MultiChange3D: A Multi-Scene, Multi-Sensor Dataset for Benchmarking 3D Geometric Change Detection

Zhaoyi Wang[1], Paweł Trybała[2], Andreas Wieser[1], Fabio Remondino[2]

[1] Chair of Geosensors and Engineering Geodesy (GSEG), ETH Zurich, Zurich, Switzerland
Email: <zhaoyi.wang><andreas.wieser>@geod.baug.ethz.ch
[2] 3D Optical Metrology (3DOM) Unit, Bruno Kessler Foundation (FBK), Trento, Italy
Email: <ptrybala><remondino>@fbk.eu

Keywords: 3D change detection, Point cloud, Deep learning, Multi-scene evaluation, Generalization

Abstract

3D change detection is essential for monitoring infrastructure, environmental dynamics, and natural hazards. However, existing algorithms are often evaluated on single-scene datasets, and their generalization across varied real-world scenes remains largely unexplored due to the absence of a universal benchmark. To address this issue, we propose **MultiChange3D**, a multi-scene, multi-sensor 3D change detection dataset for identifying geometric changes in 3D space. The dataset provides registered pairs of point clouds with ground-truth geometric change labels, enabling standardized evaluation across different methods. To demonstrate the use of the **MultiChange3D** dataset, we benchmark an initial set of approaches on a subset of the dataset. The evaluated methods include classical Euclidean distance-based methods (C2C, M3C2), 3D displacement estimation-based approaches (F2S3, Landslide-3D), and deep learning-based classification methods (KPConv, EF-KPConv, PGN3DCD). Quantitative and qualitative analyses indicate the strengths and limitations of the evaluated methods, highlighting the challenges in cross-scene generalization under variations in point density, scale, and types of changes. The full dataset and evaluation code is openly available at: `https://github.com/3DOM-FBK/multichange3d`.

1. Introduction

Change detection (CD) is the process of identifying differences in the state of objects or phenomena by comparing observations acquired at different time epochs (Singh, 1989). Three-dimensional change detection (3DCD), as a specific branch of CD, focuses on detecting such differences using 3D data representations, such as point clouds, meshes, digital elevation models (DEMs), or other forms of 3D data (Qin et al., 2016). 3DCD plays a crucial role in various applications, including indoor motion tracking (Fehr et al., 2017; Nikoohemat et al., 2018), construction progress recording (Meyer et al., 2022; Huang et al., 2022), urban change identification (Chaabouni-Chouayakh et al., 2010; Stilla and Xu, 2023), and natural disaster monitoring (Peppa et al., 2019; Mazzanti et al., 2021).

With the rapid development of active and passive 3D sensing techniques, various types of 3D data have become easily accessible. Typical examples include LiDAR-based point clouds from airborne (ALS), mobile (MLS), and terrestrial laser scanning (TLS), photogrammetric point clouds reconstructed from images acquired from an airplane or an unmanned aerial vehicle (UAV), and RGB-D point clouds (Vosselman and Maas, 2010; Shan and Toth, 2018; Liu et al., 2020; Szeliski, 2022). Due to differences in sensing principles, the resulting 3D data vary significantly in density and noise characteristics and are often collected under highly diverse scene conditions (*e.g.*, indoor environments, urban streets, or natural slopes). However, existing 3DCD datasets mostly focus on a single scene (*e.g.*, indoor or urban), limiting the ability to assess how algorithms generalize across data captured by different sensors and scenes. Consequently, numerous 3DCD methods have been developed to handle specific data domains. Classical Euclidean distance-based approaches such as Cloud-to-Cloud (C2C) (Girardeau-Montaut et al., 2005) and Multiscale Model-to-Model Cloud Comparison (M3C2) (Lague et al., 2013) directly compute spatial distances between corresponding points in two point clouds from different epochs. Handcrafted feature-based (Tran et al., 2018) methods perform CD and classification using manually designed features, *e.g.*, the point distribution and terrain height, while 3D displacement estimation-based methods (Gojcic et al., 2021; Wang et al., 2026) estimate 3D displacements between two epochs derived from pointwise correspondences established using pretrained feature descriptors. More recently, deep learning-based classification methods (Wang et al., 2023; de Gélis et al., 2023; de Gélis et al., 2024; Zhan et al., 2024) have emerged, which learn to automatically predict and classify 3D geometric changes (*e.g.*, removed and added subsets of points) from annotated datasets, showing promising performance when trained and tested on the same type of scenes.

Despite the remarkable progress, most existing methods are evaluated only within a single data domain. This raises an open question: How well can current 3DCD algorithms generalize across different types of 3D data and scenes? The issue is particularly critical for the deep learning-based classification methods, whose generalization to unseen scenes without retraining (zero-shot) remains largely unexplored. Although retraining or fine-tuning of the models with task-specific data is possible, this process often requires costly and time-consuming manual annotations. Furthermore, real-world 3DCD applications exhibit substantial variation in scale (*e.g.*, indoor furniture vs. urban buildings vs. terrain morphology) and types of changes (*e.g.*, object movement vs. deformation), with data differing in density, occlusions, coverage, and noise distribution. However, to the best of our knowledge, no prior work has systematically evaluated the generalizability of 3DCD methods across multiple data sources and scene types, and no dataset currently exists to support such evaluation.

To fill these gaps, we present **MultiChange3D**, a multi-scene, multi-sensor 3DCD dataset that includes diverse environments

This contribution has been peer-reviewed. The double-blind peer-review was conducted on the basis of the full paper.
https://doi.org/10.5194/isprs-annals-XI-2-2026-639-2026

captured by both active and passive 3D sensors. The dataset provides manually annotated geometric change labels, covering indoor, urban, and natural scenes, and enables cross-scene and cross-sensor evaluation of 3DCD algorithms. Based on this dataset, we provide an initial benchmark of a selection of 3DCD methods, including those based on Euclidean distance, 3D displacement estimation, and deep learning classification. Our experiments provide the first analysis of cross-domain generalization in 3DCD, offering insights into the feasibility of employing the examined algorithms in various real-world applications.

2. Related works and datasets

Several publicly available datasets have been proposed to support research on 3DCD, mainly covering indoor, street, and urban scenes, as summarized in Table 1. At the indoor scale, the ICRA2017-CD (Fehr et al., 2017) dataset captures small indoor environments and focuses on furniture changes in three different rooms. At the street scale, the Business District (BD) dataset (Yew and Hee Lee, 2021) contains two street scenes acquired two months apart, covering roughly 0.3 km^2. It primarily reflects construction and demolition activities, as well as other localized changes such as tree removal and vehicle movement. Similarly, the SZTAKI-Change3D (Nagy et al., 2021) and SZTAKIBudapest (Orkény et al., 2022) datasets provide LiDAR point cloud pairs captured by a car-mounted RIEGL sensor in downtown Budapest, Hungary, where typical changes involve pedestrians and vehicles. The SLPCCD (Wang et al., 2023) dataset also focuses on street-level 3DCD and defines three change categories based on small-scale samples.

Dataset	Scene	RGB	Intensity	Sensor
ICRA2017-CD	Indoor	✓	✗	RGB-D
BD-CD	Street	✓	✗	Camera (ground)
SZTAKI-Change3D	Street	✗	✓	MLS
SZTAKIBudapest	Street	✗	✓	MLS
SLPCCD	Street	✓	✗	MLS
AHN-CD	Urban	✓	✗	ALS
3DCD	Urban	✓	✗	Camera (aircraft), ALS
Urb3DCD-V2	Urban	✗	✗	Synthetic
HKCD	Urban	✓	✗	Camera (aircraft)
MultiChange3D (Ours)	Multiple	✓	✓	Multiple

Table 1. Summary of existing 3DCD datasets in comparison to our proposed **MultiChange3D** dataset.

At the urban scale, the Actueel Hoogtebestand Nederland (AHN) dataset (Sande et al., 2010), recorded using airborne LiDAR in the Netherlands, mainly documents changes in building and water infrastructure. Coletta et al. (2022) propose a dataset that combines optical images with elevation maps, enabling joint CD on both 2D and 3D domains. The Urb3DCD (de Gélis et al., 2021) dataset originally includes three classes, with synthetic data; it was later extended to Urb3DCD-V2 (de Gélis et al., 2023), featuring seven change categories and serving as a more comprehensive urban benchmark. More recently, the HKCD (Zhan et al., 2024) dataset provides photogrammetric point clouds covering approximately 8.1 km^2 of the Hong Kong urban landscape.

Despite these efforts, existing 3DCD datasets remain limited in scene diversity and environmental complexity. They are generally constructed around a single domain (*e.g.*, urban or street) and contain little data from natural outdoor environments or long-term environmental changes such as vegetation growth or geohazards (e.g., landslides). To overcome these limitations, we introduce a diverse, cross-sensor benchmark dataset designed to promote more robust, comprehensive, and generalizable 3DCD research across various sensors and a wide range of real-world scenes.

3. The MultiChange3D Dataset

3.1 Data characteristics

The proposed **MultiChange3D** dataset contains 3D data obtained with 5 different sensors, selected as commonly used devices for each specific scene. The scenes represent environments such as indoor areas, outdoor natural environments, as well as city-scale 3D reconstructions. The content of the dataset is summarized in Table 2.

In total, we provide surveys of 10 scenes of various scales, repeated from 2 to 4 times during periods of natural or induced changes. The spatial extent of the scenes ranges from approximately [3, 3, 1] m to [1029, 851, 291] m, with point cloud resolution spanning from below 1 cm to 4 cm. **MultiChange3D** contains the following subsets, grouped by the survey instrument and the scene.

The RGB-D subset contains 2 small-scale scenes, collected in an office building at the desk (*Office*) and in the open space area (*Open space (RGB-D)*). Changes in these scenes focus on induced changes of pose, addition or removal of small objects (*e.g.*, a bottle, a laptop, and furniture). The data was collected with an Intel RealSense D435i, an RGB-D camera with a resolution of 1280×720 pixels. The colored depth images were processed with BAD-SLAM (Schops et al., 2019), obtaining camera poses and a dense scene representation in the form of surfels, which were then exported as a dense point cloud.

The MLS subset contains 4 scenes, including 2 indoor scenes: *Open space (MLS)*, *Underground car parking*, and 2 outdoor scenes: *Bike parking construction* and *Vineyard*. All data was collected with a Mandeye scanner (Bedkowski, 2024), internally using a Livox Mid-360 LiDAR, featuring a Field-of-View (FoV) of 360° x 59° (h x v) and an internal 200 Hz IMU. Then, the processing was performed with PoliMap SLAM (Trybała et al., 2023) to obtain a point cloud for each epoch. In the *Open*

Sensor type	Scene	#Approx. pts.	Avg. pts. density (m)	#Epoch	#Pair	Extra feature	Scene condition
RGB-D	Office	2 M	2×10^{-3}	4	6	RGB	Indoor, cluttered
	Open space (RGB-D)	4 M	2×10^{-3}	4	6	RGB	Indoor, furniture changes
MLS	Open space (MLS)	200 k	1×10^{-2}	2	1	Intensity	Indoor, furniture changes
	Underground car parking	24 M	1×10^{-2}	3	3	Intensity	Indoor, vehicle motion
	Bike parking construction	5 M	2×10^{-2}	4	6	Intensity	Outdoor, construction, vehicle motion
	Vineyard*	5 M	2×10^{-2}	3	3	–	Outdoor, vegetation
TLS	Classroom	40 M	5×10^{-3}	2	1	Intensity and RGB	Indoor, furniture changes
	Meeting room	170 M	3×10^{-3}	2	1	Intensity and RGB	Indoor, small-scale
UAV Camera	Landslide†	20 M	4×10^{-2}	4	4	RGB	Outdoor, natural terrain
Airborne Camera	City	800 M	5×10^{-2}	2	1	RGB	Simulated changes, outdoor, large-scale urban
Airborne LiDAR	City	350 M	1×10^{-1}	2	1	Intensity and RGB	Outdoor, large-scale urban

* No ground-truth change labels; † Data from Galve et al. (2025).

Table 2. Summary of the proposed **MultiChange3D** dataset. Extra features denote information beyond 3D geometry.

This contribution has been peer-reviewed. The double-blind peer-review was conducted on the basis of the full paper.
https://doi.org/10.5194/isprs-annals-XI-2-2026-639-2026 | © Author(s) 2026. CC BY 4.0 License.

space (MLS), all the changes are induced by manually moving the objects, as the subsequent surveys were conducted in a short time span. For all other sequences, the changes are organic and typical to the specific site. *Underground car parking* contains three epochs, with the first interval of around a year, and the second of a few hours, in turn demonstrating both high and limited degree of changes of the structure and the parking space occupancy. *Bike parking construction* focuses on the structure construction progress, as well as changes in surrounding vegetation and parked bikes. Finally, the vineyard sequences were acquired throughout the year in a vineyard in Trento, Italy, with a LiDAR-based MLS system described above. The survey focuses on a set of approximately 100 m grapevine lanes. As the main changes in this dataset consist of the natural growth of plants and fruits, it is hard to provide objective ground truth labels, considering the noise level of a consumer-grade LiDAR sensor. Nevertheless, we share these data for the research community, to qualitatively test the performance of various 3DCD methods in a natural environment or perform other types of growth-related analyses.

The TLS subset contains 2 scenes collected in rooms (*Classroom* and *Meeting room*). All changes in these scenes focus on the short-term scene evolution, mainly consisting of additions or removals of various objects. The TLS data was collected with Leica RTC360 scanners, with single station point clouds co-registered to a coherent, local reference system in Leica Cyclone REGISTER 360.

The UAV Camera category contains data captured from UAV platforms and represents a landslide scene, tracking the progress of a slope failure that occurred in Southern Spain in March 2021 (Galve et al., 2025). Changes mainly focus on the landslide surface collapsing, with minor additions or removals of vehicles and vegetation. The data was collected with a DJI Phantom-4 drone equipped with Real Time Kinematics (RTK) system and a 20-megapixel camera, at a flight height of approximately 100 m. The point clouds are generated with photogrammetric image matching.

The Aerial Camera category contains data captured with a Vexcel Ultracam Osprey Prime oblique aerial camera over an urban area of Graz (Austria) of approximately 1 km^2. Changes include simulated building growth, additions or removals of buildings, and vegetation changes. The point clouds were generated using photogrammetric image matching.

The Airborne LiDAR category contains data captured with Riegel VQ680 (courtesy of AVT Airborne Sensing) over an urban Klagenfurt (Austria) area of approximately 1 km^2 within a time span of 2 years. The main differences include new buildings and vegetation changes.

3.2 Data preprocessing and annotation

We follow an easy-to-implement workflow for preprocessing and annotating change labels for the proposed **MultiChange3D** dataset. The workflow consists of data co-registration, noise filtering, and ground-truth generation. Specifically, we transform point clouds from all epochs to the local reference frame of the initial epoch (*i.e.*, epoch 0). For the scenes captured by RGB-D and MLS sensors, we use CloudCompare (V2.13.2) for co-registration: an initial alignment based on manually selected correspondences, followed by refinement with the Iterative Closest Point (ICP) algorithm (Besl and McKay, 1992). For the ICP refinement, only the manually verified stable subset of points (*i.e.*, unchanged regions) is used. For the scenes obtained by TLS scanners, the co-registration is completed using Cyclone REGISTER 360. For the UAV scenes, the provided data are already co-registered in a global reference system (see the corresponding publications for details). Then, spurious points caused by sensor noise or occlusion are manually removed using CloudCompare in side or top views. Finally, to annotate ground-truth CD labels, only significant changes that reflect the interest are assigned a “changed" label (“removed" or “added"), which means that data gaps around the boundary, tiny growth or movement of vegetation in the outdoor scenes are considered as “unchanged". To annotate the change labels, we use CloudCompare to perform manual labeling from multiple viewpoints.

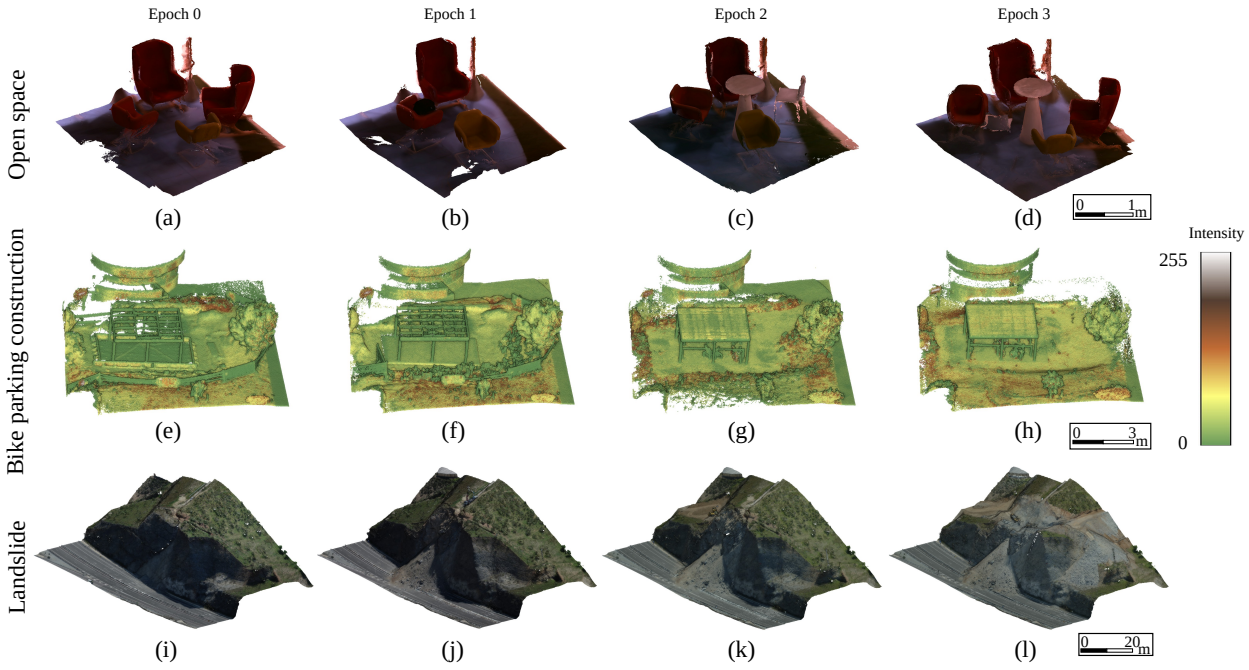

Figure 1. Visualization of three different scenes used for evaluation. The *Bike parking construction* scene is colorized by LiDAR intensity, while the other two are colorized by RGB values.

This contribution has been peer-reviewed. The double-blind peer-review was conducted on the basis of the full paper.
https://doi.org/10.5194/isprs-annals-XI-2-2026-639-2026 | © Author(s) 2026. CC BY 4.0 License.

All changes are selected by visual inspection and then cross-checked by different researchers.

4. Experiment setup

4.1 Data used for evaluation

From the proposed **MultiChange3D** dataset, we select three scenes for an initial performance evaluation of different state-of-the-art 3DCD methods (Figure 1). The representative set of scenes is selected with the variety of sensors and scene scales in mind: *Open space (RGB-D)*, *Bike parking construction* and *Landslide*. For the *Open space (RGB-D)* and *Bike parking construction*, we use the epoch pairs 1-2 and 1-3 for training, pair 2-3 for validation, and pairs 0-1, 0-2, and 0-3 for testing. For the *Landslide*, we use the epoch pairs 0-2 and 0-3 for training, pair 2-3 for validation, and pair 0-1 for testing. The splits differ across scenes due to the varying number of available epoch pairs, ensuring that models are trained on certain geometric changes while retaining unseen epochs for testing. Training and validation sets are used only for deep learning-based classification methods, while testing sets are used for all evaluated methods.

4.2 Methods selected for comparison

To assess 3DCD performance across different scenes, we select seven representative methods from three main categories for comparison. The selection is based on a review of the literature, aiming to include a diverse set of approaches ranging from simple and widely used to complex and high-performance methods, including:

Euclidean distance-based methods: C2C (Girardeau-Montaut et al., 2005) and M3C2 (Lague et al., 2013) are selected as representatives of this category, due to their simplicity and broad availability in common point cloud processing software such as CloudCompare. For each point in one point cloud, C2C computes the shortest Euclidean distance to its nearest neighbor in the other point cloud. M3C2 introduces a normal-based strategy: for each core point, it defines a cylindrical neighborhood, estimates the local surface normal, and measures the distance between the two point clouds along the normal direction.

3D displacement estimation-based methods: F2S3 (Gojcic et al., 2021) and Landslide-3D (Wang et al., 2026) are selected as representatives of this category. Specifically, F2S3 employs a feature extraction model (Poiesi and Boscaini, 2021) pretrained on an indoor dataset (Zeng et al., 2017) to extract feature descriptors of the points in point clouds. Point-to-point matching is implemented using nearest neighbor search in learned feature space, followed by single-level segmentation and binary classification as inliers or outliers. Wang et al. (2026) employ pretrained 3D local patch-based and 2D image matching models to extract features for constructing patch-to-patch matches, followed by outlier removal of low-quality patch matches and point match establishment within the remaining patch matches. Since RGB images are unavailable, we use the geometry-only version of this approach (*i.e.*, Landslide-3D) for comparison.

Deep learning-based classification methods: This category of methods directly predicts different types of change classes. For instance, Siamese KPConv (SKPConv) (de Gélis et al., 2023), Encoder Fusion Siamese KPConv (EF-SKPConv) (de Gélis et al., 2024), and PGN3DCD (Zhan et al., 2024) are representative of this category. SKPConv (de Gélis et al., 2023) employs two KPConv (Thomas et al., 2019) encoders with a weight-sharing Siamese architecture (Caye Daudt et al., 2018) to extract features from two input point clouds. These features are then fused and decoded by KPConv and passed through a fully connected layer to predict multiple change classes. Several variants of SKPConv (de Gélis et al., 2024), *e.g.*, EF-SKPConv, have been proposed to enhance 3DCD performance for certain scenes by incorporating handcrafted features or modified network structures. PGN3DCD (Zhan et al., 2024) further incorporates geometric and textural priors into the geometry-only approach, EF-KPConv (de Gélis et al., 2024), for large-scale urban 3DCD tasks. In general, these methods directly predict the classes of unchanged and changed regions (which may include several subclasses).

Different 3DCD algorithms may produce different types of outputs. For instance, Euclidean distance-based and 3D displacement estimation-based methods estimate pointwise distances rather than categorical change labels, whereas deep learning–based classification methods predict various semantic change classes. Thus, to enable consistent comparison, we first determine which points have changed and which remain unchanged. Changed points are then categorized based on their presence: points present only in the first epoch are labeled “removed”, and points present only in the second epoch are labeled “added”. Specifically, for deep learning–based classification methods, all classes other than “unchanged” are grouped into the categories of “removed” and “added”. For Euclidean distance- and 3D displacement estimation-based methods, a distance threshold proportional to the point spacing is applied to assign each point to one of these three categories: points whose nearest-neighbor distance across epochs fall within the threshold are labeled as "unchanged", while points exceeding the threshold are labeled as “removed” or “added”, depending on whether they exist in the first or second point cloud, respectively. Additionally, since 3D displacement estimation-based methods only provide valid estimates for a subset of points, to ensure full-resolution outputs for comparison, we further propagate change labels to all remaining points using nearest-neighbor-based assignment.

4.3 Evaluation metrics

Following prior works (de Gélis et al., 2023; Zhan et al., 2024; Kharroubi et al., 2025), we evaluate the performance of the 3DCD methods using several commonly adopted metrics derived from the binary confusion matrix. These metrics include Precision, Recall, F1-score, Overall Accuracy (OA), mean class Accuracy (mAcc), and mean Intersection over Union (mIoU).

4.4 Implementation details

The experiments are conducted on a Linux workstation equipped with a single NVIDIA RTX 3090 Ti GPU and an AMD Ryzen 7 5800X 8-core processor. The input point cloud pairs are spatially downsampled to 0.01, 0.05, and 0.10 m for the *Open space (RGB-D)*, *Bike parking construction*, and *Landslide* scenes, respectively, mainly due to computational memory and efficiency constraints.

The key hyperparameters are empirically fine-tuned according to the characteristics of each dataset and method, and the values used are reported in Table 3. Here, σ, s, and r denote the distance threshold, grid downsampling size, and cylindrical neighborhood radius, respectively. λ_{geo} and λ_{tex} represent the thresholds for the geometric and textural priors, respectively. Except for PGN3DCD, which produces outputs for both point clouds in a single run, all other methods are executed twice to obtain bidirectional results, *e.g.*, from epoch $0 \rightarrow 1$ and $1 \rightarrow 0$. For other hyperparameters, we adopt the default settings suggested in the corresponding publications.

Method	σ (m)	s (m)	r (m)	λ_{geo}	λ_{tex}
Open space (RGB-D)					
C2C, M3C2	0.10	-	-	-	-
F2S3, Landslide-3D	0.10	0.01	-	-	-
SKPConv, EF-SKPConv	-	0.05	1	-	-
PGN3DCD	-	0.05	1	0.1	0.6
Bike parking construction					
C2C, M3C2	0.10	-	-	-	-
F2S3, Landslide-3D	0.50	0.10	-	-	-
SKPConv, EF-SKPConv	-	0.10	5	-	-
PGN3DCD	-	0.10	5	0.2	0.6
Landslide					
C2C, M3C2	0.20	-	-	-	-
F2S3, Landslide-3D	0.50	0.10	-	-	-
SKPConv, EF-SKPConv	-	0.50	15	-	-
PGN3DCD	-	0.50	15	1.0	0.6

Table 3. Hyperparameter values used in the selected methods for three different scenes.

5. Results

To evaluate the performance of the selected 3DCD approaches, we conduct the experiments in two settings: (i) deep learning-based classification methods are trained and tested on each individual scene of three selected scenes from the **MultiChange3D** dataset (see Section 4.1); and (ii) deep learning-based classification methods are trained on urban scenes (as released by the corresponding publications) and tested on the three selected scenes. Table 4 presents the quantitative results of different methods on the three scenes, where SKPConv, EF-SKPConv, and PGN3DCD are retrained on the individual scene data. Overall, deep learning-based classification methods achieve the best quantitative scores across most of the metrics, while the Euclidean distance-based methods C2C and M3C2 exhibit competitive but inferior performance in terms of F1-score, OA, mAcc, and mIoU. Qualitative results for the three scenes are shown in Figure 2, Figure 3 and Figure 4, respectively.

Table 5 reports the cross-scene generalization results on the same test sets of the three selected scenes, without scene-specific retraining (*i.e.*, SKPConv and EF-SKPConv trained on Urb3DCD-V2, PGN3DCD on HKCD). Under this evaluation setting, the Euclidean distance-based methods (C2C, M3C2) outperform the deep learning-based classification methods, whose performance drops substantially compared to their retrained versions in terms of F1-score, OA, mAcc, and mIoU.

6. Discussion

6.1 Performance of methods in epoch-wise scenes

C2C and M3C2 achieve strong performance after threshold tuning (*e.g.*, distance thresholds), particularly in detecting changes involving additions or removals of single objects. However, they generally do not outperform the deep learning-based classification methods with models trained on the same scenes. We also observe that when changed objects overlap spatially, *e.g.*, the movement of chairs in Figure 2, Euclidean distance-based methods tend to incorrectly classify these overlapping regions as “unchanged”. This is because they compute distances either within the closest neighborhood or along surface normals, without explicitly tracking object-level motion.

For F2S3 and Landslide-3D, although their quantitative scores are relatively low (see Table 4), they can still detect relatively large changes, *e.g.*, furniture movement in Figure 2 or slope failure in Figure 4. When objects or surfaces are completely added or removed, pointwise correspondences cannot be correctly established, leading to overestimated displacements. Nonetheless, with appropriate threshold tuning, these overestimated displacements can be used to effectively identify regions with large changes. On the other hand, misclassified changes mainly occur on planar structures, such as the floor for the *Open space (RGB-D)*, the concrete base for the *Bike parking construction*, or the highway surface for the *Landslide*, where geometric features are less distinctive for reliable pointwise correspondence matching. This leads to many “unchanged” points being erroneously labeled as “added” or “removed”. Incorporating auxiliary modules, *e.g.*, ground removal, could further improve performance in these cases, which requires further investigation.

Deep learning-based classification methods trained on the same scene perform well on unseen epochs (*e.g.*, epoch 0), as shown in Table 4. This is because the models learn scene-specific change characteristics and adaptively handle thresholds, demonstrating strong capability in recognizing and interpreting new epochs within the same scene.

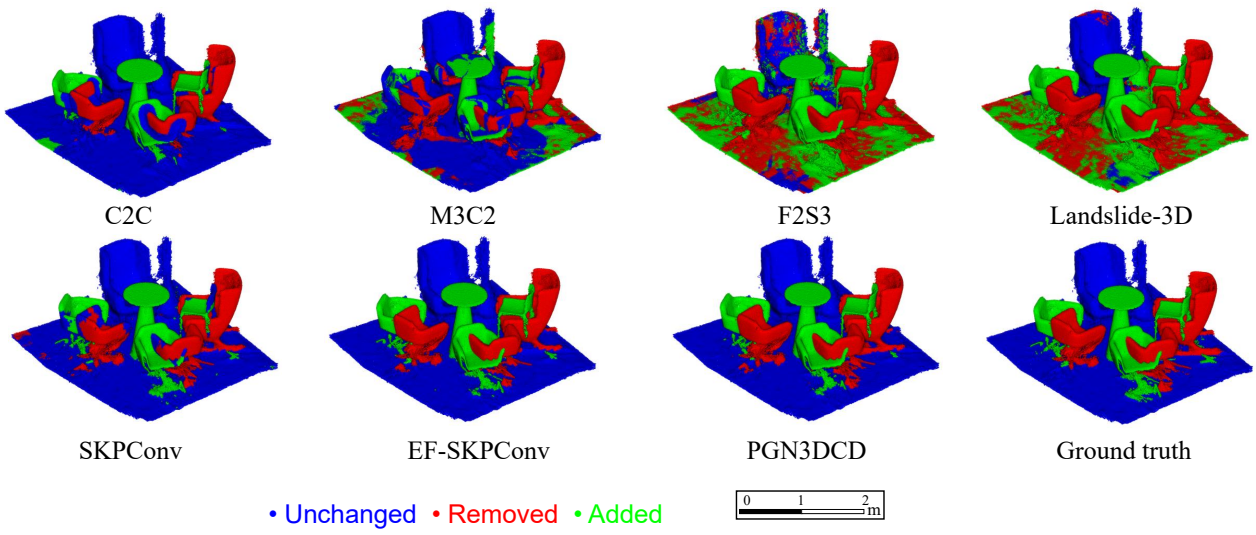

Figure 2. Qualitative results on the *Open space (RGB-D)* scene (epoch pair 0-2).

Method	Precision	Recall	F1-score	OA	mAcc	mIoU
Open space (RGB-D) (%) ↑						
C2C	95.5	62.7	75.4	88.5	80.9	73.2
M3C2	51.8	55.4	52.2	71.7	67.1	51.1
F2S3	35.8	**99.9**	51.7	46.3	61.5	29.5
Landslide-3D	40.2	99.5	55.9	54.6	67.7	38.0
SKPConv	87.5	72.5	75.6	90.2	84.0	76.1
EF-SKPConv	89.0	94.9	91.4	95.8	95.6	89.9
PGN3DCD	**97.8**	92.7	**95.0**	**97.4**	**95.8**	**93.4**
Bike parking construction (%) ↑						
C2C	51.5	50.1	49.7	78.0	68.2	54.3
M3C2	26.3	43.6	32.0	61.5	55.3	38.3
F2S3	29.0	93.3	43.8	49.1	64.8	32.0
Landslide-3D	28.0	**96.7**	43.1	46.0	64.0	29.4
SKPConv	85.8	45.9	59.2	86.3	72.0	63.4
EF-SKPConv	89.9	60.7	72.1	89.9	79.6	72.5
PGN3DCD	**94.0**	65.3	**76.8**	**91.5**	**82.1**	**76.2**
Landslide (%) ↑						
C2C	85.5	91.9	88.4	95.6	94.2	87.2
M3C2	83.8	84.6	84.2	94.5	90.6	83.1
F2S3	31.2	**98.1**	47.3	62.1	76.3	42.7
Landslide-3D	37.1	95.8	53.4	71.0	80.8	50.9
SKPConv	97.0	93.7	**95.3**	**98.4**	**96.5**	**94.5**
EF-SKPConv	**97.8**	90.6	94.0	98.0	95.1	93.2
PGN3DCD	96.1	92.4	94.1	98.0	95.8	93.3

Table 4. Quantitative results on *Open space (RGB-D)*, *Bike parking construction*, and *Landslide* scenes with SKPConv, EF-SKPConv, and PGN3DCD trained and tested on the corresponding same scenes. The best results in each column in **bold**, second-best underlined.

Method	Precision	Recall	F1-score	OA	mAcc	mIoU
Open space (RGB-D) (%) ↑						
C2C	**95.5**	62.7	**75.4**	**88.5**	**80.9**	**73.2**
M3C2	51.8	55.4	52.2	71.7	67.1	51.1
F2S3	35.8	99.9	51.7	46.3	61.5	29.5
Landslide-3D	40.2	99.5	55.9	54.6	67.7	38.0
SKPConv	30.2	**100**	45.6	30.2	50.5	15.1
EF-SKPConv	29.5	94.8	44.2	30.1	48.4	15.5
PGN3DCD	28.3	74.2	40.9	46.2	49.9	25.3
Bike parking construction (%) ↑						
C2C	**51.5**	50.1	**49.7**	**78.0**	**68.2**	**54.3**
M3C2	26.3	43.6	32.0	61.5	55.3	38.3
F2S3	29.0	93.3	43.8	49.1	64.8	32.0
Landslide-3D	28.0	96.7	43.1	46.0	64.0	29.4
SKPConv	22.0	**100**	35.5	22.0	50.1	11.1
EF-SKPConv	26.4	49.6	33.9	59.2	55.6	37.3
PGN3DCD	8.7	2.7	4.0	71.9	46.9	36.9
Landslide (%) ↑						
C2C	**85.5**	91.9	**88.4**	**95.6**	**94.2**	**87.2**
M3C2	83.8	84.6	84.2	94.5	90.6	83.1
F2S3	31.2	**98.1**	47.3	62.1	76.3	42.7
Landslide-3D	37.1	95.8	53.4	71.0	80.8	50.9
SKPConv	36.1	59.1	44.8	75.1	68.7	50.6
EF-SKPConv	19.0	55.5	28.3	51.0	52.8	31.1
PGN3DCD	49.9	96.2	65.5	81.7	87.4	63.6

Table 5. Quantitative results on *Open space (RGB-D)*, *Bike parking construction*, and *Landslide* scenes with SKPConv and EF-SKPConv trained on Urb3DCD-V2, and PGN3DCD trained on *HKCD*. The best results in each column in **bold**, second-best underlined.

6.2 Generalizability across scenes

While deep learning-based classification methods perform well when trained and tested within the same scene (single-domain setting, see Table 4), their performance degrades noticeably when applied to different scenes (cross-domain setting, see Table 5). The OA metric drops range from -16 to -65 percentage points, indicating limited and highly varied cross-scene generalization. Across the testing scenarios, precision was more consistently and severely impacted than recall. Specifically, for the deep learning-based classification methods, when the corresponding models are trained on urban datasets and directly applied to the indoor small-scale scene *Open space (RGB-D)*, and the outdoor small-scale scene *Bike parking construction*, their performance drops substantially, as shown in Table 5. This suggests that current deep learning architectures learn scene-specific characteristics and struggle to generalize to out-of-distribution scenes. For the large-scale outdoor

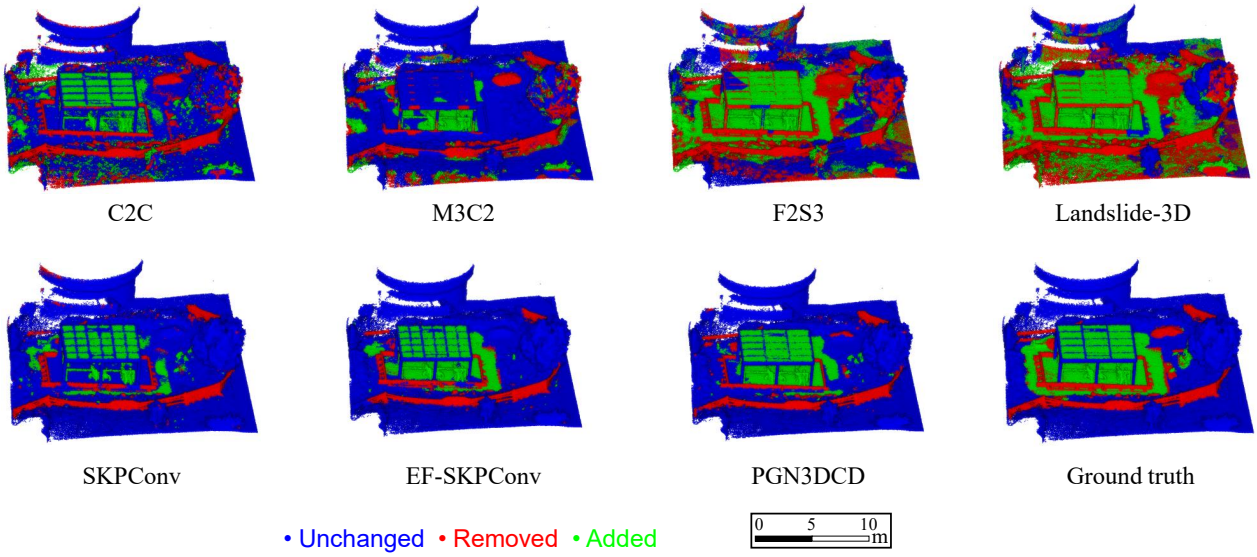

Figure 3. Qualitative results on the *Bike parking construction* scene (epoch pair 0-2).

This contribution has been peer-reviewed. The double-blind peer-review was conducted on the basis of the full paper.
https://doi.org/10.5194/isprs-annals-XI-2-2026-639-2026 | © Author(s) 2026. CC BY 4.0 License.

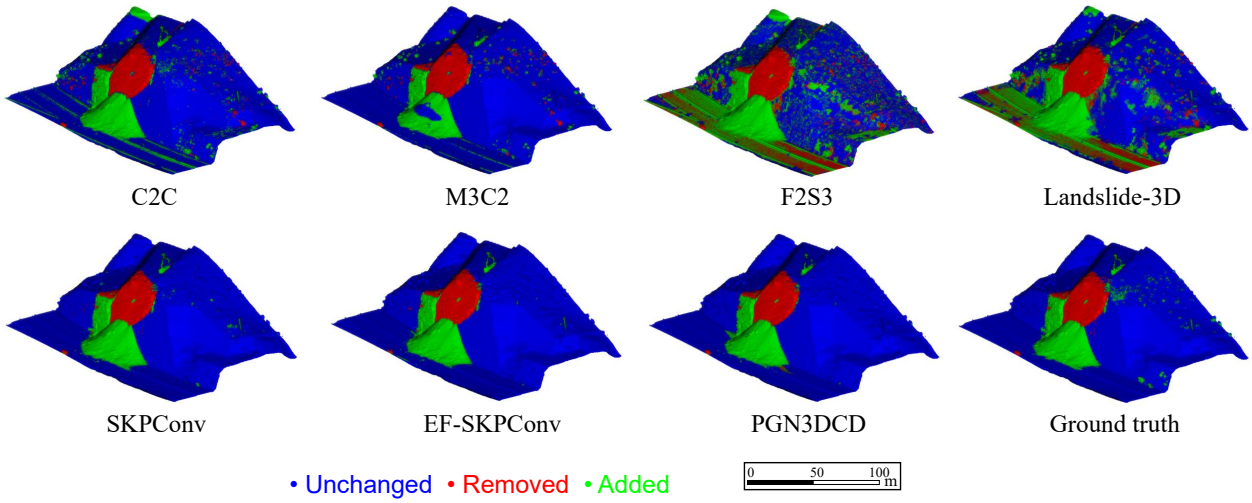

Figure 4. Qualitative results on the *Landslide* scene (epoch pair 0-1).

Landslide scene, performance degradation is less severe (e.g., PGN3DCD), suggesting better generalization across scenes of similar scale. In contrast, the Euclidean distance-based methods (particularly C2C) show more stable performance across diverse scenes, benefiting from their geometry-driven methodology that does not rely on learned features. 3D displacement estimation-based approaches show potential for detecting large changes across different scenes, while struggling with correctly detecting many “unchanged" points when the geometric structures are weak.

7. Conclusions

This paper presented **MultiChange3D**, a new multi-scene, multi-sensor 3DCD dataset for identifying geometric changes in 3D space. Performances and generalization of seven 3DCD methods from three categories across diverse real-world scenes captured by different sensors are also investigated. The key findings are:

- Euclidean distance-based methods (C2C, M3C2) achieve stable and consistent performance across scenes, benefiting from their minimal parameter requirements and independence from scene-specific characteristics, showing satisfactory generalizability across sensor and scene variations.
- 3D displacement estimation-based methods (F2S3, Landslide-3D) show potential for detecting large-scale or motion-related changes but face challenges in handling planar structures, where geometric features are less distinctive for reliable pointwise correspondences. Incorporating specific operations, such as ground removal for the *Open space (RGB-D)*, could enhance the performance of this category of methods.
- Deep learning-based classification approaches (KPConv, EF-KPConv, and PGN3DCD) perform best when trained and tested on the same scenes, but demonstrate degraded cross-scene generalization due to differences in point density, scale, and types of changes (*e.g.*, additions or removals of furniture in indoor scenes versus buildings in urban scenes).

Overall, our results highlight that while deep learning has achieved notable progress for scene-specific 3DCD, generalizable and robust detection of changes across multi-scene and multi-sensor data remains an open challenge. Future work will focus on incorporating semantic variations and developing a more generalizable approach to further advance research in 3DCD.

References

Bedkowski, J., 2024. Open source, open hardware hand-held mobile mapping system for large scale surveys. *SoftwareX*, 25, 101618.

Besl, P., McKay, N. D., 1992. A method for registration of 3-D shapes. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 14(2), 239-256.

Caye Daudt, R., Le Saux, B., Boulch, A., 2018. Fully convolutional siamese networks for change detection. *2018 25th IEEE International Conference on Image Processing (ICIP)*, 4063–4067.

Chaabouni-Chouayakh, H., Krauss, T., d’Angelo, P., Reinartz, P., 2010. 3D change detection inside urban areas using different digital surface models. *Int. Arch. Photogramm. Remote Sens. Spatial Inf. Sci.*, 38, 86–91.

Coletta, V., Marsocci, V., Ravanelli, R., 2022. 3DCD: A new dataset for 2D and 3D change detection using deep learning techniques. *Int. Arch. Photogramm. Remote Sens. Spatial Inf. Sci.*, XLIII-B3-2022, 1349–1354.

de Gélis, I., Corpetti, T., Lefèvre, S., 2024. Change Detection Needs Change Information: Improving Deep 3-D Point Cloud Change Detection. *IEEE Transactions on Geoscience and Remote Sensing*, 62, 1-10.

de Gélis, I., Lefèvre, S., Corpetti, T., 2021. Change Detection in Urban Point Clouds: An Experimental Comparison with Simulated 3D Datasets. *Remote Sensing*, 13(13).

de Gélis, I., Lefèvre, S., Corpetti, T., 2023. Siamese KPConv: 3D multiple change detection from raw point clouds using deep learning. *ISPRS J. Photogramm. Remote Sens.*, 197, 274-291.

Fehr, M., Furrer, F., Dryanovski, I., Sturm, J., Gilitschenski, I., Siegwart, R., Cadena, C., 2017. Tsdf-based change detection for consistent long-term dense reconstruction and dynamic object discovery. *2017 IEEE International Conference on Robotics and Automation (ICRA)*, 5237–5244.

Galve, J. P., Pérez-García, J. L., Ruano, P., Gomez-Lopez, J. M., Reyes-Carmona, C., Moreno-Sanchez, M. et al., 2025. Applications of UAV Digital Photogrammetry in landslide emergency response and recovery activities: the case study of a slope failure in the A-7 highway (S Spain). *Landslides*, 22(5), 1383–1396.

Girardeau-Montaut, D., Roux, M., Marc, R., Thibault, G., 2005. Change detection on points cloud data acquired with a ground laser scanner. *Int. Arch. Photogramm. Remote Sens. Spatial Inf. Sci.*, 36(3), W19.

Gojcic, Z., Schmid, L., Wieser, A., 2021. Dense 3D displacement vector fields for point cloud-based landslide monitoring. *Landslides*, 18(12), 3821–3832.

Huang, R., Xu, Y., Hoegner, L., Stilla, U., 2022. Semantics-aided 3D change detection on construction sites using UAV-based photogrammetric point clouds. *Automation in Construction*, 134, 104057.

Kharroubi, A., Remondino, F., Ballouch, Z., Hajji, R., Billen, R., 2025. Semantic and Geometric Fusion for Object-Based 3D Change Detection in LiDAR Point Clouds. *Remote Sensing*, 17(7).

Lague, D., Brodu, N., Leroux, J., 2013. Accurate 3D comparison of complex topography with terrestrial laser scanner: Application to the Rangitikei canyon (N-Z). *ISPRS J. Photogramm. Remote Sens.*, 82, 10-26.

Liu, Y., Pears, N., Rosin, P. L., Huber, P., 2020. *3D imaging, analysis and applications*. Springer.

Mazzanti, P., Caporossi, P., Brunetti, A., Mohammadi, F. I., Bozzano, F., 2021. Short-term geomorphological evolution of the Poggio Baldi landslide upper scarp via 3D change detection. *Landslides*, 18(7), 2367–2381.

Meyer, T., Brunn, A., Stilla, U., 2022. Change detection for indoor construction progress monitoring based on BIM, point clouds and uncertainties. *Automation in Construction*, 141, 104442.

Nagy, B., Kovács, L., Benedek, C., 2021. ChangeGAN: A Deep Network for Change Detection in Coarsely Registered Point Clouds. *IEEE Robotics and Automation Letters*, 6(4), 8277-8284.

Nikoohemat, S., Koeva, M., Oude Elberink, S., Lemmen, C., 2018. Change detection from point clouds to support indoor 3d cadastre. *Int. Arch. Photogramm. Remote Sens. Spatial Inf. Sci.*, 451–457.

Orkény, Z., Balázs, N., Csaba, B., 2022. Point cloud registration and change detection in urban environment using an onboard Lidar sensor and MLS reference data. *International Journal of Applied Earth Observation and Geoinformation*, 110, 102767.

Peppa, M. V., Mills, J. P., Moore, P., Miller, P. E., Chambers, J. E., 2019. Automated co-registration and calibration in SfM photogrammetry for landslide change detection. *Earth Surface Processes and Landforms*, 44(1), 287-303.

Poiesi, F., Boscaini, D., 2021. Distinctive 3d local deep descriptors. *2020 25th International Conference on Pattern Recognition (ICPR)*, 5720–5727.

Qin, R., Tian, J., Reinartz, P., 2016. 3D change detection – Approaches and applications. *ISPRS J. Photogramm. Remote Sens.*, 122, 41-56.

Sande, C. v. d., Soudarissanane, S., Khoshelham, K., 2010. Assessment of Relative Accuracy of AHN-2 Laser Scanning Data Using Planar Features. *Sensors*, 10(9), 8198–8214.

Schops, T., Sattler, T., Pollefeys, M., 2019. BAD SLAM: Bundle adjusted direct rgb-d slam. *Proc. CVPR*, 134–144.

Shan, J., Toth, C. K., 2018. *Topographic laser ranging and scanning: principles and processing*. CRC press.

Singh, A., 1989. Digital change detection techniques using remotely-sensed data. *International Journal of Remote Sensing*, 10(6), 989–1003.

Stilla, U., Xu, Y., 2023. Change detection of urban objects using 3D point clouds: A review. *ISPRS J. Photogramm. Remote Sens.*, 197, 228-255.

Szeliski, R., 2022. *Computer vision: algorithms and applications*. Springer Nature.

Thomas, H., Qi, C. R., Deschaud, J.-E., Marcotegui, B., Goulette, F., Guibas, L. J., 2019. KPConv: Flexible and deformable convolution for point clouds. *Proc. ICCV*.

Tran, T. H. G., Ressl, C., Pfeifer, N., 2018. Integrated Change Detection and Classification in Urban Areas Based on Airborne Laser Scanning Point Clouds. *Sensors*, 18(2).

Trybała, P., Kujawa, P., Romańczukiewicz, K., Szrek, A., Remondino, F., 2023. Designing and evaluating a portable LiDAR-based SLAM system. *nt. Arch. Photogramm. Remote Sens. Spatial Inf. Sci.,*, 48, 191–198.

Vosselman, G., Maas, H.-G., 2010. *Airborne and terrestrial laser scanning*. Whittles publishing.

Wang, Z., Butt, J. A., Huang, S., Medić, T., Wieser, A., 2026. Dense 3D displacement estimation for landslide monitoring via fusion of TLS point clouds and embedded RGB images. *International Journal of Applied Earth Observation and Geoinformation*, 146, 105093.

Wang, Z., Zhang, Y., Luo, L., Yang, K., Xie, L., 2023. An End-to-End Point-Based Method and a New Dataset for Street-Level Point Cloud Change Detection. *IEEE Transactions on Geoscience and Remote Sensing*, 61, 1-15.

Yew, Z. J., Hee Lee, G., 2021. City-scale scene change detection using point clouds. *2021 IEEE International Conference on Robotics and Automation (ICRA)*, 13362–13369.

Zeng, A., Song, S., Nießner, M., Fisher, M., Xiao, J., Funkhouser, T., 2017. 3DMatch: Learning local geometric descriptors from rgb-d reconstructions. *Proc. CVPR*, 199–208.

Zhan, W., Cheng, R., Chen, J., 2024. PGN3DCD: Prior-Knowledge-Guided Network for Urban 3-D Point Cloud Change Detection. *IEEE Transactions on Geoscience and Remote Sensing*, 62, 1-15.