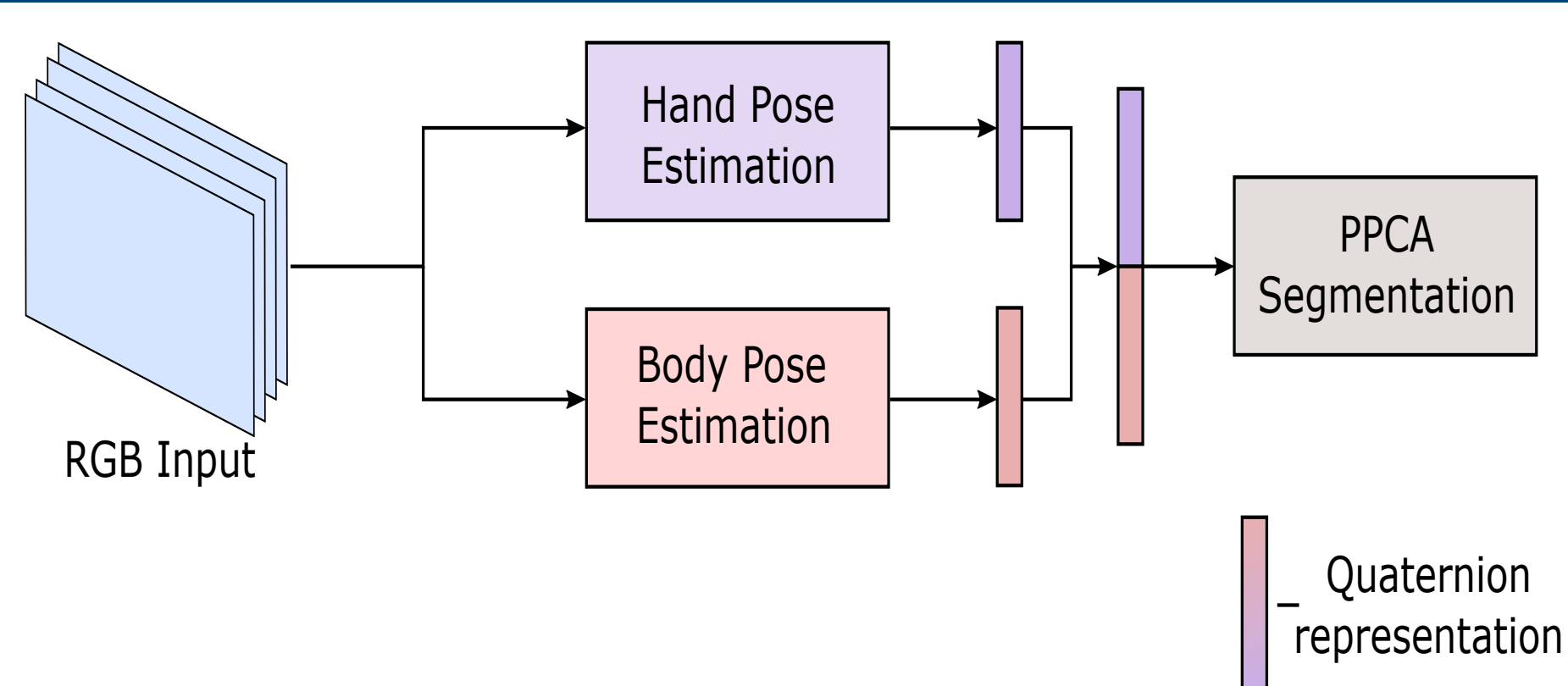


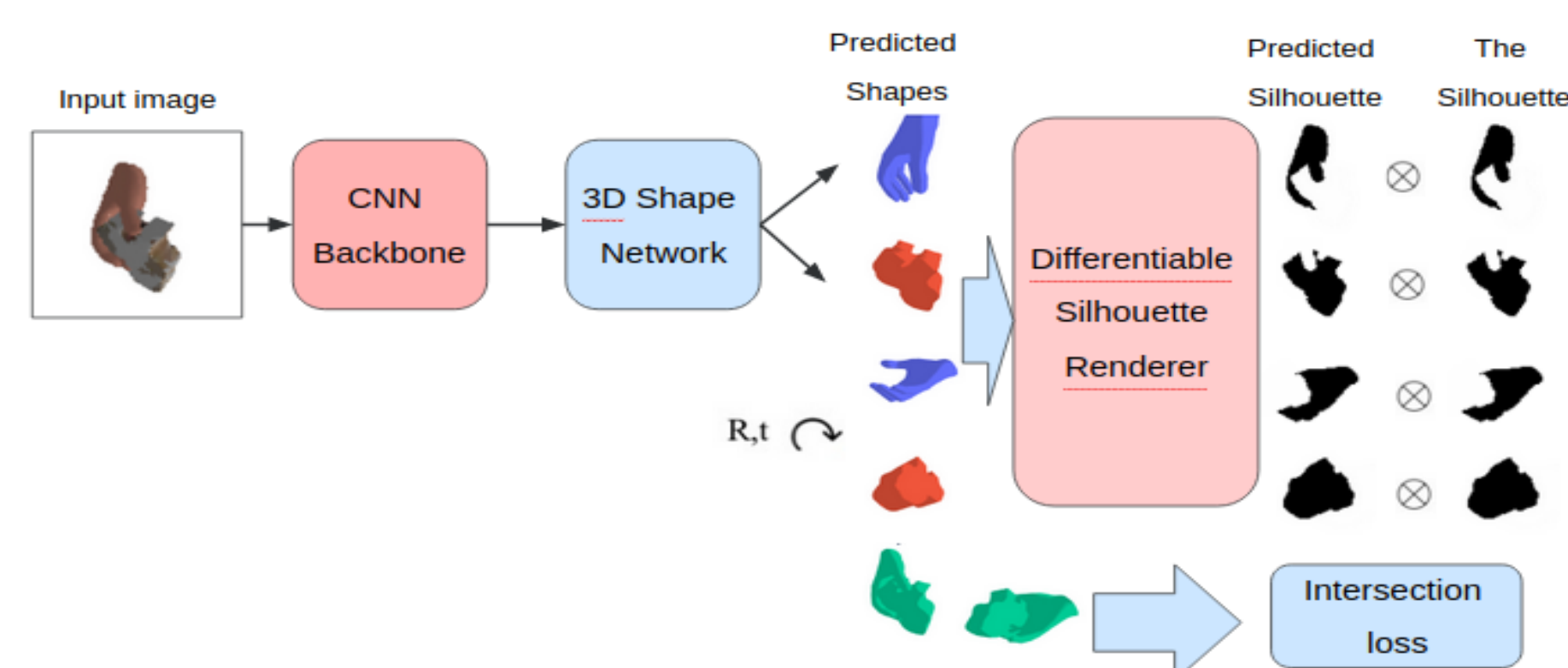
Bevezetés

- ▶ Az emberi kommunikáció megértése elengedhetetlen az ember és gép közötti interakcióhoz. A kézhez kapcsolódó kommunikációs tevékenység a tárgy manipulálása és a gesztikuláció.
- ▶ A kézmozdulatok és a beszéd együtt vannak jelen a nyelvi rendszeren belül [1]. A beszédfejlődés során bizonyos körülmények, mint például az autizmus, befolyásolhatják a kéz használatának módját.
- ▶ Kutatásunk ezen kommunikációs készségek megfigyelését célozza meg.
- ▶ Többféle technológiát ötvözünk a test és kéz póz meghatározásához. Ebből aztán egy PPCA alapú szegmentáló eljárással terveink szerint képesek leszünk megkülönböztetni az egyes gesztusokat.
- ▶ Mesh alapú póz becslő technológia fejlesztésével kéz és objektum együttes helyzetét igyekszünk meghatározni, így detektálva az objektum manipulációt, tárgy átadás-átvétel folyamatát.
- ▶ Az öt-tíz éves gyerekek autizmus diagnosztizálására kifejlesztett Autism Diagnostic Observation Schedule 2 (ADOS-2) tesztsorozat videóit nyújtják a feldolgozandó példákat.
- ▶ Az ADOS-2 több részfeladatból áll, amelyek kimutatják az autizmus különböző jellemzőit. Mi a testbeszédet érintő jegekre koncentrálnunk.

Áttekintés



1. ábra. Az előfeldolgozás és szegmentáció lépései, melyek során a bejövő videófelvételt részletekre bontjuk az elkülönülő gesztusok szerint.



2. ábra. Az algoritmusunk egyszerre becsli a kéz és objektum 3D reprezentációját, így pontosabb képet kaphatunk azok térbeli kapcsolatáról. Tanítás során nem használunk fel térbeli felügyelő információt, így megkönnyebbítve az annotációt.

Ipari alkalmazások és együttműködések

- ▶ Rush Medical School
- ▶ Argus Cognitive, Inc.
- ▶ „Memory Game” HumanE-AI mikroprojekt
- ▶ „AI @ Clinic” EU project proposal, NTNU, Bern Univ, Lyon Univ, Radboud Univ



3. ábra. Emberi, főleg gyermekek társas viselkedés vizsgálatát segítjük a mesterséges intelligencia eredményeinek alkalmazásával



4. ábra. Alkalmazási lehetőség pl. megfigyelés üzletben

Alkalmazott módszerek

- ▶ A példafeladatok egyike: Fogmosás, Arcmosás
- ▶ A feladat megoldásának a lépései az ábra 2 alatt tekinthetők meg. A nyers videó bemenetet feldolgozza a mesh alapú Frankmocap modell [2] és a test ízületeinek 3D koordinátáit becslő algoritmus [3].
- ▶ Az így számszerűsített értékek mátrixa reprezentálja az egymást átfedő frém sorozatokat. A reprezentáción dolgozik PPCA szegmentáló algoritmusunk, ami Barbic modelljére [4] épül.
- ▶ Kéz objektum becslő betanításához szintetikus előállított példákat generálunk, amivel gyorsan és olcsón tudunk előállítani nagyszámú, változatos tanító adatot.
- ▶ Az ADOS-2 teszt során használt tárgyakhoz elkészítettük azok 3D mesh reprezentációját fotogrammetria eljárással.
- ▶ Modellünk tanításához nincs szükség 3D tanító információra. Elég a tanító kép mellé a tárgy és kéz síkbeli sziluettje. Csak 2D információt felhasználva a modellünk képes 3D reprezentáció becslésére.

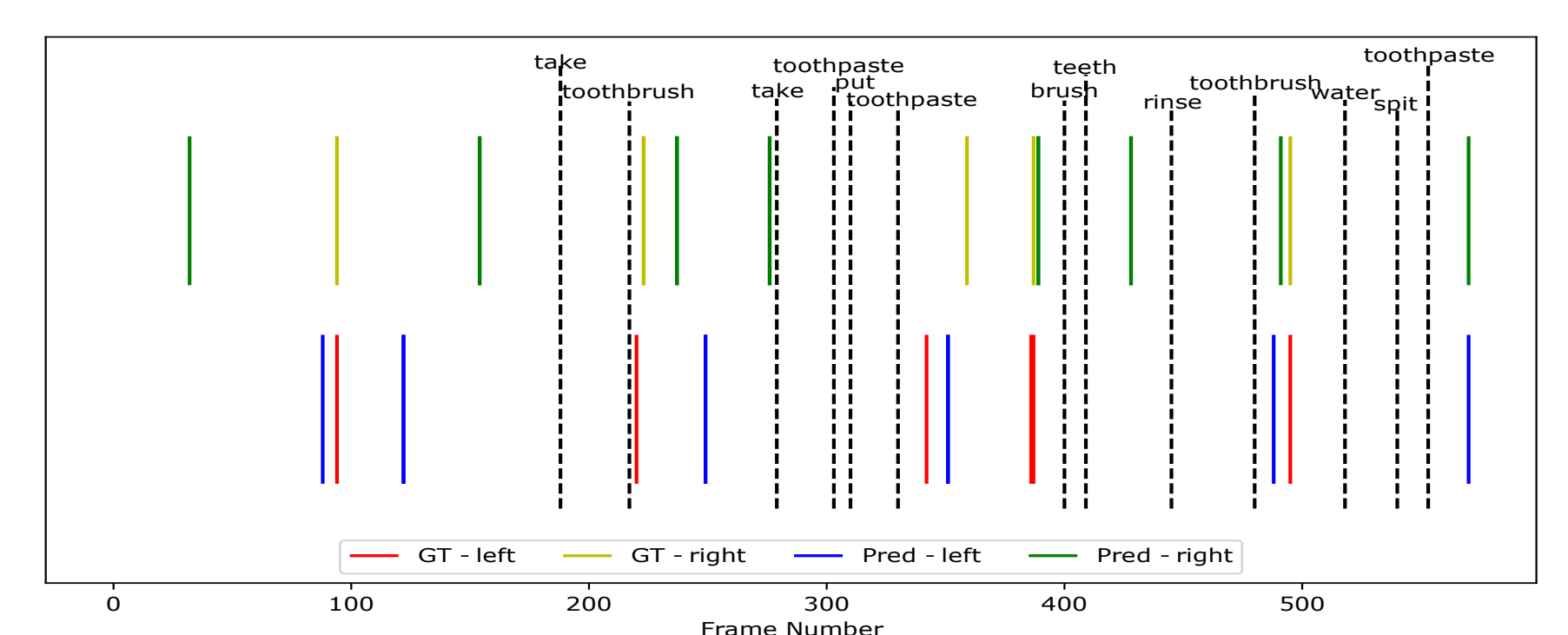


5. ábra. Mesh reprezentáció készítéséhez felállított stúdió



6. ábra. Térbeli objektum-kéz becslő algoritmusunk eredményei szintetikus adatokon

Eredmény



7. ábra. Becsült pillanata a gesztikuláció változásának amit a jobb kéz (felső vonalak) illetve a bal kéz (alsó vonalak) alapján számítunk. A videofelvétel alatt a feladat végrehajtásával kapcsolatos kulcsszavak hangzanak el, amik az ábra tetején olvashatóak. Ezek időpillanatát a függőleges szaggatott vonalak jelzik. A kulcsszavakból tudunk következtetni, mikor követik egymást a részfeladatok.



8. ábra. Egy példa sorozat hossza 90 frém, amiből az ábrán minden tizediket jelenítünk meg. A második sor második képén látható az a pillant, amikor változást detektálunk a megfigyelt gesztusokban.

[1] Adam Kendon. *Gesticulation and Speech: Two Aspects of the Process of Utterance*, pages 207–228. De Gruyter Mouton, 1980.

[2] Yu Rong, Takaaki Shiratori, and Hanbyul Joo. Frankmocap: Fast monocular 3d hand and body motion capture by regression and integration. *arXiv preprint arXiv:2008.08324*, 2020.

[3] M. Véges and A. Lőrincz. Temporal smoothing for 3d human pose estimation and localization for occluded people. In *Neural Information Processing*, pages 557–568, 2020.

[4] Jernej Barbic, Alla Safonova, Jia-Yu Pan, Christos Faloutsos, Jessica Hodgins, and Nancy Pollard. Segmenting motion capture data into distinct behaviors. volume 2004, pages 185–194, 01 2004.