Зборник радова Факултета техничких наука, Нови Сад

UDK: 4.8
DOI: https://doi.org/10.24867/30BE35Gluhovic

ДЕТЕКЦИЈА КРОВНИХ ПОВРШИНА СА САТЕЛИТСКИХ СЛИКА УПОТРЕБОМ РАЧУНАРСКЕ ИНТЕЛИГЕНЦИЈЕ

DETECTION OF ROOF SURFACES FROM SATELLITE IMAGES USING ARTIFICITAL INTELLIGENCE

Михаило Глуховић, *Факултет техничких наука, Нови Сад*

Област – ЕЛЕКТРОТХНИКА И РАЧУНАРСТВО

Кратак садржај – *Циљ овог рада био је да се обави истраживање на пољу детекције и сегментације кровних површина са сателитских слика употребом различитих алгоритама базираних на употреби вештачке интелигенције. Рад описује поступак и проблеме креирања реалног скупа за обучавање модела, као и проблеме имплементације и тренирања различитих модела базираних на конволуционим неуронским мрежама. На крају су приказани резултати предикције модела и дискусија добијених резултата.*

Кључне речи: *детекција, сегментација, обучавајући скуп, конволуционе неуронске мреже, YOLO, SegNet, U-Net, DeepLabv3*

Abstract – *The aim of this paper was to conduct an study in the field of detection and segmentation of roof surfaces from satellite images using various artificial intelligencebased algorithms. The paper describes the process and challenges of creating a real dataset for model training, as well as the issues related to the implementation and training of different models based on convolutional neural networks. Finally, the results of the model's predictions are presented, along with a discussion of the obtained results.*

Keywords: *detection, segmentation, training dataset, convolutional neural networks, YOLO, SegNet, U-Net, DeepLabv3*

1. УВОД

Основна идеја овог рада била је да се пронађу готови, већ обучени, модели за сегментацију кровова са сателитских слика. С' обзиром да потрага за истима није била успешна, за потребе рада било је неопходно тестирати што већи број потенцијално употребљивих модела. Конкретно, имплементирани су U-Net, DeepLabv3, YOLO и SegNet. Сви ови модели се заснивају на конволуционим неуронским мрежама [1].

За обуку ових модела неопходни су обучавајући скуп и одређена процесорска снага.

НАПОМЕНА:
Овај рад проистекао је из мастер рада чији ментор је био др Филип Кулић, ред. проф.

Карактеристике коришћене радне машине су биле: Intel(R) Core(TM) i5-8300H CPU @ 2.30GHz процесор са 8 Gb RAM-а и NVIDIA GeForce GTX 1050 графичку картицу са 4 Gb RAM-а.

Скупови за обучавање се састоје од улазне сателитске слике и резултујуће бинарне маске, на којој су белим пикселима означени кровови а црном позадина. Квалитет базе података за обуку има значајан утицај на коначне перформансе модела. Међутим, поред ограниченог приступа квалитетним моделима, постоји и недостатак одговарајућих и обимних база података, са изузетком неколико ресурса.

2. ОДАБИР МОДЕЛА

Главно ограничење пројекта била је мала процесорска снага која директно доприноси дужем времену обучавања модела.

Поред дужег времена обучавања снага рачунара директно утиче и на то које моделе је уопште могуће тренирати за прихватљиво време. Постоји велики број различитих модела заснованих на конволуционим неуронским мрежама намењеним за обраду слика (нпр. YOLO, U-Net, SegNet, DeepLab, DeepMask, TensorMask, PANet...). У оквиру овог рада успешно су имплементирани следећи модели.

U-Net [2] је конволуциона мрежа дизајнирана за семантичку сегментацију са енкодер-декодер архитектуром, где енкодер примењује конволуционе операције и max-pooling да извуче карактеристике са слика и смањи их, а декодер користи up-sampling и конкатенацију са претходним слојевима за реконструкцију детаља слике.

DeepLabv3 [3] је модел за семантичку сегментацију који комбинује atrous конволуцију и просторни пирамидални pooling (енгл. Atrous Spatial Pyramid Pooling - ASPP).

YOLO [4] је модел намењен за детекцију објеката у једном пролазу преко слике, третирајући овај задатак као проблем регресије. Мрежа дели слику на мрежу ћелија и предвиђа позиције објеката, граничне оквире и класе, омогућавајући брзу и тачну детекцију објеката.

SegNet [5] је мрежа за семантичку сегментацију са енкодер-декодер архитектуром која користи max-pooling индексе за реконструкцију мапа обележја, за разлику од U-Net који памти целе мапе обележја.

3. ПРИПРЕМА ОБУЧАВАЈУЋЕГ СКУПА

Једина јавно доступна база [6] погодна за овај рад била је намењена детекцију површине крова која је повољна за постављање соларних панела. Што значи да је сваки нпр. димњак рупа у масци. Поред тога слике су биле малих димензија и лоше резолуције (слика 3.1).

Слика 3.1 Пример из пронађене базе [1]

Резултати предикције U-Net модела обученог на наведеној бази били су незадовољавајући. Разлог томе је могла бити лоша база или лош модел, оба су узета у обзир.

За потребе спремања квалитетне реалне базе неопходно је било обезбедити сателитску слику високе резолуције и пронаћи алат у коме би се ручно означили кровови.

Уз помоћ python библиотека leafmap и samgeo омогућен је приступ разним сателитима од којих се на око издвојио Google-ов. Google Satellite нуди слике резолуције од 15 м до 15 цм по пикселу. На мапи је означена површина од 300 хектара која обухвата целу Сремску Каменицу и сачувана као tif датотека. Каменица има пуно различитих објеката за детектовати и слика изнад Каменице је резолуције 15 цм по пикселу што је чини идеалном за потребе обучавајућег скупа.

Слика је заузимала 2 Gb и потребно ју је било поделити на 50 мањих слика како би могли да се обрађују у алатима за маскирање. Употребљен алат за анотирање јесте superviesly [7] са својим „mask pen tool“ који омогућава прецизно обележавање кровова. Наизглед лак задатак наилази на доста препрека, неке кровове је лако означити, док су неки прекривени крошњама, сакривени у хладу или просто веома сличне боје као околни бетон и пут (слика 3.2).

Слика 3.2 Примери проблематичних ситуација

Креирана база имала је 50 слика димензија 2281x2625 пиксела и 50 одговарајућих маски са означеним крововима. Већина модела не може да обрађује толико велике слике због тога је било потребно поново их поделити, овог пута на слике димензија 640x640 пиксела. Да би димензије свих слика биле исте слике димензија различитих од 640x640 пиксела су избачени, тако да је база на крају садржала 303 слике и 303 одговарајуће маске.

У жељи да се база прошири за што мање времена, обзиром да је за једну од 50 почетних слика било потребно око 2 сата да се цела анотира, једина преостала опција јесу аугментације, односно трансформације и модификације већ постојећих слика. Две главне врсте аугментација су коришћене: аугментације геометрије, које су у овом случају битније и аугментације боје и контраста.

Аугментације геометрије су скалирање и ротирање. Скалирање је у овом случају непотребно пошто су формиране слике истих димензија. Ротирање за „нецеле“ углове, односно ротације за углове који нису облика $k * \pi/2$ доводе до појаве mirroring, односно ивице слике се пресликавају као у огледалу како би одржале жељене димензије (слика 3.3).

Слика 3.3 Примери mirroring-а

Из тог разлога коришћене су аугментације: целе ротације уз hotizontal flip (ротирање по вертикалној оси), vertical flip (ротирање по хоризонталној оси) и комбинација та два уз аугментације боје и контраста (слика 3.3).

Слика 3.3 Пример из проширеног скупа

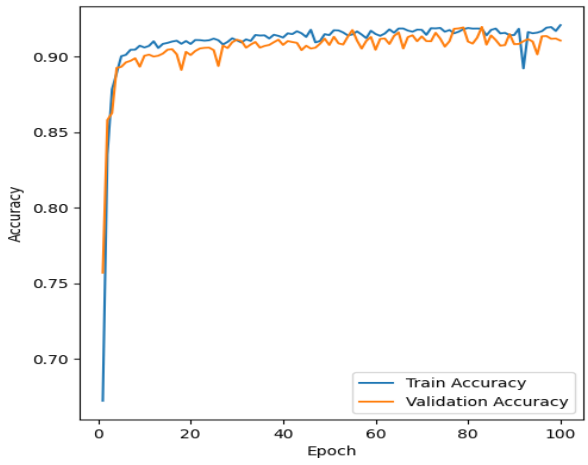
[1] Преузето из [6]

4.ОБУКА НЕУРОНСКИХ МРЕЖА

Тренирање модела је процес који захтева озбиљну количину рачунарске снаге, поготово у случају сегментације слика. Обрада графичких података (слика) за рачунар представља један од најзахтевнијих послова, јер треба да ради са великим датотекама које троше велику количину ресурса. Управо је то разлог зашто се за тренирање користе графичке картице које су оптимизоване за паралелну обраду података.

На расположивој машини која има Intel(R) Core(TM) i5-8300H CPU @ 2.30 GHz процесор са 8 Gb RAM-а и NVIDIA GeForce GTX 1050 графичку картицу са 4 Gb RAM-а, тренирање је било прилично захтевно. NVIDIA пружа софтвер CUDA [11] који је компатибилан са библиотеком pytorch помоћу које су имплементирани сви модели, и помоћу pytorch + cu118 библиотеке су тренирани. Тренирање је у одређеним случајевима било од 2 до 10 пута брже на графичкој картици него на процесору. Па шта је онда проблем? Проблем је што модел почне да се тренира и у неком моменту стане због немогућности алоцирања меморије на графичкој картици.
Један начин да се ово превазиђе јесте подешавање величина улазних слика, што је и урађено. Све слике, након подешавања су биле 640x640 пиксела и заузимале 1 Mb.

Тренирање има много параметара, од којих се у овом случају издвојио batch size, који одређује колико узорака скупа података пролази кроз мрежу у једном пролазу пре него што се изврши ажурирање тежина. Типична вредност овог параметра је од 32 па докле год машина може да изнесе. Због величине меморије коришћеног рачунара вредност овог параметра морала је бити или 2 или 4. То значи да ако 242 слике чине тренинг скуп, једна епоха тренирања ће имати 121(batch size = 2) или 61 (batch size = 4) итерацију. Мањи batch size утиче директно на време и квалитет обуке.



Слика 4.1 Промена тачности модела током обуке

Како се са слике 4.1 (тренинг SegNet модела) види модел почне да осцилује после 30те епохе. Узимајући то у обзир као и да је тренинг трајао скоро 2 дана, остали модели су тренирани на 30 до 40 епоха и трајањем у просеку 10-12 сати за U-Net, SegNet и DeepLabv3.

YOLO је модел који се најбрже тренира, али њега је и са најмањом вредности batch size параметра немогуће тренирати на графичкој картици. Архитектура YOLO алгоритма је дизајнирана за брзину, користи мање слојева и мањи број параметара од осталих модела. YOLO је уједно и једини модел који није било потребе тренирати од нуле већ је могуће преузимање и дотрениравање (енгл. fine-tune) модела “yolo8v-seg.pth”.Тренирање овог модела на процесору трајало је краће него тренирање осталих модела на графичкој картици. Из тог разлога YOLO је једини модел који је трениран на оригиналном скупу и на скупу проширеном аугментованим сликама. Међутим како YOLO у току тренинга сам врши аугментације улазних података, никаква разлика није примећена.

5. РЕЗУЛТАТИ

За потребе тестирања модела, на исти начин као са Каменицом, снимљен је Телеп, део Новог Сада који релативно личи на Каменицу. Слика обухвата око 100 хектара и као таква била је превелика за моделе. Слика је подељена на делове 640x640 пиксела, делови су провучени кроз модел и затим се од резултата реконструисала маска оригиналне слике.

На слици 5.1 приказана је слика и излазна бинарна маска из сваког модела. Стварна маска (енгл. ground turth) није направљена тако да није могуће дати бројеве за тачност предикције, али се са слика види да су модели успено сегментисали већину кровова.

Оригинална слика

Резултат предикције SegNet модела

Резултат предикције U-Net модела

Резултат предикције YOLO модела

Резултат предикције DeepLabv3 модела

Слика 5.1 Резултати предикција модела

6. ЗАКЉУЧАК

Успешно су имплементирани различити модели за сегментацију кровова са сателитских слика и направљена је реална база за тренирање тих модела. Сваки модел поседује одређене добре стране, али из резултата рада се издвајају SegNet и U-Net као лошије опције за решавање датог проблема. Оба модела представљају енкодер-декодер архитектуру где U-Net боље чува мапе обележја од SegNet-а, али као што се у резултатима види долазило је до погрешних класификација где је пут класификован као кров. Ове моделе је боље користити у ситуацијама где нема великих варијација у величинама и облицима објеката за сегментацију.

YOLO је показао одличне, очекиване, резултате обзиром да је то веома популаран модел који се користи за разне проблеме детекције и сегментације.

Коначно DeepLabv3 се показао као најбољи избор за конкретни проблем, са највише детекција и без погрешних класификација. DeepLabv3 је модел са најсложенијом архитектуром и најдужим временом обуке.

Правац даљег рада је повећање тачности детекције применом захтевнијих модела на јачем хардверу и са већим скуповима обуке.

7. ЛИТЕРАТУРА

[1] Ghosh, A., Sufian, A., Sultana, F., Chakrabarti, A., & De, D. (2017). *Fundamental concepts of convolutional neural network*. Department of Computer Science, University of Gour Banga, India; A.K. Choudhury School of Information Technology, University of Calcutta, India; Department of Computer Science & Engineering, M.A.K.A.U.T., India.

2] Ronneberger, O., Fischer, P., & Brox, T. (2015). *U-Net: Convolutional networks for biomedical image segmentation*. Computer Science Department and BIOSS Centre for Biological Signalling Studies, University of Freiburg, Germany.

[3] Chen, L.-C., Papandreou, G., Schroff, F., & Adam, H. (2017). *Rethinking atrous convolution for semantic image segmentation*. Google Inc.

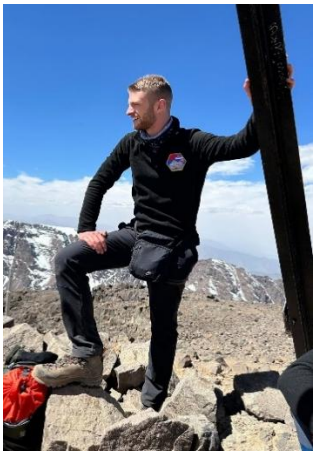
[4] Redmon, J., Divvala, S., Girshick, R., & Farhadi, A. (2016). *You only look once: Unified, real-time object detection*. University of Washington, Allen Institute for AI, Facebook AI Research.

[5] Badrinarayanan, V., Kendall, A., & Cipolla, R. (2017). *SegNet: A deep convolutional encoder-decoder architecture for image segmentation*. IEEE Transactions on Pattern Analysis and Machine Intelligence.

[6] https://drive.google.com/drive/folders/1v_VmwMEud1fcvdHRdgF8IWglBtXE8hbe

[7] https://app.supervisely.com/

Кратка биографија:



Михаило Глуховић рођен је 2001. године у Суботици. Основну школу завршио је у Новом Саду. Похађао је гимназију „J. J. Змај“. 2019. године уписује Факултет техничких наука, смер Рачунарство и аутоматика. Мастер академске студије наставља на смеру Аутоматика и управљање система. Тренутно је запослен на факултету као сарадник у настави. Контакт: gluhovic.mihailo@uns.ac.rs