

Slovenská technická univerzita v Bratislave
Fakulta informatiky a informačných technológií

FIIT-XXXX-XXXXX

Bc. Katarína Juhásová

REGISTRÁCIA OBJEKTOV NA MRAČNE BODOV
POMOCOU HLBOKÝCH NEURÓNOVÝCH SIETÍ

Priebežná správa o riešení DP2

Vedúci práce: RNDr. Andrej Lúčny, PhD.

Január 2022

Slovenská technická univerzita v Bratislave
Fakulta informatiky a informačných technológií

FIIT-XXXX-XXXXX

Bc. Katarína Juhásová

REGISTRÁCIA OBJEKTOV NA MRAČNE BODOV
POMOCOU HLBOKÝCH NEURÓNOVÝCH SIETÍ

Priebežná správa o riešení DP2

Študijný program:	Inteligentné softvérové systémy
Študijný odbor:	Informatika
Miesto vypracovania:	Ústav informatiky, informačných systémov a softvérového inžinierstva
Vedúci práce:	RNDr. Andrej Lúčný, PhD.
Január 2022	

Návrh zadania diplomovej práce

Finálna verzia do diplomovej práce¹

Študent:

Meno, priezvisko, tituly: Katarína Juhásová, Bc.
Študijný program: Inteligentné softvérové systémy
Kontakt: katkajuhasova8@gmail.com

Výskumník:

Meno, priezvisko, tituly: Andrej Lúčný, RNDr. PhD.

Projekt:

Názov: Registrácia objektov na mračne bodov pomocou hlbokých neurónových sietí
Názov v angličtine: Object Registration in 3D Point Clouds Using Deep Neural Networks
Miesto vypracovania: Ústav informatiky, informačných systémov a softvérového inžinierstva, FIIT STU
Oblasť problematiky: hlboké učenie

Text návrhu zadania²

Mračná bodov sú významnou štruktúrou v oblasti počítačového videnia, avšak výskum ich spracovania pomocou hlbokého učenia je stále v počiatočnom štádiu. Pri ich spracovaní je potrebné brať do úvahy komplikácie vyplývajúce zo štruktúry, ktorá sa vyznačuje svojou nepravidelnosťou a nesúmernosťou. Doterajší výskum však ukázal, že hlboké neurónové siete umožňujú vysporiadanie sa s týmito problémami, čím je umožnené vykonávať rôzne úlohy nad mračnami bodov, napríklad detekciu objektov, ich klasifikáciu alebo registráciu, t.j. nájdenie zadaných objektov v danom mračne.

Naštudujte problematiku spracovania 3D skenov a hĺbkových máp v miere potrebnej pre vytvorenie vhodnej sady dát. Spracujte poskytnuté dáta získané anotáciou mračen bodov zodpovedajúcim nasnímaným pneumatikám. Využite na to knižnice OpenCV, prípadne PCL. Postup vytvorenia sady dát popíšte v práci.

Oboznámte sa s problematikou hlbokého učenia. Uvedte práce, ktoré ideovo predchádzali sieťam spracúvajúcim mračná bodov (ako je napríklad PointNet) v miere potrebnej pre pochopenie činnosti týchto sietí.

Vyskúšajte vyriešiť problém registrácie objektov pomocou hlbokkej neurónovej siete vhodnej architektúry a zhodnoťte kvalitu tohto riešenia. Porovnajte možnosti riešenia založeného na hlbokom učení s klasickými metódami počítačového videnia. Na implementáciu riešenia použite prostriedky ako GPU, TensorFlow + Keras, Python, OpenCV a ďalšie.

¹ Vytlačiť obojstranne na jeden list papiera

² 150-200 slov (1200-1700 znakov), ktoré opisujú výskumný problém v kontexte súčasného stavu vrátane motivácie a smerov riešenia

Literatúra³

- Brownlee, J., 2019. Deep Learning for Computer Vision: Image Classification, Object Detection, and Face Recognition in Python. Machine Learning Mastery.
- Qi, C.R., Su, H., Mo, K. and Guibas, L.J., 2017. PointNet: Deep Learning on Point Sets for 3D Classification and Segmentation. In Proceedings of the IEEE conference on computer vision and pattern recognition (pp. 652-660).

Vyššie je uvedený návrh diplomového projektu, ktorý vypracoval(a) Bc. Katarína Juhásová, konzultoval(a) a osvojil(a) si ho RNDr. Andrej Lúčny, PhD. a súhlasí, že bude takýto projekt viesť v prípade, že bude pridelený tomuto študentovi.

V Bratislave dňa 28.1.2021

Podpis študenta

Podpis výskumníka

Vyjadrenie garanta predmetov Diplomový projekt I, II, III

Návrh zadania schválený: áno / nie⁴

Dňa:

Podpis garanta predmetov

³ 2 vedecké zdroje, každý v samostatnej rubrike a s údajmi zodpovedajúcimi bibliografickým odkazom podľa normy STN ISO 690, ktoré sa viažu k téme zadania a preukazujú výskumnú povahu problému a jeho aktuálnosť (uvedte všetky potrebné údaje na identifikáciu zdroja, pričom uprednostnite vedecké príspevky v časopisoch a medzinárodných konferenciách)

⁴ Nehodiace sa prečiarknite

ČESTNÉ PREHLÁSENIE

Čestne vyhlasujem, že som túto prácu vypracovala samostatne, na základe konzultácií a s použitím uvedenej literatúry.

V Bratislave, 3. január 2022

.....
Bc. Katarína Juhásová

POĎAKOVANIE

Ďakujem môjmu školiteľovi RNDr. Andrejovi Lúčnemu, PhD. za ochotu, odbornú pomoc, užitočné pripomienky a cenné rady, ktoré mi poskytol pri vypracovaní diplomovej práce.

Anotácia

Slovenská technická univerzita v Bratislave

FAKULTA INFORMATIKY A INFORMAČNÝCH TECHNOLOGIÍ

Študijný program: Inteligentné softvérové systémy

Autor: Bc. Katarína Juhásová

Diplomová práca: Registrácia objektov na mračne bodov pomocou
hlbokých neurónových sietí

Vedúci diplomovej práce: RNDr. Andrej Lúčny, PhD.

Január 2022

Využitie hlbokého učenia je vo sfére počítačového videnia veľmi rozšírené, prevažne však ide o spracovanie 2D dát, zatiaľ čo spracovanie 3D dátových štruktúr zaostáva. Príkladom takejto štruktúry sú mračná bodov vyznačujúce sa svojou nepravidelnosťou, nesúmernosťou a nekonzistentnou hustotou bodov, ktoré nie sú prirodzene utriedené. Vo väčšine prípadov sú tieto problémy riešené transformáciou dát na pravidelnú volumetrickú štruktúru ako voxely alebo na sekvenciu 2D snímok. Doterajší výskum však ukázal, že hlboké neurónové siete umožňujú riešenie týchto problémov aj bez špeciálnych transformácií. Vďaka tomu je možné vykonať úlohy ako klasifikácia objektov alebo ich detekcia priamo nad mračnami bodov. Predložená priebežná správa o riešení DP1 obsahuje analýzu rôznych prístupov k spracovaniu mračien bodov pomocou hlbokého učenia, pričom špeciálny dôraz je kladený na model PointNet [1]. V rámci tohto modelu boli úspešne vyriešené viaceré vyššie uvedené problémy, čo viedlo k efektívnemu získavaniu vektorovej reprezentácie črt zo vstupných mračien bodov. V závere správy je opísaný návrh riešenia problému registrácie objektov v mračne bodov pomocou vlastnej hlbokéj neurónovej siete. Vo výskumnej časti práce budú použité mračná bodov predstavujúce 3D skeny pneumatík, v ktorých budú hľadané objekty ako písmená, čísla a ďalšie znaky.

Annotation

Slovak University of Technology Bratislava

FACULTY OF INFORMATICS AND INFORMATION TECHNOLOGIES

Degree Course: Intelligent Software Systems

Author: Katarína Juhásová

Master's Thesis: Object Registration in 3D Point Clouds Using Deep
Neural Networks

Supervisor: Dr. Andrej Lúčný

2022, January

Deep learning has various applications in the computer vision domain, however, it is primarily used for image processing, while the processing of 3D data structures is left behind. An example of such a structure is the point cloud. Numerous point cloud characteristics such as their irregularity, asymmetry and inconsistent density are considered problematic, since they may cause issues during feature extraction. Therefore, these issues are in most cases solved by transformation of the data into regular volumetric structures, such as voxels or sequences of 2D images. Nevertheless, in recent years it has been proven that deep neural networks are capable of solving these issues without performing specific transformations, which results in the possibility to perform tasks like classification or object detection directly in the point clouds. The presented DP1 report contains an analysis of different deep learning point cloud processing methods, with a separate chapter focused on the model PointNet [1]. PointNet architecture provided solutions for several of the above mentioned problems, which led to an efficient acquisition of a vector feature representation of point clouds. At the end of the report, there is proposed a deep neural network architecture which should be able to perform object registration directly in point clouds. The point clouds that shall be used represent 3D tire scans. Surface of the tires is usually partially covered with letters, numbers and other characters. Proposed neural network for registration should be able to match these symbols.

Obsah

1	Úvod	1
2	Spracovanie mračien bodov hlbokým učením	2
2.1	Spracovanie transformovaných mračien bodov	3
2.2	Spracovanie mračien bodov priamo	4
3	PointNet	6
3.1	Vstupné dáta	6
3.2	Architektúra PointNet	7
3.2.1	Aproximácia symetrickej funkcie	8
3.2.2	Agregácia reprezentácie lokálnych a globálnych črt	9
3.2.3	Vnorené siete na zarovnanie vstupu a extrahovaných črt	10
3.3	Klasifikácia a segmentácia pomocou PointNet	11
4	Registrácia objektov v mračných bodov	13
4.1	PointNetLK	13
4.2	CorsNet	14
5	Návrh riešenia	17
6	Implementácia modelu a evaluácia	20
6.1	PointNet implementácia	20
6.2	PointNet trénovanie a evaluácia	21
6.3	Predspracovanie dát pre CorsNet	22
6.4	CorsNet implementácia	22
6.5	CorsNet trénovanie	23
6.6	CorsNet experimenty	24
6.6.1	Experimenty - predspracovanie dát	25
6.6.2	Experimenty - augmentácia dát	26
6.6.3	Experimenty - transfer learning	27

6.7	Vstupné dáta a predspracovanie	27
7	Zhodnotenie analýzy	30
7.1	Ďalšia práca	31
7.1.1	PointNet - ďalší vývoj	31
7.1.2	CorsNet - ďalší vývoj	31
	Literatúra	33
	Príloha A: Plán práce na riešení projektu	A-1
A.1	DP1	A-1
A.2	DP2	A-2
A.3	DP3	A-2

Zoznam použitých skratiek

CNN	Convolutional Neural Network
FCN	Fully Convolutional Network
MLP	Multilayer Perceptron
RPN	Region Proposal Network
SVD	Singular Value Decomposition
MSE	Mean Squared Error
RMSE	Root Mean Squared Error

1 Úvod

Metódy hlbokého učenia v dnešnej dobe riešia široké spektrum problémov, avšak zatiaľ čo určité oblasti, ako spracovanie obrazu sú vo veľkom preskúmané, iné oblasti zaostávajú. Príkladom je spracovanie mračien bodov, ktoré sú ako dátová štruktúra významné, ale výskum ich spracovania pomocou hlbokého učenia je stále v počiatočnom štádiu. Pri práci s mračnami bodov je potrebné brať do úvahy viaceré komplikácie vyplývajúce primárne z ich štruktúry, ktorá sa vyznačuje svojou nepravidelnosťou a nesúmernosťou. Hlboké neurónové siete však umožňujú riešiť tieto problémy a následne vykonať rôzne úlohy nad mračnami bodov.

Vo veľkej časti existujúcich modelov hlbokého učenia sú problémy vyplývajúce zo štruktúry mračien bodov riešené transformáciou mračien na 2D obrázky alebo 3D voxely. Pri takýchto transformáciách však často dochádza k zbytočnému zväčšeniu objemu vstupných dát, pričom sa stráca aj určitá presnosť a detaily zachytené v pôvodných mračnách bodov. Alternatívnym prístupom je teda spracovanie mračien priamo, pričom však musí byť architektúra neurónovej siete prispôbena na spracovanie takýchto dát a získanie vhodnej reprezentácie črt. Príkladom takejto hlbkej neurónovej siete je PointNet [1], ktorá dokáže úspešne extrahovať lokálne globálne črty pre mračná bodov reprezentované ako neusporiadané množiny bodov. Nad získanými reprezentáciami črt potom umožňuje vykonať klasifikáciu a segmentáciu 3D objektov. Spôsob extrakcie črt použitý v modeli PointNet bude implementovaný aj v neurónovej sieti, ktorá bude výstupom tejto diplomovej práce.

Cielom práce je vyriešiť problém registrácie objektov v mračnách bodov pomocou hlbokého učenia. Úloha bude riešená na datasete skenov pneumatík a vzorov na ich stranách. Pri registrácii objektov je známe, aké objekty by sa mali vo vstupných dátach nachádzať a úlohou neurónovej siete je ich nájsť. V prípade, že požadovaný objekt nebude nájdený, znamená to, že sa v mračne nenachádza, prípadne je výrazne zdeformovaný alebo nekompletný. Výsledná neurónová sieť by sa však mala vedieť vysporiadať s deformáciami spôsobenými zakrivením povrchu pneumatiky a mala by byť schopná nájsť aj týmto spôsobom zdeformované objekty.

2 Spracovanie mračien bodov hlbokým učením

Modely hlbokého učenia sú používané na aproximáciu funkcií, ktoré buď nie je možné priamo vyjadriť analyticky, alebo ich vyjadrenie je príliš náročné a neintuitívne. Hlboké neurónové siete je možné opísať ako viacvrstvové výpočtové modely využívajúce nelineárne prechody medzi vrstvami na extrahovanie črt zo vstupných dát. Medzi vrstvami platí, že vysokoúrovňové črty sú definované pomocou nízkoúrovňových, pričom tento hierarchický vzťah je označovaný ako hlboká architektúra [2]. Vstupné dáta sú pri prechode sieťou transformované na každej vrstve, pričom výstupy z konkrétnej vrstvy sú kombináciou výstupov z predošlej. Týmto spôsobom je na každej vrstve získaná reprezentácia vstupu s rôznymi úrovňami abstrakcie, pričom na vrstvách hlbšie v sieti je miera tejto abstrakcie vyššia. Nad abstraktnými reprezentáciami vstupu je ďalej možné vykonávať rôzne úlohy v závislosti od požadovanej funkcionality siete, napríklad klasifikáciu, segmentáciu alebo detekciu objektov.

Pri riešení spomenutých úloh nad 3D mračnami bodov je potrebné sa vysporiadať s určitými problémami, ako nepravidelnosť dát alebo variabilná hustota bodov v mračnách. Často v závislosti od spôsobu, akým boli dáta získané sú mračná bodov nepravidelné a miestami veľmi riedke, napríklad pri LiDAR skenoch z ulíc, alebo naopak príliš husté, čo je často prípad pri dátach z industriálneho prostredia [3]. V takej situácii môže hustota bodov skomplikovať identifikáciu hrán medzi segmentami pri segmentácii objektov. Metódy získavania dát vedia taktiež ovplyvniť prítomnosť šumu alebo iných nepresností a chýb spôsobených meraním. Podľa [4] existujú tri hlavné spôsoby prístupu k riešeniu problému nepravidelnosti mračien bodov a ich spracovaniu pomocou hlbokého učenia:

1. transformácia mračien bodov na voxely alebo sekvenciu obrázkov a následné spracovanie pomocou CNN,
2. spracovanie mračien bodov priamo pomocou MLP,
3. spojením týchto dvoch metód [5, 6].

2.1 Spracovanie transformovaných mračien bodov

Prvým častým prístupom k spracovaniu mračien bodov je ich transformácia na pravidelné 3D voxely. VoxelNet [7] rieši problém detekcie objektov v priestore. Vstupom pre model síce sú mračná bodov, avšak v prvom kroku sú jednotlivé body z mračien rozdelené do 3D mriežky (voxelov), kde pre každý voxel je vypočítaná jeho vektorová reprezentácia opisujúca obsiahnutú priestorovú informáciu. Voxely sú teda reprezentované ako 4D tenzory, z ktorých je ďalej konvolučnými vrstvami získaná agregovaná priestorová informácia a následne pomocou plne konvolučnej RPN (angl. *Region Proposal Network*) vykonaná detekcia.

Podobným spôsobom sú transformované mračná bodov na voxely aj pre model na detekciu objektov Vote3Deep [8]. Každá bunka v 3D mriežke s nenulovým počtom priradených bodov je reprezentovaná vektorom črt obsahujúcim štatistiky z bodov patriacich do danej bunky. Okrem týchto štatistík obsahuje aj binárny indikátor prítomnosti jedného a viac bodov a aj pozíciu v 3D mriežke. Pre bunky s nulovým počtom bodov nie je uložená žiadna vektorová reprezentácia, čo má za následok riedku reprezentáciu vstupu.

Podobne aj v prácach [9, 10, 11, 12] sú mračná bodov najskôr transformované na voxely a tie sú ďalej spracované hlbokým učením. V týchto prácach umožnila transformácia mračien bodov na 3D voxely využitie konvolučných neurónových sietí (angl. *Convolutional Neural Network, CNN*) [13]. V mnohých prípadoch však pri tejto transformácii došlo k zbytočnému nafúknutiu objemu vstupných dát, pričom bola stratená určitá presnosť zachytená v pôvodných mračnách bodov.

Alternatívnou možnosťou reprezentácie mračien bodov pomocou štruktúry s fixnou veľkosťou je transformácia na 2D obrázky. Výhodami tohto prístupu je, že môžu čerpať poznatky z oblasti spracovania obrazu, ktorá je výrazne hlbšie preskúmaná na rozdiel od spracovania mračien bodov. Ďalšou výhodou je, že spracovanie 2D dát je menej výpočtovo náročné v porovnaní s 3D dátami, čo je často významná vlastnosť hlavne pri využití v autonómnych vozidlách. Pang a kolektív [3] využili tieto vlastnosti a predstavili CNN na detekciu objektov v mračnách bodov. 3D problém detekcie objektov transformovali na 2D problém tak, že z pôvodného 3D vstupu bolo vytvorených niekoľko hĺbkových máp (angl. *Depth Map*) z rôznych uhlov pohľadu a detekcia bola vykonaná nad každou mapou zvlášť.

Výstupy z jednotlivých hĺbkových máp sú na záver zlúčené na určenie *bounding boxu* v pôvodnom 3D priestore. Transformácie špecificky na hĺbkové mapy sú použité v práci [14], v ktorej bola použitá CNN na klasifikáciu znakov posunkovej reči.

Transformáciu objektu z mračna bodov na 2D snímky s viacerých uhlov pohľadu využíva aj hlboká neurónová sieť na rozpoznanie a klasifikáciu objektov od Su a kolektív [15]. V rámci modelu sú najskôr pomocou paralelných CNN extrahované črty pre každú snímku samostatne. Následne sú tieto črty agregované do globálnej reprezentácie objektu, nad ktorou je vykonaná finálna klasifikácia.

Vstupné dáta sú transformované na obrázky aj pre Complex-YOLO [4], ktorý umožňuje detekciu 3D objektov. Architektúra tohto modelu vychádza zo *state-of-the-art* konvolučnej neurónovej siete (CNN) na detekciu 2D objektov YOLOv2 [16] v reálnom čase. Vstupné dáta pre Complex-YOLO sú získavané LiDAR laserovými skenermi a pred spracovaním sieťou sú transformované na RGB snímky z vtáčej perspektívy, pričom jednotlivé farebné kanály sú mapované na výšku, hustotu a intenzitu vráteného laserového signálu. Detekcia je vykonávaná nad jedným takýmto obrázkom, čo prispelo k efektívnosti modelu a spracovaniu vstupov v reálnom čase.

Rovnakým spôsobom sú mračná bodov transformované aj pre plne konvolučnú sieť (angl. *Fully Convolutional Network*, FCN) PIXOR [17], ktorá tiež rieši úlohu detekcie objektov na 3D dátach transformovaných do snímky z vtáčej perspektívy. Vylepšením oproti Complex-YOLO je, že výsledné *bounding boxy* určujú nie len pozíciu objektov, ale aj smer ich natočenia.

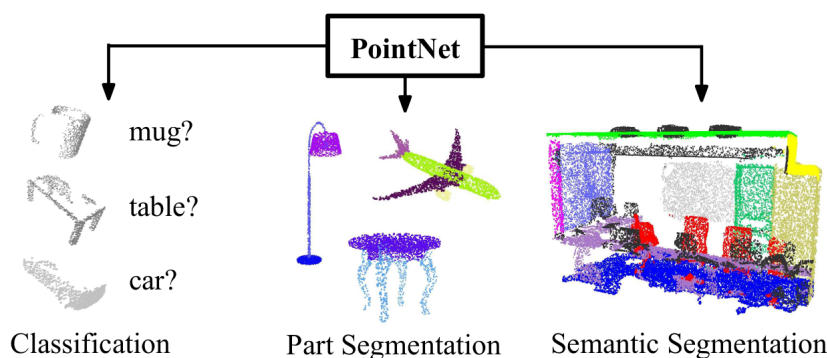
2.2 Spracovanie mračien bodov priamo

Príkladom modelu pracujúceho priamo s mračnami bodov PointNet++ [1]. Tento hierarchický model rekurzívne delí množinu bodov na menšie podskupiny a z nich získava črty s rôznymi úrovňami abstrakcie. Na každej úrovni je z podmnožiny bodov vytvorená nová skupina obsahujúca menší počet elementov. Extrahovanie každej úrovne abstrakcie sa skladá z troch typov vrstiev: *vzorkovacia vrstva*, *zgrupovacia vrstva* a *PointNet vrstva* inšpirovaná nižšie opísaným modelom PointNet [1]. Nad získanými komplexnými reprezentáciami mračien bodov je následne možné vykonať klasifikáciu alebo segmentáciu. Spôsob získavania reprezentácie črt z PointNet++ bol použitý aj v modeli na detekciu objektov opísanom v práci [18].

Ďalším modelom, ktorý umožňuje vykonávať rôzne úlohy priamo nad mračnami bodov je PointNet [1]. Jadrom jeho architektúry sú viacvrstvové perceptróny (MLP) a hoci tento model nepodporuje priamo registráciu objektov, slúžil ako základ na extrakciu črt pre viaceré modely vykonávajúce registráciu objektov. Model PointNet je podrobnejšie opísaný v ďalšej kapitole, keďže bol kľúčový aj pri návrhu nášho riešenia pre registráciu objektov.

3 PointNet

PointNet [1] je hlboká neurónová sieť, ktorá umožňuje priamo spracovať mračná bodov. Nad nimi dokáže vykonať úlohy ako klasifikácia 3D objektov, segmentácia častí 3D objektov alebo sémantická segmentácia scény. Príklady výstupov siete je možné vidieť na obrázku 3.1. Architektúra tejto siete bola navrhnutá tak, aby umožnila spracovanie mračen bodov bez transformácie na pravidelnú dátovú štruktúru vo fáze predspracovania. Model PointNet sa zároveň dokáže úspešne vysporiadať s chybami a šumom vo vstupných dátach, podobne ako aj s riedkymi mračenami.



Obr. 3.1: *PointNet* funkcionalita: klasifikácia, segmentácia častí objektov a sémantická segmentácia scény. [1]

3.1 Vstupné dáta

Vstupné dáta pre PointNet sú reprezentované ako množina bodov, v ktorej každý bod pozostáva z $[x, y, z]$ súradníc a ďalších črt ako farba alebo normála. Táto množina má určité vlastnosti, ktoré bolo potrebné zohľadniť v architektúre siete:

- *množina je neusporiadaná*: pri N počte bodov je potrebné, aby neurónová sieť bola invariantná voči $N!$ permutáciám usporiadaní bodov v rámci množiny.
- *medzi bodmi v množine existujú vzťahy*: body sa nachádzajú v Euklidovskom priestore, teda nie sú izolované, vďaka čomu je možné určiť ich vzájomné

vzdialenosti, pričom blízke body vytvárajú zo sémantického hľadiska podmnožiny. Z toho dôvodu by mal navrhnutý model vedieť lokálne zachytiť menšie štruktúry a ich vzájomné vzťahy.

- *množiny bodov sú invariantné vzhľadom na určité transformácie*: transformácie ako rotácia, zrkadlenie alebo škálovanie by nemali ovplyvniť výstupy siete pri klasifikácii ani segmentácii objektov.

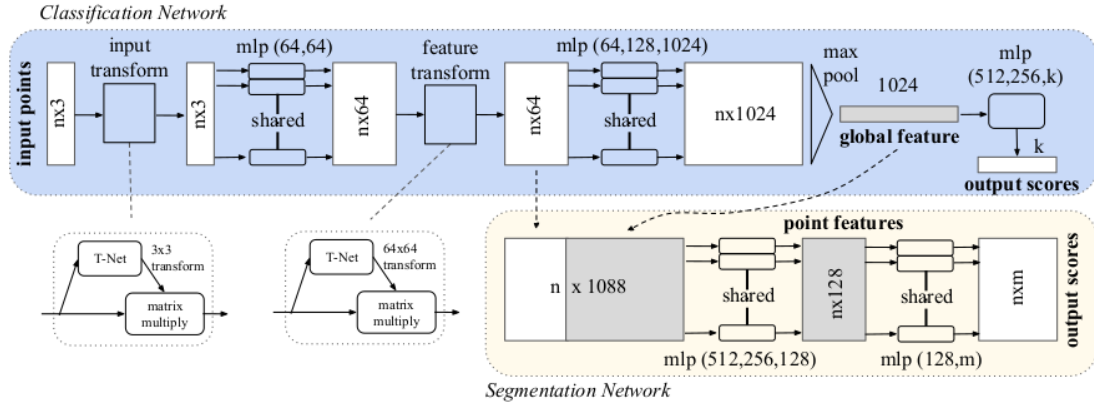
3.2 Architektúra PointNet

Hlboké neurónové siete sú založené na učení reprezentácií. Z toho dôvodu je ich dôležitou vlastnosťou aby vytvorené reprezentácie boli podobné pre príbuzné objekty, napríklad objekty patriace do jednej triedy, a zároveň aby sa reprezentácie rôznych objektov, teda z rôznych tried, odlišovali. Pri vstupoch s presne definovanou štruktúrou a poradím vstupných elementov, ako sú obrázky, je zachovanie tejto vlastnosti jednoduchšie v porovnaní s neusporiadanými množinami bodov. V tomto prípade je extrahovanie črt (reprezentácií) problematickejšie, keďže model sa nemôže spoliehať na vzťahy medzi susediacimi bodmi v množine.

Pre správne fungovanie modelu PointNet bolo potrebné vytvoriť architektúru, ktorá umožní extrahovať črty a identifikovať vzťahy medzi bodmi susediacimi v Euklidovskom priestore, pričom tieto body nemusia byť susedné aj vo vstupnej množine. Navrhnutá architektúra by mala zároveň umožniť identifikáciu črt zo vstupu takým spôsobom, aby aj dáta získané skenovaním rovnakého objektu z rôznych uhlov pohľadu mali takmer zhodné reprezentácie. Architektúru je tiež potrebné prispôbiť úlohe, ktorú má model riešiť. Na klasifikáciu objektov je potrebná globálna reprezentácia objektu, zatiaľ čo pre segmentáciu častí objektov alebo scény sú potrebné nie len globálne informácie ale aj lokálne. Kompletná architektúra siete PointNet je znázornená na obrázku 3.2.

Architektúra neurónovej siete PointNet je založená na 3 kľúčových častiach:

1. aproximácia symetrickej funkcie,
2. agregácia reprezentácie lokálnych a globálnych črt,
3. vnorené siete na zarovnanie vstupu a extrahovaných črt.



Obr. 3.2: PointNet architektúra pre klasifikáciu (Classification Network) a segmentáciu (Segmentation Network). Vstupná množina n bodov je transformovaná, následne sú získané lokálne črty pre body, ktoré sú opäť transformované a použité na získanie globálnych črt. Tieto črty agregované pomocou max pooling sú spracované s MLP na získanie skóre pre klasifikáciu do každej z k tried. V rozšírení siete pre segmentáciu sú zlúčené lokálne a globálne črty, ktorých spracovaním je získané skóre pre každú z m možných kategórií pre všetky vstupné body. Čísla v zátvorkách reprezentujú veľkosti vrstiev v MLP. [1]

3.2.1 Aproximácia symetrickej funkcie

Keďže vstupom pre model je neusporiadaná množina bodov, pri N počte bodov musí byť model invariantný voči $N!$ permutáciám vstupu. Vhodným riešením je práve použitie alebo aproximácia symetrickej funkcie. Z definície symetrickej funkcie vyplýva, že pri n vstupných parametroch x_1, x_2, \dots, x_n je výstup funkcie rovnaký pre všetky permutácie parametrov.

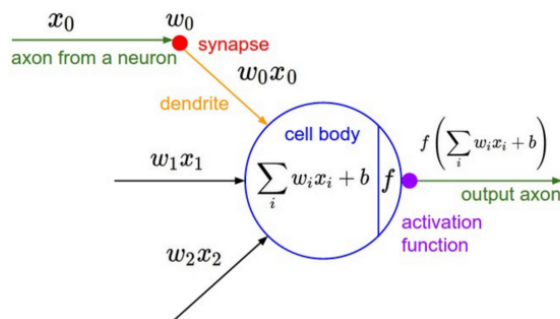
Cielom pri návrhu siete bolo, aby neurónová sieť vedela aproximovať funkciu f aplikovaním symetrickej funkcie na transformované body z množiny:

$$f(x_1, \dots, x_n) \approx g(h(x_1), \dots, h(x_n)), \quad (3.1)$$

kde $f : 2^{\mathbb{R}^N} \rightarrow \mathbb{R}$, $h : \mathbb{R}^N \rightarrow \mathbb{R}^K$ a $g : \underbrace{\mathbb{R}^K \times \dots \times \mathbb{R}^K}_n \rightarrow \mathbb{R}$ je symetrická funkcia.

Funkcia h je aproximovaná pomocou viacvrstvového perceptrónu (angl. *Multilayer Perceptron*, *MLP*) a pre g je použitá kombinácia symetrickej funkcie max pooling a funkcie s jednou premennou h . Aplikovaním rôznych funkcií h je možné extrahovať rôzne črty zo vstupu.

Model PointNet obsahuje niekoľko viacvrstvových perceptrónov (MLP), ktoré tvoria jadro architektúry. MLP sa skladajú zo vstupnej vrstvy, výstupnej vrstvy a jednej alebo viac skrytých vrstiev neurónov medzi nimi. Vrstvy MLP sú plne prepojené, čiže výstup každého neurónu z vrstvy L_n prispieva do vstupov každého z neurónov na vrstve L_{n+1} . Výstupom neurónu je lineárna kombinácia niekoľkých vstupných reálnych čísel, ktorú je možné modifikovať pomocou nelineárnej aktivačnej funkcie. Model jedného neurónu s aktivačnou funkciou je znázornený na obrázku 3.3. Vlastnosti a správanie MLP pri rôznych počtoch skrytých vrstiev a neurónov na nich sú podobnejšie opísané v práci [19].



Obr. 3.3: *Matematický model neurónu. Aktivačná funkcia $f()$ je aplikovaná na vážený súčet vstupov do daného neurónu spolu s pripočítaným bias (koeficientom posunu).* [20]

3.2.2 Agregácia reprezentácie lokálnych a globálnych črt

V časti vyššie je opísané, ako získať pomocou K rôznych funkcií reprezentáciu globálnych črt pre vstupné dáta ako vektor $[f_1, \dots, f_K]$. Takáto reprezentácia je dostačujúca pre klasifikáciu, avšak pre segmentáciu objektov aj scény je potrebná kombinácia lokálnych a globálnych črt, aby bolo možné zachytiť aj informácie o blízkych bodoch vytvárajúcich sémantické podmnožiny. Na obrázku 3.2 je možné vidieť ako je klasifikačná časť siete rozšírená pre segmentáciu objektov. Lokálne a globálne črty získané v rámci pôvodnej klasifikačnej časti sú pre účely segmentácie zlúčené. Následne sú identifikované vlastnosti pre body na základe zlúčených črt.

3.2.3 Vnorené siete na zarovnanie vstupu a extrahovaných črt

Predchádzajúce dva princípy sú založené na extrakcii viacúrovňových črt zo vstupných dát. Pritom však nebol vyriešený problém, že tieto črty by mali byť rovnaké aj keď objekt opísaný bodmi podstúpi rôzne geometrické transformácie ako rotáciu alebo zrkadlenie. Z toho dôvodu sú v modeli PointNet obsiahnuté dve „mini“ siete použité na zarovnanie vstupných bodov a lokálnych črt.

Cieľom týchto transformačných sietí je aproximovať maticu pre afinnu transformáciu samostatne pre vstup a neskôr aj pre jeho lokálne vlastnosti. Na obrázku 3.2 sú tieto siete označené ako *T-Net*. T-net časti modelu sa skladajú z konvolučných vrstiev, pooling vrstiev a plne prepojených vrstiev, čím pripomínajú klasické konvolučné neurónové siete (angl. *Convolutional Neural Network*, *CNN*). CNN sa vyznačujú vlastnosťou, že do aktivácie vypočítanej z konkrétneho bodu neprispieva iba bod sám, ale aj jeho okolie. Týmto spôsobom je možné extrahovať črty z blízkeho okolia bodu na skorších vrstvách a komplexnejšie globálne črty na neskorších vrstvách.

Na konvolučných vrstvách je výstup počítaný zo vstupu pomocou konvolúcie. Pri konvolúcii je aplikovaný posuvný filter alebo kernel (matica veľkosti $n \times m$) postupne na všetky súvislé podmnožiny veľkosti $n \times m$ zo vstupu. Počet filtrov aplikovaných na jednotlivých vrstvách predstavuje počet neurónov na vrstve a určujú počet dimenzií výstupu. Dôležitou vlastnosťou týchto vrstiev je, že pri procese konvolúcie je rovnaký filter posúvaný po celom vstupe, a teda sa snaží získať rovnakú informáciu zo všetkých častí vstupu. Vďaka tomu je možné identifikovať črty bez ohľadu na to, či sú reprezentované na začiatku alebo na konci vstupu. Veľkosť okolia, z ktorého majú byť črty získané pre konkrétny bod je možné regulovať veľkosťou kernelu. Napríklad pri veľkosti kernelu 3×3 je do výpočtu aktivácie pre konkrétny bod zahrnutá hodnota bodu samotného spolu s jeho ôsmimi susediacimi bodmi. V modeli PointNet sú použité konvolúcie 1×1 .

Využitie 1×1 konvolúcie v neurónovej sieti PointNet zabezpečuje extrakciu črt samostatne pre jednotlivé body. Po spracovaní vstupu konvolučnými vrstvami je každý bod reprezentovaný ako vektor hodnôt (črt), pričom tieto hodnoty boli pre všetky body vypočítané rovnakými kernelmi. Na takúto reprezentáciu vstupu je aplikovaná symetrická funkcia max pooling, vďaka ktorej sú do ďalších vrstiev

posunuté vždy rovnaké hodnoty bez ohľadu na zmenu poradia bodov v pôvodnom vstupe. Výstup z pooling vrstvy je ďalej spracovaný plne prepojenými vrstvami a následne použitý na získanie afinnej transformačnej matice veľkosti 3×3 , ktorá je v hlavnom modeli použitá priamo na zarovnanie vstupu.

Podobným spôsobom je získaná aj matica na zarovnanie lokálnych črt. Transformačná sieť pre lokálne črty je komplexnejšia v porovnaní s tou pre vstupné dáta a transformačná matica má v tomto prípade veľkosť 64×64 , čo výrazne zvýšilo výpočtovú náročnosť optimalizácie siete. Z toho dôvodu bola pridaná regularizácia pre výpočet stratovej funkcie (angl. *loss function*):

$$L_{reg} = \|I - AA^T\|_F^2, \quad (3.2)$$

kde A je transformačná matica predikovaná „mini“ sieťou *T-Net*. Nastavené obmedzovanie zaručilo, že výsledná transformačná matica pre lokálne črty bude blízka ortogonálnej matici. Ortogonálne matice sú z definície štvorcové matice, ktorých riadky a stĺpce sú ortogonálne (pravouhlé) vektory. Pre takéto matice platí vzťah:

$$Q^T Q = Q Q^T = I, \quad (3.3)$$

kde Q je pôvodná matica, Q^T je transponovaná matica a I je matica identity. Ortogonálna transformačná matica zachováva skalárny súčin vektorov, vďaka čomu umožňuje vykonanie unitárnych transformácií ako rotácia alebo zrkadlenie.

3.3 Klasifikácia a segmentácia pomocou PointNet

Model PointNet umožňuje klasifikáciu 3D objektov, segmentáciu častí 3D objektov a sémantickú segmentáciu scény. Z architektúry siete na obrázku 3.2 vidno, že architektúra modelu sa líši v závislosti od riešenej úlohy, kde segmentačná sieť predstavuje nadstavbu nad pôvodným čisto klasifikačným modelom.

Štruktúra vstupných dát pre všetky tri úlohy je rovnaká, avšak zatiaľ čo pre klasifikáciu a segmentáciu častí objektu reprezentujú vstupné mračná bodov jeden objekt, pre segmentáciu scény je zachytená v mračnách snímka časti scény. Výstupy siete sa líšia v závislosti od úlohy, ktorá má byť riešená. Výstupom klasifikácie objektu je jeden vektor pre celý vstup obsahujúci skóre pre každú z k možných tried.

Pre úlohy segmentácie s n množstvom bodov v mračne a m možnými sémantickými kategóriami je výstupom modelu matica veľkosti $n \times m$ obsahujúca skóre pre každú z m možných kategórií pre všetky body.

Pri klasifikácii objektov sú vstupné body najskôr transformované afínnou transformačnou maticou získanou z prvej *T-Net* siete. Zo zarovnaných transformovaných dát sú následne získané lokálne črty. Na obrázku architektúry 3.2 je táto časť siete označaná ako dva viacvrstvové perceptróny s veľkosťou 64 a zdieľanými váhami medzi neurónmi v rámci vrstiev. Formálne sú tieto vrstvy implementované ako 1×1 konvolučné vrstvy, keďže práve tie umožňujú spracovanie každého bodu samostatne rovnakými váhami. Takýmto spôsobom získané lokálne črty sú zarovnané transformačnou maticou z druhej *T-Net* „mini“ siete. Transformované črty sú opäť spracované pomocou vrstiev označených ako MLP so zdieľanými váhami, čím sú extrahované globálne črty, ktoré sú ďalej agregované pomocou max pooling funkcie. Posledná časť klasifikačnej siete je tvorená tromi MLP, ktoré z globálnych črt vstupu určia výstup klasifikácie.

Neurónová sieť na segmentáciu je rozšírením pôvodnej klasifikačnej siete. Na začiatku tejto nadstavby sú zlúčené lokálne a globálne črty získané v rámci klasifikácie. Následne sú použité vrstvy so zdieľanými váhami na extrakciu komplexných črt pre body. V závere je vykonaná segmentácia opäť pomocou vrstiev so zdieľanými váhami, po ktorej je možné jednotlivým bodom priradiť sémantickú kategóriu, do ktorej patria.

4 Registrácia objektov v mračnách bodov

Pri registrácii objektov je známe, aké objekty by sa mali vo vstupných dátach nachádzať a úlohou neurónovej siete je ich nájsť. Vstupom pre sieť sú hľadaný vzor a mračno bodov, v ktorom má byť daný vzor nájdený. Pri úspešnej registrácii je možné pomocou afínných transformácií zarovnať obe vstupné mračná a identifikovať tak hľadaný vzor v dátach. V prípade, že požadovaný objekt nebude nájdený a zarovnanie nebude uspokojujúce, znamená to, že hľadaný vzor sa v mračne nenachádza, prípadne že sa nachádza v dátach výrazne zdeformovaný alebo nekompletný.

Opäť aj pre registráciu objektov existujú riešenia, ktoré transformujú riešený problém do 2D priestoru. Príkladom je model z práce [21], ktorý vykonáva registráciu objektov na základe porovnávania lokálnych 2D črt. Medzi ďalšie práce zamerané na registráciu objektov v mračnách bodov patria [22, 23, 24, 25, 26] a stručný prehľad o existujúcich metódach spracovali Zhang a kolektív [27].

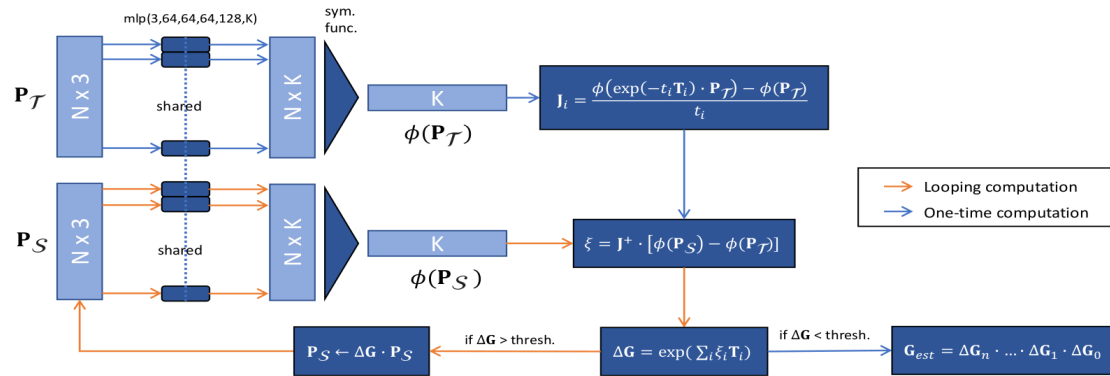
Tejto kapitole sú opísané viaceré existujúce riešenia registrácie objektov pomocou hlbokého učenia, pričom špeciálny dôraz je kladený na model CorsNet [28], ktorý bude dôležitý pre výskumnú časť predloženej práce.

4.1 PointNetLK

Problematika registrácie objektov v mračnách bodov je výrazne menej riešená v porovnaní s klasifikáciou alebo detekciou objektov v mračnách. V jednej z prác, ktoré sa venujú tejto problematike je odprezentovaný model PointNetLK [29]. Architektúra tohto modelu je znázornená na obrázku 4.1. PointNetLK využíva na získanie reprezentácie črt prístup založený na poznatkoch z PointNet a na registráciu je použitý modifikovaný Lucas-Kanade (LK) algoritmus implementovaný ako rekurentná neurónová sieť.

Lucas-Kanade algoritmus slúži na určenie vektoru posunutia (u, v) porovnaním dvoch obrázkov zachytávajúcich scénu alebo objekt v rôznom čase, pričom sa predpokladá, že časový rozdiel Δt medzi obrázkami je malý. Tento vektor posunutia je vypočítaný pre vybrané významné body obrázku. Prehľadné zhrnutie algoritmu je

opísané v článku [30]. LK algoritmus využíva okolie významných bodov na určenie ich vektoru posunutia, čo však má za následok komplikácie v prípade, že gradient v okolí je blízky nule, prípadne že bod sa nachádza na hrane a gradient je výrazne väčší v jednom smere. Pri sieti PointNetLK by však táto vlastnosť nemala spôsobovať problémy, keďže namiesto obrázkov sú použité 2D vektorové reprezentácie mračien bodov, v ktorých je výskyt takéhoto správania málo pravdepodobný.



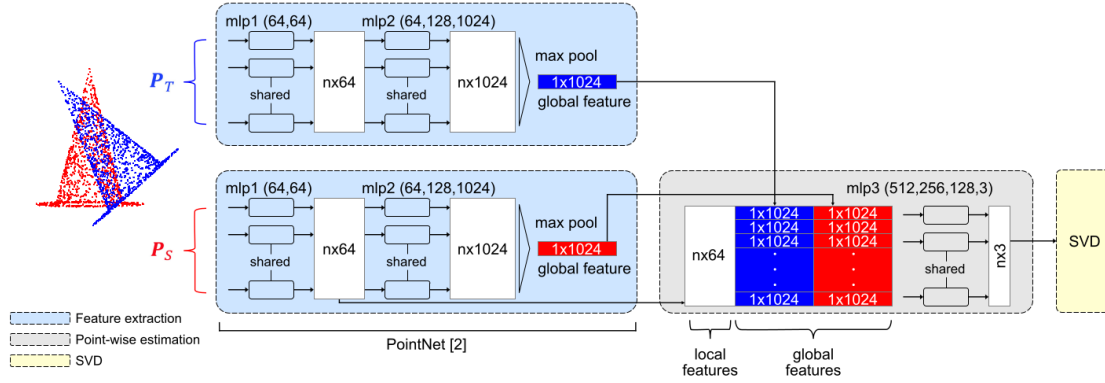
Obr. 4.1: *PointNetLK architektúra. Z mračien bodov pre vzor (template) \mathbf{P}_T a vstupný objekt (source) \mathbf{P}_S sú extrahované vektorové reprezentácie globálnych črt $\phi(\mathbf{P}_T)$ a $\phi(\mathbf{P}_S)$. Jacobiho matica je pre vzor vypočítaná iba raz a ďalej je používaná pri výpočte parametrov otočenia. Pozícia pôvodného vstupného objektu \mathbf{P}_S je následne iteratívne upravovaná, pričom prebieha aj prepočet jeho vektorovej reprezentácie [29].*

4.2 CorsNet

Lepšie výsledky ako PointNetLK dosahuje end-to-end hlboká neurónová sieť CorsNet [28]. Na rozdiel od PointNetLK, ktorý pracuje iba s globálnou reprezentáciou črt, tento model používa zlúčenú reprezentáciu lokálnych a globálnych črt. Vstupom pre CorsNet sú dve mračná bodov reprezentujúce vstupný objekt a šablónu, pričom úlohou modelu je predikovať transformačnú maticu umožňujúcu zarovnanie vstupného objektu na šablónu. Architektúru je možné vidieť na obrázku 4.2. Model tejto siete je pozostáva z troch častí:

- získanie globálnych črt z oboch vstupných mračien bodov,
- zlúčenie reprezentácií globálnych črt s lokálnymi črtami vstupného objektu a odhad súhlasnosti extrahovaných črt,

- určenie transformačnej matice pomocou singulárneho rozkladu (angl. *Singular Value Decomposition, SVD*).



Obr. 4.2: CorsNet architektúra. Mračná bodov pre vzor (template) a vstupný objekt (source) sú označené ako P_T a P_S . Architektúra je rozdelená na 3 segmenty: získanie globálnych črt (Feature extraction), odhad súhlasnosti (Point-wise estimation) a metódy singulárneho rozkladu SVD. [28]

Extrakcia globálnych črt. Mračná bodov pre vstupný objekt (P_S) a šablónu (P_T) majú rovnakú veľkosť a sú reprezentované maticami o veľkosti $n \times 3$, kde n predstavuje počet bodov opisujúcich objekt, pričom každý bod je daný trojicou $[x, y, z]$ súradníc. Na získanie lokálnych a globálnych črt z takto reprezentovaných mračen bodov sú použité dve PointNet siete. Kompletná PointNet architektúra na klasifikáciu objektov je opísaná v kapitole 3.2, avšak pri registrácii objektov je potrebná iba časť na extrakciu črt končiaca *max pooling* vrstvou.

Odhad súhlasnosti extrahovaných črt. Prvá vrstva v druhej časti siete slúži na zlúčenie reprezentácií oboch vstupných objektov. Lokálne črty pre každý bod zo vstupného objektu sú zlúčené s globálnou reprezentáciou šablóny aj vstupného objektu. Následne je použitá sekvencia MLP na odhadnutie posunutia medzi šablónou a zdrojovým objektom. Výstupná matica z posledného MLP má opäť veľkosť $n \times 3$ a ak ju označíme ako ΔP_S , mal by platiť vzťah

$$\hat{P}_T = P_S + \Delta P_S, \quad (4.1)$$

kde \hat{P}_T je predikovaná transformácia zdrojového objektu na šablónu. Výstup druhej

časti siete $\Delta \mathbf{P}_S$ však nie je žiadnym spôsobom obmedzený a transformovaný objekt $\hat{\mathbf{P}}_T$ môže byť zdeformovaný a úlohou modelu nemá byť nájdenie ľubovoľného posunutia pre každý bod zvlášť tak, aby bol v konečnom dôsledku každý bod vstupného objektu namapovaný na šablónu, ale aby celý objekt ako taký bol premietnutý na šablónu, pričom nemusí platiť, že každý bod z \mathbf{P}_S sa premietne na špecifický bod v \mathbf{P}_T . Z toho dôvodu je predikovaná transformácia $\hat{\mathbf{P}}_T$ použitá na určenie rigidnej transformačnej matice $\mathbf{G} \in SE(3)$, pre ktorú by malo platiť

$$\mathbf{G} \cdot \mathbf{P}_S = \hat{\mathbf{P}}_T.$$

SVD. Výpočet predikovanej transformačnej matice \mathbf{G}_{est} je v poslednej časti siete vykonaný pomocou SVD. Ako prvé sú určené centroidy pre \mathbf{P}_S a $\hat{\mathbf{P}}_T$ ako

$$\overline{\mathbf{P}}_S = \frac{1}{n} \sum_{i=1}^n \mathbf{P}_S \quad \text{a} \quad \overline{\hat{\mathbf{P}}_T} = \frac{1}{n} \sum_{i=1}^n \hat{\mathbf{P}}_T.$$

Kovariančná matica \mathbf{H} daná vzťahom

$$\mathbf{H} = \sum_{i=1}^n (\hat{\mathbf{P}}_T - \overline{\hat{\mathbf{P}}_T})(\mathbf{P}_S - \overline{\mathbf{P}}_S)^T$$

je následne rozložená SVD algoritmom na získanie $\mathbf{U}, \mathbf{V} \in SO(3)$:

$$[\mathbf{U}, \mathbf{S}, \mathbf{V}] = SVD(\mathbf{H}).$$

Predikovaná rotácia $\mathbf{R}_{est} \in SO(3)$ a translácia $\mathbf{t}_{est} \in \mathbb{R}^3$ sú definované vzťahmi

$$\mathbf{R}_{est} = \mathbf{V}\mathbf{U}^T \quad \text{a} \quad \mathbf{t}_{est} = -\mathbf{R}_{est} \cdot \overline{\hat{\mathbf{P}}_T} + \overline{\mathbf{P}}_S,$$

pričom kompletná predikovaná rigidná transformácia $\mathbf{G}_{est} \in SE(3)$ je daná ako

$$\mathbf{G}_{est} = \begin{pmatrix} \mathbf{R}_{est} & \mathbf{t}_{est} \\ 0 & 1 \end{pmatrix}.$$

Hoci na výpočet transformačnej matice pomocou SVD sú použité čisto lineárne operácie bez prítomnosti trénovateľných parametrov, pre ucelenosť riešenia je aj tento výpočet implementovaný ako súčasť end-to-end neurónovej siete CorsNet.

5 Návrh riešenia

Zadaním práce je vyriešiť problém registrácie 3D objektov pomocou hlbkej neurónovej siete. Jej vstupom by mali byť dve mračná bodov reprezentujúce vstupný objekt a šablónu, pričom na základe výstupu by malo byť možné zarovnať tieto objekty. Mračná bodov na vstupe budú opisovať písmená, číslice a ďalšie znaky z datasetu dostupného pre riešenia projektu. Požadovaným formátom výstupu je určitá transformácia, ktorá premietne vstupný objekt na šablónu. V prechádzajúcej kapitole boli opísané existujúce modely na registráciu 3D objektov a ich postup je možné zovšeobecniť do troch krokov:

1. získanie reprezentácií oboch vstupných objektov pomocou modelu PointNet,
2. spracovanie zlúčených reprezentácií objektov plne prepojenými vrstvami,
3. transformácia výstupu poslednej vrstvy do požadovaného formátu.

Opísaný postup viedol vo viacerých prípadoch k úspešným výsledkom a z toho dôvodu bol použitý ako základ aj pri návrhu nášho riešenia. Navrhnutá architektúra vychádza z modelu CorsNet, ktorý podľa článku [28] dosahoval výrazne lepšie výsledky v porovnaní s modelom PointNetLK [29] pri riešení registrácie objektov nad datasetom ModelNet40. Kľúčovou časťou práce na tomto projekte bude otestovanie správania modelu CorsNet na našich dátach a následná modifikácia architektúry v závislosti od priebežných zistení a výsledkov dosiahnutých pre rôzne úrovne komplexnosti transformácií. Nakoľko však neexistuje verejne dostupná implementácia, v prvom kroku bude potrebné reimplementovať a natrénovať tento model. S takto vytýčeným jadrom riešenia je možné celkový plán práce na výskumnej časti rozdeliť na nasledovné úlohy:

1. Reimplementácia modelu PointNet.
2. Reimplementácia modelu CorsNet.
3. Predspracovanie datasetu ModelNet40 pre CorsNet.
4. Natrénovanie a evaluácia modelov na pôvodných dátach.
5. Predspracovanie našich dát na formát podobný pôvodným dátam.

6. Vývoj modelu CorsNet na našom datasete.
7. Iteratívne zvyšovanie komplexnosti datasetu a vylepšenie modelu.
8. Vyhodnotenie chyby medzi šablónou a naskenovanými dátami.

CorsNet reimplementácia a trénovanie. Prvé štyri body zahŕňajú reimplementáciu modelov na základe dostupných zdrojov a ich natrénovanie na pôvodnom datasete ModelNet40. Architektúry oboch modelov sú opísané v kapitolách 3.2 a 4.2. Pri rovnakom predspracovaní dát a nastavení hyperparametrov siete ako bolo opísané v článkoch [28, 29] by malo byť možné dosiahnuť podobné výsledky a overiť tak fungovanie modelu a správnosť našej implementácie. Trénovaním modelov na pôvodných dátach sme chceli overiť, do akej miery je model schopný vykonávať registráciu a nad akými vstupnými dátami, aby sme mohli v ďalších krokoch predspracovať aj náš dataset podobným spôsobom.

Predspracovanie dát. Po úspešnej registrácii ModelNet40 objektov bude nasledovať predspracovanie našich dát a pre ich ďalšie využitie na trénovanie vlastného CorsNet modelu. V tejto fáze budú dáta predspracované takým spôsobom, aby reprezentácie objektov aj výstupné transformácie (translácia a rotácia) čo najvernejšie kopírovali charakter pôvodných tréovacích dát. Vstupný objekt a šablóna budú reprezentované množinou 3D bodov a výstupom bude transformačná matica umožňujúca premietnutie vstupného objektu na šablónu.

Vývoj modelu CorsNet na našom datasete. Keďže nami vygenerovaný dataset písmen a znakov by sa nemal výrazne líšiť od datasetu ModelNet40, implementovaný model CorsNet by mal byť schopný úspešne vykonať registráciu objektov aj nad týmito dátami, tento predpoklad však bude potrebné otestovať. V závislosti od výsledkov tréovania môžu byť zvážené malé zmeny v architektúre a optimalizácia hyperparametrov.

Iteratívne zvyšovanie komplexnosti datasetu a vylepšenie modelu. Ak bude natrénovaný model schopný vykonať dostatočne presnú registráciu jednotlivých písmen a znakov, v ďalšom kroku budú na tieto objekty aplikované komplexnejšie transformácie a mierne deformácie. Nakoľko pôvodnou myšlienkou za vznikom projektu bola možnosť registrovať vzory na zakrivenom povrchu pneumatiky, podobnú situáciu by sme chceli simulovať aj vo vstupných dátach. Okrem rozšírenia transformácií o deformácie bude zvážené zvýšenie komplexnosti samotných objektov. V takom

prípade by boli vstupné objekty generované ako dvojice písmen, neskôr trojice, prípadne aj celé slová. Komplexnosť transformácií aj objektov bude zvyšovaná postupne a v závislosti od priebežne dosahovaných výsledkov bude zvažované aj pridávanie ďalších vrstiev.

Ďalším aspektom pri predspracovaní dát môže byť odstránenie častí objektu, napríklad odrezanie nožičky písmena A alebo vynechanie časti písmena O. Mračná bodov by stále mali rovnakú veľkosť zodpovedajúcu veľkosti vstupu do siete, avšak týmto spôsobom by boli simulované chybné písmená. V ideálnom prípade by mal byť model stále schopný vykonať registráciu a správne zarovnať písmeno A s chýbajúcou nožičkou na kompletne A.

Vyhodnotenie chyby medzi šablónou a naskenovanými dátami. V prípade, že náš model bude dosahovať požadované výsledky, v rámci post-processingu by bolo možné implementovať funkciu na porovnanie šablóny a vstupného objektu a určenie ich percentuálnej zhody. Porovnanie nebude možné vykonať priamo nad šablónou a transformovaným vstupným objektom, nakoľko oba objekty sú opísané bodmi v 3D priestore, ktoré neumožňujú priamy výpočet objemu alebo percentuálnej mieru prekryvu. Potencionálnym riešením by mohlo byť vytvorenie polygónovej siete (angl. *mesh*) z bodov, následné vypočítanie obsahu prieniku týchto dvoch 3D objektov a určenie pomeru tomto obsahu voči celkovému obsahu šablóny. V ideálnom prípade, kde mračno bodov po predikovanej transformácii opisuje kompletný objekt presne zarovnaný na šablónu, tento pomer by mal byť blízky 100%. Naopak ak by bol objekt s chýbajúcou časťou správne zarovnaný, z výsledného pomeru by sme vedeli aká veľká časť objektu chýba. Táto funkcionalita však predpokladá úspešné fungovanie modelu na registráciu objektov.

6 Implementácia modelu a evaluácia

Základom implementovaného riešenia je neurónová sieť na registráciu 3D objektov Corsnet. Jadrom práce na projekte v zimnom semestri bola práve reimplementácia modelu a jeho natrénovanie na datasete ModelNet40. PointNet časť modelu na získanie reprezentácie vstupných objektov a rovnako aj kompletný CorsNet model bol úspešne implementovaný, pričom výsledky dosiahnuté v oboch prípadoch sa blížili k výsledkom odprezentovaných v článkoch [1, 28]. V nasledujúcich častiach sú uvedené detaily reimplementácie a evaluácia modelov.

6.1 PointNet implementácia

Hlboká neurónová sieť PointNet je podrobne opísaná v kapitole 3. Základná architektúra siete umožňuje extrahovanie lokálnych a globálnych črt zo vstupu reprezentovaného neusporiadanými množinami bodov a následne využíva vektorové reprezentácie črt na vykonanie klasifikácie. Na registráciu objektov sú potrebné iba časti na extrakciu črt, avšak nakoľko chceme mať možnosť použiť *transfer learning* pri tréovaní modelu CorsNet, bude potrebné implementovať a natrénovať celý klasifikačný model.

Architektúra reimplementovaného modelu PointNet kopíruje schému z obr. 3.2. Vstupný objekt je reprezentovaný $n \times 3$ maticou, kde n predstavuje počet bodov, pričom každý bod je daný trojicou $[x, y, z]$ súradníc. Neurónová sieť obsahuje dve transformačné *T-Net* siete na zarovnanie lokálnych a neskôr globálnych črt. Medzi nimi sa nachádzajú sekvencie rôzne veľkých MLP a v závere sú reprezentácie črt agregované pomocou funkcie *max pooling* a spracované tromi MLP, ktorých výstupom je klasifikácia objektu. Medzi tieto posledné MLP boli pridané *dropout* vrstvy za účelom regularizácie tréovania.

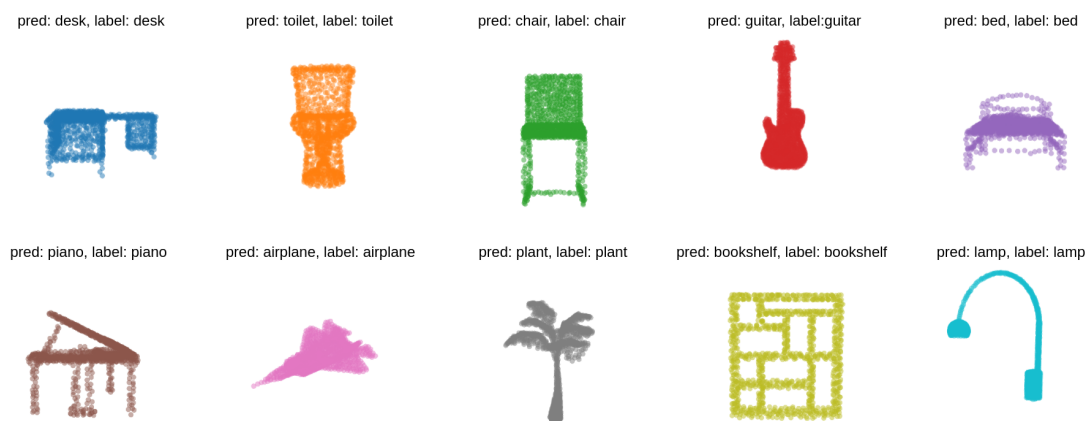
Reimplementácia modelu PointNet bola vykonaná pomocou Python knižnice Keras a vychádzala z verejne dostupného zdrojového kódu pre pôvodný model, pri ktorom bola použitá priamo knižnica TensorFlow.

6.2 PointNet tréovanie a evaluácia

Implementovaný model PointNet bol natrénovaný na klasifikáciu objektov na datasete ModelNet40. Dataset obsahuje 12 311 objektov patriacich do 40 tried a rozdelených na tréovaciu sadu (9 843 objektov) a testovaciu (2 468 objektov). Vstupné mračná bodov s veľkosťou 1024 sú rovnomerne navzorkované z povrchu objektov a normalizované do jednotkovej kocky. V rámci augmentácie počas tréovania bol pridaný šum a objekty boli rotované okolo zvislej osi.

Hyperparametre boli nastavené rovnakým spôsobom ako vo verejne dostupnej PointNet implementácii s batch veľkosťou 32, optimalizátorom Adam s hodnotou *learning rate* nastavenou na 0.001 a počtom epoch limitovaným na 250.

Celková presnosť (angl. *accuracy*) najlepšieho natrénovaného modelu bola 86.89%, čo sa blížilo k hodnote 89.22%, ktorú dosiahli Qi a kolektív [1]. Na koľko našim cieľom nie je klasifikácia objektov a model PointNet má byť využitý iba na extrakciu črt, dosiahnuté výsledky potvrdili správnosť reimplementácie a boli dostatočné pre ďalšiu prácu na projekte. Príklady klasifikácie natrénovanej siete je možné vidieť na obr. 6.1.



Obr. 6.1: *PointNet* klasifikácia.

6.3 Predspracovanie dát pre CorsNet

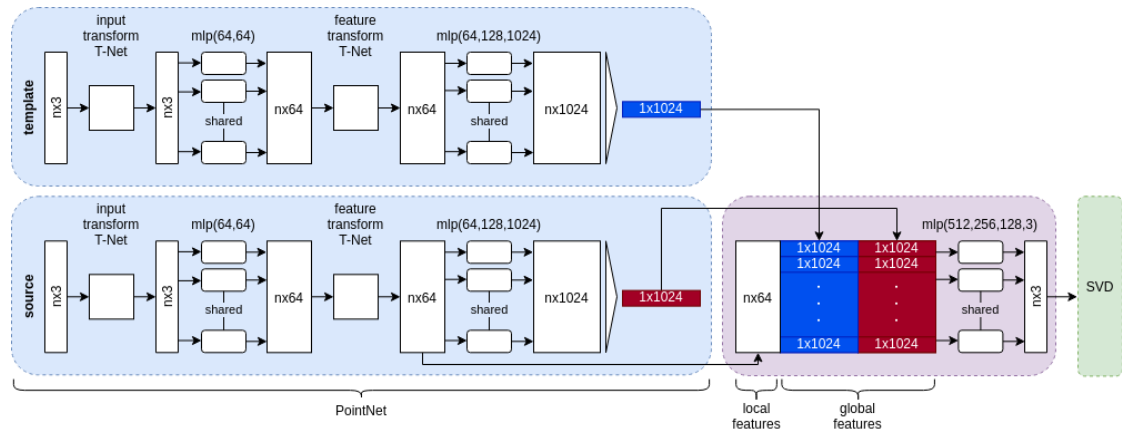
Kurobe a kolektív [28] uviedli, že dataset pre CorsNet bol vytvorený z ModelNet40, avšak spôsob predspracovania nebol veľmi detailne opísaný. Na trénovanie modelu bolo potrebné vytvoriť dvojice vstupných objektov (šablónu a vstupný objekt) a k nim prislúchajúcu transformačnú maticu \mathbf{G}_{gt} , ktorá umožňuje premietnutie vstupného objektu na šablónu.

Mračná bodov pre vstupné objekty boli vygenerované rovnakým spôsobom ako pre PointNet. Z povrchu objektov bolo rovnomerne navzorkovaných 1024 bodov, ktoré boli následne normalizované do jednotkovej kocky. Transformačné matice boli náhodne vygenerované s rotáciami okolo ľubovoľných osí z intervalu $< 0, 45 >$ stupňov a transláciami z intervalu $< 0, 0.8 >$ pre všetky tri osi. Šablóny pre príslušné vstupné objekty boli vytvorené ako skalárny súčin šablóny a transformačnej matice \mathbf{G}_{gt} . Takto pripravené trojice boli následne použité na natrénovanie modelu na registráciu objektov.

6.4 CorsNet implementácia

Na rozdiel od modelu PointNet, pre CorsNet nebola verejne dostupná implementácia a preto sme primárne vychádzali z náčrtu architektúry, ktorý je spolu s ďalšími detailami o fungovaní modelu opísaný v kapitole 4.2. Model sa skladá z dvoch PointNet sietí, jednej vrstvy na zjednotenie reprezentácie oboch vstupných objektov, niekoľko MLP na predikciu transformácie ako rozdielu medzi objektami a SVD vrstvy, ktorej výstupom je predikovaná transformačná matica. Podobne ako pôvodný CorsNet, aj naša reimplementácia je end-to-end neurónová sieť s SVD časťou implementovanou ako samostatnou vrstvou bez trénovateľných parametrov.

Naša neurónová sieť až na niekoľko detailov kopíruje architektúru zobrazenú na obr. 4.2. Zo schémy CorsNet modelu sa javí, že model PointNet nebol použitý kompletný a transformačné časti *T-Net* boli vynechané. V našom modeli je na extrakciu lokálnych a globálnych črt použitá nezmenená PointNet architektúra ako je opísaná v časti 3.2. Schéma implementovanej neurónovej siete je znázornená na obr. 6.2.



Obr. 6.2: CorsNet architektúra (reimplementácia).

6.5 CorsNet tréovanie

Na tréovanie modelu boli použité ako vstupy predspracované dvojice mračien bodov reprezentované maticami o veľkosti 1024×3 a úlohou modelu bolo predikovať rigidnú transformačnú maticu na premietnutie mračna bodov pre vstupný objekt na šablónu.

Na určenie chyby predikovanej transformačnej matice \mathbf{G}_{est} voči *ground truth* matice \mathbf{G}_{gt} bola použitá vlastná stratová funkcia (angl. *loss function*). Dáta boli predspracované tak, aby pre vstupný objekt \mathbf{P}_S a šablónu \mathbf{P}_T platilo

$$\mathbf{P}_T = \mathbf{G}_{gt} \cdot \mathbf{P}_S.$$

Stratová funkcia bola z toho dôvodu počítaná pomocou vzťahu

$$loss = ||(\mathbf{G}_{est})^{-1} \cdot \mathbf{G}_{gt} - \mathbf{I}_4||_F.$$

Túto funkciu je potrebné vyjadriť ako MSE alebo RMSE funkciu aby sa zabránilo explodovaniu funkcie do záporných čísel. Hoci tieto detaily implementácie neboli priamo uvedené, Kurobe a kolektív [28] prebrali myšlienku pre funkciu z modelu PointNetLK [29], pre ktorý je verejne dostupná implementácia, kde stratová funkcia vyzerá nasledovne:

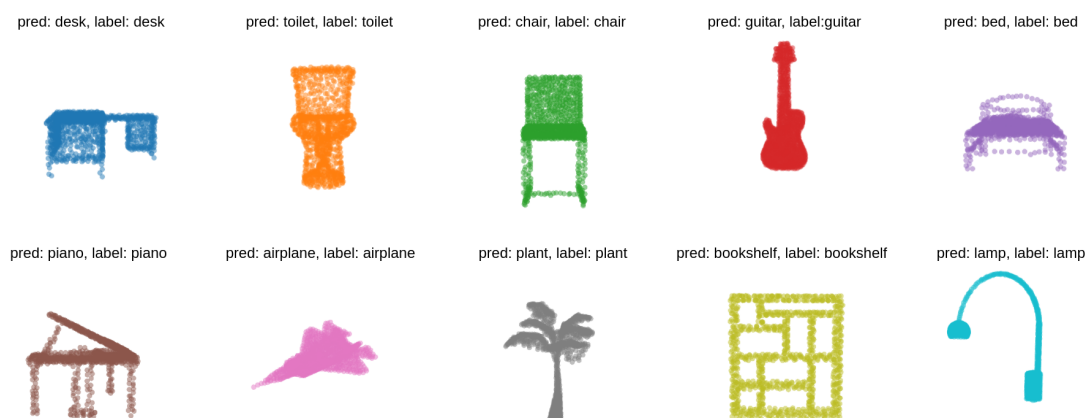
$$loss = 16 \cdot MSE((\mathbf{G}_{est})^{-1} \cdot \mathbf{G}_{gt} - \mathbf{I}_4).$$

Vynásobenie výsledku MSE konštantou pri našich experimentoch viedlo v rýchlejšej konvergencii funkcie. Úspešnosť predikcií bola meraná predovšetkým implementovanou stratovou funkciou. Okrem jej hodnoty boli sledované aj metriky presnosť (angl. *accuracy*) a MSE. Nastavenia hyperparametrov uvedené v článku [28] zahŕňalo použitie optimalizátora Adam s počiatočnou hodnotou *learning rate* 0.0001. Počet epoch bol nastavený na 300, pričom hodnota *learning rate* bola pri 75., 150. a 200. epoche vydelená 10.

6.6 CorsNet experimenty

Pre originálny model CorsNet neboli dostupné žiadne informácie o augmentácii dát počas tréovania alebo prípadnom použití *transfer learning* s predtrénovanými PointNet váhami a práve tieto charakteristiky boli prvým predmetom výskumu. Testované boli rôzne spôsoby augmentácie dát počas tréovania a drobné zmeny v predspracovaní objektov. Sledovaný bol taktiež vplyv použitia predtrénovaných váh na extrakciu črt vstupných objektov a obmedzenie tréovania váh na určitých vrstvách. Pri nastavení hyperparametrov sme sa snažili držať detailov uvedených v časti 6.5, avšak pri viacerých experimentoch bolo potrebné tieto nastavenia pozmeniť predovšetkým z dôvodu hardvérových obmedzení. Modely hlbokého učenia boli tréované na zariadení s 16 GB RAM a NVIDIA GeForce GTX 1650 Max-Q GPU 4GB.

V tejto fáze projektu nebolo cieľom dosiahnuť čo najlepšie výsledky siete. Dôraz bol kladený na overenie fungovania modelu a identifikáciu najvhodnejšej kombinácie predspracovania a augmentácie dát. Nájdená optimálna kombinácia bude slúžiť ako základ pri predspracovaní vlastného datasetu. Čas tréovania bol skrátený z pôvodných 300 epoch na 200. Napriek tomuto obmedzeniu dosahovali viaceré natréované modely úspešné výsledky. Na evaluáciu boli použité matriky presnosť, MSE na celkovú predikovanú transformačnú maticu a podobne ako pre originálny CorsNet aj RMSE samostatne pre rotáciu \mathbf{R} a transláciu \mathbf{t} .



Obr. 6.3: Výsledky registrácie objektov (**modrá**: vstupný objekt, **oranžová**: šablóna, **zelená**: transformácia).

6.6.1 Experimenty - predspracovanie dát

Spôsob generovania transformačných matíc bolo ponechaný z pôvodného predspracovania s náhodnými rotáciami z intervalu $< 0, 45 >$ stupňov a transláciami z intervalu $< 0, 0.8 >$ pre všetky tri osi. Modifikovaný bol spôsob získavania mračien bodov z 3D objektov. Pôvodne boli vstupný objekt aj šablóna tvorené rovnakým mračnom bodov a líšili sa iba transformáciou v priestore. Model PointNet by však mal byť schopný vytvoriť podobné reprezentácie globálnych črt pre rôzne mračná bodov opisujúce rovnaký objekt a preto by malo byť možné vykonať aj registráciu nad vstupným objektom a šablónou tvorených rôznymi bodmi z povrchu jedného objektu.

Predspracované dáta, ktoré použili Qi a kolektív [1] na tréning modelu PointNet obsahujú objekty opísané 2048 bodmi rovnomerne navzorkovanými z povrchu objektov. Pôvodne boli pre tréning vstupné mračná bodov tvorené iba polovicou týchto bodov. My sme navrhli alternatívny spôsob predspracovania, pri ktorom je vstupný objekt tvorený prvými 1024 bodmi a šablóna zvyškom.

Dve neurónové siete boli natrénované so všetkými nastaveniami hyperparametrov a augmentácie identickými s jediným rozdielom v predspracovaní. Evaluácia bola vykonaná nad datasetom tvoreným rovnakými mračnami bodov pre oba vstupné objekty a z tabuľky 6.1 je vidno, že pri použití rôznych mračien bodov pre šablónu

	Presnosť	MSE	RMSE (R)	RMSE (t)
Rovnaké mračná bodov	95.60%	0.00574	0.09015	0.06725
Rôzne mračná bodov	96.10%	0.00501	0.08627	0.05537

Tabuľka 6.1: *Evaluácia rôznych metód predspracovania.*

a vstupný objekt boli celkovo dosiahnuté lepšie výsledky.

6.6.2 Experimenty - augmentácia dát

Inšpirovaní augmentáciou aplikovanou pre PointNet, aj v tomto prípade bol počas tréningu k obom vstupným mračnám bodov pridaný šum z normálnej distribúcie s hodnotou $\sigma = 0.01$ a hornou hranicou 0.05. Táto perturbácia dát sa už pri prvých natrénovaných modeloch ukázala ako užitočná a bola použitá pri všetkých modeloch spomenutých vo výsledkoch.

PointNet architektúra by však mala zabezpečiť minimálne zmeny v reprezentácii objektov bez ohľadu na usporiadanie bodov vo vstupe. Na potvrdenie tohto predpokladu bolo v rámci augmentácie dát pri tréningu otestované zamiešanie bodov reprezentujúcich vstupné objekty nezávisle od seba. Týmto spôsobom sa aj pri použití rovnakých mračien bodov pre šablónu aj vstupný objekt malo zabrániť situáciám, kedy by na registráciu objektov nebola potrebná neurónová sieť, ale postačoval by iba lineárny výpočet aplikovaný na poslednej vrstve siete.

	Presnosť	MSE	RMSE (R)	RMSE (t)
Bez premiešania bodov	95.51%	0.00507	0.08355	0.06900
Premiešanie bodov	95.60%	0.00574	0.09015	0.06725

Tabuľka 6.2: *Evaluácia rôznych metód augmentácie dát.*

Z výsledkov na tabuľke 6.2 nie je možné jednoznačne tvrdiť, či premiešanie bodov prispelo k zmenšeniu odchýlky predikovaných transformácií. Porovnávané metriky však ukázali, že táto forma augmentácie dát výrazne nezhoršilo výsledky siete a prispelo k univerzálnosti funkcie siete pre dáta z reálneho sveta.

6.6.3 Experimenty - transfer learning

S predtrénovaným modelom PointNet na klasifikáciu bolo možné otestovať využitie *transfer learning* a jeho vplyv na predikcie siete. Natrénované boli tri modely s rôznymi kombináciami nastavení (tab. 6.3). Ak bola použitá metóda *transfer learning*, váhy v PointNet častiach modelu CorsNet boli inicializované na hodnoty z najlepšieho natrénovaného modelu dosahujúceho presnosť klasifikácie 86.89%. Otestovaná bola aj alternatíva, kedy boli PointNet váhy fixné a trénovateľné boli iba váhy v časti siete na určenie posunutia medzi vstupným objektom a šablónou.

Výsledky z tabuľky 6.4 ukazujú, že obmedzenie trénovateľných parametrov viedlo k najhorším výsledkom. Prekvapivým správaním bolo, že aplikácia metódy *transfer learning* neprispela k lepším výsledkom siete v porovnaní s náhodnou inicializáciou váh.

	Transfer learning	Trénovateľný PointNet
CorsNet-v1	✓	✓
CorsNet-v2	✓	-
CorsNet-v3	-	✓

Tabuľka 6.3: Nastavenie modelov na testovanie prínosu transfer learning.

	Presnosť	MSE	RMSE (R)	RMSE (t)
CorsNet-v1	95.51%	0.00507	0.08355	0.06900
CorsNet-v2	85.75%	0.03894	0.20493	0.28242
CorsNet-v3	96.56%	0.00336	0.07041	0.03570

Tabuľka 6.4: Evaluácia prínosu transfer learning.

6.7 Vstupné dáta a predspracovanie

Vstupné dáta pre túto prácu obsahujú písmená, číslce a ďalšie znaky získané z 58 obojstranných skenov pneumatík. Celkovo je k dispozícii 116 vyrovnaných skenov, z ktorých boli detekciou objektov získané prítomné znaky. Príklady detegovaných písmen a číslíc sú znázornené na 6.4. Zároveň sú k dispozícii ďalšie súbory obsahujúce 2D hĺbkovú mapu pôvodných skenov vyrovnaných do obdĺžnikov a súradnice

určujúce umiestnenie každého zo znakov na vyrovnanom povrchu pneumatiky. Tieto informácie budú dôležité pre prípadné skladanie slov z písmen v neskorších fázach projektu, prvotné zameranie však bude na registráciu jednotlivých písmen a čísiel.



Obr. 6.4: *Vstupné dáta - ASCII znaky.*

Na separáciu jednotlivých znakov bola použitá detekcia na základe hrán, čo má za dôsledok mierne nedokonalosti vo vstupných dátach. Príkladom sú chýbajúce bodky nad písmenami *i* a *j* alebo rozdelenie grafiky zo strán pneumatík na jednotlivé geometrické útvary. Dostupné dáta obsahujú okrem písmen a čísiel aj ďalšie ASCII znaky ako matematické symboly alebo zátvorky. Prítomné sú však aj kompletne menšie obrázky alebo ich časti a ďalšie non-ASCII znaky. Príklady týchto objektov sú uvedené na obr. 6.5.



Obr. 6.5: *Vstupné dáta - non-ASCII znaky.*

Dostupné dáta už boli vopred spracované a pre účely projektu nebude potrebné zložité predspracovanie. Celkový postup pozostáva zo štyroch krokov:

1. odstránenie objektov opisujúcich non-ASCII znaky,
2. normalizácia objektov,
3. navzorkovanie dvojíc šablóna-vstup z každého objektu,
4. vygenerovanie transformačnej matice a transformácia šablóny.

Na odstránenie non-ASCII objektov bude potrebné manuálne pretriediť celý dataset a odstrániť problematické objekty, ktorých príklady sú uvedené na obr.

6.5. Prvotným plánom je odstrániť všetky non-ASCII objekty, je však možné, že bude zvážené aj ponechanie kompletných obrázkov. Samostatné bodky, geometrické tvary a bližšie nešpecifikovateľné objekty nebudú použité pri trénovaní modelu.

Po odstránení nevhodných objektov bude aplikovaná rovnaká forma normalizácie ako pri doterajšom výskume a vstupné objekty budú zarovnané do jednotkovej kocky. Následne budú z každého objektu náhodne navzorkované dvojice šablóna-vstup, ktoré budú reprezentovať vstupné mračná bodov pre neurónovú sieť. V časti 6.6 bolo ukázané, že model model PointNet skutočne prispieva k úspešnej registrácii vstupných objektov, ktoré síce so šablónou opisujú identický útvar, avšak body použité na jeho reprezentáciu nemusia byť rovnaké. Táto vlastnosť modelu bude využívaná naďalej v snahe simulovať situácie z reálneho prostredia, kedy nie je možné očakávať, že mračná bodov získané zo skenov toho istého objektu budú sú rovnaké. Z toho dôvodu budú pre vstupné dvojice objektov navzorkované dve rôzne mračná bodov z rovnakého znaku alebo písmena. Doteraz vyvíjané modely spracovávali komplexnejšie objekty a na vstupe očakávali mračná bodov o veľkosti 1024. V našom prípade však ide o jednoduchšie vstupy a experimenty budú vykonané aj nad menšími vstupmi.

Predspracovanie bude zakončené vygenerovaním transformačnej matica a je použitím na transformáciu mračna bodov reprezentujúceho šablónu. V prvej fáze budú transformácie generované rovnakým spôsobom ako pri doterajšej práci, kedy boli použité rotácie okolo ľubovoľných osí z intervalu $< 0, 45 >$ stupňov a translácie z intervalu $< 0, 0.8 >$. Ďalšie kroky predspracovania ako mierne deformovanie objektov a vytváranie komplexnejších vstupov z viacerých objektov bude navrhnuté v závislosti od priebežne dosiahnutých výsledkov.

7 Zhodnotenie analýzy

V predloženej správe o riešení DP1 je poskytnutá analýza problematiky spracovania mračien bodov pomocou hlbokého učenia. Analyzované sú rôzne metódy predspracovania a reprezentácie mračien bodov spolu s rôznymi typmi architektúr neurónových sietí použitých na riešenie úloh ako klasifikácia, segmentácia alebo detekcia objektov. Špeciálny dôraz bol kladený na hlbokú neurónovú sieť PointNet, ktorej bola venovaná celá kapitola 3. Prínosom tohto modelu je, že umožňuje získanie globálnej reprezentácie črt pre mračná bodov reprezentované na vstupe ako neusporiadané množiny bodov. Od poznatkov získaných pri analýze tohto modelu sa bude odvíjať aj architektúra nášho riešenia. Ďalšia časť analýzy je venovaná špeciálne problematike registrácie objektov v mračnách bodov metódami hlbokého učenia. Opísané boli viaceré práce, pričom najvýznamnejšia z nich vychádzala zo spôsobu získavania črt aplikovanom v práci PointNet.

Posledná kapitola predloženej správy obsahuje návrh predspracovania dát a prístupu k vytvoreniu architektúry neurónovej siete, ktorá bude schopná riešiť problém registrácie objektov v mračnách bodov. V rámci diplomovej práce bude kľúčové overiť, že pomocou navrhnutého prístupu alebo jeho alternatív je možné riešiť problém registrácie objektov v mračnách bodov. Ak výsledky neurónovej siete nebudú ani na konci práce uspokojivé, bude možné vyhodnotiť, že zvolený prístup nie je vhodný na riešenie zadaného problému a ďalší čas sa pravdepodobne oplatí investovať do prieskumu iných metód.

V prípade, že dosiahnuté výsledky budú uspokojivé, navrhnutú neurónovú sieť bude možné využiť priamo v produkčnom prostredí v gumárenskom alebo automobilovom priemysle. Kontrola povrchu pneumatík (angl. *Tire Surface Inspection*) vrátane kontroly kvality textu na stranách je v súčasnosti vykonávaná tréningovými inšpektormi, keďže žiadna dostupná technológia neumožňuje vykonanie tejto kontroly.

7.1 Ďalšia práca

V kapitole 5 je opísaných osem bodov práce na projekt, z ktorých bola prvá polovica riešená počas zimného semestra. Najbližšími krokmi bude teda predspracovanie našich dát a vývoj modelu na vlastnom datasete. Plány pre ďalší vývoj modelov CorsNet a jeho súčasti PointNet sú opísané v samostatných častiach.

7.1.1 PointNet - ďalší vývoj

Pre vylepšenie modelu samotného modelu PointNet nie je veľa možností pre vylepšenie. PointNet je vo všeobecnosti považovaný za *state-of-the-art* riešenie pre získavanie vektorovej reprezentácie priamo z neusporiadaných mračien bodov a veľa iných modelov v oblasti registrácie objektov v 3D mračnách bodov využíva tento model podobným spôsobom ako CorsNet, práve na extrakciu lokálnych a globálnych črt. Predpokladáme, že aj na optimalizáciu hyperparametrov bolo vykonané veľké množstvo experimentov a týmto spôsobom výrazné zlepšenie nedosiahneme. Pre naše použitie siete je však možné rozšíriť augmentáciu dát a okrem pridávania šumu a rotácie okolo zvislej osi by mohlo byť vhodné pridať rotáciu okolo ľubovoľnej osi, prípadne rotáciu naraz vo viacerých smeroch. Vývoj samotného modelu PointNet však nie je cieľom tejto práce a spomenuté zmeny môžu byť aplikované aj pre model CorsNet, ktorého vylepšenie bude jadrom práce v nasledujúcom semestri.

7.1.2 CorsNet - ďalší vývoj

Okrem optimalizácie hyperparametrov a trénovania modelu na vlastnom datasete bude otestované ešte niekoľko možností vylepšenia siete. Model CorsNet bol trénovaný so štyrmi rôznymi verziami stratových funkcií, pričom v doterajších experimentoch sme pracovali iba s prvým typom. V ďalších krokoch bude určite zvážené použitie aj zvyšných typov. Podobne ako aktuálne používanú stratovú funkciu, aj zvyšné tri bude potrebné implementovať.

Ďalšia možná modifikácia siete vychádza z obrázku architektúry modelu CorsNet. Pri porovnaní PointNet časti tohto modelu a originálnej PointNet architektúry je možné vidieť, že pôvodný PointNet obsahuje 2 transformačné siete *T-Net*, zatiaľ čo v nákrese architektúry modelu CorsNet chýbajú. Úlohou týchto mini sietí je

aproximovať afínnu transformačnú maticu, čo má pomôcť eliminovať geometrické transformácie objektu ako rotácia alebo zrkadlenie. Pri klasifikácii je táto časť siete esenciálna, nakoľko je cieľom správne klasifikovať objekt bez ohľadu na jeho otočenie v priestore. Prínos tejto časti siete pre CorsNet však nemusí byť taký podstatný. Nakoľko cieľom CorsNet modelu je predikcia transformačnej matice, ktorá pozostáva aj z rotácie, transformovaním objektu v rámci PointNet časti sa môžu strácať informácie o rotácii a translácii v priestore. Odstránením *T-Net* častí siete by sa výrazne zmenšil počet trénovateľných parametrov, čo by umožnilo použitie väčších *batch* a celkové urýchlenie trénovania.

Posledným potencionálnym vylepšením, ktoré bude otestované je implementácia PointNet častí modelu ako siamskej neurónovej siete. Týmto spôsobom by bolo zaistené, že lokálne a globálne črty budú extrahované rovnakým spôsobom zo vstupného objektu aj šablóny.

Literatúra

1. QI, C. R. et al.: Pointnet: Deep learning on point sets for 3d classification and segmentation. In: *Proceedings of the IEEE conference on computer vision and pattern recognition*. 2017, s. 652–660.
2. DENG, L., YU, D.: Deep learning: methods and applications. *Foundations and Trends in Signal Processing*. 2014, vol. 7, no. 3–4, s. 197–387.
3. PANG, G., NEUMANN, U.: 3D point cloud object detection with multi-view convolutional neural network. In: *2016 23rd International Conference on Pattern Recognition (ICPR)*. 2016, s. 585–590.
4. SIMONY, M. et al.: Complex-yolo: An euler-region-proposal for real-time 3d object detection on point clouds. In: *Proceedings of the European Conference on Computer Vision (ECCV) Workshops*. 2018, s. 0–0.
5. QI, C. R. et al.: Frustum pointnets for 3d object detection from rgb-d data. In: *Proceedings of the IEEE conference on computer vision and pattern recognition*. 2018, s. 918–927.
6. WANG, Y. et al.: Dynamic graph cnn for learning on point clouds. *Acm Transactions On Graphics (tog)*. 2019, vol. 38, no. 5, s. 1–12.
7. ZHOU, Y., TUZEL, O.: Voxelnet: End-to-end learning for point cloud based 3d object detection. In: *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*. 2018, s. 4490–4499.
8. ENGELCKE, M. et al.: Vote3deep: Fast object detection in 3d point clouds using efficient convolutional neural networks. In: *2017 IEEE International Conference on Robotics and Automation (ICRA)*. 2017, s. 1355–1361.
9. HUANG, J., YOU, S.: Point cloud labeling using 3d convolutional neural network. In: *2016 23rd International Conference on Pattern Recognition (ICPR)*. 2016, s. 2670–2675.
10. ZHOU, Y. et al.: End-to-end multi-view fusion for 3d object detection in lidar point clouds. In: *Conference on Robot Learning*. 2020, s. 923–932.

11. LI, Y. et al.: MVF-CNN: Fusion of multilevel features for large-scale point cloud classification. *IEEE Access*. 2019, vol. 7, s. 46522–46537.
12. ZHI, S. et al.: Toward real-time 3D object recognition: A lightweight volumetric CNN framework using multitask learning. *Computers & Graphics*. 2018, vol. 71, s. 199–207.
13. BROWNLEE, J.: *Deep Learning for Computer Vision: Image Classification, Object Detection, and Face Recognition in Python*. Machine Learning Mastery, 2019.
14. KANG, B., TRIPATHI, S., NGUYEN, T. Q.: Real-time sign language fingerspelling recognition using convolutional neural networks from depth map. In: *2015 3rd IAPR Asian Conference on Pattern Recognition (ACPR)*. 2015, s. 136–140.
15. SU, H. et al.: Multi-view convolutional neural networks for 3d shape recognition. In: *Proceedings of the IEEE international conference on computer vision*. 2015, s. 945–953.
16. REDMON, J., FARHADI, A.: YOLO9000: better, faster, stronger. In: *Proceedings of the IEEE conference on computer vision and pattern recognition*. 2017, s. 7263–7271.
17. YANG, B., LUO, W., URTASUN, R.: Pixor: Real-time 3d object detection from point clouds. In: *Proceedings of the IEEE conference on Computer Vision and Pattern Recognition*. 2018, s. 7652–7660.
18. QI, C. R. et al.: Deep hough voting for 3d object detection in point clouds. In: *Proceedings of the IEEE/CVF International Conference on Computer Vision*. 2019, s. 9277–9286.
19. PANCHAL, G. et al.: Behaviour analysis of multilayer perceptrons with multiple hidden neurons and hidden layers. *International Journal of Computer Theory and Engineering*. 2011, vol. 3, no. 2, s. 332–337.
20. ZENG, H.: Towards better understanding of deep learning with visualization. *The Hong Kong University of Science and Technology*. 2016.

21. LIN, C.-C. et al.: A novel point cloud registration using 2D image features. *EURASIP Journal on Advances in Signal Processing*. 2017, vol. 2017, no. 1, s. 1–11.
22. LU, W. et al.: DeepICP: An end-to-end deep neural network for 3D point cloud registration. *arXiv preprint arXiv:1905.04153*. 2019.
23. GROSS, J., OŠEP, A., LEIBE, B.: Alignnet-3d: Fast point cloud registration of partially observed objects. In: *2019 International Conference on 3D Vision (3DV)*. 2019, s. 623–632.
24. PEREZ-GONZALEZ, J., LUNA-MADRIGAL, F., PIÑA-RAMIREZ, O.: Deep learning point cloud registration based on distance features. *IEEE Latin America Transactions*. 2019, vol. 17, no. 12, s. 2053–2060.
25. ELBAZ, G., AVRAHAM, T., FISCHER, A.: 3D point cloud registration for localization using a deep neural network auto-encoder. In: *Proceedings of the IEEE conference on computer vision and pattern recognition*. 2017, s. 4631–4640.
26. WANG, Y., SOLOMON, J. M.: Deep closest point: Learning representations for point cloud registration. In: *Proceedings of the IEEE/CVF International Conference on Computer Vision*. 2019, s. 3523–3532.
27. ZHANG, Z., DAI, Y., SUN, J.: Deep learning based point cloud registration: an overview. *Virtual Reality & Intelligent Hardware*. 2020, vol. 2, no. 3, s. 222–246.
28. KUROBE, A. et al.: CorsNet: 3D point cloud registration by deep neural network. *IEEE Robotics and Automation Letters*. 2020, vol. 5, no. 3, s. 3960–3966.
29. AOKI, Y. et al.: Pointnetlk: Robust & efficient point cloud registration using pointnet. In: *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*. 2019, s. 7163–7172.
30. ROJAS, R.: Lucas-kanade in a nutshell. *Freie Universit at Berlinn, Dept. of Computer Science, Tech. Rep*. 2010.

Príloha A: Plán práce na riešení projektu

A.1 DP1

- 01.03.2021 - 07.03.2021: Zber a triedenie ďalších materiálov k téme diplomovej práce.
- 08.03.2021 - 14.03.2021: Inštalácia a spustenie modelu PointNet.
- 15.03.2021 - 21.03.2021: Podrobná analýza architektúry PointNet z vydanej práce a dostupného zdrojového kódu.
- 22.03.2021 - 04.04.2021: Podrobný opis funkcionality, architektúry a spôsobu reprezentácie vstupných dát pre model PointNet.
- 05.04.2021 - 18.04.2021: Predspracovanie dát pre vlastný model na registráciu objektov v mračnách bodov.
- 19.04.2021 - 02.05.2021: Písanie častí práce zameraných na spracovanie mračien bodov hlbokým učením a registrácie objektov v mračnách.
- 03.05.2021 - 09.05.2021: Pridanie kapitol 1, 5, 7 a plánu práce pre jednotlivé fázy riešenia diplomovej práce.
- 10.05.2021 - 16.05.2021: Finalizácia priebežnej správy o riešení DP1 a zapracovanie pripomienok.
- 17.05.2021: Odovzdanie priebežnej správy o riešení DP1.

V úvode semestra bol harmonogram práce dodržiavaný podľa plánu, čo sa však postupne s nárastom iných študijných povinností skomplikovalo. Písanie práce sa celkovo presunulo až do druhej polovice semestra. Oproti pôvodnému plánu boli vykonané určité obmeny, v rámci ktorých bolo vynechané predspracovanie dát, pričom ale bola napísaná rozsiahlejšia analýza prístupov k spracovaniu mračien bodov hlbokým učením, než bolo na začiatku predpokladané. Aj napriek výskytu niekoľkých komplikácií a posunu časového harmonogramu bola priebežná správa o riešení DP1 úspešne odovzdaná v požadovanom termíne.

A.2 DP2

Pri finalizácii priebežnej správy o riešení DP1 nie je možné presne určiť časový harmonogram plánu práce pre fázu DP2, nakoľko je plánované vykonať pedspracovanie dát a prípadne aj začať s vývojom vlastného modelu hlbokého učenia počas skúškového obdobia a pred začatím ďalšieho semestra. Body práce, ktorým sa budem venovať v rámci riešenia DP2 budú zahŕňať nasledovné:

- Zapracovanie pripomienok k častiam práce odovzdaným v rámci priebežnej správy o riešení DP1.
- Dokončenie pedspracovanie dát na trénovanie vlastného modelu hlbokého učenia v prípade, že tak nebolo urobené do začiatku semestra.
- Implementácia modelu inšpirovaného prácou PointNet na získanie komplexnej reprezentácie črt zo vstupných dát.
- Evaluácia a optimalizácia časti na získanie reprezentácie črt zo vstupných dát.
- Rozšírenie modelu o registráciu objektov v mračnách bodov.
- Evaluácia a optimalizácia vlastného natrénovaného modelu.
- Zhrnutie výstupov modelu a získaných poznatkov do priebežnej správy o riešení DP2.

A.3 DP3

Konkrétny plán práce pre finálnu fázu diplomovej práce DP3 bude vyvodený na základe výsledkov dosiahnutých vo fáze DP2. Predpokladaná náplň práce bude zahŕňať nasledovné body:

- Zapracovanie pripomienok k častiam práce odovzdaným v rámci priebežnej správy o riešení DP2.
- Vývoj, optimalizácia a ďalšie úpravy modelu hlbokého učenia implementovaného v predchádzajúcej fáze.
- Vyskúšanie alternatívnych prístupov aplikovateľných na náš problém registrácie objektov v mračnách bodov.
- Finalizácia písania diplomovej práce.