

Slovenská technická univerzita v Bratislave  
Fakulta informatiky a informačných technológií

FIIT-XXXX-XXXXX

Bc. Katarína Juhásová

REGISTRÁCIA OBJEKTŮ NA MRAČNE BODOV  
POMOCOU HLBOKÝCH NEURÓNŮVÝCH SIETÍ

Priebežná správa o riešení DP1

Vedúci práce: RNDr. Andrej Lúčný, PhD.

Máj 2021

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 DP1

Študijný program:     Inteligentné softvérové systémy  
Študijný odbor:       Informatika  
Miesto vypracovania:   Ústav informatiky, informačných systémov  
                             a softvérového inžinierstva  
Vedúci práce:         RNDr. Andrej Lúčny, PhD.  
Máj 2021

# Návrh zadania diplomovej práce

*Finálna verzia do diplomovej práce <sup>1</sup>*

## 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 zadania<sup>2</sup>

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 hlbkových 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 hlbkej 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.

<sup>1</sup> Vytlačiť obojstranne na jeden list papiera

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

### Literatúra<sup>3</sup>

- 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 / nie<sup>4</sup>

Dňa: .....

---

Podpis garanta predmetov

---

<sup>3</sup> 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)

<sup>4</sup> Nehodiace sa prečiarknite

## **ČESTNÉ PREHLÁSENIE**

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

V Bratislave, 16. máj 2021

.....  
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.

Máj 2021

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

2021, May

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

---

<b>1</b>	<b>Úvod</b>	<b>1</b>
<b>2</b>	<b>Spracovanie mračien bodov hlbokým učením</b>	<b>2</b>
2.1	Spracovanie transformovaných mračien bodov . . . . .	3
2.2	Spracovanie mračien bodov priamo . . . . .	4
<b>3</b>	<b>PointNet</b>	<b>6</b>
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
<b>4</b>	<b>Registrácia objektov v mračnách bodov</b>	<b>13</b>
<b>5</b>	<b>Návrh riešenia</b>	<b>16</b>
5.1	Vstupné dáta a predspracovanie . . . . .	16
5.2	Architektúra hlbkej neurónovej siete . . . . .	17
<b>6</b>	<b>Zhodnotenie analýzy</b>	<b>18</b>
	<b>Literatúra</b>	<b>19</b>
	<b>Príloha A: Plán práce na riešení projektu</b>	<b>A-1</b>
A.1	DP1 . . . . .	A-1
A.2	DP2 . . . . .	A-2
A.3	DP3 . . . . .	A-2

## Zoznam použitých skratiek

---

<b>CNN</b>	Convolutional Neural Network
<b>FCN</b>	Fully Convolutional Network
<b>MLP</b>	Multilayer Perceptron
<b>RPN</b>	Region Proposal Network
<b>SVD</b>	Singular Value Decomposition

# 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 počiatkoch. 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 vtácej 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 vtácej perspektívy. Vylepšením oproti Complex-YOLO je, že výsledné *bounding boxy* určujú nie len pozíciu objektov, ale aj smer ich natočenia.

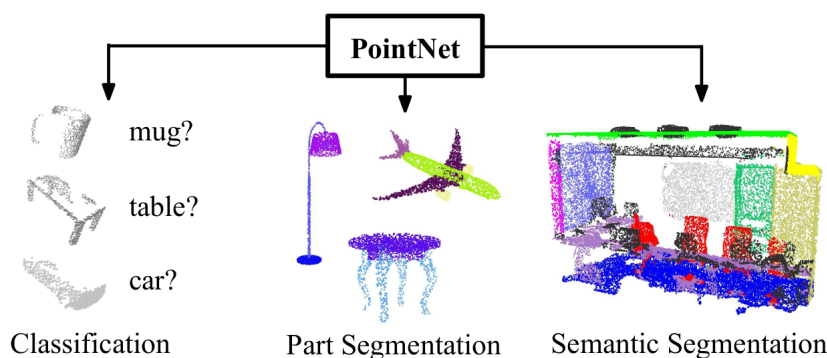
## 2.2 Spracovanie mračien bodov priamo

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 mračnami.



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.

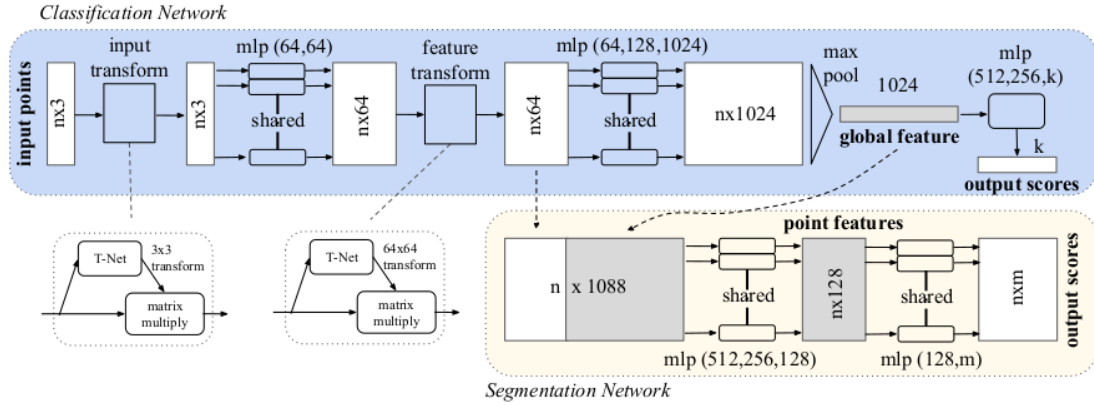
## 3.2 Architektúra PointNet

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.

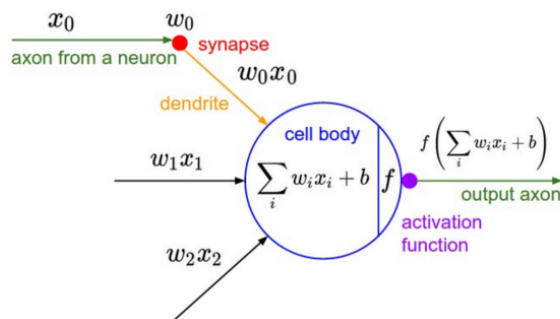
Cielom pri návrhu siete bolo, aby neurónová sieť vedela aproximovať funkciu  $f$  aplikovaním symetrickej funkcie na transformované body z množiny:

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

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 blízkyh 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  $3 \times 3$  je do výpočtu aktivácie pre konkrétny bod zahrnutá hodnota bodu samotného spolu s jeho ôsmimi susediacimi bodmi. V modeli PointNet sú použité konvolúcie  $1 \times 1$ .

Využitie  $1 \times 1$  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  $3 \times 3$ , 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  $64 \times 64$ , 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.

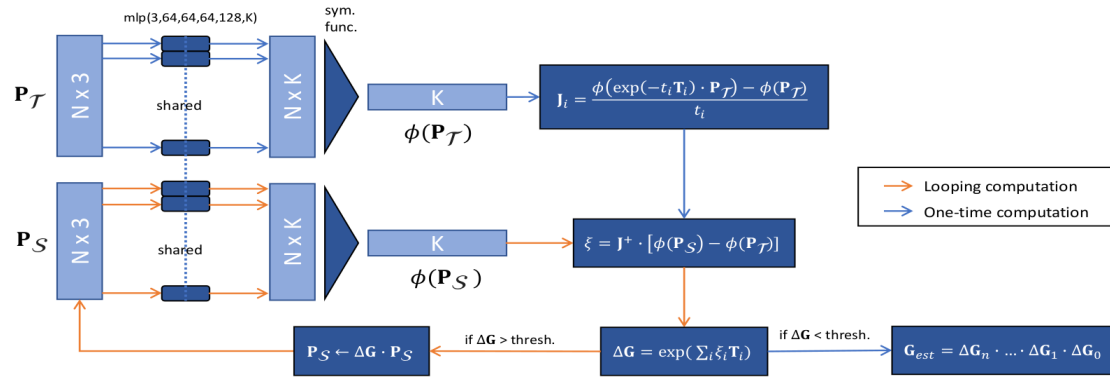
Pre úlohy segmentácie s  $n$  množstvom bodov v mračne a  $m$  možnými sémantickými kategóriami je výstupom modelu matica veľkosti  $n \times m$  obsahujúca skóre pre každú z  $m$  možných kategórií pre všetky body.

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  $1 \times 1$  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 uspokojivé, znamená to, že hľadaný vzor sa v mračne nenachádza, prípadne že sa nachádza v dátach výrazne zdeformovaný alebo nekompletný.



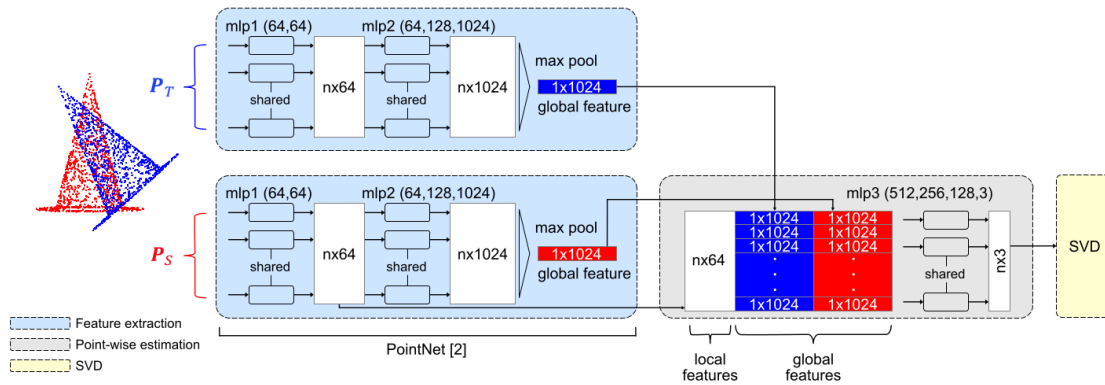
Obr. 4.1: *PointNetLK* architektúra. Z mračien bodov pre vzor (template)  $\mathbf{P}_{\mathcal{T}}$  a vstupný objekt (source)  $\mathbf{P}_{\mathcal{S}}$  sú extrahované vektorové reprezentácie globálnych črt  $\phi(\mathbf{P}_{\mathcal{T}})$  a  $\phi(\mathbf{P}_{\mathcal{S}})$ . Jacobiho matica je pre vzor vypočítaná iba raz a ďalej je používaná pri výpočte parametrov otočenia. Pozícia pôvodného vstupného objektu  $\mathbf{P}_{\mathcal{S}}$  je následne iteratívne upravovaná, pričom prebieha aj prepočet jeho vektorovej reprezentácie [21].

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 [21]. 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  $\Delta t$  medzi obrázkami je malý. Tento vektor posunutia je vypočítaný pre vybrané významné body obrázku. Prehľadné zhrnutie algoritmu je opísané v článku [22]. 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.

Lepšie výsledky ako PointNetLK dosahuje model hlbokého učenia CorsNet [23]. Na rozdiel od PointNetLK, ktorý pracuje iba s globálnou reprezentáciou črt, tento model používa zlúčenú reprezentáciu lokálnych a globálnych črt. Architektúru siete je možné vidieť na obrázku 4.2. Model tejto siete je pozostáva z troch častí:

- získanie globálnych črt zo vzoru a vstupného mračna bodov,
- 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. [23]



Opäť aj pre registráciu objektov existujú riešenia, ktoré transformujú riešený problém do 2D priestoru. Príkladom je model z práce [24], ktorý vykonáva registráciu objektov na základe porovnávania lokálnych 2D črt. Medzi ďalšie práce zamerané na registráciu objektov v mračnách bodov patria [25, 26, 27, 28, 29] a stručný prehľad o existujúcich metódach spracovali Zhang a kolektív [30].

## 5 Návrh riešenia

---

### 5.1 Vstupné dáta a predspracovanie

Aktuálne dostupné vstupné dáta pre túto prácu obsahujú 66 skenov pneumatík z oboch strán (dokopy 132 skenov) spolu s hľadaným vzorom a ďalšími súbormi užitočnými pre predspracovanie dát. Pre každý zo skenov sú k dispozícii nasledovné súbory:

- referencia, tzv. recept alebo vzor, ktorý má byť nájdený na skene pneumatiky,
- referencia zredukovaná z pôvodnej šírky 65536 na 240000 pixelov,
- vyrovnaný 3D sken pneumatiky, v ktorom je pneumatika vyrovnaná do obdĺžnika, pričom je zachované zakrivenie vo vertikálnom smere,
- sken pneumatiky vyrovnaný do 2D hĺbkovej mapy pomocou Jacobiho iteračného algoritmu,
- vyrovnaná 2D hĺbková mapa s veľkosťou zredukovanou z pôvodných rozmerov  $65536 \times 1600$  na  $24000 \times 1600$ ,
- polozenie referencie na vyrovnané meranie (sken) spolu s offsetom posunutia merania voči hľadanému vzoru,
- polozenie referencie na vyrovnané meranie s farebne odlíšenými časťami referencie (*targets*), kde každý objekt ako písmeno alebo ilustračný prvok je považovaný za samostatný *target*,
- finálna registrácia určená z polozenia referencie na meranie, kde sa však nachádzajú aj chyby.

V rámci predspracovania dát bude potrebné vykonať nasledovné kroky:

- utriedenie zle anotovaných obrázkov s chybnou finálnou registráciou,
- transformácia polozenia referencie na meranie na binárny obraz polozenia, kde bude pre jednotlivé body určené, či sa nachádzajú v hľadanom vzore alebo nie,
- zredukovanie veľkosti dát takým spôsobom, aby bolo zachované iba zaujímavé časti obsahujúce vzor a orezané časti predstavujúce čistý povrch pneumatiky,

- anotácia dát takým spôsobom, aby bola pre každý bod dostupná informácia o objekte, ktorý sa na danom mieste nachádza, prípadne hodnota 0 alebo  $-1$  pre body nepatriacich do referencie.

## 5.2 Architektúra hlbkej neurónovej siete

Architektúra neurónovej siete navrhutej v rámci tejto práce bude primárne vychádzať z architektúry modelu PointNet [1], ktorý bol podrobne opísaný v kapitole 3 s architektúrou načrtnutou na obrázku 3.2. Táto architektúra sa preukázala ako úspešná pri spracovaní mračen bodov priamo bez špeciálneho predspracovania a transformácie. Implementovaným modelom by malo byť možné získať reprezentáciu globálnych črt samostatne pre hľadanie referencie a aj pôvodný sken pneumatiky. Následne by malo byť potrebné iba zrovnať tieto dve reprezentácie. Takýto prístup aplikovali aj Kurobe a kolektív [23] pri vývoji modelu CorsNet, ktorý je opísaný v kapitole 4. V tejto hlbkej neurónovej sieti sú lokálne a globálne črty extrahované pomocou PointNet prístupu a na registráciu objektov je použitých niekoľko MLP. Z architektúry CorsNet na obrázku 4.2 je vidno, že zo získaných reprezentácií boli následne zlúčené globálne črty pre vzor a lokálne aj globálne črty pre vstupné mračno bodov a pomocou niekoľkých MLP a singulárneho rozkladu (SVD) bola určená transformačná matica pre pôvodný vstup. Predpokladáme, že takýto alebo podobný postup by mohol priniesť úspešné výsledky aj pri riešení problému registrácie nad našimi dátami. Podľa dosiahnutých výsledkov bude ďalej zvažované použitie iných aproximačných metód alebo algoritmov na registráciu objektov.

Alternatívnym prístupom by mohla byť transformácia problému na segmentačný problém, na čo by mohol byť využitý model implementovaný podľa PointNet. Identifikované segmenty by následne mohli byť porovnané s pôvodným vzorom.

## 6 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.

## 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. 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.
22. ROJAS, R.: Lucas-kanade in a nutshell. *Freie Universit at Berlinn, Dept. of Computer Science, Tech. Rep*. 2010.
23. 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.
24. 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.
25. LU, W. et al.: DeepICP: An end-to-end deep neural network for 3D point cloud registration. *arXiv preprint arXiv:1905.04153*. 2019.
26. 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.
27. 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.
28. 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.
29. 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.
30. 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.

## 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, 6 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 predspracovanie 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 predspracovanie dát na tréningovanie 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.