

1 Automatic facies classification using convolutional
2 neural network for 3D outcrop data:
3 Application to the outcrop of the mass-transport
4 deposit

5
6 Ryusei Sato, Kazuki Kikuchi, Hajime Naruse
7

9 **ACKNOWLEDGMENTS**

10 This study was conducted as a joint research project between Kyoto University and INPEX
11 Corporation. The authors appreciate Kobacon Corporation for supporting the construction
12 of the 3D point cloud data. We also appreciate Sediment Dynamics Research Consortium
13 for funding this study.

ABSTRACT

Recent advancements in three-dimensional (3D) facies modeling of outcrop have enabled the prompt reconstruction of outcrops through drone photogrammetry. Such 3D outcrop models facilitate the understanding of spatial lithofacies distributions and serve as valuable tools for quantitative analysis. Nevertheless, challenges persist in identifying facies within the digital models of large-scale outcrops. To address this issue, this study proposes a method for automatically recognizing outcrop facies, using a convolutional neural network (CNN) model applied to a 3D point cloud of an outcrop. The process involves constructing a 3D outcrop point cloud through drone photogrammetry and translating the point cloud into a series of two-dimensional images containing the color and roughness properties of the outcrop surface. The CNN model was trained using manually annotated datasets with five distinct classes: pebbly mudstone, sedimentary block, vegetation, beach, and topsoil. The trained model accurately estimated outcrop facies in unknown test images with a probability as high as 86.4% in terms of F_1 -score. Furthermore, experimental results revealed that considering roughness information significantly improved lithofacies classification accuracy. Our method was applied to the outcrop of a mass-transport deposit exposed in the Upper Cretaceous to Paleocene Akkeshi Formation along the Esashito coast of Hokkaido Island, northern Japan. The trained CNN model effectively reconstructed the 3D facies model, which substantially agreed with the actual spatial distribution in visual assessments. This novel approach offers a swift and precise means of reconstructing 3D facies models even in large-scale or inaccessible outcrops, paving the way for quantitative analyses across diverse regions.

1 INTRODUCTION

Recently, three-dimensional (3D) facies models have been used to investigate the architecture of ancient deposits. A 3D facies model is a geometry composed of a point cloud, meshes, or voxels representing the spatial distributions of lithofacies and has been utilized for various purposes including interpretation of outcrops. For example, Kirsch et al. (2019) obtained 3D distributions of rock physical properties using a hyperspectral camera and identified deformation architectures caused by earthquakes. Thomas et al. (2021) reconstructed a 3D facies model representing porosity and permeability by outcrop observations and clarified the 3D distribution of hydrocarbon reservoirs. Furthermore, Wenbiao et al. (2016) reconstructed a subsurface 3D facies model from core datasets using geostatistics.

3D facies models of large-scale outcrops would be helpful, especially for the analysis of mass-transport deposits (MTDs). MTDs are generally composed of poorly sorted muddy matrices and deformed sedimentary blocks. The size distribution and spatial variation of sedimentary blocks are critical for the characterization and reconstruction of gravity flows from which MTDs are deposited (Naruse and Otsubo, 2011; Sobiesiak et al., 2016; Otsubo et al., 2018; Cardona et al., 2020). However, the measurement of the sizes and shapes of sedimentary blocks irregularly distributed in outcrops is a time-consuming task because the sedimentary architecture of outcrop images is to be manually traced (Naruse and Otsubo, 2011; Sobiesiak et al., 2016; Cardona et al., 2020). Thus, automated recognition of the internal structures of MTDs is critical for the quantitative investigation of MTDs in large-scale outcrops.

Recent developments in drone photogrammetry have made it easy to acquire 3D digital

outcrop models from large-scale outcrops. Previously, Jorry et al. (2003) reconstructed a 3D facies model using the fence diagram method with wellbore cores and outcrop observations. However, excavating wellbore cores requires significant amounts of labor and time, especially in large-scale outcrops. In contrast, a 3D point cloud can be easily obtained by photogrammetry based on outcrop photos taken at different distances and angles (Aicardi et al., 2016). Moreover, drones enable the acquisition of 3D point clouds of cliffs too steep for researchers to reach (Colomina and Molina, 2014).

As previously stated, the acquisition of a 3D digital outcrop model has become more accessible; however, significant challenges persist in interpreting this model to construct a detailed 3D facies model. To reconstruct 3D facies models, earlier research employed facies classification through *in situ* observations within outcrops (Jorry et al., 2003; Thomas et al., 2021). However, manually classifying facies in outcrops is a time-consuming and labor-intensive process (Liu et al., 2020). Furthermore, classification relying on naked-eye observations necessitates specialized expertise from researchers and may result in undesirable classification accuracy variations. Kirsch et al. (2019) used a hyperspectral camera for this purpose, but the requirement for stable footing limits data acquisition, which is contingent on researchers' access to the outcrop. Although facies classification within an outcrop represents a crucial step in fieldwork, enhancements in efficiency and objectivity are still required to optimize the process.

To overcome these challenges, this study used a machine learning technique to enhance the efficiency and objectivity of facies classification. Machine learning is a method for constructing prediction or detection models using data. Supervised learning is a type of training in which correct values or labels are assigned to outputs corresponding to specific inputs (training data). A convolutional neural network (CNN) is a neural network with

convolution layers beneficial to grasp features in images and is widely used in the fields of image classification and semantic segmentation (Simonyan and Zisserman, 2014; Krizhevsky et al., 2017). Several studies have employed CNNs to classify lithofacies in outcrop images (Ran et al., 2019; Liu et al., 2020; Pascual et al., 2021; Malik et al., 2022). However, CNNs have not been used to identify the lithofacies distribution in 3D outcrop models.

This study proposes a method for reconstructing a 3D facies model using a CNN for a 3D point cloud of outcrop. The CNN model in this study is trained with a dataset subsampled from the entire outcrop that includes both the color and roughness properties of the outcrop surface. When researchers observe outcrops, they refer to their surface irregularities and colors because the roughness of sedimentary rocks reflects their weathering resistance due to rock properties. Therefore, the CNN model is expected to identify facies more precisely when trained using the color and roughness properties of outcrops.

As a case study, the trained CNN model was applied to the MTD exposed in an outcrop along the Esashito coast located in Hamanaka city, southern eastern Hokkaido, northern Japan. The outcrop exposes the Upper Cretaceous to the Paleocene Akkeshi Formation, the Nemuro Group. The MTD crops out over 1.6 km and consists of a pebbly mudstone matrix containing sedimentary blocks of alternating beds of sandstones and mudstones (Naruse and Otsubo, 2011). In this study, a 3D point cloud was generated by photogrammetry, with outcrop images captured by a drone along the Esashito coast. Then, a machine learning model was established to reconstruct a 3D facies model.

2 GEOLOGICAL SETTINGS

The Upper Cretaceous to Paleocene Nemuro Group is primarily composed of turbidites, hemipelagites, and slump deposits (Kiminami, 1978). Naruse (2003) performed sedimentological analysis of the Group and interpreted it as a submarine fan deposit in the forearc basin of the Paleo–Kuril Arc. The Group has a total thickness of over 3000 m, and is unconformably overlain by the Eocene Urahoro Group (Naruse, 2003). The Nemuro Group crops out in two separated regions: the Shiranuka Hill area and the coastal area ranging from Kushiro to the western Kuril Islands, eastern Hokkaido, northern Japan (Figure 1). In the coastal area, the upper 2000 m of the Group crops out along the coastal cliff. It has not experienced significant tectonic deformation and is homoclinic (Kiminami, 1978). It strikes in the NE–SW direction and dips 10°–25° southward (Kiminami, 1978), whereas the Group is characterized by strong tectonic deformations with folds and faults in the Shiranuka Hill area.

The Akkeshi Formation comprises the upper part of the Nemuro Group, which is distributed in the coastal area. The Formation consists of alternating beds of sandstones and mudstones, thick coarse sandstones, and pebbly mudstones containing sedimentary blocks (Kiminami, 1979). This formation is 900–1300 m thick and has been interpreted as submarine channel-levee complexes with MTDs (Naruse, 2003). The deposits of the channel-levee complexes comprise two types of successions: channel-fill deposits consisting of pebbly mudstones and small-scale MTDs, and levee and overbank deposits composed of mud-prone alternating beds of thin sandstones and mudstones. Large-scale MTDs can also be observed in the Akkeshi Formation.

This study surveyed the 1.6-km continuous and 10-m high outcrop on the Esashito

coast, Hamanaka city, where the large-scale MTD bed of the Akkeshi Formation is well exposed. The coastline extends ENE–WSW that is nearly parallel to the strike of the Akkeshi Formation in this region, indicating that the same horizon of the MTD bed is exposed along the outcrop. Through paleo-stress analysis, Naruse and Otsubo (2011) clarified that this MTD was deposited from the gravity flow that migrated from the northern west to the southern east. Thus, the eastern part of the outcrop along the Esashito coast exposes the downstream part of the MTD.

The MTD in this study area primarily consists of poorly sorted pebbly mudstone as the matrix, which contains various sizes of sedimentary blocks composed of alternating beds of sandstones and mudstones. The diameters of these blocks generally range from 1–10 m, but they can reach 100 m. Naruse and Otsubo (2011) divided the MTD into three facies according to the size, mode of accumulation, and degree of deformation of sedimentary blocks. Facies A is characterized by matrix-supported small blocks sparsely scattered in the pebbly mudstone matrix. Facies B comprises clast-supported blocks. In this facies, the blocks have moderate diameters (10–30 m) and exhibit extensive soft-sediment deformation (i.e., slump folding). Facies C consists of large blocks nearly 100 m in diameter. Facies A and B are distributed in the upstream part of the MTD, while Facies C occurs in the distal end of the bed. The multiple inverse method for reconstructing paleo-stress fields was applied to mesoscale faults in deformed blocks, revealing that the mass-transport process initially expanded radially on the submarine slope and was compressed parallel to the paleocurrent direction during deposition (Naruse and Otsubo, 2011; Otsubo et al., 2018).

3 METHODOLOGY

In this study, a 3D point cloud was constructed using outcrop images captured by a drone along the Esashito coast. A limited portion of the 3D point cloud data was translated into a set of two-dimensional (2D) images to generate the training dataset for the CNN model. The images had four channels exhibiting RGB colors and roughness properties. Labels for the outcrop facies were manually assigned to the pixels on these 2D images to generate the training dataset for teaching the CNN model. It is important to note that the term “outcrop facies” includes vegetation, beaches, and topsoil, which are not usually referred to as facies. Actual 3D outcrop models inevitably include these surface states. In addition to the lithofacies, an essential task of the model proposed in this study is to remove these elements that cover outcrops for geological analyses efficiently. A U-net-type architecture with residual connections was adopted for the CNN model to conduct semantic segmentation of the outcrop images. The trained model was finally applied to the 2D images covering the entire outcrop, and the classes of outcrop facies predicted by the CNN were translated to the original 3D point cloud, producing a 3D facies model of the outcrop.

Construction of 3D Outcrop Point Cloud

A 3D outcrop model was constructed from outcrop images captured by a drone (DJI Phantom 4 Pro). The images were taken in a range of 700 m of the coastal cliff parallel to the Esashito coast extending from the northern east to the southern west

(43.166°N, 145.270°E). In this study, 4235 outcrop images were taken, whose resolution was 5472×3648 pixels. Each image covered a region approximately 20 m wide and 10 m high. These images were taken at different points and angles, overlapping each other by almost 80%.

A 3D point cloud representing the outcrop surface was constructed from the outcrop images by photogrammetry using ContextCapture produced by Bentley Systems, Incorporated. Each point in the point cloud had information about its coordinates (x , y , and z) and RGB 8-bit values.

Translation from 3D Point Cloud to 2D Images

In this study, the point cloud exhibiting the outcrop was segmented and translated into a set of 2D images with 224×224 pixels in resolution. The translated images had four channels. Three channels in the image datasets were used to record RGB color values. The other channel recorded the roughness distance h , defined as the distance of the outcrop surface from the fitted plane. The procedure for obtaining this roughness distance is explained below. The distribution of the roughness distance contains information such as bulges of gravel or protrusion of sandstone layers on the outcrop surface.

The procedures for translating the 3D point cloud to 2D images were as follows.

(1) To reduce the data size and remove noise, a 3D median filter with a radius of 2 cm was applied to the point cloud. The filter outputted medians of x , y , and z coordinates and RGB values. In this study, the distance of the filter movement was set to 2 cm, and thus, the procedure resulted in a smoothed point cloud in which the

distance between the points was also 2 cm.

(2) Subsets of the smoothed point cloud were extracted using a spheric sampling window with a diameter of $\sqrt{3}$ m (Figures 2A, B, and C). The sampling interval was set to 1 m so that the sampled subsets partly overlapped. These sampling window sizes and intervals were designed to sample all points at least once.

(3) A fitted plane to the outcrop surface was obtained from each subset of the point cloud using the least-squares method (Figure 2D). Then, each point in the point cloud subset was projected to the fitted plane, and the distance h from the point to the plane was calculated. This distance h reflects the regional roughness of the outcrop; thus, h is referred to as a roughness distance.

(4) An image of 224×224 pixels covering a $\sqrt{3}$ m square area of the outcrop was generated from the projected points in each subset of the point cloud (Figure 3). The fitted outcrop plane was gridded into 224×224 pixels, and the projected RGB values and roughness distance were interpolated at the pixel grid points. The interpolation was performed using the radial basis function (RBF) method with the Gaussian function. Because the interpolated images occasionally contained outliers, values exceeding or lower than the prescribed threshold values were rounded down or up so that all values fell within a certain range. Thresholds for RGB values ranged from 0–255, and those for roughness distances ranged from -0.2 to 0.2 m based on naked-eye observations of the outcrop. Finally, all values of the four channels of these 2D images were normalized in each channel from -1 to 1 using the minimum and maximum values.

Open3D ver.0.8.0.0 (Zhou et al., 2018), a Python package, was used to process 3D

point cloud data. The interpolation using the RBF method was computed using
scipy ver.1.7.3 (Virtanen et al., 2020).

Labeling of Training Datasets

Labeled images were generated by manually annotating labels for the outcrop
facies to the pixels on the 2D images translated from the 3D point cloud (Figure 4).
In this study, five classes of outcrop facies were assigned to the images: 1) pebbly
mudstone, 2) sedimentary block, 3) vegetation, 4) beach, and 5) topsoil. Classes 1
and 2 were lithofacies, while others were areas in which the MTD was not exposed.
The roughness distance h was not referred to when these classes were manually
labeled.

The pebbly mudstones were identified as blue-gray gravels scattered in the
mudstone matrix in naked-eye observations. The sedimentary blocks represent the
well-defined beddings of sandstones and mudstones eroded by weathering. The
vegetation area can be recognized as green to light green mottled patterns in the
images. The beach was a red-brownish monotonous region. The topsoil was a
brown region with a rough surface representation.

In this study, the training images were mainly selected from two regions of the
outcrop (Figure 5). The first region was 14 m wide and 30 m high and was located
approximately 350 m from the western end of the outcrop (Figure 10A and C). This
region was split into 1706 images (Figure 5A). The second region was 2 m wide and
30 m high and was located approximately 480 m from the western end of the outcrop
(Figure 10A and C). This region was split into 322 images (Figure 5C). This

continuous region sampling makes it easy to match the actual outcrop facies when labeling the classes in small 2D images. These two regions were selected because they include all classes evenly. In addition to these two continuous sampling regions, 192 images were randomly taken from the entire outcrop region.

In total, 2220 images of 224×224 pixels were labeled. Among these labeled images, 1776 and 444 were randomly selected for the training and validation of the CNN model, respectively.

Training of CNN Model

This study employed a CNN as the model for automated classification of the outcrop facies. A CNN is one of the machine learning models that is frequently applied to image classification and semantic segmentation problems (Lecun et al., 1998). For semantic segmentation of regions of classes in 2D images, we constructed a U-net-type neural network with residual connections called ResUNet (Zhang et al., 2018; Diakogiannis et al., 2020; Shang et al., 2021) that used ResNet (He et al., 2016) in its encoder and decoder networks. This study adopted the structure of ResNet50 (He et al., 2016-06), and the initial weights of the ResUNet encoder in this study were copied from those of ResNet50 previously trained on the ImageNet dataset (transfer learning). In this study, the ResUNet model was built by Python 3.9 with Tensorflow 2.8.2 (Abadi et al., 2015) and Keras 2.8.0 (Chollet et al., 2015).

The categorical cross-entropy loss E was adopted as a loss function to train the CNN model, which takes the following form:

$$E = -\sum_c \sum_k t_{k,c} \log y_{k,c} \quad (1)$$

265

266 where $y_{k,c}$ denotes an output of the CNN model, indicating a probability that the k -
 267 th pixel is classified into the c -th class. When the k -th pixel is classified into the
 268 class c in the training image, the parameter $t_{k,c}$ is set to 1; otherwise, it is set to 0. The
 269 CNN model was trained by optimization of weight coefficients to minimize this loss
 270 function. The stochastic gradient descent (SGD) method (Duchi et al., 2011) was
 271 adopted as the optimization method in this study because the original literature
 272 proposing ResNet employed it (He et al., 2016). SGD is a standard optimization
 273 method used to train neural networks, which divides the training data into small
 274 subsets called mini-batches, and the weight coefficients of the network are modified
 275 using the gradient of the loss function for each mini-batch to optimize the network
 276 for the analysis. The batch size, defined as the number of training data included in
 277 the subset, was set to 8 in this study. Each mini-batch was given to the model to
 278 optimize the weight coefficients of the network, and the combination of the datasets
 279 in the mini-batches was shuffled after all the training data were fed to the CNN
 280 model. This cycle is called an epoch. In this study, the model was trained on 700
 281 epochs in total. The other hyperparameters used to train the ResUNet model are
 282 listed in Table 1. The initial learning rate and its decay rate of SGD were manually
 283 optimized, searching the values between 0.01–0.19 and $0.0-2.0^{-5}$, respectively.

284 Categorical accuracy A was used to evaluate the performance of the model after
 285 training. This metric is defined as follows:

$$A = \frac{\sum TP_i}{\sum (TP_i + FN_i)} \quad (2)$$

287

288 where TP_i and FN_i denote the number of pixels predicted to be or not to be the i -th
289 class, respectively, and they were classified into the i -th class in the training image.

290 Categorical accuracy A becomes 0 when the model fails to classify all pixels and
291 is unity when all pixels are classified correctly.

292

Test of Models

Two conditions for model training were tested to examine the significance of the outcrop roughness property in recognizing the classes of outcrop facies using the model. The model was trained on images with only RGB values (Condition 1) and on images with both RGB values and roughness distance h (Condition 2). The test dataset containing 200 outcrop images was randomly extracted from the entire outcrop region, and the facies labels were obtained by manual tracing. This dataset was not used in the training process. The performance of the models trained under Conditions 1 and 2 was evaluated using the test dataset in terms of *recall*, *presicion* and F_1 -score, defined as follows:

$$recall_i = \frac{TP_i}{TP_i + FN_i} \quad (3)$$

$$precision_i = \frac{TP_i}{TP_i + FP_i} \quad (4)$$

$$F_{1i} = \frac{2TP_i}{2TP_i + FN_i + FP_i} \quad (5)$$

where FP_i denotes the number of pixels predicted to be the i -th class even though they were not classified into the i -th class in the training image. *Recall* describes how many samples are correctly classified in all the actual positive samples. *Precision* describes how many samples are actually positive in all the samples predicted as positive.

Transcript of Predicted Labels

The classes of outcrop facies predicted from the 2D images using the CNN model were inversely transcribed to the 3D point cloud to reconstruct 3D facies models. This transcriptional process is the opposite of how the 3D outcrop point cloud was projected onto the fitted outcrop plane. The relation between the subsets of point clouds and the generated 2D images was recorded during the initial transcribing procedure. This record was used to determine to which pixel the location projected from each 3D point on the fitted plane belongs. Then, the class of the outcrop facies on the pixel was transcribed to the 3D point to reconstruct the 3D distribution of the classes in the outcrop. Figures 5B and D show examples of the 3D distribution of the outcrop facies classes reconstructed from manually labeled images.

4 RESULTS

Characteristics of 3D Outcrop Data of MTD on Esashito Coast

A 3D outcrop model of the Esashito coast was constructed from photographs captured by the drone (Figure 7). The density of the points was approximately 13000 points/m² on the outcrop surface. The constructed 3D point cloud covered the 717-m-wide region of the outcrop.

The characteristics of the MTD can be observed in this 3D point cloud. In the range of 0–200 m from the southwestern end of the outcrop, sedimentary blocks with a diameter of approximately 30 m are observed (Figure 8). The blocks typically exhibit slump folding and pile up with each other. The MTD 200–350 m from the southern western end of the outcrop is primarily composed of pebbly mudstones (Figure 8). Blocks with a diameter of 1 m on

average, rarely with a diameter of 10 m, are scattered in the matrix in this region. In the remaining areas of the outcrop, four large blocks with a diameter approaching 100 m are observed. Three of the four blocks on the southwest side are deposited with a horizontal internal bedding surface. The rest at the northeastern end of the outcrop is deposited with a bedding surface approximately vertically.

Training of CNN Model

The CNN model used in this study was trained under two conditions: the training dataset without the roughness distance (Condition 1) and that with the distance (Condition 2). Figure 9 depicts the training histories of the CNN model under these conditions in 700 epochs. The loss and evaluation function values for the training and validation datasets varied consistently during the training epochs, indicating that no significant overlearning occurred. The evaluation function for the validation dataset resulted in high values (greater than 0.95 in the median for the last 100 epochs) for both conditions (Table 2).

The performance of the CNN models trained under Conditions 1 and 2 was evaluated using the unseen test dataset (Table 3). The results show that the CNN model trained with roughness distance information (Condition 2) is superior in terms of outcrop lithofacies prediction (pebbly mudstones and sedimentary blocks) to the model trained without roughness distance information (Condition 1). The averaged F_1 -scores weighted by the number of pixels in each class were nearly identical for Conditions 1 (86.9%) and 2 (87.1%). However, the F_1 -score for pebbly mudstones and sedimentary blocks were 53.5% and 43.9% for the model trained under Condition 1, respectively. Meanwhile, they were 63.4% and 55.6% for the model trained under Condition 2, respectively. Recall and precision of

lithofacies were highly improved in Condition 2 except for the precision of sedimentary blocks. In contrast, the F_1 -score for the vegetation area class was higher under Condition 1 (95.3%) than Condition 2 (91.2%).

Reconstruction of 3D Facies Model

The 3D facies models of the outcrop along the Easashito coast were reconstructed using the CNN models trained under Conditions 1 and 2 (Figure 10). A comparison with naked-eye observations proved that the lithofacies were well predicted for the entire outcrop region, and thus, the facies distribution in the outcrop can be recognized appropriately in the model. In both models, many large sedimentary blocks were embedded in the pebbly mudstone matrix of the MTD, but the large regions of the sedimentary blocks were erroneously classified into pebbly mudstones in the 3D facies model of Condition 1 (no outcrop roughness information). The identification of the region of blocks was significantly improved in the facies model of Condition 2 (with outcrop roughness information).

5 DISCUSSION

Significance of the Roughness Properties of Outcrops for Lithofacies

Identification

This study performed semantic segmentation of the outcrop facies based on the color and roughness properties of the outcrop. The roughness property significantly improved lithofacies classification (pebbly mudstone and sedimentary blocks) (Table 3; Figure 10).

Under Condition 1, lithofacies classes were incorrectly identified compared with other classes (Table 3). These mispredictions could have been caused by color differences caused by weathering between the two lithofacies. However, by focusing on surface irregularity rather than color properties, lithofacies classes had more consistent features. The pebbly mudstone consisted of gravels scattered in the muddy matrices, which exhibited irregular rough surfaces. The sedimentary blocks comprised alternating beds of sandstones and mudstones, whose exposed surfaces were smooth but protruded with a plate-like shape. Conversely, under Condition 2 (with roughness property in the training data), the CNN model predicted pebbly mudstone and sedimentary blocks more accurately than the model under Condition 1.

In contrast to lithofacies classes, the roughness properties were not necessarily helpful in identifying the other classes. The vegetation areas had a monotonous appearance without any informative change in surface morphology. The CNN model identified this class with significantly higher accuracy under the training condition that considered only the color properties of the outcrop. The F_1 -score for the class of the vegetation area was lower under Condition 2 than Condition 1, probably because the vegetation area significantly varied in roughness properties.

The beach area's homogeneity in color and surface roughness exhibited slightly higher F_1 -scores under Condition 2, implying that the roughness property improves classification accuracy for classes with consistent surface features. The topsoil has low F_1 -scores under both conditions. The topsoil is observed overlying other lithofacies and distributed discontinuously throughout the outcrop, with the appearance varying depending on whether the topsoil is overlaying the pebbly mudstone or the sedimentary blocks.

In summary, considering the outcrop roughness properties improves the performance of 3D facies models. Lithofacies exhibit characteristic features of surface irregularity because of their resistance to outcrop weathering. Although other areas where deposits are not exposed might not be characterized by surface roughness, the accuracy of lithofacies prediction is highly improved from the viewpoint of geological applications.

Comparison with Previous Studies

Several research endeavors have been directed toward reconstructing 3D facies models of outcrops to explore the spatial arrangement of various lithofacies. Kirsch et al. (2019) used support vector machine, machine learning technique, to discriminate lithofacies and construct 3D facies models. This method leveraged data on surface topography collected through photogrammetry of drone imagery, which is similar to the approach used in this study. However, their study did not consider the surface morphology of outcrops for lithofacies identification. Rather, their investigation focused on probing the physical attributes of rocks captured using a hyperspectral camera, using them as indicators for lithologic classification. Although prediction accuracy was not explicitly quantified, the employed technique provided a satisfactory representation of the lithologic distribution within the outcrops. A limitation of their method lies in the time-intensive nature of data collection using hyperspectral cameras and the inherent difficulty in detecting areas where rocks are not exposed. In contrast, our study uses only drone imagery for data acquisition, conferring a distinct advantage in terms of ease of application and overall efficiency.

A study that parallels our approach was conducted by Malik et al. (2022), focusing on semantic segmentation (pixel-by-pixel discriminant analysis) of images for lithofacies

identification. In their investigation, Malik et al. (2022) applied semantic segmentation to 2D outcrop images to discern different lithofacies, constructing an effective lithofacies discrimination model. The model achieved accuracy exceeding 80%, demonstrating the validity of using a U-net-type model similar to that in our study. Notably, however, their work targeted 2D images, unlike the 3D point cloud data analyzed in our research. This finding indicates that CNN models, such as the one employed in this study, can be universally applied to lithofacies identification. Our results further reveal that outcrop roughness is a sufficient indicator for lithofacies recognition. Hence, it is anticipated that incorporating information about outcrop surface morphology into the dataset trained by Malik et al. (2022) could lead to the development of CNN models with even higher lithofacies identification accuracy.

Possible Applications to Geological Analyses

Our method can be applied to various research purposes, in particular, characterizing the chaotic internal structure of the MTD. Naruse and Otsubo (2011) demonstrated that the interior of the MTD is highly heterogeneous. They revealed that the size of the blocks contained within the MTD indicates a power-law distribution and that the concentration of blocks varies periodically. However, quantifying the size and shapes of blocks in MTDs by manually tracing outcrops is labor-intensive, and thus, there have been few studies of the same kind. Because our method can be easily applied to large outcrops, significant progress in elucidating the internal structure of MTDs can be expected in the future.

Examining the internal structure of deposits exposed in large outcrops is also essential for assessing the reservoir quality of sedimentary rocks. For example, the 3D outcrop facies

model helps conduct flow simulations to assess the hydrological properties of the sandstones. Siddiqui et al. (2019) reconstructed a 3D model of an outcrop using light detection and ranging (LiDAR) data, complemented by outcrop photographs, to determine the distribution of four distinct types of sandstone facies. Although their study relied on manual interpretation of sandstone facies by visual inspection, implementing automated lithofacies recognition using the ResUNet model proposed in this research enhances their method, making it more readily adaptable to more extensive outcrops. It should be noted that the CNN model developed in this study can provide the probability of each class for every pixel. Thus, the stochastic algorithm considering the uncertainty of facies for the voxels can be adopted for such flow simulations using the 3D subsurface model.

Model Limitations and Future Improvements

One limitation of the approach used in this study, which projects the outcrop surface onto a plane, is that it cannot handle situations in which the outcrop is overhung relative to the projection plane. Such a situation may occur in the alternating sandstone and mudstone beds, where the sandstone beds protrude from the outcrops at an oblique angle. In such an overhanging case, multiple points on the outcrop surface are plotted against the projection plane. As a result, regions in the produced outcrop 2D images will show anomalous noise in color and roughness information because of the interpolation of the extremely fluctuated values of the regions (Figure 11). To solve this problem, the colors and roughness of the outcrop surfaces should be considered in three-dimensional space, which is possibly achieved by a neural network such as PointNet (Qi et al., 2017) that handles 3D point clouds

as inputs instead of projecting them onto 2D images, as done in this study, for outcrops with many obliquely protruding objects.

Another issue in this study is the accumulation of training data. This study used 1776 images created by artificial tracing as data. Although this training dataset was sufficient to determine the lithology of the outcrops investigated in this study, new training data will likely be required when interpreting other outcrops. In the future, constructing a CNN model that can be universally used for facies discrimination by accumulating training data on various outcrops would be desirable.

6 CONCLUSIONS

This study proposed a method for reconstructing 3D facies models using a CNN for 3D point clouds obtained by photogrammetry of outcrop images captured by a drone. The model developed in this study decomposes 3D outcrop data into a series of 2D images, identifies their facies, and reconstructs them into a 3D model. Our method is unique in that it retains not only the surface color but also the outcrop surface roughness information in the image data and uses both to discriminate the facies.

The target selected for training and testing the model developed in this study is the outcrop of the MTD in the Cretaceous–Paleocene Akkeshi Formation in eastern Hokkaido, northern Japan. Spanning 1.6 km in length, this outcrop exposes a single MTD bed comprising sedimentary blocks of sandstone–mudstone alternations and pebbly mudstone matrices. The primary aim of the model proposed in this study was to precisely discern areas of the exposed MTD within the outcrop and distinguish between the two types of lithofacies in those areas.

The procedures for decomposing a 3D outcrop to 2D images of 224×224 pixels were as follows. (1) A 3D median filter was applied to the 3D point cloud. (2) Regional point clouds were extracted at regular intervals. (3) A plane was fitted to each extracted point cloud using the least-squares method, and the points were projected onto the fitted plane. (4) The RGB values and roughness distance of each point from the projected plane were interpolated at the pixels of a 2D image. Thus, the RGB colors and roughness distance of the 3D point cloud were translated into channels of 2D images.

A CNN model was developed to classify the facies of the outcrop from decomposed 2D images, using the ResUNet framework for the model structure. From the 2D images of the outcrop, 1776 images were selected for training, 444 images were used for validation, and 200 images were chosen for testing. The interpretation results of the test dataset using the trained model were then compared with those obtained through artificial tracing. With an average F-1 score of 86.4%, the performance of the model is sufficiently practical. Furthermore, this study explored the impact of including or excluding outcrop roughness information on the discrimination ability of the model. Notably, incorporating roughness information into outcrop lithofacies identification markedly improved the performance of the model.

Finally, a 3D facies model was constructed by applying the trained model to the entire outcrop exposed at the Esashito coast and translating the semantic segmentation results from the 2D images into a 3D point cloud. This process verified that the model can accurately identify the facies across the entire outcrop.

As the techniques developed in this study are implemented across various areas in the future, they can make significant advancements in reconstructing sedimentary processes and

assessing reservoir rock quality. Using drones to gather data from expansive outcrops, which have previously been challenging for direct human surveys, simplifies the process considerably. Applying our findings to previously under-surveyed areas will facilitate the acquisition of sedimentological data, thereby expanding our understanding and capability in the field.

REFERENCES CITED

Abadi, M. et al., 2015, Tensorflow: Large-scale machine learning on heterogeneous systems, accessed July 25, 2023, <https://www.tensorflow.org/>.

Aicardi, I., F. Chiabrando, N. Grasso, A. M. Lingua, F. Noardo, and A. Spanò, 2016, Uav photogrammetry with oblique images: First analysis on data acquisition and processing: The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences, v. 41, doi:10.5194/isprs-archives-XLI-B1-835-2016.

Cardona, S., L. J. Wood, B. Dugan, Z. Jobe, and L. J. Strachan, 2020, Characterization of the Rapanui mass-transport deposit and the basal shear zone: Mount Messenger Formation, Taranaki Basin, New Zealand: *Sedimentology*, v. 67, no. 4, p. 2111–2148, doi:10.1111/sed.12697.

Chollet, F. et al., 2015, Keras, accessed July 25, 2023, <https://keras.io>.

Colomina, I. and P. Molina, 2014, Unmanned aerial systems for photogrammetry and remote sensing: A review: *ISPRS Journal of Photogrammetry and Remote Sensing*, v. 92, p. 79–97, doi:10.1016/j.isprsjprs.2014.02.013.

Diakogiannis, F. I., F. Waldner, P. Caccetta, and C. Wu, 2020, ResUNet-a: A deep learning framework for semantic segmentation of remotely sensed data: *ISPRS Journal of Photogrammetry and Remote Sensing*, v. 162, p. 94–114, doi:10.1016/j.isprsjprs.2020.01.013.

Duchi, J., E. Hazan, and Y. Singer, 2011, Adaptive subgradient methods for online learning and stochastic optimization: *Journal of Machine Learning Research*, v. 12, no. 61, p. 2121–2159.

Geological Survey of Japan, National Institute of Advanced Industrial and Science Technology, 2023, Seamless digital geological map of Japan V2 (original edition): Tsukuba, Japan, Geological Survey of Japan, National Institute of Advanced Industrial and Science Technology, scale 1:200,000, 1 sheet.

He, K., X. Zhang, S. Ren, and J. Sun, 2016, Deep residual learning for image recognition: *Proceedings of the Institute of Electrical and Electronics Engineers Conference on Computer Vision and Pattern Recognition*, p. 770–778.

Jorry, S., E. Davaud, and B. Caline, 2003, Controls on the distribution of nummulite facies: A case study from the Late Ypresian El Garia Formation (Kesra Plateau, central Tunisia): *Journal of Petroleum Geology*, v. 26, no. 3, p. 283–306.

Kiminami, K., 1978, Stratigraphic re-examination of the Nemuro Group: *Chikyu-Kagaku (Earth Science)*, v. 32, no. 3, p. 120–132.

Kiminami, K., 1979, Sedimentary petrological research of the Nemuro Group: *Chikyu-Kagaku (Earth Science)*, v. 33, no. 3, p. 152–162.

Kirsch, M. et al., 2019, Hyperspectral outcrop models for palaeoseismic studies: *The Photogrammetric Record*, v. 34, no. 168, p. 385–407, doi:10.1111/phor.12300.

Krizhevsky, A., I. Sutskever, and G. E. Hinton, 2017, ImageNet classification with deep convolutional neural networks: *Communications of the ACM*, v. 60, no. 6, p. 84–90, doi:10.1145/3065386.

LeCun, Y., L. Bottou, Y. Bengio, and P. Haffner, 1998, Gradient-based learning applied to document recognition: *Proceedings of the Institute of Electrical and Electronics Engineers*, v. 86,

no. 11, p. 2278–2324, doi:10.1109/5.726791.

Liu, X., H. Wang, H. Jing, A. Shao, and L. Wang, 2020, Research on intelligent identification of rock types based on faster R-CNN method: Institute of Electrical and Electronics Engineers Access, v. 8, p. 21804–21812, doi:10.1109/ACCESS.2020.2968515.

Malik, O. A., I. Puasa, and D. T. C. Lai, 2022, Segmentation for multi-rock types on digital outcrop photographs using deep learning techniques: *Sensors*, v. 22, no. 21, p. 8086, doi:10.3390/s22218086.

Naruse, H., 2003, Cretaceous to Paleocene depositional history of North-Pacific subduction zone: Reconstruction from the Nemuro Group, eastern Hokkaido, northern Japan: *Cretaceous Research*, v. 24, no. 1, p. 55–71, doi:10.1016/S0195-6671(03)00024-7.

Naruse, H. and M. Otsubo, 2011, Heterogeneity of internal structures in a mass-transport deposit, upper Cretaceous to Paleocene Akkeshi Formation, Hokkaido island, northern Japan, *in* R. C. Shipp, P. Weimer, and H. W. Posamentier, eds., *Mass-Transport Deposits in Deepwater Settings: SEPM Special Publication 96*, p. 279–290, doi:10.2110/sepmsp.096.279.

Otsubo, M., H. Naruse, and A. Miyakawa, 2018, Temporal changes in the internal stresses and pore pressures in a large-scale submarine mass transport deposit: *Progress in Earth and Planetary Science*, v. 5, no. 59, doi:10.1186/s40645-018-0219-0.

Pascual, A. D., K. Mcisaac, and G. Osinski, 2021, Autonomous and real-time rock image classification using convolutional neural networks, preprint, doi:10.20944/preprints202109.0285.v1.

Qi, C. R., H. Su, K. Mo, and L. J. Guibas, 2017, Pointnet: Deep learning on point sets for 3D classification and segmentation: *Proceedings of the Institute of Electrical and Electronics Engineers Conference on Computer Vision and Pattern Recognition*, p. 652–660.

Ran, X., L. Xue, Y. Zhang, Z. Liu, X. Sang, and J. He, 2019, Rock classification from field image

589 patches analyzed using a deep convolutional neural network: *Mathematics*, v. 7, no. 8, 755 p.,
590 doi:10.3390/math7080755.

591 Shang, G., G. Liu, P. Zhu, J. Han, C. Xia, and K. Jiang, 2021, A deep residual u-type
592 network for semantic segmentation of orchard environments: *Applied Sciences*, v. 11, no. 1, 322
593 p., doi:10.3390/app11010322.

594 Siddiqui, N. A., M. Ramkumar, A. H. A. Rahman, M. J. Mathew, M. Santosh, C. W. Sum, and D.
595 Menier, 2019, High resolution facies architecture and digital outcrop modeling of the Sandakan
596 formation sandstone reservoir, Borneo: Implications for reservoir characterization and flow
597 simulation: *Geoscience Frontiers*, v. 10, no. 3, p. 957–971, doi:10.1016/j.gsf.2018.04.008.

598 Simonyan, K. and A. Zisserman, 2014, Very deep convolutional networks for large-scale image
599 recognition, preprint, <https://arxiv.org/abs/1409.1556v6>.

600 Sobiesiak, M. S., B. Kneller, G. I. Alsop, and J. P. Milana, 2016, Internal deformation and
601 kinematic indicators within a tripartite mass transport deposit, NW Argentina: *Sedimentary*
602 *Geology*, v. 344, p. 364–381, doi:10.1016/j.sedgeo.2016.04.006.

603 Thomas, H. et al., 2021, Contribution of drone photogrammetry to 3D outcrop modeling of facies,
604 porosity, and permeability heterogeneities in carbonate reservoirs (Paris Basin, Middle Jurassic):
605 *Marine and Petroleum Geology*, v. 123, 104772 p., doi:10.1016/j.marpetgeo.2020.104772.

606 Virtanen, P. et al., 2020, SciPy 1.0: Fundamental algorithms for scientific computing in python:
607 *Nature Methods*, v. 17, no. 3, p. 261–272, doi:10.1038/s41592-019-0686-2.

608 Wenbiao, Z., D. Taizhong, L. Zhiqiang, Y. Shujin, L. Yu, and X. Huaming, 2016, Application
609 of multi-point geostatistics in deep-water turbidity channel simulation: A case study of Plutonio
610 oilfield in Angola: *Petroleum Exploration and Development*, v. 43, no. 3, p. 443–450,
611 doi:10.1016/S1876-3804(16)30051-9.

612 Zhang, Z., Q. Liu, and Y. Wang, 2018, Road extraction by deep residual u-net: Institute of
613 Electrical and Electronics Engineers Geoscience and Remote Sensing Letters, v. 15, no. 5, p.749–
614 753, doi:10.1109/LGRS.2018.2802944.
615 Zhou, Q.-Y., J. Park, and V. Koltun, 2018, Open3D: A modern library for 3D data processing,
616 preprint, <https://arxiv.org/abs/1801.09847v1>.

617

Vitae

Ryusei Sato is a student of a master's degree at Kyoto University (Japan). His research mainly focuses on sedimentological analysis and methodological developments using image analysis for field surveys.

Kazuki Kikuchi is an assistant professor at Chuo University (Japan). His research focuses on ichnology of ancient deep-marine deposits, especially on relationships between variation in diversity of trace fossils and paleoenvironmental changes. Recently he has been working on core image analysis using convolutional neural networks.

Hajime Naruse is a professor at Kyoto University (Japan). His research focuses on the morphodynamics and sedimentology of tsunamis and sediment gravity flows based on field surveys and numerical or physical experiments. Recently, he has been working on the inverse analysis of the depositional process using machine learning methods.

FIGURE AND TABLE CAPTIONS

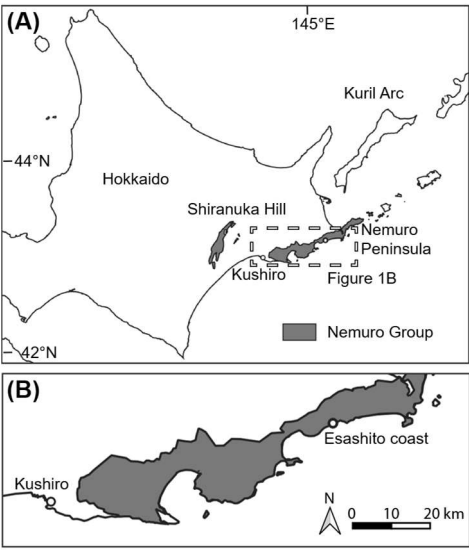
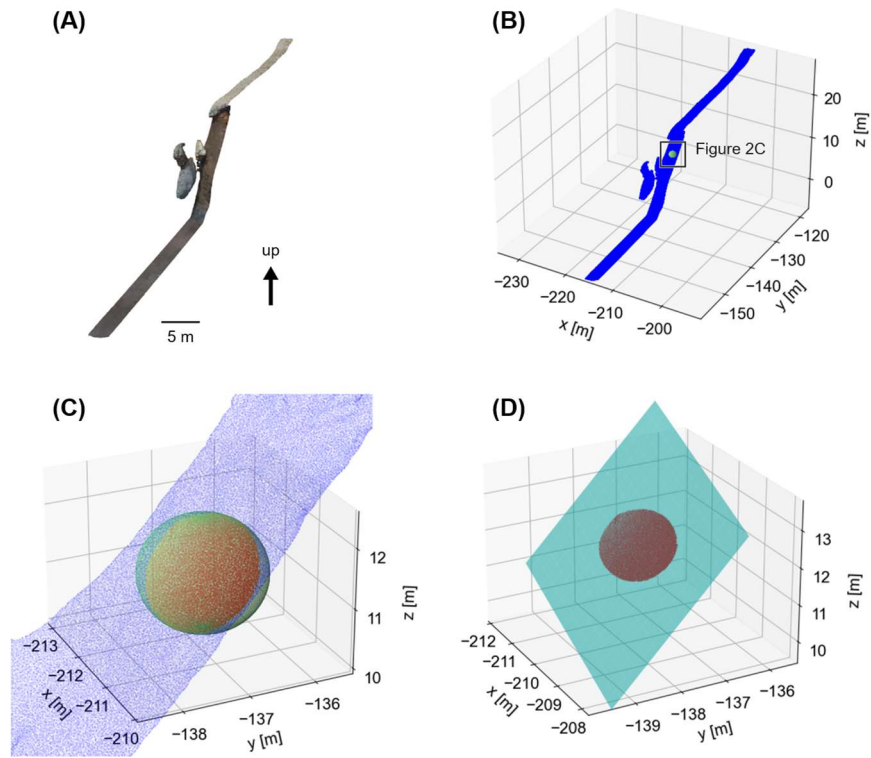
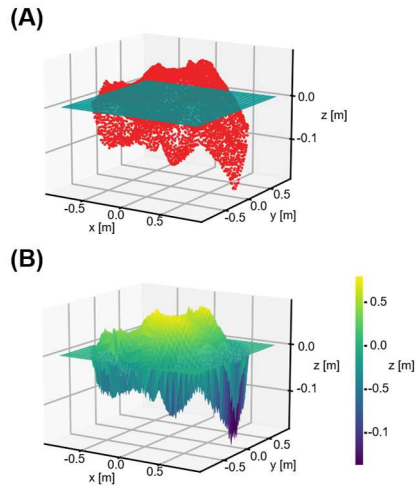


Fig. 1. Index map of the study area. A. Location of the study area. B. Distribution of the Nemuro Group and the location of the Esashito coast (43.166°N, 145.270°E). Maps are modified after Geological Survey of Japan, National Institute of Advanced Industrial and Science Technology (2023).



639
 640 Fig. 2. Schematic of the procedures for processing point cloud data. A. Outcrop point cloud with
 641 original colors segmented into approximately 2 m of width. The outcrop region was extracted from
 642 approximately 350 m from the western end of the outcrop (See Figures 5A and 10C). B. The same
 643 point cloud as in Figure 2A. (The colors are converted into blue.) C. Enlarged view of Figure 2B.
 644 Point cloud (red color) extracted by a green sphere with a diameter of $\sqrt{3}$ m. D. Plane
 645 approximating the average outcrop surface obtained from the point cloud.

646



647
 648 Fig. 3. Schematic of the procedures for interpolation of point cloud data. A. Plane approximating
 649 the outcrop surface obtained from a red point cloud, as shown in Figure 2D. B. Surface plot of the
 650 interpolated roughness distance h according to the point cloud of Figure 3A, with topographic color
 651 bands.
 652

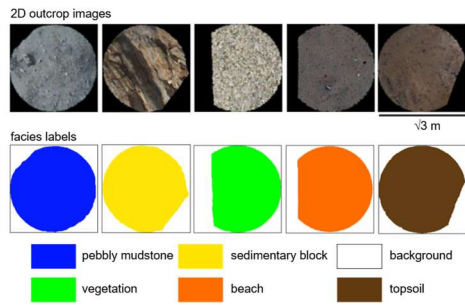
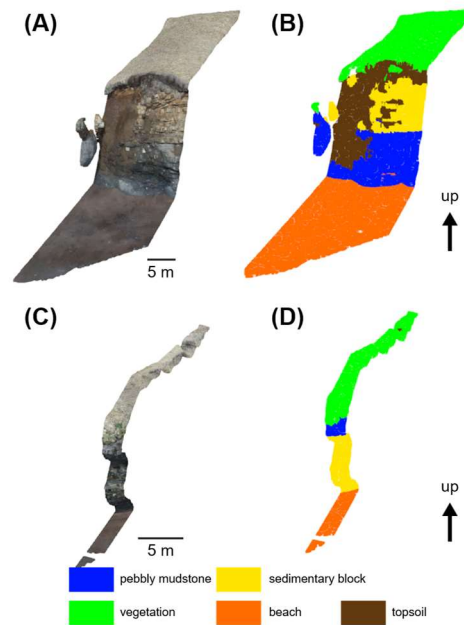


Fig. 4. Two-dimensional outcrop images transformed from three-dimensional point cloud and manually labeled images. The image size is 224×224 pixels, which corresponds to $\sqrt{3}$ m square region in the outcrop. The labeled region of the outcrop facies exhibits a circular shape because the spherical region of a point cloud was subsampled.



659

660

661

662

663

664

665

Fig. 5. Three-dimensional outcrop point clouds used as training data for the convolutional neural network (CNN) model. A, C. Original point cloud used as training data for the CNN model. B, D. Labeled outcrops of Figures 5A and 5C, respectively. Labeled outcrops are reconstructed from manually labeled two-dimensional images. The outcrop in Figure 2A corresponds to the left side region of Figure 5A.

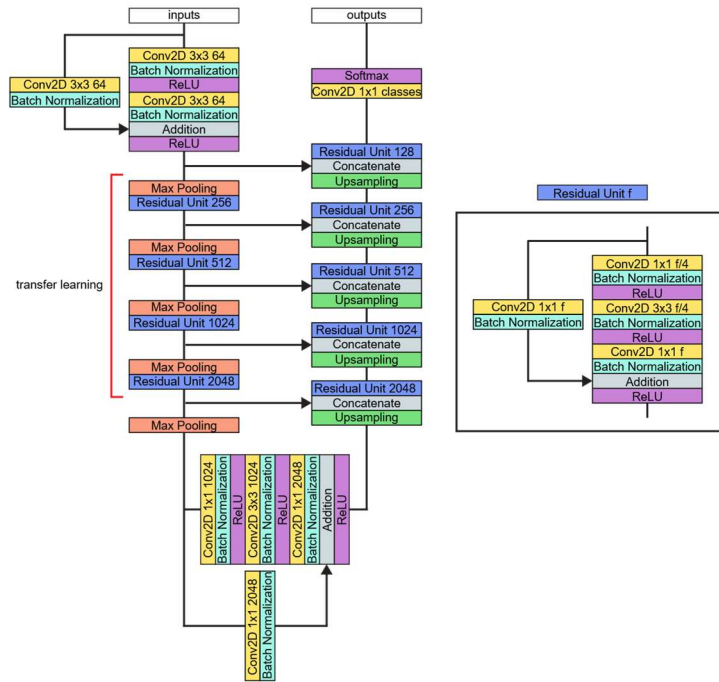
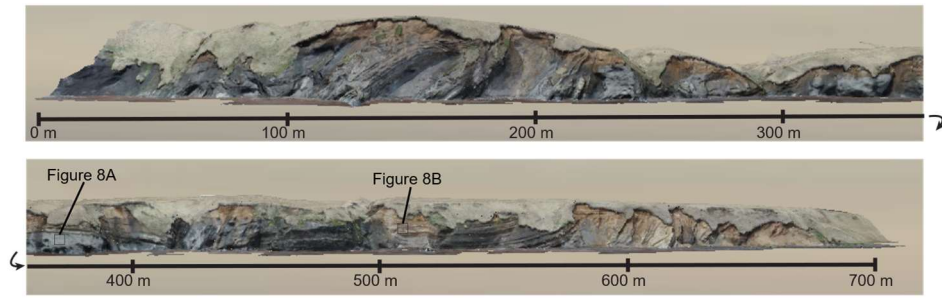


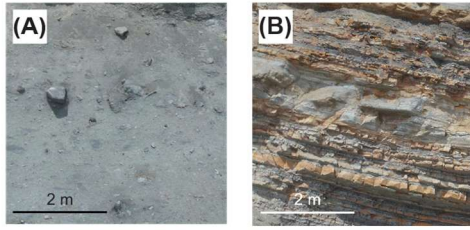
Fig. 6. Architecture of ResUNet. Transfer learning was applied to the encoder network.



669

670 Fig. 7. Three-dimensional outcrop model of the Esashito coast extending 717 m.

671



672
673 Fig. 8. Facies of mass-transport deposit (MTD) on the Esashito coast. A. Pebbly mudstone
674 comprising the MTD matrix. B. Alternating beds of sandstones and mudstones in a sedimentary
675 block embedded in the MTD matrix. The images were taken from the three-dimensional outcrop
676 model of Figure 7.

677

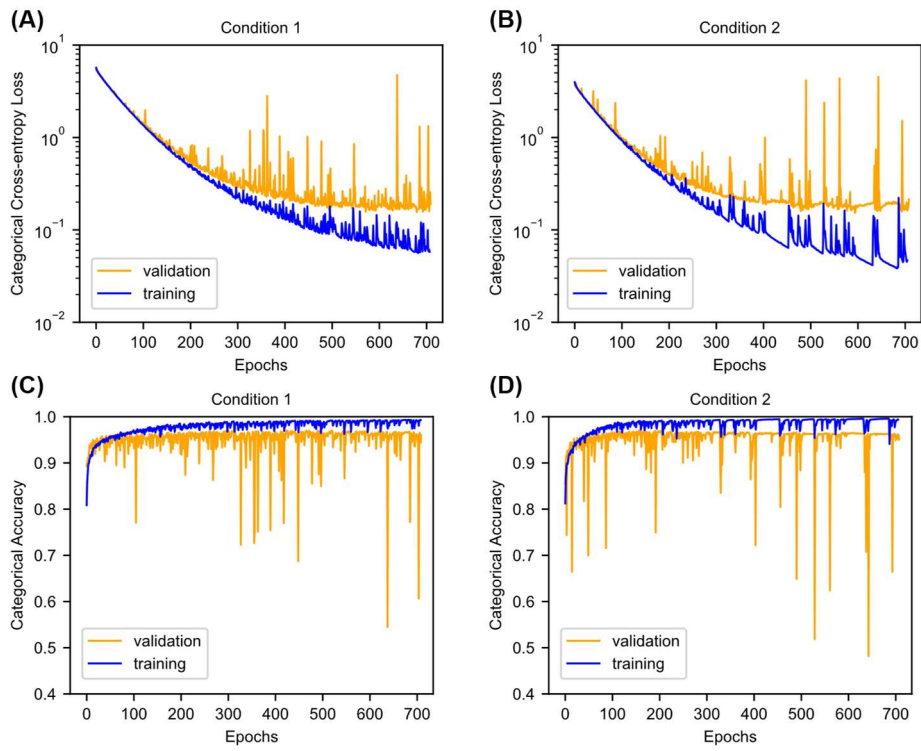


Fig. 9. Histories of training and validation loss and accuracy.

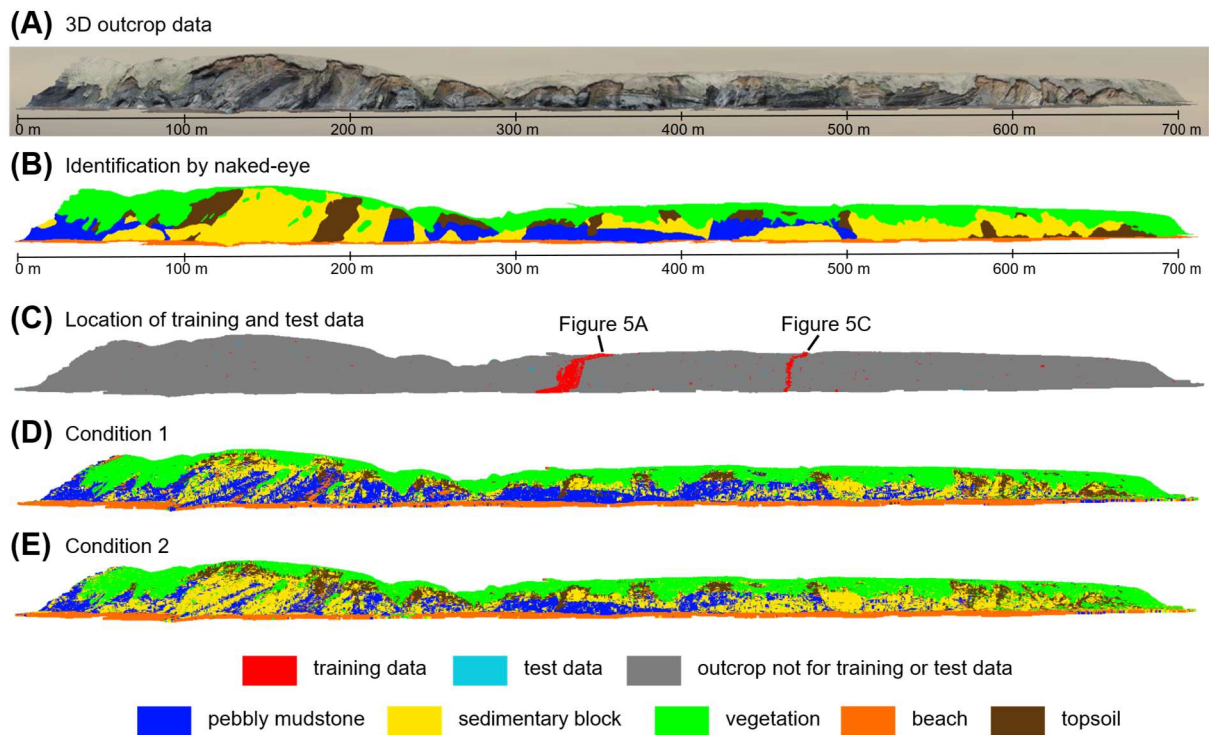
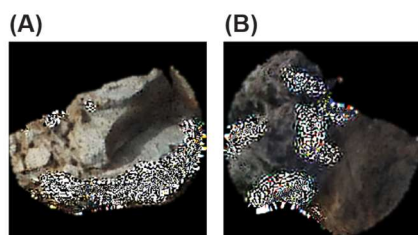


Fig. 10. Outcrop models. A. Original three-dimensional (3D) outcrop model. B. Sketch drawn by naked-eye observation of the original 3D outcrop model. C. Location of outcrops used for training and test data. D. 3D facies model reconstructed under Conditions 1. E. 3D facies model reconstructed under Condition 2.



687

688 Fig. 11. Images showing problems in the interpolation of RGB values. A. Sedimentary block. B.

689 Vegetation and topsoil. The sides of the images correspond to $\sqrt{3}$ m.

690

Epochs	Batch Size	Loss Function	Optimizer	Learning Rate	Decay
700	8	Categorical Cross-entropy Loss	SGD	0.18	0.00001

691

692 Table 1. Hyperparameters for training the convolutional neural network model. The same
693 hyperparameters were used under Conditions 1 and 2.

694

Condition	Training Loss	Training Accuracy	Validation Loss	Validation Accuracy
1	0.066	0.991	0.180	0.962
2	0.046	0.995	0.190	0.963

Table 2. Median values of the categorical cross-entropy loss and categorical accuracy of the convolutional neural network models in the last 100 training epochs.

Condition	Evaluation Metric	vegetation	beach	topsoil	mudstone	block	background	average
condition 1	True Positive	2940870 (94.3)	784700 (89.4)	332993 (45.0)	412771 (81.9)	298926 (38.3)	3946544 (98.3)	
	False Negative	176114 (5.7)	92955 (10.6)	407765 (55.0)	90933 (18.1)	482054 (61.7)	68575 (1.7)	
	False Positive	116006 (1.7)	166137 (1.8)	103820 (1.1)	626157 (6.6)	281214 (3.0)	25062 (0.4)	
	True Negative	6802210 (98.3)	8991408 (98.2)	9190622 (98.9)	8905339 (93.4)	8973006 (97.0)	5995019 (99.6)	
	recall	94.3	89.4	45.0	81.9	38.3	98.3	86.9
	precision	96.2	82.5	76.2	39.7	51.5	99.4	88.5
	F1-score	95.3	85.8	56.6	53.5	43.9	98.8	86.9
	True Positive	2706983 (86.8)	755489 (86.1)	348777 (47.1)	425973 (84.6)	521480 (66.8)	3918748 (97.6)	
	False Negative	410001 (13.2)	122166 (13.9)	391981 (52.9)	77731 (15.4)	259500 (33.2)	96371 (2.4)	
	False Positive	111793 (1.6)	92654 (1.0)	144719 (1.6)	414155 (4.3)	574219 (6.2)	20210 (0.3)	
condition 2	True Negative	6806423 (98.4)	9064891 (99.0)	9149723 (98.4)	9117341 (95.7)	8680001 (93.8)	5999871 (99.7)	
	recall	86.8	86.1	47.1	84.6	66.8	97.6	86.5
	precision	96.0	89.1	70.7	50.7	47.6	99.5	88.9
	F1-score	91.2	87.6	56.5	63.4	55.6	98.5	87.1

699

700 Table 3. Confusion matrix (number of pixels), recall, precision and F_1 -score showing the classification accuracy of the outcrop facies
701 by the trained models. The averaged *recall*, *presicion* and F_1 -score are weighted by the number of pixels. (mudstone: pebbly mudstone;
702 block: sedimentary block).