2026 IEEE International Conference on Robotics and Automation (ICRA 2026)
June 1-5, 2026. Vienna, Austria

iVISION-2DCD: A Long-Term Change Detection Dataset for Large-Scale Outdoor Construction Monitoring

Dayou Mao[1], Yuchen Lin[2], Ashkan Ebadi[1], John Zelek[2], Alexander Wong[2], Yuhao Chen[2]

Abstract— **Automation in construction is essential for reducing costs and human errors in large-scale projects. We approach the construction progress monitoring from the aspect of detecting changes in construction sites. As construction buildings continue to evolve in geometry and appearance over time, change detection need to be performed from *arbitrary* camera viewpoints. This necessitates developing 2D Change Detection (2DCD) algorithms that operate robustly across diverse camera perspectives at construction sites. While developing and evaluating such systems is data-intensive, no open-source benchmark dataset exists at the intersection of 2D change detection and construction automation research. Data collection using Unmanned Aerial Vehicles (UAVs) is gaining its popularity in outdoor large-scale surveying. However, in active construction sites conducting drone missions equipped with high-end sensors imposes safety concerns. Flight trajectory and collected camera viewpoints can be significantly limited. To address this critical gap, we introduce iVISION-2DCD, a large-scale synthetically generated dataset from dense LiDAR point clouds with photorealistic input images and accurate ground truth annotations. Our dataset formally defines the problem of viewpoint-robust 2DCD at construction sites and captures the inherent complexities of real-world deployment. In this paper, we present our systematic methodology for synthetic data generation, developing novel view synthesis techniques to overcome bi-temporal alignment and viewpoint diversity challenges, and implementing semi-automated semantic segmentation with change label generation while preserving challenging real-world cases. Benchmark evaluations using state-of-the-art 2DCD algorithms demonstrate that iVISION-2DCD poses novel research challenges for the computer vision and robotics communities.**

I. INTRODUCTION

The construction industry, representing approximately USD 10 trillion annually and comprising 13% of global GDP, faces severe inefficiencies where over 53% of projects experience delays and 66% exceed budgets due to inadequate monitoring approaches [8], [9]. Manual monitoring methods contribute to schedule overruns of up to 20% and cost escalations reaching 80% above estimates [10], [11], highlighting the urgent need for automated solutions. Change Detection (CD) algorithms enable automated progress monitoring by providing objective, quantitative measurements that circumvent time-consuming manual inspections and facilitate continuous monitoring across large spatial extents impractical for traditional survey methods [2], [4], [3]. Unmanned Aerial Vehicles (UAVs) provide efficient large-scale data acquisition through cost-effective monitoring [3], enhanced safety by eliminating ground-based personnel requirements [8], and affordable 3D reconstruction with rich semantic information [6]. While 3D Change Detection (3DCD) offers geometric advantages, 2D Change Detection (2DCD) remains the practical choice for widespread deployment: RGB-only drones are significantly more cost-effective and lightweight compared to LiDAR-equipped systems, reducing safety constraints and operational expertise requirements that limit accessibility for construction teams. In this paper, we focus on 2DCD for construction progress monitoring.

While deep learning approaches have shown promise in this domain, they require substantial training data. Despite this need, existing datasets exhibit critical limitations: temporal sampling inadequacies with studies spanning merely 2.5-4 months and 3-5 timestamps, spatial resolution constraints limiting detection precision, and environmental bias toward controlled indoor scenarios that fail to capture outdoor complexities including seasonal variations and weather conditions [3], [6], and single camera pose from top-down view captured by satellite imageries [15], [12], [26]. There exists no comprehensive dataset for 2DCD in long-term, large-scale outdoor construction environments that addresses these fundamental limitations.

To address this critical gap in 2DCD datasets for construction monitoring, we introduce iVISION-2DCD and present a comprehensive methodology for curating this large-scale dataset. Our data acquisition employs a DJI Matrice 350 drone equipped with RGB cameras and LiDAR sensors, enabling dense multi-modal capture. Flight trajectories are planned to balance operational constraints and data quality. The dataset curation addresses the challenge of diverse viewpoint sampling through novel view synthesis from dense LiDAR point clouds, systematic camera generation, and pixel coverage optimization, creating bi-temporally aligned image pairs from arbitrary viewing angles. We algorithmically generate change labels from semantic class annotations to overcome manual annotation costs, while preserving dataset authenticity by retaining challenging real-world cases: terrain changes invisible in RGB, dynamic object artifacts from construction equipment, and systematic projection errors. Our methodology transforms raw multi-temporal acquisitions into a structured benchmark with diverse viewpoints and accurate change labels, capturing construction dynamics from subtle material additions to major structural transformations across varying environmental conditions.

Our resulting dataset distinguishes itself along the following six dimensions: (i) high-fidelity data acquisition utilizing

*Corresponding Authors:
Dayou Mao (daniel.mao@nrc-cnrc.gc.ca)
Ashkan Ebadi (ashkan.ebadi@nrc-cnrc.gc.ca)
Yuhao Chen (yuhao.chen1@uwaterloo.ca)
[1]National Research Council Canada, Ottawa, ON, Canada
[2]Vision and Image Processing Research Group, Systems Design Engineering, University of Waterloo, Waterloo, ON, Canada

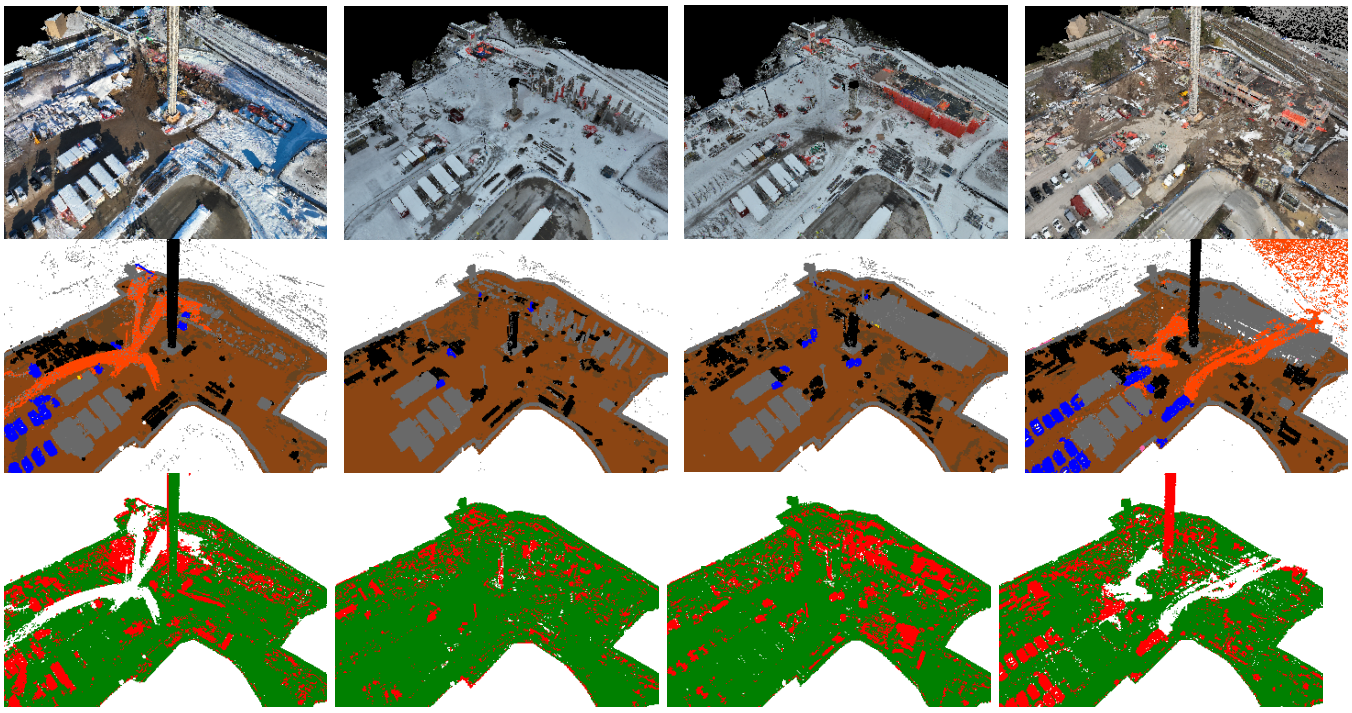

Fig. 1. Example scenes (not consecutive) rendered from a distant camera: Dec-03-2024, Jan-19-2025, Feb-08-2025, Mar-12-2025, from left to right. First row: rendered RGB image from the LiDAR point cloud. Second row: rendered 3D segmentation point cloud from manual annotation. Third row: rendered 3D change point cloud compared from immediate previous scene. In the change maps, green represents unchanged points; red represents changed points; white represents ignored points.

state-of-the-art UAV platforms with advanced imaging sensors; (ii) synthetic 2DCD data generated from synthetically generated diverse camera angles; (iii) long-term temporal coverage with over 50 acquisition sessions spanning one full year, capturing seasonal variations including winter snow conditions; (iv) extensive spatial coverage spanning large construction sites (32.3K m^2); (v) authentic complexity from active construction environments; and (vi) multi-modal data incorporating both geometric and textural information. We establish performance baselines for state-of-the-art 2DCD algorithms, demonstrating that iVISION-2DCD introduces novel challenges at the intersection of computer vision, robotics, and construction automation.

Our contribution in this paper is 3-fold:

- We present a systematic methodology for generating synthetic 2DCD data from dense LiDAR point clouds, overcoming operational constraints through novel view synthesis and semi-automated annotation.
- We introduce iVISION-2DCD, a novel benchmark dataset that captures the unique complexities of real-world construction monitoring scenarios.
- We demonstrate that iVISION-2DCD establishes new research challenges on vision-based construction progress monitoring systems through comprehensive benchmark comparison of existing 2DCD algorithms and open-sourced datasets.

This work represents an ongoing research effort. The work in this paper utilizes data from November 27, 2024 to March 12, 2025. Raw drone data from September 2024 to August 2025 has been made public.

II. RELATED WORK

A. Construction Datasets

Datasets supporting research in automation in construction progress monitoring exhibit critical limitations across three dimensions: temporal sampling inadequacies that miss construction dynamics, environmental bias toward controlled indoor settings, and single modality dependencies that ignore complementary sensor information.

Temporal Sampling Inadequacies. Current outdoor construction datasets demonstrate critically sparse temporal sampling that fails to capture construction dynamics adequately. UAV-based road construction monitoring spans only 2.5 months with merely 3 temporal acquisitions [3], while UAV photogrammetric construction site studies cover 4 months with only 5 timestamps [6]. This temporal sparsity prevents detection of gradual construction changes and completely misses seasonal environmental variations essential for robust outdoor monitoring algorithms.

Environmental Bias and Limited Outdoor Applicability. A significant portion of construction change detection research demonstrates problematic bias toward indoor environments that fails to address outdoor complexities [2], [1]. Indoor construction monitoring faces constraints including limited views, weak scanning geometries, and occlusions while demanding higher geometric resolution [2], yet these controlled conditions cannot capture environmental challenges inherent in outdoor sites including seasonal vegetation

variations, dynamic shadow patterns, and weather-induced visual changes.

Single Modality Dependencies. Many studies rely on single data modalities - either RGB images only, point clouds only, or SAR imagery only - missing complementary information available through multi-modal sensing approaches that combine geometric and radiometric information for robust environmental adaptation [3], [7], [5].

B. Change Detection

We provide a brief literature review of Change Detection (CD) starting with the traditional bi-temporal framework, then examining existing benchmark datasets and their limitations, followed by two improved learning paradigms that address specific limitations: single-temporal methods for reducing annotation costs and 3D approaches for incorporating geometric information.

Bi-Temporal CD. Typical formulation of (binary) 2DCD task is: given bi-temporal, RGB images I_1 and I_2 captured by the same camera (with same camera intrinsics and pose), we are to perform dense classification to predict class 0 to unchanged pixels, and class 1 to changed pixels [20], [21], [22], [24], [25]. Neural networks can be trained with supervision from ground-truth per-pixel change labels. However, the initial research focus has two limitations - high costs in dense bi-temporal change labels as compared to single-temporal labels; and lack of depth information in 2D maps, which are respectively address by single-temporal CD and 3DCD research, as we describe in the following paragraphs.

2DCD Datasets. Existing 2DCD benchmarks include satellite datasets (OSCD [15], Sentinel-2, Landsat) with coarse resolution (10-30m), aerial imagery datasets (LEVIR-CD [13], SYSU-CD [26], WHU-CD) with higher resolution (0.5-2m) for urban monitoring, and disaster assessment datasets (xBD [17], AirChange [16]). These datasets share a critical limitation: exclusive reliance on nadir-view imagery from remote sensing platforms, leaving vertical structures unobserved, volumetric changes unquantified, and creating occlusion blind spots. The absence of oblique or street-level perspectives severely limits their applicability to real-world scenarios requiring multi-angle change analysis.

Single-Temporal CD. Single-temporal CD addresses the prohibitive expense of pairwise labeling bi-temporal images by reformulating CD from requiring temporally-paired images with pixel-perfect alignment to learning from arbitrary unpaired images, where any two single-temporal images can form a training pair regardless of their temporal or spatial correspondence [36], [38]. The mainstream approach constructs pseudo bi-temporal pairs by artificially introducing changes into single-temporal images and leveraging existing semantic annotations to automatically generate change labels, transforming the expensive bi-temporal annotation problem into a more tractable single-temporal semantic labeling task [36], [37], [38], [39], [40].

3DCD. A sub-area of CD [41], [42], [43], [44] design neural networks on point cloud CD datasets [18], [19] due to the well-known problem of lack of elevation information from 2D views. In urban building construction application, this problem is especially pronounced. Change map generation in 3D space is necessary before projecting into 2D for curating our iVISION-2DCD dataset.

III. Synthetic Data Generation

Obtaining 2DCD data for construction monitoring faces three fundamental challenges: UAV missions cannot achieve pixel-perfect bi-temporal alignment, operational constraints limit viewpoint diversity, and manual change annotation proves prohibitively expensive. We develop novel view synthesis techniques to solve the bi-temporal alignment and viewpoint diversity challenges (Section III-A), and implement semi-automated semantic segmentation with change label generation to overcome the manual annotation cost challenge (Section III-B). This pipeline transforms dense LiDAR point clouds into co-registered photorealistic imagery with accurate ground truth change labels capturing construction dynamics from perspectives impossible during actual drone missions.

A. Novel View Synthesis

Operational UAV missions must continuously optimize trajectories around evolving construction geometry - structures rise, and site layouts transform between captures. This makes it impossible to repeat exact flight paths and camera poses across temporal acquisitions. Yet typical 2DCD algorithms require pixel-perfect bi-temporal alignment with identical camera poses, fundamentally necessitating synthetic rendering from our LiDAR reconstructions. However, synthetic rendering introduces two additional challenges we must address: limited viewpoint diversity from safety restrictions in real captures (Section III-A.1) and pixel coverage gaps inherent in point cloud rendering (Section III-A.2).

1) Synthetic Camera Generation: While synthetic rendering solves temporal alignment, it inherits the viewpoint limitations of our real captures. Operational safety constraints prohibit UAV flights inside construction fence boundaries and below 65m crane height during weekdays, restricting our LiDAR missions to predominantly nadir perspectives captured from safe altitudes. Yet robust 2DCD algorithms require training and evaluation on diverse viewpoints - including ground-level views and oblique angles - to generalize across deployment scenarios. To transcend these physical limitations, we implement three complementary strategies addressing altitude restrictions, interior access constraints, and close-range imaging limitations: (1) Vertical sampling: Generates cameras at varying altitudes below operational flight levels, perpendicular to the ground plane; (2) Mirror sampling: Creates interior viewpoints by reflecting exterior camera positions across site boundaries; and (3) Forward sampling: Produces close-range perspectives by translating cameras toward construction elements. Examples are shown in Fig. 2. This cascaded synthesis pipeline generates 32 synthetic viewpoints from each real camera position, effectively mitigating the viewpoint limitations imposed by operational safety constraints.

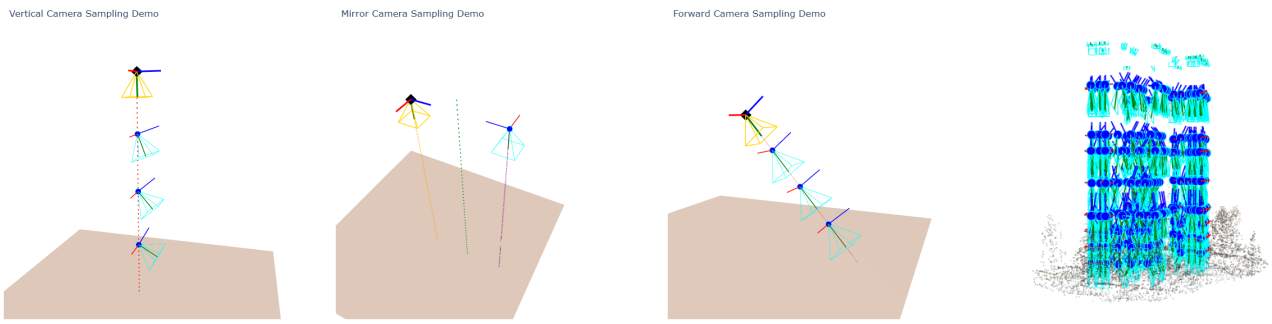

Fig. 2. From left to right: Illustration of vertical, mirror, forward sampling, and visualization of complete synthetic camera generation for 20 randomly sampled real cameras in November 27, 2024 scene. Camera right, forward, and up directions are visualized in red, green, and blue, respectively. Ground plane is visualized in saddle brown. Dashed lines in vertical sampling and mirror sampling represent the vertical projection of the camera position to the ground plane. Dashed line in forward sampling represent the movement (forward) direction.

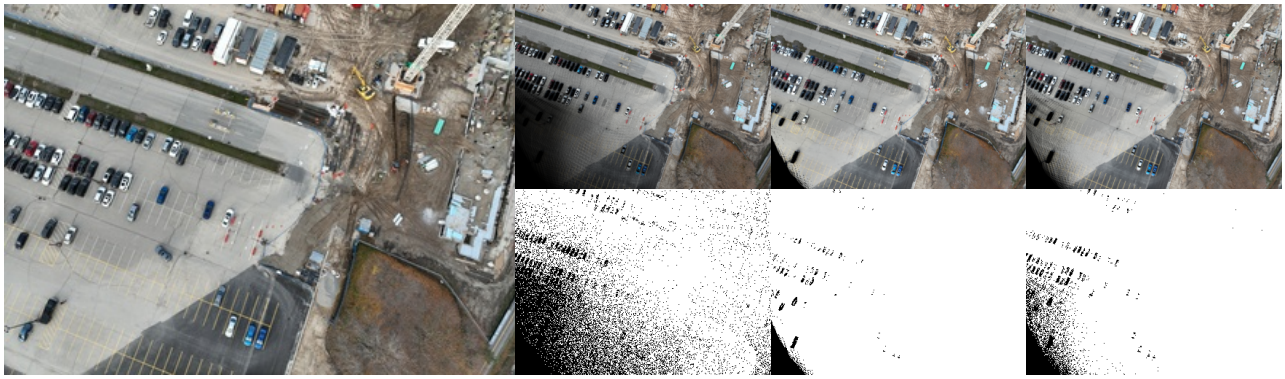

Fig. 3. Left: original capture of the first image in the November 27 LiDAR mission. Right: synthetic rendering at the same scene, camera intrinsics and extrinsics, using (original resolution, 1 pixel point size), (1/8 resolution, 1 pixel point size), and (original resolution, 3 pixels point size), from left to right respectively. The respective pixel hit masks (white for pixel hit and black for pixel miss) are shown in the second row.

2) Maximizing Pixel Coverage: Despite our exceptionally dense LiDAR point clouds (200-800M points), their discrete, non-volumetric nature creates a fundamental rendering challenge - each point projects to isolated pixels rather than continuous surfaces. When projected to 2D at original resolution, we achieve only ∼77% pixel coverage, leaving ∼23% of pixels filled with zeros. These coverage gaps require explicit masking during neural network training and degrade synthetic image quality. Existing rendering solutions prove inadequate for our scale: open-source libraries either fail computationally on our massive point clouds [48] or require mesh representations that introduce additional preprocessing overhead and potential reconstruction artifacts [49].

We implement a pure PyTorch-based rendering pipeline that provides complete control over the synthesis process while fully exploiting our high-precision LiDAR data. Our analysis quantifies pixel coverage improvements through two straightforward yet effective strategies. The first strategy to increase pixel coverage is to simply render at a lower resolution (downscale the camera intrinsics when projecting points to image plane). The second strategy is to make a simple model for volumetric points by letting each point occupy pixels in a disk of diameter 3 pixels.

Fig. 3 shows that although rendering at original resolution and 1 pixel point size (77.29% pixel coverage) provides close-to-real texture on the top-right diagonal, the bottom-left diagonal has dark regions because of insufficient pixel coverage, and the missed pixels were assigned 0 values, which makes black colors. Rendering at 1/8 scale with 1 pixel point size (97.46% pixel coverage) or rendering at original resolution and 3 pixels point size (94.05% pixel coverage) both significant clears up the missing pixels due to point cloud *sparsity*. However, regions of point cloud *incompleteness* cannot be solved. Our image resolution is at (5280, 3956) width and height in the raw data, (5645, 4082) in the undistorted images from COLMAP [50], and (706, 510) after 1/8 downscale, which is still higher than most existing 2DCD datasets [12], [26] and what most algorithms process [20], [23], [25]. In our iVISION-2DCD generation, we stick to using 1/8 resolution and 1 pixel point size.

B. Semantic Segmentation and Change Labels

Generating accurate ground truth labels for construction 2DCD datasets faces three fundamental challenges: 2D images lack geometry information necessary for distinguishing structural changes from perspective variations [41], [18], dense bi-temporal change annotation proves prohibitively expensive at scale [36], [38], and construction datasets require domain-specific filtering to exclude irrelevant environmental noise. To address these challenges, we perform

semantic segmentation directly on 3D point clouds to preserve full geometric information, follow the single-temporal paradigm to annotate single-temporal semantic labels then deriving changes through bi-temporal comparison, and implement construction-focused annotation to filter irrelevant noise. This subsection details our semantic segmentation and change detection labeling methods.

1) Semantic Segmentation Labels: We use DJI Modify [46] for initial automated point cloud labeling, followed by manual refinement to add construction-specific categories and correct labeling around dynamic objects. Additionally, we designate all regions outside the construction fence boundary as "Ignore" class, along with irrelevant temporal variations such as airborne snow, vegetation growth, and activities beyond site perimeters, ensuring our dataset focuses exclusively on construction-specific changes within the active construction zone. This approach achieves highly accurate per-point semantic labels with fine-grained class distinctions, leading to a dataset specifically tailored for construction automation research. Examples are shown in Fig. 1.

2) Change Detection Labels: Point clouds inherently produce asymmetric change labels due to varying point distributions [18], [19] - changes detected from time t_1 to t_2 differ from those detected from t_2 to t_1. Yet typical 2DCD algorithms typically require symmetric change masks where pixel (i, j) shows identical change regardless of temporal direction [12], [13], [26]. We resolve this fundamental mismatch through bi-directional 3D semantic comparisons, computing changes in both temporal directions, projecting and rasterizing each to image space, then taking union (logical OR) of the two resulting change maps as the final 2DCD ground-truth label. This ensures the resulting 2D maintain the symmetry that canonical 2DCD training demands.

IV. Dataset Characteristics

We present the visual quality challenge descriptions and examples and dataset statistics in this section.

A. Challenging Cases

Our dataset exhibits 4 challenging cases: Snow, Ignore Ground, Dynamic Objects, and RGB Projection Errors.

Snow creates annotation ambiguity as accumulated snow intermixes with construction elements while airborne snow remains separable. We assign accumulated snow the semantic class of underlying objects and label airborne particles as "Snow". Examples are shown in Fig. 4.

Ignore Ground. Subtle terrain variations from material staging and excavation are invisible in RGB but indicate construction progress through micro-topographical changes. Examples are shown in Fig. 5.

Dynamic Objects create point cloud artifacts: noise clusters from slow objects (cranes, workers), trailing artifacts from vehicles, and missing data from high-speed traffic ($\geq$ 60 km/h). Examples are shown in Fig. 6.

RGB Projection Errors. The LiDAR reconstruction pipeline introduces systematic complexities where thin elevated structures (cranes, poles) project incorrect colors onto ground surfaces, creating confounding visual patterns. Examples are shown in Fig. 7.

B. Dataset Statistics

For the temporal window spanning November 27, 2024 to March 12, 2025 (17 scenes), we generate 16 consecutive scene pairs, yielding 32 bi-directional change maps. Our drone LiDAR missions typically contains 300-350 images per scene. Therefore, we downsample the $32N$ synthetic cameras generated from N original cameras to 300 per scene, resulting in $300 \times 16 = 4800$ synthetic cameras.

V. Experiments

A. Setup

Selected Methods and Datasets. We evaluate 18 representative 2D change detection methods: FC-Siam-diff, FC-Siam-conc, FC-EF [20], DSAMNet [26], DSIFN [27], HANet [29], SNUNet-ECAM [21], TinyCD [30], RFL-CDNet [31], ChangeFormer [23], BiFA [32], CDX-Former [33], Changer [34], FTN [22], HCGMNet [35], ChangeMamba [25], ChangeNext, DsferNet [28]. We benchmark these methods on six datasets: our oblique-view construction dataset iVISION-2DCD alongside five established nadir-view benchmarks - LEVIR-CD [13], SYSU-CD [26], OSCD [15], CDD [14], and AirChange [16]. These methods and datasets were selected based on their popularity, reported performance, and source code availability.

Training. We set up our training following [36], [20], [38]. We assign class weights inverse-proportionally to the number of pixels in the dataset. Throughout all experiments, we use SGD optimizer with initial learning rate 10^{-3}, momentum 0.9, and weight decay 10^{-4}. We use polynomial learning rate scheduler with power 0.9. Batch size was set to 4 throughout. Standard data augmentations including random rotation, random flip, random color jitter, and random crop has been applied on all existing datasets. Differently, iVISION-2DCD synthetically samples diverse close-up views towards the construction site. Random rotation is no longer a reasonable data augmentations as it would create clearly out-of-distribution samples, e.g., an upside-down crane, and hence removed from the list.

B. Results

Table I reports mIoU scores across all methods and datasets, revealing that iVISION-2DCD presents fundamentally different challenges compared to existing benchmarks.

Performance Degradation. All methods experience significant performance drops on iVISION-2DCD compared to established benchmarks. State-of-the-art methods like ChangeMamba decline from 84.67% (LEVIR-CD) and 91.35% (CDD) to 52.56% on our dataset, while BiFA drops from 81.11% (LEVIR-CD) to 50.55%.

Ranking Inversions. More critically, method rankings exhibit dramatic reversals between datasets. Methods achieving top performance on existing benchmarks often fail on iVISION-2DCD: ChangeFormer ranks 1st on OSCD (50.38%) and 2nd on AirChange (75.07%) but drops to

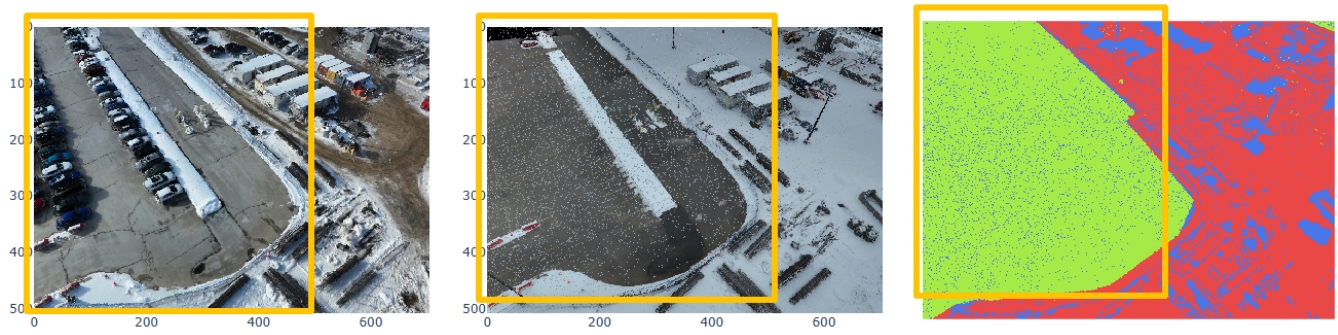

Fig. 4. Example of airborne snow captured in the point clouds. Left column: time t_1 input images. Middle column: time t_2 input images. Right column: change labels. Changes due to airborne snow. This has been re-mapped to "Ignore" class and not present in our generated ground-truth changes.

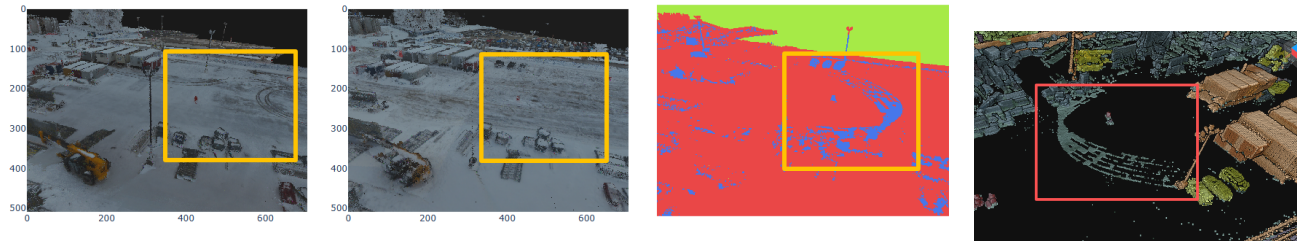

Fig. 5. Example of ignore ground semantic changes hard to recognize from RGB inputs. From left to right: time t_1 input image, time t_2 input image, ground truth change map, and semantic segmentation map (with "Ground" class made invisible).

Fig. 6. Examples of dynamic elements in the construction site. Figures shown by DJI Modify to amplify the dynamic points. From left to right: moving crane, moving string attached to the crane, moving vehicle, and moving construction workers.

Fig. 7. Examples of RGB projection error in the reconstructed LiDAR point clouds. Figures shown by DJI Modify to amplify the error points. From left to right: red color of the crane falsely projected onto the ground; crane frameworks grey colors projected onto the ground; light pole in construction site texture projected onto the ground.

TABLE I

BENCHMARK RESULTS OF WIDELY-STUDIED AND STATE-OF-THE-ART METHODS ON OUR IVISION-2DCD DATASET AND FIVE OTHER WELL-ESTABLISHED BENCHMARK DATASETS. ALL SCORES ARE MEAN INTERSECTION OVER UNION (MIOU) IN PERCENTAGES. NUMBERS IN PARENTHESES INDICATE RANKINGS ON EACH DATASET. TOP THREE METHODS ON EACH DATASET ARE HIGHLIGHTED IN BOLD. METHODS ARE SORTED BY THEIR PERFORMANCE ON IVISION-2DCD IN DESCENDING ORDER.

Method	LEVIR-CD [13]	SYSU-CD [26]	OSCD [15]	CDD [14]	AirChange [16]	iVISION-2DCD (Ours)
FTN	33.05 (16)	**72.24 (2)**	25.73 (16)	86.18 (4)	33.53 (18)	**53.77 (1)**
ChangeMamba	**84.67 (1)**	**72.26 (1)**	**48.97 (2)**	**91.35 (1)**	**73.87 (3)**	**52.56 (2)**
HCGMNet	68.44 (5)	70.58 (4)	47.93 (13)	**88.72 (3)**	71.98 (4)	**52.05 (3)**
BiFA	**81.11 (2)**	61.03 (10)	48.62 (8)	44.31 (14)	**75.79 (1)**	50.55 (4)
DsferNet	32.00 (17)	**72.03 (3)**	48.70 (5)	**91.14 (2)**	50.81 (15)	46.56 (5)
RFL-CDNet	62.47 (6)	65.64 (8)	48.63 (7)	50.77 (9)	63.84 (10)	44.32 (6)
CDXFormer	**76.55 (3)**	68.05 (6)	48.62 (9)	84.56 (5)	68.39 (6)	43.49 (7)
ChangeFormer	54.58 (8)	63.21 (9)	**50.38 (1)**	44.91 (13)	**75.07 (2)**	41.79 (8)
SNUNet-ECAM	48.75 (11)	68.69 (5)	**48.70 (3)**	57.31 (8)	70.24 (5)	41.71 (9)
FC-Siam-diff	6.82 (18)	55.43 (12)	2.93 (18)	43.99 (15)	61.81 (12)	40.98 (10)
ChangeNext	49.16 (9)	52.70 (16)	46.66 (14)	61.72 (7)	57.20 (13)	40.36 (11)
FC-Siam-conc	48.23 (12)	54.01 (14)	48.70 (4)	43.81 (16)	64.54 (9)	40.10 (12)
HANet	68.65 (4)	66.18 (7)	48.11 (12)	45.27 (11)	68.11 (7)	39.82 (13)
TinyCD	43.44 (14)	54.85 (13)	22.34 (17)	65.82 (6)	54.44 (14)	39.41 (14)
FC-EF	40.38 (15)	53.36 (15)	48.62 (11)	43.23 (17)	67.27 (8)	39.16 (15)
DSIFN	48.99 (10)	40.12 (17)	48.62 (10)	44.92 (12)	49.81 (16)	38.73 (16)
DSAMNet	47.46 (13)	31.79 (18)	48.70 (6)	39.64 (18)	41.75 (17)	35.73 (17)
Changer	54.88 (7)	59.30 (11)	40.14 (15)	48.60 (10)	61.90 (11)	29.78 (18)

8th place (41.79%) on ours. DsferNet ranks 2nd on CDD (91.14%) and 3rd on SYSU-CD (72.03%) yet only 5th on ours (46.56%) while failing catastrophically on LEVIR-CD (17th, 32.00%). Conversely, methods performing poorly elsewhere succeed on our dataset: FTN consistently ranks near bottom positions (18th on AirChange, 16th on LEVIR-CD and OSCD) yet achieves top performance (53.77%) on iVISION-2DCD, while FC-Siam-diff catastrophically fails on LEVIR-CD (18th, 6.82%) and OSCD (18th, 2.93%) but manages 10th place (40.98%) on ours.

These dramatic ranking inversions indicate that iVISION-2DCD captures unique visual complexities from oblique-view construction scenes that fundamentally differ from nadir-view patterns in existing aerial and satellite benchmarks.

VI. CONCLUSION

In this paper, we have presented iVISION-2DCD, a comprehensive synthetic 2DCD dataset for construction monitoring generated from dense LiDAR point clouds. We developed novel view synthesis techniques to overcome fundamental challenges in 2DCD data curation: achieving pixel-perfect bi-temporal alignment through synthetic rendering, transcending operational viewpoint limitations via systematic camera generation, and optimizing pixel coverage to 97% through strategic resolution choices. Our semi-automated semantic segmentation pipeline leverages single-temporal labeling paradigms to generate accurate change maps while preserving challenging real-world cases including snow covering, terrain changes invisible in RGB, dynamic object artifacts, and systematic projection errors. The dataset spans over one year with 50+ temporal captures, providing both geometric and textural richness essential for robust algorithm development. Through extensive experiments with widely-studied and state-of-the-art 2DCD algorithms, we demonstrated that iVISION-2DCD introduces novel challenges specific to construction monitoring applications. Our work establishes a new benchmark for viewpoint-robust 2DCD in construction environments, opening advancement opportunities at the intersection of computer vision, robotics, and construction automation.

ACKNOWLEDGMENT

This work was supported by National Research Council of Canada (NRC) through the Construction Sector Digitalization and Productivity Challenge program, project number CSDP-014-1.

REFERENCES

[1] Czerniawski, T., Ma, J. W., & Leite, F. (2021). Automated building change detection with amodal completion of point clouds. Automation in construction, 124, 103568.

[2] 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.

[3] Han, D., Lee, S. B., Song, M., & Cho, J. S. (2021). Change detection in unmanned aerial vehicle images for progress monitoring of road construction. Buildings, 11(4), 150.

[4] Suh, J. W., Zhu, Z., & Zhao, Y. (2024). Monitoring construction changes using dense satellite time series and deep learning. Remote Sensing of Environment, 309, 114207.

[5] Yang, C. H., Pang, Y., & Soergel, U. (2017). Monitoring of building construction by 4D change detection using multi-temporal SAR images. ISPRS Annals of the Photogrammetry, Remote Sensing and Spatial Information Sciences, 4, 35-42.

[6] 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.

[7] Attard, L., Debono, C. J., Valentino, G., & Di Castro, M. (2018). Vision-based change detection for inspection of tunnel liners. Automation in Construction, 91, 142-154.

[8] Ham, Y., Han, K. K., Lin, J. J., & Golparvar-Fard, M. (2016). Visual monitoring of civil infrastructure systems via camera-equipped Unmanned Aerial Vehicles (UAVs): a review of related works. Visualization in Engineering, 4(1), 1.

[9] Agarwal, R., Chandrasekaran, S., & Sridhar, M. (2016). Imagining construction's digital future. McKinsey & Company, 24(06), 1-13.

[10] Golparvar-Fard, M., Pena-Mora, F., & Savarese, S. (2015). Automated progress monitoring using unordered daily construction photographs and IFC-based building information models. Journal of Computing in Civil Engineering, 29(1), 04014025.

[11] Yang, J., Park, M. W., Vela, P. A., & Golparvar-Fard, M. (2015). Construction performance monitoring via still images, time-lapse photos, and video streams: Now, tomorrow, and the future. Advanced Engineering Informatics, 29(2), 211-224.

[12] Ji, S., Wei, S., & Lu, M. (2018). Fully convolutional networks for multisource building extraction from an open aerial and satellite imagery data set. IEEE Transactions on geoscience and remote sensing, 57(1), 574-586.

[13] Chen, H., & Shi, Z. (2020). A spatial-temporal attention-based method and a new dataset for remote sensing image change detection. Remote sensing, 12(10), 1662.

[14] Lebedev, M. A., Vizilter, Y. V., Vygolov, O. V., Knyaz, V. A., & Rubis, A. Y. (2018). Change detection in remote sensing images using conditional adversarial networks. The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences, 42, 565-571.

[15] Daudt, R. C., Le Saux, B., Boulch, A., & Gousseau, Y. (2018, July). Urban change detection for multispectral earth observation using convolutional neural networks. In IGARSS 2018-2018 IEEE International Geoscience and Remote Sensing Symposium (pp. 2115-2118). Ieee.

[16] Benedek, C., & Szirányi, T. (2009). Change detection in optical aerial images by a multilayer conditional mixed Markov model. IEEE Transactions on Geoscience and Remote Sensing, 47(10), 3416-3430.

[17] Gupta, R., Hosfelt, R., Sajeev, S., Patel, N., Goodman, B., Doshi, J., ... & Gaston, M. (2019). xbd: A dataset for assessing building damage from satellite imagery. arXiv preprint arXiv:1911.09296.

[18] 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), 2629.

[19] 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.

[20] Daudt, R. C., Le Saux, B., & Boulch, A. (2018, October). Fully convolutional siamese networks for change detection. In 2018 25th IEEE international conference on image processing (ICIP) (pp. 4063-4067). IEEE.

[21] Fang, S., Li, K., Shao, J., & Li, Z. (2021). SNUNet-CD: A densely connected Siamese network for change detection of VHR images. IEEE Geoscience and Remote Sensing Letters, 19, 1-5.

[22] Yan, T., Wan, Z., & Zhang, P. (2022). Fully transformer network for change detection of remote sensing images. In Proceedings of the Asian Conference on Computer Vision (pp. 1691-1708).

[23] Bandara, W. G. C., & Patel, V. M. (2022, July). A transformer-based siamese network for change detection. In IGARSS 2022-2022 IEEE International Geoscience and Remote Sensing Symposium (pp. 207-210). IEEE.

[24] Zhang, H., Chen, K., Liu, C., Chen, H., Zou, Z., & Shi, Z. (2024). CD-Mamba: Incorporating Local Clues into Mamba for Remote Sensing Image Binary Change Detection. arXiv preprint arXiv:2406.04207.

[25] Chen, H., Song, J., Han, C., Xia, J., & Yokoya, N. (2024). ChangeMamba: Remote sensing change detection with spatiotemporal state space model. IEEE Transactions on Geoscience and Remote Sensing, 62, 1-20.

[26] Shi, Q., Liu, M., Li, S., Liu, X., Wang, F., & Zhang, L. (2021). A deeply supervised attention metric-based network and an open aerial image dataset for remote sensing change detection. IEEE transactions on geoscience and remote sensing, 60, 1-16.

[27] Zhang, C., Yue, P., Tapete, D., Jiang, L., Shangguan, B., Huang, L., & Liu, G. (2020). A deeply supervised image fusion network for change detection in high resolution bi-temporal remote sensing images. ISPRS Journal of Photogrammetry and Remote Sensing, 166, 183-200.

[28] Chang, S., Kopp, M., Ghamisi, P., & Du, B. (2024). Dsfer-Net: A deep supervision and feature retrieval network for bitemporal change detection using modern Hopfield networks. IEEE Transactions on Geoscience and Remote Sensing.

[29] Han, C., Wu, C., Guo, H., Hu, M., & Chen, H. (2023). HANet: A hierarchical attention network for change detection with bitemporal very-high-resolution remote sensing images. IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing, 16, 3867-3878.

[30] Codegoni, A., Lombardi, G., & Ferrari, A. (2023). TINYCD: A (not so) deep learning model for change detection. Neural Computing and Applications, 35(11), 8471-8486.

[31] Gan, Y., Xuan, W., Chen, H., Liu, J., & Du, B. (2024). RFL-CDNet: Towards accurate change detection via richer feature learning. Pattern Recognition, 153, 110515.

[32] Zhang, H., Chen, H., Zhou, C., Chen, K., Liu, C., Zou, Z., & Shi, Z. (2024). Bifa: Remote sensing image change detection with bitemporal feature alignment. IEEE Transactions on Geoscience and Remote Sensing, 62, 1-17.

[33] Wu, Z., Ma, X., Lian, R., Lin, Z., & Zhang, W. (2024). CDXFormer: Boosting Remote Sensing Change Detection with Extended Long Short-Term Memory. arXiv e-prints, arXiv-2411.

[34] Fang, S., Li, K., & Li, Z. (2023). Changer: Feature interaction is what you need for change detection. IEEE Transactions on Geoscience and Remote Sensing, 61, 1-11.

[35] Han, C., Wu, C., & Du, B. (2023, July). HCGMNet: A hierarchical change guiding map network for change detection. In IGARSS 2023-2023 IEEE International Geoscience and Remote Sensing Symposium (pp. 5511-5514). IEEE.

[36] Zheng, Z., Ma, A., Zhang, L., & Zhong, Y. (2021). Change is everywhere: Single-temporal supervised object change detection in remote sensing imagery. In Proceedings of the IEEE/CVF international conference on computer vision (pp. 15193-15202).

[37] Zheng, Z., Zhong, Y., Ma, A., & Zhang, L. (2024). Single-temporal supervised learning for universal remote sensing change detection. International Journal of Computer Vision, 132(12), 5582-5602.

[38] Chen, H., Song, J., Wu, C., Du, B., & Yokoya, N. (2023). Exchange means change: An unsupervised single-temporal change detection framework based on intra-and inter-image patch exchange. ISPRS journal of photogrammetry and remote sensing, 206, 87-105.

[39] Shu, Q., Pan, J., Zhang, Z., & Wang, M. (2022). MTCNet: Multi-task consistency network with single temporal supervision for semi-supervised building change detection. International Journal of Applied Earth Observation and Geoinformation, 115, 103110.

[40] Du, Q., Peng, J., Chen, X., He, Q., He, L., Nie, Q., ... & Wang, C. (2024). Single-temporal supervised remote change detection for domain generalization. arXiv preprint arXiv:2404.11326.

[41] de Gélis, I., Lefèvre, S., & Corpetti, T. (2023). Siamese KPConv: 3D multiple change detection from raw point clouds using deep learning. ISPRS Journal of Photogrammetry and Remote Sensing, 197, 274-291.

[42] 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), 448.

[43] Zhan, W., Cheng, R., & Chen, J. (2024). PGN3DCD: Prior-knowledge-guided network for urban 3D point cloud change detection. IEEE Transactions on Geoscience and Remote Sensing.

[44] Lu, J., Dai, C., Zhang, Z., Liu, X., Zhou, R., Ji, S., ... & Wang, H. (2025). Ms-DANet: Multi-scale Difference-aware Network for 3D Point Cloud Change Detection. IEEE Transactions on Geoscience and Remote Sensing.

[45] DJI. (2025). DJI Terra (Version 5.0.2) [Computer software]. https://enterprise.dji.com/dji-terra

[46] DJI. (2025). DJI Modify (Version 1.4.0) [Computer software]. https://enterprise.dji.com/modify

[47] Ravi, N., Reizenstein, J., Novotny, D., Gordon, T., Lo, W. Y., Johnson, J., & Gkioxari, G. (2020). Accelerating 3d deep learning with pytorch3d. arXiv preprint arXiv:2007.08501.

[48] Zhou, Q. Y., Park, J., & Koltun, V. (2018). Open3D: A modern library for 3D data processing. arXiv preprint arXiv:1801.09847.

[49] Blender Development Team. (2022). Blender (Version 3.1.0) [Computer software]. https://www.blender.org

[50] Schonberger, J. L., & Frahm, J. M. (2016). Structure-from-motion revisited. In Proceedings of the IEEE conference on computer vision and pattern recognition (pp. 4104-4113).