Zbornik radova Fakulteta tehničkih nauka, Novi Sad

UDK: 004.9
DOI: https://doi.org/10.24867/23BE05Dakic

ДЕТЕКЦИЈА И СЕГМЕНТАЦИЈА КАЈАКАША УПОТРЕБОМ КОНВОЛУЦИОНИХ НЕУРОНСКИХ МРЕЖА

DETECTION AND SEGMENTATION OF KAYAKERS USING CONVOLUTIONAL NEURAL NETWORKS

Никола Дакић, *Факултет техничких наука, Нови Сад*

Област – СОФТВЕРСКО ИНЖЕЊЕРСТВО И ИНФОРМАЦИОНЕ ТЕХНОЛОГИЈЕ

Кратак садржај – *У раду је представљен систем за детекцију кајакаша на видео снимку. Систем врши парсирање видео записа и обрађује сваки фрејм. На сликама се детектују и сегментирају инстанце кајакаша. За решавање наведених задатака, коришћен је Mask R-CNN метод са конволуционом неуронском мрежом, ResNet101 архитектуре. Модел је направљен употребом технике преносног учења. Наведена техника преносног учења користи Mask R-CNN модел који је претходно обучен на Microsoft COCO скупу података. Као резултат система генерисан је излазни видео снимак на којем је детектован и сегментиран кајакаш.*

Кључне речи: *Детекција и сегментација објеката, кајакаш, Mask R-CNN, конволуционе неуронске мреже*

Abstract – *The paper presents a system for detecting kayakers on video. The system parses the video and processes each frame. In the images, instances of kayakers are detected and segmented. To solve the mentioned tasks, the Mask R-CNN method is used with ResNet101 architecture. The model was created using the transfer learning technique. The transfer learning technique uses a Mask R-CNN model previously trained on the Microsoft COCO dataset. As a result of the system, an output video was generated with the detected and segmented kayaker.*

Keywords: *Object detection and segmentation, kayakers, Mask R-CNN, convolutional neural networks*

1. УВОД

Детекција објеката је изозован задатак рачунарске визије, који је првобитно подразумевао предвиђање где се објекат тачно налази као и о ком типу објекта се ради. Нагли раст заједнице која се бавила овим проблемима, довела је до унапређења детекције објеката, који осим претходно споменута два задатка, подразумева и сегментацију инстанци. Сегментација инстанци је задатак који подразумева тачну детекцију свих објеката на слици, као и прецизну сегментацију сваке инстанце. Потпуније речено сегментација инстанци подразумева комбиновање класичних задатака детекције објеката у којем је циљ да се класификују појединачни објекти а затим да се они локализацију употребом граничних оквира, и класификовањем сваког пиксела у фиксни скуп категорија тако да се добију засебне инстанце објеката. Примена овог задатка је разнолика и неки од најчешћих употреба су: детекција објеката у малопродаји, аутономна вожња, детекција животиња у пољопривреди, откривање људи у безбедности, детекција возила у транспорту итд.

У овом раду приказана је имплементација фазе детекције и фазе сегментације појединачних објеката. Наведене фазе имплементиране су употребом сложених архитектура конволуционих неуронских мрежа.

За решавање проблема детекције у овом раду коришћен је Mask R-CNN [1] метод са конволуционом неуронском мрежом ResNet101 архитектуре. Модел је направљен употребом технике преносног учења. Наведена техника преносног учења користи Mask R-CNN модел који је претходно унапред обучен на Microsoft COCO [2] скупу података. Такав унапред обучен модел се затим користи као основа за генерисање модела над ручно прикупљеном скупу података. Финални модел представљен.у овом раду постиже *0.933 mAP* над тест скупом података.

2. ПРЕТХОДНА РЕШЕЊА

Компјутерски вид је мултидисциплинарно поље које је добило много пажње претходних година, наглим развојем конволуционих неуронских мрежа (енгл. Convolutional Neural Networks - CNN), а аутомобили који се сами возе (енгл. self-driving cars) заузимају централно место у овој области. Саставни део компјутерског вида је детекција објеката. Детекција објеката помаже у процени позе објекта, детекцији возила, надзору и сличним задацима. Разлика између алгоритама за детекцију објеката и алгоритама за класификацију је у томе што код алгоритама за детекцију покушавамо да нацртамо гранични оквир око објекта од интереса, како бисмо га лоцирали унутар слике. Такође на једној слици може постојати више граничних оквира који представљају различити објекте од интереса, што значи да не знамо унапред њихов тачан број. Стога главни разлог зашто се задаци детекције објеката не могу решити стандардом изградњом конволуционих неурноских мрежа, праћеним потпуно повезаним слојем, (енгл. fully connected layer) је тај што је дужина излазног слоја промењива. Број појављивања објеката од интереса није фиксан.

Наивни приступ решавању овог проблема би био да се из слике узму различити интересни региони и да се

НАПОМЕНА:
Овај рад проистекао је из мастер рада чији ментор је био проф. др Александар Ковачевић.

затим примени CNN ради класификације присуства објекта унутар региона. Проблем са овим приступом лежи у томе што објекти од интереса могу имати различите просторне локације унутар слике као и различите пропорције, што последично доводи до огромног броја региона које треба процесуирати. Стога су развијени алгоритми попут R-CNN [3] (Region Based Convolutional Neural Networks), YOLO [4] (You Only Look Once) и други, који ефикасно бирају и процесуирају регионе од интереса.

R-CNN метод превазилази проблем одабира огромног броја региона тако што уз помоћ селективне претраге издваја само 2000 региона из слике, и они се називају предложени региони (енгл. region proposals). Ови предложени региони се затим претварају у квадрате који се уносе у конволуциону неуронску мрежу која производи вектор карактеристика као излаз. CNN се понаша као екстрактор карактеристика а њен излаз се затим уноси у неки бинарни класификатор попут SVM (енгл. Support Vector Machine) ради класификације присуства објекта унутар предложених региона. Мане овог приступа су што и даље треба пуно времена за обуку неуронске мреже, јер се мора класификовати 2000 региона по слици, из чега следи да се не може применити у реалном времену.

Такође још једна значајна мана овог приступа јесте што је алгоритам селективне претраге фиксан алгоритам. Не постоји никакво учење у овој фази, што може довести до лошег генерисања самих предложених региона.

Fast R-CNN представља унапређење R-CNN методе, и то је постигнуто тако што се конволуционој неуронској мрежи шаље улазна слика уместо предложених региона, на основу које CNN генерише конволуциону мапу карактеристика. Разлог зашто је Fast R-CNN бржи од R-CNN методе је тај што се не мора сваки пут прослеђивати 2000 предложених региона конволуционој неуронској мрежи. Уместо тога, операције конволуције се ради само једном по слици и из ње се генерише мапа карактеристика.

Faster R-CNN представља унапређење Fast R-CNN методе. Унапређење је постигнуто тако што се избацује селективна претрага као начин генерисање предложених региона, јер она представља спор и дуготрајан процес који утиче на перформансе мреже. Уместо ње, користи се засебна мрежа за предвиђање предложених региона.

YOLO алгоритам за детекцију објеката представља алгоритам који се разликује од претходно наведених алгоритама. Сви претходних алгоритми користе регионе да локализују објекат унутар слике, и не гледају комплетну слику већ само делове слике који имају велику вероватноћу да садрже објекат. Код YOLO алгоритма, једна конволуциона неуронска мрежа предвиђа граничне оквире и њихове вероватноће да садрже циљни објекат. Из тог разлога YOLO алгоритам је доста бржи од претходно наведених алгоритама, ал и има потешкоћа са детекцијом малих објеката унутар слике.

Mask R-CNN представља унапређење Faster R-CNN методе, тако што поред предвиђања класе и оквира објекта, које Faster R-CNN даје као излаз, проширују излаз са додатном граном која предвиђа маску објекта.

3. МЕТОД

У наредним поглављима изложен је скуп података, начин креирања модела, начин евалуације решења и резултат система.

3.1 Скуп података

Скуп података коришћен у овом раду се састоји од 65 слика на којима се налазе кајакаши. Слике за потребе овог рада су скупљане ручно помоћу претраживача *Google, Bing*, и *Yandex*. Улазни видео снимак, који се обрађује у овом раду, снимљен је дроном из птичије перспективе, те су за потребе обучавања модела узимане углавном слике кајакаша из птичије перспективе.

С обзиром да је задатак Mask R-CNN модела да класификује објекте, предвиди њихове граничне оквире, као и да предвиди маске за детектоване објекте, неопходно је да ручно прикупљени скуп података који се користи за обучавање модела поседује координате свих полигона (полигон - многоугао који окружује сваког кајакаша на слици). За сваку слику која је коришћена приликом обучавања модела, уз помоћ *VIA* [5] софтвера, генерисан је адекватан *json* објекат. Сваки генерисани *json* објекат садржи информације које прецизно указују где се тачно налази кајакаш на слици. Сви *json* објекти су груписани и сачувани у склопу *annotations.json* фајла.

Прикупљени и анотирани скуп податак је подељен на скуп за тренирање и скуп за тестирање модела. Скуп за тренирање модела се састоји од 50 слика кајакаша и њихових анотација. Скуп за тестирање модела се састоји од 15 слика кајакаша и њихових анотација.

3.2 Креирање модела за детекцију и сегментацију кајакаша

Систем представљен у овом раду се базира на *Matterport Inc. Mask R-CNN* имплементацији [6]. Наведена имплементација даје основу за изградњу система детекције кајакаша.

Систем се може поделити у два модула:

- Модул за обраду видео снимка
- Модул за детекцију и сегментацију инстанце

3.2.1 Модул за обраду видео снимка

Модул за обраду видео снимка се бави учитавањем и парсирањем улазног видео снимка, као и генерисањем излазног. Модул је имплементиран употребом *OpenCV* библиотеке, и састоји се од *mark_kayaker* и *predict_kayaker* метода. Задатак *mark_kayaker* методе је да исцрта граничне оквире око сваке инстанце, текстуално прикаже називе детектованих класа, исцрта маску сваке инстанце и нумерички прикаже поузданост предикција.

Задатак *predict_kayaker* методе је да учита улазни видео снимак, издели улазни видео снимак на фрејмове, генерише и сачува излазни видео снимак.

3.2.2 Модул за детекцију и сегментацију инстанце

За решавање проблема детекције кајакаша и његове сегментације коришћен је Mask R-CNN модел и

техника преносног учења. Техника преносног учења користи Mask R-CNN модел претходно обучен на Microsoft COCO скупу података, за генерисање новог модела.

За кичму (енгл. backbone) Mask R-CNN модела која је задужена за екстракцију карактеристика објеката из слике, одабрана је ResNet-101 архитектура. За потребе сегментација инстанци коришћен је стохастички градијент и Адамов оптимизатор.

Модул за детекцију и сегментацију инстанце се састоји од метода *train* и *evaluate_model*. Задатак *train* методе је да покрене процес тренирања модела у зависност од одабраних параметара. Задатак *evaluate_model* методе је да евалуира колико су добре предикције модела.

4. ЕВАЛУАЦИЈА РЕШЕЊА И РЕЗУЛТАТИ

За евалуацију резултата модела коришћена је *mAP* (енгл. *mean Average Precision*) метрика. То је уобичајена метрика за проблем детекције. *mAP* се ослања на метрике прецизности (енгл. *precision*) и одзива (енгл. *recall*) који се рачунају по формули:

$$Precision = \frac{TP}{TP + FP}$$

$$Recall = \frac{TP}{TP + FN}$$

Где је *TP* број *True Positive* (тачно позитивних) детекција, *FP* број *False Positive* (нетачно позитивних) детекција, і *FN* број *False Negative* (нетачно негативних) детекција.

Прецизност одговара на питање колико од тога што је детектовано је релевантно, а одзив на питање колико од тога што је релевантно је детектовано.

Average Precision (AP) метрика односи се на једну класу и рачуна се тако што се узму све регије од интереса које је модел одредио за посматрану класу (укупно *n* регија). Регије се потом сортирају по сигурности тј. по *IoU* (енгл. *Intersection over Union*) детектоване регије (енгл. *RoI*) и истините лабеле (енгл. *GT- ground truth label*) по формули:

$$IoU = \frac{RoI \cap GT}{RoI \cup GT}.$$

На основу сортираних регија формира се график на ком x оса представља одзиве од 0, 0.1, 0.2 …, 1 (узимајући у обзир нулту регију) прву регију, прве две регије, …, свих *n* детектованих регија), а *y* оса прецизност детекција за одређену вредност одзива.

Нпр. за вредност одзива 0.5 рачуна се прецизност детекције првих 50% регија, где се детекција сматра успешном уколико је *IoU* већи од неке границе (нпр. 0.5). Од добијеног графика се потом формира нови график, који за сваку вредност одзива узима највећу вредност прецизности.

Коначно *Average Precision* се рачуна на основу графика као вредност површине испод зелене криве подељено са 11, а *mean Average Precision* као просечна вредност *Average Precision* метрике примењена над свим класама.

У табели 1. приказана је детаљна конфигурација параметара, коришћена приликом тренирања Mask R-CNN модела.

За потребе овог рада, тренирано је више модела са различитим параметрима. Модели и њихови резултати су приказани у табели 2.

Табела 1. *Приказ конфигурационих параметара који су коришћени приликом обучавања модела*

CONFIGURATIONS	
BACKBONE	resnet101
BACKBONE_STRIDES	[4, 8, 16, 32, 64]
BATCH_SIZE	2
BBOX_STD_DEV	[0.1 0.1 0.2 0.2]
COMPUTE_BACKBONE_SHAPE	None
DETECTION_MAX_INSTANCES	1
DETECTION_MIN_CONFIDENCE	0.7
DETECTION_NMS_THRESHOLD	0.3
FPN_CLASSIF_FC_LAYERS_SIZE	1024
GPU_COUNT	1
GRADIENT_CLIP_NORM	5.0
IMAGES_PER_GPU	2
IMAGE_CHANNEL_COUNT	3
IMAGE_MAX_DIM	512
IMAGE_META_SIZE	14
IMAGE_MIN_DIM	512
IMAGE_MIN_SCALE	0
IMAGE_RESIZE_MODE	square
IMAGE_SHAPE	[512 512 3]
LEARNING_MOMENTUM	0.9
LEARNING_RATE	0.001
LOSS_WEIGHTS	{'rpn_class_loss': 1.0, 'rpn_bbox_loss': 1.0, 'mrcnn_class_loss': 1.0, 'mrcnn_bbox_loss': 1.0, 'mrcnn_mask_loss': 1.0}
MASK_POOL_SIZE	14
MASK_SHAPE	[28, 28]
MAX_GT_INSTANCES	100
MEAN_PIXEL	[123.7 116.8 103.9]
MINI_MASK_SHAPE	(56, 56)
NAME	kayaker_cfg
NUM_CLASSES	2
POOL_SIZE	7
POST_NMS_ROIS_INFERENCE	1000
POST_NMS_ROIS_TRAINING	2000
PRE_NMS_LIMIT	6000
ROI_POSITIVE_RATIO	0.33
RPN_ANCHOR_RATIOS	[0.5, 1, 2]
RPN_ANCHOR_SCALES	(32, 64, 128, 256, 512)
RPN_ANCHOR_STRIDE	1
RPN_BBOX_STD_DEV	[0.1 0.1 0.2 0.2]
RPN_NMS_THRESHOLD	0.7
RPN_TRAIN_ANCHORS_PER_IMAGE	256
STEPS_PER_EPOCH	100
TOP_DOWN_PYRAMID_SIZE	256
TRAIN_BN	False
TRAIN_ROIS_PER_IMAGE	32
USE_MINI_MASK	True
USE_RPN_ROIS	True
VALIDATION_STEPS	5
WEIGHT_DECAY	0.0001

Поред наведене конфигурације, модели који су обучавани за потребе овог рада, разликују се по почетним тежинама које су одабране, слојевима који се обучавају и укупном броју епоха који су постављени

приликом обучавања модела. Приказ одабраних параметара модела се може видети у табели 2, као и њихови резултати на тренинг и тест скупу података.

Табела 2. *Приказ модела и њихових резултата*

Модел	Почетне тежине	Тренирани слојеви	Број епоха	Тренинг скуп података *mAP*	Тест скуп података *mAP*
M1	COCO	heads	5	0.844	0.800
M2	COCO	heads	10	0.824	0.867
M3	COCO	heads	15	0.844	0.867
M4	COCO	heads	20	0.844	0.800
M5	COCO	all	5	0.865	0.867
M6	COCO	all	10	0.888	0.933
M7	COCO	all	15	0.865	0.867
M8	COCO	all	20	0.888	0.933
M9	M8	heads	5	0.865	0.933
M10	M8	heads	10	0.885	0.933
M11	M8	heads	15	0.865	0.867
M12	M8	heads	20	0.885	0.933

На основу резултата из табеле 2, може се уочити да најслабије резултате имају модели *M1, M2, M3 и M4*. Приликом обучавања ових модела, коришћена је техника преносног учења, а за почетне тежине одабран је претходно трениран модел који је обучаван на *MS COCO* скупу података. Наведеним моделима су обучаване само њихове главе а под тим се подразумева да је трениран део за класификовање објекта, део за одређивање граничног оквира и део за генерисање маске.

За почетне тежине *M5, M6, M7 и M8* модела је такође одабран модел који је обучаван на *MS COCO* скупу података, само су код њих тренирани сви слојеви модела. Овако тренирани модели су дали боље резултате из разлога што су тренирани и слојеви *Backbone*, *RPN RoIAlign*, који су задужени за екстракцију карактеристика објеката из слика.

На основу претходних резултата може се закључити да је за потребе успешне детекције и сегментације кајакаша, приликом тренирања модела, неопходно поново истренирати и делове модела који су задужени за екстракцију карактеристика.

За почетне тежине модела *M9, M10, M11 и M12* одабран је модел *M8* који је претходно дао најбоље резултате предвиђања. Приликом обучавања наведених модела, тренирани су само делови *M8* модела који се баве класификацијом, детекцијом и сегментацијом инстанци, са циљем добијања бољих резултата, међутим њихово поновно тренирање није резултовало побољшањем резултата модела.

5. ЗАКЉУЧАК

У овом раду представљен је систем који се бави детекцијом и сегментацијом кајакаша на видео снимку. За потребе овог рада коришћен је ручно прикупљен скуп података, и *Matterport Inc. Mask R-CNN* модел конволуционе неуронске мреже.

Приликом прављења модела коришћена је метода преносног учења. Најбољи резултати су се показали код модела који су за основу користили MS COCO скуп података и код којих су тренирани свих слојеви Mask R-CNN модела.

Представљено решење се може побољшати проширењем постојећег скупа података као и додатним подешавањем хиперпараметара модела.

6. ЛИТЕРАТУРА

[1] Abdulla, W.: Mask R-CNN for object detection and instance segmentation on keras and tensorflow. https://github.com/matterport/Mask RCNN (2017), accessed 07-Oct-2020

[2] T. Lin, M. Maire, S. Belongie, J. Hays, P. Perona, D. Ramanan, P. Doll´ar, and C. L. Zitnick, “Microsoft COCO: Common objects in context,” in ECCV, 2014.

[3] Girshick, Ross & Donahue, Jeff & Darrell, Trevor & Malik, Jitendra. (2015). Region-Based Convolutional Networks for Accurate Object Detection and Segmentation. IEEE Transactions on Pattern Analysis and Machine Intelligence. 38. 1-1. 10.1109/TPAMI.2015.2437384.

[4] Handalage, Upulie & Kuganandamurthy, Lakshini. (2021). Real-Time Object Detection Using YOLO: A Review. 10.13140/RG.2.2.24367.66723.

[5] VGG Image Annotator (VIA) https://www.robots.ox.ac.uk/~vgg/software/via/via_demo.html [приступљено 24.11.2022.]

[6] Matterport Inc. Mask R-CNN https://github.com/matterport/Mask_RCNN [приступљено 24.11.2022.]

Кратка биографија:

Никола Дакић рођен је 1994. године у Новом Саду. Основне академске студије завршио је 2019. године на Факултету техничких наука у Новом Саду. Мастер рад је одбранио 2022. године из области Електротехнике и рачунарства, смер Софтверско инжењерство и информационе технологије - модул Интелигентни системи.
контакт: ndakic94@gmail.com