Computer Vision–Based Estimation of Flood Depth in Flooded-Vehicle Images

Somin Park[1]; Francis Baek[2]; Jiu Sohn[3]; and Hyoungkwan Kim[4]

Abstract: This study proposes a vision-based method for flood depth estimation using flooded-vehicle images with a ground-level view. The proposed method is comprised of three main processes: segmentation of vehicle objects, cross-domain image retrieval, and estimation of flood depth. First, Mask region-based convolution neural network (R-CNN) is used to detect flooded vehicles in flooding images. Second, on the basis of feature maps from VGGNets, dynamic feature space selection is employed to select a three-dimensional (3D) rendered car image most similar to the flooded object using the metric of cosine distance. Finally, the flood depth is calculated through a comparison of the flooded object and the 3D rendered image. The feature maps from Pooling layer 4 of VGG19, under the condition of a cosine distance of <0.55, produces an average error of 7.51 pixels, corresponding to 9.40% of the tire height. A total of 500 flooding images are used to validate the method. **DOI: 10.1061/(ASCE)CP.1943-5487.0000956.** © *2020 American Society of Civil Engineers.*

Author keywords: Disaster management; Flood depth estimation; Deep learning; Computer vision; Image retrieval.

Introduction

Among the various extreme weather events, flooding is a natural disaster that occurs frequently around the world, causing significant harm to humans and material damage. According to the World Disaster Report (IFRC 2018), between 2008 and 2017 there were a total of 1,522 floods, and floods were the most common weather-related hazards. Additionally, in this period, there were 50,312 flood-related fatalities, and 730 million people required immediate assistance during periods of emergency (IFRC 2018). Urban areas with dense populations and infrastructures are particularly susceptible to flooding. The land use extends to rivers or coastal areas where there is a risk of flooding, and the proportion of impermeable surfaces has increased because of deforestation and urbanization. Disasters in urban areas where economic and social resources are concentrated can lead to large-scale direct damage, as well as social-order disruption resulting from various types of indirect damage. Therefore, measures to reduce flood damage in cities are vital. Government agencies provide flood warning services to manage flood risks, and the services deal with events that are imminent or already occurring (WMO 2013). Government agencies provide people with information regarding where and when risks may occur to reduce the negative impacts of disasters. However, owing to climate change and excessive urbanization, the uncertainty of disasters is increasing. Therefore, to reduce the damage caused by disasters, it is necessary to accurately evaluate the disaster potential in a timely manner.

In urban areas where the population has increased and industrial activities are concentrated, the importance of transportation is growing, resulting in higher road density in more developed regions (UN 2016). Flooding is prevalent in road infrastructure because road surfaces with low water infiltration rates are preferential paths for storm water. During and after the flood event, driving in flooding conditions may lead to serious accidents, as drivers may lose control or collide with objects. In many countries, vehicle-related accidents are serious in the event of flood. For the period 1959–2005 in the US, 63% of flood-related fatalities were vehicle-related (Terti et al. 2017). In Greece, 40% of flood deaths are vehicle-related (Diakakis and Deligiannakis 2017) and more than 3,000 vehicles were damaged in 2017 by flash flood in Bangkok, the capital of Thailand (Samui Times 2017). To provide warning from flooding, flooding warning systems are operated in manual systems or automatic systems in the US (WMO 2013). In manual systems, rain gauges or stream gauges are used for a simple-to-use flood forecast procedures based on tables, graphs, or charts. Conversely, automatic systems consist of computer software and various sensors, such as water-level sensors or hydrologic sensors. Because for both systems data collection devices cannot cover everywhere in districts, flood damage information for specific road segments may not be collected. In addition, not all states have a standard inspection method for flooded roads (Medina et al. 2016). Considering the risk of flooding on road pavements, it is necessary to collect detailed information about the flooded road area for accurate and timely distribution of warning.

With the proliferation of mobile devices, individuals can instantly record and share the situation they are in via text and images. Social media data have the advantages of being real-time and shareable and have exhibited potential for utilization in a variety of recent flood-related studies. Image data include visual information, such as the severity of damage and contextual information supporting "the interpretation and verification of derived information" (Fohringer et al. 2015), which can be used to analyze flood events. This meaningful information can enhance the capacity of disaster response or disaster recovery. Vehicles are at risk in flood situations and can directly demonstrate the degree of severity of

[1]Graduate Student, School of Civil and Environmental Engineering, Yonsei Univ., Seoul 03722, Korea. Email: somin109@yonsei.ac.kr

[2]Graduate Student, School of Civil and Environmental Engineering, Yonsei Univ., Seoul 03722, Korea. Email: fbaek@yonsei.ac.kr

[3]Graduate Student, School of Civil and Environmental Engineering, Yonsei Univ., 50, Seoul 03722, Korea. Email: jiujohn@yonsei.ac.kr

[4]Professor, School of Civil and Environmental Engineering, Yonsei Univ., Seoul 03722, Korea (corresponding author). Email: hyoungkwan@yonsei.ac.kr

© ASCE 04020072-1 J. Comput. Civ. Eng.

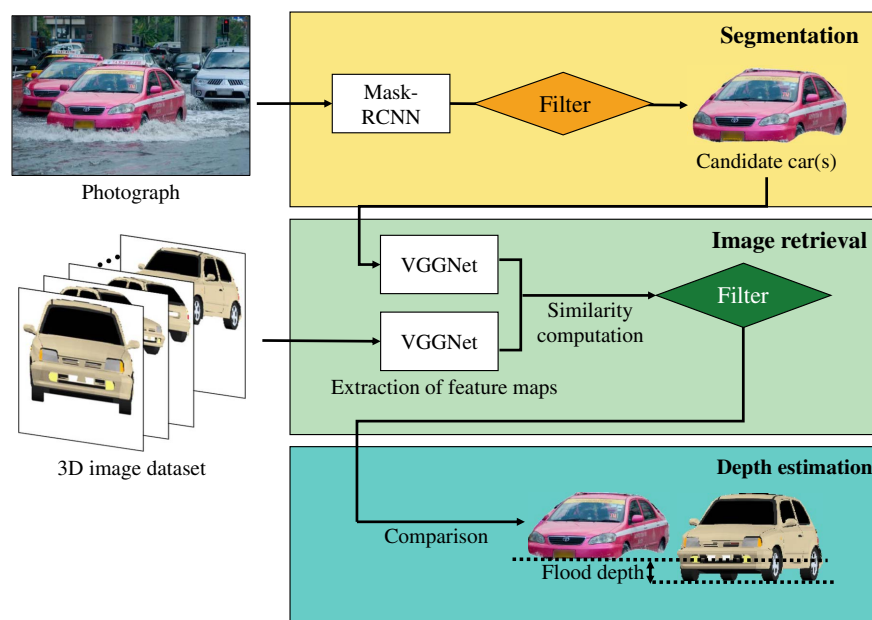J. Comput. Civ. Eng., 2021, 35(2): 04020072

Fig. 1. Framework of the proposed method for flood depth estimation. (Car image by Thor Jorgen Udvang/Shutterstock.com; 3D-rendered images by Somin Park.)

flooding on roads. For analyzing impacts of flooding on road traffic disruption, previous studies have used flooded depth of vehicles as a stability criterion or floating condition (Pregnolato et al. 2017; Kramer et al. 2016). Monitoring of vehicles exposed to flooding can be useful for evaluating flood damage. Flood impacts on particular local road segments can be identified from flooded vehicle images. During flood events, people can bypass the flooded road segment by obtaining the flood depth information. Even after flood events, the flood depth information can be used for identifying and repairing the specific road segment. However, because images are unstructured data, a person (manager) must manually check each image to derive the visual information.

We propose a vision-based method to analyze vehicles exposed to flooded roads in public image data to identify the flood depth for an intelligent transportation system (ITS). Fig. 1 illustrates the framework of the proposed method. In this study, two forms of data are used: vehicle images of various sizes and types obtained from the public domain, and images rendered from a three-dimensional (3D) vehicle model. The proposed method consists of three main processes. First, vehicle objects in the image are detected using a deep learning–based object segmentation network Mask-region-based convolution neural network (Mask-RCNN) (He et al. 2017). Second, cross-domain image retrieval is performed using a deep learning–based classification network [VGGNet (Simonyan and Zisserman 2014)] to select the 3D rendered image most similar to the vehicle object. Both processes involve data filtering to select the appropriate data to be used for the final purpose of flood depth estimation. Finally, the flood depth is calculated by identifying the difference between the paired images.

Related Works

Disaster management includes the prevention, preparedness, response, and recovery phases and has been the focus of various studies. In the subject of disaster preparedness, Kim and Ham (2019) proposed a geospatial localization method for distant objects, and Zheng et al. (2015) suggested a planning method for emergency relief transportation. In the subject of disaster response, Choi et al. (2018) proposed a disaster simulation platform that can be flexibly applied to diverse disaster situations, Wang and Zlatanova (2019) designed and developed a safe routing algorithm in the presence of moving obstacles, and Xu et al. (2018) suggested a fleet management solution dealing with the imbalances between rescue demands and vehicle supplies. Damage assessment can be used in the disaster response and recovery stages. Zhou et al. (2019) proposed a method for evaluating building damage using remotely sensed data, and Kashani et al. (2014) used terrestrial laser scanning to measure the wall and roof damage of buildings. Ploeger et al. (2015) presented an integrated tool for rapid structural data collection to support disaster risk assessment. For the disaster recovery phase, Radmehr and Araghinejad (2014) and El-Anwar et al. (2015) introduced tools for hazard map generation and optimized reconstruction, respectively, to assist the decision makers in postdisaster planning.

For rapid disaster response and recovery, studies have been actively conducted to automatically detect disaster events. With a focus on floods, disaster detection methods using various data have been proposed. Remote images from synthetic aperture radar (SAR) have been used to detect flooded areas (Mason et al. 2014; Long et al. 2014). Double scattering was used to detect flooded areas in SAR layover regions (Mason et al. 2014), and a change detection and thresholding technique was applied to SAR images to identify flooded areas (Long et al. 2014). The SAR data have the capability to cover all-weather and large damaged area within a short time; however, it is difficult for the SAR data to provide detailed and accurate information regarding floods, which may not be visible owing to radar shadowing and layover caused by buildings (Mason et al. 2014). Road surface recognition based on laser radar was performed to distinguish six cases including flooded conditions (Aki et al. 2016). With the advent of mobile devices and the Internet, the use of large-scale public data has exhibited potential for fast information delivery (D'Andrea et al. 2015; Ma et al. 2017). Lin et al. (2018) proposed a method for filtering disaster-related textual data using crowdsourcing, and Fohringer et al. (2015) proposed a textual mining method with an additional

© ASCE   04020072-2   J. Comput. Civ. Eng.

J. Comput. Civ. Eng., 2021, 35(2): 04020072

functionality to generate a flood inundation depth map. A hierarchical classification method was proposed for disaster images using a public dataset containing text and image information (Yang et al. 2011).

Image processing has recently achieved substantial progress, and computer-vision and deep learning approaches are being proposed for analysis of flood images. Studies using normal camera images have been conducted for tasks such as flood event detection, floodwater detection, and flood depth estimation. First of all, there are recent works to use deep-learning models for distinguishing between flood and nonflood images. Lopez-Fuentes et al. (2017) proposed a multimodal deep-learning approach for flooding classification. Features extracted from the text data through bidirectional long short-term memory (LSTM) (Hochreiter and Schmidhuber 1997) were concatenated with the image features extracted using InceptionV3 (Szegedy et al. 2016) for flood detection. Said et al. (2018) used a combination of object- and scene-level features of images; the features were extracted from multiple deep-learning models [AlexNet (Krizhevsky et al. 2012), VGGNet (Simonyan and Zisserman 2014), and ResNet50 (He et al. 2016)]. Cycle-consistent generative adversarial networks (CycleGANs) (Zhu et al. 2017) were used to improve the performance of flood detectors (Pouyanfar et al. 2019). The CycleGANs improved the classifier performance by applying style transfers in blurry, rainy, and night contexts.

In addition to classifying images into flood and nonflood, there have also been attempts to extract flood area from images. A pipeline was proposed to identify inundated area using an approach of dry-flooded image pairing (Witherow et al. 2018; Cetin et al. 2019). This method used various image processing techniques such as water edge detection and inpainting, and finally flooded area was detected by registering the pair images. Successful results of 87% accuracy were yielded, but there were many hand-selected parameters such as threshold values. Sazara et al. (2019) proposed a deep learning approach to automatically detect flood area. A fully convolutional network (Long et al. 2015) was used to identify the regions of flooding and scored a f1-score of 91%. Although these three studies achieved high accuracy to detect flood area, flood depth information could not be identified by the suggested methods.

To deal with the problem mentioned previously, research has been conducted to estimate flood depth of various objects. Nair and Rao (2016) surveyed methods of flood depth estimation based on image processing. The surveyed methods detected water level indicators: water level characters on a bridge (Lin et al. 2013), reference line in a floodwall (Ortigossa et al. 2015), and four circle markings on an open channel (Hies et al. 2012). The proposed studies were effective only when using images taken from fixed positions to detect the indicators covered by each study. Narayanan et al. (2014) analyzed partially submerged building images, by using scale invariant feature transform (SIFT) to find corresponding feature points of pair images. This method required a reference image and the height value for each building to extract flood information, and the flooded building image should be taken with a similar orientation and location as the reference image. Geetha et al. (2017) suggested a method using human objects for identifying the extent of flood. This method relied on front-facing humans and water with brown color, making it difficult to apply to various situations. Meng et al. (2019) proposed a method to detect human keypoints by Mask-RCNN and utilized inferred average heights of detected persons. However, flood depth could only be obtained from humans with few obstacles because keypoints were not accurately detected when there were obstacles. Chaudhary et al (2019) considered five categories (person, car, bus, bicycle, and house) whose dimensions were approximately known. The proposed method predicted water level for objects in images by Mask-RCNN, and finally computed a global water level in a single image, achieving an average error of 8.56 cm. It was innovative in that the presented model recognized instance objects and at the same time predicted their water levels. However, this method is limited because of their use of the average human height and the classification approach; the height of a person can be significantly different from the average height, and the classification approach can result in an error when the height of the object is close to a classification boundary.

Previous studies for flood management used various types of data, which can be obtained from a range of devices, including personal mobile devices and radars installed on satellites. These studies involved flood event detection, flood area segmentation, and flood depth estimation. However, further research is needed to advance image analysis for flood depth estimation. With the widespread use of camera-equipped devices, such as smartphones, the number of public images has increased rapidly. There is a strong need to capitalize on the vast image data to assess each city district's vulnerability to flood damage. Therefore, we propose a deep learning–based method for flood depth estimation based on public flood image data.

Methodology

Segmentation and Filter

The proposed method uses Mask R-CNN, which is a state-of-the-art method for object detection and segmentation (He et al. 2017), for instance on the level recognition of flooded vehicles. Mask R-CNN is an improved version of Faster R-CNN (Ren et al. 2015), with the addition of segmentation functionality. The structure of Mask R-CNN comprises a backbone architecture for feature extraction and a network head for object recognition. The backbone architecture, i.e., the first half of the network, extracts feature maps from the input image, and the second half, i.e., the network head, performs classification, regression, and mask prediction using the extracted feature maps. The three tasks of the network head are performed in parallel for the regions of interest generated by the region proposal network. RoIAlign—a method for converting integer pixel unit into real values using interpolation of relevant weights—is used for Mask R-CNN to perform accurate localization of instances. In the experiment of He et al. (2017), excellent results were achieved with regard to accuracy and speed when a combination of ResNet and a feature pyramid network (FPN) (Lin et al. 2017) were used as the backbone architecture. The proposed method in this paper relies on Mask R-CNN with ResNet101-FPN to obtain vehicle objects for flood depth estimation.

The vehicles detected by Mask R-CNN are filtered according to three conditions and used in the next step of the proposed method. First, the vehicle is not used when it is occluded by another object. Second, the vehicle is not used when only a part of it is extracted because it is located at the edge of the image. Finally, the vehicle is not used when it has a width-to-height ratio lower than that of the 3D vehicle model. Filtering is performed according to the bounding-box information and segmentation information of the objects from Mask R-CNN. A bounding box is an axis-aligned rectangular box surrounding an object, and the position and size of the object are represented by the coordinates of the upper left and the coordinates of the lower right. Segmentation refers to the prediction result in pixel units in the bounding box.

Fig. 2 illustrates the filtering conditions using the values obtained by the bounding box and segmentation information.
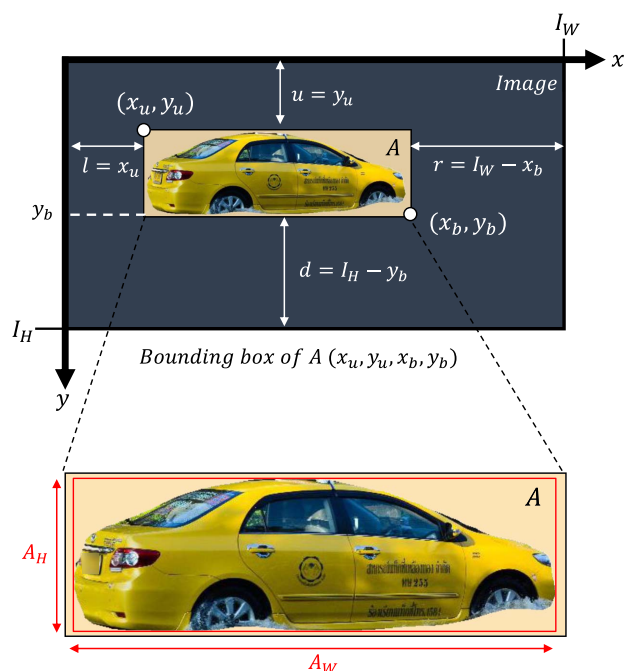
© ASCE   04020072-3   J. Comput. Civ. Eng.

J. Comput. Civ. Eng., 2021, 35(2): 04020072

Fig. 2. Parameters used for filtering the dataset. (Image by suteenakin/Shutterstock.com.)

Fig. 3. Assumption for the flooded part of the vehicle based on the 3D-rendered image that is similar with regard to the shape and orientation of the flooded vehicle: (a) flooded vehicle image (image by Thor Jorgen Udvang/Shutterstock.com); (b) 3D-rendered image (image by Somin Park); and (c) combined image for flood depth estimation.

First, to deal with occlusion situations, an occluded vehicle object must be removed using the bottom line values (y_b) of the two bounding boxes. The object with the upper bottom line is considered to be covered by the other object. Thus, a vehicle object with smaller y_b of two overlapping objects is excluded from the next filtering. This filtering determines whether a vehicle object is occluded by all kinds of objects detectable by Mask-RCNN. Second, the distance between the bounding box and the border of the image is considered to determine whether the object is at the edge of the image. If the smallest value of u, l, r, and d is zero, it is excluded from the analysis. Finally, to identify the objects that can provide flood depth information, the width-to-height ratios ($r_{w/h}$) of the objects are calculated. To use the correct ratio of the object, instead of the height and width of the bounding box, the width (A_W) and height (A_H) values of the segmented area are used (Fig. 2).

The overall width of the vehicle is maintained when the vehicle is partially flooded; however, the height of the vehicle tends to decrease as the flood depth increases. Therefore, it is assumed that the difference in the height of the vehicle between the two images indicates the depth of the flood when the width of the selected pair images is equal. Fig. 3 illustrates this assumption. Fig. 3(a) depicts an example where the height of the vehicle has been affected by flooding, and Fig. 3(b) depicts a 3D rendered image matched with it. When the widths of the vehicles in the two images are adjusted equally, the flooded part of the vehicle is the same as the lower part of Fig. 3(b). Thus, the whole view of the vehicle is the same as the composite of the two images, as illustrated in Fig. 3(c). When the two images have the same width, the height of the flooded vehicle should be lesser than the height of the 3D rendered image, and in this study, the difference is suggested to be the flood depth. Thus, if the $r_{w/h}$ of the flooded-vehicle image is greater than or equal to that of the 3D rendered image, this information can indicate the flood depth. To illustrate this, Fig. 4 depicts two image pairs in which the vehicles have the same orientation but different shapes. The flooded vehicles depicted in Figs. 4(a and b) have the same orientation as the 3D rendered images depicted in Figs. 4(c and d)
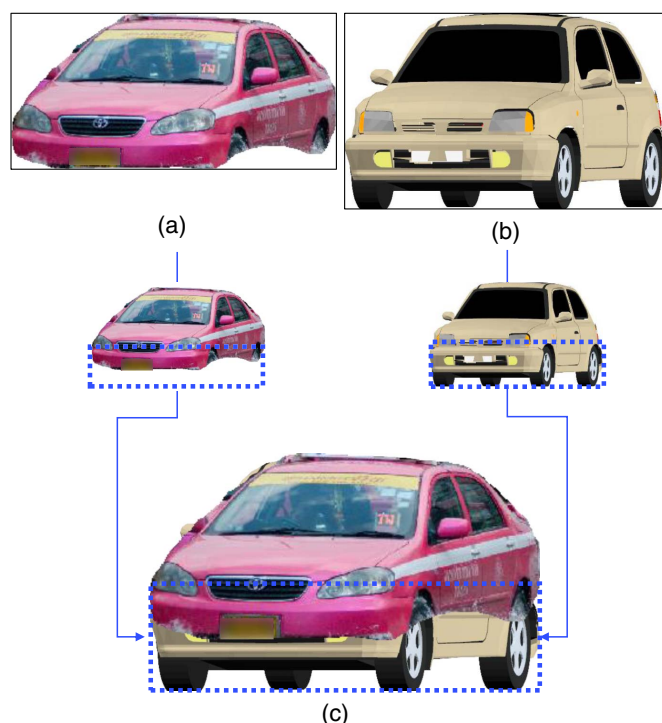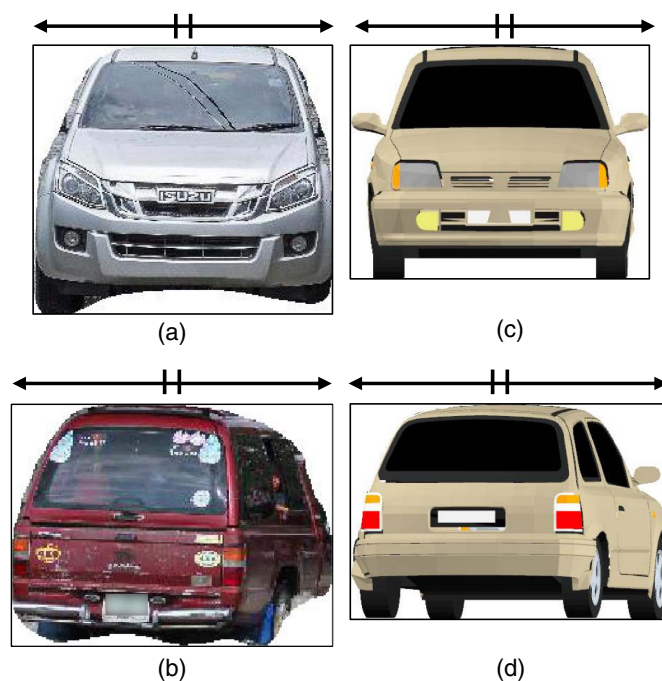

Fig. 4. Two pairs of images: (a and b) flooded vehicle images; and (c and d) 3D-rendered images that are similar to the vehicle images with regard to the vehicle orientation but different with regard to the vehicle shape. [Car image in (a) by PongMoji/Shutterstock.com; car image in (b) by Charlie Edward/Shutterstock.com; 3D-rendered images in (c) and (d) by Somin Park.]

© ASCE   04020072-4   J. Comput. Civ. Eng.

J. Comput. Civ. Eng., 2021, 35(2): 04020072

(but a different shape). Because the $r_{w/h}$ of the flooded vehicle is lesser than that of the corresponding 3D rendered image, the height of the flooded vehicle is greater than the 3D rendered image. Therefore, the flooded vehicles in Fig. 4 are found to be of a different type than the 3D vehicle model and are eliminated from the dataset of the proposed method. The value of $r_{w/h}$ is also used to select the appropriate flood vehicle images. As depicted in Fig. 4, among the various angled 3D rendered images, the front-facing image has the smallest value of $r_{w/h}$. Therefore, if the $r_{w/h}$ value of the detected flooded vehicle is less than that of the front 3D image, the vehicle type differs between the two images. Thus, regardless of the shape and orientation, the flood depth information cannot be provided. This data is removed without being sent to the next step of the proposed method. Vehicles, with visually different features from the 3D vehicle model but not excluded by the $r_{w/h}$, will be processed in the next section of filter.

Image Retrieval and Filter

Feature maps extracted from the convolutional neural network (CNN) were utilized to evaluate the similarity between the flooded-vehicle images and the 3D rendered images. The CNN extracts the characteristics of the input data by using multiple layers stacked with convolutional and pooling layers. The convolutional layer processes the input data through a fixed-size filter and channels, and the feature map outputs through the layer are used as inputs to the next layer. The pooling layer extracts the main features while reducing the size of the convolutional layer output. Ha et al. (2018) proposed a method for vision-based indoor localization through cross-domain image retrieval using the feature maps extracted from the pooling layer. The method proposed evaluates the similarity between two images by vectorizing the feature maps of two images that pass through a pretrained CNN and calculating the cosine distance of the two vectors. Park et al. (2019) presented a method for generating 3D rendered images with different perspectives. By exploiting the methods (Ha et al. 2018; Park et al. 2019), the 3D rendered image most similar with regard to the direction and viewpoint of the flooded vehicle is retrieved from the dataset created from the 3D vehicle model. The proposed method uses Pooling layers 3 and 4 of VGG16 and VGG19, which yield the best results (Ha et al. 2018), to retrieve the 3D rendered image with the most

similar shape and orientation to the flooded vehicle. In detail, the feature maps of the flooded-vehicle images and rendered 3D images are vectorized and compared using the cosine distance; the rendered 3D image with the lowest cosine distance value is retrieved.

The proposed method uses a modified version of the technique proposed by Ha et al. (2018) for analyzing the flooded-vehicle images by changing the format of the input images. Because the input size of the pretrained VGGNet is fixed at 224×224, resizing of all images is required. To resize an image while maintaining its width-to-height ratio, the image width should be fixed, and the height should be adjusted according to the original ratio. Both the 3D rendered and flooded-vehicle images were scaled to 212 pixels in width, and the resized image with a resolution of 224×224 had a white margin of 6 on the top, left, and right.

For image retrieval, a new method called dynamic feature space selection (DFSS) is applied, in which only a portion of the feature map is used to avoid including the white area underneath the input image generated owing to the constraint of the width-to-height ratio. Fig. 5 depicts the range of the feature maps used for the image retrieval in DFSS. The upper portion of the rectangular prism in the feature maps in Fig. 5(a) represents the relevant information corresponding to the vehicle image in Pooling layer 3; the bottom portion of the feature maps is not relevant, because it corresponds to the white area of the input image. Similarly, Fig. 5(b) illustrates the strategy for feature extraction in Pooling layer 4. As depicted in Fig. 5(a), the spatial size of the output of Pooling layer 3 is 28, and the number of channels is 256. When n is the sum of the height of the flooded vehicle and the top margin of the input image, only the top 1/8 portion along the height axis is used for the calculation. The reason for dividing n by 8 is as follows: when the input data passes through Pooling layer 3, it becomes 28×28 in size, which is 1/8 of its original size. Similarly, as depicted in Fig. 5(b), when the feature map that passed through Pooling layer 4 is used, because the spatial size of the feature map is 1/16 of the input data size, only the top 1/16 of the feature map is used for the similarity calculation. The range of the feature maps determined using the flooded-vehicle images also applies to the 3D rendered images. The height value of the flooded vehicles determines the feature-map dimension of the 3D rendered images used for the similarity calculation.
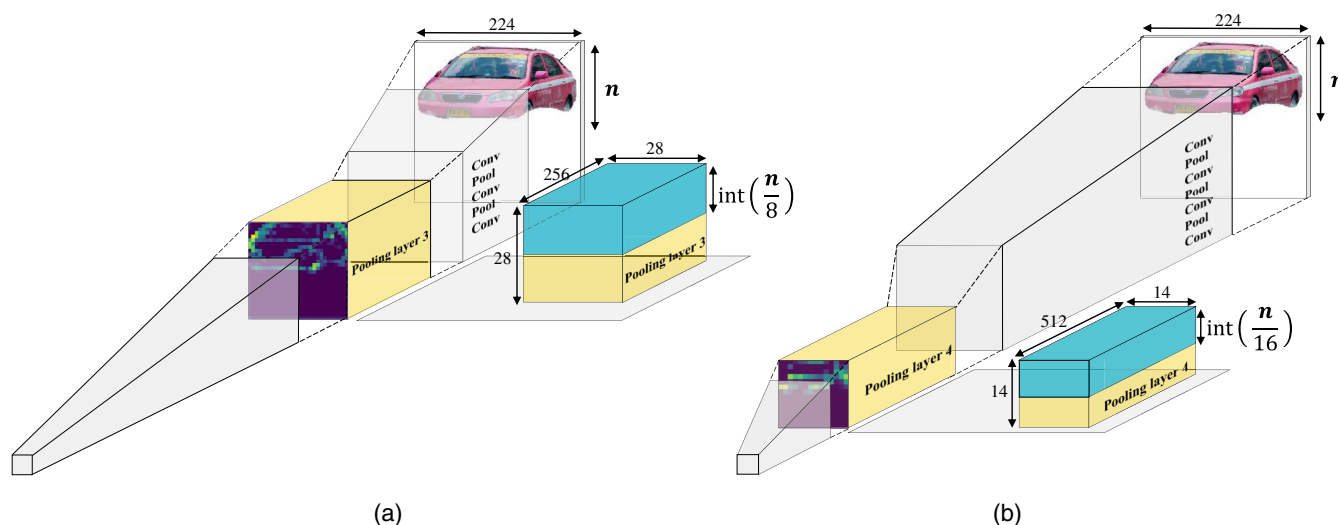

Fig. 5. Feature maps used for image retrieval in the DFSS method: (a) Pooling layer 3 of VGGNet; and (b) Pooling layer 4 of VGGNet. Conv and Pool = convolutional and pooling layers, respectively; numbers denote the spatial resolution and channels; and n = height of the flooded car plus the space at the top. (Car image by Thor Jorgen Udvang/Shutterstock.com.)

© ASCE   04020072-5   J. Comput. Civ. Eng.

J. Comput. Civ. Eng., 2021, 35(2): 04020072

The process of image retrieval excludes data that are not suitable for flood depth estimation. First, the width-to-height ratio condition is applied for the selection of appropriate images. The flooded-vehicle image is excluded when its $r_{w/h}$ is smaller that of the retrieved 3D image. Second, problematic images that are not excluded by the bounding box and segmentation information are excluded from the dataset. Problematic images include objects that are not vehicles or are difficult to discern, vehicles that are only partly detected, and multiple vehicles that are detected as one vehicle. Because problematic images have different visual characteristics from the 3D rendered images, the cosine distance calculated from the comparison is expected to be large. The large cosine distance provides a basis for the removal of such images. Some vehicle images appear as normal vehicle images but in fact are significantly different from the 3D vehicle model used in this study. These vehicle images may provide inaccurate estimation of the flood depth. Thus, the use of the $r_{w/h}$ and cosine-distance criteria also plays the role of eliminating seemingly normal vehicle images with the potential to generate inaccurate estimation of the flood depth.

Flood Depth Estimation

The height difference between the flooded vehicle and the retrieved 3D rendered image is used to estimate the flood depth. The following assumption is made for flood depth estimation: when the vehicle is partially flooded, the original width of the vehicle is maintained, and thus, the difference in vehicle height between the two images of the same width represents the flood depth. However, when flooding is severe, the width of the vehicle in the flooded-vehicle image does not match that of the original vehicle. The non-flooded parts of such images tend to be magnified, as depicted in Fig. 6. Fig. 6(a) depicts the estimated flood depth (d_1) for a flooded-vehicle image in the case of severe flooding, under the assumption that the pair of images have the same width. The area represented by the flooded-vehicle image in Fig. 6(a) is significantly larger than the corresponding part in the retrieved 3D image, and it may not be valid to compare the height between the two images. To address this problem, in the proposed method, the height of the flooded-vehicle image is finely adjusted using the retrieved image information. The width of the flooded vehicle is resized to match the longest horizon value (w_{long}) in the retrieved 3D image; w_{long} is the largest value along the horizontal axis within the vertical range of 0 to the height of the flooded vehicle (h_1), as depicted

in Fig. 6(b). Thus, the enlarged part of the vehicle is appropriately adjusted, resulting in a more accurate estimation of the flood depth (d_2).

Experimental Results

Data Preparation

Two forms of data—flooding images and 3D rendered images—were used in this study. A total of 988 flooding images were obtained from the Shutterstock website. On the website, flooding images containing vehicles were searched for using keywords such as "flooding," "flood road," "flood vehicle," and "flood car." The 988 images were then divided into Datasets 1 and 2. Dataset 1, which consisted of 488 images, was used to determine the conditions for selecting the data suitable for the proposed method, and Dataset 2, which consisted of the remaining 500 images, was used to validate the method.

A total of 44 3D rendered images were generated from a 3D vehicle model. As illustrated in Fig. 7(a), 3D rendered images were created by rotating the 3D vehicle model on a horizontal plane. In this way, 36 3D rendered images were generated from the 3D vehicle model rotated 360° in 10° increments. Figs. 7(b and c) depict the 3D rendered images obtained from the side and front perspectives, respectively. While the difference in appearance among
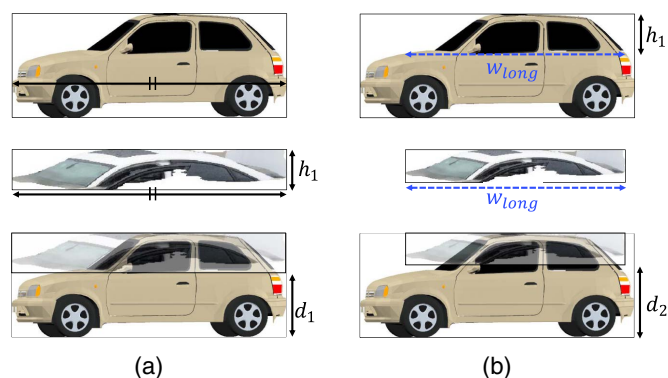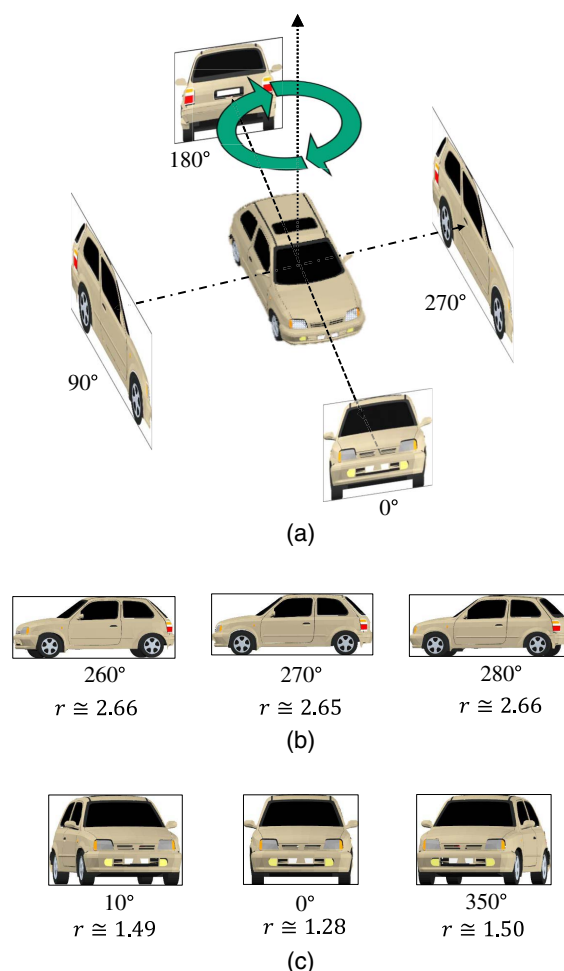

Fig. 6. Comparison for flood depth estimation: (a) images without fine adjustment; and (b) images with fine adjustment, resized according to the longest horizon. Variable w_{long} represents the longest horizon in the h_1 range; and h_x and d_x represent the height of the flooded-vehicle image and the flood depth, respectively. (Car image by Trong Nguyen/Shutterstock.com; 3D-rendered images by Somin Park.)

Fig. 7. Generation process for the 3D rendered images: (a) 3D car model; (b) three side images; and (c) three front images, where r represents the width-to-height ratio. (Images by Somin Park.)
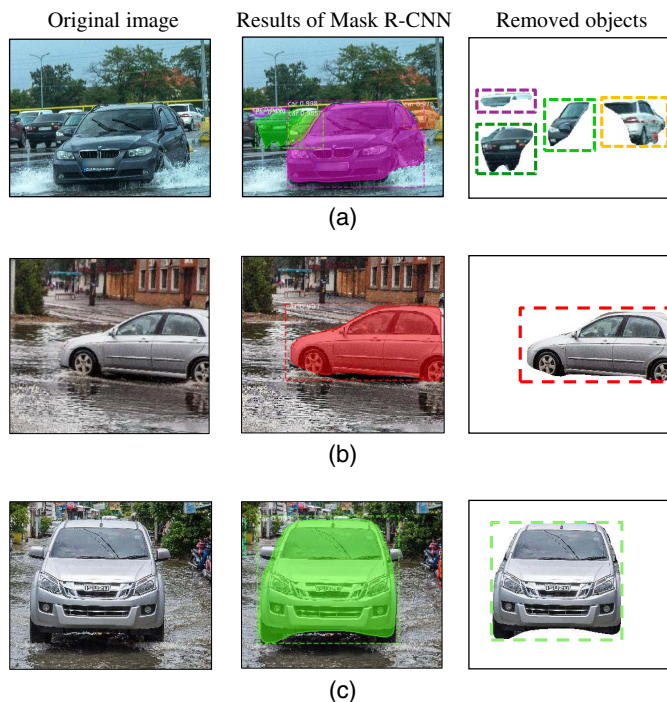
© ASCE 04020072-6 J. Comput. Civ. Eng.

J. Comput. Civ. Eng., 2021, 35(2): 04020072

Original image   Results of Mask R-CNN   Removed objects

(a)

(b)

(c)

Fig. 8. Three filtering conditions: (a) occlusion; (b) location at the edge of the image; and (c) width-to-height ratio of the object. [Images in (a and b) by A_Lesik/Shutterstock.com; image in (c) by PongMoji/Shutterstock.com.]

images obtained by rotating 10° from the side was not large, the appearance of the rendered images rotated 10° exhibited a significant difference. This difference is represented by the ratio of the width to the height, which is depicted at the bottom of the 3D rendered images in Figs. 7(b and c). To obtain accurate results for image retrieval and flood depth estimation, the proposed method obtained additional 3D rendered images at 5° intervals from the front and back. Using the cylindrical coordinate system depicted in Fig. 7(a), four additional images were rendered between 340° and 20° and between 160° and 200°, respectively, resulting in a total of 44 3D rendered images.

Segmentation and Filter

The Mask R-CNN used in this study was based on a ResNet101-FPN backbone and was pretrained with the COCO dataset (Lin et al. 2014), using 80,000 training data. The vehicle detection in the flooding images was based on the Mask R-CNN implementation by Matterport Inc., which was released under an MIT License (Abdulla 2017). The COCO dataset consisted of 80 classes, and only the objects of the "car" class were used for flood depth estimation. When Dataset 1 (composed of 488 images) was fed into the Mask R-CNN, a total of 1,892 objects were detected as "car." The 1,892 objects were then filtered to remove objects inappropriate for the remainder of the proposed method. As mentioned previously, three types of objects were not suitable for the flood depth estimation: vehicle images occluded by other objects, vehicle images located at the edge of the image, and vehicle images with a $r_{w/h}$ value smaller than that of the front 3D image [approximately 1.28, as depicted in Fig. 7(c)]. Filtering out these objects left 709 of the 1,892 objects. Fig. 8 depicts examples of the original images, the Mask R-CNN results, and the results with the objects removed. Fig. 8(a) depicts the exclusion of four vehicles occluded by other objects. In Fig. 8(b), the object located at the right edge of the image is excluded. In Fig. 8(c), the object is excluded because its $r_{w/h}$ value is smaller than that of the front 3D image.

Image Retrieval and Filter

The cosine distance computed in the image retrieval process was used to filter the dataset before performing flood depth estimation. The purpose of this filtering was to select a dataset suitable for flood depth estimation. The proper dataset should not contain problematic objects and should have visual features sufficiently similar to those of the 3D vehicle model. Fig. 9 depicts examples of problematic objects. The images in Fig. 9(a) were detected as a vehicle, but it was difficult to determine whether a vehicle was present. Fig. 9(b) depicts that the objects blocking the vehicle were not detected; thus, the vehicle images were not filtered out. In some cases, multiple vehicles were detected as a single vehicle, as depicted in Fig. 9(c).

The cross-domain comparison matched the 709 vehicle objects to their corresponding 3D rendered images, and the aforementioned problematic images were excluded by comparing the $r_{w/h}$ values. The number of images selected in this manner depended on the CNN model and pooling layer used; it ranged between 532 and 543. Table 1 presents the distributions of the numbers of vehicles and problematic objects determined by the cosine-distance value, which was used as the criterion for image retrieval. With the same value of the cosine distance, more vehicle objects and problematic objects were identified using Pooling layer 3 than when Pooling layer 4 was used. The data distribution in Table 1 indicates that 10 cases (combined conditions of cosine distance and pooling
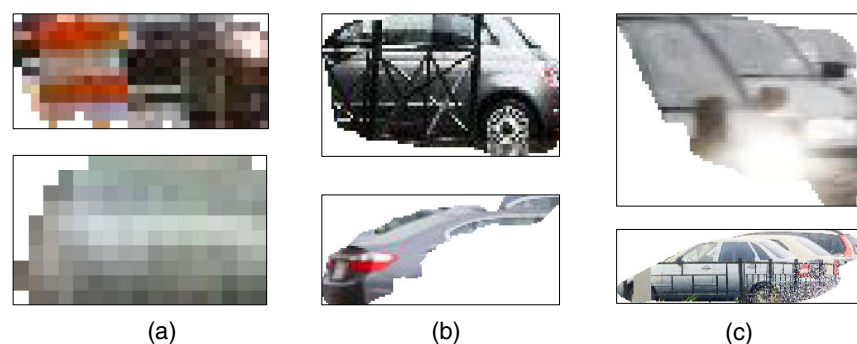

(a)   (b)   (c)

Fig. 9. Examples of problematic objects. [Images in (a) by ungvar and John Gomez/Shutterstock.com; images in (b) by Stefano Guidi and alenka2194/Shutterstock.com; images in (c) by A_Lesik and alenka2194/Shutterstock.com.]

Table 1. Data distribution based on the cosine distance and the layers for feature extraction in Dataset 1

CNN model	Layer for feature extraction	Number of vehicle objects (number of problematic objects)									
		$d_{cosine} < 0.30$	$d_{cosine} < 0.35$	$d_{cosine} < 0.40$	$d_{cosine} < 0.45$	$d_{cosine} < 0.50$	$d_{cosine} < 0.55$	$d_{cosine} < 0.60$	$d_{cosine} < 0.65$	$d_{cosine} < 0.70$	$d_{cosine} \leq 1.00$
VGG16	Pooling layer 3	—	—	—	1 (0)	13 (1)	63 (2)	251 (6)	497 (86)	542 (116)	543 (118)
	Pooling layer 4	—	—	—	2 (0)	8 (0)	**31 (0)**	**93 (0)**	245 (1)	412 (11)	541 (100)
VGG19	Pooling layer 3	2(0)	34 (1)	196 (2)	458 (66)	535 (109)	540 (114)	540 (114)	540 (114)	540 (114)	540 (114)
	Pooling layer 4	—	—	—	1 (0)	6 (0)	**30 (0)**	**101 (0)**	267 (1)	421 (16)	532 (90)

Note: d_{cosine} = cosine distance. Four cases in boldface type were used to highlight the layer for feature extraction and represent threshold values of cosine distance. These threshold values were applied to Dataset 1 and Dataset 2 to estimate the flood depth in this paper.
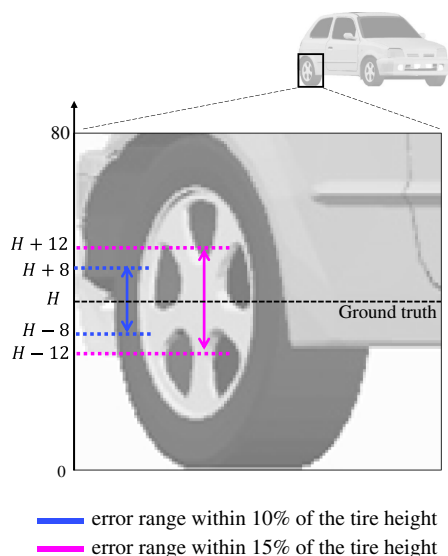
Fig. 10. Error ranges in flood depth estimation. (Images by Somin Park.)

layer) did not contain problematic objects. Because six of the 10 cases had insufficient data, four cases were used to determine the layer for feature extraction and the threshold value of the cosine distance. The four cases are in boldface type in Table 1.

Flood Depth Estimation

To evaluate the flood depth estimation, the height of all the retrieved 3D rendered images was changed to 200 pixels, and the size of the flooded-vehicle image was changed via the method described in the section "Flood Depth Estimation." The height difference between the two resized images was the estimated flood depth. In the 3D rendered images, which were 200 pixels in height, the tire was 80 pixels tall. Errors of 8 and 12 pixels corresponded to 10% and 15% of the tire height, respectively. Fig. 10 graphically shows the pixel-level error that occurs in flood depth estimation.

When the ground truth of the flood depth is H, the estimated flood depth with an error of <8 pixels lies between $H-8$ and $H + 8$. The longer arrow indicates the extent to which the estimated flood depth exists with an error of <12 pixels.

Table 2 presents the average error resulting from the flood depth estimation for four cases in three values: the error in pixels, the error in centimeters, and the ratio of the error to the tire height. To calculate the error in centimeters, tire height in centimeters was determined by examining the tire heights for 154 vehicle types (Automobiledimension.com 2020). The ground-truth values for the vehicle flood depth were determined via manual photographic readings. The best of the four cases was the combination of Pooling layer 4 of VGG19 with a cosine distance of <0.55 (V19-p4-d55). The remaining three results yielded similar mean errors; the largest mean error was observed for Pooling layer 4 of VGG19 with a cosine distance of <0.60 (V19-p4-d60). When image retrieval was performed with Pooling layer 4 of VGG16, the case with a threshold of 0.55 for the cosine-distance value (V16-p4-d55) had less data but a larger average range of the error than the case with a cosine distance of 0.60 (V16-p4-d60).

Dataset 2

Dataset 2, which consisted of 500 additional flooding images, was used in the process from segmentation to flood depth estimation for validation of the proposed method. A total of 1,898 vehicles were extracted from the 500 images, and 695 objects were used for retrieval after the first filtering. Table 3 presents the flood estimation results based on Dataset 2. In this paper, we used the four cases (V16-p4-d55, V16-p4-d60, V19-p3-d55, and V19-p4-d60) that were found to be the best for flood depth estimation with Dataset 1. The V19-p4-d55, which exhibited the best results with Dataset 1, also exhibited the smallest mean error corresponding to 9.40% of the tire height.

Fig. 11 depicts the frequency and cumulative percentage of the error generated by the flood depth estimation. The vertical blue bars in the histograms indicate the number of objects with a given error, i.e., the error in flood depth estimation (in pixel units). The rightmost bar corresponds to the largest error for the case. The V16-p4-d60, which exhibited the largest mean error among the four cases, also exhibited the largest error value (40 pixels) among the four

Table 2. Error in estimating the flood depth according to the longest horizon in Dataset 1

CNN model	Layer for feature extraction	$\text{avg}\Delta H \mid \text{avg}\Delta H \cdot \dfrac{\text{tire height in cm}}{\text{tire height in pixels}} \mid \dfrac{\text{avg}\Delta H}{\text{tire height in pixels}} \mid$ number of vehicle objects	
		$d_{cosine} < 0.55$	$d_{cosine} < 0.60$
VGG16	Pooling layer 4	9.65 pixels | 8.32 cm | 12.06% | 31	9.54 pixels | 11.93 cm | 11.93% | 93
VGG19	Pooling layer 4	8.43 pixels | 7.27 cm | 10.54% | 30	9.76 pixels | 8.42 cm | 12.20% | 101

Note: H = ground truth of flood depth; h = estimated flood depth; $\Delta H = |H - h|$, error; $\text{avg}\Delta H$ = average of ΔH; and d_{cosine} = cosine distance.

© ASCE   04020072-8   J. Comput. Civ. Eng.

J. Comput. Civ. Eng., 2021, 35(2): 04020072

Table 3. Error in estimating the flood depth according to the longest horizon in Dataset 2

CNN model	Layer for feature extraction	avgΔH|avg$\Delta H \cdot \dfrac{\text{tire height in cm}}{\text{tire height in pixels}}$|$\dfrac{\text{avg}\Delta H}{\text{tire height in pixels}}$| number of vehicle objects	
		$d_{\text{cosine}} < 0.55$	$d_{\text{cosine}} < 0.60$
VGG16	Pooling layer 4	9.59 pixels | 8.27 cm | 11.99% | 27	11.01 pixels | 9.50 cm | 13.76% | 110
VGG19	Pooling layer 4	7.52 pixels | 6.49 cm | 9.40% | 29	10.51 pixels | 9.06 cm | 13.14% | 112

Note: H = ground truth of flood depth; h = estimated flood depth; $\Delta H = | - h|$, error; avgΔH = average of ΔH; and d_{cosine} = cosine distance.

cases, as depicted in Fig. 11(b). The largest error (25 pixels) of V19-p4-d55, which exhibited the smallest average error, was the smallest among the four cases, as depicted in Fig. 11(c). The line graphs indicate the cumulative percentage of the objects at or below a given error. The four graphs show where the 8 and 12 pixels (errors of 10% and 15%, respectively, of the tire height) are located in the cumulative curves. For example, as depicted in Fig. 11(c), the cumulative rates 68.97% and 86.21% indicate that approximately 69% of the data have an error of ≤8 pixels and approximately 86% of the data have an error of ≤12 pixels, respectively. These two percentages are the largest among the corresponding values of the four cases. This data distribution results in the V19-p4-d55 case containing relatively little data, with extremely large errors and the smallest mean error, as presented in Table 3.

Discussion

Image Retrieval and Flood Depth Estimation

In this study, only the four cases of Pooling layer 4 of VGGNet were used, among all the combinations for flood depth estimation.

The reason for this selection was that the threshold values (cosine distance) of the cases with Pooling layer 4 in Dataset 1 could completely exclude problematic objects; with Pooling layer 3, one or more problematic objects were included in the subset determined by the smaller threshold values. These problematic objects were either small enough to be difficult to discern, or had a very simple structure. Fig. 12 illustrates the feature maps for three pairs of images extracted from Pooling layers 3 and 4 of VGG19. As indicated by the feature maps in Figs. 12(a–c), Pooling layers 3 and 4 both represent structural characteristics in the images. However, the feature maps extracted from Pooling layer 3 provide more detailed information than those extracted from Pooling layer 4. This difference is explained as follows: the area represented by one grid in the feature maps of the later pooling layer is larger than the area represented by one grid in the feature maps of the front layer. Because the feature maps in Pooling layer 4 generate larger approximations for the same area, the real image must have a clear structural similarity to have a high similarity to the 3D rendered image. It is interpreted that the problematic objects did not satisfy the similarity condition for the feature maps in Pooling layer 4, leading to large cosine distances.
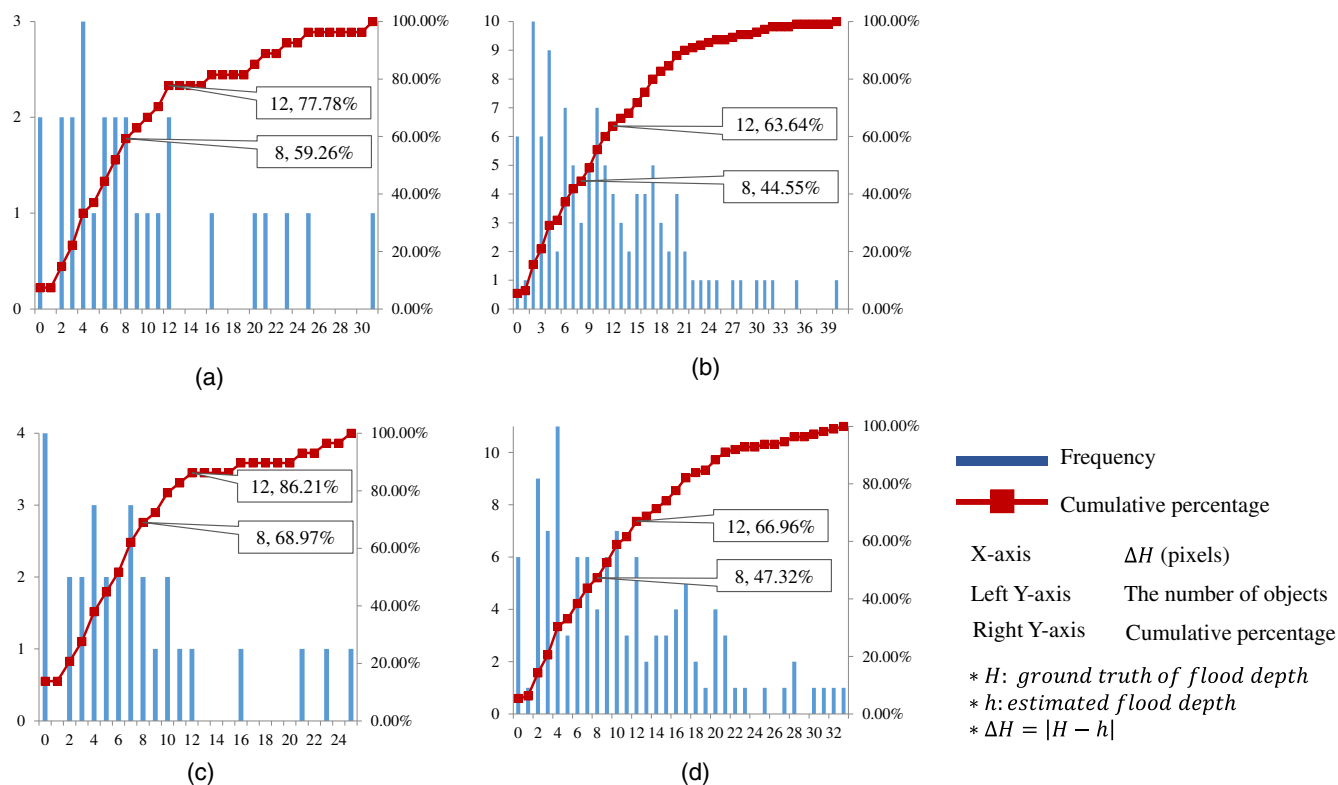

Fig. 11. Histograms and cumulative histograms of the flood depth estimation error derived from (a) V16-p4-d55; (b) V16-p4-d60; (c) V19-p4-d55; and (d) V19-p4-d60.
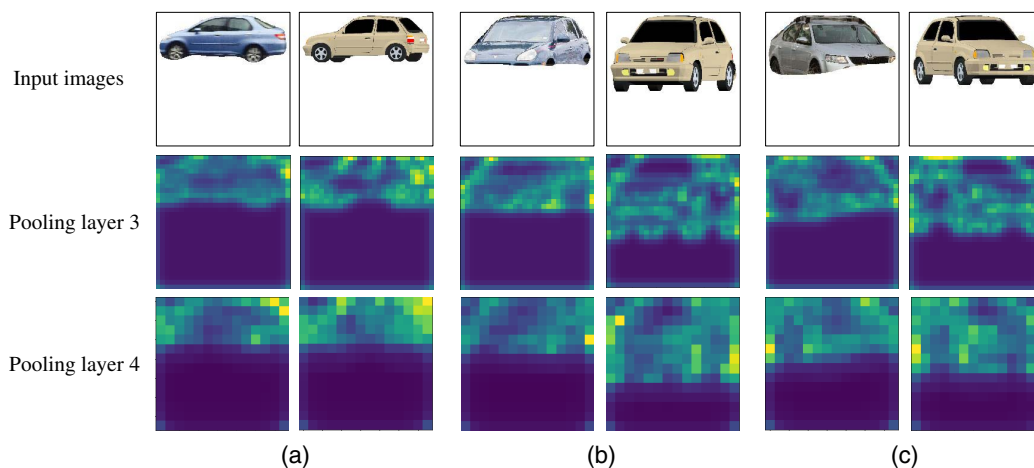
Fig. 12. Visualization of the VGG19 network feature maps. [Car image in (a) by Kae B Yuki/Shutterstock.com; car image in (b) by michelmond/Shutterstock.com; car image in (c) by Pinhead Studio/Shutterstock.com; 3D-rendered images by Somin Park.]

In Table 3, two cases, under the condition of a cosine distance of <0.55, showed smaller mean error values than the other two cases with a cosine distance of <0.60. For the V16-p4-d55 and V19-p4-d55, the mean error values of 8.27 and 6.49 cm were calculated, respectively. The V16-p4-d60 and theV19-p4-d60 had mean error values of 9.50 and 9.06 cm, respectively. The dataset selected by the cosine distance of <0.55 had a higher similarity to the 3D rendered images than the dataset filtered by the cosine distance of <0.60. It can be explained that a dataset with a higher similarity to 3D rendered images contributed to improving the accuracy of flood depth estimation. Therefore, the image retrieval section, which includes the processes of comparing two types of images and deriving the similarity between them, plays an important role in the proposed method for flood depth estimation.

Causes of Flood Depth Estimation Error

Errors in flood depth estimation occurred for the following reasons: incorrect segmentation of vehicles, approximate feature extraction in image retrieval, and varied locations for estimating the flood depth. Fig. 13 depicts the three reasons for the large errors in the case of V19-p3-d55. First, the flood depth of the vehicle in Fig. 13(a) was estimated to be smaller than the actual value, owing to incorrect segmentation. The polygon in Fig. 13(a) represents a portion of a vehicle reflected on the flooding water. The similarity between the reflected part on the water and the corresponding part of the vehicle caused the segmentation result to be incorrect and the flood depth to be estimated inaccurately. Second, it is surmised that excessively simplified feature maps impaired the similarity comparison between the two images in the retrieval process. Vehicles of different types from the 3D vehicle model are used, and as a result, a large error can occur. Fig. 13(b) depicts an example of a vehicle with a measured flood depth greater than the actual value. The upper portion of the rectangular prism in Fig. 13(b) indicates the range of the feature maps used to calculate the cosine distance between the two images. Owing to the usage of only the top four grids of the feature map, the paired images were included in the dataset for flood depth estimation; they would have been excluded from the dataset with the use of features having sufficient details. Finally, the different locations for flood depth estimation caused errors. Fig. 13(c) depicts an example where the flood depth was measured to be larger than the actual value. In the flooded-vehicle image in Fig. 13(c), the flood height was determined by the rear tire, in comparison with the 3D rendered image; however, the

ground truth of the flood depth was determined by the front tire part, which hardly flooded, as indicated by the boxes around the tires in Fig. 13(c). Owing to the difference between the two values, a large error occurred in the flood depth estimation. In the proposed
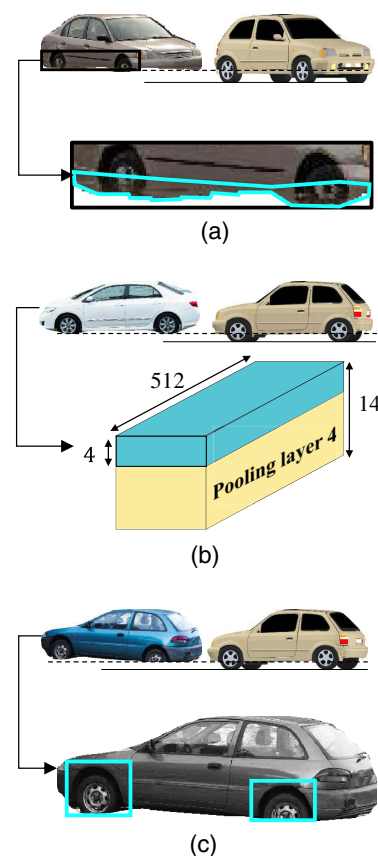

Fig. 13. Three causes of flood depth estimation errors: (a) incorrect segmentation; (b) simplified feature maps; and (c) varied locations for flood depth estimation. In each pair of images, the distance between the dotted line and solid line indicates the estimated flood depth. The numbers denote the spatial resolution and channels. [Car image in (a) by Marc Bruxelle/Shutterstock.com; car image in (b) by Bubbers BB/Shutterstock.com; car image in (c) by Dariush M/Shutterstock.com; 3D-rendered images by Somin Park.]

method, the ground truth of the flood depth was determined according to the least flooded portion of the flooded vehicle. Thus, estimation errors were generated when the height difference was not measured for the least flooded part.

The first of the three causes of error was related to the performance of the deep-learning model for segmentation. This problem can be solved by training the deep-learning model with additional data to distinguish features between the nonflooded part of the car and the reflections in the flooding water. For addressing the other two causes, further works are needed to better assess the similarity between the two images. As an extension of the proposed method, feature maps extracted from multiple layers can be analyzed for the comparison analysis. A new metric can be devised to replace the cosine distance for improved comparison of two vectorized feature maps.

Conclusion

This paper proposed a deep learning–based method for flood depth estimation using flooded-vehicle images and 3D rendered vehicle images. Mask R-CNN was used to detect vehicle objects, and VGGNets were used to search for a 3D rendered image similar to the flooded-vehicle image. A comparison between the two images after the scale adjustment enabled the estimation of the flood depth. Additionally, the results from both networks were used as information to filter out problematic objects. A total of 988 flooding images were divided into two datasets: Dataset 1 for the model development, and Dataset 2 for the validation. With Dataset 1, four candidate models consisting of different cosine-distance criteria and VGGNets were selected for feature extraction and flood depth estimation. With Dataset 2, the best result was obtained by Pooling layer 4 of VGG19, with a cosine distance of <0.55; the flood depth estimation resulted in an average error value corresponding to 9.40% of the tire height.

The contributions of this study are as follows. This is the first study in which the flood depth is automatically detected using ground-level flood images. The proposed method can be used to analyze flooding images in a timely and cost-efficient manner because no additional information is required to determine the flood depth. The method is useful for dealing with the large amount of image data resulting from the proliferation of smartphones and the increasing use of social media. Second, the DFSS method was suggested to accurately analyze flooded-vehicle images. In the DFSS, the range of feature maps were dynamically determined according to the size of the flooded-vehicle image. Owing to DFSS, the proposed method can easily be applied to other classes included in the flood images, such as bicycles, buses, and people. The method can also be used in other research fields to obtain information from images with occluded objects.

The proposed method has the following limitations. First, it depends on the visual characteristic of objects. When a detected vehicle is covered by objects not dealt with by the deep-learning model for segmentation, even if both the segmentation and the image retrieval are successfully performed on the detected object, the flood depth information may be inaccurate. Second, flooding information from a single image may be insufficient. If most of a vehicle body is covered by splash water but the vehicle is not actually flooded, the situation is still considered extremely serious. Third, only one type of 3D vehicle model is used in this method. Therefore, if different vehicle types such as trucks and SUVs are detected in flood scenes, a certain amount of error can occur in the flood depth estimation.

According to these limitations, in future studies, it is recommended to exploit large amounts of data for fine-tuning of deep learning models. Additional 3D rendered images for different types of vehicles will help to reduce errors of flood depth estimation. Generative adversarial networks (GANs) can be leveraged for increasing the amount of data available. If the data generation capability of GANs is successfully harnessed, the flood depth estimation may be possible using the generated image of entire vehicle. Finally, in order to deal with impacts of hydrodynamic forces, use of video data for tracking a vehicle can be considered. The use of successive frames of video can increase the accuracy of flood depth estimation by detecting the effect of hydrodynamic forces.

In addition to the future studies, the use of vehicle objects and other urban objects together in estimating flood depth may be considered. The vehicle is accessible on any road section as a moveable object, and this feature suggests that the road range covered by the vehicle images is very wide. However, the impacts of the hydrodynamic forces acting on the vehicle, as mentioned in the previous paragraph, can affect the accuracy of the water depth estimation. This challenge can be dealt with by using other urban objects as references for flood depth estimation. For example, urban infrastructures such as hydrants and traffic signs are known for their appropriate heights. Unlike the vehicle object, they can avoid the impacts of hydrodynamic forces as immoveable objects. However, these objects may not be identified in all flooded road images because they are installed apart from each other along the side of the road. If all these objects are analyzed together, the pros depending on the objects can be utilized and the cons depending on the objects can be complemented. With upgrades, the improved method is expected to yield accurate flood depth information in an efficient manner.

Data Availability Statement

All models or code that support the findings of this study are available from the corresponding author upon reasonable request.

Acknowledgments

References

Abdulla, W. 2017. "Mask R-CNN for object detection and instance segmentation on Keras and TensorFlow." Accessed January 1, 2020. https://github.com/matterport/Mask_RCNN.

Aki, M., T. Rojanaarpa, K. Nakano, Y. Suda, N. Takasuka, T. Isogai, and T. Kawai. 2016. "Road surface recognition using laser radar for automatic platooning." *IEEE Trans. Intell. Transp.* 17 (10): 2800–2810. https://doi.org/10.1109/TITS.2016.2528892.

Automobiledimension.com. 2020. "New car dimensions in the European market with the photo of each automobile size showing length, width and height." Accessed July 28, 2020. https://www.automobiledimension.com.

Cetin, M., K. Iftekharuddin, C. Sazara, J. Goodall, and Y. Shen. 2019. *Estimating road inundation levels due to recurrent flooding from image data.* Springfield, VA: National Technical Information Service.

Chaudhary, P., S. D'Aronco, M. Moy de Vitry, J. P. Leitão, and J. D. Wegner. 2019. "Flood-water level estimation from social media images." *ISPRS Ann. Photogramm. Remote Sens. Spatial Inf. Sci.* IV-2/W5 (2/W5): 5–12. https://doi.org/10.5194/isprs-annals-IV-2-W5-5-2019.

Choi, M., R. Starbuck, S. Lee, S. Hwang, S. Lee, M. Park, and H.-S. Lee. 2018. "Distributed and interoperable simulation for comprehensive disaster response management in facilities." *Autom. Constr.* 93 (May): 12–21. https://doi.org/10.1016/j.autcon.2018.05.007.

D'Andrea, E., P. Ducange, B. Lazzerini, and F. Marcelloni. 2015. "Real-time detection of traffic from Twitter stream analysis." *IEEE Trans. Intell. Transp.* 16 (4): 2269–2283.

Diakakis, M., and G. Deligiannakis. 2017. "Flood fatalities in Greece: 1970–2010." *J. Flood Risk Manage.* 10 (1): 115–123. https://doi.org/10.1111/jfr3.12166.

El-Anwar, O., J. Ye, and W. Orabi. 2015. "Efficient optimization of post-disaster reconstruction of transportation networks." *J. Comput. Civ. Eng.* 30 (3): 04015047. https://doi.org/10.1061/(ASCE)CP.1943-5487.0000503.

Fohringer, J., D. Dransch, H. Kreibich, and K. Schröter. 2015. "Social media as an information source for rapid flood inundation mapping." *Nat. Hazard. Earth Syst.* 15 (12): 2725–2738. https://doi.org/10.5194/nhess-15-2725-2015.

Geetha, M., M. Manoj, A. S. Sarika, M. Mohan, and S. N. Rao. 2017. "Detection and estimation of the extent of flood from crowd sourced images." In *Proc., 2017 Int. Conf. on Communication and Signal Processing (ICCSP)*, 603–608. New York: IEEE.

Ha, I., H. Kim, S. Park, and H. Kim. 2018. "Image retrieval using BIM and features from pretrained VGG network for indoor localization." *Build. Environ.* 140 (May): 23–31. https://doi.org/10.1016/j.buildenv.2018.05.026.

He, K., G. Gkioxari, P. Dollár, and R. Girshick. 2017. "Mask R-CNN." In *Proc., IEEE Int. Conf. on Computer Vision*, 2961–2969. New York: IEEE.

He, K., X. Zhang, S. Ren, and J. Sun. 2016. "Deep residual learning for image recognition." In *Proc., IEEE Conf. on Computer Vision and Pattern Recognition*, 770–778. New York: IEEE.

Hies, T., P. Babu, Y. Wang, R. Duester, H. S. Eikaas, and T. Meng. 2012. "Enhanced water-level detection by image processing." In *Proc., 10th Int. Conf. on Hydroinformatics*. Hamburg, Germany: HIC.

Hochreiter, S., and J. Schmidhuber. 1997. "Long short-term memory." *Neural Comput.* 9 (8): 1735–1780. https://doi.org/10.1162/neco.1997.9.8.1735.

IFRC (International Federation of Red Cross and Red Crescent Societies). 2018. *World disaster report*. Geneva: International Federation of Red Cross and Red Crescent Societies.

Kashani, A. G., P. S. Crawford, S. K. Biswas, A. J. Graettinger, and D. Grau. 2014. "Automated tornado damage assessment and wind speed estimation based on terrestrial laser scanning." *J. Comput. Civ. Eng.* 29 (3): 04014051. https://doi.org/10.1061/(ASCE)CP.1943-5487.0000389.

Kim, H., and Y. Ham. 2019. "Participatory sensing-based geospatial localization of distant objects for disaster preparedness in urban built environments." *Autom. Constr.* 107 (Sep): 102960. https://doi.org/10.1016/j.autcon.2019.102960.

Kramer, M., K. Terheiden, and S. Wieprecht. 2016. "Safety criteria for the trafficability of inundated roads in urban floodings." *Int. J. Disaster Risk Reduct.* 17 (Aug): 77–84. https://doi.org/10.1016/j.ijdrr.2016.04.003.

Krizhevsky, A., I. Sutskever, and G. E. Hinton. 2012. "Imagenet classification with deep convolutional neural networks." In *Proc., Advances in Neural Information Processing Systems*, 1097–1105. Stateline, NV: Neural Information Processing Systems Foundation.

Lin, F., W. Y. Chang, L. C. Lee, H. T. Hsiao, W. F. Tsai, and J. S. Lai. 2013. "Applications of image recognition for real-time water level and surface velocity." In *Proc., IEEE Int. Symp. on Multimedia*, 259–262. New York: IEEE.

Lin, T.-Y., P. Dollár, R. Girshick, K. He, B. Hariharan, and S. Belongie. 2017. "Feature pyramid networks for object detection." In *Proc., IEEE Conf. on Computer Vision and Pattern Recognition*, 2117–2125. New York: IEEE.

Lin, T.-Y., M. Maire, S. Belongie, J. Hays, P. Perona, D. Ramanan, P. Dollár, and C. L. Zitnick. 2014. "Microsoft COCO: Common objects in context." In *Proc., European Conf. on Computer Vision*, 740–755. New York: Springer.

Lin, W.-Y., T.-H. Wu, M.-H. Tsai, W.-C. Hsu, Y.-T. Chou, and S.-C. Kang. 2018. "Filtering disaster responses using crowdsourcing." *Autom.*

Constr. 91 (Mar): 182–192. https://doi.org/10.1016/j.autcon.2018.03.016.

Long, J., E. Shelhamer, and T. Darrell. 2015. "Fully convolutional networks for semantic segmentation." In *Proc., IEEE Conf. on Computer Vision and Pattern Recognition*, 3431–3440. New York: IEEE.

Long, S., T. E. Fatoyinbo, and F. Policelli. 2014. "Flood extent mapping for Namibia using change detection and thresholding with SAR." *Environ. Res. Lett.* 9 (3): 035002. https://doi.org/10.1088/1748-9326/9/3/035002.

Lopez-Fuentes, L., J. van de Weijer, M. Bolanos, and H. Skinnemoen. 2017. "Multi-modal deep learning approach for flood detection." In *Proc., MediaEval, CEUR Workshop*. Aachen, Germany: CEUR-WS.

Ma, T. Y., G. Motta, and K. X. Liu. 2017. "Delivering real-time information services on public transit: A framework." *IEEE Trans. Intell. Transp.* 18 (10): 2642–2656. https://doi.org/10.1109/TITS.2017.2656387.

Mason, D. C., L. Giustarini, J. Garcia-Pintado, and H. L. Cloke. 2014. "Detection of flooded urban areas in high resolution synthetic aperture radar images using double scattering." *Int. J. Appl. Earth Obs. Geoinf.* 28 (May): 150–159. https://doi.org/10.1016/j.jag.2013.12.002.

Medina, R. A., R. B. Mallick, P. Kirshen, L. M. McCarthy, and J. S. Daniel. 2016. "Decision-tree based approach to making post-flooding road opening and closure decisions for transportation agencies." In *Proc., 95th Annual Meeting of the Transportation Research Board*. Washington, DC: Transportation Research Board.

Meng, Z., B. Peng, and Q. Huang. 2019. "Flood depth estimation from web images." In *Proc., 2nd ACM SIGSPATIAL Int. Workshop on Advances on Resilient and Intelligent Cities*, 37–40. New York: Association for Computing Machinery.

Nair, B. B., and S. Rao. 2016. "Flood water depth estimation—A survey." In *Proc., IEEE Int. Conf. on Computational Intelligence and Computing Research*, 1–4. New York: IEEE.

Narayanan, R., V. M. Lekshmy, S. Rao, and K. Sasidhar. 2014. "A novel approach to urban flood monitoring using computer vision." In *Proc., 5th Int. Conf. on Computing, Communications and Networking Technologies (ICCCNT)*, 1–7. New York: IEEE.

Ortigossa, E. S., F. Dias, J. Ueyama, and L. G. Nonato. 2015. "Using digital image processing to estimate the depth of urban streams." In *Proc., Workshop of Undergraduate Works in Conjunction with Conf. on Graphics, Patterns and Images (SIBGRAPI)*. Porto Alegre, Brazil: Brazilian Computer Society.

Park, S., F. Baek, J. Sohn, and H. Kim. 2019. "Deep learning-based vehicle image matching for flooding damage estimation." In *Proc., Creative Construction Conf. 2019*, 38–41. Budapest, Hungary: Budapest Univ. of Technology and Economics & Diamond Congress.

Ploeger, S., M. Sawada, A. Elsabbagh, M. Saatcioglu, M. Nastev, and E. Rosetti. 2015. "Urban RAT: New tool for virtual and site-specific mobile rapid data collection for seismic risk assessment." *J. Comput. Civ. Eng.* 30 (2): 04015006. https://doi.org/10.1061/(ASCE)CP.1943-5487.0000472.

Pouyanfar, S., Y. Tao, S. Sadiq, H. Tian, Y. Tu, T. Wang, S.-C. Chen, and M.-L. Shyu. 2019. "Unconstrained flood event detection using adversarial data augmentation." In *Proc., 2019 IEEE Int. Conf. on Image Processing (ICIP)*, 155–159. New York: IEEE.

Pregnolato, M., A. Ford, S. M. Wilkinson, and R. J. Dawson. 2017. "The impact of flooding on road transport: A depth-disruption function." *Transp. Res. Part D: Transp. Environ.* 55 (Jun): 67–81. https://doi.org/10.1016/j.trd.2017.06.020.

Radmehr, A., and S. Araghinejad. 2014. "Developing strategies for urban flood management of Tehran city using SMCDM and ANN." *J. Comput. Civ. Eng.* 28 (6): 05014006. https://doi.org/10.1061/(ASCE)CP.1943-5487.0000360.

Ren, S., K. He, R. Girshick, and J. Sun. 2015. "Faster R-CNN: Towards real-time object detection with region proposal networks." In *Proc., Advances in Neural Information Processing Systems*, 91–99. Stateline, NV: Neural Information Processing Systems Foundation.

Said, N., K. Pogorelov, K. Ahmad, M. Riegler, N. Ahmad, O. Ostroukhova, P. Halvorsen, and N. Conci. 2018. "Deep learning approaches for flood classification and flood aftermath detection." In *Proc., MediaEval, CEUR Workshop*. Aachen, Germany: CEUR-WS.

Samui Times. 2017. "More than 3,000 cars 'drown' in latest Bangkok flooding—Insurance companies ordered to be 'just.'" Accessed July 28,

© ASCE

04020072-12

J. Comput. Civ. Eng.

J. Comput. Civ. Eng., 2021, 35(2): 04020072

2020. https://www.samuitimes.com/3000-cars-drown-latest-bangkok-flooding-insurance-companies-ordered-just/.

Sazara, C., M. Cetin, and K. M. Iftekharuddin. 2019. "Detecting floodwater on roadways from image data with handcrafted features and deep transfer learning." In *Proc., 2019 IEEE Intelligent Transportation Systems Conf. (ITSC)*, 804–809. New York: IEEE.

Simonyan, K., and A. Zisserman. 2014. "Very deep convolutional networks for large-scale image recognition." Preprint, submitted September 4, 2014. http://arxiv.org/abs/1409.1556.

Szegedy, C., V. Vanhoucke, S. Ioffe, J. Shlens, and Z. Wojna. 2016. "Rethinking the inception architecture for computer vision." In *Proc., IEEE Conf. on Computer Vision and Pattern Recognition*, 2818–2826. New York: IEEE.

Terti, G., I. Ruin, S. Anquetin, and J. J. Gourley. 2017. "A situation-based analysis of flash flood fatalities in the United States." *Bull. Am. Meteorol. Soc.* 98 (2): 333–345. https://doi.org/10.1175/BAMS-D-15-00276.1.

UN (United Nations). 2016. *Policies on spatial distribution and urbanization: Data booklet, statistical papers—United Nations (Ser. A). Population and vital statisticsreport.* New York: UN.

Wang, Z., and S. Zlatanova. 2019. "Safe route determination for first responders in the presence of moving obstacles." *IEEE Trans. Intell. Transp.* 21 (3): 1044–1053. https://doi.org/10.1109/TITS.2019.2900858.

Witherow, M. A., C. Sazara, I. M. Winter-Arboleda, M. I. Elbakary, M. Cetin, and K. M. Iftekharuddin. 2018. "Floodwater detection on roadways from crowdsourced images." *Comput. Methods Biomech. Biomed. Eng. Imaging Vis.* 7 (5–6): 529–540. https://doi.org/10.1080/21681163.2018.1488223.

WMO (World Meteorological Organization). 2013. *Flood forecasting and early warning. APFM Technical Document No. 19—Integrated Flood Management.* Geneva: WMO.

Xu, G. Y., J. W. Wang, G. Q. Huang, and C. H. Chen. 2018. "Data-driven resilient fleet management for cloud asset-enabled urban flood control." *IEEE Trans. Intell. Transp.* 19 (6): 1827–1838. https://doi.org/10.1109/TITS.2017.2740438.

Yang, Y., H.-Y. Ha, F. Fleites, S.-C. Chen, and S. Luis. 2011. "Hierarchical disaster image classification for situation report enhancement." In *Proc., 2011 IEEE Int. Conf. on Information Reuse and Integration*, 181–186. New York: IEEE.

Zheng, Y. J., M. X. Zhang, H. F. Ling, and S. Y. Chen. 2015. "Emergency railway transportation planning using a hyper-heuristic approach." *IEEE Trans. Intell. Transp.* 16 (1): 321–329. https://doi.org/10.1109/TITS.2014.2331239.

Zhou, Z., J. Gong, and X. Hu. 2019. "Community-scale multi-level post-hurricane damage assessment of residential buildings using multi-temporal airborne LiDAR data." *Autom. Constr.* 98 (Nov): 30–45. https://doi.org/10.1016/j.autcon.2018.10.018.

Zhu, J.-Y., T. Park, P. Isola, and A. A. Efros. 2017. "Unpaired image-to-image translation using cycle-consistent adversarial networks." In *Proc., IEEE Int. Conf. on Computer Vision*, 2223–2232. New York: IEEE.

© ASCE 04020072-13 J. Comput. Civ. Eng.

J. Comput. Civ. Eng., 2021, 35(2): 04020072