# Combination of Images and Point Clouds in a Generative Adversarial Network for Upsampling Crack Point Clouds

**NHUNG HONG THI NGUYEN**[1], (Student Member, IEEE),
**STUART PERRY**[1], (Senior Member, IEEE), **DON BONE**[1], (Member, IEEE),
**HA LE THANH**[2], (Member, IEEE), **MIN XU**[1], (Member, IEEE), AND **THUY THI NGUYEN**[3]

[1]Faculty of Engineering and Information Technology, University of Technology Sydney, Ultimo, NSW 2007, Australia
[2]VNU University of Engineering and Technology, Hanoi, Vietnam
[3]Faculty of Information Technology, Vietnam National University of Agriculture, Hanoi 12400, Vietnam

Corresponding author: Nhung Hong Thi Nguyen (nhunghongthi.nguyen@student.uts.edu.au)

**ABSTRACT** Point cloud data of cracks can be used for various purposes such as crack detection, depth calculation and crack segmentation. Upsampling low-density point clouds can help to improve the performance of those tasks. Building on existing methods that upsample point clouds from low-resolution point cloud input, to improve feature definition, this paper proposes a new method for upsampling low-density point clouds using a combination of these point clouds and corresponding 2D images of the original objects as input data. We use an architecture based on Generative Adversarial Networks (GAN) for training input point clouds with additional information from the corresponding 2D images. The key idea is to exploit features from both 2D images and point clouds to enrich point clouds in both the training and testing phases. Our method takes advantage of the combination of 2D images and point clouds using a GAN framework. Experimental results show our proposed method achieves a higher effectiveness compared with previous upsampling methods.

**INDEX TERMS** Point cloud upsampling, generative adversarial network, crack point cloud, enrichment point cloud, combining point clouds and images.

## I. INTRODUCTION

### A. RESEARCH BACKGROUND AND CHALLENGES

Crack detection and segmentation are essential for maintaining civil constructions, such as roads, bridges, and buildings. The application of digital imaging methods to this problem has resulted in significant advances. However, 2D crack images do not contain the rich information of three dimensional (3D) data such as point cloud data when examining complex structure or thin cracks.

Reconstruction techniques that facilitate the creation of high-resolution 3D data allow for in-depth and highly actionable investigations of the development of cracks [1]. Crack features such as crack width and crack depth can be characterized by using 3D data from a structured light scanner [2], but any reconstruction technique requires high-resolution point clouds to maximize utility.

The associate editor coordinating the review of this manuscript and approving it for publication was Jinjia Zhou.

Using 3D data has the potential to significantly improve crack detection and segmentation [3]. One of the difficulties when processing point clouds is that the low density of points limits the achievable resolution of features required for good detection. Upsampling of the point cloud data offers the possibility of addressing this problem, but current upsampling methods struggle to achieve the required fidelity.

Point cloud upsampling is a topic that challenges researchers in computer vision, and has attracted increasing interest in recent years [4]–[7]. In these papers, the authors focus on upsampling low-resolution point clouds by learning features from similar point clouds. The results from the above work are effective for upsampling many kinds of point cloud objects. However, there are some limitations that still need to be addressed to make them useful for the analysis of surfaces with cracks. Of particular importance is an ability to handle very sparse point clouds particularly those containing high spatial gradients or significantly non-uniform sampling exhibiting voids.
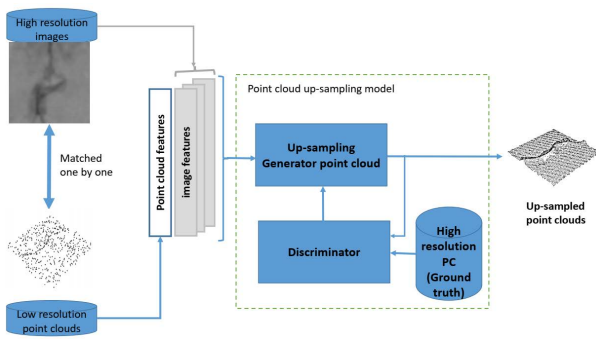
**FIGURE 1.** Proposed architecture for upsampling a point cloud by combining a low-resolution point cloud with a 2D image. The input is the combined data from high-resolution images and their matched point clouds. The point cloud upsampling model is based on a GAN framework that contains both generative and discriminative sub-models. The expected output is a high-resolution point cloud.

This paper focuses on proposing a new method to create high-resolution crack point clouds from the fusion low-resolution point clouds and high resolution 2D images. Our method supports many applications in civil engineering that need to use high-resolution 3D crack data.

## B. APPROACH AND CONTRIBUTIONS

In this paper, we combine information in low-resolution point clouds with features derived from high-resolution 2D images of the same subject matter to produce an upsampled version of the low-resolution point clouds. Several different approaches to the upsampling were explored, such as combining at the point-pixel level, or using transfer learning to combine at the feature level. Apart from improvements in the fidelity of the upsampled point cloud, the method also seeks to increase the uniformity of the sampling and fill voids in the point cloud data.

Crack point clouds are considered as 2.5D point clouds as they take the form of almost planar surfaces disturbed principally by the cracks. Similar 2.5D point clouds are important in a range of different applications, from the construction industry [8], [9] to health monitoring. However, high-density point clouds or 3D data are expensive and sometimes impractical to collect, so collecting low-density point clouds and then upsampling them offers advantages in practice.

There are two main types of 3D scanning technologies used for crack analysis. The first type is light amplification by stimulated emission of radiation (laser) triangulation 3D scanning technology and the second type is structured light 3D scanning technology. Crack 3D data can be collected using terrestrial laser scanning [10], [11] or a mobile laser scanning [12]. The laser technique is limited in resolution, so 3D data from laser data sources is used to detect surface distresses more than 1 cm wide [10]. Therefore, cracks that are smaller than 1 cm should be detected on upsampled data point clouds. The ground truth 3D data used in this paper are collected by a scanner using the structured light 3D scanning technology that has accuracy at 0.1 mm. The structured light 3D scanner has an advantage in resolution, but setting up

these systems for capture can sometimes be impractical in real world applications. Therefore, collecting high-resolution point clouds for training and using upsampled point clouds for real world applications has beneficial applications in real world scenarios.

Figure 1 shows our idea for applying transfer learning between point clouds and images in a type of GAN architecture with two parts; the *generative model* and *discriminative model* [13] for crack point cloud upsampling. The point cloud samples and image samples are created and matched one by one. Image features, extracted from images by a crack detection model, are combined with point cloud features before becoming input data for a GAN architecture. The output is a high-resolution uniformly sampled point cloud.

Our main contributions are:

- We propose a novel approach based on GAN and transfer learning that is effective for upsampling sparse point clouds.
- We show that 2D images can enrich the low-resolution point clouds, and the performance of this architecture is superior to architectures that use point clouds only.
- Our model achieves a superior performance compared to prior art point cloud upsampling approaches, demonstrating that our specifically tailored combination of 2D images and point clouds takes advantage of features from both images and point clouds.
- We present a new dataset of concrete crack point clouds and their corresponding images that will be made available to the research community.

## II. RELATED WORK

Significant advances have been made in the upsampling of point cloud data in recent years [5]–[7]. There are several methods for upsampling point cloud, including traditional methods and more recent methods based on convolutional neural networks.

## A. TRADITIONAL METHODS

Traditional methods such as interpolation between input points, have some disadvantages, arising principally from the fact that point clouds often do not have any spatial order or a regular structure [5]. In a method for extraction of break-lines and ground points from LiDAR point clouds [14], experiments showed that interpolation errors are mainly distributed around the break-lines. This strongly suggests that interpolation methods may not be effective for crack point clouds processing. Some previous experiments [5], [15] also indicated that methods based on neural networks demonstrated superior performance compared with the traditional methods. Hence, we focus on these advanced methods using Convolutional Neural Networks (CNNs) for upsampling point clouds.

## B. METHODS BASED ON CONVOLUTIONAL NEURAL NETWORKS

Methods based on CNNs have been used by many researchers for upsampling point clouds [5]–[7], [16]–[18]. These

methods often have two parts. The first part extracts features from point clouds, and the second part reconstructs points from feature expansion and optimizes the output by comparison with ground truth point clouds.

PU-Net [5] used a network architecture that has four components; patch extraction, point feature embedding, feature expansion, and coordinate reconstruction. The advantage of this model is that it can learn both local features and global features of a point cloud. The PU-Net is considered a multi-stage technique that notably improves the result, however this model shows more artifacts in high ratio upsampled results. This problem may be because the PU-Net model tries to correct the mistakes generated in earlier stages.

PU-GAN [6] is a proposal for upsampling point clouds using a GAN framework and demonstrates improvements in the quality of the resulting point clouds compared to earlier approaches. The PU-GAN method was evaluated using experiments that were implemented mostly on synthetic scanned data. On real point clouds such as LiDAR point clouds, the method of PU-GAN cannot fill some holes and cannot produce a high level of uniformity in the upsampled results.

A multi-step upsampling network is proposed by Yifan *et al.* [17]. The key idea of Yifan *et al.* is the use of a multistep patch-based network. The patch size is adaptive to the present step. The experiments in that work were implemented on synthetic point cloud data and not real scanned data. The multi-step method upsamples the point clouds in many steps, so has the disadvantage of also requiring many levels of ground truth resolution.

The key idea of Yifan *et al.* is using a multi-step patch-based network. The patch size is adaptive to the present step. The experiments in this paper are implemented on point clouds data that are made by software and are not real scanned data. The multi-step method upsamples point clouds in many steps, so this method also needs many levels of ground truth resolution. It is a disadvantage of Yifan's method.

In [7], a method called Point Cloud Super Resolution (PCSR) used Adversarial Residual Graph Networks for upsampling point clouds. Their experiments show that the residual blocks are effective in offering better performance and stable training. However, this approach has some disadvantages. These models cannot fill large holes or missing parts, and are also not effective for very small structures. The method we propose, based on CNN and the combination of point clouds and images, improves the quality of upsampled point clouds and solves the problem of filling voids in the point cloud data. We implement our method on real scanned data.

### C. METHODS COMBINING 2D IMAGES AND 3D DATA

There are existing point cloud processing methods using a combination of 2D images and 3D data. Based on the level of features used for combination, these works can be divided into two main types: Low-level feature combination and high-level feature combination.

#### 1) LOW-LEVEL FEATURE COMBINATION

In this method of feature combination, the low-level features collected by simple transforms are combined with point clouds or 3D data. Image features such as grayscale pixels and entropy values have been combined with depth information from point clouds for upsampling point clouds [19], [20]. In these works, the depth values for locations in the point cloud are assigned to corresponding pixels of an image, and then the upsampled point clouds are created by interpolation using the local entropy of pixels around the current pixel being processed. The simple combination of the depth value of each point and the grayscale image data may lead to aliasing errors due to the inconsistencies in the directions of gradients in the depth data.

#### 2) HIGH-LEVEL FEATURES COMBINATION

High-level image features can be extracted by convolutional neural networks and used to enrich point clouds. In a proposal for LiDAR Point Cloud Segmentation [21], an effective fusion method of RGB data and LiDAR was developed to combine features from color images with features from point clouds to segment the 2.5D point clouds. The idea of combining features from images and 3D data was used for object reconstruction [22]. In this work, Yang *et al.* used images for reconstructing 3D objects, and showed that features from images could be concatenated with features from 3D data. The combination of different kinds of features such as image features and point cloud features can be considered as a form of *transfer learning* [23].

This paper proposes two methods of point cloud and image combination. One method uses low-level image features and the other uses high-level image features.

### D. CRACK 3D DATA PROCESSING

Crack point clouds are used for crack detection and segmentation [24]–[29]. However, the effectiveness of these methods depends on the quality of point clouds. To enhance the point cloud quality and improve the accuracy of the crack detection and segmentation, our approach uses upsampling of the crack point cloud.

### III. PROPOSED METHOD

Our proposed method aims to create high-density point clouds from sparse point clouds and their corresponding images. We require the output point clouds to contain a high number of points, and have a uniform distribution of points in order to support the subsequent crack analysis. To do that, we propose an architecture based on GAN and use a combination of point clouds and images as input data.

In this section, we focus on two main parts. The first part is the proposed method for combining images and point clouds. We present in detail how images and point clouds are aligned and combined to become input data to a GAN model. The second part is the proposed GAN model based which uses the combined data from images and point clouds to upsample

the low-resolution point clouds. The three major components of this model are the generative model, discriminative model, and the loss functions. Each of these components are illustrated in this section. The generative model takes input from one of two feature combination methods; point-pixel and feature-feature combination and generates an upsampled version of the point cloud. The discriminative model ensures the fidelity of the upsampled point cloud by comparing the characteristics of the upsampled point cloud to that of the ground truth. Both the generative and discriminative models work together through the formulation of an appropriate loss function as will be described below.

### A. COMBINING IMAGES AND POINT CLOUDS

In this work, we propose two ways to combine image data and point cloud data. The first approach we call "point-pixel combination", and the second we call "feature-feature combination".

#### 1) POINT-PIXEL COMBINATION

Point clouds often have no regular structure, however, images have a regular order, and if their matched point clouds belong to a 2.5D surface such that the implicit surface is single-valued, then there is an alignment of the point cloud and the image such that each point in the point cloud can be matched to a pixel in the image unambiguously. An image has no explicit information about the depth of an object, but the image pixel's intensity often contains implicit depth information. In the case of crack images darker areas in a crack often correspond to the deepest parts of the crack. For this reason, we expect that the information from an image can be combined with point cloud information as an additional channel. The additional channel can be built from the image grayscale value or can be created from other features derived from the image.

For *Point-pixel combination*, we create a four-channel point cloud from each low-resolution point cloud input and its corresponding image. From each point $(x_k^{pc}, y_k^{pc}, z_k^{pc})$ in the input point cloud sample, we find the matched pixel $(x_k^{im}, y_k^{im})$ (where the grayscale value is $img[x_k^{im}, y_k^{im}]$) in the corresponding image. The new point contains four channels, $(x_k^{pc}, y_k^{pc}, z_k^{pc}, img[x_k^{im}, y_k^{im}])$.

We have also implemented additional channel using other image features. The first kind of feature is the Sobel gradient feature. Sobel features are produced by applying a Sobel operator [30] to the image. Gamma features extracted from the image using Gamma filters [31], [32] are also used in our experiments with $\gamma = 0.2$. Gamma features are strong features utilized in image enhancement. The last kind of image feature we use are Difference of Gaussians (DoG) features [33], [34]. DoG is used in various methods of image processing such as edge detection and image matching.

#### 2) FEATURE-FEATURE COMBINATION

To implement feature-feature combination, we transfer knowledge from the image domain, and then combine it with knowledge from the point cloud domain by a layer concatenation operation. We use two different models to extract features from point clouds and from images. Point cloud features are extracted from the same model that is used from *Point-pixel combination*. Image features are extracted from a crack detection model that was pre-trained in a previous model for crack detection [35]. This model is better suited for the method presented in this paper than other existing models for image features extraction because it was trained from a crack dataset. To take image features from the crack detection model, we freeze the first layers and only retrain the features from the last max-pooling layer.

The features from images provide extra useful information for the point cloud features. The image feature extraction component extracts features that can enhance the point cloud feature extraction such as crack edge features. Using transfer knowledge from a crack detection model has another advantage, it saves training time, because it uses the concept that has already been learnt from an existing model and learns from the last layer features with more complex representations specifically suited to the upsampling problem.

While the "point-pixel combination" method can be considered as the combination of point clouds and the *local* image features, the "feature-feature combination" method is the combination of global point cloud features and the *global* image features.

Figure 2 shows the architecture for feature extraction. The workflow in figure 2a is used for point-pixel combination, and figure 2b shows our proposed network for combining point clouds and images at the feature level.

In the *point pixel combination* architecture as shown in figure 2a, the input can be considered as a 4-channel point cloud. The first three channels are from the point cloud, and the fourth channel is an additional channel that comes from the matched image such as the grayscale value or other features. Point cloud features are extracted from the 4-channel point clouds by a convolutional neural network. We use twelve convolutional layers with an increment of the number of kernels. There are four concatenated layers placed after each three-layer of CNN to improve the global features by combining with local features.

Figure 2b shows the combination of point cloud features and image features. The point cloud samples and their matched 2D images are processed separately by the point cloud extractor and the 2D image extractor. Then, we combine global point cloud features and global image features following the two feature extraction models. The output of the *feature-feature combination* is the combined features with a diversity of the features from both point clouds and 2D images.
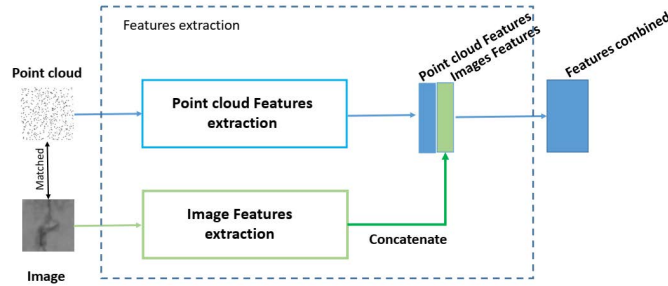
### B. UPSAMPLING MODEL BASED ON GAN
#### 1) GENERATIVE MODEL

In a general GAN architecture, the generative model aims to generate new data from input data. While some other
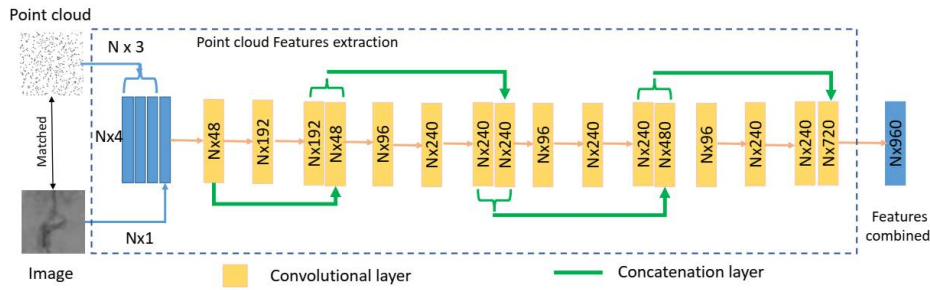
(a) Point-pixel combination.



(b) Point cloud feature - Image feature combination.

**FIGURE 2.** Two proposed methods for point cloud and image combination. (a) Point-pixel combination combines image and point clouds to create 4-channel input data and then extracts features from the 4-channel data. (b) Point cloud feature - Image feature combination combines point clouds and images by concatenating features that have been extracted separately from point clouds and images.
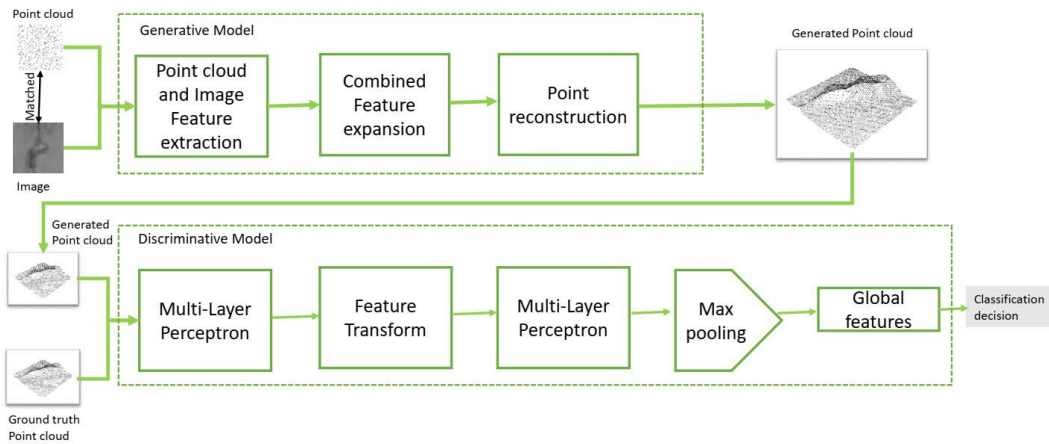


**FIGURE 3.** Point cloud upsampling framework using the combination of point clouds and images based on a GAN architecture. The generative model takes point clouds and images as inputs and comprises components for feature extraction, feature expansion and point reconstruction. The output of the generative model and the ground truth point clouds are inputs to the discriminative model that comprises a feature transform layer, Multi-Layer Perceptrons (MLPs) module, max pooling and global features extraction module.

upsampling data proposals based on GAN use only the low-resolution data as input data [6], [36], we use the combination of point clouds and images as the input. The upper part of figure 3 shows our model for the generator.

There are three transformations in our generative model: data feature extraction, feature expansion, and data reconstruction. A three-transformation generative model was also used in other works for upsampling point clouds such as in

PU-GAN [6], and for upsampling images [37], [38]. We use convolutional neural networks for three operations. The main role of each part is:

- *Feature extraction* extracts features from low-resolution point clouds. Each feature is represented by a high-dimensional vector. All high-dimensional vectors combine together and produce feature maps $P_{Features}$. The

number of features in $P_{Features}$ is equal to the number of points in a low-resolution input point cloud.

- *Feature expansion* based on deconvolution operation is a transformation that expands the feature maps $P_{Features}$ to a set of new feature maps $P_{Expanded_Features}$ with a higher number of features. The number of features in $P_{Expanded_Features}$ is similar to the number of points in the target point cloud.

- *Point Reconstruction* is the last phase in the generative model that regresses and aggregates features from $P_{Expanded_Features}$ to build a high-resolution point cloud with a given ratio. In this part, we use the Farthest Point Sampling method [39], [40] to optimize the uniformity of points in the output point cloud. The expected output point cloud is a high-density uniform point cloud.

### 2) DISCRIMINATIVE MODEL

The discriminative model aims to distinguish the ground truth point clouds and the generated high-resolution point clouds, so it works as a classification model. The discriminator uses the ground truth point clouds as positive samples, and the generated point clouds as negative samples. We use the MLP [41] for point cloud feature extraction. The MLP is also used in other proposals for point cloud upsampling [6] and point cloud classification [4]. We also use a feature transform block to improve the accuracy of point cloud classification [4]. Before the last activation layer, a max-pooling layer is used for collecting global features. The output of the discriminative model is a decision that indicates whether an input point cloud is a real or artificially generated point cloud. The lower part of figure 3 shows the basic framework for the discriminative model.

In the training phase, the generator and discriminator are trained alternately. In each epoch of the training phase, the generator is kept constant while the discriminator is trained, and then the discriminator is kept constant while the generator is trained. The distribution of the ground-truth point cloud is $x$, and the input data is a low-resolution point cloud and an image, $(z, img)$. Then, the generator space is $G((z, img), \theta_g)$, where G is a generative function represented by a network with parameters $\theta_g$, and the discriminator space is $D(y, \theta_d)$, where $D(x)$ is the probability that $x$ came from the real data rather than $(z, img)$ and $D(x)$ is represented by a network with parameters $\theta_d$.

While the generator tries to minimize the difference between the created point cloud and the ground truth point cloud, the discriminator tries to maximize the probability of assigning the correct label to ground truth examples and samples from $G$. Finally, the value function $V(D, G)$ is optimized as equation 1:

$$\min_{G} \max_{D} V(D, G)$$
$$= E_{x \sim p_x(x)} [log D(x)]$$
$$+ E_{(z,img) \sim p_{input}(z,img)} [log(1 - D(G(z, img)))] \quad (1)$$

### 3) LOSS FUNCTIONS

Our method aims to upsample point clouds using a GAN architecture, wherein the output point clouds should be uniform, and each point should be on the underlying object surfaces. So, the final loss function has contributions from three terms corresponding to each of the above three goals.

*Least Squares loss function for Generative Adversarial Networks.* Least-squared loss is a loss function for GAN architectures that was proposed by [42]. Least-squared loss avoids the problem of vanishing gradients. Equation 2 and 3 show the loss functions for the generator part and discriminator parts of our system.

$$L_{gan}(G) = \frac{1}{2}[D(Q) - 1]^2 \quad (2)$$

$$L_{gan}(D) = \frac{1}{2}[D(Q)^2 + (D(\tilde{Q}) - 1)^2] \quad (3)$$

where $D(Q)$ is is the confidence value predicted by $D$ from the generator output $Q$, and $\tilde{Q}$ is the fake sample.

*Uniform loss* aims to increase the uniformity of the distribution of points in the generated point cloud. We require uniformity of the output point cloud both globally and locally. To optimise the global uniformity, we take a number of random points, and from each random point we generate an area (denoted as $A_i$) on the underlying surface, and then optimise the number of points on each area. To optimise the local uniformity in each small area $A_i$, we optimise the nearest distance from point to point.

*Reconstruction loss* aims to encourage the generated points to lie on the target surface. Consider $Q$ as the ground truth point cloud. Then $Q$ has the same number of points as the fake point cloud $\tilde{Q}$. The reconstruction loss uses the Earth Mover's [43] distance as per equation 4.

$$d(Q, \tilde{Q}) = \min_{\varphi: Q \to \tilde{Q}} \sum_{p_i \subset Q} \|p_i - \varphi(p_i)\| \quad (4)$$

where $\varphi : Q \to \tilde{Q}$.

Finally, the model is trained by minimizing $L_{gan}(G)$ which combines the adversarial, uniform distribution and reconstruction loss terms with weights $\omega_{gan}, \omega_{uni}, \omega_{rec}$, respectively as shown in equation 5.

$$L_{gan} = \omega_{gan}L_{gan}(G) + \omega_{uni}L_{uni} + \omega_{rec}L_{rec} \quad (5)$$

## IV. EXPERIMENTS AND RESULTS
### A. DATA PREPARATION

There is no available crack point cloud dataset or dataset with point clouds and corresponding images. We implement our proposed method on a crack dataset we have collected ourselves. This dataset has two parts. The first part consists of point clouds, and the second part consists of images corresponding to each point cloud. Point clouds and images were collected in our laboratory. We used concrete blocks with dimensions $10cm \times 10cm \times 30cm$ from which we captured point clouds using an EinScan-SD scanner, and captured 2D images using a Canon EOS 5D Mark IV camera.

### 1) GROUND TRUTH POINT CLOUDS

The original point clouds from the scanner are very large. We divide the original 20 scans of individual concrete blocks into nearly 2000 point clouds for training and 200 point clouds for testing. Each divided point cloud contains 4096 points and one or more cracks.

### 2) INPUT POINT CLOUDS

The ground truth point clouds are down-sampled randomly to sparse point clouds that each contain 512 points. Our method aims to upsample these down-sampled point clouds by a factor of 8 to match the sampling of the original point cloud.

### 3) IMAGES

From the original images and original point clouds, we used Meshlab [44] to align the images with the corresponding point clouds. The output of the alignment process is a point cloud and its matched image. For each point cloud tile obtained by dividing the original large point cloud, a corresponding tile of the pre-aligned image of the concrete block is extracted by cropping.

Some examples of point clouds and their corresponding images are shown in figure 4.

### B. EVALUATION METRICS

We used two evaluation metrics to assess and compare our method with other point cloud upsampling methods.

**Chamfer distance (CD)** is an evaluation metric for assessing the similarity of two point clouds. It was defined and used in previous work [43], PU-GAN [6]. We used this distance measure to compare point clouds and assess the model.

Given two point sets $P = \{p_1, p_2, \ldots, p_n\}$ and $Q = \{q_1, q_2, \ldots, q_m\}$, the CD is calculated as equation 6.

$$CD(P, Q) = \frac{1}{P} \sum_{p \in P} \min_{q \in Q} ||p - q||_2^2 + \frac{1}{Q} \sum_{q \in Q} \min_{p \in P} ||p - q||_2^2 \tag{6}$$

**Hausdorff distance (HD)** was introduced by Hausdorff [45]. Berger [46] describes the application of HD to surface reconstruction in an upsampling proposal, and HD was also used to evaluate a method for upsampling point clouds [6]. HD measures how far two non-empty subsets are from each other. HD is the greatest of all the distances from a point in one set to the closest point in the other set. In point cloud comparison, given two point sets $P = \{p_1, p_2, \ldots, p_n\}$ and $Q = \{q_1, q_2, \ldots, q_m\}$. Hausdorff distance from A to B is defined as in equation 7.

$$H(P, Q) = \max_{p \in P} \min_{q \in Q} ||p - q|| \tag{7}$$

**Uniformity measure** To calculate the uniformity of generated point clouds from all methods, we calculate the standard deviation of the number of points belonging to equal areas on the point cloud surface. To do that, for each generated point cloud we randomly sample 1000 points. Then from each point $O_j$, we count the number of points inside a circle defined

by ($O_j$, *radius*). A point cloud has a higher uniformity if the standard deviation of the counted number of points is small. We evaluate the uniformity of point clouds with two values of *radius, r = 0.4 and r = 0.5*. Our method achieves a better uniformity by this measure and this is qualitatively evident in the visualization shown in figure 6.

Based on the metrics presented above, we compare the performance of our method to that of a number of other established point cloud upsampling methods [5]–[7]. The results of this comparison are summarised in table 1. The lower values of CD, HD, and uniformity of our method clearly demonstrate the superior performance of our approach compared to these methods. Figure 5 shows some examples of the image input, low-resolution point cloud input, ground truth point cloud, and the generated point clouds from the PU-Net, PU-GAN, PCSR, and our method.

The CD and HD value of the PU-Net model is very large as compared to other methods as shown in table 1. The PU-Net model used a feature aggregation method that does not consider spatial information when performing the feature expansion operation [5]. We speculate that this is the reason that the PU-Net model produces eight disjoint point clusters when used to upsample our dataset, resulting in very poor performance (the fourth column in figure 5). As a result, the PU-Net model obtained a very high value of CD and HD. So, we do not further compare the PU-Net model with other methods. Figure 5 shows our results when using *point-pixel combination* with DoG features from images.

Our point cloud data are 2.5D point sets that describe the surface of concrete blocks. However, the generated point clouds from the PCSR [7] are watertight in some areas, resulting in the cracks of the generated point clouds from PCSR being filled over with points, so they do not show the correct surface of the concrete in the vicinity of the cracks.

The PU-GAN model [6] achieved good results in our experiments. However, the fusion of high-resolution image data achieved by our approach significantly improves upon the results from the PU-GAN method. By combining point cloud and image data at two levels (point-pixel level and feature-feature level) we achieve better results against the CD and HD distance metrics. Our method also generates point clouds with more uniformly distributed samples.

Our method is better in filling the gaps on the point cloud surface as shown in figure 7. We take examples from PU-GAN, the PCSR method, and our method using "point-pixel combination" with DoG features. Two sides of a crack are often sparser than other areas. We compare the areas surrounded by black circles in the figure. On the ground truth surface, this area is sparse, on the point cloud generated by PU-GAN, there is a hole. Our method can generate this area with a higher point density.

In the "point-pixel combination", we use grayscale and other features from images such as Sobel, Gamma, DoG features. In comparison to the "feature-feature combination", the "point-pixel combination" is the better combination method. Among the "point-pixel combination" options, the
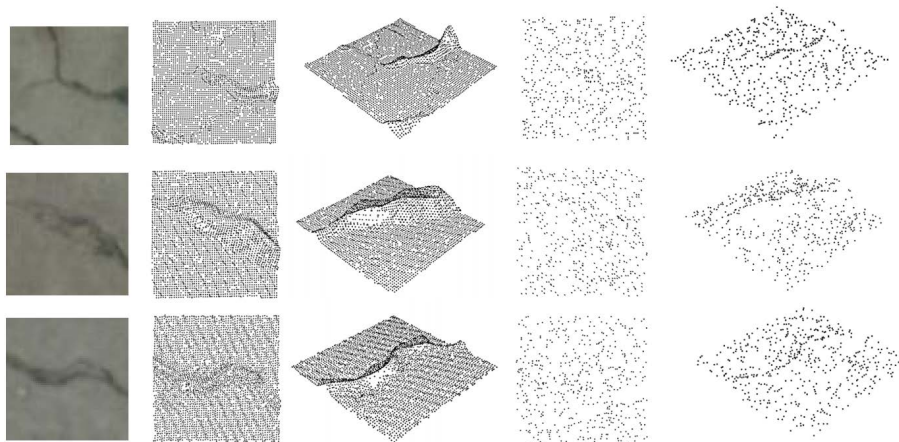
**FIGURE 4.** Examples of images and their corresponding point clouds. The first column shows the original images. The second and the third columns show ground-truth point clouds viewed from the top and an oblique angle respectively. The last two columns show input low-resolution point clouds viewed from the top and an oblique angle respectively.

**TABLE 1.** Comparisons of chamfer distance and hausdorff distance measurements.

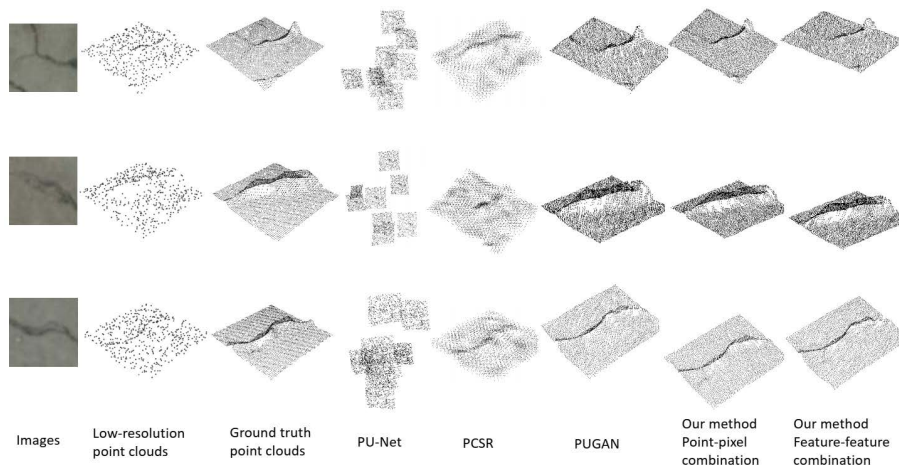| Methods | | CD $(10^{-3})$ | HD $(10^{-3})$ | Uniformity | |
| --- | --- | --- | --- | --- | --- |
| | | | | r = 0.4 | r = 0.5 |
| PU Net | | 328.86 | 1650.18 | - | - |
| PCSR | | 0.59 | 14.47 | 8.32 | 12.01 |
| PU-GAN | | 0.33 | 6.07 | 4.08 | 5.9 |
| Our methods | Point-Pixel with grayscale | **0.30** | **5.74** | **3.61** | **5.6** |
| | Point-Pixel with Sobel | **0.29** | **5.25** | **3.39** | **5.1** |
| | Point-Pixel with Gamma | **0.28** | **4.96** | **3.42** | **5.2** |
| | Point-Pixel with DoG | **0.28** | **4.95** | **3.17** | **4.9** |
| | Feature - Feature | **0.31** | **5.75** | **3.8** | **5.6** |



**FIGURE 5.** Examples of generated point clouds from five different methods and three different sets of input data. The two first columns contain input images and input low-resolution point clouds. The third column contains ground truth point clouds. The subsequent three columns are results from three previous methods, PU-Net, PCSR and PU-GAN. The last two columns are results from the proposed method with the two different approaches for combining image and point cloud information in the GAN architecture.

combination of point cloud with a channel of DoG features extracted from the images is the best combination for upsampling.

To assess the impact of the point cloud's resolution on point cloud classification, we implement the Pointnet architec-

ture [4] for classifying the crack point clouds and non-crack point clouds in low-resolution and high-resolution. Positive samples are point clouds that contain crack points, negative samples do not contain crack points. The Pointnet network takes a point cloud as input, then followed by a feature
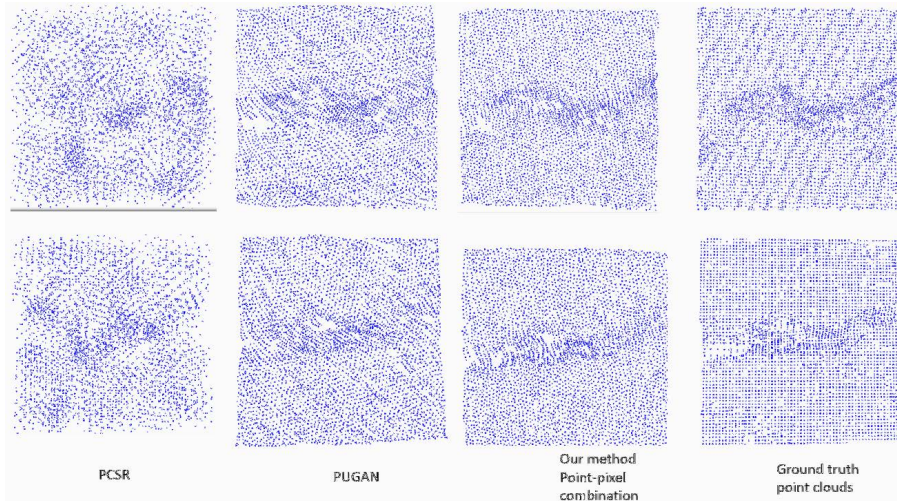
**FIGURE 6.** Examples of generated point clouds with different uniformities created using two existing methods and our method compared with the ground truth. The first column shows results from the PCSR method, the second column shows results from the PU-GAN method. The third column shows our results using point-pixel combination, and the last column shows the ground truth point clouds.

**TABLE 2.** Classification results of low-resolution and high-resolution crack point clouds.

| Data | | Pr | Re | F1-score | Accuracy (%) |
|---|---|---|---|---|---|
| Low-resolution point clouds | | 0.90 | 0.59 | 0.71 | 63 |
| High-resolution point clouds from | Ground truth | **0.99** | **0.73** | **0.85** | **81** |
| | PCSR | 1 | 0.5 | 0.66 | 50 |
| | PU-GAN | 0.98 | 0.61 | 0.75 | 67 |
| | Our method | **0.94** | **0.69** | **0.80** | **76** |

transform between two MLPs and then aggregates point features by a max-pooling layer. The classification output is the score for the two classes of crack point cloud or non-crack point cloud. The low-resolution point clouds contain 512 points as the input of the upsampling model. The high-resolution point clouds containing 4096 points are the ground truth point clouds, and the generated point clouds from the different upsampling methods. Table 2 shows our results for the classification of point clouds. We use precision (Pr), recall (Re), F-score, and accuracy to evaluate the classification model. The results indicate that the high-resolution point clouds can be classified better than the low-resolution point clouds. The generated point clouds from our method are better for classification than other upsampling methods.

The experiments in this paper show that the number of points in the point cloud affects the accuracy of crack detection. The experiment with a small number of points (512 points for each cloud) got a lower accuracy (63%). The experiment with a higher number of points (4096 points for each cloud) got a higher accuracy (up to 81%). In recent multispectral LiDAR point cloud classification research [47], the experiments also show that the higher number of points, the better the accuracy of the point cloud classification. Point cloud data that has a high uniformity where the number of points is large enough to represent small details will give the best results. The number of points should be optimized depending on the dataset and the number of points should

accurately capture the shape of the object [4]. Our ground truth data and our upsampled data display a high uniformity and sufficient density to achieve a reasonable accuracy in crack detection.

## C. DISCUSSION

While there is some existing work that combines 2D images and 3D data for other applications, there are no contributions using the combination of 2D and 3D data for upsampling point clouds. We have proposed such a method and demonstrate significantly better results than existing point cloud upsampling methods.

A GAN architecture is employed for the combination in two ways: the first one integrates points from 3D space and the corresponding pixels from 2D space, the second one integrates point cloud features and image features that are transferred from different models. Our proposed method also illustrates that images and point clouds can be combined at both the low-level and high-level feature stages. The generated high-resolution point clouds show that the combination of low-level features first, followed by the extraction of high-level features, is more effective than combining the high-level features that have been extracted separately.

The experiments in this paper generate high-resolution point clouds with high uniformity. We use the Chamfer and Hausdorff standard distance measures for evaluating point clouds and the proposed method obtains superior results
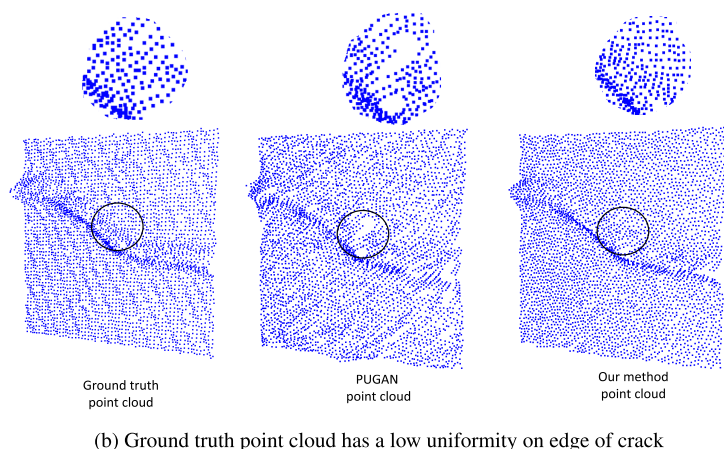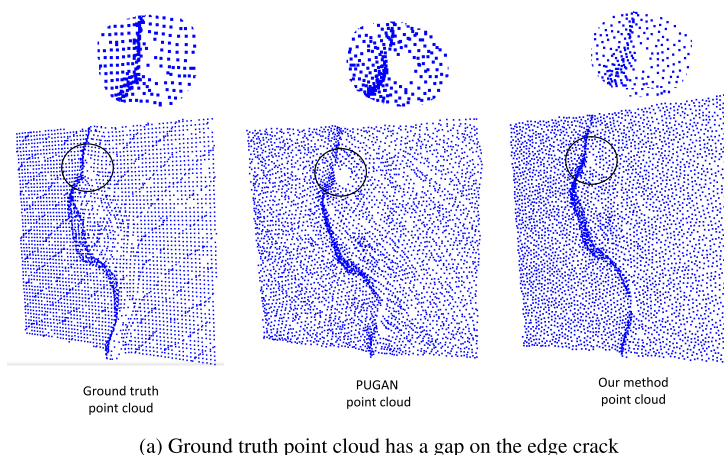
(a) Ground truth point cloud has a gap on the edge crack



(b) Ground truth point cloud has a low uniformity on edge of crack

**FIGURE 7.** Comparing results from our method and other methods in terms of filling a gap in the point clouds. The first column shows ground truth point clouds, the second shows results obtained by PU-GAN, and the last column shows results obtained from the proposed method. The ground truth in figure (a) shows a visible gap on the edge of the crack, while the ground truth in figure (b) shows a sparse crack edge.

compared with current state of the art methods. In terms of the detection of cracks in the point clouds, the up-scaled point clouds from our method perform better in terms of classification compared to other methods.

The high-resolution point clouds are generated from low-resolution point clouds, so they save cost in the collection and storage of the point clouds before evaluating and allow for the use of less expensive and more practical scanning equipment. The generated point clouds are also shown to improve the detection of cracks in the point clouds. The proposed method can be used in different 2.5D point clouds of other constructions and surfaces such as concrete or steel bridges and tunnels in both upsampling and classification. We feel this method has the potential to contribute significantly to civil engineering applications.

The proposed method is tailored for 2D images and associated 2.5D point clouds. In this paper, we have not investigated combining images and full 3D point clouds. The extension to 3D point clouds with complex structure is complicated by the difficulty of capturing the necessary structural information in a 2D image. However, with further work, this method

could be adapted to more general 3D point clouds if the 3D object can be treated as a set of 2.5D surfaces with associated point clouds and images. In future work we will consider the extension of our approach to point clouds of more complex 3D structures.

## V. CONCLUSION

We proposed a new method for upsampling crack point clouds using a combination of point clouds and two dimensional images as input data. We use a GAN architecture and combine point clouds and images in two ways: the first way is concatenating points from 3D space and the corresponding pixels from 2D space. The second way is concatenating point cloud features and image features that are transferred from different models. Comparisons of our approach to other established methods based on both distance metrics and a measures of uniformity of the point samples in the point cloud demonstrates that the combination of point clouds and images can improve the performance of upsampling of very low density point clouds. Our focus in this paper is on crack point clouds such as those from concrete bridges, and road

pavements, but the approach could be generalised to other 2.5D point cloud datasets and with further work may be applied to more general 3D data sets. We will make the dataset used in this paper available to the research community.

## REFERENCES

[1] A. Rovinelli, M. D. Sangid, H. Proudhon, Y. Guilhem, R. A. Lebensohn, and W. Ludwig, "Predicting the 3D fatigue crack growth rate of small cracks using multimodal data via Bayesian networks: *In-situ* experiments and crystal plasticity simulations," *J. Mech. Phys. Solids*, vol. 115, pp. 208–229, Jun. 2018.

[2] Z. Zhuo, C. Zhu, C.-S. Tang, H. Xu, X. Shi, and V. Mark, "3D characterization of desiccation cracking in clayey soils using a structured light scanner," *Eng. Geol.*, vol. 299, Mar. 2022, Art. no. 106566.

[3] C. Jiang and Y. J. Tsai, "Enhanced crack segmentation algorithm using 3D pavement data," *J. Comput. Civil Eng.*, vol. 30, no. 3, May 2016, Art. no. 04015050.

[4] R. Q. Charles, H. Su, M. Kaichun, and L. J. Guibas, "PointNet: Deep learning on point sets for 3D classification and segmentation," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jul. 2017, pp. 652–660.

[5] L. Yu, X. Li, C.-W. Fu, D. Cohen-Or, and P.-A. Heng, "PU-Net: Point cloud upsampling network," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit.*, Jun. 2018, pp. 2790–2799.

[6] R. Li, X. Li, C.-W. Fu, D. Cohen-Or, and P.-A. Heng, "PU-GAN: A point cloud upsampling adversarial network," in *Proc. IEEE/CVF Int. Conf. Comput. Vis. (ICCV)*, Oct. 2019, pp. 7203–7212.

[7] H. Wu, J. Zhang, and K. Huang, "Point cloud super resolution with adversarial residual graph networks," 2019, *arXiv:1908.02111*.

[8] M. R. Jahanshahi and S. F. Masri, "Adaptive vision-based crack detection using 3D scene reconstruction for condition assessment of structures," *Autom. Construct.*, vol. 22, pp. 567–576, Mar. 2012.

[9] S. Zhao, F. Kang, J. Li, and C. Ma, "Structural health monitoring and inspection of dams based on UAV photogrammetry with image 3D reconstruction," *Autom. Construct.*, vol. 130, Oct. 2021, Art. no. 103832.

[10] Z. Feng, A. E. Issaoui, M. Lehtomäki, M. Ingman, H. Kaartinen, A. Kukko, J. Savela, H. Hyyppä, and J. Hyyppä, "Pavement distress detection using terrestrial laser scanning point clouds–accuracy evaluation and algorithm comparison," *ISPRS Open J. Photogramm. Remote Sens.*, vol. 3, Jan. 2022, Art. no. 100010.

[11] X. Xu and H. Yang, "Intelligent crack extraction and analysis for tunnel structures with terrestrial laser scanning measurement," *Adv. Mech. Eng.*, vol. 11, no. 9, 2019, Art. no. 1687814019872650.

[12] M. Zhong, L. Sui, Z. Wang, and D. Hu, "Pavement crack detection from mobile laser scanning point clouds using a time grid," *Sensors*, vol. 20, no. 15, p. 4198, Jul. 2020.

[13] I. Goodfellow, J. Pouget-Abadie, M. Mirza, B. Xu, D. Warde-Farley, S. Ozair, A. Courville, and Y. Bengio, "Generative adversarial nets," in *Proc. Adv. Neural Inf. Process. Syst.*, vol. 27, 2014, pp. 1–9.

[14] B. Yang, R. Huang, Z. Dong, Y. Zang, and J. Li, "Two-step adaptive extraction method for ground points and breaklines from LiDAR point clouds," *ISPRS J. Photogramm. Remote Sens.*, vol. 119, pp. 373–389, Sep. 2016.

[15] R. Li, X. Li, P.-A. Heng, and C.-W. Fu, "Point cloud upsampling via disentangled refinement," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2021, pp. 344–353.

[16] L. Yu, X. Li, C.-W. Fu, D. Cohen-Or, and P.-A. Heng, "EC-Net: An edge-aware point set consolidation network," in *Proc. Eur. Conf. Comput. Vis. (ECCV)*, 2018, pp. 386–402.

[17] W. Yifan, S. Wu, H. Huang, D. Cohen-Or, and O. Sorkine-Hornung, "Patch-based progressive 3D point set upsampling," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2019, pp. 5958–5967.

[18] R. Sharma, T. Schwandt, C. Kunert, S. Urban, and W. Broll, "Point cloud upsampling and normal estimation using deep learning for robust surface reconstruction," 2021, *arXiv:2102.13391*.

[19] H. Kwon, K. Kim, and J. Dolne, "Improving 3D registration by upsampling of sparse point cloud through fusion with high-resolution 2D image," *Proc. SPIE*, vol. 10410, Sep. 2017, Art. no. 104100I.

[20] H. Kwon and K. Kim, "Three dimensional model generation using heterogeneous 2D and 3D sensor fusion," U.S. Patent 10 614 579, Apr. 7, 2020.

[21] G. Krispel, M. Opitz, G. Waltner, H. Possegger, and H. Bischof, "FuseSeg: LiDAR point cloud segmentation fusing multi-modal data," in *Proc. IEEE Winter Conf. Appl. Comput. Vis. (WACV)*, Mar. 2020, pp. 1874–1883.

[22] S. Yang, M. Xu, H. Xie, S. Perry, and J. Xia, "Single-view 3D object reconstruction from shape priors in memory," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2021, pp. 3152–3161.

[23] S. Bozinovski and A. Fulgosi, "The influence of pattern similarity and transfer learning upon the training of a base perceptron B2," in *Proc. Symp. Inf.*, 1976, pp. 3–121.

[24] Y.-C.-J. Tsai and F. Li, "Critical assessment of detecting asphalt pavement cracks under different lighting and low intensity contrast conditions using emerging 3D laser technology," *J. Transp. Eng.*, vol. 138, no. 5, pp. 649–656, May 2012.

[25] Q. Li, D. Zhang, Q. Zou, and H. Lin, "3D laser imaging and sparse points grouping for pavement crack detection," in *Proc. 25th Eur. Signal Process. Conf. (EUSIPCO)*, Aug. 2017, pp. 2036–2040.

[26] A. Zhang, K. Wang, B. Li, E. Yang, X. Dai, and Y. Peng, "Automated pixel-level pavement crack detection on 3D asphalt surfaces using a deep-learning network," *Comput.-Aided Civil Infrastruct. Eng.*, vol. 32, no. 10, pp. 805–819, 2017.

[27] A. Zhang, K. C. P. Wang, Y. Fei, Y. Liu, S. Tao, C. Chen, J. Q. Li, and B. Li, "Deep learning–based fully automated pavement crack detection on 3D asphalt surfaces with an improved CrackNet," *J. Comput. Civil Eng.*, vol. 32, no. 5, Sep. 2018, Art. no. 04018041.

[28] Y. Fei, K. C. P. Wang, A. Zhang, C. Chen, J. Q. Li, Y. Liu, G. Yang, and B. Li, "Pixel-level cracking detection on 3D asphalt pavement images through deep-learning- based CrackNet-V," *IEEE Trans. Intell. Transp. Syst.*, vol. 21, no. 1, pp. 273–284, Jan. 2020.

[29] A. Zhang, K. C. P. Wang, Y. Fei, Y. Liu, C. Chen, G. Yang, J. Q. Li, E. Yang, and S. Qiu, "Automated pixel-level pavement crack detection on 3D asphalt surfaces with a recurrent neural network," *Comput.-Aided Civil Infrastruct. Eng.*, vol. 34, no. 3, pp. 213–229, Mar. 2019.

[30] F. G. Irwin, "An isotropic 3×3 image gradient operator," Presentation at Stanford AI Project, Tech. Rep., 2014.

[31] C. A. Poynton, "Rehabilitation of gamma," *Proc. SPIE*, vol. 3299, pp. 232–249, Jul. 1998.

[32] R. C. Gonzalez and R. E. Woods, *Digital Image Processing*, 2002.

[33] D. Marr and E. Hildreth, "Theory of edge detection," *Proc. Roy. Soc. London. B, Biol. Sci.*, vol. 207, no. 1167, pp. 187–217, Feb. 1980.

[34] T. Lindeberg, "Image matching using generalized scale-space interest points," *J. Math. Imag. Vis.*, vol. 52, no. 1, pp. 3–36, 2015.

[35] N. H. Nguyen, S. Perry, D. Bone, H. T. Le, and T. T. Nguyen, "Two-stage convolutional neural network for road crack detection and segmentation," *Expert Syst. Appl.*, vol. 186, Dec. 2021, Art. no. 115718.

[36] R. Dahl, M. Norouzi, and J. Shlens, "Pixel recursive super resolution," in *Proc. IEEE Int. Conf. Comput. Vis. (ICCV)*, Oct. 2017, pp. 5439–5448.

[37] C. Dong, C. C. Loy, K. He, and X. Tang, "Image super-resolution using deep convolutional networks," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 38, no. 2, pp. 295–307, Feb. 2015.

[38] X. Hu, W. Ren, J. LaMaster, X. Cao, X. Li, Z. Li, B. Menze, and W. Liu, "Face super-resolution guided by 3D facial priors," in *Proc. Eur. Conf. Comput. Vis.* Cham, Switzerland: Springer, 2020, pp. 763–780.

[39] T. F. Gonzalez, "Clustering to minimize the maximum intercluster distance," *Theor. Comput. Sci.*, vol. 38, pp. 293–306, Jan. 1985.

[40] C. Moenning and N. A. Dodgson, "Fast marching farthest point sampling," Univ. Cambridge, Comput. Lab., Cambridge, U.K., Tech. Rep., 2003.

[41] L. B. Almeida, "C1. 2 multilayer perceptrons," in *Handbook of Neural Computation*, vol. 1, 1997.

[42] X. Mao, Q. Li, H. Xie, R. Y. K. Lau, Z. Wang, and S. P. Smolley, "Least squares generative adversarial networks," in *Proc. IEEE Int. Conf. Comput. Vis. (ICCV)*, Oct. 2017, pp. 2794–2802.

[43] H. Fan, H. Su, and L. Guibas, "A point set generation network for 3D object reconstruction from a single image," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jul. 2017, pp. 605–613.

[44] P. Cignoni, M. Callieri, M. Corsini, M. Dellepiane, F. Ganovelli, and G. Ranzuglia, "MeshLab: An open-source mesh processing tool," in *Proc. Eurograph. Italian Chapter Conf.*, V. Scarano, R. D. Chiara, and U. Erra, Eds. The Eurographics Assoc., 2008.

[45] H. Blumberg, "Hausdorff's Grundzüge der Mengenlehre," *Bull. Amer. Math. Soc.*, vol. 27, no. 3, pp. 116–129, 1920.

[46] M. Berger, J. A. Levine, L. G. Nonato, G. Taubin, and C. T. Silva, "A benchmark for surface reconstruction," *ACM Trans. Graph.*, vol. 32, no. 2, pp. 1–17, Apr. 2013.

[47] Z. Jing, H. Guan, P. Zhao, D. Li, Y. Yu, Y. Zang, H. Wang, and J. Li, "Multispectral LiDAR point cloud classification using SE-PointNet++," *Remote Sens.*, vol. 13, no. 13, p. 2516, Jun. 2021.

**NHUNG HONG THI NGUYEN** (Student Member, IEEE) received the B.S. and M.S. degrees in electronics and telecommunications from the Hanoi University of Technology and Science, Hanoi, Vietnam, in 2008 and 2010, respectively. She is currently pursuing the Ph.D. degree in computer science with the Faculty of Engineering and Information Technology, University of Technology Sydney, Australia.

She has been working as a Lecturer with the University of Transport and Communications (UTC), Hanoi, since 2011. Her research interests include the digital image processing, machine learning, computer science, and the applications of artificial intelligent (AI) in transport.

Mrs. Nguyen's awards include a Scholarship with the University of Technology Sydney in the Ph.D. joint program with Vietnam National University.

**STUART PERRY** (Senior Member, IEEE) received the Ph.D. degree from the University of Sydney, in 1999.

He began his research career studying the application of image processing to the detection of underwater objects in sector-scan sonar imagery for the Maritime Operations Division, Defense Science and Technology Organization (DSTO) for the Ph.D. degree. He has over 20 years of experience conducting research into image processing, psychophysics, signal processing, image quality, and models for the quantification of image preference and aesthetics for both government, industry, and academia. During this time, he represented Australia in Computer-Aided Detection and Classification Specialist Group, Technical Panel 13 (Mine Warfare and High-Frequency Acoustics), Maritime Group, The Technical Cooperation Program (TTCP). From 2003 to 2016, he worked at Canon Information Systems Research Australia (CiSRA), a Canon group company and one of the Canon Group's largest research and development facilities outside of Japan. During this time, he worked on camera white balancing technologies, led research teams working on print quality measurement, document security, and perceptual quality measurement for various consumer devices. In 2016, he joined the Perceptual Imaging Laboratory (PILaboratory), Faculty of Engineering and IT, conducting research into color and perceptual quality in 3-D environments.

**DON BONE** (Member, IEEE) received the Ph.D. degree in physics from ANU, in 1988. He worked at ANU for a couple of years developing optical diagnostics for hyper-velocity flows. During his Ph.D. degree and in subsequent years, he developed or helped develop several seminal methods in fringe analysis which have important applications in interferometry, structured light, and X-ray phase contrast imaging. For most of his career, he has worked on image and video processing problems, in government and industry research organizations (CSIRO, Mediaware Solutions Pty Ltd, CiSRA), though he has interspersed this with other activities doing contract research and starting and operating several businesses. In 2016, he joined the Faculty of Engineering and Information Technology, UTS, as an Industry Associate Professor, a position from which he has recently retired. He is currently an Adjunct Fellow at UTS. He is a member of the Association of Computing Machinery (ACM), Optica, and the International Society for Optics and Photonics (SPIE).

**HA LE THANH** (Member, IEEE) received the B.S. and M.S. degrees in information technology from the College of Technology, Vietnam National University, Hanoi, and the Ph.D. degree from Korea University, in 2010. In 2005, he received a Korean Government Scholarship for Ph.D. program with the Department of Electronics Engineering, Korea University. After graduation, he joined the Faculty of Information Technology, University of Engineering and Technology, Vietnam National University, as an Associate Professor. His research interests include image/video analysis and processing, satellite image processing, and computer vision. He has deep experiences in teaching digital image processing, computer vision, and multimedia communication courses for both undergraduate and postgraduate programs. He has also been the principle investigator and the main investigator of many fundamental research and technology development projects funded by both domestic and international organizations. He also makes contributions in serving many domestic and international ICT academic conferences, including KSE, NICS, ATC, SoICT, and ICEIC. In addition, he is a member of the Institute of Electronics, Information and Communication Engineers (IEICE), and the Vietnamese Association for Pattern Recognition (VAPR).

**MIN XU** (Member, IEEE) received the B.E. degree from the University of Science and Technology of China, the M.S. degree from the National University of Singapore, and the Ph.D. degree from The University of Newcastle, Australia. She is currently an Associate Professor with the University of Technology Sydney. She has authored or coauthored more than 150 research publications in top international journals and conferences. Her research interests include multimedia data analytics, computer vision, and machine learning.

**THUY THI NGUYEN** received the M.Sc. degree in information technology from the Hanoi University of Science and Technology, in 2002, and the Ph.D. degree in computer science from the Graz University of Technology, Austria, in 2009.

She has been a Lecturer and a Researcher with the Faculty of Information Technology, Vietnam National University of Agriculture, since 1998, and the Head of the Department of Computer Science, since 2011. She was a Specially Appointed Associate Professor during her visit to Osaka University, Japan, in 2016. She has published a number of papers in prestigious international journals and conferences and the coauthor of three patents in the aforementioned fields. She joined program committee and being a reviewer of several national and international journals and conferences. Her research interests include computer vision, machine learning, and pattern recognition.

- - -