

УНИВЕРЗИТЕТ У БЕОГРАДУ МАТЕМАТИЧКИ ФАКУЛТЕТ



МАСТЕР РАД
на катедри за
Рачунарство и информатику

на тему

Примена неуронских поља зрачења у рендеровању

Коста Грујчић

Београд, 20. март 2023.

Ментор:

проф. др Младен Николић
Универзитет у Београду, Математички факултет

Чланови комисије:

проф. др Младен Николић
Универзитет у Београду, Математички факултет

проф. др Младен Николић
Универзитет у Београду, Математички факултет

проф. др Младен Николић
Универзитет у Београду, Математички факултет

Датум одбране: 20. март 2023.

Ceojoj majyu

Наслов мастер рада: Примена неуронских поља зрачења у рендеровању

Резиме:

Кључне речи: машинско учење, неуронска поља, рендеровање

Списак слика

2.1	Интерпретација тачкастог модела камере	8
4.1	Илустрација NeRF модела према [1]	16
4.2	Поређење NeRF и mip-NeRF модела према [2]	18
4.3	Приказ интегралне Фуријеизације према [2]	22
4.4	Приказ сферних хармоника и сферне Фуријеизације према [3]	24
4.5	Поређење mip-NeRF и Ref-NeRF архитектура	26
5.1	По један поглед из сваког скупа података	28
6.1	Приказ резултата LEGO скупа података	33
6.2	Приказ резултата MATERIALS скупа података	34
6.3	Приказ резултата DRUMS скупа података	35
6.4	Приказ резултата CHAIR скупа података	36
6.5	Приказ резултата SHIP скупа података	37

Списак табела

6.1 Резулати за LEGO скуп података	30
6.2 Резулати за MATERIALS скуп података	30
6.3 Резулати за DRUMS скуп података	31
6.4 Резулати за CHAIR скуп података	31
6.5 Резулати за SHIP скуп података	31
6.6 Време обучавања за сваки модел	31

Садржај

1 Увод	3
2 Основни појмови рачунарске графике	5
2.1 Светлост и боја	5
2.2 Камера	7
2.2.1 Тачкасти модел камере	7
2.3 Рендеровање	9
2.3.1 Запреминско рендеровање	9
3 Основни појмови машинског учења	11
3.1 Неуронске мреже	12
4 Неуронска поља зрачења	15
4.1 NeRF	15
4.1.1 Параметризација	16
4.1.2 Рендеровање	16
4.1.3 Обучавање	17
4.2 Mip-NeRF	18
4.2.1 Интегрална Фуријеизација	19
4.2.2 Параметризација и обучавање	22
4.3 Ref-NeRF	23
4.3.1 Сферна Фуријеизација	23
4.3.2 Параметризација и обучавање	25
4.3.3 Израчунавање вектора нормале	25
5 Скупови података	27
6 Експерименти	29
6.1 Метрике	29
6.2 Резултати	30
6.3 Време обучавања	31
6.4 Квалитет резултата	33
6.5 Детаљно поређење	37
6.5.1 NeRF и mip-NeRF	38
6.5.2 NeRF и Ref-NeRF	38
6.5.3 mip-NeRF и Ref-NeRF	38

Глава 1

Увод

Рендеровање је проблем стар колико и рачунарска графика. Постоји небројано пуно различитих алгоритама, приступа, техника и хеуристика за решавање овог важног проблема, а управо је рендеровање покретач низа хардверских унапређења у деценијама иза нас. С друге стране, машинско учење и вештачка интелигенција уопште, су изузетно актуелна област истраживања и примене. Употреба машинског учења у домену рачунарског вида је довела до неслуђених успеха и унапређења чак и у свакодневном животу. Тако се природно дошло до *неуронског рендеровања*, односно, рендеровања сцене употребом неуронских мрежа које су засигурно најпознатија фамилија модела машинског учења.

Неуронска поља зрачења су један од водећих новитета машинског учења, нарочито у оквиру подобласти рачунарског вида. У питању су врсте неуронских мрежа посебно намење за учење векторских поља. Будући да је рендеровање центар интесирања, реч је о специфичној врсти поља којима се описује боја сваке тачке. Имају широку, пре свега, генеративну примену и представљају изузетно живу научну тему. Отуда се и јавила идеја темељног заснивања ове врсте модела као и њиховог упоређивања на основу добијених резултата.

Како је у рачунарству од велике важности и пратећа имплементација, значај овог рада можемо видети и у чињеници да су сви модели у потпуности имплементирани. То за последицу има једноставан поступак обучавања као и употребе.

Један од споредних циљева овог рада је превођење израза који су се безразложно одомаћили у српској научној заједници као англицизми. С друге стране, скраћенице остају у извornом облику и то на латиници, што читаоцу пружа могућност за непосредно претраживање навода.

Битно је напоменути да слике овог рада нису ауторске, већ су преузете из одговарајућих научних радова који су јавно доступни. Њихов изворни облик је задржан, уз превођење појединих делова уколико је за тим било потребе.

Глава 2

Основни појмови рачунарске графике

2.1 Светлост и боја

Светлост представља електромагнетно зрачење чија је таласна дужина у сегменту од око 350 до 700 nm, које побуђује визуелни систем човека. То значи да људи *не виде* светлост осталих таласних дужина, тако да ћемо убудуће под видљивим спектром светлости мислити управо на овај, видљив човеку.

Потребно је увести физичке величине којима се могу квантификовати основна физичка својства светлости. Прво ћемо дефинисати радиометријске величине, а онда одговарајуће фотометријске.

Дефиниција 2.1.1. *Фотон је квант електромагнетног зрачења. Енергија фотона је:*

$$E = \frac{hc}{\lambda} [\text{J}],$$

где је h Планкова константа, c брзина светлоти, а λ таласна дужина фотона.

Дефиниција 2.1.2. *Укупна енергија зрачења извора зрачења је:*

$$Q_e = \int_{S^2} E d\Omega [\text{J}].$$

Дефиниција 2.1.3. *Флукс зрачења је укупна енергија зрачења која доспе на неку површину по јединици времена:*

$$\Phi_e = \frac{\partial Q_e}{\partial t} [\text{W}].$$

Дефиниција 2.1.4. *Означеност је укупан флукс зрачења по јединици површине:*

$$E_e = \frac{\partial \Phi_e}{\partial A} \left[\frac{\text{W}}{\text{m}^2} \right].$$

Дефиниција 2.1.5. *Јачина зрачења је укупан флукс зрачења у неком смеру по јединичном просторном углу:*

$$I_{e,\Omega} = \frac{\partial \Phi_e}{\partial \Omega} \left[\frac{\text{W}}{\text{sr}} \right].$$

Дефиниција 2.1.6. Зрачење је је укупан флукс зрачења у неком смеру по јединици површине и јединичном просторном углу:

$$L_{e,\Omega} = \frac{\partial^2 \Phi_e}{\partial \Omega \partial A} \left[\frac{\text{W}}{\text{sr} \cdot \text{m}^2} \right].$$

Спектралне величине се дефинишу у односу на таласну дужину. На пример, спектрално зрачење је $L_{e,\Omega,\lambda} = \partial L_{e,\Omega}/\partial \lambda$. Аналогно и за остале.

Након дефинисања мноштва физичких величина, коначно долазимо и до фотометријских, или како се још називају и *визуелне*. Кључна разлика у односу на радиометријске, или како се још називају и *енергетске*, је што се у овом случају у обзир узима и спектрална осетљивост посматрача. То одговара интуитивном поимању светлости – човек светлост разликује на основу боје, што је у директној кореспонденцији са таласном дужином.

У основи спектралне зависности посматрача је *функција релативне светлосне осетљивости* V . Помоћу ове функције изражавамо просечну осетљивост човека на светлост одређене таласне дужине. Просечна је у смислу да може варирати у популацији, али представља врло добру апроксимацију у општем случају, поготово имајући у виду да је стандардизована од стране Међународног комитета за осветљење.

Дефиниција 2.1.7. Светлосни флукс је укупна енергија која протекне кроз неку површину у јединици времена:

$$\Phi_v = K \int_0^\infty \Phi_{e,\lambda} V(\lambda) d\lambda [\text{lm}].$$

За вредност онстанте K из 2.1.7 се узима $683.002 \frac{\text{lm}}{\text{W}}$. Реч је о још једној примеру стандардизације у области фотометрије.

Дефиниција 2.1.8. Јачина светлости је укупна снага коју емитује извор светлости у одређеном смеру по јединичном просторном углу:

$$I_v = K \int_0^\infty I_{e,\lambda} V(\lambda) d\lambda [\text{cd}].$$

Дефиниција 2.1.9. Осветљеност је укупан светлосни флукс на некој површини:

$$E_v = K \int_0^\infty I_{e,\lambda} E_{e,\lambda} V(\lambda) d\lambda [\text{lx}].$$

Дефиниција 2.1.10. Сјајност је укупан светлосни флукс који напушта, пролази или пада на површину по јединичном просторном углу и по ортогоналној пројекцији јединичне површине:

$$L_v = \frac{\partial^2 \Phi_v}{\partial \Omega \partial A \cos \theta} \left[\frac{\text{cd}}{\text{m}^2} \right].$$

Сјајност је једина фотометријска величина коју човек непосредно опажа. Она представља мерило за субјективни утисак о мањој или већој сјајности светлеће или осветљене површине.

Већ смо поменули да људи разликују неке таласне дужине светlostи, односно виде светlost одређене боје. Ту способност нам дају три врсте чепића који се налазе у жутој мрљи мрежњаче ока. Према томе, кроз ове рецепторе, наш мозак прима свега три врсте сигнала за сваки очни надражај. Међутим, људско око није идеалан спектрометар, па поједине таласе просто види као светlost исте боје.

Зато се боја на рачунару представља тројкама (R, G, B) где су координате удео црвене, плаве и зелене боје редом, које се узимају за основне. Постоје и други системи боја попут HSV или CMYK, али о њима неће бити речи јер ћемо надаље користити искључиво RGB систем.

На овај начин је могуће чувати тачно 255^3 боја, што је и више него доволно будући да људско око разликује до 10 милиона боја. Даља практична ограничења се тичу квалитета монитора као и графичког процесора, али се тиме нећемо бавити.

2.2 Камера

Можемо сматрати да нам је појам камере познат из стварног света. У овој глави ћемо строго дефинисати камеру и дати један од многобројних начина њеног моделовања у домену рачунарске графике.

Дефиниција 2.2.1. Камера је пресликавање из \mathbb{R}^3 у \mathbb{R}^2 .

У питању је врло општа, готово бескорисна дефиниција. Међутим, ако размислимо о томе да се стваран свет врло добро може представити као простор димензије 3, а да се фотографија може схватити као раван, можемо увидети да је у питању заиста исправна формулатија. Оно што овом дефиницијом није обухваћено јесте како се тачно од света око нас долази до слике. Зато има смисла говорити о моделима камере као различитим парадигмама изведбе поменутог пресликавања. У овом раду ће највише бити коришћен тачкасти модел камере тако да ћемо у наставку поглавља дати његову прецизну дефиницију.

2.2.1 Тачкасти модел камере

Посматрајмо канонски Еуклидски простор димензије 3 и раван $z = f$ коју ћемо звати **раван слике**. У овом моделу камере се произвољна тачка $\mathbf{x}_w = (x, y, z)$ из простора пресликава у тачку $\mathbf{x}_p = (u, v)$ која је тачка пресека равни слике и праве која спаја \mathbf{x}_w и центар камере \mathbf{c} , који ћемо за сад поставити у координатни почетак. Другим речима, у питању је централна пројекција са центром у координатном почетку. Тривијалном применом сличности троуглова долазимо до

$$(x, y, z) \mapsto (fx/z, fy/z). \quad (2.1)$$

Права која пролази кроз центар камере и нормална је на раван слике називамо **главном осом**, а тачка у којој се главна оса и раван слике секу називамо **главном тачком**.

Приметимо још једну особину централног пројектовања – све тачке праве која пролази кроз центар камере се пројектују у исту тачку равни слике. Зато ћемо увести хомогене координате. Отуда можемо писати $\mathbf{x}_p = P\mathbf{x}_w$. Овакво пресликање се може преписати и у матричном облику

$$\begin{bmatrix} x \\ y \\ z \\ 1 \end{bmatrix} \mapsto \begin{bmatrix} fx \\ fy \\ fz \\ 1 \end{bmatrix} = \begin{bmatrix} f & 0 & 0 & 0 \\ 0 & f & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} x \\ y \\ z \\ 1 \end{bmatrix}. \quad (2.2)$$

Ради конзистентности са наставком, раставићемо P на производ

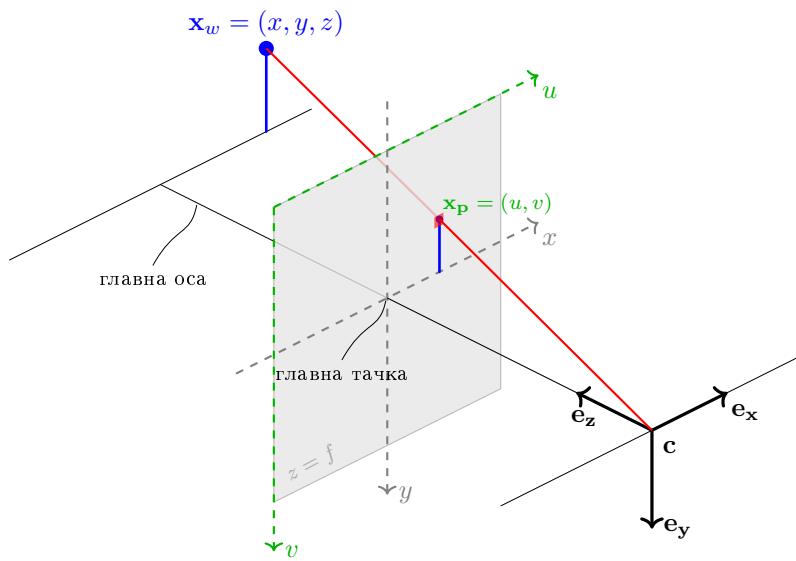
$$K [R \quad t], \quad (2.3)$$

где је $K \in M_{3,4}(\mathbb{R})$, $R \in M_{3,3}(\mathbb{R})$ и $t \in M_{3,1}(\mathbb{R})$. Уклањајући последњу колону из 2.2 добијамо матрицу K из овако изменjenог пресликања.

Уколико уопштимо положај главне тачке и њене координате означимо са (p_x, p_y) , матрица K поприма облик

$$\begin{bmatrix} f & 0 & p_x \\ 0 & f & p_y \\ 0 & 0 & 1 \end{bmatrix}. \quad (2.4)$$

Конечно, можемо уопштити и положај центра камере. Тада уочавамо два координатна система - онај с почетка, канонски, у ком нам је лако да баратамо и други, са центром камере као координатним почетком. Није тешко увидети да се кретањем (ротацијом и трансляцијом) репер камере може довести до канонског. На тај начин употребљавамо пресликање P . На слици 2.1 је дата интерпретација овог пресликања.



Слика 2.1: Интерпретација тачкастог модела камере

Постоји укупно 9 слободних параметара – по 3 за сваку од матрица K , R и t . Матрица K се назива **матрица калибрације**, а њене вредности **унутрашњим параметрима** камере. Вредности матрице $[R \quad t]$ се називају **спољашњим параметрима** камере.

Поменимо још једну терминолошку конвенцију. Координатни систем из којег вршимо пројекцију се назива и **светским координатним системом**, а координатни систем индукован положајем камере **координатни систем камере**. **Координатни систем слике** је онај везан за раван слике.

Корисно је дати додатни коментар у вези са нулом у првом реду матрице калибрације. У питању је коефицијент смицања којим је могуће уопштити раван слике – правоугаоник постаје паралелограм.

2.3 Рендеровање

Када смо говорили о камери, видели смо да слика није ништа друго до дводимензиона пројекција стварног света. Слично је и када комплексну сцену, овог пута виртуелног, света желимо да прикажемо на монитору. Мора доћи до пројектовања света, а слику коју видимо је уствари слика добијена посматрањем света кроз виртуелну камеру. Тада процес се назива **рендеровање**. Дакле, рендеровањем се од сцена долази до конкретне репрезентације слике у пикселима.

Пре него што се упустимо у рендеровање, вратимо се корак уназад. Светлост и камеру смо, на први поглед, увели врло неповезано. Светлост смо дефинисали више експериментално ослањајући се на физичке законе, док смо камеру увели строго геометријски. Сада ћемо оба појма ставити у контекст и објаснити њихову везу.

Праве које смо уочавали у моделу камере су светлосни зраци. Раван слике у практици није бесконачна, већ има своју ширину и дужину којима се одређују димензије слике. Такође, услед физичких ограничења слика се мора дискретизовати. Посматрамо ли слику на тај начин, видећемо да је она матрица чије елементе називамо **пикселима**. Како ће који пиксел да изгледа, односно које ће боје бити, одређује количина зрачења која допре до камере. Зрачење смо увели у зависности од правца, што у комбинацији са пројективним особинама модела камере одређује који ће пиксел бити погођен. Такође, природно је увести ограничење по питању дела простора који камера може да опажа. Величину слике смо већ поменули, али не и параметар по z оси. Ради једноставности геометрије, уводимо **предњу и задњу раван одсецања**. Све испред предње и иза задње равни одсецања не утиче на резултујућу слику.

2.3.1 Запреминско рендеровање

Посматрајмо шта се дешава са светлости приликом проласка кроз неку средину. Може доћи до:

- упијања – средина упија фотоне светлосног зрака приликом чега се ослобађа топлота или неки други вид енергије.
- емисије – како је светлост зрачење, пролазак светла кроз средину је загрева. Када средина достигне одређену температуру, може доћи до емитовања светlosti.

- расипања – део фотона напушта правац зрака што доводи до мешања фотона са различитих правца.

Према томе, више чинилаца утиче на коначно зрачење које ће путем неког светлосног зрака доћи до камере. Тачке неког светлосног зрака с почетком у \mathbf{o} и правцем \mathbf{d} једнозначно одређујемо као $\mathbf{r}_{\mathbf{o},\mathbf{d}}(t) = \mathbf{o} + t\mathbf{d}$.

Нека су предња и задња раван одсецања редом $z = t_n$ и $z = t_f$, а светлосне зраке посматрамо из положаја камере \mathbf{c} . Према томе, до камере дуж зрака $\mathbf{r}_{\mathbf{c},\mathbf{d}}$ допире следећа количина зрачења

$$\int_{t_n}^{t_f} T(t) \sigma(\mathbf{r}_{\mathbf{c},\mathbf{d}}(t)) L_{e,\Omega} \Big|_{\mathbf{r}_{\mathbf{c},\mathbf{d}}(t)} dt, \quad (2.5)$$

где је $T(t) = \exp\left(-\int_{t_n}^t \sigma(\mathbf{r}_{\mathbf{c},\mathbf{d}}(s))ds\right)$ акумулирана пропусност зрака од t_n до t . Ова величина представља вероватноћу да светлосни зрак на путу од t_n до t не удари нити у једну препреку. Напоменимо да се интеграција врши искључиво по фотонима који се налазе на светлосном зраку од интереса, или другим речима, интеграли се само по расутој светлости.

Како нам је боја пиксела крајњи циљ, једначину 2.5 ћемо преписати у колориметријском облику

$$C(\Pi_{\mathbf{r}_{\mathbf{c},\mathbf{d}}}) = \int_{t_n}^{t_f} T(t) \sigma(\mathbf{r}_{\mathbf{c},\mathbf{d}}(t)) C(\mathbf{r}_{\mathbf{c},\mathbf{d}}(t)) dt, \quad (2.6)$$

где је $\Pi_{\mathbf{r}_{\mathbf{c},\mathbf{d}}}$ тачка пресека светлосног зрака и равни слике, а C поље које сваку тачку пресликава у њену RGB боју.

Приказани поступак одређивања боје пиксела се назива *запреминско рендеровање*. У питању је само један од многобројних алгоритама за рендеровање који ће бити коришћен у наставку.

Глава 3

Основни појмови машинског учења

У овој глави ћемо формулисати теоријски оквир неопходан за разматрање и примену машинског учења.

Машинско учење је област *вештачке интелигенције*. Неформално говорећи, машинско учење обухвата алгоритме изведене из података или како се то често говори – научених из података. То значи да нема експлицитног програмирања, а врло често ни контроле процеса учења, већ се алгоритми дефинишу низом операција параметризованих на основу података који су изложени процесу обучавања. Процес обучавања, дакле, представља одређивање поменутих параметара којима се операције од значаја израчунавају на што исправнији начин. Приметимо да је та исправност прилично неодређен појам и зависи од примене, података и циља обучавања. Такође, врло се често не може ни квантifikovati што проблем обучавања чини утолико тежим.

Имајући у виду досег овог рада, потребно је увести појам надгледаног учења. Реч је о врсти учења у којој су уз податке присутни и додатни подаци којима је директно могуће утврдити да ли је излаз алгоритма исправан или није. На пример, ако је потребно утврдити да ли је на датој слици мачка, уз слику би постојао једнобитни податак који то недвосмислено потврђује. Према томе, скуп података можемо посматрати као скуп одбира $\mathcal{D} = \{(x_i, y_i)\}$, где x_i представља улаз, а y_i одговарајући излаз. То нам омогућава да посматрамо расподелу таквих одбира, односно густину расподеле $p(x, y)$. Природно се намеће потреба за што приближнијем одређивању поменуте расподеле, што подразумева одређивање функције f којом успостављамо везу између одговарајућих парова x и y . Кандидата за f има несагледиво много, а нама је потребна она *најбоља*, при чему се овог пута то мора формално дефинисати.

Претпоставимо да су y_i узорковани из метричког простора. Зато можемо дефинисати *функцију грешке* \mathcal{L} којом меримо квалитет апроксимације y_i вредношћу $f(x_i)$. Вреднујући одбирке у складу са својом густином дефинишемо *ризик*

$$R(f) = \mathbb{E}(\mathcal{L}(y, f(x))) = \int \mathcal{L}(y, f(x)) p(x, y) dx dy. \quad (3.1)$$

Проблем надгледаног учења покушава да дође до f за које се ризик минимизује. Начелно, обучавање се може извести по свим могућим функцијама f . Како је то практично немогуће, претрага функција се усмерава увођењем додатних претпоставки, односно рестрикција, скупа претраге. Ми ћемо посматрати функције параметризоване скупом параметара Θ , тако да минимизовање ризика посматрамо као

$$\min_{\Theta} R(f_{\Theta}). \quad (3.2)$$

У општем случају не мора постојати само један исправан излаз за конкретан улаз, што оправдава дефинисање ризика коришћењем заједничке расподеле $p(x, y)$. То за наше потребе неће бити неопходно. Наиме, посматраћемо условну расподелу $p(y|x)$.

Расподела $p(y|x)$ је ретко кад позната. Зато се спроводи стандардни статистички третман – ризик се мења емпириским ризиком:

$$ER(f_\Theta, \mathcal{D}) = \frac{1}{|\mathcal{S}|} \sum_{i=1}^{|\mathcal{S}|} \mathcal{L}(y_i, f_\Theta(x_i)), \quad (3.3)$$

где је \mathcal{S} узорак скупа \mathcal{D} .

Претпоставимо да се скуп параметара састоји од само једног вектора димензије свега 2, тј. $\Theta = \{\mathbf{w}\} = \{(w_1, w_2)\}$, а да је $x_i \in \mathbb{R}^1$. Имајући у виду израз 3.2, за кандидате функције f , између остalog, имамо $\sin(x_i w_1 + w_2)$, $\log(x_i w_1 w_2)$, $\exp(w_2 x_i^{w_1})$. Иако је проблем постављен као веома једноставан, а параметарском рестрикцијом начињен још једноставнијим, и даље се чини као веома тежак будући да немамо начин да за разумно коначно времена претражимо све такве кандидате. Зато је потребно посматрати тачно одређене класе функција f за које је то могуће, а међу најпознатијим су свакако *неуронске мреже*. Заправо, потребно је фиксирати конкретну архитектуру модела, а потом вршити оптимизацију параметара у складу са одабраном метриком.

3.1 Неуронске мреже

Дефиниција 3.1.1. *Неуронска мрежа је функција облика*

$$f_\Theta(\mathbf{x}) = \left(\prod_{i=1}^L a_i (\beta_i + W_i) \right) (\mathbf{x}),$$

где је $\Theta = \{W_i\}_{i=1}^L$ скуп параметара који се обучава, $W_i \in \mathbb{R}^{d_i \times d_{i-1}}$ при чему је $d_0 = \dim(\mathbf{x})$ и $\{a_i\}_{i=1}^L$ функције које се примењују члан по члан. Функције $\{a_i\}_{i=1}^L$ називамо **активационим функцијама**. Параметар L називамо **бројем слојева неуронске мреже**.

Активационе функције могу бити произвољне, докле год поштују услове димензионалности. Међутим, уколико су оне линеарне, тада неуронска мрежа постаје линеарна, односно постаје линеарна регресија. У циљу добијања што разноврснијих модела, за функције активације се узимају нелинеарне функције. Један од популарних избора је $\text{ReLU}(x) = \max\{0, x\}$ [4].

Неуронске мреже су познате и под именом *вишеслојни перцептрони*, па ће убудуће бити коришћена и скраћеница MLP (енг. *multilayer perceptron*).

Значајан теоријски резултат даје следећа теорема [5]

Теорема 3.1.1. *За сваку функцију $F : \mathbb{R}^n \rightarrow \mathbb{R}^m$ интеграбилну у Бахнеровом смислу и свако $\epsilon > 0$, постоји неуронска мрежа f_Θ са ReLU активационим функцијама, тако да је $\{d_i = \max\{n+1, m\}\}_{i=1}^L$ и важи*

$$\int_{\mathbb{R}^n} \|f_\Theta(\mathbf{x}) - F(\mathbf{x})\| dx < \epsilon.$$

Ово за последицу има да се готово свака функција може произвољно добро апоксимирати неуронском мрежом, али како доказ није конструктиван није очигледно како таква неуронска мрежа изгледа.

Уколико су активационе функције диференцијабилне функције, то ће бити и неуронска мрежа. Зато се поступак минимизације емпиријског ризика спроводи градијентним спустом. Поступак се спроводи у две етапе – прво се неуронска мрежа примени над подацима пропагацијом унапред, а потом се пропагацијом уназад, користећи правилом ланца, изврши ажурирање параметара.

Како је градијентни спуст оптимизација првог реда, није гарантовано да ће пронађени локални минимум бити и глобални. У пракси се показује да градијентни спуст даје веома добра решења.

Када је реч о имплементацији неуронских мрежа и конкретим детаљима њиховог обучавања, важно је разматрати хардверска ограничења. Наиме, ови модели могу бити веома комплексни где се ред величине броја параметара креће и до неколико милиона, па чак и до неколико милијарди. Зато се приликом обучавања прибегава разнородним техникама, често нумеричке или хеуристичке природе, којима се смањује укупно време обучавања или смањује утрошак меморије. Еклатантан пример је *стохастички градијентни спуст* [6] – параметри се ажурирају на основу само једног одбира уместо целог скупа података. Овај метод је сушта супротност конвенционалном градијентном спусту, тако да се углавном врши компромис.

Глава 4

Неуронска поља зрачења

Ову главу почињемо са две важне дефиниције.

Дефиниција 4.1. Поље је пресликавање $F : \mathbb{R}^n \rightarrow \mathbb{R}^m$. Специјално, поље је скаларно за $m = 1$.

Дефиниција 4.2. Неуронско поље је поље макар делimiично параметризовано неуронском мрежом.

Неуронско поље зрачења је посебан случај неуронског поља за $n = 5$ и $m = 3$. У питању је параметризација пресликавања које свакој тачки (x, y, z) придружује зрачење и пропусност и то за сваки правац одређен угловима θ и ϕ . Пропусност се може видети и као вероватноћа да се зрак *зауставља* у тој тачки.

Размотримо шта добијамо оваквом поставком. Претпоставимо да је неуронско поље савршено обучено. То значи да можемо утврдити боју пиксела из сваког могућег угла гледања. Тако се сцена рендерије и из погледа који се нису нашли у скупу за обучавање.

Како је таква неуронска параметризација непозната *a priori*, имамо слободу у виду дизајнирања архитектуре параметризације. Подробно ћемо обрадити оригинални NeRF модел и његово проширење mip-NeRF, а потом се осврнути на специфичан Ref-NeRF који је прилагођен тачно одређеним условима.

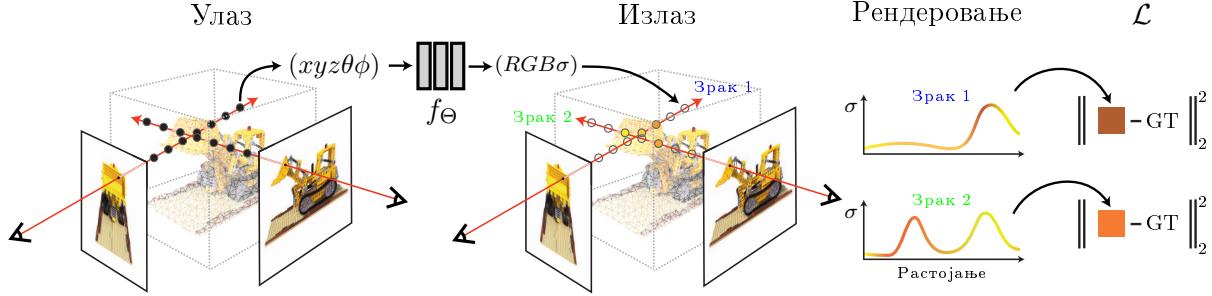
Оно што на први поглед издаваја неуронска поља зрачења од осталих параметризација неуронским мрежама јесте обучавање. Узмимо за потребе илустрације ImageNet [7] скуп података. Поменути скуп је сачињен од великог броја троканалних слика, тако да је сваки узорак изворног скupa једноставан за представити. Међутим, скуп података којим ми баратамо је сачињен од различитих погледа, а уз сваки поглед је придружена одговарајућа изрендерована слика. Имајући у виду улаз неуронског поља зрачења, скупу података је неопходно увести још један слој гранулације – зрак са сваки угао гледања у одређеном погледу. Дакле, један одбирај из скupa података над којим се обучава неуронско поље зрачења се може видети као један светлосни зрак и то за један поглед уз пратећи рендер.

4.1 NeRF

NeRF (енг. *Neural Radiance Field*) је најстарији модел у фамилији модела који ћемо у овом раду обрадити. Заснива се на запреминском рендериовању и прва је

изведба неуронских поља зрачења овог типа.

На слици 4.1 се може видети илустрација овог модела.



Слика 4.1: Илустрација NeRF модела према [1]

4.1.1 Параметризација

Неуронско поље зрачења се параметризује вишеслојним перцептроном, али на благо неубичајен начин. Јасно је да σ не зависи од угла гледања већ искључиво од положаја \mathbf{x} . С друге стране, боја пиксела зависи од положаја као и од пропусности. Зато ReLU MLP од 8 слојева са по 256 неурона на улазу добија само \mathbf{x} , док је излаз предвиђена вредност σ и вектор димензије 256. Тада вектор се спаја са параметрима угла гледања и пропушта кроз један слој са 128 неурона и ReLU активационом функцијом. На овај начин се добија боја пиксела која је условљена погледом. Поменуте неуронске мреже називамо редом *просторни* и *усмерени* MLP.

Чак и уз овако дефинисан поступак добијања коначног излаза, улаз и даље има малу димензију. Да би се превазишао тај проблем, улази се прво пресликају у простор веће димензије. Према томе, неуронску мрежу F_Θ можемо видети као композицију $F'_\theta \circ \gamma$ где се $\gamma : \mathbb{R} \rightarrow \mathbb{R}^{2D}$ и то конкретно

$$\gamma(x) = (\sin(2^0 \pi x), \cos(2^0 \pi x), \dots, \sin(2^{D-1} \pi x), \cos(2^{D-1} \pi x)). \quad (4.1)$$

Исто пресликање се користи за све улазе које MLP добија, с тим што се за положај узима $D = 10$, а за поглед $D = 4$. Овај начин *подизања* у простор веће димензије није случајан и инспирисан је [8], а могуће варијације су објашњене у [9]. Описани поступак се назива *Фуријеизација*.

4.1.2 Рендеровање

У основи је једначина 2.6. За потребе апроксимације интеграла биће коришћен Гаусов метод. Уколико користимо детерминистички приступ, одбирачи који учествују у апроксимацији ће увек бити исти. То је неповољан приступ будући да би у том случају неуронско поље зрачења било увек узорковано за исти скуп параметара. Уместо тога, увешћемо еквидистантну поделу сегмента $[t_n, t_f]$, а потом насумице одабрати по један број из сваког елемента поделе. То формално исказујемо наредним изразом

$$t_i \sim \mathcal{U} \left[t_n + \frac{i-1}{N}(t_f - t_n), t_n + \frac{i}{N}(t_f - t_n) \right]. \quad (4.2)$$

Препишимо једначину 2.6 у дискретном облику, у складу са 4.2

$$\hat{C}(\Pi_{\mathbf{r}_{\mathbf{c}, \mathbf{d}}}) = \sum_{i=1}^N T_i (1 - \exp(-\sigma_i \Delta_i)) \mathbf{c}_i, \quad (4.3)$$

где је $T_i = \exp \left(-\sum_{j=1}^{i-1} \sigma_j \Delta_j \right)$, парови (\mathbf{c}_i, σ_i) одговарајући у складу са избором t_i , а Δ_i ширина елемента поделе.

Једначине 2.6 и 4.3 се видно разликују по начину на који учествује пропусност. Наиме, уколико би Гаусов метод применили директно, појавио би се линеаран члан $\sigma_i \Delta_i$. Тај члан је је у апроксимацији замењен експоненцијалним $1 - \exp(-\sigma_i \Delta_i)$. У наставку ћемо објаснити и зашто [10].

Како узимамо да су на сваком елементу поделе боја и пропусност константни, једначину 2.6 расписујемо у складу са тим. Погледајмо како произвољни елемент поделе учествује у апроксимацији.

$$\begin{aligned} \hat{C}_i(\Pi_{\mathbf{r}_{\mathbf{c}, \mathbf{d}}}) &= \int_{t_i}^{t_{i+1}} T(t) \sigma_i c_i dt \\ &= \int_{t_i}^{t_{i+1}} \sigma_i c_i \exp \left(- \int_{t_n}^t \sigma(\mathbf{r}_{\mathbf{c}, \mathbf{d}}(s) ds) \right) dt \\ &= \sigma_i c_i \int_{t_i}^{t_{i+1}} \exp \left(- \int_{t_n}^{t_i} \sigma(\mathbf{r}_{\mathbf{c}, \mathbf{d}}(s) ds) \right) \exp \left(- \int_{t_i}^{t_{i+1}} \sigma_i ds \right) dt \\ &= T_i \sigma_i c_i \int_{t_i}^{t_{i+1}} \exp(-\sigma_i(t - t_i)) dt \\ &= T_i \sigma_i c_i \frac{\exp(-\sigma_i(t - t_i))}{-\sigma_i} \Big|_{t_i}^{t_{i+1}} \\ &= T_i (1 - \exp(-\sigma_i \Delta_i)) c_i. \end{aligned} \quad (4.4)$$

4.1.3 Обучавање

Имплементација инспирисана описаним поступком рендеровања је врло неефикасна – празан простор и прикривени региони не доприносе рендерираној слици, а непрестано се узоркују. Из тог разлога се користе два неуронска поља, за *фина* и *груба* предвиђања. Прво се узоркује N_c одбирача на начин објашњен у 4.2 који се потом дају грубом пољу. На основу грубих предвиђања се врши још једно узорковање, али овог пута боље навођено. Фино узорковање је пристрасније релевантним деловима простора.

Да би то спровели у дело, излаз грубог неуронског поља записујемо као збир боја дуж зрака пондерисаног непрозирношћу

$$\hat{C}_c(\Pi_{\mathbf{r}_{\mathbf{c}, \mathbf{d}}}) = \sum_{i=1}^{N_c} w_i c_i, \quad w_i = T_i (1 - \exp(-\sigma_i \Delta_i)). \quad (4.5)$$

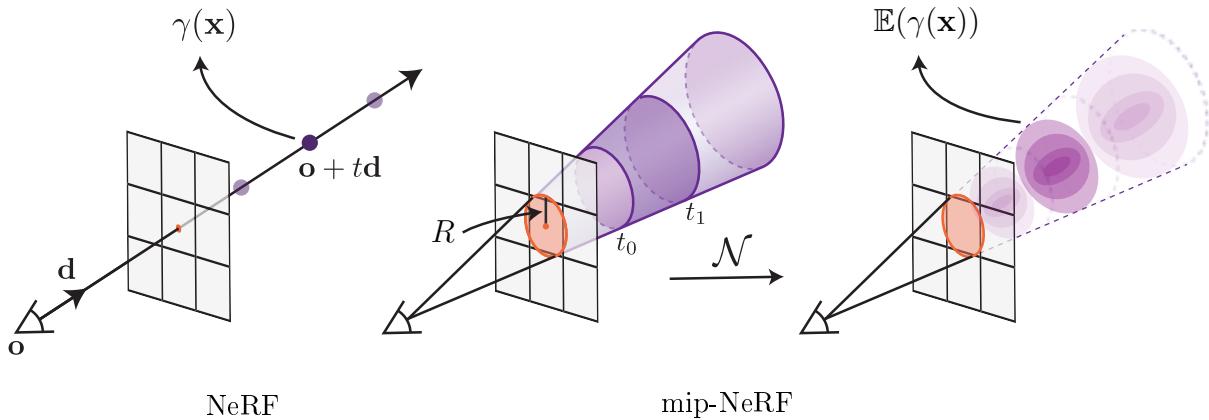
Скалирањем тежина $w_i = w_i / \sum_{j=1}^{N_c} w_j$, добија се део-по-део константна густина расподеле дуж зрака. Из ове расподеле се (инверзно) узоркује N_f одбирача, а потом финомеуронско поље израчунато у свим $N_c + N_f$ одбирача на начин описан у једначини 4.3. У свим експериментима се за N_c узима 64, а за N_f 128. Да би означили о каквом узорковању је реч, увешћемо додатне ознаке f и c .

Функција губитка која се користи за обучавање NeRF-а је

$$\mathcal{L} = \sum_{\mathbf{r}_{c,d} \in \mathcal{B}} \left(\left\| \hat{C}_c \left(\Pi_{\mathbf{r}_{c,d}^c} \right) - C \left(\Pi_{\mathbf{r}_{c,d}^c} \right) \right\|_2^2 + \left\| \hat{C}_f \left(\Pi_{\mathbf{r}_{c,d}^f} \right) - C \left(\Pi_{\mathbf{r}_{c,d}^f} \right) \right\|_2^2 \right). \quad (4.6)$$

4.2 Mip-NeRF

Као што смо могли да видимо у претходној глави, NeRF је далеко од идеалног теоријског модела и показује извесне недостатке. То су постојање два модела који се упоредо обучавају, узорковање дуж само једног зрака за сваки пиксел, а као што ћемо видети касније у табели 6.6, обучавање траје данима. Овом архитектуром покушаћемо да превазиђемо или макар ублажимо наведено.



Слика 4.2: Поређење NeRF и mip-NeRF модела према [2]

Кренућемо од кључног недостатка – постојање само једног зрака по пикселу. Природно проширење је конусно емитовање зрака. Конус ћемо ограничити двема равнима чиме добијамо зарубљену купу која је једнозначно одређена центром \mathbf{o} , осом симетрије \mathbf{d} и полу пречником круга R који је пресек поменуте купе и равни слике. Скуп тачака које припадају зарубљеној купи на сегменту параметра $[t_0, t_1]$ је одређен индикаторском променљивом

$$F(\mathbf{x}, \mathbf{o}, \mathbf{d}, R, t_0, t_1) = I \left(t_0 < \frac{\mathbf{d}^T (\mathbf{x} - \mathbf{o})}{\|\mathbf{d}\|_2^2} < t_1 \right) \wedge \left(\frac{\mathbf{d}^T (\mathbf{x} - \mathbf{o})}{\|\mathbf{d}\|_2 \|\mathbf{x} - \mathbf{o}\|_2} > \frac{1}{\sqrt{1 + (R/\|\mathbf{d}\|_2)^2}} \right). \quad (4.7)$$

4.2.1 Интегрална Фуријеизација

Један од, наизглед, не нарочито битних одлика NeRF-а је допринос Фуријеизације улаза на начин објашњен у 4.1. Зато ћемо посебну пажњу посветити управо томе.

Главни циљ је задржати изворну Фуријеизацију. Најједноставнији приступ је одређивање очекиване изворне Фуријеизације по свим тачкама које припадају ку- пи зрака. Тако добијамо

$$\gamma^*(\mathbf{o}, \mathbf{d}, R, t_0, t_1) = \frac{\int \gamma(\mathbf{x}) F(\mathbf{x}, \mathbf{o}, \mathbf{d}, R, t_0, t_1) d\mathbf{x}}{\int F(\mathbf{x}, \mathbf{o}, \mathbf{d}, R, t_0, t_1) d\mathbf{x}}. \quad (4.8)$$

Међутим, интеграл који се јавља у овој једначини не умемо да израчунамо. Зато ћемо зарубљену купу представити нормалном расподелом која нам то омогућава. У ту сврху, потребно је одредити очекивање и варијансу функције $F(\mathbf{x}, \cdot)$. То за последицу има да је зарубљена купа апроксимирана елипсоидом, а расподела одређена са пет вредности – \mathbf{o} , \mathbf{d} , μ_t (очекивано растојање дуж зрака), σ_t^2 (варијанса дуж зрака) и σ_r^2 (варијанса дуж нормале зрака).

У наставку ћемо извести засад непознате вредности параметара μ_t , σ_t и σ_r . Помагајмо зарубљену купу чија је оса симетрије нормална, односно, паралелна са свим осама. Употребом сферних координата добијамо параметризацију

$$(x, y, z) = \phi(r, t, \theta) = (rt \cos \theta, rt \sin \theta, t), \quad (4.9)$$

за $\theta \in [0, 2\pi]$, $t \geq 0$, $|r| \leq R$. Јакобијан је $rt^2 dr dt d\theta$.

Запремина зарубљене купе је

$$V = \int_0^{2\pi} \int_{t_0}^{t_1} \int_0^R rt^2 dr dt d\theta = \pi R \frac{t_1^3 - t_0^3}{3}. \quad (4.10)$$

Расподела у којој се тачке узоркују равномерно из зарубљене купе има густину rt^2/V , у шта се можемо уверити интеграцијом Јакобијана.

Први момент по t је

$$\begin{aligned} \mathbb{E}(t) &= \frac{1}{V} \int_0^{2\pi} \int_{t_0}^{t_1} \int_0^R t \cdot rt^2 dr dt d\theta \\ &= \frac{1}{V} \pi R^2 \frac{t_1^4 - t_0^4}{4} \\ &= \frac{3(t_1^4 - t_0^4)}{4(t_1^3 - t_0^3)}. \end{aligned} \quad (4.11)$$

Први момент по x (као и по y због симетрије је)

$$\mathbb{E}(x) = \frac{1}{V} \int_0^{2\pi} \int_{t_0}^{t_1} \int_0^R rt \cos \theta \cdot rt^2 dr dt d\theta = 0. \quad (4.12)$$

Други момент по t је

$$\begin{aligned} \mathbb{E}(t^2) &= \frac{1}{V} \int_0^{2\pi} \int_{t_0}^{t_1} \int_0^R t^2 \cdot rt^2 dr dt d\theta \\ &= \frac{1}{V} \pi R^3 \frac{t_1^5 - t_0^5}{5} \\ &= \frac{3(t_1^5 - t_0^5)}{5(t_1^3 - t_0^3)}. \end{aligned} \quad (4.13)$$

Други момент по x (као и по y због симетрије) је

$$\begin{aligned}\mathbb{E}(x^2) &= \frac{1}{V} \int_0^{2\pi} \int_{t_0}^{t_1} \int_0^R (rt \cos \theta)^2 \cdot rt^2 dr dt d\theta \\ &= \frac{1}{V} \frac{R^4}{4} \frac{t_1^5 - t_0^5}{5} \pi \\ &= \frac{3R^2(t_1^5 - t_0^5)}{20(t_1^3 - t_0^3)}.\end{aligned}\tag{4.14}$$

Када смо израчунали све ове моменте, коначно можемо одредити параметре ко-нусне расподеле који су нам непознати. Очекивање дуж зрака μ_t је први момент по t . Варијансу дуж зрака σ_t^2 добијамо као $\text{Var}(t) = \mathbb{E}(t^2) - \mathbb{E}(t^2)$. И коначно, $\sigma_r^2 = \mathbb{E}(x^2)$.

У претходним изразима се појављују разлике виших степена, што врло често проузрокује нумеричке потешкоће када су t_0 и t_1 близки. Да то заиста јесте од значаја потврђује чињеница да је апроксимација коју спроводимо прецизна једино када је зарубљена купа ниска, односно, када је елипсоид нормалне расподеле пљоснат. Зато се уводе смене $t_\mu = (t_0 + t_1)/2$ и $t_\delta = (t_1 - t_0)/2$. Коначно, добијамо

$$\begin{aligned}\mu_t &= t_\mu + \frac{2t_\mu t_\delta^2}{3t_\mu^2 + t_\delta^2} \\ \sigma_t^2 &= \frac{t_\delta^2}{3} - \frac{4t_\delta^4(12t_\mu^2 - t_\delta^2)}{15(3t_\mu^2 + t_\delta^2)^2} \\ \sigma_r^2 &= R^2 \left(\frac{t_\mu^2}{4} + \frac{5t_\delta^2}{12} - \frac{4t_\delta^4}{15(3t_\mu^2 + t_\delta^2)} \right).\end{aligned}\tag{4.15}$$

На овај начин смо потпуно одредили нормалну расподелу, али у координатном систему равни слике. Међутим, није тешко доћи до трансформације у светски координатни систем

$$\begin{aligned}\mu &= \mathbf{o} + \mu_t \mathbf{d} \\ \Sigma &= \sigma_t^2 (\mathbf{d} \mathbf{d}^T) + \sigma_r^2 \left(\mathbf{I} - \frac{\mathbf{d} \mathbf{d}^T}{\|\mathbf{d}\|_2^2} \right).\end{aligned}\tag{4.16}$$

У досадашњем разматрању смо припремили терен за дефинисање *интегралне Фуријеизације* која се прилагођава новоуведеном конусном емитовању зрака. Подсећићемо се идеје још једном – израчунати очекивану Фуријеизацију на елипсоиду нормалне расподеле. Дакле, потребно је пресликати улаз у простор више димензије, а потом у том простору израчунати очекивање. Да би се та два корака ускладила, другачије ћемо записати 4.1

$$P = \begin{bmatrix} 1 & 0 & 0 & 2 & 0 & 0 & \dots & 2^{L-1} & 0 & 0 \\ 0 & 1 & 0 & 0 & 2 & 0 & \dots & 0 & 2^{L-1} & 0 \\ 0 & 0 & 1 & 0 & 0 & 2 & \dots & 0 & 0 & 2^{L-1} \end{bmatrix}^T \text{ и } \gamma(\mathbf{x}) = \begin{bmatrix} \sin(P\mathbf{x}) \\ \cos(P\mathbf{x}) \end{bmatrix}.\tag{4.17}$$

Параметри расподеле се мењају у складу са новим простором, тако да добијамо $\boldsymbol{\mu}_\gamma = P\boldsymbol{\mu}$. Облик матрице коваријансе следи из њене линеарности. Наиме, како је $\text{Cov}(A\mathbf{x}, B\mathbf{y}) = A \text{Cov}(\mathbf{x}, \mathbf{y}) B^T$, то је $\Sigma_\gamma = P\Sigma P^T$.

Преостаје да се израчуна очекивање у складу са вишедимензионом расподелом. Потребно је и доволно је показати понашање очекивања $\gamma(x)$. Прво ћемо показати очекивање у случају једнодимензионе нормалне расподеле, а потом и у случају који је нама потребан.

$$\begin{aligned}\mathbb{E}_{x \sim \mathcal{N}(\mu, \sigma^2)}(\sin x) &= \mathbb{E}(\text{Im} \exp(ix)) \\ &= \text{Im} \mathbb{E}(\exp(ix)) \\ &= \text{Im}(\phi_{x \sim \mathcal{N}(\mu, \sigma^2)}(1)) \\ &= \text{Im} \left(\exp \left(i\mu - \frac{\sigma^2}{2} \right) \right) \\ &= \sin(x) \exp \left(-\frac{\sigma^2}{2} \right)\end{aligned}\tag{4.18}$$

где је функција $\phi_{x \sim \mathcal{N}(\mu, \sigma^2)}(t)$ је карактеристична функција нормалне расподеле.

Аналогно, добијамо $\mathbb{E}_{x \sim \mathcal{N}(\mu, \sigma^2)}(\cos x) = \cos(x) \exp(-\sigma^2/2)$. Уколико уопштимо претходно израчунавање, добијамо

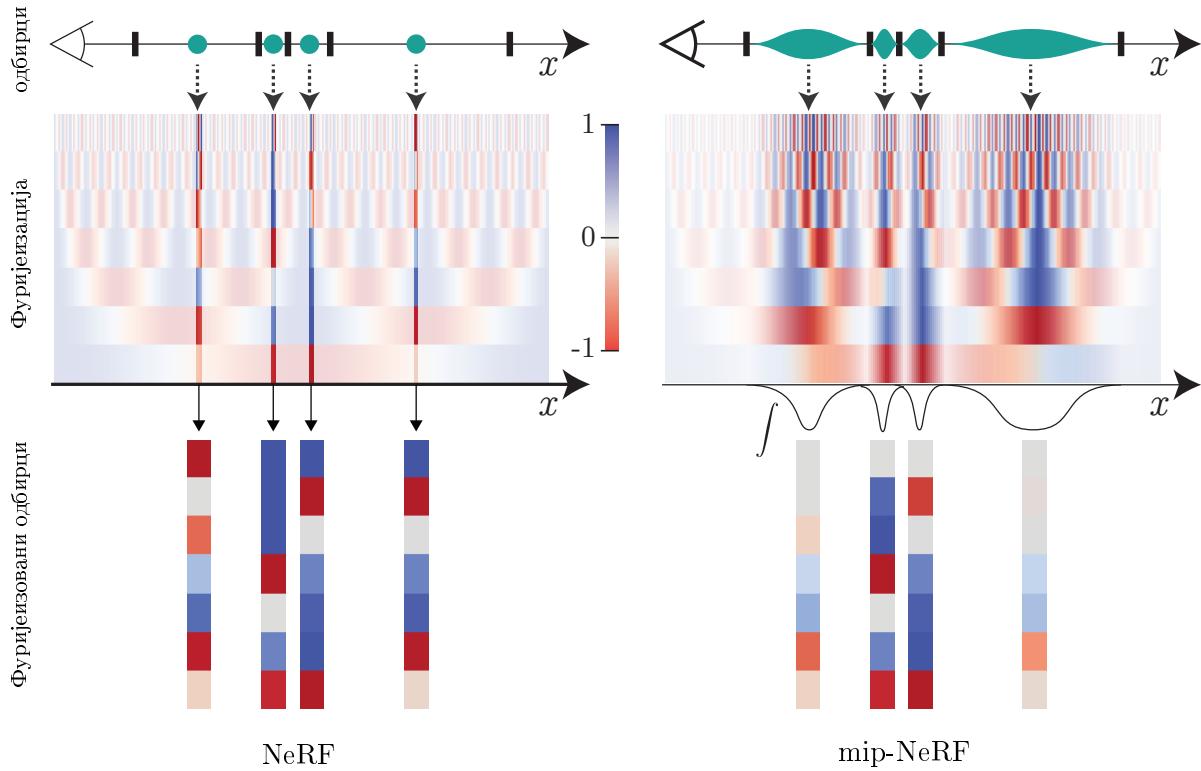
$$\begin{aligned}\gamma(\boldsymbol{\mu}, \Sigma) &= \mathbb{E}_{x \sim \mathcal{N}(\mu, \sigma^2)}(\gamma(x)) \\ &= \begin{bmatrix} \sin(\mu_\gamma) \odot \exp(-(1/2) \text{diag}(\Sigma_\gamma)) \\ \cos(\mu_\gamma) \odot \exp(-(1/2) \text{diag}(\Sigma_\gamma)) \end{bmatrix},\end{aligned}\tag{4.19}$$

где је са \odot означено Адамарово, члан-по-члан множење.

Начин на који се изводи множење у запису 4.19 подразумева употребу само дијагоналних елемената матрице Σ_γ . Израчунавање пуне матрице изискује пуно рачунских операција и места, за чим уопште нема потребе. С тим у вези, дијагоналне чланове ћемо израчунати директно $\text{diag}(\Sigma_\gamma) = [\text{diag}(\Sigma), 4 \text{diag}(\Sigma), \dots, 4^{L-1} \text{diag}(\Sigma)]^T$ при чему је

$$\text{diag}(\Sigma) = \sigma_t^2(\mathbf{d} \odot \mathbf{d}) + \sigma_r^2 \left(I - \frac{\mathbf{d} \odot \mathbf{d}}{\|\mathbf{d}\|_2^2} \right)\tag{4.20}$$

На слици 4.3 је приказана разлика у Фуријеизацији у случају NeRF и mip-NeRF модела.



Слика 4.3: Приказ интегралне Фуријеизације према [2]

4.2.2 Параметризација и обучавање

Иако mip-NeRF зраке емитује по конусу, обучавање се практично изводи на исти начин као и у случају NeRF-а. Уместо узорковања \$n\$ одбираца дуж зрака, узоркује их се \$n + 1\$ јер ће се узастопно у паровима користити за границе зарубљене купе. Након примене Фуријеизације, прослеђују се MLP-у и добијају парови \$(\sigma, c)\$. MLP је по структури идентичан оном који се користи за параметризацију NeRF-а.

Међутим, mip-NeRF не захтева употребу финог и грубог модела, већ се цео поступак спроводи употребом само једног. Одговор лежи у употреби конусног емитовања зрака и интегралне Фуријеизације.

Са \$\mathbf{k}_{\mathbf{o}, \mathbf{d}, R}\$ ћемо означити зарубљену купу са врхом у \$\mathbf{o}\$, осе симетрије \$\mathbf{d}\$ и полу-пречника пресека купе и равни слике са \$R\$. Поменути пресек означавамо са \$\Pi_{\mathbf{k}_{\mathbf{o}, \mathbf{d}, R}}\$. Ознаке \$f\$ и \$c\$ имају исто значење као и у 4.6. Функција губитка која се користи за обучавање mip-NeRF-а је

$$\mathcal{L} = \sum_{\mathbf{f}_{\mathbf{c}, \mathbf{d}, R} \in \mathcal{B}} \left(\left\| \hat{C} \left(\Pi_{\mathbf{k}_{\mathbf{c}, \mathbf{d}, R}^f} \right) - C \left(\Pi_{\mathbf{k}_{\mathbf{c}, \mathbf{d}, R}^f} \right) \right\|_2^2 + \lambda \left\| \hat{C} \left(\Pi_{\mathbf{k}_{\mathbf{c}, \mathbf{d}, R}^c} \right) - C \left(\Pi_{\mathbf{k}_{\mathbf{c}, \mathbf{d}, R}^c} \right) \right\|_2^2 \right). \quad (4.21)$$

Параметар \$\lambda\$ контролише утицај грубе поделе на функцију губитка јер се само један модел користи за обе врсте поделе.

Врши се укупно по 128 узорковања за грубу и фину поделу.

4.3 Ref-NeRF

Иако претходне имплементације дају веома добре резултате, за одређене сцене можемо видети њихова ограничења. Посматрањем резултата на сликама и 6.2 и 6.3 можемо уочити проблем са учењем *сјајности*. Где год постоји изразита спекуларна светлост, NeRF и mip-NeRF нису у стању да је доћарају верно, већ се стиче утисак дифузне рефлексије – другим речима, површина делује *обрушено*.

Водећи рачуна о томе како је вектор \mathbf{d} усмерен, његову рефлексију око \mathbf{n} рачунамо

$$\mathbf{d}_r = \frac{\mathbf{d}}{\|\mathbf{d}\|} - 2 \left(\frac{\mathbf{d}}{\|\mathbf{d}\|} \cdot \mathbf{n} \right) \mathbf{n}. \quad (4.22)$$

о начину добијања поменуте нормале биће више речи касније.

4.3.1 Сферна Фуријеизација

Поново се као кључна разлика јавља другачија врста Фуријеизације, с тим што је овог пута она статистички заснована. У наставку ћемо објаснити и зашто.

Одаберимо неку тачку на сфери произвољно. Уколико је осветљење сфере у окolini те тачке дифузно, тада је густина одбијених зрака равномернија јер не долази до нагле разлике у осветљењу. С друге стране, ако је у питању спекуларна рефлексија, тада је расподела одбијених зрака готово Диракова – приметно је нагомилавање густине у једној тачки.

Отуда идеја да посматрамо расподелу која добро моделује другачију концентрацију густине у различитим тачкама на сферама. Зато бирајмо *фон Мизес-Фишерову* расподелу. Даћемо формални опис ове расподеле у случају димензије 3.

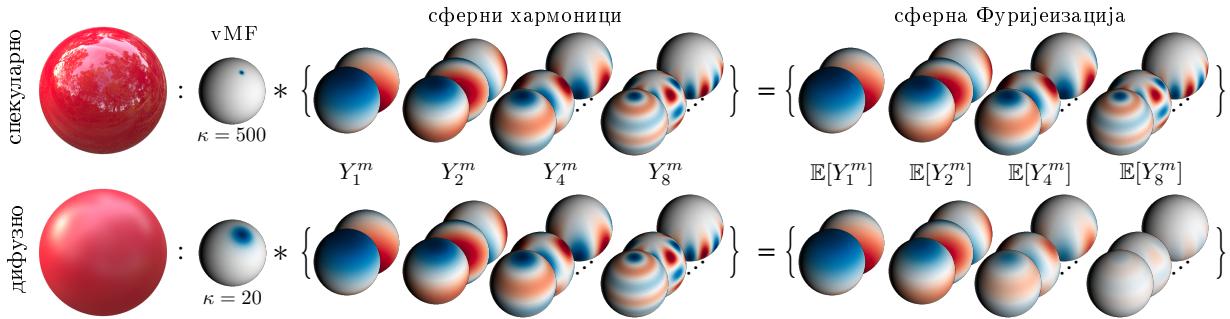
Дефиниција 4.3.1. Густина *фон Мизес-Фишерове* расподеле је дата са

$$f(\mathbf{x}) = \frac{\kappa}{4\pi \sinh \kappa} \exp(\kappa \boldsymbol{\mu}^T \mathbf{x}).$$

Параметри $\boldsymbol{\mu}$ и κ се редом зову *очекивање* и *концентрација*.

У претходним изведбама смо за потребе пресликовања у простор више димензије користили Фуријеове коефицијенте, односно, *хармонике*. Мада се, строго говорећи, хармоники јављају у мало другачијем облику него што је то случај у Фуријеовом развоју.

Како смо за потребе боље анализе осветљења објекта природно прешли на сферу, а у складу са тим и на одговарајућу расподелу, хармонике је потребно заменити сферним еквивалентом. Зато уводимо *сферне хармонике*, чији је приказ дат на слици 4.4.



Слика 4.4: Приказ сферних хармоника и сферне Фуријеизације према [3]

Дефиниција 4.3.2. Сферни хармоници су решења Лапласове диференцијалне једначине $\nabla^2 f = 0$ у сферним координатама, где је $f : \mathbb{R}^3 \rightarrow \mathbb{C}$ два пута диференцијабилна функција. Сферни хармоник степена l и реда m је комплексна функција на јединичној сferи у ознаки Y_l^m .

Следећу теорему наводимо без доказа.

Теорема 4.3.1. Општи облик сферног хармоника $Y_l^m(\theta, \phi)$ је $N e^{im\phi} P_l^m(\cos \theta)$, где је P_l^m одговарајући Лежандров полином, а N коефицијент нормализације.

Коефицијент нормализације можемо одредити из ортогоналности. Може се показати да из $\int_{S^2} Y_l^m Y_{l'}^{m'} dS^2 = 0$, следи

$$N = \sqrt{\frac{2l+1}{4\pi} \cdot \frac{(l-m)!}{(l+m)!}}.$$

Претходно уведене појмове ћемо обједнити *сферном Фуријеизацијом* и то

$$\gamma(\mathbf{d}_r, \kappa) = \{\mathbb{E}_{\omega \sim vMF(\mathbf{d}_r, \kappa)}(Y_l^m(\boldsymbol{\omega})) : (l, m) \in \mathcal{M}_L\}, \quad (4.23)$$

где је $\mathcal{M}_L = \{(l, m) : l = 1, \dots, 2^L \wedge m = 0, \dots, l\}$.

Остаје питање израчунавања очекивања $\mathbb{E}_{\omega \sim vMF(\mathbf{d}_r, \kappa)}(Y_l^m(\boldsymbol{\omega}))$. Показује се [3] да је поменути израз приближно једнак

$$\exp\left(-\frac{l(l+1)}{2\kappa}\right) Y_l^m(\mathbf{d}_r). \quad (4.24)$$

Иако је изложени теоријски модел расподеле у потпуности оправдан, није јасно како долазимо до параметра концентрације. Приметимо да је у питању величина обрнуто пропорционална грубости материјала ρ . Дифузно осветљење је карактеристично за делове предмета који нису доволно сјајни, и обратно. Зато користимо идентитет $\kappa = 1/\rho$. За параметар ρ је задужен просторни MLP, о чему ће бити више речи мало касније.

4.3.2 Параметризација и обучавање

Неуронске мреже које чине параметризацију Ref-NeRF-а су у нешто измењеном облику него што смо до сада имали прилике да видимо. Непосредни излаз просторног MLP-а је сада обогаћен новим величинама. У питању су сви они параметри на које смо се до сада позивали, али и неке које до сад нисмо имали прилике да поменемо. Према томе, величине које просторни MLP треба да научи су:

- \mathbf{c}_s , \mathbf{c}_d – спекуларна и дифузна боја светlostи,
- \mathbf{s} – нијанса спекуларне светlostи,
- $\hat{\mathbf{n}}$ – нормала у датој тачки \mathbf{x} ,
- \mathbf{b} – латентна репрезентација положаја \mathbf{x} .

Напоменимо још једном да баратамо са посебним величинама које се тичу дифузног и спекуларног осветљења, тј. боја има своју дифузну и спекуларну компоненту. Остаје да их објединимо у циљу добијања коначне боје. Како дифузно осветљење зависи само од положаја, користићемо

$$\mathbf{c} = \Gamma(\mathbf{c}_d + \mathbf{s} \odot \mathbf{c}_s), \quad (4.25)$$

где је Γ функција преноса у стандардизовани RGB систем [11].

У основи овог модела се налази mip-NeRF. За просторни MLP се користи идентична архитектура, али се за потребе усмереног MLP-а користи 8 слојева са по 256 неурона и ReLU функцијом активације. Приказ архитектуре и редоследа операција је дат на слици 4.5.

Обучавање се спроводи на исти начин као и у случају mip-NeRF-а, тако да то нећемо описивати поново.

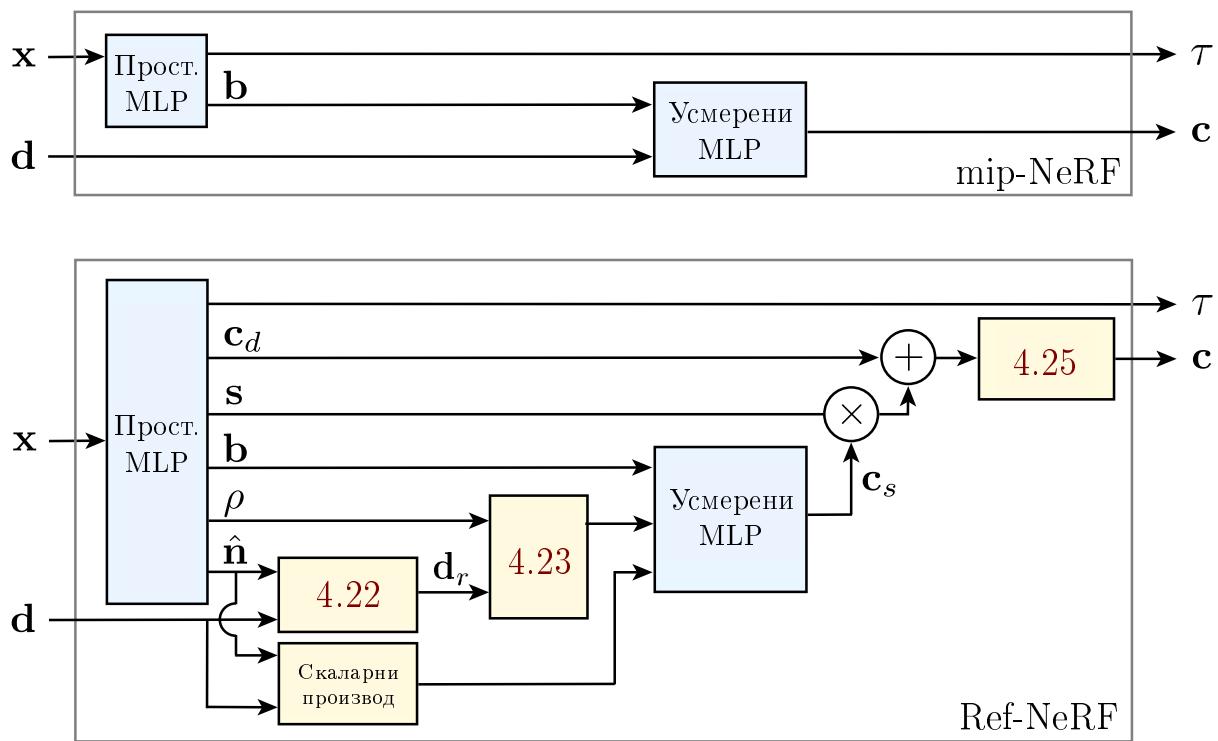
Међутим, као главна разлика у погледу параметризације између поменутих модела је у већем броју параметара у корист Ref-Nerf-а. У табели 6.6 је приказан њихов однос.

4.3.3 Израчунавање вектора нормале

Вектор нормале у некој тачки се тривијално рачуна из усмереног MLP-а. Може се користити идентитет

$$\hat{\mathbf{n}}(\mathbf{x}) = -\frac{\nabla \tau(\mathbf{x})}{\|\nabla \tau(\mathbf{x})\|}. \quad (4.26)$$

Међутим овај метод је изузетно непрецизан јер се од самог почетка ослања на недовољно обучен усмерени модел.



Слика 4.5: Поређење mip-NeRF и Ref-NeRF архитектура

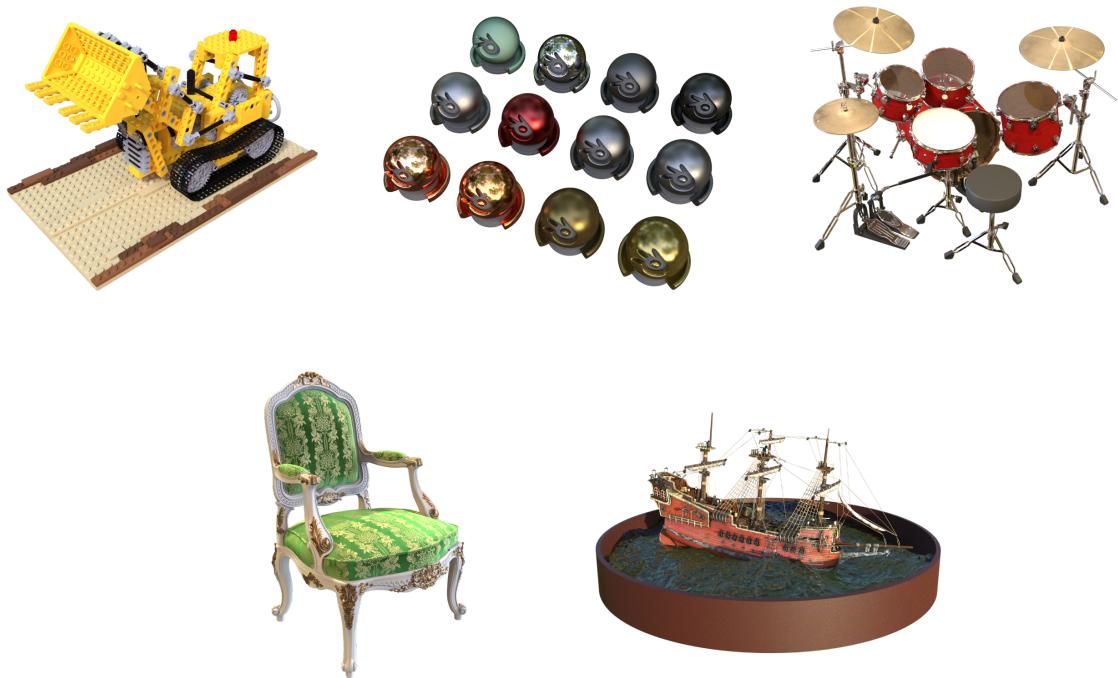
Глава 5

Скупови података

Скупови података у области неуронских поља зрачења се уграбо могу поделити на две групе – да ли је позната калибрација камере или не. У овом раду смо се определили за прву групу. Разлози су пре свега техничке природе. Проблем калибрације камере је један од централних проблема рачунарског вида и као такав је добро испитан. Може се рећи да је овај проблем *лоше условљен*. Готово незнатне разлике могу довести до значајних грешака у пројекцијама, а врло прецизна калибрација изискује и скупу опрему. Олакшавајућа околност је да су познати поступци одређивања унутрашњих и спољашњих параметара камере, али јасно је да они не могу бити ни близу прецизни као што је, на пример, роботска калибрација. С тим у вези, мишљења смо да је оправдано користити синтетички генерисане скупове података у којима ће камера бити постављена на унапред одређено место, а сви њени параметри ће бити познати по дефиницији.

За потребе овог рада, коришћени су LEGO, MATERIALS, DRUMS, CHAIR и SHIP. У питању су познати, референтни скупови података. Модели ће бити упоређени на сваком од њих у истом окружењу. Конкретно, у питању су редом багер од Лего коцки, 16 лоптица различитих материјала, бубњеви, столица и брод. На слици 5.1 је приказан по један поглед за сваки од скупова.

Сваки скуп података је подељен на три подскупа – за обучавање, проверу и тестирање. У подскуповима за обучавање и проверу се налази тачно 100 газличитих погледа, док се за тестирање користи 200. Рендери су синтетички, резолуције 800×800 пиксела без радијалне и тангенцијалне дисторзије.



Слика 5.1: По један поглед из сваког скупа података

Глава 6

Експерименти

Модели ће бити упоређени по свакој од метрика које ћемо навести и објаснити у наставку. Такође, како је дужина трајања обучавања од великог значаја, посебну пажњу ћемо посветити и том аспекту.

Обучавање је вршено употребом сервиса AzureML са избором параметара у складу са изворним радовима, са само једним изузетком – параметар `batch_size` је мањи услед просторних ограничења. Графички процесор који је коришћен носи ознаку NVIDIA Tesla K80 са 24GB сопствене меморије. Имплементација је изведена у програмском језику Python користећи PyTorch [12].

Поштовани су изворни избори параметара, са само једним изузетком – параметар `batch_size` је мањи услед просторних ограничења.

6.1 Метрике

PSNR (енг. *Peak Signal-to-Noise Ratio*). Нека је дата монокроматска слика I и њена априксимација K . Средњеквадратно одступање ове две слике је

$$\text{MSE} = \frac{1}{mn} \sum_{i=0}^{m-1} \sum_{j=0}^{n-1} (I_{i,j} - K_{i,j})^2. \quad (6.1)$$

Однос сигнала и шума је

$$\text{PSNR} = 10 \log_{10} \left(\frac{I_{\max}^2}{\text{MSE}} \right) [\text{dB}], \quad (6.2)$$

где је I_{\max} највећа могућа вредност коју пиксел на слици може имати. Како се канал слике обично представља једним бајтом, ова вредност у већини случајева износи 255.

Виште вредности ове метрике указују на бољу априксимацију.

LPIPS (енг. *Learned Perceptual Image Patch Similarity*) [13]. Идеја је имати метрику која опонаша људску процену сличности две слике. У ту сврху се користи модел \mathcal{F} обучен на сликовном скупу података. Не постоје никаква ограничења у погледу архитектуре, па чак ни скупа података, али се показује да VGG [14] и AlexNet [15] у комбинацији са ImageNet-ом врло добро раде у пракси.

Означимо са I троканалну слику, а са K њену апроксимацију. Иако слојеви модела \mathcal{F} не морају имати ни улаз ни излаз димензије 2, у овом тренутку ће бити лакше да претпоставимо да то јесте случај. Пре свега из рачунских разлога. Уколико \mathcal{F} захтева тензоре неке друге димензије, увек се томе можемо прилагодити једноставним преобличавањем.

Нека је $\alpha_{i,j,k}^I$ резултат примене активационе функције у слоју k у врсти i и колони j за улазну слику I . Ове вредности су нормализоване канал по канал. Аналогно имамо и $\alpha_{i,j,k}^K$.

$$\text{LPIPS} = \sum_{k=1}^L \frac{1}{N_i N_j} \sum_{i=0}^{N_i-1} \sum_{j=0}^{N_j-1} \|w_k \cdot (\alpha_{i,j,k}^I - \alpha_{i,j,k}^K)\|^2. \quad (6.3)$$

Ниже вредности ове метрике указују на бољу апроксимацију.

SSIM (енг. *Structual Similarity Index Measure*) [16]. Нека је дата монохроматска слика I и њена апроксимација K .

$$\text{SSIM} = \frac{(2\mu_I\mu_K + c_1)(2\sigma_{I,K} + c_2)}{(\mu_I^2 + \mu_K^2 + c_1)(\sigma_I^2 + \sigma_K^2 + c_2)}, \quad (6.4)$$

где су μ_I и μ_K узорачке средине, σ_I^2 и σ_K^2 узорачке дисперзије, а $\sigma_{I,K}$ узорачка коваријанса између I и K . Вредности c_i су дефинисане као $(k_i I_{\max})^2$, где је су k_i константе и то обично редом 0.01 и 0.03.

У пракси се показује да овако дефинисана метрика неће дати увек задовољавајуће резултате. Зато се она ретко примењује на целој слици, већ се слика дели на мање делове, а резултати метрике потом упросече.

Више вредност ове метрике указују на бољу апроксимацију.

6.2 Резултати

модел	PSNR \uparrow	LPIPS \downarrow	SSIM \uparrow
NeRF	34.38	0.032	0.973
mip-NeRF	35.80	0.026	0.978
Ref-NeRF	35.79	0.024	0.979

Табела 6.1: Резултати за LEGO скуп података

модел	PSNR \uparrow	LPIPS \downarrow	SSIM \uparrow
NeRF	30.45	0.055	0.956
mip-NeRF	30.58	0.052	0.958
Ref-NeRF	35.71	0.028	0.985

Табела 6.2: Резултати за MATERIALS скуп података

модел	PSNR \uparrow	LPIPS \downarrow	SSIM \uparrow
NeRF	25.28	0.080	0.929
mip-NeRF	25.52	0.079	0.932
Ref-NeRF	25.52	0.073	0.934

Табела 6.3: Резултати за DRUMS скуп података

модел	PSNR \uparrow	LPIPS \downarrow	SSIM \uparrow
NeRF	34.93	0.029	0.979
mip-NeRF	35.20	0.028	0.981
Ref-NeRF	35.84	0.022	0.984

Табела 6.4: Резултати за CHAIR скуп података

модел	PSNR \uparrow	LPIPS \downarrow	SSIM \uparrow
NeRF	29.94	0.161	0.878
mip-NeRF	30.53	0.154	0.884
Ref-NeRF	29.50	0.164	0.871

Табела 6.5: Резултати за SHIP скуп података

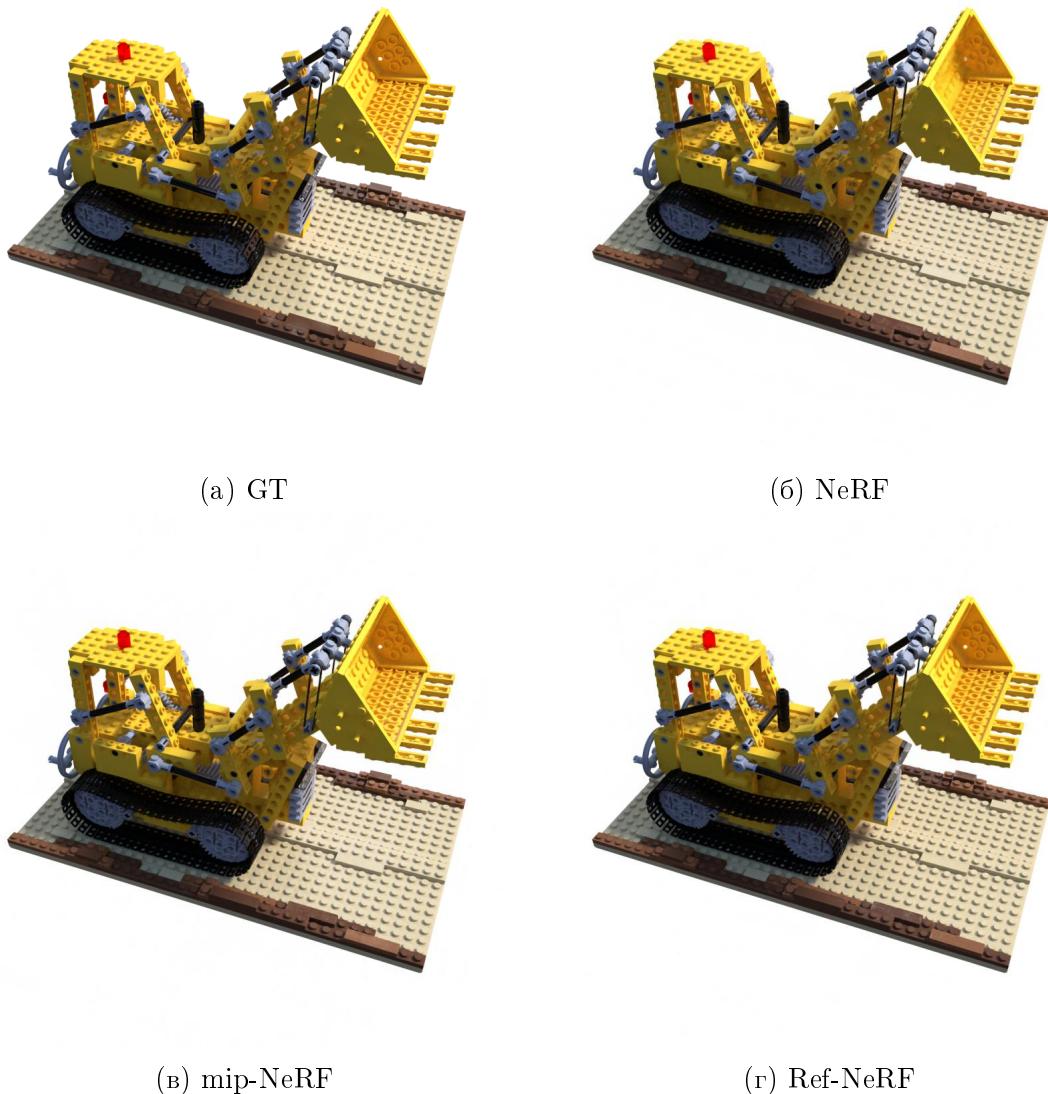
6.3 Време обучавања

Дужина трајања обучавања не зависи од скупа података, а сви модели су обучавани тачно одређен број итерација.

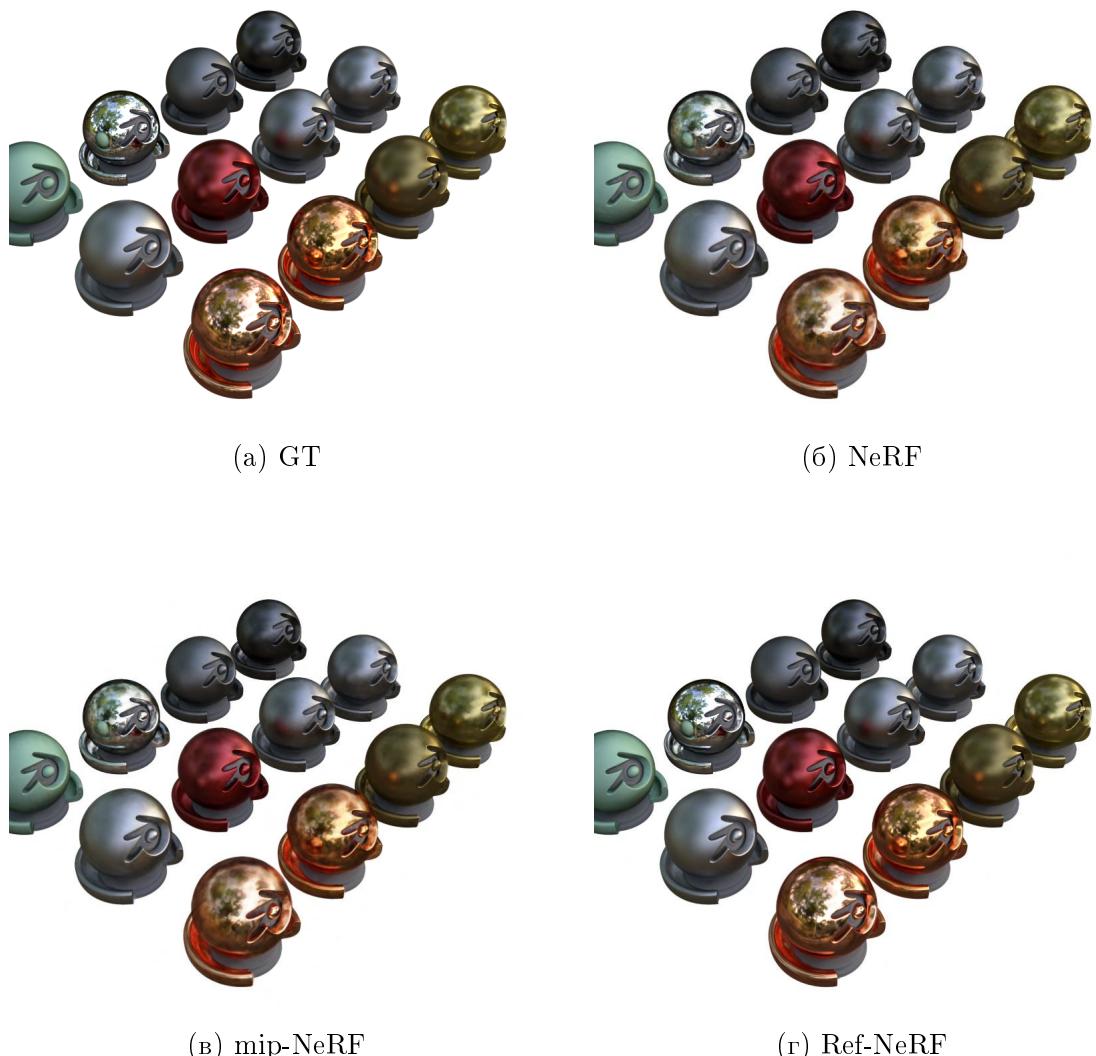
модел	време обучавања (у сатима)	број итерација	број параметара
NeRF	120	1000000	1200000
mip-NeRF	127	1000000	612000
Ref-NeRF	135	1000000	1100000

Табела 6.6: Време обучавања за сваки модел

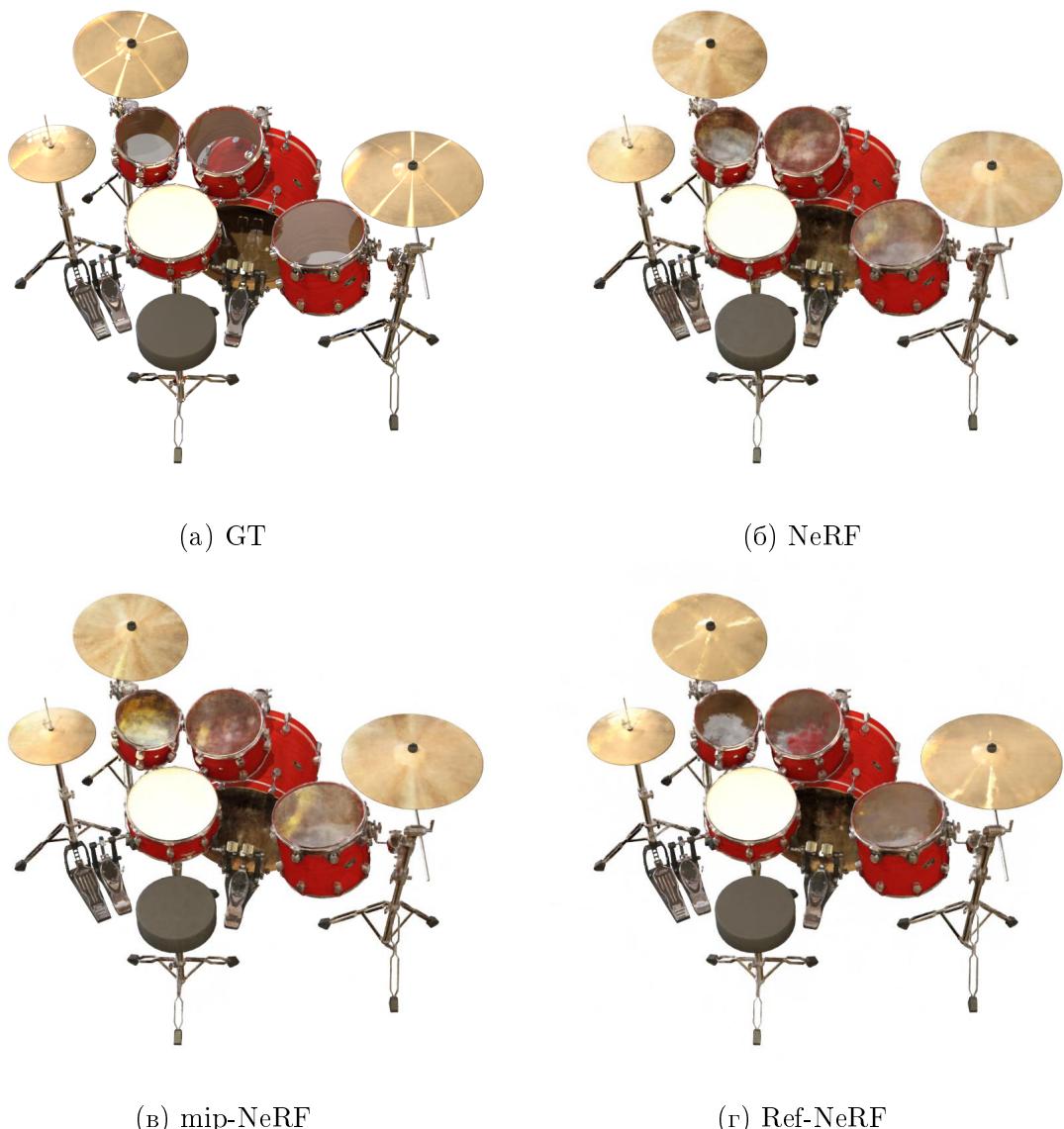
6.4 Квалитет резултата



Слика 6.1: Приказ резултата LEGO скупа података



Слика 6.2: Приказ резултата MATERIALS скупа података



Слика 6.3: Приказ резултата DRUMS скупа података



(a) GT



(б) NeRF

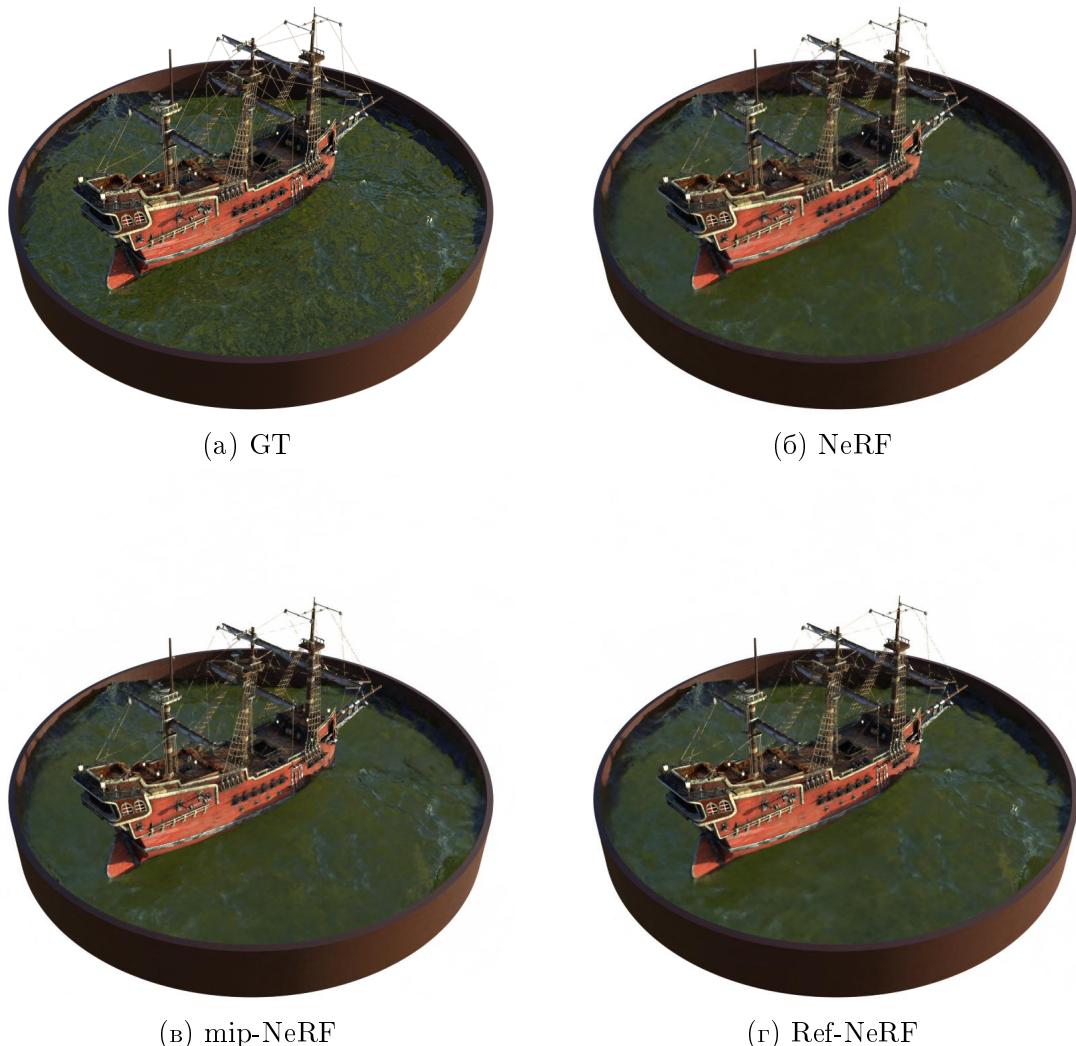


(в) mip-NeRF



(г) Ref-NeRF

Слика 6.4: Приказ резултата CHAIR скупа података



Слика 6.5: Приказ резултата SHIP скупа података

6.5 Детаљно поређење

Само гледањем резултата у претходној глави се поставља значајно питање – постоје ли очите разлике у квалитету? На основу метрика можемо видети да нема разлике у реду величине квалитета, али да очито постоје оне суптилније. Зато ћемо се позабавити појединачним поређењем парова модела, а примере у којима то није изражено нећемо ни наводити.

- 6.5.1 NeRF и mip-NeRF
- 6.5.2 NeRF и Ref-NeRF
- 6.5.3 mip-NeRF и Ref-NeRF

Глава 7

Закључак

У овом раду су неуронска поља зрачења у потпуности теоријски заснована. Објашњени су основни физички и математички концепти, уз одговарајућа образложења и разлоге увођења. Сви модели су успешно имплементирани, а искоришћен је и пун потенцијал решења у облаку када је реч о њиховом обучавању. Резултати су приказани, а модели упоређени. Можемо рећи да су постигнути резултати у складу са ауторским радовима.

Можемо рећи да је Ref-NeRF однео победу и да се намеће као први избор. Иако је за обучавање овог модела потребно највише времена, у питању је разлика од највише 10%. Побољшања су више него приметна без готово икаквих погоршања. Преостала два модела нису *значајно* лошија, тако да и њихово коришћење може дати задовољавајуће резултате.

Даљи рад се може сагледати на основу недостатаца које су обрађени модели недвосмислено показали.

- Неопходна је огромна рачунарска моћ. Обучавање траје дugo, а резултати још увек нису бољи, па чак ни једнаки, у поређењу са традиционалним приступима.
- Нити један од обрађених модела не може да ради у реалном времену.
- Неопходно је имати веома прецизну поставку сцене, а добијени модел је обучен за само једну сцену.

На основу свега досад изреченог, можемо закључити да неуронска поља зрачења имају светлу будућност како у погледу истраживања, тако и у погледу примене. У питању је млада област која има велики потенцијал.

Библиографија

- [1] B. Mildenhall, P. P. Srinivasan, M. Tancik, J. T. Barron, R. Ramamoorthi, and R. Ng, “Nerf: Representing scenes as neural radiance fields for view synthesis,” in *ECCV*, 2020.
- [2] J. T. Barron, B. Mildenhall, M. Tancik, P. Hedman, R. Martin-Brualla, and P. P. Srinivasan, “Mip-nerf: A multiscale representation for anti-aliasing neural radiance fields,” *ICCV*, 2021.
- [3] D. Verbin, P. Hedman, B. Mildenhall, T. Zickler, J. T. Barron, and P. P. Srinivasan, “Ref-NeRF: Structured view-dependent appearance for neural radiance fields,” *CVPR*, 2022.
- [4] K. Fukushima, “Cognitron: A self-organizing multilayered neural network,” *Biological Cybernetics*, vol. 20, pp. 121–136, 1975.
- [5] J. L. Sejun Park, Chulhee Yun and J. Shin, “Minimum width for universal approximation,” in *Proceedings of American Federation of Information Processing Societies: 1977 National Computer Conference*, ICLR, 2021.
- [6] J. Kiefer and J. Wolfowitz, “Stochastic estimation of the maximum of a regression function,” *The Annals of Mathematical Statistics*, vol. 23, pp. 462–466, 1952.
- [7] J. Deng, W. Dong, R. Socher, L.-J. Li, K. Li, and L. Fei-Fei, “Imagenet: A large-scale hierarchical image database,” in *2009 IEEE conference on computer vision and pattern recognition*, pp. 248–255, Ieee, 2009.
- [8] A. Vaswani, N. Shazeer, N. Parmar, J. Uszkoreit, L. Jones, A. N. Gomez, L. Kaiser, and I. Polosukhin, “Attention is all you need,” 2017.
- [9] M. Tancik, P. P. Srinivasan, B. Mildenhall, S. Fridovich-Keil, N. Raghavan, U. Singhal, R. Ramamoorthi, J. T. Barron, and R. Ng, “Fourier features let networks learn high frequency functions in low dimensional domains,” 2020.
- [10] N. L. Max and M. Chen, “Local and global illumination in the volume rendering integral,” in *Scientific Visualization: Advanced Concepts*, 2010.
- [11] M. Anderson, R. Motta, S. Chandrasekar, and M. Stokes, “Proposal for a standard default color space for the internet—srgb,” 1996.

- [12] A. Paszke, S. Gross, F. Massa, A. Lerer, J. Bradbury, G. Chanan, T. Killeen, Z. Lin, N. Gimelshein, L. Antiga, A. Desmaison, A. Kopf, E. Yang, Z. DeVito, M. Raison, A. Tejani, S. Chilamkurthy, B. Steiner, L. Fang, J. Bai, and S. Chintala, “Pytorch: An imperative style, high-performance deep learning library,” in *Advances in Neural Information Processing Systems 32*, pp. 8024–8035, Curran Associates, Inc., 2019.
- [13] R. Zhang, P. Isola, A. A. Efros, E. Shechtman, and O. Wang, “The unreasonable effectiveness of deep features as a perceptual metric,” in *2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 586–595, 2018.
- [14] K. Simonyan and A. Zisserman, “Very deep convolutional networks for large-scale image recognition,” in *International Conference on Learning Representations*, 2015.
- [15] A. Krizhevsky, I. Sutskever, and G. E. Hinton, “Imagenet classification with deep convolutional neural networks,” in *Advances in Neural Information Processing Systems* (F. Pereira, C. Burges, L. Bottou, and K. Weinberger, eds.), vol. 25, Curran Associates, Inc., 2012.
- [16] Z. Wang, A. Bovik, H. Sheikh, and E. Simoncelli, “Image quality assessment: from error visibility to structural similarity,” *IEEE Transactions on Image Processing*, vol. 13, no. 4, pp. 600–612, 2004.