

Élesztő sejtek és egyéb objektumok
idő-szimmetrikus követése, pozíciós torzítás
kiküszöbölése referenciaadatokon



Szabó Gergely

A PhD disszertáció tézisei

témavezető:

Dr. Horváth András

Pázmány Péter Katolikus Egyetem
Információs Technológiai és Bionikai Kar
Roska Tamás Műszaki és Természettudományi Doktori Iskola

Budapest, 2024

Bevezetés

PhD-tanulmányaim során elsődleges céloom egy újszerű, több objektumot szegmentálni és követni képes architektúra fejlesztése volt, amely mikroszkópos videofelvételeken képes az osztódó élesztősejteket szegmentálni és követni. Ez a kezdetben erősen specializált feladat olyan általánosítható megoldás kidolgozásához vezetett, amely applikálható jelentősen eltérő környezetekben, feladatokban is. Az általam tervezett architektúra, amelyet eredetileg élesztősejtek detektálására, példányalapú szegmentálására és követésére szántam, jelentős előnyöket mutatott a meglévő eszközökkel szemben, mind teljesítmény, mind alkalmazkodóképesség szempontjából. Ez az eredmény túlnő az eredeti célkitűzésen, lehetővé téve a megoldás alkalmazását számos képalkotási modalitás esetében. Átfogó értékelések során vizsgáltam az architektúra erősségeit és korlátait különféle szintetikus és részben szintetikus adathalmazokon. Ezek az eredmények képezik az 1. téziscsoport alapját.

Az architektúra tervezése közben egy korábban nagyrészt figyelmen kívül hagyott torzítást azonosítottam a konvolúciós neurális hálókból (CNN-ek), különösen objektumok pozicionálásával kapcsolatban, amelyek széles körben használt képfeldolgozási referenciaadatokban jelennek meg. Ez a torzítás, amely még az részhalmozokban is kimutatható, messzemenő következményekkel jár az olyan feladatok esetében, amelyek CNN-eket használnak objektumfelismerésre és szegmentálásra. A torzítással kapcsolatos vizsgálataim — adatmanipulációs technikák és architekturális módosítások segítségével — egyszerű megoldásokra mutattak rá, amelyek szinte teljes mértékben kiküszöbölik a torzítást, valamint a CNN-ek peremfeltételeit jelölik meg, mint a probléma valószínűsíthető forrását. Ezen eredmények képezik a 2. téziscsoport alapját.

Fő hozzájárulásaim mellett, amelyeket első szerzős publikációim tükröznek, számos tanulmányban vettem részt másod- vagy későbbi szerzőként. Bár ezek nem tartoznak szorosan sem a tézispontjaimhoz, sem a PhD-tanulmányaim témájához, az átláthatóság érdekében ezen munkák listáját is megadom.

Új tudományos eredmények

Tézis 1a

*Egy új, mélytanuláson alapuló, több objektum példányalapú szegmentálására és követésére alkalmas architektúrát fejlesztettem ki osztódó élesztősejtek videomikroszkópos felvételeihez. Az IFOM által gyűjtött élesztősejt követő adathalmazon az architektúra IoU-alapú szegmentálási és követési F -értékeket ért el $[0,918 \pm 0,019, 0,917 \pm 0,016]$, amelyek felülmúlták az adott feladatra tervezett korszerű eszközök teljesítményét, specifikusan a *Phylocell* $[0,881 \pm 0,020, 0,878 \pm 0,020]$ és a *YeaZ* $[0,818 \pm 0,022, 0,807 \pm 0,023]$ eredményeit.*

Kapcsolódó publikáció: [\[Ar1\]](#)

Sejtek szegmentálása és követése videomikroszkópos felvételeken számos biológiai kutatási területen alapvető fontosságú, beleértve a gyógyszerkutatást, a génexpresszió monitorozását és másokat. Bár léteznek automatikus vagy félautomatikus megoldások az élesztősejtek követésére, még a gépi tanuláson alapuló megközelítések is gyakran specifikus architekturális feltételezésekre támaszkodnak, mint például a klasszikus mintázat- és pozícióillesztési modellek, illetve az időbeli előre irányuló, markovi követés. Míg ezek a megoldások előnyösek lehetnek olyan területeken, mint az autóipar, ahol a predikció sebessége kulcsfontosságú, a videomikroszkópos követési alkalmazásokban a predikciós teljesítmény — különösen a stabil követés — sokkal fontosabb. Továbbá, ezek a korlátok csökkentik a modellek rugalmasságát, akadályozva azok adatvezérelt alkalmazkodóképességét. Az International Foundation of Medicine (IFOM) megkeresésére céloim egy olyan új, több objektumot szegmentáló és követő (MOTS) architektúra kifejlesztése volt, amely támogatja az adatvezérelt általánosítást, valamint a predikciós teljesítményt és a fokozott követési stabilitást helyezi előtérbe a predikció sebességével szemben. Az architektúra részleteire kitérek a Tézis 1b-ben.

A MOTS-architektúrához használt modelleket az IFOM által manuálisan korrigált adatokon tanítottam. A végső kiértékeléseket az IFOM-ban végezték olyan mintákon, amelyeket a modelltervezés és -tanítás során nem osztottak meg

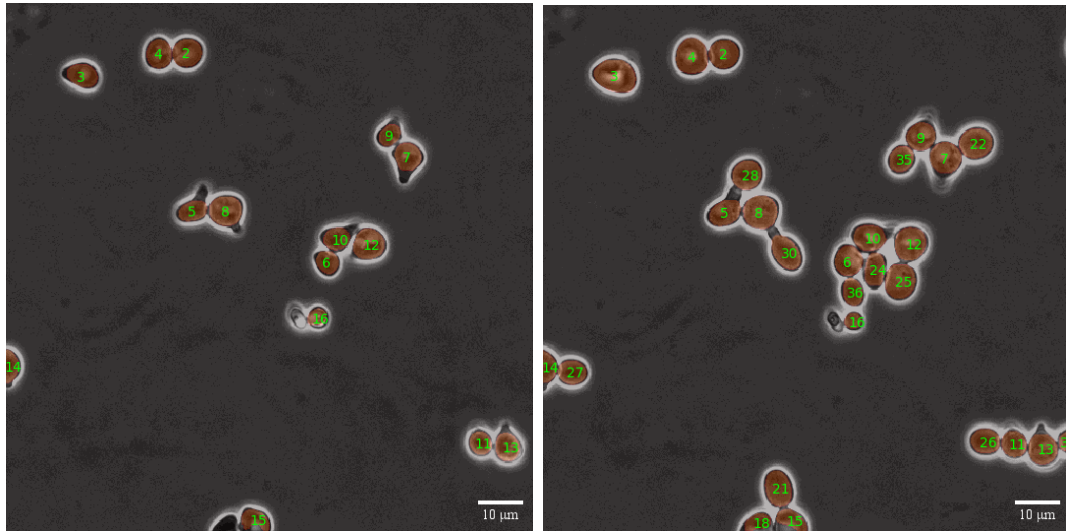


Figure 1: **Követési mintaeredmények**

Élesztő sejtek szegmentálása és követése 8 képkockányi időbeli különbséggel. Az új sejtek megfelelően új azonosítókat kaptak, miközben a meglévő sejtek követése konzisztens maradt.

velem. A Fig. 1 ábrán egy ilyen szegmentálási és követési predikció látható, amely az empirikus értékelés alapján szinte tökéletes eredményeket szemléltet.

Az összehasonlító értékeléshez a PhyloCell és a YeaZ eszközöket választottam. A PhyloCell azért került kiválasztásra, mert ezt használták az adat-címkék ("ground truth label") létrehozására, amelyeket később manuálisan korrigáltak, míg a YeaZ az általam tervezett architektúrához legnagyobb hasonlóságot mutató eszköz volt, bár még mindig jelentősen eltérő architektúrát használt. A példányalapú szegmentálási F-érték mérése Jaccard-index (IoU) alapján történt, míg a követési F-érték a szomszédos képkockák közötti objektum egyedkapcsolatok szerint lett értékelve. A kapcsolatillesztés esetében egy predikció akkor számít pozitívnak, ha mindkét szegmentációs példány helyesen lett detektálva, és a detektált példányok követési azonosítói mind a kiértékelő címkék, mind a predikció oldalán azonosak voltak, míg a hamis predikciók közé tartozik az objektum elvesztése mind a kiértékelő címke, mind a predikció oldalán, valamint a követési azonosító cseréje.

A szegmentálási és követési teljesítmény a látómezőt elhagyó pályák engedélyezésével és tiltásával lett értékelve a címkézett adatokra. Az előbbi pozitív torzítást eredményezett a PhyloCell javára, amely tervezői döntés szerint követi a

látómezőt elhagyó objektumokat, míg az utóbbi elfogulatlan értékelést nyújtott. A Fig. 2 ábrán bemutatott eredmények szerint az elfogulatlan scenárióban a javasolt architektúra messze felülmúlja mind a Phylocell $[0,881 \pm 0,020, 0,878 \pm 0,020]$, mind a YeaZ $[0,818 \pm 0,022, 0,807 \pm 0,023]$ teljesítményét szegmentálási és követési F-értékek tekintetében $[0,918 \pm 0,019, 0,917 \pm 0,016]$, sőt, a látómezőt elhagyó pályák engedélyezése esetén is jobb teljesítményt nyújt.

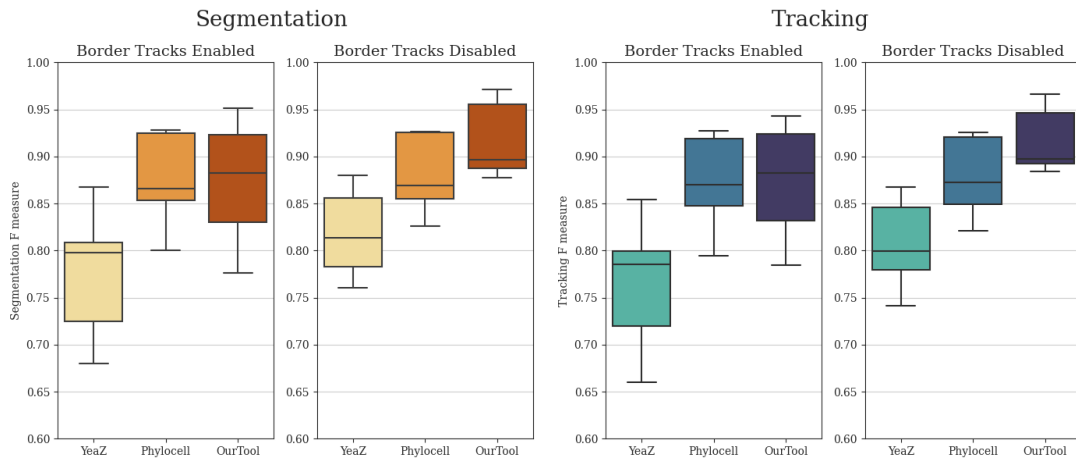


Figure 2: **Eszköz összehasonlítás**

A Phylocell, YeaZ és az általam tervezett architektúra ("OurTool") szegmentálási és követési teljesítményének összehasonlító értékelése. Az értékelés az IoU alapú szegmentálási és az időbeli kapcsolatillesztések F-értékein alapul, a látóteret elhagyó predikciók engedélyezett ("border tracks enabled") és letiltott ("border tracks disabled") állapotaiban. A letiltott határpályák elfogulatlan összehasonlítást biztosítanak az eszközök között, míg az engedélyezett határpályák a nyers kimeneteket tükrözik.

Tézis 1b

Az architektúra — időszimmetrikus követési megközelítésének köszönhetően — természetesen képes a töredezett pályák rekonstruálására, jelentősen javítva a követési konzisztenciát, ami kritikus az öröklődési kapcsolatok pontos meghatározásához. Minden ötödik objektumpéldány véletlenszerű eltávolítása esetén a követésen alapuló rekonstrukció a követési F -értéket $0,404 \pm 0,016$ -ról $0,888 \pm 0,013$ -ra javította. Kapcsolódó publikációk: [Ar1, Ar2]

Az architektúra két fő részből áll: példányalapú szegmentálásból és követésből. A követési szegmens tovább bontható a következő lépésekre: időbeli lokális követés, globális hozzárendelés teljes gráfbejárással, valamint egy utófeldolgozási lépés a kihagyott példányok interpolálására. A makro-architektúra adatáramlás szempontjából a Fig. 3 ábrán látható. Az architektúra újdonsága az egyedi időszimmetrikus lokális követés és globális hozzárendelési lépések együttes alkalmazása. A megközelítés alapja minden objektum követése időben lokális környezetében TR távolságra, mind negatív mind pozitív irányba. A lokális követési predikciók közötti metrikus hasonlóság az időbeli eltolás figyelembevételével számítható, amint azt a Fig. 4 ábra szemlélteti, majd a hierarchikus sorrendben csökkenő időbeli eltolással végrehajtott optimális globális hozzárendelés a "magyar módszer" alkalmazásával történik. Ezt követően az architektúra azon képessége, hogy $2TR$ tartományon belül természetesen újra összekapcsolja a töredezett pályákat, lehetővé teszi az elmulasztott objektumpéldányok azonosítását, amelyeknek interpolált pozíciója a következőképpen számítható:

$$\Delta c(x, y) = \frac{(t - t_{\text{last}})c_{\text{next}}(x, y) + (t_{\text{next}} - t)c_{\text{last}}(x, y)}{t_{\text{next}} - t_{\text{last}}} \quad (1)$$

$$S(x, y) = S_{\text{last}}(x, y) + c(x, y) \quad (2)$$

Ahol $S(x, y)$ és t az interpolált szegmentációt és annak megfelelő időpontját; $S_{\text{last}}(x, y)$, $c_{\text{last}}(x, y)$ és t_{last} az adott azonosítóval rendelkező objektum utolsó előfordulásának szegmentációját, centrális koordinátáját és időpontját; $c_{\text{next}}(x, y)$ és t_{next} az adott azonosítóval rendelkező objektum következő előfordulásának centrális koordinátáját és időpontját jelöli.

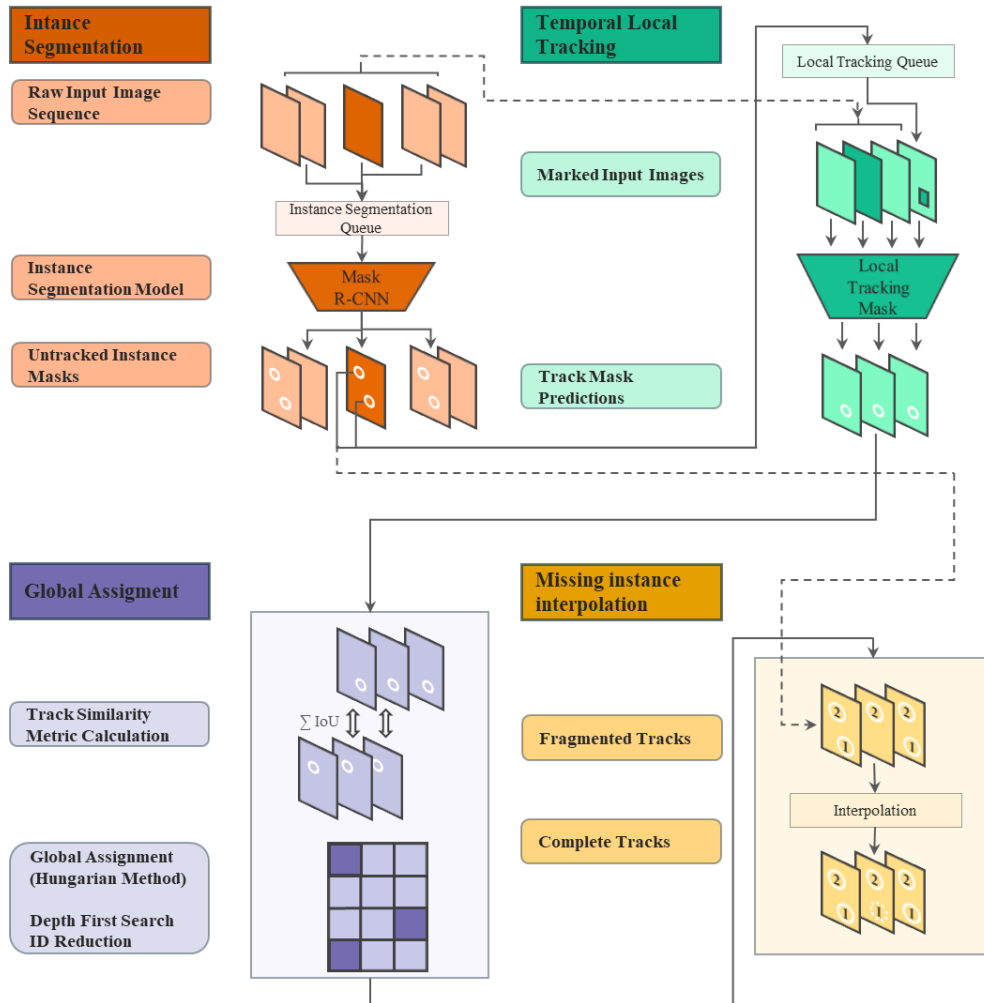


Figure 3: **Vázlatos makro-architektúra**

Az architektúra adatfolyamának szemléltetése a nyers bemeneti képsorozattól a végleges útvonal-predikciókig.

Az architektúra rekonstruktív képességeit legjobban azok a kísérletek mutatják be, amelyekben véletlenszerűen, egyenletes eloszlásban minden 1:15 vagy 1:5 objektum el lett távolítva, vagy minden 1:5 objektum véletlenszerűen, de folytonos blokkokban lett eltávolítva. Az eredmények, amelyeket a Fig. 5 és Fig. 6 ábrák mutatnak be, szemléltetik, hogy az architektúra hatékonyan őrzi meg a követési folytonosságot, különösen megfelelően megválasztott lokális követési tartomány (TR) modellparaméter esetén.

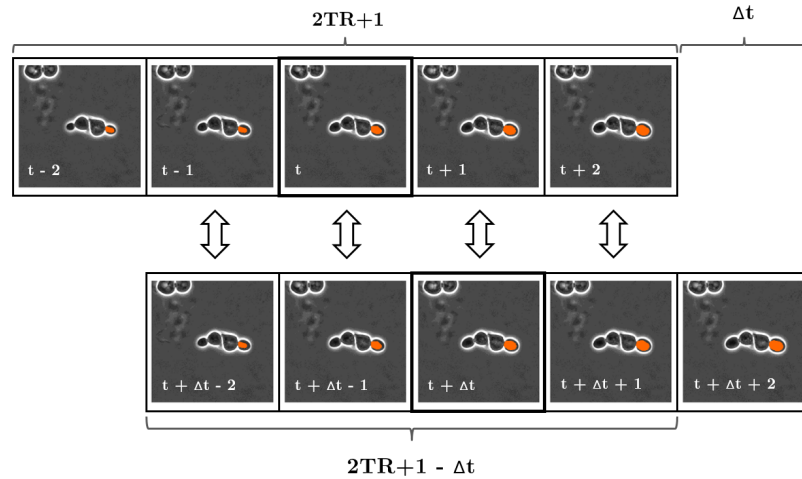


Figure 4: **Metrikus hasonlóság szerkezete**

A metrikus hasonlóság mérésének lépését bemutató vázlatos ábra, amely $2TR + 1$ hosszúságú, különböző sejtpéldányokhoz tartozó lokális pályák között vizsgálja a hasonlóságot a Δt időbeli távolságú képkockákon. A folytonos vonalak a központi szegmentációs példányokat jelölik, amelyek Δt időbeli távolsággal párosíthatók, míg a nyilak az egyes időblokkokhoz tartozó szegmentációs becslések közötti hasonlósági metrikát jelölik. Az egyes metrikus eredmények ezt követően átlagolva lesznek, ami egy számmal reprezentálható hasonlósági mértéket eredményez.

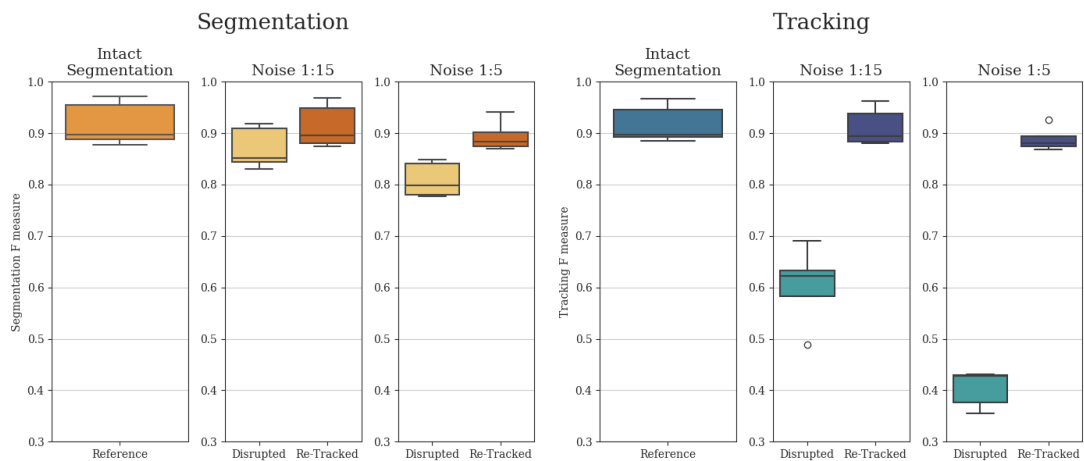


Figure 5: **Egyenletes véletlenszerű zaj rekonstrukciója**

A szegmentálási és követési F-értékek összehasonlító értékelése Intakt, Zaj 1:15 és Zaj 1:5 esetekben. Mind a szegmentálási, mind a követési eredmények tartalmazzák az adott zaj által módosított alapértékeket, valamint a módosított szegmentációból újra követett eredményeket.

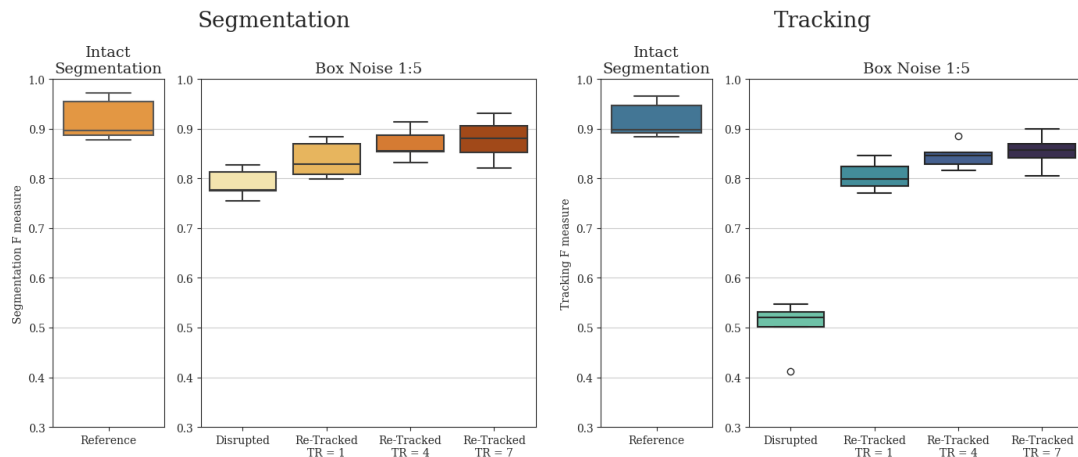


Figure 6: **Blokk zaj rekonstrukciója**

A szegmentálási és követési F-értékek összehasonlító értékelése Intakt és Blokk Zaj 1:5 esetekben. Mind a szegmentálási, mind a követési eredmények tartalmazzák az alapértékeket, amelyeket a zaj módosított, valamint a módosított szegmentációval újra követett eredményeket, lokális nyomon követési tartomány (TR) értékei alapján: 1, 4 és 7.

Tézis 1c

Az architektúrát különböző szintetikus és félszintetikus adathalmazokon értékeltem, bemutattva annak robusztusságát sejt követési környezetek és más természetes követési feladatok, például gyalogos követés kihívásaival szemben. Az eredmények igazolják az architektúra sokoldalúságát, az új követési megközelítés megbízhatóságát, és útmutatóként szolgálhatnak más adathalmazokon várható teljesítményéhez.

Kapcsolódó publikációk: [[Ar1](#), [Ar2](#)]

A különböző sejt követési környezetek szimulációjához és benchmark-értékeléséhez öt szintetikus adathalmazt hoztam létre, mivel nem állnak rendelkezésre jó minőségű, nagy mintaszámú, nyilvánosan elérhető sejt-szegmentálási és követési adathalmazok.

- "S. Arrows" gyorsan mozgó nyilakat szimulál, amelyek a mozgás irányát jelképezik és "szilárd" objektumok.
- "S. Amoeboids" gyorsan mozgó, részben véletlenszerű alakú objektumokat szimulál, amelyek folyamatos elasztikus változásokon mennek keresztül, Perlin-zajjal modellezve.
- "S. Amoeboids-PC" a "S. Amoeboids"-hoz hasonló objektumokat tartalmaz, de fáziskontraszt mikroszkóp hatását szimulálja.
- "S. Amoeboids-PCC" a "S. Amoeboids-PC"-re építve módosított mozgási viselkedést vezet be az összetapadás elősegítésével.
- "S. Amoeboids-PCCA" hasonló a "S. Amoeboids-PCC"-hez, de kiegészítve egy statikus, erősen zavaró zajjal, amely videónként véletlenszerű.

Ezenkívül minden szintetikus minta tartalmaz képenkénti és videónkénti véletlenszerű háttérzajt. Példa predikciók a Fig. 7 ábrán láthatók, míg a szegmentálási és követési F-értékek az egyes adathalmaz-kategóriákra vonatkozóan a Tab. 1 táblázatban található. Az eredmények azt mutatják, hogy még lényegesen bonyolultabb feladatok esetén is, mint az élesztősejt-követés, az architektúra megbízható objektumkövetést ér el. Ugyanakkor a zajos vagy más módon kihívást jelentő felvételeknél az észlelési és példányalapú szegmentálási teljesítmény észrevehetően csökken.

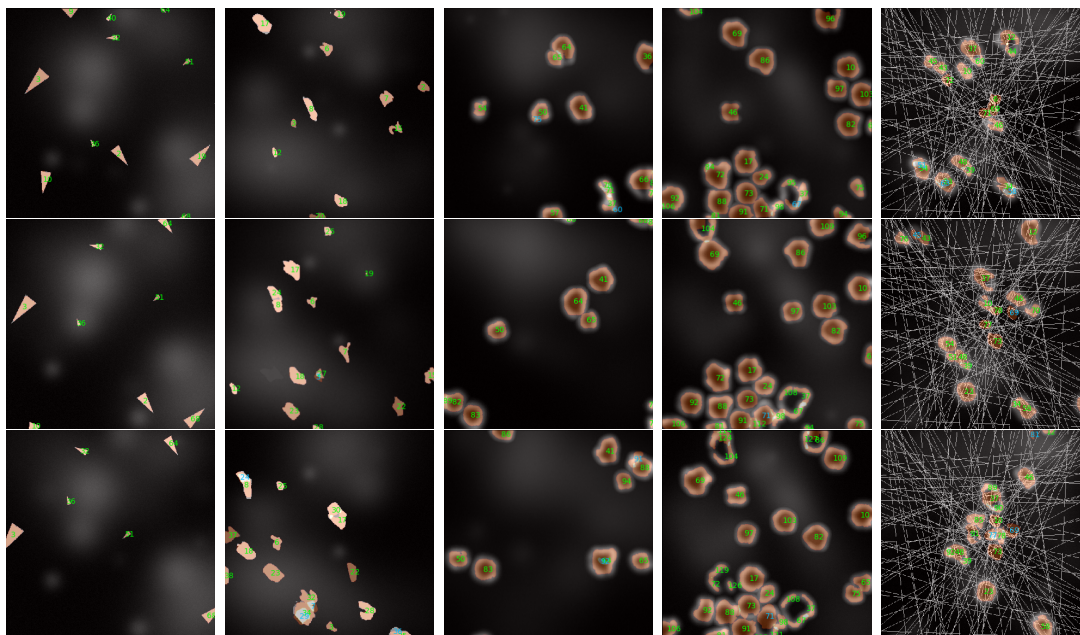


Figure 7: **Szintetikus adathalmazok követési eredményei**

Szegmentálási és követési eredmények különböző szintetikus adathalmazokon, amelyek erősen eltérő objektumviselkedéseket mutatnak. Balról jobbra sorrendben: "S. Arrows," "S. Amoeboids," "S. Amoeboids-PC," "S. Amoeboids-PCC," és "S. Amoeboids-PCCA." A megjelenített egymást követő képek mindössze 4 képkocka távolságra vannak egymástól, hogy érthető eredményeket mutassanak még rendkívül gyorsan mozgó objektumok esetén is.

	Segmentation F score	Tracking F score
S. Arrows	0.9185 ± 0.0057	0.8990 ± 0.0076
S. Amoeboids	0.7137 ± 0.0082	0.6605 ± 0.0093
S. Amoeboids-PC	0.6861 ± 0.0089	0.6662 ± 0.0098
S. Amoeboids-PCC	0.7726 ± 0.0101	0.7693 ± 0.0111
S. Amoeboids-PCCA	0.5078 ± 0.0129	0.5022 ± 0.0140
Yeast Reference	0.9234 ± 0.0136	0.9202 ± 0.0138

Table 1: **Szintetikus adathalmazok numerikus eredményei**

A teljes architektúra szegmentálási és követési F-értékei különböző szintetikus adathalmazokon, amelyek lényegesen eltérő objektumviselkedéseket és kihívásokat mutatnak a természetes élesztősejt-felvételekhez képest.

Részben a "S. Arrows" és a "S. Amoeboids" alapjaira építve, két további kísérleti környezetet hoztam létre, amelyek mindegyike több adathalmazból áll, és különböző erősségű jellegzetes viselkedéseket mutat be. A "Vizuális jelzés"

szcenárióban a szintetikus nyilak gyors forduló viselkedést mutatnak, amely csak vizuális jelek alapján jósolható meg. Ez a viselkedés analóg például autók indexlámpája által részben megjósolható mozgással vagy gyalogosok mozgásával, amely a nézésük irányából és egyéb vizuális jelekből következtethető. A "Részben véletlenszerű pozicionálás" scenárióban a szintetikus amőboidok elasztikus változásai kikapcsolásra kerülnek, az amőboidok azonban félig véletlenszerű, teljesen kiszámíthatatlan mozgási mintázatot mutatnak. E környezet célja alaptesztek nyújtása olyan potenciális valós esetekhez, ahol a vizuális jelek és az objektumok mozgása összefügg, vagy ahol az objektumok stabil követése legalább részben morfológiai vagy más vizuális jellemzőkön alapulhat, nem csupán pozíció deriváltakon. Ezekre az adathalmazokra vonatkozó összehasonlító predikciós eredmények a Tézis 1d táblázatban találhatóak.

A teljesen szintetikus adathalmazokon és az élesztőkövetési benchmarkon túl a javasolt architektúrát teszteltem a MOTSynth-MOTS-CVPR22 benchmark adathalmaz egyik változatán is. Ezen az adathalmazon a modell tanítása GTA V videojáték segítségével generált szintetikus mintákon, míg a kiértékelést kézzel annotált valós gyalogos követési és szegmentálási adatokon történik. Az adathalmaz rendkívül nagy mérete miatt kiterjedt hiperparaméter-hangolás nélkül is az architektúra **48,56** HOTA értéket ért el, amely közel azonos a 2022-ben benyújtott 48,8 benchmark értékkel. Ez a benchmark P. Bermann és munkatársai "Tracking Without Bells and Whistles" (ICCV, 2019) című, széles körben használt korszerű architektúráján alapult. Ezenkívül a javasolt architektúra által elért **82,39** AssA pontszám messze meghaladja a 44,6 benchmark értéket, jelezve az architektúra kiemelkedő objektum követési képességét. Ez az eredmény arra utal, hogy potenciálisan lényegesen magasabb HOTA értékek érhetők el csupán a szegmentációs modell továbbfejlesztésével.

Tézis 1d

Ablációs vizsgálatot végeztem az architektúrán, különválasztva a mozgásalapú és morfológián alapuló követés hozzájárulását, amelyeket az eredeti architektúra természetesen integrál. Az eredmények feltárták a mozgás és a morfológia egyéni hatását a követési teljesítményre az értékelt scenáriókban, és kiemelték az olyan követési módszerek előnyeit, amelyek mindkettőt ötvözik. Továbbá a javasolt architektúra minden scenárióban következetesen felülmúlta vagy megegyezett a széles körben használt Kalman-szűrő teljesítményével. Különösen a részben randomizált objektummozgású scenárióban az architektúra 3,49-szeres javulást ért el az asszociációs F -értékben, köszönhetően annak, hogy képes volt az összes időben lokális képalkotási információt felhasználni a követéshez.

Kapcsolódó publikáció:[Ar2]



Figure 8: „Vizuális jelzés” metrikus eredményei

Átlagos metrikus eredmények a követőmodellek (*Kalman*, *TS*, *TS-L2* és *TS-Shape*) esetében a „Vizuális jelzés” scenárióhoz tartozó *Synthetic Arrows*, *Synthetic Arrows TR-1* és *Synthetic Arrows TR-2* adathalmazokon.

Az architektúra moduláris kialakítása lehetővé tette különböző követési modellek integrált megvalósítását. Két ablatív változatot implementáltam, amelyek ugyanazokat a lokális követő modelleket használták az adott adathalmazon, de módosították a globális hozzárendelési lépést. Az egyik változatban csak a centroidok közötti euklideszi távolságot (L2) vettem figyelembe a hozzárendelés során, míg a

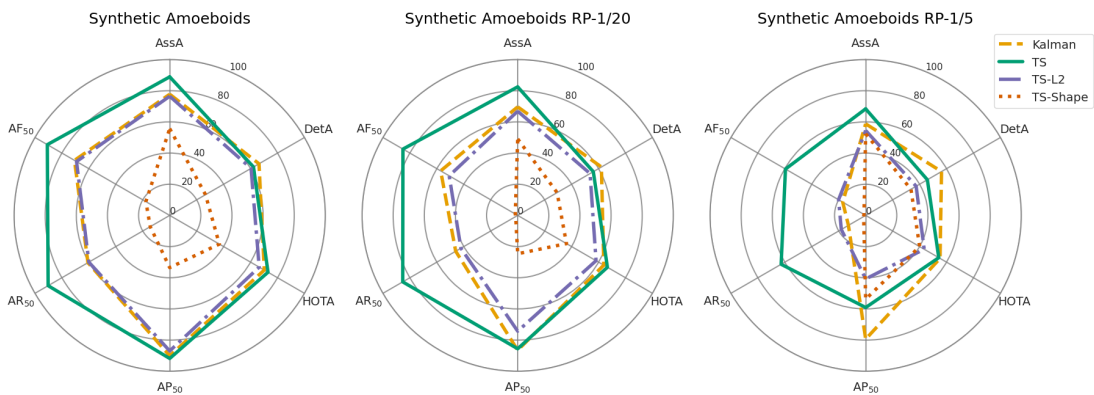


Figure 9: **„Részben véletlenszerű pozicionálás” metrikus eredményei**
 Átlagos metrikus eredmények a követőmodellek (*Kalman*, *TS*, *TS-L2* és *TS-Shape*) esetében a „Részben véletlenszerű pozicionálás” szcenárióhoz tartozó *Synthetic Amoeboids*, *Synthetic Amoeboids RP-1/20* és *Synthetic Amoeboids RP-1/5* adathalmazokon.

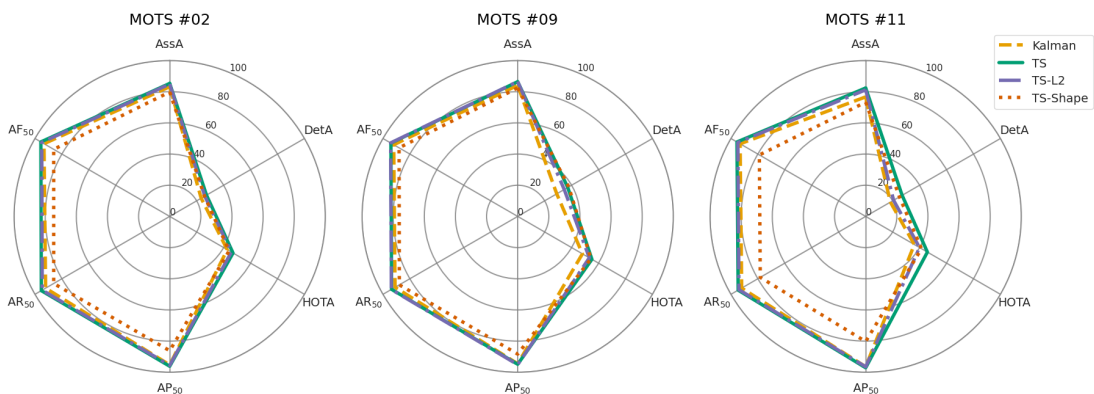


Figure 10: **MOTS adathalmaz metrikus eredményei**
 Átlagos metrikus eredmények a követőmodellek (*Kalman*, *TS*, *TS-L2* és *TS-Shape*) esetében a MOTs adathalmaz mintáin.

másikban csak a centroidhoz igazított IoU (alak-alapú, Shape) játszott szerepet. Emellett implementáltam a széles körben használt Kálman-szűrőt is.

A „Vizuális jelzés” és a „Részben véletlenszerű pozicionális” szcenáriók, valamint a MOTsSynth-MOTS-CVPR22 adathalmaz variáns értékelése a javasolt időszimmetrikus követési architektúrával (TS), annak két ablatív változatával (TS-L2 és TS-Shape), valamint a Kálman-szűrővel történt. Ezekhez az értékelésekhez egy továbbfejlesztett metrikarendszert alkalmaztam, amely eltér a sejt követés során

használt kapcsolat-alapú F-értéktől. Míg a kapcsolat-alapú F-értékek hatékonyan jelzik a követési teljesítményt, erősen korrelálnak a szegmentálási teljesítménnyel, így szükségessé válik a két metrika kombinált értelmezése. Ezzel szemben ezekhez a scenáriókhöz asszociációs teljesítményt mértem, ahol a kapcsolatokat csak akkor vettem figyelembe, ha a szegmentálás mindkét oldalon sikeres volt, kiküszöbölve a korrelációt, hogy jobban elkülöníthető legyen a követés egyéni hozzájárulása. A mért metrikák közé tartozott az asszociációs precizitás (AP_{50}), visszahívás (AR_{50}) és F-érték (AF_{50}), amelyeket mind 50%-os IoU küszöbön mértem a szegmentálási helyességhez, és 100-zal skáláztam a többi statisztikával való konzisztencia érdekében. Ezek a metrikák részletes betekintést nyújtanak a követési teljesítménybe. Emellett mértem az ipari szabványnak számító "bounding box"-alapú DetA, AssA és HOTA pontszámokat is, hogy általános teljesítményértékelést nyújtsak a detektálás, az asszociáció és az összteljesítmény szempontjából, összevethető módon a nyilvános benchmark eredményekkel, például a MOTSynth-MOTS-CVPR22 adatbázison.

A „Vizuális jelzés” scenárió eredményei a Fig. 8-ban láthatók, részletezve az eseteket, amikor nincs véletlenszerű fordulás, ritka a fordulás (TR-1), és gyakori a fordulás (TR-2). A „Részben véletlenszerű pozicionálás” scenárió eredményei a Fig. 9-ben találhatóak, lefedve az eseteket, amikor nincs véletlenszerű mozgás, kismértékű a véletlenszerű mozgás (RP-1/20), és nagymértékű a véletlenszerű mozgás (RP-1/5). A MOTSynth-MOTS-CVPR22 adatbázis eredményei a Fig. 10-ben láthatók, három manuálisan annotált értékelési minta egyedi eredményeit bemutatva. Ezek az eredmények kiemelik a követő architektúra azon képességét, hogy integrálja a pozíciós és vizuális jelzéseket, és demonstrálják a Kálman-szűrővel szembeni fölényét, még olyan feladatok esetében is, amelyekre sem specifikusan nem tervezték, sem jelentősen nem optimalizálták. Különösen a „Részben véletlenszerű pozicionálás” scenárióban a javasolt architektúra 3,49-szer magasabb asszociációs F-értéket (AF_{50}) ért el, amely jelentősen jobb követési teljesítményt tükröz ezen konkrét feladat esetében.

Tézis 2a

Bemutattam, hogy a széles körben használt referencia adathalmazok, például az MS-COCO, objektum-pozicionálási torzítást mutatnak, amely erősen kedvez azoknak az objektumoknak, amelyek a kép középpontjához közel helyezkednek el. Ez a torzítás akár több mint öt nagyságrenddel csökkentheti a predikciós teljesítményt a kép szélein, még akkor is, ha az objektumok teljesen láthatóak. A torzítás hatásait szegmentációs, detekciós és osztályozási feladatok esetében elemeztem, részletes lokalizációval a predikciós teljesítmény eredményei és aktivációs térképek alapján. Kapcsolódó publikáció: [Ar3]

A modern képfeldolgozó neurális hálózatok alapja a konvolúció művelete, amely inherensen eltolás-invariáns tulajdonságú. Ez a tulajdonság lehetővé teszi a konvolúciós neurális hálózatok (CNN-ek) eltolás-invariáns viselkedését. Bár a publikáció időpontjában már ismert volt, hogy a CNN-ek feltételezett eltolás-invariáns viselkedése nem mindig érvényesül, ennek a hatásnak az alapvető oka nem volt mélyen vizsgálva, és az esetleges intenzitás és hatás nem volt egyértelműen bemutatva. Az általam végzett kezdeti mérések kimutatták, hogy a képek középpontjához közel elhelyezett objektumokkal végzett tanítás erősen módosíthatja a predikciós teljesítményt a látótér széleinek közelében. Ezek a tesztek több mint 16,000-szeres teljesítménycsökkenést mutattak, még akkor is, ha az objektumok teljes mértékben a látómezőben maradtak.

Ezen hatás jelentőségét tovább növeli a népszerű és széles körben használt benchmarkok erős objektum-pozicionálási torzítása. Ez a torzítás, amely jelen van mind a tanítási, mind a tesztelési adathalmazokban, az értékelés során észrevétlen marad. Jelenléte csak akkor válik nyilvánvalóvá, amikor a tanított modellek részlegesen kudarcot vallanak olyan valós alkalmazásokban, ahol a torzítás nem létezik.

A torzítás mértékének és lokalizációjának alapos vizsgálata érdekében részletesebb méréseket végeztem különböző képfeldolgozási feladatokban. Ezekhez a kísérletekhez több U-net modellt tanítottam kontrollált adathalmazon, az ImageNet mintáit háttérként felhasználva, hogy a feladat ne legyen triviális. A kézzel rajzolt MNIST számokat ezekre a képekre helyeztem szegmentálási és osztályozási célból, úgy, hogy pozíciójukat korlátoztam a tanítási és értékelési fázisok során

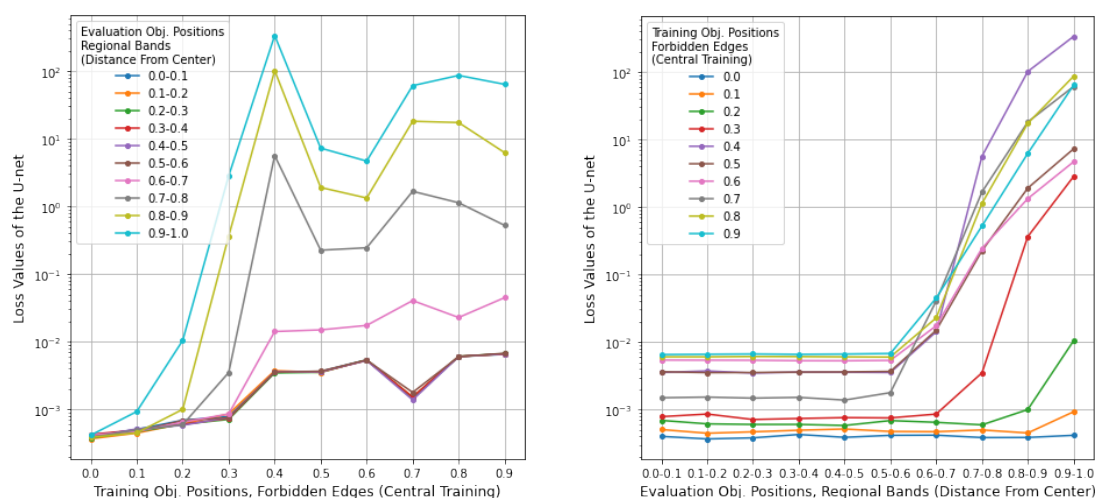


Figure 11: **Pozícionálási torzítás eloszlásai**

U-Netek átlagos veszteségértékei korlátozott régiókban tanítva és értékelve. A veszteségértékek drasztikus növekedése kezdődik a 0,5-0,6 értékelési sávnál, kivéve, ha az objektumokat kifejezetten a szélekhez közel helyezték el a tanítás során. Ez azt jelzi, hogy a képek több mint 64 százaléka lényegesen rosszabb teljesítménnyel lesz szegmentálva és osztályozva az elfoglalt terület szempontjából.

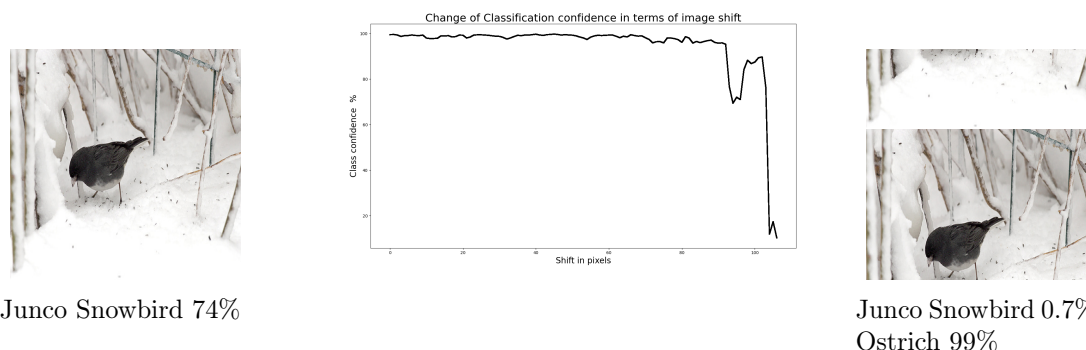


Figure 12: **Osztályozási torzítás példája**

"Junco Snowbird" példa, amely bemutatja az objektumok kép határára történő eltolásának drasztikus hatását az osztályozó modellekre, a VGG-16 előtanított verziójával és ImageNet mintákkal. Az első oszlop az eredeti, helyesen osztályozott képeket mutatja. A második oszlop azt ábrázolja, hogyan változik az eredeti osztály osztályozási magabiztossága a kép eltolásával. Az utolsó oszlop a legnagyobb vizsgált eltolást mutatja. Az eredeti osztály és az újonnan prediktált osztály osztályozási magabiztossága az egyes képek alatt szerepel az értékekkel együtt.

meghatározott régiókra. A tanítás során az objektumok csak központosított régiókban helyezkedhettek el, míg az értékelés során a középponttól meghatározott távolságra lévő sávokra korlátozódtak. Ezen kísérletek eredményei a Fig. 11 ábrán

láthatók.

Ezen túlmenően, előtanított modellekkel végzett példány szegmentálási és osztályozási feladatokat is vizsgáltam, ahol az objektumokat manuálisan a kép széleihez közel helyeztem el. Egy példát az osztályozási eredményekre a Fig. 12 ábrán mutatok be. Az eredmények megerősítik, hogy a predikciós teljesítmény drasztikusan csökken a képek szélein minden vizsgált feladat esetében, kivéve, ha az objektumokat a kép széleihez közel helyezték el a tanítás során.

A torzítás további lokalizációja és esetleges okainak feltárása érdekében aktivációs térképeken alapuló különbségképeket számoltam ki, amelyek az objektumok eltolódott pozíciói alapján készültek. Ezekhez a modellekhez az előzőekben leírt adathalmazon tanítottam, ahol az ImageNet minták szolgálták háttérként, az MNIST minták pedig szegmentálandó és osztályozandó objektumként. Példaeredmények a Fig. 13 ábrán találhatóak.

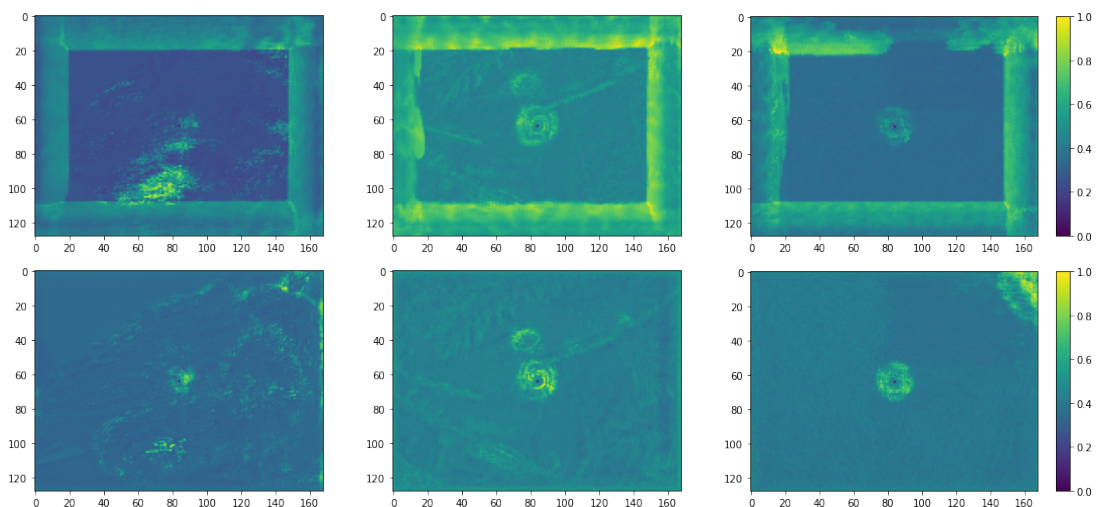


Figure 13: **Aktivációs térképek különbségei**

Normalizált aktivációs térkép-különbségek az eltolási vektorok alapján, ahol a nulla eltolás a középpontban helyezkedik el. A felső sor azokat az eredményeket mutatja, amelyknél az objektumok középen helyezkedtek el a tanítás során, a kép központi 30 százalékán. Az alsó sor azoknak az eredményeit mutatja, amelyknél az objektumokat a széleken helyezték el a tanítás során, a kép központi 70 százalékán kívül. Mint látható, a csak középen található objektumokat tartalmazó képeken tanított CNN-ek markáns eltéréseket mutatnak aktivációs térképeikben, ha az objektumok a széleken jelennek meg. Ezzel szemben az objektumokat a széleken tartalmazó képeken tanított CNN-ek nem mutatnak ilyen különbségeket.

Tézis 2b

Architektúrával és adatmanipulációval kapcsolatos megoldásokat javasoltam az objektum-pozicionálási torzítás csökkentésére. A leghatékonyabb megközelítés az volt, hogy az összes konvolúciós rétegben a null-kiegészítést toroidális határfeltételekkel helyettesítettem. Ez a módosítás több mint 37,000-szeres teljesítménynövekedést eredményezett a legszélsőségesebb esetekben.

Kapcsolódó publikáció: [Ar3]

A Tézis 2a-ban leírt mérési eredmények alapján több potenciális megoldást javasoltam az objektum-pozicionálási torzítás csökkentésére. Ezek közül kettő mutatta a legjobb teljesítményt: a tanítóminták véletlenszerű eltolása, még az objektumok részleges elvesztésének kockázatával is, valamint a toroidális határfeltételek bevezetése minden konvolúciós rétegben. Az utóbbi szinte semmilyen hátránnyal nem jár, eltekintve az architekturális módosítás szükségességétől.

	MS-COCO Orig	MS-COCO Shifted
Box mAP Orig.	33.6%	15.6%
Box mAP Shifted	31.7%	28.4%
Seg mAP Orig	31.4%	17.4%
Seg mAP Shifted	29.4%	25.3%

Table 2: **Detekciós torzítás kiküszöbölése**

Mask R-CNN átlagos precíziós (mAP) eredmények, ResNet-50 modell gerinccel és ROI-align-al rendelkező "feature pyramid" hálózattal az MS-COCO adathalmazon. Kétféle adathalmaz került kiértékelésre: az eredeti MS-COCO adathalmaz (MS-COCO Orig) és egy módosított változat, ahol egy véletlenszerűen kiválasztott objektum mindig a kép szélére lett eltolva (MS-COCO Shifted). Az mAP értékek "bounding box" objektumdetekciós (Box) és példány-szegmentációs (Seg) feladatokhoz lettek rögzítve, két különböző tanító-környezetben. Az (Orig.) jelöléssel ellátott sorok az eredeti MS-COCO adathalmazon elért eredményeket mutatják, míg a (Shifted) jelölésű sorok olyan hálózatok mAP-értékeit tartalmazzák, amelyek olyan adathalmazon lettek tanítva, ahol egy véletlenszerű objektum mindig a kép szélére került.

Az osztályozási, példány-szegmentálási és szemantikus szegmentálási feladatok eredményeit a Tézis 2a-ban leírtak szerint értékeltem, akár egyenként, akár kombinációban alkalmazva. Ezek az eredmények a Tab. 3, a Tab. 2 és a Fig. 14

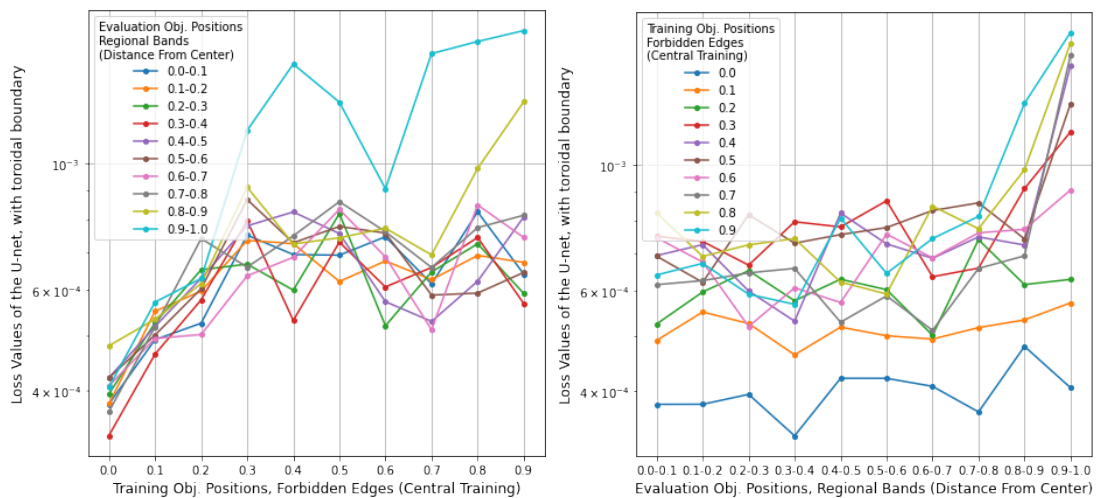


Figure 14: **Pozicionálási torzítás eloszlásai - Toroidális határfeltétel**
 U-Netek átlagos veszteség értékei, amelyek korlátozott régiókon lettek tanítva, minden konvolúciós rétegben toroidális határfeltételekkel. Az Fig. 11 ábrán látható értékekhez képest a tanító- és kiértékelési pozíciók alapján mért veszteségi értékek relatív növekedése minimális.

ábrákban található. Az eredmények azt mutatják, hogy a toroidális határfeltételek használata kissé jobb teljesítményt nyújt, mint a véletlenszerű eltolások. Ez az előny valószínűleg abból fakad, hogy a nagy véletlenszerű eltolások jelentős artefaktokat hozhatnak létre, például az objektumok részleges eltávolítását a látómezőből. Mindazonáltal mindkét megoldás messze felülmúlja az eredeti modelleket, és szinte teljesen megszünteti a torzítás hatásait.

Architecture	ImageNet	ImageNet Boundary
DenseNet121 Orig.	75.1%	3%
DenseNet121 Shifted	69.3%	66%
DenseNet121 Toroidal	74.2%	68%

Table 3: **Osztályozási torzítás kiküszöbölése**

A DenseNet121 architektúra top-1 teszt pontossági eredményei az eredeti ImageNet teszt halmazon (első oszlop) és egy manuálisan létrehozott kis elemszámú mintacsoporton, amelyben az objektumok a kép szélére lettek pozicionálva (második oszlop). A modell három változatban lett értékelve: az eredeti ImageNet tanítóhalmazon null-kiegészítéssel tanítva (Orig.), egy olyan adathalmazon tanítva, ahol az objektumok a kép szélére lettek tolvá (Shifted), és az eredeti adathalmazon toroidális határfeltételekkel (Toroidal).

Tézisekhez kapcsolódó publikációim

- [Ar1] G. Szabó, P. Bonaiuti, A. Ciliberto, and A. Horváth, “Enhancing cell tracking with a time-symmetric deep learning approach,” *arXiv preprint arXiv:2308.03887*, 2023. (document)
- [Ar2] G. Szabó, Z. Molnár, and A. Horváth (2024). "Post-Hoc MOTS: Exploring the Capabilities of Time-Symmetric Multi-Object Tracking", *arXiv preprint arXiv:2412.08313*, 2024 (document)
- [Ar3] G. Szabó and A. Horváth, “Mitigating the bias of centered objects in common datasets,” in *2022 26th International Conference on Pattern Recognition (ICPR)*, pp. 4786–4792, IEEE, 2022. (document)

Egyéb publikációim

- [Au1] R. Bagdy-Bálint, G. Szabó, Ö. H. Zováthi, B. H. Zováthi, Á. Somorjai, C. Köpenczei, and N. K. Rózsa, “Accuracy of automated analysis in cephalometry,” *Journal of Dental Sciences*, Elsevier, 2024.
- [Au2] D. Babicz, S. Kontár, M. Peto, A. Fulop, G. Szabó, and A. Horváth, “Receptive field size optimization with continuous time pooling,” in *Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision*, pp. 1449–1458, 2021.
- [Au3] W. A. Bosbach, B. Németh, R. Zelei, J. F. Senge, B. Pasztor, L. Ebner, M. Szabo, S. Anderson, G. Szabo, P. Dlotko, A. Horvath, K. Daneshvar, and J. Heverhagen, “SCR 2023 poster presentation: An open access AI-based pattern recognition tool for application in MSK imaging – suitability and limitations of resources today [Poster],” *Swiss Congress of Radiology*, Zenodo, 2023. DOI: 10.5281/zenodo.7958527.
- [Au4] G. Szabó, “Time-symmetric generalization of multi-object tracking,” *PhD Proceedings Annual Issues of the Doctoral School Faculty of Information Technology and Bionics*, vol. 19, pp. 189–192, 2024. [Online]. Available: <https://m2.mtmt.hu/api/publication/35446463>.

- [Au5] G. Szabó, “Time symmetric tracking of yeast cells using convolutional neural networks,” *PhD Proceedings Annual Issues of the Doctoral School Faculty of Information Technology and Bionics*, vol. 18, pp. 145–148, 2023. [Online]. Available: <https://m2.mtmt.hu/api/publication/34176533>.
- [Au6] G. Szabó, “The effects and mitigation of object centering bias in common datasets,” *PhD Proceedings Annual Issues of the Doctoral School Faculty of Information Technology and Bionics*, vol. 17, pp. 191–194, 2022. [Online]. Available: <https://m2.mtmt.hu/api/publication/35446504>.
- [Au7] G. Szabó, “Investigation of cell motility,” *PhD Proceedings Annual Issues of the Doctoral School Faculty of Information Technology and Bionics*, vol. 16, pp. 169–172, 2021. [Online]. Available: <https://m2.mtmt.hu/api/publication/35446512>.