

# Prototipi značilik za adaptivno zaznavanje ovir na vodni površini

Lojze Žust, Matej Kristan

Fakulteta za računalništvo in informatiko, Univerza v Ljubljani, Večna pot 113, 1000 Ljubljana

E-pošta: {lojze.zust, matej.kristan}@fri.uni-lj.si

## Feature Prototypes For Adaptive Maritime Obstacle Detection<sup>1</sup>

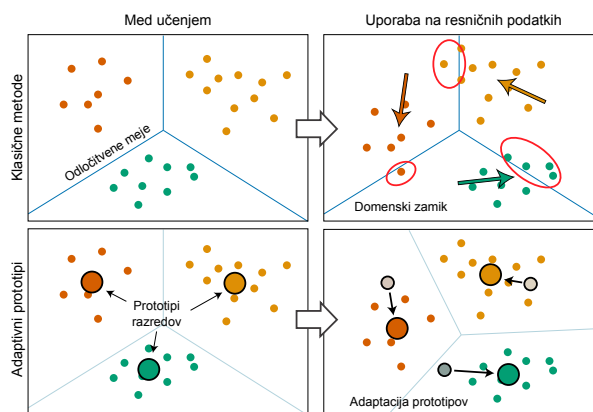
Unmanned surface vehicles (USV) rely on robust perception methods for obstacle detection. Current segmentation-based state-of-the-art methods lack the desired robustness and generalization capabilities required to adapt to new situations. To address this, we design WaSR-AD, a network with an explicit adaptation capability based on class prototypes. Initial prototypes are extracted during training and adapted during inference in an online fashion. The adapted prototypes are used to enrich the image features with additional adaptive context. Evaluation on the MODS benchmark reveals that such explicit adaptation of the prototypes significantly improves the detection performance, achieving 14% lower water segmentation error and 3.6% F1-score increase inside the critical 15m danger-zone area around the boat, with a negligible cost in inference time.

## 1 Uvod

Avtonomna plovila so hitro razvijajoče raziskovalno področje in imajo velik potencial na področju čezoceanskega transporta tovarov, nadzora obalnega okolja in avtomatskega pregledovanja umetnih konstrukcij. Njihova avtonomija je odvisna od zmožnosti zaznavanja, ki je še posebej zahtevno v zamejenih okoljih, kot priobalne vode, marine, mestni kanali in reke. Tu je lahko videz plovne površine (vode) zelo raznolik. Med drugim se na mirnih vodah pojavljajo zrcalni odsevi kopnega in ovir, barva in prosojnost vode se spreminjata z vremenom, pri plutju proti soncu pa je vidljivost zelo omejena, pojavljajo pa se odbleski na vodni površini. Prav tako je tudi videz ovir zelo raznolik, v grobem pa jih običajno razdelimo na statične (npr. obala, pomoli) in dinamične (npr. čolni, plavalci, boje).

Trenutno najboljši pristopi za zaznavanje ovir na vodni površini uporabljajo metode semantične segmentacije na osnovi globokega učenja [1, 8], ki prek učnih primerov zajamejo znatno variacijo v videzu vode in ovir. Žal pa te metode še vedno ne dosegajo zahtevane izjemno visoke robustnosti in prilagodljivosti, ki sta potrebni za praktične aplikacije avtonomnih čolnov.

<sup>1</sup>Ta projekt je bil finančno podprt s strani ARRS iz programa P2-0214 in projekta J2-2506

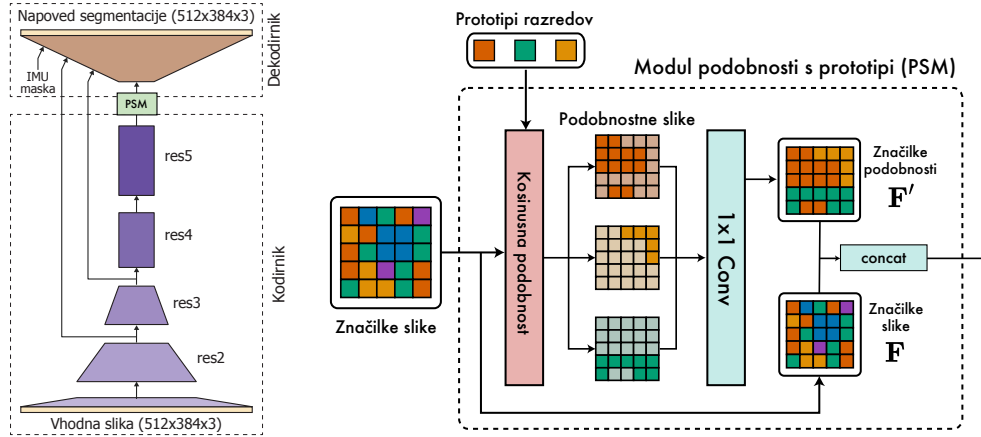


Slika 1: Med učenjem v prostoru značilnic nastanejo odločitvene meje (zgoraj levo). Pri uporabi na resničnih podatkih se lahko zaradi domenskega zamika porazdelitev značilk znatno spremeni, kar vodi do slabih napovedi (zgoraj desno). Da bi naslovili ta problem, naša metoda med učenjem oceni prototipe posameznih razredov (spodaj levo), med inferenco pa prototipe sproti adaptira, da se bolje prilagajajo novi distribuciji značilk (spodaj desno).

Ena od težav, s katerimi se soočajo trenutne metode, je vrzel med učnimi scenariji in uporabo v resničnih scenarijih. Učne slike lahko zajamejo le omejeno informacijo o možnem videzu ovir in vode. Pogosto se zgodi, da opazovani podatki v resničnem svetu padejo izven opazovane učne porazdelitve ali pa se nekoliko razlikujejo zaradi sprememb pogojev, kar lahko vodi v slabo natančnost metod (glej Sliko 1).

Metode za domensko adaptacijo skušajo ta problem reševati s poravnavo značilk ciljne domene, za boljše ujemanje z izvirno učno distribucijo značilk [10]. Te metode med učenjem običajno zahtevajo dostop do primerov ciljne domene, kar pa je pogosto nepraktično za resnične scenarije, saj je zajemanje celotne možne variacije resničnega sveta praktično neizvedljivo. Namesto tega želimo, da bi bile metode sposobne sprotne prilagajanja na nove situacije, tudi v času napovedovanja, ne zgolj v postopku učenja. V tem delu želimo narediti korak proti taki vrsti sprotne prilagodljivosti.

V ta namen predstavljamo novo mrežo za vodno domeno WaSR-AD, ki izvaja segmentacijo na podlagi prototipov razredov z izračunom podobnosti med značilkami



Slika 2: Arhitektura WaSR-AD (levo) z dodanim Modulom podobnosti s prototipi, oz. PSM (desno). PSM izračuna podobnosti med značilkami slike in vsakim prototipom razreda, kar proizvede podobnostne slike. Le te dodatno zakodira z uporabo  $1 \times 1$  konvolucije, nastale značilke pa združi z značilkami originalne slike ter jih pošlje v dekodirnik.

slike in prototipi razredov. Prototipe razreda ocenimo med učenjem, med inferenco pa jih lahko prilagajamo z upoštevanjem napovedi modela, da se le-ti bolje prilagajo opazovani porazdelitvi značilk (glej Sliko 1). Eksperimentalna evalvacija kaže, da tako prilagajanje znatno izboljša natančnost zaznavanja ovir na testni množici MODS [3], zlasti znotraj kritičnega 15-metrskega nevarnega območja neposredno pred čolnom (+3,2 F1), hkrati pa ima zanemarljiv vpliv na čas izvajanja metode.

## 2 Sorodna dela

### 2.1 Zaznavanje ovir na vodi

Za skupno obravnavo tako statičnih kot dinamičnih ovir v enotnem omrežju novejša metode [1, 11] predstavijo problem zaznavanja ovir prek problema segmentacije anomalij, kjer slikovne elemente slike pripišemo enemu izmed treh razredov: voda, nebo ali ovira. Vendar pa se zaradi zgoraj omenjenih specifik vodne domene metode splošne semantične segmentacije, razvite za avtonomne avtomobile, ne prenesejo dobro v vodno domeno [4, 2, 3], kljub nedavnemu povečanju števila zbirk podatkov za pomorsko segmentacijo [2, 8, 7]. Zaradi tega je bilo v zadnjih letih predstavljenih več novih arhitektur in modifikacij [1, 7, 12, 13], ki zmanjšujejo vrzel med domena avtonomnih vozil in plovil. Ti pristopi dosegajo obetavne rezultate na testnih zbirkah zaznavanja ovir v vodni domeni [3], vendar še vedno obstaja več težav, ki preprečujejo uporabo zaznavanja ovir v resničnih situacijah. Metodam manjka predvsem robustnosti na nove situacije in nove variacije ovir.

### 2.2 Prototipi značilk

Prototipi značilk so preprost način za opis skupkov podatkovnih točk v prostoru značilk. Iz tega razloga se pogosto uporabljajo pri problemih, kjer je podana informacija zelo omejena, kot je na primer segmentacija z nekaj primeri (angl. few-shot segmentation) [9, 6]. Te metode iz majhnega števila primerov izločijo prototipe razredov, ki se nato uporabijo za napovedovanje segmentacije preostalih slik prek računanja podobnosti med značilkami

slike in prototipi razreda. Podoben pristop je bil uporabljen tudi za shranjevanje prototipov značilk celotne zbirke podatkov [5], kar eksplicitno zagotavlja kontekst značilk, ki presega zgolj trenutno sliko, in posledično izboljša učinkovitost standardnih metod semantične segmentacije. Vsi ti pristopi nakazujejo moč prototipov značilk pri vnašanju dodatnega konteksta v napovedni proces mreže. V tem delu želimo to idejo peljati še korak dlje. Predlagamo postopek za sprotno prilagajanje naučenih prototipov med inferenco, kar omogoča spreminjanje konteksta prototipov na trenutno situacijo.

## 3 Adaptivna prototipna mreža

### 3.1 Osnovna struktura

Arhitektura WaSR-AD temelji na uveljavljenem modelu za segmentacijo ovir v vodni domeni WaSR [1] in sledi strukturi kodirnik-dekodirnik (glej Sliko 2). Kodirnik iz vhodne slike  $\mathbf{X} \in \mathbb{R}^{H \times W \times 3}$  najprej izloči visoko-nivojske značilke  $\mathbf{F} \in \mathbb{R}^{\frac{H}{8} \times \frac{W}{8} \times C}$ , kjer je  $C$  število kanalov značilk. Visokonivojske značilke nato vstopijo v modul podobnosti s prototipi (poglavje 3.2), ki značilkam slike doda kontekst iz adaptiranih prototipov razredov  $\mathbf{A}$  prek računanja podobnosti. Dobljene značilke  $\mathbf{F}' \in \mathbb{R}^{\frac{H}{8} \times \frac{W}{8} \times C}$  združimo z originalnimi značilkami slike  $\mathbf{F}$  ter jih posredujemo dekodirju, ki napove končno segmentacijo.

### 3.2 Modul podobnosti s prototipi

Naloga modula podobnosti s prototipi (angl. Prototype Similarity Module - PSM) je, da v značilke slike vnese kontekst iz prototipov. S tem poskrbimo, da so napovedi modela delno odvisne od prototipov razredov, ki jih želimo prilagajati med inferenco. Za razred  $i$  najprej izračunamo kosinusno podobnost značilk v  $\mathbf{F}$  z adaptiranim prototipom razreda  $\mathbf{A}_i \in \mathbb{R}^C$  in dobimo sliko podobnosti. Na lokaciji  $x$  je slika podobnosti določena kot

$$\mathbf{S}_i^x = \frac{\mathbf{F}^x \cdot \mathbf{A}_i}{\|\mathbf{F}^x\| \|\mathbf{A}_i\|}. \quad (1)$$

Nastale slike podobnosti  $\mathbf{S} \in \mathbb{R}^{\frac{H}{8} \times \frac{W}{8} \times M}$ , kjer je  $M$  število prototipov oz. razredov, pošljemo skozi učljivo

$1 \times 1$  konvolucijo, da se uskladimo število kanalov z originalnimi značilkami slike  $\mathbf{F}$  in dobimo značilke  $\mathbf{F}' \in \mathbb{R}^{\frac{H}{8} \times \frac{W}{8} \times C}$ . S tem uravnotežimo prispevek originalnih značilk in značilk, ki smo jih pogojili z adaptivnimi prototipi. Značilke  $\mathbf{F}$  in  $\mathbf{F}'$  spnemo po kanalih in jih posredujemo dekodirju.

### 3.3 Izbor prototipov razredov

Med postopkom učenja želimo na podlagi celotne zbirke podatkov čim bolj oceniti prototipe razredov  $\mathbf{P}$ , ki bodo uporabljeni kot izhodišče za adaptivne prototipe pri inferenci. Enostaven način za določanje prototipov iz podatkov je izračun povprečne značilke vseh pikslov v zbirki podatkov, ki pripadajo določenemu razredu. Toda tega pristopa ne moremo direktno uporabiti, saj za pravilno učenje dekodirnika že v postopku učenja potrebujemo prototipe, le-ti pa so prav tako odvisni od kodirnika, ki se med učenjem spreminja. Zaradi tega med postopkom učenja za ocenjevanje prototipov  $\mathbf{P}$  uporabimo naslednjo sprotno metodo.

V učnem koraku  $t$  najprej izračunamo povprečno značilko  $\mathbf{p}_{t,i}$  za vsak razred  $i$  iz trenutnega učnega paketa. Nato globalne prototipe  $\mathbf{P}$  posodobimo z uporabo eksponentnega drsečega povprečja (angl. exponential moving average - EMA)

$$\mathbf{P}_t = \alpha \mathbf{p}_t + (1 - \alpha) \mathbf{P}_{t-1}, \quad (2)$$

kjer je  $\mathbf{P}_{t-1}$  prejšnja globalna ocena prototipov razredov in  $\alpha$  faktor glajenja. Na koncu postopka učenja dobimo končno oceno prototipov razredov  $\mathbf{P}$ , ki jih uporabimo za osnovo adaptivnih prototipov med inferenco.

### 3.4 Adaptacija prototipov razredov

Med inferenco segmentacijske oznake niso na voljo. Namesto tega prototipe ocenimo z uporabo postopka gruččenja, ki temelji na metodi K povprečij in izmenjuje naslednja dva koraka: (i) korekcija prototipov na podlagi napovedane segmentacije, in (ii) napovedovanje segmentacije na podlagi izboljšanih prototipov.

V praksi ločeno sledimo množici adaptiranih prototipov  $\mathbf{a}$ , ki jih na začetku posnetka inicializiramo iz medučenjem-ocenjenih prototipov ( $\mathbf{a}_0 = \mathbf{P}$ ). Nato lahko za opazovano sliko v koraku inference  $t$  ocenimo ustrezne predvidene prototipe  $\hat{\mathbf{p}}_t$  kot uteženo povprečje značilk slike

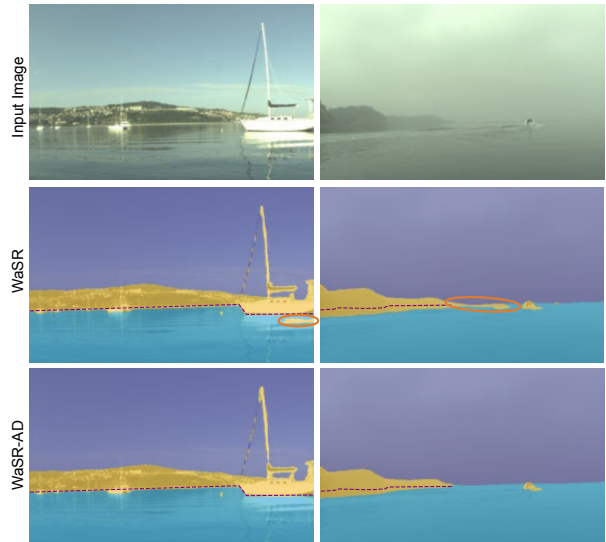
$$\hat{\mathbf{p}}_{t,i} = \frac{\sum \hat{y}_{t,i} \mathbf{F}_t}{\sum \hat{y}_{t,i}}, \quad (3)$$

kjer za utež uporabimo napovedano verjetnost razreda  $\hat{y}_{t,i}$  in kjer so  $\mathbf{F}_t$  značilke slike na koraku  $t$ . Z ocenjenimi prototipi lahko posodobimo adaptirane prototipe z uporabo eksponentnega drsečega povprečja

$$\mathbf{a}_t = \alpha_a \hat{\mathbf{p}}_t + (1 - \alpha_a) \mathbf{a}_{t-1}, \quad (4)$$

kjer je  $\alpha_a$  faktor glajenja. Da preprečimo, da bi se prilagodljivi prototipi preveč oddaljili od izhodišča  $\mathbf{P}$ , izrazimo končne adaptirane prototipe kot uteženo kombinacijo izhodiščnih prototipov  $\mathbf{P}$  in trenutnih adaptivnih prototipov  $\mathbf{a}_t$

$$\mathbf{A}_{t+1} = \beta \mathbf{a}_t + (1 - \beta) \mathbf{P}, \quad (5)$$



Slika 3: Kvalitativna primerjava WaSR in WaSR-AD. Adaptivna mreža WaSR-AD je bolj robustna na odseve in raznolike vremenske pogoje (npr. megla).

kjer  $\beta$  določa moč prilagajanja. Če je  $\beta = 0$ , se adaptacija ne izvaja in bodo med sklepanjem uporabljeni samo izhodiščni prototipi  $\mathbf{P}$ . Če pa je  $\beta = 1$ , prilagajanja prototipov ne omejujemo.

## 4 Eksperimenti

### 4.1 Podrobnosti implementacije

WaSR-AD izhaja iz arhitekture WaSR [1] in za kodirnik uporablja ResNet-101. Da ohranimo strukturo dekodirnika čim bolj nedotaknjeno, na konec enkoderja dodamo dodaten projekcijski sloj ( $1 \times 1$  konvolucija), ki prepolovi število kanalov značilk. Tako se po združevanju značilk  $\mathbf{F}$  in  $\mathbf{F}'$  v PSM število kanalov ujema s pričakovani dekodirnika.

Tudi pri postopku učenja smo sledili WaSR, s cenilno funkcijo ločevanja vode, hiperparametri, optimizacijsko metodo, spreminjanjem hitrosti učenja in bogatenjem podatkov. WaSR-AD smo učili 50 epoh na NVIDIA RTX 2080Ti z velikostjo paketa 4. Med učenjem je bila adaptacija onemogočena ( $\beta = 0$ ). Parametre EMA pa smo nastavili na  $\alpha = \alpha_a = 0,9$ . Vse modele v eksperimentih smo učili na zbirki MaSTr1325 [2].

### 4.2 Postopek evalvacije

WaSR-AD smo evalvirali na testni množici MODS [3] za detekcijo ovir za avtonomna plovila. MODS vsebuje približno 100 označenih sekvenc, ki so bile zajete v različnih pogojih. Sledimo evalvacijskemu protokolu MODS [3], ki vrednoti aspekte segmentacije, ki so pomembni za praktično navigacijo USV, in ločeno ocenjuje natančnost detekcije meje med statičnimi ovirami in vodo (*t.j.* vodni rob) in detekcijo dinamičnih ovir. Napaka segmentacije roba vode ( $\mu$ ) se izračuna kot povprečna kvadratna napaka (razdalja) med ocenjenim robom in pravim robom, medtem ko se detekcija dinamičnih ovir ovrednoti prek števila resnično pozitivnih (TP), lažno poziti-

Tabela 1: Rezultati na testni množici MODS za detekcijo ovir na vodi, merjeni v natančnosti segmentacije vodnega roba ( $\mu$ ) in F1 natančnosti detekcije ovir (skupaj in v nevarnem območju). Parameter  $\beta$  določa moč adaptacije.

|                           | $\mu$       | F1          | F1 <sub>d</sub> | FPS |
|---------------------------|-------------|-------------|-----------------|-----|
| WaSR                      | 15.2        | 93.5        | 87.6            | 4.4 |
| WaSR-AD ( $\beta = 0$ )   | 14.1        | 93.1        | 89.9            | 4.3 |
| WaSR-AD ( $\beta = 0.1$ ) | <b>13.0</b> | <b>93.9</b> | <b>90.8</b>     | 4.3 |

tivnih (FP) in lažno negativnih (FN) detekcij, kar povzame memo z oceno F1, natančnostjo (Pr) in priklicem (Re). Dinamična ovira se šteje kot pravilno zaznana (TP), če je zadosten delež znotraj očitnega okvirja ovire segmentiran kot ovira (več kot 50%), v nasprotnem primeru pa se ovira šteje kot nezaznana (FN). Napovedani skupki ovir izven očitanih okvirjev pa se štejejo za lažno pozitivne zaznave (FP). Učinkovitost zaznavanja poročamo po celotnem plovnem območju ter ločeno znotraj 15 m nevarnega območja od USV, kjer je zmogljivost detekcije kritična za preprečevanje trkov.

### 4.3 Evalvacija na MODS

Rezultati evalvacije na MODS so predstavljeni v Tabeli 1. WaSR-AD že z uporabo zgolj izhodiščnih prototipov razreda naučenih med postopkom učenja (brez adaptacije,  $\beta = 0$ ) zmanjša napako segmentacije vodnega roba za 1,1 točke in izboljša rezultat detekcije F1 znotraj nevarnega območja za 2,3 točke, vendar pa doseže 0,4 točke slabše rezultate v skupni oceni F1. Rezultat nakazuje, da že eksplicitna ekstrakcija prototipov razredov iz učne množice vnaša širši kontekst pri odločitvenem procesu in koristi generalizaciji modela, tudi brez prilagajanja, podobno kot je bilo pokazano v [5].

Vključitev sprotne adaptacije prototipov razreda med inferenco ( $\beta = 0, 1$ ) še znatno poveča natančnost detekcije in segmentacije po vseh metrikah (dodatno zmanjšanje  $\mu$  za 1,1, +0,8 skupen F1 in +0,9 F1 znotraj nevarnega območja). Kvalitativni rezultati na Sliki 3 kažejo, da izboljšave adaptivne mreže izvirajo predvsem iz zmanjšanja števila lažno pozitivnih detekcij na svetlih odbleskih in v prizorih z veliko megle, ki jih mreža ni videla med postopkom učenja.

Za praktično uporabo metod za detekcijo vodnih ovir je zelo pomembna tudi hitrost izvajanja. V ta namen primerjamo hitrost izvajanja metod WaSR in WaSR-AD, ki jo merimo v številu obdelanih slik v sekundi (FPS). Meritve hitrosti smo izvedli na osebнем računalniku z grafično kartico NVIDIA GTX 1660 Ti. Zaradi razmeroma preproste strukture predlaganega PSM modula in enostavnega koraka adaptacije prototipov ugotavljamo, da ima adaptacija zanemarljiv vpliv na hitrost sklepanja.

## 5 Zaključek

Predstavili smo WaSR-AD, adaptivno metodo detekcije ovir v vodni domeni z uporabo prototipov razredov. Že brez prilagajanja ekstrakcija prototipov razredov WaSR-AD izboljša natančnost zaznavanja na MODS. Poleg tega

smo pokazali, da lahko zelo preprosta shema adaptacije prototipov v nekaterih primerih izboljša generalizacijo in sposobnost zaznavanja.

Potrebne so nadaljnje raziskave glede vpliva moči adaptacije ( $\beta$ ) na zmogljivost modela. Poleg tega naše nedavno delo [12] nakazuje, da je za natančno zajemanje variacije razreda ovir potrebnih več prototipov. V našem nadaljnjem delu se bomo zato posvetili uporabi večjih prototipov za posamezen razred in naprednejšemu adaptacijskemu postopku z uporabo transformerjev.

## Literatura

- [1] Borja Bovcon and Matej Kristan. WaSR—A Water Segmentation and Refinement Maritime Obstacle Detection Network. *IEEE Transactions on Cybernetics*, pages 1–14, July 2021.
- [2] Borja Bovcon, Jon Muhovič, Janez Perš, and Matej Kristan. The MaSTr1325 dataset for training deep USV obstacle detection models. In *2019 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, pages 3431–3438, 2019.
- [3] Borja Bovcon, Jon Muhovič, Duško Vranac, Dean Mozetič, Janez Perš, and Matej Kristan. MODS – A USV-oriented object detection and obstacle segmentation benchmark. *IEEE Transactions on Intelligent Transportation Systems*, May 2021.
- [4] Tom Cane and James Ferryman. Evaluating deep semantic segmentation networks for object detection in maritime surveillance. In *Proceedings of AVSS 2018 - 2018 15th IEEE International Conference on Advanced Video and Signal-Based Surveillance*, 2019.
- [5] Zhenchao Jin, Tao Gong, Dongdong Yu, Qi Chu, Jian Wang, Changhu Wang, and Jie Shao. Mining Contextual Information Beyond Image for Semantic Segmentation. In *ICCV 2021*, August 2021.
- [6] Gen Li, Varun Jampani, Laura Sevilla-Lara, Deqing Sun, Jonghyun Kim, and Joongkyu Kim. Adaptive Prototype Learning and Allocation for Few-Shot Segmentation. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 8334–8343, April 2021.
- [7] Dalei Qiao, Guangzhong Liu, Wei Li, Taizhi Lyu, and Juan Zhang. Automated Full Scene Parsing for Marine ASVs Using Monocular Vision. *Journal of Intelligent & Robotic Systems*, 104(2):1–20, 2022.
- [8] L. Steccanella, D. D. Bloisi, A. Castellini, and A. Farinelli. Waterline and obstacle detection in images from low-cost autonomous boats for environmental monitoring. *Robotics and Autonomous Systems*, 124, 2020.
- [9] Boyuand Yang, Changand Liu, Bohaoand Li, Jianbinand Jiao, and Qixiang Ye. Prototype Mixture Models for Few-Shot Semantic Segmentation. In *European Conference on Computer Vision*, pages 763–778. Springer International Publishing, 2020.
- [10] Yanchao Yang and Stefano Soatto. FDA: Fourier domain adaptation for semantic segmentation. In *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, pages 4084–4094. IEEE Computer Society, 2020.
- [11] L Yao, D Kanoulas, Z Ji, and Y Liu. ShorelineNet: An Efficient Deep Learning Approach for Shoreline Semantic Segmentation for Unmanned Surface Vehicles. In *Proceedings of the 2021 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, 2021.
- [12] Lojze Žust and Matej Kristan. Learning Maritime Obstacle Detection from Weak Annotations by Scaffolding. In *Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision*, pages 955–964, January 2022.
- [13] Lojze Žust and Matej Kristan. Temporal Context for Robust Maritime Obstacle Detection. In *2022 IEEE/RJS International Conference on Intelligent Robots and Systems (IROS)*, 2022.