SACAIR 2024

Proceedings
of the Fifth Southern African Conference for Artificial Intelligence Research

December 2024
Bloemfontein, South Africa

Editors:

Aurona Gerber
Jacques Maritz
Anban Pillay

SACAIR
Southern African Conference for Artificial Intelligence Research

ISBN : 978-0-7961-6069-0

Proceedings of the 5th Southern African Conference for Artificial Intelligence Research

SACAIR 2024

ISBN: 978-0-7961-6069-0 (e-book)

Editors

Aurona Gerber

University of the Western Cape
Cape Town, South Africa

Jacques Maritz

University of the Free State
Bloemfontein, South Africa

Anban Pillay

University of KwaZulu-Natal
Durban, South Africa

The SACAIR 2024 Proceedings Editors attest as follows:
All conference paper submissions that appear in these proceedings have been double-blind peer reviewed by an international programme committee prior to acceptance and publication.

Published Online by

The SACAIR Steering Committee
Private Bag X20, Hatfield, 0028

December 2024

Preface

Foreword from the Conference Chairs

Dear authors and readers,

It is with great pleasure that we write this foreword to the Proceedings of the fifth Southern African Conference for Artificial Intelligence Research (SACAIR 2024) held in-person at the University of the Free State, Bloemfontein, South Africa from 2 to 6 December 2024. The programme included an unconference for students on 2 December 2024 (a student-driven event that included talks and interactions with industry), a day of tutorials on 3 December, and the main conference from 4-6 December 2024.

SACAIR 2024 is the fifth conference in the series of trans- and inter-disciplinary conferences administered by the SACAIR Steering Committee, an affiliate of the Centre for AI Research (CAIR), South Africa. The Centre for AI Research (CAIR)[1] is a South African distributed research network that was established in 2011 with the aim of building world-class Artificial Intelligence research capacity in Southern Africa. CAIR conducts foundational, directed and applied research into various aspects of AI through its various research groups based at higher education institutions in South Africa.

Although still a young conference, SACAIR is quickly establishing itself as a premier artificial intelligence conference in the Southern African region. The fifth conference builds on the success of previous conferences. The inaugural CAIR conference, the Forum for AI Research (FAIR 2019), was held in Cape Town, South Africa, in December 2019. SACAIR 2020 was held in February 2021 after being postponed due to the Covid pandemic and SACAIR 2021 was an online event hosted by the University of KwaZulu-Natal in December 2021. The 2022 conference edition was held in Stellenbosch, Western Cape, and SACAIR 2023 was held at the 26 Degrees South venue, Muldersdrift, Gauteng, South Africa, from 4-8 December 2023.

We are pleased that SACAIR 2024 continued to enjoy the support of the South African artificial intelligence research community. The conference, held under the theme of *Artificial Intelligence for Societal Impact*, brought together a diverse group of researchers and practitioners. Artificial Intelligence has indeed made significant strides in sifting information and offering targeted solutions to real-world challenges. The advancement of responsible AI research is of paramount importance, as the role of AI in shaping our future societies cannot be overstated. However, the conversion of how these advancements are developing into tangible outcomes in our immediate contexts is often not achieved. This is particularly relevant in areas where societal hardships persist due to the complex interplay of socio-political, economic, historical, and environmental realities. This year's conference is therefore centred on exploring the actual societal

[1] https://www.cair.org.za/

impact delivered by artificial intelligence beyond its potential promises and from all scientific perspectives.

The conference attracted support from both authors, who submitted high-quality research papers, as well as researchers who supported the conference by serving on the international program committee. SACAIR 2024 brought together a diverse group of researchers and practitioners in the fields of Computer Science, Information Systems, Knowledge Representation and Reasoning, Law, and Philosophy of AI. The conference was organized as a multi-track conference that would cover broad areas of Artificial Intelligence namely:

- Algorithmic and Data-Driven AI (Computer Science).
- Symbolic AI (Knowledge Representation and Reasoning).
- Socio-technical and human-centred AI (Information Systems).
- Responsible and Ethical AI (Philosophy and Law / Humanities).

The accepted papers include contributions from symbolic AI and those from data-driven AI, as the focus on knowledge representation and reasoning remains an important ingredient of studying and extending human intelligence. In addition, important contributions from the fields of socio-technical and human-centred AI, as well as responsible and ethical AI are included in this volume. We expect this multi- and interdisciplinary conference to grow into the premier AI conference in Southern Africa as it brings together nationally and internationally established and emerging researchers from across various disciplines including Computer Science, Mathematics, Statistics, Informatics, Philosophy and Law. The conference is also focused on cultivating and establishing a network of talented students working in AI from across Africa.

A conference of this nature is not possible without the hard work and contributions of many stakeholders. We extend our sincere gratitude to our sponsors: the Artificial Intelligence Journal (AIJ), the National Institute of Computational Sciences (NiTHeCS), the University of the Free State (UFS) and the Centre for Artificial Intelligence Research (CAIR). These sponsors have made it possible to offer generous scholarships to students and emerging academics to participate in the conference. We sincerely thank the technical chairs for their work in overseeing the technical aspects of the conference and the publication of the two volumes of the proceedings, the international panel of reviewers, our keynotes, authors, and participants for their contributions. Finally, we extend our gratitude to the track chairs, the local organising committee, the student organisers, and the conference organiser for their substantive contributions to the success of SACAIR 2024.

Katinka de Wet and Herkulaas Combrink
Organising Chairs: SACAIR 2024
December 2024

The program committee comprised 156 members (representing some 73 research institutions), 29 of whom were from outside Southern Africa. Each paper was sent for review to at least three members of the program committee in a rigorous, double-blind peer-review process. Most papers received at least three reviews, often followed by a meta-review by the respective track chairs. Great care was taken to ensure the conference's integrity, including careful attention to avoid conflicts of interest. The following criteria were used to rate submissions and to guide decisions: relevance to SACAIR, significance, technical quality, scholarship, and presentation (including quality and clarity of writing).

We received 153 abstracts, and after submission and a first round of evaluation, 101 submissions were sent to our SACAIR program committee for review. The papers consisted of 70 in the CS track, 12 in the IS track, 5 in the KRR track and 14 in the P&L track. Twenty-one full research papers were selected for publication in this online volume (acceptance rate 21%). The total acceptance rate for publication in the two SACAIR 2024 proceedings volumes was 50% for reviewed submissions. In total, two papers from the Responsible and Ethical AI track, four papers from the Socio-technical and Human-Centered AI track, and 13 papers from the Algorithmic and Data-Driven AI track were accepted for publication in this online SACAIR volume.

Thank you to all the authors who submitted work of an exceptional standard to the conference and congratulations to the authors whose work was accepted for publication. We place on record our gratitude to the Program Committee members, whose thoughtful and constructive comments were well received by the authors.

Aurona Gerber, Jacques Maritz and Anban Pillay
Technical Chairs: SACAIR 2024
December 2024

SACAIR2024 Attestation

The 2024 Southern African Conference on AI Research (SACAIR 2024) held as a hybrid-online and in-person event at the University of the Free State, Bloemfontein, South Africa from 4 to 6 December 2024.

We received 153 abstracts, and after submission and a first round of evaluation, 101 submissions were sent to our SACAIR program committee for review. The papers consisted of 70 in the CS track, 12 in the IS track, 5 in the KRR track and 14 in the P&L track. All the papers were also double-blind and peer-reviewed by at least two technical reviewers from the program committee.

The program committee comprised 156 members (representing some 73 research institutions), 29 of whom were from outside Southern Africa. Each paper was reviewed by at least two members of the program committee in a rigorous, double-blind process. Great care was taken to ensure the integrity of the conference including careful attention to avoid conflicts of interest. The following criteria were used to rate submissions and to guide decisions: relevance to SACAIR, significance, technical quality, scholarship, and presentation, which included quality and clarity of writing.

Twenty-one full research papers were selected for publication in this online volume (acceptance rate 21%). This is the second SACAIR 2024 proceedings volume, the first volume is published as a volume of Springer CCIS (CCIS 2326). The total acceptance rate for publication in the two SACAIR 2024 proceedings volumes was 50% for reviewed submissions. In total, two papers from the Responsible and Ethical AI track, four papers from the Socio-technical and Human-Centered AI track, and 13 papers from the Algorithmic and Data-Driven AI track were accepted for publication in this online SACAIR proceedings volume with the ISBN number `978-0-7961-6069-0` (e-book).

Authors of accepted papers were affiliated with various national and international universities. The table below indicates the percentage of authors from the respective institutions for this online proceedings volume.

We can thus confirm that more than 75% of the papers were authored by researchers from different universities.

Institution	Author Percentage
CSIR	2
North-West University	2
Stellenbosch University	4
University of Cape Town	19
University of Johannesburg	6
University of KwaZulu-Natal	10
University of Pretoria	15
University of the Free State	6
University of the Western Cape	8
International Institutions	23

Aurona Gerber, Jacques Maritz and Anban Pillay
Technical Chairs: SACAIR 2024
December 2024

Aurona Gerber, Emma Ruttkamp-Bloem, Anban Pillay, Edgar Jembere
SACAIR Steering Committee
December 2024

SACAIR2024 Organisation

Local Organising Committee

Katinka de Wet (General Co-chair)	University of the Free State, South Africa
Herkulaas Combrink (General Co-chair)	University of the Free State, South Africa
Renee le Roux (Conference Organiser)	Mongoose Communications & Design, South Africa

Technical Committee

Susan Brokensha	University of the Free State, South Africa
Aurona Gerber	University of the Western Cape, South Africa
Anban Pillay	University of KwaZulu-Natal, South Africa
Jacques Maritz	University of the Free State, South Africa
Wynand Nel	University of the Free State, South Africa

Program Committee

Algorithmic and Data-Driven AI (Computer Science)

Track Chairs

Akinyelu, Andronicus	University of the Free State, South Africa
Davel, Marelie	North-West University, South Africa
Jembere, Edgar	University of KwaZulu-Natal, South Africa
Nyirenda, Clement	University of the Western Cape, South Africa
Pillay, Anban	University of KwaZulu-Natal, South Africa
Van Zyl, Terence	University of Johannesburg, South Africa

PC Members

Abate, Demmelash	Wolaita Sodo University, Ethiopia
Achary, Thimershen	Independent, South Africa
Agbobli, Gabriel	University of Ghana, Ghana
Amankwatia, Manuel	Kwame Nkrumah University of Science and Technology, Ghana
Amuda, Kamorudeen	University of Ibadan, Nigeria
Amukwatse, Jean	Uganda Technology and Management University, Uganda
Anderson, George	University of Botswana, Botswana
Asaju, Christine	Thswane University of Technology, South Africa
Atemkeng, Marcellin	Rhodes University, South Africa
Barnard, Etienne	North-West University, South Africa
Batani, John	University of KwaZulu-Natal, South Africa
Bopape, Lebogang	Durban University of Technology, South Africa
Bosman, Anna Sergeevna	University of Pretoria, South Africa
Brink, Willie	Stellenbosch University, South Africa
Brown, Dane	Rhodes University, South Africa
Buys, Jan	University of Cape Town, South Africa
Chilipa, Richard	Malawi University of Science and Technology, Malawi
Daramola, Olawande	University of Pretoria, South Africa
De Freitas, Allan	University of Pretoria, South Africa
Derks, Iena	University of Pretoria, South Africa
Dlamini, Johnson	University of Pretoria, South Africa
Dlamini, Sibonelo	University of KwaZulu-Natal, South Africa
Du Toit, Tiny	North-West University, South Africa
Eiselen, Roald	North-West University, South Africa

Engelbrecht, Emile	Stellenbosch University, South Africa
Ezeji, Ijeoma	University of Zululand, South Africa
Ezugwu, Absalom	North-West University, South Africa
Fabris-Rotelli, Inger	University of Pretoria, South Africa
Fashoto, Stephen	University of Swaziland, Eswatini
Fendji, Louis	University of Ngaoundere, Cameroon
Formanek, Claude	University of Cape Town, South Africa
Harley, Charis	University of Johannesburg, South Africa
Haskins, Bertram	Nelson Mandela University, South Africa
Helberg, Albert	North-West University, South Africa
Hounkanrin, Anicet	University of Cape Town, South Africa
Isingizwe, Frederic	University of the Western Cape, South Africa
James, Steven	University of the Witwatersrand, South Africa
Jeewa, Asad	University of KwaZulu-Natal, South Africa
Kankuzi, Bennett	Malawi University of Science and Technology, Malawi
Karim, Zainoolabadien	University of the Witwatersrand, South Africa
Khangamwa, Gift	University of the Witwatersrand, South Africa
Klein, Richard	University of the Witwatersrand, South Africa
Kogeda, Paul Okuthe	University of KwaZulu-Natal, South Africa
Kotze, Michelle	University of Pretoria, South Africa
Kotzé, Eduan	University of the Free State, South Africa
Laidi, Amel	Abou Bekr Belkaid University, Algeria
Louw, Aby	Council for Scientific and Industrial Research, South Africa
Makhoere, Lucky	Sol Plaatje University, South Africa
Malan, Katherine	University of South Africa, South Africa
Marais, Patrick	University of Cape Town, South Africa
Masethe, Dan	Tshwane University of Technology, South Africa
Mathebula, Miehleketo	University of Pretoria, South Africa
Mathonsi, Topside	Tshwane University of Technology, South Africa
Mauwa, Hope	University of Mpumalanga, South Africa
Melatagia Yonta, Paulin	University of Yaounde, Cameroon
Modipa, Thipe	University of Limpopo, South Africa
Modupe, Abiodun	University of Pretoria, South Africa
Mogale, Hope	North-West University, South Africa
Moodley, Deshendran	University of Cape Town, South Africa
Moore, Stephen	University of Cape Coast, Ghana
Motsoehli, Moseli	University of Hawaii at Manoa, USA
Mthethwa, Sthembile	Council for Scientific and Industrial Research, South Africa
Muite, Benson	University of Tartu, Estonia
Mulaudzi, Rudzani	University of the Witwatersrand, South Africa
Mungloo-Dilmohamud, Zahra	University of Mauritius, Mauritius
Mutanga, Murimo Bethel	Mangosuthu University of Technology, South Africa
Mutembei, Leonard	University of Johannesburg, South Africa

Nandutu, Irene	Uganda Technology and Management University, Uganda
Nemakhavhani, Ndamulelo	University of Pretoria, South Africa
Ngxande, Mkhuseli	Stellenbosch University, South Africa
Nicolls, Fred	University of Cape Town, South Africa
Nogwina, Mnoneleli	Sol Plaatje University, South Africa
Nzomo, Mbithe	University of Cape Town, South Africa
Okuboyejo, Damilola	Durban University of Technology, South Africa
Olusanya, Michael	Sol Plaatje University, South Africa
Oluyide, Oluwakorede Monica	University of the Western Cape, South Africa
Onuegbu, Joseph	Ariel University, Israel
Otoo, Joseph	C. K. Tandem University of Technology and Applied Sciences, Ghana
Rananga, Seani	University of Pretoria, South Africa
Schlippe, Tim	IU International University of Applied Sciences, Germany
Sefara, Tshephisho Joseph	University of Limpopo, South Africa
Semwayo, Thembinkosi	University of the Witwatersrand, South Africa
Senekane, Makhamisa	University of Johannesburg, South Africa
Shiabaw, Dawit	University of Gondar, Ethiopia
Shock, Jonathan	University of Cape Town, South Africa
Singh, Yuvika	University of KwaZulu-Natal, South Africa
Tachiona, Terence	Midlands State University, Zimbabwe
Tapamo, Jules-Raymond	Univesity of KwaZulu-Natal, South Africa
Tom, Sheethal	Cape Peninsula University of Technology, South Africa
Tuyishimire, Emmanuel	North-West University, South Africa
Vadapalli, Hima Bindu	University of Johannesburg, South Africa
Wang, Hairong	University of the Witwatersrand, South Africa
Watson, Bruce	National Security Centre of Excellence, Canada
Wegner, Jan Dirk	University of Switzerland, Switzerland
Yibre, Abdulkerim Mohammed	Bahir Dar Institute of Technology, Bahir Dar University, Ethiopia

Socio-technical and Human-Centred AI (Information Systems)

Track Chairs

Eybers, Sunet	University of South Africa, South Africa
Hinkelmann, Knut	FHNW University of Applied Sciences and Arts, Northwestern Switzerland

PC Members

Abdullah, Hanifa	University of South Africa, South Africa
Botha, Benjamin	University of the Free State, South Africa
Callisto De Donato, Massimo	University of Camerino, Italy
Fornari, Fabrizio	University of Camerino, Italy
Hanslo, Ridewaan	University of Johannesburg, South Africa
Howard, Grant Royd	University of South Africa, South Africa
Härer, Felix	University of Fribourg, Switzerland
Laurenzi, Emanuele	FHNW University of Applied Sciences and Arts, Northwestern Switzerland
Mahlaza, Zola	University of Cape Town, South Africa
Martin, Andreas	FHNW University of Applied Sciences and Arts, Northwestern Switzerland
Matthee, Machdel	University of Pretoria, South Africa
Parry, Douglas	Stellenbosch University, South Africa
Piangerelli, Marco	University of Camerino, Italy
Pretorius, Henk	University of Pretoria, South Africa
Price, Catherine S.	University of KwaZulu-Natal, South Africa
Smit, Danie	University of Pretoria, South Africa
Smuts, Hanlie	University of Pretoria, South Africa
Van Belle, Jean-Paul	University of Cape Town, South Africa
Van Staden, Corné	University of South Africa, South Africa

Responsible and Ethical AI (Philosophy, Law and Humanities)

Track Chairs

Ruttkamp-Bloem, Emma	University of Pretoria, South Africa
Tollon, Fabio	University of Edinburgh, UK

PC Members

Basu, Kaberi	Trinity College Dublin, Ireland
Benhammou, Morraya	The Hague University of Applied Sciences, Netherlands
Benton, Paige	University of Pretoria, South Africa
Bock, Ivan	Stellenbosch University, South Africa
Claassen, Kristy	University of Twente, Netherlands
Formánek, Arzu	University of Vienna, Austria
Friedman, Cindy	Utrecht University, Netherlands
Garforth, James	University of Edinburgh, UK
Robertson, Helen Sarah	University of the Witwatersrand, South Africa
Schulz, Maria	Tilburg University, Germany
Vorster, Ethan	University of Johannesburg, South Africa
Xivuri, Khensani	University of Johannesburg, South Africa
de Villiers-Botha, Tanya	Stellenbosch University, South Africa

Knowledge Representation and Reasoning (Symbolic AI)

Track Chairs

Casini, Giovanni	Institute of Information Science and Technologies, National Research Council of Italy, Italy
Heyninck, Jesse	Open Universiteit, Netherlands

PC Members

Booth, Richard	Cardiff University, UK
Haldimann, Jonas Philipp	Vienna University of Technology, Austria
Giordano, Laura	Università del Piemonte Orientale, Italy
Leenen, Louise	University of the Western Cape, South Africa
Meyer, Thomas	University of Cape Town, South Africa
Pino Perez, Ramon	Université d'Artois, France
Rens, Gavin	Stellenbosch University, South Africa

Table of Contents

Deep Learning for Cleaning Cultural Heritage Point Clouds

Luc Hayward[1 [0000−0003−3667−961X]], Patrick Marais[1 [0000−0001−8747−7765]], and Jan Dirk Wergner[2 [0000−0002−0290−6901]]

[1] University of Cape Town, Cape Town, South Africa
hywluc001@myuct.ac.za
[2] University of Zurich, Switzerland
jandirk.wegner@uzh.ch

Abstract. Laser scanning technology is often used in the Cultural Heritage domain to capture the 3D structure of a site, with each scan consisting of a set of 3D point coordinates — a point cloud. Before these point clouds can be utilsied to build a complete 3D surface model, unwanted points must be removed. While manual *point cloud cleaning* is time-consuming, previous work has shown promise in automating parts of the process. This study builds on a previous approach which interprets point cloud cleaning as a segmentation task accomplished via binary *point classification*, applied to individual point clouds. This approach uses a basic Random Forest (RF) classifier with hand-crafted features, is designed to clean scans one by one via incremental per scan training, and requires a complex graph-based post-processing step to achieve acceptable results. By contrast, we leverage modern *point-based* deep learning models to directly learn useful features, and develop a framework that processes the fully registered set of point clouds, rather than cleaning scans individually. Our method focuses on purely geometric attributes, uses a few-shot fine-tuning approach and, unlike the single scan method, does not require segmentation post-processing to improve results. Under this scheme, users label $2.5 - 50\%$ of an unlabelled scan, and a model is trained to label the rest. We assess three deep learning point-based models (Pointnet++, KPConv, Point Transformer) along with a baseline Random Forest model, focusing on speed, accuracy, and the reduction of *total labelling effort*. Our findings reveal that modern deep learning requires minimal human labelling, with up to 85% reduction in total labelling effort. KPConv stands out for its efficiency with less human input, while Random Forests work best for simpler scenes. This study highlights deep learning's effectiveness in reducing manual labour in point cloud cleaning in the cultural heritage domain.

Keywords: Point Clouds · Deep Learning · Cultural Heritage

1 Introduction

Cultural Heritage sites throughout the world are a source of religious, cultural, historical and archaeological importance. Increased tourism can lead to some of

Fig. 1: Comparison of point cloud (red = discard, blue = keep) and real world environment for Bagni di Nerone

these sites rapidly eroding whilst natural disasters may result in permanent damage or complete destruction. To aid in the documentation of a site or structure one can acquire a series of laser scans around the site of interest, which densely sample surfaces to yield one or more point clouds (see Figure 1). This point cloud set can be aligned (registered) and then used to generate a 3D surface model. Before the point cloud can be meshed to create a 3D model, unwanted points must be removed from each scan in a process known as *point cloud cleaning*. For simple cases such as ground detection or foliage removal simpler "semantic segmentation" algorithms can be used to partition the point cloud into groups of semantically related points. Unwanted point categories (such as foliage) can then be discarded. This labelling task is complicated in the cultural heritage setting where objects may be deemed undesirable due to modelling/aesthetic choices. This is a key problem for training a general machine learning model as each scene may differ drastically in its modelling and cleaning requirements [22].

Previous work in the segmentation of point clouds has focused on single scans and used multi-class labels covering different semantic groups such as trees, ground, and walls. An alternative approach [9], the basis for this expanded study, changed the cleaning problem to one of *binary point classification*. Under this approach, every point is labelled as either "keep" or "discard" with the latter corresponding to points to be removed during cleaning. This approach is based on non-deep learning methods, uses handcrafted features, and operates on individual scans rather than fully registered point clouds [11, 9, 15]. Although the approach works surprisingly well, it does not utilise a fully registered set of scans or attempt to assess the utility of point-based deep learning (DL) models for the cleaning problem. It also requires post-processing steps to increase the quality of the segmentation. A significant issue in the use of machine learning for this task is the potential for highly imbalanced datasets. Class label disparity can be as high as 99 : 1, despite the change to a binary labelling scheme. Bias can occur towards either "keep" or "discard" labels and is scene dependent.

Modern DL models have shown promise in the semantic segmentation of large benchmark datasets, without the need for human-engineered features [27, 28]. The principal aim of this work is to investigate the usefulness of such state-of-the-art point-only deep learning models under this binary segmentation point cloud cleaning approach.

This work logically continues from the single-scan approach of Marais et al. [9] to fully registered point clouds and removes the requirement for explicit feature engineering as well as segmentation post-processing, which was integral to their better results.

More specifically, we investigate several open-source deep learning models: Pointnet++, KPConv, and Point Transformer. We also examine a baseline Random Forest (RF) to allow for easier comparison with the original study. Each of these models (where applicable) is tested on our datasets both by training from scratch as well as fine-tuning after pre-training on the Stanford 3D indoor benchmark (S3DIS). This paper explores the performance of these models across a range of training set sizes and scenes, as well as methods to minimise the amount of human labelling needed to train the model for a given scene. To evaluate the performance of the models we make use of the mean Intersection over Union (mIoU).

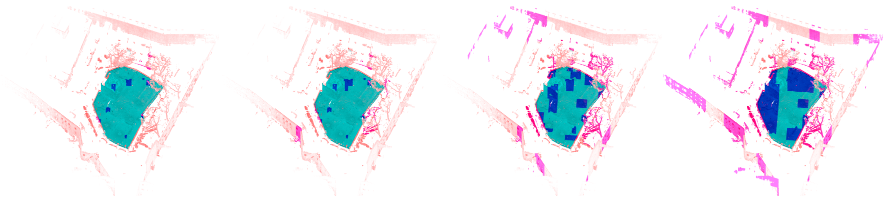
Scope and Limitations: This work focuses on methods for reducing the amount of time a human needs to spend labelling during the cleaning process. We chose to focus on more established models in Pointnet++ and KPConv, in addition to the Random Forest baseline. The Point Transformer model was added since, at the time of testing, it was the best performing model with a stable implementation on GitHub. We forego a direct comparison with manual labelling as it would require extensive user studies and access to experts.

2 Background and Related Work

Point cloud semantic segmentation can be accomplished either with or without recourse to machine learning (ML). Non-ML techniques include model fitting (e.g. RANSAC [4] for plane extraction [23]), region growing [13, 16, 26], and clustering [5]; although these may be limited in their ability to handle complex scenes [27]. Since our interest is specifically focused on machine learning based methods, we do not give further consideration to non-ML approaches. Several existing tools are available for manual segmentation using 2D polygon selection and brushes, as well as semi-automated approaches for specific semantic classes such as ground plane removal (often via RANSAC) and foliage removal, although this has been shown to mislabel heritage structures [15].

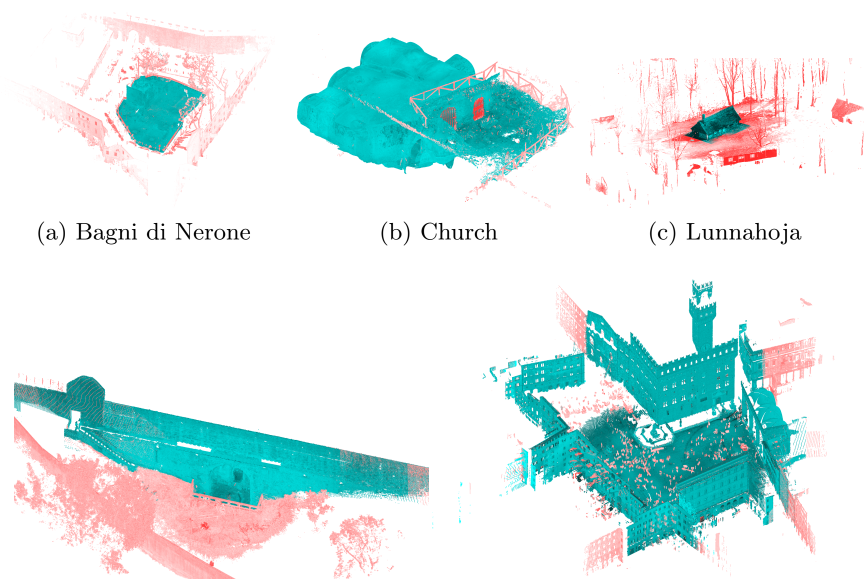
Pointnet [17] marked the start of viable DL models for the segmentation and classification of large 3D point clouds. Pointnet++ [18] improves on Pointnet by taking into account local neighbourhood information with hierarchical spatial structures, accounting for both local and global contexts. Whilst many future models use Pointnet++ as a backbone, its performance suffers compared to newer, larger models described below.

KPConv is an efficient CNN-based approach for point cloud classification and segmentation tasks [25]. At release, it achieved state-of-the-art (SOTA) results on several point segmentation and classification benchmarks, including S3DIS *Area*



(a) 2.5% (b) 5% (c) 25% (d) 50%

Fig. 2: Bagni di Nerone: Each training split is a strict subset of the next. Dark blue/pink indicates the training set.



(a) Bagni di Nerone (b) Church (c) Lunnahoja

(d) Montelupo (e) Piazza

Fig. 3: Overview of each scene (red is discard, blue is keep)

of 2cm to reduce the effect of noisy or non-uniform point densities inspired by KPConv. The labels in the subsampled point cloud are assigned to the majority label from the nearest neighbours in the original cloud.

Table 1: Number of points and percentage "discard" labels for each point cloud as well as reprojection accuracy. The central columns representing the number of points and the percentage labelled for discard each contain two measurements seperated by a "/". These are the measurements on the subsampled and the original point clouds.

Scene	Subsampled / Original		Reprojection Accuracy	Area m^2
	Num. Points	Discard (%)		
Bagni di Nerone	5.4 / 55.4M	2.4 / 6.8M (44.8% / 12.3%)	99.5%	19k
Church	1.0 / 21.9M	0.05 / 0.3M (4.6% / 1.4%)	100%	4.7k
Lunnahoja	2.2 / 30.7M	1.6 / 12.4M (70.5% / 40.6%)	96.9%	11k
Montelupo	3.1 / 115.4M	1.9 / 52.6M (62.1% / 45.6%)	99.6%	2.9k
Monument	2.8 / 49.4M	2.7 / 40.9M (97.4% / 82.7%)	99.9%	3.5k
Piazza della Signoria	16.2 / 55.1M	0.7 / 1.4M (4.2% / 2.5%)	99.8%	52k

We determined empirically in Table 1 that the subsampled point clouds can be reprojected onto the original point clouds using nearest neighbour reprojection with at least 98% accuracy across all datasets. Second, using a $1m^2$ grid in the XY plane, any cells containing fewer than 100 isolated points are removed. This removes less than 0.1% of points in all scenes without incorrectly discarding any points.

For each scene, we select 5%, 10%, 25%, and 50% train-test splits (with each training set being a strict subset of the next, see Figure 2) using the column grid method described in Section 4.1. These splits are determined by area, rather than point density, to better approximate labelling effort. The reasoning behind these splits, rather than a larger, 75%/25% training / test split for instance, is that we are aiming to emulate the human labelling process. Given the aim of reducing human labelling effort we chose to use these smaller training set sizes.

In several scenes, there are inconsistent labellings in the ground truth labels which likely negatively impacted the model training. Figure 5a illustrates an example of this in which regions of the ground points (centre of the figure) are incorrectly labelled "discard" when they should be marked as "keep".

The Stanford 3D Indoor Scene (S3DIS) is a large-scale indoor dataset which is commonly used for benchmarking models. We use it as a pretraining dataset as the sampling density, object sizes, and data format are similar to our point clouds. Both SemanticKITTI [1] and Paris Lille [21] were considered as alternative pretraining datasets. However, the car-mounted LiDAR scans produce different point density distributions — in the form of repeating circular patterns — compared to the TLS-based heritage scans. Following the completion of this study, and so not used as part of our experiments, the authors were made aware of the ArCH dataset [14]. This is a collection of cultural heritage buildings scanned using terrestrial laser scanning, UAVs, and photogrammetry methods. Unlike S3DIS which was evaluated in the original papers of our three deep learning models, to the best of our knowledge ArCH does not have published results on either KPConv or Point Transformer.

4 Experimental Design and Implementation

Here we detail the implementation of the models used and explain issues related to data preparation/pre-processing and experimental setup.

4.1 Pre-processing and data prepararation

Point Features: Although several scenes contain additional RGB and Intensity features, we use only the XYZ coordinates. RGB data is unreliable in heritage environments due to changing lighting conditions [9]. Furthermore, initial testing with both pretrained randomly initialised Pointnet++ models indicated that the laser return intensity did not reliably improve mIoU scores ($\overline{x} = 0, \sigma = 0.17$)).

Dataset tiling: Drawing inspiration from PointPillars [8], the scenes are split into $1 \times 1m^2$ columns in the XY plane. This size preserves geometric detail whilst keeping points per sample small. To maintain consistency between the models, we precompute the splits when creating the different training set sizes such that there is no overlap with the test data (see Section 3). We use this sampling strategy for Pointnet++ and retain the original spherical sampling strategy for both KPConv and Point Transformer.

Data Augmentation: The default augmentations of Z-rotation, scaling, and random perturbation are used.

4.2 Exploratory experiments

It is important to reiterate that model development and hyperparameter tuning are performed exclusively on the Church scene (cf. Table 2).

Pointnet++ was the main model explored during the exploratory study, although training was erratic and struggled to converge on the imbalanced class labels. We tuned Pointnet++'s hyperparameters using a grid search over several hyperparameters: learning rate, weight decay, number of points, whether to augment the points, and use of intensity or xyz features. The largest improvement came from increasing the weight decay parameter in the Adam optimizer [6] from

Table 2: Descriptions of datasets used for experiments

Scene - Description	**Discarded data and labelling inconsistencies**
Bagni di Nerone Ancient Roman bath site	Structures around the area of interest, vegetation, railings, cars, trees, and people, noisy ground labels and parts of railing, tree foliage and people
Church Underground Church & stepped courtyard	Railings, gates, interior scaffolding, gate frames and ground labels
Lunnahoja Wood and stone cabin in the woods	Trees and surrounding buildings, points inside the main building, ground labels around the perimeter of the main building are noisy
Montelupo Church site with clutter and foliage, including a deep alcove with irregular geometry	Vegetation around the area of interest and unintuitive labelling of people and wall in one area
Piazza della Signoria Busy Piazza in Florence	People and vehicles, scattered data due to window glass, buildings at the periphery of the scene

0.0001 to 0.01, at the cost of slower convergence. We found increased weight decay to be particularly useful for all models in smoothing the learning.

KPConv's default hyperparameters produced immediately useable results. Based on suggestions by the author , we use class balanced sampling to reduce the effect of the imbalanced scenes, the non-deformable network for faster training, and a smaller initial radius to better pick up fine detail structures such as gates and scaffolding. Whilst class balanced sampling was not used on Pointnet++ nor Point Transformer — a possible avenue for future work — the relative performance of the models is broadly consistent across the scenes with both imbalanced and balanced class labels.

Point Transformer did not have an original implementation available, we make use of the reimplementation from POSTECH CVLab at Pohang University. Experimentation with this model is constrained by the need for a GPU with at least 12GB of VRAM in order to train effectively on our dataset. We use a Bayesian hyperparameter optimization via Weights and Biases Sweeps to tune the hyperparameters against the test loss on the Church dataset. These following parameters are tuned: the learning rate, power, scheduler type, max points per sample, warmup length, weight decay, and whether to freeze the network body during training. Unlike the previous two models which are relatively stable, the Point Transformer is highly sensitive to the chosen hyperparameters with different training set sizes requiring different hyperparameters.

Random Forest and XGBoost utilise Scikit-Learn's RF and the default XGBoost implementation enabling a train-predict loop of between ten and one hundred seconds depending on the scene and amount of training data.

4.3 Experimental Design

Our preferred metric is mIoU (rather than accuracy) as it provides a more informative measure for our imbalanced binary labelling task. Accuracy is only used to ease comparisons with our baseline from Marais et. al. [9]. To quantify the utility of the model's predictions, we introduce a secondary metric Effective mIoU: the effective performance over the entire scene taking into account the "perfect" labelling of the training set. We use the formula $Predicted\ mIoU * (1 - labelling\%) + labelling\%$ where $labelling\%$ is the percentage of the scene, by area, labelled by the user.

Early experiments showed that it was infeasible to create a third representative split for validation or to randomly select samples for the training set: scenes are typically focused around a central point of interest and often have imbalanced label and geometry distributions. We select the splits by hand into training and test splits, ensuring the labels are not visible during selection. This aims to mimic a real-world interaction where the labeller is asked to select "representative and informative" regions for labelling. It is important to note that we only tune the models on the Church scene. The remaining scenes are tested blind to not unduly bias the results.

5 Results and Discussion

We focus our discussion on two aspects of model performance. Firstly, the *Total Labelling Effort* needed to fully label the scene is derived from the mIoU: $100\% - \text{mIoU} \times \text{prediction labels } \%$. Intuitively this is the proportion of the scene which must be labelled for fine-tuning plus the amount of corrections which must still be done on the model's outputs. We also compare the running time and accuracy of our results to the baseline, acknowledging that for imbalanced datasets accuracy may be difficult to compare. Improvements in the runtime of the RF models are due to the change in implementation rather than the use of faster hardware than the baseline. Both of these are important metrics for understanding the real world applicability of the models.

It is important to reiterate that Marais et. al. [9] examined single scans which exhibit different point densities and geometric representations. Whilst this makes for a less direct comparison, this study also aims to determine whether the shift to a registered point cloud can directly improve results.

5.1 Total Labelling Effort

A key focus of this study was reducing the human labelling effort. Figure 4 shows the median total labelling effort of each of the models across each of our scenes. , with the core summary results available in Appendix A.

Our experiments demonstrate that a 5% initial labelling is, on average, the most efficient approach, with the lowest mean total labelling effort of 25.85% achieved by the 5% Randomly Initialised KPConv model. This is followed closely by the Pretrained Point Transformer at 29.86%.

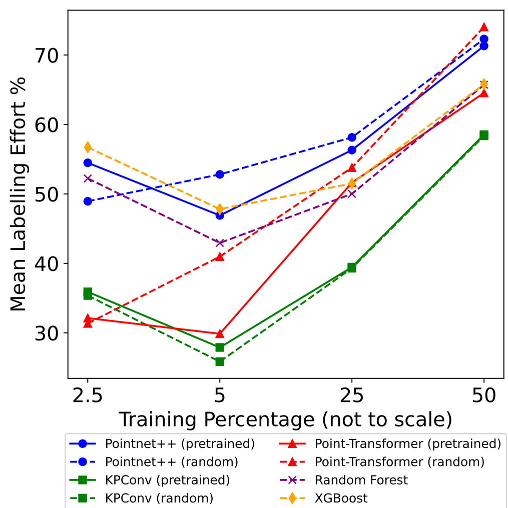
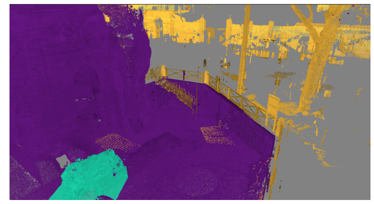
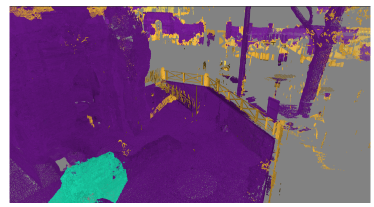


Fig. 4: Mean Total Labelling Effort for each model across the scenes.



(a) 2.5% Ground Truth Back Corner Railings



(b) 2.5% Predictions Back Corner Railings

Fig. 5: Bagni di Nerone: KPConv Pretrained prediction results with 2.5% training data. Green: training keep, Purple: test/predict keep, Yellow: test/predict discard.

On Bagni Nerone for instance, KPConv achieved an effective mIoU of 93.2% with just 5% of the scene labelled, translating to a total labelling effort of 11.8%. This demonstrates the potential for these models. Across the scenes tested, KPConv trained on 5% achieves a total labelling effort of as low as 11.8% on Church with a maximum of 42% on Piazza. This demonstrates a significant reduction in labelling that must otherwise be done by a human to complete a scene.

5.2 Accuracy and Runtime

In Table 3, the accuracy and runtimes from our baseline (Marais et. al. [9]) on each scene are compared with our highest accuracy DL and RF models, under 25% training data, on each scene. For the baseline we report the average overall accuracy across all the scans in a given scene. The runtime includes the end-to-end pipeline for training and inference including data processing. Lunnahoja is not included in these results as there are no result for the baseline and we do not feel our accuracy alone meaningfully contributes.

The accuracy of the models varied significantly across different scenes and training conditions. For instance, in the Church scene, KPConv with 5% training data achieved the highest average mIoU. However, in terms of overall accuracy,

Table 3: Comparison of model accuracy and runtime for our best deep learning and tree based models (under 5% training data) vs the baseline on each of the shared scenes.

Scene	**Model**	**Overall Accuracy (%)**	**Time (minutes)**
Bagni Nerone	Marais RF Post Processed	98.1	75
	Point Transformer 2.5% Random Init.	99.5	67
	RF 5% (ours)	94.4	0.4
Church	Marais RF	99.1	14.75
	Point Transformer 5% Pretrained	99.6	63
	RF 2.5%	98.3	1.7
Montelupo	Marais Post Processed	93.0	81
	Point Transformer 2.5% Random Init.	95.7	60
	RF 5% (ours)	90.3	0.2
Piazza	Marais RF Post Processed	97.9	298
	Point Transformer 2.5% Random Init.	97.8	67
	RF 2.5% (ours)	84.8	1.2

the Point Transformer model with 5% pretrained data achieved a 99.6%. The training and inference times also varied notably. RF models were significantly faster, taking as little as two minutes for training, while deep learning models required several hours. This presents a clear trade-off between the accuracy and time efficiency of the models. Moreover, whilst we can broadly compare our results against the chosen baseline, one should be careful of placing undue trust in these results where they may seem to contradict the mIoU numbers. In deeply imbalanced scenes such as Church and Piazza, a high accuracy is not necessarily indicative of real world labelling effectiveness. Due to the way in which mIoU is calculated, a slight decrease in overall accuracy can still translate into a signficant increase in mIoU, as the average class accuracy improves.

5.3 Qualitative Results on Bagni di Nerone

Figure 5 compares the target and predicted results when fine tuned on just 2.5% of the scene. When fine tuned with 2.5% of the scene, KPConv appears to label "thin" objects as discard, and to keep the larger surfaces mostly. This works well in the main site area but leads to trees and the buildings on the periphery being poorly labelled. Additionally, we see an example of KPConv correctly handling the ground regions that are (incorrectly) labelled as "discard" in the ground truth. However, even with 50% for fine-tuning some areas are poorly handled. In particular, regions which were unable to be scanned correctly or contain many holes — such as the front face of the main building and one of the trees — are predicted incorrectly.

5.4 Trends Between models

In general, the RF and PointNet++ models achieve higher mIoU scores on simple scenes, while KPConv and Point Transformer perform better on larger and more complex scenes. Additionally, on scenes in which there were significant inconsistencies in the ground truth labelling we found KPConv — and Point Transformer to a lesser extent — to accurately correct for these in its predictions.

Pretrained vs Randomly Initialised Weights
Pretraining was generally beneficial for maximising mIoU over an unseen test set. This is particularly evident for the Point Transformer at lower training percentages on our scenes with an average improvement of 11%. However, on some less complex scenes in which the labelling decision was based largely on the position within the scene pretraining unexpectedly reduced model performance. KPConv was found to be more variable in whether or not pretraining improved performance, though on average the randomly initialised models slightly outperformed on our data.

Model efficiency can be split into time optimised, and human labelling optimised results. Unsurprisingly, pre-labelling 50% of the scene results in the least post-prediction corrections, with an average effective mIoU of 90.3%. Our analysis shows that labelling 2.5-5% of the scene as fine-tuning labels is the optimal proportion to minimize total labelling effort after correcting the resulting predictions.

The RF models are an order of magnitude faster to train and predict labels for the remainder of the scene compared to the deep learning models. On relatively simple scenes (Bagni di Nerone, Montelupo, and Monument), it is more efficient to correct the additional $5-10\%$ of labels than to wait for a deep learning model. This is particularly useful if cleaning must be done in the field without access to powerful GPU computing. XGBoost generally performed worse than our RF models in our experiments. Although XGBoost is often the "default", it can require more extensive tuning. It is possible that further tuning and additional precomputed features as in Marais et. al. [9] could improve the model's performance. Both models suffer when the geometry and/or decision boundaries are not axis-aligned. This is particularly evident on Lunnahoja in which both models perform worse than random guessing or even majority class prediction. We attribute this to the cabin's diagonal orientation to the XY axes. Furthermore, both models can often have their prediction performance improved by going up to 25% training data. The additional effort could be considered to be offset by the far faster training and inference times of these models compared to any of the three deep learning models.

Deep learning models, in contrast, are most useful when reducing the labelling effort is more important than the increased training and inference times. The Randomly initialized KPConv model consistently requires the least total labelling effort. During training, KPConv quickly achieves a reasonable level of performance before taking a long time to reach the maximum mIoU for an experiment.

Point Transformer, despite often requiring the least labelling effort, is highly inconsistent in the amount of training data needed and the effect pertaining will have. Additionally, we observed the Point Transformer model to occasionally produce a low total labelling effort, but with a far lower than expected average class accuracy compared to the corresponding KPConv model with a similarly low total labelling effort. Conversely, KPConv, particularly when randomly initialized and trained on 5% of the scene, produces consistently good results with a maximum total labelling effort of 43%, averaging at $25 \pm 11\%$. KPConv is consistently within the top two models tested. The trade-off for this consistency is longer training times, notably three to four hours compared to Point Transformer's one hour, while RF models train in under two minutes.

Finally, our baseline deep learning model, Pointnet++, is not particularly fast to train or particularly effective. It is often unable to learn anything meaningful in more complex scenes, whilst on simple scenes the more traditional RF is within a few percentage points for total labelling effort and is far faster to train.

Recent Advances, such as PointNext [19] and Stratified Transformer [7], have been released since the conclusion of our experiments which improve upon the Point Transformer S3DIS benchmark. We believe that the improvements of the newer models would not significantly affect the outcome of our experiments, although these could be explored in future works. The changes proposed in PointnNeXT are of particular interest as they demonstrate the majority of their performance via improved data augmentation strategies.

6 Conclusion

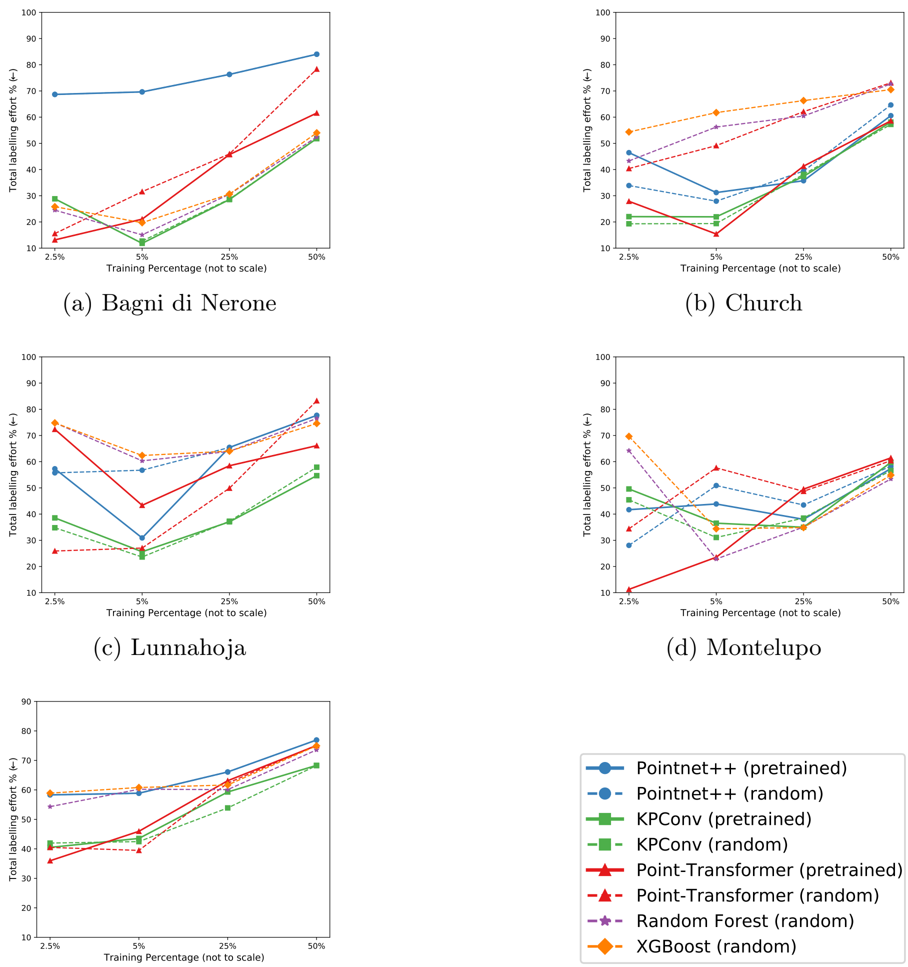
This work compares modern deep learning methods with traditional machine learning approaches in a point cloud cleaning setting, building on the binary classification approach proposed by [9]. We show the potential for deep learning models to reduce the human labelling effort by as much as 85% across a range of scene types, including highly imbalanced class distributions.

Model performance can vary significantly based on the dataset. KPConv and RF most consistently require the least labelling effort and training/prediction time respectively. RF models are efficient in simpler scenes but less capable in complex geometries. Point Transformer occasionally required the least labelling effort, however, it was too inconsistent to recommend for real-world use.

Future Work: The major limitation of the study is the lack of further scenes to evaluate the models across a broader range of environments. Potential directions for future work include leveraging RF's fast training for real-time labelling assistants or exploring a combination of whole scene input and human-engineered features [9].

SUPPLEMENTARY MATERIAL

A Total Labelling Effort Detailed Results



(a) Bagni di Nerone (b) Church

(c) Lunnahoja (d) Montelupo

(e) Piazza (f) Legend

Fig. 6: Total Labelling Effort for each tested model on each scene in our data.

References

1. Behley, J., Garbade, M., Milioto, A., Quenzel, J., Behnke, S., Stachniss, C., Gall, J.: SemanticKITTI: A dataset for semantic scene understanding of LiDAR se-

quences. Proceedings of the IEEE International Conference on Computer Vision **2019-Octob**(iii), 9296–9306 (2019). https://doi.org/10.1109/ICCV.2019.00939
2. Breiman, L.: Random forests. Machine learning **45**(1), 5–32 (2001)
3. Chen, T., Guestrin, C.: Xgboost: A scalable tree boosting system. In: Proceedings of the 22nd acm sigkdd international conference on knowledge discovery and data mining. pp. 785–794 (2016)
4. Fischler, M.A., Bolles, R.C.: Random sample consensus: a paradigm for model fitting with applications to image analysis and automated cartography. Communications of the ACM **24**(6), 381–395 (1981)
5. Grilli, E., Menna, F., Remondino, F.: A review of point clouds segmentation and classification algorithms. The International Archives of Photogrammetry, Remote Sensing and Spatial Information Sciences **42**, 339 (2017)
6. Kingma, D.P., Lei Ba, J.: ADAM: A METHOD FOR STOCHASTIC OPTIMIZATION. In: International Conference on Learning Representations (2015)
7. Lai, X., Liu, J., Jiang, L., Wang, L., Zhao, H., Liu, S., Qi, X., Jia, J.: Stratified Transformer for 3D Point Cloud Segmentation. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. pp. 8500–8509 (2022)
8. Lang, A.H., Vora, S., Caesar, H., Zhou, L., Yang, J., Beijbom, O.: PointPillars: Fast Encoders for Object Detection From Point Clouds. In: 2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). pp. 12689–12697 (jun 2019). https://doi.org/10.1109/CVPR.2019.01298, https://github.com/nutonomy/second.pytorch
9. Marais, P., Dellepiane, M., Cignoni, P., Scopigno, R., Dellepiane, .M., Cignoni, P., Scopigno, R.: Semi-automated Cleaning of Laser Scanning Campaigns with Machine Learning. ACM Journal on Computing and Cultural Heritage **12**(3), 16 (2019). https://doi.org/10.1145/3292027, https://doi.org/10.1145/3292027
10. Maturana, D., Scherer, S.: Voxnet: A 3D convolutional neural network for real-time object recognition. In: 2015 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS). pp. 922–928. IEEE (2015)
11. Mulder, R.L., Marais, P.: Accelerating point cloud cleaning. In: Catalano, C.E., Luca, L.D. (eds.) 2016 Eurographics Workshop on Graphics and Cultural Heritage, GCH 2016. pp. 211–214. The Eurographics Association (2016). https://doi.org/10.2312/gch.20161410
12. Mulder, R.: Accelerating Point Cloud Cleaning. Master's thesis, University of Cape Town (2017)
13. Nguyen, A., Le, B.: 3D point cloud segmentation: A survey. In: 2013 6th IEEE conference on robotics, automation and mechatronics (RAM). pp. 225–230. IEEE (2013)
14. Pierdicca, R., Paolanti, M., Matrone, F., Martini, M., Morbidoni, C., Malinverni, E.S., Frontoni, E., Lingua, A.M.: Point Cloud Semantic Segmentation Using a Deep Learning Framework for Cultural Heritage pp. 1–23 (2020). https://doi.org/10.3390/rs12061005
15. Pocock, C.: 3D Scan Campaign Classification with Representative Training Scan Selection. Master's thesis, University of Cape Town (2019)
16. Pu, S., Vosselman, G., Others: Automatic extraction of building features from terrestrial laser scanning. International Archives of Photogrammetry, Remote Sensing and Spatial Information Sciences **36**(5), 25–27 (2006)
17. Qi, C.R., Su, H., Mo, K., Guibas, L.J.: PointNet: Deep learning on point sets for 3D classification and segmentation. In: Proceedings - 30th IEEE Conference on Computer Vision and Pattern Recognition, CVPR 2017. vol. 2017-Janua, pp. 77–85 (2017). https://doi.org/10.1109/CVPR.2017.16

18. Qi, C.R., Yi, L., Su, H., Guibas, L.J.: PointNet++: Deep hierarchical feature learning on point sets in a metric space. Advances in Neural Information Processing Systems **2017-Decem**, 5100–5109 (2017)
19. Qian, G., Li, Y., Peng, H., Mai, J., Hammoud, H.A.A.K., Elhoseiny, M., Ghanem, B.: PointNeXt: Revisiting PointNet++ with Improved Training and Scaling Strategies pp. 1–18 (jun 2022), http://arxiv.org/abs/2206.04670
20. Riegler, G., Osman Ulusoy, A., Geiger, A.: Octnet: Learning deep 3D representations at high resolutions. In: Proceedings of the IEEE conference on computer vision and pattern recognition. pp. 3577–3586 (2017)
21. Roynard, X., Deschaud, J.E., Goulette, F.: Paris-Lille-3D: A large and high-quality ground-truth urban point cloud dataset for automatic segmentation and classification. The International Journal of Robotics Research **37**(6), 545–557 (2018). https://doi.org/10.1177/0278364918767506
22. Rüther, H., Held, C., Bhurtha, R., Schroeder, R., Wessels, S.: From Point Cloud to Textured Model, the Zamani Laser Scanning Pipeline in Heritage Documentation. South African Journal of Geomatics **1**(1), 44 (2012)
23. Schnabel, R., Wahl, R., Klein, R.: Efficient RANSAC for point-cloud shape detection. In: Computer graphics forum. vol. 26, pp. 214–226. Wiley Online Library (2007)
24. Stucker, C., Richard, A., Wegner, J.D., Schindler, K.: Supervised outlier detection in large-scale mvs point clouds for 3d city modeling applications. ISPRS Annals of the Photogrammetry, Remote Sensing and Spatial Information Sciences **IV-2**, 263–270 (2018). https://doi.org/10.5194/isprs-annals-IV-2-263-2018, https://isprs-annals.copernicus.org/articles/IV-2/263/2018/
25. Thomas, H., Qi, C.R., Deschaud, J.E., Marcotegui, B., Goulette, F., Guibas, L.J.: KPConv: Flexible and Deformable Convolution for Point Clouds. In: Proceedings of the IEEE/CVF International Conference on Computer Vision. pp. 6411–6420 (2019)
26. Vo, A.V., Truong-Hong, L., Laefer, D.F., Bertolotto, M.: Octree-based region growing for point cloud segmentation. ISPRS Journal of Photogrammetry and Remote Sensing **104**, 88–100 (2015)
27. Xie, Y., Tian, J., Zhu, X.X.: Linking Points With Labels in 3D. IEEE Geoscience and Remote Sensing Magazine **8**(March), 38–59 (2020). https://doi.org/10.1109/MGRS.2019.2937630, https://ieeexplore.ieee.org/stamp/stamp.jsp?tp=&arnumber=9028090
28. Yang, S., Hou, M., Li, S.: Three-dimensional point cloud semantic segmentation for cultural heritage: A comprehensive review. Remote. Sens. **15**, 548 (2023), https://api.semanticscholar.org/CorpusID:256183521
29. Zhao, H., Jiang, L., Jia, J., Torr, P.H.S., Koltun, V.: Point transformer. In: Proceedings of the IEEE/CVF International Conference on Computer Vision. pp. 16259–16268 (2021)