ZAZNAVANJE PODVODNIH OBJEKTOV Z UPORABO GENERATIVNIH MODELOV

SANDRA RODRÍGUEZ DOMÍNGUEZ,[1,2] JANEZ PERŠ[2]

[1] Universidad Politécnica de Madrid, Madrid, Španija
sandra.rodriguez.dominguez@alumnos.upm.es
[2] Univerza v Ljubljani, Fakulteta za elektrotehniko, Ljubljana, Slovenija
sandra.rodriguez.dominguez@alumnos.upm.es, janez.pers@fe.uni-lj.si

V podvodnih okoljih predstavljajo spremenljiva osvetlitev, motnost vode in biološka raznolikost občutne ovire, zaradi katerih tradicionalne metode računalniškega vida ne delujejo dobro. Tudi učljive metode delujejo le, če uporabimo dovolj raznoliko zbirko podatkov, ki vsebuje vso pričakovano variabilnost podvodnega sveta. Zaradi narave samega podvodnega okolja pa je to lahko težavno, drago ali celo nemogoče, vsekakor pa zahteva veliko delovnih ur za označevanje objektov v učni množici. Ta problem smo naslovili z razvojem nove metodologije, ki na podlagi izjemno majhnega nabora sintetično generiranih slik objektov (10 v našem primeru) in večjega nabora ozadij brez objektov zanimanja (nekaj 100 slik) izdela učno bazo poljubne velikosti, primerno za učenje globokih metod zaznavanja objektov, ki ne zahteva nobenega ročnega označevanja. V našem primeru smo metodologijo uporabili za detekcijo ribe Acanthurus leucosternon, katere podobo za učenje smo generirali s pomočjo orodij DALL-E in Stable Diffusion. Naučen model smo preizkusili na realnih posnetkih tropskih koralnih grebenov z algoritmom zaznavanja objektov YoloV8, pri čemer dosežemo F1=0.6, ne da bi algoritem videl eno samo realistično sliko objekta v času učenja.

DOI
https://doi.org/10.18690/um.feri.1.2024.4

ISBN
978-961-286-837-6

Ključne besede:
globoke nevronske mreže,
detekcija objektov,
augmentacija,
generativni modeli ,
podvodni posnetki

Univerzitetna založba Univerze v Mariboru

DOI
https://doi.org/
10.18690/um.feri.1.2024.4

ISBN
978-961-286-837-6

DETECTION OF UNDERWATER OBJECTS USING GENERATIVE MODELS

SANDRA RODRÍGUEZ DOMÍNGUEZ,[1,2] JANEZ PERŠ[2]

[1] Universidad Politécnica de Madrid, Madrid, Spain
sandra.rodriguez.dominguez@alumnos.upm.es
[2] University of v Ljubljana, Faculty of Electrical Engineering, Ljubljana, Slovenia
sandra.rodriguez.dominguez@alumnos.upm.es, janez.pers@fe.uni-lj.si

Keywords:
deep neural networks, object detection, augmentation, generative models, underwater images

In underwater environments, variable lighting, water turbidity, and biodiversity present significant obstacles that cause traditional computer vision methods to perform poorly. Even learning-based methods only work if one uses a sufficiently diverse dataset that contains all the expected variability of the underwater world. However, due to the nature of the underwater environment itself, this can be difficult, expensive or even impossible, and it certainly requires many man-hours to annotate objects in the training dataset. We addressed this problem by developing a new methodology that, based on an extremely small set of synthetically generated object images (10 in our case) and a larger, diverse set of backgrounds without objects of interest (a few 100 images), produces a training dataset of arbitrary size, suitable for training deep object detection methods, without the need for any manual annotation. In our case, we used the methodology to detect the fish Acanthurus leucosternon, whose training images were generated using DALL-E and Stable Diffusion tools. We tested the learned training on real images of tropical coral reefs with the YoloV8 object detection algorithm, achieving F1=0.6 without the algorithm seeing a single realistic image of the object during learning.

University of Maribor Press

1 Uvod

Podvodna okolja, bogata z biotsko raznovrstnostjo in kompleksnostjo, predstavljajo pomembne izzive za zaznavanje in identifikacijo objektov. Metode računalniškega vida pod vodo se uporabljajo v različne namene, vključno z raziskovanjem morskih virov (Han, 2020), podvodno navigacijo (Xie, 2018), podvodnim videonadzorom (Shkurti, 2012), ocenjevanjem populacij morskih vrst, preučevanjem ekosistemov, ohranjanjem morskih vrst, ribolovom, odkrivanjem neeksplodiranih ubojnih sredstev pod vodo in podvodno arheologijo.

Zaradi pomanjkanja osvetlitve, motnosti vode in raznolikosti oblik, barv ter velikosti morskih organizmov tudi moderni algoritmi globokega učenja odpovedo, razen če jim damo na voljo zelo raznoliko učno bazo slik, ki vsebuje vse možne variacije v podvodnem okolju. Po drugi strani pa je podvodno okolje v primerjavi s kopnim bistveno bolj nevarno tako za ljudi kot za robote (Aldhaheri, 2022). Zajem raznolikih podatkov, ki bi omogočali obsežno učenje globokih modelov je torej drag, dolgotrajen in nevaren, če pa gre za redko videne živalske ali rastlinske vrste pa sploh nemogoč.

V tem članku smo se problema lotili z uporabo generativnih modelov za izdelavo slik na podlagi tekstovnih opisov, kot jih poznamo iz orodij DALL-E (Ramesh, 2021) in Stable Diffusion. Osnovna ideja našega pristopa je v tem, da uporabimo majhno bazo segmentiranih slik pridobljenih z enim od teh orodij, potem pa izvedemo ekstremno augmentacijo: slike objektov transformiramo tako geometrijsko kot barvno, jih postavimo na naključno mesto v naključno izbrani sliki ozadja brez objektov, in jim dodamo šum. Posebnost našega pristopa je, da zaradi specifike podvodnega okolja izvedemo *prenos svetlobnega vira* iz slike ozadja na objekt, kar zahteva obsežno, vendar avtomatsko predobdelavo slik ozadja, s katero ocenimo parametre osvetlitve.

2 Sorodna dela

V razvoju algoritmov za detekcijo objektov so konvolucijske nevronske mreže (CNN) pokazale izjemno učinkovitost pri prepoznavanju in lokalizaciji objektov na slikah. Uporaba nevronskih mrež se je razširila tudi na podvodne aplikacije (Villon, 2016; Yang, 2020) z uporabo mrež Yolo in Faster RCNN. Avtorji so se ukvarjali

tudi s prilagoditvijo algoritmov na podvodno okolje, vključno z uporabo generativnih modelov in tehnik obnavljanja slik (Wang, 2020; Chen, 2020; Liu, 2020).

3 Metodologija

Predlagana metoda sloni na naslednjih komponentah:

- Model globoke nevronske mreže za detekcijo ali segmentacijo objektov. V našem primeru smo uporabili mrežo YoloV8[1].
- Večje število (npr. nekaj 100) slik ozadja, ki pa morajo vsebovati vso pričakovano raznolikost morskega okolja. Anotacije niso potrebne, z izjemo potrditve, da na slikah ni objektov, ki jih hočemo detektirati.
- Manjše število (npr. 10) slik objekta, pridobljenih z generativnimi slikovnimi modeli (v našem primeru DALL-E 2 in Stable Diffusion), ter segmentacijske maske za vsako sliko. Ker je število slik majhno, je potreben vložek dela za pridobitev mask zanemarljiv.

3.1 Ocena svetlobnega vira slik ozadja

Ocena svetlobnega vira (angl. illuminant) je korak, pri katerem ocenimo, *kakšne transformacije* kanalov RGB slike so potrebne, da iz slike raznolikih, živih barv (slika A), dobimo sliko, ki je barvno in svetlostno čimbolj podobna sliki ozadja (B), ki jo imamo pred sabo. Z drugimi besedami, oceniti želimo parametre funkcije, ki nam iz slike A naredi sliko, ki je naj čimbolj barvno podobna B:

$$B \approx f(A; \theta_1, \theta_2, \ldots, \theta_{12}) \tag{1}$$

Določitev parametrov $\theta_1 \ldots \theta_{12}$ je optimizacijski problem, pri katerem optimiziramo kriterijsko funkcijo podobnosti med sliko $f(A; \theta_1, \theta_2, \ldots, \theta_{12})$ in sliko B.

$$\theta_1, \theta_2, .., \theta_{12} = \underset{\theta_1, \theta_2, .., \theta_{12}}{\operatorname{argmin}} d\big(B, f(A; \theta_1, \theta_2, .., \theta_{12})\big) \tag{2}$$

[1] https://docs.ultralytics.com/

pri čemer so parametri $\theta_1 ... \theta_{12}$ parametri splošne linearne preslikave za vsakega od kanalov slike A:

$$B_{\text{red}} = \theta_3 + \frac{\theta_4 - \theta_3}{\theta_2 - \theta_1}(A_{\text{red}} - \theta_1) \qquad B_{\text{green}} = \theta_5 + \frac{\theta_6 - \theta_5}{\theta_8 - \theta_7}\left(A_{\text{green}} - \theta_7\right) \quad (3)$$

$$B_{\text{blue}} = \theta_9 + \frac{\theta_{10} - \theta_9}{\theta_{12} - \theta_{11}}(A_{\text{blue}} - \theta_{11})$$

Razdalja d je kar razdalja Bhattacharaya med barvnima, 3D histogramoma slik A in B:

$$d_{Bhattacharyya}(h_A, h_B) = -ln\left(\sum_i \sqrt{h_A(i) \cdot h_B(i)}\right) \quad (4)$$

Optimizacijski postopek izvedemo za vsako sliko ozadja posebej, pri čemer za sliko A uporabimo eno sliko ozadja, ki je lepih, živih barv. V trenutni implementaciji uporabimo optimizacijo z metodo simpleksov, ki ne potrebuje gradientov kriterijske funkcije (Lagarias, 1998). Postopek je časovno potraten, vendar ga je treba za vsako sliko ozadja izvesti le enkrat, shranjene parametre $\theta_1 ... \theta_{12}$ pa lahko seveda uporabimo večkrat

3.2 Prenos svetlobnega vira

Ocenjene parametre $\theta_1, \theta_2, \ldots, \theta_{12}$ shranimo in jih lahko kadarkoli uporabimo za to, da na poljubno sliko živih barv *prenesemo svetlobni vir* z uporabo enačb (2).

Slika 1: Učinek prenosa svetlobnega vira
Vir: lasten

Primeri slik A, B ter transformirane slike $f(A; \theta_1, \theta_2, \ldots, \theta_{12})$ so prikazani na sliki 1. Od leve proti desni vidimo vhodno sliko živih barv, sliko ozadja, kjer je osvetlitev močno v prid modrozeleni barvi in na skrajni desni rezultat po prenosu osvetlitve iz druge slike na prvo sliko.

3.3 Generiranje slik objektov

Za generiranje slik objektov smo uporabili orodji DALL-E (verzijo 2) in Stable Diffusion. Optimalne poizvedbe se razlikujejo od orodja do orodja. DALL-E 2 deluje dobro s prostim tekstom, Stable diffusion pa potrebuje ključne besede. Za naš primer ribe Acanthurus leucosternon dobimo dobre rezultate orodja DALL-E 2 s preprosto poizvedbo »*powder blue tang on black background*«, ki nam da štiri slike ribe na črnem ozadju. Slike je potrebno segmentirati od črnega ozadja, kar lahko izvedemo relativno enostavno s poljubnim risarskim programom, ki omogoča izrezovanje poligonov. Ugotovili smo, da se avtomatska segmentacija (ozadje je črno) ne splača.

Slika 2: Naravna slika ribe Acanthurus leucosternon in trije generirani primerki.
Vir: lasten

Slika 2 prikazuje primer naravne slike ribe in tri sintetično generirane primerke. Po vrsti od leve zgoraj desno in navzdol: naravna slika, rezultat DALL-E 2, rezultat Stable Diffusion, rezultat DALL-E 3 (Microsoft Bing image creator).

3.4 Generiranje učne množice

Vsako sliko iz učne baze generiramo po naslednjem postopku:

1. Naključno izberemo sliko ozadja in preberemo shranjene parametre osvetlitve $\theta_1, \theta_2, \ldots, \theta_{12}$
2. Naključno izberemo sliko objekta, preberemo tudi binarno segmentacijsko masko.
3. Sliko objekta zmanjšamo na velikost, ki smo jo določili naključno.
4. Na sliko objekta prenesemo osvetlitev ozadja po formulah (3)
5. Z enakomerno verjetnostjo izberemo kot rotacije v območju od -45° do -45° in sliko objekta ter masko rotiramo.
6. Z verjetnostjo 1/8 sliko objekta prezrcalimo po višini in z verjetnostjo 1/2 po širini.
7. Naključno z verjetnostjo 1/2 izberemo ali bomo sliko »stisnili« po višini, enako po širini. Koeficient zmanjšanja širine ali višine izberemo naključno med 0.2 in 0.8.
8. Naključno izberemo amplitudo naključnega belega šuma in ga prištejemo sliki objekta.
9. Z Gaussovim filtrom zgladimo prej binarno sliko maske in z glajeno masko uteženo sliko objekta prilepimo na naključno izbrano lokacijo v sliki ozadja
10. Iz transformacij, ki smo jih izbrali v prejšnjih točkah generiramo sintentično anotacijo objekta na sliki.

4 Eksperimenti in rezultati

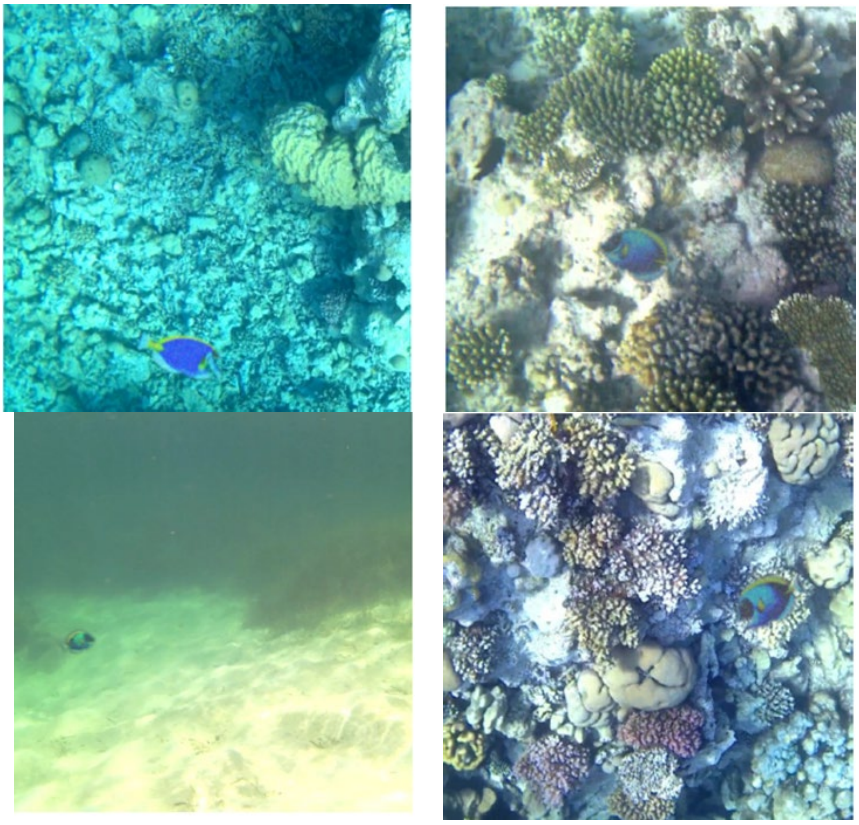
Za eksperiment smo uporabili zasebno podatkovno zbirko s posnetki tropskih morij iz treh različnih lokacij (Rdeče morje, Karibi, Indijski ocean).

4.1 Testna množica

V zbirki smo označili odseke, kjer ni bilo vidne ribe Acanthurus leucosternon in tiste odseke, kjer je bila riba prisotna. Na ta način smo pridobili veliko množico slik ozadja ter primerno veliko testno množico, ki smo jo ročno anotirali.

4.2 Učna množica

Po postopku iz poglavja 3.4 smo generirali učno množico 10.000 učnih slik s pripadajočimi anotacijami in jo razdelili na 6000 učnih slik in 4000 validacijskih slik. Edini označeni objekt na slikah je bila riba Acanthurus leucosternon. Ni nas motilo, če so bile na slikah ozadja tudi ribe drugih vrst. Nekaj učnih slik prikazuje slika 3.



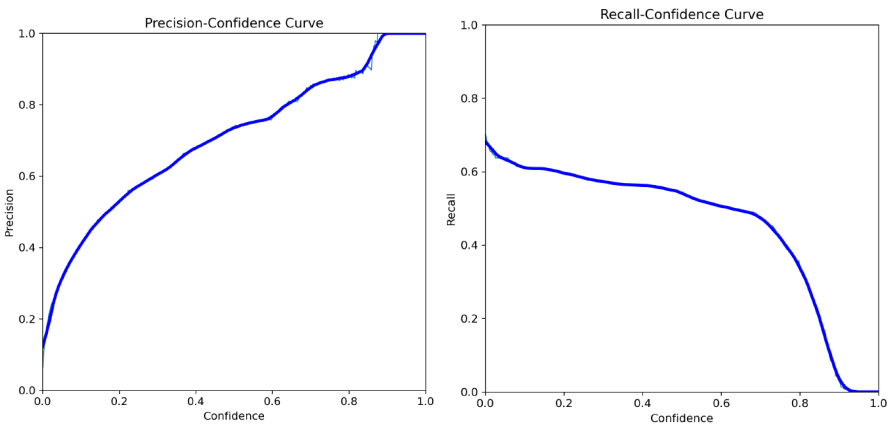
Slika 3: Štiri od slik iz sintetično generirane učne baze
Vir: lasten

4.3 Učenje detektorja objektov

Uporabili smo že prednaučeno arhitekturo YoloV8x, ki smo jo na učni bazi samo doučili z enorazrednimi anotacijami (razred »riba«, ki je predstavljal ribo Acanthurus leucosternon). Uporabili smo privzete parametre učenja.

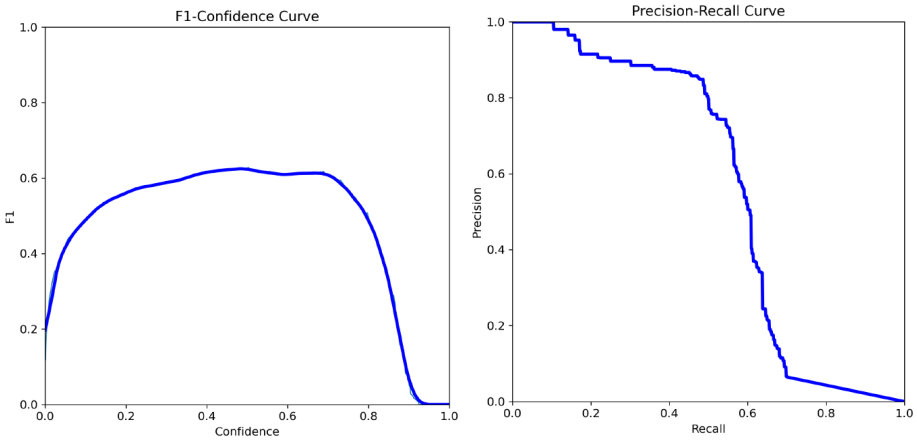
4.4 Rezultati

Uporabili smo knjižnico Ultralytics YoloV8 in vanjo vgrajena orodja za analizo uspešnosti detektorja objektov. Opazovali smo natančnost (ang. precision), priklic (ang. recall) in mero F1. Mere so odvisne od praga zaupanja (angl. confidence threshold), ki se giblje med 0 in 1, zato jih prikazujemo v obliki grafikonov (sliki 4 in 5) . Vse mere so bile izračunane na podlagi predpisanega praga Jaccardovega indeksa 0.5 (ang. Intersection over Union, IoU). Tako učenje kot testiranje smo izvajali na slikah velikosti 512x512 slikovnih elementov. Na večini od 1464 testnih slik ni bilo rib, tako da je bilo v testni množici 345 rib.



Slika 4: Rezultati: natančnost (precision) in priklic (recall) na realni testni množici, prikazana glede na prag zaupanja.

Vir: lasten



Slika 5: Rezultati: mera F1 in krivulja natančnost/priklic na realni testni množici
Vir: lasten

5 Zaključek

Pokazali smo, da so generativni slikovni modeli, tudi v obliki splošno in brezplačno dostopnih orodij, presenetljivo koristen pripomoček za generiranje velikih količin učnih podatkov. Naši rezultati na sicer omejenem problemu detekcije tropske ribe v okolju koralnih grebenov kažejo, da je mogoče doseči mero F1 ki presega 0.6 (ob priklicu in natančnosti 0.6), kar je izrazito dober rezultat, če upoštevamo, da algoritem med učenjem ni videl nobene realistične slike opazovane ribe, testiran pa je bil samo na realističnih slikah.

Viri in literatura

Han, M., Lyu, Z., Qiu, T., & Xu, M. (2020). A review on intelligence dehazing and color restoration for underwater images. IEEE Transactions on Systems, Man, and Cybernetics: Systems, 50, 1820-1832, doi: 10.1109/TSMC.2017.2788902.

Xie, K., Pan, W., & Xu, S. (2018). An underwater image enhancement algorithm for environment recognition and robot navigation. Robotics, 7, 14. doi: doi.org/10.3390/robotics7010014

Shkurti, F., Xu, A., Meghjani, M., Higuera, J. C. G., Girdhar, Y., Giguère, P., et al. (2012). Multi-domain monitoring of marine environments using a heterogeneous robot team. In Proc. IEEE/RSJ Int. Conf. Intell. Robots Syst, 1747-1753. doi: 10.1109/IROS.2012.6385685.

Aldhaheri, S., De Masi, G., Pairet, È., & Ardón, P. (2022). Underwater Robot Manipulation: Advances, Challenges and Prospective Ventures. OCEANS 2022 – Chennai doi:10.1109/OCEANSChennai45887.2022.9775489

Ramesh, A., Pavlov, M., Goh, G., Gray, S., Voss, C., Radford, A., Chen, M., & Sutskever, I. (2021). Zero-Shot Text-to-Image Generation. arXiv:2102.12092 [cs.CV]. doi: 10.48550/arXiv.2102.12092

Yang, H., Liu, P., Hu, Y., & Fu, J. (2020). Underwater object recognition based on yolov3. ICUS 2021, doi: 10.1109/ICUS52573.2021.9641489

Wang, Z., Liu, C., Wang, S., Tang, T., Tao, Y., Yang, C., Li, H., Liu, X., & Fan, X. (2020). UDD: An underwater open-sea farm object detection dataset for underwater robot picking. 2020.

Chen, X., Lu, Y., Wu, Z., Yu, J., & Wen, L. (2020). Reveal of Domain Effect: How Visual Restoration Contributes to Object Detection in Aquatic Scenes. arXiv:2003.01913 [cs.CV]. doi: 10.48550/arXiv.2003.01913

Liu, H., Song, P., & Ding, R. (2020). Towards domain generalization in underwater object detection. 2020. ICIP 2020. doi: 10.1109/ICIP40778.2020.9191364.

Lagarias, J. C., Reeds, J. A., Wright, M. H., & Wright, P. E. (1998). Convergence Properties of the Nelder-Mead Simplex Method in Low Dimensions. SIAM Journal of Optimization, 9(1), 112-147.