

Primerjava segmentacijskih metod globokega učenja za detekcijo ovir na vodi

Borja Bovcon, Matej Kristan

University of Ljubljana, Faculty of Computer and Information Science, Večna pot 113, Ljubljana, Slovenia

borja.bovcon@fri.uni-lj.si, matej.kristan@fri.uni-lj.si

Izvleček

Razvoj segmentacijskih metod globokega učenja za detekcijo ovir na vodi je v precejšnjem zaostanku v primerjavi z razvojem na sorodni domeni avtonomnih vozil (AGV). Do nedavnega je bil glavni razlog za to pomanjkanje ustreznih podatkovnih zbirk ter dejstvo, da metode razvite za AGV niso primerne za aplikacijo na vodno okolje zaradi domenskih specifik. Trenutno ni jasno katere arhitekture so najprimernejše za vodno domeno. Zato smo izbrali tri popularne arhitekture metod globoke semantične segmentacije (U-Net, PSPNet, DeepLab2), jih učili na podatkovni zbirki MaSTr1325 ter evalvirali njihovo delovanje na ločeni podatkovni zbirki Modd2. Eksperimentalna analiza je pokazala pomembnost globokega ogrodja arhitekture, ki omogoča učenje različnih izgledov vode. Kot zahteven problem so se izkazali artefakti v vodi (morska pena, odbleski in odsevi), ki prožijo lažne alarme in s tem motijo postopek avtonomne navigacije.

Ključne besede: detekcija ovir, globoko učenje, morski okoliš, primerjava metod, semantična segmentacija

Abstract

Development of deep segmentation methods for water surface obstacle detection lags far behind the development on a related domain of autonomous ground vehicles (AGVs). The main reasons is that much fewer curated training datasets exist and that methods developed for the AGVs rely on domain-specific assumptions and cannot be readily applied to the marine environment. It is currently unclear which architectures are most appropriate for marine environments. To address this, we selected three common state-of-the-art deep segmentation architectures (U-Net, PSPNet, DeepLab2), trained them on MaSTr1325 dataset and evaluated their performance on a separate test set Modd2. Experimental analysis shows a significant importance of the capacity of the deep backbone architecture, required to cope with the varying water appearance. Artefacts, such as foam fragments, glitter and reflections, present a leading cause of false positive detections which disturb autonomous navigation of the USV.

Keywords: benchmark, deep learning, marine environment, obstacle detection, semantic segmentation

1 UVOD

Razvoj morske robotike je privedel do vzpostavitve novega razreda avtonomnih plovil (USV), t.j. robotskih čolnov. Ta plovila so cenovno dostopna in prenosljiva za plovbo v plitvih vodah in ozkih marinah. Uporabljajo se predvsem za priobalno patroljiranje in oddaljeni pregled težko dostopnih okolji, ki lahko predstavljajo nevarnost za človeka. To zahteva visoko stopnjo avtonomije, ki je predvsem odvisna od pravočasnega odkrivanja in izogibanja bližnjim oviram ter plavajočim odpadkom. Za odkrivanje ovir ter boljše razumevanje okolice in njenega plovnega območja lahko uporabimo številne robotske senzorje

(RADAR [Almeida et al., 2009, Onunka and Bright, 2010], SONAR [Heidarsson and Sukhatme, 2011], LIDAR). Omejitve plovnosti ter napajanja USV-jev pa omejujejo namestitev potencialnih senzorjev. Zato se kamere kot lahki in informativni senzorji, vedno bolj uveljavljajo kot primerni mehanizmi za odkrivanje ovir.

Na sorodni domeni avtonomnih vozil (AGV) je bilo razvitih že veliko algoritmov računalniškega vida za detekcijo ovir s pomočjo kamere. Kot najobsebnije so se izkazale metode semantične segmentacije, ki temeljijo na globokem učenju [Alvarez et al., 2012b, Alvarez et al., 2012a, Levi et al., 2015, Olivei-

ra et al., 2016]. Metode globokega učenja zahtevajo ogromne količine natančnih in reprezentativnih učnih vzorcev. Domena AGV ima na voljo več takšnih ustreznih podatkovnih zbirk kot so KITTI [Geiger et al., 2012], Cityscapes [Cordts et al., 2016] in BDD100k [Yu et al., 2018]. Razvoj metod semantične segmentacije na področju USV-jev pa precej zaostaja. Zaradi pomanjkanja učnih podatkovnih zbirk pomorskega okoliša, je bilo izvedenih le nekaj poskusov uporabe globokega učenja. Lee et al. [Lee et al., 2018] so sestavili lastno podatkovno zbirko sedmih tipov ladij ter uporabili Faster R-CNN [Ren et al., 2015] za detekcijo le-teh. Vendar njihova metoda ne omogoča detekcije poljubnih ovir, ki niso bile videne v učni fazi. Cane et al. [Cane and Ferryman, 2018] so preizkusili delovanje treh metod globoke semantične segmentacije na pomorski domeni. Metode so učili na vsesplošni podatkovni zbirki ADE20k, ki pa vizualno ni reprezentativna za okolje, v katerem USV deluje. Rezultat je relativno slaba segmentacija.

V tem članku evalviramo tri popularne arhitekture metod globoke semantične segmentacije, ki so se izkazale uspešne na področju AGV-jev na podatkovni zbirki, ki je reprezentativna za robotska plovila.

2 METODE SEMANTIČNE SEGMENTACIJE

Izbrali smo tri popularne arhitekture metod globoke semantične segmentacije, ki dosegajo izjemne rezultate in predstavljajo osnovo velikemu številu naprednih metod semantične segmentacije:

- 1) *Arhitektura U-Net* [Ronneberger et al., 2015] je bila uspešno aplicirana na širok spekter različnih problemov. Sestavljata jo kodirnik, ki zajema kontekst, ter simetrični dekodirnik, ki poskrbi za natančno lokalizacijo. Korespondenčni nivoji kodirnika in dekodirnika so med seboj povezani, kar preprečuje problem izginjanja gradienta tekom učenja.
- 2) *Arhitektura PSPNet* [Zhao et al., 2017] je dosegla izjemne rezultate segmentacije urbanega okoliša. Ogradje arhitekture temelji na konceptu popularne mreže ResNet [He et al., 2016] z razširjenimi konvolucijami. Mreža vsebuje tudi modul piramidnega združevanja (*Pyramid-Pooling Module*) za povečavo vizualnega konteksta.
- 3) *Arhitektura DeepLab2* [Chen et al., 2018a] podobno temelji na ogradi ResNet [He et al., 2016] z razširjenimi konvolucijami. Kombinirana je z modulom *Atrous Spatial Pyramid Pooling* (ASPP), ki omogoča zajem konteksta na različnih skalah. Pogojno slu-

čajno polje (Conditional Random Field - CRF) poskrbi za izpopolnitev segmentacije. DeepLab2 implementira dve različici arhitekture — enoslojno in večslojno. Večslojna različica sprejme kot vhod sliko na različnih skalah ter vsako vzporedno in neodvisno segmentira. Rezultate segmentacij sešteje ter jih izpopolni z modulom CRF. V nadaljnjem bomo označili enoslojno različico DeepLab2 z DL2s, večslojno pa z DL2m.

3 EVALVACIJA

3.1 Podatkovno zbirka in podrobnosti implementacija

Izbrane arhitekture (Poglavje 2) smo učili na podatkovni zbirki MaSTr1325 [Bovcon et al., 2019], ki vključuje 1325 reprezentativnih, pomorskih slik, zajetih z USV-jem. Za zajem slik je bil uporabljen stereo sistem kamer *Vrmagic VRmMFC*, ki ga sestavljata dva senzorja CCD *Vrmagic VRmS-14/C-COB* z medsebojno oddaljenostjo 0,3 metra, lečo *Thorlabs MVL-4WA* z goriščno razdaljo 3,5 milimetra ter vidnim kotom 132,1°. Slike podatkovne zbirke so označene natančno do posameznega slikovnega elementa za tri semantične komponente — nebo, morje ter okoliš z ovirami. Pred učenjem smo izvedli povečevanje zbirke z naključnimi rotacijami (5, 15 stopinj), zrcaljenjem preko vertikalne osi ter ustreznim barvnim prenosom [Reinhard et al., 2001]. Po povečevanju je učna množica vsebovala 53000 učnih slik.

Za učenje metod smo uporabili cenilno funkcijo *softmax cross-entropy* ter optimizator *momentum*. Parameter začetne stopnje učenja smo nastavili na 1-4 ter nanj aplicirali funkcijo polinomskega upada s parametrom 0,9. Za inicializacijo ograda metod PSPNet in DeepLab2 smo uporabili uteži, pred naučene na podatkovni zbirki ImageNet [Deng et al., 2009]. Postopek finega učenja metod na podatkovni zbirki MaSTr1325 smo izvajali 13 epoh.

Evalvacijo smo opravili na ločeni podatkovni zbirki Modd2 [Bovcon et al., 2018], ki je bila zajeta z enakim USV-jem kot učna podatkovna zbirka. To je trenutno najzahtevnejša in največja testna podatkovna zbirka morskega okoliša. Vsebuje 28 raznolikih sekvenc z zahtevnimi vremenskimi pogoji kot so megla, nizko ležeče sonce, itd. Slike v podatkovni zbirki so sinhronizirane z meritvami senzorja inercij (IMU). Za evalvacijo kakovosti segmentacije smo uporabili dva tipa metrik: (i) klasične segmentacijske metrike

(*mean pixel accuracy, mean intersection-over-union in frequency-weighted intersection-over-union*) ter (ii) metriko za detekcijo ovir, kot so jo predstavili Kristan *et al.* [Kristan *et al.*, 2016].

Ovire v neposredni bližini USV-ja predstavljajo večjo nevarnost kot oddaljene. Za natančnejšo evaluacijo bližnje nevarnosti, smo s pomočjo meritev IMU v slikah testne množice določili t.i. kritični pas. To

je krožno območje z radijem 15 metrov, centrirano na trenutni lokaciji USV-ja. Radij krožnega območja smo določili pod predpostavko, da se USV giblje s povprečno hitrostjo 1.5 m/s. Za izogibanje oviram, ki so oddaljene več kot 15 metrov, je na voljo potemtakem vsaj deset sekund, kar ne predstavlja nevarnosti. Oviram, ki so bližje od 15 metrov, pa se moramo izmakniti v manj kot deset sekundah.

Tabela 1: Klasične segmentacijske metrike (levo) in metrike za detekcijo ovir (desno). Napaka približne ocenitve roba morja je označena z *ped* in merjena v pikslih, število pozitivnih detekcij s TP (true positives), število lažnih detekcij s FP (false positives), število zgrešenih ovir s FN (false negatives) ter splošna F-mera, merjena v procentih. Število detekcij TP je obratno sorazmerno s številom detekcij FN, F-mera pa je definirana kot harmonična sredina priklica ($TP/(TP+FN)$) in natančnosti ($TP/(TP+FP)$), izračunana z uporabo enačbe $F = 2TP / (2TP + FP + FN)$. V oglatih oklepajih so zapisane detekcije znotraj nevarnostnega območja.

Architecture	Mean PA	Mean IOU	fw-IOU	TP	FP	FN	F-mera
U-Net (Ronneberger <i>et al.</i> , 2015)	91,96	96,80	90,65	592 (208)	3706 (3337)	87 (43)	23,8 (11,0)
PSPNet (Zhao <i>et al.</i> , 2017)	90,98	94,50	88,22	346 (126)	54 (39)	333 (125)	64,1 (60,6)
DL2s (Cheb <i>et al.</i> , 2018a)	92,65	97,49	91,48	369 (167)	108 (36)	310 (84)	63,8 (73,6)
DL2m (Chen <i>et al.</i> , 2018b)	92,61	97,41	91,38	304 (147)	65 (14)	375 (104)	58,2 (71,4)

3.2 Analiza rezultatov

Rezultati, pridobljeni s klasičnimi segmentacijskimi metriki, se nahajajo v Tabeli 1. Opazimo, da DL2s doseže konsistentno najboljše rezultate v vseh treh segmentacijskih metrikah. Zgolj na podlagi rezultatov klasičnih segmentacijskih metrik ne moremo sklepati katera od metod bi bila najprimernejša za avtonomno navigacijo. Dodatno potrebno informacijo nam prikrbi metrika za detekcijo ovir (Tabela 1). Rob morja najboljše ocenita DL2s in DL2m, sledita jima pa U-Net ter PSPNet, pri čemer je napaka ocene roba morja metode PSPNet dvakrat večja. To je razvidno na spodnji vrstici Slike 1. Največje število pravih detekcij, ter s tem posledično tudi najmanjše število detekcij FN, doseže U-Net. Po natančnem vizualnem pregledu smo ugotovili, da je metoda U-Net zelo občutljiva na artefakte v vodi (Slika 1 srednja vrstica). Rezultat je veliko pravih detekcij, ampak še večje število lažnih detekcij, ki so posledica morske pene, valov, odbleskov ter odsevov. Najboljši rezultat F-mere doseže PSPNet, kateremu sledita DL2s in DL2m. Znotraj kritičnega območja se najboljše obnese DL2s, sledijo pa mu DL2m, PSPNet in U-Net. Podrobnejša analiza je pokazala, da večstopenjska različica DL2 zabeleži manj detekcij, saj se manjše ovire in artefakti izgubijo na skalah nižje ločljivosti (Slika 1 zgornja vrstica). To posledično privede do zmanjšanja števila detekcij FP in TP ter istočasno do povečanja števila detekcij FN,

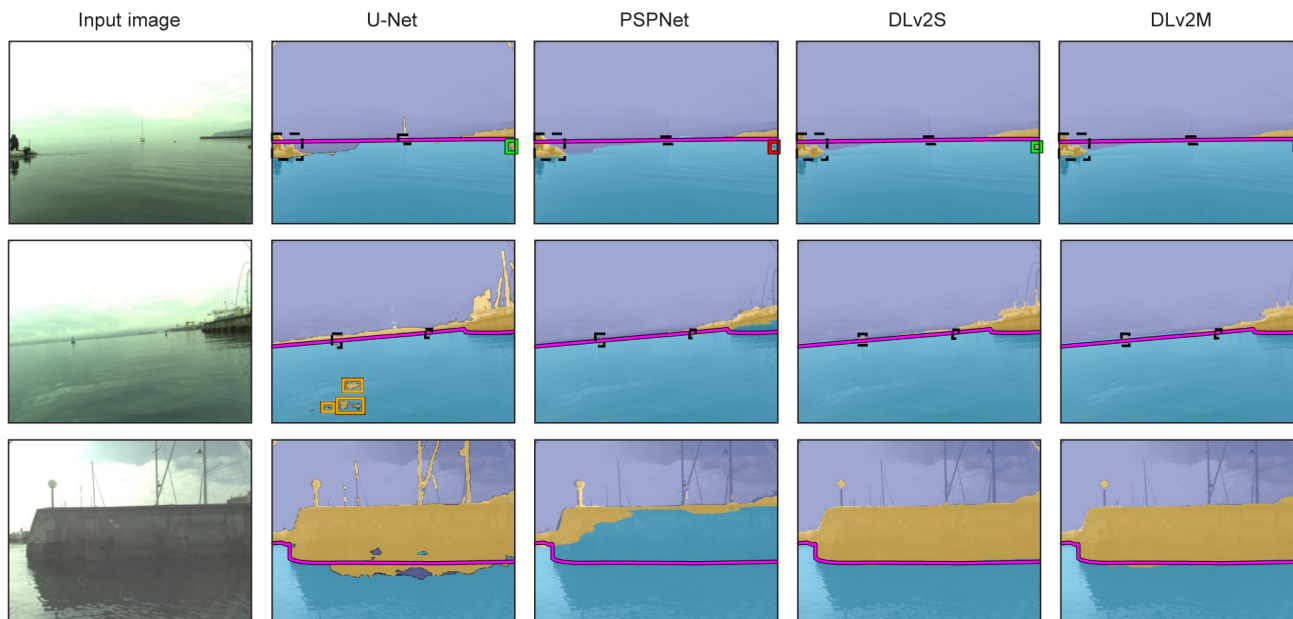
kar predstavlja nevarnost plovilu. Na podlagi izpostavljenih rezultatov doseže DL2s najboljše razmerje med natančnostjo ocene roba morja, številom pravih detekcij ter številom lažnih alarmov.

4 SKLEP

Predstavili smo primerjavo treh popularnih arhitektur za globoko semantično segmentacijo. Izbrane metode smo učili na podatkovni zbirki MaStr1325 morskega okoliša, njihovo kakovost detektiranja ovir pa testirali na ločeni podatkovni zbirki Modd2.

Rezultati so pokazali, da je problem segmentacije vode zelo težak, saj njen vizualni izgled zelo variira glede na vremenske pogoje in morsko dno. Za uspešno segmentacijo je potrebno globoko ogrodje arhitekture, kot je ResNet, ki omogoča mreži naučiti se različnih izgledov vode.

Empirični rezultati kažejo, da lažne detekcije, ki jih povzročajo artefakti (morska pena, odbleski in odsevi), predstavljajo zahteven problem za avtonomno navigacijo, saj plovilu prožijo lažne alarme, ki ovirajo nemoteno plovo. Nevarnejši problem predstavljajo zgrešene detekcije, saj lahko vodijo do trka in poškodbe avtonomnega plovila. V prihodnjem delu nameravamo nasloviti problem lažnih detekcij z vpeljavo časovne komponente, problem zgrešenih detekcij pa z ustreznimi spremembami arhitekture mreže.



Slika 1: Kvalitativna primerjava delovanja metod. Nebo, ovire in voda so označene s temno-plavo, rumeno ter cian barvo. Zlati standard roba morja je označen z roza črto, zlati standard ovir pa s črkanim pravokotnikom. Lažne detekcije označujejo rdeči pravokotniki, pravilne detekcije pa zeleni pravokotniki.

LITERATURA

- [1] [Almeida et al., 2009] Almeida, C., Franco, T., Ferreira, H., Martins, A., Santos, R., Almeida, J. M., Carvalho, J., and Silva, E. (2009). Radar based collision detection developments on USV ROAZ II. In *OCEANS - EU*, pages 1–6.
- [2] [Alvarez et al., 2012a] Alvarez, J. M., Gevers, T., LeCun, Y., and Lopez, A. M. (2012a). Road scene segmentation from a single image. In *European Conference on Computer Vision*, pages 376–389. Springer.
- [3] [Alvarez et al., 2012b] Alvarez, J. M., LeCun, Y., Gevers, T., and Lopez, A. M. (2012b). Semantic road segmentation via multi-scale ensembles of learned features. In *European Conference on Computer Vision*, pages 586–595. Springer.
- [4] [Bovcon et al., 2019] Bovcon, B., Muhovič, J., Perš, J., and Kristan, M. (2019). The mastr1325 dataset for training deep usv obstacle detection models. In *IROS*.
- [5] [Bovcon et al., 2018] Bovcon, B., Perš, J., Kristan, M., et al. (2018). Stereo obstacle detection for unmanned surface vehicles by IMU-assisted semantic segmentation. *Robotics and Autonomous Systems*, 104:1–13.
- [6] [Cane and Ferryman, 2018] Cane, T. and Ferryman, J. (2018). Evaluating deep semantic segmentation networks for object detection in maritime surveillance. In *2018 15th IEEE International Conference on Advanced Video and Signal Based Surveillance (AVSS)*, pages 1–6. IEEE.
- [7] [Chen et al., 2018a] Chen, L.-C., Papandreou, G., Kokkinos, I., Murphy, K., and Yuille, A. L. (2018a). Deeplab: Semantic image segmentation with deep convolutional nets, atrous convolution, and fully connected crfs. *IEEE TPAMI*, 40(4):834–848.
- [8] [Chen et al., 2018b] Chen, L.-C., Zhu, Y., Papandreou, G., Schroff, F., and Adam, H. (2018b). Encoder-decoder with atrous separable convolution for semantic image segmentation. *arXiv preprint arXiv:1802.02611*.
- [9] [Cordts et al., 2016] Cordts, M., Omran, M., Ramos, S., Rehfeld, T., Enzweiler, M., Benenson, R., Franke, U., Roth, S., and Schiele, B. (2016). The cityscapes dataset for semantic urban scene understanding. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 3213–3223.
- [10] [Deng et al., 2009] Deng, J., Dong, W., Socher, R., Li, L.-J., Li, K., and Fei-Fei, L. (2009). ImageNet: A large-scale hierarchical image database. In *Computer Vision and Pattern Recognition, 2009. CVPR 2009. IEEE Conference on*, pages 248–255. IEEE.
- [11] [Geiger et al., 2012] Geiger, A., Lenz, P., and Urtasun, R. (2012). Are we ready for autonomous driving? the kitti vision benchmark suite. In *2012 IEEE Conference on Computer Vision and Pattern Recognition*, pages 3354–3361. IEEE.
- [12] [He et al., 2016] He, K., Zhang, X., Ren, S., and Sun, J. (2016). Deep residual learning for image recognition. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 770–778.
- [13] [Heidarsson and Sukhatme, 2011] Heidarsson, H. K. and Sukhatme, G. S. (2011). Obstacle detection and avoidance for an autonomous surface vehicle using a profiling sonar. In *ICRA 2011*, pages 731–736.
- [14] [Kristan et al., 2016] Kristan, M., Kenk, V. S., Kovačič, S., and Perš, J. (2016). Fast image-based obstacle detection for unmanned surface vehicles. *IEEE TCYB*, 46(3):641–654.
- [15] [Lee et al., 2018] Lee, S.-J., Roh, M.-I., Lee, H.-W., Ha, J.-S., Woo, I.-G., et al. (2018). Image-based ship detection and classification for unmanned surface vehicle using real-time object detection neural networks. In *The 28th International Ocean and Polar Engineering Conference*. International Society of Offshore and Polar Engineers.
- [16] [Levi et al., 2015] Levi, D., Garnett, N., Fetaya, E., and Herzlyia, I. (2015). Stixelnet: A deep convolutional network for obstacle detection and road segmentation. In *BMVC*, pages 109–11.
- [17] [Oliveira et al., 2016] Oliveira, G. L., Burgard, W., and Brox, T. (2016). Efficient deep models for monocular road segmentati-

- on. In *Intelligent Robots and Systems (IROS), 2016 IEEE/RSJ International Conference on*, pages 4885–4891. IEEE.
- [18] [Onunka and Bright, 2010] Onunka, C. and Bright, G. (2010). Autonomous marine craft navigation: On the study of radar obstacle detection. In *ICCAR 2010*, pages 567–572.
- [19] [Reinhard et al., 2001] Reinhard, E., Adhikhmin, M., Gooch, B., and Shirley, P. (2001). Color transfer between images. *IEEE Computer graphics and applications*, 21(5):34–41.
- [20] [Ren et al., 2015] Ren, S., He, K., Girshick, R., and Sun, J. (2015). Faster R-CNN: Towards real-time object detection with region proposal networks. In *Advances in neural information processing systems*, pages 91–99.
- [21] [Ronneberger et al., 2015] Ronneberger, O., Fischer, P., and Brox, T. (2015). U-Net: Convolutional networks for biomedical image segmentation. In *MICCAI*, pages 234–241. Springer.
- [13] [Yu et al., 2018] Yu, F., Xian, W., Chen, Y., Liu, F., Liao, M., Madhavan, V., and Darrell, T. (2018). Bdd100k: A diverse driving video database with scalable annotation tooling. *arXiv preprint arXiv:1805.04687*.
- [14] [Zhao et al., 2017] Zhao, H., Shi, J., Qi, X., Wang, X., and Jia, J. (2017). Pyramid scene parsing network. In *EEE Conf. on Computer Vision and Pattern Recognition (CVPR)*, pages 2881–2890.

■

Borja Bovcon je magistriral iz računalništva in matematike na Univerzi v Ljubljani, Fakulteti za matematiko in fiziko ter Univerzi v Ljubljani, Fakulteti za računalništvo in informatiko leta 2017. Trenutno obiskuje doktorski študij na Univerzi v Ljubljani, Fakulteti za računalništvo in informatiko ter dela kot raziskovalec v Laboratoriju za umetne vizualne spoznavne sisteme (LUVSS). Njegovo raziskovalno področje obsega računalniški vid, razpoznavanje vzorcev, detekcijo ovir ter avtonomne sisteme.

■

Matej Kristan, izr. prof. dr., je doktoriral leta 2008 na Univerzi v Ljubljani, Fakulteti za elektrotehniko. Trenutno je član Laboratorija za umetne vizualne spoznavne sisteme (LUVSS) ter izredni profesor na Univerzi v Ljubljani, Fakulteti za računalništvo in informatiko. Njegovo raziskovalno področje obsega verjetnostne metode za računalniški vid s poudarkom na vizualnemu sledenju, dinamičnih modelih, sprotnemu učenju in strojnem vidu za mobilno robotiko.