

Tumulus Distribution Prediction in the Iwase Senzuka Tumulus Cluster with Topographic Maps

**Mengbo You¹⁾ Kouichi Konno¹⁾ Atsushi Noguchi²⁾
Ryosuke Nakamura³⁾ Yuichi Takata⁴⁾**

1) Iwate University

2) Komatsu University Research Center for Next Generation Archaeological Studies

3) The National Institute of Advanced Industrial Science and Technology (AIST)

4) Nara National Institute for Cultural Properties

Abstract

Conducting comprehensive excavations in the extensive areas surrounding ancient tumulus clusters can reveal the locations of undiscovered tumuli. However, such endeavors demand significant manpower, resources, and time. By preemptively estimating the locations of these tumuli, we can avoid unnecessary excavations and improve efficiency. This study aims to examine the correlation between the distribution of ancient tumuli and topographical features. Leveraging this correlation, our goal is to forecast the likelihood of encountering ancient tumuli in various unexplored areas around the Iwase Senzuka tumulus cluster. This paper proposes a novel method to analyze the topographic point cloud data. The method involves superimposing archaeological markings onto the derived topographic map for annotation, cropping out patches and combining them to predict the likelihood of tumulus presence. The generated 2D distribution heatmaps are integrated and mapped to a color-coded point cloud for simultaneous observation of the important feature distribution identified by the deep learning network and examination of the 3D topography.

1 Introduction

The Wakayama Prefecture Kii-fudoki-no-oka Museum of Archaeology and Folklore was established in 1971 with the aim of conserving and showcasing the national special historic site, the Iwase Senzuka Kofungun (Iwase Senzuka tumulus cluster) [1]. This tumulus cluster is one of the largest clusters of tumulus in Japan built from the 5th to the 7th centuries. There are about 430 tumuli in the premises and about 700 tumuli in total in the entire area. The tumuli in this cluster are characterized by their unique horizontal stone chambers with stone shelves and beams. Another feature is that keyhole-shaped tumuli are commonly built along the ridge line at a maximum height of 150 meters. These characteristics are likely to result in a distribution of construction sites that differs from those in other parts of the world. Although many tumuli have already been discovered, archaeologists believe that there is still a high likelihood of numerous undiscovered ancient tumuli remaining in this area. The park of Iwase Senzuka tumulus cluster covers an area of approximately 67 hectares. However, conducting excavations in a systematic, area-by-area manner would inevitably consume significant manpower, resources, and time.

This study focuses on the problem of estimating the locations of tumuli in advance to eliminate the possibility of excavating them mistakenly and improve the efficiency of the excavation process. Previous research has found that the distribution of tumuli in other places of the world is associated with various topographic features and geographical variables [2, 3, 4]. For instance, statistical analysis has demonstrated a correlation between the distribution of archaeological sites in the desert area of southern Morocco and various geo-environmental variables, including topographic elevation, slope, orientation, and proximity to water sources [5]. The geo-environmental criteria, such as the digital elevation model (DEM), slope, hillshade, fertility, distance to river, and distance to the palm grove, were used to predict archaeological site locations in the desert area of southwestern Algeria [6]. Six geo-environmental factors of distance to rivers, distance to cropland, slope, aspect, elevation, and terrain texture were used to generate a archaeological predictive map for the Bekaa valley in Lebanon [7]. However, acquiring the

geographical variables and making judgements based on these variables requires a considerable amount of manual operation and relies on experienced experts.

Different from the desert or valley area in the previous research, the tumuli in Iwase Senzuka tumulus cluster are likely to have different distribution patterns and relationships with topographic information. Instead of obtaining the geo-environmental variables, this paper proposes that the correlation between the distribution of ancient tumuli and the topographical features extracted from the map image can be used to predict the likelihood of the presence of ancient tumuli in each potential location within the unexplored areas of Iwase Senzuka tumulus cluster. Before embarking on excavations, the available information for predicting the presence of ancient tumuli at a particular site is severely limited. Specifically, ancient tumuli are buried beneath the ground, often covered by vegetation above. Even when excluding the influence of noise caused by vegetation, the present observable topographic features may differ from those at the time of the tumulus construction due to gradual changes in the landscape over time. For instance, prominent features of the terrain at the time of construction, such as bulges or other landmarks, might have served as direct indicators of tumuli placement. However, over the years, these features may become less distinct or even completely disappear. In contrast, the relative positions of valleys, ridges, or hilltops are likely factors considered during tumulus construction, which may still be helpful for prediction. Therefore, this paper tries to extract such reliable topographic features from the surrounding area of three types of topographic maps for tumuli estimation.

2 Related work

In recent years, the use of machine learning methods has yielded promising results by automating the feature extraction and category classification process for generating predictive archaeological models [8, 9, 10]. Orengo et al. [11] demonstrated the potential of machine learning-based classification of multi-sensor, multi-temporal satellite data for the remote detection and mapping of archaeological mounded settlements in arid environments. The multi-sensor, multi-temporal satellite data could serve as a bottle-

neck for other archaeological predictive models applied in different places. Trier et al. [12] applied the deep neural networks to airborne laser scanning data for semi-automatic mapping of archaeological topography. Besides, the detection framework of YOLOv3 has been adopted to detect archaeological tumuli in North-Western Iberia [13]. These applications of deep learning methods focus on extracting the topographic features from LiDAR data. Generally, massive training images are needed to produce significant results. The characteristic tumular shapes in one place may also differ from those in other places, which also makes the collection of tumuli data difficult and the existing methods cannot directly applied to our task. However, the abovementioned methods represent only a small fraction of the broader field of machine learning research. Specifically, in the task of predicting archaeological tumuli, machine learning is often employed as a tool for processing and analyzing data that has already been measured and classified manually. The content and format of the data determine the ease of learning, which is one reason why there is relatively scant literature on using machine learning to predict mound locations.

On the other hand, the mainstream research about deep learning has achieved the state-of-the-art results in the task of image classification and object detection. However, the majority of literature on image classification and object detection focuses on distinctly different category labels, such as cats versus dogs [14, 15, 16], or pedestrians versus cars [17, 18, 19], which can also be distinguished by human observation. The training process can be regarded as updating the numerical feature values at every position within the network structure through numerous images and label data, ultimately using activation functions to establish a direct correspondence between the image data and labels. The result effectively emulates the human ability to distinguish visually. As the number of training and testing images grows with updates to benchmark challenges, the models become increasingly capable of handling image noise and varied shooting conditions (angles, lighting, etc.), sometimes even surpassing human visual distinguishing abilities under extremely challenging conditions [20, 21, 22]. However, the initial phase of defining labels still focuses on categories that are visually distinguishable. The subject of this paper

is different. Even with human eyes, it is difficult to discern the locations of tumuli on topographic maps compared to areas without tumuli. Archaeological experts can predict the likely presence of tumuli based on the terrain distribution that matches ancient criteria for placing tumuli, guided by their prior experience. In fact, the placement of tumuli often follows patterns that vary by region, culture, and tradition, making it challenging for those without sufficient archaeological knowledge or experience to differentiate between locations with and without tumuli. The goal of this study is to explore this issue, which has rarely been considered within the realms of image classification or object detection. It aims to use deep learning techniques to learn the distribution patterns of tumuli, thereby predicting the likelihood of tumuli at a given location.

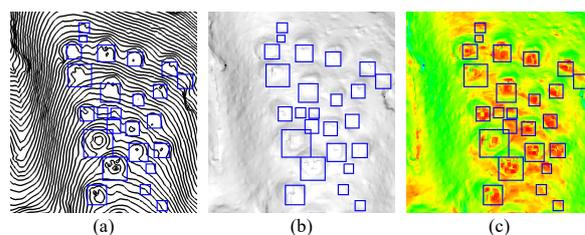


Figure 1: The examples of 3 topographic maps covering a small part of the Maeyama-A region. (a) contour line map. (b) shaded relief image. (c) slope map.

Our previous work [23] also utilized the topographic maps of Maeyama-A region (a part of the Iwase Senzuka tumulus cluster) as the research subject to estimate the locations of tumuli. This region includes 182 tumuli that have already been discovered, while other areas may still conceal undiscovered tumuli. The objective is to extract image features from the topographic maps around discovered tumuli that can distinguish areas with and without tumulus. These features are then used to predict the distribution of undiscovered tumuli in the remaining areas. The acquired topographic maps include a slope map, a contour line map, and a shaded relief image. An example of the topographic map covering a small part of the Maeyama-A region is illustrated in Fig. 1. The rectangles indicate the position of discovered tumuli. It is challenging to discern patterns in the distribution of tumuli based solely on visual observation of images. Because some tumuli in the topographical maps

are distinctly different from their surroundings, while others do not show any noticeable differences, making it difficult to predict the locations of undiscovered tombs using only the human eye. Specifically, the contour line map indicates the vertical distance in ground elevation with the contour interval between contour lines. In Fig. 1(a), some large discovered tumuli exhibit circular contour lines within the blue boxes, while others do not. The shaded relief image shows the shape of the terrain realistically by showing how the three-dimensional surface would be illuminated from a point light source. In Fig. 1(b), some discovered tumuli exhibit clear topographical elevations and depressions, while others are characterized by subtle terrain variations that blend seamlessly with the surrounding landscape. The slope map shows the steep or gentle level of a slope at any given position. In Fig. 1(c), some discovered tumuli reveal areas of red surrounded by green and yellow, while others do not show obvious differences from the surroundings. Each of the three types of topographic maps was verified to contain valuable information for predicting the locations of tumuli. However, the previous work annotated the surrounding region around discovered tumuli as non-tumulus samples, where some incorrect labels might have been mixed in. Besides, the input images of deep learning networks after processing had different scales in the previous works. This paper revises the annotations of tumulus and non-tumulus samples based on the recommendations of archaeological experts, and organizes them as three channels of the input image with the same scale to train the deep learning network. This paper also improved the visualization of prediction result, allowing for a clear comparison between the features identified by the deep learning network and those recognized by archaeological experts.

3 Method

As shown in Fig. 2, the main process of the proposed method is as follows. First, three different types of topographic maps are generated from 3D point cloud data of Maeyama-A region. Using the marked distribution map of discovered ancient tumuli, the corresponding ground truth annotations are made for the three topographic maps. These annotated regions are

then individually segmented into patches. The three topographic map patches corresponding to the same location are combined as three channels of a single image, which serves as the input for the deep learning network. The transfer learning technique is used to retrain a pre-trained network model to distinguish between “non-tumulus” and “tumulus” categories. After training, for any given patch image, the network firstly performs a forward propagation. If the confidence score for the “tumulus” category is higher than that for the “non-tumulus” category, the confidence score for the “tumulus” category is backpropagated through the last convolutional layer. The resulting gradient is then linearly combined with the feature maps from the last convolutional layer to produce a heatmap of the input patch image. This heatmap highlights the features that contributed most to the network’s classification of the patch as a “tumulus”. Finally, the heatmaps of patches from different locations are combined into a whole 2D heatmap that describes the importance of various features. To compare the features identified by the deep learning network with those used by human archaeologists, the color of 2D heatmap is mapped onto the 3D point cloud. The color-coded annotations on the 3D point cloud can be utilized in the following scenarios:

(1) By comparing the topographic and 3D heatmap distribution of known tumulus areas, archaeologists can simultaneously observe the important feature distribution identified by the deep learning network while examining the 3D topography. This allows them to verify whether the features they focus on when identifying tumuli coincide with those identified by the network. The common features can then be described in human-understandable language, improving the ability to communicate this knowledge to others and aiding in the training of future archaeologists.

(2) Archaeology novices can learn and master the knowledge needed for tumulus identification by combining the color-coded 3D point cloud, topographic data, and the feature descriptions provided by experienced archaeologists. Additionally, the results from the deep learning network can serve as a reference to assist beginners in their tumulus identification efforts.

(3) The automated patch-by-patch identification process serves as an initial automated screening for future tumulus excavation work. Areas with higher

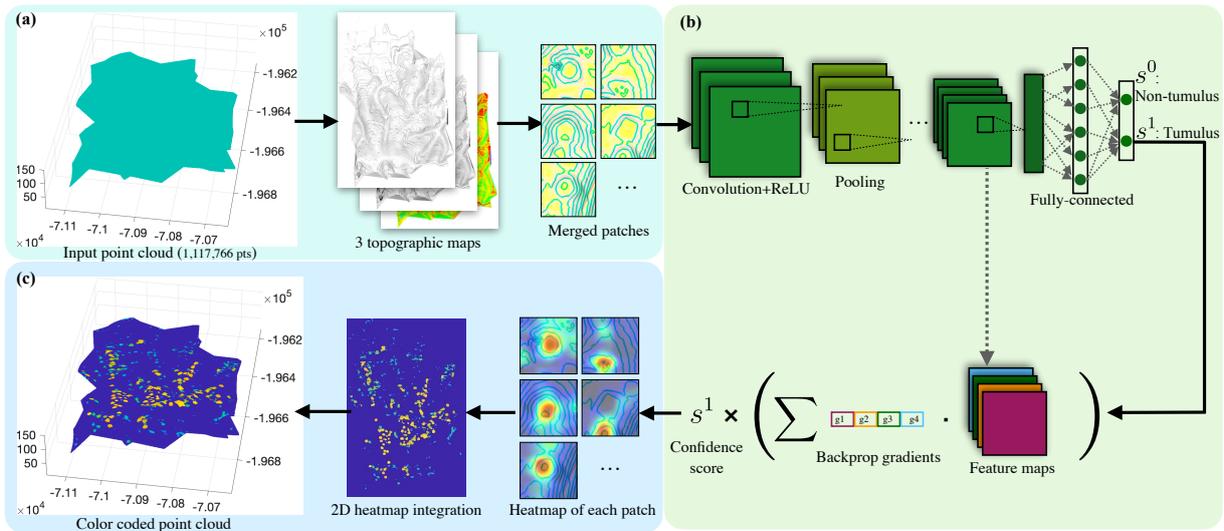


Figure 2: The processing flow of the proposed method.

heatmap values are identified as more likely to contain tumuli, allowing archaeologists to prioritize these areas for further excavation, thus saving them from manually inspecting every location.

3.1 Data Processing

The original topographic maps do not contain any location information of the discovered tumuli. To obtain the location annotation of discovered tumuli, the coordinate and size of each discovered tumuli on the image should be manually annotated to make the training data. A map with all the discovered tumuli marked in circles by archeologists is used to generate the ground-truth data. However, as shown in Fig. 3(a), this tumuli-marked map only covers a rectangular section of the Maeyama-A region, and its distance scale is different with those of the original topographic maps, as shown in Fig. 3(b). To transfer the coordinate and size of each discovered tumuli from the tumuli-marked map to the original topographic map, the image of the tumuli-marked map is resized to make them have consistent distance scales. Specifically, the image file of tumuli-marked map has a resolution of 2480×3507 and its distance scale indicates that 50 meters is represented by 325 pixels. The three original topographic maps shares the same image size of 11616×6576 and the same distance scale where 350 meters is represented by 2854 pixels. The tumuli-marked map is enlarged to 125.45% ($2854 \div 325 \times 50 \div 350$) of its original size to obtain

the consistent distance scale.

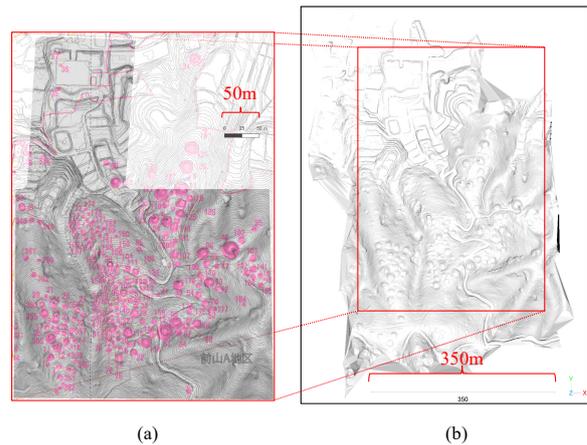


Figure 3: The scale information used for the alignment of maps to obtain ground-truth positions of discovered tumuli. (a) marked map of discovered tumuli. (b) the slope map.

Then, as shown in Fig. 4, the marked map is made semi-transparent and placed over the original topographic maps to make them overlap. Places with distinct terrain changes can be used to verify whether two images are perfectly aligned. After adjusting the relative positions, the map on the top can be changed to be different transparency to visually confirm the effects of overlap. The overlapping states are exported as images with the same resolution as the original topographic maps. Each circular region on these images is manually annotated with a bounding rectangle. The coordinate and size of the rectangle also correspond to the discovered tumuli on the original topographic

maps, and is used to cropped out the patch image only containing a single tumulus from the original topographic maps. The patch images of three topographic maps can be used independently or combined as input images for deep learning networks, for which the experimental result comparison will be discussed in the next section.

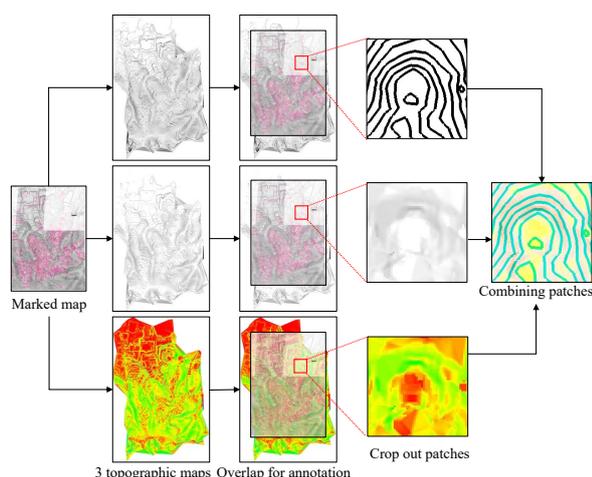


Figure 4: The image processing steps.

Although the coordinate and size of discovered tumulus patch images can be annotated according to the marking of archeologists, the non-tumulus region is not marked similarly. According to the experience of tumuli discovery by archeologists, it is highly unlikely that tumuli would be found in valleys or on slopes. It is also reasonable to choose the empty areas along the ridges as non-tumulus samples. Besides, the regions surrounding the cluster of discovered tumuli have higher possibility to be checked while discovering the tumuli. Thus, the surrounding regions has higher possibility to be non-tumuli than regions far away from discovered tumuli. Regions manually found following these rules are annotated as the position of non-tumulus. Considering the discovered tumuli have different sizes according to the actual size of the tumuli, the maximum size of all discovered tumuli regions is chosen for non-tumuli patches, and the center of the non-tumuli patch is preserved. To this end, the position and size of both tumulus and non-tumulus regions are defined and manually annotated. The annotation is utilized to crop the patch image from the three topographic maps. The patch images containing the discovered tumuli are regarded as the class of tumulus samples, and rectangular regions

labelled as “non-tumulus” are considered as the other class of non-tumuli samples. As shown in Fig. 5, tumulus samples (representing discovered tumuli) are shown by blue rectangles, while non-tumulus samples are shown in red. These samples can be cropped out separately from three distinct topographic maps by utilizing the locations of the blue and red rectangles.

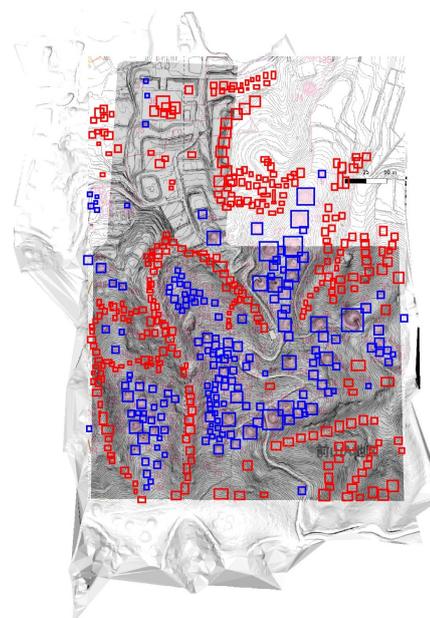


Figure 5: The ground-truth annotation for tumulus (blue rectangles) and non-tumulus (red rectangles) on the slope map.

3.2 Patch Image Prediction

The annotated data containing tumulus and non-tumulus samples can be divided into three subsets for training, validation, and testing. Deep learning architectures is trained with the training subset, simultaneously validated on the validation subset, and tested on the testing subset after training to verify the results. This data division strategy helps in selecting the optimal experimental parameter settings by comparing the experimental results in terms of the same evaluation metrics. Subsequently, the selected parameter settings are applied to train once again on all annotated data. At this point, all data is used for training without division. The generated model serves as the final prediction model. Since the aim of this study is to predict the likelihood of the existence of tumulus in the remaining locations based on the topographic

maps of already discovered locations, the final prediction model, trained with the selected optimal parameter settings, can be employed to examine the remaining locations. The confidence score obtained in each location can be outputted as the likelihood of tumulus existence, thus completing the overall prediction for the area.

Regarding the parameters that need to be selected, the first consideration is the content of input images. The three types of topographic maps may contain overlapping redundant information or exhibit unique feature representations due to different expression forms. To determine which topographic map of the three yields the best performance, the effectiveness of using three different topographic maps individually and in combination as input images can be compared. The optimal choice of input image is selected for further experiments. Besides, the choice of different deep learning models can also be optimized by evaluating the result on the same dataset.

For comparison, the pre-trained CNN models of a shallow-layer network, i.e. SqueezeNet [24], and three enhanced network models with deeper layers and more complex structures, i.e. ResNet50 [25], EfficientNet [26] and Inception-ResNet-v2 [27] are used to evaluate the performance of classifying each patch image into two classes of tumulus and non-tumulus to validate the distinguishing ability of the deep learning network architecture. An optimal model is selected according to the evaluation result for further distribution prediction. To make the pre-trained models fit the tumulus patch identification task, the input layer and the last three layers of them needs to be modified. For instance, ResNet50 is 50 layers deep and trained on more than a million images from the ImageNet database. The pre-trained model can classify images into 1000 object categories, such as keyboard, mouse, pencil, and many animals. As a result, the network has learned rich feature representations for a wide range of images. The network has an image input size of $224 \times 224 \times 3$. To make the tumulus image fit the size of input layer, all input images are resized to the same size as its input layer. Besides, the last three layers, i.e., a fully connected layer, a softmax layer, and a classification layer, are replaced by new layers that fit the tumuli classification task. In other words, the early layers are reused for image feature extraction. By freezing the parameters in the

early feature extraction layers, the training process converges much faster than training a network with randomly initialized parameters.

3.3 Heatmap for Each Patch

The output layer of the deep learning network is set to a vector with two elements, $\{s^0, s^1\}$, representing the confidence that the input patch image belongs to the “non-tumulus” and “tumulus” categories, respectively. The larger the value of confidence s^1 , the higher the probability that the input patch image belongs to the “tumulus” category. However, the area corresponding to the patch image still occupies a significant region in the original topographic map. We are more interested in knowing which specific features within this area contributed to the final decision of classifying it as a tumulus. Therefore, we seek methods to visualize feature importance, specifically Gradient-weighted Class Activation Mapping (Grad-CAM) [28]. Grad-CAM uses the gradient information flowing into the last convolutional layer of the CNN to assign importance values to each neuron for a particular decision of interest. In this context, we are only concerned with the Grad-CAM for the tumulus category. This means that when calculating the gradient information, we only take the partial derivatives of confidence s^1 with respect to the last convolutional layer. Let A^k represent the k -th feature map obtained by the last convolutional layer after training. As shown in the backpropagation gradients g_1, g_2, g_3 and g_4 in Fig. 2(b), the backpropagation gradient g_k corresponding to the k -th feature map can be expressed as follows,

$$g_k = \frac{1}{z} \sum_i \sum_j \frac{\partial s^1}{\partial A_{i,j}^k}, \quad (1)$$

where z is the number of elements in the feature map, i and j are the iterator for each elements of the feature map. By performing a linear combination of each gradient and its corresponding feature map, the output heatmap $Y_{p,q}$ of Grad-CAM can be obtained for the patch centered at (p, q) , as shown below,

$$Y_{p,q} = \sum_k g_k * \frac{A_{p,q}^k + |A_{p,q}^k|}{2}, \quad (2)$$

where the negative values of $A_{i,j}^k$ are ignored, and only the positive values are retained. The actual examples

of calculated Y are shown in the “heatmap of each patch” in Fig. 2(c), where the heatmaps are overlaid on the original patch images to achieve better visualization.

3.4 2D Heatmap Integration

Each input patch image of the deep learning network is extracted from a large, complete topographic map. Although each patch image can generate its own heatmap, a major challenge remains: how to organize these numerous heatmaps to provide a complete visual representation of the topographic map. To address this, we introduce a weighted Grad-CAM calculation method. This method adjusts the relative value of different heatmaps using the confidence value of s^1 obtained from each patch image as the weight, as shown below,

$$\begin{aligned} Y'_{p,q} &= s^1 * Y_{p,q} \\ &= s^1 \sum_k \left(\frac{A_{p,q}^k + |A_{p,q}^k|}{2z} \sum_i \sum_j \frac{\partial s^1}{\partial A_{i,j}^k} \right) \end{aligned} \quad (3)$$

By iterating over the values of p and q in $Y'_{p,q}$, we can obtain a heatmap of the adjusted important feature distribution for each patch image. In this way, the following two issues can be addressed. Firstly, comparability between patches: the $Y_{p,q}$ in Grad-CAM assigns values between 0 and 1 to represent the importance of features learned during the training process for each patch. When integrating the heatmaps of different regions into a global heatmap (with the same dimensions as the topographic map), it is necessary to account not only for the importance of features but also for the confidence score of the region being classified as the tumulus category. To achieve this, we used the confidence score s^1 for the tumulus category as the weighting factor. The weighted values were then used to create the global heatmap. Secondly, overlap between patches: If adjacent patches do not have overlapping areas, the boundary regions are excluded from being the focus of any patch. To address this, we ensured that adjacent patches had overlapping regions. In the overlapping areas, the same pixel position may be estimated multiple times in the heatmap. To reconcile these multiple estimates, the values were weighted using the tumulus confidence score. Among

the weighted estimates, the maximum value was selected as the global estimate for that pixel. To achieve this, we propose an algorithm to record the contribution distribution of features at different locations within each patch image for discrimination, as shown in Algorithm. 1. This approach also addresses the issue of fragmented heatmaps that occur when significant features appear at the edges of patches. In Algorithm. 1, the first nested loop for i and j is used to identify the maximum and minimum values along the four edges of the patch centered at (p, q) . The second loop for i and j is responsible for assigning heatmap values to the pixels in the global map that correspond to all the pixels within the patch. The condition for assigning a heatmap value is that the difference between the maximum and minimum values along the patch edges must be below the threshold γ . If the difference exceeds γ , it indicates that the most significant features are located on one of the patch edges, which reduces precision compared to when the significant features are located within the patch’s inner region. Therefore, Algorithm. 1 excludes cases of imprecise evaluation.

3.5 Colored Point Cloud

To address the issue of visualizing the integrated 2D heatmap with the 3D terrain for comparison purposes, we mapped each point in the 3D point cloud to its corresponding position on the 2D heatmap. To establish a mapping between the 2D heatmap and the 3D point cloud, we first need to ensure that the XY plane of the 3D point cloud aligns with the plane of the 2D heatmap. Next, we need to ensure that the point cloud rotates within the XY plane to match the 2D heatmap. In fact, since all three topographic maps are derived from the original 3D LiDAR data, rotations around the X, Y, or Z axes are not an issue. The remaining challenge is the scale and translation between the XY coordinates of the point cloud and the pixel positions on the 2D heatmap. Assume that any given point (x, y, z) in the point cloud corresponds to the pixel position (x', y') on the 2D heatmap. Then, the scale and translation problem can be expressed by $x = a * x' + b$ and $y = c * y' + d$. This equation can be solved for a , b , c , and d by manually annotating two corresponding pairs of points. Alternatively, annotating more than two sets of points can yield a more

Algorithm 1: Algorithm for 2D Heatmap Integration.

input : The resolution of topographic map $w \times h$. The feature maps and confidence score s^1 for each patch. The half length of patch edge e .
output : The integrated 2D heatmap of H .
parameter: The threshold to determine whether important features are distributed along the boundary γ

Initialize $H \in 0_{w \times h}$;

for $p \leftarrow e$ **to** $w - e - 1$ **do**

for $q \leftarrow e$ **to** $h - e - 1$ **do**

 Initialize $y_{max} \leftarrow 0, y_{min} \leftarrow +\infty$;

for $i \leftarrow 0$ **to** $2e - 1$ **do**

for $j \leftarrow 0$ **to** $2e - 1$ **do**

if $i == 0$ **or** $i == 2e - 1$ **or** $j == 0$ **or** $j == 2e - 1$ **then** // Find maximum & minimum values along four edges of the patch

 Calculate $Y'_{p-e+i, q-e+j}$ by Eq. (3);

$y_{max} \leftarrow \max(y_{max}, Y'_{p-e+i, q-e+j})$;

$y_{min} \leftarrow \min(y_{min}, Y'_{p-e+i, q-e+j})$;

if $y_{max} - y_{min} < \gamma$ **then** // If the most significant features are not located on any one of the patch edges

for $i \leftarrow 0$ **to** $2e - 1$ **do**

for $j \leftarrow 0$ **to** $2e - 1$ **do**

 Calculate $Y'_{p,q}$ by Eq. (3);

$H_{p,q} \leftarrow \max(H_{p,q}, Y'_{p,q})$;

accurate correlation for linear regression. Finally, the RGB values from (x', y') in the 2D heatmap are written into the corresponding position of (x, y, z) within the 3D point cloud.

4 Experimental Results

4.1 Experimental Setting

As shown in the top-left block of Fig. 2, the original point cloud data contains 1,117,766 points. The resolution of the three topographic maps, i.e. the slope map, the contour line map, and the shaded relief image, which are derived from the point cloud data, is 11616×6676 . By aligning the marked discovered tumuli map to them, there are 182 discovered tumuli and 326 non-tumuli annotated and cropped out to make 508 patch images for each of the three topographic maps. The patch images are used as ground truth data and labelled as tumulus or non-tumulus. Then, the binary classification experiment is conducted to verify whether the CNN architectures could extract distinguishable features from the topographic maps. However, the optimal model should be determined by performance evaluation. The evaluation experiment uses the same data division method, splitting all ground truth data by percentage: 70% for training, 15% for validation, and 15% for testing. All patch images are resized to the input layer size of the pre-trained model. To avoid the patches being resized to different scales, as shown in Fig. 5, the center coordinates of the ground truth annotated rectangles are used, and the largest side length of the discovered tumuli is applied as the uniform side length for each patch. In the training phase, the initial learning rate was set to 0.001. Stochastic gradient descent with a momentum optimizer was used with a minibatch size of 128, a weight decay factor of 0.0001, and momentum of 0.95. The maximum number of epochs was set to 60. After training, the testing images are considered as unknown images to verify whether the predicted label matched the ground-truth label. The matched images are considered correct prediction and the mismatched images are incorrect prediction.

4.2 Evaluation Metrics

For comparison, the prediction result is estimated using the same evaluation metrics, i.e. accuracy, F-score, precision, AUC, and prediction time. Accuracy denotes the ratio of the number of correctly labeled images over the total. In this case, it denotes the proportion of correctly predicted image numbers to the number of test images. F-score is the harmonic mean of the precision and sensitivity, where the precision is the number of correctly identified images divided by the number of images identified as positives, and the sensitivity is the number of correctly identified positives divided by the number of true positives. When calculating the evaluation metrics of accuracy, precision, and F-score, a test image is judged as positive when the confidence score is above 0.5. However, by considering varying confidence scores as judging thresholds, the receiver operating characteristic curve (ROC) is drawn to evaluate the diagnostic ability of this binary classifier system. This result indicates the performance of a classification model at varying prediction thresholds. The area under curve (AUC) denotes the area under the corresponding ROC curves. The larger AUC value indicates a better ROC curve. The metric of prediction time is the time consumed to predict the entire testing dataset, measured in a desktop PC with a Core i7-13700k CPU, Geforce RTX 3080 GPU, and the 64GB memory.

4.3 Performance Evaluation

To compare the performance of different neural network models, four different pre-trained models were employed to train four classifiers for comparison. As shown in Fig. 6, the loss values for the four models gradually converge during the training iterations. However, the convergence value for SqueezeNet is significantly higher than that of the other three models. This likely indicates that SqueezeNet is not as effective at extracting features with the same discriminative ability as the other models. The prediction results, as shown in Table. 1, indicate that EfficientNet-b0 performed the best in the first three metrics of accuracy, F-score, and precision, albeit with marginal differences compared to the other three models. This confirms that all models involved in the experiment were capable of extracting sufficient features to distinguish between tumulus and non-tumulus categories,

with EfficientNet-b0 slightly outperforming the others. However, in terms of the metric of prediction time, there was a gradual increase in the range of SqueezeNet, ResNet-50, and Inception-ResNet-v2, as the depth of the neural network models increased, aligning with the expectation that deeper networks require more computational resources for inference. However, EfficientNet-b0 was proposed to adjust the model scaling and balance network depth, width, and resolution, which is proved effective in this experiment and EfficientNet-b0 is only slightly slower than SqueezeNet. By analyzing the AUC metric from the ROC curves shown in Fig. 7, ResNet-50 slightly outperformed the other models. EfficientNet-b0 and Inception-ResNet-v2 produced similar results, while SqueezeNet performed the worst. This indicates that when using different threshold values to determine the confidence score s^1 , ResNet-50 performs the best. However, in Table. 1, where accuracy is compared, EfficientNet-b0 is the best. This suggests that when choosing the larger value between confidence scores s^0 and s^1 as the prediction result, EfficientNet-b0 performs the best.

Table 1: Evaluation of prediction results using the different pre-trained models. (Best results are highlighted in bold)

Input image	Depth	Accuracy	F-score	Precision	Prediction time (s)
SqueezeNet	18	0.7910	0.8466	0.8069	2.1977
ResNet-50	50	0.8916	0.9189	0.8914	11.0300
EfficientNet-b0	82	0.9474	0.9608	0.9245	3.3347
Inception-ResNet-v2	164	0.8877	0.9143	0.9045	21.0155

4.4 Visualization of Prediction Result

EfficientNet-b0 is selected as the optimal model to predict patches and make the 2D heatmap integration (as shown in Fig. 8) for the entire Maeyama-A region according to Alg. 1. To address the request for quantitative evaluation, the overlap ratio between the focus areas of the heatmap of Fig. 8(a) and the regions of known tumuli is calculated. In this process, a threshold for determining overlaps emerged as an important factor. Specifically, a heatmap value exceeding a certain threshold is considered existence of a tumulus. When the heatmap at the center coordinates of a known tumulus is classified as existence of a tumulus,

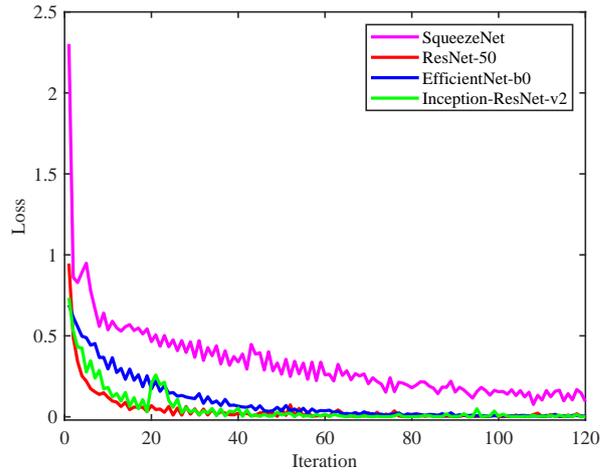


Figure 6: Comparison of the loss curves using different pre-trained models for identification experiment.

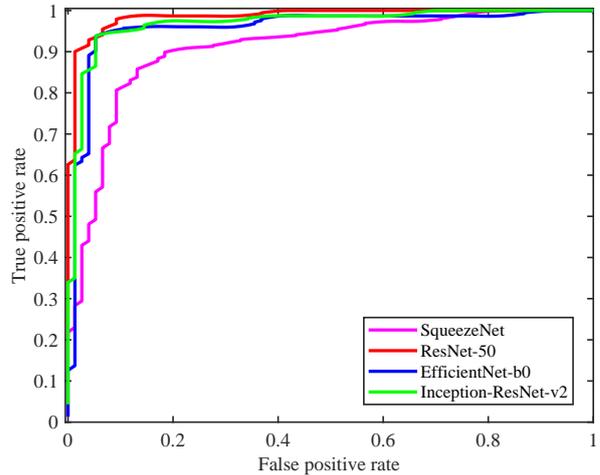


Figure 7: Comparison of the ROC curves using different pre-trained models for identification experiment.

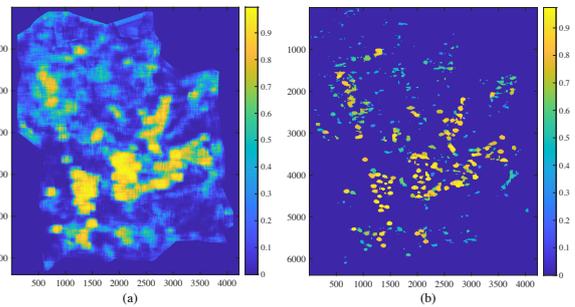


Figure 8: Result of 2D heatmap integration. (a) confidence score of s^1 . (b) weighted Grad-CAM in Eq. (3).

it is defined as an overlap. To analyze this, the threshold is incrementally increased from a minimum value of 0 up to 1, calculating the overlap rate corresponding to each threshold. This generates a curve showing how the overlap rate changes with the threshold. As shown in Fig. 9, when the threshold is set at 0.25, all known tumuli are covered by the regions classified as existence of a tumulus on the heatmap. However, this setting likely includes a significant number of false alarms. When the threshold is set at 0.5, 80.77% of the known tumuli are covered by the regions classified as existence of a tumulus.

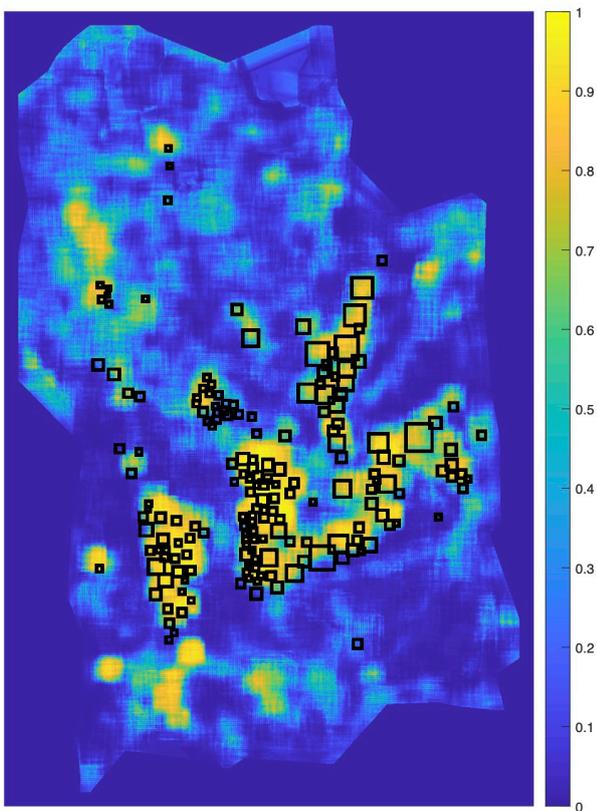


Figure 9: Overlap between the heatmap of confidence score and the groundtruth of known tumuli.

By mapping the heatmap to the point cloud data, we create a color-coded point cloud. This allows for simultaneous observation of the important feature distribution identified by the deep learning network and examination of the 3D topography. The color-coded point cloud has over 1 million points and can be examined with point cloud processing software. By rotating and dragging, users can observe the terrain of a specific location from multiple perspectives and see

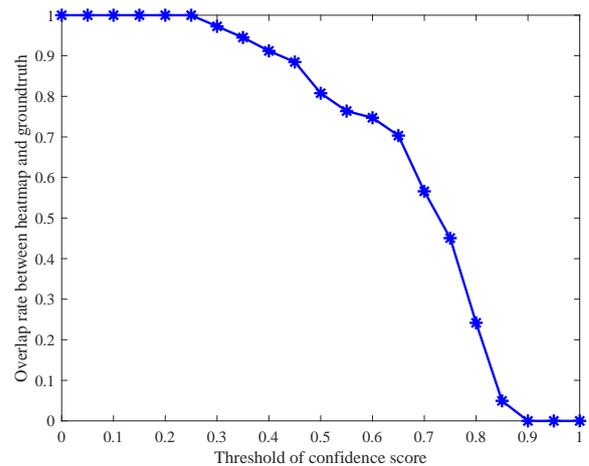


Figure 10: The overlap rate between the heatmap and the groundtruth changes with the threshold.

the features that deep learning algorithms highlight as important. Four screenshots of local parts in the point cloud are shown in Fig. 11. Regions with high numerical values, as shown by yellow areas in Fig. 11 can be translated into archaeological descriptive language to facilitate the summary of these topographic features.

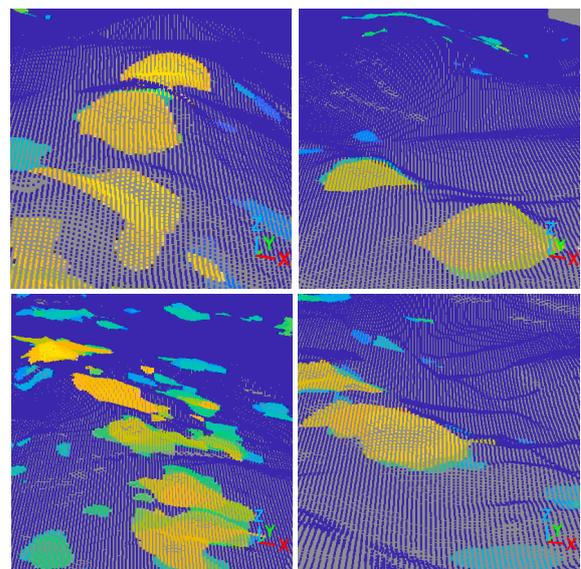


Figure 11: Four screenshots of local parts in the color-coded point cloud.

5 Conclusion

This paper proposes a novel approach to analyze the topographic point cloud data by annotating three derived topographic maps, cropping out patches and merging them to predict the likelihood of tumulus presence within each patch, and mapping the 2D heatmaps back to a color-coded point cloud. In the future, we plan to select some typical areas identified as tumulus by deep learning networks and present them to archaeologists for professional evaluation. These areas can be prioritized for future excavation efforts, thereby enhancing efficiency.

However, it remains uncertain whether the features extracted by deep learning represent the topographic features of the tumuli or the topographic features of the surrounding terrain. On one hand, since this study utilizes UAV-LiDAR data collected over the entire Maeyama-A region, it is possible that when generating topographic maps on a large scale, the internal shapes of the tumuli were overlooked. As a result, the model may not be capturing their internal features and is instead relying on surrounding topographic features for prediction. On the other hand, when visualizing the heatmap results for detected tumuli, we observe that in some cases, the heatmap's focus aligns with the center of the tumuli. However, a comprehensive comparison and statistical analysis of all known tumuli have not yet been conducted in this study. For undetected tumuli, there is currently no way to determine whether the extracted features correspond to the topographic features of the tumuli themselves or those of the surrounding terrain. Therefore, as part of future work, further investigation and comparisons will be necessary to clarify whether deep learning is primarily identifying the topographic features of the tumuli or the topographic features of the surrounding terrain.

6 Acknowledgements

The basic concept of CNN deep learning with the topographic maps part in our method has already been presented at the 15th International Conference On Frontiers Of Information Technology, Applications And Tools (FITAT2023) [23], and we extended the concept to precise prediction and 3D visualizations. Specifically, we have revised the ground truth data

based on the insights of archaeologists regarding non-tumulus areas and have used the latest CNN model for predictions and visualizations. A part of this work was supported by JSPS KAKENHI Grant Number JP22K00998 and JP23K169370.

References

- [1] “Wakayama Prefecture Kii-fudoki-no-oka Museum of Archaeology and Folklore,” accessed on April 26th, 2024. [Online]. Available: <https://www.kiifudoki.wakayama-c.ed.jp>
- [2] L. Li, Y. Li, X. Chen, and D. Sun, “A prediction study on archaeological sites based on geographical variables and logistic regression—a case study of the Neolithic era and the bronze age of Xiangyang,” *Sustainability*, vol. 14, no. 23, p. 15675, 2022.
- [3] L. Liu, F. Liu, W. Zhang, and Z. Xu, “Spatial distribution and evolution of ancient settlements from the Neolithic to the bronze age in Dalian Area, China,” *Frontiers in Earth Science*, vol. 10, p. 917520, 2022.
- [4] H. Wu, X. Wang, X. Wang, L. Zhang, and S. Dong, “Predictive modeling for neolithic settlements in the Lingnan Region, South China,” *Journal of Archaeological Science: Reports*, vol. 49, p. 103992, 2023.
- [5] A. F. Nsanziyera, H. Lechgar, S. Fal, M. Maanan, O. Saddiqi, A. Oujaa, and H. Rhinane, “Remote-sensing data-based Archaeological Predictive Model (APM) for archaeological site mapping in desert area, South Morocco,” *Comptes Rendus Geoscience*, vol. 350, no. 6, pp. 319–330, 2018, the West African Craton and its margins. A tribute to Professor André Michard for his jubilee of works in Morocco.
- [6] I. Guechi, H. Gherraz, A. Korichi, and D. Alkama, “Predicting archaeological sites locations in desert areas, using GIS-AHP-GeoTOPSIS model: Southwestern Algeria, Bechar,” *Archaeologies*, vol. 19, no. 2, pp. 471–499, 2023.

- [7] G. Abou Diwan, “Gis-based comparative archaeological predictive models: a first application to iron age sites in the Bekaa (Lebanon),” *Mediterranean Archaeology and Archaeometry*, vol. 20, no. 2, pp. 143–143, 2020.
- [8] S. H. Bickler, “Machine learning arrives in archaeology,” *Advances in Archaeological Practice*, vol. 9, no. 2, pp. 186–191, 2021.
- [9] S. Oonk and J. Spijker, “A supervised machine-learning approach towards geochemical predictive modelling in archaeology,” *Journal of archaeological science*, vol. 59, pp. 80–88, 2015.
- [10] A. Resler, R. Yeshurun, F. Natalio, and R. Giryes, “A deep-learning model for predictive archaeology and archaeological community detection,” *Humanities and Social Sciences Communications*, vol. 8, no. 1, 2021.
- [11] H. A. Orengo, F. C. Conesa, A. Garcia-Molsosa, A. Lobo, A. S. Green, M. Madella, and C. A. Petrie, “Automated detection of archaeological mounds using machine-learning classification of multisensor and multitemporal satellite data,” *Proceedings of the National Academy of Sciences*, vol. 117, no. 31, pp. 18 240–18 250, 2020.
- [12] O. D. Trier, D. C. Cowley, and A. U. Walde-land, “Using deep neural networks on airborne laser scanning data: Results from a case study of semi-automatic mapping of archaeological topography on Arran, Scotland,” *Archaeological Prospection*, vol. 26, no. 2, pp. 165–175, 2019.
- [13] I. Berganzo-Besga, H. A. Orengo, F. Lumbreras, M. Carrero-Pazos, J. Fonte, and B. Vilas-Estévez, “Hybrid msvm-based deep learning and multitemporal sentinel 2-based machine learning algorithm detects near 10k archaeological tumuli in North-Western Iberia,” *Remote Sensing*, vol. 13, no. 20, p. 4181, 2021.
- [14] Y. Y. chen, “Dog and cat classification with deep residual network,” in *Proceedings of the 2020 European Symposium on Software Engineering*, 2020, pp. 137–141.
- [15] R. Kumar, M. Sharma, K. Dhawale, and G. Singal, “Identification of dog breeds using deep learning,” in *2019 IEEE 9th International Conference on Advanced Computing (IACC)*. IEEE, 2019, pp. 193–198.
- [16] I.-H. Wang, K.-C. Lee, S.-L. Chang *et al.*, “Images classification of dogs and cats using fine-tuned VGG models,” in *2020 IEEE Eurasia Conference on IOT, Communication and Engineering (ECICE)*. IEEE, 2020, pp. 230–233.
- [17] H.-M. Jeon, V. D. Nguyen, and J. W. Jeon, “Pedestrian detection based on deep learning,” in *IECON 2019-45th Annual Conference of the IEEE Industrial Electronics Society*, vol. 1. IEEE, 2019, pp. 144–151.
- [18] G. Li, Y. Yang, and X. Qu, “Deep learning approaches on pedestrian detection in hazy weather,” *IEEE Transactions on Industrial Electronics*, vol. 67, no. 10, pp. 8889–8899, 2019.
- [19] L. Chen, S. Lin, X. Lu, D. Cao, H. Wu, C. Guo, C. Liu, and F.-Y. Wang, “Deep neural network based vehicle and pedestrian detection for autonomous driving: A survey,” *IEEE Transactions on Intelligent Transportation Systems*, vol. 22, no. 6, pp. 3234–3246, 2021.
- [20] D.-N. Zou, S.-H. Zhang, T.-J. Mu, and M. Zhang, “A new dataset of dog breed images and a benchmark for finegrained classification,” *Computational Visual Media*, vol. 6, pp. 477–487, 2020.
- [21] Z. Luo, F. Branchaud-Charron, C. Lemaire, J. Konrad, S. Li, A. Mishra, A. Achkar, J. Eichel, and P.-M. Jodoin, “MIO-TCD: A new benchmark dataset for vehicle classification and localization,” *IEEE Transactions on Image Processing*, vol. 27, no. 10, pp. 5129–5141, 2018.
- [22] M. A. Butt and F. Riaz, “CARL-D: A vision benchmark suite and large scale dataset for vehicle detection and scene segmentation,” *Signal Processing: Image Communication*, vol. 104, p. 116667, 2022.
- [23] M. You, K. Konno, A. Noguchi, R. Nakamura, and Y. Takata, “A study on predicting the distribution of Iwasesenzuka-Kofun with CNN,” in

15th International Conference On Frontiers Of Information Technology, Applications And Tools (FITAT2023), 2023.

- [24] B. Koonce and B. Koonce, “Squeezenet,” *Convolutional Neural Networks with Swift for Tensorflow: Image Recognition and Dataset Categorization*, pp. 73–85, 2021.
- [25] K. He, X. Zhang, S. Ren, and J. Sun, “Deep residual learning for image recognition,” in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2016, pp. 770–778.
- [26] M. Tan and Q. Le, “Efficientnet: Rethinking model scaling for Convolutional Neural Networks,” in *International conference on machine learning*. PMLR, 2019, pp. 6105–6114.
- [27] C. Szegedy, S. Ioffe, V. Vanhoucke, and A. Alemi, “Inception-v4, inception-resnet and the impact of residual connections on learning,” in *Proceedings of the AAAI conference on artificial intelligence*, vol. 31, no. 1, 2017.
- [28] R. R. Selvaraju, M. Cogswell, A. Das, R. Vedantam, D. Parikh, and D. Batra, “Grad-cam: Visual explanations from deep networks via gradient-based localization,” in *Proceedings of the IEEE international conference on computer vision*, 2017, pp. 618–626.

Mengbo You



is an associate professor in the Faculty of Science and Engineering at Iwate University. He received his B.S. degree in Computer Science from Northwest A&F University in 2012, and his M.S. and doctor of engineering degrees from Iwate University in 2015 and 2018, respectively. His research interests include machine learning, image processing, object detection,

and deep learning. He is a member of the Society for Art and Science and the Institute of Image Electronics Engineers of Japan.

Kouichi Konno



is a professor of Faculty of Science and Engineering at Iwate University. He received a BS in Information Science in 1985 from the University of Tsukuba. He earned his Dr.Eng. in precision machinery engineering from the University of Tokyo in 1996. He joined the solid modeling project at RICOH from 1985 to 1999, and the XVL project at Lattice Technology in 2000. He worked on an associate professor of Faculty of Engineering at Iwate University from 2001 to 2009. He has written a book Introduction to 3D shape processing. His research interests include 3D modeling, 3D surface data compression, archaeological relics restoration. He is a member of The Society for Art and Science, The Institute of Image Information and Television Engineers, Japan Association for Archaeoinformatics, Information Processing Society of Japan, and EuroGraphics.

Atsushi Noguchi



was born in 1971. Since 2023, he has been serving as a special commissioned Associate Professor at the Research Centre for Next Generation Archaeological Studies of the Komatsu University. He is also a visiting researcher at the National Institute of Advanced Industrial Science and Technology (AIST). His work focuses on archaeological research in Japan, South and West Asia, as well as international cooperation in the protection of cultural heritage.

Ryosuke Nakamura



received the Ph.D. degree in planetary sciences from Kobe University, Kobe, Japan, in 1996.,He worked with Japan Aerospace Exploration Agency (JAXA) to develop space-borne sensors and the ground-based data analysis system until 2005. He is currently with AIRC, AIST working on AI system to handle rapidly growing geoinformation, such as satellite imagery and lidar point cloud, to meet scientific and social demand. His research interests include satellite remote sensing, virtual 4-D models and their applications for smart city.

Yuichi Takata



was born in 1983. He currently serves as a Senior Researcher at the Nara National Research Institute for Cultural Properties. His research interests focus on stone material production and transportation in pre-modern periods. He is also interested in methodological research on how information technologies, such as databases, can be effectively applied to archaeology and historical studies to serve as useful information infrastructure.