

# 3D orientacijos nustatymas

Mykolas Skrodenis  
Informatikos institutas  
Matematikos ir informatikos fakultetas  
Vilnius, Lietuva  
mykolas.skrodenis@gmail.com

**Santrauka**—Šioje darbo apžvalgoje aptariama objektų aptikimo ir 3D orientacijos įvertinimo sritis, ypač naudojant "Objectron" duomenų rinkinį. Išnagrinėta, kaip iš 2D objektų aptikimo galima plėsti į 3D orientacijos nustatymą ir kokie tai turi privalumai. Remiantis "deeplake" platforma, naudojant dviejų pakopų mediapipe sprendimus su MobileNetv2 architektūra, modelis gali tiksliai įvertinti objekto dydį, padėti ir orientaciją 3D erdvėje. Ši naujausia architektūra pasižymi dideliu tikslumu ir skaičiavimo efektyvumu. Taip pat paminėta, kad "Objectron" duomenų rinkinys yra vertingas šaltinis tyrėjams ir kūrėjams, siekiantiems spręsti 3D orientacijos uždavinius. Galiausiai, aptarti šios technologijos taikymo būdai, pvz., papildytoji realybė ir robotika.

**Index Terms**—Objektų aptikimas, 3D orientacija, Objectron duomenų rinkinys, 2D objektų aptikimas, 3D objektų aptikimas

## I. ĮVADAS

Objektų aptikimas yra gerai ištirtinėta kompiuterinės regos sritis, kuri taikoma tiek praktinėje veikloje, tiek akademiniuose tyrimuose. Tačiau didžiojoje šių tyrimų dalyje daugiausia dėmesio skiriama 2D objektų nuspėjimui. Tačiau, išplėtus šią sritį į 3D, galima nustatyti objekto dydį, padėti ir orientaciją realiaame pasaulyje, o tai atveria plačias taikymo galimybes [1]. Pastaraisiais metais šioje srityje padaryta didelė pažanga, nes atsirado didelės apimties duomenų rinkiniai ir patobulinti gilaus mokymosi metodai. Tarp šių duomenų rinkinių "Objectron" duomenų rinkinys tapo vertingu šaltiniu tyrėjams ir kūrėjams, siekiantiems spręsti 3D orientacijos įvertinimo uždavotį. Šio darbo pagrindinis tikslas - ištirti 3D orientacijos įvertinimą naudojant "Objectron" duomenų rinkinį. Išnagrinėsime sunkumus, susijusius su teisingu objekto orientacijos nustatymu realiose situacijose, ir aptarsime galimus šios technologijos panaudojimo būdus. Taip pat apžvelgsime naujausiose moksliniuose tyrimuose naudotus metodus ir algoritmus šiai problemai spręsti, pažymėdami jų privalumus ir trūkumus.

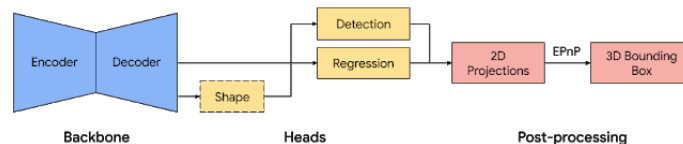
## II. METODAI

Šiai uždavotiai atlikti nagrinėjami du mediapipe sprendimai: Vieno etapo metodas, ir dviejų etapų metodas. Duomenų rinkiniui užkrauti ir naudoti, panaudota "deeplake" platforma. Užuoat atsisiuntę "Objectron" duomenų rinkinį "Python", galima be vargo matyti ir užkrauti duomenis, naudodami tik vieną kodo eilutę. Prieš pereinant prie tolesnių veiksmų, būtina konvertuoti į RGB, nes surinkti vaizdai rodomi BGR spalvų formatu.

### A. Vieno etapo metodas

Modelio pagrindą sudaro kodavimo ir dekodavimo architektūra, sukurta remiantis MobileNetv2 [3]. Taikomas kelių uždavotų mokymosi metodas, kartu numatant objekto formą ir aptikimo bei regresijos uždavotus. Formos uždavinys prognozuoja objekto formos signalus, remdamasis turima pagrindine anotacija, pavyzdžiui, segmentavimu, nors jis nėra privalomas, jei mokymo duomenyse nėra formos anotacijos. Aptikimo uždavotiai atlikti naudojamos anotuotos 3D lokalizacijos dėžės, kad atitiktų Gauso pasiskirstymą, kurio centras yra dėžės centre, o standartiniai nuokrypiai yra proporcingi dėžės dydžiui. Šis pasiskirstymas atspindi objekto centro vietą, siekiant ją tiksliai nuspėti. Regresijos uždavinys įvertina aštuonių ribojančiųjų dėžučių viršūnių 2D projekcijas, dar labiau patikslina objekto vietos nustatymą ir suteikia tikslas ribų koordinatas.

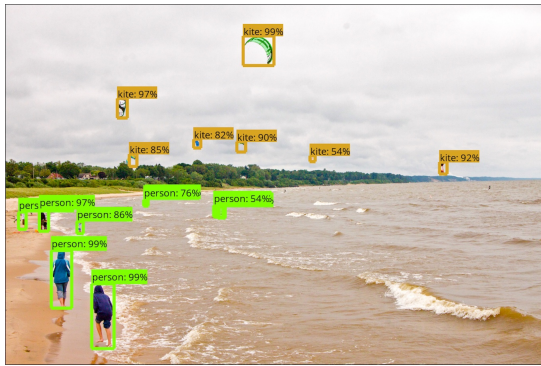
Galutinėms ribojančio lauko 3D koordinatėms gauti naudojamas gerai žinomas pozos įvertinimo algoritmas, vadinamas EPnP (Efficient Perspective-n-Point). EPnP leidžia atkurti objekto 3D ribinę dėžę net ir neturint išankstinių žinių apie objekto matmenis. Ši galimybė leidžia tiksliai įvertinti objekto pozą ir dydį, o tai padeda išsamiau suprasti objekto erdvines charakteristikas.



1 pav. – vieno etapo modelis [2]

### B. Dviejų etapų metodas

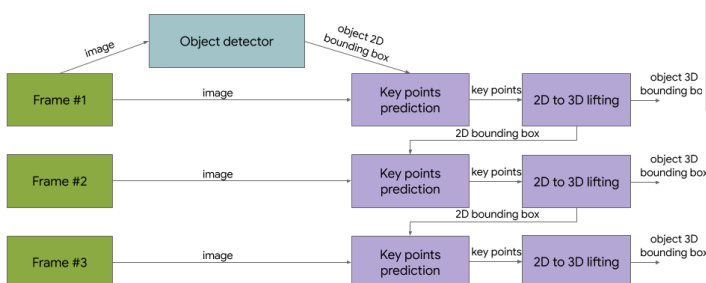
Pirmajame etape naudojamas "TensorFlow" objektų aptikimo modelis, kad būtų galima rasti 2D objekta (TensorFlow Object Detection API). Pasinaudojant modelio galimybėmis, pirmajame etape efektyviai nustatoma dominanti sritis, kurioje yra dominantis objektas. Gavus 2D apybraižą, antrasis etapas perima objekto 3D riboženklį įvertinimą. Tuo pat metu antrajame etape apskaičiuojamas ir kito kadro objekto 2D apkarpymas, taip sumažinant poreikį paleisti objekto detektorių kiekviename kadre. Šis optimizavimas gerokai padidina skaičiavimo efektyvumą, nes objekto detektorių galima apeiti vėlesniuose kadruose, kol nereikia atlikti naujo apkarpymo.



2 pav. – Tensorflow objektų lokalizavimo API pavidys [4]

Antrasis etapas, atsakingas už 3D ribinių laukų prognozavimą, yra skirtas tikslams įvertinimams atlikti remiantis pateiktu 2D apkarpytu. Naudodamasis apkarpytame regione esančia vaizdine informacija, modelis gali nustatyti objekto erdvinis matmenis ir padėti 3D erdvėje. Šis dviejų pakopų metodas padidina bendrą 3D ribinio lauko įvertinimo proceso tikslumą ir patikimumą.

Be to, antrojo etapo 3D ribinių laukų prognozavimas demonstruoja išpūdingą našumą, pasiekdamas aukštą 83 kadrų per sekundę (FPS) kadrų dažnį naudojant "Adreno 650" mobilųjį GPU [1]. "Objectron" modelio versijose naudojama dviejų pakopų architektūra yra didelė pažanga, palyginti su ankstesniu vieno etapo metodu. Pirmajame etape naudojant "TensorFlow" objektų aptikimo modelį 2D pasėliui gauti, o vėliau naudojant specialų antrąjį etapą 3D ribinių laukų įvertinimui, modelis pasiekia didesnę tikslumą ir skaičiavimo efektyvumą. Ši atnaujinta architektūra, pasižyminti dideliu kadrų dažniu mobiliuosiuose GPU, puikiai tinka realaus laiko taikomosioms programoms, kurioms reikia tikslaus objektų aptikimo ir patikimo 3D įvertinimo.



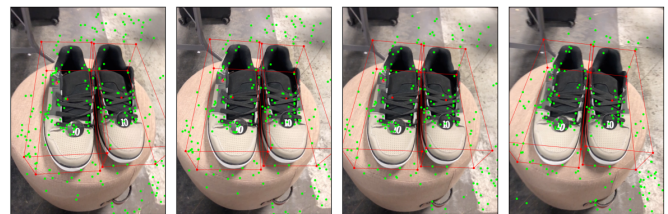
3 pav. – Dviejų etapų modelis

### III. DUOMENYS

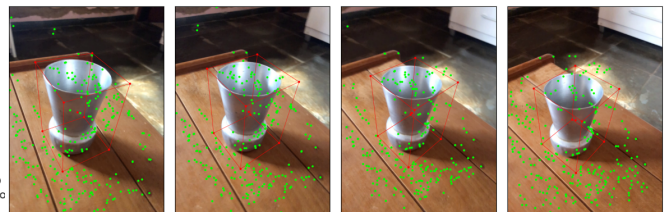
Darbe naudojamas "Objectron" duomenų rinkinys. "Objectron" duomenų rinkinį sudaro keli trumpi, į objektą orientuoti vaizdo įrašai, kuriuose kamera lėtai sukasi aplink objektą ir fiksuoja jį iš įvairių perspektyvų. Filmukuose yra rankiniu būdu anotuoti 3D ribiniai langeliai, kuriuose nurodyta kiekvieno objekto vieta, orientacija ir dydis. Informaciją taip pat

sudaro kameros orientacija, reti taškų debesys ir plokštuminių paviršių apibūdinimas iš supančios aplinkos. Visa tai yra AR seansų metaduomenys. Duomenų rinkinys buvo surinktas iš 10 šalių penkiuose skirtinguose žemynuose, taip užtikrinant geografinę įvairovę. Jį sudaro dviračiai, knygos, fotoaparatai, sausų pusryčių dėžės, kėdės, puodeliai, nešiojamieji kompiuteriai ir batai, taip pat 14 819 anotuotų vaizdo įrašų ir 4 mln. anotuotų nuotraukų. Neapdorotų duomenų rinkinio dydis - 1,9 TB (įskaitant vaizdo įrašus ir jų anotacijas). Bendras duomenų rinkinio dydis - 4,4 TB (įskaitant vaizdo įrašus, įrašus, sekas ir t. t.). Duomenys pasiekiami naudojant "gsutil", HTTP API "<https://storage.googleapis.com/objectron>" arba naudojant debesinę python saugyklą. Duomenys saugomi "bucket" pavidalu. Duomenyse yra:

- Trumpi vaizdo įrašai
- Anotacijų etiketės, kuriose yra objektų 3D lokalizavimas.
- AR metaduomenys (pvz., kameros pozicijos, taškų debesys ir plokštuminiai paviršiai).
- Apdorotas duomenų rinkinys: suskaidyti ir išmaišyti tf.įrašai su anotuotais kadrų, tf.example formatu.



4 pav. – Iš duomenų rinkinio atvaizduoti batu 4 video rėmai su anotacijomis

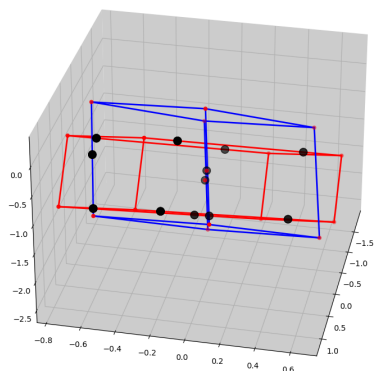


5 pav. – Puoduko 4 video rėmai

### IV. REZULTATAI

Sprendžiant užduotį, tvarkingai veikiančio modelio sukurti ir apmokyti nepavyko. Rezultatam gauti, buvo numatyta, apskaičiuojant atlikto spėjimo IoU metrikos pagalba. Šiai metrikai skaičiuoti paruoštą "from objectron.dataset import iou" biblioteka iš "<https://github.com/google-research-datasets/Objectron/blob/master/objectron/dataset/iou.py>".

Mediapipe turi laisvai prieinamą modelį. "Objectron" modelis yra galingas "MediaPipe" teikiamas įrankis 3D objektams aptikti. Jis naudoja mašininio mokymosi metodus, kad įvertintų 3D objektų orientaciją vaizduose ar vaizdo įrašuose, o tai gali būti pritaikoma įvairiai - nuo papildytosios realybės iki robotikos. Modelis geba atpažinti batus, kėdes, kamerą ir



6 pav. – Mėlyna forma yra tiesa, Raudona spėjimas

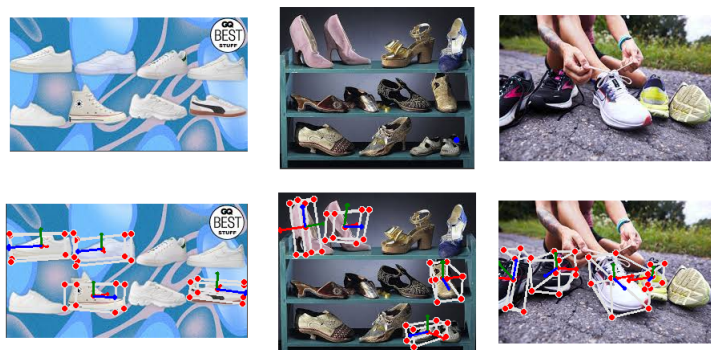
```
iou = 0.34046594050485063
iou (via sampling)= 0.3363526708914891
```

7 pav.

puodelius Tikrinant modelio veikimą, galima pastebėti modelio veikimą, net sudėtingas nuotraukas su daug objektų, gana tiksliai įvardija objektų orientacijas.

#### LITERATŪRA

- [1] Artsiom Ablavatski Jianing Wei Matthias Grundmann Adel Ahmadyan, Liangkai Zhang. Objectron: A large scale dataset of object-centric videos in the wild with pose annotations. *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2021.
- [2] Adel Ahmadyan, Tingbo Hou, Jianing Wei, Liangkai Zhang, Artsiom Ablavatski, and Matthias Grundmann. Instant 3d object tracking with applications in augmented reality, 2020.
- [3] Tingbo Hou, Adel Ahmadyan, Liangkai Zhang, Jianing Wei, and Matthias Grundmann. Mobilepose: Real-time pose estimation for unseen objects with weak shape supervision. *CoRR*, abs/2003.03522, 2020.
- [4] Jonathan Huang, Vivek Rathod, Chen Sun, Menglong Zhu, Anoop Korattikara, Alireza Fathi, Ian Fischer, Zbigniew Wojna, Yang Song, Sergio Guadarrama, and Kevin Murphy. Speed/accuracy trade-offs for modern convolutional object detectors. *CoRR*, abs/1611.10012, 2016.



8 pav. – Tiramas Mediapipe objectron sprendimas