Alkalmazott Informatikai és Alkalmazott Matematikai Doktori Iskola

Óbudai Egyetem

Mély neurális hálózatok új alkalmazásai a robotirányítás és rendszer-felügyelet terén

Ph.D. Tézisfüzet

Károly István Artúr

Témavezető:
Dr. Galambos Péter

Budapest
2023

Motiváció

A gépi tanulást alkalmazó megoldások egyre inkább éreztetik jelenlétüket számos iparágban, gyakran észrevétlenül hatva mindennapi életünkre, szolgáltatásainkra és katonai alkalmazásokra. A modern ipari kihívásokra nyújtott megoldások szempontjából különösen előnyösek ezek az Ipar 4.0 koncepciók lehetőségeit feltáró fejlesztések.

Az gépi tanulási megközelítések közül a mély tanulás (Deep Learning/DL) témaköréhez kapcsolhatók az elmúlt évek számos jelentősebb és nagyobb hatással bíró fejlesztései. A mély tanulási módszerek fejlődésük során innovatív előrelépéseket hoztak a robotika különböző területein is [1]. Annak ellenére, hogy a DL alapú megoldások széles körben alkalmazhatók, jellemzőek rájuk a magas számítási komplexitás és az eredmények kiszámíthatatlansága miatt jelentkező kihívások [2, 3]. Biztonságkritikus rendszerekben, mint például az önvezető autók és ipari robotok, soha nem használják a DL módszereket önállóan, és kimenetüket mindig bizonytalansággal kezelik. Következésképpen, ezeket a DL módszereket gyakran olyan tesztekkel validálják, amelyek képesek a modell robosztusságának kiértékelésére [4, 5, 6, 7]. A DL megoldásokat alkalmazó rendszerek magas számítási komplexitása és időigényes tanítási folyamata miatt alternatív modell architektúrák és tanítási stratégiák is bevezetésre kerültek, mint például a mély konvolúciós neurális hálózatok [8, 9] és a transfer learning [10, 11, 12], melyek lehetővé teszik robosztus modellek tanítását korlátozott erőforrásokkal is.

Magán a tanítási folyamaton kívül, a mély tanulás számára az adatgyűjtési folyamat és a tanító adathalmaz előkészítése is jelentős erőforrásokat igényel, különösen akkor, ha az adathalmaz előkészítése manuálisan történik [3]. Ez a probléma kifejezetten fontos a robotikában, ahol az adatgyűjtési folyamat gyakran magába foglalja a valós robottal való mozgást/feladat végrehajtást [13]. Mivel az adatgyűjtés akár hónapokig is tarthat még több robottal is, az adthalmaz előkészítés jelentős költségeket vonhat maga után. Ezért, az új mély tanulási megoldások a tanításhoz szükséges címkézett adatmennyiség minimalizálása érdekében előszerettel használják a felügyelet nélküli tanítás valamint a transfer learning módszereit [14, 15]. Ezen kívül bizonyos megközelítések az erőforrásigény csökkentését szimulációban történő adatgyűjtés segítségével valósítják meg [16, 17, 18].

Ez a disszertáció főként a percepció szintű problémákra összpontosít, ideértve az objektum felismerést, szegmentálást és más kapcsolódó feladatokat, amelyek elengedhetetlenek a robotos manipuláció és a mobil robot navigáció során. Bár

korábbi tanulmányok azt mutatták, hogy a mély tanulás módszerei ígéretes eredményeket nyújthatnak a robotikában, ugyanakkor jelentős kihívásokra is rámutattak, melyeket a magas számítási komplexitás és a nagy mennyiségű tanító adat szükségessége okoz. Tekintettel az erőforrás-hatékonyság jelentőségére a robotikában, az ilyen kihívások megfelelésére irányuló új, mély tanuláson alapuló, megoldások keresése kritikus fontossággal bír. Ezért ez a disszertáció a felügyelet nélküli tanítás, a transfer learning és az automatizált adathalmaz készítési módszerek lehetőségeit vizsgálja a robotika területén.

1. Tézis

Valósidejű klaszterezésen alapuló állapot és anomália felismerés

Bevezetés

A Support Vector Machine (SVM) módszerrel lineáris, bináris klasszifikációt hajthatunk végre a

$$f(\mathbf{x}) = \mathbf{w}^T\mathbf{x} + b,$$

formában meghatározott diszkriminancia függvénnyel, ahol $\mathbf{x}$ a bementi vektor, $\mathbf{w}$ a súly vektor és b az úgynevezett torzítás (bias) [19].

A predikciók az $f(\mathbf{x})$ értékén alapulnak, ahol $f(\mathbf{x}) \geq 0$ esetén a predikció $y = 1$, és $f(\mathbf{x}) < 0$ esetén a predikció $y = -1$ lesz [19]. Míg a hagyományos SVM-ek a döntési függvény paramétereit ($\mathbf{w}$ és b) egy címkézett tanító adthalmaz segítségével határozzák meg, az egyosztályú SVM (One-Class Support Vector Machine vagy OCSVM) kifejezetten az egy osztályba tartozó minták az összes többi osztálytól való megkülönböztetésére lett tervezve [20, 21]. Ennek eredményeként az OCSVM felügyelet nélküli tanítása megvalósítható egy olyan címkézetlen adathalmaz használatával, amelynek minden mintáját ugyanahoz az osztályhoz tartozónak feltételezünk. Ez a módszer alkalmas anomáliák detektálására, amikor ritkán előforduló mintákat kell elkülöníteni a többi, normálisnak vélt mintától.

Két széles körben használt módszer létezik az OCSVM-ek megvalósítására. Az egyik módszer egy a tanító mintákat az origótól elválasztó hipersíkot definiál a "feature" térben, majd a hipersík és az origó távolságát maximalizálja [20]. A másik módszer a tanító minták "feature" térbeli hipergömbbel történő bekerítésén, és a hipergömb térfogatának minimalizálásán alapul [21].

A döntési függvényt, amely egy adott mintát ($\mathbf{x}$) a felismert osztály tagjának minősít, a maximalizálás vagy minimalizálás problémájának Lagrange multiplikátor módszer segítségével való megoldásával kaphatjuk meg (speciális megoldó algoritmusok, például az SMO használatával [22, 23]). A feature térben mért távolságok meghatározásához a kernel módszer használható [21], amely transzformálja a problémát egy magasabb dimenziós térbe, ahol a minták lineárisan elválaszthatóvá válnak.

A Gauss kernel (Radial Basis Function vagy RBF kernel) a leggyakrabban használt kernel függvény. A számítása két adatpont, $\mathbf{x}_i$ és $\mathbf{x}_j$, esetén a következőképpen adódik:

$$K(\mathbf{x}_i, \mathbf{x}_j) = \exp\left(-\frac{\|\mathbf{x}_i - \mathbf{x}_j\|^2}{2\sigma^2}\right),$$

ahol a kernel paraméter σ befolyásolja a kernel függvény érzékenységét.

A Gauss kernel sorba fejtése egy olyan végtelen sort eredményez amelyben a

$$\langle \mathbf{x}_i, \mathbf{x}_j \rangle, \langle \mathbf{x}_i, \mathbf{x}_j \rangle^2, \langle \mathbf{x}_i, \mathbf{x}_j \rangle^3 \dots$$

tagok szerepelnek, amik önmaguk is kernel függvények. Ez rugalmasságot biztosít az osztályozó tetszőleges dimenziójú térben történő megtervezéséhez, lehetővé téve nemlineáris döntési határok kialakítását a feature térben.

Az OCSVM-ekkel végzett felügyelet nélküli tanítás elméletére alapozva egy felügyelet nélküli tanítást alkalmazó valós idejű algoritmust hoztam létre (Algoritmus 1.), amely egy dinamikusan képzett OCSVM modellekből álló együttes (ensemble, vagy $\mathbf{E}$) segítségével képes robotalkalmazások állapotának és anomáliáinak automatikus észlelésére. Az algoritmus bemenete egy adatfolyam ($\mathcal{S}$). $\mathcal{S}$-ben az adatpontok ($\mathbf{X}_i | i \in \{t, t-1, t-2, ...\}$, ahol $\mathbf{X}_t$ a legfrissebb adat a t időlépésben) az f (Hz) mintavételi ráta szerint állnak rendelkezésre, amit a robotvezérlő beállítása határoz meg. Az OCSVM predikciókhoz és tanításához az algoritmus csúszó ablak mintavételezést használ. A predikciók átlapolódó csúszó ablakokon ($\mathbf{W}^t$) történnek, biztosítva a valós idejű működést. Másrészről, a tanító halmaz ($\mathbf{T}$) nem átlapolódó mintákból ($\mathbf{W}^t_{train}$) áll, hogy elkerüljük a túltanítást és csökkentsük a számítási igényeket. Az algoritmust a w és n paraméterekkel lehet beállítani, amelyek meghatározzák a csúszó ablakok szélességét és a tanításhoz használt minták számát. Az algoritmus paramétereitől függő számításigényének meghatározásához képleteket definiáltam, lineáris illetve nemlineáris (RBF) kernel függvények használatát feltételezve.

Az algoritmus egy kontingencia táblát hoz létre ($\mathbf{C}$), amely tartalmazza, hogy az $\mathbf{E}$-ben lévő OCSVM-ek milyen gyakran aktiválódtak együtt. Ezt az információt aztán arra használjuk fel, hogy OCSVM csoportokat hozzunk létre azzal a céllal, hogy felismert állapotok hierarchiáit építsük fel (Algoritmus 2). A hierarchia kialakítása offline módon történik, az adatfolyam egy rögzített szegmensén ($\mathcal{R}$), annak érdekében, hogy biztosítsuk, hogy az összes OCSVM ugyanazt a bemenetet kapja.

Továbbá megmutattam, hogy a javasolt és megvalósított klaszterező algoritmus a generatív gépi tanulási modellek (például a Generative Adversarial Neural Networks/GAN) kiértékelésére és összehasonlítására is használható, a szintetizált kimenetek statisztikai elemzésével. Az kísérleti eredmények azt mutatják,

hogy ez a módszer nagyszerű alternatívát nyújthat olyan felhasználási esetekben, ahol a szintetizált adatok szemantikájához nem állnak rendelkezésre előtanított feture extractor hálózatok. Ugyanakkor az algoritmus önmagában képtelen a mode collapse azonosítására, és használata egy diverzitást mérő mutatóval kombinálva ajánlott.

Új tudományos eredmények

1. Tézis

Bemutatok egy új klaszterező algoritmust (Algoritmus 1.) robotalkalmazások állapotainak automatikus online és valós idejű osztályozására és anomáliák detektálására, valamint generatív gépi tanulási modellek kiértékelésére véges dimenziós állapotleírást, ill. folytonos numerikus jellemzőket felhasználva. A javasolt módszer hatékonyságát egy reprezentatív kollaboratív robotalkalmazáson végzett kiértékeléssel, valamint egy valós ipari környezetben való sikeres alkalmazással igazoltam.

Algorithm 1: Clustering algorithm

```
input: S, w, n, stopping_criterion, cd
/* Initialize internal variables                                   */
W^t = [], W^t_train = [], T = [] E = [];
C = [[]];
i = 0, count = 0;
stop = False, no_train_step = 0;
on new X_t in S:
    /* Update data structures                                      */
    i += 1;
    W^t.append(X_t);
    if W^t.size() > w then
        W^t.remove(0);
    if i == w then
        W^t_train=W^t;
        i = 0;
        T.append(W^t_train);
        if T.size() > n then
            T.remove(0);
    else
        W^t_train=W^{t-1}_train;
    /* Perform predictions                                         */
    p = [];
    for OCSVM in E do
        p.append(OCSVM.predict(W^t));
    /* Train a new OCSVM if needed                                 */
    if !stop and (all(p==-1) or E.size() == 0) and T.size() == n then
        if count < cd then
            count += 1;
        else
            E.append(OCSVM.train(T));
            count = 0;
    else
        no_train_step += 1;
        if no_train_step == stopping_criterion then
            stop = True;
    /* Update contingency table                                    */
    C_update(p, C);
```

1.1. Altézis

Megmutattam, hogy egy dinamikusan felépített OCSVM-együttes egy kontingenciatáblázattal együtt felhasználható elemi állapotok többszintű hierarchiájának felépítésére egy alulról felfelé irányuló hierarchiaépítő stratégia segítségével (Algoritmus 2.).

Algorithm 2: Bottom-up hierarchy building strategy

```
input : E, C, R, th
/* Initialize internal variables                                  */
H = [], G = [];
N = R.size();
/* Calculate entropies                                            */
for OCSVM in E do
    H.append(Entropy(OCSVM.predict(R)));
H /= max(H);
for h in H do
    /* Find the index of the minimal entropy OCSVM                */
    i_h = argmin(H);
    if i_h in G then
        H[i_h] = 2 ;                  // OCSVM i_h is already in a group
    else
        /* Create a new group                                     */
        G.append([i_h]);
        I = [];
        for j = 0; j < C[i_h].size(), j++ do
            I.append(Info_Gain(C,(i_h, j), N)); // Compute Information Gain
              for i_h
        I[i_h] = -1;
        for j = 0; j < I.size(); j++ do
            /* Start with the most similar OCSVM                  */
            i_g = argmax(I/max(I));
            if C[i_h][i_g]/C[i_h][i_h] > th then
                G[-1].append(i_g) ;       // Add i_g to the current group
                I[i_g] = -1;
            else
                break;
        H[i_h] = 2;
return: G
```

1.2. Altézis

Reprezentatív példákon keresztül megmutattam, hogy a bemeneti adatfolyamban a nem átfedő csúszóablakok használata a tanító minták elkészítésére jelentősen csökkenti a számítási időt anélkül, hogy jelentősen rontaná a predikciók minőségét. A számítási követelmények meghatározásához képleteket definiáltam (1) és (2) a módszer paramétereitől függően.

$$t_{train} \propto N^{OCSVM} w \mathcal{O}_{train}$$

linear:

$$\mathcal{O}_{train} = O(dn) \rightarrow t_{train} \propto N^{OCSVM} wdn \approx N^{OCSVM} \mathcal{T} fd \tag{1}$$

non-linear:

$$\mathcal{O}_{train} = O(dn^2) \rightarrow t_{train} \propto N^{OCSVM} wdn^2 \approx N^{OCSVM} \mathcal{T} fdn$$

ahol $\mathcal{O}_{train}$ egy OCSVM közvetlenül az adatpontokon történő betanításának számítási követelménye.

Az inference módban az algoritmus számítási követelménye

$$t_{inference} \propto N^{OCSVM} w \mathcal{O}_{inference}$$

linear:

$$\mathcal{O}_{inference} = O(d) \rightarrow t_{inference} \propto N^{OCSVM} wd \tag{2}$$

non-linear:

$$\mathcal{O}_{inference} = O(dn) \rightarrow t_{inference} \propto N^{OCSVM} wdn \approx N^{OCSVM} \mathcal{T} fd$$

ahol $\mathcal{O}_{train}$ egy OCSVM közvetlenül az adatpontokon történő betanításának számítási követelménye.

1.3. Altézis

Igazoltam, hogy az eredetileg robotikai alkalmazásokhoz tervezett OCSVM alapú anomália-detektáló megközelítés hatékonyan alkalmazható generatív gépi tanulási modellek kiértékelésére a szintetizált kimenetek statisztikai elemzésével. A jelenlegi kiértékelési módszerektől eltérően ez a megközelítés képes a modellek kiértékelésére a kimeneti adatszemantikától függetlenül, mivel nem igényel előtanított feature extractor hálózatot.

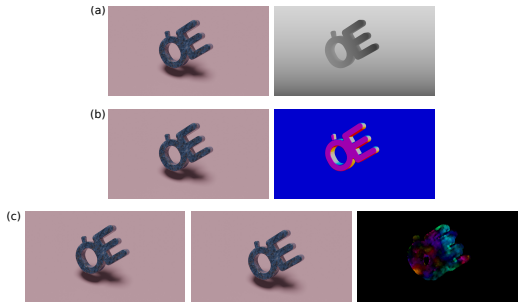
Kapcsolódó publikációk: [KA1, KA2, KA3].

2. Tézis

Kereszt-modális leképzésen alapuló transfer learning RGB feature extractor hálózatokhoz

Bevezetés

Sok gépi látáson alapuló robotikai megoldásban, például mozgó akadályok észlelésénél mobil robotikában, előnyös lehet a tipikusan használt RGB jellemzőkön kívül további modalitásokat is felhasználni [24, 25]. A leggyakrabban használt nem-RGB, de látással kapcsolatos modalitások a mélység, a felületi normálisok és az optikai áramlás. Mivel ezek a modalitások szorosan kapcsolódnak a vizuális információhoz, mindegyiknek megvan a megfelelő vizuális reprezentációja, értelmezése, amelyeket általában megjelenítési célokra szokás használni. Az 1. ábra néhány ilyen reprezentáció példáját mutatja be.



Ábra 1: Nem-RGB modalitások képi reprezentációja. **a**: RGB kép és a hozzá tartozó mélységadatok szürkeárnyalatos képként, **b**: RGB kép és a hozzá tartozó felületi normálisok RGB képként, **c**: Egymást követő RGB képkockák és a hozzájuk tartozó optikai áramlás RGB képként

A mobil robotikában a mozgó objektumok szegmentálásához egy adott specifikus feladatra szabott, nyilvánosan elérhető adathalmaz beszerzése jelentős kihívást jelenthet. Ezért gyakran szükség lehet egy, a feladatra szabott, tanító adathalmaz összeállítására minden egyes konkrét felhasználási esetre. A szükséges tanító minták számának csökkentése érdekében előtanított feature extractor hálózatokat lehet használni transzfer learning [12, 11] segítségével. A leggyakrabban használt, előtanított feature extractor hálózatokat azonban általában nagyméretű RGB kép adathalmazok segítségével tanítják be, így nem képesek

közvetlenül más modalitásokat feldolgozni.

Az általam javasolt módszer a bemenetek megfelelő leképzése segítségével lehetővé teszi a korábban RGB képeken előtanított feature extractor hálózatok hasznosítását nem RGB modalitások (például optikai áramlás) feldolgozására. A bemenetek leképzését kereszt-modális leképzésnek nevezik.

Egy új mély neurális hálózatot hoztam létre, Optical Flow Segmentation Network (OFSNet) néven, ami mozgó objektumok szegmentálását hajtja vére videoszekvenciákban, mobil robotok navigációjához. Az OFSNet modell a népszerű U-Net architektúrán [26] alapul, és az Inception v3 [27] feature extractor hálózatot használja (1. táblázat). Ezen a hálózaton demonstrálom, hogy a javasolt kereszt-modális leképzés milyen módon használható optikai áramlás modalitás feldolgozására, RGB képeken előtanított feature extractor hálózat használatával.

Table 1: Az OFSNet modell struktúrája, az Inception v3 feature extractor hálózatot magába foglalva (#1-től #12-ig). A #13-tól to #18-ig terjedő struktúra a saját kontribúcióm. A hálózat tanítása során kizárólag ezeknek a rétegenek a paraméterei módosultak.

#	**type**	**patch size/stride** or remarks	**input size**
Layers from Inception v3 model			
1	conv	3x3/2	299x299x3
2	conv	3x3/1	149x149x32
3	conv padded	3x3/1	147x147x32
4	pool	3x3/2	147x147x64
5	conv	3x3/1	73x73x64
6	conv	3x3/2	71x71x80
7	conv	3x3/1	35x35x192
8	3 x Inception	Inception block	35x35x288
9	5 x Inception	Inception block	17x17x768
10	2 x Inception	Inception block	8x8x1280
11	pool	8x8	8x8x2048
12	linear	Inception v3 features	1x1x2048
Layers for segmentation			
13	transposed conv	3x3/2	1x1x2048
14	transposed conv	4x4/2	3x3x1280
15	skip connection	#9+#14	8x8x1280
16	transposed conv	16x16/2	8x8x1280
17	linear	logits	30x30x1
18	sigmoid	classifier	30x30x1

Az OFSNet modell tanítása önfelügyelt módon történt, az Unsupervised Non-

Local Consensus Voting (uNLC) [28] módszer segítségével, melynek kimenetei biztosították a tanító adathalmaz alap-igazság (ground-truth) szegmentációit.

A mozgó objektumok gyakran viszonylag kicsinek tűnnek a képeken, ami osztály kiegyensúlyozatlansághoz vezet a pozitív és negatív pixelek számát illetően. Ez a kiegyensúlyozatlanság akadályozhatja a tanítási folyamat konvergenciáját azáltal, hogy befolyásolja a számított veszteség és így a gradiensek értékét is.

A kereszt entrópia veszteség (Cross Entropy Loss) egy tipikus veszteségfüggvény a szegmentációs modellek esetén

$$L_{CE} = -\frac{1}{N}\sum_{i=1}^{N} y_i log_2(p_i) + (1-y_i) log_2(1-p_i),$$

ahol L_{CE} a kereszt entrópia veszteség egy képkockára, és N a képen lévő pixelek számát jelenti. A helyes címke pixelenként y_i, ahol 0 a hátteret, 1 pedig az objektumot jelenti. A prediktált valószínűsége, hogy az i-edik pixel az objektumhoz tartozik p_i-vel van jelölve.

Egy másik gyakran használt veszteségfüggvény a Soft Dice loss [29, 30]

$$\begin{aligned} SoftDice &= \frac{2\sum_{i=1}^{N} y_i p_i}{\sum_{i=1}^{N} p_i + \sum_{i=1}^{N} y_i}, \\ L_{SD} &= 1 - SoftDice. \end{aligned} \tag{3}$$

ahol L_{SD} a Soft Dice loss, N a pixelszámot jelöli, és y_i az i-edik pixel valós címkéje, ahol 0 a hátteret 1 pedig az objektumot jelenti. A prediktált valószínűsége, hogy az i-edik pixel az objektumhoz tartozik p_i-vel van jelölve.

Mivel a kereszt entrópia veszteség jobban bünteti a hamis pozitív predikciókat, és a Soft Dice loss jobban bünteti a hamis negatívokat, egy új veszteségfüggvényt javasoltam, (compound loss néven), a tanító adathalmazban jelentkező osztály kiegyensúlyozatlanság okozta kihívások kiküszöbölésére. A compound loss a kereszt entrópia veszteség és a Soft Dice loss dinamikusan állítható lineáris kombinációját használja.

Új tudományos eredmények

2. Tézis
Kidolgoztam egy Deep Learning modellt (OFSNet) és ennek megfelelő veszteségfüggvényt videoszekvenciákban történő mozgó objektum szegmentálásra, lehetővé téve a mozgó akadályok elkerülését beltéri környezetben a mobil robot navigációhoz. A javasolt módszert egy valós ipari mobil robot (AGV) rendszer prototípusával validáltam.

2.1. Altézis
Igazoltam, hogy az RGB képeken előtanított feature extractor hálózatok megfelelő formázással (kereszt-modális leképzéssel) (4) általánosíthatók kombinált optikai áramlás és szürkeárnyalatos képekből álló bemeneti adatokra. Ezt a következtetést alátámasztották a DAVIS 2016 adathalmzazon végzett kísérletek és egy ipari AGV rendszer prototípusából nyert valós adatok.

$$
\begin{aligned}
\hat{R} &= \mathcal{F}_{::1} + abs(min(\mathcal{F}_{::1}))\mathbf{J}^{w\times h}) \\
\hat{G} &= \mathcal{F}_{::2} + abs(min(\mathcal{F}_{::2}))\mathbf{J}^{w\times h}) \\
\mathcal{I}_{::1}^{RGB} &= \frac{\hat{R}}{max(\hat{R})} \\
\mathcal{I}_{::2}^{RGB} &= \frac{\hat{G}}{max(\hat{G})} \\
\mathcal{I}_{::3}^{RGB} &= \frac{\mathcal{I}^{Y}}{max(\mathcal{I}^{Y})}
\end{aligned}
\tag{4}
$$

2.2. Altézis
Bevezettem egy összetett veszteségfüggvényt (5) és egy kapcsolódó empirikus tanítási eljárást, amely a Cross-Entropy Loss és a Soft Dice Loss függvények dinamikus lineáris kombinációját használja, hogy kiküszöbölje azok ellentétes hatásait. Ennek a veszteségfüggvénynek és a tanítási stratégiának a hatékonyságát az OFSNet modell betanításával demonstráltam. Bizonyítottam, hogy az α paraméter dinamikus állításával biztosítható a legjobb átfedés az annotáció és a prediktált szegmentációs maszkok között, még kiegyensúlyozatlan adatok esetén is.

$$
\begin{aligned}
L &= (1-\alpha)L_{CE} + \alpha L_{SD} \\
\alpha &= \frac{\left(\frac{j}{n_e}\right)^4}{1.6}
\end{aligned}
\tag{5}
$$

Kapcsolódó publikációk: [KA4, KA5, KA6].

3. Tézis

Automatikus nagyléptékű adathalmaz generálás

Bevezetés

Az adathalmaz előkészítés és az annotáció idő és erőforrás igényessége jelentős akadályai az új mély tanulási (DL) megoldások fejlesztésének. Szintetikus adatok használata esetén az adatgyűjtési és annotálási eljárás könnyen automatizálható. A szintetikus adathalmazon való tanítás esetén azonban biztosítani kell, hogy a betanított modellek képesek legyenek a valós adatokat is eredményesen feldolgozni [16, 17, 18]. Ezt a problémát gyakran a „valóságszakadék áthidalásának" (bridging the reality gap) nevezik, ami a szintetikus és a valós adattartományok különbségeiből adódó kihívások lekűzdésére utal. Valós mintákon történő tanítás esetében ez a modell adaptációs probléma nem jelentkezik, viszont az adatgyűjtési folyamat kevésbé automatizálható. Ebben a téziscsoportban két módszer kerül bemutatásra vizuális robotikai kihívásokhoz használt automatizált adathalmaz generálására.

Az első módszer lehetővé teszi valós képekből álló objektum szegmentációs adathalmazok automatikus annotálását, kifejezetten a robotikai alkalmazásokra fókuszálva. A módszer a robotika azon egyedülálló tulajdonságait használja ki, melyek lehetővé teszik a robot póz információihoz való hozzáférést és ezáltal a kamera valós térbeli 3D elhelyezkedésének pontos ismeretét. Ezenkívül felhasználja azt a tényt, hogy az ismert geometriájú objektumok előre meghatározott pózokban helyezhetők el a robot munkaterében. Ezen feltevések alapján a módszer segítségével nagy pontossággal leírható a teljes környezet digitális mása, ami lehetővé teszi egy benne elhejezett virtuális kamera segítségével az ismert objektum geometria képsíkbeli vetületeinek kiszámítását. A virtuálisan levetített szegmentációs maszkok a tényleges képeket annotálják, így létrehozva a tanító adathalmazt ami felügyelt tanítással használható. Ez a megközelítés a digitális iker paradigmát terjeszti ki a DL adathalmazok létrehozásának és a DL modellek validálásának területére.

A szegmentációs maszkok létrehozásához a módszer az objektum geometria képsíkra történő perspektivikus vetületeit számítja ki. Egy 3D pont ${}^{w}\mathbf{X} = ({}^{w}X, {}^{w}Y, {}^{w}Z, 1)^{\top}$ (a "world" koordináta rendszerben megadva) perspektivikus vetülete $\bar{\mathbf{x}} = (u, v, 1)^{\top}$ az

$$\bar{\mathbf{x}} = \mathbf{K}\mathbf{\Pi}\ {}^{c}\mathbf{T}_{w}\ {}^{w}\mathbf{X}, \tag{6}$$

összefüggéssel írható le, ahol $\mathbf{K}$ a kamera mátrix, ami a kamera intrinsic paramétereit

tartalmazza, melyek a kamera kalibráció során határozhatók meg [31]. A projekciós mátrix $\mathbf{\Pi}$ formája $[\mathbf{I}|\mathbf{0}]$, ahol $\mathbf{I}$ egy 3×3-as egységmátrix és $\mathbf{0}$ egy nullákból álló három elemű oszlopvektor. Végül, ${}^{c}\mathbf{T}_{w}$ a 4×4 méretű homogén transzformációs mátrix ami a world és a kamera koordináta rendszerek közötti transzformációt írja le.

A javasolt módszert az 3. Algoritmus mutatja be. A probléma szimbolikus leírásához $P({}^{w}\mathbf{X})$ jelöli a ${}^{w}\mathbf{X}$ pont perspektivikus vetületét, F egy olyan felületet, amelyet a határoló pontjainak halmaza határoz meg ($F = \{{}^{w}\mathbf{X}_1, {}^{w}\mathbf{X}_2, ..., {}^{w}\mathbf{X}_n\}$), $R({}^{w}\mathbf{X})$ egy sugár, amely a kamera koordináta rendszerének origójából ered és áthalad a ${}^{w}\mathbf{X}$ ponton, és ${}^{w}\mathbf{X}^{\mathcal{O}}_{all}$ az $\mathcal{O}$ objektum felületén található összes lehetséges 3D pont halmaza.

A szegmentációs maszkok létrehozásához az egyes objektumok felületén egy véges ponthalmazt kell definiálni: ${}^{w}\mathbf{X}^{\mathcal{O}} = \{{}^{w}\mathbf{X} | {}^{w}\mathbf{X}\ \mathcal{O}\text{ felületén}\}$, ${}^{w}\mathbf{X}^{\mathcal{O}} \subseteq {}^{w}\mathbf{X}^{\mathcal{O}}_{all}$. A $\mathbb{P}({}^{w}\mathbf{X}^{\mathcal{O}})$ hatványhalmaz tartalmazza az összes lehetséges (nem feltétlenül értelmes) felületet az $\mathcal{O}$ objektum számára, egy adott ponthalmaz ${}^{w}\mathbf{X}^{\mathcal{O}}$ esetén. Egy felület minden pontjának a képsíkba történő levetítésével sokszögek hozhatók létre : $Poly^{F} = \{P({}^{w}\mathbf{X}_i)$, ahol ${}^{w}\mathbf{X}_i \in F\}$, és a pontok vetületeit használjuk a sokszög csúcsaiként. Egy olyan felületekből álló halmazt ($F^{\mathcal{O}} \subseteq \mathbb{P}({}^{w}\mathbf{X}^{\mathcal{O}})$) kell definiálnunk az $\mathcal{O}$ objektum számára, amelyben lévő felületekre teljesül, hogy az összes $P({}^{w}\mathbf{X}_j)$ által megadott vetület, ahol ${}^{w}\mathbf{X}_j \in {}^{w}\mathbf{X}^{\mathcal{O}}_{all}$, legalább az egyik $Poly^{F_k}$ sokszögbe esik, ahol $F_k \in F^{\mathcal{O}}$, de a $P({}^{w}\mathbf{X})$ vetületek, ahol $R({}^{w}\mathbf{X})$ nem metszi az objektumot a 3D térben, nem esnek egyik $Poly^{F_k}$ sokszögbe sem, ahol $F_k \in F^{\mathcal{O}}$.

A második módszer szintetikus adathalmazok létrehozását és azok automatikus annotációját teszi lehetővé. A szintetikus adathalmaz generáló eljárás alapját a Blender 3D programcsomag képzi [32]. A Blender, egy nyílt forráskódú szoftver széleskörű funkcionalitással, amely a robotos alkalmazásokhoz gyakran használt szimulátorokkal ellentétben nem korlátozódik egy specifikus felhasználási tartományra. Az elsősorban számítógépes grafikához tervezett Blender, eszközök széles skáláját kínálja a vizuális környezetek (scene) manipulálásához. Ezen eszközök közé tartozik a 3D objektummodellezés, a megvilágítás és a kamera konfigurációja, a geometria módosítása, textúrák és shaderek definiálása, képek utófeldolgozása stb. A Blender képes fotorealisztikus renderelt képeket generálni, és fizikai szimulációt is tartalmaz a Bullet fizikai engine segítségével. Ezenkívül a Blender, Python API-jának köszönhetően, jól integrálható DL tanítási munkafolyamatokkal, mivel a legtöbb DL keretrendszer támogatja a Python programozási nyelvet. Kifejlesztettem egy új Python-alapú addont, a Blender Annotation Tool-t (BAT), amellyel a szintetikus adathalmazokhoz tartozó szeg-

mentációs maszkokat tartalmazó annotációk automatikusan elkészíthetők.

A szintetikus adathalmaz alapú megközelítés hatékonyságát három kísérlettel validáltam: egy valós robotos pick-and-place feladaton, egy benchmarkon (OpenLORIS-Object [7]), amely a folytonos tanítási módszerek kiértékelésére használható, és egy grasp detection feladaton. A pick-and-place kísérletek eredményei a javasolt szintetikus adathalmaz generáló módszernek köszönhető teljesítmény javulásra világítanak rá, a vizuális objektum detekcióra használt DL modellek esetében. Az Open-LORIS-Object benchmarkon végzett kísérletek megmutatták, hogy a szintetikus adatok jelentősen növelhetik az experience replay módszert használó folytonos tanítási módszerek forward transfer metrikáját. Ezen túlmenően a grasp detection kísérletek azt bizonytják, hogy a szintetikus adathalmaz generálási eljárás nagy rugalmassága és alacsony erőforrásigénye miatt a feladatspecifikus információk felhasználásával a grasp detection feladatok esetén is javulást lehet elérni.

A szintetikus adatok felhasználása során felmerülő "valóságszakadék" leküzdésének kihívására egy olyan megközelítést vezettem be, amely a javasolt automatizált valós adathalmaz annotálási módszert kombinálja a szintetikus adathalmaz generáló folyamattal. Ez a Filling The Reality Gap (FTRG) nevű megközelítés a valós környezet teljes szintetikus reprezentációját használja. A javasolt módszerek segítségével szegmentációs maszkokat készíthetünk mind a valós, mind a szintetikus környezetekhez. A módszer célja a szintetikus és a valós adatok közötti szakadék áthidalása a valós és szintetikus elemek egyetlen képen belüli sima átmenetű keverésével. A pick-and-place problémával kapcsolatos kísérleti eredmények azt mutatják, hogy az FTRG-módszer alkalmazása a tanító adathalmaz készítésére jobb teljesítményt eredményez mint a legelterjedtebb megközelítések, mint például a fotorealisztikus szintetikus adatok használata vagy a domain randomization technikája.

A forward transfer a folytonos tanítást használó modellek predikcióinak pontosságát mérő metrika [7]. Az értéke megmutatja, hogy egy modell mennyire jól alkalmazkodik az új feladatokhoz a korábbi feladatokon elvégzett tanítás után. Az automatizált szintetikus adathalmaz generáló folyamat segítségével létrehoztam az OpenLORIS-Object adathalmaz egy részhalmazának szintetikus megfelelőjét (SynLORIS-adatadathalmazt), és bizonyítottam, hogy az experience replay módszert használó folytonos tanulási módszerek [33] forward transfer metrikája jelentősen javítható, ha a tanításuk során szintetikus mintákat is felhasználnak.

A javasolt szintetikus adathalmaz generáló folyamat segítségével egy olyan eljárást dolgoztam ki, amely automatikusan generál feladat-specifikus grasp de-

tection adathalmazokat robotos manipulációhoz. Ez a módszer lehetővé teszi a Grasp Quality Convolutional Neural Network hálózatok (GQCNN-ek) [34] finomhangolását adott szerelési feladatokhoz, ismert objektum és környezeti geometriákat, valamint az összeállítási sorrend ismeretét feltételezve. A cél a GQCNN modellek betanítása olyan megvalósítható robotos megfogások detektálására, amelyek esetében a megvalósíthatóság nem csupán az objektum és a megfogó geometriáját és fizikai tulajdonságait veszi figyelembe, hanem az összeszerelés során az objektumok későbbi elhelyezését is. Az aszimmetrikus illesztés típusú feladaton végzett szimulált grasp detection kísérletek azt mutatják, hogy a javasolt módszerrel létrehozott adathalmazokat használva a GQCNN-ek finomhangolása jelentősen javítja a sikeres feladat végrehajtás valószínűségét.

Új tudományos eredmények

3. Tézis

Két eljárást hoztam létre objektum szegmentációs adathalmazok automatikus összeállítására és címkézésére. Megmutattam, hogy ezek az adathalmazok felhasználhatók mély tanulási modellek tanítására a vizuális robotos manipuláció feladataihoz, mint például a scene recognition vagy az object/grasp detection. Az első módszer a vetítési algoritmust (3.) használja, hogy példány szegmentációs maszkokat generáljon valós képekhez ismert geometriát feltételezve. A második módszer számítógépes grafikát használ szintetikus renderelt képek automatikus generálására és címkézésére.

Algorithm 3: Projection algorithm

input : *Image shape:* $[w, h, 3]$*, List of objects:* $\mathbf{O} = [\mathcal{O}_1, \mathcal{O}_2, \dots]$

```
/* Init annotation as black image                                    */
Init: M = zeros((w, h, 5));
for O ∈ O do
    for T ∈ O.triangles do
        /* Projection as in (6)                                      */
        v_1^i, v_2^i, v_3^i = Project(T.vertices);
        temp_img = zeros((w, h));
        /* Get internal pixels of the triangle                       */
        P = Where(DrawTriangle(temp_img, (v_1^i, v_2^i, v_3^i), color=1) == 1);
        for p ∈ P do
            if M[p][0 : 3] == [0, 0, 0] then
                /* It was background before                          */
                M[p][0 : 3] = O.color_id;
                M[p][3] = O.id;
                M[p][4] = T.id;
            else if M[p][0 : 3] == O.color_id then
                /* It is the same object                             */
                Pass;
            else
                T̃ = O.GetTriangle(M[p][3], M[p][4]);
                if IsOccluded(T, by = T̃) then
                    /* T̃ occludes T                                  */
                    Pass;
                else
                    M[p][0 : 3] = O.color_id;
                    M[p][3] = O.id;
                    M[p][4] = T.id;
return: M
```

3.1. Altézis

Megmutattam, hogy a szintetikus minták hozzávétele az experience replay módszert alkalmazó folytonos tanítást használó modellek tanítási folyamata során jelentősen javíthatja azok forward transfer metrikáját. Ennek az állításnak az érvényességét az OpenLORIS Object adathalmaz egy részhalmazának szintetikus másánán bizonyítottam, melyen összehasonlítottam két folytonos tanulási modellt: az egyik szintetikus adatokat is felhasznált a tanításhoz,a másik pedig nem.

3.2. Altézis

Bevezettem egy megoldást a „valóság szakadék" kihívásának kezelésére, amikor a szintetikus adatokon tanított mély tanulási modelleket a valós adatokon használjuk. Ez az FTRG (Filling The Reality Gap) névre keresztelt megoldás magában foglalja a javasolt valós és szintetikus adatok automatikus annotációs technikáit, hogy egyetlen képen belül sima átmenetet tegyen lehetővé a szintetikus és a valódi kép elemek között. Mask R-CNN modellek összehasonlító elemzésével, melyek különböző módszereket használtak a „valóság szakadék" leküzdésére (ide értve a randomizációt és a fotorealisztikus szintetiksu adatok használatát), demonstráltam, hogy az FTRG módszerrel jobb eredményeket érhetünk el mint más modern megközelítésekkel.

3.3. Altézis

Egy feladat-specifikus grasp detection adathalmaz létrehozására alkalmas módszert javasoltam robotos szerelési feladatokhoz, amely figyelembe veszi, hogy bizonyos megfogási pózokhoz, nem feltétlenül rendelhetők ütközésmentes elhelyezések. Ez a módszer automatizált szintetikus adatgenerálást és címkézést, valamint mintavételezésen alapuló grasp detection technikákat foglal magába, kihasználva az ismert objektum és összeállítás geometriákat és az összeállítási sorrendet. A módszer hatékonyságát egy GQCNN hálózat finomhangolásával demonstráltam egy szintetikus adathalmazon, és megmutattam, hogy a finomhangolt GQCNN felülmúlja az eredetit egy aszimmetrikus illesztés típusú robotos szerelési feladatban.

Kapcsolódó publikációk: [KA7, KA8, KA9, KA10, KA11].

A TÉZISEKHEZ KAPCSOLÓDÓ PUBLIKÁCIÓK

[KA1] A. I. Károly, R. Fullér, and P. Galambos, "Unsupervised clustering for deep learning: A tutorial survey," *Acta Polytechnica Hungarica*, vol. 15, no. 8, pp. 29–53, 2018.

[KA2] A. I. Károly, J. Kuti, and P. Galambos, "Unsupervised real-time classification of cycle stages in collaborative robot applications," in *2018 IEEE 16th World Symposium on Applied Machine Intelligence and Informatics (SAMI)*. IEEE, 2018, pp. 000 097–000 102.

[KA3] A. I. Károly, M. Takács, and P. Galambos, "OCSVM-based Evaluation Method for Generative Neural Networks," in *2019 International Joint Conference on Neural Networks (IJCNN)*. IEEE, 2019, pp. 1–6.

[KA4] A. I. Károly, P. Galambos, J. Kuti, and I. J. Rudas, "Deep learning in robotics: Survey on model structures and training strategies," *IEEE Transactions on Systems, Man, and Cybernetics: Systems*, vol. 51, no. 1, pp. 266–279, 2020.

[KA5] A. I. Károly, R. N. Elek, T. Haidegger, K. Széll, and P. Galambos, "Optical flow-based segmentation of moving objects for mobile robot navigation using pre-trained deep learning models," in *2019 IEEE International Conference on Systems, Man and Cybernetics (SMC)*. IEEE, 2019, pp. 3080–3086.

[KA6] A. I. Károly, R. N. Elek, T. Haidegger, and P. Galambos, "Moving Obstacle Segmentation with an Optical Flow-based DNN: an Implementation Case Study," in *2021 IEEE 25th International Conference on Intelligent Engineering Systems (INES)*. IEEE, 2021, pp. 000 189–000 194.

[KA7] A. I. Károly and P. Galambos, "Automated Dataset Generation with Blender for Deep Learning-based Object Segmentation," in *2022 IEEE 20th Jubilee World Symposium on Applied Machine Intelligence and Informatics (SAMI)*. IEEE, 2022, pp. 000 329–000 334.

[KA8] A. I. Károly, Á. Károly, and P. Galambos, "Automatic Generation and Annotation of Object Segmentation Datasets Using Robotic Arm," in *2022 IEEE 10th Jubilee International Conference on Computational Cybernetics and Cyber-Medical Systems (ICCC)*. IEEE, 2022, pp. 000 063–000 068.

[KA9] A. I. Károly, S. Tirczka, T. Piricz, and P. Galambos, "Robotic Manipulation of Pathological Slides Powered by Deep Learning and Classi-

cal Image Processing," in *2022 IEEE 22nd International Symposium on Computational Intelligence and Informatics and 8th IEEE International Conference on Recent Achievements in Mechatronics, Automation, Computer Science and Robotics (CINTI-MACRo)*. IEEE, 2022, pp. 000 387–000 392.

[KA10] A. I. Károly and P. Galambos, "Task-Specific Grasp Planning for Robotic Assembly by Fine-Tuning GQCNNs on Automatically Generated Synthetic Data," *Applied Sciences*, vol. 13, no. 1, p. 525, 2023.

[KA11] A. I. Károly, S. Tirczka, H. Gao, I. J. Rudas, and P. Galambos, "Increasing the Robustness of Deep Learning Models for Object Segmentation: A Framework for Blending Automatically Annotated Real and Synthetic Data," *IEEE Transactions on Cybernetics*, 2023.

HIVATKOZÁSOK

[1] H. A. Pierson and M. S. Gashler, "Deep Learning in Robotics: A Review of Recent Research," *CoRR*, vol. abs/1707.07217, 2017. [Online]. Available: http://arxiv.org/abs/1707.07217

[2] I. Goodfellow, Y. Bengio, and A. Courville, *Deep Learning*. MIT Press, 2016, http://www.deeplearningbook.org.

[3] Y. LeCun, Y. Bengio, and G. Hinton, "Deep learning," *nature*, vol. 521, no. 7553, pp. 436–444, 2015.

[4] M. Everingham, L. Van Gool, C. K. I. Williams, J. Winn, and A. Zisserman, "The Pascal Visual Object Classes (VOC) Challenge," *International Journal of Computer Vision*, vol. 88, no. 2, pp. 303–338, Jun. 2010.

[5] F. Perazzi, J. Pont-Tuset, B. McWilliams, L. Van Gool, M. Gross, and A. Sorkine-Hornung, "A Benchmark Dataset and Evaluation Methodology for Video Object Segmentation," in *The IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, June 2016, pp. 724–732.

[6] A. Geiger, P. Lenz, C. Stiller, and R. Urtasun, "Vision meets robotics: The KITTI dataset," *The International Journal of Robotics Research*, vol. 32, no. 11, pp. 1231–1237, 2013.

[7] Q. She, F. Feng, X. Hao, Q. Yang, C. Lan, V. Lomonaco, X. Shi, Z. Wang, Y. Guo, Y. Zhang *et al.*, "OpenLORIS-Object: A robotic vision dataset and

benchmark for lifelong deep learning," in *2020 IEEE international conference on robotics and automation (ICRA)*. IEEE, 2020, pp. 4767–4773.

[8] Y. LeCun, P. Haffner, L. Bottou, and Y. Bengio, *Object Recognition with Gradient-Based Learning*. Berlin, Heidelberg: Springer Berlin Heidelberg, 1999, pp. 319–345. [Online]. Available: https://doi.org/10.1007/3-540-46805-6_19

[9] A. Krizhevsky, I. Sutskever, and G. E. Hinton, "ImageNet classification with deep convolutional neural networks," in *Advances in neural information processing systems*, 2012, pp. 1097–1105.

[10] Y. Bengio, "Deep learning of representations for unsupervised and transfer learning," in *Proceedings of ICML Workshop on Unsupervised and Transfer Learning*, 2012, pp. 17–36.

[11] M. Huh, P. Agrawal, and A. A. Efros, "What makes ImageNet good for transfer learning?" *CoRR*, vol. abs/1608.08614, 2016. [Online]. Available: http://arxiv.org/abs/1608.08614

[12] S. J. Pan and Q. Yang, "A Survey on Transfer Learning," *IEEE Transactions on Knowledge and Data Engineering*, vol. 22, no. 10, pp. 1345–1359, 2010.

[13] S. Levine, P. Pastor, A. Krizhevsky, J. Ibarz, and D. Quillen, "Learning hand-eye coordination for robotic grasping with deep learning and large-scale data collection," *The International journal of robotics research*, vol. 37, no. 4-5, pp. 421–436, 2018.

[14] P. Baldi, "Autoencoders, unsupervised learning, and deep architectures," in *Proceedings of ICML Workshop on Unsupervised and Transfer Learning*, 2012, pp. 37–49.

[15] A. Radford, L. Metz, and S. Chintala, "Unsupervised Representation Learning with Deep Convolutional Generative Adversarial Networks," *CoRR*, vol. abs/1511.06434, 2015. [Online]. Available: http://arxiv.org/abs/1511.06434

[16] J. Tremblay, A. Prakash, D. Acuna, M. Brophy, V. Jampani, C. Anil, T. To, E. Cameracci, S. Boochoon, and S. Birchfield, "Training deep networks with synthetic data: Bridging the reality gap by domain randomization," in *Proceedings of the IEEE conference on computer vision and pattern recognition workshops*, 2018, pp. 969–977.

[17] A. Prakash, S. Boochoon, M. Brophy, D. Acuna, E. Cameracci, G. State, O. Shapira, and S. Birchfield, "Structured domain randomization: Bridging the reality gap by context-aware synthetic data," in *2019 International*

Conference on Robotics and Automation (ICRA). IEEE, 2019, pp. 7249–7255.

[18] M. Roberts, J. Ramapuram, A. Ranjan, A. Kumar, M. A. Bautista, N. Paczan, R. Webb, and J. M. Susskind, "Hypersim: A photorealistic synthetic dataset for holistic indoor scene understanding," in *Proceedings of the IEEE/CVF International Conference on Computer Vision*, 2021, pp. 10 912–10 922.

[19] C. Cortes and V. Vapnik, "Support-vector networks," *Machine learning*, vol. 20, no. 3, pp. 273–297, 1995.

[20] B. Schölkopf, R. C. Williamson, A. Smola, J. Shawe-Taylor, and J. Platt, "Support vector method for novelty detection," *Advances in neural information processing systems*, vol. 12, 1999.

[21] D. M. Tax and R. P. Duin, "Support vector data description," *Machine learning*, vol. 54, no. 1, pp. 45–66, 2004.

[22] Z.-Q. Zeng, H.-B. Yu, H.-R. Xu, Y.-Q. Xie, and J. Gao, "Fast training support vector machines using parallel sequential minimal optimization," in *2008 3rd international conference on intelligent system and knowledge engineering*, vol. 1. IEEE, 2008, pp. 997–1001.

[23] C.-C. Chang and C.-J. Lin, "LIBSVM: A library for support vector machines," *ACM transactions on intelligent systems and technology (TIST)*, vol. 2, no. 3, pp. 1–27, 2011.

[24] J. Kao, D. Tian, H. Mansour, A. Vetro, and A. Ortega, "Moving object segmentation using depth and optical flow in car driving sequences," in *2016 IEEE International Conference on Image Processing (ICIP)*, Sep. 2016, pp. 11–15.

[25] T. Zhou, S. Wang, Y. Zhou, Y. Yao, J. Li, and L. Shao, "Motion-attentive transition for zero-shot video object segmentation," in *Proceedings of the AAAI conference on artificial intelligence*, vol. 34, no. 07, 2020, pp. 13 066–13 073.

[26] O. Ronneberger, P. Fischer, and T. Brox, "U-Net: Convolutional Networks for Biomedical Image Segmentation," *Computing Research Repository (CoRR)*, vol. abs/1505.04597, 2015, visited on 2019-04-15. [Online]. Available: http://arxiv.org/abs/1505.04597

[27] C. Szegedy, V. Vanhoucke, S. Ioffe, J. Shlens, and Z. Wojna, "Rethinking the inception architecture for computer vision," in *Proceedings of the IEEE*

conference on computer vision and pattern recognition, 2016, pp. 2818–2826.

[28] A. Faktor and M. Irani, "Video segmentation by non-local consensus voting." in *BMVC*, vol. 2, no. 7, 2014, p. 8.

[29] T. Sørensen, "A method of establishing groups of equal amplitude in plant sociology based on similarity of species and its application to analyses of the vegetation on danish commons," *Biol. Skr.*, vol. 5, pp. 1–34, 1948.

[30] L. R. Dice, "Measures of the amount of ecologic association between species," *Ecology*, vol. 26, no. 3, pp. 297–302, 1945.

[31] K. M. Dawson-Howe and D. Vernon, "Simple pinhole camera calibration," *International Journal of Imaging Systems and Technology*, vol. 5, no. 1, pp. 1–6, 1994.

[32] B. O. Community, *Blender - a 3D modelling and rendering package*, Blender Foundation, Stichting Blender Foundation, Amsterdam, 2018. [Online]. Available: http://www.blender.org

[33] D. Rolnick, A. Ahuja, J. Schwarz, T. Lillicrap, and G. Wayne, "Experience replay for continual learning," in *Advances in Neural Information Processing Systems*, H. Wallach, H. Larochelle, A. Beygelzimer, F. d'Alché-Buc, E. Fox, and R. Garnett, Eds., vol. 32. Curran Associates, Inc., 2019. [Online]. Available: https://proceedings.neurips.cc/paper/2019/file/fa7cdfad1a5aaf8370ebeda47a1ff1c3-Paper.pdf

[34] J. Mahler, J. Liang, S. Niyaz, M. Laskey, R. Doan, X. Liu, J. A. Ojea, and K. Goldberg, "Dex-net 2.0: Deep learning to plan robust grasps with synthetic point clouds and analytic grasp metrics," *arXiv preprint arXiv:1703.09312*, 2017.