

FELADATKIÍRÁS

Napjaink egy igen aktívan fejlődő technológiai területe a 3D adatok rekonstrukciója és feldolgozása. Egy lehetséges alkalmazás például telefonnal vagy járműre szerelt kamerával készített képek sorozata alapján virtuális színteret rekonstruálni, ami később bejárható és módosítható egy képernyőn vagy virtuális valóság környezetben keresztül. Az elmúlt években egyre nagyobb teret nyernek a neurális hálókra alapuló rekonstrukciós módszerek, azon belül is az ún. neurális radiancia mezők (Neural Radiance Fields (NeRF)) módszere, ami a térfogati képalkotás inverzióján alapul és jelentős előrelépést jelentett az elérhető rekonstrukciós minőség tekintetében.

A gyakorlatban a geometria és a látvány rekonstrukcióján felül szükség lehet a színtérben látható egyes elemek szegmentálása és osztályozása. Például a személyiségi jogok védelme megkívánhatja, hogy a színtérből eltávolítsuk a felvételeken szereplő emberi alakokat, amihez a bemeneti képeken, vagy a rekonstruált 3D jeleneten kell szemantikus szegmentációt végezni, majd bizonyos kategóriákhoz tartozó halmazokat eltávolítani. A NeRF által rekonstruált térfogati reprezentáció által az ilyen jellegű feladatokat is könnyebbé válhatnak.

A hallgató feladatai:

- Tanulmányozza a téma szakirodalmát. Ismerkedjen meg a NeRF módszerek alapjaival.
- Alkalmazzon egy NeRF-alapú rekonstrukciós eljárást szimulált és/vagy valós adatokból származó képekből történő 3D rekonstrukcióra.
- Vizsgálja meg a NeRF révén előállt 3D adatok szemantikus szegmentálásának lehetőségeit.
- Vizsgálja meg a térfogati rekonstrukció manuális vagy automatikus szerkesztésének lehetőségeit.



Budapesti Műszaki és Gazdaságtudományi Egyetem
Villamosmérnöki és Informatikai Kar
Irányítástechnika és Informatika Tanszék

Horváth Viktor András

**3D REKONSTRUKCIÓ ÉS
SZEGMENTÁLÁS NEURÁLIS
RADIANCIA MEZŐKKEL**

MSc

KONZULENS

Dr. Vaitkus Márton

BUDAPEST, 2023

Tartalomjegyzék

Összefoglaló	7
Abstract.....	8
1 Bevezetés	9
1.1 3D rekonstrukció.....	9
1.1.1 Aktív megközelítés	9
1.1.2 Passzív megközelítés	9
1.2 Neurális háló alapú rekonstrukció	10
1.3 A dolgozat felépítése	11
2 Elméleti áttekintő, irodalomkutatás.....	12
2.1 Neurális hálózatok	13
2.1.1 Neurális hálózatok felépítése	13
2.1.2 Hiba függvény.....	15
2.1.3 Optimalizáció.....	16
2.1.4 Regularizáció	17
2.2 Radiancia mezők.....	18
2.2.1 Neurális radiancia mezők.....	19
2.2.2 Gauss halmazok	21
2.3 Radiancia mező alapú szegmentáció	23
2.3.1 Neural Semantic Field	24
2.3.2 Instance Neural Radiance Field	24
2.3.3 3D-s Gauss halmazok szegmentálása	25
3 Technológiai háttér	27
3.1 Python	27
3.1.1 PyTorch.....	28
3.1.2 NumPy	29
3.1.3 Open3D.....	29
3.2 CUDA	30
3.3 CARLA.....	30
4 Megvalósítás	32
4.1 Szintetikus adathalmaz	32
4.2 3D jelenet rekonstrukciója	34

4.2.1 Az adathalmaz formázása	35
4.2.2 Tanítás.....	36
4.3 Címkezés a szegmentált képek vetítésével	37
4.3.1 Kovarianciamátrix kiszámítása.....	37
4.3.2 Ellipszoid megjelenítése: Ray casting	38
4.3.3 Kamera pozíció és orientáció.....	40
4.3.4 Sugarak inicializálása maszkolt pixelek irányában	45
4.3.5 Címkezés maszkolt kép pixelein keresztül indított sugarakkal	46
4.3.6 Többféle címke rendelése egy ellipszoidhoz	48
4.3.7 Szobor eltávolítása.....	49
4.4 Sugárkövetés optimalizálása.....	50
4.4.1 Képek számának csökkentése	50
4.4.2 Átlátszóság figyelembevétele	50
4.4.3 Képek nagy felbontásának kezelése	51
4.4.4 Ellipszoidok szűrése a kamera síkjával	51
4.4.5 Sugár körüli szűrés.....	52
4.5 Szegmentáció utófeldolgozása.....	53
4.5.1 Smoothing.....	53
4.5.2 Zajsűrés	55
4.5.3 Field alapú felosztás.....	55
5 Eredmények.....	58
5.1 3D rekonstrukció.....	58
5.2 3D szegmentáció.....	60
6 A közúti forgalomban alkalmazott rekonstrukció jogi és etikai aspektusai	64
6.1 Fogalmak	64
6.1.1 Alapvető jog.....	64
6.1.2 Személyes, érzékeny és bizalmas adat.....	65
6.1.3 Résztvevők.....	66
6.1.4 AI rendszerek kockázati besorolása.....	67
6.2 Morális és jogi aggályok.....	68
6.2.1 Adatvédelem megsértése	68
6.2.2 Hozzájárulás.....	68
6.2.3 Környezeti hatások	69
6.2.4 Megfigyelés és nyomon követés.....	69

6.2.5 Diszkrimináció és elfogultság.....	70
6.3 Ellenintézkedések	70
6.3.1 Rugalmas hozzájárulás	70
6.3.2 Szegmentáció	70
6.3.3 Energia csökkentés	71
7 Összefoglaló	73
7.1 Továbbfejlesztési lehetőségek	73
Irodalomjegyzék.....	75
Függelék.....	77

HALLGATÓI NYILATKOZAT

Alulírott **Horváth ViktorAndrás**, szigorló hallgató kijelentem, hogy ezt a diplomatervet meg nem engedett segítség nélkül, saját magam készítettem, csak a megadott forrásokat (szakirodalom, eszközök stb.) használtam fel. Minden olyan részt, melyet szó szerint, vagy azonos értelemben, de átfogalmazva más forrásból átvettem, egyértelműen, a forrás megadásával megjelöltem.

Hozzájárulok, hogy a jelen munkám alapadatait (szerző(k), cím, angol és magyar nyelvű tartalmi kivonat, készítés éve, konzulens(ek) neve) a BME VIK nyilvánosan hozzáférhető elektronikus formában, a munka teljes szövegét pedig az egyetem belső hálózatán keresztül (vagy hitelesített felhasználók számára) közzétegye. Kijelentem, hogy a benyújtott munka és annak elektronikus verziója megegyezik. Dékáni engedéllyel titkosított diplomatervek esetén a dolgozat szövege csak 3 év eltelte után válik hozzáférhetővé.

Kelt: Budapest, 2024. 01. 15.

.....
Horváth Viktor András

Összefoglaló

Az utóbbi évtizedekben végbement technológiai előrelépéseknek és az egyre inkább bővülő minőségi adatbázisoknak köszönhetően a gépi tanulás új lehetőségek tárházát nyitotta meg számos eddig megoldatlan kihívás abszolválására. A mélytanulás és a számítógépes grafika területén megjelent algoritmusok lehetővé tették egy újfajta megközelítést az alakzatrekonstrukciónak, amely szerint a korábbi módszerekkel ellentétben a teret egy sűrű, folytonos vagy diszkrét színes közegeként kell értelmezni. Ezt nevezik radiancia mezőnek.

A kutatásom célja a radiancia mező alapú rekonstrukciók szegmentációjának a vizsgálata. Elmélyedtem a téma szakirodalmában, ismerkedtem és összehasonlítottam különböző radiancia mezőre épülő rekonstrukciós eljárásokat, melyeket a szegmentáció szempontjából is megvizsgáltam, valamint implementáltam saját megoldást a problémára. Ehhez egy szimulált környezetben, egy körforgalmat körbejáró önvezető jármű kamerája által előállított képei alapján rekonstruáltam a háromdimenziós jelenetet, majd egy olyan eljárást fejlesztettem, amely a jelenet szegmentációjára alkalmas.

Továbbá a mesterséges intelligenciára épülő alkalmazásoknak meg kell felelniük az egyén egészségére, biztonságára és alapvető jogaira jelentett kockázati szintjük alapján megfogalmazott előírásoknak. A diplomamunkám során megvizsgáltam a közúti forgalomban alkalmazott rekonstrukciót jogi és etikai aspektusokból, valamint a szegmentáció gyakorlati felhasználását a különböző jogi előírások betartására.

A dolgozatomban röviden ismertetem a neurális hálózatok működésének elvét, különböző neurális radiancia mező alapú módszereket és részletezem az általam felhasznált adathalmazt, az adathalmaz rekonstrukciójának, illetve szegmentációjának feladatára implementált megoldásom környezetét és fontosabb lépéseit. Végül bemutatom és értékelem az elért eredményeket.

Abstract

Thanks to the technological advances that have taken place in recent decades and the ever-expanding quality databases, machine learning has opened up a wealth of new possibilities for solving many hitherto unsolved challenges. Algorithms published in the field of deep learning and computer graphics enabled a new kind of approach to shape reconstruction, according to which, in contrast to previous methods, the space must be interpreted as a dense, continuous or discrete colored medium. This is called the radiance field.

My research aims to examine the segmentation of radiance field-based reconstructions. During my diploma work, I delved into the literature on the topic, got to know and compared different reconstruction procedures based on the radiance field, which I also examined from the point of view of segmentation, and implemented my solution to the problem. To do this, I reconstructed the three-dimensional scene based on the images produced by the camera of a self-driving vehicle going around a roundabout in a simulated environment and then developed a procedure that is suitable for scene segmentation.

Furthermore, applications based on artificial intelligence must comply with the regulations based on their level of risk to the health, safety and fundamental rights of the individual. During my thesis, I examined the reconstruction applied in road traffic from legal and ethical aspects, as well as the practical use of segmentation to enforce various legal regulations.

In my thesis, I briefly describe the principle of the operation of neural networks, and different neural radiance field-based methods and detail the data set I used, the environment, and the most important steps of my solution implemented for the task of data set reconstruction and segmentation. Finally, I present and evaluate the achieved results.

1 Bevezetés

A napjainkban megfigyelhető 3D-modellezés és animáció iránti növekvő kereslet, különösen a filmiparban, számítógépes játékokban és az egyre inkább elterjedt virtuális valóság-alkalmazásokban, amely erősen hozzájárul a háromdimenziós rekonstrukció technológiai fejlődéséhez. A szórakoztatóipar mellett az alakzatrekonstrukció térnyerése az egészségügyi szektorban is növekszik, különösen a sebészeti tervezés, diagnosztika és személyre szabott kezelések terén. [1]

1.1 3D rekonstrukció

A háromdimenziós rekonstrukció mindig is egy fontos és nehezen teljesíthető cél volt. Az eljárás segítségével meghatározható bármilyen objektum háromdimenziós profilja, illetve a profilt alkotó ponthalmaz koordinátái. Elmondható, hogy az alakzatok térbeli rekonstrukciója számos tudományterületet foglalkoztató probléma, ilyen például a számítógépes grafika, számítógépes animáció, számítógépes látás, orvosi képalkotás, virtuális valóság és digitális média. Az eljárás kutatása és fejlesztése napjainkban nagy jelentőséggel bír.

A számítógépes látás és számítógépes grafika területén a 3D rekonstrukció az a folyamat, amely valós objektumok alakjának és megjelenésének megtalálásával foglalkozik. A folyamatot megvalósító módszereket a működésük szerint két csoportba soroljuk: aktív és passzív. [2]

1.1.1 Aktív megközelítés

Az aktív megközelítés lényege, hogy a mélységi térkép ismeretében numerikus közelítéssel végzik az alakzatok rekonstruálását. Ezek a módszerek aktívan beavatkoznak a rekonstruálás folyamatába mechanikusan vagy radiometriailag távolságmérők segítségével, hogy megszerezzék a mélységi térképet, például strukturált fény, lézeres távolságmérő vagy egyéb aktív érzékelési technika alkalmazásával.

1.1.2 Passzív megközelítés

A 3D-s rekonstrukció passzív módszerei az aktívakkal ellentétben nem nyúlnak az objektumhoz, csak az objektum felületéről visszavert vagy kibocsájtott sugárzás mérésére használnak egy szenzort, ami alapján a tárgy térbeli szerkezetére tudnak

következtetni. Ez a szenzor tipikusan egy kamera, amely az érzékelő alapján a beérkező fotonokat különböző intenzitású pixelekre képezi le. Az ilyen esetekben kép alapú rekonstrukcióról beszélünk, ahol az eljárás bemeneti adata a jelentről készített videó vagy képek sorozata és a kimenet a háromdimenziós modell.



1.1. ábra 2D kép alapú alakzatrekonstrukció. [3]

Az aktív megközelítés az egyszerűségéből adódóan a passzív megközelítéshez képest rugalmasabb és a gyakorlatban több esetben alkalmazható.

1.2 Neurális háló alapú rekonstrukció

Utóbbi években bekövetkezett technológiai fejlődések lehetővé tették a különböző eszközök számítási kapacitásának megnövelését. Ennek köszönhetően ismét teret nyernek a magas számításigényű gépi tanuló algoritmusok melyek egyik speciális esete a neurális háló, melynek lényege, hogy kiértékeli a bemenete és a kimenete közötti összetett, többváltozós matematikai függvényt anélkül, hogy bármilyen előismerete lenne az adatok közötti kapcsolatokról (black-box jellegű). A bemeneteből a neurális hálót alkotó, rétegekbe szervezett neuronok által képzett lineáris kombinációt egy nemlinearitáson vezetik keresztül. Egy ilyen neurális hálózat az univerzális approximáció tétele alapján bármilyen függvényt tetszőlegesen közelíthet. [4]

Neurális hálók nyújtotta lehetőségeket egyre több területen igyekeznek kiaknázni, ilyen terület többek között a 3D rekonstrukció. A mély tanulás alapú megközelítéshez elő kell állítani be- és kimeneti adat párokat, amellyel az adott háló optimalizálja a súlyait annak megfelelően, hogy a lehető legjobb általánosítóképesség mellett a legkisebb hiba keletkezzen a becsült és az elvárt kimenet között.

A háromdimenziós rekonstrukciónak több olyan megközelítése is létezik, melyek alapjául szolgál a neurális hálózat alkalmazása. A be- és kimeneti adat párok megválasztása azon múlik, hogy a feladat során milyen információ adott és milyen módon

vannak reprezentálva a tér háromdimenziós objektumai. Egy lehetséges módszert jelent, ha kétdimenziós színes képek és a hozzájuk tartozó kamera pozíciók alapján áll elő a választott módon reprezentált háromdimenziós tér, melyet a neurális háló a renderelt és az eredeti képek eltéréséből számolt hiba alapján optimalizál.

Fontos kérdés az objektumok reprezentációja, amely lehet implicit, parametrikus vagy például a jelenleg egyik legnépszerűbb kutatási terület témáját képző radiancia mező alapú.

1.3 A dolgozat felépítése

A bevezetést követő fejezetben a dolgozat elméleti háttéréről lesz szó, ebben a fejezetben a témában való irodalomkutatás során érintett, illetve megtanult főbb témakörök kerülnek részletezésre. Ilyenek például a mély tanulás eszközei, a neurális hálók, a radiancia mezők és azok különböző reprezentációi, illetve az azokra épülő eljárások a szegmentálás feladatának elvégzésére.

A harmadik fejezet a vizsgált neurális radiancia alapú alkalmazások és a saját implementáció elkészítéséhez szükséges technológiai háttérrel ismerteti, milyen környezetben, milyen eszközöket és fontosabb könyvtárakat felhasználva készült el.

Az ezt követő fejezetben a megoldáshoz vezető, szükség esetén kódrészletekkel, ábrákkal, képletekkel kiegészített kulcsfontosságú tervezői döntések lesznek tárgyalva.

Az ötödik fejezet bemutatja és értelmezi az elért eredményeket, valamint kiértékeli az eljárás hatékonyságát.

A hatodik fejezet betekintést nyújt a témába etikai és jogi aspektusokból. Megvizsgálja az eljárás kockázati tényezőit és az azok alapján megfogalmazott elvárásokat, majd bemutatja a szegmentáció gyakorlati alkalmazását a személyiségi jogok védelmére.

Végül a dolgozat utolsó fejezete összefoglalja a kutatás során megszerzett tapasztalatokat és kitér a javítási és továbbfejlesztési lehetőségekre.

2 Elméleti áttekintő, irodalomkutatás

Ahogy más területeken, úgy az alakzatrekonstrukció területén is új megközelítési lehetőségeket teremtett a mesterséges intelligencia, pontosabban annak egy alterülete, a gépi tanulás fejlődése és térhódítása. A gépi tanulás célja, hogy bizonyos feladatokra emberi intelligenciával felérő vagy azt meghaladó megoldásokat képes legyen megtalálni.

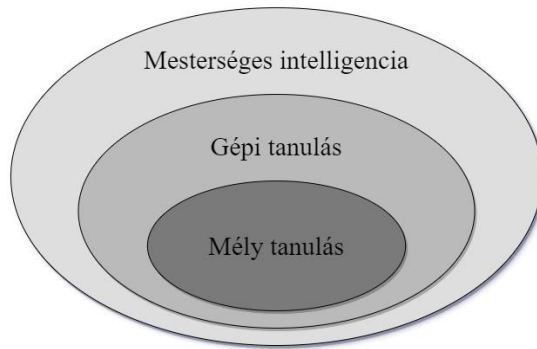
A gépi tanulás olyan strukturált számítógépes algoritmusok tanulmányozása, melyek tapasztalat révén automatikusan javulnak. Ezek az algoritmusok az átadott adat alapján olyan modelleket építenek fel, melyek képesek felfedni az adatok közötti háttérben meghúzódó, rejtett összefüggéseket.

Az eljárás sikeressége, tehát a modell általánosító-képessége erősen függ az adatok minőségétől és mennyiségétől, tehát az adatok előkészítése és előfeldolgozása egy rendkívül kritikus fázis.

A gépi tanulásnak alapvetően három típusa van: [5]

- Felügyelt tanulás: A felügyelt tanulás célja az, hogy az előre definiált bemeneti és kimeneti adatpárok közötti összefüggést leíró függvényt megtalálja vagy közelítse.
- Felügyelet nélküli tanulás: A módszer célja, hogy valamilyen előre definiált hasonlósági metrika alapján klaszterekbe rendezi az adatot úgy, hogy a felügyelt változatával ellentétben, a bemeneti adathoz nincsenek elvárt kimeneti értékek rendelve.
- Megerősítéssel tanulás: Olyan módszer, amelynek a célja cselekvések, a cselekvések által kiváltott jutalmak és megfigyelések sorozata alapján kiismerni a környezetet a jutalom maximalizálásával.

A gépi tanulás egy speciális formája a mély tanulás, melynek eszköze a neurális hálózat.



2.1. ábra A mesterséges intelligencia, a gépi tanulás és a mély tanulás kapcsolata.

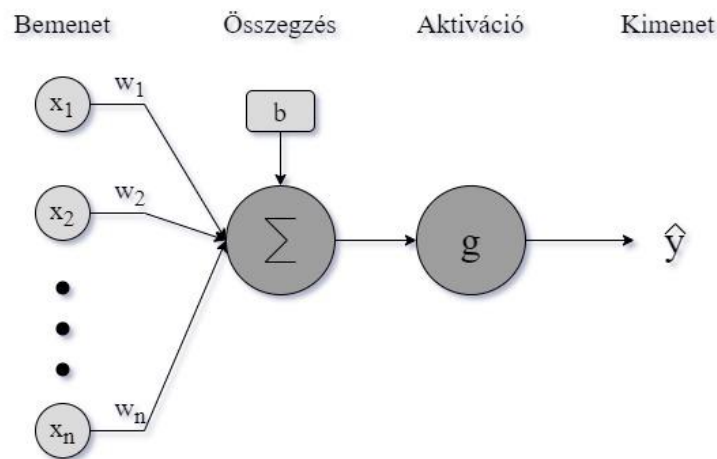
2.1 Neurális hálózatok

A neurális hálózatot struktúrába rendezett gépi tanuló algoritmusok alkotják és a működtetéséhez az adott feladathoz rendelkezésre álló adatokat jellemzően három halmazba szokták sorolni: tanító, validációs és teszt adathalmaz. Mindegyik halmaz be- és a hozzájuk tartozó kimeneti adat párokat tartalmazza.

A neurális hálózatok működtetése két fő fázisból áll: tanítás és tesztelés. A tanítás során a neurális háló megtanulja az adatok közötti rejtett összefüggéseket a tanító és a validációs halmaz segítségével, majd a tesztelési fázis során ellenőrzésképpen kiértékeli a modell teljesítményét még soha nem látott adatokon, a teszt adatokon.

2.1.1 Neurális hálózatok felépítése

A neurális hálózatokat alkotó legfőbb építőelemek, a neuronok a biológiában megfigyelhető idegsejtekhez hasonlóan feldolgozzák és továbbítják a beérkező információt. Míg az idegsejtek dendritekkel és axonnal rendelkeznek, addig a modellezett változatuk be- és kimenetekkel. [6]



2.2. ábra A neuron működése.

A neuron a bemenetén kapott értékekből álló $x = \begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{pmatrix}$ vektor elemeinek veszi a $w = \begin{pmatrix} w_1 \\ w_2 \\ \vdots \\ w_n \end{pmatrix}$ súlyvektorral vett súlyozott összegét, majd hozzáad egy b eltolást. Az így kapott értéket a kezelhetőség érdekében egy intervallumra képezi le a választott $g(\cdot)$ aktivációs függvénnyel, amely nemlinearitást visz bele a számítási folyamatba.

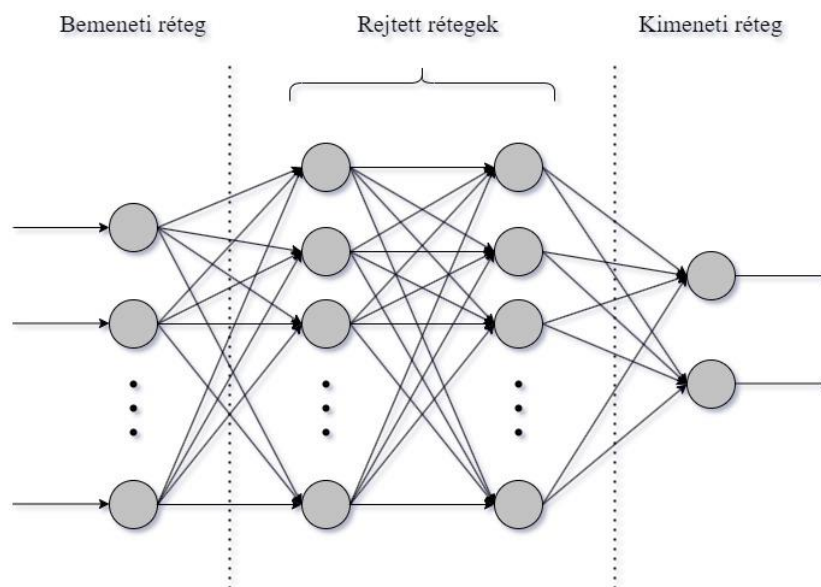
Vektoros formában így írható fel a neuron egy n darab attribútummal rendelkező bemeneti adataira adott válasza:

$$\hat{y} = g(x \cdot w + b)$$

A neuronok, más szóval perceptronok magukban nem elégségesek ahhoz, hogy egy komplex, gyakorlatban is előforduló feladatot megoldjanak, viszont ezek a perceptronok rétegekbe rendezve felhasználhatóak többek között osztályozásra, az adat jellegzetességeinek felismerésére, előrejelzésre vagy approximációra. Ennek a rétegekből felépített struktúrának az elnevezése Multi-Layer Perceptron, ismertebb nevén neurális hálózat.

Az egyes rétegek neuronjai a szomszédos rétegek neuronjaival állhatnak összeköttetésben. Klasszikus értelemben vett MLP jellemzően egy teljesen összekötött hálózat, ahol az egymással szomszédos rétegek minden neuronja összeköttetésben áll egymással. A neurális hálózatok rétegeinek három típusa van: [7]

- **Bemeneti réteg:** Erre érkezik a feldolgozandó adat. Dimenzióját a bemeneti adat és attribútumainak száma határozza meg.
- **Kimeneti réteg:** a kimeneti réteg neuronjainak számát a feladat célja határozza meg. Egy bináris klasszifikáció esetén két neuron alkotja a kimeneti réteget, melyek értéke reprezentálja a minta adott osztályhoz való tartozásának a valószínűségét.
- **Rejtett réteg:** Az előző két réteg között tervezői döntés alapján meghatározott számú rejtett réteg fordulhat elő, melyek a velük szomszédos rétegekkel állnak közvetlen kapcsolatban.



2.3. ábra Két rejtett rétegű bináris klasszifikációt végző MLP architektúrája [8]

Egynél több rejtett réteget tartalmazó neurális hálózatokra mély neurális hálózatként hivatkoznak a szakirodalomban. A rejtett rétegek száma egy hiperparaméter, melyet a feladathoz, illetve annak komplexitásához érdemes igazítani.

2.1.2 Hiba függvény

A neurális hálózat x_i bemeneti adatokra adott válaszát valamilyen módon értékelni kell. Erre az értékelésre szolgál a költség függvény, más szóval hiba függvény, melynek bemenete az $f(x_i)$ hálózat által generált kimenet és az y_i elvárt kimenet.

Az egyik leggyakrabban használt hiba függvény az átlagos négyzetes hiba, rövidítve MSE (Mean Squared Error) kiszámítja a becült és az elvart kimenetek közötti négyzetes eltérések átlagát. Ezzel a módszerrel a nagyobb eltéréseket jobban bünteti, mint a kisebb hibákat:

$$MSE = \frac{\sum_i^n (f(x_i) - y_i)^2}{n - 1}$$

Az MSE kétdimenziós formában végigiterálhat képek pixelein és így elvégezhető az összehasonlítás.

A neurális háló alapú háromdimenziós rekonstrukció helyességének kiértékelése történhet egy teszt adathalmazból kikerülő kép és a hozzá tartozó kamerapozícióból a térbeli rekonstrukció alapján renderelt kép összehasonlításával. A rekonstruált vagy más módon előállított képek minőségének becslése elvégezhető különböző, erre a célra kifejlesztett metrikával.

Egyik ilyen metrika az MSE-re épülő jel-zaj arány, rövidebb nevén PSNR (Peak Signal-to-Noise Ratio). A PSNR a referencia kép maximumához (MAX_I^2) viszonyítja a hibát logaritmikus skálát alkalmazva.

$$PSNR = 10 \cdot \log_{10} \left(\frac{MAX_I^2}{MSE} \right)$$

Egy másik említésre méltó gyakran alkalmazott, képminőség jellemzésére használt metrika a strukturális hasonlóság, rövidítve SSIM (Structural Similarity) az összehasonlítás során figyelembe veszi súlyozott módon a kép luminanciáját, kontrasztját, valamint a struktúráját.

2.1.3 Optimalizáció

A hiba függvény a nevéből adódóan a neurális háló kimenetének a hibáját méri, tehát minél kisebb ez az érték annál jobban teljesít a hálózat. [8]

$$\min_{\theta} L(f_{\theta}(x), y)$$

Az optimalizáció célja a több rétegnek köszönhetően nagyon bonyolult, θ paraméterek alapján számolt tetszőlegesen megválasztott L veszteségfüggvény minimalizálása, melyet a súlyok finomhangolásával tehet meg.

A súlyok frissítésére már számos algoritmus létezik, melyek alapját a gradiens módszer képezi. A gradiens kiszámítható az L költségfüggvény W súlymátrixszal vett parciális deriváltja segítségével. A derivált a meredekséget mutatja meg adott pontban, és ebben a kontextusban azt jelenti, hogy milyen módon kell változtatni a súlyokat, hogy a hiba a lehető legnagyobb mértékben növekedjen. Mivel a hiba minimalizálása a cél, a gradienssel ellentétes irányban kell elmozdulni egy előre definiált η lépésközzel vagy más kifejezéssel tanulási rátával.

$$\Delta w_{ij} = \left(\eta \times -\frac{\partial L}{\partial w_{ij}} \right)$$

A gradiens módszernek ma már számos továbbfejlesztett variánsa van, melyek képesek adaptív működésre is. Fontos megjegyezni, hogy az optimalizációs technikák között akadnak népszerűbbek, de nincsen egy globálisan legjobban teljesítő algoritmus. Általában feladatfüggő, hogy melyik teljesít a legjobban.

2.1.4 Regularizáció

A neurális hálózatok feladata, hogy a tanítóhalmaz adatai alapján tanult összefüggéseket új, még soha nem látott adatokon is fel tudja használni. Viszont a neurális hálózatok különösen hajlamosak túlilleszkedésre. A túlilleszkedés, más szóval túltanulás egy gyakran előforduló jelenség, amely akkor lép fel, amikor a hálózat túlságosan rátanul a tanító adathalmaz karakterisztikájára. A jelenség megfigyelhető, amikor a hálózat a tanítóhalmazon mért pontossága javul ennek ellenére a tesztelésre szánt adatokon egyre rosszabb eredményeket ér el. [8]

A túltanulás ellen alapvetően a hálózat komplexitásának csökkentésével lehet védekezni, melyre számos bevált technika létezik:

- **Architektúra egyszerűsítése:** A hálózat paramétereinek számának csökkentése, rétegek elhagyásával vagy méretük csökkentésével.
- **Korai leállítás:** A hálózat tanításának megállítása azon a ponton, amikor a prediktív teljesítmény egy dedikált validációs halmazon mérve már nem nő tovább.
- **Zaj hozzáadása:** A bemeneti adatokhoz való zaj hozzáadása bizonyos esetekben javítja az általánosítóképességet.

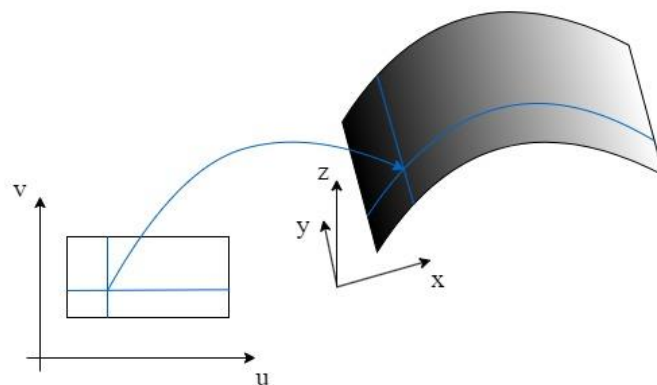
- Dropout: A dropout eljárás során a réteg véletlenszerűen kiválasztott neuronjai figyelmen kívül maradnak. A dropout hatása olyan, mintha több kisebb hálózatot tanítottak volna és azoknak vennék az aggregát becslését.
- Súlycsökkentés: A neurális hálózat költségfüggvénye figyelembe veszi a súlyok értékét, így azokat is minimalizálja.

2.2 Radiancia mezők

A háromdimenziós alakzatrekonstrukciónak alapvető kérdése, hogy milyen módon lehet leírni a térbeli alakzatokat. A neurális háló alapú rekonstrukciós eljárások esetén is többféle lehetőség létezik az alakzatokat reprezentálására.

Egyik ilyen megközelítés, az implicit leírási mód, melynek lényege, hogy a rekonstruálni kívánt objektum felületét egy háromváltozós térbeli skalárfüggvény szintfelülete reprezentálja. Ekkor a neurális háló bemenete egy ponthoz tartozó koordináta, a kimenete pedig az adott ponthoz tartozó távolság értéke. Az előnye, hogy könnyen alkalmazható és rugalmas, viszont a szintér által tartalmazott textúrákat és megvilágítást nehéz kinyerni.

Egy másik megoldást jelenthet a számítógéppel segített geometriai tervezésben leggyakrabban használt megközelítés, a paraméteres leírás. Ebben az esetben a neurális háló bemenete egy síkbeli (u, v) koordináta pár, kimenete a pontnak megfelelő háromdimenziós pozíció. Az implicit leírással ellentétben a színek és anyagjellemzők könnyen rekonstruálhatóak. [9]

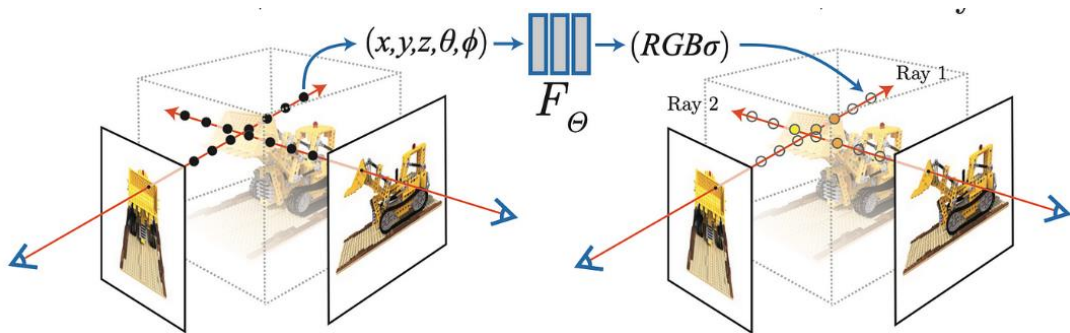


2.4. ábra Az ábra a paraméteres leírást szemlélteti.

Egy harmadik megoldás alapja, hogy a teret egy olyan közeg tölti ki, ami a tér jellegétől függően elnyeli és megtöri a rajta áthaladó fényt. Ezt a reprezentációt illetik radiancia mező néven, melyben a radiancia szó arra utal, hogy a pontok kibocsájtanak egy színt egy adott irányban és a mező szó arra, hogy a tér felosztható ilyen pontokra. Ekkor a kamera pixeleibe érkező fény színét és intenzitását a közegen áthaladó fénysugár által metszett pontok hatásának integrálásával kapjuk. Ez a számítógépes grafikában is ismert ray casting eljárás. Rekonstrukció esetében ez a folyamat fordítva játszódik le:

- bemenet: egy 5D koordináta, amely tartalmazza a térbeli pont (x, y, z) koordinátáját és a (θ, ϕ) nézeti irányt
- kimenet: az RGB szín és σ sűrűség.

Az összerendelést egy neurális hálózat is megvalósíthatja, ez az úgynevezett NeRF, teljes nevén Neural Radiance Field, ami jelenleg az egyik legnépszerűbb kutatási téma a neurális hálózatok és az alakzatrekonstrukció területén.



2.5. ábra NeRF által megvalósított fordított ray casting. [10]

2.2.1 Neurális radiancia mezők

Ahogy korábban kifejtettem, a 3D rekonstrukciós problémára számos megoldás létezik, viszont egy komoly áttörést jelentett a NeRF módszer, mely nem foglalkozik az alakzatok felületének explicit leírásával, hanem a teret egy színes, sűrű, világító, közegként kezeli, ez a radiancia mező. Ez a megközelítés rendkívül rugalmas, a korábbi módszerekkel ellentétben képes megtanulni a nézőpont függő megvilágítást, anélkül, hogy a fényforrásról ismereteink lennének.

A számítógépes grafikában jól ismert ray casting módszert felhasználva a NeRF a jelenetről készített bemeneti képek interpolációjával rendereli a teljes jelenetet. A kimenet egy olyan térfogat, amelynek színe és sűrűsége a nézet irányától és a kibocsátott fény sugárzástól függ. Minden sugárhoz kapunk egy kimeneti térfogatot, és ezek a

térfogatok alkotják az összetett jelenetet. Ebben a formában a statikus jelenetet egy folytonos 5D-s függvény írja le, amelyet egy teljesen összekapcsolt neurális hálóval közelítünk úgy, hogy az 5D koordinátából visszafelé haladva következtet a térfogatsűrűségre és színre. [10]



2.6. ábra NeRF módszer folyamatábrája. [10]

A teljesen összekötött MLP (Multi-Layer Perceptron) alapú módszer a gradiens süllyesztés segítségével minimalizálja a teszt és renderelt képek között számított veszteségfüggvényt.

Egy primitív objektumot reprezentáló függvényt, mely helyettesíthető egy neurális hálóval neurális grafikus primitívnek nevezünk. Jellemzően egy pozíciót és mellette olyan plusz paramétereket kap bemenetként, amelyek az adott objektum leírásához szükségesek. Kimenete pedig a grafikus primitívet jellemző tulajdonságok az adott pozícióban. Máshogy fogalmazva a neurális grafikus primitív egy olyan objektum, amelynek alakját és megjelenését leíró attribútumai átadhatóak egy neurális hálónak. Ilyen objektum például a NeRF is, melyre az Instant-ngp épül. [11]

Az Instant-ngp egyik legnagyobb előnye, hogy a renderelő és tanító algoritmusai optimalizáltak, feladatspecifikus GPU implementációk segítségével akár százszor gyorsabb számításra képes, mint a naiv tenzor alapú megközelítések. A hatékonyabb renderelés a sugár mentén nem mindig azonos lépésközzel mintavételez. Az objektum elmetszéséig átugorja az üres teret és az objektum után sem végez sűrű mintavételezést, ezáltal sokkal kevesebb lépést kell végrehajtani.

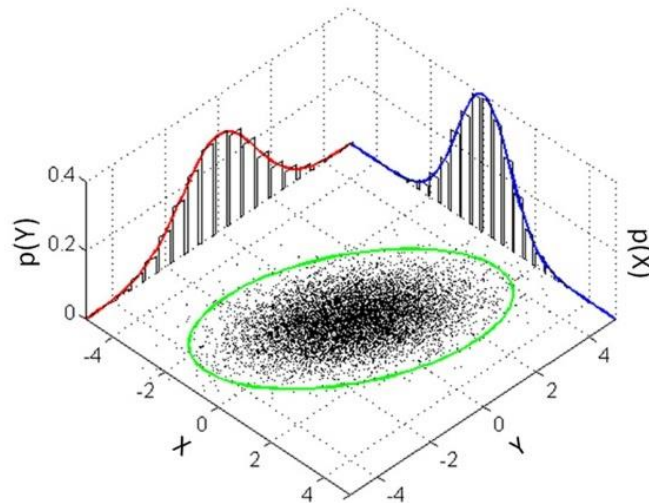
Alapvetően egy kevés neuron alkotta teljesen összekötött neurális hálóról van szó, melyet kis mérete miatt gyorsabban lehet tanítani. Fejlesztői a hangsúlyt a háló komplexitásának növelése helyett az optimális bemenet enkódolásra helyezték. Speciális térfogat-hashelésnek köszönhetően a memória-kihasználtság mellett a sebesség és a minőség is javul.

A klasszikus MLP alapú NeRF egy rendkívül innovatív megoldás az alakzatrekonstrukcióra, viszont az architektúrájából adódóan nagyon lassú a betanítása és más jelenetekre nem alkalmazható. Az Instant-NGP ehhez képest egy sokkal gyorsabb és rugalmasabb megoldást nyújtott, melyen kívül még számos másik NeRF alapú eljárás született a közelmúltban.

2.2.2 Gauss halmazok

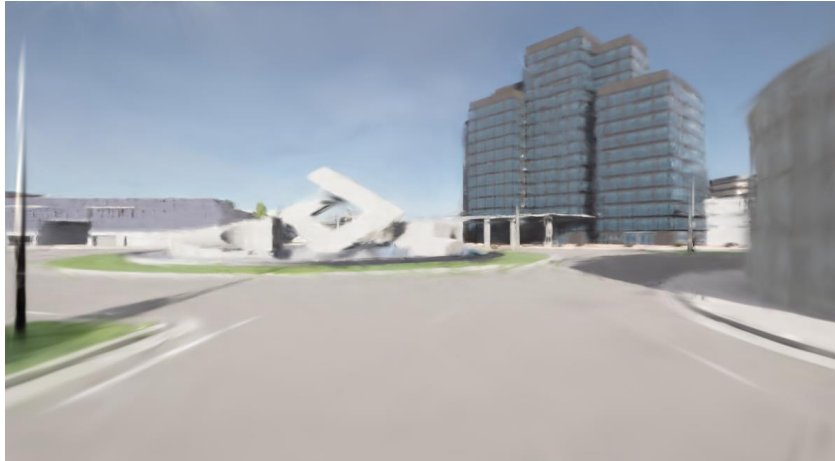
A radiancia mező alapú módszerek forradalmasították a fotósorozattal vagy videóval rögzített jelenetek rekonstrukcióját. Ezekkel a módszerekkel a jelenetet új perspektívából is meg lehet figyelni a megvilágítást és árnyékokat megőrizve. Ugyanakkor a minőség megőrzéséhez szükséges neurális hálózatok tanítása és renderelése rendkívül időigényes, így meg kell találni a minőség és a sebesség közötti arany középutat.

A radiancia mező a teret, a valósághoz hasonlóan részecske szinten közelíti meg, színes sűrű közegként. A radiancia mező reprezentálható egy neurális hálózat segítségével, ez a korábbi fejezetben részletezett NeRF. Viszont egy másik megközelítés szerint ezeknek a részecskének lehet kiterjedése is. Ezek a kiterjedéssel rendelkező részecskék a Gauss-halmazok, melyek geometriája megfelel az ellipszoidéhoz.



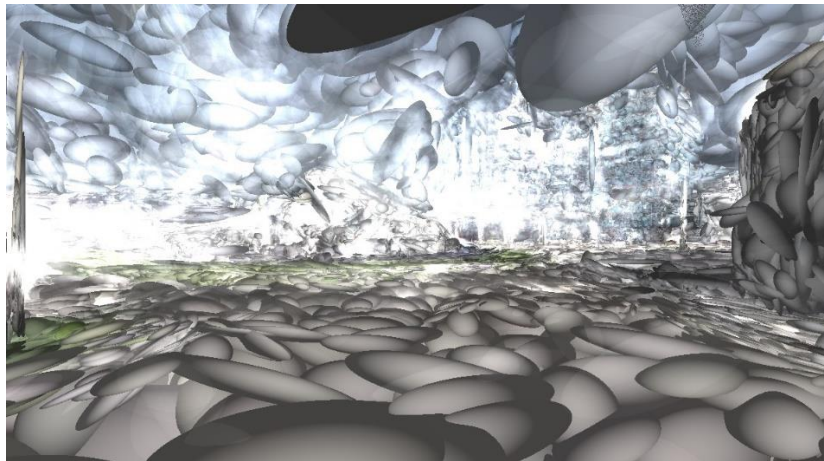
2.7. ábra Két 1D-s Gauss eloszlás feszíti ki a 2D-s Gauss halmazt. [12]

Az ellipszoidokat Gauss eloszlások feszítik ki, értelemszerűen 3 egymásra merőleges tengely mentén. Hasonlóan rendelkeznek a szükséges attribútumokkal, ilyen a nézőpont függő szín és áttetszőség. Ezt a koncepciót feldolgozó eljárás a Gaussian Splatting.



2.8. ábra Egy körforgalom rekonstrukciójáról renderelt kép.

A Gaussian Splatting úgy tekinti, hogy a teret különböző Gauss eloszlások által definiált ellipszoidok alkotják.

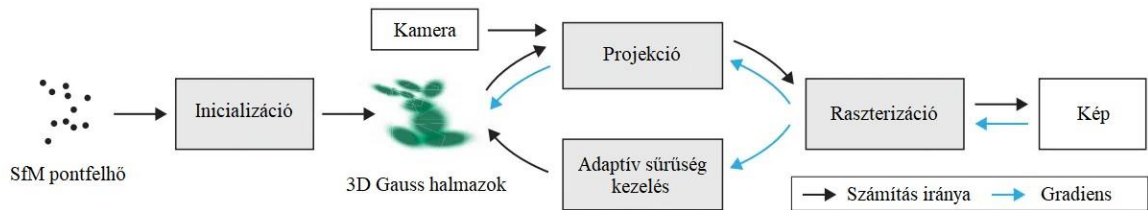


2.9. ábra A rekonstruált körforgalom megjelenítése ellipszoidokkal.

Első lépésként COLMAP könyvtár segítségével az egymást követő képeken előforduló jellemzőket detektál és azok alapján felépít egy háromdimenziós pontfelhőt. Ezzel a pontfelhővel inicializálja a jelenetben található 3D-s Gauss halmazokat.

Ezt követően PyTorch-ban implementált gépi tanulást alkalmazó optimalizációt hajt végre a modell. Az egyes ellipszoidok optimalizált paraméterei:

- a 3D pozíciója,
- az alakját és orientációját leíró kovariancia mátrix,
- a színe,
- és áttetszősége.



2.10. ábra Gauss Splatting algoritmus architektúrája. [13]

Az optimalizáció az adathalmaz egy képe és a hozzá tartozó kamera pozícióból renderelt kép eltéréséből számolt hiba alapján történik. A Gaussian Splatting egyik nagy előnye, hogy a rendereléshez nem használ neurális hálózatot, hanem egy megfelelően kioptimalizált raszterizációt hajt végre a jeleneten. A gyors sebességében közrejátszik az is, hogy a tér üres területeire nem helyez háromdimenziós Gauss halmazokat, ezáltal sokkal ritkább a radiancia mezője, mint a folytonos NeRF-nek.

2.3 Radiancia mező alapú szegmentáció

A gépi látás az a terület, amely lehetővé teszi, hogy a számítógépek képesek legyenek feldolgozni és értelmezni a vizuális információt, például színes, RGB képek pixeleit. Napjainkban az internet forgalmának jelentős részét képek és videók teszik ki, emiatt mondani sem kell mekkora jelentőséggel bír a számítógépes látás. A terület három fő feladata az osztályozás, a detektálás és a szegmentáció.

Az osztályozás címkéket rendel a képekhez vagy a képek egyes régióijához, a detektálás lokalizálja a keresett objektumot és a szegmentálás minden pixelhez szemantikus címkét rendelve felosztja a képet régiókra. A dolgozat témája az utóbbi, így a továbbiakban a szegmentáció kapja a fókuszot.

A szegmentálás elvégezhető térben is, ekkor alap esetben a tér pontjaihoz lesz rendelve egy szemantikus címke. A radiancia mező alapú megközelítések is alkalmasak a tér szegmentálására, azzal a módosítással, hogy nem egy nézeti iránytól függő szint ad vissza, hanem a ponthoz tartozó szegmens színét áttetszőség nélkül. A szemantikusan szegmentált térből ezután könnyen eltávolíthatóak a feladat szempontjából nem releváns, zavaró szegmensek, részletek.

A neurális radiancia mezőkön végzett szegmentálás egy aktív kutatási terület, ahol folyamatosan jelennek meg új eredmények. Az igen nagy kihívást jelentő feladat

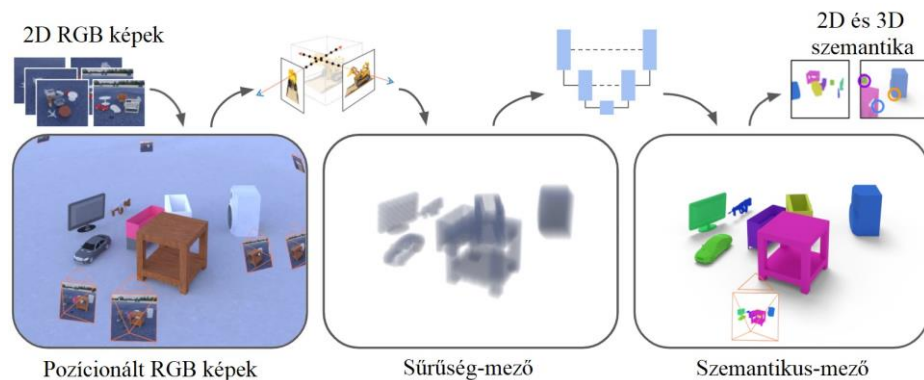
sikerességét meghatározhatja a felcímkezett adathalmaz, a szegmentációt végző hálózat struktúrája vagy a jelenet komplexitása.

Egyik lehetséges módja a szegmentálásnak, amikor előre felcímkezett adathalmazon történik a neurális háló tanítása. Ez magával vonja a háló optimalizálását a becsült és a tényleges címkék eltéréséből számolt hiba szerint. A szegmentáció történhet a kapott radiancia mezők sűrűség és szín értékei vagy az abból renderelt képeken végzett szegmentációból visszaszámolt pozíciók alapján is, de ezek mellett még számos módja van.

2.3.1 Neural Semantic Field

A NeSF (Neural Semantic Fields) megközelítés egyszerű, beállított RGB képeket felhasználva állítja elő a 3D szemantikus mezőt, amely a tér minden pontjához hozzárendel egy kategóriát jelölő címkét. [14]

A képeket az azokból például a COLMAP segítségével kinyert kameraadatoknak megfelelően helyezik el a jelenetben. Ez alapján a NeRF előállítja a jelenet geometriájának térfogati reprezentációját, ami a tér háromdimenziós pontjaihoz egy sűrűség értéket rendel (sűrűség mező). Majd egy szemantikus következtető modell a 3D pontokhoz valószínűségi eloszlást rendel, ami a pontnak adott osztályba való tartozásának a valószínűségét takarja, ez a szemantikus mező. A szemantikus mező a tér minden pontjára értelmezve van, és a sűrűség mezővel kombinálva tetszőleges kamera pozícióból renderelhető szegmentáció RGB információ nélkül is.



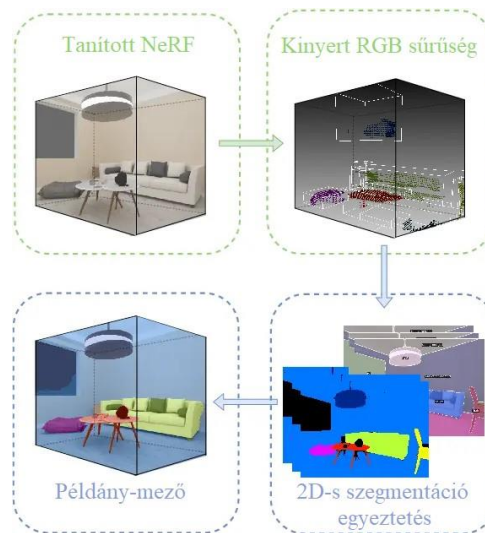
2.11. ábra NeSF működési elve.

2.3.2 Instance Neural Radiance Field

A 3D-s példány szegmentálás problémájára egy új lehetőséget kínál a 2D-s képekből történő 3D-s radiancia mező előállítása. A 2D-s képeken történő műveletek

hatékonyságával szemben a 3D-s szegmentálás sikeressége erősen függ az adatok minőségétől és mennyiségétől. A korábbi megoldásokkal ellentétben, melyek mélységi értéket is mérő szenzorok által előállított adatokon dolgoztak az Instance NeRF egy olyan megoldást kínál, amelynek nincs szüksége mélységi, illetve szegmentációs információra. [15]

Az Instance NeRF a bemeneteként egy betanított NeRF modellt kap. Ebből a NeRF modellből a NeRF-RCNN neurális háló segítségével előállítja a diszkrét 3D maszkolt jelenetet, majd ennek a 3D jelenetnek a projekcióját hasonlítja össze a NeRF modellből renderelt, ugyanolyan beállítású 2D képen készített maszkkal. Az így előállított szegmentációs maszkkal felügyeli a példány mező tanítását.



2.12. ábra Instance-NeRF működési elve.

2.3.3 3D-s Gauss halmazok szegmentálása

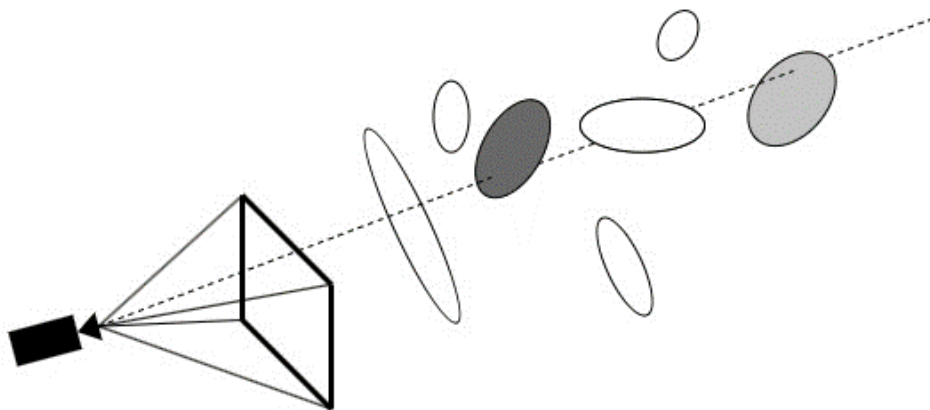
A szemantikus szegmentálás elvégezhető a háromdimenziós Gauss halmazok által reprezentált radiancia mezőn is. Ekkor nem a tér pontjaihoz vannak hozzárendelve a szegmenseket jelölő címkék, hanem a jelenetet alkotó ellipszoidokhoz. Az egyes szegmensekhez egy egyediséget biztosító RGB színt lehet rendelni, amely megfelel a szegmenst jelölő címkének. Majd az ellipszoidokat a címkéjüknek megfelelő színnel beszínezve vizualizálható a már szegmentált tér. A Gaussian Splatting felhasználása mellett a feladat elvégzésére több lehetőség is van.

Egy lehetséges megközelítés, ha a jelenetről készített eredetileg RGB képsorozatok szemantikusan szegmentált változataikkal történik a Gaussian Splatting tanítása. Ebben az esetben a rekonstrukció eredménye egy olyan színes térfogat, melyben

az ellipszoidok színe hasonló, de nem feltétlenül azonos az egyes szegmensek egyedi színeivel. Az egyes ellipszoidok színe több nézőpontból is kiértékelt színek átlaga alapján lesz meghatározva, majd a Voronoi diagram alapján felosztott RGB térben a legközelebbi szomszéd algoritmus megkeresi az adott színhez tartozó szegmens színt.

Egy másik módszer, ha a jelenetről készült RGB képsorozatok alapján rekonstruálja a teret a Gaussian Splatting, majd az RGB képek szemantikusan szegmentált változataikkal történik a jelenetet alkotó ellipszoidok címkézése. Ez elvégezhető a szegmentált képek jelenetre való projekciójával:

1. Sugarak indítása az adott képet készítő kamera pozícióból a kép pixeleibe.
2. Ray casting algoritmus alapján legközelebbi metszéspont kiszámítása.
3. A legközelebbi metszésponthoz tartozó ellipszoid címkézése a szemantikusan szegmentált kép megfelelő pixele alapján.



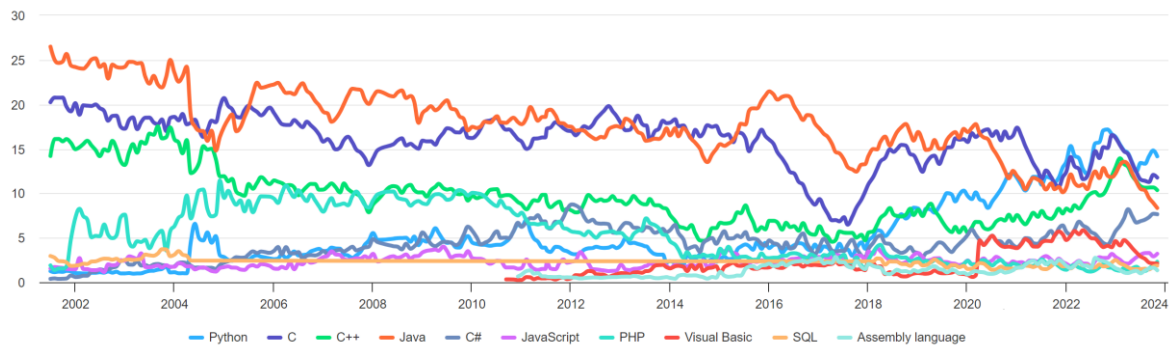
2.13. ábra Ray casting az ellipszoidokon.

3 Technológiai háttér

Ezen fejezet a dolgozat megoldásához szükséges technológiai háttérét ismerteti. Kitér az implementáció programozási nyelvére, röviden bemutatja a leggyakrabban használt könyvtárakat és azok sajátosságait.

3.1 Python

A Python egy több, mint 30 éves nyelv, amely az évtizedek alatt egyre nagyobb népszerűségnek örvend. A mély tanulás témakörben a legnagyobb áttörés 2012-ben történt, amikor az ImageNet verseny legjobb eredményét egy mély neurális háló produkálta.



3.1. ábra Programozási nyelvek népszerűsége TIOBE index alapján. [16]

A gépi tanulás, illetve az adattudomány területén a Python a leggyakrabban használt programozási nyelv. Ezeken a területeken elkerülhetetlen komplex algoritmusok fejlesztése, használata és a Python egyszerűen olvasható és könnyen értelmezhető szintaxisának köszönhetően a programozó nagyobb figyelmet fordíthat rájuk. [17]

Ezekon a területeken gyakori feladat az adatok gyűjtése, előfeldolgozása, az adatok manipulálása, mély neurális hálózatok tanítása és kiértékelése, amelyekhez a Python rengeteg könyvtárat, számos adatstruktúrát és algoritmust biztosít.

A legnépszerűbb gépi tanulást segítő keretrendszerek:

- PyTorch
- TensorFlow
- Keras

- Theano

3.1.1 PyTorch

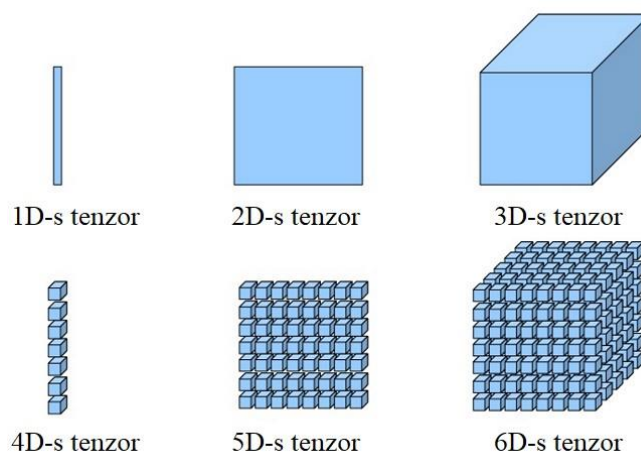
Az előző fejezetben felsorolt mély tanulást segítő keretrendszerek között nincs egy globálisan legjobb, általában feladattól függ, hogy melyik teljesít jobban. A NeRF alapú rekonstrukció területén való kutatás során PyTorch implementáció fordult elő a legnagyobb számban, ami nem véletlen. [18]

A PyTorch a részletes, könnyen követhető dokumentációja és oktatóanyagai miatt könnyen tanulható, melyet az aktív, segítőkész fórumok is segítenek.

Az egyik legnagyobb előnye, hogy jól strukturáltságának köszönhetően nagyobb rugalmasságot és magasabb szabadságfokot biztosít a fejlesztői számára.

Ezen felül a PyTorch támogatja a CPU és GPU használatát a számítási folyamatok párhuzamosítására. A radiancia alapú rekonstrukciós eljárásoknak nagyon magas számításigénye a valós idejű futáshoz szinte kötelezővé teszi az egyszerre több GPU alkalmazását. A számítási folyamatok ily módon való párhuzamosítása felgyorsítja a renderelés és tanítás idejét.

Tipikus adatstruktúrája a tenzor, ami hasonlóan a többdimenziós tömbökhöz képes a modell be- és kimeneti adatait tárolni, valamint különböző műveleteket végrehajtani rajtuk.



3.2. ábra PyTorch tenzorok. [19]

PyTorch kialakításának köszönhetően zökkenőmentesen integrálható a Pythonnal és népszerű könyvtáraival, mint például a NumPy.

3.1.2 NumPy

A tudományos számítási feladatok elvégzéséhez a NumPy egy alapvető, széles körben használt Python csomag. A NumPy lehetővé tesz több dimenziós tömbökön végzett matematikai vagy logikai műveleteket, a tömbök alakjának manipulálását, rendező algoritmusok végrehajtását, különböző adatbázisban használt műveleteket, tömbök fájlból történő beolvasását vagy kimentését. [20]

A Numpy alap objektuma az úgynevezett ndarray, amely a létrehozásakor fix mérettel rendelkezik és annak módosításakor egy új ndarray objektum jön létre és a régi törlődik.

Hasonlóak, mint a PyTorch tenzorok, azzal a fő különbséggel, hogy a tenzorok képesek a GPU-kon futni, ezzel gyorsítva a számításokat. A két adatstruktúra közötti átjárás rugalmas, az ndarray objektumról tensorra való konvertálás a következő egyszerű kóddal elvégezhető:

```
import torch
import numpy as np

numpy_arr = np.ndarray([1, 2, 3])
tensor_arr = torch.from_numpy(numpy_arr)
```

Hasonlóan egyszerű az átváltás fordított irányban is:

```
tensor_arr = torch.tensor([1, 2, 3])
numpy_arr = tensor_arr.numpy()
```

A feladat megvalósítása során egyaránt fel lettek használva különböző célokra a korábban részletezett objektumok.

3.1.3 Open3D

Az Open3D egy nyílt forráskódú modern könyvtár a háromdimenziós adatok feldolgozására, amely elérhető Python és C++ nyelven fejlesztők számára is. [21]

Az Open3D különböző adat struktúrák használatát teszi lehetővé, melyek közül a feladat megvalósítása során a PointCloud reprezentálta a jelenet pontfelhőjét. A PointCloud struktúrának három fő mezője van:

- PointCloud.points: a pontfelhő pontjainak 3D-s koordinátái.
- PointCloud.normals: a pontokhoz tartozó normálvektorok.
- PointCloud.colors: a pontokhoz tartozó színek.

A három adatmező közül az első a fő, a másik kettő csak akkor tekinthető helyesnek, ha a dimenziójuk megfelel a pontok dimenziójával. Ezen felül az Open3D direkt memória hozzáférést biztosít a PointCloud adatmezőihez NumPy tömbbel.

A pontfelhő főleg a számítások ellenőrzésére és a részeredmények vizualizálására szolgáltak.

3.2 CUDA

A neurális radiancia mezőket alkalmazó módszerek egyik legnagyobb hátránya a megnövekedett számításigény.

Erre megoldást jelent az algoritmusok párhuzamosítása, melyet a felépítésükből adódóan nagy teljesítményű GPU-k segítségével lehet kivitelezni. Az NVIDIA párhuzamos működésű grafikus processzorainak programozására használható legnépszerűbb környezet a CUDA.

A CUDA (Compute Unified Device Architecture) programozási modell a C és C++ nyelvek kiegészítése, amely plusz könyvtárakkal és nyelvi elemekkel lehetővé teszi a GPU hatékony kihasználását a programok párhuzamosításával. A CUDA segítségével a fejlesztők optimálisabb működést érhetnek el és ezáltal gyorsabb radiancia alapú megoldások születhetnek. [22]

3.3 CARLA

A CARLA egy nyílt forráskódú önvezetésre szolgáló szimulátor, amely könnyen elérhető a fejlesztők és kutatók számára. A célja, hogy az önvezetésre épülő kutatási és fejlesztési feladatokat elősegítse és támogassa. A szimulátor felépítése egy jól skálázható kliens-szerver architektúrát valósít meg. [23]

A szerver felelős mindenért, ami a szimulációhoz kapcsolódik: szenzor rendereléséért, fizika számításáért, szimulált világ állapotának, illetve szereplőinek frissítéséért és még sok egyéb feladatért. Mivel a célja az, hogy realisztikus környezetet teremtsen érdemes egy arra dedikált GPU-n futtatni a szervert, különösen, ha gépi tanulásról van szó.

A kliens oldal olyan modulokat tartalmaz, melyek a jelenet szereplőinek logikáját és a szimulált világ kondícióit valósítják meg. A Python vagy C++ nyelven is elérhető

CARLA API egy olyan réteget biztosít, amely folyamatosan fejlődve új funkciókat biztosít a szerver és a kliens közötti kommunikációra.

A CARLA nagy szabadságfokot biztosít a felhasználók számára a választékos technikai eszközeivel:

- Forgalomirányítás: Egy beépített rendszer, amely kezeli a jelenet járműveit, hogy a valósághoz hű scenáriókat készítsen.
- Szenzorok: A járművek a beállított szenzorok alapján adnak információt arról, hogy mit észlelnek a környezetükből. A CARLA-ban lehetőség van a járművekhez csatolni szenzorokat, melyek a scenárió során készített mérései tárolhatóak és visszakereshetőek. Jelenleg többféle szenzor használata is támogatott, kamerától kezdve a radarokon át a lidarokig.
- Nyomon követés: A CARLA rögzíti a szimuláció szereplőinek állapotait az idő előrehaladtával, így lehetőség van tetszőleges időpillanatba visszaugrani és megvizsgálni a jelenetet.
- Új asset-ek: A CARLA biztosít a felhasználók számára különböző környezeteket és időjárási viszonyokat, de ezek az elemek testreszabhatóak és az irányelvek betartásával újakat is lehet generálni.
- Scenárió futtatás: A járművek tanításának megkönnyítése érdekében a CARLA biztosít különböző helyzeteket és szituációkat leíró útvonalakat.

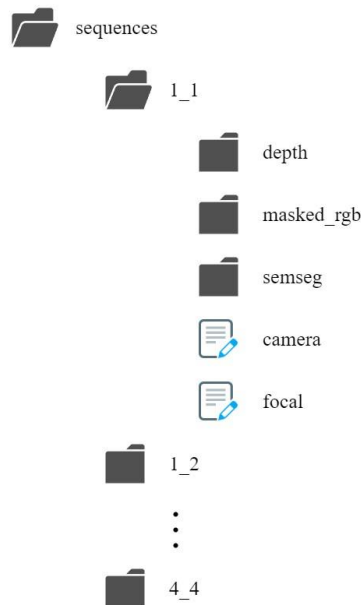
4 Megvalósítás

A következő fejezetekben a diplomamunka gyakorlati oldala van részletezve. Az előre kitűzött célok közé tartozott a NeRF módszer alkalmazása egy példán. Ez a gyakorlati példa megfelelő környezet és eszközök hiányában szintetikus úton lett előállítva a CARLA szimulátor felhasználásával. Az ilyen módon generált adatok zajmentesebbek, kevesebb hibát tartalmaznak, mintha valós úton történt volna az előállításuk. Majd ezt az adathalmazt használtam fel a radiancia alapú rekonstrukcióra és a szegmentációs lehetőségek vizsgálatára.

4.1 Szintetikus adathalmaz

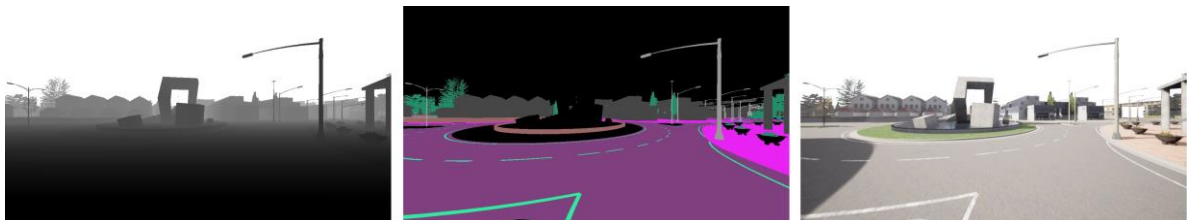
A feladat célja egy önjáró autó által ismert világ-részlet alapján egy körforgalom térbeli rekonstrukciója. A világ-részletet kamerával alkotott képekből rakja össze. A szintetikus adathalmaz a CARLA szimulátorban lett előállítva.

A jármű által körbejárt körforgalomnak négy bejárata van. Mindegyik bejáraton több előre meghatározott útvonalat is megtett a jármű, melyeken többnyire a körforgalom közepén elhelyezett műalkotás látszik. A megegyező útvonalakon készített képsorozatok abban térnek el, hogy a kamera kezdeti pozíciója különbözik. A kamerák inicializációja az önvezető jármű irányára merőleges síkon, bizonyos küszöbértékek között véletlenszerűen lettek generálva. A kamera ebből a kezdeti pozícióból követte az autó mozgását és készítette a képeket bizonyos időközönként úgy, hogy a kamera orientációja megegyezik a jármű adott pillanatban mért orientációjával. Összesen 4500 darab kamera pozícióból készültek képek.



4.1. ábra Az adathalmaz mappa struktúrája.

Külön almappákba (1_1, 1_2, ..., 4_4) kerültek az egyes útvonalak során készült képek. Minden kamerapozícióból készült egy mélységi kép, egy szemantikusan szegmentált kép és egy mélységi térkép alapján maszkolt kép, melyen az előre definiált értéket meghaladó távolságokhoz tartozó pixel alfa értéke 0, ami azt jelenti, hogy teljesen áttetsző. Utóbbit kizárólag a PNG (Portable Network Graphics) képformátum támogatja.



4.2. ábra Ugyanazon kamera pozícióhoz tartozó mélységi kép, szemantikusan szegmentált kép és maszkolt RGB kép.

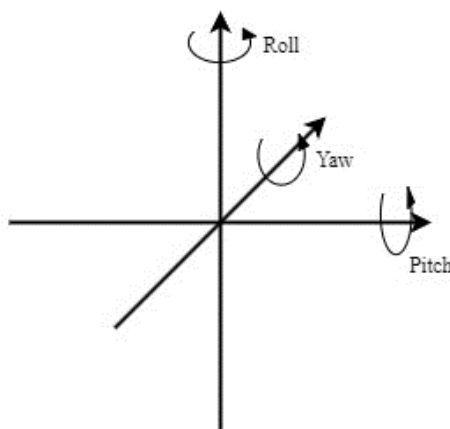
A mélységi információ alapján generált RGB képek mögötti intuíció az volt, hogy ily módon a távoli értékek nem lesznek zavaróak a tanítás során és a lényeges, közeli objektumok pontosabbak lesznek. Utólag kiderült, hogy az elv nem volt teljesen hibás, de a megválasztott távolsági küszöb mellett készített maszkolt képek pont ellenkezőleg hatottak a tanítás eredményére.

Az egyes útvonalakhoz tartozó könyvtárak a különböző típusú képeket tartalmazó mappák mellett tartalmaznak további két szöveges fájlt:

- camera.txt

- focal.txt

Az utóbbi a kamera fókusz távolságának értékét tárolja. Az előbbi az öt tartalmazó útvonalon készített képek kamera információit tartalmazza. Annyi sora van amennyi képből áll az adott sorozat és minden sorának hat koordinátája van. Az első három a kamera háromdimenziós transláció vektorának értékei, a második három pedig a kamera orientációját meghatározó roll yaw és pitch, melyek az egyes tengelyek körüli elforgatásoknak felelnek meg.



4.3. ábra Tengelyek menti elforgatások szemléltetése.

Megfigyelhető, hogy a sorok harmadik koordinátájának értéke csak minimálisan változik. Ez annak köszönhető, hogy ez a koordináta érték mutatja a jármű magasságát. A roll és pitch nullához közeli értékek, mivel a kamera főleg a függőleges tengely körül forgott.

```
3.261429214477539062e+01 -2.791510581970214844e+00 1.835505366325378418e+00 -4.272460355423390865e-04 -1.791074676513671875e+02 1.707547227852046490e-03
3.017135238647460938e+01 -2.842333316802978516e+00 1.835375905036926270e+00 -4.150390159338712692e-03 -1.789016113281250000e+02 7.854716968722641468e-04
2.770692634582519531e+01 -2.952233791351318359e+00 1.834720611572265625e+00 -2.612304501235485077e-02 -1.778777465820312500e+02 6.761886761523783207e-04
2.514418601989746094e+01 -3.415596961975097656e+00 1.832604289054870605e+00 -9.603880345821380615e-02 -1.722281494140625000e+02 -1.010867999866604805e-03
2.261042213439941406e+01 -4.448024749755859375e+00 1.832610130310058594e+00 -9.558103978633880615e-02 -1.625724334716796875e+02 -1.372867962345480919e-03
2.028378295898437500e+01 -5.890097618103027344e+00 1.833059668540954590e+00 -8.068847656250000000e-02 -1.530509185791015625e+02 -1.072339597158133984e-03
```

4.4. ábra Kamera információk.

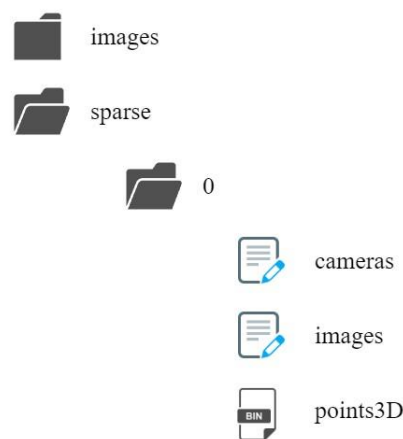
4.2 3D jelenet rekonstrukciója

A rendelkezésre álló adathalmazt olyan formára kell hozni, amit a gépi tanuló algoritmus megkíván. A feladat esetében ez a Gaussian Slpating volt.

4.2.1 Az adathalmaz formázása

Az adathalmaz formázásának első lépése az egyes útvonalak adatainak összevonása. Ennek az abszolválására implementáltam a `CarlaDataMerger` Python nyelven implementált projektet, melynek a célja az adatok összevonása. A kód végigiterál az útvonalak mappáin és a kívánt típusú képeket a célkönyvtárba másolja egyediséget biztosító nevekkel. Ezzel párhuzamosan a kamera információkat tartalmazó szöveges fájlok sorait egyetlen szöveges fájlba másolja.

Az összevont adathalmaz formázásához meghívtam a `CARLA2NeRF` nevű Git repository `carla2nerf`, majd a `sample_gs_data` Python szkripteket, melynek eredményeként előállt a Gaussian Splatting eljárás által feldolgozható adathalmaz. [24]



4.5. ábra Gaussian Splatting elvárt bemenete.

Az előállt adathalmaz az eredeti képek megadott százalékát tartalmazza. A képek az `images` nevű mappába kerültek. A képek mellett az adathalmaz tartalmaz még további három fájlt. A `cameras` szöveges fájlban megtalálhatóak a kamera tulajdonságait leíró paraméterek:

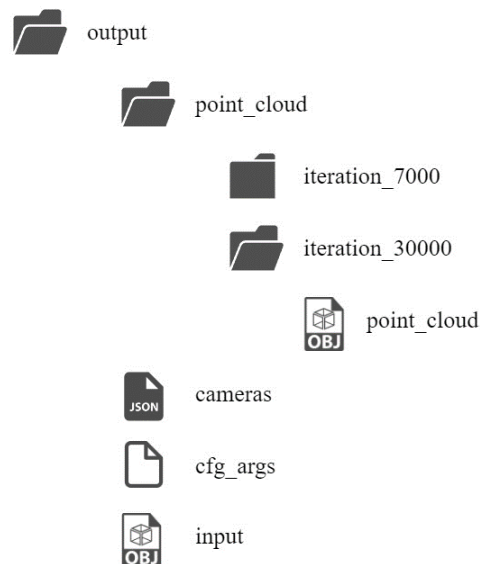
- kamera típusa,
- kép szélessége pixelekben mérve,
- kép magassága pixelekben mérve,
- kamera fókusztávolsága,
- kamera vízszintes és függőleges látószöge.

Az images fájl tartalmazza az egyes képek azonosítóját és nevét, valamint a képeket előállító kamera pozícióit és irányát meghatározó 4 elemű vektorokat.

A points3D a háromdimenziós Gauss halmazok kezdeti középpontjait tárolja, melyet a COLMAP generált le az egymást követő képek jellegzetességeinek lokalizálása alapján.

4.2.2 Tanítás

A Gaussian Splatting tanítása az elkészült adathalmazzal történt. A tanítás percek alatt lefutott mind a szemantikusan szegmentált képekből, mind a mélységi információ alapján maszkolt RGB képekből álló adathalmazon.



4.6. ábra Gaussian Splatting kimenete.

A Gaussian Splatting kimenete a tanításhoz felhasznált képek és a kezdeti pontfelhő alapján számított térbeli rekonstrukció. Az optimalizáció 30000 iteráción keresztül történt és kétszer végzett mentést: először a 7000. iteráció után és egyszer a teljes tanítást követően. A Gaussian Splatting egyik előnye a gyors konvergencia, ennek köszönhetően a 7000. iterációt követően is korrekt eredményt produkált, hasonló veszteséggel, mint a végén.

A Gaussian Splatting eredménye az ellipszoidokból álló háromdimenziós rekonstrukció. A teret a Polygon formátumú point_cloud írja le, melynek attribútumai a következők:

- az ellipszoidok középpontjának 3D-s koordinátája

- az ellipszoid három elemű DC komponense, amely az ellipszoid diffúz színét reprezentálja
- a nézeti irány függő színét meghatározó 45 elem, melyet egy szferikus harmonikus segítségével lehet kinyerni
- az áttetszőség értéke
- az egyes tengelyek menti skálázás
- az elforgatás négy elemű vektora

Az eljárás kimenete tartalmazza továbbá a pontfelhő rendereléséhez szükséges `cfg_args` nevű konfigurációs fájlt és a tanításra szánt képekhez tartozó kamera információkat. A körforgalmas adathalmaz esetében 206857 ellipszoid feszíti ki a teret.

4.3 Címkezés a szegmentált képek vetítésével

A térfogat szemantikus szegmentálásához szükséges eljárás a PyCharm fejlesztői környezetben lett implementálva Python nyelven. A pontfelhő megjelenítéséhez és a különböző részfeladatok helyességének ellenőrzésére szolgáló vizualizációkhoz az Open3D könyvtár nyújtott segítséget.

4.3.1 Kovarianciamátrix kiszámítása

A jelenetben szereplő háromdimenziós Gauss halmazok geometriája megfelel az ellipszoidéval, így a továbbiakban ellipszoidként is lesznek hivatkozva. Az ellipszoidokat reprezentáló 3x3-as kovarianciamátrix kiszámításához szükség van az $s = (s_0, s_1, s_2)$ tengelyek menti skálázásra, valamint az elforgatás $q = (q_0, q_1, q_2, q_3)$ kvaternió vektorára.

A tengelyek menti skála értékeknek venni kell az exponenciálisát és betölteni egy 3x3-as egység mátrix főátlójába:

$$L = \begin{bmatrix} e^{s_0} & 0 & 0 \\ 0 & e^{s_1} & 0 \\ 0 & 0 & e^{s_2} \end{bmatrix}.$$

Az orientációt meghatározó q kvaternió vektorból konstruálni kell egy 3x3-as mátrixot a következőképpen:

$$R = \begin{bmatrix} 1 - 2(q_2^2 + q_3^2) & 2(q_1q_2 - q_0q_3) & 2(q_1q_3 - q_0q_2) \\ 2(q_1q_2 - q_0q_3) & 1 - 2(q_1^2 + q_3^2) & 2(q_2q_3 - q_0q_1) \\ 2(q_1q_3 - q_0q_2) & 2(q_2q_3 - q_0q_1) & 1 - 2(q_1^2 + q_2^2) \end{bmatrix}.$$

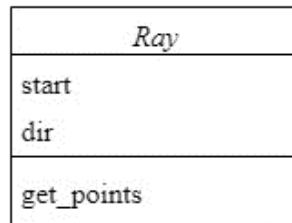
Ezt követően az azonos dimenzióra hozott mátrixok szorzatát meg kell szorozni a szorzatuk transzponáltjával és így előáll az ellipszoid geometriáját leíró kovariancia mátrix:

$$K = (R \times L) \times (R \times L)^T.$$

A kovariancia mátrix egy nagyon tömör, könnyen használható reprezentáció, amely hatékonyabbá teszi az optimalizációt. Ezután az ellipszoid középpontjához tartozó 3D-s koordinátájával való eltolással elhelyezhető megfelelő helyre a térben.

4.3.2 Ellipszoid megjelenítése: Ray casting

A sugár működése a Ray osztályban van implementálva. A sugár megfelel egy egyenesnek, melynek képlete megadható az irányvektorával és egy ponttal. A Ray osztálynak két attribútuma van: kezdőpont (start) és irány (dir). A sugár vizualizációjához készült egy függvény, amely meghatározott lépésközzel visszaadja a pontjait a megadott intervallumon.



4.7. ábra Ray osztálydiagramja.

A Gauss halmazok működését az Ellipsoid osztály tartalmazza. Minden Ellipsoid példánynak van középpontja (center) és az előző fejezetben részletezett módon kiszámított kovariancia mátrixa (cov_mat). Legfontosabb módszere a compute_intersection, amely a paraméterül megkapott sugárral való metszéspontját adja vissza.

Az ellipszoid felületének p pontjait a következő egyenlettel lehet megkapni:

$$(p - c)^T \times K^{-1} \times (p - c) = 1,$$

ahol a c az ellipszoid középpontját jelöli. Az ellipszoid felületét leíró egyenletbe a sugár egyenletének behelyettesítése a következő összefüggéshez vezet:

$$(s + d \cdot t - c)^T \times K^{-1} \times (s + d \cdot t - c) = 1.$$

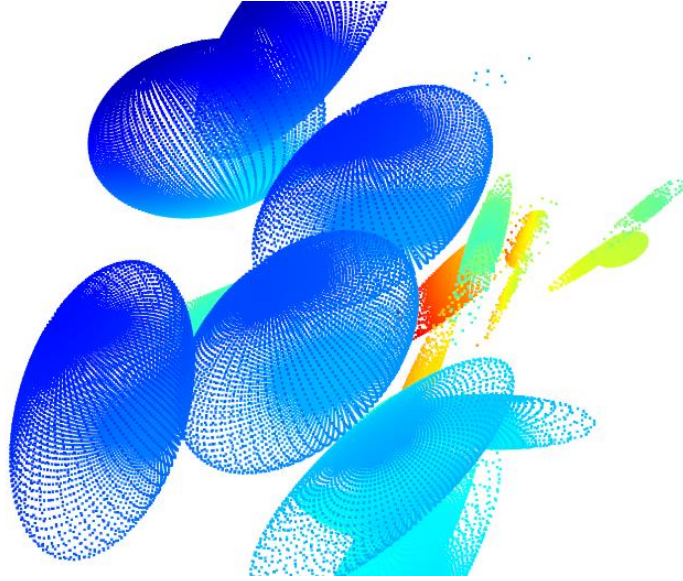
Az egyenlet rendezését követően egy másodfokú egyenletet kapunk, melynek D diszkriminánsa meghatározza a metszés módját:

- $D < 0$ esetén a sugár nem metszi az ellipszoidot.
- $D = 0$ esetén egy megoldáshoz jutunk. Ekkor a sugár az érintője az ellipszoidnak.
- $D > 0$ esetén két helyen is metszi az ellipszoidot a sugár. Ekkor a sugár kezdőpontjától mérve a közelebbi pont lesz a kívánt metszéspont.

<i>Ellipsoid</i>
center
cov_mat
get_points
compute_intersection

4.8. ábra Ellipszoidot reprezentáló osztály diagramja.

A metszéspontok helyességének ellenőrzésére készült egy adott ellipszoid pontjait megjelenítő `get_points` függvény, amely az ellipszoid középpontja körül inicializál egy gömböt, majd ezen gömbön mintavételezett pontokból indít sugarakat az ellipszoid középpontja felé. A sugarak és az ellipszoid metszéspontjait egy pontfelhőbe menti el és ezek kerülnek kirajzolásra. A függvény több, egymáshoz közeli ellipszoidra történő meghívásával született az alábbi kép.



4.9. ábra Távolság alapján színezett Gauss halmazok.

Látszik, hogy a rekonstrukció ellipszoidjai különböző alakot és orientációt vesznek fel. Előfordul, hogy az ellipszoidok egymásba csúsznak, ami nehezíti a ray casting-on alapuló eljárást és rontja annak pontosságát.

4.3.3 Kamera pozíció és orientáció

A tanítóhalmaz képeit alkotó kamera információkat a CameraDatas osztály tárolja. A kamera fontosabb adatai közé tartozik a kép felbontása (width és height), a vízszintes és függőleges irányú látószög (fovX, fovY), az egyes kamera pozíciók 3D-s koordinátája, a kamerák orientációját leíró négy elemű vektorok és az elkészített képek egyedi nevei. Az adatokat a Gaussian Splatting által előállított adathalmaz JSON (JavaScript Object Notation) kiterjesztésű cameras fájljából olvassa be a load_from_json metódussal.

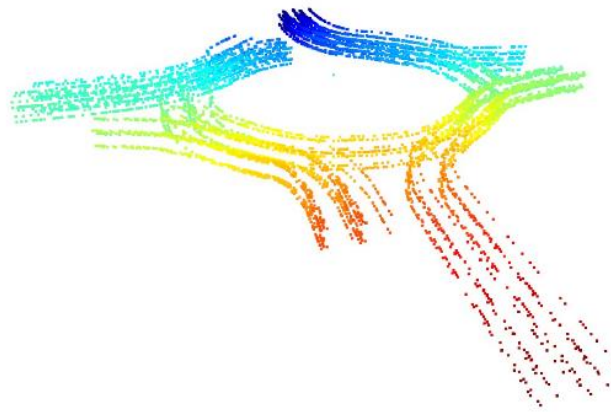
A nézeti irányt leíró kvaternió vektor a fájlból beolvasott rotációs mátrixból konstruált 4x4-es K mátrix sajátvektoraiból és sajátértékeiből van kiszámítva, melyet a directions_from_rotmats és a rotmat_to_qvec metódusok valósítanak meg. A rotációs mátrix a Z tengely menti egységvektort (0, 0, 1) forgatja a megfelelő irányba.

A focal_to_fov függvény kiszámítja a vertikális és horizontális látószöget a kép szélességéből (w) és magasságából (h), illetve a kamera fókusztávolságából (f) a következő összefüggéssel:

$$fovX = 2 \cdot \tan^{-1} \frac{w}{2f}$$

$$fovY = 2 \cdot \tan^{-1} \frac{h}{2f}$$

Ahhoz, hogy a képeken való végigiterálás kézenfekvőbb legyen, létre lett hozva egy `get_by_index` metódus, ami index alapján visszatér a kamera pozíciójával, a kamera orientációjával és az onnan készített kép egyedi nevével, amely alapján a képet be lehet tölteni.



4.10. ábra Távolság alapján színezett kamera pozíciók.

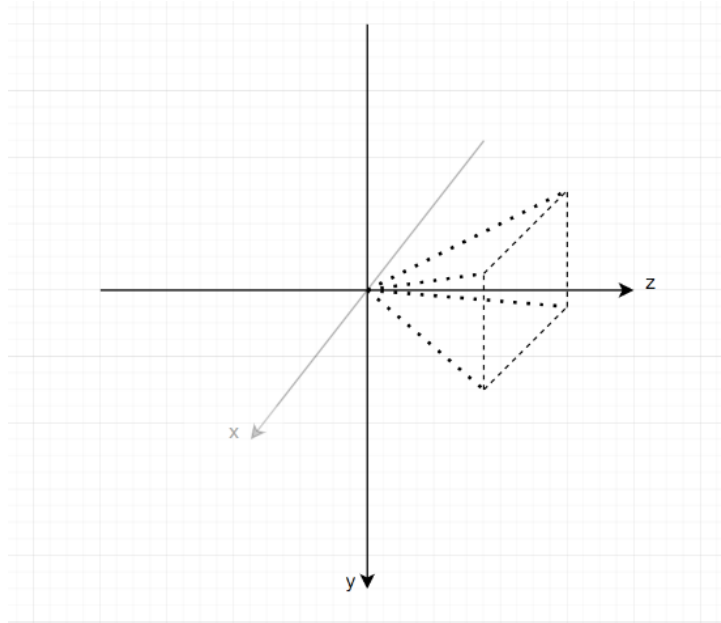
Összesen 1575 kép alapján lett tanítva a Gaussian Splatting model és ezen képeket készítő kamerák pozícióiból álló pontfelhő a fentebbi ábrán látható. Jól kivehető a körforgalom helyzete és hogy ugyanazon útvonalon több kamera pozíció is lett inicializálva a korábban említett módon.

<i>CameraDatas</i>
width
height
fovX
fovY
positions
directions
image_name
load_from_json
rotmat_to_qvec
directions_from_rotmats
focal_to_fov
set_corners
get_by_index

4.11. ábra CameraDatas osztály fontosabb attribútumai és függvényei.

A kamerák számontartását a CameraDatas végzi, viszont az egyes képekhez tartozó képkereteket a CameraWindow osztály valósítja meg, melynek célja, hogy a kamerák helyzetéből a jelenetbe kilőtt sugarak könnyen lekérdezhetőek legyenek.

A képkeretet a kamera pozíciója (center) és a sarokpontjai (corners) definiálják, amelyekhez szükséges információt a CameraDatas get_by_index függvény szolgáltatja. Inicializálásakor létrejön egy kezdeti képkeret, melynek sarokpontjai a Z tengely mentén egy egységgel eltolva, arra merőleges síkon vannak inicializálva a kamera fókusz távolságának megfelelően.



4.12. ábra Kamera ablak inicializálása.

A példány konstruktora a center attribútumát beállítja a kamera pozíciójának megfelelően és a kezdeti sarokpontokat a set_corners metódussal transzformálja a megfelelő helyre a térben.

A set_corners függvény először kiszámítja az X, Y és Z tengely menti elforgatások szögét, az axis_with_angles metódussal, ami paraméterül megkapja az elforgatás kvaternió vektorát. A kapott pitch, roll és yaw felhasználásával konstruálható a három tengely körüli elforgatás mátrix:

$$R_y(pitch) = \begin{bmatrix} \cos(pitch) & 0 & \sin(pitch) \\ 0 & 1 & 0 \\ -\sin(pitch) & 0 & \cos(pitch) \end{bmatrix}$$

$$R_x(roll) = \begin{bmatrix} 1 & 0 & 0 \\ 0 & \cos(roll) & -\sin(roll) \\ 0 & \sin(roll) & \cos(roll) \end{bmatrix}$$

$$R_z(yaw) = \begin{bmatrix} \cos(yaw) & -\sin(yaw) & 0 \\ \sin(yaw) & \cos(yaw) & 0 \\ 0 & 0 & 1 \end{bmatrix}$$

Az egymástól eltérő koordináta rendszerek közötti konverzió nem mindig triviális, így bonyodalmat okozhat a tengelyek és elforgatások párosítása. Kis kísérletezés után kiderült, hogy a yaw érték az Y, a pitch érték a Z és a roll érték az X tengely mentén forog. Az így kapott forgatás mátrixokat összeszorozva előáll a rotációs mátrix:

$$R = R_y(pitch) \cdot R_x(roll) \cdot R_z(yaw)$$

Az R mátrix segítségével a rotate függvény elforgatja a kezdeti sarokpontokat, majd a translate függvény eltolja őket a kamera pozíciójába mutató helyvektorral.

<i>CameraWindow</i>
center
corners
axis_with_angles
rotate
translate
set_corners
get_ray_by_pixel

4.13. ábra Kamera ablakának osztályát reprezentáló osztály.

Az osztály legfontosabb függvénye a get_ray_by_pixel, paraméterként megkapja a kép adott pixelének koordinátáját, ami alapján visszatér a térben egy megfelelően irányított és pozicionált Ray példánnyal. Ez a sugár a kamera középpontjából indul és a megadott pixel irányába mutatva halad át a téren.

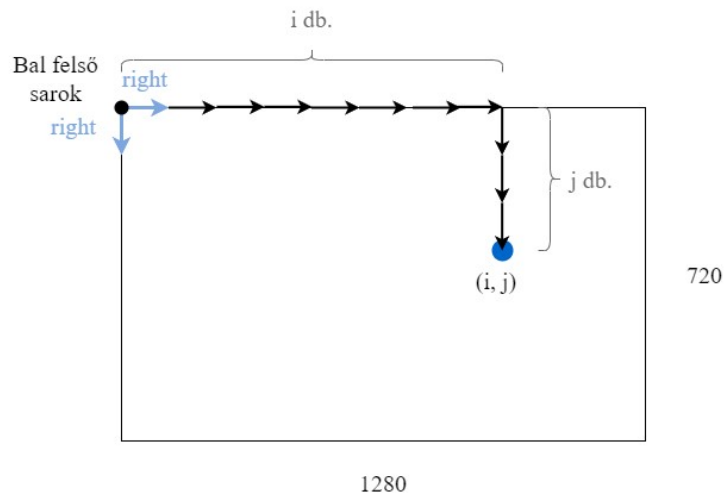
A függvény a sarokpontokba mutató vektorok alapján számítja ki a megfelelő irányt. Ehhez első lépésként definiál két egy pixel nagyságú vektort:

- right: jobbra mutató vektor
- down: lefelé mutató vektor

A jobbra mutató vektor a bal felső sarokba mutató vektor és a jobb felső sarokba mutató vektor különbségének a kép szélességének értékével való osztásával kapható meg. Hasonló az eljárás a lefelé mutató vektor kiszámításához, ekkor a bal alsó sarokba mutató vektor és a bal felső sarokba mutató vektor különbségét kell osztani a kép pixelben mért magasságával.

Ezután a két irányvektornak a paraméterül átadott, megfelelő koordinátákkal vett lineáris kombinációjához hozzá kell adni a kamera pozíciójából a bal felső sarokba mutató vektort (BF) és így előáll a pixelbe mutató irány:

$$\text{sugár iránya} = BF + (i * \text{right} + j * \text{down})$$



4.14. ábra Adott pixel irányába mutató sugár számítása.

A sugár iránya egy egység hosszú vektor, amely megkapható a vektornak a hosszával való elosztását követően.

4.3.4 Sugarak inicializálása maszkolt pixelek irányában

A jelenetben való sugár metszés először egy könnyen átlátható példán lett ellenőrizve.



4.15. ábra Véletlenszerűen választott szemantikusán szegmentált kép.

A szemantikusán szegmentált képekből álló tanító halmazból véletlenszerűen ki lett választva egy kép, melyen a növényzet alapján lett végrehajtva a maszkolás. Ennek eredménye, hogy minden növényzet címkéjű pixel intenzitása megmaradt és bármely másik osztályba sorolt pixel intenzitása zérus lett.



4.16. ábra Szegmentált kép növényzet alapján készített maszkja.

A program végigiterál a kép pixelein és a nem nulla intenzitású pixelek irányába a képet készítő kamera adatai alapján inicializált CamerWindow példány visszaad egy sugarat.

Azért a növényzet lett kiválasztva, mert kisebb számban és sok helyen fordul elő, amivel gyors és egyértelmű eredmény születhet.

4.3.5 Címkézés maszkolt kép pixelein keresztül indított sugarakkal

A részfeladat célja, hogy egy kép alapján azonosított növényzet címkéjű pixelekhez tartozó ellipszoidokat megtalálja a program a térben. A következő ábrán látható az előző fejezetben említett szemantikusan szegmentált kép RGB változata, ami alapján megbecsülhető a növényzet térbeli elhelyezkedése.

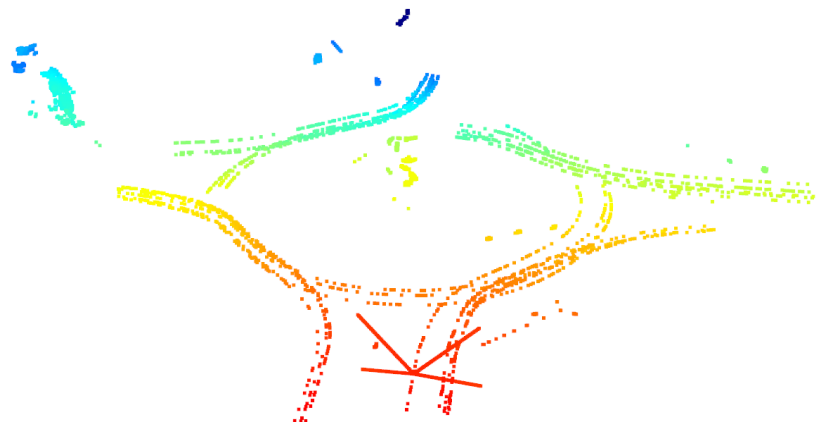


4.17. ábra A kiválasztott kép RGB változata.

Ahhoz, hogy a metszéspontok térbeli elhelyezkedését értelmezni lehessen, a kamerák pozíciója is meg lett jelenítve, melyet követően már érzékelhetőek a távolságok és az irányok.

A kamerák pozíciója mellett még meg lett jelenítve négy egyeneshez tartozó pontthalmaz. Ezek az egyenesek a képet előállító kamera adatai alapján létrehozott CameraWindow példány középpontját kötik össze a képkeret négy sarkának pontjával.

A kép betöltése után elkészült a maszkolt változata, amin a növényzethez nem tartozó pixelek intenzitása zérus. Minden másik pixel irányában indult egy-egy sugár és az azok által elmetszett pontok kerültek a kirajzolt pontfelhőbe, ami távolság alapján van színezve.



4.18. ábra Metszéspontokat, kamera pozíciókat és a kamerát tartalmazó pontfelhő.

A maszkolt képen a növényzethez tartozó pixeleket három csoportba lehet rendezni. A bal oldalon lévő fák, a szobor mögötti növényzet és a kilenc egymás mellett sorban elhelyezkedő cserepes növény.

A kirajzolt pontfelhőn megfigyelhető ez a három csoport. A baloldali pontok viszonylag egy helyre estek, ami jót jelent. A szobor mögött lévő növényzet irányába tartó sugarak nagy része elakadt a körforgalom középpontjában elhelyezkedő szoborban, illetve a magasabb metszéspontok térben jó helyre kerültek. A jobb oldalon látható növények is hasonló módon félsikerrel lettek megtalálva, voltak sugarak, melyek elakadtak a szoborban.

4.3.6 Többféle címke rendelése egy ellipszoidhoz

Az egyszerűsített példa sikerességét követően a több kép és többféle címke alapján történő címkézés következett. A futás időigényessége miatt a tanítás, illetve címkézés több lépésben, a program többszöri futtatásával történik. Emiatt szükség van a címkék fájlba való elmentésére és betöltésére.

A címkézéshez létre lett hozva annyi üres lista (`label_lists`), amennyi ellipszoid található a jelenetben. A pixelek intenzitásaihoz az `assign_label` metódus rendel egy-egy címkét. Mivel az egyes szegmensek színének összege egyedi, ezért kézenfekvő az összegük alapján végezni a hasonlítást. A címkézés úgy működik, hogy a pixelen áthaladó sugár az elmetszett objektumhoz tartozó listához hozzáadja a pixel alapján meghatározott címkét.

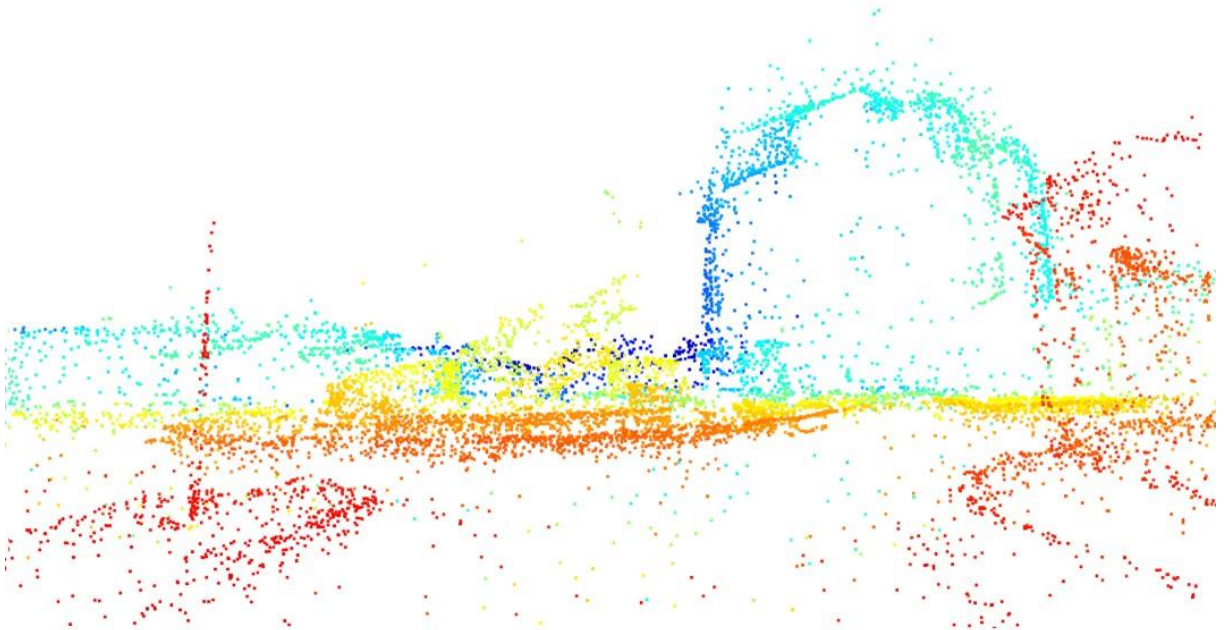
```
def assign_label(color):
    if np.sum(np.array([107,142,35]))==np.sum(color): return 1 #
    color_vegetation
    elif np.sum(np.array([0,0,0]))==np.sum(color): return 2 # color_statue
    elif np.sum(np.array([153, 153, 153]))==np.sum(color): return 3 #
    color_pole
    elif np.sum(np.array([102, 102, 156]))==np.sum(color): return 4 #
    color_wall
    elif np.sum(np.array([70,70,70]))==np.sum(color): return 5 #
    color_building
    elif np.sum(np.array([0,0,255]))==np.sum(color): return 6 #
    color_vehicle
    elif np.sum(np.array([244,35,232]))==np.sum(color): return 7 #
    color_sidewalk
    elif np.sum(np.array([128,64,128]))==np.sum(color): return 8 #
    color_road
    elif np.sum(np.array([157,234,50]))==np.sum(color): return 9 #
    color_roadline
    elif np.sum(np.array([190,153,153]))==np.sum(color): return 10 #
    color_fence
    else:
        return -1
```

A CARLA szimulátorban előállított szemantikusan szegmentált képeken 9 féle objektum van megkülönböztetve: növényzet, szobor, villanyoszlop, fal, épület, jármű, járda, útfestés és kerítés.

Megeshet, hogy egy ellipszoidhoz több címke is fog tartozni ebben az esetben a címke lista leggyakoribb eleme, azaz a módusza adja meg az adott 3D-s Gauss halmaz végleges címkéjét.

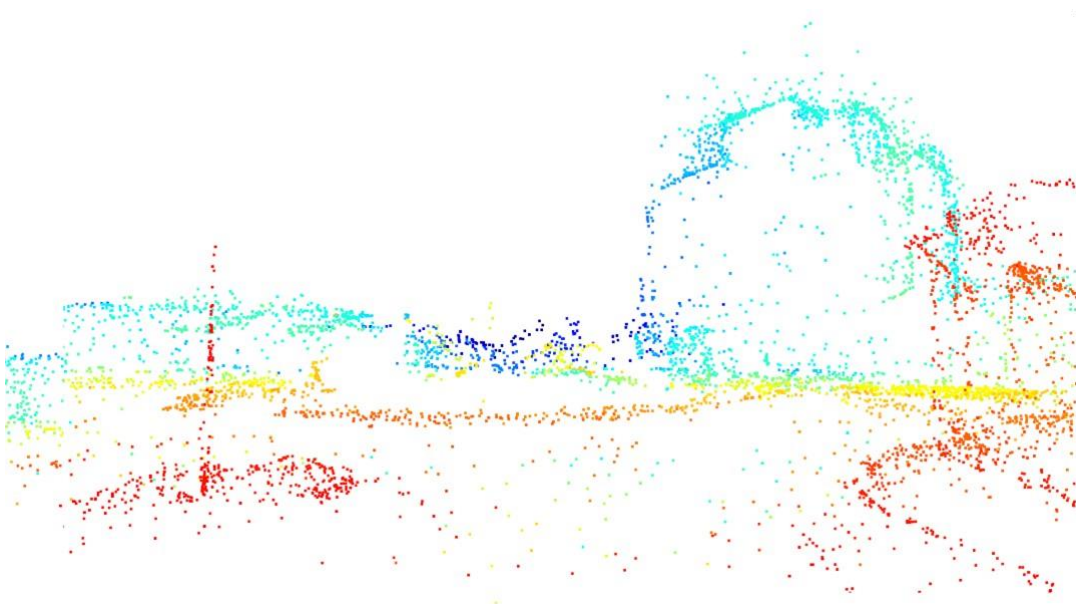
4.3.7 Szobor eltávolítása

A körforgalom közepén lévő szobor címkéjével ellátott pontok elhagyásával kivehető a címkézés eredménye. Az első képen minden elmetszett, bármilyen címkével ellátott ellipszoid középpontja látható távolság szerint színezve.



4.19. ábra Bármilyen címkéjű elmetszett ellipszoid középpontja.

A második képen a szobor címkéjű pontokat leszámítva minden másik pontból alkotta pontfelhő látható.



4.20. ábra Az összes, kivéve a szobor címkéhez tartozó elmetszett ellipszoid középpontja.

A kirajzolt pontfelhő színezzhető a címkéje alapján is, melyhez a PointCloud colors nevű mezőjében az elmesztett ellipszoidok címke listájának móduszához tartozó szint kell beállítani. Fontos, hogy a színek dimenziója megegyezzen a pontok dimenzió számával.



4.21. ábra A címkék alapján színezett ellipszoid középpontok.

Megfigyelhető, hogy a körforgalom közepén elhelyezkedő szobor eltávolítása után megjelennek az addig kitakart objektumok.

4.4 Sugárkövetés optimalizálása

Az ellipszoidok mennyiségéből, a kamerák felbontásából és a képek számának nagyságából adódóan a ray casting számítási igénye rendkívül magas és futási ideje nem órákban, hanem napokban mérhető. Emiatt a címkézés egyszerűsítésére és gyorsítására kisebb módosítások lettek bevezetve, melyek ebben a fejezetben kerülnek részletezésre.

4.4.1 Képek számának csökkentése

Az adathalmazt alkotó képeket nem véletlenszerű pozíciókban inicializált kamerák készítették, hanem egy a körforgalomban végig haladó jármű. Emiatt az egymást követő képek közötti eltérés minimális. A jelenetet felcímkéző sugármetszésen alapuló kód gyorsítása érdekében csak minden tizedik kamera által készített kép alapján történik a címkézés. Ezzel a módosítással kevésbé lesz precíz az eredmény, de annak ellenére, hogy a futási idejét sokat javít a minőségéből mégsem veszít annyit.

4.4.2 Átlátszóság figyelembevétele

A rekonstruált háromdimenziós jelenetet alkotó Gauss halmazok száma magas, nagyságrendileg több százezer ellipszoid alkotja. Ez nagymértékben befolyásolja a sugármetszés számítási igényét, mivel az egyes sugaraknak több ellipszoiddal is ellenőrizni kell a metszéspontját.

Ennek a problémának a kiküszöbölésére, illetve javítására megoldást nyújt, ha az egyes halmazokhoz tartozó, bizonytalansági szintnek is tekinthető átlátszósági értékek is figyelembe vannak véve. Így csak azokban az esetekben ellenőrizzük a sugár és ellipszoid metszéspontját vagy pontjait, ha az ellipszoidhoz tartozó átlátszósági érték egy előre meghatározott küszöbértéket meghalad.

4.4.3 Képek nagy felbontásának kezelése

A képek felbontása 720x1280, ami egy meglehetősen nagy felbontás a sugármetszés feladat szempontjából. A felbontásból adódó nehézség áthidalására két féle megoldás is van:

- kép felbontásának csökkentése,
- minden x . pixel figyelembevétele.

A felbontás csökkentése azért lehet problémás ebben az esetben, mert ilyenkor több pixel értéke átlagolódik, ami után a szegmentált kép pontatlan lesz és az adott pixelhez tartozó sugár az eredeti képhez képest eltérő címkével fogja ellátni az elemzett objektumot.

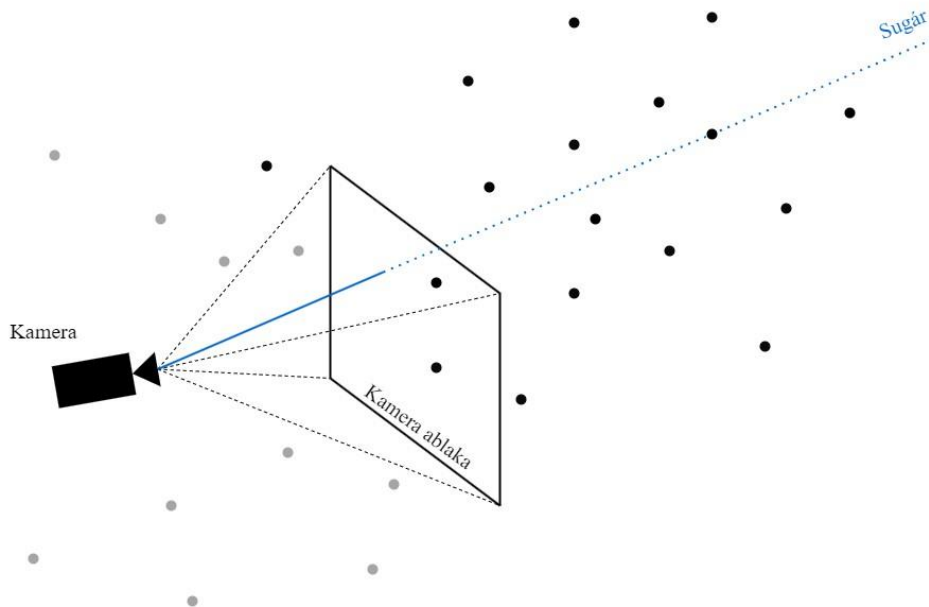
Emiatt a második megoldás lett kiválasztva és a vízszintes és függőleges irányban is csak minden huszadik pixelre végződik el a sugármetszés. Ezzel az egyszerűsítéssel élve képenként 36x64 sugarat indít a program, ami egy kép esetén jelentősen rontja a címkézés minőségét, viszont sok képre elvégezve és azok aggregát eredményét véve elfogadható lesz a végeredmény.

4.4.4 Ellipszoidok szűrése a kamera síkjával

Ahogy korábban már említésre került, a rekonstruált háromdimenziós jelenetet meglehetősen sok Gauss halmaz alkotja, ami megnehezíti a ray casting feladatát. Fontos, hogy ezeknek a száma valamilyen módon csökkenjen az átlátszóságuk figyelembevételéhez hasonlóan.

Egy másik ilyen lehetőség, ha a pozíciójuk szerint vannak kiválogatva az ellipszoidok. Az algoritmus először végig iterál a képeken és minden kép esetén indít 36x64 sugarat. Sokkal optimálisabb, ha a kamera síkjának az irány szerinti negatív oldalán elhelyezkedő ellipszoidokat nem vesszük figyelembe. Mivel adott a kamera pozíciója és a nézeti iránya, így kiszámolható a kamera síkja és így meghatározható,

melyik pontok esnek a rossz oldalára. Ezen pontok kiszűrésével jelentősen kevesebb ellipszoidra kell ellenőrizni a sugárral való metszéspont számítást és sokat lehet gyorsítani az eljáráson.

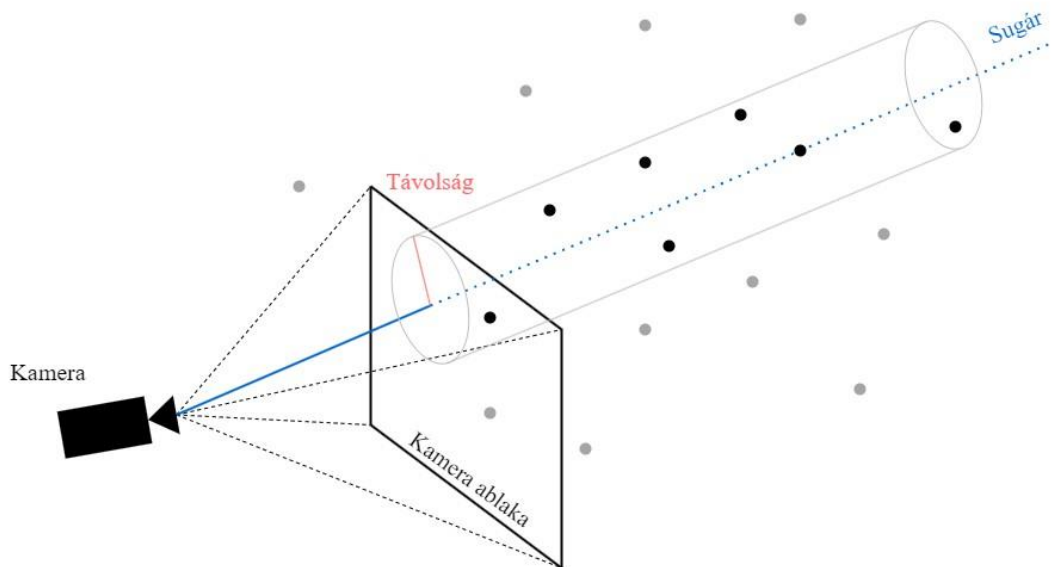


4.22. ábra Kamera mögé eső ellipszoidok szűrése.

4.4.5 Sugár körüli szűrés

Egy másik lehetőség a Gauss halmazok pozíciója alapján történő szűrésének a sugártól való távolság vizsgálata. Ennek során a sugárhoz közel eső halmazok megmaradnak és csak ezekre hívódik meg a sugármetszés számítás (`compute_intersection` metódus). A sugártól távolabbra eső halmazok, melyeket valószínűleg nem érint a sugár figyelmen kívül maradnak.

Az ellipszoidok középpontjának a sugártól való távolságához végig kell iterálni a megmaradt ellipszoidokon. A vizsgált ellipszoid távolságát a középpontja és a középpontjának a sugárra projektált képének a távolsága határozza meg. A sugártól megadott küszöbértéket meghaladó távolságban lépő ellipszoidokat ideiglenesen eldobjuk és csak a megmaradt ellipszoidokra végezzük el a metszéspontszámítást.



4.23. ábra Ellipszoidok szűrése a sugár mentén.

Korábbi fejezetben említésre került, hogy minden kép esetén történik egy előszűrés, ami a kép síkjának rossz oldalára eső Gauss halmazokat eliminálja. A jó oldalra eső ellipszoidok indexét pedig lementi egy listába. Az egyes sugarak inicializálásakor újra szűrésen esnek keresztül a megmaradt indexek, a hozzájuk tartozó középpont sugártól való távolsága alapján.

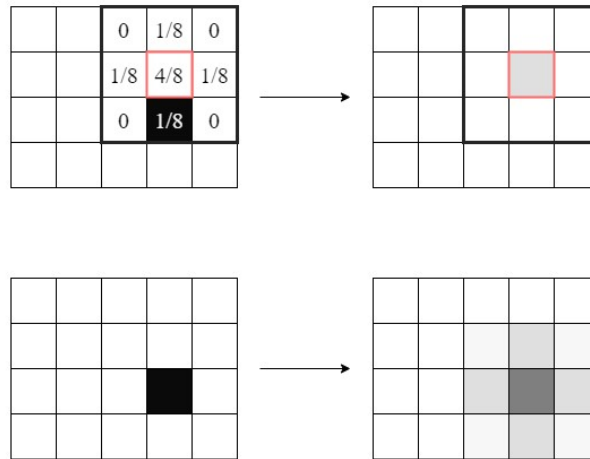
4.5 Szegmentáció utófeldolgozása

A címkézett pontfelhőn, a klasszikus képfeldolgozó algoritmusokhoz hasonlóan el lehet végezni különböző műveleteket, amelyek javítják a szegmentált pontfelhő minőségét, tisztaságát és értelmezhetőségét.

4.5.1 Smoothing

Egyik gyakran alkalmazott utófeldolgozási eljárás a simítás. A klasszikus képfeldolgozásban a simítás során a kép kiugró, zajos értékei eliminálódnak vagy tompulnak. Képek esetén egy meghatározott méretű csúszóablak megy végig adott lépésközzel a képen és számítja ki az átlagot a mezői súlya alapján.

$$0*255+1/8*255+0*255+1/8*255+4/8*255+1/8*255+0*255+1/8*0+0*255 = 223$$



4.24. ábra Kép simítása Gauss kernellel.

A művelet hasonlóan elvégezhető térben is. Csúszóablak helyett egy távolságot kell meghatározni, amely alapján el lehet dönteni, hogy két ellipszoid egymás szomszédságában van-e. Ezután minden ellipszoidra ki kell számítani akár a távolság alapján súlyozott szomszédok átlagos címkéjét. Ennek eredményeképpen a kiugró, zajos értékek eliminálódnak és kis környezetben egységesebb szegmentálást kapunk. Az eljárás azért működhet, mert a Gaussian Splatting által eredményül adott ellipszoidokkal kitöltött térben, a jellegéből adódóan az egymáshoz közel eső ellipszoidok általában ugyanazon objektumhoz tartoznak. Nem jellemzőek a kiugró értékek.



4.25. ábra Zajos szegmentáció.

A képen az ellipszoidok középpontja alkotta pontfelhő látható. A pontfelhő színezése pedig az ellipszoidok színezését kapta a szegmentációt követően. Megfigyelhetők olyan részletek, ahol vegyesebben fordulnak elő különböző szegmensek kis térfogaton belül. Ilyen zajosabb részletet lehet látni például a villanyoszlopoknál.



4.26. ábra Szegmentáció simítás után.

A simítás elvégzése után sokkal egységesebb lokálisan összefüggő szegmens eloszlás figyelhető meg. A korábban említett zajos villanyoszlopok teljesen egységessé váltak. Vannak részletek, melyek hibásan vesztek el, ez abból fakad, hogy az eljárás nem veszi figyelembe a közeg sűrűségét.

4.5.2 Zajsűrés

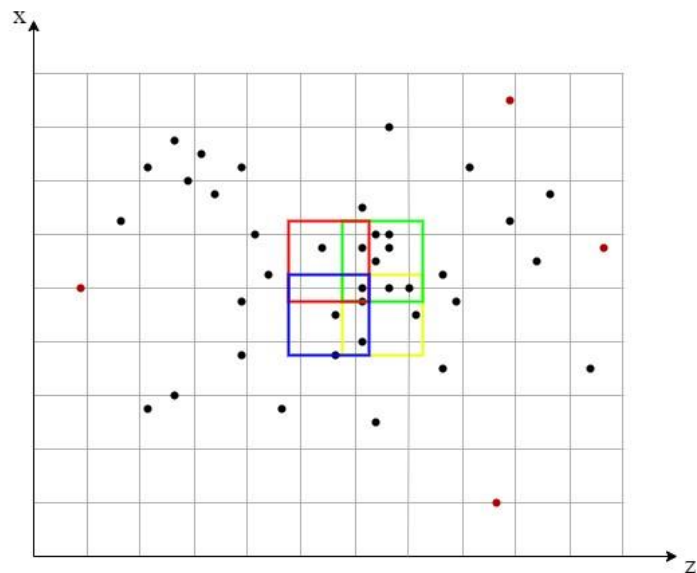
Egy másik lehetőség a bizonytalan vagy hibás metszéspontok törlése. Ezen pontok jellemzően olyan közegben vannak, ahol ritka a pontok eloszlása. A rekonstrukció ilyen jellegű hibája javítható az előző eljáráshoz hasonlóan, azzal a különbséggel, hogy nem a szomszédok címkéjét kell vizsgálni, hanem a szomszédok számát. Adott távolságon belül kevés szomszéd esetén a pontot el lehet dobni. Az eljárás a közeg sűrűségét figyelembe véve élesítheti a szegmensek határát és javíthatja a rekonstrukció minőségét.

4.5.3 Field alapú felosztás

A szegmentáció javítását végző utófeldolgozások során ki kell számítani az ellipszoidok közötti távolságokat. Az ellipszoidok száma miatt nem célravezető minden

ellipszoid esetén minden másik ellipszoidot megvizsgálni, ezért érdemes valamilyen terület alapú megközelítést bevezetni.

A körforgalomra alkalmazott rekonstrukció esetén az ellipszoidok jellemzően a vízszintes sík mentén helyezkednek el, melyet az X és a Z tengelyek feszítenek ki. A pontok területekre való felosztása kizárólag ezen tengelyek mentén történt.

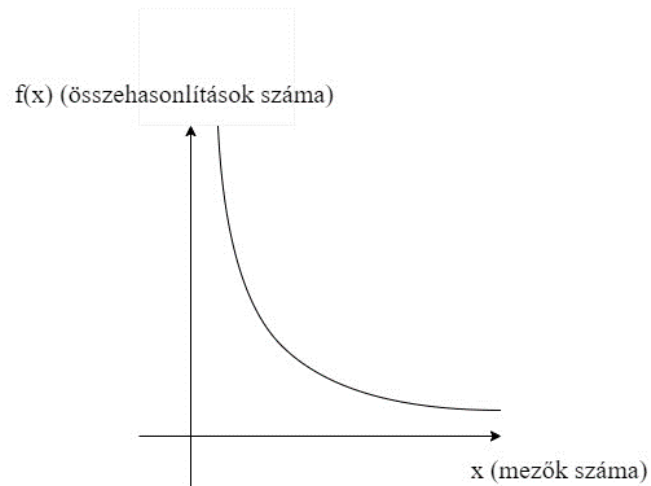


4.27. ábra Az ábra a pontok felosztásának módját ábrázolja..

A felosztáshoz első lépésként meg kellett keresni a legnagyobb és legkisebb X és Z koordinátákkal rendelkező pontokat az általuk lefedett terület meghatározásához. Ezután el kellett dönteni, hogy mekkora területekre legyen felosztva a jelenet. n darab ellipszoid mellett csak a szomszédság vizsgálatát figyelembe véve a simítás, illetve a zajszűrés során n^2 összehasonlítás történik alap esetben. Egyenletes eloszlást feltételezve az összehasonlítások száma x számú mező esetén:

$$f(x) = x \cdot \left(\frac{n}{x}\right)^2.$$

A függvény karakterisztikájából adódóan minél több mező van, annál kevesebb összehasonlítást kell végezni.

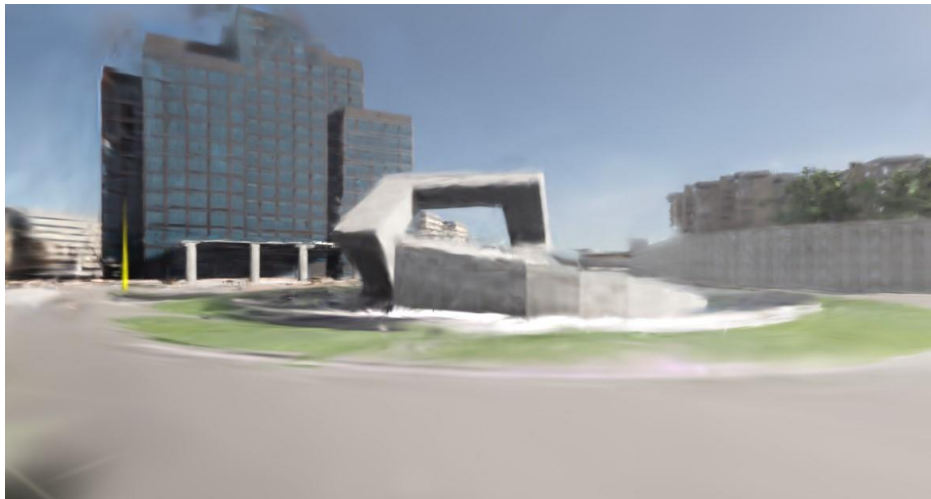


4.28. ábra A függvény jellege pozitív x -ekre értelmezve.

A mezők száma egy kellően magas értéket kapott: 1024. Így a jelenetet 32×32 mező fedti le, melyek átlapolódnak a szomszédok vizsgálatokor használt távolsági küszöbértékkel. Az ellipszoidok a középpontjuk szerint lettek felosztva, így vannak olyan pontok melyekhez egy, kettő vagy négy mezőt azonosító `field_id` lett rendelve. Ezt követően a simítás és zajszűrés során az egyes ellipszoidok szomszédsága csak az adott mezőn belül történt.

5 Eredmények

Ebben a fejezetben a kutatás eredményei kerülnek részletezésre. A feladatok közé tartozott egy példán alkalmazni radiancia mező alapú rekonstrukciós eljárást, valamint a rekonstruált jeleneten szemantikus szegmentációt végezni.

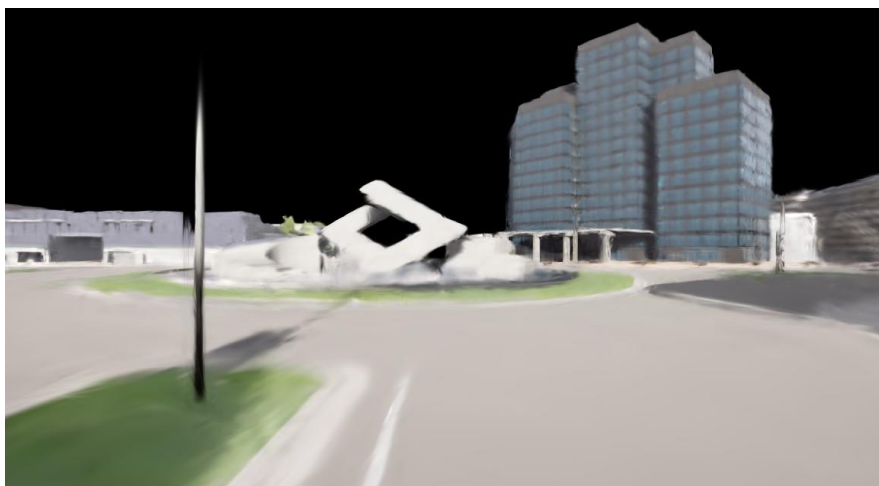


5.1. ábra A körforgalom rekonstrukciója.

5.1 3D rekonstrukció

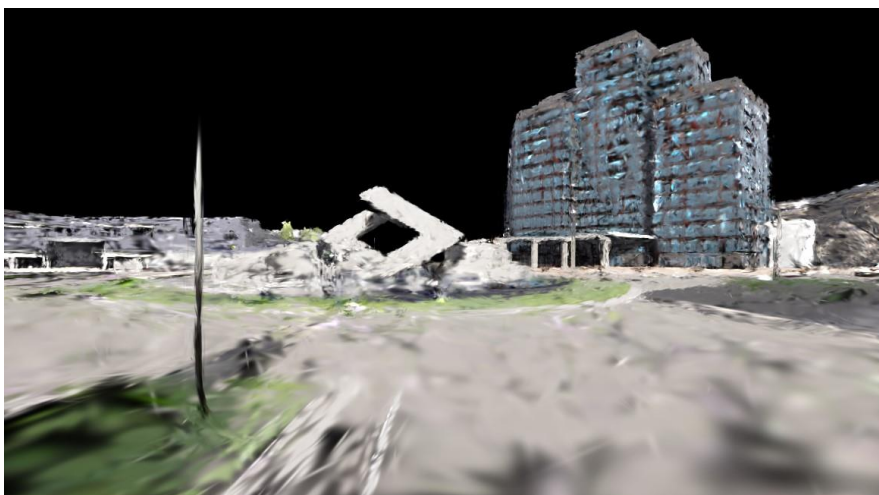
A tanítás során a modell a rekonstrukción renderelt és a valós képek alapján számolt PSNR metrikát tartotta számon. A tanítás 30000 iterációból állt és készített egy mentést a 7000., illetve az utolsó iterációt követően, melyek közül a pontosabb eredmények érdekében az utóbbi lett felhasználva.

A Gaussian Splatting által előállított PLY kiterjesztésű fájl valós idejű interaktív renderelését a SIBR (System for Image-Based Rendering) alkalmazás valósítja meg. A SIBR-ben a felhasználónak lehetősége van navigálni a jelenetben, mind a pozíció mind az orientáció változtatásával. Lehet videót rögzíteni, módosítani a megjelenítés módját vagy akár a Gauss halmazok méretét a skála érték módosításával.



5.2. ábra Az RGB rekonstrukció.

Az ábrán megfigyelhető a radiancia mező alapú rekonstrukciós eljárások egyik legnagyobb előnye, melyre más megközelítések nem, vagy csak részben alkalmasak. Ez nem más, mint a jelenet fényviszonyainak és hatásainak megőrzése. Ennek hatására a rekonstruált jelenet új kamerapozícióból való renderelésekor képes visszaadni az eredeti jelenetben található fényforrásokat, árnyékokat, de olyan effektusokat is, mint például a fényvisszaverődés. Jól kivehető a kép bal oldalán látható villanyoszlop vagy a jobb oldali épület által vetett árnyék az úton.



5.3. ábra Az RGB rekonstrukció csökkentett méretű Gauss halmazokkal.

Az interaktív megjelenítőben lehetőség van a Gauss halmazok skála értékeinek módosítására. A skála csökkentésével a jelenetet alkotó ellipszoidok könnyen kivehetővé válnak. A geometriájuknak köszönhetően az ellipszoidok képesek visszaadni az olyan finomabb részleteket, mint például az épület oldala.



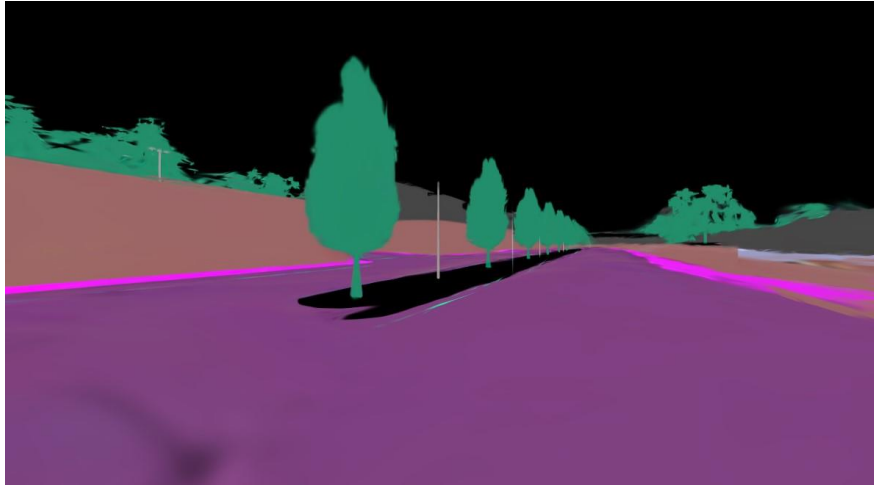
5.4. ábra Az RGB rekonstrukció pontszerűre csökkentett Gauss halmazokkal.

A csökkentés mértékének növelésével gyakorlatilag megkapható az ellipszoidok középpontjaiból álló pontfelhő.

Az adathalmaz RGB képei a hozzájuk tartozó mélységi információ alapján lettek maszkolva. Ennek következménye, hogy a távolra eső objektumok le lettek vágva a képről. Bizonyos helyzetekben viszont előfordul, hogy adott irányban egy távoli objektum nem látszik a képen, de a jármű előrehaladtával, pár képkockával később már megjelenik a képeken a korábban nem látott objektum. Ez a megközelítés pontatlan eredményekhez vezet, a Gaussian Splatting nem fogja tudni megfelelően lefedni a teret és a rekonstrukció elmosódott, zajos lesz.

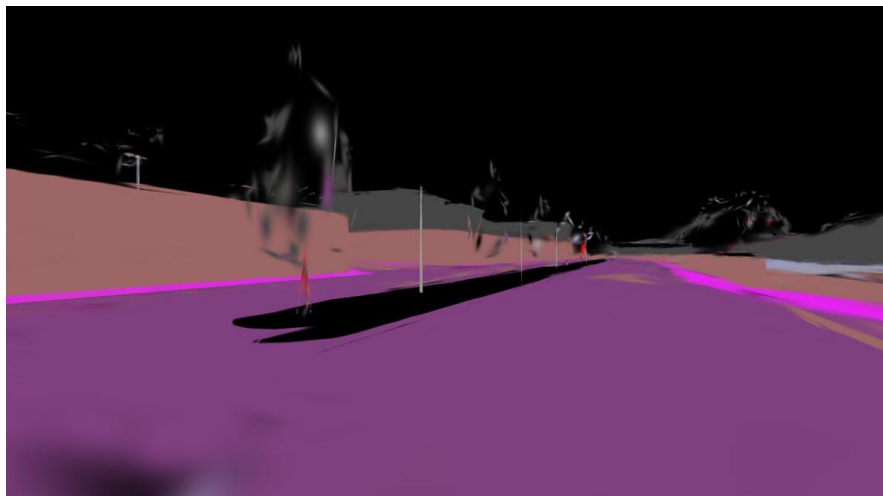
5.2 3D szegmentáció

A háromdimenziós szegmentáció eredménye is hasonlóan látványos. A képen földből kiemelkedő körül járható objektumok, mint például a fák és a villanyoszlopok, szépen egységesen elkülönülnek a környezetüktől.



5.5. ábra 3D-s szegmentáció.

A háromdimenziós szegmentációban a kevés adat miatt az objektumok nincsenek rendesen körbejárva, emiatt a rekonstrukció zajos. Vannak olyan holtterek, ahova a kamera nem nézett soha, viszont az egyes objektumok eltávolítását követően ezek láthatóvá válnak.

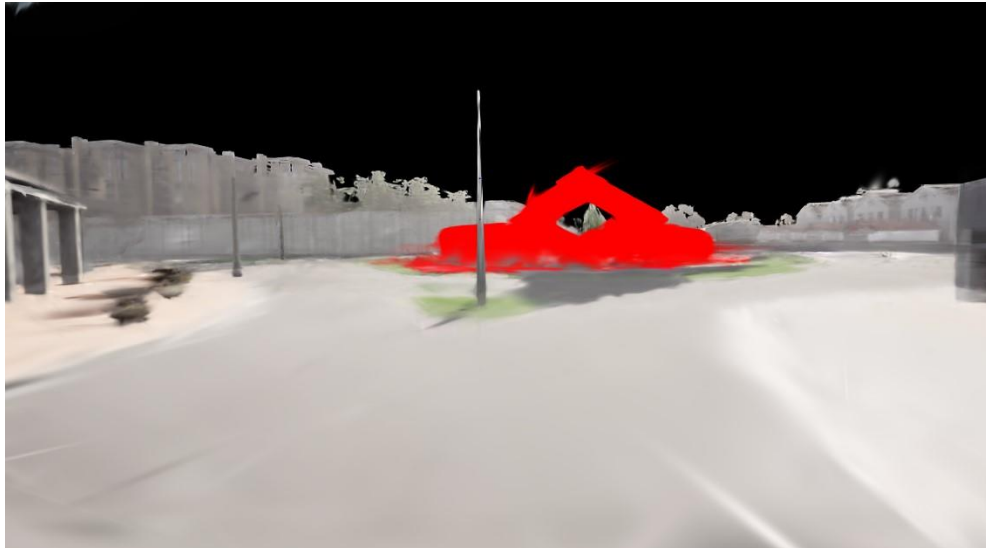


5.6. ábra Növényzethez tartozó ellipszoidok eliminálása a térből.

Ebből látszik, hogy kismennyiségű adat esetén a jeleneten végzett simítás és zajmentesítés elengedhetetlen a tiszta egységes szegmentációhoz.

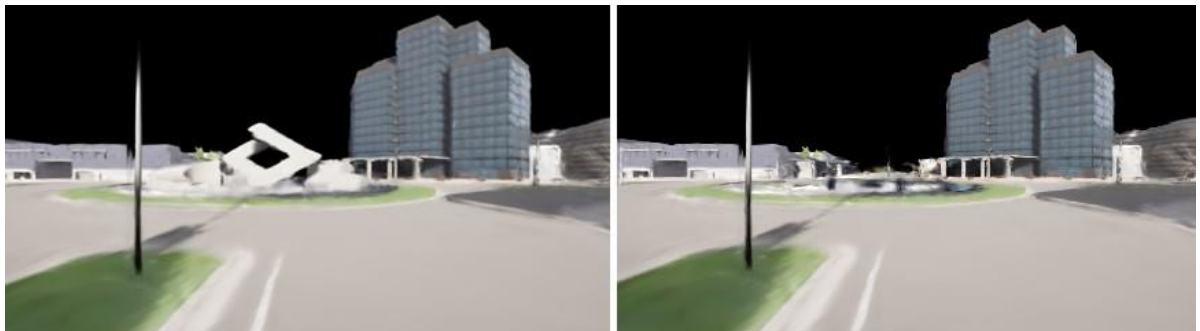
A rekonstruált jelenet körforgalmának közepén elhelyezkedő műalkotáshoz tartozó ellipszoidok a szemantikusan szegmentált képekből álló adatbázis felhasználásával lettek azonosítva, majd a finomítás és zajszűrés segítségével lettek pontosítva.

A szemléltetés miatt a detektált ellipszoidok diffúz színét meghatározó komponensei piros RGB kódolású vektorra (255, 0, 0) lettek módosítva. A nézeti irány alapú színezés kiszámításához szükséges 45 elemű komponens értékei 0-ra lettek állítva.



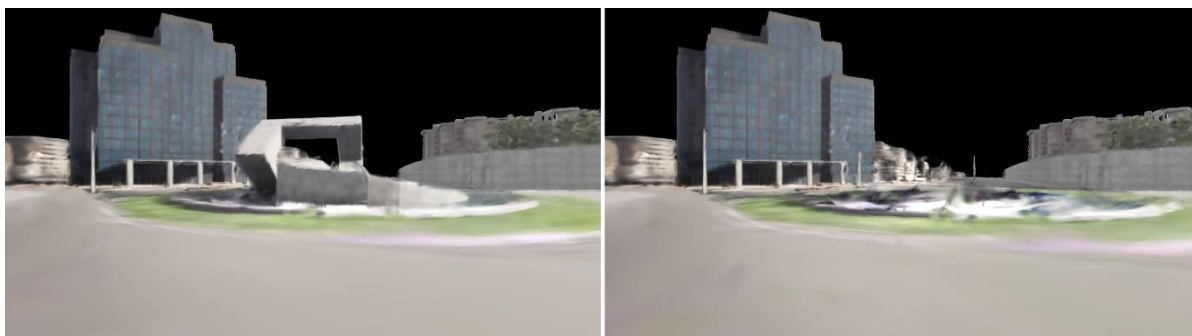
5.7. ábra A szoborhoz tartozó ellipszoidok piros színezéssel.

A szegmentálásra szükség lehet amikor a feladat szempontjából nem releváns objektumok szerepelnek a jelenetben. Ebben az esetben nem elégséges a maszkolásuk, viszont az eltávolításukkal megoldható a probléma.



5.8. ábra A szoborhoz tartozó ellipszoidok kivágása.

A szegmens eltávolítását követően könnyen láthatóvá válnak adott nézőpontból korábban nem látott objektumok.



5.9. ábra A szoborhoz tartozó ellipszoidok kivágása.

A képeken megfigyelhető, hogy a távolabbi részletek homályosabbak, elmosódottabbak. Ez a jelenség annak tudható be, hogy a távoli objektumok kevésbé lettek körbejárva, mint a jelenet közepén elhelyezkedő szobor.



5.10. ábra A szoborhoz tartozó ellipszoidok kivágása.

Érdekesség, hogy az objektum eltávolítását követően az általa vetett árnyék a térben maradt, ami bizonyos esetekben zavaró lehet. NeRF-ök esetén a rekonstrukcióból eltávolított objektum jellemzői a környezetre terjednek ki, annak eltávolítását követően. Ilyen szempontból a Gaussian Splatting-gel egyszerűbben elvégezhető a szegmentálás.

6 A közúti forgalomban alkalmazott rekonstrukció jogi és etikai aspektusai

A neurális radiancia mező alapú rekonstrukciós eljárás egy rendkívül népszerű és gyorsan fejlődő kutatási terület, amely egyre több feladatot képes ellátni. Leggyakrabban objektumok modellezésére használják, de ahogy a diplomamunka során is bebizonyosodott, kisebb kiterjedéssel rendelkező terek rekonstruálására is alkalmasak lehetnek. Ebből kifolyólag könnyen elképzelhető, hogy a jövőben városok/város részletek rekonstruálását fogják végezni.

A technológia képes lehet önvezető járművek navigációs rendszerét támogatni és különböző helyzeteket virtuálisan körbejárni. Számos előnye mellett viszont etikai, jogi és erkölcsi aggályokat is felvet.

A probléma megoldásához egyensúlyra van szükség a technológiai innováció és az egyének jogainak védelme között. Ehhez szükség van az adatvédelmi szabályzatok és az etikai irányelvek, valamint a vonatkozó törvények és előírások betartására. Ezenkívül elengedhetetlen a rendszer résztvevői közötti folytonos és szoros együttműködés a felmerülő kihívások kezeléséhez.

6.1 Fogalmak

A közúti forgalomban alkalmazott neurális radiancia alapú rekonstrukciós eljárások etikai és jogi aspektusainak megvizsgálásához és megértéséhez szükséges definiálni néhány fontos alapfogalmat. [25]

6.1.1 Alapvető jog

Alapvető jogot minden személy örököl, függetlenül a nemzetiségétől, etnikumától, nemétől vagy például a vallásától. Az alapvető jog abban tér el az abszolút jogtól, hogy bizonyos esetekben korlátozhatóak. A korlátozás lehetősége két tényezőtől függ:

- szükségesség
- arányosság

Szükségesség, amikor a jog korlátozására szükség van, valamilyen okból kifolyólag. Ez az ok lehet a jólét fenntartása vagy akár egy másik alapvető jog gyakorlása.

Az arányosság szempontjából figyelembe kell venni az adott jog korlátozásának pozitív és negatív következményeit, melyek közül az előbbinek kell hangsúlyosabbnak lennie.

A magánélet és a személyes adatok védelme az Európai Unióban alapvető jognak minősülnek.

6.1.2 Személyes, érzékeny és bizalmas adat

A személyes adatnak minősül minden olyan információ, amely egy azonosított vagy azonosítható természetes személyhez kapcsolódik. A személy azonosított, ha valamilyen szempontok mentén egy csoportból kitűnt és azonosítható, amennyiben ez lehetséges. [25]

Az azonosító lehet fizikális, pszichológiai, genetikai, mentális, gazdasági, kulturális vagy más jellemzője egy természetes személynek.

A személyes adat egy speciális kategóriája az érzékeny adat. Érzékeny adatnak minősül az a személyes adat, amelyből a személy etnikai hovatartozására, politikai nézeteire, vallási vagy filozófiai irányzatára, genetikájára vagy szexuális irányultságára lehet következtetni. Ezen adatok tárolása és feldolgozása korlátozott.

A bizalmas adat egy tág fogalom, olyan információk összefoglaló neve, melyek kereskedelmi értékkel rendelkeznek. Ezek nyilvánosságra hozatala, módosítása vagy elvesztése az adat birtokosának versenyhelyzetének károsításával jár.



6.1. ábra A személyes, érzékeny és bizalmas adat kapcsolata.

A GDPR (General Data Protection Regulation), általános adatvédelmi szabályzat elsődleges célja, hogy az egyéneknek nagyobb befolyásuk legyen a személyes adataik kezelésére. A GDPR szerint minden személyes adat azonosítónak minősül, melyek képesek azonosítani az érintett személyt.

Az azonosítóknak két típusa létezik:

- direkt és
- indirekt azonosító.

A direkt azonosítók képesek magukban azonosításra. A meglétük nem szükséges, de elégséges feltétele a személyes adatnak. Az indirekt vagy kvázi azonosítók együtt képesek egy személyt azonosíthatóvá tenni.

Az azonosíthatóság kérdésének megválaszolásához a támadó motiváltságát és a támadás sikeres kivitelezésének valószínűségét kell megvizsgálni. Ezek a fogalmak erősen kontextus függőek, nincsenek expliciten előre definiálva a határvonalaik.

6.1.3 Résztvevők

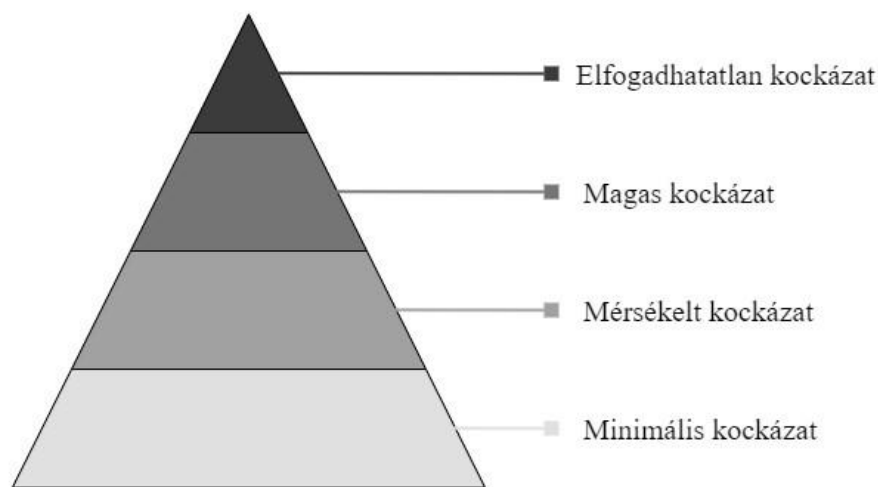
A személyes adatokat gyűjtő, feldolgozó rendszereknek alapvetően három szereplője van: [25]

- Adatkezelő, aki meghatározza a személyes adatok felhasználásának a célját és módját.

- Adatfeldolgozó dolgozza fel az adatot az adatkezelő nevében és felelősséggel tartozik a rendszer másik két szereplőjének.
- Érintett, olyan természetes élő személy, akinek az adatait feldolgozzák.

6.1.4 AI rendszerek kockázati besorolása

Az Európai Unió mesterséges intelligenciájáról szóló törvény tervezete (AI Act) a minimálistól az elfogadhatatlanig négy kockázati szintbe sorolja a mesterséges intelligenciát alkalmazó rendszereket, melyek használatára vonatkozó előírásokat minden tagállam köteles betartani. [26]



6.2. ábra MI rendszerek kockázati szintjei.

Az emberek biztonságát, jogait vagy megélhetését egyértelműen veszélyeztető MI alkalmazásokat az elfogadhatatlan kockázati kategóriába soroljuk, melyek használata tiltott. Ilyenek például a manipulatív rendszerek.

Az elfogadhatatlan kockázatú besorolást a magas kockázati szint követi, amely az emberek jogait csorbítja és biztonságukat mérsékeltten, de veszélyezteti. Ide tartoznak a forgalmat, önvezetést támogató vagy a biometrikus azonosításra alkalmas adatokat felhasználó mesterséges intelligencia rendszerek.

A harmadik kockázati szint a mérsékelt vagy limitált, melyek eleget tesznek a megszabott speciális átláthatósági kötelezettségeknek. Például chat bot-ok esetén a felhasználónak tisztában kell lennie azzal, hogy egy géppel kommunikál.

Az utolsó kategória a minimális vagy kockázattal egyáltalán nem járó mesterséges intelligenciát felhasználó rendszerek. Ide sorolhatóak például az MI kompatibilis

videójátékok vagy spam-szűrők. Jelenleg az MI rendszerek túlnyomó többsége ebbe a kockázati szintbe sorolható.

6.2 Morális és jogi aggályok

A gépi tanuló algoritmusok fejlesztésekor és használatakor gyakran a feladatok megoldására és nagyobb teljesítmény elérésére fókuszálva figyelmen kívül maradnak olyan tényezők, amik morális vagy jogi problémákhoz vezethet. Fontos, hogy a fejlesztési ciklus minden szakaszában megfelelő hangsúlyt kapjanak ezek a kérdések is.

Első pillantásra úgy tűnhet, hogy az alakzatrekonstrukció nem jár kockázattal, de számos szempontot figyelembe kell venni, mielőtt egy ilyen keretrendszert alkalmaznának a közúti forgalomban.

6.2.1 Adatvédelem megsértése

A neurális radiancia mezők tanítása a rekonstruálni kívánt jelenetről készült felvételek alapján történik. A közúti forgalomban alkalmazott NeRF módszerek megkövetelik, hogy a járművekre szerelt kamerák folyamatosan készítsenek képeket, melyek számos, problémához vezető jelenetet megörökíthetnek.

A tanítóhalmaz képei tartalmazhatnak olyan információt, amelyről be lehet azonosítani az érintett személyeket. A GDPR szerint személyes adatnak minősül minden direkt vagy kvázi azonosító. Ilyen például az emberi arc, amely alapján magas valószínűséggel lehet következtetni az érintett személy szexuális beállítottságára is.

Az érzékeny adatok feldolgozásával korlátozhatják a személyes adatok védelméről szóló alapvető emberi jogot. Minden alapvető jog megsértésekor meg kell vizsgálni az eljárást a szükségesség és arányosság szempontjából.

6.2.2 Hozzájárulás

A személyes adatok feldolgozásának módja három fő szempontot kell teljesítsen:

- igazságosság
- törvényesség
- átláthatóság

A személyes adatok feldolgozása akkor minősül törvényesnek, ha az érintett hozzájárulásával történik. Az adatkezelő szükség esetén köteles demonstrálni, hogy érvényes hozzájárulással rendelkezik.

A hozzájárulás az, amikor az érintetteknek valódi választási lehetőséget és felügyeletet biztosítanak azzal kapcsolatban, hogy az adataik milyen módon lesznek feldolgozva.

Valódi választási lehetőség hiányában a hozzájárulás érvénytelen. A beleegyezés visszautasítása esetén sem érheti kár az érintett személyt, aki a hozzájárulás megadása után is könnyen visszavonhatja azt.

Előfordulhat, hogy az utcákon rögzített egyének nem adtak beleegyezést a lefényképezésükhöz, amely kérdéseket vet fel a képek készítésének és felhasználásának etikusságát illetően. A közterületen tartózkodó valamennyi személy hozzájárulásának a megszerzése nem kis kihívást jelentő feladat.

6.2.3 Környezeti hatások

Az radiancia mező alapú rekonstrukció hatalmas mennyiségű adat gyűjtését és feldolgozását követeli meg, amely jelentős mértékben megnöveli a fogyasztást. Nem szabad megfedkezni az adathalmaz képeit előállító járművek által kibocsátott szennyező anyagokról sem.

Az energiafogyasztás megnövekedése pedig előidézhet negatív környezeti hatásokat. A NeRF technológia etikus módon történő felhasználásához nem szabad figyelmen kívül hagyni a rendszer ökológiai lábnyomát sem.

6.2.4 Megfigyelés és nyomon követés

A jelenet rekonstruálásához szükséges adatok magukban vagy más információval kombinálva képesek a jelenetben található személyek vagy járművek azonosítására. Az azonosítás segítségével az eljárás alkalmas lehet a nyomon követési és megfigyelési feladatok ellátására.

A tömeges megfigyelés felhasználható közbiztonság javítására, viszont a visszaélés lehetősége aggodalomra adhat okot. A GDPR törvénytelennek minősíti az olyan érintettek nyomon követését, akik nem adták hozzájárulásukat.

6.2.5 Diszkrimináció és elfogultság

Napjainkban a bőrszín, etnikai származás, nem, genetikai tulajdonságok, vallási vagy politikai nézetek, kor, szexuális irányultság alapján történő megkülönböztetés különösen tiltott. Ennek ellenére a rekonstrukció adatai lehetőséget adnak az azok diszkriminatív módon történő feldolgozására.

A felsorolt megkülönböztetésekre alkalmas védett attribútumok alapján az egyének vagy közösségek profilalkotása illegálisnak minősül.

6.3 Ellenintézkedések

A közúti forgalomban történő rekonstrukcióhoz kapcsolódó, előző fejezetben ismertetett morális és jogi problémák kezeléséhez etikai, politikai és technológiai ellenintézkedések összessége szükséges.

6.3.1 Rugalmas hozzájárulás

Az aggályok enyhítésének egyik módja, egy olyan protokoll fejlesztése, amely a képeken szereplő személyes adatokat és rendszám táblákat valamilyen módon kivehetetlenné tesz és értesíti az érintetteket.

Az érintettek ezután hozzájárulhatnak ahhoz, hogy az adataik feldolgozásra kerülhessenek vagy ne.

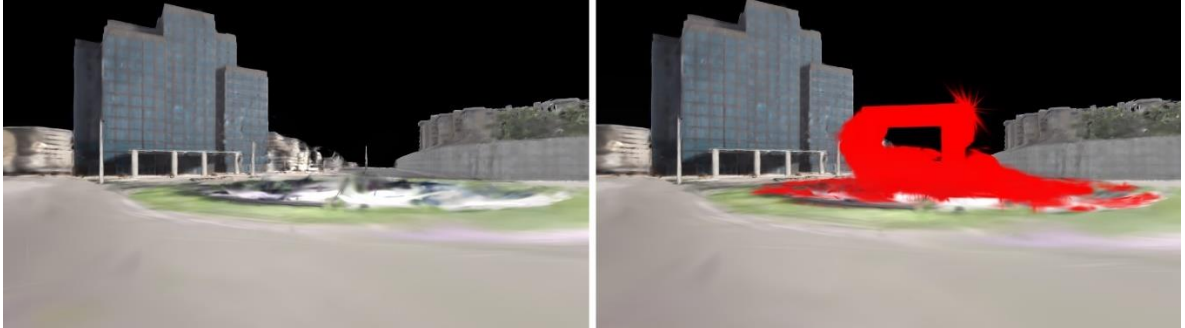
6.3.2 Szegmentáció

Az érzékeny adatok eliminálásának egyik módja a diplomamunka során fejlesztett szegmentációs eljárásra épül, melynek segítségével a nem releváns, zavaró vagy személyes adatok védelmét sértő információt tartalmazó objektumok redukálhatóak a radiancia mező alapú rekonstrukcióból.

A szegmentáció egy kézenfekvő megoldást nyújt az olyan alapvető emberi jogok betartására, mint például a személyes adatok védelme, anonimitás, megelőzi a hozzájárulásból eredő problémákat, valamint az adatok diszkriminatív módon történő felhasználását is ellehetetleníti.

A szegmentáció, illetve a maszkolás elvégezhető a tanítóhalmaz képein is, vagy a jelenet rekonstrukcióján. A maszkolás esetén az adott objektumhoz tartozó pixelek, illetve az objektum háromdimenziós rekonstrukcióját alkotó ellipszoidok színe egy nem

nézeti irányfüggő diffúz színnel van ellátva. Ez a megoldás csak bizonyos védelmet biztosít, mert az alakból egyértelműen lehet következtetni a jármű típusára és ezáltal azonosíthatóvá válhat.



6.3. ábra Piros színnel maszkolt objektum a 3D-s rekonstrukcióban.

A szimulátorban előállított adathalmaz nem tartalmazott gyalogosokat, így a szemléltetés miatt a körforgalom közepén elhelyezkedő objektumot kell személyes adatnak tekinteni. A maszkolást követően csak az objektum alakja kivehető.



6.4. ábra Piros színnel maszkolt objektum a 3D-s rekonstrukcióban.

A maszkolásnál egy biztonságosabb megoldás lehet az szegmentált objektum eliminálása, ami megnehezítené az alak alapján történő járművek azonosítását. Fontos megjegyezni, hogy az azonosítás továbbra sem lehetetlen csak elhanyagolható valószínűségű, ugyanis az eltávolított objektum árnyéka továbbra is megmarad a rekonstrukcióban.

6.3.3 Energia csökkentés

A negatív környezeti hatások gyakran a technológia által megkövetelt nagy mértékű energiafogyasztás következtében lépnek fel, melyek csökkentésére több megoldás is létezik. [27]

- Az adatok tárolását és feldolgozását adatközpontokban végzik, ahol a szerverek hűtése és működtetése sok energiát igényel. Ezek optimalizálásával, a hatékonyság növelésével és megújuló energiaforrások használatával javítható az energia kihasználás.
- Az adathalmaz képeit előállító járművek is jelentősen hozzájárulnak az összesített energiafogyasztáshoz. A körülményeket javítaná, ha olyan járművek alkalmaznak, amelyeknek gazdaságosabb a fogyasztása. Egy másik megoldás lehet, a rendszer résztvevői önkéntes jelleggel készítenék a felvételeket, a járművük használata közben.

7 Összefoglaló

A diplomamunka során megvizsgáltam a háromdimenziós rekonstrukció és a rekonstruált jelenben való szegmentáció lehetőségeit radiancia alapú megközelítéseket felhasználva. Ehhez a szakirodalmi kutatásom elsődleges célja a gépi tanuló algoritmusok tanulmányozása, valamint rekonstrukcióra és szegmentációra alkalmazható hatékony módszerek kutatása és összehasonlítása volt a NeRF-ök területén.

A kiválasztott gyakorlati példa alapját egy szimulátorban generált szintetikus úton előállított adathalmaz képezte. A tanítóhalmaz egy körforgalmat különböző bejárataiból megközelítő önvezető jármű kamerája által készített képekből és a hozzájuk tartozó kamera pozíciójára és orientációjára vonatkozó kamera információkból áll.

A feladatom célja volt a körforgalom rekonstruálása egy radiancia alapú eljárással. A választott módszer a Gaussian Splatting, amely a teret geometriailag ellipszoidoknak megfelelő háromdimenziós Gauss halmazokkal tölti ki. Az eljárás az ellipszoidok középpontján, rotációján, a tengelyek menti skála értékeken, áttetszőségén és a nézeti iránytól függő színén végzett optimalizációval közelíti a jelenet objektumainak felszínét.

Az ismertett technológiai háttérrel rendelkező, szegmentációt végző, ray casting-ra épülő implementáció megvalósítása során és az elért eredményekből levont következtetésekből sokat tanultam a NeRF témaköréről és a Python programozási nyelvről.

Összességében elmondható, hogy a radiancia alapú rekonstrukció és szegmentáció egy működőképes koncepció, amely nem véletlen egy rendkívül felkapott kutatási terület, hetente megjelenő új eredményekkel és eljárásokkal. A kutatás számos továbbfejlesztési lehetőség előtt áll még, melyek az eljárás hatékonyságát tovább növelnék.

7.1 Továbbfejlesztési lehetőségek

Az eredmények javításához elengedhetetlen egy megfelelően generált adathalmaz. A jelenleg használt adathalmaz RGB képei a hozzájuk tartozó mélységi információ alapján vannak maszkolva. Ennek következtében előfordulnak olyan objektumok, melyek néha láthatóak és néha nem, ami problémát okozhat a radiancia

mező optimalizálásakor, ugyanis az inkonzisztens megjelenítést a modell zajjal próbálja kompenzálni. Célravezetőbb lenne egy olyan adathalmaz, amelyen csak az éghez tartozó pixelek vannak kivágva. Az adathalmaz képeinek változatossága is jelentősen javítaná az elért eredményeket, viszont a feladat realiztikusságának megőrzése érdekében ez nem kivitelezhető.

A szegmentáláshoz használt színek eloszlása nem optimális. Lehet, hogy emberi szem színérzékenysége miatt az egyes szegmens színek látványosan elkülönülnek egymástól, viszont a gépi látás csak a kapott adatokra támaszkodik. Pontosabb szegmentálást eredményezne, ha a szegmensek színei az RGB térben mérve egymástól a lehető legmesszebb helyezkednének el.

A címkézés elvégezhető ray casting helyett rasterizációval is. Ehhez felhasználható a CUDA kóddal már megfelelően kioptimalizált Gaussian Splatting implementáció kis módosításokkal. Inicializációkor a maszkolt RGB képek alapján létrehozott rekonstrukciót kapná, melyet a szemantikusan szegmentált képek felhasználásával tanítana tovább úgy, hogy az ellipszoidok geometriáját befagyasztva kizárólag a színüket módosítsa. A színhez egy plusz attribútummal kell ellátni a Gauss halmazokat, amely egy olyan hosszú feature vektor ahány különböző szegmens fordul elő a jelenetben. Minden értéke az adott osztályhoz való tartozásának valószínűségét tükrözi. A jelenet renderelésekor a pixelek színét az ellipszoidok ezen attribútuma határozza meg és a tanítás során ezek a vektorok vannak optimalizálva. Fontos, hogy az eredeti implementáció hiba számításához használt függvényt is le kell cserélni egy erre a feladatra alkalmas veszteségfüggvényre, mint például a kereszt-entrópia.

Utófeldolgozáshoz használt simítás és zajszűrés statikus jelleggel működik. Egy előre meghatározott távolság érték alapján vannak meghatározva a szomszédságok, amely nem veszi figyelembe a közeg sűrűségét. Az adaptív jellegű működés bevezetésével a sűrűbben, egymáshoz közelebb elhelyezkedő Gauss halmazok esetén finomabban történne a simítás és durvábban a ritkább közegekben. Ennek következtében az utófeldolgozás alatt nem vesznének el a jelenet apróbb részletei.

Irodalomjegyzék

- [1] Medium: *3D Reconstruction Technology Market Size, Growth, Forecast 2023-2030*, <https://medium.com/@nettieboyle84/3d-reconstruction-technology-market-size-growth-forecast-2023-2030-89e272a5c6c8> (Elérés időpontja: 2023.11.29.)
- [2] Theo Moons, Luc Van Gool and Maarten Vergauwen (2010): *3D Reconstruction from Multiple Images Part 1: Principles*
- [3] 3DSCANEXPERT.: https://3dscanexpert.com/wp-content/uploads/realitycapture_review_header.jpg (Elérés időpontja: 2023.11.29.)
- [4] Balázs Csanád Csáji (2001): *Approximation with Artificial Neural Networks*, <https://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.101.2647&rep=rep1&type=pdf> (Elérés időpontja: 2023.11.29.)
- [5] Microsoft Azure: *Mi a gépi tanulás?*, <https://azure.microsoft.com/hu-hu/resources/cloud-computing-dictionary/what-is-machine-learning-platform/> (Elérés időpontja: 2023.11.30.)
- [6] Towards Data Sciences: *Neural Network: The Essence of Artificial Neural Network*, <https://towardsdatascience.com/neural-network-the-essence-of-artificial-neural-network-c605eb32de56#:~:text=The%20structure%20of%20an%20artificial,signals%20received%20by%20the%20dendrite> (Elérés időpontja: 2023.11.30.)
- [7] Dr. Gyires-Tóth Bálint: *Backpropagation*, Deep learning a gyakorlatban Python és LUA alapon (BMEVITMAV45)
- [8] Dr. Boglár Bence Márton: *Neurális hálózatok 1*, Gépi tanulás (BMEVIMIMA05)
- [9] Dr. Juhász Imre: *Görbék és felületek modellezése*, <https://gyires.inf.unideb.hu/KMITT/c10/ch07.html> (Elérés időpontja: 2023.12.05.)
- [10] Ben Mildenhall, Pratul P. Srinivasan, Matthew Tancik, Jonathan T. Barron, Ravi Ramamoorthi, Ren Ng: *NeRF: Representing Scenes as Neural Radiance Fields for View Synthesis*, <https://arxiv.org/pdf/2003.08934.pdf> (Elérés időpontja: 2023.12.05.)
- [11] Towards Data Sciences: *Paper Explained – Instant Neural Graphics Primitives with a Multiresolution Hash Encoding*, <https://towardsdatascience.com/paper-explained-instant-neural-graphics-primitives-with-a-multiresolution-hash-encoding-8e5a05865378> (Elérés időpontja: 2023.12.06.)
- [12] Reshot AI: *3D Daussian Splatting*, <https://www.reshot.ai/3d-gaussian-splatting> (Elérés időpontja: 2023.12.06.)
- [13] Medium: *3D Gaussian Splatting: Real-Time Rendering of Photorealistic Scenes*, <https://medium.com/axinc-ai/3d-gaussian-splatting-real-time-rendering-of-photorealistic-scenes-f7f1a47f060> (Elérés időpontja: 2023.12.06.)
- [14] Suhani Vora, Noha Radwan, Klaus Greff, Henning Meyer, Kyle Genova, Mehdi S. M. Sajjadi, Etienne Pot, Andrea Tagliasacchi, Daniel Duckworth: *NeSF: Neural Semantic Fields for Generalizable Semantic Segmentation of 3D Scenes*, <https://openreview.net/pdf?id=ggPhsYCsm9> (Elérés időpontja: 2023.12.06.)

- [15] Yichen Liu, Benran Hu, Junkai Huang, Yu-Wing Tai, Chi-Keung Tang: *Instance Neural Radiance Field*, <https://arxiv.org/pdf/2304.04395.pdf> (Elérés időpontja: 2023.12.06.)
- [16] TIOBE: *TIOBE Index for November 2023*, <https://www.tiobe.com/tiobe-index/> (Elérés időpontja: 2023.12.07.)
- [17] GitHub Blog: *Why Python keeps growing, explained*, <https://github.blog/2023-03-02-why-python-keeps-growing-explained/> (Elérés időpontja: 2023.12.07.)
- [18] NVIDIA: *PyTorch*, <https://www.nvidia.com/en-us/glossary/data-science/pytorch/> (Elérés időpontja: 2023.12.07.)
- [19] Medium: *10 most common Maths Operation with Pytorch Tensor*, <https://medium.com/@anoorasfatima/10-most-common-maths-operation-with-pytorchs-tensor-70a491d8cafd> (Elérés időpontja: 2023.12.07.)
- [20] NumPy: *What is NumPy?*, <https://numpy.org/doc/stable/user/whatisnumpy.html> (Elérés időpontja: 2023.12.07.)
- [21] Qian-Yi Zhou, Jaesik Park, Vladlen Koltun: *Open3D: A Modern Library for 3D Data Processing*, <https://arxiv.org/pdf/1801.09847.pdf> (Elérés időpontja: 2023.12.07.)
- [22] ELTE: *A CUDA programozási nyelv*, <http://nyelvek.inf.elte.hu/leirasok/CUDA/index.php?chapter=1> (Elérés időpontja: 2023.12.07.)
- [23] CARLA Simulator: *CARLA*, https://carla.readthedocs.io/en/latest/start_introduction/#:~:text=It%20grants%20access%20to%20any,simulator%20within%20other%20learning%20environments. (Elérés időpontja: 2023.12.07.)
- [24] Zhu Morui, Szász Erik, Dr. Szántó Mátyás, Dr. Szemenyei Márton: *CARLA2NeRF*, <https://github.com/VisionLab-IIT/CARLA2NeRF> (Elérés időpontja: 2023.12.08.)
- [25] Pejó Balázs: *Legal*, Személyes adatok védelme (BMEVIHIAV35)
- [26] European Commission: *Regulatory framework proposal on artificial intelligence*, <https://digital-strategy.ec.europa.eu/en/policies/regulatory-framework-ai> (Elérés időpontja: 2023.12.15.)
- [27] Utilities One: *The energy Consumption of Data Centers Environmental Concerns*, <https://utilitiesone.com/the-energy-consumption-of-data-centers-environmental-concerns> (Elérés időpontja: 2023.12.15.)

Függelék

1. Background

This document concerns MSc theses dealing with the design, development, deployment and/or use of AI-based systems or techniques.¹

Examples. The uses of AI in military environments could include: decision and planning support, collaborative combat, cybersecurity and digital influence, logistics and operational, robotics and autonomy, support services and target identification and engaging.

The manner in which an AI solution is deployed or used may change the ethical characteristics of the system. It is therefore important to ensure ethics compliance even in cases where your thesis does not develop itself an AI based system/technique.

A proposal for a regulation laying down harmonised rules on AI (Artificial Intelligence Act) is currently pending adoption by the EU legislator. This regulation, when it enters into force, will have effect on all AI-related activities. Before its adoption and entry into force, we strongly encourage the use of the Assessment List for Trustworthy Artificial Intelligence (ALTAI) to develop procedures to detect, assess the level and address potential risks.

2. How to address the issues

Your activities must comply with ethics, notably:

- highest ethical standards;
- applicable international, EU and national law (in particular, the principles and values enshrined in the EU Charter of Fundamental rights and the EU Treaties).

This requires specific ethically-focused approach during the design, development, deployment, and/or use of AI-based solutions.

Any use of AI systems or techniques should be clearly described in the thesis and you must demonstrate their technical robustness and safety (they must be dependable and resilient to changes).

The approach must be built upon the following key prerequisites for ethically sound AI systems²:

Human agency and oversight — AI systems must support human autonomy and decision-making, enabling users to make informed autonomous decisions regarding the AI systems. This is particularly relevant for AI systems that can affect human behaviour by guiding, influencing or supporting humans in decision-making processes (*e.g. recommendation systems, predictive algorithms, disease diagnosing tools*). The right to human agency should be safeguarded by setting up appropriate oversight mechanisms to prevent possible adverse effects and uphold human autonomy.

- AI systems must not subordinate, coerce, deceive or manipulate people, and should not create attachment or stimulate addiction.
- The development of lethal autonomous weapons without the possibility of meaningful human control over selection and engagement decisions when carrying out strikes against humans are prohibited.³

Privacy and data governance — AI systems must guarantee privacy and data protection throughout the system's lifecycle. The principles of privacy by design and by default must be taken into account in the process of designing, developing, selecting and using AI. The quality, integrity and security of data should be rigorously checked and adequately managed. Data minimisation and data protection should never be leveraged to hide or obscure bias, and these should be addressed without harming privacy rights.

¹This assessment is based on EU Grants – How to complete your ethics self-assessment: V2.0 – 13.07.2021

²As identified by the Independent High Level Expert Group on AI set up by the European Commission in the Ethics guidelines for trustworthy AI. See also: Assessment List for Trustworthy Artificial Intelligence (ALTAI) for self-assessment

³Article 10(6) of EDF Regulation 2021/697.

Transparency — All data sets and processes associated with AI decisions must be well communicated and appropriately documented. AI systems must be explainable and open in the communication about their limitations. The principle of transparency is closely linked to the principles of tractability and explicability and facilitates the implementation of human agency, data governance and human oversight. It includes all elements relevant to an AI system (*e.g. the data, the system and the processes by which it is designed, deployed and operated*).

Fairness, diversity and non-discrimination — Best possible efforts should be made to avoid unfair bias (*e.g. stemming from the used data sets or the ways the AI is developed*). AI systems should be user-centric and whenever relevant, designed to be usable by different types of end-users with different abilities. AI systems should avoid functional bias by offering the same level of functionality and benefits to end-users with different abilities, beliefs, preferences and interests, to the extent possible. Inclusion and diversity must be enabled during the entire life cycle of the AI system. Ensure objectivity and inclusiveness of the developed systems/approaches.

Societal and environmental well-being — The impact of the developed and/or used AI system/technique on the individual, society and environment must be carefully evaluated and any possible risk of harm must be avoided. Increased vigilance is needed for solutions that may potentially have significant negative social or environmental impact. Sustainability and ecological responsibility of AI systems should be encouraged, and research should be fostered into AI solutions addressing areas of global concern, for instance the Sustainable Development Goals. Overall, AI should be used to bring positive transformative changes to the society, environment or the economy. AI systems should serve to maintain and foster democratic processes and respect the plurality of values and life choices of individuals; they must not undermine democratic processes, human deliberation or democratic voting systems or pose a systemic threat to society at large.

Examples of social impact: *negative impact on human rights, democratic processes, functioning of media and mass communication, labour and labour market; educational choices; consumer interests and consumer protection, social cohesion and social exclusion, cultural diversity and cultural heritage, international co-operation, mass surveillance.*

Accountability — Requires that the actors involved in the development or operation take responsibility for the way that these applications function and for the resulting consequences. Accountability requires and presupposes certain levels of transparency as well as oversight. To be held to account, developers or operators of AI systems must be able to explain how and why a system exhibits particular characteristics or results in certain outcomes.

This implies that, amongst others, the developed/used AI solutions must:

- ensure that people are aware they are interacting with an AI system and are informed (in a language and terms understandable by all) about its abilities, limitations, risks and benefits – the manner in which this is done must be described in the thesis;
- The manner in which information is provided should not depend on particular educational backgrounds, technical knowledge, or other skills which cannot be assumed of all people.
- prevent possible limitations on human rights and freedoms (*e.g. freedom of expression, access to information, freedom of movement etc.*);
- not be designed in a way that may lead to objectification, dehumanization, subordination, discrimination, stereotyping, coercion, manipulation of people or creation of attachment or addiction;
- be able to demonstrate compliance with the principles of data minimisation and privacy by design and by default when processing personal data – the principles of lawfulness, transparency and fairness of the data processing must be respected at all times; for more information, please consult the Guidance on ethics and data protection in research projects;
- must be designed in a way to avoid bias in both input data and algorithm design – the system should be able to prevent potential discrimination, stigmatisation or any other adverse effects on the individual related to the use of the developed/deployed AI system/technique – the manner in which this is done must be described in your thesis;

- must address the potential impact on the individual, society or the environment. An evaluation of the potential negative individual, societal and/or environmental impacts must be carried out and be included in the thesis along with the measures to be set in place to mitigate any potential adverse effect;
- The ethics risk assessment and risk mitigation measures must cover the design, development, deployment and post-deployment phases.
- must not reduce the safety and wellbeing of the individuals. Whenever relevant, the safety of the developed/used systems must be demonstrated in the thesis;
- should be developed in a way that enables human oversight (human-in-the-loop, human-on-the-loop, human-in-command), traceability and auditability – whenever possible, explanation on how decisions are taken by the developed/used AI along with the logic behind it should be provided.

For further detailed requirements, please consult the Assessment List for Trustworthy Artificial Intelligence (ALTAI).

At the design and development stage, the implementation of the key requirements for ethically sound AI systems can be ensured by adopting the ‘ethics by design’ approach. The latter is aimed at preventing ethics issues from occurring by integrating ethical values based requirements into the design of the developed/used AI solution. The ethics by design approach will greatly facilitate your ethics compliance. For more information, please consult Guidelines on ethics by design for AI.

Some types of objectives, methodologies, system architecture or design may be inherently problematic (due to serious ethical non-compliance). This is the case for instance for AI systems that risk to:

- limit human rights, subordinate, deceive or manipulate people, violate bodily or mental integrity, create attachment or addiction, or hide the fact people are interacting with an AI system;
- cause people to be disadvantaged socially or politically, reduce the power that they have over their lives, or result in discrimination, either by the system, or by the way it will be used;
- cause people to suffer physical, psychological or financial harm, cause environmental damage, or significantly damage social processes and institutions (*for example, by contributing to misinformation of the public*).

3. Ethics issues checklist

Activity	Y/N	Information to be provided in the thesis	Documents to be provided as appendix
<i>Does this activity involve the development, deployment and/or use of Artificial Intelligence-based systems?</i>	Y	1) Explanation as to how the participants and/or end-users will be informed about: their interaction with an AI system/technology (if relevant); the abilities, limitations, risks and benefits of the proposed AI system/technique; the manner in which decisions are taken and the logic behind them (if relevant). 2) Details on the measures taken to avoid bias in input data and algorithm design. 3) Explanation as to how the respect to fundamental human rights and freedoms (e.g. human autonomy, privacy and data protection) will be ensured. 4) Detailed explanation on the potential ethics risks and the risk mitigation measures.	1) Detailed risk assessment accompanied by a risk mitigation plan (if relevant). These must cover the development, deployment and postdeployment phases. 2) Copies of ethics approvals (if relevant).

Activity		Y/N	Information to be provided in the thesis	Documents to be provided as appendix
Could the AI based system/technique potentially stigmatise or discriminate against people (e.g. based on sex, race, ethnic or social origin, age, genetic features, disability, sexual orientation, language, religion or belief, membership to a political group, or membership to a national minority)?		Y	1) Detailed explanation of the measures set in place to avoid potential bias, discrimination and stigmatisation.	
Does the AI system/technique interact, replace or influence human decision-making processes (e.g. issues affecting human life, health, well-being or human rights, or economic, social or political decisions)?		N	1) Detailed explanation on how humans will maintain meaningful control over the most important aspects of the decision-making process. 2) Explanation on how the presence/role of the AI will be made clear and explicit to the affected individuals.	1) Information sheets / Template Informed consent forms (if relevant).
Does the AI system/technique have the potential to lead to negative social (e.g. on democracy, media, labour market, freedoms, educational choices, mass surveillance) and/or environmental impacts either through intended applications or plausible alternative uses?		Y	1) Justification of the need for developing/using this particular technology. 2) Assessment of the ethics risks and detailed description of the measures set in place to mitigate the potential negative impacts during the research, development, deployment and postdeployment phase.	1) For serious and/or complex cases: Algorithmic impact assessment/human right assessment. These must cover the development, deployment and postdeployment phases.
Does this activity involve the use of AI in a weapon system?		N		
I F Y E S	Is it possible to establish which specific function/functions are automated/autonomous in the weapon system?	-	1) Justification for the need. 2) Detailed explanation on how humans will maintain meaningful control.	1) Detailed overview of the automated functions.
	If the weapon system has AI-enabled functions, could these functions render the weapon system indiscriminate?	-	1) Justification for the need. 2) Detailed explanation on how humans will maintain meaningful control.	1) Description of the automated navigation and its ability to discriminate targets.
	Does the design include the possibility of an autonomous mode for selfprotection? If yes, can the system reliably distinguish between targets (threats) and non-targets?	-	1) Justification for the need. 2) Detailed explanation on how humans will maintain meaningful control.	1) Detailed explanation on how the potential ethics algorithmic assessment will work.
Does the AI to be developed/used in the project raise any other ethical issues not covered by the questions above (e.g., subliminal, covert or deceptive AI, AI that is used to stimulate addictive behaviours, lifelike humanoid robots, etc.)?		N	1) Detailed explanation on how the potential ethics issues will be addressed and the measures set in place to mitigate ethics risks.	1) Detailed risk assessment accompanied by a risk mitigation plan. These must cover the development, deployment and postdeployment phases.

In case it is not possible to identify the potential risks related to the AI system/techniques at this stage, describe the procedure you intend to use to detect, assess and address potential ethics issues (or explain why such a procedure is not needed).

4. Background documents & further reading

Artificial intelligence

1. Proposal for an EU Regulation on a European approach for Artificial Intelligence
2. Ethics guidelines for trustworthy AI, Independent High Level Expert Group on AI
3. Assessment List for Trustworthy Artificial Intelligence (ALTAI) for self-assessment
4. Guidelines on ethics by design/operational use for Artificial Intelligence
5. EU White Paper on Artificial intelligence

Defence

1. United Nations, Convention on Certain Conventional Weapons, Group of Governmental Experts, Lethal Autonomous Weapon Systems, CCW GGE LAWS 11 guiding principles

Ethical use of generative AI in academic writing

1. The ethics of using AI in research and scientific writing, Paperpal. November 16, 2022
2. Guidance on the Appropriate Use of Generative Artificial Intelligence in Graduate Theses, University of Toronto, October 12, 2023