

Bulanık Sualtı Görüntülerinde Derin Öğrenme Tabanlı Balık Tespiti

Deep Learning-Based Fish Detection in Turbid Underwater Images

Tansel AKGÜL, Nurullah ÇALIK, Behçet Uğur TÖREYİN
İstanbul Teknik Üniversitesi, Bilişim Enstitüsü, İstanbul, Türkiye
{akgul16,ncalik,toreyin}@itu.edu.tr

Özetçe —Derin öğrenme modellerinin birçok alanda verdiği başarılı sonuçlar, su altı çalışmalarının zorlu şartlarında uğraşılacak problemler için çıkış kapısı olmuştur. Bu problemlerin başında, bulanık ve arkaplan gürültüsünün yoğun olduğu görüntülerde balık tespiti yapılması gelmektedir. Dolayısıyla bulanık ve arka plan gürültüsünün yoğun olduğu sularda balıkların tespit edilmesi, tanınması ve yollarının izlenmesi aşılması gereken önemli bir eşik olarak ortada durmaktadır. Buna yönelik olarak çalışmamızda, Kahramanmaraş Ceyhan Havzası Tekir Çayında bulunan Çataloluk HES'teki dikey yarık balık geçidinde iki farklı kamera ile videolar çekilmiştir. Ardından elde edilen videolar üzerinden 400 resimli balık tespitine yönelik özgün bir veri kümesi oluşturulmuştur. Bu görüntüler üzerinden, nesne tespitinde en ileri modeller arasında olan YOLO-V2, YOLO-V3, YOLO-V3 Tiny ve MobileNet-SSD ağları eğitilmiş bir sinir ağına ağırlıklarını alarak ve aynı alandaki veriler üzerinde eğitilen yeni bir modelin başlatma (fine-tuning) stratejisi ile eğitilmiş ve sonrasında kesinlik, duyarlılık ve ortalama hassasiyet başarımları üzerinden karşılaştırılarak MobileNet-SSD modelinin %50 eğitim verisi kullanıldığında %88.07 ortalama hassasiyet elde edilmiştir.

Anahtar Kelimeler—Su altı görüntüleri, balık tespiti, görüntü iyileştirme, derin öğrenme, yapay sinir ağları

Abstract—The successful results of the deep learning models in many areas have been the exit gateway to the problems faced in the challenging conditions of underwater studies. One of these problems is the detection of fish in images with a high turbid and background noise. Therefore, the detection of fish in turbid and background noisy water is an important threshold to be overcome to classify them and track their paths. In this study, videos were taken from the reservoir basin in Kahramanmaraş Ceyhan region with two different cameras. Then, a novel data set is presented which contains 400 images for the detection of fish in the wild. By using these data set, the state-of-the-art detection models, YOLO-V2, YOLO-V3, YOLO-V3 Tiny and MobileNet-SSD networks are trained with fine-tuning strategy, and then they are compared over the precision, recall and mean Average Precision (mAP) performances.

Keywords—Underwater images, fish detection, image improvement, deep learning, neural networks

I. GİRİŞ

Su altından elde edilen video görüntüleri, ışık zayıflaması, arka plan gürültüsü, renk kaybolması gibi zorlu etkenlere maruz kalmaktadır. Ayrıca, doğal ortamda bulunan çamur ve balçık, suyun bulanıklığına neden olmaktadır. Su altı koşulları, ışığın su altında ilerlerken kırılma, saçılmaya uğraması ve farklı renk tonlarının farklı miktarda soğurulması nedeniyle kara ortamına göre daha zorlu problemleri bünyesinde barındırır. Işık tayfında varolan farklı dalga boylarının suyun içinde yoğun miktarda bulunan parçacıklara çarpması, renk bozulmalarına sebep olmaktadır. Tüm bu problemlerin aşılarak balıkların tespit edilmesi, su altı çalışmalarının diğer konuları olan balıkların tanınması, popülasyon tespiti, balık rota kestirimi gibi önemli alanların ilk basamağını oluşturmaktadır.

Su altı videoları, balık yaşamını, davranışını ve su altı ekosistemini incelemek için kullanılır [1], [2]. Bu videolar oldukça uzundur ve insan gözü ile analiz edilmeleri zordur. Bu nedenle, videoların analizi için yapay zeka yöntemlerinin kullanılması elzemdir. Zorlu bir problem olan balık tespiti için Derin Öğrenme (DÖ) modellerinin kullanılması, bu alan için önemli katkılar ve fırsatlar sunar [3], [4]. Fakat, DÖ yöntemleri ekosistem-bağımsız değildir. Bu nedenle, başka bir ekosistem için eğitilen modeller, farklı yerlerde kullanıldığında istenilen sonuçları verememektedir [5]. Buradan hareketle, su altı çalışmalarında DÖ kullanmak isteyen araştırmacılar, kendi hazırladıkları veri kümeleri üzerinden modeller geliştirmişlerdir [6]. Bu çalışmada da Kahramanmaraş Ceyhan Havzası Tekir Çayında bulunan Çataloluk HES'teki dikey yarık balık geçidinden elde edilen video içeriklerinden seçilen 400 resim etiketlenerek balık tespitine yönelik özgün bir veri kümesi oluşturulmuştur. Nesne tespitinde, Evrişimsel Sinir Ağları (ESA) tabanlı yöntem kullanılarak model karşılaştırılmaları yapılmıştır.

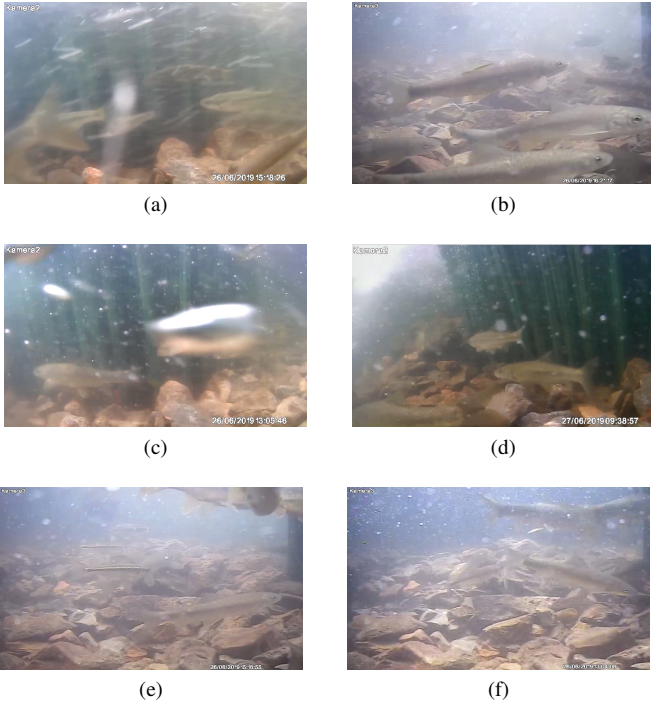
A. Literatür Özeti

Su altında ışık alanını simüle etmek birçok nedenle zordur. Işığın sadece su molekülleriyle değil aynı zamanda suda bulunan organik ve organik olmayan maddelerle girdiği etkileşim gibi birçok faktörler etkili olur [7]. Her bir bileşenin ışık konsantrasyonu (floresan veya Raman saçılması) bu etkileşimleri

büyük ölçüde etkiler ve çok farklı sonuçlara yol açar.

Su altı imgelerindeki bozukluğun temel nedenlerinden biri; ışığın dalga boyuna bağlı olarak nesnenin görüntüsünün bozulmasıdır. Özellikle derin sularda uzun dalga boyu (kırmızı) zayıflamaktadır. Daha kısa dalga boyları (mavi ve yeşil) daha kolay iletilmektedir [8]. Bulanık kıyılarda, sudaki bileşenler sudaki görünürlüğü azaltır ve yeşil renklerin geçişini daha kolay arttırmaktadır [8].

Su altı bulanıklığını giderme için birçok çalışma yapılmıştır. Bunlardan biri kontrastlı ve gri video çerçevesi dengelenmiş uygulama yöntemidir [9]. Bu yöntem; yeşil bulanık sularda iyi bir performans göstermiştir. Yöntem ile karanlık bölgelerde yeterli sonuç alınmadığı da gözlenmiştir. Su altı görüntülerini derinlik haritası kullanarak zayıflama katsayılarını tahmin etmek, bulanıklığı gidermek ve görüntüyü onararak yeniden işlemek için farklı yöntemler önerilmiştir [10], [11]. Su altı balık tespiti için genel kullanıma açık Fish4Knowledge [12] veri kümeleri kullanarak, Wenwei Xu [13] hidroelektrik santrallerinde DÖ yöntemlerinden YOLO modelini kullanarak su altı balık tespitinde analizler yapmışlardır. Malte Pedersen [14] deniz canlılarının tespiti ve sınıflandırma sistemleri için bilgisayarlı görme tabanlı yöntemler geliştirmişlerdir. Bu çalışmada daha çok deniz yıldızı, küçük balıklar, deniz anası gibi canlı tiplerini sınıflandırmak için benzer DÖ yöntemleri kullanmışlardır. Ülkemizin kendine has doğal su altı yaşam formuna uygun DÖ modelleri üzerinde özgün veri kümeleriyle deneysel çalışmalar yapılmaktadır.



Şekil 1. Su altı ortam görüntülerindeki zor imge örnekleri: (a),(b) Bulanık ve akıntılı su altı görüntülerinde farklı büyüklükteki balıklar (c) Balık görünümü su kabarcığı (d) Yapay sazlık bloklarının dinamik arka plan dokusu (e),(f) Parlak kabarcıklar ve taşlarla benzer renkte kamufle olmuş balıklar.

II. VERİ KÜMESİ TANITIMI VE NESNE TESPİT MODELLERİ

Özgün bir veri kümesi hazırlamak için Kahramanmaraş Ceyhan Havzası Tekir Çayında bulunan Çataloluk HES'teki dikey yarıklı balık geçidinden TÜBİTAK ARDEB 117M151 numaralı proje kapsamında alınan videolar kullanılmıştır. Videolardan oluşturulan görüntü veri kümesi üzerinde DÖ modelleri karşılaştırılmıştır. Veri kümesi hazırlanırken nesne tespiti açısından hayli zorlu ortamları barındıran görüntüler toplanmıştır (Şekil 1).

A. Ceyhan Havzası Tekir Çayı Çataloluk HES Video Veri Kümesi

Kahramanmaraş Ceyhan Havzası Tekir Çayında bulunan Çataloluk HES'te TÜBİTAK ARDEB 117M151 numaralı proje kapsamında oluşturulan dikey yarıklı balık geçidinde kameralar kullanılmıştır. Kameralar, balık davranışını incelemek için farklı açılarda ve pozlarda konumlandırılmıştır. Kameralardan alınan videolar farklı ışık koşulları, farklı bulanıklık, farklı su hızları gibi niteliklere göre 5 dakikalık, videolar haline getirildi. Farklı zamanlarda ve 2 farklı pozda Kamera2 ve Kamera3'den alınan her biri 5 dakikalık 53 videodan rastgele 400 adet video karesi seçilerek imge veri kümesi oluşturulmuştur. Bu veri kümesi %80-%20 ve %50-%50 eğitim ve test kümesi haline getirilmiştir. Su altı ortamında bulunan kayalar, bitkiler, balık kümelenmesi, hava kabarcığı ışık yansıması gibi zor koşulları içeren görüntüler (Şekil 1), balık tespitinde beklenmedik sonuçlara yol açabilmektedir. Su altı ortamının zor koşulları balık tespitinde başarıyı azaltan etkenler olarak düşünülmüştür.

B. Nesne Tespitinde Kullanılan Modeller

Çalışma kapsamında, YOLO gerçek zamanlı nesne algılama algoritmasından yararlanılmıştır. YOLO imgeye evrişimsel tek bir sınır ağı uygular ve bölgelere böler. Bölünmüş her bölge için sınırlayıcı kutuları ve olasılıkları tahmin eder. Nesne tespitinde sıklıkla kullanılan YOLO'nun türevleri; YOLO-V2, YOLO-V3, YOLO-V3 Tiny ve MobileNet-SSD (Single Shot Detector) modelleri, veri kümesi üzerinde sınanmıştır. Modellerin tasarımları, içerdikleri katman çeşitleri ve sayıları bakımından farklılıklar göstermektedir.

1) *YOLO-V2 ve V3:* YOLO 9000 olarak da isimlendirilen YOLO-V2, Faster R-CNN'ini [15] hem hız hem de performans açısından geçen bir ağ olarak önerilmiştir [16]. YOLO-V1'den farklı olarak, ağ içinde yığın normalleştirme (batch normalization) [17] kullanılması başarıma önemli katkılar sunmuştur. YOLO-V2 modelinin bir diğer yenilikçi yönü ise çapa (anchor) kutularının veri kümesi üzerinden k-ortalama ile kümelenmesi üzerinden seçilmesidir. Girdi olarak 448×448 boyutlarında görüntü kabul eden YOLO-V2 modeli, görüntüyü 13×13 'lük ızgaralara böler ve bu ızgaralar üzerinden sınıflandırma ve yerleştirme (localization) aynı ağ içinde yapılır. YOLO-V2, 19 evrişimsel ve 5 en-yüksek havuzlama katmanlarını içeren Darknet 19 modelini kullanmaktadır [16].

YOLO-V2 her ne kadar iyi sonuçlar verse de ağ tasarımında kalıntı (residual) katmanlar ve üst-örnekleme gibi yenilikçi özellikleri barındırmamaktadır. Bu eksiklikler üzerine YOLO-V3 modeli önerilmiştir [18]. V3 modelinin V2'ye göre en önemli farklılıkları arasında kalıntı bloklarının yanı

sıra Öznitelik Piramidi Ağı (Feature Pyramid Network) [19] yapısını kullanması gelmektedir. Bu sayede 3 farklı ölçekte sınıflandırma ve nesne tespiti yapılmaktadır. 9 çapa kutusunun kullanıldığı YOLO-V3 modeli de V2'de olduğu gibi eğitim kümesi üzerinden k-ortalama algoritmasını kullanarak kutuların büyüklüğüne karar vermektedir. Ayrıca sınıflandırma için softmax yerine her sınıf için lojistik sınıflandırıcı kullanılmıştır. YOLO-V3 mimari olarak Darknet 53 yapısını kullanmaktadır [18]. Kısıtlı ortamlar için kullanılan YOLO-V3 Tiny YOLO-V3'ün küçük bir sürümüdür. Daha hızlıdır ancak daha az doğrudur.

2) *MobileNet-SSD (Single Shot Detector) v2*: Gerçek zamanlı nesne tespitinde yaygın kullanılan bir diğer model ise SSD'dir [20]. Modele "*Single Shot*" denilmesinin nedeni, nesne tespitinin ve sınıflandırmanın tek bir ileri yönlü hesaplamada olmasıdır. Ayrıca *inception* tarzında iç içe bloklardan oluşan çoklu-kutu (*multibox*) üzerinden sınıflandırma yapabilmesi SSD'nin önemli özelliklerindendir. SSD ağı taban olarak, [21]'de önerilen ve transfer öğrenimi çalışmalarında yoğun olarak kullanılan VGG-16 ağını kullanmaktadır. Fakat, VGG-16 gerçek zamanlı nesne tespiti için elverişli olmadığı için [22] çalışmasında önerilen *MobileNet-V2* üzerine kurulu bir SSD modeli yine [22]'de önerilmiştir. MobileNet, ağı tasarımıyla kullanılan ayrıştırılabilir evrimsel katmanı (separable convolutional layer) ile hesaplama yükünü önemli derecede azaltırken, başarımlarını parametre açısından büyük ölçekli diğer ağlardan çok geride kalmamaktadır. Kayıp fonksiyonu olarak sınıflandırmada *softmax* kullanılırken yerelleştirme (*localization*) için Huber kaybı (*smooth ℓ_1*) kullanılmıştır.

III. BAŞARIM DEĞERLENDİRMESİ VE TEST SİSTEMATİĞİ

Yapılan çalışmada veri kümesi %80 eğitim ve %20 test; %50 eğitim ve %50 test oranlarında bir kez bölünmüştür. Modeller eğitilmiş bir sinir ağının ağırlıklarını alarak ve aynı alandaki veriler üzerinde eğitilen yeni bir modelin başlatma (fine-tuning) stratejisi ile probleme uyarlanmıştır. Tüm modeller Tablo I'de verilen parametreler doğrultusunda eğitilmiştir. YOLO ve MobileNet-SSD modellerin ön-eğitiminde (pre-trained) kullanılan veri kümeleri ise sırasıyla ImageNet [23] ve Pascal VOC [24]'dir. Ağların eğitiminde kullanılan donanım olarak Intel Core i7 işlemcili, 2GB GT 650M ram işlemci hızı bir bilgisayarda eğitilmiştir.

TABLO I. EĞİTİM İÇİN KULLANILAN ÜST-PARAMETRE AYARLARI

Üst-Parametre Ayarları	
Ağırlık En İyileme	Adam
İlk Öğrenme Oranı	0.001
Eğitim Adımı	10.000
Mini-Yığın Miktarı	5

A. Modellerin Farklı Eşik Değerlerine Göre İrdelenmesi

Modeller ile balık tespit sonuçları, ortalama hassasiyet (mAP) metriği üzerinden yapılmıştır. Veri kümesi olarak kullanılan video karelerinde tespit edilen balık sayısı, hassasiyet metriğine göre tahminlerinin doğruluk oranı belirlenmiştir.

TABLO II. MODELLERİN %50 EĞİTİM %50 TEST ORANI İÇİN MAP (%) BAŞARIMLARI

Model	Eşik Değerleri				
	0,10	0,30	0,50	0,70	0,90
YOLO-V2	78,61	75,12	73,82	70,68	54,96
YOLO-V3 (Tiny)	77,98	73,95	67,49	59,98	44,00
YOLO-V3	71,95	66,93	62,92	56,06	46,74
MobileNet-SSD	88,07	87,00	80,68	72,47	48,45

Çalışmalardan elde edilen sonuçlara göre, Tablo II'de 0.1 eşik değerinde %20-%80 eğitim ve test verisinde en düşük ortalama hassasiyetinin YOLO-V3 olduğu gözlemlenmiştir. YOLO-V3, YOLO-V2'den daha gelişmiş bir model olmasına rağmen daha düşük bir değer ölçülmüştür. Bunun nedeni, YOLO-V3'ün katman sayısının YOLO-V2'den daha fazla olması ve sınırlı veri kümesinde beklenen değerin daha düşük çıkması olarak açıklanabilir. Tablo III'de ise %80-%20 eğitim ve test sonuçlarında en iyi değeri YOLO-V3 Tiny, en düşük değeri YOLO-V3 olduğu gözlenmiştir. Bu sonuçlara göre, SSD-MobileNet modelinin diğer modellere göre daha iyi bir performans sergilediği sonucuna varılmıştır. Eğitim veri kümesinde kullanılan YOLO-V2, YOLO-V3'den daha az parametreye sahip olmasına rağmen tespit sonuçları daha iyidir. %50-%50 test eğitim sonuçlarından ise daha dengeli sonuçlar elde edilmiştir. YOLO-V2 ve YOLO-V3 Tiny değerleri 10.000 iterasyonda ve 0.1 eşik değerinde yakın sonuçlar ile karşılaştırılmıştır. YOLO-V3 ise %20-%80 veri kümesine göre nispeten bir iyileşme izlenmiştir. Tablo II ve Tablo III'de MobileNet-SSD model sonuçlarının YOLO-V2, YOLO-V3 Tiny, YOLO-V3 modellerine göre daha başarılı sonuçlar aldığı gözlenmiştir.

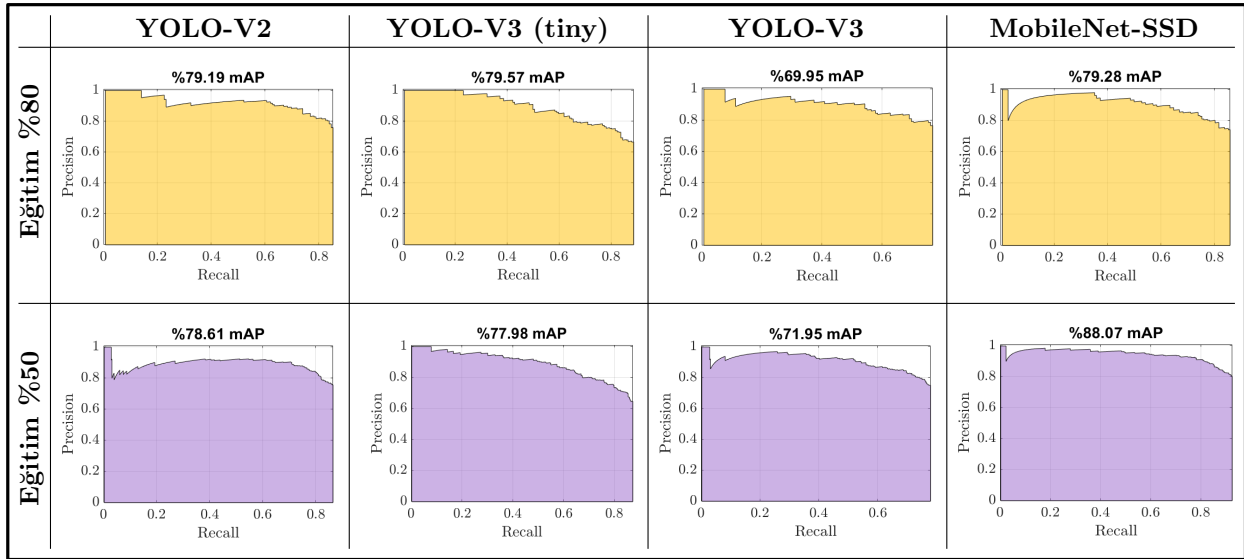
Şekil 2'de ise modellerin mAP değerlerinin detayları incelenmiştir. Genel olarak modellerin hassasiyet başarımları çok yüksektir. Bu modellerin balık tanıma noktasında tespit ettikleri nesnenin gerçekten balık olmasının yüksek olduğu anlamına gelmektedir. Diğer taraftan veri kümesi içinde arka plan ve diğer gürültüler altında bulunan balıkların tespiti noktasındaki yetersizliklerden dolayı mAP düşük çıkmaktadır.

TABLO III. MODELLERİN %80 EĞİTİM %20 TEST ORANI İÇİN MAP (%) BAŞARIMLARI

Model	Eşik Değerleri				
	0,10	0,30	0,50	0,70	0,90
YOLO-V2	79,19	77,54	74,08	69,92	55,24
YOLO-V3 (Tiny)	79,57	72,55	65,38	54,84	32,88
YOLO-V3	69,95	60,23	51,84	43,54	21,67
MobileNet-SSD	79,28	77,69	71,01	61,97	47,24

IV. SONUÇ

Bu çalışmada nesne tespitinde evrimsel sinir ağı (ESA) tabanlı 4 farklı model ile su altı balık tespiti eğitimi ve testleri yapılmıştır. YOLO-V2, YOLO-V3, YOLO-V3 Tiny ve MobileNet-SSD modelleri kullanılmıştır. Her biri 5 dakika süreli, 53 adet gerçek su altı yaşamdan alınmış videodan rastgele seçilen 400 adet video karesi alınmıştır. Ülkemiz su



Şekil 2. Nesne tespit yöntemlerinin mAP grafikleri üzerinden karşılaştırılması

altı yaşam formuna uygun gerçek imgelerden özgün bir etiketli veri kümesi oluşturulmuştur. Bu veri kümesi, eğitilmiş bir sinir ağının ağırlıklarını alarak ve aynı alandaki veriler üzerinde eğitilen yeni bir modelin başlatma (fine-tuning) stratejisi ile eğitilmiş ve sonrasında doğruluk (accuracy), hassasiyet (precision) ve ortalama hassasiyet (mean Average Precision) başarımları üzerinden karşılaştırılarak MobileNet-SSD modelinin %50 eğitim verisi kullanıldığında %88.07 ortalama hassasiyet elde edilmiştir.

BİLGİLENDİRME

Bu çalışma, TÜBİTAK ARDEB 117M151 ve 119E248 numaralı projeler ile İTÜ - BAP MAB-2020-42314 numaralı proje kapsamında desteklenmiştir.

KAYNAKLAR

- [1] Y. Yildirim, et al., "Image analysis based fish tail beat frequency estimation for fishway efficiency," *2018 26th European Signal Processing Conference (EUSIPCO), Rome*, pp. 1790–1794, 2018.
- [2] Kucukali, et al., "Flow structure and fish passage performance of a brush-type fish way: a field study in the İyidere River, Turkey," *Marine and Freshwater Research* 70.11, pp. 1619–1632, 2019.
- [3] M. Sung, et al., "Vision based real-time fish detection using convolutional neural network," in *OCEANS 2017 - Aberdeen*, 2017, pp. 1–6.
- [4] Salman, et al., "Automatic fish detection in underwater videos by a deep neural network-based hybrid motion learning system," *ICES Journal of Marine Science*, 2019.
- [5] Sun, et al., "Transferring deep knowledge for object recognition in low-quality underwater videos," *Neurocomputing*, vol. 275, pp. 897–908, 2018.
- [6] Sébastien Villon, et al., "A deep learning method for accurate and fast identification of coral reef fishes in underwater images," *Ecological Informatics*, vol. 48, pp. 238 – 244, 2018.
- [7] Gutierrez, et al., "Visualizing underwater ocean optics," in *Computer Graphics Forum*, vol. 27/2. Wiley Online Library, 2008, pp. 547–556.
- [8] Lythgoe, et al., "Problems of seeing colours under water," in *Vision in fishes*. Springer, 1975, pp. 619–634.
- [9] C. Ancuti, et al., "Enhancing underwater images and videos by fusion," in *2012 IEEE Conference on Computer Vision and Pattern Recognition*, June 2012, pp. 81–88.
- [10] P. Drews Jr, et al., "Transmission estimation in underwater single images," in *2013 IEEE International Conference on Computer Vision Workshops*, Dec 2013, pp. 825–830.
- [11] P. Drews, et al., "Automatic restoration of underwater monocular sequences of images," in *2015 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, Sep. 2015, pp. 1058–1064.
- [12] Boom, et al., "A research tool for long-term and continuous analysis of fish assemblage in coral-reefs using underwater camera footage," *Ecological Informatics*, vol. 23, pp. 83–97, 2014.
- [13] Wenwei Xu, et al., "Underwater fish detection using deep learning for water power applications," 2018.
- [14] Pedersen, et al., "Detection of marine animals in a new underwater dataset with varying visibility," in *The IEEE Conference on Computer Vision and Pattern Recognition (CVPR) Workshops*, June 2019.
- [15] Ren, et al., "Faster r-cnn: Towards real-time object detection with region proposal networks," in *Advances in Neural Information Processing Systems*, 2015, pp. 91–99.
- [16] Redmon, et al., "Yolo9000: better, faster, stronger," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2017, pp. 7263–7271.
- [17] Ioffe, et al., "Batch normalization: Accelerating deep network training by reducing internal covariate shift," *arXiv:1502.03167*, 2015.
- [18] Redmon, et al., "Yolov3: An incremental improvement," *arXiv:1804.02767*, 2018.
- [19] Lin, et al., "Feature pyramid networks for object detection," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2017, pp. 2117–2125.
- [20] Liu, et al., "Ssd: Single shot multibox detector," in *European Conference on Computer Vision*. Springer, 2016, pp. 21–37.
- [21] Simonyan, et al., "Very deep convolutional networks for large-scale image recognition," *arXiv:1409.1556*, 2014.
- [22] Sandler, et al., "Mobilenetv2: Inverted residuals and linear bottlenecks," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2018, pp. 4510–4520.
- [23] Deng, et al., "Imagenet: A large-scale hierarchical image database," in *2009 IEEE Conference on Computer Vision and Pattern Recognition*. Ieee, 2009, pp. 248–255.
- [24] Mark Everingham, et al., "The pascal visual object classes challenge: A retrospective," *International Journal of Computer Vision*, vol. 111, pp. 98–136, 2014.