97,92	97,93
	264	97,97	97,88	97,94	97,96	97,93	97,64	97,63	97,71	97,64	97,65
	506	97,99	97,96	97,92	97,95	97,94	97,92	97,88	97,92	97,89	97,90
	1012	97,99	97,98	97,87	97,90	97,93	97,98	97,85	97,93	97,84	97,92
	2002	98,06	97,99	97,98	97,94	97,95	97,83	97,78	97,72	97,74	97,77
	4004	98,11	97,98	98,03	98,01	98,06	97,60	97,73	97,61	97,51	97,69

Поређењем резултата које дају различите политике на истом скупу, долази се до закључка да *easy + different* даје најбољи резултат на MLRSNet-Hard скупу, а *easy* на MLRSNet-Holdout. При томе, на MLRSNet-Holdout је разлика између резултата које дају *easy* и *easy + different* прилично мала, тачније не прелази 0,5%. Међутим, било би превише рано да се закључи да је једна од поменутих политика најбољи избор у општем случају. Пожељно је да детектор буде испитан на још скупова слика – оних који су добијени даљинском детекцијом, али и на скуповима из других домена. Осим тога, вриједи испитати како колекција слика из које се узимају узорци који одступају од расподјеле утиче на перформансе.

Сви до сада изложени резултати, као и они који ће бити дати у наставку рада, се добијају полазећи од критеријума за детекцију датог изразом (3.18). У табелама 5.10 и 5.11 је показано да се, при различитим политикама бирања, бољи резултати добијају када критеријум за детекцију представља комбинација три карактеристична односа (3.18) умјесто два (3.17).

Табела 5.10: Поређење два OE метода за детекцију OOD узорака на MLRSNet-Hard скупу.

	different		*hard*		*easy*		*hard + different*		*easy + different*	
	(3.17)	(3.18)	(3.17)	(3.18)	(3.17)	(3.18)	(3.17)	(3.18)	(3.17)	(3.18)
88	94,92	95,97	93,17	94,24	94,17	94,65	94,55	95,53	95,87	96,23
132	95,39	96,37	95,05	95,28	94,52	94,67	95,28	95,85	95,52	96,55
264	95,79	96,33	94,81	95,52	94,87	95,27	95,70	96,30	94,75	96,20
506	95,46	96,64	95,86	96,07	94,52	94,94	94,86	95,93	95,85	96,06
1012	96,11	96,46	95,54	96,09	94,97	95,32	95,65	96,02	96,49	96,94
2002	95,96	96,74	96,26	96,67	94,97	95,58	96,21	96,74	96,30	96,93
4004	96,05	96,74	96,23	96,75	95,03	95,80	96,18	96,82	96,15	96,52

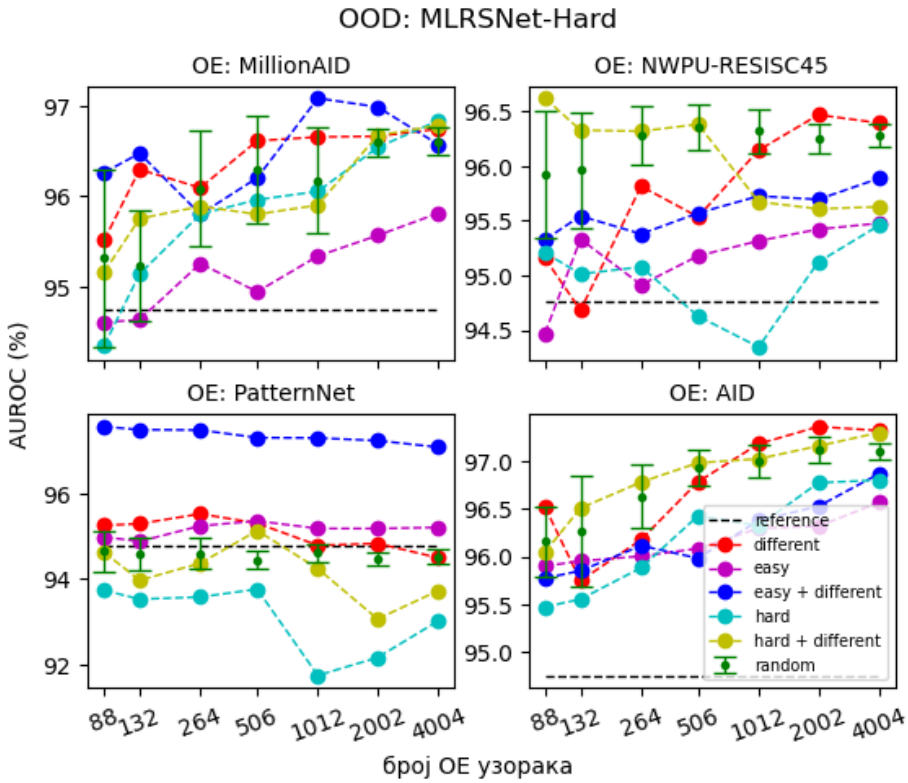
Табела 5.11: Поређење два OE метода за детекцију OOD узорака на MLRSNet-Holdout скупу.

	different		*hard*		*easy*		*hard + different*		*easy + different*	
	(3.17)	(3.18)	(3.17)	(3.18)	(3.17)	(3.18)	(3.17)	(3.18)	(3.17)	(3.18)
88	96,19	96,70	95,63	95,69	97,50	98,04	96,04	96,80	97,41	97,70
132	96,95	97,00	95,84	96,44	97,54	97,93	96,77	97,16	97,40	97,92
264	96,77	97,13	96,11	96,66	97,48	97,96	96,65	96,91	97,26	97,64
506	96,92	97,39	96,70	97,12	97,33	97,95	95,75	96,76	97,20	97,89
1012	96,81	97,40	96,78	97,28	97,30	97,90	96,76	97,21	97,12	97,84
2002	96,70	97,56	97,12	97,66	97,05	97,94	96,65	97,64	96,87	97,74
4004	96,59	97,69	97,02	97,79	97,19	98,01	97,09	97,82	96,91	97,51

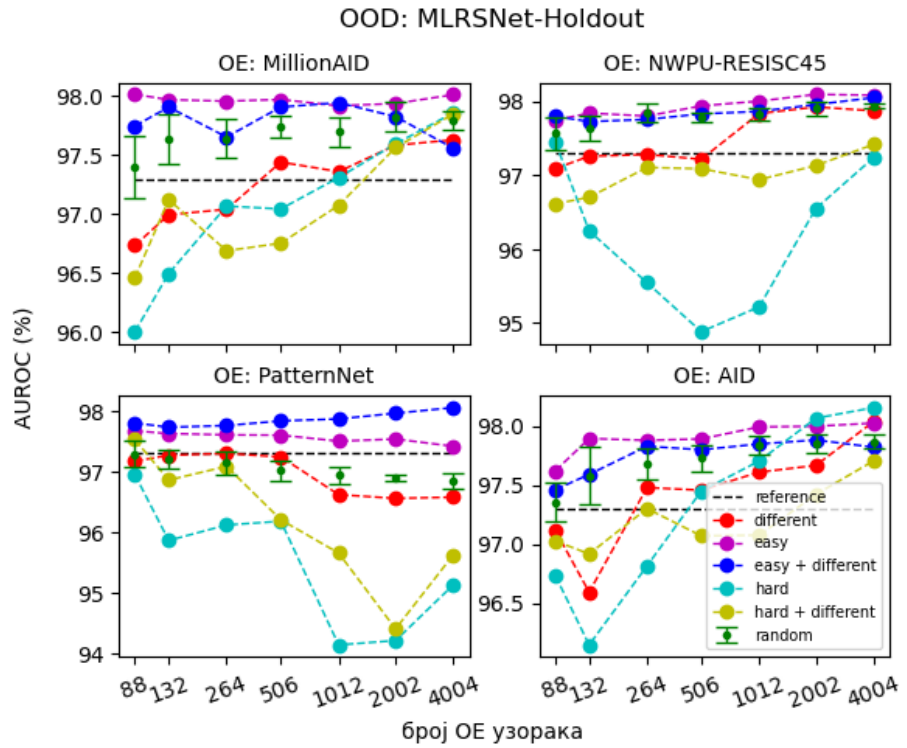
5.4.2. Утицај скупа из којег се бирају узорци ван расподјеле

Осим из MillionAID колекције, узорци који одступају од расподјеле се бирају и из других скупова добијених даљинском детекцијом: NWPU-RESISC45, PatternNet и AID, али и из скупова из других домена: ImageNet100 и Food5K. Узорци се могу бирати и из скупова у којима се налазе слике из различитих домена, што се у раду демонстрира на скуповима добијеним конкатенацијом AID и ImageNet100, као и MillionAID и Food5K скупова. Детектор је у сваком поменутом случају тестиран на два OOD скупа: MLRSNet-Hard и MLRSNet-Holdout и добијене AUROC вриједности су приказане графички на сликама 5.9-5.16.

Поређење резултата које дају различите политике бирања узорака на MLRSNet-Hard и MLRSNet-Holdout скуповима је дато на сликама 5.9 и 5.10, при различитим OE колекцијама. Уочава се да при свим разматраним OE колекцијама слика *hard + different* и *easy + different* у односу на *hard* и *easy* дају боље резултате на MLRSNet-Hard OOD скупу. На MLRSNet-Holdout скупу *different* политика у комбинацији са *hard* или *easy* доноси побољшање само у случају бирања узорака из појединих OE колекција. Генерално, и када се побољшање употребом *different* није остварило, разлике у перформансама су минималне. Из тог разлога у наставку рада тестирање није ни вршено коришћењем политика *hard* и *easy*, без *different* компоненте. Треба истаћи да се, у случају када се узорци који одступају од расподјеле бирају из PatternNet и NWPU-RESISC45 скупова, политикама *hard* и *hard + different* не остварује побољшање детекције. Дакле, резултати су лошији у односу на базни случај, тј. када се на располагању нема ниједан узорак који одступа од ID расподјеле. Једино се политикама *easy* и *easy + different* на оба тестна OOD скупа остварује побољшање перформанси детектора у сваком разматраном случају. За те двије политике је карактеристично да се перформансе метода не мијењају значајно са промјеном броја OE узорака. Дакле, величина одабраног OE скупа нема велики утицај на перформансе, што се нарочито може примјетити на MLRSNet-Holdout скупу, без обзира на колекцију слика из које се бирају OE узорци. Додатно, са Слике 5.9 се примјећује да се бирањем OE узорака из AID колекције, на MLRSNet-Hard скупу, остварују AUROC вриједности веће за 3%. У односу на базни случај, уочено побољшање је изненађујуће значајно, нарочито кад се узме у обзир величина AID скупа, али и специфична резолуција припадајућих слика.



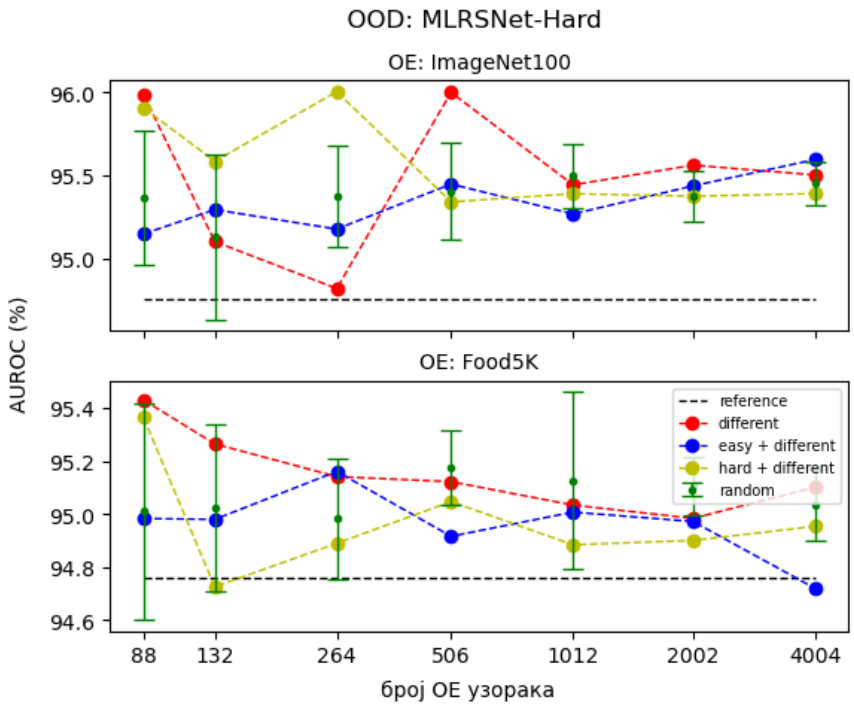
Слика 5.9: AUROC вриједности на тестном MLRSNet-Hard скупу при различитим RS колекцијама слика из којих се бирају узорци ван расподјеле.



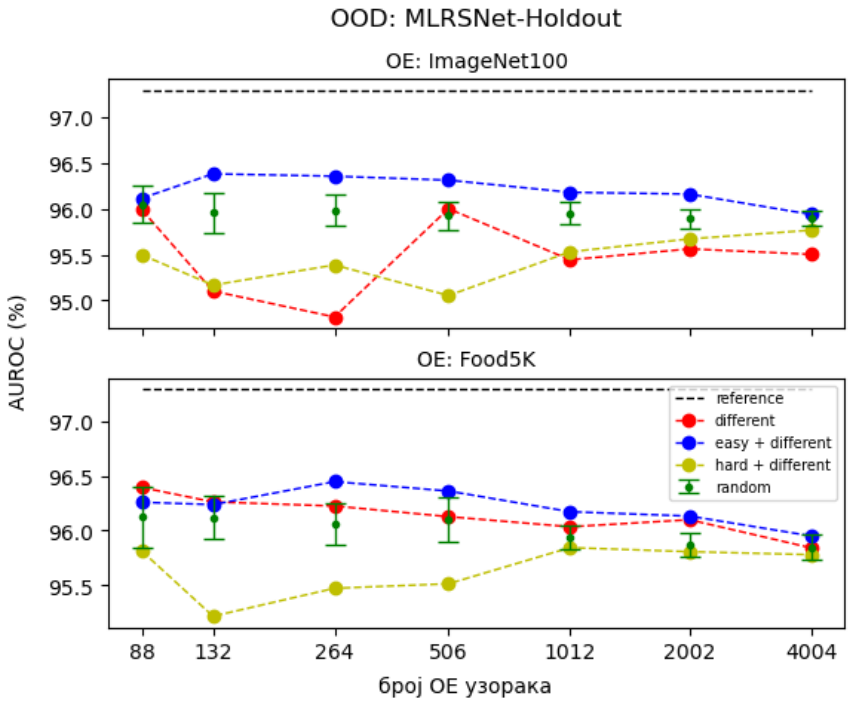
Слика 5.10: AUROC вриједности на тестном MLRSNet-Holdout скупу при различитим RS колекцијама слика из којих се бирају узорци ван расподјеле.

Очекивано, ако се узорци бирају из скупова слика које нису добијене даљинском детекцијом, остварују се лошији резултати на оба тестна OOD скупа. Графици са слика 5.11-5.12 приказују AUROC вриједности детекције OOD узорака при различитим политикама бирања узорака, у случају да се OE узорци бирају из ImageNet100 и Food5K колекција слика. На MLRSNet-Hard скупу се и на овај начин остварује побољшање AUROC вриједности, које достиже нешто изнад 1%. Перформансе детектора на MLRSNet-Holdout скупу су такве да примјена свих испитаних политика бирања резултује лошијим резултатима у односу на случај без узорака ван расподјеле.

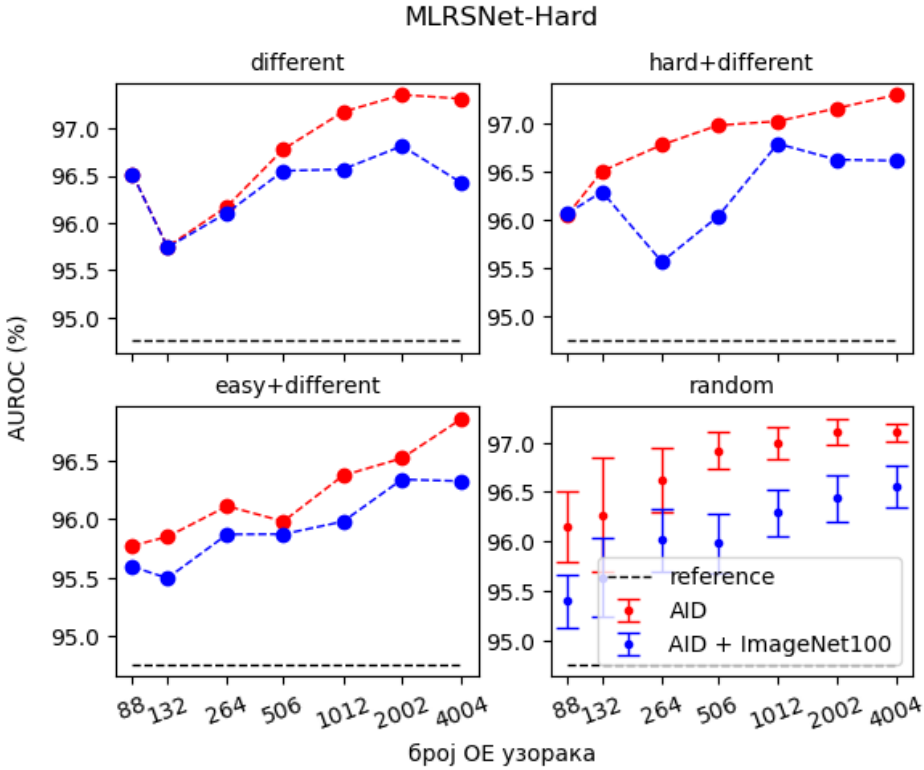
На Сликама 5.13-5.16 су приказане AUROC вриједности добијене тестирањем детектора на MLRSNet-Hard и MLRSNet-Holdout скуповима, уколико су узорци који одступају од расподјеле бирани из скупова добијених конкатенацијом једног произвољно изабраног скупа чије су слике добијене даљинском детекцијом, те једног произвољно изабраног скупа из другог домена. Конкретно, извршена је конкатенација AID скупа са ImageNet100, а MillionAID са Food5K. Упоредо су дате AUROC вриједности добијене при бирању узорака из скупа слика из домена даљинске детекције и из скупа добијеног конкатенацијом. На тај начин се испитује утицај присуства слика које нису добијене даљинском детекцијом у колекцији слика из које се узимају узроци ван расподјеле. Експерименти су показали да располагање узорцима који одступају од ID расподјеле, а који нису из домена даљинске детекције, утиче негативно на резултате на OOD скупу који је сличнији ID подацима, тј. на MLRSNet-Hard скупу. Негативан утицај се нарочито види када се узорци бирају из скупа добијеног конкатенацијом MillionAID и Food5K, на Слици 5.15, кад се за 506 и 1.012 изабраних OE узорака добија мања AUROC вриједност за више од 1%. Са друге стране, на MLRSNet-Holdout скупу се бирањем OE узорака који нису добијени даљинском детекцијом може остварити и побољшање, али са Слика 5.14 и 5.16 се види да не постоји драстична разлика.



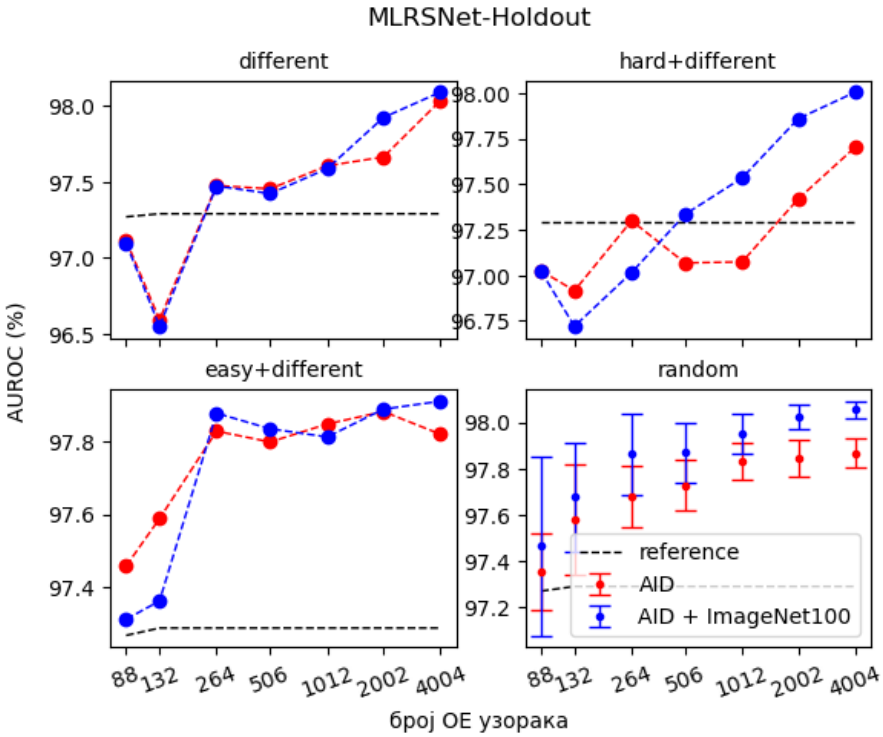
Слика 5.11: AUROC вриједности на тестном MLRSNet-Hard скупу при различитим NRS колекцијама слика из којих се бирају узорци ван расподјеле.



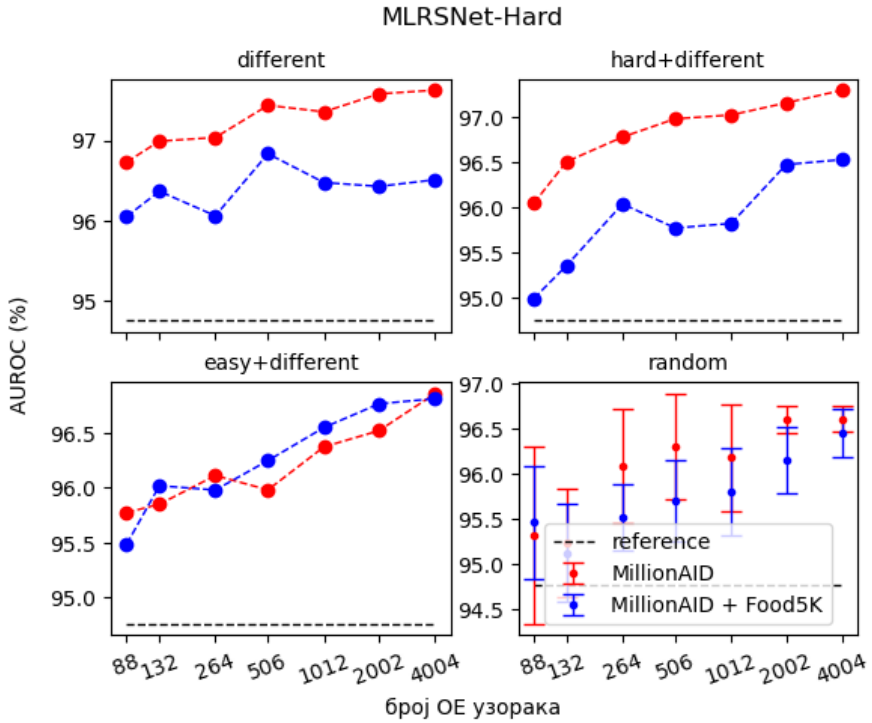
Слика 5.12: AUROC вриједности на тестном MLRSNet-Holdout скупу при различитим NRS колекцијама слика из којих се бирају узорци ван расподјеле.



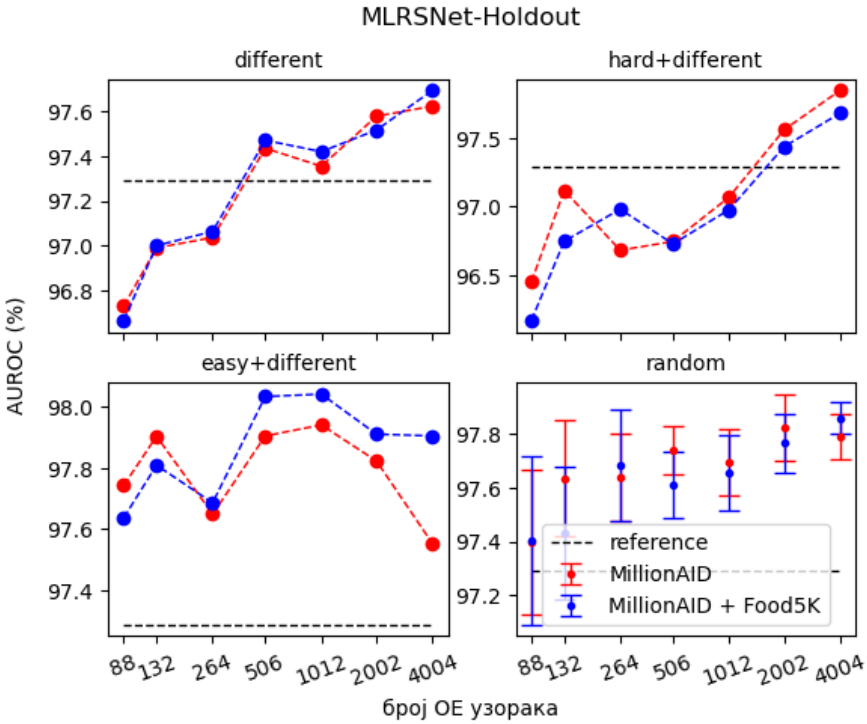
Слика 5.13: Резултати добијени на MLRSNet-Hard у случају бирања узорака ван расподјеле из AID скупа и скупа добијеног конкатенацијом AID и Imagenet100.



Слика 5.14: Резултати добијени на MLRSNet-Holdout у случају бирања узорака ван расподјеле из AID скупа и скупа добијеног конкатенацијом AID и Imagenet100.



Слика 5.15: Резултати добијени на MLRSNet-Hard у случају бирања узорака ван расподјеле из MillionAID скупа и скупа добијеног конкатенацијом MillionAID и Food5K.



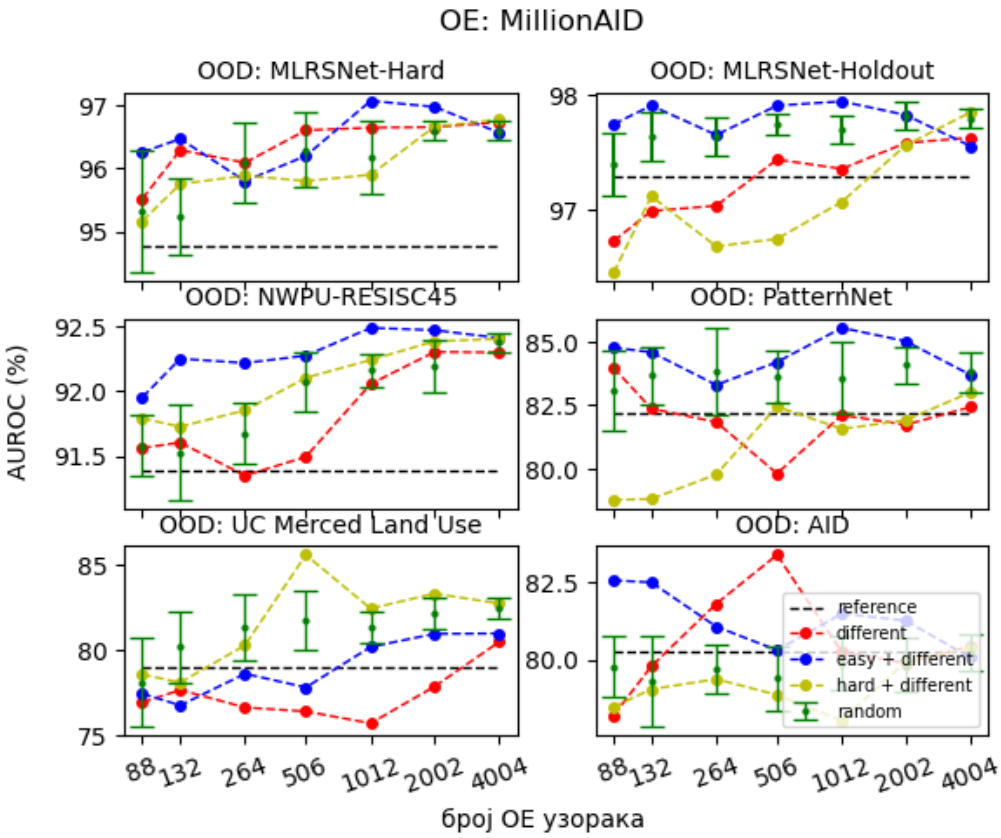
Слика 5.16: Резултати добијени на MLRSNet-Holdout у случају бирања узорака ван расподјеле из MillionAID скупа и скупа добијеног конкатенацијом MillionAID и Food5K.

5.4.3. Перформансе метода на различитим OOD скуповима

На путу ка изношењу приједлога о најбољој политици бирања узорака који одступају од расподјеле, детектор је додатно тестиран на неколико OOD скупова који одступају од расподјеле ID података. Из колекција слика NWPU-RESISC45, PatternNet, UC Merced Land Use и AID су елиминисане класе које припадају ID расподјели, те су преостали дијелови поменутих скупова кориштени за тестирање. У овом случају, узорци који одступају од расподјеле су све вријеме бирани из MillionAID скупа. Резулттати су дати на Слици 5.17.

У пет од шест испитаних сценарија политика *easy + different* даје најбољи резултат. Уколико се узму у обзир и сви раније урађени експерименти, може се примјетити да се управо ова политика показала најбољом, у смислу да је често давала највеће AUROC вриједности и да ни у једном случају није покварила перформансе полазног детектора. Са друге стране, на три OOD скупа, у већини испитаних тачака величине изабране OE колекције, политика *hard + different* даје лошије резултате него детектор који не користи OE узорке. Исто се може констатовати и за резултат који даје политика *different* na UC Merced Land Use OOD скупу.

Са Слике 5.17 се може уочити да повећање броја изабраних OE узорака не гарантује боље перформансе детектора. На половини OOD скупова кориштених у експерименту (MLRSNet-Holdout, PatternNet и AID) се *easy + different* политиком остварује боља AUROC вриједност употребом 88 OE узорака, у односу на 4.004 OE узорка. На PatternNet и AID скуповима се, уз само њих 88, има за отприлике 2,5% већа вриједност AUROC у односу на базни случај, када детектор не користи OE узорке.



Слика 5.17: AUROC вриједности на различитим OOD скуповима из домена даљинске детекције

Од свих испитаних OOD скупова, AID је онај на којем се примјећује најслабији успјех детектора. Разлог за то може бити специфична величина слика из AID скупа, од 600×600 пиксела. Трансформација њене величине у 224×224 резултује губитком квалитета, те се на тај начин отежава задатак детектору.

Резултати добијени помоћу релативно једноставног и тако мало захтјевног метода су упоредиви са онима добијеним у раду [18]. Наиме, у поменутом раду се на скуповима MLRSNet-Hard и MLRSNet-Holdout прикупљањем података који одступају од расподјеле прије тестне фазе и модификацијом класификатора остварује побољшање AUROC вриједности од 2,4% и 0,0%, респективно. Предложени метод у овом раду, на истим скуповима података, претходно прикупљеним узорцима који одступају од расподјеле, без модификације класификатора и при оптималној политици бирања OE узорака остварује побољшања од приближно 3% и 1%. Треба узети у обзир и чињеницу да је у поменутом раду искоришћена цијела fmoW [72] колекција од милион слика, само умањена за слике које припадају ID класама, док је у овом раду максимални број искоришћених OE узорака само 4.004. Осим тога, треба имати у виду да је у [18] OE библиотека попуњавана и у тестној фази, OOD узорцима из MLRSNet скупа, а у овом раду једино OOD узорцима из других RS колекција. Стога, правично је поредити резултате добијене у овом раду са резултатима из помињаног рада, али само ако у [18] не узимамо у обзир резултате који се односе на MLRSNet OOD скупове. С тим у вези, уочава се још једна предност новопредложеног метода. Наиме, повећање броја коришћених OE узорака углавном резултује побољшањем перформанси детекције и перформансе су, за сваку испитану величину OE колекције, боље у односу на базни случај (без иједног OE узорка). Одступање уочених резултата детекције у оквиру овог

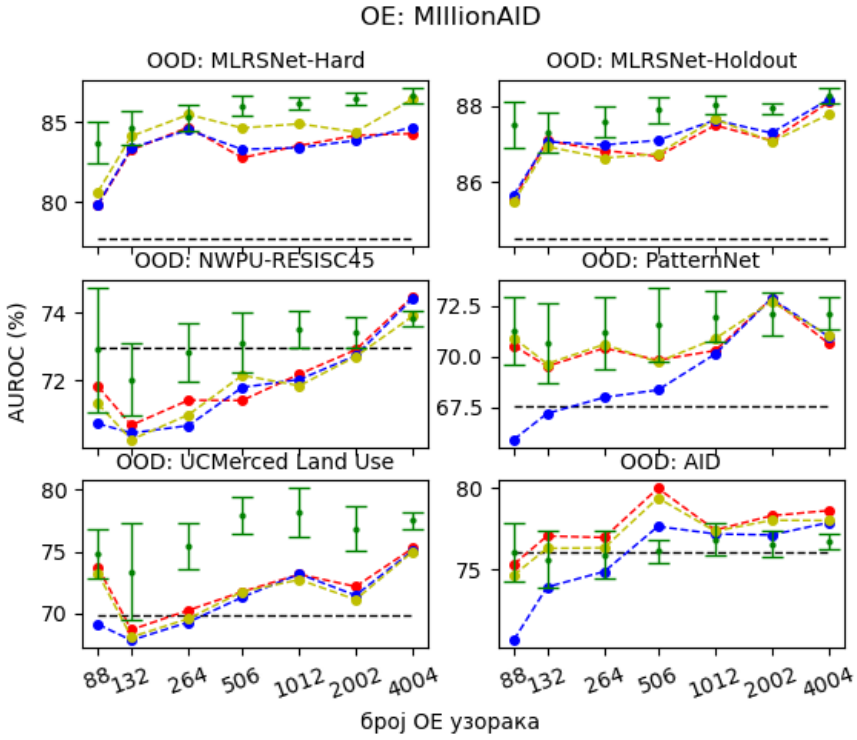
рада, који су лошији у односу на базни случај, је занемариво мало у односу на оно које се на скупу RSI-CB256, са 264 изабрана OE узорка, јавља у [18] и које премашује чак 4%.

5.4.4. Утицај величине тренинг скупа

Да би се испитао утицај величине тренинг скупа на перформансе метода, детектор је тестиран на истим OOD скуповима као и у претходном одјељку, при чему је основни класификатор обучен на само 440 тренинг слика из ID дијела MLRSNet скупа. Тренинг слике су униформно расподијељене у 22 ID класе и представљају подскуп тренинг скупа који је кориштен у ранијим експериментима. И у овом експерименталном дијелу су узорци који одступају од расподјеле бирани из MillionAID скупа слика. На Слици 5.18 су дати резултати које дају различите политике бирања у случају мањег тренинг скупа.

У пет од испитаних шест сценарија *random* избор узорака ван расподјеле резултује највећим AUROC вриједностима. Поменуто води ка закључку да је у случају обучавања основног класификатора на малом тренинг скупу препоручљиво узорке бирати насумично, а не употребом неке од предложених политика, које подразумјевају да се узорци бирају на основу правила. Да би се отклонила сумња да овај закључак узрокује само специфичан избор узорака који одступају од расподјеле, експерименти су поновљени за десет различитих избора. Из тог разлога су резултати који се односе на *random* политику дати помоћу средњих вриједности и стандардних девијација.

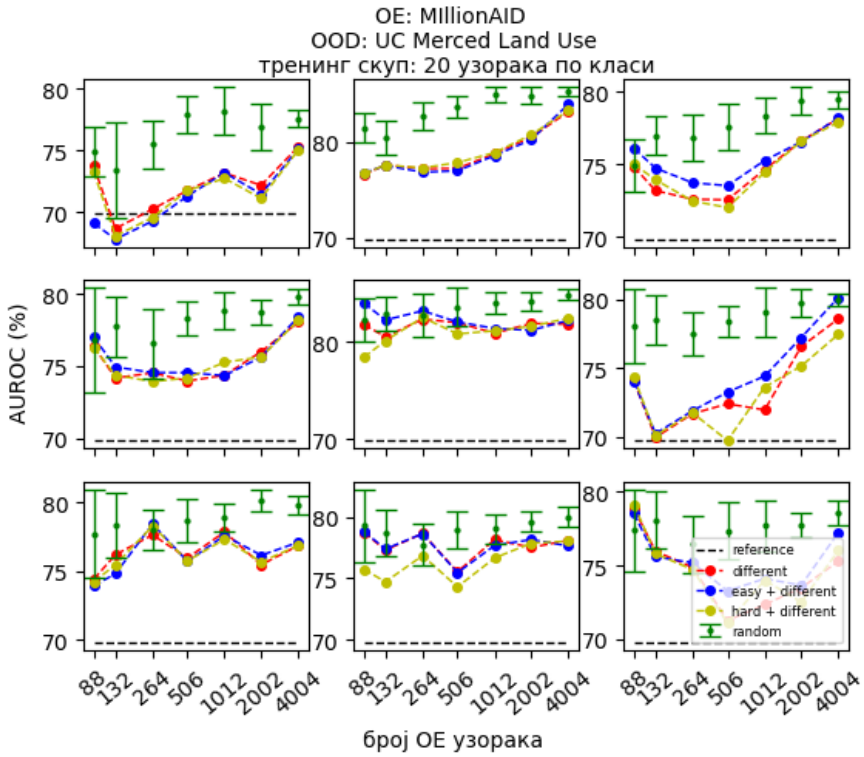
У поређењу са перформансама немодификованог детектора, дакле без употребе узорака који одступају од расподјеле, предложеним методом се остварује веће апсолутно побољшање резултата детекције уколико се на располагању има мањи број узорака за обучавање мреже. За исти број узорака који одступају од расподјеле у случају мањег тренинг скупа, постиже се побољшање AUROC вриједност од чак и преко 10%, док у случају



Слика 5.18: AUROC вриједности на различитим OOD скуповима из домена даљинске детекције у случају обучавања класификатора на малом тренинг скупу

полазног тренинг скупа не прелази 6%. При томе, поменуто побољшање од 6% је примјећено на Слици 5.17, на UC Merced Land Use колекцији, кад је изабрано 506 OE узорака, и то *hard + different* политиком. Генерално, за све посматране веичине OE колекције и на различитим OOD скуповима, у случају да се користи цијели тренинг скуп, побољшања су знатно мања и не прелазе 3%. Дакле, уочено говори да је предложени метод детекције нарочито препоручљиво користити у ситуацијама у којима није могуће основни класификатор обучити на великом броју слика.

Да би се остварила независност резултата од избора тренинг скупа, намеће се идеја да би експерименте било добро поновити више пута, за различите скупове кориштене у процесу обучавања. Како се највећа предност *random* политике види на UC Merced Land Use скупу, згодно би било на њему испитати колику разлику прави промјена тренинг скупа. Резултат је дат на Слици 5.19, за девет различитих скупова кориштених за обучавање основног класификатора. Сваки тренинг скуп величине од 440 слика је насумично одабрани подскуп скупа кориштеног за обучавање у почетним експериментима. На основу приложених графика се види да је надмоћ *random* политике над осталима, за неке специфичне изборе тренинг скупа, изражена у слабијој мјери. Ипак, резултати показују да је, за сваки од испитаних избора тренинг скупа, најбоље насумично бирати узорке који одступају од расподјеле.



Слика 5.19: AUROC вриједности на UC Merced Land Use скупу у случају обучавања класификатора на малом тренинг скупу

6. Закључак

Поређењем резултата добијених примјеном базног (MSP) метода и пет метода базираних на мјерењу удаљености (KNN, NCM, NNDR, MD, RMD) уочено је да детекцију OOD узорака из MLRSNet-Hard и MLRSNet-Holdout скупова најуспјешније врши RMD метод. Осим тога, и на већем дијелу осталих RS и NRS скупова на којима су методи тестирани, поменути метод даје највећу AUROC и најмању FPR вриједност. Генерално, значајно боље перформансе свих метода су уочене на скуповима који нису из RS домена, али лошије на RS скуповима чија расподјела одступа од расподјеле базног MLRSNet скупа.

Ограничења разне природе која могу да се јаве у пракси у процесу прикупљања слика су била основна мотивација за испитивање утицаја величине тренинг скупа на перформансе метода. Наиме, смањење обима тренинг скупа резултује слабљењем перформанси свих метода. Међутим, динамика којом перформансе појединих метода слабе у незанемаривој мјери зависи од метода, те поредак резултата које дају поједини методи не остаје исти при различитим величинама тренинг скупа. Мали број тренинг узорака узрокује проблем при процјени матрица коваријанси потребних за реализацију RMD метода. У том случају, за отприлике 10 пута мањи тренинг скуп (20 тренинг узорака по класи) добија се за чак 12% мања AUROC вриједност на MLRSNet-Hard скупу. Поређења ради, при истом смањењу обима тренинг скупа и на истом тестном скупу, AUROC вриједност које даје NCM метод се смањи за само 4%. Узевши све добијене резултате у обзир, NCM и KNN уз косинусну метрику су се показали као најмање осјетљиви на величину тренинг скупа, па је њихова примјена прикладна у ситуацијама у којима није могуће прикупити велики скуп слика за обучавање.

За испитивање утицаја архитектуре којом се издвајају обиљежја, у улози основног класификатора су искориштене и двије трансформаторске архитектуре. Показано је да сви методи, сем NNDR са Еуклидовом метриком, дају боље резултате са бар једним трансформатором у односу на случај када се обиљежја издвајају помоћу ResNet50 конволуционе мреже. Посебно је изражена разлика у резултатима који се добијају методом заснованом на мјерењу Махаланобисових удаљености. Уколико се обиљежја издвајају ResNet50 мрежом, добијене AUROC вриједности MD методом на базним OOD скуповима су мање од 80%, док у случају издвајања обиљежја трансформаторским архитектурама оне прелазе чак 95%.

На крају, показано је да предложени метод добијен модификацијом NNDR има способност да побољша резултате детекције употребом ограничене колекције изабраних слика које одступају од расподјеле тренинг података, и то прикупљених прије тестне фазе и без модификације основног класификатора. Испитивањем разних политика бирања узорака који одступају од расподјеле, из различитих колекција слика, показано је да је најповољније OE библиотеку попуњавати оним узорцима који имају највећи однос удаљености у ID домену, али и да треба водити рачуна о томе да изабрани узорци буду равномјерно распоређени по простору обиљежја (*easy + different* политика). Ипак, уколико се на располагању има мали број тренинг узорака, експерименти су показали да је OE узорке је најбоље бирати насумично (*random* политиком).

Овај рад оставља могућност да се предложени метод за детекцију OOD узорака надогради у будућности, и то: (1) предлагањем нових, интелигентнијих, политика бирања узорака који одступају од расподјеле, (2) генерисањем и коришћењем виртуелних OE узорака, добијених примјеном разних операција на реалне OE узорке или комбинујући тренинг узорке са реалним OE, (3) прикупљањем узорака који одступају од ID расподјеле у реалном времену,

у тестној фази, при чему модификација основног класификатора не би била неопходна и слично.

Литература

[1] Campbell, James B., and Randolph H. Wynne. *Introduction to remote sensing*. Guilford press, 2011.

[2] Cheng, Gong, et al. "Remote sensing image scene classification meets deep learning: Challenges, methods, benchmarks, and opportunities." *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing* 13 (2020): 3735-3756.

[3] O'Shea, Keiron, and Ryan Nash. "An introduction to convolutional neural networks." *arXiv preprint arXiv:1511.08458* (2015).

[4] Krizhevsky, A., I. Sutskever, and G. Hinton. "Imagenet classification with deep convolutional networks." *Proceedings of the 26th Annual Conference on Neural Information Processing Systems (NIPS)*.

[5] Simonyan, Karen, and Andrew Zisserman. "Very deep convolutional networks for large-scale image recognition." *arXiv preprint arXiv:1409.1556* (2014).

[6] He, Kaiming, et al. "Deep residual learning for image recognition." *Proceedings of the IEEE conference on computer vision and pattern recognition*. 2016.

[7] Szegedy, Christian, et al. "Going deeper with convolutions." *Proceedings of the IEEE conference on computer vision and pattern recognition*. 2015.

[8] Howard, Andrew G., et al. "Mobilenets: Efficient convolutional neural networks for mobile vision applications." *arXiv preprint arXiv:1704.04861* (2017).

[9] Tan, Mingxing, and Quoc Le. "Efficientnet: Rethinking model scaling for convolutional neural networks." *International conference on machine learning*. PMLR, 2019.

[10] Dosovitskiy, Alexey, et al. "An image is worth 16x16 words: Transformers for image recognition at scale." *arXiv preprint arXiv:2010.11929* (2020).

[11] Yang, Jingkang, et al. "Generalized out-of-distribution detection: A survey." *arXiv preprint arXiv:2110.11334* (2021).

[12] Blaschke, Thomas, and Josef Strobl. "What's wrong with pixels? Some recent developments interfacing remote sensing and GIS." *Zeitschrift für Geoinformationssysteme* (2001): 12-17.

[13] Hendrycks, Dan, Mantas Mazeika, and Thomas Dietterich. "Deep anomaly detection with outlier exposure." *arXiv preprint arXiv:1812.04606* (2018).

[14] Zhang, Jingyang, et al. "Mixture outlier exposure: Towards out-of-distribution detection in fine-grained environments." *Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision*. 2023.

[15] Roy, Abhijit Guha, et al. "Does your dermatology classifier know what it doesn't know? detecting the long-tail of unseen conditions." *Medical Image Analysis* 75 (2022): 102274.

[16] Fort, Stanislav, Jie Ren, and Balaji Lakshminarayanan. "Exploring the limits of out-of-distribution detection." *Advances in Neural Information Processing Systems* 34 (2021): 7068-7081.

[17] Ghoting, Amol, Srinivasan Parthasarathy, and Matthew Eric Otey. "Fast mining of distance-based outliers in high-dimensional datasets." *Data Mining and Knowledge Discovery* 16 (2008): 349-364.

[18] Inkawhich, Nathan, et al. "Improving out-of-distribution detection by learning from the deployment environment." *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing* 15 (2022): 2070-2086.

[19] Kim, Jaeyoung, et al. "Pseudo Outlier Exposure for Out-of-Distribution Detection using Pretrained Transformers." *arXiv preprint arXiv:2307.09455* (2023).

[20] Thulasidasan, Sunil, et al. "A simple and effective baseline for out-of-distribution detection using abstention." (2020).

[21] Zhang, Hongyi, et al. "mixup: Beyond empirical risk minimization." *arXiv preprint arXiv:1710.09412* (2017).

[22] Vladimir Risojević: „Multimedijalni sistemi“, Univerzitet u Banjoj Luci, Elektrotehnički fakultet (2018)

[23] Lowe, David G. "Distinctive image features from scale-invariant keypoints." *International journal of computer vision* 60 (2004): 91-110.

[24] Haralick, Robert M., Karthikeyan Shanmugam, and Its' Hak Dinstein. "Textural features for image classification." *IEEE Transactions on systems, man, and cybernetics* 6 (1973): 610-621.

[25] Jain, Anil K., Nalini K. Ratha, and Sridhar Lakshmanan. "Object detection using Gabor filters." *Pattern recognition* 30.2 (1997): 295-309.

[26] Ojala, Timo, Matti Pietikainen, and Topi Maenpaa. "Multiresolution gray-scale and rotation invariant texture classification with local binary patterns." *IEEE Transactions on pattern analysis and machine intelligence* 24.7 (2002): 971-987.

[27] Swain, Michael J., and Dana H. Ballard. "Color indexing." *International journal of computer vision* 7.1 (1991): 11-32.

[28] Dalal, Navneet, and Bill Triggs. "Histograms of oriented gradients for human detection." *2005 IEEE computer society conference on computer vision and pattern recognition (CVPR'05)*. Vol. 1. Ieee, 2005.

[29] Oliva, Aude, and Antonio Torralba. "Modeling the shape of the scene: A holistic representation of the spatial envelope." *International journal of computer vision* 42 (2001): 145-175.

[30] Jégou, Hervé, et al. "Aggregating local image descriptors into compact codes." *IEEE transactions on pattern analysis and machine intelligence* 34.9 (2011): 1704-1716.

[31] Yang, Yi, and Shawn Newsam. "Bag-of-visual-words and spatial extensions for land-use classification." *Proceedings of the 18th SIGSPATIAL international conference on advances in geographic information systems*. 2010.

[32] Yan, Le Cun, B. Yoshua, and H. Geoffrey. "Deep learning." *nature* 521.7553 (2015): 436-444.

[33] Dridi, Salim. "Supervised learning-a systematic literature review." (2021).

[34] Vaswani, Ashish, et al. "Attention is all you need." *Advances in neural information processing systems* 30 (2017).

[35] Deng, Jia, et al. "Imagenet: A large-scale hierarchical image database." *2009 IEEE conference on computer vision and pattern recognition*. Ieee, 2009.

[36] Ridnik, Tal, et al. "Imagenet-21k pretraining for the masses." *arXiv preprint arXiv:2104.10972* (2021).

[37] Hinton, Geoffrey, Oriol Vinyals, and Jeff Dean. "Distilling the knowledge in a neural network." *arXiv preprint arXiv:1503.02531* (2015).

[38] Chollet, François. "Xception: Deep learning with depthwise separable convolutions." *Proceedings of the IEEE conference on computer vision and pattern recognition*. 2017.

[39] Hendrycks, Dan, and Kevin Gimpel. "A baseline for detecting misclassified and out-of-distribution examples in neural networks." *arXiv preprint arXiv:1610.02136* (2016).

[40] Lee, Kimin, et al. "A simple unified framework for detecting out-of-distribution samples and adversarial attacks." *Advances in neural information processing systems* 31 (2018).

[41] Ren, Jie, et al. "A simple fix to mahalanobis distance for improving near-ood detection." *arXiv preprint arXiv:2106.09022* (2021).

[42] Kamoi, Ryo, and Kei Kobayashi. "Why is the mahalanobis distance effective for anomaly detection?." *arXiv preprint arXiv:2003.00402* (2020).

[43] Denouden, Taylor, et al. "Improving reconstruction autoencoder out-of-distribution detection with mahalanobis distance." *arXiv preprint arXiv:1812.02765* (2018).

[44] Yang, Yijun, Ruiyuan Gao, and Qiang Xu. "Out-of-distribution detection with semantic mismatch under masking." *European Conference on Computer Vision*. Cham: Springer Nature Switzerland, 2022.

[45] Ristin, Marko, et al. "Incremental learning of ncm forests for large-scale image classification." *Proceedings of the IEEE conference on computer vision and pattern recognition*. 2014.

[46] Bendale, Abhijit, and Terrance Boult. "Towards open world recognition." *Proceedings of the IEEE conference on computer vision and pattern recognition*. 2015.

[47] Chen, Guangyao, et al. "Learning open set network with discriminative reciprocal points." *Computer Vision–ECCV 2020: 16th European Conference, Glasgow, UK, August 23–28, 2020, Proceedings, Part III 16*. Springer International Publishing, 2020.

[48] Back to the Basics: Revisiting Out-of-Distribution Detection Baselines

[49] Sun, Yiyou, et al. "Out-of-distribution detection with deep nearest neighbors." *International Conference on Machine Learning*. PMLR, 2022.

[50] Dimitrić, Dajana, Vladimir Risojević, and Mitar Simić. "Nearest Neighbor Based Out-of-Distribution Detection in Remote Sensing Scene Classification." *2023 22nd International Symposium INFOTEH-JAHORINA (INFOTEH)*. IEEE, 2023.

[51] Fix, Evelyn, and Joseph Lawson Hodges. "Discriminatory analysis. Nonparametric discrimination: Consistency properties." *International Statistical Review/Revue Internationale de Statistique* 57.3 (1989): 238-247.

[52] Cover, Thomas, and Peter Hart. "Nearest neighbor pattern classification." *IEEE transactions on information theory* 13.1 (1967): 21-27.

[53] Mendes Júnior, Pedro R., et al. "Nearest neighbors distance ratio open-set classifier." *Machine Learning* 106.3 (2017): 359-386.

[54] Yun, Sangdoo, et al. "Cutmix: Regularization strategy to train strong classifiers with localizable features." *Proceedings of the IEEE/CVF international conference on computer vision*. 2019.

[55] Verma, Vikas, et al. "Manifold mixup: Better representations by interpolating hidden states." *International conference on machine learning*. PMLR, 2019.

[56] Fort, Stanislav, Jie Ren, and Balaji Lakshminarayanan. "Exploring the limits of out-of-distribution detection." *Advances in Neural Information Processing Systems* 34 (2021): 7068-7081.

[57] Qi, Xiaoman, et al. "MLRSNet: A multi-label high spatial resolution remote sensing dataset for semantic scene understanding." *ISPRS Journal of Photogrammetry and Remote Sensing* 169 (2020): 337-350.

[58] Cheng, Gong, Junwei Han, and Xiaoqiang Lu. "Remote sensing image scene classification: Benchmark and state of the art." *Proceedings of the IEEE* 105.10 (2017): 1865-1883.

[59] Zhou, Weixun, et al. "PatternNet: A benchmark dataset for performance evaluation of remote sensing image retrieval." *ISPRS journal of photogrammetry and remote sensing* 145 (2018): 197-209.

[60] Xia, Gui-Song, et al. "AID: A benchmark data set for performance evaluation of aerial scene classification." *IEEE Transactions on Geoscience and Remote Sensing* 55.7 (2017): 3965-3981.

[61] URL http://weegee.vision.ucmerced.edu/datasets/landuse.html

[62] Long, Yang, et al. "On creating benchmark dataset for aerial image interpretation: Reviews, guidances, and million-aid." *IEEE Journal of selected topics in applied earth observations and remote sensing* 14 (2021): 4205-4230.

[63] Singla, Ashutosh, Lin Yuan, and Touradj Ebrahimi. "Food/non-food image classification and food categorization using pre-trained googlenet model." *Proceedings of the 2nd International Workshop on Multimedia Assisted Dietary Management*. 2016.

[64] URL https://www.kaggle.com/datasets/ambityga/imagenet100

[65] Jeremy Howard. imagenette, 2019, URL https://github.com/fastai/imagenette

[66] He, Kaiming, et al. "Deep residual learning for image recognition." *Proceedings of the IEEE conference on computer vision and pattern recognition*. 2016.

[67] URL https://huggingface.co/google/vit-base-patch16-224

[68] URL https://huggingface.co/google/vit-large-patch16-224

[69] Levin, G., et al. "Terrapattern: open-ended, visual query-by-example for satellite imagery using deep learning." (2016).

[70] Simonyan, Karen, and Andrew Zisserman. "Very deep convolutional networks for large-scale image recognition." *arXiv preprint arXiv:1409.1556* (2014).

[71] https://whueducn-my.sharepoint.com/personal/longyang_whu_edu_cn/_layouts/15/onedrive.aspx?ga=1&id=%2Fpersonal%2Flongyang%5Fwhu%5Fedu%5Fcn%2FDocuments%2FMillion%2DAID%2FMillion%2DAID%2Ftrain

[72] Christie, Gordon, et al. "Functional map of the world." *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*. 2018.

Биографија аутора

Дајана Димитрић (дјевојачки Јованић) је рођена 21. 9. 1998. године. Основну школу „Бранко Ћопић“ је завршила 2013. године као носилац Вукове дипломе., а Гимназију „Свети Сава“ је завршила 2017. године, у Приједору. 2017. године је уписала Електротехнички факултет Универзитета у Бањој Луци, студијски програм Електроенергетика и аутоматика. Основне студије је заврпила 2021. године, одбраном рада под називом „Технике управљања погонским претварачем у електромоторним погонима“ и са просјечном оцјеном у току студија 9,49, те стекла звање дипломираног инжењера електротехнике. У току основног, средњег и високог образовања учествовала је и била награђивана на многим такмичењима. Други циклус студија на Електротехничком факултету Универзитета у Бањој Луци, студијски програм Електроенергетски и индустријски системи, је уписала 2021. године. Тренутно је запослена као асистент на Катедри за општу електротехнику Електротехничког факултета Универзитета у Бањој Луци. До сада има објављен један научни рад. Удата, мајка једне дјевојчице.

Прилог 2.

УНИВЕРЗИТЕТ У БАЊОЈ ЛУЦИ
ПОДАЦИ О АУТОРУ ОДБРАЊЕНОГ МАСТЕР/МАГИСТАРСКОГ РАДА

Име и презиме аутора мастер/магистарског рада: **Дајана Димитрић**

Датум, мјесто и држава рођења аутора: **21. 9. 1998.**

Назив завршеног факултета/Академије аутора и година дипломирања:

Електротехнички факултет Универзитета у Бањој Луци, 2021. година

Датум одбране завршног/дипломског рада аутора: **24. 9. 2021.**

Наслов завршног/дипломског рада аутора: **Технике управљања погонским претварачем у електромоторним погонима**

Академско звање коју је аутор стекао одбраном завршног/дипломског рада:

дипломирани инжењер електротехнике – 240 ECTS

Академско звање које је аутор стекао одбраном мастер/магистарског рада: **Мастер електротехнике – 300 ECTS – Електроенергетски и индустријски системи**

Назив факултета/Академије на коме је мастер/магистарски рад одбрањен:

Електротехнички факултет Универзитета у Бањој Луци

Наслов мастер/магистарског рада и датум одбране: **Детекција узорака који одступају од расподјеле у класификацији слика добијених даљинском детекцијом, 29. 2. 2024.**

Научна област мастер/магистарског рада према CERIF шифрарнику: **T 121**

Имена ментора и чланова комисије за одбрану мастер/магистарског рада:

проф. др Зденка Бабић, предсједник
проф. др Владимир Рисојевић, ментор
доц. др Славица Гајић, члан

У Бањој Луци, дана 20. 2. 2024. године

Декан

Прилог 3.

Изјава 1

ИЗЈАВА О АУТОРСТВУ

Изјављујем да је
мастер/магистарски рад

Наслов рада: Детекција узорака који одступају од расподјеле у класификацији слика добијених даљинском детекцијом

Наслов рада на енглеском језику: Out-of-distribution detection in remote sensing scene classification

- ☒ резултат сопственог истраживачког рада,
- ☒ да мастер/магистарски рад, у цјелини или у дијеловима, није био предложен за добијање било које дипломе према студијским програмима других високошколских установа,
- ☒ да су резултати коректно наведени и
- ☒ да нисам кршио/ла ауторска права и користио интелектуалну својину других лица.

У Бањој Луци 20. 2. 2024. године

Потпис кандидата

Зорана Ђекић

Изјава 2

Изјава којом се овлашћује Електротехнички факултет факултет Универзитета у Бањој Луци да мастер рад учини јавно доступним

Овлашћујем Електротехнички факултет Универзитета у Бањој Луци да мој мастер рад, под насловом

Детекција узорака који одступају од расподјеле у класификацији слика добијених даљинском детекцијом

који је моје ауторско дјело, учини јавно доступним.

Мастер рад са свим прилозима предала сам у електронском формату, погодном за трајно архивирање.

Мој мастер рад, похрањен у дигитални репозиторијум Универзитета у Бањој Луци, могу да користе сви који поштују одредбе садржане у одабраном типу лиценце Креативне заједнице *(Creative Commons)*, за коју сам се одлучио/ла.

1. Ауторство
2. Ауторство - некомерцијално
3. Ауторство - некомерцијално - без прераде
4. Ауторство - некомерцијално - дијелити под истим условима
5. Ауторство - без прераде
6. Ауторство - дијелити под истим условима

(Молимо да заокружите само једну од шест понуђених лиценци, кратак опис лиценци дат је на полеђини листа).

У Бањој Луци 20. 2. 2024. године

Потпис кандидата

Изјава 3

Изјава о идентичности штампане и електронске верзије мастер/магистарског рада

Име и презиме аутора: Дајана Димитрић

Наслов рада Детекција узорака који одступају од расподјеле у класификацији слика добијених даљинском детекцијом

Ментор: проф. др Владимир Рисојевић

Изјављујем да је штампана верзија мог мастер рада идентична електронској верзији коју сам предала за дигитални репозиторијум Универзитета у Бањој Луци.

У Бањој Луци 20. 2. 2024. године

Потпис кандидата

Дајана Димитрић

УНИВЕРЗИТЕТ У БАЊОЈ ЛУЦИ
ЕЛЕКТРОТЕХНИЧКИ ФАКУЛТЕТ
Патре 5
78000 Бања Лука

Др Зденка Бабић, редовни професор
Универзитет у Бањој Луци, Електротехнички факултет

Др Владимир Рисојевић, ванредни професор
Универзитет у Бањој Луци, Електротехнички факултет

Др Славица Гајић, доцент
Универзитет у Бањој Луци, Електротехнички факултет

НАУЧНО – НАСТАВНОМ ВИЈЕЋУ ЕЛЕКТРОТЕХНИЧКОГ ФАКУЛТЕТА УНИВЕРЗИТЕТА У БАЊОЈ ЛУЦИ

Одлуком Научно–наставног вијећа Електротехничког факултета Универзитета у Бањој Луци број 20/3.11-7/24 од 19.01.2024. године, именовани смо за чланове Комисије за завршни рад другог циклуса студија кандидата Дајане Димитрић, дипл. инж. ел., под називом „Детекција узорака који одступају од расподјеле у класификацији слика добијених даљинском детекцијом“. Након прегледа приложеног рада, подносимо сљедећи

ИЗВЈЕШТАЈ

1. БИОГРАФСКИ ПОДАЦИ КАНДИДАТА

Дајана Димитрић је рођена 21. 9. 1998. године. Основну школу „Бранко Ћопић“ је завршила 2013. године, а Гимназију „Свети Сава“ 2017. године, у Приједору. Исте године је уписала Електротехнички факултет Универзитета у Бањој Луци, студијски програм Електроенергетика и аутоматика. Основне студије је завршила 2021. године, одбраном рада под називом „Технике управљања погонским претварачем у електромоторним погонима“ и са просјечном оцјеном у току студија 9,49, те стекла звање дипломираног инжењера електротехнике. Други циклус студија на Електротехничком факултету Универзитета у Бањој Луци, студијски програм Електроенергетски и индустријски системи, је уписала 2021. године и положила све испите предвиђене планом и програмом. Тренутно је запослена као асистент на Катедри за општу електротехнику Електротехничког факултета Универзитета у Бањој Луци. До сада је објавила један рад:

1. Dajana Dimitrić, Vladimir Risojević, Mitar Simić, “Nearest Neighbor Based Out-of-Distribution Detection in Remote Sensing Scene Classification”, In *Proceedings of the 2023 22nd International Symposium INFOTEH-JAHORINA (INFOTEH)*, IEEE, ISSN: 2767-9470 (Online), 15-17 March 2023, pp. 1-6

2. ОСНОВНИ ПОДАЦИ О РАДУ

Завршни рад другог циклуса студија кандидата Дајане Димитрић, дипл. инж. електротехнике, под називом "Детекција узорака који одступају од расподјеле у класификацији слика добијених даљинском детекцијом" има обим од 79 страница. Садржи насловну страну на српском и енглеском језику, информације о ментору и раду на српском и енглеском језику, садржај, посвету, списак табела (укупно 12 табела), списак слика (укупно 42 слике) и списак скраћеница. Рад је организован у седам поглавља:

1. Увод
2. Класификација слика
3. Детекција слика које одступају од расподјеле
4. Материјал и методологија
5. Експериментални резултати и анализа
6. Закључак

На крају рада дати су преглед коришћене литературе (72 референце) и биографија кандидата.

У првој глави завршног рада је дефинисан проблем класификације слика добијених даљинском детекцијом и дат кратак преглед модела дубоког учења и приступа за њихово обучавање који достижу најбоље перформансе на поменутом проблему. Уочено је, међутим, да иако модели могу да остваре велику тачност класификације за тестне слике које припадају истој расподјели као слике на којима је модел обучен, питање препознавања узорака који не припадају ни једној од класа виђених током обучавања остаје неријешено. У овој глави је изложена и структура самог рада и наведени доприноси рада.

У другој глави је формално дефинисан проблем класификације слика, а затим су описани основни елементи модерних архитектура за класификацију слика базираних на неуронским мрежама, конкретно конволуционих неуронских мрежа и трансформатора за рачунарски вид.

У трећој глави је дефинисан проблем детекције слика које одступају од расподјеле и дат преглед најзначајнијих приступа за његово рјешавање, конкретно, приступи засновани на класификацији, процјени расподјеле вјероватноће података из тренинг скупа, приступи засновани на реконструкцији, те приступи засновани на мјерењу удаљености. Пошто су приступи засновани на удаљености од највећег интереса за овај завршни рад, та група приступа је детаљније обрађена. На основу анализе постојеће литературе, дат је преглед предности и мана метода заснованих на: алгоритму к-најближих сусједа, удаљености од најближег центроида, односу удаљености, Махаланобисовој удаљености и релативној Махаланобисовој удаљености. Такође је описана могућност побољшања перформанси детектора излагањем узорцима који одступају од расподјеле и предложена модификација метода детекције заснованог на рачунању односа удаљености који користи и примјере узорака који одступају од расподјеле.

Материјал и методологија експеримената извршених у оквиру практичног дијела рада су описани у четвртој глави. У експериментима су коришћене колекције слика добијених даљинском детекцијом: MLRSNet, NWPU-RESISC45, PatternNet, AID, UC Merced Land Use, MillionAID и три колекције слика из других домена: Food5K, ImageNet100 и Imagenette. За класификацију слика су коришћена два репрезентативна

модела: конволуциона неуронска мрежа, ResNet50, и два трансформатора за рачунарски вид, ViT-base-patch16-224 и ViT-large-patch16-224. Сви модели су фино подешени за класификацију слика добијених даљинском детекцијом на подскупу слика из MLRSNet колекције. Слике из другог дијела MLRSNet колекције, са дисјунктним скупом класа, коришћене су као тестне слике које одступају од расподјеле. Слично, за тестове су коришћени и други скупови слика добијених даљинском детекцијом, NWPU-RESISC45, PatternNet, те скупови слика из других домена. За оцјену перформанси детектора коришћене су двије метрике: удио лажно позитивних слика (False Positive Rate) и површина испод радне карактеристике пријемника (Area Under the Receiver Operating Characteristic).

Експериментални резултати су приказани у петој глави. Сви коришћени модели су претренирани на сликама свакодневних објеката и фино подешени за класификацију слика добијених даљинском детекцијом на подскупу класа из MLRSNet колекције. У експериментима су упоређене перформансе различитих метода заснованих на мјерењу удаљености и испитана је зависност перформанси метода детекције од броја тренинг узорака. Такође, пошто су коришћена три различита модела за класификацију слика, испитан је и утицај архитектуре основног класификатора на перформансе детектора. У наставку је имплементирана предложена модификација детектора заснованог на односу удаљености која укључује излагање детектора узорцима који одступају од расподјеле и анализиране су његове перформансе на истом скупу тестова који је коришћен у првој групи експеримената. Анализиран је утицај броја узорака који одступају од расподјеле као и политике њиховог бирања на перформансе детектора. Такође, размотрен је и утицај скупа из којег се бирају узорци ван расподјеле. Коначно, дате су смјернице за практичну имплементацију детектора.

У шестој глави су, на основу експерименталних резултата, изведени закључци. Изложене су главне предности и недостаци имплементираних детектора слика које одступају од расподјеле, те потенцијални правци за даље истраживање у предметној области.

3. АНАЛИЗА И НАЈВАЖНИЈИ ДОПРИНОСИ РАДА

Разматрајући завршни рад другог циклуса студија кандидата Дајане Димитрић, Комисија је закључила да својим садржајем, постигнутим резултатима и закључцима задовољава критеријуме који се постављају пред завршни рад другог циклуса студија. Рад у цјелини има добро систематизовану структуру и план излагања. Наслов рада је јасно формулисан, разумљив, прецизно описује предмет истраживања и у потпуности указује на садржај рада.

Свеобухватном теоријском анализом као и конкретним експерименталним радом, кандидат Дајана Димитрић је показала зрелост и способност да савлада и систематизује знања из једне истраживачке области.

Имајући у виду значај проблема детекције слика које одступају од расподјеле, те актуелност истраживања имплементације робусних модела за класификацију у области машинског учења, овај завршни рад обухвата актуелна истраживања и представља допринос стању у области.

Комисија констатује да је рад написан у складу са образложењем у пријави теме, као и да су остварени сви резултати који су били и планирани у образложењу пријаве теме:

А. Имплементација метода за детекцију узорака који одступају од расподјеле заснованих на мјерењу удаљености у простору обиљежја

У раду су имплементирани методи за детекцију узорака који одступају од расподјеле засновани на: алгоритму к-најближих сусједа, удаљености од најближег центроида, односу удаљености, Махаланобисовој удаљености и релативној Махаланобисовој удаљености. Имплементирани методи су тестирани на проблему детекције слика добијених даљинском детекцијом које одступају од расподјеле слика из скупа за обучавање. Као мјера перформанси кориштени су удио лажно позитивних слика (False Positive Rate) и површина испод радне карактеристике пријемника (Area Under the Receiver Operating Characteristic). Испитан је и утицај архитектуре основног класификатора на перформансе детектора. Експериментални резултати показују да метод заснован на релативној Махаланобисовој удаљености даје најбоље резултате у већини експеримената, али да перформансе метода заснованих на алгоритму к-најближих сусједа и односу удаљености не заостају много. Уочено је, међутим, да смањење броја тренинг узорака више негативно утиче на метод заснован на релативној Махаланобисовој удаљености него на метод заснован на алгоритму к-најближих сусједа. Тестови са различитим архитектурама основног класификатора су показали да конволуциона неуронска мрежа, ResNet-50, и трансформатор за рачунарски вид, ViT-large-patch16-224, резултују сличним перформансама детекције.

Б. Приједлог побољшања перформанси детектора излагањем узорцима који одступају од расподјеле

Предложена је модификација метода детекције заснованог на рачунању односа удаљености која користи и примјере узорака који одступају од расподјеле. Предложена модификација је имплементирана и тестирана је њена ефективност на колекцијама слика кориштеним у претходним експериментима. Уочено је да излагање детектора примјерима узорака који одступају од расподјеле побољшава перформансе детектора.

В. Приједлог начина избора узорака за побољшање перформанси излагањем детектора узорцима који одступају од распоцјеле

Узорци ван расподјеле треба да на што бољи начин моделују узорке који одступају од расподјеле и који се могу јавити у тестној фази. Притом, треба имати на уму да се у тренутку бирања не располаже било каквим информацијама о тестним узорцима, сем једном - да су слике које долазе на улаз детектора углавном из домена слика добијених даљинском детекцијом. Анализиране су три различите политике бирања као и њихове комбинације. Уочено је да се бољи резултати добијају ако су изабрани узорци што равномјерније распоређени у простору обиљежја, при чему се позиција узорка процјењује на основу његовог положаја у односу на положаје тренинг узорака. Поред тога, испоставило се да је боље бирати узорке чија детекција као узорака који одступају од расподјеле није тешка.

4. ЗАКЉУЧАК И ПРИЈЕДЛОГ

Комисија сматра да завршни рад другог циклуса студија под називом „Детекција узорака који одступају од расподјеле у класификацији слика добијених даљинском детекцијом“, кандидата Дајане Димитрић, дипл. инж. електротехнике, садржи све потребне елементе и резултате којима су остварени постављени циљеви истраживања, те предлаже Научно-наставном вијећу Електротехничког факултета Универзитета у Бањој Луци да усвоји извјештај Комисије и одобри заказивање јавне усмене одбране.

Бања Лука, 05.02.2024. године

Комисија:

Проф. др Зденка Бабић, предсједник

Проф. др Владимир Рисојевић, ментор

Доц. др Славица Гајић, члан