

OTDK-DOLGOZAT

2024

**Milton Friedman Egyetem
Módszertani és Informatikai tanszék**

**A számítógépes látásalapú objektumdetektálás és képfeldolgozás kihívásai és
korlátjai a vízfelszín alatt**

**Challenges and limitations of computer vision-based object detection and
image processing under the water surface**

Babenyecz Bálint

Gazdaságinformatikus BSc

Konzulens tanár:

Dr. Bárkányi Pál, PhD

Informatika Tanszék

Kézirat lezárása:

2024. JÚLIUS. 22.

TARTALOMJEGYZÉK

1	Bevezetés.....	4
2	Szakirodalmi körkép	7
2.1	A két fő szakirodalom bemutatása.....	8
2.2	A "Deep Learning based Underwater Object Detection" disszertáció és az „Underwater object detection using deep learning” mesterdolgozat összehasonlító elemzése.....	10
3	Hipotézis felállítása a szakirodalom alapján	19
4	A víz alatti objektumfelismerés optimalizálása mélytanulási technikákkal: Kihívások, megoldások és kísérleti eredmények.....	21
4.1	Adatelőkészítés és -tisztítás	24
4.1.1	Zajcsökkentés és Kontrasztjavítás.....	24
4.1.2	Színszabályozás Technikák.....	26
4.1.3	Fény Csökkenés Adaptáció	28
4.2	Adathalmaz Kiterjesztése (Data Augmentation)	28
4.2.1	Képek Módosításának Módszerei	29
4.3	Mélytanulási Alapú Javítások.....	29
4.3.1	Autoencoderek és GAN-ek Alkalmazása.....	30
4.4	Transfer Learning	31
4.4.1	Előre Kiképzett Modellek Használata.....	32
4.4.2	Transfer Learning alkalmazásának előnyei.....	32
4.5	Figyelem Mechanizmusok.....	34
4.5.1	Figyelem-alapú Mechanizmusok	34
4.5.2	Releváns Jellemzők Kiemelése	36
4.6	Hibrid Modellek.....	36
4.6.1	CNN-ek és RNN-ek Kombinálása	37
4.6.2	Transformer Modellek Integrálása	38

4.7	Regularizáció és Finomhangolás	38
4.7.1	Regularizációs Technikák Alkalmazása.....	38
4.8	A kísérleti adatmodell terveinek megalkotása és fejlesztése	41
4.8.1	A kísérleti modell felépítése.....	43
4.8.2	A kísérleti modell teljesítmények értékelése.....	46
5	Összefoglalás.....	56
5.1	Megvalósult Célok Értékelése	56
5.2	Meg Nem Valósult Célok Elemzése.....	57
5.3	A Továbbfejlesztés Lehetőségei	58
	Fogalomtár.....	60
	Irodalomjegyzék	63
	Absztrakt.....	67
	Abstract.....	68
	Ábrák és táblázatok jegyzéke	3
	Mellékletek	69

ÁBRÁK ÉS TÁBLÁZATOK JEGYZÉKE

1. ábra - Különböző fényszóródások változó vízfelszíni kondíciók esetén [2].	22
2. ábra - A látható fény színtartományai különböző vízmélységekben [4]	23
3. ábra - SeaTrue modell szíinkorrekciója [8]	27
4. ábra- A kísérleti adatmodell „Tanítási” és „validációs” pontossága ("Training accuracy" és "Validation accuracy")	46
5. ábra - A kísérleti adatmodell „Tanítási” és „validációs” vesztesége ("Training loss" és "Validation loss")	47
6. ábra - A kísérleti adatmodell Grad-CAM vizualizációi	52
1. táblázat - A kísérleti adatmodell klasszifikációs riport táblázata	49

1 BEVEZETÉS

A gépi látáson és objektumfelismerésen alapuló technológiákat, kritikus jelentőségű felhasználási területeken alkalmazzák és ezen területek folyamatosan bővülnek. Ilyen alkalmazási lehetőségek például a mélytengeri bányászat vagy a víz alatti építkezések, amelyek potenciális gazdasági növekedési lehetőségeket kínálnak. A gépi látás és objektumfelismerés a globális ökoszisztéma fenntartásának és védelmének szempontjából is nélkülözhetetlen technológiává válhat, azáltal, hogy ezen technológiák segítségével pontosabb információt kaphatunk az egyes ökoszisztémák működéséről, állapotáról. A Föld, mint vízbolygó, és az óceánok gazdag biológiai sokfélesége különleges figyelmet érdemelnek ebben a kontextusban. A nemzetközi halászati iparág, mely 2016-ban összesen 143 milliárd dolláros bevételt termelt, jól példázza a víz alatti környezet gazdasági és ökológiai jelentőségét [1].

A vízfelszín alatti objektumdetektálás kiemelten fontos és egyben kihívást jelentő terület a tengeri kutatásokban és víz alatti technológiákban. Ez a probléma nem csupán a tengerbiológia és a régészet számára releváns, hanem létfontosságú a kereskedelmi hajózás, olaj- és gázkitermelés, valamint a haditengerészet számára is. A víz alatti környezet specifikus kihívásai – mint az alacsony fényviszonyok, a víz fizikai tulajdonságai vagy a tengeri élőlények jelenléte – jelentősen megnehezítik az objektumok azonosítását és klasszifikálását.

A víz alatti környezetben a gépi objektumfelismerés a mesterséges intelligencia (MI) alkalmazása számos kihívással szembesül, melyek közül az egyik legösszetettebb kihívás a vizuális adatok torzulása. A víz fizikai tulajdonságai, mint a látótávolság, a fénytörés vagy az elszórt fény, jelentősen befolyásolják a képfelvételek minőségét. A mélyebb vízrétegekben a fényviszonyok gyengülése miatt a képek sötétek és homályosak lehetnek, mely komoly kihívást jelent az objektumok pontos azonosításában. Ezen kívül, a vízben lebegő részecskék és az élőlények mozgása tovább bonyolítja a képi adatok elemzését, mivel ezen törmelékek gyakran összetéveszthetőek a keresett objektumokkal.

A másik jelentős akadály a mesterséges intelligencia alkalmazásában a víz alatti környezet változékonysága. A tengeri áramlatok, a víz hőmérsékletének és ezzel együtt viszkozitásának ingadozásai, valamint az időjárási viszonyok hatása mind befolyásolják a víz alatti képek minőségét. Ez megnehezíti a mesterséges intelligencia modellek tanítását és alkalmazását, mivel a modellnek képesnek kell lennie alkalmazkodni a változó körülményekhez. Emellett, a víz alatti adatgyűjtés magas költségei és a nehezen hozzáférhető környezet szintén korlátozzák a nagy mennyiségű és változatos adatok gyűjtését, amelyek elengedhetetlen a hatékony gépi tanulási modellek fejlesztéséhez.

Dolgozatom témájaként részletesen elemzem a víz alatti képfeldolgozás lehetőségeit, kihívásait, illetve az azokra adható válaszokat, megoldásokat. A kutatás keretében összehasonlítom a különböző szakirodalmi forrásokat, hogy bemutassam, milyen algoritmikus megoldásokat, megközelítéseket alkalmaznak a terület szakértői, kutatói. Részletesen tárgyalom a víz alatti környezet specifikus kihívásait, mint például az optikai torzulásokat, az alacsony fényviszonyokat és a víz fizikai tulajdonságait, ezután megvizsgálom a kihívásokra adható válaszokat. Emellett egy kísérleti modellen keresztül igyekszem elemezni az egyes megadásokat, figyelembe véve a dolgozat technikai határait. Dolgozatomban elsődlegesen a meglévő technológiai megoldások elemzésével, továbbfejlesztésével foglalkozok, a dolgozat tartalmi keretei és témája okán elsődleges szoftveres adatkészletekre, megoldásokra és az ezekkel kapcsolatos kísérletekre fókuszálok.

A dolgozat elsődlegesen a gépi látás szakirodalmi vizsgálatára összpontosít, majd kiterjeszti a fókusz a víz alatti képfeldolgozás specifikus kihívásaira, különös tekintettel a látható fény tartományában történő képelemzésre. A kutatás keretében nem tárgyalom részletesen a hardveres követelményeket; helyette az algoritmusok és modellek felé fordítom a figyelmet. Ez a megközelítés lehetővé teszi, hogy mélyrehatóan vizsgáljam azokat a szoftveres megoldásokat, melyek jelentősen javíthatják a víz alatti képek értelmezését és analízisét. A kutatás ilyen módon történő strukturálásával és gyakorlati kísérletezéssel a dolgozat holisztikus értékelést kíván adni a gépi látás és objektum felismerés technikák lehetőségeiről és korlátairól a víz alatti képfeldolgozás kontextusában. E célkitűzés érdekében a dolgozat a következő fő kutatási kérdésre keresi a lehetséges választ: Hogyan optimalizálhatók a

mélytanulási technikák és algoritmusok a víz alatti objektumfelismerés pontosságának és megbízhatóságának javítására, optimalizálására, figyelembe véve a víz alatti környezet egyedi kihívásait, mint például az optikai torzulásokat, az alacsony fényviszonyokat és a víz fizikai tulajdonságait? A kérdés megválaszolása során a dolgozat részletesen vizsgálja a jelenlegi technológiai megoldásokat, azok továbbfejlesztési lehetőségeit, valamint egy kísérleti modellen keresztül elemzi az egyes megközelítések hatékonyságát.

2 SZAKIRODALMI KÖRKÉP

A gépi számítógépes látásalapú objektumdetektálással kapcsolatos szakirodalmak áttekintése során különböző forrásokat vizsgáltam meg, amelyek a képfeldolgozás és objektumdetektálás számos aspektusát tárgyalják. Ezek a szakirodalmak a technológia alapvető elméletétől kezdve, a gyakorlati alkalmazásokig terjednek, és különböző nézőpontokat kínálnak a számítógépes látás fejlődésének és alkalmazásainak megértéséhez.

A szakirodalmak egy része az alapvető képfeldolgozási technikákra és objektumdetektálási módszerekre koncentrál, bemutatva, hogyan használhatók ezek az eszközök az adatok értelmezéséhez és a vizuális információk feldolgozásához. Más források a számítógépes látás fejlett alkalmazásait elemzik, például az arcfelismerés, a járműfelismerés vagy a víz alatti objektumdetektálás területén. Ezek a publikációk bemutatják a legújabb technológiai fejlesztéseket, például a mélytanulási módszereket és a konvolúciós neurális hálózatokat, amelyek forradalmasították a számítógépes látás képességeit.

Kutatásaim során leginkább azon szakirodalmakat elemeztem melyek részletesen bemutatnak specifikus problémaköröket a víz alatti objektum detektálással kapcsolatban. Az elemezett szakirodalmak döntő többsége ezáltal olyan szakcikk mely konkrét, gyakran gyakorlati problémamegoldásokra koncentrál. Ezen cikkek nagyrésze a Cornwell egyetem, az IEE konferencia sorozat és a más konferenciák archív anyagaként elérhető, ezáltal tartalmukat tekintve gyakran egy konkrét technológiai megoldást taglalnak.

2.1 A két fő szakirodalom bemutatása

Ez a fejezet rész áttekintést ad a dolgozatban végzett empirikus kutatás alapját képező két alapvető kutatási dolgozatról. Az első szakirodalom, Long Chen "Deep Learning based Underwater Object Detection" című doktori disszertációja (Chen, 2022), míg a második, Moutaz Mohamed Fathy Ahmed Aboseada "Underwater object detection using deep learning" mesterdolgozata (Aboseada, 2020). Ez a két munka a jelenlegi kutatás kontextusában a legrelevánsabbnak és a legbefolyásosabbnak tekinthető, mivel mindkettő konkrétan a jelen tézishoz kapcsolódó problématerületeket más-más nézőpontból tárgyalja.

A kutatási munkák elemzése elsősorban a hipotézisekre és a hozzájuk kapcsolódó eredményekre, illetve lehetséges megoldásokra fókuszál. Az egyes tézisek tartalmát kutatási témák és eredmények szerint vizsgálom, követve az egyes művek tematikus felépítését.

A két tézis felépítésénél és hipotéziséknél fogva megfelelő összehasonlítási és további kutatási lehetőségeket kínál az összehasonlító elemzés eredményei alapján. Bár a két munka terjedelmében és mélységében eltér egymástól, a vizsgált téma-specifikus tanulmányok közül a kutatási célok és módszertan tekintetében ezek mutatják a leginkább hasonló szerkezetet.

A "Deep Learning based Underwater Object Detection" című doktori disszertáció egy átfogó tanulmányt mutat be a fejlett mélytanulási rendszerek fejlesztéséről a víz alatti objektumészlelés (UOD) egyedi kihívásainak kezelésére. A kutatás elsődleges célja olyan robusztus mélytanulási keretrendszer kifejlesztése, amely jelentősen javítja a víz alatti objektumok észlelési pontosságát, miközben biztosítja a robusztusságot a víz alatti képekben rejlő kedvezőtlen körülményekkel szemben. A dolgozat hat fő fejezetből áll, amelyek egy bevezetést, szakirodalmi áttekintést, víz alatti képjavítást mélytanulással és fizikai prioritásokkal, víz alatti objektumok észlelését zajos adatkészletekben, víz alatti objektumok észlelését kiegyensúlyozatlan adatkészletekben, valamint következtetéseket és perspektívákat tartalmaznak.

Másrészt a "Underwater object detection using deep learning" című mesterdolgozat a mélytanulási technikák alkalmazására összpontosít a víz alatti objektumok észlelésének pontosságának és hatékonyságának javítására. A dolgozat különböző mélytanulási architektúrákat tár fel, mint például a konvolúciós neurális hálózatokat (CNN) és a régióalapú CNN-eket (R-CNN), valamint ezek adaptációját a víz alatti környezethez. A munka azt is vizsgálja, hogy az adatbővítési technikák és a transzfertanulás milyen hatással van a víz alatti objektumészlelési modellek teljesítményére. A doktori disszertáció nagy hangsúlyt fektet a víz alatti környezet által támasztott kihívások – például súlyos képtorzulás, zajos adatok és kiegyensúlyozatlan adathalmazok – kezelésére, új mélytanulási keretrendszerek és algoritmusok javaslatával. Ezzel szemben a mesterdolgozat inkább a meglévő mélytanulási technikák gyakorlati alkalmazására, illetve a víz alatti objektumfelderítési feladatokra való optimalizálására koncentrálnak.

2.2 A "Deep Learning based Underwater Object Detection" disszertáció és az „Underwater object detection using deep learning” mesterdolgozat összehasonlító elemzése

Long Chen doktori disszertációja átfogó tanulmányt mutat be a fejlett mélytanulási rendszerek fejlesztéséről a víz alatti objektumok észlelésének (UOD) egyedi kihívásainak kezelésére. A kutatás elsődleges célja olyan robusztus mélytanulási keretrendszer kifejlesztése, amely jelentősen javítja a víz alatti objektumok észlelési pontosságát. Másrészt Moutaz Mohamed Fathy Ahmed Aboseada mesterdolgozata a mélytanulási technikák alkalmazására fókuszál a víz alatti objektumok észlelésének pontosságának és hatékonyságának javítására, feltárva a különböző mélytanulási architektúrákat és azok adaptációját a víz alatti környezethez.

A kutatási célok és módszertanok összehasonlító elemzése

A doktori értekezés és a mesterdolgozat kutatási céljaiban és módszertanában különbözik, tükrözve az egyes munkák terjedelmét és mélységét. Long Chen doktori disszertációja a víz alatti környezet által támasztott alapvető kihívásokra kíván megoldást találni, mint például a súlyos képtorzulás, a zajos adatok és a kiegyensúlyozatlan adatkészletek. E cél elérése érdekében a szerző új mélytanulási keretrendszereket és algoritmusokat javasol, amelyeket kifejezetten ezeknek a problémáknak a kezelésére terveztek. A doktori disszertációban alkalmazott kutatási módszertan az elméleti elemzést, az algoritmusfejlesztést és az empirikus értékelést ötvözi, biztosítva a probléma átfogó megközelítését.

Ezzel szemben a mesterdolgozat a meglévő mélytanulási technikák gyakorlati alkalmazására és azok optimalizálására fókuszál a víz alatti objektumfelderítési feladatokhoz. Ennek a munkának az elsődleges célja a különböző mélytanulási architektúrák, például a konvolúciós neurális hálózatok (CNN-ek) és a régióalapú CNN-ek (R-CNN-ek) hatékonyságának feltárása a víz alatti objektumok észlelésével összefüggésben. A mesterdolgozatban szereplő kutatási módszertan ezen technikák empirikus értékelésére helyezi

a hangsúlyt, az adatbővítés és a transzfer tanulás hatását vizsgálva a víz alatti objektumdetektáló modellek teljesítményére.

A kutatási eredmények és következtetések összehasonlító elemzése

A doktori értekezés és a mesterdolgozat eltérő kutatási eredményeket és következtetéseket mutat be, amelyek eltérő kutatási célkitűzéseikből és módszertanukból fakadnak. Long Chen doktori disszertációja, amely a víz alatti objektumok észlelésének alapvető kihívásait kívánja megválaszolni, számos újszerű mélytanulási keretrendszert és algoritmust javasol. A legfigyelemreméltóbbak közé tartozik a HybridDetectionGAN keretrendszer a víz alatti képjavításhoz, a Sample-Weighted HyPER Network (SWIPENET) a kis tárgyak észleléséhez zajos környezetben, valamint a FAGR algoritmus az osztálykiegyensúlyozatlansági problémák megoldására.

A doktori disszertációban bemutatott kísérleti eredmények jelentős előrelépést mutatnak a legmodernebb megközelítésekhez képest. A javasolt módszerek hatékonyan kezelik a súlyos képtorzulás, a zajos adatok és a kiegyensúlyozatlan adatkészletek jelentette kihívásokat, ami fokozott észlelési pontosságot és robusztusságot eredményez a valós víz alatti környezetekben. A szerző arra a következtetésre jut, hogy ezen újszerű technikák egységes keretbe integrálása nagymértékben előre mozdíthatja a víz alatti objektumok észlelését.

Ezzel szemben a mesterdolgozat a meglévő mélytanulási technikák empirikus értékelésére és a víz alatti környezethez való adaptálására fókuszál. A kutatási eredmények kiemelik a konvolúciós neurális hálózatok (CNN-ek) és a régióalapú CNN-ek (R-CNN-ek) hatékonyságát a víz alatti objektumok észlelésének pontosságának és hatékonyságának javításában. A szerző bemutatja, hogy az adatbővítési technikák és a transzfertanulás alkalmazása tovább javíthatja ezeknek a modelleknek a teljesítményét, különösen az olyan felvételeken, ahol a megjegyzésekkel ellátott víz alatti adatkészletek elérhetősége korlátozott.

A mesterdolgozatban levont következtetések a mélytanulási technikák alkalmazásának gyakorlati megvalósíthatóságát hangsúlyozzák a víz alatti objektumfelderítési feladatokban. A

szerző azt sugallja, hogy ezen technikák optimalizálása, az adatkiegészítés és az átviteli tanulás használatával párosulva, jelentős javulást eredményezhet az észlelési pontosságban és hatékonyságban. A mesterdolgozat ugyanakkor elismeri a jelenlegi megközelítések korlátait is, és rávilágít a további kutatások szükségességére a víz alatti környezet által támasztott sajátos kihívások kezelésére.

A kutatási hatások és jövőbeli irányok összehasonlító elemzése

A doktori disszertáció és a mesterdolgozat külön betekintést nyújt a víz alatti tárgydetektálás kutatásának jövőbeli irányába. Long Chen doktori disszertációja hangsúlyozza az újszerű mélytanulási keretrendszerek és algoritmusok fejlesztésének fontosságát, amelyek kifejezetten a víz alatti környezet egyedi kihívásainak kezelésére készültek. A szerző azt javasolja, hogy a jövőbeni kutatásoknak az alacsony szintű képjavító technikák és a magas szintű objektumészlelési módszerek egységes végpontok közötti keretrendszerbe való integrálására kell összpontosítaniuk. Emellett a disszertáció rávilágít az adatkiegészítésszerűség kezelésére szolgáló technikák további feltárására, valamint a víz alatti környezet dinamikus viszonyaihoz alkalmazkodó modellek kidolgozására.

A mesterdolgozat ezzel szemben a meglévő mélytanulási technikák kiaknázásában és a víz alatti tartományhoz való adaptálásában rejlő lehetőségeket hangsúlyozza. A szerző azt javasolja, hogy a jövőbeni kutatásoknak ezeknek a technikáknak a finomítására és optimalizálására kell összpontosítaniuk, különös tekintettel arra, hogy képesek-e kezelni a víz alatti képek egyedi jellemzőit. A dolgozat kiemeli a mélytanulási modellek képzését és értékelését megkönnyítő nagyobb és változatosabb, annotált víz alatti adatkészletek kifejlesztésének fontosságát.

Mindkét munka hangsúlyozza az interdiszciplináris együttműködés jelentőségét a víz alatti objektumok felderítésének területén. A szerzők hangsúlyozzák, hogy szoros együttműködésre van szükség a számítógépes látás, a mélytanulás, a tengertudomány és a víz alatti robotika szakértői között annak érdekében, hogy átfogó megoldásokat dolgozzanak ki, amelyek a víz alatti környezet összetett kihívásait kezelik.

A kutatási eredmények és következtetések összehasonlító elemzése

A doktori értekezés és a mesterdolgozat eltérő kutatási eredményeket és következtetéseket mutat be, amelyek eltérő kutatási célkitűzéseikből és módszertanukból fakadnak. Long Chen doktori disszertációja, amely a víz alatti objektumok észlelésének alapvető kihívásait kívánja megválaszolni, számos újszerű mélytanulási keretrendszert és algoritmust javasol. A legjelentősebb javaslatok a következők:

1. HybridDetectionGAN keretrendszer: Ez a keretrendszer egy hibrid víz alatti képszintézis modellt és egy észlelési-percepció javítási modellt kombinál. A hibrid szintézis modell mind a fizikai prioritásokat, mind az adatvezérelt jelzéseket magában foglalja, hogy valósághű víz alatti képeket hozzon létre, míg a továbbfejlesztett modell észlelést elősegítő képeket generál az észlelés pontosságának javítása érdekében.

2. Mintasúlyozott HyPER hálózat (SWIPENET): A SWIPENET-et arra tervezték, hogy megbirkózzon a kis tárgyak észlelésével járó kihívásokkal zajos víz alatti környezetben. Kihasználja a nagy felbontású és szemantikában gazdag Hyper Feature Maps funkciót, és mintával súlyozott észlelési veszteséget alkalmaz a zajos minták hatásának szabályozására.

3. „Factor-agnostic gradient re-weighting” (FAGR) algoritmus: A FAGR a víz alatti objektumok észlelésének osztálykiegyensúlyozatlansági problémáját a különböző osztályok gradienseinek újrásúlyozásával az észlelési pontosságuk alapján kezeli. Ez a megközelítés hatékonyabbnak bizonyul, mint a hagyományos adatelosztáson alapuló újraegyensúlyozási módszerek.

A doktori disszertációban bemutatott kísérleti eredmények jelentős előrelépést mutatnak a legmodernebb megközelítésekhez képest. Például a HybridDetectionGAN keretrendszer 5%-kal növeli az átlagos pontosságot (mAP) és 3%-kal növeli az Intersection over Union (IoU) értékét a képjavítás nélküli képzéshez képest. A SWIPENET+CMA keretrendszer az URPC2018 és a ChinaMM adatkészleteken felülmúlja a többi kis objektum-észlelési

módszert, például a DSSD-t, a RetinaNet-et és az FCOS-t. A FAGR algoritmus hatékonyan csökkenti a különböző osztályok közötti pontossági eltéréseket, ami kiegyensúlyozottabb észlelési teljesítményt eredményez.

Ezzel szemben a mesterdolgozat a meglévő mélytanulási technikák empirikus értékelésére és a víz alatti környezethez való adaptálására fókuszál. A szerző különféle CNN-architektúrák, például VGG16, ResNet50 és ResNet101, valamint régióalapú CNN-modellek, például Faster R-CNN teljesítményét vizsgálja. A kutatási eredmények rávilágítanak e modellek hatékonyságára a víz alatti tárgyak észlelésének pontosságának és hatékonyságának javításában. Például a ResNet101-es gerinchálózattal rendelkező Faster R-CNN 70,9%-os mAP-t ér el a ChinaMM-adatkészleten, ami felülmúlja a többi CNN-alapú modellt.

A Moutaz Mohamed Fathy Ahmed Aboseada bemutatja az adatkiegészítési technikák, mint például a vízszintes átfordítás, a függőleges átfordítás és a véletlenszerű vágás előnyeit a mélytanulási modellek teljesítményének javításában. A kísérletek azt mutatják, hogy az adatbővítés akár 3%-kal javíthatja az eredményeket az URPC2018 adatkészleten. Ezen túlmenően a dolgozat feltárja az átviteli tanulás lehetőségét a víz alatti adatkészleteken előre betanított modellek finomhangolásával, ami hatékonyak bizonyul olyan képeken, ahol korlátozott a jegyzett víz alatti adatok elérhetősége.

A kutatási hatások és jövőbeli irányok összehasonlító elemzése

A doktori disszertáció és a mesterdolgozat külön betekintést nyújt a víz alatti tárgydetektálás kutatásának jövőbeli irányába. Long Chen doktori disszertációja hangsúlyozza az újszerű mélytanulási keretrendszerek és algoritmusok fejlesztésének fontosságát, amelyek kifejezetten a víz alatti környezet egyedi kihívásainak kezelésére készültek. A szerző azt javasolja, hogy a jövőbeni kutatásnak a következőkre kell összpontosítania:

1. Az alacsony szintű képjavítás és a magas szintű objektumészlelés integrálása egy egységes végpontok közötti keretrendszerbe.

2. Az adatkiegyensúlyozatlanság kezelésére szolgáló technikák további feltárása, mint például a javasolt FAGR algoritmus.

3. Olyan modellek kidolgozása, amelyek képesek alkalmazkodni a víz alatti környezet dinamikus viszonyaihoz, például változó víz zavarosságához és fényviszonyokhoz.

Moutaz Mohamed Fathy Ahmed Aboseada mesterdolgozata ezzel szemben a meglévő mélytanulási technikák kiaknázásában és a víz alatti tartományhoz való adaptálásában rejlő lehetőségeket hangsúlyozza. A szerző azt javasolja, hogy a jövőbeni kutatásnak a következőkre kell összpontosítania:

1. CNN-alapú modellek finomítása és optimalizálása víz alatti objektumok észlelésére, különösen abból a szempontból, hogy képesek legyenek kezelni a víz alatti képek egyedi jellemzőit, például a színtorzítást és a homályos hatásokat.

2. Nagyobb és változatosabb, annotált víz alatti adatkészletek kidolgozása a mélytanulási modellek képzésének és értékelésének megkönnyítésére. A szerző kiemeli a víz alatti környezetek, objektumosztályok és képminőségek széles skáláját lefedő adatkészletek szükségességét.

3. Fejlett adatkiegészítési technikák (például GAN-ok) feltárása valósághű víz alatti képek szintetizálására és a képzési adatok bővítésére.

Mindkét munka hangsúlyozza az interdiszciplináris együttműködés jelentőségét a víz alatti objektumok felderítésének területén. A szerzők hangsúlyozzák, hogy szoros együttműködésre van szükség a számítógépes látás, a mélytanulás, a tengertudomány és a víz alatti robotika szakértői között annak érdekében, hogy átfogó megoldásokat dolgozzanak ki, amelyek a víz alatti környezet összetett kihívásait kezelik.

Összefoglalva, a doktori disszertáció és a mesterdolgozat összehasonlító elemzése feltárja, hogyan járulnak hozzá a mélytanulást alkalmazó víz alatti tárgyak észleléséhez. A doktori disszertáció olyan újszerű keretrendszerek és algoritmusok fejlesztésére összpontosít, mint a HybridDetectionGAN, a SWIPENET és a FAGR, valamint a víz alatti objektumok észlelésének alapvető kihívásainak megválaszolására. A mesterdolgozat ezzel szemben a meglévő technológia gyakorlati alkalmazását és optimalizálását helyezi előtérbe, mint például a CNN-ek és az adatbővítés. Mindkét mű értékes betekintést nyújt a tudományterület jelenlegi állásába, és útmutatást ad a jövőbeli kutatási irányokhoz, kiemelve az interdiszciplináris együttműködés fontosságát és a folyamatos innováció szükségességét.

Következtetés: Eredmények és meg nem valósult célok a víz alatti objektumdetektálás kutatásban

1. A doktori disszertáció javaslatot tett a HybridDetectionGAN keretrendszerre, amely egy hibrid víz alatti képszintézis modellt és egy észlelési-percepció javítási modellt kombinál. A keretrendszer 5%-kal javítja az átlagos pontosságot (mAP) és 3%-kal az Intersection over Union (IoU) értéket a képjavítás nélküli képzéshez képest.

2. A doktori disszertációban bemutatott Sample-Weighted HyPER Network (SWIPENET) hatékonyan kezeli a kisméretű tárgyak észlelésének kihívását zajos víz alatti környezetben. A SWIPENET az URPC2018 és a ChinaMM adatkészleteken felülmúlja a többi korszerű módszert, például a DSSD-t, a RetinaNet-et és az FCOS-t.

3. A doktori disszertációban javasolt „Factor-agnostic Gradient Re-weighting” (FAGR) algoritmus a víz alatti objektumok észlelésének osztálykiegyensúlyozatlanságával foglalkozik. A FAGR sikeresen csökkenti a különböző osztályok közötti pontossági eltéréseket, kiegyensúlyozottabb észlelési teljesítményt eredményezve.

4. A mesterdolgozat bemutatja a meglévő mélytanulási technikák, mint például a konvolúciós neurális hálózatok (CNN-ek) és a régió alapú CNN-ek (R-CNN-ek) hatékonyságát a víz alatti objektumok észlelésének pontosságának és hatékonyságának javításában. A disszertáció rávilágít arra is, hogy az adatbővítés és az átviteli tanulás milyen előnyökkel jár a modellek teljesítményének javításában.

A két tézis átfogó áttekintése alapján mindkét munka jelentős mértékben hozzájárult a mélytanulást alkalmazó víz alatti tárgyak észlelésének területéhez. Azonban még mindig vannak korlátok és lehetséges további kutatási területek. A következőkben olyan jövőbeli fejlesztésekre és irányokra teszek javaslatot, amelyeket a két szerzőtől eltérően értékelek.

1. Az alacsony szintű képjavítás és a magas szintű objektumészlelés integrálása egy egységes végpontok közötti keretrendszerbe: Míg a doktori disszertáció külön modelleket javasolt a képjavításra és az objektumészlelésre, egy integrált keretrendszert javaslok, amely együttesen optimalizálja ezeket a feladatokat az észlelési pontosság és hatékonyság további javítása érdekében.

2. A dinamikus víz alatti környezethez alkalmazkodó modellek kidolgozása: Mindkét munka elismeri a víz alatti környezet változó feltételeiből adódó kihívásokat, mint például a változó víz zavarossága vagy a rossz fényviszonyok. A javasolt módszerek azonban nem foglalkoznak kifejezetten ezekkel a problémákkal. A jövőbeli kutatásoknak olyan adaptív modellek kifejlesztésére kell összpontosítaniuk, amelyek dinamikusan képesek alkalmazkodni ezekhez a változó körülményekhez.

3. Nagyméretű, változatos, annotált víz alatti adatkészletek létrehozása: Bár mindkét munka meglévő víz alatti adatkészleteket használ, mint például az URPC2017, URPC2018 és a ChinaMM, még mindig hiányoznak a nagyszabású, átfogó adatkészletek, amelyek a víz alatti környezetek széles skáláját lefedik, mind az objektumosztályok és a képminőségek tekintetében. Az ilyen adatkészletek fejlesztése kulcsfontosságú a robusztusabb és általánosítható mélytanulási modellek képzéséhez és értékeléséhez.

4. Fejlett adatkiegészítési technikák feltárása: Míg a mesterdolgozat bemutatja az alapvető adatkiegészítési technikák előnyeit, mint például a lapozás és a vágás, addig a fejlettebb módszereket, mint például a „Generative Adversarial Networks” (GAN-ok) alkalmazását még nem implementálták teljesen a víz alatti objektumok észlelésére. A jövőbeli kutatásoknak meg kell vizsgálniuk e technikákban rejlő lehetőségeket a valósághű víz alatti képek szintetizálásában és a képzési adatok bővítésében.

5. Valós idejű víz alatti objektum-detektálás: Bár a doktori disszertáció és a mesterdolgozat jelentős előrelépéseket tett a víz alatti objektumok észlelésének pontosságának javításában, a felsorolt módszerek valós idejű teljesítményét még nem vizsgálták alaposan. A jövőbeli kutatásnak a javasolt keretrendszerek és algoritmusok optimalizálására kell összpontosítania a valós idejű következtetésekhez, ami döntő fontosságú a gyakorlati alkalmazásokhoz, például az autonóm víz alatti járművekhez és a tengeri megfigyelőrendszerekhez.

Összefoglalva, Long Chen doktori disszertációja és Moutaz Mohamed Fathy Ahmed Aboseada mesterdolgozata jelentős mértékben hozzájárult a tudományterülethez, megválaszolva a kulcsfontosságú kihívásokat, és újszerű megoldásokat javasolt. A jövőbeli kutatások számára azonban számos meg nem valósult cél és terület maradt, beleértve a képjavítás és az objektumészlelés integrációját, az adaptív modellek fejlesztését, a nagyszabású adathalmazok létrehozását, a fejlett adatkiegészítési technikák feltárását és a valós hasznosítás optimalizálását.

2.3 Hipotézis felállítása a szakirodalom alapján

Long Chen doktori disszertációja és Moutaz Mohamed Fathy Ahmed Aboseada mesterdolgozata összehasonlító elemzése alapján a következő hipotézist és lehetséges kutatások javasolhatóak a víz alatti objektumok területén végzett további kutatásokhoz. A feltételezés szerint a két szerző átfogói munkái ellenére öt olyan terület látható melyeken tovább optimalizálhatóak a víz alatti objektumok detektálási megoldásai. Figyelembe véve az eddigi eredmények és meg nem valósult célokat ezek a területek a következők:

1. Fejlett képjavító technikák integrálása:

Bár mindkét mű a képjavítással foglalkozott, láthatóan nem fektettek kellő hangsúlyt a színvisszaállításra. A víz alatti környezet egyedi kihívások elé állítja a színeket, például a vörös hullámhosszok elnyelését, ami kékes vagy zöldes árnyalatokat eredményez. A fejlett színpontozási algoritmusok, például a fehéregyensúly finomhangolása vagy az adaptív színhőmérséklet-beállítás javíthatják az észlelési teljesítményt. A legkorszerűbb képjavító technikák, például a „Generative Adversarial Networks”-ök (GAN) vagy a mélytanuláson alapuló módszerek közvetlenül az észlelési folyamatba történő integrálása lehetővé tenné a modellek számára, hogy megtanulják az észlelés szempontjából kedvező reprezentációkat, növelve a pontosságot és a hatékonyságot.

2. Adaptív mélytanulási architektúrák:

A víz alatti környezet dinamikus és változó természetéből adódó kihívások kezelésére olyan adaptív mélytanulási architektúrákat lehet kifejleszteni, amelyek automatikusan finomítják paramétereiket és struktúráikat a bemeneti adatok jellemzői alapján. Figyelem mechanizmusok vagy neurális architektúra keresési (NAS) technikák beépítése lehetővé teszi a modell számára, hogy a kép leginformatívabb területeire összpontosítson, és dinamikusan igazítsa összetettségét az adott víz alatti viszonyokhoz.

3. Tanulás átvitele kiterjedt földi adatkészletekből:

Tekintettel a nagyméretű, megjegyzésekkel ellátott víz alatti adatkészletek korlátozott elérhetőségére, átviteli tanulási technikákat kell alkalmazni a kiterjedt földi objektum-észlelési adatkészletekből, például a COCO-ból vagy az ImageNet-ből származó ismeretek kiaknázására. Az előre betanított modellek finomhangolása ezeken adatkészleteken valamint a tartomány-adaptációs módszerek, például a „Generative Adversarial Networks”-ök alkalmazása tovább növelheti az objektumok detektálásnak hatásfokát. Ezen módszerek együttes alkalmazásával a víz alatti objektumdetektorok teljesítménye jelentősen javítható anélkül, hogy hatalmas mennyiségű annotált víz alatti adatra lenne szükség.

4. CNN és „Transformer” modellek kombinációja:

A CNN és a „Transformer” architektúrák kombinálása ígéretes megközelítés a víz alatti objektumok észlelésére. Míg Chen és Aboseada elsősorban CNN-eket alkalmazott, addig a Transformer modellek, mint például a „Vision Transformer”-ek (ViT), kiválóak a hosszú távú függőségek és a globális kontextus modellezésében. A hibrid CNN-Transformer architektúra lehetővé tenné a helyi és globális jellemzők együttes kihasználását, ami pontosabb és robusztusabb észlelést eredményez.

5. Figyelem mechanizmusok:

Figyelem mechanizmusok, például nem lokális neurális hálózatok vagy a Grad-CAM technika beépítése jelentősen javíthatja a mélytanulási modellek teljesítményét a víz alatti objektumok észlelésében. Ezek a mechanizmusok lehetővé teszik a modellek számára, hogy dinamikusan hozzárendeljék a fókusz és a számítási erőforrásokat a bemeneti adatok legrelevánsabb részeihez, ami pontosabb és hatékonyabb eredményeket eredményez. Míg Chen és Aboseada nem alkalmazott kifejezetten figyelemmechanizmusokat, ezeknek a javasolt keretrendszerükbe történő integrálása tovább javíthatja az észlelési képességeket.

3 A VÍZ ALATTI OBJEKTUMFELISMERÉS OPTIMALIZÁLÁSA MÉLYTANULÁSI TECHNIKÁKKAL: KIHÍVÁSOK, MEGOLDÁSOK ÉS KÍSÉRLETI EREDMÉNYEK

A víz alatti tárgyak felismerése és észlelése számos kihívás elé állítja a mélytanulási modelleket. Ebben a fejezetben részletesen feltérképezem ezeket a kihívásokat, és olyan technológiákat és módszereket mutatok be, amelyek segíthetnek leküzdeni ezeket a problémákat, ezáltal javítva a víz alatti objektumok észlelésének hatékonyságát és pontosságát. Továbbá, a fejezetben konkrét megoldásokat és megvalósítási ötleteket tárok fel egy kísérleti adatmodell tervezése, fejlesztése, majd kiértékelése folyamán, ezáltal tesztelve a megoldási irányokat és lehetőségeket. Az előző fejezetben bemutatott hipotézis és javasolt fejlesztések alapján öt kulcsfontosságú területre összpontosítottam, amelyek további fejlesztésekre adhatnak lehetőséget a víz alatti objektumok felismerésének optimalizálása szempontjából:

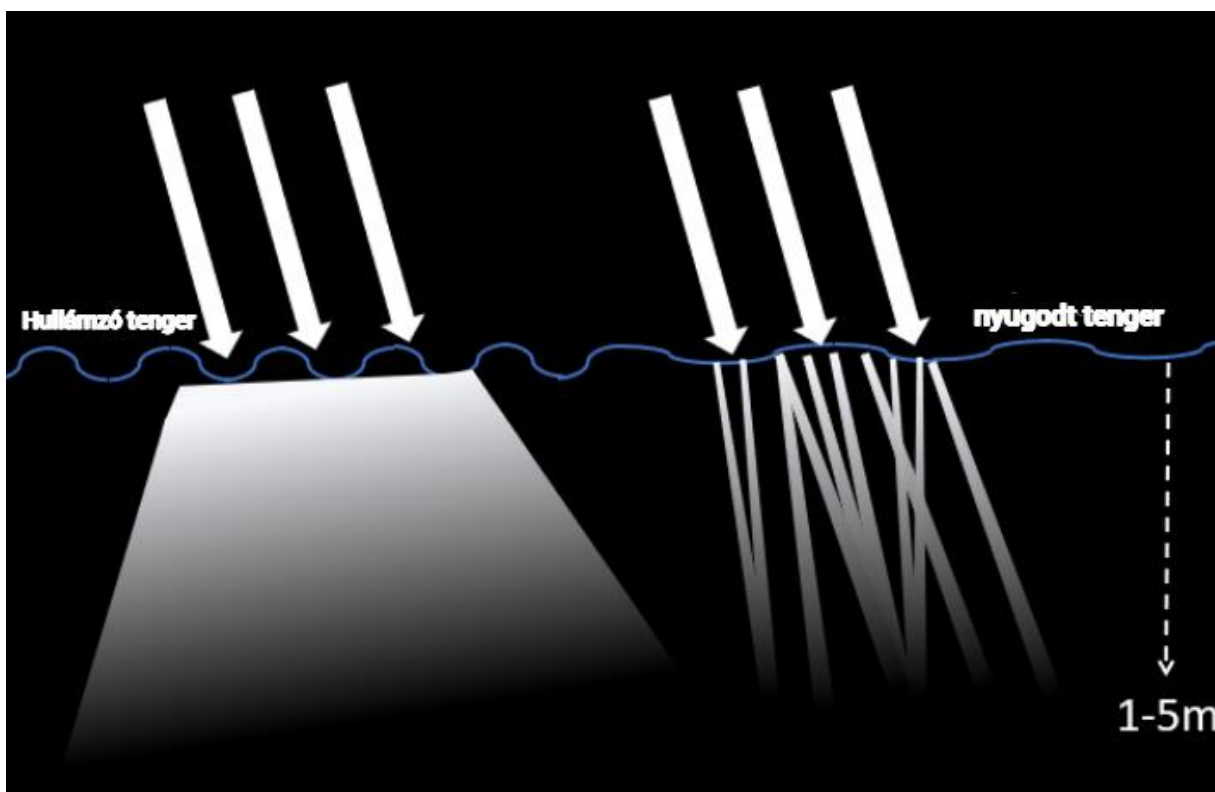
1. Fejlett képjavító technikák integrálása
2. Adaptív mélytanulási architektúrák alkalmazása
3. Tanulás átvitele kiterjedt földi adatkészletekből
4. CNN és „Transformer” modellek kombinációja
5. Figyelem mechanizmusok alkalmazása

Ebben a fejezetben konkrét megoldásokat és megvalósítási ötleteket tárok fel ehhez az öt területhez, a víz alatti képfeldolgozás és mélytanulás területén elért legújabb kutatási és technológiai vívmányokra támaszkodva.

Mielőtt specifikus megvalósításokat elemzenék az egyes technológiák és módszerek tekintetében, amelyek segíthetnek optimalizálni a víz alatti objektumok felismerését, elengedhetetlen, hogy megértsük a víz alatti környezet által támasztott egyedi kihívásokat és problémákat. A víz alatti képalkotás számos kihívással néz szembe a víz és a fény közötti fizikai kölcsönhatás miatt, beleértve a fény elnyelődését és szóródását, a lecsökkent

láthatóságot, az alacsony kontrasztot, az egyenetlen megvilágítást, a zajt, az elmosódást és a színromlást.

A víz alatti környezet specifikus kihívásokat, mint például a diffúziós hatásokat és a nem egyenletes megvilágítást is jelenthet. A víz alatti fényterjedés különleges csillogó mintázattokat hozhat létre, amelyeket a hullámok által okozott felszíni egyenetlenségek tovább erősíthetnek, mely diffúz fényhatást eredményezhet. A nap helyzete, a napszak és az évszakok szintén befolyásolják a visszaverődő fény mennyiségét, ami jelentősen ronthatja a láthatóságot. A látható képeket és formákat továbbá a víz alatti környezetben lévő részecskék és üledékek is befolyásolhatják. A 1. ábra szemlélteti, hogy a behatoló fény függőlegesen polarizált, ami csökkentheti a tárgyak fényességét, míg a vízszintes polarizáció visszaverődik, ami azt is mutatja, hogy a fény hogyan szóródik különböző élővízfelszínek esetén.

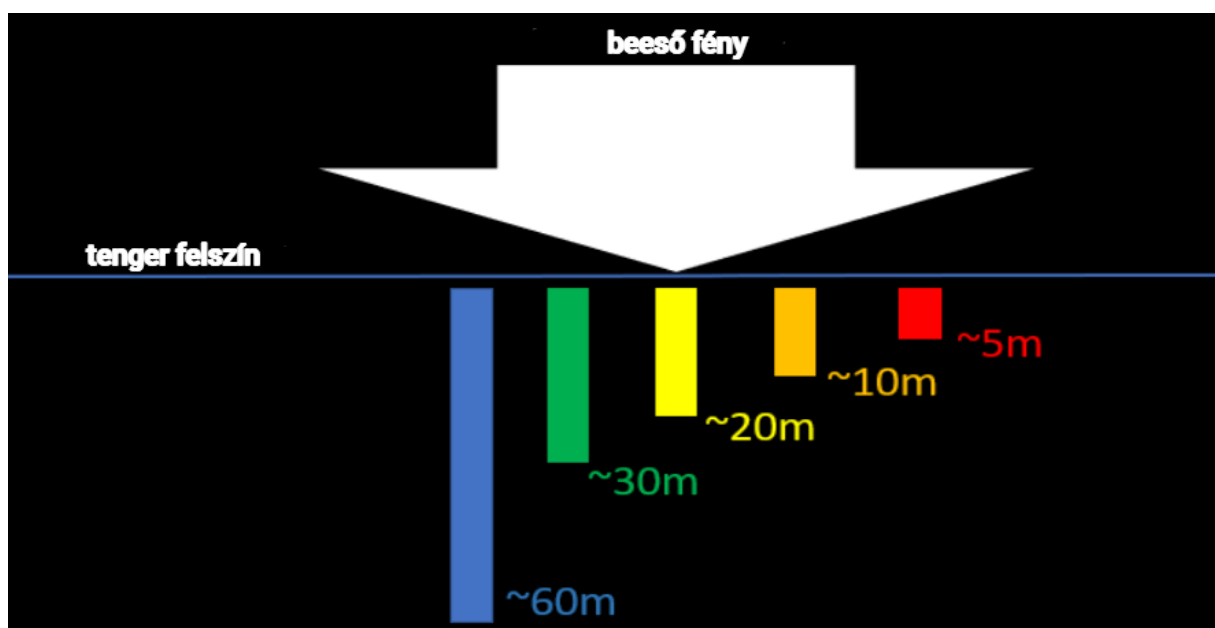


1. ábra - Különböző fényszóródások változó vízfelszíni kondíciók esetén [2].

Ahhoz, hogy ezen fényviszony módosító faktorok egy objektív mérési mód szerint is mérhetővé váljanak, érdemes lehet az élővizek besorolása azok belső és megfigyelt optikai

jellemzői alapján. A belső optikai jellemzők tanulmányozása magában foglalja a víz közegének a fényre gyakorolt hatásainak vizsgálatát, ahol a fény elnyelése és szóródása a fényabszorpcióval és -szóródással kapcsolatos. A megfigyelt optikai jellemzők szintén figyelembe veszik, hogyan befolyásolja a közeg a fény irányát, beleértve a fényforrás geometriai elrendezésének vizsgálatát is. Nils Jerlov által kifejlesztett vízkategóriák rendszere az első olyan kvantitatív osztályozási rendszer, amely a megfigyelt optikai jellemzőkre épül. A Jerlov által definiált élővíz-kategóriák további felosztása partközeli és nyílt óceáni vizekre lehetséges, amelyeket további csoportokba lehet szervezni [3].

A láthatóság objektív mérése és osztályozása azonban további paramétereket igényel, melyek nem csak a vizuális kontúrokra, hanem a színhúségekre is fókuszálnak jelent teljeskörű megoldást az egyes élővíztípusok vizualitással kapcsolatos viszonyainak ismertetésére és skálázásra. További mérési szempont, hogy az egyes objektumok vizualitása mennyire adja vissza azok valódi színét- A látható fény tartományok a 600 nm, 525 nm és 475 nm hullámhosszokon általában a vörös, zöld, és kék színekben érzékelhetőek. A vörös fény nyelődik el a legintenzívebben, ahogy a 2. ábra is mutatja.



2. ábra - A látható fény színtartományai különböző vízmélységekben [4]

Ahogy az 1. és 2. ábrán látható, az élővizetekbe való hatékony objektumdetektálás nem csak az eddigiekben tárgyalt akadályok miatt jelent kihívásokat, hanem látható fénytartományok alapvető fiziológiai kölcsönhatásai miatt is.

3.1 Adatelőkészítés és -tisztítás

A hipotézis és javasolt fejlesztések első pontjával, „A fejlett képjavító technikák integrációjával” közvetlenül foglalkozom ebben az adat-előkészítés és -tisztítás című alfejezetben. A fejlett képjavító technikák integrálása kulcsfontosságú a víz alatti objektumok felismerésének optimalizálása szempontjából, mivel segít a képek minőségének és konzisztenciájának javításában, mielőtt azokat mélytanulási modellekkel feldolgozzák. Ez a rész különféle módszereket és technológiákat tárgyal, amelyek segítségével javítható a víz alatti képek láthatósága, biztosítva, hogy az objektumfelismerő folyamatba betáplált adatok a lehető legjobb minőségűek legyenek.

Mielőtt az objektumfelismerés technikáival detektálunk egy tárgyat vagy élőlényt, a gépi képfeldolgozás általános módszereit vehetjük igénybe annak érdekében, hogy a feldolgozott felvételek minősége és tartalma konzisztens minőséget érjen el az adatok feldolgozása előtt. A gyakorlatban ezeket általános képjavító algoritmusokat jelentik, amelyeket hétköznapi körülmények között is széles felhasználói rétegek használnak a mindennapokban. Például az okostelefon kamerákba épített kontraszt- és zajszűrések, mesterséges színjavítások mostanra általános funkciókká váltak minden okoseszközben. A vízfelszín alatti felvételek esetében azonban további képjavító megoldásokra lehet szükség, mint például a speciális színszabályozási technikák alkalmazása.

3.1.1 Zajscsökkentés és Kontrasztjavítás

Az adatelőkészítés és -tisztítás alapvető lépéseket foglal magában, amelyek célja az adatok minőségének javítása, így biztosítva, hogy az adatelemzés vagy a gépi tanulási modellek alkalmazása során megbízható eredmények születhessenek. Ezen eljárások különösen fontosak

képadatok esetében, ahol a zaj és a kontraszt szintjének optimalizálása elengedhetetlen a képek értelmezhetősége és a modell pontosságának növelése szempontjából.

A képek minőségének javítása két fő módszer alkalmazásával történik: zajcsökkentéssel és kontrasztjavítással. Ezek a technikák kritikus jelentőségűek a képek előkészítésében, különösen olyan speciális környezetekben, mint a víz alatti képfeldolgozás, ahol a képek gyakran szenvednek alacsony kontraszttól és magas zajszinttől.

A zajcsökkentés egy olyan folyamat, amely célja a képeken lévő nem kívánt információ, vagyis a zaj csökkentése anélkül, hogy a kép fontos részletei elvesznének. A zaj típusai változatosak lehetnek, beleértve a szenzor zajt, a szkennert zajt, vagy az alacsony fényviszonyokból adódó zajt. A zajcsökkentés módszerei közé tartozik a lineáris szűrők, mint a Gauss-szűrő, és a nem lineáris szűrők, például a medián szűrő alkalmazása. Ezen technikák intelligens alkalmazása segít a zaj mértékének csökkentésében, miközben megőrzi a kép lényegi tulajdonságait [5].

A kontrasztjavítás egy másik fontos eljárás, amely célja a képek dinamikus tartományának növelése, így javítva azok vizuális értékelhetőségét. Ez különösen hasznos lehet olyan képek esetén, amelyek homályosak vagy túl sötétek. A kontrasztjavításnak számos módszere létezik, többek között az adaptív hisztogram-egyenlítés, amely lokálisan növeli a kontrasztot a kép különböző részein. Egy továbbfejlesztett technika, mint például az adaptív szín- és kontrasztnövelés, valamint egy zajcsökkentési (ACCE-D) keretrendszer, kifejezetten javítja a víz alatti képek minőségét, ahol a környezeti feltételek rendkívül változatosak lehetnek.

Ezen eljárások alkalmazása jelentősen javíthatja a képek minőségét, ami növeli a gépi tanulási modellek pontosságát és javítja az adatelemzés megbízhatóságát. Az ACCE-D keretrendszer kiváló példája annak, hogyan lehet specifikus kihívásokra szabott megoldásokat alkalmazni a képfeldolgozás területén, így biztosítva a magasabb szintű adatminőséget és analitikai pontosságot [6].

3.1.2 Színszabályozás Technikák

A víz alatti képek javításának technológiai kulcsfontosságúak az olyan környezetekben készült képek minőségének növelésében, ahol a fény elnyelődése és szóródása jelentősen befolyásolja a képek minőségét. A kihívást jelentő fényviszonyok miatt ezek a képek gyakran szenvednek alacsony kontraszttól, színeltorzulástól és részletek elmosódásától. Az effajta problémák kezelésére kifejlesztett technológiák két fő kategóriába sorolhatók: színszabályozáson alapuló és szuperpixel sötét csatorna prioritáson alapuló módszerek.

A színszabályozáson alapuló technikák célja, hogy korrigálják a víz alatti környezetben készült képek színhibáit. Ezek a módszerek elsősorban azokat a színeltéréseket célozzák meg, amelyek a víz fizikai tulajdonságai miatt jönnek létre. A színtorzió egyik leggyakoribb megközelítése az automatikus fehéregyensúly-kiegyenlítés és a hisztogram-egyenlítésen alapuló módszerek alkalmazása, amelyek segítségével javítható a képek általános színhűsége és kontrasztja. Azonban vízfelszín alatti felvételeknél a színek torzulása olyan mértékű lehet, amelyre önmagában nem jelent megoldást a fehéregyensúly- kiegyenlítés. A színszabályozási technikák sajátossága, hogy az eredeti színek semmilyen látható tartományban nem kerülnek rögzítésre, ezáltal a valós színek kalkulálása és modellezése is különösen nehéz feladat. A felvételek színoptimalizációs előkészítéshez alkalmazható a Sea-Thru modell, mely képes a vízfelszín alatti felvételek utólagos színtorziójára, a modell a színtorzió helyességét egy speciális tanulmányokon alapuló módszerrel éri el melynek lényege, hogy képmélységi adatokat is figyelembe vesz. A Sea-Thru modell kifejlesztésénél egy úgynevezett RGBD tanuló adathalmazt használtak a modell fejlesztői, amelyben a képek mélységi információi is megtalálhatóak, ez alapján a modell képes általános kétdimenziós víz alatti képe színeit megfelelően megjeleníteni [7]. A 3. ábrán látható a Sea-Thru modell által javított felvétel, a jobb oldali képen látható a modell színtorziós képessége mely minőségében javítja a színtorzulásból adódó problémákat.



3. ábra - SeaTrue modell színeskorrekciója [8]

Szuperpixel Sötét Csatorna Prioritáson alapuló technikák

A szuperpixel sötét csatorna prioritáson alapuló módszerek a képek egy másik jellemzőjét, a sötét csatornát használják fel a víz alatti képek javítására. Ez a technika a képek azon részeit elemzi, amelyek alacsony intenzitásúak a három színcsatornán, és ezeket az információkat használja fel a kép általános minőségének javítására. Ez a módszer különösen hatékony lehet a víz alatti képek színeinek és kontrasztjának javításában, ahol a fény szóródása és elnyelődése miatt a képek gyakran tűnnek elmosódottnak vagy túl kéknek [9].

A színszabályozáson és a szuperpixel sötét csatorna prioritáson alapuló technikák egyaránt jelentős előrelépést jelentenek a víz alatti képek minőségének javításában. Mindkét megközelítés javítja a képek színét és kontrasztját, bár különböző módon. A színszabályozás közvetlenül a színelterések korrigálására összpontosít, míg a sötét csatorna prioritás a képek alacsony intenzitású részeinek elemzésével javítja a kép általános minőségét.

3.1.3 Fény Csökkenés Adaptáció

A fény csökkenésének adaptációja a víz alatti környezetben a konvolúciós rétegek finomhangolásával érhető el, amely a fény spektrális csillapításának kompenzálására szolgál. A mélytanulás és a konvolúciós neurális hálózatok (CNN) alkalmazása lehetővé teszi az algoritmusoknak, hogy a bemeneti képekből közvetlenül tanuljanak alapvető paramétereket, javítva ezzel a képek minőségét azáltal, hogy kezelik a fény csillapítását és javítják a kontrasztot. Az ilyen típusú modellek előnye, hogy a képjavítási folyamatot fizikai koncepciókhoz igazíthatják, bár ez kihívásokkal is járhat a modell általánosító képességének szempontjából [10].

3.2 Adathalmaz Kiterjesztése (Data Augmentation)

Az adathalmazok kiterjesztése szorosan kapcsolódik a hipotézis és a javasolt fejlesztések második pontjához: az „Adaptív mélytanulási architektúrák alkalmazása” -hoz. Az adatkiegészítési technikák sokrétűbb és reprezentatívabb adatkészletek létrehozására használhatók, amelyek elősegíthetik az adaptív mélytanulási architektúrák jobb tanulását és általánosítását a különféle vízfelszín alatti körülményekre.

Minden adatmodell, így a képalkotáshoz használt modellek is akkor a legpontosabbak, ha megfelelő mennyiségű és minőségi tanuló adat alapján alkotják meg őket. Azonban a megfelelő objektumdetektáláshoz nem csak kvantitatív de kvalitatív feltételek is társulnak. A tanuló adathalmaz méretén túl, fontos továbbá, hogy az adathalmazban lévő felvételek mennyire változatosak, illetve milyen minőségűek. Az adatok diverzitása elengedhetetlen ahhoz, hogy a létrejött CNN modell objektum felismerési képességei a lehető legpontosabbak legyenek. Nagy mennyiségű, de egymáshoz hasonló felvételek esetén feltehetőleg a CNN modell túl specifikus tulajdonságokat rögzít ezáltal létrejön az overfitting (túlillesztés) jelensége. A túlillesztés elkerülésére érdekében tehát változatos felvételeken kell az adott gépi tanulás modell tanítását elvégezni. A vízfelszín alatti felvételek számossága általában limitált, hiszen az ilyen típusú felvételek készítése időigényes, továbbá komplex és speciális képalkotó

eszközöket igényel, ezen feltételekből adódóan pedig magasak az elkészítés bekerülési költségei.

3.2.1 Képek Módosításának Módszerei

Amennyiben egy adott számú felvételtől kell minél változatosabb tanuló adathalmazt készíteni lehetőségünk van képek manipulálásra, azáltal, hogy megváltoztatjuk az adathalmaz felvételt, mennyiségileg is növeljük a tanuló adathalmaz méretét. A mennyiségi növekedésen túl ezáltal megelőzhetjük a túlillesztés vagyis az overfitting jelenségét is, mivel egy adott felvételt a modell több különböző verzióban is detektálhat saját tanulási folyamatában. Manipulációs lehetőségek lehetnek a képek elforgatása, ezzel új képeket kapunk melyeken ugyan azon objektumok láthatóak, mégis újabb rész elemei a tanuló adathalmaznak. További kép módosítási lehetőségek a színek változtatása, ebben az esetben a tanuló adathalmaz bővítésén túl, lehetőségünk van megelőzni, hogy a gépi tanulás modellek túltanulják az egyes objektumokat azok színei alapján. A felvételek módosításának egy másik megoldása, ha az elkészület felvételekből interpoláció segítségével új pixelértékeket adunk meg a meglévő pixelek alapján, az eredmény tekintetében ez lehet akár egy egyszerű képbe nagyítás, vagy fókuszváltás. Az adatok manipulálásnak, illetve az adathalmazok bővítésének módszereit az angol nyelvű szakirodalom „data augmentation” -ként nevezi [11].

3.3 Mélytanulási Alapú Javítások

A gépi tanulás modellek számára a megfelelő mennyiségű és minőségű képanyag elérésére nem csak a tanuló adathalmaz bővítésével lehetséges. A bemutatott hagyományos kontrasztjavítások, fehéregyensúly helyreállításokon és mélytanulás alapú színjavításokon túl lehetőségünk további modellek használatára melyek szintén képesek javítani a felvételek minőségén, detektálhatóságán. A képjavító mélytanulás alapú algoritmusok és modellek két kiemelkedő csoportjába tartoznak az úgynevezett „autoencoder”-ek és GAN modellek melyeket specifikusan használhatunk felvételek javítására.

3.3.1 Autoencoderek és GAN-ek Alkalmazása

Az „Autoencoders” és a „Generative Adversarial Networks” (GAN) alkalmazása a képminőség javítására és tisztítására a mélytanulási technológiák területén élvonalbeli megközelítést képvisel. Ezek a módszerek a mélytanulási modellek képességeit alkalmazzák a képek szerkezeti integritásának és vizuális tisztaságának jelentős javítására, ami különösen előnyös a víz alatti képjavítás kontextusában.

Az „autoencoder”-ek olyan neurális hálózatok, amelyek megtanulják a bemeneti adatokat alacsonyabb, kompaktabb formába kódolni, majd visszakódolni az eredeti formájukba. Ez a folyamat lehetővé teszi a képekből értelmes, magas szintű jellemzők kinyerését, például szerkezeti információkat, amelyek elengedhetetlenek a kép alapvető jellemzőinek megőrzéséhez a javítási folyamat során.

A „Generatív Adversarial Networks”-ök (GAN-ok) ezt folyamatot egy lépéssel tovább viszik azáltal, hogy két hálózatot használnak: egy generátort réteget, amely képeket készít, és egy megkülönböztető réteget, amely kiértékeli azokat. Ezáltal tulajdonképpen egy kompetitív folyamatot hoznak létre, mely minőségében javított új képek létrehozását eredményezi. Konkrét példa erre a CycleGAN, amely figyelemre méltó sikereket ért el a színekorekcióban és a víz alatti képek helyreállításában [12].

Ezenkívül az „Auto-Embedding Generative Adversarial Network” (AEGAN) az „autoencoder”-ek és a GAN-ok kifinomult keverékét példázza. Az AEGAN az automatikus kódolók által kinyert kis felbontású beágyazásokat használja fel, hogy gazdagítsa a GAN modellek teljesítményét, a valósághűen javított mintaképek létrehozásában. A gazdag szerkezeti információt tartalmazó beágyazások kihasználásával az AEGAN javítja a képgenerátorok azon képességét, hogy értelmes és vizuálisan koherens képeket készítsen [13].

Összefoglalva, a mélytanulás, különösen a CNN-ek és GAN-ok alkalmazása a víz alatti képjavításban létfontosságú a pontos és hatékony vizuális információ-helyreállítás eléréséhez. Ezeknek a modelleknek az alkalmazása nem csak a vízfelszín alatti képek javításában, de a tanuló adathalmaz növelésében is számos előnnyel járhat, mivel további diverzitást lehet velük eszközölni az adathalmazon belül.

3.4 Transfer Learning

Az empirikus kutatások alapján a transzfertanulás megfelelő eszköz lehet a hipotézis harmadik pontjához és a javasolt fejlesztésekhez: „Tanulás átvitele kiterjedt földi adatkészletekből”. Ez az alfejezet közvetlenül foglalkozik az előre betanított modellekből származó ismeretek nagyméretű földi adatkészleteken történő felhasználásának fontosságával és előnyeivel a víz alatti objektumfelismerő rendszerek teljesítményének javítása érdekében.

Mint minden informatikai feladatban, így a mélytanuló modellek fejlesztése is elsődleges szempont az egyes megoldások újra felhasználhatósága, illetve az ezáltal elérhető fejlesztési idő lerövidítése, valamint a költségek csökkentése. A mélytanuló modellek esetén is egy hatékony fejlesztési paradigma a már meglévő modellek újra felhasználása, illetve konfigurálása a jelenlegi fejlesztési feladatokhoz. Ez a megközelítés különösen előnyös, ha az új feladathoz korlátozott a tanuló adatok mennyisége, vagy ha a modell nulláról történő betanítása számítási szempontból költséges. Egy nagy adathalmazra előre betanított modell felhasználása egy hasonló feladatra, költséghatékonyan kivitelezhető viszonylag kisebb módosításokkal és finomhangolással. Ez nemcsak jelentősen csökkenti a modell képzési idejét, valamint szükséges számítási erőforrásokat, hanem gyakran a modell teljesítményének javulását is eredményezheti az új feladaton. Például egy képfelismerésre előképzett neurális hálózat újra felhasználható speciális feladatokra, például vízfelszín képek diagnosztizálására vagy specifikus objektumok, élőlények elemzésére. Az modellek újra felhasználását az angol szakirodalom „transfer learning”-nek vagyis „transzfer tanulásnak” nevezi.

3.4.1 Előre Kiképzett Modellek Használata

Az előre betanított konvolúciós neurális hálózatok (CNN-ek), mint például a VGG és a ResNet alkalmazása a mélytanulási projektekben elterjedt gyakorlattá vált, ami jelentős előnyöket kínál a tanulási folyamat felgyorsításában és a modell hatékonyságának növelésében. Ez különösen előnyös olyan adatkészleteknél, ahol a betanítási adatok mennyisége korlátozott, mint például vízfelszín alatti felvételeken alapuló adathalmazok.

Az olyan modellek, mint a Simonyan és Zisserman által bevezetett VGG, valamint a He és munkatársai által kifejlesztett ResNet, fontos szerepet játszanak a mélytanuláson alapuló projektek előmozdításában. A kiterjedt, általános adatkészletek, például az ImageNet-re kiképzett modellek súlyokkal és funkciókkal vannak felszerelve, amelyek képesek felismerni a képadatok összetett mintáit. Mély rétegiek és kifinomult architektúráik ideálissá teszik őket képfelismerési feladatokra. Például a ResNet alkalmazása víz alatti képelemzésre lehetőséget biztosíthat a tengeri élőlények vagy a víz alatti objektumok észlelését és osztályozását [14].

Előre kiképzett modellek használata esetén, előnyeik ellenére döntő fontosságú a túlillesztés kockázatának kezelése, különösen kisebb vagy specifikusabb adatkészletek esetén. A probléma forrása, hogy olyan előre kiképzett generális adathalmazzal dolgozó modellek, mint a ResNet, ugyan kiváló eredményeket produkálnak általános alkalmazásban, de finomhangolás nélkül speciális környezetben lehetséges, hogy túltanulnak bizonyos objektum formákat. A finomhangolási technikák segítségével, ahol a modell végső rétegét egy adott feladatra áttanítják, miközben a korábbi rétegek súlyát rögzítették, hatékonyan csökkentik az túlillesztés (overfitting) kockázatát [15].

3.4.2 Transfer Learning alkalmazásának előnyei

Az transzfer tanulás előnyei, mint például a képzési adatok iránti csökkentett igény és a gyorsabb konvergencia, különösen nyilvánvalóak a víz alatti képfelismerés területén. A transzfertanulás egy mélytanulási technika, amely lehetővé teszi a már meglévő

modellismeretek alkalmazását olyan kapcsolódó feladatok megoldására, amelyek eltérnek az eredeti feladattól, amelyre a modellt kidolgozták. Ez a megközelítés jelentős előnyökkel jár a korlátozott képzési adatokkal rendelkező forgatókönyveknél, és célja a modell betanítási idejének csökkentése.

Gyorsabb konvergencia

A transzfertanulás nemcsak csökkenti a szükséges képzési adatok mennyiségét, hanem felgyorsítja a konvergencia rátáját a képzés során. Mivel az előre betanított modellek már rendelkeznek alapvető ismeretekkel a képek általános jellemzőiről, a modell egy adott feladatra történő finomhangolása kevesebb iterációt igényel a kívánt teljesítményszint eléréséhez. Ez jelentős időmegtakarítást eredményez, különösen a szoros határidőkkel rendelkező projekteknél és a kutatás-fejlesztési prototípuskészítés korai szakaszában.

Hatékonyság növelése

Egy eredetileg általános képfelismerésre kiképzett ResNet modell adaptálása a víz alatti felvételek speciális kihívásaihoz – mint például a halfajok felismerése vagy a víz alatti építmények észlelése – lehetővé teszi a fejlesztők számára, hogy gyorsabban és kevesebb erőforrás felhasználásával érjenek el nagy pontosságot.

Összefoglalva, az olyan előre betanított modellek, mint a VGG és a ResNet, jelentősen növelik a mélytanulási projektek hatékonyságát a képzési idő és a szükséges erőforrások csökkentésével. A túlillesztés (overfitting) kockázatának figyelemmel kísérése és a megfelelő technikák, például a finomhangolás és az adatbővítés alkalmazása azonban továbbra is kulcsfontosságú a modellek optimális teljesítményének biztosításához. Ez különösen fontos az olyan speciális területeken, mint a víz alatti képszintézis, ahol egyedi környezeti kihívásokkal kell foglalkozni. Az transzfer tanulás előnyei, mint a kevesebb tanítóadat szükségessége és a gyorsabb konvergencia, kulcsfontosságúak a mélytanulási projektek sikeréhez, különösen adatkorlátozott környezetben, mint a vízfelszín alatti felvételek.

3.5 Figyelem Mechanizmusok

Ez az alfejezet közvetlenül foglalkozik a figyelemmechanizmusok, például a nem helyi neurális hálózatok és a Grad-CAM technika beépítésének fontosságával és előnyeivel a mélytanulási modellek teljesítményének javítása érdekében a víz alatti objektumfelismerési feladatokban.

Minden elkészült felvételen találhatóak az adott képfelvétel szempontjából irreleváns információk, ahogy az emberi felismerés is általában egy adott kép fókuszpontját detektálja először. A releváns és nem releváns információk szeparálása egy természet evolúciós folyamat, a fizikai térben is először a hozzá közelebb eső tárgyat detektálja az emberi szem. A vizuálisan érzékelt ingerek szeparálásának oka, hogy az emberi érzékelés és memorizálás tehát a figyelem kapacitásai limitáltak. A mélytanulás figyelmen alapuló mechanizmusa egy olyan számítási modell, amely utánozza az emberi figyelem folyamatát, és lehetővé teszi a neurális hálózatok számára, hogy a bemeneti adatok meghatározott részeire összpontosítsanak, amelyek relevánsabbak az adott feladat szempontjából. Ez a mechanizmus jelentősen javítja a különböző gépi tanulási feladatok teljesítményét, beleértve a CNN modelleket is.

3.5.1 Figyelem-alapú Mechanizmusok

A figyelem alapú mechanizmus alapvető működése magában foglalja a bemeneti adatok minden részének súlyozását, amely az adott rész fontosságát vagy relevanciáját jelenti az adat szempontjából. Például a képfeldolgozás kontextusában a mechanizmus nagyobb súlyt rendelhet azokhoz a pixelekhez, amelyek egy érdekes objektum részét képezik, lehetővé téve a modell számára, hogy jobban összpontosítson ezekre a pixelekre. A fontos részletek kiválasztását jellemzően egy tanítható pontozási funkció határozza meg, amely a figyelem súlyait számítja ki. Ez a funkció a modell sajátos architektúrájától és a feladat jellegétől függően többféleképpen is megtervezhető. A kiszámított figyelemsúlyokat ezután a bemeneti adatok súlyozott összegének létrehozására használják, amely kontextusvektort alkot. Ez a kontextusvektor a bemeneti adatok reprezentációja, amely kiemeli a legrelevánsabb részeket, hatékonyan lehetővé téve a modell számára, hogy „figyeljen” rájuk.

Ahhoz, hogy a legrelevánsabb részelemek ne kapjanak túl nagy súlyozást egy modellen belül, a figyelem mechanizmusok általában normalizálják a kapott súlyokat. A figyelemsúlyokat általában egy „softmax” funkcióval normalizálják, biztosítva, hogy összegük egy legyen. Ez a normalizálás lehetővé teszi a modell számára, hogy hatékonyan ossza el fókusz a bemeneti adatok között, és a nagyobb súlyok nagyobb fókuszú területeket jelezenek, de a terület mérete is arányos maradjon.

A „Non-local Neural Networks” figyelmi mechanizmusok alkalmazását a szövegfelismerés mellett, hasonló koncepcióját alkalmazzák a kép- és videófeldolgozási feladatokra. Ezek a hálózatok azonosítják a vizuális adatok távoli szegmenseit, mégis azok releváns kapcsolataira összpontosítanak, lehetővé téve számukra a hosszú távú függőségek rögzítését, amelyeket a hagyományos konvolúciós neurális hálózatok (CNN) gyakran figyelmen kívül hagynak. Például a képbesorolás vagy a víz alatti képeken belüli objektumészlelés során a „Non-local Neural Networks”-ök a teljes jelenetre fókuszálhatnak, hogy észleljék és osztályozzák az objektumokat, még olyan kihívást jelentő körülmények között is, ahol rosszak a láthatósági paraméterek. A kép vagy videó keret távoli részei közötti kapcsolatok kiemelésével a „Non-local Neural Networks”-ök javítják a modell képességeit a vizuális tartalom átfogó értelmezésében.

Lényegében a figyelem alapú hálózatok alkalmazzák a figyelemmechanizmusok sokoldalúságát és hatékonyságát a mélytanulásban. Azáltal, hogy lehetővé teszik a modellek számára, hogy dinamikusan hozzárendeljék a fókusz és a számítási erőforrásokat a bemeneti adatok legrelevánsabb részeihez, ezek a mechanizmusok jelentős fejlesztéseket tesznek lehetővé az összetett adatszerkezetek feldolgozásában és megértésében, vizuális formában [16].

3.5.2 Releváns Jellemzők Kiemelése

A mélytanuló modellek „figyelmének” egy másik hatékony eszköze lehet a relevancia alapú megkülönböztetés, például a „Grad-CAM” technika. A megoldás lényege, hogy bármely célkonceptió (például egy CNN által detektált osztály) gradiensének felhasználásával, amely a végső konvolúciós rétegbe áramlik, egy durva lokalizációs térkép létrehozása. Ez a lokalizációs térkép kiemeli a kép fontos régióit egy konceptió előrejelzéséhez. A Grad-CAM hőképet állít elő úgy, hogy kiszámítja az adott osztály pontszámának gradiensét egy konvolúciós réteg jellemzőtérkép-aktiválásaihoz képest, majd a gradienseket a szélességi és magassági dimenziók mentén összevonja (globális átlag, vagyis „pooling”), hogy megkapja a fontossági súlyokat minden egyes létrejött tereptérképhez. Ez a folyamat vizuális magyarázatokat eredményez a CNN-ek döntéseihöz, így könnyebben megérthető, hogy a kép mely részei befolyásolták a modell előrejelzését [17].

3.6 Hibrid Modellek

A hibrid modellek technológiájával megvalósítható a CNN és transzformátor modellek kombinációja, ezáltal megfelelő eszközt biztosít a hipotézis megvizsgáláshoz. Ez az alfejezet hibrid modellek használatának előnyeit tárgyalja, mint például a konvolúciós neurális hálózatok (CNN-ek) kombinálása a visszatérő neurális hálózatokkal (RNN). Az RNN és CNN modellek kombinálásán túl bemutatom a „transformers” modellek hibrid alkalmazását is.

A hibrid modellek használata a víz alatti tárgyak észlelésére a képeken rendkívül előnyös lehet a víz alatti környezet összetett és változatos természete miatt. Ez a megoldás lehetővé teszi az objektumok robusztusabb észlelését az egyes modelltípusok erősségeit kihasználva. Például homályos vagy rosszul látható vízviszonyok között a hibrid modellek jobban felismerik a tárgyakat a háttérből a kontextuális és szekvenciális jelzések integrálásával, ami pontosabb és megbízhatóbb észlelési eredményeket eredményez.

3.6.1 CNN-ek és RNN-ek Kombinálása

A konvolúciós neurális hálózatok (CNN-ek) és az ismétlődő neurális hálózatok (RNN-ek) kombinálása egy összetett „robosztus” módszert kínál a képadatok összetett jellemzőinek értelmezésére. Ez a megközelítés mindkét hálózattípus erősségeit kihasználja: a CNN-ek hatékonyan nyerik ki és dolgozzák fel a képjellemzőket, míg az RNN-ek kiválóak az időbeli vagy szekvenciális adatok kezelésében, lehetővé téve a képsorozatokban vagy videókban található szekvenciafüggő információk értelmezését. Ez a hibrid módszer különösen előnyös olyan alkalmazásoknál, amelyekben szükség van a mozgások és események időbeli kontextusának megértésére, a képsorozatok közötti objektumok követésére, valamint részletes képleírások vagy videótartalom-elemzés létrehozására, ahol a folytonosság vizsgálata döntő fontosságú. A térbeli elemzéshez szükséges CNN-ek és az időbeli megértést szolgáló RNN-ek integrálásával ez a kombinált megközelítés javítja a vizuális felismerési, leírási és szemantikai igazítási feladatok pontosságát [18].

A CNN és RNN modellek együttes alkalmazása azonban leginkább a videofelvételek esetén lehet hasznos, hiszen az RNN-ek időbeli dinamikákat is képesek detektálni. A videofelvételek kapcsán elért eredményektől függetlenül az önálló képek esetén is ígéretes eredményeket érhetnek el a hibrid modellek. Az RNN és CNN modellek kimagasló példája az úgynevezett hosszú távú ismétlődő konvolúciós hálózat (LRCN), amely a konvolúciós és az ismétlődő hálózatokat kombinálja a vizuális felismeréssel járó feladatokhoz. Az LRCN-ek hatékonyan kezelnek változó hosszúságú vizuális bemeneti szekvenciákat és ez a hibrid megközelítés lehetővé teszi a modell számára, hogy a térbeli jellemzőket megtanulja a CNN-n keresztül, míg az időbeli dinamikát az RNN-eken keresztül, ez a megközelítés ígéretes eredményeket mutat a videofelismerési, képfelismerési feladatokban [19].

3.6.2 Transformer Modellek Integrálása

A „Transformer” modellek integrálása a konvolúciós neurális hálózatokkal (CNN-ekkel) egy másik lehetséges megközelítés a mélytanulásban, amelynek célja, hogy javítsa a hosszú távú függőségek és a komplex kapcsolatok modellezését a képi adatokban. Ez a kombináció kihasználja a „transformer” modellek erős képességeit a hosszú távú kontextusok és az összefüggések érzékelésére, míg a CNN-ek hatékonyan dolgozzák fel a térbeli információkat. Hasonlóan az LRCN modellekhez, a „transformer” modellek képesek mélyebb, kontextusban gazdagabb reprezentációkat létrehozni, ami növeli a modellek pontosságát és értelmezési képességét [20].

3.7 Regularizáció és Finomhangolás

Ebben a fejezetben egy alternatív lehetőséget vázolok fel a „Fejlett képjavító technikák integrációja” és az „Adaptív mélytanulási architektúrák alkalmazása” hipotézis pontokra. Ez a rész a regularizációs technikák és finomhangolási módszerek alkalmazásának fontosságát tárgyalja a túlillesztés megelőzésére, valamint a mélytanulási modellek stabilitásának és általánosítási teljesítményének javítására a víz alatti objektumfelismerési feladatokban.

A CNN-modellek úgynevezett túlillesztése akkor fordul elő, ha a modell túl jól megtanulja a betanítási adatokat, például egy adott felvételen a zajt vagy más kiugró értékeket, ami rossz általánosításhoz vezethet új, még nem látott adatok esetében. A túlillesztés jelenségének eredménye, hogy bizonyos mintákat nagyon jó értékekkel ismer fel az adott modell, azonban más alakzatok esetén az detektálás fals eredményeket mutat.

3.7.1 Regularizációs Technikák Alkalmazása

A regularizációs technikák alkalmazása kulcsfontosságú a mélytanulási modellek túlillesztésének megelőzésében és a modell detektálási stabilitásának növelésében. A szabványosítás, vagyis regularizáció egy olyan technika, amely megakadályozza a túlilleszkedést a modell bemeneti értékeinek megváltoztatásával, például egy szorzótag

hozzáadásával a veszteségi függvényhez, vagy olyan speciális rétegek használatával, amelyek véletlenszerűen kihagyják a jellemzők egy részét a képzés során. A regularizációs technikák azáltal is kezelhetik a modellek komplexitását, hogy felszoroznak nagy súlyértékeket, ezzel javítva a modellek általánosítási teljesítményét. Ezek a módszerek segítenek abban, hogy a modell megtanuljon robusztus, jól általánosító jellemzőket, ahelyett, hogy csak memorizálná a betanítási adatokat.

L1 és L2 regularizációk

Az L1 és L2 regularizáció, vagyis a Ridge és Lasso regularizációk, olyan technikák, amelyeket a gépi tanulási modellekben alkalmaznak az „overfitting”, vagyis a túlillesztés megelőzésére. Ezek a módszerek az egyes súlyokra úgynevezett „nehezítő értékeket” alkalmaznak, ezzel korlátozva azok nagyságát és elősegítve egy egyszerűbb modell létrehozását, ami erősebben általánosít az ismeretlen adatokra.

Az L1 regularizáció, vagy Lasso (Least Absolute Shrinkage and Selection Operator) a súlyok abszolút értékének összegét adja hozzá a veszteségfüggvényhez. Ez a típusú nehezítés arra kényszeríti a modellt, hogy csökkentse a súlyok értékét, valamint néhány súly esetében akár pontosan nullára is csökkentheti azokat. Ennek eredményeképpen a Lasso regularizáció hozzájárulhat a modell "ritkítéséhez" azzal, hogy kiküszöböli azokat a jellemzőket, amelyek nem járulnak hozzá jelentős mértékben a kimenet előrejelzéséhez. Ez a jelenség különösen hasznos lehet, amikor szeretnénk meghatározni a bemeneti változók közül melyek a legfontosabbak a kimenet szempontjából, javítva ezzel a modell interpretálhatóságát.

Az L2 regularizáció, vagy Ridge a súlyok négyzetének összegét adja hozzá a veszteségfüggvényekhez. Ez a típusú „nehezítés” nem vezet a súlyok ritkítéséhez, ahogy az L1 teszi, hanem inkább arra ösztönzi a modellt, hogy oszlassa el a súlyokat egyenletesen a jellemzők között, ahelyett, hogy néhány nagyon nagy súlyra támaszkodna. Az L2 regularizáció hatására a modell súlyai kisebbek lesznek, de nem válnak pontosan nullává. Ez segít a túlillesztés csökkentésében anélkül, hogy jelentős jellemzők lennének teljesen kizárva a modelltől [21].

A gépi tanulás kontextusában a „nehezítés” a veszteségfüggvényhez hozzáadott további értékre utal, amelyet a modell a képzés során minimálisra kíván csökkenteni. Ez a megoldás hatékonyan befolyásolja a modellt, hogy a kisebb vagy egyszerűbb súlyértékű megoldásokat részesítse előnyben. A „nehezítés” hozzáadásának elsődleges célja a túlillesztés megelőzése azáltal, hogy meg akadályozza a modellt abban, hogy túl szorosan illeszkedjen az edzési adatokhoz, különösen azáltal, hogy összetett vagy nagy súlyú konfigurációkra hagyatkozik, amelyek esetleg nem általánosíthatók jól a nem látható adatokra.

A „nehezítések” beépítése megváltoztatja azt az optimalizálási folyamatot és problémát, amelyet a modell a képzés során megold. Szabályozás nélkül a modell kizárólag az eredeti veszteségfüggvény alapján optimalizál, jellemzően az előre jelzett és a tényleges értékek közötti különbség csökkentésére törekszik. A regularizációval a modellnek a nehezítésekből származó eltérések mértékét is minimalizálnia kell, ami azt jelenti, hogy egyensúlyba kell hoznia az adatok jó illesztését (minimalizálja az előrejelzési hibákat) a súlyok kis értéken való tartásával (a büntetések minimalizálásával).

Ez a fajta „egyensúlyozás” olyan modellekhez vezet, amelyek kisebb valószínűséggel illeszkednek túl a tanuló adatokhoz, mivel bonyolultságuk miatt büntető értékeket kapnak (a súlyuk nagyságával mérve). A regularizálás lényegében az egyszerűbb modellek irányába torzítást vezet, azzal a várákozással, hogy az egyszerűbb modellek jobban általánosítanak új adatokra, mint a túl bonyolultak.

Gyakorlatilag a nehezítés mértékét általában egy „hyperparameter” szabályozza, minél nagyobb ennek a „hyperparameter”-nek az értéke, annál erősebb a büntetés, és annál inkább az egyszerűség felé tolódik a modell. A „hyperparameter”-nek a megfelelő érték kiválasztása kulcsfontosságú az alul- és túlillesztés közötti jó egyensúly eléréséhez.

„Early Stopping” vagyis korai leállítás

Az „Early Stopping” vagyis korai leállítás a gépi tanulási modellek, különösen a mély neurális hálózatok betanításában használt szabályozási technika a túlillesztés megelőzésére. A korai leállítás technikája ezt úgy kezeli, hogy a modell teljesítményét a betanítási adatoktól különálló validációs adatkészleten figyeli. A képzési folyamat leáll, ha a modell teljesítménye a validációs halmazon már nem javul, feltételezve, hogy a továbbképzés csak túlillesztéshez vezet. A korai leállítás oka abban a megfigyelésben gyökerezik, hogy a képzés kezdeti szakaszában a modell megtanul általánosítani a betanítási adatokból, de a képzés előrehaladtával egyre inkább elkezdi memorizálni a betanítási adatokat, beleértve a zajt és a kiugró értékeket is. Ez a fajta memorizálás a modell azon képességének romlásához vezet, hogy jól teljesítsen a nem látott adatokon [22].

A „dropout” vagyis elhagyási rétegek alkalmazása

A „dropout” rétegek a neurális hálózatok túlilleszkedésének csökkentésére szolgáló elsődleges módszerként szolgálnak a modern mélytanulási korszakban. Ezek a rétegek ideiglenesen kihagyják a hálózat bizonyos csomópontjait a képzés során, megakadályozva ezzel a csomópontok közötti túlzott adaptációt. A „dropout” rétegek alkalmazása jelentős mértékben javítja a hálózatok általánosítási képességét, különféle architektúrák esetén is, csökkentve a túlilleszkedés problémáját [23].

3.8 A kísérleti adatmodell terveinek megalkotása és fejlesztése

Annak érdekében, hogy a disszertáció keretein belül egy gyakorlatias és megvalósítható víz alatti objektum-észlelési modellt építhessek, egy olyan optimális megközelítésre összpontosítok, amely hatékonyan kezeli a probléma megvalósítható aspektusait, figyelembe véve a rendelkezésre álló erőforrásokat és időkorlátokat. Ennek a fókuszált megközelítésnek az a célja, hogy szilárd alapot biztosítson a modell későbbi továbbfejlesztésére és további megoldások tesztelésére.

A javasolt modell az EfficientNetB7 architektúrát fogja használni, amely kivételes teljesítményt mutatott a képosztályozási feladatokban, miközben viszonylag alacsony számítási költséget igényel [24]. Az EfficientNetB7 optimalizált architektúrája, amely egyensúlyban tartja a hálózat mélységét, szélességét és felbontását, ideális választássá teszi célzott víz alatti objektum-észlelési modellemhez. Ennek az architektúrának a kihasználásával nagy pontosságot kívánok elérni a víz alatti objektumok észlelésében és osztályozásában, miközben minimalizálom a képzéshez és a következtetésekhez szükséges számítási erőforrásokat.

A modell betanításához és értékeléséhez a „Sea Animals Image Dataset”-et használom, amely víz alatti képek sokféle gyűjteménye és kifejezetten a tengeri állatokra és korallokra összpontosít. Ez az adatkészlet különféle fényviszonyok között rögzített képeket tartalmaz, jól megvilágított és gyengén megvilágított képeket egyaránt. Az objektumtípusok egy meghatározott halmazára koncentráló adatkészlet használatával hatékonyan fejleszthetem és tesztelhetem a modellt a dolgozat keretein belül, miközben csökkentem a képzési folyamat bonyolultságát és számítási követelményeit.

Transzfer tanulási technikákat fogok alkalmazni, hogy az EfficientNetB7 modell előzetes betanításából szerzett tudást hasznosítsam olyan nagyméretű adatkészleteken, mint az ImageNet. Az előre betanított súlyok finomhangolása a „Sea Animals Image Dataset”-en lehetővé teszi, hogy jelentősen csökkentsem a képzési időt miközben javítom a modell teljesítményét. Ez a megközelítés lehetővé teszi a modell számára, hogy kihasználja a forrástartományból tanult általános képi jellemzőket és adaptálja azokat a víz alatti objektumok észlelésének sajátos jellemzőihez.

A modell teljesítményének és robusztusságának további javítása érdekében adatbővítési technikákat alkalmazok, beleértve a véletlenszerű kivágást, átfordítást és elforgatást. Ezen technikák segítenek növelni a képzési minták sokféleségét és csökkenteni a túlillesztést, lehetővé téve a modell számára, hogy jobban általánosítson a víz alatti képekre. Ezen túlmenően implementálok regularizációs módszereket, mint például az L2-regularizáció és az

„early stopping” technikája, hogy megelőzzük a túlillesztést és javítsuk a modell általánosító képességét.

Fontos hangsúlyozni, hogy a disszertáció célja egy célzott és hatékony víz alatti objektum-észlelési modell kidolgozása EfficientNetB7 architektúra, transzfertanulás és adatkiegészítési technikák felhasználásával. Míg a legösszetettebb technikák, például a színvisszaállítás és a figyelemmechanizmusok nem szerepelnek a jelenlegi megvalósításban a kivitelezési korlátok miatt, az ebből a fókuszált megközelítésből nyert meglátások értékes alapot jelentenek a jövőbeli kutatáshoz és fejlesztéshez a víz alatti objektumok észlelése terén.

Összefoglalva, megközelítesemben prioritásként kezelem a dolgozat keretein belül egy praktikus és hatékony víz alatti objektum-észlelési modell kidolgozását. Az EfficientNetB7 architektúrára, az átviteli tanulásra, az adatbővítésre és a „Sea Animals Image Dataset”-re összpontosítva nagy pontosságot és robusztusságot kívánok elérni a víz alatti objektumok észlelésében és osztályozásában.

3.8.1 A kísérleti modell felépítése

A javasolt modell az EfficientNetB7 architektúrát fogja használni, amely egy nagymértékben optimalizált konvolúciós neurális hálózat (CNN) és kivételes teljesítményt mutatott a képosztályozási feladatokban, miközben viszonylag alacsony számítási költséggel jár [24]. Az EfficientNetB7 összetett skálázási módszert alkalmaz, amely egyenletesen skálázza a hálózat mélységét, szélességét és felbontását, kiegyensúlyozott és hatékony architektúrát eredményezve. Ennek a legmodernebb architektúrának a kihasználásával nagy pontosságot kívánok elérni a víz alatti objektumok észlelésében és osztályozásában, miközben minimalizálom a képzéshez és a következtetésekhez szükséges számítási erőforrásokat.

A modell rétegei

1. Bemeneti réteg: A bemeneti réteg 224x224 pixel fix felbontású RGB képeket fogad, ami kompatibilis az EfficientNetB7 architektúrával.

2. EfficientNetB7 „Base”: Az előre betanított EfficientNetB7 modellt fogok használni a jellemzők jelölésére, kihasználva azt a képességet, hogy megkülönböztető jellemzőket rögzítsen a bemeneti képekből. Az előre betanított modell súlyait a képzés kezdeti szakaszában lefagyasztom, hogy megőrizték a tanult mintákat.

3. „Global average pooling layer”: Egy „Global average pooling” réteg kerül hozzáadásra az EfficientNetB7 alap rétegére, hogy ezáltal csökkentsük az adatok térbeli méreteit és hogy fix hosszúságú jellemzővektort kapjanak az értékek.

4. Dense layers: Két teljesen összekapcsolt „dense” réteget adok a modellhez ReLU aktiválással, hogy megtanulják a magas szintű reprezentációkat és rögzítsék a kivont jellemzők közötti összetett kapcsolatokat. Az első „dense” réteg 128, míg a második 256 egységneuronból áll.

5. Dropout: A túlillesztés elkerülése és a modell általánosítási képességének javítása érdekében minden sűrű réteg után 0,45-ös elhagyási rétegeket adok hozzá. A „Dropout” véletlenszerűen nullára állítja a bemeneti egységek egy részét a tanulási folyamat során, ami segít a modellnek robusztusabb funkciók megismerésében.

6. Kimeneti réteg: Az utolsó réteg egy „dense” réteg „softmax” aktiválási funkcióval, amely valószínűségi eloszlást hoz létre az adatkészletben található különböző tengeri állatkategóriáknak megfelelő 23 osztály között.

A modellfejlesztés lépései

A víz alatti objektum-észlelési modell fejlesztése szisztematikus megközelítést követ, biztosítva a tengeri állatok képadatkészletének hatékony hasznosítását és a legjobb gyakorlatok megvalósítását a mélytanulásban. A modellfejlesztési folyamat legfontosabb lépései a következők:

1. Adatok előfeldolgozása: A „Sea Animals Image Dataset” előfeldolgozásra kerül az EfficientNetB7 architektúrával való konzisztencia és kompatibilitás biztosítása érdekében. A képeket 224x224 pixelre méretezem át, és az EfficientNetB7-re jellemző megfelelő előfeldolgozási funkcióval normalizálom.

2. Adatfelosztás: Az adatkészlet képzési és tesztelési készletekre lesz felosztva, az adatok 80%-a betanításra, a fennmaradó 20%-a pedig tesztelésre lesz fenntartva. Ez a felosztás lehetővé teszi a modell teljesítményének értékelését nem látott adatokon, és segít felmérni általános képességeit.

3. Adatbővítés: A képzési adatok sokféleségének növelése és a modell robusztusságának javítása érdekében adatkiegészítési technikákat alkalmazok. Ezek a technikák magukban foglalják a véletlenszerű vízszintes tükrözést, az elforgatást, a nagyítást és a kontraszt beállítását. Az adatbővítés segít a modellnek megtanulni az invariáns jellemzőket, és csökkenti a túlillesztés jelenségét.

4. Transzfertanulás: A transzfertanulást az EfficientNetB7 modell ImageNet adatkészleten történő előzetes betanítása során szerzett ismeretek hasznosítására alkalmazom. Az előre betanított súlyok a modellbe kerülnek, és az alaprétegeket kezdetben lefagyasztom. Ez a megközelítés lehetővé teszi a modell számára, hogy minél pontosabb predikciókat adjon a nagyszabású adatkészletből tanult jellemzőábrázolások alapján.

5. Modell összeállítás: A modell egy optimalizálóval, mint például „Adam” függvénnyel és a kategorikus keresztentropia veszteségfüggvénnyel kerül összeállításra. A tanulási rátát kis

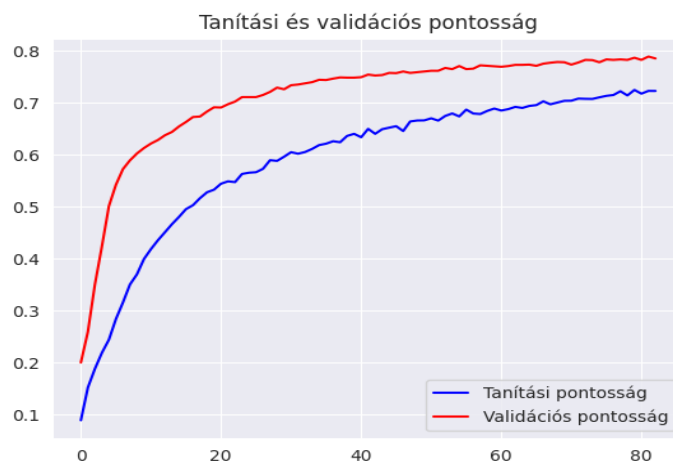
értékre állítom be (pl. 0,00001), hogy biztosítsam a stabil tanulási folyamatot és megakadályozzuk az optimális súlyok túllépését.

6. Modellképzés: A modellt betanítására megfelelő mennyiségű tanulási kört (epoch) alkalmazok (pl. 100). Korai leállítást és modellellenőrzést alkalmazok a túlillesztés megelőzése és a legjobban teljesítő modell elérése érdekében.

7. A modell értékelése: A betanított modellt megfelelő mérőszámok segítségével értékelem a teszt adatokon, illetve megvizsgálom és értelmezem a Grad-Cam vizualizáció eredményeit.

3.8.2 A kísérleti modell teljesítmények értékelése

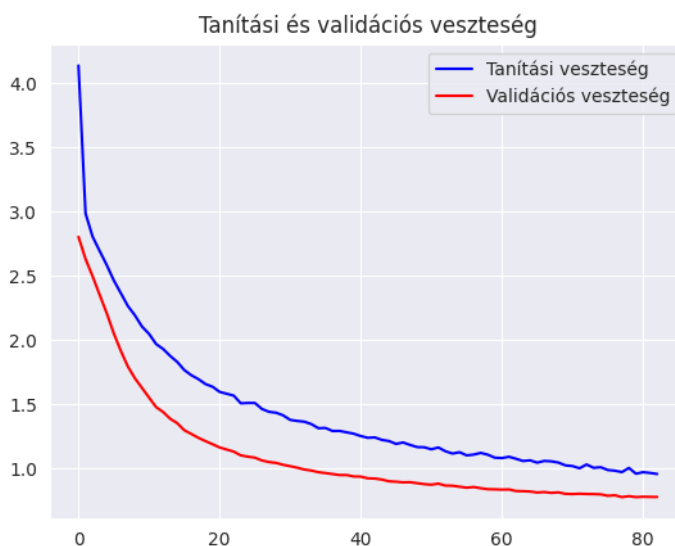
A következőkben megvizsgálom a kísérleti adatmodell milyen eredményeket ért el a „Sea Animals Image Dataset”-en végzett betanulás utáni predikciók tekintetében.



4. ábra- A kísérleti adatmodell „Tanítási” és „validációs” pontossága ("Training accuracy" és "Validation accuracy")

Az 4. ábra a modell pontosságát szemlélteti a képzési „epoch”-ok során. A tanítási pontosság (kék vonal) és a validációs pontosság (piros vonal) konvergál, ami azt jelzi, hogy a

modell nemcsak a betanítási adatokra, hanem a nem látott validációs adatokra is jól általánosítható a betanítás során. A pontossági értékek az „epoch”-ok számával nőnek mindkét halmaz esetében, jelezve a modell fejlődését a tanulási folyamatban.



5. ábra - A kísérleti adatmodell „Tanítási” és „validációs” vesztesége ("Training loss" és "Validation loss")

Az 5. ábra „Tanítási és validációs veszteség” a modell veszteségi vagy hibaértékét mutatja be az egyes „epoch”-ok során. A tanítási veszteség (kék vonal) és a validációs veszteség (piros vonal) a képzés előrehaladtával csökken, ami jelzi a modell javulását. Az a tény, hogy a validációs veszteség nem növekszik szignifikánsan a „epoch”-ok során, jó jel, mivel arra utal, hogy a modell nem illeszkedik túl a képzési adatokhoz.

Ezen megfigyelések alapján több következtetést is levonhatunk:

1. A modell idővel folyamatosan javul, amit a fokozatosan növekvő pontosság bizonyít mind a képzési, mind az érvényesítési halmazok esetében.
2. A betanítás és a validálás pontossága közötti kis különbség azt jelzi, hogy a modell jól általánosít, és nem mutat jelentős túlillesztést.

3. A veszteséggörbék alakja azt mutatja, hogy a modell tanul és csökkenti az iterációk során fellépő hibákat, valamint nincs lényeges eltérés a betanítási és validációs veszteség között, ami a jó általánosítási képesség másik jele.

Gyakorlati szempontból ezek az eredmények azt sugallják, hogy a modell valószínűleg jól teljesít új, nem látott adatokon. Ha a validációs veszteség növekedni kezd, miközben a képzési veszteség tovább csökken, az egyértelmű jele lenne a túlillesztésnek, ami azt jelzi, hogy a modell a képzési adatokban lévő zajt tanulja meg, nem pedig az általános mintákat. Ebben az esetben azonban ilyen viselkedés nem figyelhető meg, ami azt jelenti, hogy a modell jól teljesít a validációs adatokon is.

A következő 1. számú táblázatban láthatóak a modell egyes mérőszámai a különböző tengeri állatkategóriákra vonatkozóan. A táblázat számos mérőszámot tartalmaz: pontosság, „visszahívás” (recall), F1-pontszám és „támogatás” (support).

Osztály	Pontosság	Visszahívás (recall)	F1-pontszám	Támogatás (support)
Kagylók	0.52	0.62	0.57	92
Korallok	0.48	0.81	0.60	99
Rákok	0.92	0.98	0.95	101
Delfin	0.75	0.86	0.80	146
Angolna	0.88	0.56	0.68	117
Halak	0.74	0.72	0.73	112
Medúzák	0.92	0.95	0.93	166
Homár	0.84	0.60	0.70	109
Tengeri meztelenkagylók	0.75	0.77	0.76	112
Polip	0.56	0.41	0.47	122
Vidra	0.98	1.00	0.99	94
Pingvin	0.87	0.73	0.80	94
Pufferhalak	0.77	0.65	0.71	100
Tengeri sugár	0.82	0.70	0.75	86
Tengeri sünök	0.95	0.98	0.97	107
Csikóhal	0.57	0.73	0.64	88
Fóka	0.90	0.89	0.90	74
Cápák	0.75	0.74	0.74	118
Garnélarák	0.54	0.54	0.54	99
Tintahal	0.64	0.62	0.63	87
Tengeri csillagok	0.94	1.00	0.97	105
Teknősök	0.93	0.97	0.95	391
Bálna	0.74	0.65	0.70	124
Pontosság			0.78	2743
Makró átlag	0.77	0.76	0.76	2743
Súlyozott átlag	0.79	0.78	0.78	2743

1. táblázat - A kísérleti adatmodell klasszifikációs riport táblázata

A pontosság méri a valódi pozitív előrejelzések arányát az összes pozitív előrejelzés között. Más szavakkal, megválaszolja a kérdést: "Az összes olyan eset közül, amelyet a modell egy adott osztályhoz tartozónak jósolt, hány volt valójában helyes?" Például a "Vidra" osztály pontossága 0,98, ami azt jelzi, hogy amikor a modell vidraként jósol egy képet, akkor az esetek 98%-ában helyes.

A „Recall” ezzel szemben számszerűsíti a modell azon képességét, hogy azonosítani tudja az összes pozitív példányt. Megválaszolja a következő kérdést: "Az összes példány közül, amelyek valóban egy adott osztályhoz tartoznak, hányat azonosított helyesen a modell?" A "Tengeri csillagok" osztályt nézve 1,00-as „recall”-t láthatunk, ami azt jelenti, hogy a modell sikeresen azonosítja az adatkészletben lévő összes tengeri csillagot, a jellegzetes formából adódóan ez egy kifejezetten könnyű mintázat.

Az F1-pontszám a pontosság és a „recall” harmonikus átlaga, amely kiegyensúlyozott értékelést nyújt a modell teljesítményéről. A magas F1-pontszám azt jelzi, hogy a modell jó egyensúlyt ér el a pontosság és a „recall”-ok között. A „Medúzák” osztály például 0,93-as F1-pontszámmal rendelkezik, ami arra utal, hogy a modell jól teljesít a medúza azonosításában és a hamis pozitív eredmények elkerülésében.

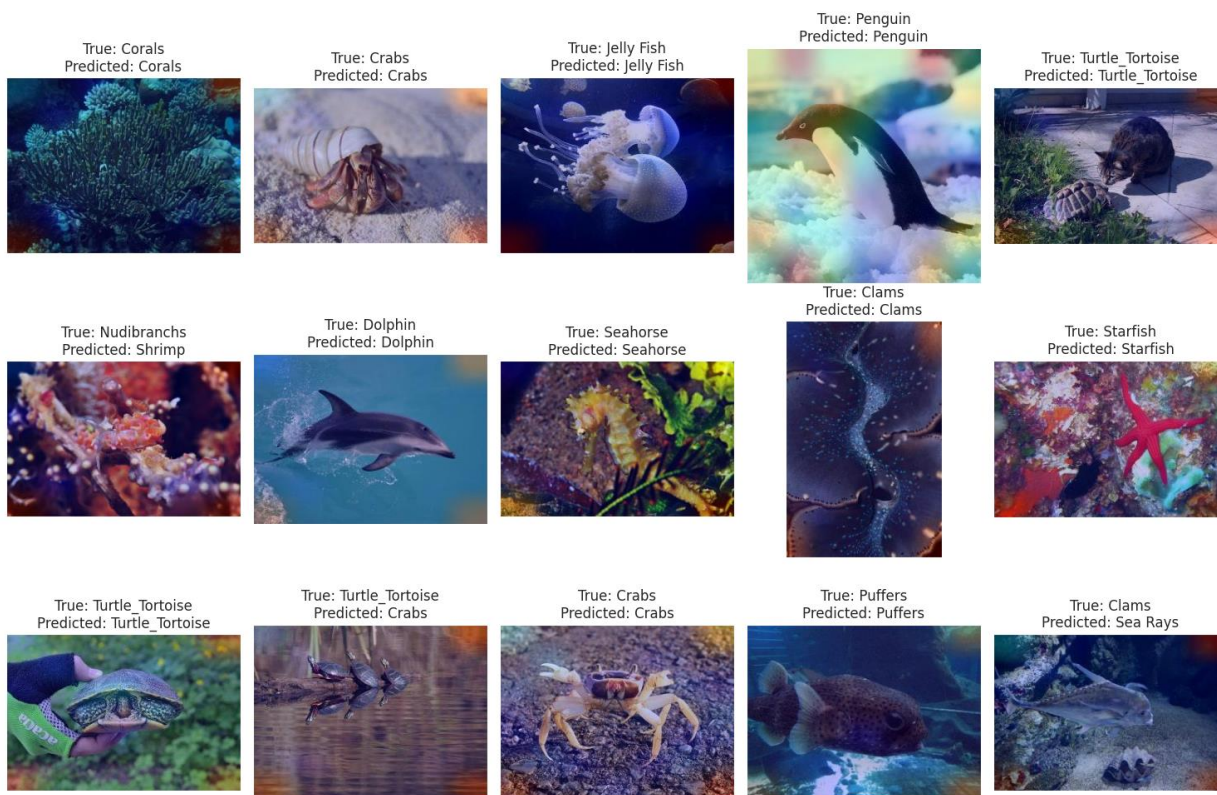
A „support” az adatkészlet egyes osztályaihoz tartozó valódi példányok számát jelenti. Segít megérteni az osztályeloszlást, és azonosítja a lehetséges osztálykiegyensúlyozatlansági problémákat. Ebben az esetben a „Teknősök” osztály a legmagasabb, 391-es, míg a "Fóka" osztály rendelkezik a legalacsonyabb, 74-es „support”-al.

A makróátlag minden osztályra külön-külön kiszámítja a pontosságot, a „recall”-t és az F1-pontszámot, majd a súlyozatlan átlagot veszi, és minden osztályt egyformán kezel. A makró átlagos F1-pontszám 0,76, ami azt jelzi, hogy a modell átlagosan észszerűen jól teljesít minden osztályban.

A súlyozott átlag ezzel szemben figyelembe veszi az egyes osztályok „support” értékeit, nagyobb jelentőséget tulajdonítva a magasabb „support” értékű osztályoknak. A 0,78-as súlyozott átlagos F1-pontszám arra utal, hogy a modell valamivel jobban teljesít, ha figyelembe vesszük az adathalmaz osztályeloszlását.

Az egyes osztálymutatókat tekintve több megfigyelést is tehetünk. A modell nagy teljesítményt (F1-pontszám > 0,90) ér el az olyan osztályokban, mint a „Rákok”, „Vidra”, „Tengeri sünök”, „Tengeri csillag” és a „Teknősök”. Ez azt jelzi, hogy a modell különösen hatékony ezen tengeri állatkategóriák azonosításában és osztályozásában.

A modell azonban gyengébben teljesít bizonyos osztályoknál, mint például a „kagylók” (F1-pontszám: 0,57), a „korallak” (F1-pontszám: 0,60), a „polipok” (F1-pontszám: 0,47) vagy a „rákok” (F1) -pontszám: 0,54) esetében. Ezek az alacsonyabb pontszámok azt sugallják, hogy a modellnek nehézségei lehetnek ezen állatok megkülönböztetésében másoktól, vagy nagy lehet az osztályon belüli variabilitás, ami kihívást jelent a modell számára, hogy megtanulja a meghatározó jellemzőket.



6. ábra - A kísérleti adatmodell Grad-CAM vizualizációi

A 6. ábrán látható Grad-CAM vizualizációkban megfigyelhetjük, hogy a modell hogyan fókuszál a bemeneti képek meghatározott régióira, hogy ezáltal előrejelzéseket készítsen. A hőterképek kiemelik azokat a területeket, amelyek a leginkább hozzájárulnak a modell besorolási döntéseihez.

Az első öt képen a modell sikeresen azonosítja és lokalizálja a tengeri állatok releváns részeit, például a teknősök és kagylók páncélját, a cápák uszonyait és a rákok ollóit. A hőterképek ezekre a megkülönböztető jellemzőkre koncentrálnak, jelezve, hogy a modell megtanulta felismerni ezen állatok legfontosabb vizuális jellemzőit.

A sorozat 6. (jobbról második sor elsőelem) képén a „rák” előrejelzésnél azonban láthatunk egy példát, ahol a képminőség kihívás elé állítja a modellt. Úgy tűnik, hogy a kép rosszul van megvilágítva, és zöldes-kékes árnyalatú, ami valószínűleg a víz alatti környezetnek

köszönhető. A képhez tartozó Grad-CAM hő térkép szétszórtabb és kevésbé fókuszál az állat sajátosságaira. Úgy tűnik, hogy a modell inkább a háttérre és a környező területekre fókuszál, nem pedig magának a tengeri lénynek a sajátosságaira. Ez a példa kiemeli a képminőség hatását a modell teljesítményére, vagyis az alkalmazott technológiák nem minden esetben elégségesek a víz alatti objektumdetektáláshoz.

A modell eredményeinek és teljesítményének összefoglalása

A modell teljesítménye a „Sea Animals Image Dataset”-en bizonyítja a modell hatékonyságát a különböző tengeri állatok, például rákok, medúzák, vidrák, tengeri sünek, tengeri csillagok valamint teknősök pontos észlelésében és osztályozásában, nagy pontossággal, és magas F1 pontértékekkel. A betanítási és validálási pontossági görbék konvergálnak, jelezve a modell azon képességét, hogy jól általánosít a nem látott adatokra, míg a veszteségi görbék mind a betanítási, mind a validációs tengelyeken következetes csökkennek, tovább erősítve a modell tanulási képességét.

Az elemzés azonban rávilágít azokra a kihívásokra is, amelyeket a gyengébb minőségű víz alatti képek jelentenek, amint az a Grad-CAM vizualizációkból is kitűnik. Azokban az esetekben, ahol a képek gyenge megvilágítástól, színtorzulástól vagy zavarosságtól szenvednek, a modell fókusza általában szétszórtabb, valamint kevésbé releváns vagy félrevezető vizuális jelzésekre is fókuszál. Ez a megállapítás hangsúlyozza a további kép-előfeldolgozási technikák szükségességét, mint például a színvisszaállítási és a zajcsökkentési technikák, hogy ezáltal tovább javuljon a víz alatti képek vizuális minősége, mielőtt betáplálásra kerülnek a modellbe.

A modell teljesítményének és robusztusságának további javítása érdekében kihívást jelentő víz alatti környezetben több kutatási irány is feltárható. Először is, a legkorszerűbb szín-helyreállítási módszerek, mint például a hipotetikus optimális modell leírásban (1. melléklet) említett Sea-thru modell beépítése jelentősen javíthatja a víz alatti képek vizuális minőségét, és megkönnyítheti a pontosabb objektumfelismerést. A Sea-thru modell alkalmazásával a víz

alatti képeken gyakran előforduló színtorzulások és kékes-zöld árnyalatok korrigálhatók, ami természetesebb és kiegyensúlyozottabb képeket eredményezhet.

Másodszor, az adaptív architektúrák, például a Neural Architecture Search (NAS) technikák vizsgálata automatikusan optimalizálhatja a hálózati struktúrát és a „hyperparameter”-eket a bemeneti adatok jellemzői alapján, lehetővé téve a modell számára, hogy jobban kezelje a víz alatti környezet változó körülményeit. Ezenkívül a figyelemmechanizmusok objektumészlelő hálózatba történő integrálása segíthet a kép leginkább informatív területeire összpontosítani és figyelmen kívül hagyni a nem releváns háttérinformációkat.

Ezenkívül a fejlett objektumészlelési architektúrák, például a YOLO (You Only Look Once) vagy az SSD (Single Shot MultiBox Detector) alkalmazásával potenciálisan javulhat a modell valós idejű teljesítménye és pontossága. Ezek az architektúrák kiváló teljesítményt mutattak a földi objektum-észlelési feladatokban és a fent említett színhelyreállítási és figyelemmechanizmusok beépítésével a víz alatti adatkészletekhez is jól illeszthetők [27].

Ezek a strukturális fejlesztéseken túlmenően, ha a képzési adatkészletet sokféle víz alatti képpel bővítjük, különféle vízviszonyokat, objektumkategóriákat és képminőségeket rögzítünk, ez tovább javíthatja a modell általánosítási képességét. Egy nagyszabású víz alatti adatkészlet, például a hipotetikus (1. melléklet) modellben említett URPC10K adatkészlet alkalmazása és annotálása elegendő képzési adatot szolgáltat a mélytanulási modellekhez, lehetővé téve a robusztus és általánosítható jellemzők megtanulását.

Összefoglalva, ez a fejezet átfogó elemzést ad a mélytanulási technikák segítségével a víz alatti objektumok észlelésének kihívásairól és lehetséges megoldásairól. Az olyan kulcsfontosságú területekre összpontosítva, mint a fejlett képjavító technikák, az adaptív mélytanulási architektúrák, a kiterjedt földi adatkészletekből való tanulás átvitele, a CNN és a transzformátor modellek kombinációja, valamint a figyelemmechanizmusok alkalmazása. Továbbá a fejezet különféle stratégiákat vizsgál a teljesítmény és a robusztusság optimalizálásának céljából. Az EfficientNetB7 architektúrán, az átviteli tanuláson és az

adatkiegészítési technikákon alapuló kísérleti adatmodell ígéretes eredményeket mutatott a tengeri állatok pontos észlelésében és osztályozásában a „Sea Animals Image Dataset”-ből. A fejezet azonban feltárja azokat a kihívásokat is, amelyeket a rosszabb minőségű víz alatti képek esetén továbbra is fennállnak. E kihívások kezelése és a modell teljesítményének további javítása érdekében jövőbeli kutatási irányokat javaslok, beleértve a legmodernebb színvisszaállítási módszereket, például a Sea-thru modell integrálását, az adaptív architektúrák, például a „Neural Architecture Search” (NAS) technikák alkalmazását, valamint figyelemmechanizmusok és fejlett objektumészlelési architektúrák, például YOLO vagy SSD beépítését. Az azonosított kihívások kezelésével és a javasolt fejlesztések beépítésével a jövőbeli kutatás jelentősen javíthatja a kísérleti adatmodell teljesítményét és robusztusságát.

4 ÖSSZEFOGLALÁS

4.1 Megvalósult Célok Értékelése

A dolgozat keretein belül sikeresen feltártam a víz alatti számítógépes látás alapú objektumdetektálás technológiai kihívásait, valamint mélytanulási módszereket és algoritmusokat értékeltem. Az elvégzett kutatás értékes betekintést nyújtott a jelenlegi legmodernebb technikákba és azok hatékonyságába a víz alatti környezet által támasztott egyedi kihívások kezelésében.

A dolgozat sikeresen azonosította és elemezte azokat a kulcsfontosságú tényezőket, amelyek befolyásolják az objektumészlelési modellek teljesítményét víz alatti körülmények között, mint például a képminőség romlása, a színtorzulás és az annotált adatkészletek szűkössége. Az átfogó szakirodalmi áttekintésen és empirikus kísérletezésen keresztül rávilágítottam a fejlett képjavító technikák, az adaptív mélytanulási architektúrák és a kiterjedt földi adatkészletekből származó tanulás átvitelének fontosságára.

Továbbá a dolgozat keretében sikeresen kifejlesztésre és értékelésre került egy kezdetleges kísérleti adatmodellt, amely az EfficientNetB7 architektúrán, transzfertanuláson és adatkiegészítési technikákon alapul. A modell ígéretes eredményeket mutatott a különböző tengeri állatok pontos észlelésében és osztályozásában a „Sea Animals Image Dataset”-ből, bemutatva a mélytanulási módszerekben rejlő lehetőségeket a víz alatti objektumfelderítési feladatokban. Továbbá, a kutatás értékes betekintést nyújtott a modell döntéshozatali folyamatának a Grad-CAM technikák alkalmazásával történő megjelenítésébe és értelmezéséhez is. Ez javította a modell viselkedésének megértését és a releváns képi régiók azonosítását, hozzájárulva az osztályozási döntésekhez.

Összességében a dolgozat sikeresen elérte elsődleges célját, a technológiai kihívások feltárását és a mélytanulási módszerek értékelését a víz alatti tárgyak észleléséhez. A kutatás eredményei és megfigyelései hozzájárulnak a fejlesztés kutatás további elmélyítéséhez a mellékelt hipotetikus optimális modell elemeinek további kidolgozásával.

4.2 Meg Nem Valósult Célok Elemzése

A dolgozatban elért jelentős előrelépés ellenére bizonyos célkitűzések a kutatási folyamat során felmerülő kihívások és korlátok miatt nem valósultak meg. Ezek a kihívások elsősorban a víz alatti környezet összetett természetéből és a nagyméretű, változatos adatkészletek megszerzésének és annotálásának velejáró nehézségeiből adódnak. A kutatás során az egyik legnagyobb kihívás az átfogó, annotációkkal ellátott víz alatti képadatkészletek korlátozott elérhetősége volt. Bár a „Sea Animals Image Dataset” értékes forrást nyújtott a modellképzéshez és -értékeléshez, még mindig hiányzik belőle a víz alatti állapotok és tárgykategóriák széles skálájának teljes körű rögzítéséhez szükséges sokféleség is. Az ilyen kiterjedt adatkészletek létrehozását a víz alatti adatgyűjtéssel és annotációval járó logisztikai nehézségek és magas költségek túlmutattak ezen dolgozat lehetőségein.

Egy másik kihívás, amely megakadályozta bizonyos célok teljes megvalósítását, a mélytanulási modellek nagyméretű adathalmazokon történő képzésének számítási bonyolultsága volt. A dolgozat az EfficientNetB7 architektúrát használta, amely bár rendkívül hatékony, de jelentős számítási erőforrásokat és betanítási időt igényel. A számítási infrastruktúra korlátjai és az időkorlátok limitálják a fejlettebb architektúrák feltárását és a „hyperparameter”-ek finomhangolását az optimális teljesítmény érdekében. Továbbá a dolgozat elsősorban a látható fény spektrumára összpontosított a víz alatti objektumok észlelésére. A hipotetikus optimális modellben (1. melléklet) említett multispektrális és mélységi adatok integrálása azonban nem valósult meg teljes mértékben a kutatás keretein belül. A további adatkezelési módok beépítése javíthatja a modell teljesítményét azáltal, hogy kiegészítő információkat biztosít, és kezeli a víz alatti környezet jelentette kihívásokat.

Az értekezés felismerte a tartomány-specifikus transfertanulás és az előre betanított modellek víz alatti tartományra történő adaptálásának fontosságát is. Míg a kutatás kimutatta a földi adatkészletekből történő tanulás átvitelének hatékonyságát, a finomhangolási folyamatot tovább lehetne optimalizálni a víz alatti képek egyedi jellemzőinek jobb rögzítése érdekében.

A víz alatti specifikus előképzési adatkészletek fejlesztése és a fejlett tartomány-adaptációs technikák alkalmazása továbbra is feltáratlan kutatási területe marad.

Összefoglalva, bár a dolgozat jelentős előrelépést tett a mélytanulás segítségével a víz alatti objektumok észlelésének kihívásainak kezelésében, bizonyos célkitűzések a probléma összetett természete és a felmerülő korlátok miatt nem valósultak meg. A meg nem valósult célkitűzések elemzése rávilágít a folyamatos kutatási és fejlesztési erőfeszítések szükségességére az egyes kihívások leküzdése és a víz alatti számítógépes objektum detektálás határainak kutatása érdekében.

4.3 A Továbbfejlesztés Lehetőségei

A kutatás eredményei és megfigyelései alapján számos ígéretes út javasolható a jövőbeli fejlesztésekhez. Először is, a fejlett színhelyreállítási technikák, például a hipotetikus optimális modellben említett Sea-thru modell (1. melléklet) integrálása jelentős potenciált rejt magában a víz alatti objektumészlelő rendszerek teljesítményének javításában. A víz közeg által okozott színtorzulások és csillapítások hatékony kezelésével ezek a technikák jelentősen javíthatják a víz alatti képek vizuális minőségét, megkönnyítve a pontosabb és megbízhatóbb tárgyfelismerést. A további fejlesztésnek és kutatásnak az ilyen színhelyreállító algoritmusok fejlesztésére és finomítására kell összpontosítania, amelyek kifejezetten a víz alatti környezet kihívásaihoz igazodnak.

Másodszor, az adaptív mélytanulási architektúrák, például a „Neural Architecture Search„ (NAS) technikákon alapuló megoldások izgalmas utakat kínálnak a modell további fejlesztésére. Ezek az architektúrák képesek automatikusan optimalizálni szerkezetüket („hyperparameters”) a bemeneti adatok jellemzői alapján, lehetővé téve, hogy jobban alkalmazkodjanak a víz alatti képek változó feltételeihez. A NAS technikák kihasználásával az újabb modellváltozatokat úgy lehet megtervezni, hogy dinamikusan igazítsák összetettségüket és detektálási képességeiket, ami jobb teljesítményt és robusztusságot eredményez a különböző víz alatti környezetekben.

Harmadszor, a figyelemmechanizmusok objektumészlelési folyamatba való beépítése ígéretes irányt kínál a modell azon képességének fokozására, hogy a kép leginkább informatív területeire fókuszáljon. Az olyan technikák, mint az önfigyelem, a csatorna szerinti figyelem és a térbeli figyelem használhatók a fontossági súlyok dinamikus hozzárendeléséhez a különböző jellemzőkhöz és régiókhoz, lehetővé téve a modell számára, hogy a legrelevánsabb információkat priorizálja a pontos objektumészlelés érdekében. A figyelemmechanizmusok integrálása potenciálisan javíthatja a modell azon képességét, hogy kezelje a szóródásokat és az objektum léptékének változásait, amelyek gyakori kihívást jelentenek a víz alatti képeken.

Ezenkívül a fejlett objektumészlelési architektúrák, például a YOLO (You Only Look Once) vagy az SSD (Single Shot MultiBox Detector) alkalmazása potenciálisan javíthatja a víz alatti objektumészlelési rendszerek valós idejű teljesítményét és pontosságát. Ezek az architektúrák kiváló teljesítményt mutattak a földi objektum-észlelési feladatokban, és a fent említett színhelyreállítási technikák és figyelemmechanizmusok beépítésével a víz alatti felvételekhez illeszthetők. A tengeri környezet egyedi jellemzőire optimalizált víz alatti objektumészlelési architektúrák fejlesztése olyan terület, amely további kutatást és fejlesztésre ad lehetőséget.

Végül, a felügyelet nélküli és félig felügyelt tanulási technikák fejlesztése a víz alatti objektumok észlelésére ígéretes irány a következő kutatás számára. Tekintettel a nagyméretű, annotált víz alatti adatkészletek szükségére, a felügyelet nélküli tanulási megközelítések, például az „autoencoder”-ek vagy a GAN-ok kihasználása lehetővé teheti a modell számára, hogy értelmes reprezentációkat tanuljon meg címkézetlen adatokból. A félig felügyelt tanulási technikák, amelyek kis mennyiségű annotált adatot címkézetlen adatok nagy készletével kombinálnak, szintén segíthetnek csökkenteni a kiterjedt kézi annotálásra való hagyatkozást. Hatékony, felügyelt és félig felügyelt tanulási stratégiák kifejlesztése a víz alatti objektumok észlelésére jelentősen enyhítheti az annotálások szűk keresztmetszetét, és megkönnyítheti ezen rendszerek valós helyzetekben való alkalmazását.

FOGALOMTÁR

AlexNet - az első mély konvolúciós neurális hálózat, amely jelentősen javította a képosztályozási teljesítményt [23]

R-CNN - Region-based Convolutional Neural Networks,

Fast R-CNN - az R-CNN gyorsított és továbbfejlesztett változata [23]

Faster R-CNN - valós idejű objektumdetektálásra optimalizált, az R-CNN-en alapuló hálózat [23]

YOLO - You Only Look Once - valós idejű objektumdetektáló algoritmus [28]

SSD - Single Shot MultiBox Detector - valós idejű objektumdetektáló algoritmus [27]

RetinaNet - egy egylépéses objektumdetektáló hálózat, amely a fókuszvesztés kezelésével kezeli az osztály-egyensúlytalanság problémáját [25]

CenterNet - objektumdetektáló hálózat, amely az objektumok középpontját és méretét becsüli [23]

ViT - Vision Transformers - képosztályozásra és detektálásra használt transzformátor modell [26]

CNN - Convolutional Neural Network [5]

UOD - Underwater Object Detection - víz alatti objektumdetektálás [25]

SWIPENET - Sample-Weighted Hyper Network - zajos környezetben történő objektumfelismerésre tervezett hálózat [25]

FAGR - Factor-Agnostic Gradient Re-weighting - osztálykiegyensúlyozatlansági problémák kezelésére szolgáló algoritmus [25]

UIS - Underwater Image Synthesis - víz alatti képszintézis [25]

CycleGAN - képjavításra használt generatív hálózat [12]

WaterGAN - víz alatti képek generálására és javítására használt GAN modell.

ChinaMM - víz alatti képadatkészlet [25]

CMA - Curriculum Multi-class Adaboost - mélytanulási detektorok sorozatos képzésére használt módszer [25]

NECMA - CMA zajcsökkentő szakasza [25]

NLCMA - CMA zajtanulási szakasza [25]

URPC2017 - vízfelszín alatti képfelvételek adatkészlete [25]

URPC2018 - vízfelszín alatti képfelvételek adatkészlete [25]

URPC2019 - vízfelszín alatti képfelvételek adatkészlete [25]

NR - Noise Removal - zajeltávolítás [25]

mAP - mean Average Precision - objektumdetektálás értékelési metrikája [25]

IoU - Intersection over Union - objektumdetektálás értékelési metrikája [25]

ACCE-D keretrendszer - Adaptive Color and Contrast Enhancement with Denoising - adaptív szín- és kontrasztjavítás zajcsökkentéssel [6]

GAN - Generative Adversarial Network – generatív hálózat [23]

AEGAN - Auto-Embedding Generative Adversarial Network – automata kódolókat használó GAN [13]

VGG - konvolúciós neurális hálózataarchitektúra [14]

ResNet - konvolúciós neurális hálózataarchitektúra [15]

EfficientNet - hatékony konvolúciós hálózatmodell képosztályozásra [24]

Sea Animals Image Dataset – tengeri élővilágot ábrázoló képgyűjtemény, standardizált paraméterekkel [29]

IRODALOMJEGYZÉK

[1] Aboseada, M. M. F. A. (2020): Underwater Object Detection Using Deep Learning (Víz alatti objektumfelismerés mélytanulással). (MSc szakdolgozat) Információtechnológiai Kar, Tallinni Műszaki Egyetem. Témavezető: Tuhtan, J. A., Konzulens: Remmas, M. W. 13. oldal.

[2] Aboseada, M. M. F. A. (2020): Underwater Object Detection Using Deep Learning (Víz alatti objektumfelismerés mélytanulással). (MSc szakdolgozat) Információtechnológiai Kar, Tallinni Műszaki Egyetem. Témavezető: Tuhtan, J. A., Konzulens: Remmas, M. W. 18.

[3] Johnson, L.J. (2015): The Underwater Optical Channel. DOI:10.13140/RG.2.1.1295.7283.
https://www.researchgate.net/publication/280050464_The_Underwater_Optical_Channel.
(Letöltés ideje: 2023-11-12) 13-16.

[4] Aboseada, M. M. F. A. (2020): Underwater Object Detection Using Deep Learning (Víz alatti objektumfelismerés mélytanulással). (MSc szakdolgozat) Információtechnológiai Kar, Tallinni Műszaki Egyetem. Témavezető: Tuhtan, J. A., Konzulens: Remmas, M. W. 20.

[5] Guang Deng, L.W. Cahill (1993): An adaptive Gaussian filter for noise reduction and edge detection. Proceedings of the 1993 IEEE Conference Record Nuclear Science Symposium and Medical Imaging Conference. DOI: 10.1109/NSSMIC.1993.373563.

[6] Li, X., Hou, G., Li, K. & Pan, Z. (2021): Enhancing Underwater Image via Adaptive Color and Contrast Enhancement, and Denoising. arXiv:2104.01073.

[7] Akkaynak, D. (2019): Sea-Thru: A Method for Removing Water From Underwater Images. Proceedings of the 2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). DOI: 10.1109/CVPR.2019.00178. ISBN: 978-1-7281-3293-8, 978-1-7281-3294-5. ISSN: 2575-7075, 1063-6919.

[8] Yuval, M. (2019): [Kép az "Sea-thru" algoritmus használatáról, amely feltárja az igazi színeket a víz alatt]. Business Insider. <https://www.businessinsider.com/sea-thru-algorithm-reveals-true-colors-underwater-2019-11> (Letöltés ideje: 2024.04.10)

[9] Yang, F., Sun, Q., Jin, H. & Zhou, Z. (2020): "Superpixel Segmentation with Fully Convolutional Networks." Proceedings of the 2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). 13961-13970. DOI: 10.1109/CVPR42600.2020.01398.

[10] Deng, X., Liu, T., He, S., Xiao, X., Li, P. & Gu, Y. (2023): An underwater image enhancement model for domain adaptation. Frontiers in Marine Science. Vol. 10. DOI: 10.3389/fmars.2023.1138013

[11] Mikołajczyk, A. & Grochowski, M. (2018): Data augmentation for improving deep learning in image classification problem. Proceedings of the 2018 International Interdisciplinary PhD Workshop (IIPHDW). DOI: 10.1109/IIPHDW.2018.8388338.

[12] Du, R., Li, W. & Chen, S., Li, C. (2021): Unpaired Underwater Image Enhancement Based on CycleGAN. Information. Vol. 13, No. 1. DOI: 10.3390/info13010001.

[13] Guo, Y., Chen, Q., Chen, J., Wu, Q., Shi, Q. & Tan, M. (2019): Auto-Embedding Generative Adversarial Networks For High Resolution Image Synthesis. IEEE Transactions on Multimedia. Vol. 21, No. 11. 2726-2737. DOI: 10.1109/TMM.2019.2908352.

[14] Simonyan, K. & Zisserman, A. (2014): Very Deep Convolutional Networks for Large-Scale Image Recognition. Proceedings of the 3rd International Conference on Learning Representations, ICLR 2015. DOI: 10.48550/arXiv.1409.1556.

[15] He, K., Zhang, X., Ren, S. & Sun, J. (2016): Deep Residual Learning for Image Recognition. Proceedings of the 2016 IEEE Conference on Computer Vision and Pattern

Recognition (CVPR). 770-778. DOI: 10.1109/CVPR.2016.90. ISBN: 978-1-4673-8851-1, 978-1-4673-8852-8. ISSN: 1063-6919.

[16] Wang, X., Girshick, R., Gupta, A. & He, K. (2018): Non-local Neural Networks. Proceedings of the 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition. 7794-7803. DOI: 10.1109/CVPR.2018.00813. ISBN: 978-1-5386-6420-9, 978-1-5386-6421-6. ISSN: 2575-7075, 1063-6919.

[17] Selvaraju, R.R., Cogswell, M., Das, A., Vedantam, R., Parikh, D. & Batra, D. (2017): Grad-CAM: Visual Explanations from Deep Networks via Gradient-Based Localization. Proceedings of the 2017 IEEE International Conference on Computer Vision (ICCV). 618-626. DOI: 10.1109/ICCV.2017.74. ISBN: 978-1-5386-1032-9, 978-1-5386-1033-6. ISSN: 2380-7504.

[18] Karpathy, A. & Fei-Fei, L. (2016): Deep Visual-Semantic Alignments for Generating Image Descriptions. IEEE Transactions on Pattern Analysis and Machine Intelligence. Vol. 39, No. 4. 664-676. ISSN: 0162-8828, 2160-9292, 1939-3539. DOI: 10.1109/TPAMI.2016.2598339.

[19] Donahue, J., Hendricks, L.A., Rohrbach, M., Venugopalan, S., Guadarrama, S., Saenko, K. & Darrell, T. (2016): Long-Term Recurrent Convolutional Networks for Visual Recognition and Description. IEEE Transactions on Pattern Analysis and Machine Intelligence. Vol. 39, No. 4. 677-691. ISSN: 0162-8828, 2160-9292, 1939-3539. DOI: 10.1109/TPAMI.2016.2599174.

[20] Dosovitskiy, A., Beyer, L., Kolesnikov, A., Weissenborn, D., Zhai, X., Unterthiner, T., Dehghani, M., Minderer, M., Heigold, G., Gelly, S., Uszkoreit, J. & Houlsby, N. (2021): An Image is Worth 16x16 Words: Transformers for Image Recognition at Scale. Proceedings of the 9th International Conference on Learning Representations, ICLR 2021. DOI: 10.48550/arXiv.2010.11929.

[21] Van Otten, N. (2023): "L1 And L2 Regularization Explained, When To Use Them & Practical How To Examples." Data Science, Machine Learning. <https://spotintelligence.com/2023/05/26/11-l2-regularization/> (Letöltés ideje: 2024. 03.09).

[22] Pramoditha, R. (2022): "Using Early Stopping to Reduce Overfitting in Neural Networks." Data Science 365. <https://medium.com/data-science-365/using-early-stopping-to-reduce-overfitting-in-neural-networks-7f58180caf5b> (Letöltés ideje: 2024. 03.09).

[23] Yadav, H. "Dropout in Neural Networks: A Deep Dive into the Go-To Method for Reducing Overfitting in the Modern Era of Deep Learning," Towards Data Science, 2022.

[24] Tan, M. & Le, Q.V. (2019): EfficientNet: Rethinking Model Scaling for Convolutional Neural Networks. Proceedings of the 36th International Conference on Machine Learning, PMLR. Vol. 97. 6105-6114.

[25] Chen, L. (2022): Deep Learning based Underwater Object Detection. (PhD disszertáció) School of Computing and Mathematical Sciences, University of Leicester.

[26] Khan, S. H., Naseer, M., Hayat, M., Zamir, S. W., Khan, F. & Shah, M. (2021): Transformers in Vision: A Survey. DOI: 10.48550/arXiv.2101.01169. 1-5.

[27] Liu, W., Anguelov, D., Erhan, D., Szegedy, C., Reed, S., Fu, C-Y. & Berg, A.C. (2016): "SSD: Single Shot MultiBox Detector." Proceedings of the European Conference on Computer Vision (ECCV). DOI:10.1007/978-3-319-46448-0_2, <https://arxiv.org/abs/1512.02325> (Letöltés ideje: 2023-11-12)

[28] Chandan, G., Jain, A., Jain, H. & Mohana (2018): Real-Time Object Detection and Tracking Using Deep Learning and OpenCV. Proceedings of the 2018 International Conference on Inventive Research in Computing Applications (ICIRCA). DOI: 10.1109/ICIRCA.2018.8597266.

[29] Vencer, L. (2023): Sea Animals Image Dataset. <https://www.kaggle.com/datasets/vencerlanz09/sea-animals-image-dataset>. (Letöltés ideje: 2024-07-23)

ABSZTRAKT

Ez a szakdolgozat a számítógépes látásalapú objektumdetektálás területén belüli aktuális technológiák és módszerek vizsgálatára összpontosít, különös tekintettel a víz alatti környezetben való alkalmazásukra. A kutatás célja a mélytengeri számítógépes látásalapú objektumdetektálás technológiai kihívásainak feltárása, és az ehhez kapcsolódó mélytanulási módszerek és algoritmusok hatékonyságának értékelése. A módszertan magában foglalja a legújabb szakirodalmi áttekintést, esettanulmányokat és saját fejlesztéseket a mélytanulás területén. A fő eredmények között szerepelnek a különböző adatelőkészítési, képfeldolgozási technikák hatékonyságának összehasonlítása, valamint az új, víz alatti környezetre optimalizált modellek bemutatása. A szakdolgozat arra a következtetésre jut, hogy a számítógépes látás és mélytanulás területén történő fejlesztések jelentős előrelépést jelentenek a víz alatti számítógépes látásalapú objektumdetektálásban, nyitva hagyva az utat a további kutatások és innovációk számára ezen a területen.

Kulcsszavak: számítógépes látás, mélytanulás, víz alatti számítógépes látásalapú objektumdetektálás, mélytengeri képfeldolgozás, adatelőkészítés, képfeldolgozási technikák.

ABSTRACT

This thesis focuses on the examination of current technologies and methods in computer vision-based object detection, with a special emphasis on their application in underwater environments. The research aims to explore the technological challenges of deep-sea object recognition and evaluate the effectiveness of related deep learning methods and algorithms. The methodology includes a comprehensive review of the latest literature, case studies, and original developments in the field of deep learning. Key findings include a comparison of the effectiveness of various data preparation and image processing techniques, as well as the introduction of new models optimized for underwater environments. The thesis concludes that advancements in computer vision and deep learning significantly enhance underwater object detection, paving the way for further research and innovation in this field.

Keywords: computer vision, deep learning, underwater object detection, deep-sea image processing, data preparation, image processing techniques.

MELLÉKLETEK

1. számú melléklet.

Optimális hipotetikus modell felépítésének és fejlesztésének leírása

Ez a mellékelt felvázol egy lehetséges hipotetikus optimális modellt mely alkalmazza a hipotézis megvalósítandó pontjait és az arra adott tényleges megoldási javaslatokat, mint optimalizációt.

A modell céljainak bemutatása

A javasolt hipotetikus modell célja, hogy átfogó és adaptív megközelítést biztosítson a fejlett víz alatti objektumok észleléséhez és alkalmazza az empirikus kutatás során felvázolt optimalizálási javaslatokat, például színvisszaállítási módszerek, adaptív architektúrák és figyelemmechanizmusok integrálásával. A modell célja, hogy jelentősen javítsa a víz alatti objektumérzékelő rendszerek pontosságát és robusztusságát, lehetővé téve azok alkalmazását különféle kihívást jelentő víz alatti környezetben és alkalmazásokban.

A módszertan bemutatása

A javasolt modell fejlesztéséhez és értékeléséhez az URPC10K adatkészletet használhatjuk fel. Az URPC10K adatkészlet egy nagyméretű, annotációkkal ellátott vízfelszín felvételekből összeállított adatkészlet, amelyet kifejezetten objektumészlelési feladatokhoz terveztek. 10 000 nagy felbontású víz alatti képből áll, amelyeket különféle vízviszonyok között rögzítettek,

beleértve a tiszta, közepesen zavaros és erősen zavaros környezetet is. Az adatkészlet a víz alatti objektumok széles skáláját fedi le, például halakat, korallokat és tengeri törmeléket, a megfelelő határolókeret-annotációkkal.

Az URPC10K adatkészlet különösen alkalmas a javasolt modellhez a víz alatti állapotok és objektumkategóriák sokfélesége miatt. Az adatkészlet nagy mérete és jó minőségű annotációi elegendő képzési adatot biztosítanak a mély tanulási modellekhez, lehetővé téve számukra a robusztus és általánosítható funkciók megtanulását. Ezenkívül a különböző vízviszonyokból származó képek felvétele lehetővé teszi a modell számára, hogy alkalmazkodjon a víz alatti környezet különböző kihívásaihoz, például a szintorzulásokhoz és a korlátozott láthatósághoz.

A javasolt modell robusztus kiértékelésének biztosítása érdekében az URPC10K adatkészlet képzési és tesztelési részhalmazokra lesz felosztva. A képzési készlet a teljes adatkészlet 80%-át, míg a tesztelési készlet a fennmaradó 20%-ot fogja tartalmazni. Ez a felosztás hatékony modellképzést és elfogulatlan teljesítményértékelést tesz lehetővé.

Az adatok modellbe való betáplálása előtt előfeldolgozási lépéseket alkalmaznak a víz alatti képek minőségének javítása érdekében. A hipotézis megvalósításában tárgyalt Sea-thru színhelyreállítási modellt a víz alatti képek színeinek optimalizálására alkalmazzuk. A Sea-thru modell egy speciális RGBD képzési adatkészletet használ, amely mélységi információkat tartalmaz, lehetővé téve a víz alatti képek színeinek pontos visszaállítását. A Sea-thru modell alkalmazásával a víz alatti képeken gyakran előforduló szintorzulás és kékes-zöld árnyalat korrigálásra kerül, ami természetesebb és tetszetősebb képeket eredményez.

A javasolt modellt az átviteli tanulás és a finomhangolási technikák kombinációjával képezzük. A transfertanulást az előre betanított modellekből, például a VGG-ből vagy a ResNetből szerzett ismeretek hasznosítására alkalmazzák majd, amelyeket nagyszabású földi adatkészleteken, például az ImageNet-en képeztek ki. Az előre betanított modelleket funkciókivonóként fogják használni, súlyukat pedig az URPC10K adatkészleten finomhangoljuk, hogy alkalmazkodjanak a víz alatti képek sajátos jellemzőihez.

A betanítási folyamat során adatbővítési technikákat, például véletlenszerű kivágást, átfordítást és elforgatást alkalmaz a modell a képzési készlet bővítésére és a modell általánosítási képességének javítására. A modellt a megfelelő veszteségi függvények, például az objektumészlelés fókuszveszteségének használatával képezzük ki, optimalizáló algoritmusoként Adam-ot vagy az SGD-t alkalmazzuk.

A képzési folyamatot megfelelő értékelési mérőszámok segítségével fogják nyomon követni, beleértve az átlagos pontosságot (mAP), az Intersection over Union (IoU) és az F1-pontszámot, hogy értékeljék a modell teljesítményét a képzési halmazon. Korai megállást és modellellenőrzést alkalmazhatunk a túlillesztés elkerülése és a legjobban teljesítő modell kiválasztása érdekében.

Modellarchitektúra

A javasolt modellarchitektúra három fő összetevőből áll: egy színhelyreállító modulból, egy adaptív objektumészlelési modulból és egy figyelemmechanizmus-modulból.

A színhelyreállító modul a Sea-thru modellt fogja használni a bemeneti víz alatti képek színeinek javítására. A Sea-thru modellt előzetesen betaníthatjuk az RGBD adatkészletre, és finom hangoljuk az URPC10K adatkészlet egy részhalmazán, hogy alkalmazkodjanak a víz alatti környezet speciális színjellemzőihez.

Az adaptív objektumészlelési modul a legmodernebb objektumészlelési architektúrán fog alapulni, mint például a YOLO vagy az SSD. Az objektumészlelési hálózatot úgy módosítjuk, hogy olyan adaptív architektúrákat építsen be, mint például a Neural Architecture Search (NAS) technikák, hogy automatikusan optimalizálják a hálózati struktúrát és a

hiperparamétereket a bemeneti adatok jellemzői alapján. Ez az alkalmazkodóképesség lehetővé teszi a modell számára, hogy jobban kezelje a víz alatti környezet változó feltételeit.

A figyelemmechanizmus modul integrálva lesz az objektumészlelő hálózatba, hogy a kép leginformatívabb területeire összpontosítson, és elnyomja a nem releváns háttérinformációkat. Figyelem mechanizmusok, mint például a csatorna szerinti és a térbeli figyelem, a kép különböző jellemzőinek és régióinak fontosságának súlyozására kerülnek alkalmazásra, javítva a modell azon képességét, hogy pontosan észlelje a tárgyakat.

A színhelyreállító modul, az adaptív objektumészlelési modul és a figyelemmechanizmus modul egy egységes architektúrába épül be, lehetővé téve a végpontok közötti képzést és következtetést. A bevitt víz alatti képek először a színhelyreállító modulon haladnak át vizuális minőségük javítása érdekében, ezt követi az adaptív objektumészlelő modul figyelem mechanizmusokkal a víz alatti objektumok pontos észlelésére és lokalizálására.

Következtetések

A javasolt hipotetikus modell átfogó és adaptív megközelítést mutat be a fejlett víz alatti objektumok észleléséhez a legmodernebb mély tanulási technikák integrálásával, beleértve a színvisszaállítási módszereket, az adaptív architektúrákat és a figyelemmechanizmusokat. Az URPC10K adatkészlet kihasználásával, amely sokféle víz alatti állapotot és objektumkategóriát biztosít, a modell célja, hogy jelentősen javítsa a víz alatti objektumérzékelő rendszerek pontosságát és robusztusságát.

Egy ilyen modell kifejlesztése gondos adat-előfeldolgozást igényel, beleértve a színvisszaállítást a Sea-thru modell használatával, valamint egy jól megtervezett oktatási folyamatot, amely magában foglalja az átviteli tanulást, a finomhangolást és az adatkiegészítési technikákat. A színhelyreállító modulból, egy adaptív objektumészlelési modulból és egy figyelemmechanizmus-modulból álló modellarchitektúra lehetővé teszi a modell számára, hogy hatékonyan kezelje a víz alatti környezet kihívásait, és pontosan észlelje az érdeklődésre számot tartó objektumokat.

E modell potenciális előnyei, beleértve a víz alatti ökoszisztémák jobb megértését, az autonóm víz alatti felmérések lehetővé tételét és a tengeri erőforrások fenntartható kezelésének elősegítését, értékes és izgalmas kutatási irányt jelentenek, amely hatással lehet a tudományterület további fejlődésére.