www.itcon.org - Journal of Information Technology in Construction - ISSN 1874-4753

MSFD-NET: A MULTISCALE SPATIAL FEATURE DESCRIPTOR NETWORK FOR SEMANTIC SEGMENTATION OF LARGE-SCALE BRIDGE POINT CLOUDS

SUBMITTED: September 2024
REVISED: November 2024
PUBLISHED: December 2024
EDITOR: Robert Amor
DOI: 10.36680/j.itcon.2024.040

M. Saeed Mafipour, Dr.-Ing.
Chair of Computational Modeling and Simulation, Technical University of Munich
ORCID: https://orcid.org/0000-0002-2076-8653
saeed.mafipour@gmail.com

Simon Vilgertshofer, Prof. Dr.-Ing.
HM Munich University of Applied Sciences
ORCID: https://orcid.org/0000-0003-4271-2076
simon.vilgerthsofer@hm.edu

André Borrmann, Prof. Dr.-Ing.
Chair of Computational Modeling and Simulation, Technical University of Munich
ORCID: https://orcid.org/0000-0003-2088-7254
andre.borrmann@tum.de

SUMMARY: *Digital Twins (DTs) provide a promising solution for bridge operation, thanks to their ability to mirror the physical conditions into a digital representation. At the core of the DTs is a geometric-semantic model. The modeling process for existing bridges, however, requires extensive manual effort. Given the high number of bridges in operation worldwide, there is an urgent need for automating this process. Available low-effort capturing methods, including laser-scanning and photogrammetry, generate raw point cloud data (PCD) that requires further processing to achieve a high-quality model. This paper focuses on the semantic segmentation of the PCD, which is the essential first step in an automated processing pipeline. A novel deep learning model, called multi-scale spatial feature descriptor network (MSFD-Net), is proposed for the semantic segmentation of PCD. The model is tested using the PCD of six bridges in Bavaria, Germany. The results show that MSFD-Net can automate semantic segmentation of bridges with mean accuracy (mAcc) of 98.29 % and mean intersection over union (mIoU) of 93.57 %.*

KEYWORDS: *Digital Twin, Bridge Information Modeling, Semantic Segmentation, Deep Learning, Point Cloud Data.*

REFERENCE: *M. Saeed Mafipour, Simon Vilgertshofer, André Borrmann (2024). MSFD-Net: A Multiscale Spatial Feature Descriptor Network for Semantic Segmentation of Large-Scale Bridge Point Clouds. Journal of Information Technology in Construction (ITcon), Vol. 29, pg. 914-934, DOI: 10.36680/j.itcon.2024.040*

1. INTRODUCTION

In the transportation system of most countries, there is a large stock of aging bridges that require substantial attention to enable long-term operation. The recent ASCE report card (ASCE, 2021) illustrates that the amount of structurally deficient bridges in poor condition (46,154 or 7.5 % of ca. 617,000) is increasing as the deterioration rate exceeds the rate of repair, rehabilitation, and replacement. The National Bridge Inspection Standards (NBIS) require transportation agencies to evaluate the structural status of existing bridges periodically over the structure's service life. Currently, most of the processes involved with the inspection, operation, and maintenance of bridges are insufficiently supported by digital methods, thus increasing the costs associated with bridge evaluation and quality assurance (Sacks et al., 2016, 2018a, 2018b). Hence, the corresponding transportation operators are currently faced with heightened pressure to provide automated mechanisms for efficient coverage of the existing bridges.

Bridge information modeling (BrIM) is a technique that generates the digital counterpart of bridges that represents the physical and functional characteristics of the structure. BrIM is capable of providing a visual management system for Accelerated Bridge Construction (ABC), Virtual Design and Construction (VDC), and structural analysis (McGuire et al., 2016; Jeong et al., 2017). BrIM can be employed in the pre-construction, under-construction, and post-construction phases to facilitate the operation and management processes of bridges (Marzouk and Hisham, 2012; Shim et al., 2011; Rashidi and Karan, 2018). A bridge information model can also be extended and expressed as a digital twin (DT) through an access point and a link to handle bidirectional updates between the physical and the digital counterparts (Brilakis et al., 2019). Besides inheriting all the features of a bridge information model, the DT of a bridge also reflects geometric and semantic changes to the structure and the damages and deteriorations occurring over time (Vilgertshofer et al., 2022). In the current practice, technical drawings are commonly the primary resource for creating the geometric model underlying BrIM and DT. Despite the feasibility of 3D modeling based on conventional technical drawings, the drawings of most older bridges are not digitized and, in the worst cases, are not even available. Also, the geometry of a bridge might change due to aging or undocumented alterations and can, therefore, differ from the existing technical drawings.

Photogrammetry and Terrestrial Laser Scanning (TLS) are two primary geodetic techniques that can capture existing bridges with comparatively low effort. Both methods generate point cloud data (PCD) that intuitively shows bridges' rough geometric and partial semantic information. PCD can be seen as an alternative or complement to drawings for the geometric modeling of bridges, especially in the operational phase of the structure (Mohammadi et al., 2021). However, the main challenge is that the resulting PCD needs to be enriched and abstracted to a geometric model fulfilling the requirements of the DT or BrIM. This process is performed manually and laboriously today, requiring effort that cannot be invested in thousands of bridges. To handle this challenge, this paper aims to automate the processing of PCD.

Semantic segmentation is an essential first step in the pipeline for the automated 3D geometric modeling of bridges from PCD (Mafipour et al., 2023). This process differentiates the point cloud of a bridge into separate point clouds of the bridge’s elements, thereby paving the way for further extraction of geometric and semantic features. Despite the existence of tools and software programs for manual semantic segmentation of bridge point clouds, conducting this step is still error-prone and costly. To address this issue, this paper proposes a novel deep learning model, coined Multiscale Spatial Feature Descriptor (MSFD-Net), to automate the semantic segmentation process of bridge point clouds. MSFD-Net consists of various modules to extract local and global features and provides a mechanism for describing the pair-wise relationships between points. It also expresses the features in multiple scales to leverage the low-level and high-level features in the prediction process of labels. Thanks to the random sub-sampling strategy in the subsequent layers of the network and a U-shaped autoencoder,

The key contributions of the paper are summarized as follows:

- Proposal of MSFD-Net, a deep learning model to automate the semantic segmentation process of bridge PCD.
- Introduction of various modules for effective spatial encoding of points based on local, global, and pair-wise features.
- The description of the extracted features in various scales to benefit from the low-level and high-level features in classifying points.

- Analysis of the performance of MSFD-Net in terms of prediction accuracy and training time.

This paper is structured as follows: Section 2 outlines related works in the semantic segmentation of point clouds. Section 3 provides an overview of MSFD-Net, introduces the modules employed in the model, and describes them in detail. Section 4 presents the dataset of bridge point clouds and the statistical metrics for validation of the model and configures the hyperparameters. Section 5 reports the conducted experiments on MSFD-Net and demonstrates a numerical and visual comparison of this model with a state-of-the-art deep learning model. Section 6 evaluates the proposed modules' impact on the network performance. The paper ends with a conclusion in Section 7, discussing the development of our research, the significant findings, including known limitations and possible generalizations, and topics for future research.

2. BACKGROUND AND RELATED WORK

Various methods have been proposed to automate semantic segmentation of the bridge point clouds. The proposed methods can be categorized as bottom-up, top-down, and deep learning-based. This section provides a short overview of these approaches.

2.1 Bottom-up semantic segmentation

The bottom-up methods transform the low-level features into high-level features and leverage them to generate a more complex system at successively higher levels (Borenstein and Ullman, 2008). The low-level features are generally the raw attributes of PCD, such as the x, y, and z coordinates of points and the RGB color codes. The high-level features are typically the surface normal, meshes, surface planes/patches, non-uniform B-Spline surfaces, and voxels (Palop et al., 2010; Sampath and Shan, 2009; Marton et al., 2009; Zhang and Tang, 2015; Dimitrov et al., 2016; Dimitrov and Golparvar-Fard, 2015). There are multiple algorithms that are capable of utilizing these features for primitive detection and model reconstruction. Region growing (RG), RANdom SAmple Consensus (RANSAC), Hough-Transform (HT), and Octree-Based (OB) can be mentioned as dominant algorithms in this category. Many of these methods have been devoted to generating surface-based primitives, especially planar surfaces, as they exist more commonly in PCD (Patraucean et al., 2015). However, they have also been extended to other primitive shapes, such as the sphere, cylinder, cone, and torus (Schnabel et al., 2007), which can be defined mathematically in a closed-form formulation. The bottom-up algorithms have been used principally or partly in bridges' geometric digital twinning process.

Truong-Hong and Lindenbergh (2022) proposed a cell- or voxel-based region growing (CRG/VRG) algorithm to extract the planar surfaces in the PCD of RC bridges. Qin et al. (2021) used the local and global density of points for semantic segmentation of bridge PCD and fitted cylindrical and cuboid shapes to the segmented point cloud of elements. Lee et al. (2020) detected planar surfaces using M-estimator SAmple Consensus (MSAC). They extracted the value of parameters for specific types of bridge decks by measuring the distance between each pair of planes.

2.2 Top-down semantic segmentation

Contrary to bottom-up, top-down methods start from a complex system and decompose it to subordinate systems or elements that are simpler to interpret (Kokkinos et al., 2006). Top-down approaches require hierarchical planning and deep insight into the system so that no coding can be started without a sufficient level of detail at least for some parts. The top-down approach is typically a heuristic approach based on a set of domain-specific information/criteria. This knowledge-based approach can leverage the existing information such as predefined constraints, parameters, object types/instances, and topological relationships to break down the initial complex model to a set of basic elements (Dore and Murphy, 2014; Pu and Vosselman, 2009; Koppula et al., 2011; Ahmed et al., 2014). These basic elements can be further refined in more detail and in many additional subsystem levels. A pioneering study using a top-down approach was REFAB (Reverse Engineering FeAture-Based)(Thompson and Owen, 1999), which uses parametric geometric constraints such as parallelism, connectivity, perpendicularity, and symmetry to convert points to mechanical solid models.

The top-down approach has also been used for the geometric digital twinning of bridges. Lu et al. (2019) proposed a heuristic top-down approach for semantic segmentation of RC bridges based on directional geometric cuts and relative distance of points. Girardet and Boton (2021) described a visual programming approach to foster the parametric modeling of bridges. Zhao et al. (2019) defined a heuristic algorithm using the horizontal histogram of density and a support vector machine (SVM) for portioning structural elements in RC bridges. Yan and Hajjar (2021) proposed a top-down heuristic method for semantic segmentation of steel girder bridges through the existing geometric and topological constraints in these bridges. Pan et al. (2019) described a top-down method based on graph construction and combined it with a rule-based classification algorithm to segment the main elements of heritage bridges.

2.3 Deep learning-based semantic segmentation

Following the breakthrough results of applying deep learning (DL) models to images, there has been a strong interest to adapt such models to 3D geometric data such as meshes or point clouds. Contrary to images, a raw point cloud is an unstructured dataset without any underlying grid. Thus, its features cannot be extracted simply by passing through the convolution and pooling layers (Wang et al., 2019). To address this issue, the initial DL models mostly relied on transforming input point clouds to either multi-view images or volumetric data that are readable by 2D and 3D convolutions (Zhou and Tuzel, 2018; Xie et al., 2015; Su et al., 2015). However, the complexity and memory issues arising from the intermediate representations restricted the capability of these models.

PointNet (Qi et al., 2017a) is the first 3D DL model with the ability to process and extract features from points directly. This architecture uses a transformer to make the model invariant to rigid transformations and a symmetric function to achieve permutation stability to conquer the unordered point cloud input. From this model onward, 3D DL architectures have been developed in three parts: sampling strategy, feature extraction, and encoding-decoding process (Landrieu and Simonovsky, 2018; Qi et al., 2017b; Zhao et al., 2021; Thomas et al., 2019; Lin et al., 2020; Li et al., 2020; de G´elis et al., 2023; Shinde et al., 2021). Thanks to the sub-sampling and up-sampling in subsequent layers of a U-shaped autoencoder (Ronneberger et al., 2015), the recent architectures are even capable of inferring per-point semantics for large-scale point clouds (Hu et al., 2020). The encoder of these models has been also used for real-time semantic segmentation and model-to-cloud fitting of primitive shapes (Li et al., 2019). Most recently, DL models have been also employed in the geometric digital twinning of bridges (Jing et al., 2022; Mafipour et al., 2022).

Hu et al. (2021) employed a multi-view convolutional neural network (CNN) to extract features from photogrammetry and to link it with a multi-layer perceptron (MLP) to segment the point cloud of a bridge. Lee et al. (2021) added contextual features by kd-tree and K-nearest neighbors (KNN) search to PointNet (Qi et al., 2017a) and deep graph convolutional network (DGCNN) architectures (Wang et al., 2019) and improved the performance of these models for the semantic segmentation of bridges. Xia et al. (2022) used the local reference frame (LRF) of points and proposed a local descriptor to extract the features of points and to segment the PCD of bridges. Jing et al. (2022) augmented the point cloud dataset of bridges with synthetic point clouds and proposed Bridge-Net as a 3D DL model for semantic segmentation of arch bridge point clouds. They also used RANSAC for model-to-cloud fitting and to extract the parameters' values. Yang et al. (2022) also augmented the point cloud dataset of bridges with synthetic data and developed a DL model based on the super point graph (SPG) architecture (Landrieu and Simonovsky, 2018) for semantic segmentation of bridge PCD.

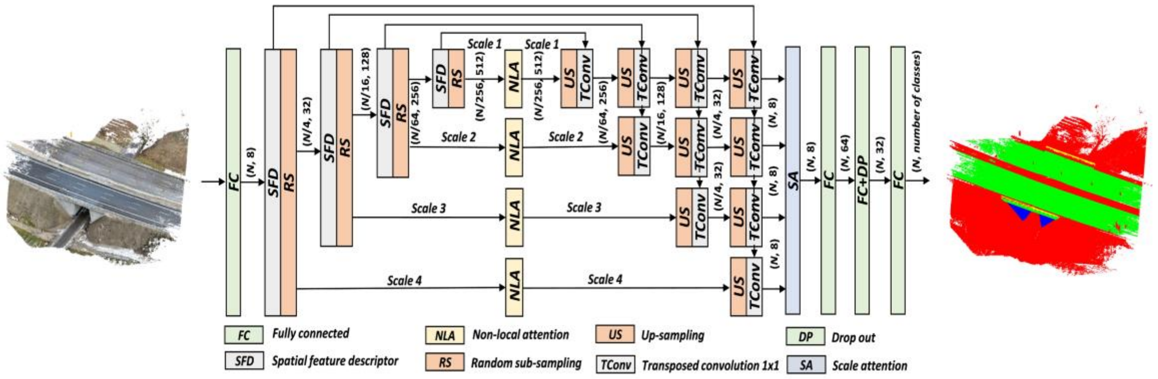


Figure 1: Architecture of MSFD-Net.

However, these approaches are limited as they can be equipped with cutting-edge AI modules/layers such as non-local attention, self-attention, and contextual attention for ecient encoding of points. Also, most existing models only encode local neighborhoods and cannot provide the existing spatial relationships between points located far from each other.

3. MSFD-NET ARCHITECTURE

MSFD-Net has been designed to capture the local and global spatial relations among points and model the existing patterns in various scales. It requires the coordinates of the points (x,y,z) and their corresponding normal vectors (nx, ny, nz) as input to determine the semantic label of the bridge point clouds. Other features, such as RGB color codes, density, and even learned latent features, can also be added and processed by this network. As shown in Figure 1, the architecture of MSFD-Net is basically an auto-encoder (U-Net) expressed in several scales. The encoder of the model includes a spatial features descriptor (SFD) module followed by random sub-sampling in sequential layers. Leveraging the sub-sampling and aggregation mechanisms, MSFD-Net can process large-scale point clouds.

The SFD module receives point features to generate a geometric representation of the local spatial neighborhoods. It also aggregates the generated local features and global features and increases the receptive field of points. The bottleneck is a non-local attention module providing global awareness about the spatial features in the latent space. The decoder has also been expressed in multi-scales to fuse the high-level feature map of points with the basic features generated by the initial and intermediate layers. The first scale of this model is a plain decoder collecting features from all the layers of the network, while the next scales ignore some of the intermediate blocks to provide a better perception of the initial layers. The decoded spatial features are attentively fused in the end to generate a feature map from all the scales. These modules are further described in the following sections.

3.1 Spatial features descriptor (SFD)

SFD aims to extract and learn the spatial geometric features of points. It also combines the local geometric features with the global input features to provide a more comprehensive description of the scene. The framework of the SFD has been inspired by the local feature aggregation module proposed in RandLA-Net (Hu et al., 2020). However, its spatial encoder is different and consists of two blocks named local spherical representation (LSpR) and local surface representation (LSuR). The details of these blocks are elaborated in the next sections, then, a general overview of the feature aggregation module is provided.

3.1.1 Local spherical representation (LSpR)

Existing point-based models generally adopt Cartesian coordinates as the input, while the relative position of points is highly sensitive to the possible transformations and normalization in this coordinate system. On the contrary, the spherical coordinate system can simply represent the position of a point by two angles limited in the range of [0,2π] and a radius/distance. Figure 2a shows the local neighborhood of the query point pi and its K neighbors $\{p^1{}_i,p^2{}_i,\ldots,p^K{}_i\}$ obtained by the K-nearest neighbors (KNN) algorithm. The relative position of pi with respect to its neighbors can be expressed in the spherical coordinate system $(r_i{}^k,\phi^k{}_i,\theta_i{}^k)$ as below:

$$r_i^k = \sqrt{x_i^{k^2} + y_i^{k^2} + z_i^{k^2}} \tag{1}$$

$$\phi_i^k = arctan(\frac{y_i^k}{x_i^k}) \tag{2}$$

$$\theta_i^k = arctan(\frac{z_i^k}{\sqrt{x_i^{k^2} + y_i^{k^2}}}), \tag{3}$$

where (x_i^k, y_i^k, z_i^k) are the relative coordinates of p_i^k in the Cartesian coordinate system with center p_i.

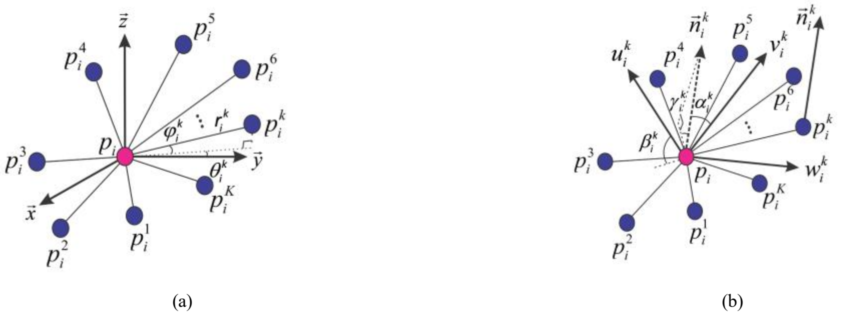


Figure 2: Local representation of a neighborhood: (a) local spherical representation (LSpR); (b) local surface representation (LSuR).

3.1.2 Local surface representation (LSuR)

The generated features by LSpR describe the relative position of points in a local neighborhood. However, they can not provide adequate information about the representative surface by the points. To handle this problem, a local Darboux frame can be defined, and the underlying surface on which the points are positioned is approximated (Rusu et al., 2008). A Darboux frame is a dynamic/moving frame constructed on a surface. It is the analog of the Frenet–Serret frame and can represent the curvature, normal curvature, and relative torsion of the surface. Figure 2b illustrates the query point p_i with the normal vector $\vec{n}_i$ as well as the neighboring points $\{p_i^1, p_i^2, ..., p_i^K\}$ and normals $\{\vec{n}^1{}_i, \vec{n}^2{}_i, ..., \vec{n}^K{}_i\}$. A Darboux frame with the axes $(\vec{u}^k{}_i, \vec{v}_i^k, w_i^{\vec{k}})$ can be defined for each pair of $(p_i, p^k{}_i)$ as below (Rusu et al., 2008):

$$\vec{u}_i^k = \vec{n}_i, \quad \vec{v}_i^k = \vec{u}_i^k \times \frac{p_i^k - p_i}{||p_i^k - p_i||_2}, \quad \vec{w}_i^k = \vec{u}_i^k \times \vec{v}_i^k, \qquad (4)$$

where $||.||_2$ is the L2 norm of the vectors connecting the centroid point to the neighboring points.

Based on the defined Darboux frame, the dependencies between each pair (n_i, n_i^k) can be defined by three angles α_i^k, β_i^k, andγ_i^k as follows (Rusu et al., 2008):

$$\vec{\alpha}_i^k = \vec{v}_i^k . n_i^k, \quad \vec{\beta}_i^k = \vec{u}_i^k . \frac{p_i^k - p_i}{||p_i^k - p_i||_2}, \quad \vec{\gamma}_i^k = arctan(\frac{\vec{w}_i^k . \vec{n}_i^k}{\vec{u}_i^k . \vec{n}_i^k}), \qquad (5)$$

3.1.3 Feature aggregation

As shown in Figure 3, the SFD module receives a point cloud containing N points, each described by d features (coordinates, RGB color, etc.), i.e. $\mathsf{F} = \{f_1, ..., f_i, ..., f_N\}$, where $\mathsf{F} \in \mathsf{R}^{N \times d}$. This feature set is passed through a shared MLP (M) to provide a global description of the input features ($\mathsf{M}(f_i)$). In parallel, the neighboring points are gathered by a simple KNN and the global features are distributed over the local neighborhoods (f_i^k). The neighboring points and normals are also sent to the LSpR and LSuR blocks to generate the local features. These features are concatenated, and a shared MLP is applied as below:

$$r_i^k = \mathcal{M}(r_i^k \oplus \phi_i^k \oplus \theta_i^k \oplus \vec{\alpha}_i^k \oplus \vec{\beta}_i^k \oplus \vec{\gamma}_i^k), \qquad (6)$$

where $\oplus$ is the concatenation operation.

The feature set r_i^k has been mostly defined based on rotation angles that, in turn, aid the network in learning local features and achieving more robust performance in practice. The local features r_i^k and the global features f_i^k are concatenated eventually ($f_i^k \oplus r_i^k$) to generate the new set $\hat{F}_i = \{\hat{f}_i^1, ..., \hat{f}_i^k, ..., \hat{f}_N^K\}$ describing the spatial features of points in each neighborhood. Symmetric functions such as mean and max can be applied to summarize the features of the neighborhoods.

However, to emphasize the more important features, the weighted summation of features is calculated while the weights are also learnable parameters for the network. This process, which is called attentive pooling, can be described as (Hu et al., 2020):

$$\tilde{f}_i = \sum_{k=1}^{K} (\hat{f}_i^k . g(\hat{f}_i^k, W)), \tag{7}$$

where $g()$ consists of a 1×1 convolution layer followed by a softmax to generate attention scores, and W is the learnable weights of the shared MLP. Contrary to RandLA-Net, which generates the weights W through a fully connected layer, MSFD-Net uses a convolution layer. This simple step can eliminate the need for changing the shape of tensors and speed up the performance of the network to a large extent.

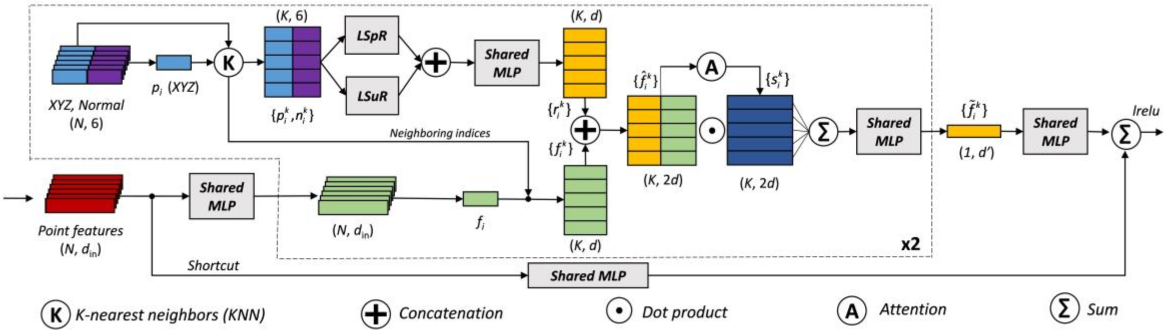


Figure 3: Spatial feature descriptor (SFD).

The SFD module leads to a set of features that encapsulates the geometric and semantic features of points. To collect residuals and preserve the lowersemantic information incoming to the module, a skip connection is used and two sets of features are stacked in each layer. As a result, the receptive field of points increases, and the generated features can be propagated over the neighborhoods.

3.2 Global awareness

The SFD module represents every point based on a set of local and global features collected from its local neighborhood. It also stacks the features and propagates them to the surrounding neighborhoods to increase the receptive fields. To preserve the computation efficiency of the model, the number of stacks is generally limited and cannot be increased. This limitation gradually reduces the impact radius of points, and as a result, the pairwise dependencies (global awareness) cannot be expressed completely for the points placed at large distances.

Non-local attention is a module first proposed to encode the spatial dependencies among pixels in images (Wang et al., 2018). This block has also been recently used in the spatial encoding of point clouds as well (Liu et al., 2020). Non-local networks generally consist of three 1 × 1 convolution layers to provide different representations of the input in the latent space of the problem. As shown in Figure 4, the non-local attention module of MSFD-Net receives the input feature map $F_{in} \in \mathrm{R}^{N\times C}$ and generates the output feature map $F_{out} \in \mathrm{R}^{N\times C}$ that contains not only the initial inputs but also the spatial dependencies.

This non-local attention module only generates two representations, F1 and F2, through two shared MLPs containing a 1×1 convolution layer followed by batch normalization and a Rectified Linear Activation (ReLU) function. These convolution layers generate a single feature for each point representing its spatial position in the latent space, i.e., F1, F2, $\in \mathrm{R}^{N\times 1}$. The batch normalization layer provides more stability, and the ReLU activation function adds non-linearity and limits the weights to positive values. The first feature map, F1, is multiplied by the transposed of the second feature map, F2, and the attention matrix $\tilde{F} = F_1 \otimes F_2^T$ (where $\tilde{F} \in \mathrm{R}^{N\times N}$) is obtained that encapsulates the pairwise relationship between points. This matrix is passed through a softmax function to be normalized and further multiplied element-wise by the input feature map $F_{in,}$ leading to the feature map $\hat{F} = Softmax(\tilde{F}) \odot F_{in}$. To tune the impact of $\hat{F}$ on the input feature map F_{in}, it is passed through a 1 × 1 convolution layer, and the resulting feature map is summed with F_{in} to generate the output feature map F_{out}. This implementation of the non-local attention module can provide more stability as the attention matrix $F^{\sim}$ is directly multiplied by the input feature map F_{in}. It also provides a more efficient mechanism to alleviate or augment the pair-wise relationships through a convolution layer before the final summation.

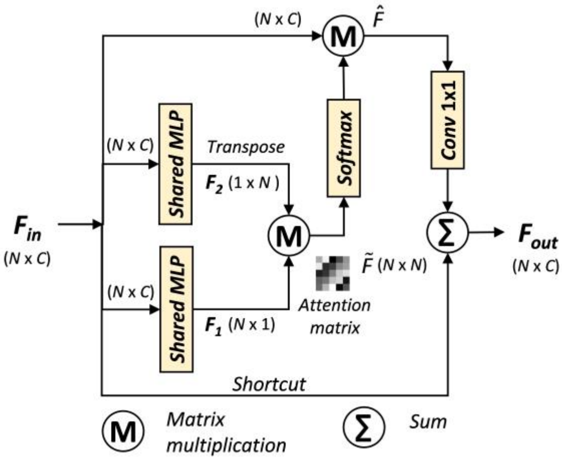


Figure 4: Non-local attention layer.

3.3 Multi-scale network fusion

The decoder of MSFD-Net consists of different scales, and the number of scales changes with respect to the number of layers. These scales provide various feature maps that not only contain high-level features but also low-level features (Qiu et al., 2021; Liu et al., 2020). As shown in Figure 1, the first scale of MSFD-Net collects the extracted features from all the layers, while the next scales ignore some of the intermediate SFD blocks to reduce the level of details. To increase the global awareness of points, the extracted features from all the scales are passed through the non-local attention block. The resulting feature map is concatenated with the extracted features from the corresponding encoder layer through a skip link. These features are up-sampled by nearest interpolation and then passed through a 1 × 1 transposed convolution layer. To emphasize the more important scales in the learning process of MSFD-Net, the weighted summation of the feature maps is calculated while the weights are also learnable parameters. Given the feature maps F = $\{F_1,...,F_i,...F_S\}$, Fi $\in \mathrm{R}^{N\times d}$, where S is the total number of scales, N is the number of points, and d is the number of features in the last layer, the weighted summation of the feature maps can be calculated as follows:

$$\tilde{F} = \sum_{i=1}^{S} (\hat{F}_i . g(F_i, W)), \tag{8}$$

where $F^{\sim}$ is the final feature map, $g()$ is a function consisting of a 1×1 convolution layer followed by a softmax to generate scale attention scores and W is the learnable weights of the shared MLP.

RandLA-Net has approximately 1.2 million parameters to be trained. With 700.000 parameters, MSFD-Net has a significantly lower number of parameters. The network achieves efficient segmentation by reducing the number of parameters significantly compared to other deep learning models used for similar tasks. This efficiency is due to the use of local feature aggregation, random point sampling, and efficient attention mechanisms, which help process large point clouds while maintaining performance and minimizing memory and computation requirements.

4. EXPERIMENTAL SETUP

The point cloud data of six single-span RC highway bridges, located in Bavaria, Germany, is used for evaluating the proposed model. This dataset has been acquired through UAV photogrammetry and processed by Structure from Motion (SfM). As shown in Figure 5, the samples comprise a bridge deck, abutments, railings, and background; thus, these four classes are considered for semantic segmentation. For comparing the performance of the new architecture with that of an existing one, RandLA-Net (Hu et al., 2020), a well-known deep learning model, is also trained on the bridge samples.

4.1 Data preparation

The raw point cloud of the bridges contains *x,y,z* coordinates, and RGB color codes. To reduce the processing load, each sample is subsampled by a uniform grid subsampling method with a grid size of 5 cm which is a good compromise between accuracy and reducing the amount of data in this scenario. To process the point clouds with MSFD-Net, normals are also calculated through a kd-tree and KNN search algorithm. Note that the normals are only used in the LSuR block of the network. All the bridge samples are translated to the origin of the coordinate system and all the input values are normalized in the range of zero to one. Due to the different number of points in each class (imbalanced dataset), class weights are computed based on the number of points over all the training samples in each class. Subsequently, a higher weight is assigned to the classes with a lower number of points and vice versa.

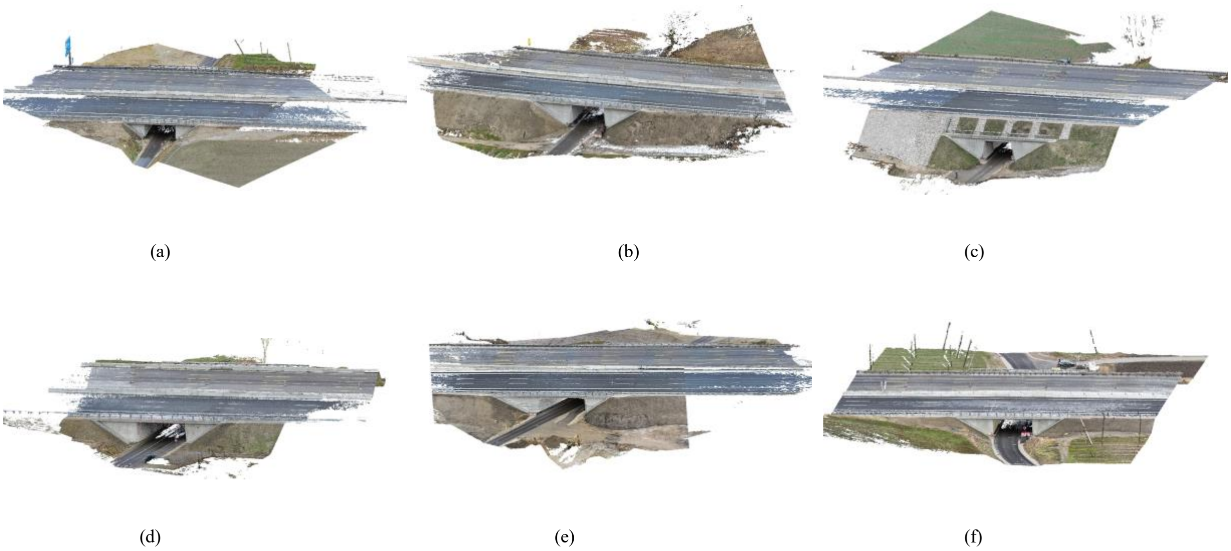


Figure 5: Photogrammetric PCD of six single-span RC bridges.

4.2 Validation method

To evaluate the performance of the model completely, extensive tests are conducted on MSFD-Net and RandLA-Net (Hu et al., 2020) simultaneously through the leave-one-out cross-validation (LOOCV) method. Six folds are adopted for training each model. One sample is held out for testing in every fold, and the models are trained on the five remaining samples. Accuracy (Acc) and intersection over union (IoU) are reported for each class of the unseen sample. Assuming the number of *N* classes, the Acc and IoU of the class *i* can be calculated as below:

$$Acc_i = \frac{TP_i}{TP_i + FN_i} \tag{9}$$

$$IoU_i = \frac{TP_i}{TP_i + FN_i + FP_i}, \tag{10}$$

where *TP*, *FN*, and *FP* are the total number of true positive, false negative, and false positive predictions by the model, respectively.

In addition to the class-wise evaluation, the overall performance of the models can be summarized through mean accuracy (mAcc) and mean intersection over union (mIoU). These metrics are simply the average of Acc and IoU over classes as follows:

$$mAcc = \frac{1}{N}\sum_{i=1}^{N} Acc_i \quad , \tag{11}$$

$$mIoU = \frac{1}{N}\sum_{i=1}^{N} IoU_i \quad , \tag{12}$$

where *N* is the total number of classes and *i* refers to points.

4.3 Hyperparameters

MSFD-Net and RandLA-Net both require a set of hyperparameters to be configured. These parameters include the number of sampling layers and their ratio, the number of neighbors for KNN, and the dimension of features in the input layer. The remaining parameters, such as batch size, the maximum number of points in each batch, and the learning rate, are associated with the training phase of the model. All these hyperparameters are obtained through an empirical process, and the values leading to the best results are reported. Both of the models showed their best performance with the same value of hyperparameters. Four sampling layers with a ratio of 1/4 are considered so that only 25 % of the points in each layer are retained. 16 neighbors are selected for the KNN algorithm, and a number of eight features are adopted for the input layer. A batch size of 2 with a maximum of 60,000 points with 100 steps per epoch is considered so that the samples can be evaluated more precisely. The models are trained by the Adam optimizer (Kingma and Ba, 2014) algorithm with a learning rate of 0.001 for 300 epochs.

5. SEMANTIC SEGMENTATION RESULTS

MSFD-Net and RandLA-Net are trained on a single GPU (RTX 3080) with 16 GB RAM. The performance of both models is tested on exactly the same unseen samples through the LOOCV method and their Acc, IoU, mAcc, and mIoU are compared. Both of the models need to be capable of processing largescale point clouds as most of the samples still have more than two million points after the initial uniform grid sub-sampling ($N \geq 2 \times 10^6$).

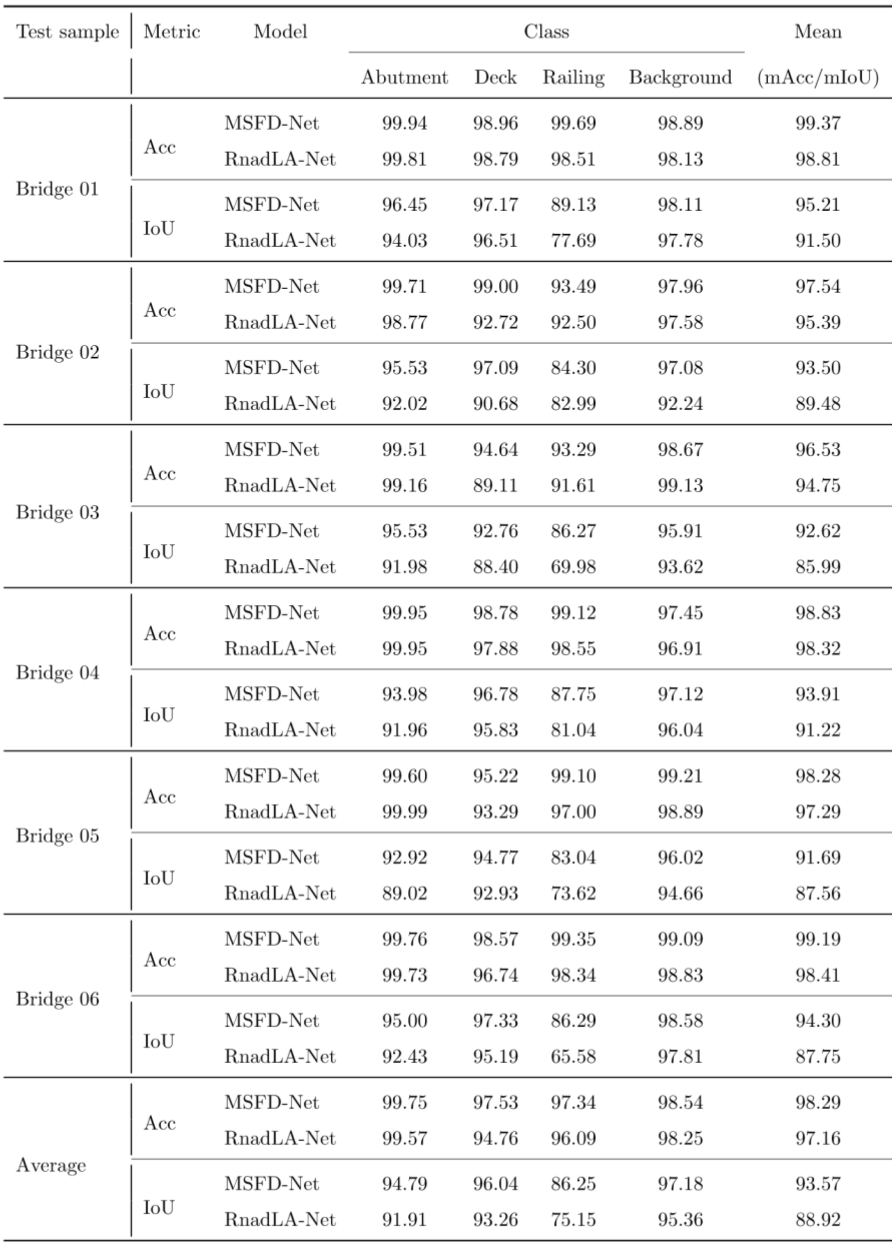
Table 1 demonstrates the comparative results of MSFD-Net (ours) and RandLA-Net (Hu et al., 2020) in terms of Acc and IoU per class throughout the LOOCV method. The last column of this table shows the value of mAcc and mIoU resulting from the models in each test. The last row also illustrates the average of all the tests describing the overall performance of the models in the LOOCV test. Considering this row, MSFD-Net has been capable of gaining the overall mAcc 98.29 % and mIoU 93.57 %. RandLA-Net has also obtained the mAcc 97.16 % and mIoU 88.92 %.

This conveys that MSFD-Net has improved the overall value of mAcc by Δ = 1.12 % and mIoU by Δ = 4.65 %. This row also shows a summary of the resulting values of Acc and IoU per class. As can be seen, MSFD-Net has obtained higher values in all the classes. The Acc of classes abutment, deck, railing, and background has increased from 99.57 % to 99.75 % (Δ = 0.18 %), 94.76 % to 97.53 % (Δ = 2.77 %), 96.09 % to 97.34 % (Δ = 1.25 %), and

98.25 % to 98.54 % (Δ = 0.30 %), respectively. Their IoU has also enhanced from 91.91 % to 94.79 % (Δ = 2.88 %), 93.26 % to 96.04 % (Δ = 2.78 %), 75.15 % to 86.25 % (Δ = 11.10 %), and 95.36 % to 97.18 % (Δ = 1.83 %), respectively.

The lower values of Acc and IoU obtained for the class *Railing* show that it has been the most difficult class for both of the models to predict. This is mainly due to the lower number of points this class has compared to the other classes. Although applying weights can solve this problem to some extent, the models would have a lower chance to observe and learn the spatial features of the local neighborhoods belonging to this class. MSFD-Net has also obtained the highest improvement for the class railing implying this model can learn the local neighborhoods more precisely, thanks to its LSpR and LSuR modules. These modules aid the model to highlight the underlying surface at places where the railings are connected to the bridge deck. Thus, the model can recognize the intersection parts and segment the elements more properly.

Table 1: Comparative results of MSFD-Net and RandLA-Net in LOOCV (%).

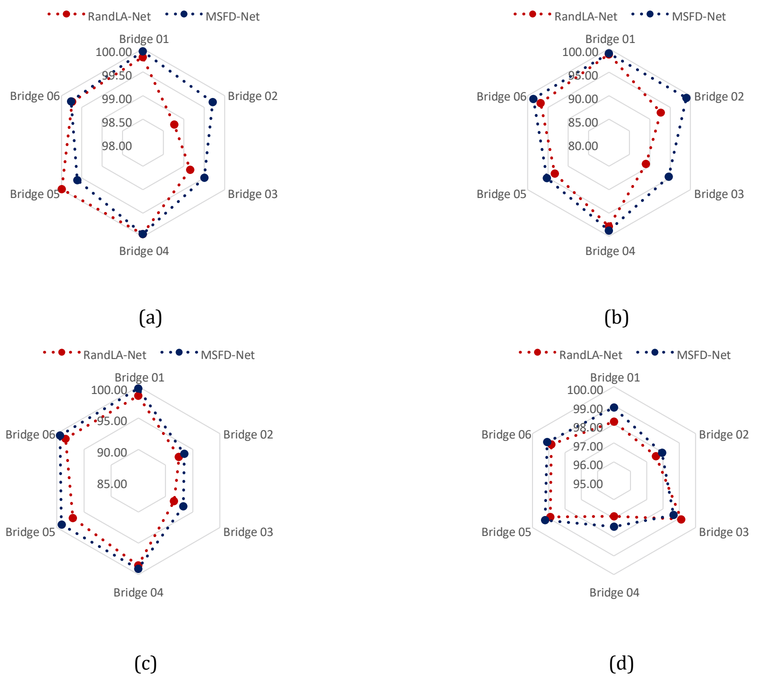


Test sample	Metric	Model	Class				Mean
			Abutment	Deck	Railing	Background	(mAcc/mIoU)
Bridge 01	Acc	MSFD-Net	99.94	98.96	99.69	98.89	99.37
		RnadLA-Net	99.81	98.79	98.51	98.13	98.81
	IoU	MSFD-Net	96.45	97.17	89.13	98.11	95.21
		RnadLA-Net	94.03	96.51	77.69	97.78	91.50
Bridge 02	Acc	MSFD-Net	99.71	99.00	93.49	97.96	97.54
		RnadLA-Net	98.77	92.72	92.50	97.58	95.39
	IoU	MSFD-Net	95.53	97.09	84.30	97.08	93.50
		RnadLA-Net	92.02	90.68	82.99	92.24	89.48
Bridge 03	Acc	MSFD-Net	99.51	94.64	93.29	98.67	96.53
		RnadLA-Net	99.16	89.11	91.61	99.13	94.75
	IoU	MSFD-Net	95.53	92.76	86.27	95.91	92.62
		RnadLA-Net	91.98	88.40	69.98	93.62	85.99
Bridge 04	Acc	MSFD-Net	99.95	98.78	99.12	97.45	98.83
		RnadLA-Net	99.95	97.88	98.55	96.91	98.32
	IoU	MSFD-Net	93.98	96.78	87.75	97.12	93.91
		RnadLA-Net	91.96	95.83	81.04	96.04	91.22
Bridge 05	Acc	MSFD-Net	99.60	95.22	99.10	99.21	98.28
		RnadLA-Net	99.99	93.29	97.00	98.89	97.29
	IoU	MSFD-Net	92.92	94.77	83.04	96.02	91.69
		RnadLA-Net	89.02	92.93	73.62	94.66	87.56
Bridge 06	Acc	MSFD-Net	99.76	98.57	99.35	99.09	99.19
		RnadLA-Net	99.73	96.74	98.34	98.83	98.41
	IoU	MSFD-Net	95.00	97.33	86.29	98.58	94.30
		RnadLA-Net	92.43	95.19	65.58	97.81	87.75
Average	Acc	MSFD-Net	99.75	97.53	97.34	98.54	98.29
		RnadLA-Net	99.57	94.76	96.09	98.25	97.16
	IoU	MSFD-Net	94.79	96.04	86.25	97.18	93.57
		RnadLA-Net	91.91	93.26	75.15	95.36	88.92

Comparing the values of mAcc and mIoU resulting from each test sample also shows the superior performance of MSFD-Net. The value of mAcc has improved in bridge samples 01 to 06 from 98.81 % to 99.37 % (Δ = 0.56 %), 95.39 % to 97.54 % (Δ = 2.15 %), 94.75 % to 96.53 % (Δ = 1.77 %), 97.29 % to 98.28 % (Δ = 0.99 %), 98.41 % to 99.19 % to (Δ = 0.78 %), and 97.16 % to 98.24 % (Δ = 1.08 %), respectively. The value of mIoU has also increased from 91.50 % to 95.38 % (Δ = 3.88 %), 89.48 % to 93.50 % (Δ = 4.02 %), 85.99 % to 92.62 % (Δ = 6.62 %), 91.22 % to 93.91 % (Δ = 2.69 %), 87.56 % to 91.69 % (Δ = 4.13 %), and 87.75 % to 94.30 % (Δ = 6.54 %), respectively.

To evaluate the Acc and IoU per class of the test samples, the radar diagrams of both models are depicted in Figures 6 and 7. As shown in these figures, the radar diagram of MSFD-Net has spread more and is placed out of RandLA-Net's diagram. This demonstrates the higher Acc and IoU of MSFD-Net in classifying the points. Comparing the radar diagrams also shows that MSFD-Net has not seen sharp changes in prediction conveying the more stable performance of this network. Among the classes, the highest difference in Acc can be seen in the class deck, more specifically the bridge samples 2 and 3 with around 5 % difference. The IoU of the models has shown even a higher difference in Figure 7. This case is more notable for the classes abutment and railing where the difference between the diagrams is around 10 % for some of the samples.

The closest performance in terms of IoU can be seen in the class background. This can be due to the higher number of points in this class which increases the chance of learning. The summary of the diagrams can be seen in Figure 8 representing the mAcc and mIoU of the bridge samples. Despite the competitive performance in mAcc, MSFD-Net has significantly improved the value of mIoU in prediction.



(a)

(b)

(c)

(d)

Figure 6: Class-wise Acc of the models: (a) abutment; (b) deck; (c) railing; (d) background.

The numerical differences resulting from the models can be observed visually as well. Figure 9 illustrates a visual comparison between MSFD-Net and RandLA-Net in which the wrongly predicted parts have been shown in a circle. As can be seen, MSFD-Net has been more successful than RandLA-Net in predicting the correct label of the points. The prediction results of MSFD-Net throughout the LOOCV test have been shown in Figure 10 in which most of the points have been labeled correctly.

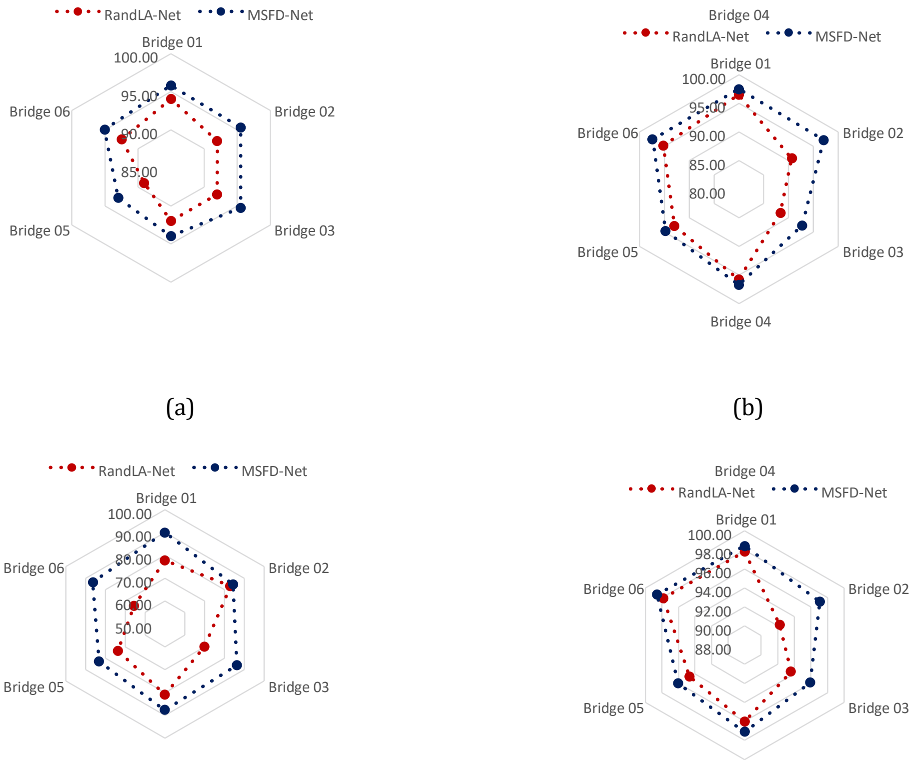


Figure 7: Class-wise IoU of the models: (a) abutment; (b) deck; (c) railing; (d) background.

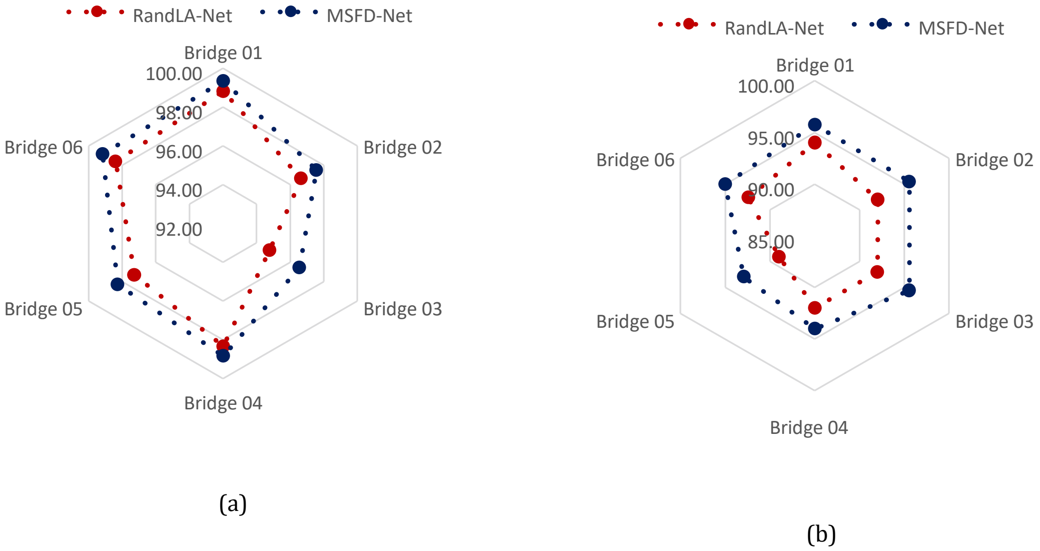


Figure 8: mAcc and mIoU of MSFD-Net and RandLA-Net shown by radar diagrams: (a) mAcc; (b) mIoU.

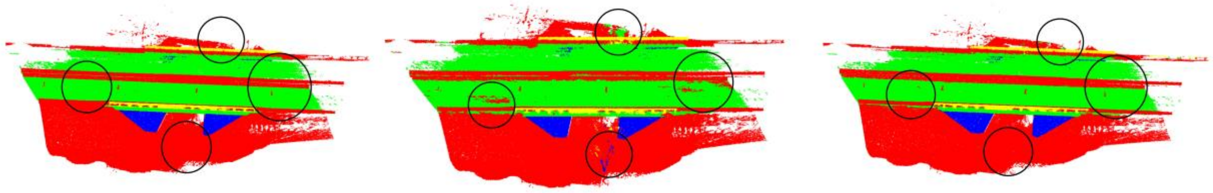


Figure 9: Visual comparison of MSFD-Net with RandLA-Net: ground truth (left); RandLA-Net (middle); MSFD-Net(ours) (Right).

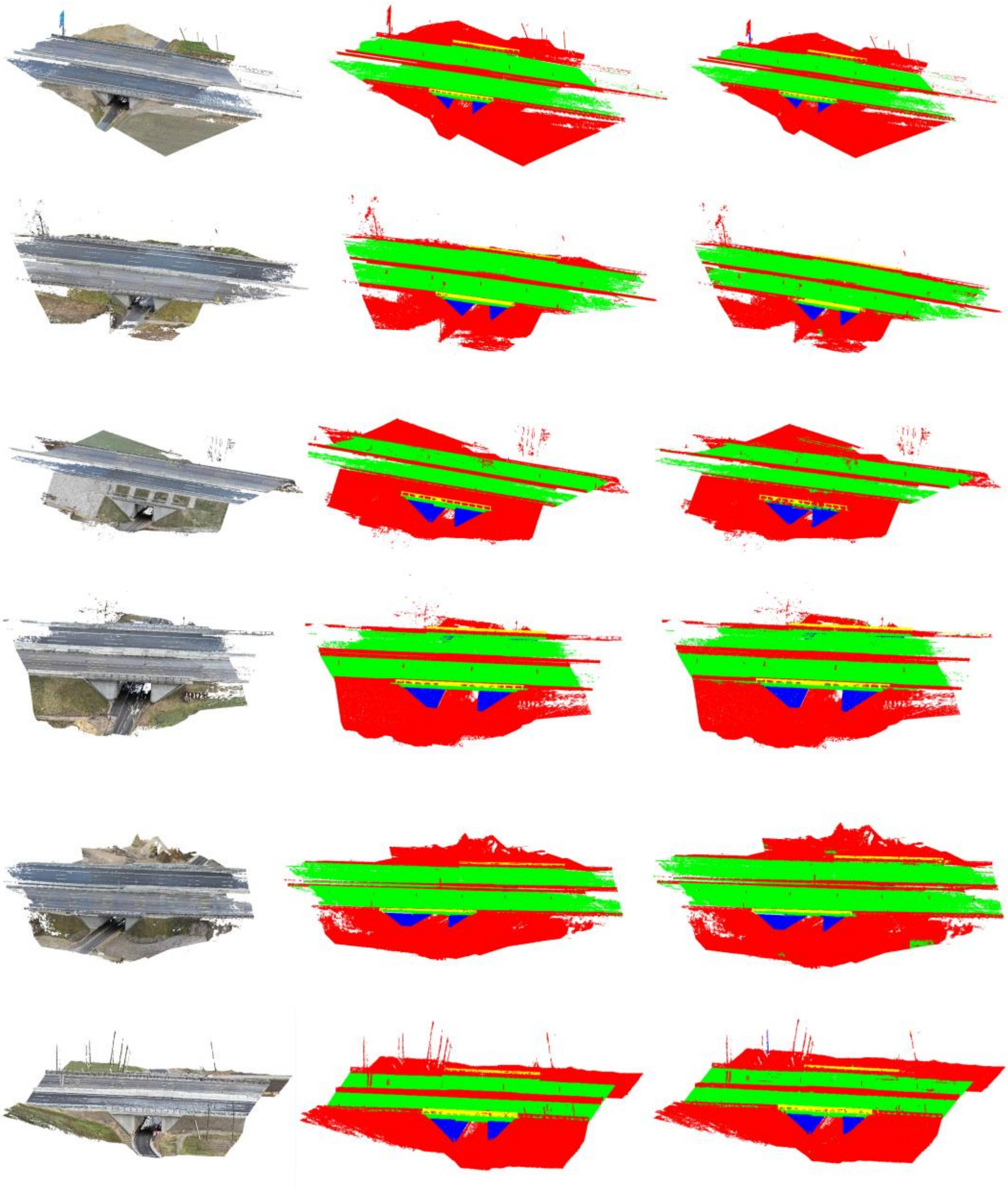
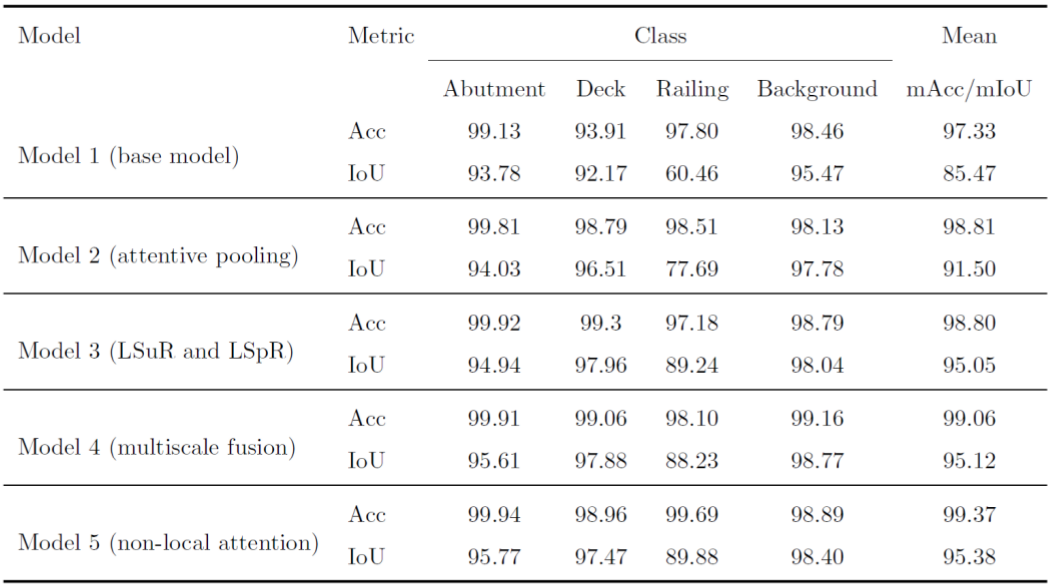


Figure 10: Prediction results of MSFD-Net on all the six samples: original point cloud (left); ground truth (middle); prediction (right). Class color: abutments(blue); railings(yellow); background(red); bridge deck(green).

6. PARAMETRIC STUDY

To analyze the impact of the proposed modules on the performance of the network, a base model is created and each module is added to this model. Then, the performance of the models is tested on an unseen bridge sample. In total five models are proposed to observe the effectiveness of the modules. The first model (Model 1) consists of an architecture similar to the first scale of MSFD-Net, however, with a plain shared MLP in the bottleneck and a spatial feature encoder similar to RandLA-Net but with a simple mean function for pooling. The second model (Model 2) is RandLA-Net and benefits from the attentive pooling layer instead of the mean function in the base model. Model 3 is different than model 2 in the encoding process, as it uses the proposed LSuR and LSpR modules in MSFD-Net. Model 4 is the expression of Model 3 in multiscales fused by the scale attention layer. Model 5 also contains the nonlocal attention module in the bottleneck and represents MSFD-Net. Considering the added modules to the based model, Model 2-5 evaluate the impact of the attentive pooling layer, LSuR as well as LSpR, multiscale fusion, and non-local attention layer, respectively.

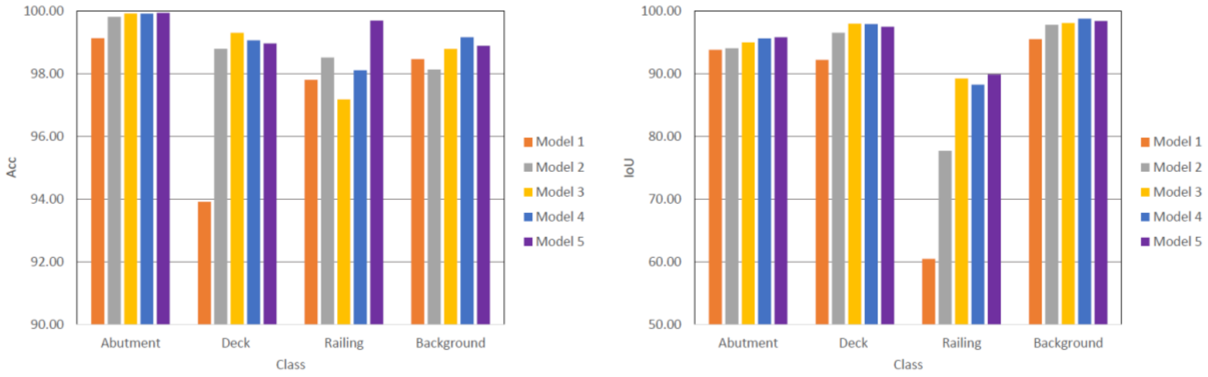
Table 2: The impact of the proposed modules on the performance of the base model.



Model	Metric	Class				Mean
		Abutment	Deck	Railing	Background	mAcc/mIoU
Model 1 (base model)	Acc	99.13	93.91	97.80	98.46	97.33
	IoU	93.78	92.17	60.46	95.47	85.47
Model 2 (attentive pooling)	Acc	99.81	98.79	98.51	98.13	98.81
	IoU	94.03	96.51	77.69	97.78	91.50
Model 3 (LSuR and LSpR)	Acc	99.92	99.3	97.18	98.79	98.80
	IoU	94.94	97.96	89.24	98.04	95.05
Model 4 (multiscale fusion)	Acc	99.91	99.06	98.10	99.16	99.06
	IoU	95.61	97.88	88.23	98.77	95.12
Model 5 (non-local attention)	Acc	99.94	98.96	99.69	98.89	99.37
	IoU	95.77	97.47	89.88	98.40	95.38

Table 2 shows the results of the models in terms of Acc, mAcc, IoU, and mIoU in semantic segmentation of the same unseen bridge sample (Bridge 01). Each row of the table represents a model and the module added to the prior model. As can be seen, the value of mAcc and mIoU has been increased from 97.33 % to 99.37 % (Δ = 2.05 %) and 85.47 % to 95.38 % (Δ = 9.91 %), respectively, after adding all the proposed modules to the base model (from Model 1 to Model 5). The results of the table show that adding attentive pooling has enhanced the value of mAcc and mIoU from 97.33 % to 98.81 % (Δ = 1.49 %) and 85.47 % to 91.50 % (Δ = 6.03 %), respectively. This implies that attentive pooling can provide a better mechanism for describing the local neighborhoods than a simple mean or max pooling function. Comparing the results of Model 2 with Model 3 demonstrates almost the same values of mAcc while the value of mIoU has been increased from 91.50 % to 95.05 % (Δ =3.54 %). This improvement shows that using LSuR and LSpR in the encoder of RandLA-Net can improve the value of mIoU. Note that the encoder of RandLANet encodes 10 features comprising the absolute coordinate of the neighboring points and the centroid point, the relative position of points in the Cartesian coordinate system, and the distance from the centroid point to the neighbors while MSFD-Net encodes only six features, five out of six have been expressed in angle. This conveys that the local neighborhoods can be expressed more properly with an even lower number of features. Model 4 shows the impact of multiscale fusion by Δ = 0.26 % improvement in the value of mAcc and a slight improvement in the value of mIoU by Δ = 0.08 %. This shows that low-level features generated in the initial and intermediate layers of the network can aid the network to learn the scene more accurately. Model

5 also represents the effectiveness of the non-local attention module. As can be seen, this module has also enhanced the mAcc and mIoU from 99.06 % to 99.37 % (Δ = 0.31 %) and 95.12 % to 95.38 % (Δ = 0.26 %).



(a)

(b)

Figure 11: Class-wise comparison of the models: (a) Acc; (b) IoU.

To evaluate the class-wise performance of the models, Figure 11 has been depicted. As can be seen, Model 5 (MSFD-Net) has shown a higher performance in terms of IoU and Acc than Model 2 (RandLA-Net) in all the classes. From Model 1 to 5 the improvement in the performance of the model can be seen in the classes, especially the classes of abutments and railing. From Model 3 onward, the performance of the models is more competitive, and each model has been able to perform very well in the prediction of a class. For instance, Model 4 with a simple shared MLP in the bottleneck has been capable of achieving the highest statistical metrics in the prediction of points belonging to the class Background. Nonetheless, MSFD-Net is still outperforming the models in most classes and in terms of mIoU and mAcc.

In has to be notited that the investigated bridges are not part of the same highway. However, highway bridges in Germany are, to a certain extent, standardized with respect to the construction types and general design. This fact results in similarities between the bridges, which is the main motivation for applying a machine-learning approach here. At same time, bridges 2 and 3 show deviations from these quasi-standards that result in a lower performance of the network.

The performance of the models can be compared in terms of training time as well. To this end, three metrics of the number of iterations per second (iter/sec), training time per epoch, and the total training time of the models are measured. Table 3 shows the obtained results for the models after training. As can be seen, MSFD-Net has achieved a faster performance than RandLA-Net based on all three metrics. This is mainly due to the used attentive pooling layer in MSFDNet whose weights are generated by a simple 1 × 1 convolution layer.

Table 3: Comparing the training time of RandLA-Net and MSFD-Net.

Model	Iter/sec	Training time per epoch (sec)	Total training time (min)
MSFD-Net	2.54	19.23	101.15
RandLA-Net	2.15	23.12	120.63

7. CONCLUSION

This paper presented a novel deep learning model, coined multiscale spatial feature descriptor network (MSFD-Net). It has been proposed to automate the semantic segmentation process of bridge point clouds. MSFD-Net benefits from a multiscale descriptor that not only highlights the high-level features of the points but also the low-level features generated in the initial and intermediate layers of the network. This model also represents the underlying surface of the local neighborhoods which in turn increases the accuracy of the model in practice. Comparing the results of this model with those of RandLA-Net (Hu et al., 2020) throughout the leave-one-out cross-validation (LOOCV) method shows the superior performance of MSFD-Net with 1.12 % improvement in the value of mean accuracy (mAcc) and 4.65 % in the value of mean intersection over union (mIoU). Evaluation of the modules and the required training time also shows the efficient and fast performance of the model in processing largescale point clouds. As a result, the proposed model can be used for automating the semantic segmentation of bridge point clouds. However, this model can be improved and, moreover, it could be combined with a module for creating the geometric model of the bridge elements. The employed subsequent random sampling in MSFD-Net results in the fast performance of this network in processing large-scale point clouds. Nonetheless, this sampling strategy might drop out important features in the training process of the model, especially in classes with a low number of points. In the future, a module is proposed to sample points based on the extracted features from the SFD layers. In doing so, the network can not only learn the spatial features but also recognize the points boosting the learning process. Also, the performance of the model will be tested on more complex bridges such as multi-span curved bridges with more classes to cover a wider variety of bridge types.

ACKNOWLEDGEMENTS

We thank the German Federal Ministry for Digital and Transport (BMDV) for funding this research in the scope of the TwinGen project.

REFERENCES

Turk Z (1991). Integration of existing programs using frames. Wagter H, Spoonamore J (ed.); The computer integrated future. CIB Seminar, 16-17 September, 1991. Eindhoven University of Technology, Calibre, The Netherlands (ISSN: 2706-6568), http://itc.scix.net/paper/w78-1991-22

Ahmed, M.F., Haas, C.T., Haas, R., 2014. Automatic detection of cylindrical objects in built facilities. Journal of Computing in Civil Engineering 28, 04014009. doi:http://dx.doi.org/10.1061/(ASCE)CP.1943-5487.0000329.

ASCE, 2021. ASCE's 2021 Infrastructure Report Card. Report. https://infrastructurereportcard.org/wp-content/uploads/2020/12/National_IRC_2021-report.pdf. Date Accessed 27 Oct 2024.

Borenstein, E., Ullman, S., 2008. Combined top-down/bottom-up segmentation. IEEE Transactions on pattern analysis and machine intelligence 30, 2109– 2125. doi:http://dx.doi.org/10.1109/TPAMI.2007.70840.

Brilakis, I., Pan, Y., Borrmann, A., Mayer, H.G., Rhein, F., Vos, C., Pettinato, E., Wagner, S., 2019. Built environment digital twining, International Workshop on Built Environment Digital Twinning

Dimitrov, A., Golparvar-Fard, M., 2015. Segmentation of building point cloud models including detailed architectural/structural features and mep systems. Automation in Construction 51, 32–45. doi:http://dx.doi.org/10.1016/j.autcon.2014.12.015.

Dimitrov, A., Gu, R., Golparvar-Fard, M., 2016. Non-uniform b-spline surface fitting from unordered 3d point clouds for as-built modeling. Computer-Aided Civil and Infrastructure Engineering 31, 483–498. doi:http://dx.doi.org/ 10.1111/mice.12192.

Dore, C., Murphy, M., 2014. Semi-automatic generation of as-built bim fa¸cade geometry from laser and image data. Journal of Information Technology in Construction (ITcon) 19, 20–46.

de Gelis, I., Lefevre, 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. doi:https: //doi.org/10.1016/j.isprsjprs.2023.02.001.

Girardet, A., Boton, C., 2021. A parametric bim approach to foster bridge project design and analysis. Automation in Construction 126, 103679. doi:http://dx.doi.org/10.1016/j.autcon.2021.103679.

Hu, F., Zhao, J., Huang, Y., Li, H., 2021. Structure-aware 3d reconstruction for cable-stayed bridges: A learning-based method. Computer-Aided Civil and Infrastructure Engineering 36, 89–108. doi:http://dx.doi.org/10.1111/ mice.12568.

Hu, Q., Yang, B., Xie, L., Rosa, S., Guo, Y., Wang, Z., Trigoni, N., Markham, A., 2020. Randla-net: Efficient semantic segmentation of large-scale point clouds, in: Proceedings of the IEEE/CVF Conference on

Computer Vision and Pattern Recognition, pp. 11108–11117. doi:http://dx.doi.org/ 10.1109/CVPR42600.2020.01112.

Jeong, S., Hou, R., Lynch, J.P., Sohn, H., Law, K.H., 2017. An information modeling framework for bridge monitoring. Advances in engineering software 114, 11–31. doi:http://dx.doi.org/10.1016/j.advengsoft.2017.05.009.

Jing, Y., Sheil, B., Acikgoz, S., 2022. Segmentation of large-scale masonry arch bridge point clouds with a synthetic simulator and the bridgenet neural network. Automation in Construction 142, 104459. doi:http://dx.doi.org/ 10.1016/j.autcon.2022.104459.

Kingma, D.P., Ba, J., 2014. Adam: A method for stochastic optimization. arXiv preprint arXiv:1412.6980 doi:https://doi.org/10.48550/ arXiv.1412.6980.

Kokkinos, I., Maragos, P., Yuille, A., 2006. Bottom-up & top-down object detection using primal sketch features and graphical models, in: 2006 IEEE Computer Society Conference on Computer Vision and Pattern Recognition, (CVPR'06), IEEE. pp. 1893–1900. doi:http://dx.doi.org/10.1109/CVPR. 2006.74.

Koppula, H., Anand, A., Joachims, T., Saxena, A., 2011. Semantic labeling of 3d point clouds for indoor scenes. Advances in neural information processing systems 24.

Landrieu, L., Simonovsky, M., 2018. Large-scale point cloud semantic segmentation with superpoint graphs, in: Proceedings of the IEEE conference on computer vision and pattern recognition, pp. 4558–4567. doi:http: //dx.doi.org/10.1109/CVPR.2018.00479.

Lee, J.H., Park, J.J., Yoon, H., 2020. Automatic bridge design parameter extraction for scan-to-bim. Applied Sciences 10, 7346. doi:http://dx.doi. org/10.3390/app10207346.

Lee, J.S., Park, J., Ryu, Y.M., 2021. Semantic segmentation of bridge components based on hierarchical point cloud model. Automation in Construction 130, 103847. doi:http://dx.doi.org/10.1016/j.autcon.2021.103847.

Li, L., Sung, M., Dubrovina, A., Yi, L., Guibas, L.J., 2019. Supervised fitting of geometric primitives to 3d point clouds, in: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, pp. 2652–2660. doi:http://dx.doi.org/10.1109/CVPR.2019.00276.

Li, X., Wang, L., Wang, M., Wen, C., Fang, Y., 2020. Dance-net: Densityaware convolution networks with context encoding for airborne lidar point cloud classification. ISPRS Journal of Photogrammetry and Remote Sensing 166, 128–139. doi:https://doi.org/10.1016/j.isprsjprs.2020.05.023.

Lin, Y., Vosselman, G., Cao, Y., Yang, M.Y., 2020. Active and incremental learning for semantic als point cloud segmentation. ISPRS journal of photogrammetry and remote sensing 169, 73–92. doi:https://doi.org/10.1016/j.isprsjprs.2020.09.003.

Liu, K., Gao, Z., Lin, F., Chen, B.M., 2020. Fg-net: Fast large-scale lidar point clouds understanding network leveraging correlated feature mining and geometric-aware modelling. arXiv preprint arXiv:2012.09439 doi:https:// doi.org/10.48550/arXiv.2012.09439.

Lu, R., Brilakis, I., Middleton, C.R., 2019. Detection of structural components in point clouds of existing rc bridges. Computer-Aided Civil and Infrastructure Engineering 34, 191–212. doi:http://dx.doi.org/10.1111/mice. 12407.

Mafipour, M., Vilgertshofer, S., Borrmann, A., 2022. Digital twinning of bridges from point cloud data by deep learning and parametric models, in: Proc. of European Conference on Product and Process Modeling 2022.

Mafipour, M. S., Vilgertshofer, S., & Borrmann, A. (2023). Automated geometric digital twinning of bridges from segmented point clouds by parametric prototype models. Automation in Construction, 156, Article 105101. https://doi.org/10.1016/j.autcon.2023.105101

Marton, Z.C., Rusu, R.B., Beetz, M., 2009. On fast surface reconstruction methods for large and noisy point clouds, in: 2009 IEEE international conference on robotics and automation, IEEE. pp. 3218–3223. doi:http: //dx.doi.org/10.1109/ROBOT.2009.5152628.

Marzouk, M., Hisham, M., 2012. Bridge information modeling in sustainable bridge management. pp. 457–466. doi:http://dx.doi.org/10.1061/ 41204(426)57.

McGuire, B., Atadero, R., Clevenger, C., Ozbek, M., 2016. Bridge information modeling for inspection and evaluation. Journal of Bridge Engineering 21, 04015076. doi:http://dx.doi.org/10.1061/(ASCE)BE.1943-5592. 0000850.

Mohammadi, M., Rashidi, M., Mousavi, V., Karami, A., Yu, Y., Samali, B., 2021. Quality evaluation of digital twins generated based on uav photogrammetry and tls: Bridge case study. Remote Sensing 13, 3499. doi:http: //dx.doi.org/10.3390/rs13173499.

Palop, J.J., Mucke, L., Roberson, E.D., 2010. Quantifying biomarkers of cognitive dysfunction and neuronal network hyperexcitability in mouse models of alzheimer's disease: depletion of calcium-dependent proteins and inhibitory hippocampal remodeling, in: Alzheimer's Disease and Frontotemporal Dementia. Springer, pp. 245–262. doi:http://dx.doi.org/10.1007/ 978-1-60761-744-0_17.

Pan, Y., Dong, Y., Wang, D., Chen, A., Ye, Z., 2019. Three-dimensional reconstruction of structural surface model of heritage bridges using uavbased photogrammetric point clouds. Remote Sensing 11, 1204. doi:http://dx.doi.org/10.3390/rs11101204.

Patraucean, V., Armeni, I., Nahangi, M., Yeung, J., Brilakis, I., Haas, C., 2015. State of research in automatic as-built modelling. Advanced Engineering Informatics 29, 162–171. doi:https://doi.org/10.1016/j.aei.2015.01. 001.

Pu, S., Vosselman, G., 2009. Knowledge based reconstruction of building models from terrestrial laser scanning data. ISPRS Journal of Photogrammetry and Remote Sensing 64, 575–584. doi:http://dx.doi.org/10.1016/j. isprsjprs.2009.04.001.

Qi, C.R., Su, H., Mo, K., Guibas, L.J., 2017a. Pointnet: Deep learning on point sets for 3d classification and segmentation, in: Proceedings of the IEEE conference on computer vision and pattern recognition, pp. 652–660.

Qi, C.R., Yi, L., Su, H., Guibas, L.J., 2017b. Pointnet++: Deep hierarchical feature learning on point sets in a metric space. Advances in neural information processing systems 30. doi:https://doi.org/10.48550/arXiv.1706.02413.

Qin, G., Zhou, Y., Hu, K., Han, D., Ying, C., 2021. Automated reconstruction of parametric bim for bridge based on terrestrial laser scanning data. Advances in Civil Engineering 2021, 1. doi:http://dx.doi.org/10.1155/ 2021/8899323.

Qiu, S., Anwar, S., Barnes, N., 2021. Semantic segmentation for real point cloud scenes via bilateral augmentation and adaptive fusion, in: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, pp. 1757–1767. doi:https://arxiv.org/abs/2103.07074v2.

Rashidi, A., Karan, E., 2018. Video to brim: Automated 3d as-built documentation of bridges. Journal of performance of constructed facilities 32, 04018026. doi:http://dx.doi.org/10.1061/(ASCE)CF.1943-5509.0001163.

Ronneberger, O., Fischer, P., Brox, T., 2015. U-net: Convolutional networks for biomedical image segmentation, in: Medical Image Computing and Computer-Assisted Intervention – MICCAI 2015: 18th International Conference, Munich, Germany, October 5-9, 2015, Proceedings, Part III 18, Springer. pp. 234–241.

Rusu, R.B., Blodow, N., Marton, Z.C., Beetz, M., 2008. Aligning point cloud views using persistent feature histograms, in: 2008 IEEE/RSJ international conference on intelligent robots and systems, IEEE. pp. 3384–3391. doi:http: //dx.doi.org/10.1109/IROS.2008.4650967.

Sacks, R., Kedar, A., Borrmann, A., Ma, L., 2016. SeeBridge information delivery manual (IDM) for next generation bridge inspection. ISARC. doi:http://dx.doi.org/10.22260/ISARC2016/0100.

Sacks, R., Kedar, A., Borrmann, A., Ma, L., Brilakis, I., Hu¨thwohl, P., Daum, S., Kattel, U., Yosef, R., Liebich, T., et al., 2018a. Seebridge as next generation bridge inspection: overview, information delivery manual and model view definition. Automation in Construction 90, 134–145. doi:http://dx.doi.org/10.1016/j.autcon.2018.02.033.

Sacks, R., Kedar, A., Borrmann, A., Ma, L., Brilakis, I., Hu¨thwohl, P., Daum, S., Kattel, U., Yosef, R., Liebich, T., 2018b. Seebridge as next generation bridge inspection: overview, information delivery manual and model view definition. Automation in Construction 90, 134–145. doi:10.1016/j.autcon.2018.02.033.

Sampath, A., Shan, J., 2009. Segmentation and reconstruction of polyhedral building roofs from aerial lidar point clouds. IEEE Transactions on geoscience and remote sensing 48, 1554–1567. doi:http://dx.doi.org/10.1109/TGRS. 2009.2030180.

Schnabel, R., Wahl, R., Klein, R., 2007. Efficient ransac for point-cloud shape detection, in: Computer graphics forum, Wiley Online Library. pp. 214–226. doi:http://dx.doi.org/10.1111/j.1467-8659.2007.01016.x.

Shim, C., Yun, N., Song, H., 2011. Application of 3d bridge information modeling to design and construction of bridges. Procedia Engineering 14, 95–99. doi:http://dx.doi.org/10.1016/j.proeng.2011.07.010.

Shinde, R.C., Durbha, S.S., Potnis, A.V., 2021. Lidarcsnet: A deep convolutional compressive sensing reconstruction framework for 3d airborne lidar point cloud. ISPRS Journal of Photogrammetry and Remote Sensing 180, 313–334. doi:https://doi.org/10.1016/j.isprsjprs.2021.08.019.

Su, H., Maji, S., Kalogerakis, E., Learned-Miller, E., 2015. Multi-view convolutional neural networks for 3d shape recognition, in: Proceedings of the IEEE international conference on computer vision, pp. 945–953. doi:http: //dx.doi.org/10.1109/ICCV.2015.114.

Thomas, H., Qi, C.R., Deschaud, J.E., Marcotegui, B., Goulette, F., Guibas, L.J., 2019. Kpconv: Flexible and deformable convolution for point clouds, in: Proceedings of the IEEE/CVF international conference on computer vision, pp. 6411–6420. doi:https://doi.org/10.48550/arXiv.1904.08889.

Thompson, W., Owen, J., 1999. Feature-based reverse engineering of mechanical parts. IEEE Transactions on robotics and automation 15, 57–66. doi:http: //dx.doi.org/10.1109/70.744602.

Truong-Hong, L., Lindenbergh, R., 2022. Automatically extracting surfaces of reinforced concrete bridges from terrestrial laser scanning point clouds. Automation in Construction 135, 104127. doi:http://dx.doi.org/10.1016/j.autcon.2021.104127.

Vilgertshofer, S., Mafipour, M., Borrmann, A., Martens, J., Blut, T., Becker, R., Blankenbach, J., Gl¨obels, A., Beetz, J., Celik, F., 2022. Twingen: Advanced technologies to automatically generate digital twins for operation and maintenance of existing bridges, in: Proc. of European Conference on Product and Process Modeling 2022.

Wang, X., Girshick, R., Gupta, A., He, K., 2018. Non-local neural networks, in: Proceedings of the IEEE conference on computer vision and pattern recognition, pp. 7794–7803. doi:http://dx.doi.org/10.1109/CVPR.2018.00813.

Wang, Y., Sun, Y., Liu, Z., Sarma, S.E., Bronstein, M.M., Solomon, J.M., 2019. Dynamic graph cnn for learning on point clouds. Acm Transactions On Graphics (tog) 38, 1–12. doi:http://dx.doi.org/10.1145/3326362.

Xia, T., Yang, J., Chen, L., 2022. Automated semantic segmentation of bridge point cloud based on local descriptor and machine learning. Automation in Construction 133, 103992. doi:http://dx.doi.org/10.1016/j.autcon. 2021.103992.

Xie, J., Fang, Y., Zhu, F., Wong, E., 2015. Deepshape: Deep learned shape descriptor for 3d shape matching and retrieval, in: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 1275–1283. doi:http://dx.doi.org/10.1109/TPAMI.2016.2596722.

Yan, Y., Hajjar, J.F., 2021. Automated extraction of structural elements in steel girder bridges from laser point clouds. Automation in Construction 125, 103582. doi:http://dx.doi.org/10.1016/j.autcon.2021.103582.

Yang, X., del Rey Castillo, E., Zou, Y., Wotherspoon, L., Tan, Y., 2022. Automated semantic segmentation of bridge components from large-scale point clouds using a weighted superpoint graph. Automation in Construction 142, 104519. doi:http://dx.doi.org/10.1016/j.autcon.2022.104519.

Zhang, C., Tang, P., 2015. Visual complexity analysis of sparse imageries for automatic laser scan planning in dynamic environments. Computing in Civil Engineering , 271–279doi:http://dx.doi.org/10.1061/9780784479247.034.

Zhao, H., Jiang, L., Jia, J., Torr, P.H., Koltun, V., 2021. Point transformer, in: Proceedings of the IEEE/CVF International Conference on Computer Vision, pp. 16259–16268. doi:http://dx.doi.org/10.1109/ICCV48922.2021. 01595.

Zhao, Y.P., Wu, H., Vela, P.A., 2019. Top-down partitioning of reinforced concrete bridge components, in: Computing in Civil Engineering 2019: Smart Cities, Sustainability, and Resilience. American Society of Civil Engineers Reston, VA, pp. 275–283. doi:http://dx.doi.org/10.1061/9780784482445. 035.

Zhou, Y., Tuzel, O., 2018. Voxelnet: End-to-end learning for point cloud based 3d object detection, in: Proceedings of the IEEE conference on computer vision and pattern recognition, pp. 4490–4499. doi:http://dx.doi.org/10.1109/CVPR.2018.00472.