HOKKAIDO UNIVERSITY

Title	Snow- or ice-covered road detection in winter road surface conditions using deep neural networks
Author(s)	Moroto, Yuya; Maeda, Keisuke; Ogawa, Takahiro et al.
Citation	Computer-aided civil and infrastructure engineering, 39(19), 2935-2950 https://doi.org/10.1111/mice.13233
Issue Date	2024-06-03
Doc URL	https://hdl.handle.net/2115/95313
Rights	This is the peer reviewed version of the following article: Moroto, Y., Maeda, K., Ogawa, T., & Haseyama, M. (2024). Snow- or ice-covered road detection in winter road surface conditions using deep neural networks. Computer-Aided Civil and Infrastructure Engineering, 1-16., which has been published in final form at https://doi.org/10.1111/mice.13233. This article may be used for non-commercial purposes in accordance with Wiley Terms and Conditions for Use of Self-Archived Versions. This article may not be enhanced, enriched or otherwise transformed into a derivative work, without express permission from Wiley or by statutory rights under applicable legislation. Copyright notices must not be removed, obscured or modified. The article must be linked to Wiley's version of record on Wiley Online Library and any embedding, framing or otherwise making available the article or pages thereof by third parties from platforms, services and websites other than Wiley Online Library must be prohibited.
Type	journal article
File Information	CACAIE_Moroto.pdf

Hokkaido University Collection of Scholarly and Academic Papers : HUSCAP

Received: Added at production | Revised: Added at production | Accepted: Added at production

DOI: xxx/xxxx

ARTICLE TYPE

Snow- or Ice-covered Road Detection in Winter Road Surface Conditions using Deep Neural Networks

Yuya Moroto[1] | Keisuke Maeda[2] | Takahiro Ogawa[3] | Miki Haseyama*[3]

[1]Graduate School of Information Science and Technology, Hokkaido University, Hokkaido, Japan
[2]Data-Driven Interdisciplinary Research Emergence Department, Hokkaido University, Hokkaido, Japan
[3]Faculty of Information Science and Technology, Hokkaido University, State name, Japan

Correspondence
*Miki Haseyama, Faculty of Information Science and Technology, Hokkaido University, Email: mhaseyama@lmd.ist.hokudai.ac.jp

Present Address
N-14, W-9, Kita-ku, Sapporo, Hokkaido, 060-0814, Japan

Summary

Traffic accidents occur frequently in cold and snow- or ice-covered regions due to weather changes that occur during the winter season. To detect the snow- or ice-covered roads in road surface conditions, road surface images captured using fixed-point cameras installed along the route are sufficient. This paper proposes a snow- or ice-covered road detection method that uses the deep convolutional autoencoding Gaussian mixture model (DCAGMM) with structural similarity (SSIM). The DCAGMM method, which is an unsupervised anomaly detection method, is unaffected by imbalance in the training data. In addition, the end-to-end convolutional neural network implemented in the DCAGMM enables the capture of the unique characteristics of the road surface images. Finally, by reconstructing the input images as normal images, the comparison of the input and reconstructed images enables identification of snow- or ice-covered road areas without requiring pixel-level annotations. Furthermore, the road surface images include complex characteristics for reconstruction, and the SSIM-based reconstruction error allows us to preserve the image quality of the reconstructed image. Experimental results obtained on real-world road surface images demonstrate the effectiveness of the proposed method.

KEYWORDS:
Winter road surface conditions, anomaly detection, deep learning, Gaussian mixture model, convolutional neural network, structural similarity.

1 | INTRODUCTION

Snow and ice on roads cause unpredictable traffic accidents, and drivers have no control over their vehicles in such accidents. Road damage detection and classification methods have been studied to protect drivers from accidents (Maeda, Kashiyama, Sekimoto, Seto, & Omata 2021; Maeda, Sekimoto, Seto, Kashiyama, & Omata 2018; Saha, Arya, & Sekimoto 2024); however, these studies are not applicable to roads covered by snow or ice. Thus, the another study on the prevention of snow and ice on roads is necessary. In Japan, cold and snow- or ice-covered regions account for approximately 60% of the land area, and traffic accidents occur frequently due to changing winter weather conditions. In addition, it has been reported that approximately 90% of winter accidents are slip accidents caused by the snow or ice on the road surface, such as freezing and snow accumulation (Nakai et al. 2019). Under these circumstances, as one way of preventing slip accidents, snow removal and applying antifreeze are implemented after detecting the snow- or ice-covered roads in road surface conditions (Nakai et al. 2019). The process of treating snow or ice is summarized as follows: First, road managers determine whether treatments are required. If the risks associated with winter road surface conditions are increasing, the target roads are closed and vehicles are not permitted to pass. During this closure period, the treatments, e.g., snow removal and antifreeze applications, are performed to reduce risk of the

road surface conditions. Snow- or ice-covered road detection has been studied under such circumstances to avoid missing the timing of the closure. In particular, when targeting highways, an early decision on whether the road surface is at risk is required, and this paper focuses on such highways.

Several previous studies have investigated the classification of road surface conditions for snow- or ice-covered road detection (Ohiro, Takakura, Sakuraba, Hanatsuka, & Hagiwara 2019; Saida et al. 2016; Uchida, Gotou, & Okamoto 2000; Yamada, Tanizaki, Ueda, Horiba, & Sugie 2000). In a previous study (Saida et al. 2016), road surface conditions were detected using the heat balance theory based on digital geographical data; however, this study required digital geographical data for all routes to be analyzed, and it is difficult to collect such data. In another study (Uchida et al. 2000), road surface conditions were detected using data acquired from fixed-point cameras and sensors installed along the route. In addition, in one study (Yamada et al. 2000), detection methods were developed using a subtractive method based on road surface images acquired using fixed-point cameras installed along the route and climate information. However, with the subtractive method, the reference image must be updated manually because the road surface conditions and roadside objects may change. Thus, a method that can classify road surface images without manual labor is required to detect road surface conditions efficiently. From a different perspective, the vibration waveform obtained from acceleration sensors installed on the surface of tires on snow and ice patrol vehicles can be used to classify the road surface conditions (Ohiro et al. 2019). However, this method requires a snow and ice patrol vehicle to pass through the target route. Thus, the remaining challenge is constructing a model that can classify road surface images automatically based on only visual information obtained from fixed-point cameras. In addition, the classification results are a poor rationale for decision making. Thus, in addition to classification results, it is expected to make more informed decisions using the knowledge and experience owned by road managers by presenting possible areas of snow or ice.

With the advancements of deep learning technologies (LeCun, Bengio, & Hinton 2015), image processing techniques have progressed considerably and made significant contributions in many tasks. For example, deep convolutional neural networks (CNN), which are a type of deep neural network, can be used to improve the accuracy of image classification (Krizhevsky, Sutskever, & Hinton 2012; Rawat & Wang 2017; Sultana, Sufian, & Dutta 2018) and semantic segmentation (Garcia-Garcia, Orts-Escolano, Oprea, Villena-Martinez, & Garcia-Rodriguez 2017; Hao, Zhou, & Guo 2020; Simonyan & Zisserman 2014) for natural images by extracting features that are suitable for each task automatically. In fact, the applications of deep learning are wide-ranging and have produced remarkable results in various research fields (Rafiei & Adeli 2016; Rafiei, Khushefati, Demirboga, & Adeli 2017; Ruiz, Díaz, González, & Cavas 2023; Urdiales, Martín, & Armingol 2023). To train CNNs, sufficient training data are required for each class to realize effective image classification results, and pixel-level annotations are required to perform semantic segmentation. However, to classify images related to real-world applications, e.g., road surface images, the amount of available training data for each class may be insufficient. In addition, to assign pixel-level annotations to images from real-world applications, we cannot use crowdsourcing services, e.g., Amazon Mechanical Turk[1], because expert knowledge is required, and a sufficient amount of training data with pixel-level annotations may not be prepared.

For snow- or ice-covered road detection in road surface conditions, labels (e.g., dry and sherbet) assigned to road surface images can be imbalanced in the training data, which means that there is bias in the data distribution. Such a bias is well-known to lead to accuracy limitations; however, we cannot control the distribution in road surface conditions due to the weather conditions. Thus, a method that can treat the imbalance in training data is required to detect the snow- or ice-covered roads in road surface conditions. In addition, differing from natural images that are generally used for image classification tasks in the computer vision field, road surface images exhibit unique characteristics. Thus, a mechanism that can capture the characteristics that are specific to road surface images is required for an effective detection method. Furthermore, the important tasks include identification of snow- or ice-covered road areas in the road surface images in addition to the classification of the road surface images. However, as mentioned previously, we cannot apply a semantic segmentation technique based on machine learning due to the difficulty associated with pixel-level annotation. In summary, the requirements for an effective deep learning-based method that uses road images are as follows.

Requirement 1)
Treatment of imbalance in training data
Requirement 2)
Capturing the image representations of road surface images
Requirement 3)
Identification of snow- or ice-covered road areas in images without pixel-level annotation

This paper proposes a snow- or ice-covered road detection method in winter road surface conditions that uses deep anomaly detection based on a deep neural network. Anomaly detection, which is a major task in the machine learning field, attempts to detect anomaly samples that deviate from the distribution of normal samples. General deep learning methods

[1] https://www.mturk.com/

tend to learn the characteristics of images belonging to the class with the largest number of samples if the number of samples in each class is imbalanced. In addition, anomaly detection methods are trained on only normal samples, and the difference from the distribution of normal samples is calculated as the abnormality when abnormal samples are input. Specifically, in this paper, road surface conditions that allow for safe driving are considered the normal conditions, and the anomaly detection method is trained using road surface images belonging to normal conditions.

For anomaly detection, we propose the use of deep convolutional autoencoding Gaussian mixture model (DCAGMM) based on the DAGMM (Zong et al. 2018), which is an existing anomaly detection method. The DCAGMM comprises two networks, i.e., a deep convolutional autoencoder (DCAE) (Masci, Meier, Cireşan, & Schmidhuber 2011) for image reconstruction and a GMM for abnormality calculation. Unlike the DAGMM, the DCAGMM adopts the DCAE rather than the autoencoder (AE) in the DAGMM to compress the input image. The DCAE allows the DCAGMM to capture unique image characteristics from road surface images, even though road surface images are inherently different from those of natural images. Note that such an encoder-decoder structure has also been used in a previous study on road modeling (F. Wang, Chen, Wijnands, & Guo 2020). Applications of DAGMM and DCAGMM have been realized in the medical and materials science fields (Togo, Saito, Ogawa, & Haseyama 2019; Togo, Watanabe, Ogawa, & Haseyama 2020), and the results have demonstrated that the DAGMM and DCAGMM can handle the characteristics of the data in a specific field effectively. Note that the images used in these previous studies were grayscale images (Togo et al. 2019 2020). In contrast, the current study considers color images. The information in color images is more complex than that of grayscale images, and the reconstruction network may overlook detailed characteristics. Consequently, the output image is blurred, and its details are unclear; therefore, the quality of the reconstructed image may be degraded even if the model detects abnormal samples successfully. Research in other fields has required successful anomaly detection; thus, the quality degradation of the reconstructed image is acceptable (Togo et al. 2019 2020). However, we must satisfy requirement 3 and cannot accept such image quality degradation. In fact, pixel-level detection has been conducted in previous studies on crack detection (J. Chen & He 2022; Siriborvornratanakul 2023; A. A. Zhang et al. 2022) in addition to the bounding box-based detection method (C. Zhang, Chang, & Jamshidi 2020).

In the flow of DCAGMM, the encoded values and reconstruction error are used as the low-dimensional representation. Then, to improve the quality of the reconstructed image, we add the structural similarity (SSIM) (Z. Wang, Bovik, Sheikh, & Simoncelli 2004) as the reconstruction error, whereas the original DAGMM simply calculates the L2 distance. The reconstruction error based on the per-pixel L2 distance fails to find structural discrepancies between the original and reconstructed images when the respective pixel values are roughly consistent. Thus, by introducing the SSIM into the DCAE-based reconstruction network, we can evaluate the perceptual similarity that is less sensitive to edge information and capture interdependencies between the local pixel areas based on statistical analysis. Thus, SSIM can improve the accuracy of DCAE-based anomaly detection (Bergmann, Löwe, Fauser, Sattlegger, & Steger 2018). As discussed previously, road surface images have more complex characteristics than images used in other fields, which can lead to quality degradation in the reconstructed images. For accurate identification of snow- or ice-covered road areas in images without pixel-level annotation, we add the SSIM to the reconstruction error and preserve the image quality of the reconstructed images.

The proposed method is an end-to-end model that uses images as input to capture unique characteristics from road surface images that are inherently different from general natural images. Thus, the proposed method is expected to realize highly accurate anomaly detection of road surface conditions. The following properties allow the proposed method to satisfy the above requirements.

Property 1)
We use only normal images for training; thus, the proposed method is unaffected by imbalance in the training data.

Property 2)
The proposed method adopts an end-to-end training process from the input images to capture the image representations of the road surface images effectively.

Property 3)
By reconstructing the input images as normal images, the comparison of the input and reconstructed images enables identification of snow- or ice-covered road areas without requiring pixel-level annotations.

The reconstructed image can be generated as a normal image even if the input image is an abnormal sample. In addition, the difference between the input and reconstructed images may identify snow- or ice-covered road areas, as shown in Fig. 1. Thus, the proposed method realizes snow- or ice-covered road detection in winter road surface conditions with less manual labor using the DCAGMM with only normal road surface images.

The primary contributions of this study are summarized as follows.

- An effective anomaly detection model is newly constructed to detect snow- or ice-covered roads existing in road surface images.

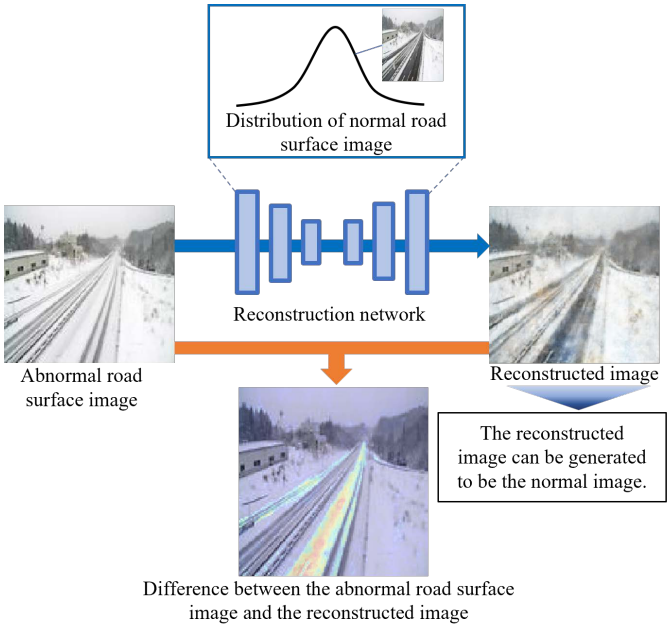


Figure 1 One of the main contributions of the proposed method. The proposed method generates the reconstructed image from the input image as a normal image even if the input image is abnormal. By calculating the structural discrepancies between the original and reconstructed images, the proposed method can identify snow- or ice-covered road areas in the input image without pixel-level annotations given by experts.

- The differences between the reconstructed and input images are used to realize sufficient explainability of the detection results without requiring pixel-level annotations.
- Experimental results obtained using real-world road surface images demonstrate the effectiveness of the proposed detection method for road surface images.

2 | ROAD SURFACE IMAGES

Road surface images are acquired using fixed-point cameras installed along the road route. This study attempts to detect abnormal road surface conditions and identify snow- or ice-covered road areas using only road surface images. Thus, it is essential to determine whether the road surface condition is abnormal and where the snow- or ice-covered road areas are located. Note that the target area is only the driving part in the road surface image, and more than half of the area in the image is unrelated to the target area, e.g., objects along the roadside or the road not for driving. In addition, the driving part is the area that is considered the background in natural images rather than the main object, e.g., an apple or a cat in natural images. This makes it more challenging to detect abnormal road surface conditions from images. Therefore, road surface images possess unique characteristics, and we must discover the changes in the driving parts using road surface images.

There are two main types of road surface images, i.e., images captured from the far field (Fig. 2) and those captured from the near field (Fig. 3). In addition, the labels assigned to road surface images differ in many cases. For example, the "dry," "sherbet," and "snow" labels are assigned to road surface images captured from the far field, as shown in Fig. 2, and the "dry," "wetness," "black sherbet," "white sherbet," "fallen snow," "freeze," and "compacted snow" labels are assigned to road surface images captured from the near field, as shown in Fig. 3. These road surface images were classified and labeled by experienced road managers from the East Nippon Expressway Company Limited, and they divided the annotation task and assigned the labels through a visual inspection process. It is well-known that the generalization ability of deep learning models, e.g., CNNs, is improved by training with a variety of images; thus, it can be effective to use various road surface images acquired at different angles or those that show different routes, as shown in Fig. 4. Under these circumstances, we divide the road surface images into "abnormal" or "normal" such that various road surface images with different types of labels can be used in the same learning process. The number of normal road surface images is considered to be much larger than that of the abnormal ones due to the effects of weather changes. This implies that anomaly detection, which detects abnormal samples among many normal samples, is suitable for various road surface images.

Thus, the proposed method can capture the unique characteristics of road surface images due to the CNN-based feature extraction module with end-to-end training. As training images, we use various road surface images and realize sufficient robustness to different view angles and routes.

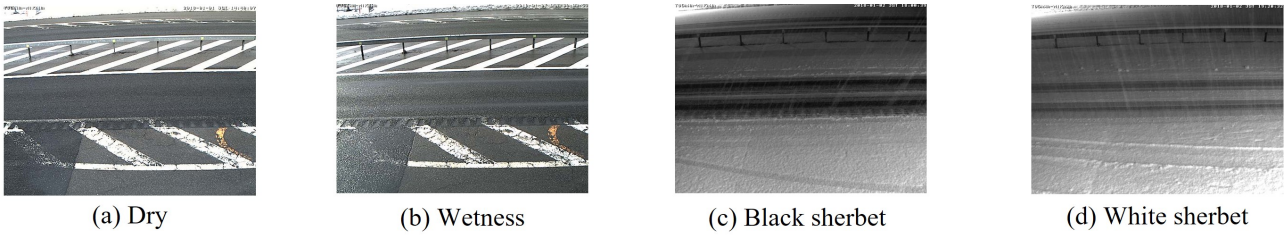
3 | PROPOSED ANOMALY DETECTION METHOD

In this section, we describe the proposed detection method based on DCAGMM with SSIM for road surface images in detail. Figure 5 shows an overview of the proposed method. In the DCAGMM, we introduce convolutional layers and SSIM-based loss into the DAGMM (G. Huang, Liu, Van Der Maaten, & Weinberger 2017), i.e., the unsupervised anomaly detection method. We train the DCAGMM using only normal road surface images, and the proposed method can detect abnormalities and identify snow- or ice-covered road areas if the abnormal road surface image is input in the test phase. The



(a) Dry (b) Sherbet (c) Snow

Figure 2 Example road surface images and corresponding labels.



(a) Dry (b) Wetness (c) Black sherbet (d) White sherbet

(e) Fallen snow (f) Freeze (g) Compacted snow

Figure 3 Additional examples of road surface images and corresponding labels.



(a) (b) (c)

Figure 4 Examples of road surface images taken at various angles or reflecting various routes. (a) and (b) are taken at various routes; (a) and (c) are taken at various angles.

proposed method comprises the DCAE-based compression network (Section 3.1) and the GMM-based estimation network (Section 3.2). Note that the compression and estimation networks are trained simultaneously using the objective function discussed in Section 3.3.

3.1 | Compression Network

In the following, we explain the compression network, which is employed to perform the dimension compression and reconstruction of the input images. In addition, the proposed method incorporates the DCAE (Masci et al. 2011), which comprises an encoder Enc(·) and decoder Dec(·) composed of convolutional layers as the compression network. The encoder

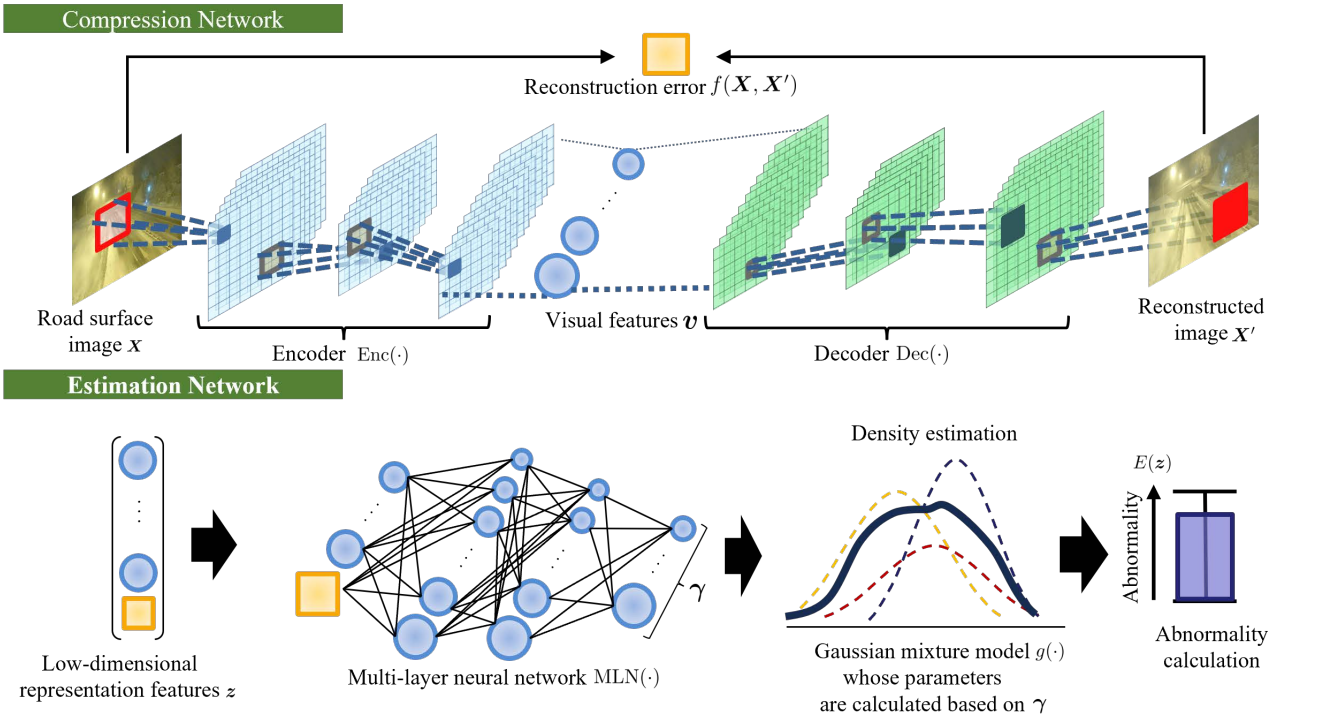


Figure 5 Overview of the proposed method. The proposed method performs anomaly detection based on the DCAGMM with the compression and estimation networks. These networks are optimized simultaneously to realize end-to-end training.

compresses images and calculates the visual features $\boldsymbol{v} \in \mathbb{R}^{d_v}$ (where d_v is the dimension of the visual features) from the input road surface image $\boldsymbol{X}$ as follows:

$$\boldsymbol{v} = \mathrm{Enc}(\boldsymbol{X}; \boldsymbol{\theta}_{\mathrm{enc}}), \tag{1}$$

where $\boldsymbol{\theta}_{\mathrm{enc}}$ is the set of training parameters in the encoder. Note that the pixel values of road surface image $\boldsymbol{X}$ is transformed to [0,1]. Then, to obtain the visual features, including sufficient information to represent the input image $\boldsymbol{X}$, the decoder calculates the reconstructed image $\boldsymbol{X}'$ as follows:

$$\boldsymbol{X}' = \mathrm{Dec}(\boldsymbol{v}; \boldsymbol{\theta}_{\mathrm{dec}}), \tag{2}$$

where $\boldsymbol{\theta}_{\mathrm{dec}}$ is the set of the training parameters in the decoder. In addition, the DCAGMM extracts the low-dimensional representation features $\boldsymbol{z} \in \mathbb{R}^{d_z}$ (where d_z represents the dimensions of the low-dimensional representation features) to calculate the abnormality by performing vector concatenation of the visual features and reconstruction errors as follows:

$$\boldsymbol{z} = [\boldsymbol{v}^{\top}, \boldsymbol{f}(\boldsymbol{X}, \boldsymbol{X}')]^{\top}. \tag{3}$$

Here, the $\boldsymbol{f}(\cdot)$ function calculates several types of reconstruction errors. The original DAGMM adopts the L2 distance or the combined use of the L2 distance and cosine similarity between the input and reconstructed images as the function that calculates the reconstruction error. Compared using only the L2-distance, the performance of the DCAE-based anomaly detection technique is improved by introducing the SSIM (Bergmann et al. 2018). Thus, the proposed method implements the L2 distance, cosine similarity, and SSIM. The reconstruction error is calculated as follows:

$$\boldsymbol{f}(\boldsymbol{X}, \boldsymbol{X}') = \left[\mathrm{L2}(\boldsymbol{X}, \boldsymbol{X}'); \mathrm{CS}(\boldsymbol{X}, \boldsymbol{X}'); \mathrm{SSIM}(\boldsymbol{X}, \boldsymbol{X}')\right], \tag{4}$$

where L2(·), CS(·), and SSIM(·) represent the functions of the L2 distance, cosine similarity, and SSIM, respectively. In the proposed method, we employ the relative L2 distance to align the value range rather than the original L2 distance as follows:

$$\mathrm{RL2}(\boldsymbol{X}, \boldsymbol{X}') = \frac{||\boldsymbol{X} - \boldsymbol{X}'||_F}{||\boldsymbol{X}||_F}, \tag{5}$$

where $||\cdot||_F$ is the Frobenius norm. Thus, we calculate the reconstruction error as follows:

$$\boldsymbol{f}(\boldsymbol{X}, \boldsymbol{X}') = \left[\mathrm{RL2}(\boldsymbol{X}, \boldsymbol{X}'), \mathrm{CS}(\boldsymbol{X}, \boldsymbol{X}'), \mathrm{SSIM}(\boldsymbol{X}, \boldsymbol{X}')\right]. \tag{6}$$

In addition, the cosine similarity is calculated as follows:

$$\mathrm{CS}(\boldsymbol{X}, \boldsymbol{X}') = \frac{\boldsymbol{X} \cdot \boldsymbol{X}'}{||\boldsymbol{X}||_F \, ||\boldsymbol{X}'||_F}. \tag{7}$$

Then, the SSIM is calculated as follows:

$$\mathrm{SSIM}(\boldsymbol{X}, \boldsymbol{X}') = l(\boldsymbol{X}, \boldsymbol{X}')^{\alpha} c(\boldsymbol{X}, \boldsymbol{X}')^{\beta} s(\boldsymbol{X}, \boldsymbol{X}')^{\gamma}, \tag{8}$$

where α, β, and γ are hyperparameters, l is a comparison of the luminance values between images, c is a comparison of the contrast between images, and s is the structural correlation between images. These parameters are calculated as follows:

$$l(\boldsymbol{X}, \boldsymbol{X}') = \frac{2\mu_{\boldsymbol{X}}\mu_{\boldsymbol{X}'} + C_1}{\mu_{\boldsymbol{X}}^2 + \mu_{\boldsymbol{X}'}^2 + C_1}, \tag{9}$$

$$c(\boldsymbol{X}, \boldsymbol{X}') = \frac{2\rho_{\boldsymbol{X}}\rho_{\boldsymbol{X}'} + C_2}{\rho_{\boldsymbol{X}}^2 + \rho_{\boldsymbol{X}'}^2 + C_2}, \tag{10}$$

$$s(\boldsymbol{X}, \boldsymbol{X}') = \frac{2\rho_{\boldsymbol{X},\boldsymbol{X}'} + C_2}{2\rho_{\boldsymbol{X}}\rho_{\boldsymbol{X}'} + C_2}. \tag{11}$$

Here, $[\mu_{\boldsymbol{X}}, \rho_{\boldsymbol{X}'}]$ and $[\mu_{\boldsymbol{X}'}, \rho_{\boldsymbol{X}}]$ represent the mean and variance of $\boldsymbol{X}$ and $\boldsymbol{X}'$, respectively, and $\rho_{\boldsymbol{X},\boldsymbol{X}'}$ is the covariance between $\boldsymbol{X}$ and $\boldsymbol{X}'$. In addition, C_1 and C_2 are constant values to stabilize the evaluation value when the denominator reaches a minimum value. In addition, by setting $\alpha = \beta = \gamma = 1$, the SSIM can be rewritten as follows:

$$\mathrm{SSIM}(\boldsymbol{X}, \boldsymbol{X}') = \frac{(2\mu_{\boldsymbol{X}}\mu_{\boldsymbol{X}'} + C_1)(2\rho_{\boldsymbol{X},\boldsymbol{X}'} + C_2)}{(\mu_{\boldsymbol{X}}^2 + \mu_{\boldsymbol{X}'}^2 + C_1)(\rho_{\boldsymbol{X}}^2 + \rho_{\boldsymbol{X}'}^2 + C_2)}. \tag{12}$$

It has been reported that $C_1 = 0.01$ and $C_2 = 0.03$ are effective values (Z. Wang et al. 2004). Thus, the proposed method calculates the low-dimensional representation features from the road surface image for anomaly detection using the DCAE and SSIM. Here, the DCAE captures the unique characteristics of the road surface image, and the SSIM improves the image quality of the reconstructed image, which is a key factor in terms of identifying snow- or ice-covered road areas without requiring pixel-level annotations. As a result, the DCAE using SSIM calculates the significant representation to reconstruct the normal road surface images while guaranteeing high image quality by training on only normal samples, which helps detect abnormal samples and identify snow- or ice-covered road areas.

3.2 | Estimation Network

Here, we describe the GMM-based estimation network used to calculate the abnormality from the low-dimensional representation features of the input road surface images. The GMM is a probabilistic model based on several multivariate Gaussian distributions and their weights for normal road surface images. The abnormal samples are detected based on the difference between the low-dimensional representation features of the test data and the distributions obtained from normal data in the GMM. Specifically, GMM $g(\boldsymbol{z})$ is represented as follows:

$$g(\boldsymbol{z}) = \sum_{k=1}^{K} w_k \mathcal{N}(\boldsymbol{z}|\boldsymbol{\mu}_k, \boldsymbol{\Sigma}_k), \tag{13}$$

where w_k is the weight of k (=1, 2, … , K; K is the number of multivariate Gaussian distributions) th multivariate Gaussian distribution. Then, the probability density function of the kth multivariate Gaussian distribution $\mathcal{N}(\boldsymbol{z}|\boldsymbol{\mu}_k, \boldsymbol{\Sigma}_k)$ is calculated as follows:

$$\mathcal{N}(\boldsymbol{z}|\boldsymbol{\mu}_k, \boldsymbol{\Sigma}_k) = \frac{1}{\sqrt{(2\pi)^{d_z}|\boldsymbol{\Sigma}_k|}} \exp\left(-\frac{1}{2}\bar{\boldsymbol{z}}_k^\top \boldsymbol{\Sigma}_k^{-1} \bar{\boldsymbol{z}}_k\right), \tag{14}$$

$$\bar{\boldsymbol{z}}_k = \boldsymbol{z} - \boldsymbol{\mu}_k, \tag{15}$$

where $\boldsymbol{\mu}_k$ and $\boldsymbol{\Sigma}_k$ are the mean and covariance for the probability density function of the kth multivariate Gaussian distribution, respectively, and $|\cdot|$ denotes the determinant of a matrix. Generally, the expectation-maximization algorithm (Moon 1996) is employed to optimize the parameters of the GMM (Zimek, Schubert, & Kriegel 2012). However, DCAGMM performs this optimization using a multilayer neural network. Thus, the estimation network uses GMM, including the K multivariate Gaussian distributions and multilayer neural network that calculates the probability belonging to each multivariate Gaussian distribution. In the estimation network, the abnormality $E(\boldsymbol{z})$ for the road surface image $\boldsymbol{X}$ is calculated as follows:

$$E(\boldsymbol{z}) = -\log\left(\sum_{k=1}^{K} w_k \frac{\exp\left(-\bar{\boldsymbol{z}}_k^\top \boldsymbol{\Sigma}_k^{-1} \bar{\boldsymbol{z}}_k/2\right)}{\sqrt{|2\pi\boldsymbol{\Sigma}_k|}}\right). \tag{16}$$

The proposed method detects an abnormal road surface image if the abnormality $E(\boldsymbol{z})$ is a high value.

Given a training normal image $\boldsymbol{X}_n$ ($n = 1, 2, \ldots, N$, where N is the number of training normal images), the mixing ratio, mean, and covariance for the kth Gaussian distribution are calculated as follows:

$$w_k = \sum_{n=1}^{N} \frac{\gamma_{k,n}}{N}, \tag{17}$$

$$\boldsymbol{\mu}_k = \frac{\sum_{n=1}^{N} \gamma_{k,n}\boldsymbol{z}_n}{\sum_{n=1}^{N} \gamma_{k,n}}, \tag{18}$$

$$\boldsymbol{\Sigma}_k = \frac{\sum_{n=1}^{N} \gamma_{k,n}(\boldsymbol{z}_n - \boldsymbol{\mu}_k)(\boldsymbol{z}_n - \boldsymbol{\mu}_k)^\top}{\sum_{n=1}^{N} \gamma_{k,n}}, \tag{19}$$

where $\gamma_{k,n}$ represents the probability of belonging to the kth Gaussian distribution for the training image $\boldsymbol{X}_n$, and $\boldsymbol{\gamma}_n = [\gamma_{1,n}, \gamma_{2,n}, \ldots, \gamma_{K,n}]^\top \in \mathbb{R}^K$ are calculated by inputting the low-dimensional representation features $\boldsymbol{z}_n$ to the multi-layer neural network $\mathrm{MLN}(\cdot)$ as follows:

$$\boldsymbol{\gamma}_n = \phi\big(\mathrm{MLN}(\boldsymbol{z}_n; \boldsymbol{\theta}_{\mathrm{mln}})\big). \tag{20}$$

Here, $\phi(\cdot)$ is the activation function, and $\boldsymbol{\theta}_{\mathrm{mln}}$ is the set of training parameters in the multilayer neural network. Note that $\boldsymbol{z}_n$ is the low-dimensional representation features obtained from the training image $\boldsymbol{X}_n$. As a result, the estimation network learns the feature distribution of the normal samples and can detect abnormal samples.

3.3 | Objective Function

Here, we explain the objective function employed in the proposed method. The compression network (Section 3.1) and the estimation network (Section 3.2) are trained simultaneously, and their parameters $\boldsymbol{\theta} = \{\boldsymbol{\theta}_{\text{enc}}, \boldsymbol{\theta}_{\text{dec}}, \boldsymbol{\theta}_{\text{mln}}\}$ are optimized as follows:

$$\hat{\boldsymbol{\theta}} = \arg\min_{\boldsymbol{\theta}} \mathcal{J}(\boldsymbol{\theta}). \tag{21}$$

Here, $\mathcal{J}(\boldsymbol{\theta})$ is the loss function of the proposed method, which is calculated as:

$$\mathcal{J}(\boldsymbol{\theta}) = \frac{1}{N}\sum_{n=1}^{N}\left(\mathcal{L}_{\text{rec}}(\boldsymbol{X}_n, \boldsymbol{X}'_n) + \lambda_1 E(\boldsymbol{z}_n)\right) + \lambda_2 P(\boldsymbol{\Sigma}), \tag{22}$$

where $\mathcal{L}_{\text{rec}}(\boldsymbol{X}_n, \boldsymbol{X}'_n)$ is the reconstruction error as follows.

$$\mathcal{L}_{\text{rec}}(\boldsymbol{X}, \boldsymbol{X}') = \text{L2}(\boldsymbol{X}, \boldsymbol{X}') - \text{CS}(\boldsymbol{X}, \boldsymbol{X}') - \text{SSIM}(\boldsymbol{X}, \boldsymbol{X}'), \tag{23}$$

In addition, the function $P(\boldsymbol{\Sigma})$ suppresses the divergence of the covariance $\boldsymbol{\Sigma}_k$ and is calculated as follows:

$$P(\boldsymbol{\Sigma}) = \sum_{k=1}^{K}\sum_{j=1}^{d_z}\frac{1}{\boldsymbol{\Sigma}_k(j,j)}, \tag{24}$$

where $\boldsymbol{\Sigma}_k(j,j)$ represents the jth diagonal element of $\boldsymbol{\Sigma}_k$, and λ_1 and λ_2 are hyperparameters. Thus, the end-to-end training process is realized using the objective function comprising the reconstruction error calculated by the compression network and the abnormality calculated by the estimation network.

In summary, the proposed method provides the following three contributions. 1) The training images used in Eqs. (17)–(19) indicate that the DCAGMM is trained on only normal road surface images to handle imbalance in the training data. 2) The objective function in Eq. (22) realizes the end-to-end training process to capture the specific characteristics of the road surface images. 3) The reconstructed image obtained using Eq. (reconctruc) is generated as a normal image, and different areas between the original and reconstructed images are calculated to identify snow- or ice-covered road areas without requiring pixel-level annotations.

4 | EXPERIMENT

In this section, we discuss the experimental settings and results, which validate the efficiency of the proposed method for detecting snow- or ice-covered roads in winter road surface conditions.

4.1 | Experimental Settings

In this experiment, we adopted three datasets, i.e., road surface images captured from the far field (hereafter referred to as the far dataset), as shown in Fig. 2, road surface images captured from the near field (hereafter referred to as the near dataset), as shown in Fig. 3, and a dataset containing both types of images (hereafter referred to as the mixed dataset). Note that all data in these datasets were provided by the East Nippon Expressway Company Limited. Figures 2-4 show examples of the road surface images used in this experiment. As can be seen, we conducted experiments on several types of road surface images and confirmed the effectiveness of the proposed method. Here, we assumed that the trained proposed method was not used for the road surface images taken by a location not included in the training data. Each image in the far dataset was assigned a label ("dry," "sherbet," or "snow"). Here, the "dry" label was considered "normal," and these images were used for training in the anomaly detection scenario. The "sherbet" and "snow" labels were considered "abnormal," and these images contained the snow- or ice-covered road areas. These labels were assigned by experienced road managers at the East Nippon Expressway Company Limited, and they divided the annotation task and assigned the labels through a visual inspection process. In addition, each image in the near dataset was assigned one label ("dry," "wetness," "black sherbet," "white sherbet," "fallen snow," "freeze," or "compacted snow"). Here, the "dry" and "wetness" labels were considered "normal," and these images were used for training in the anomaly detection scenario. The other labels, i.e., "black sherbet," "white sherbet," "fallen snow," "freeze," and "compacted snow" were considered "abnormal," and these images contained snow- or ice-covered road areas. Table 1 shows the correspondence between the labels used in each dataset and the normal and abnormal conditions, and Table 2 shows the number of images in each dataset. In this experiment, the images in the far dataset were captured at one-minute intervals from February 1, 2019 to February 8, 2019 using a fixed-point camera at three locations. The images in the near dataset were captured at 20-minute intervals from January 1, 2018 to January 31, 2018 using a fixed-point camera at a single location. The images with the normal labels were divided randomly into training and test images, and the images with abnormal labels were only used in the testing phase.

To compare the proposed method, we implemented and evaluated the following anomaly detection methods.

1. One-class support vector machine (OC-SVM) (Schölkopf, Platt, Shawe-Taylor, Smola, & Williamson 2001)
   The OC-SVM is a conventional but effective anomaly detection method based on the discrimination hyperplane in an SVM (Vapnik 1963). In addition, we introduced the radial basis function kernel into the OC-SVM to improve the discrimination ability.

Table 1 Correspondence between labels in each dataset and normal/abnormal

	Abnormal	Normal
Far dataset	Sherbet, Snow	Dry
Near dataset	Black sherbet, White sherbet, Fallen snow, Freeze, Compacted snow	Dry, Wetness

Table 2 Number of images in each dataset. Note that training images belonging to "abnormal" do not exist, and they are represented as "-" because the proposed and compared methods are unsupervised anomaly detection methods.

	Training		Test		Total Sum	
Dataset	Abnormal	Normal	Abnormal	Normal	Abnormal	Normal
Far dataset	-	7,907	1,279	1,976	1,279	9,883
Near dataset	-	723	479	181	479	904
Mix dataset	-	8,630	1,758	2,157	1,758	10,787

2. DAGMM (Zong et al. 2018)
   The DAGMM is a state-of-the-art unsupervised anomaly detection method and the original version of DCAGMM. Here, the L2 distance and cosine similarity are used in the loss function employed in the DAGMM method.

3. DCAE (Z. Chen, Yeo, Lee, & Lau 2018)
   The DCAE is a reconstruction-based unsupervised anomaly detection method. With this method, the anomalous sample was detected by reconstructing the input image and calculating the difference, and we adopted the same loss function as that used in the proposed method. The effectiveness of DCAE has been confirmed in terms of detecting anomalous infrastructures (Chow et al. 2020).

Note that the dimension of the image is excessively large when directly input to the OC-SVM and DAGMM methods, and the learning process may become unstable. In contrast, the effectiveness of the DAGMM method with the image features extracted from images as the input has been verified in the materials information field (Togo et al. 2019). Thus, in the experiment conducted in the current study, we calculated the following image features before inputting images into OC-SVM and DAGMM methods. Here, we used the variables obtained by applying the global average pooling to the outputs of the final convolution layer of the EfficientNetB7 model (Tan & Le 2019) pretrained using ImageNet (Deng et al. 2009). In addition to comparing the anomaly detection methods, we compared the reconstruction error and loss function in the DCAGMM and DCAE methods to evaluate the effectiveness of SSIM by adopting the L2 distance, cosine similarity, and SSIM.

In the proposed method, we set λ_1 and λ_2 to 0.1 and 0.005, respectively. The compression network comprised the encoder with seven convolution layers and the decoder with seven deconvolution layers. In the estimation network, we implemented three fully-connected layers. Here, the activation function was fundamentally the Leaky ReLU function (Maas, Hannun, Ng, et al. 2013), with the sigmoid function after the final layer of the decoder in the compression network and the softmax function after the final layer of the estimation network. We also employed the Adam (Kingma & Ba 2014) optimizer. Here, the batch size, number of epochs, learning rate, and K were set to 38, 1000, 0.0001, and 8, respectively. Then, to confirm the influence of the dimension d_v of the visual feature $\boldsymbol{v}$, we varied this dimension, i.e., $\{8, 16, 32, 64, 126, 256\}$, for the DCAGMM and DCAE methods.

The proposed and compared methods commonly output the abnormality $E(\boldsymbol{X})$ for an input image $\boldsymbol{X}$. We then divided each image into "abnormal" or "normal" according to the following threshold:

$$\text{result} = \begin{cases} \text{“Abnormal”} & (E(\boldsymbol{X}) \geq \tau) \\ \text{“Normal”} & (E(\boldsymbol{X}) < \tau) \end{cases}, \tag{25}$$

where τ is the threshold value. To evaluate the performance of each anomaly detection method, we used the area under receiver operating characteristic curve (AUROC) and area under the precision-recall curve (AUPRC) according to the literature (Ruff et al. 2021) as threshold-independent measures. Note that the baseline AUROC value is 0.5, and the baseline AUPRC value is equal to the proportion of the "abnormal" images, i.e., 1,279/3,255 for the far dataset, 479/660 for the near dataset, and 1,758/3,915 for the mixed dataset.

We implemented the proposed and compared methods on a computer running the Ubuntu 18.04 operating system with an Intel Core i7-8700 CPU and two GeForce RTX 2080 Ti GPUs. In addition, we used the Python and TensorFlow frameworks to train and test the deep learning models.

Table 3 AUROC and AUPRC results. Here "S," "C," and "L" represent the use of SSIM, cosine similarity, and L2 distance, respectively. "DCAGMM w/ S, C, L" represents the proposed method, cosine similarity, and L2-distance as the function to calculate the reconstruction error function. We set the dimension d_v of the visual features in DCAGMM and DCAE to 64.

	AUROC			AUPRC		
	Far dataset	Near dataset	Mixed dataset	Far dataset	Near dataset	Mixed dataset
DCAGMM w/ S, C, L	**0.966**	**0.965**	**0.952**	**0.951**	**0.986**	**0.921**
DCAGMM w/ C, L	0.564	0.664	0.691	0.572	0.851	0.642
DCAGMM w/ L	0.373	0.912	0.827	0.379	0.959	0.824
DCAE w/ S, C, L	0.921	0.941	0.878	0.831	0.971	0.800
DCAE w/ C, L	0.928	0.935	0.863	0.845	0.968	0.794
DCAE w/ L	0.924	0.936	0.865	0.841	0.967	0.815
DAGMM	0.180	0.288	0.231	0.435	0.191	0.422
OC-SVM	0.907	0.775	0.825	0.928	0.461	0.799

4.2 | Detection Evaluation

Table 3 shows the experimental results, where "S," "C," and "L" represent the use of SSIM, cosine similarity, and L2 distance, respectively. In addition, "DCAGMM w/ S, C, L" represents the DCAGMM method with SSIM, cosine similarity, and L2 distance as the function to calculate the reconstruction error, that is, the proposed method. Here, we used the 64-dimensional visual features in the DCAGMM and DCAE methods. As can be seen, in terms of the AUROC and AUPRC values, the proposed method, i.e., DCAGMM w/ S, C, L, outperformed the compared methods, and the experimental results confirm the effectiveness of the proposed method. By comparing the OC-SVM and DAGMM methods with the proposed DCAGMM w/ S, C, L, the feature representation learning using DCAE from the input images was more effective than using the pretrained CNN model to capture the unique characteristics of the road surface images. This reason for this finding is the difference of the visual characteristics between the ImageNet and our dataset. The images in ImageNet include distinct objects corresponding to each class, and the images in our dataset do not have such objects for each class, and this difference may result in reduced classification performance. In addition, a comparison between the proposed DCAGMM w/ S, C, L and DCAE w/ S, C, L methods demonstrates that the GMM method with DCAE is valid for detecting abnormal images from the road surface images. In terms of the difference in the reconstruction error function, a comparison between the DCAGMM w/ S, C, L, DCAGMM w/ C, L, and DCAGMM w/ L methods demonstrates that the combined use of the SSIM, cosine similarity, and L2 distance can give an adequate reconstruction error for the reconstructed images. In addition to the pixel-level reconstruction error, e.g., the cosine similarity and L2 distance, introducing an image quality assessment, e.g., SSIM, can evaluate perceptual similarity, which is less sensitive to edge information, thereby leading to performance improvements. Furthermore, by focusing on the difference in the experimental datasets, the AUROC and AUPRC values obtained by the compared methods are inconsistent depending on the dataset, and the values obtained by the proposed DCAGMM method w/ S, C, L are generally high regardless of the target dataset.

Figure 6 shows the architectural differences between the DCAGMM, DAGMM, and DCAE methods. As can be seen, the DCAGMM and DCAE methods primarily comprise convolution layers, and the DAGMM method comprises fully-connected layers. Thus, the DAGMM method cannot process the original road surface images and identify the snow- or ice-covered areas. The DCAGMM and DCAE methods outperformed the DAGMM method because they can capture the unique characteristics of the road surface images. In addition, the DCAGMM and DAGMM methods implement the GMM module for anomaly detection, while the DCAE method employs the reconstruction error as the abnormality score. Thus, the DCAE method is limited in terms of its ability to capture the distribution of the normal samples and is outperformed by the proposed method. The experimental results confirm that the proposed DCAGMM method w/ S, C, L is effective at detecting abnormal images from the road surface images from a practical point of view.

In the following, we discuss annotation costs. In a previous study on road surface anomaly detection using remote sensing data and deep convolutional networks, the number of training images with pixel-level annotations was approximately 6,000 (Xia, Sun, Zhang, & Wang 2019). In another anomaly detection study in the computer vision field, the number of training images with pixel-level annotation was approximately 5000 (Bergmann, Fauser, Sattlegger, & Steger 2019; Liu, Zhou, Xu, & Wang 2023). Compared to these studies, the proposed method, which achieved high detection performance with a similar amount of data without pixel-level annotations,

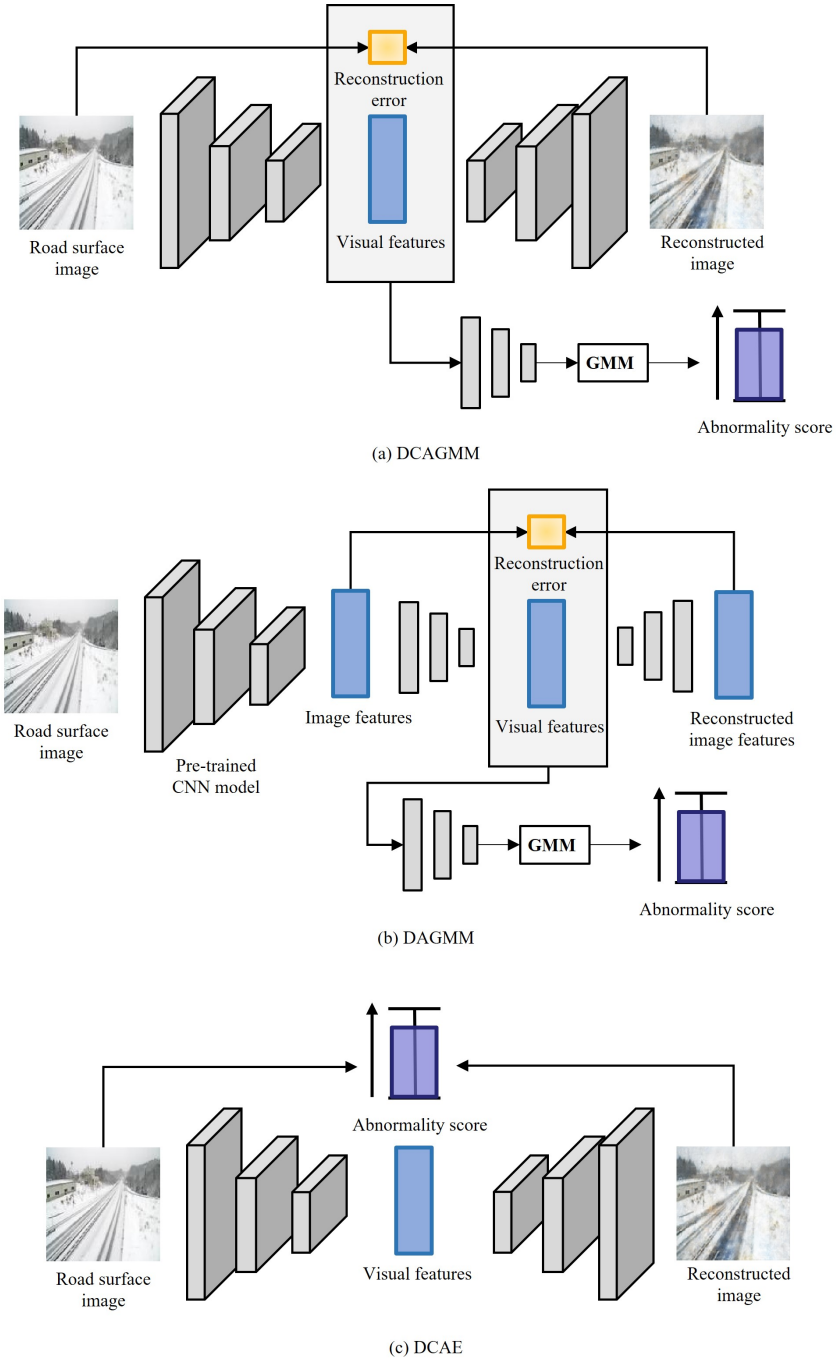
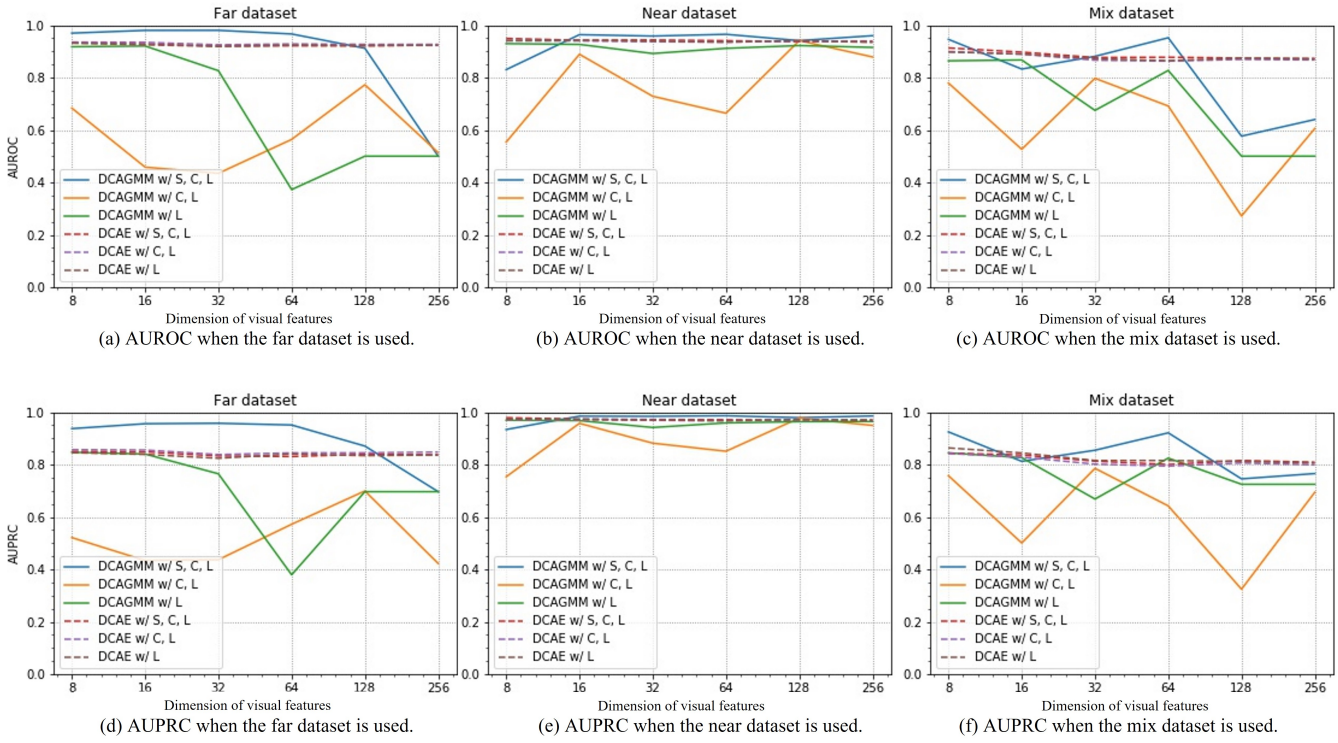


Figure 6 Architectural differences between DCAGMM, DAGMM, and DCAE. (a) DCAGMM and (c) DCAE primarily comprise the convolution layers, and (b) DAGMM comprises fully-connected layers. In addition, (a) DCAGMM and (b) DAGMM implement the GMM module for anomaly detection, and (c) DCAE employs the reconstruction error as the abnormality score.

was successful in terms of reducing annotation costs. In other words, with the proposed method, the annotation cost only involve assigning an abnormal or normal label to the nearly equal number of images utilized in comparable anomaly detection studies.

4.3 | Effects of Dimension d_v of Visual Features

The performance changes of the DCAGMM and DCAE methods when the dimension d_v of the visual features was varied in $\{8, 16, 32, 64, 128, 256\}$ are shown in Fig. 7. For all patterns, the performance of the DCAE method did not change



(a) AUROC when the far dataset is used. (b) AUROC when the near dataset is used. (c) AUROC when the mix dataset is used.

(d) AUPRC when the far dataset is used. (e) AUPRC when the near dataset is used. (f) AUPRC when the mix dataset is used.

Figure 7 Performance changes of DCAGMM and DCAE when the dimension d_v of the visual features varies.

significantly due to the dimension d_v of the visual features, whereas the DCAGMM w/ C, L and DCAGMM w/ L methods could be affected considerably by the dimension d_v of the visual features. The multilayer network for the GMM cannot be stable when only pixel-level reconstruction errors, e.g., cosine similarity and L2 distance, are used as the reconstruction error function. However, the proposed method was stable with $\{16, 32, 64\}$, and the AUROC and AUPRC values in each dataset were the highest when the proposed DCAGMM method w/S, C, L was used with 64-dimensional visual features. Consequently, the stability of the dimension d_v of the visual features is improved by introducing SSIM into the reconstruction error function in the proposed method, and the highest performance was also improved.

4.4 | Evaluation of Snow- or Ice-covered Road Area Identification

Figure 8 shows identification examples for the snow- or ice-covered road areas. By comparing the proposed DCAGMM method w/ S, C, L and DCAE w/ S, C, L in the first image from the left, the proposed DCAGMM method w/ S, C, L identified the center snow- or ice-covered road area more accurately than DCAE w/ S, C, L. In addition, for the second image from the left, the proposed DCAGMM method w/ C, L identified the excessive snow- or ice-covered road area on the right side, whereas the DCAGMM w/ S, C, L could only identify the driving part. For these examples, in terms of the fourth and fifth images from the left, the proposed DCAGMM method w/ S, C, L identified the center snow- or ice-covered road area more accurately than the DCAE method w/ S, C, L. These findings confirm that the proposed DCAGMM method w/ S, C, L can effectively identify snow-covered road areas more accurately than the compared methods. The third and sixth images from the left show that the proposed method works effectively under different weather and lighting conditions; however, there are no major differences between the compared methods. The third images from the left show that there are misidentifications under adverse weather conditions, e.g., the areas corresponding to fallen snows. Here, the proposed method performed the frame-based detection and identification, and thus additional simple processing such as median processing of multiple frames can improve robustness in the cases that there are momentary areas related to the adverse weather, e.g., rains and snows when the proposed method is used for practical situations. In addition, the semantic segmentation models (Shashaani et al. 2023; Xu et al. 2023) can easily restrict target areas by splitting driving parts although we did not do that in order to confirm the potential of the proposed method. Thus, in addition to the theoretical correctness discussed in Section 3, we have confirmed that the proposed method is effective in terms of identifying the snow-covered

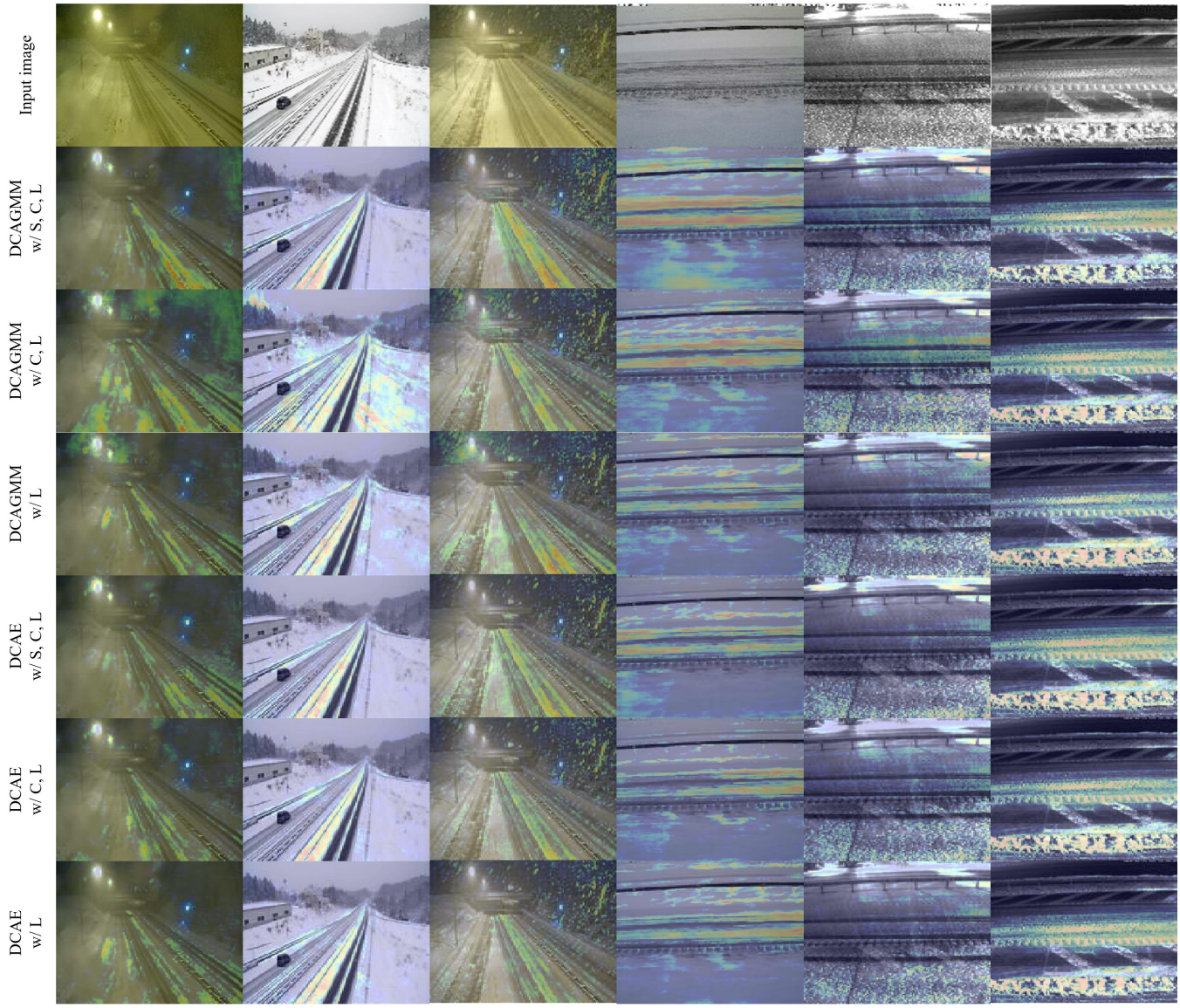


Figure 8 Identification examples of snow- or ice-covered road areas. The first, second, and third images are from the far dataset, and the fourth, fifth, and sixth images are from the near dataset.

road area in road surface images without requiring extensive pixel-level annotations.

5 | LIMITATIONS AND FUTURE WORK

In this section, we explain observed limitations and relevant future work. Although we have presented sufficient theoretical adequacy in Section 3 and practical validity in Section 4, the proposed method exhibits several limitations that must be considered. For example, we show the misdetection examples of DCAGMM w/ S, C, L under the experimental conditions where the dimension d_v of the visual features is set to 64 in Fig. 9. Here, Fig. 9(a) and Fig. 9(b) show abnormal samples detected as normal by the proposed method, and Fig. 9(c) and Fig. 9(d) show normal samples detected as abnormal. Specifically, in Figs. 9(a) and (b), the road surfaces have little snow, and most of the driving parts in the road are exposed; thus, the proposed method incorrectly recognized these samples as normal. Conversely, the road surface is wet from rain in Fig. 9(c), and the camera is blurred in Fig. 9(d); thus, the proposed method incorrectly recognized these samples as abnormal. The proposed method accurately identified the driving parts in the



(a) (b) (c) (d)

Figure 9 Misdetection examples of DCAGMM w/ S, C, L. (a) and (b) show the abnormal samples detected as normal by the proposed method. (c) and (d) show the normal samples detected as abnormal by the proposed method.

road surface images; however, incorrect recognition results were obtained when the road surface was ambiguous or the camera was dirty or otherwise impaired due to natural factors, e.g., rain. Thus, the proposed method is limited in terms of its accuracy because the images captured by cameras are sensitive to such natural factors. From the above discussion, we consider that additional information sources, e.g., changes in weather and temperature, should be introduced into the proposed method in future work.

6 | CONCLUSION

This paper has proposed a snow- or ice-covered road detection method in road surface images using the DCAGMM method and SSIM. The proposed method adopts an unsupervised anomaly detection framework, i.e., the DCAGMM method, to handle imbalance in the training data. In addition, the DCAE-based compression network implemented in the proposed method captures the unique characteristics of the road surface images by performing end-to-end learning with images as inputs. Finally, by reconstructing the input images as normal images based on the SSIM-based reconstruction error, the comparison of the input and reconstructed images enables us to identify snow- or ice-covered road areas without requiring extensive pixel-level annotations.

In the proposed method, we adopt the CNN and GMM architectures. However, several machine learning method can be explored to extend the proposed method to realize better performance, e.g., the neural dynamic classification algorithm Rafiei and Adeli (2017), dynamic ensemble learning algorithm (Alam, Siddique, & Adeli 2020), finite element machine for fast learning (Pereira, Piteri, Souza, Papa, & Adeli 2020), and self-supervised learning (Rafiei, Gauthier, Adeli, & Takabi 2022). Investigating the effects of implementing such techniques in the proposed method will be the focus of future work.

ACKNOWLEDGMENTS

In this research, we used the data provided by East Nippon Expressway Company Limited. This work was partly supported by the JSPS KAKENHI Grant Numbers JP21H03456, JP23K11211.

References

Alam, K. M. R., Siddique, N., & Adeli, H. (2020). A dynamic ensemble learning algorithm for neural networks. *Neural Computing and Applications*, *32*(12), 8675–8690.

Bergmann, P., Fauser, M., Sattlegger, D., & Steger, C. (2019). Mvtec ad–a comprehensive real-world dataset for unsupervised anomaly detection. In *Proc. IEEE/CVF Conf. Computer Vision and Pattern Recognition (CVPR)* (pp. 9592–9600).

Bergmann, P., Löwe, S., Fauser, M., Sattlegger, D., & Steger, C. (2018). Improving unsupervised defect segmentation by applying structural similarity to autoencoders. *arXiv preprint arXiv:1807.02011*.

Chen, J., & He, Y. (2022). A novel u-shaped encoder–decoder network with attention mechanism for detection and evaluation of road cracks at pixel level. *Computer-Aided Civil and Infrastructure Engineering (CACAIE)*, *37*(13), 1721–1736.

Chen, Z., Yeo, C. K., Lee, B. S., & Lau, C. T. (2018). Autoencoder-based network anomaly detection. In *Proc. IEEE Wireless Telecommunications Symposium (WTS)* (pp. 1–5).

Chow, J. K., Su, Z., Wu, J., Tan, P. S., Mao, X., & Wang, Y.-H. (2020). Anomaly detection of defects on concrete structures with the convolutional autoencoder. *Advanced Engineering Informatics*, *45*, 101105.

Deng, J., Dong, W., Socher, R., Li, L.-J., Li, K., & Fei-Fei,

L. (2009). Imagenet: A large-scale hierarchical image database. In *Proc. IEEE/CVF Int'l Conf. Computer Vision and Pattern Recognition (CVPR)* (pp. 248–255).

Garcia-Garcia, A., Orts-Escolano, S., Oprea, S., Villena-Martinez, V., & Garcia-Rodriguez, J. (2017). A review on deep learning techniques applied to semantic segmentation. *arXiv preprint arXiv:1704.06857*.

Goodfellow, I., Pouget-Abadie, J., Mirza, M., Xu, B., Warde-Farley, D., Ozair, S., … Bengio, Y. (2014). Generative adversarial nets. In *Advances in Neural Information Processing Systems (NearIPS)* (pp. 2672–2680).

Hao, S., Zhou, Y., & Guo, Y. (2020). A brief survey on semantic segmentation with deep learning. *Neurocomputing*, *406*, 302–321.

Hochreiter, S., & Schmidhuber, J. (1997). Long short-term memory. *Neural Computation*, *9*(8), 1735–1780.

Huang, G., Liu, Z., Van Der Maaten, L., & Weinberger, K. Q. (2017). Densely connected convolutional networks. In *Proc. IEEE/CVF Int'l Conf. Computer Vision and Pattern Recognition (CVPR)* (pp. 4700–4708).

Huang, G.-B., Zhu, Q.-Y., & Siew, C.-K. (2004). Extreme learning machine: a new learning scheme of feedforward neural networks. In *Proc. IEEE Int'l Joint Conf. Neural Networks (IJCNN)* (Vol. 2, pp. 985–990).

Kingma, D. P., & Ba, J. (2014). Adam: A method for stochastic optimization. *arXiv preprint arXiv:1412.6980*.

Krizhevsky, A., Sutskever, I., & Hinton, G. (2012). Imagenet classification with deep convolutional neural networks. In *Advances in Neural Information Processing Systems (NearIPS)* (pp. 1097–1105).

Kwon, T. J., Fu, L., & Melles, S. J. (2017). Location optimization of road weather information system (rwis) network considering the needs of winter road maintenance and the traveling public. *Computer-Aided Civil and Infrastructure Engineering (CACAIE)*, *32*(1), 57–71.

LeCun, Y., Bengio, Y., & Hinton, G. (2015). Deep learning. *Nature*, *521*(7553), 436–444.

Liu, Z., Zhou, Y., Xu, Y., & Wang, Z. (2023). Simplenet: A simple network for image anomaly detection and localization. In *Proc. IEEE/CVF Conf. Computer Vision and Pattern Recognition (CVPR)* (pp. 20402–20411).

Luo, J., Vong, C.-M., & Wong, P.-K. (2013). Sparse bayesian extreme learning machine for multi-classification. *IEEE Trans. Neural Networks and Learning Systems (TNNLS)*, 836–843.

Maas, A. L., Hannun, A. Y., Ng, A. Y., et al. (2013). Rectifier nonlinearities improve neural network acoustic models. In *Proc. Int'l Conf. Machine Learning (ICML)* (Vol. 30, p. 3).

Maeda, H., Kashiyama, T., Sekimoto, Y., Seto, T., & Omata, H. (2021). Generative adversarial network for road damage detection. *Computer-Aided Civil and Infrastructure Engineering (CACAIE)*, *36*(1), 47–60.

Maeda, H., Sekimoto, Y., Seto, T., Kashiyama, T., & Omata, H. (2018). Road damage detection and classification using deep neural networks with smartphone images. *Computer-Aided Civil and Infrastructure Engineering (CACAIE)*, *33*(12), 1127–1141.

Masci, J., Meier, U., Cireşan, D., & Schmidhuber, J. (2011). Stacked convolutional auto-encoders for hierarchical feature extraction. In *Proc. Int'l Conf. Artificial Neural Networks (ICANN)* (pp. 52–59).

Moon, T. K. (1996). The expectation-maximization algorithm. *IEEE Signal Processing Magazine*, *13*(6), 47–60.

Nakai, S., Kosugi, K., Yamaguchi, S., Yamashita, K., Sato, K., Adachi, S., … Sato, T. (2019). Study on advanced snow information and its application to disaster mitigation: An overview. *Bulletin of Glaciological Research*, *37*, 3–19.

Ohiro, T., Takakura, K., Sakuraba, T., Hanatsuka, Y., & Hagiwara, T. (2019). Development of advanced anti-icing spray system using automated road surface condition judgement system. *JSTE Journal of Traffic Engineering (Japanese)*, *5*(4), B_7–B_15.

Pereira, D. R., Piteri, M. A., Souza, A. N., Papa, J. P., & Adeli, H. (2020). Fema: A finite element machine for fast learning. *Neural Computing and Applications*, *32*, 6393–6404.

Rafiei, M. H., & Adeli, H. (2016). A novel machine learning model for estimation of sale prices of real estate units. *Journal of Construction Engineering and Management*, *142*(2), 04015066.

Rafiei, M. H., & Adeli, H. (2017). A new neural dynamic classification algorithm. *IEEE Trans. Neural Networks and Learning Systems (TNNLS)*, *28*(12), 3074–3083.

Rafiei, M. H., Gauthier, L. V., Adeli, H., & Takabi, D. (2022). Self-supervised learning for electroencephalography. *IEEE Trans. Neural Networks and Learning Systems (TNNLS)*.

Rafiei, M. H., Khushefati, W. H., Demirboga, R., & Adeli, H. (2017). Supervised deep restricted boltzmann machine for estimation of concrete. *ACI Materials Journal*, *114*(2), 237.

Rawat, W., & Wang, Z. (2017). Deep convolutional neural networks for image classification: A comprehensive review. *Neural Computation*, *29*(9), 2352–2449.

Ruff, L., Kauffmann, J. R., Vandermeulen, R. A., Montavon, G., Samek, W., Kloft, M., … Müller, K.-R. (2021). A unifying review of deep and shallow anomaly detection. *Proc. the IEEE*, *109*(5), 756–795.

Ruiz, L., Díaz, S., González, J. M., & Cavas, F. (2023). Improving the competitiveness of aircraft manufacturing

automated processes by a deep neural network. *Integrated Computer-Aided Engineering (ICAE)*(Preprint), 1–12.

Saha, P. K., Arya, D., & Sekimoto, Y. (2024). Federated learning–based global road damage detection. *Computer-Aided Civil and Infrastructure Engineering (CACAIE)*.

Saida, A., Fujimoto, A., Tokunaga, R., Hirasawa, M., Takahashi, N., Ishida, T., & Fukuhara, T. (2016). Verification of hfn forecasting accuracy in hokkaido using route-based forecasting model of road snow/ice conditions. *Japanese Society of Snow and Ice (Japanese)*, 4.

Schölkopf, B., Platt, J. C., Shawe-Taylor, J., Smola, A. J., & Williamson, R. C. (2001). Estimating the support of a high-dimensional distribution. *Neural Computation*, *13*(7), 1443–1471.

Shashaani, S., Teshnehlab, M., Khodadadian, A., Parvizi, M., Wick, T., & Noii, N. (2023). Using layer-wise training for road semantic segmentation in autonomous cars. *IEEE Access*.

Simonyan, K., & Zisserman, A. (2014). Very deep convolutional networks for large-scale image recognition. *arXiv preprint arXiv:1409.1556*.

Siriborvornratanakul, T. (2023). Pixel-level thin crack detection on road surface using convolutional neural network for severely imbalanced data. *Computer-Aided Civil and Infrastructure Engineering (CACAIE)*, *38*(16), 2300–2316.

Sultana, F., Sufian, A., & Dutta, P. (2018). Advancements in image classification using convolutional neural network. In *Proc. Int'l Conf. Research in Computational Intelligence and Communication Networks (ICRCICN)* (pp. 122–129).

Tan, M., & Le, Q. (2019). Efficientnet: Rethinking model scaling for convolutional neural networks. In *Proc. Int'l Conf. Machine Learning (ICML)* (pp. 6105–6114).

Togo, R., Saito, N., Ogawa, T., & Haseyama, M. (2019). Estimating regions of deterioration in electron microscope images of rubber materials via a transfer learning-based anomaly detection model. *IEEE Access*, *7*, 162395–162404.

Togo, R., Watanabe, H., Ogawa, T., & Haseyama, M. (2020). Deep convolutional neural network-based anomaly detection for organ classification in gastric x-ray examination. *Computers in Biology and Medicine*, *123*, 103903.

Uchida, M., Gotou, K., & Okamoto, J. (2000). Web systems for sensing and predicting road surface conditions in winter season. *Yokogawagiho (Japanese)*, *44*(4), 21–24.

Urdiales, J., Martín, D., & Armingol, J. M. (2023). An improved deep learning architecture for multi-object tracking systems. *Integrated Computer-Aided Engineering (ICAE)*, *30*(2), 121–134.

Vapnik, V. (1963). Pattern recognition using generalized portrait method. *Automation and Remote Control*.

Wang, F., Chen, Y., Wijnands, J. S., & Guo, J. (2020). Modeling and interpreting road geometry from a driver's perspective using variational autoencoders. *Computer-Aided Civil and Infrastructure Engineering (CACAIE)*, *35*(10), 1148–1159.

Wang, Z., Bovik, A. C., Sheikh, H. R., & Simoncelli, E. P. (2004). Image quality assessment: from error visibility to structural similarity. *IEEE Trans. Image Processing (TIP)*, *13*(4), 600–612.

Xia, W., Sun, S., Zhang, B., & Wang, L. (2019). A method for road surface anomaly detection from remote sensing data by deep convolutional networks. In *Proc. IEEE Int'l Conf. Transportation Information and Safety (ICTIS)* (pp. 861–866).

Xu, G., Zhou, W., Qian, X., Zhang, Y., Lei, J., & Yu, L. (2023). Thcanet: Two-layer hop cascaded asymptotic network for robot-driving road-scene semantic segmentation in rgb-d images. *Digital Signal Processing*, *136*, 104011.

Yamada, M., Tanizaki, T., Ueda, K., Horiba, I., & Sugie, N. (2000). A system of discrimination of the road condition by means of image processing. *IEEJ Transactions on Industry Applications (Japanese)*, *120*(8-9), 1053–1060.

Yin, X., & Liu, X. (2017). Multi-task convolutional neural network for pose-invariant face recognition. *IEEE Trans. Image Processing (TIP)*, *27*(2), 964–975.

Zhang, A. A., Wang, K. C., Liu, Y., Zhan, Y., Yang, G., Wang, G., … others (2022). Intelligent pixel-level detection of multiple distresses and surface design features on asphalt pavements. *Computer-Aided Civil and Infrastructure Engineering (CACAIE)*, *37*(13), 1654–1673.

Zhang, C., Chang, C.-c., & Jamshidi, M. (2020). Concrete bridge surface damage detection using a single-stage detector. *Computer-Aided Civil and Infrastructure Engineering (CACAIE)*, *35*(4), 389–409.

Zimek, A., Schubert, E., & Kriegel, H.-P. (2012). A survey on unsupervised outlier detection in high-dimensional numerical data. *Statistical Analysis and Data Mining*, *5*(5), 363–387.

Zong, B., Song, Q., Min, M. R., Cheng, W., Lumezanu, C., Cho, D., & Chen, H. (2018). Deep autoencoding gaussian mixture model for unsupervised anomaly detection. In *Proc. Int'l Conf. Learning Representations (ICLR)*.

How to cite this article: Moroto Y., Maeda K., Ogawa T. and Haseyama M. (2023), A regime analysis of Atlantic winter jet variability applied to evaluate HadGEM3-GC2, *Q.J.R. Meteorol. Soc.*, *2017;00:1–6*.