

# WALL STONE EXTRACTION BASED ON STACKED CONDITIONAL GAN AND MULTISCALE IMAGE SEGMENTATION

M. Sakamoto<sup>1,\*</sup>, T. Shinohara<sup>1</sup>, Y. Li<sup>1</sup>, T. Satoh<sup>1</sup>

<sup>1</sup> PASCO CORPORATION, 4-9-6 Aobadai, Meguro-ku, Tokyo 153-0042, Japan - (moittu9191, taarkh6651, yionr\_3951, tuost7017)@pasco.co.jp

**KEY WORDS:** Wall Stone Extraction, Stacked Conditional GAN, Multiscale Image Segmentation, Edge Detection, Convex Hull Fitness, pix2pix, Object-based Image Analysis

## ABSTRACT:

The historical castles (castellated walls), which are cultural heritages in Japan, require regular maintenance, and it is necessary to record the arrangement of individual wall stones in the maintenance work. Recently, image processing techniques are practiced to optimize maintenance and management of the infrastructure assets. In the previous study, we proposed an automatic method for efficiently extracting individual wall stone polygons by improved multiscale image segmentation technique. However, the problem has remained that wall stone polygons could not be extracted properly when there were no clear gaps or boundaries between stones. To address this problem, we improved the multiscale image segmentation technique used in our previous studies. The first improvement is that in the region growing process, selecting the best combination of a plurality of objects instead of two. The second improvement is the modification of the shape criterion to be used. Besides, we proposed three-stage Stacked cGAN for wall stone edge detection that enables us to complement areas with weak or broken boundaries of stone edges. This approach is composed of a coarse-to-fine based image-to-edges translation network. The edge images derived from this method are used as the additional channel in multiscale image segmentation with a higher weight compared to the other RGB channels. It was confirmed that the separation performance of individual wall stone polygons was improved by the proposed method. Furthermore, the proposed method is highly effective to reduce the difficulty in setting of the scale parameter, which is usually sensitive to segmentation results and requires trial and error.

## 1. INTRODUCTION

There are hundreds of historical castles (castellated walls) in Japan and some of which are designated as national treasures and important cultural properties. Castellated walls are positioned as tangible cultural heritage, which requires regular maintenance to preserve their original state (Fig. 1). For the demolition and repair work of the castellated wall, it is necessary to identify the individual stones constituting the wall in advance, and is expected that the workload can be significantly reduced by automated processing.



Figure 1. Example of a typical castellated wall

The following methods have been conventionally used in documentation regarding cultural heritage. One approach is to image the front view of the castellated walls using a laser scanner (Křemen et al., 2011; Vacca et al., 2012). However, it is difficult to identify individual wall stones after dismantling by this method. Another approach is to attach integrated circuits (IC) tags for identifying the individual wall stones before dismantling the castellated walls for restoration (Ryu et al., 2014). However, it is quite burdensome to attach IC tags on each stone and to capture the images individually. As a study similar to our theme, there is one to detect bricks from masonry walls (Ibrahim et al., 2019). This approach combines U-Net based brick seed localization and the Watershed algorithm for accurate instance segmentation of bricks. Though, the processing result almost depends on the extraction state of the seed region by U-Net.

In our previous study, we proposed a method for efficiently extracting individual wall stone polygons by improved multiscale image segmentation technique (Sakamoto et al., 2018). We focused on the fact that many wall stones have a convex hull shape, and we introduced the new shape criterion called *convex hull fitness* and showed that it is possible to improve the performance for the extraction of wall stone polygons.

On the other hand, even when this method was applied, the problem has still remained that wall stone polygons could not be extracted properly when clear gaps did not exist between stones or the texture between adjacent stones was very similar (shown as the red circle in Fig. 2).

In this study, we introduce the Conditional GAN (Generative Adversarial Networks), a kind of deep learning technique, to

\* Corresponding author

complement the unclear stone edges in the target image. By using these extracted edge images supplementarily, we confirmed that the separation performance of individual wall stone polygons was improved and at the same time, reduction of difficulty in setting scale parameters in image segmentation was achieved.

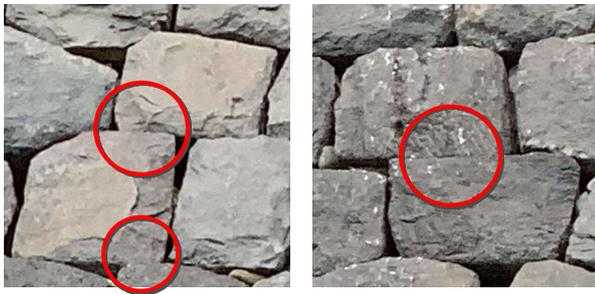


Figure 2. Examples of wall stones with unclear boundaries

## 2. PROPOSED METHOD

### 2.1 Extraction of Wall Stone Polygons by Multiscale Image Segmentation

The multiscale image segmentation is a kind of region growing approach of evaluating and merging image regions in the units referred to as objects (Baatz and Schäpe, 2000; Chen et al., 2005; Esch et al., 2008; Li et al., 2009).

Each region (object) starts from one pixel at the initial stage. The decision to merge the neighboring regions is determined based on evaluating the changes in heterogeneity including both the spectral component and shape component in objects. This relationship is formulated according to the following equation.

$$F = w_{color} \cdot F_{color} + w_{shape} \cdot F_{sha} \quad (1)$$

$$0 \leq w_{color}, w_{shape} \leq 1 \text{ and } w_{color} + w_{shape} = 1$$

where  $F_{color}$  = heterogeneity in spectral component  
 $F_{shape}$  = heterogeneity in shape component  
 $w_{color}$  = weight for  $F_{color}$   
 $w_{shape}$  = weight for  $F_{shape}$

The degree of object merge is regulated by a parameter called the scale parameter ( $SP$ ), which also affects the size of the derived objects indirectly. The merge process of neighboring regions is performed when the evaluation value  $F$  does not exceed the square value of  $SP$ .  $SP$  starts from a small value and then all the possible merging of the objects is performed. If no objects can be processed, the  $SP$  is incremented. This process is applied repeatedly until the  $SP$  reaches the predefined value.

$F_{color}$  is defined using the average and/or standard deviation calculated from each color component (channel) of pixels within the objects.

$$F_{color} = \sum_{i=1}^N w_{c_i} \left\{ n_r \cdot \sigma_{r,i} - \sum_{s \in \{p\}} n_s \cdot \sigma_{s,i} \right\} \quad (2)$$

$$0 \leq w_{c_i} \leq 1, \quad i = 1, \dots, N \text{ and } \sum_{i=1}^N w_{c_i} = 1$$

where  $p$  = an object before the merge  
 $r$  = an object after the merge

$\{p\}$  = a set of objects neighboring to object  $p$   
 $s$  = subset of  $\{p\}$   
 $n$  = number of pixels within an object  
 $\sigma_i$  = average and/or standard deviation of the  $i$ -th channel in an object  
 $w_{c_i}$  = weight for the  $i$ -th channel  
 $N$  = number of channels

$F_{shape}$  is calculated using several types of shape criteria that regulate geometric properties for the merged objects.

$$F_{shape} = \sum_{i=1}^K w_{s_i} \cdot f_i \quad (3)$$

$$0 \leq w_{s_i} \leq 1, \quad i = 1, \dots, K \text{ and } \sum_{i=1}^K w_{s_i} = 1$$

where  $f_i$  = heterogeneity in the  $i$ -th shape criterion  
 $w_{s_i}$  = weight for the  $i$ -th shape criterion  
 $K$  = number of shape criteria

This method has a very high affinity for the application to wall stone polygons extraction because of the existence of gaps between stones, differences in texture of individual stones, and geometric stability of each stone polygons. In addition, by adding a restriction on the size of the generated object in processing based on the prior knowledge of the dimensions of the stones, an excessive merging of the objects can be suppressed.

On the other hand, there is a problem that adjustment of the upper limit for the  $SP$  requires trial and error. This is because the region of a stone is divided into a plurality of polygons when the  $SP$  is small, and a plurality of stone regions are merged into one polygon when the scale parameter is large.

Based on our previous study (Sakamoto et al., 2018), it has been confirmed that the *compactness* ( $f_{cmpct}$ ) and the *convex hull fitness* ( $f_{cnv}$ ) are both effective as shape criteria, which are defined as follows.

$$f_{cmpct} = n_r \cdot \frac{l_r}{\sqrt{n_r}} - \sum_{s \in \{p\}} n_s \cdot \frac{l_s}{\sqrt{n_s}} \quad (4)$$

$$f_{cnv} = n_r \cdot \frac{l_r}{\sqrt{n_r}} \left\{ t \cdot \left( 1 - \frac{n_r}{c_r} \right) + (1-t) \cdot \left| 1 - \frac{l_r}{m_r} \right| \right\} \quad (5)$$

$$- \sum_{s \in \{p\}} \left\{ n_s \cdot \frac{l_s}{\sqrt{n_s}} \left\{ t \cdot \left( 1 - \frac{n_s}{c_s} \right) + (1-t) \cdot \left| 1 - \frac{l_s}{m_s} \right| \right\} \right\}$$

where  $l$  = boundary length of an object  
 $c$  = number of pixels in the convex hull of an object  
 $m$  = boundary length of the convex hull of an object  
 $t$  = adjustment parameter in the range of  $0 \leq t \leq 1$

The *convex hull fitness* is originally proposed in our previous work. Different from the formulas in all existing studies by various researchers, in this study, we have extended that the number of objects to be merged at one time is not limited to two. That is, in the region growing process, the best combination is selected from all possible combinations. Besides, the equation of the definition for *convex hull fitness* is modified to be more ideal index.

From our previous knowledge (Sakamoto et al., 2018), the processing parameters are adjusted as follows.

(1) Multi-stage parameter setting

The setting of the  $SP$  is divided into several stages, and the weights for the spectral evaluation index and the shape evaluation index are sequentially changed.

(2) The balance between spectral index and shape index

At the stage where the  $SP$  is small, the weight for the spectral index (color and texture information) is increased, and the setting is made so that the weight for the shape index (shape stability) gradually increases as the  $SP$  increases.

(3) Balance in shape criteria

In the initial stage, the weights for  $f_{cmap}$  and  $f_{cvx}$  are made approximately equal, and the weight of  $f_{cvx}$  is increased as the  $SP$  increases.

The processing result based on only the image segmentation is used as a baseline for evaluating the effect of newly introducing the edge images generated by the proposed Conditional GAN approach.

2.2 Edge Extraction of Wall Stones by Conditional GAN

**2.2.1 Overview:** Here we describe a methodology for the edge extraction of wall stones and its complementary usage in image segmentation using one of the Generative Adversarial Networks (GANs) (Goodfellow et al., 2014) technique. GANs learn a representation of images that tries to determine whether the generated output images are real or fake, while simultaneously training a generative model to minimize the loss of these images. Isola et al. (2017) proposed pix2pix, a kind of conditional GAN (cGAN) framework (Mirza et al., 2014), for image translation. They introduced a “U-net” based architecture (Ronneberger et al., 2015) as a generator and a convolutional classifier as a discriminator. We propose a diverted network architecture based on cGAN that enables to utilize edge detection of wall stones with unclear edge gaps or similar textures in boundaries. Therefore, our task is positioned as the “image-to-edges” translation problem.

**2.2.2 Stacked cGAN:** First, we propose three-stage Stacked cGAN for wall stone edge detection that enables an end-to-end learning architecture. Our approach is composed of a coarse-to-fine based image-to-edges translation network that consists of three stages: 1) thick edge generator, 2) refiner (generator for refinement) and 3) thin edge generator (Fig. 3). All stages follow an adversarial model, i.e. each stage consists of a generator and discriminator mechanism.

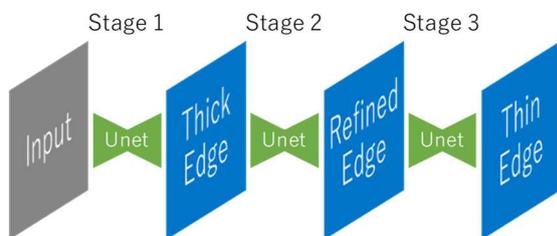


Figure 3. Conceptual diagram of proposed flow

First, the thick edge generator produces a coarse edge from the input wall stone image. Next, a refinement network modifies to fill in missing regions of coarse edges by generating more thick edges. Finally, thin edge generator sharpens reinforced coarse edges. All models in each stage are trainable and differentiable.

The real images (ground truth) in all stages are rasterized images having different thickness generated from the same vectorized edge data.

**2.2.3 Network architecture:** The proposed network architectures for each stage are consisting of the U-net based generator and Convolutional Neural Networks (CNN) based discriminator. Fig. 4 shows our generator which predicts the edge image from the input data. Original U-net is simply an encoder-decoder framework with skip connections between the encoding and decoding layers. Our U-net based generator is modified to reduce the number of pooling layers while keeping the number of convolutional layers. Original U-net makes the smallest feature maps resized into 1/32 from the size of the input image, but our generator makes the smallest one resized into 1/8. To extract features for edge detection from small objects, feature maps with high resolution are very important to preserve vanishing edge information. On the other hand, our discriminator is a very simple convolutional image classifier with a pair of images, i.e. “input image (given to generator) and generated image (derived from generator)” or “input image (given to generator) and real image (true edge image)”.

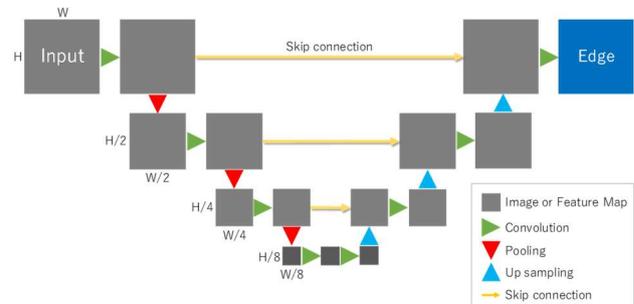


Figure 4. Diagram of the proposed network

**2.2.4 Loss functions and optimization:** Here we explain the loss functions for each step. The objective of the cGAN is finding the generator which minimizes the accuracy of real or fake predictions and also is finding the discriminator which maximizes those same predictions. The loss function of a cGAN ( $L_{cGAN}^i$ ) for each layer can be expressed as:

$$L_{cGAN}^i(D^i, G^i) = \mathbb{E}_{x^i, y^i} [\log D^i(x^i, y^i)] + \mathbb{E}_{x^i, z} [\log(1 - D^i(x^i, G(x^i, z)))] \quad (6)$$

where  $i$  = index of stage  
 $G^i$  = generator in stage  $i$   
 $x$  = input data  
 $y$  = ground truth  
 $z$  = input noise

The input noise feeds in a dropout layer.  $G^i$  tries to minimize this objective against an adversarial  $D^i$  that tries to maximize it. Additionally, our generator is expected to not only deceive the discriminator but also to be near the distance between ground truth and output in a color space. We used the loss function as the L1 distance ( $L_{L1}$ ):

$$L_{L1}^i(G^i) = \mathbb{E}_{x^i, y^i, z} [\|y^i - G^i(x^i, z)\|_1] \quad (7)$$

Our final loss function for Stacked cGAN is defined as follows.

$$\arg \min_{G^i} \max_{D^i} L_{cGAN}^i(D^i, G^i) + \lambda^i L_{L1}^i(G^i) \quad (8)$$

where  $\lambda^i$  = the weight for generator in the stage  $i$ .

We use minibatch SGD and apply the Adam solver, with a learning rate of 0.0002 and momentum parameters  $\beta_1 = 0.5$ ,  $\beta_2 = 0.999$ .

**2.2.5 Utilization of edge image:** The edge images derived from proposed Stacked cGAN are used as fourth channel of processed wall stone images in baseline method (multiscale image segmentation). The channel allocated for the edge is processed with a higher weight compared to the other RGB channels.

### 3. EXPERIMENTAL RESULTS AND DISCUSSION

#### 3.1 Training Data

The training image of wall stone edges applied to each proposed network model was obtained if the stones were different, regardless of whether there was a clear gap between the stones. The edge data were obtained in vector format and were converted to raster images by changing the line thickness as necessary.

For the training and the inference, we implemented a model, which was extended based on the network of pix2pix, to suit the processing of our proposed method. Three wall stone images and the corresponding wall stone edge images were used as training data, and another two wall stone images were used for the test.

Here, the number of images used for training is relatively small due to the following two reasons. Firstly, when a certain castellated wall is targeted, there are not so many variations on the edge of the wall stones to be extracted. Secondly, we confirmed that even if the number of images used for training was increased, there was almost no difference in the stability of model generation and the performance of wall stone extraction in cross-validation. Thus, due to utilizing the proposed method in practical situations such as wall stone extraction in different types of castellated walls, there is an advantage to prepare considerably less training data.

#### 3.2 Network Model

For comparison, we set the following three network models. Here the wall stone image to be learned is denoted by  $Img$ , the training edge image having a thickness of  $w$  pixels is denoted by  $Edge_{real}w$ , and the edge image generated so as to approach  $Edge_{real}w$  is denoted by  $Edge_{fake}w$ . Also, a flow for generating a fake image ( $fake$ ) so as to approach the conversion from the input image ( $input$ ) to the true image ( $real$ ) will be described as " $input \mapsto \langle real \rangle \mapsto fake$ ".

(a) *Type 1:*

A model that uses only stage 1 and converts " $Img \mapsto \langle Edge_{real}10 \rangle \mapsto Edge_{fake}10$ ".

(b) *Type 2:*

A model that converts " $Img \mapsto \langle Edge_{real}40 \rangle \mapsto Edge_{fake}40$ " in stage 1, " $Edge_{fake}40 \mapsto \langle Edge_{real}40 \rangle \mapsto Edge_{fake}40$ " in stage 2, and " $Edge_{fake}40 \mapsto \langle Edge_{real}10 \rangle \mapsto Edge_{fake}10$ " in stage 3.

(c) *Type 3:*

A model in which stage 2 is replaced with " $Edge_{fake}40 \mapsto \langle Edge_{real}80 \rangle \mapsto Edge_{fake}80$ " and stage 3 is replaced with " $Edge_{fake}80 \mapsto \langle Edge_{real}10 \rangle \mapsto Edge_{fake}10$ " in model *Type 2*.

#### 3.3 Processing Results by Baseline Method

Fig. 5 shows the processing results when only a pure image segmentation processing (baseline method) is used. When only a spectral component is used as the evaluation index, individual wall stone cannot be satisfactorily extracted (Fig. 5(b)). As shown in Fig. 5 (c) and (d), when the shape component is added as the evaluation index, it can be observed that the performance of wall stone polygon extraction is improved. However, some polygons have been generated across the stone boundaries or have less merging.

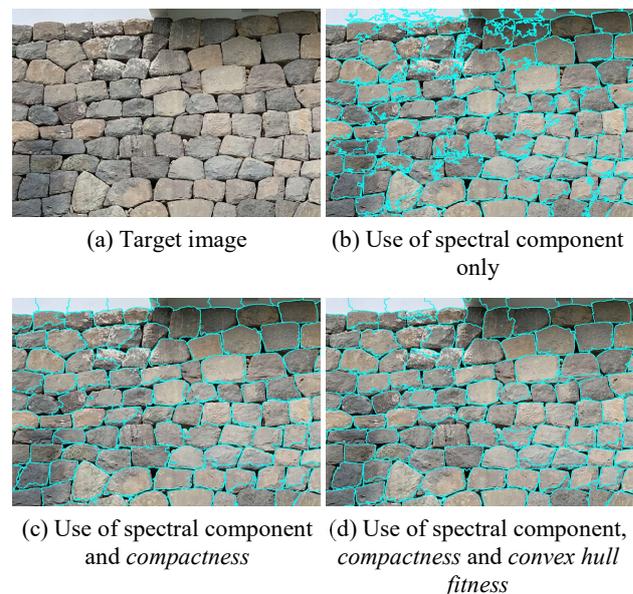


Figure 5. Image segmentation results

#### 3.4 Edge Generation Result

In the network model of *Type 1*, the training process was slightly unstable, and failed to converge during training in some cases because of the unstable gradient from discriminator. On the other hand, in both three-stage network model of *Type 2* and *Type 3*, stable learning was possible in all cross-validation results.

Fig. 6 (a) shows the edge detection result for comparison using the Sobel filter. In this result, many small edges derived from the texture of the wall stones are detected, and when the stone has no clear boundary, there is a break in the edge. Fig. 6 (b), (c) and (d) indicate the edge detection results by each stage of our proposed Stacked cGAN.

In Fig. 6 (b), that is the output result of stage 1, desirable edge detection of wall stone with little noise is realized, but interruption of the edge is also observed. In Fig. 6 (c) in which a thicker edge is output, it can be confirmed that the breakage of many edges is improved. However, at a position where the interruption of the edge is large, the connection of the edge is not well established due to a lack of context in cGAN. Fig. 6 (d) represents the thinned edges generated from stage 3. Compared with the result of the Sobel filter, it can be confirmed that the reproducibility of the edges at the boundaries of the wall stones is improved and the noise is properly suppressed.

Fig. 6 (e) and (f) indicate the edge detection results derived from the network model of *Type 1* and *Type 2*, respectively. Compared to the result of Fig. 6 (d), these results have lower edge predictions and insufficient noise removal. In other words, it is

proved that the proposed three-stage Stacked cGAN is effective for predicting edges at unclear wall stone boundaries.

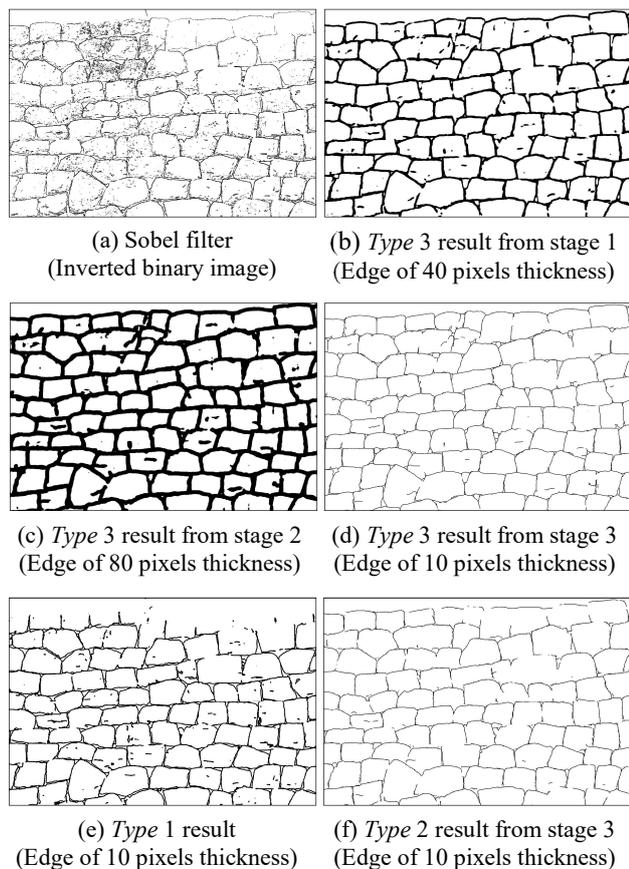


Figure 6. Edge detection results by proposed method

Fig. 7 illustrates a generation result of wall stone edges including stones not similar to given as the training data (shown in the lower half area of the image). In these areas, edges are not so extracted even though relatively clear gaps exist. This result implies that the proposed method does not simply predict the edges from the image, but generates the image after aggregating the properties of the wall stones given in the training process.

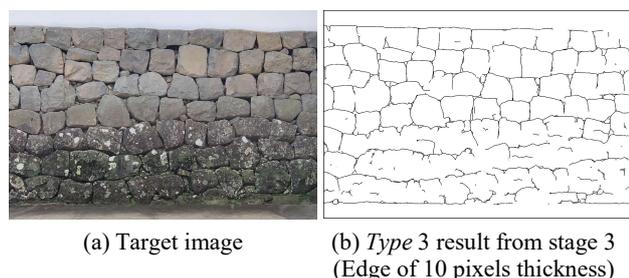


Figure 7. Processing result for test image containing stones not similar to training data

### 3.5 Extraction Results of Wall Stone Polygons

Fig. 8 is the comparison results of wall stone polygons extraction by multiscale image segmentation using an edge image generated by proposed Stacked cGAN as input data. Compared with the baseline result shown in Fig. 8(a), the processing results using the stone edge images shown in Fig. 8 (b), (c) and (d) realize the suppression of excessive merging of polygons and the improvement of separation performance of individual stones

polygons. Even in some areas where edges are not sufficiently detected, there are some places where the image segmentation achieves acceptable polygon extraction.

Table 1 shows the result of the quantitative evaluation of the number of extracted wall stones. In this table, “valid wall stones” indicate the stones successfully extracted. In addition, “undermerged wall stones” mean that the area is covering less than 90% by the extracted polygons, and “overmerged wall stones” indicate the case where one polygon is merged with the area exceeding 20% in any of a plurality of wall stones. In this table, *Baseline 1* is a method when only the *compactness* is used as shape criteria and only the conventional merging of two neighboring regions. *Baseline 2* is a method proposed in this study, that uses both *compactness* and the modified *convex hull fitness* as shape criteria, and in the region growing process selecting the best combination of a plurality of regions instead of two is allowed.

*Baseline 2* has improved the extraction performance of wall stone polygons compared to the conventional *Baseline 1*. Furthermore, we can confirm that the result of all types of proposed network models based on Stacked cGAN achieved performance exceeding the baseline methods. In particular, the improvement by *Type 3* is remarkable. In the object-based evaluation of the number for this case, 93.6% of wall stone polygons could be extracted properly. In *Type 3*, since the edge image is gradually improved as the stage progresses, it is considered that the effect of complementing the edge breakage became higher compared to other types. Therefore, it demonstrated the best performance for the extraction of the wall stone polygons.

On the other hand, it should be noted that the *Type 1* has a problem in the stability of training process and thus the processing result fluctuates. The result by *Type 1* shown here is somewhat irregular when the model happened to converge well.

When only the conventional baseline method is applied, it is difficult to finely adjust the degree of merging of wall stone polygons by setting the upper limit of *SP*. However, it was confirmed that when a wall stone edge image was used together, a stable processing result was obtained even when a relatively large variation was added to the upper limit of *SP*.

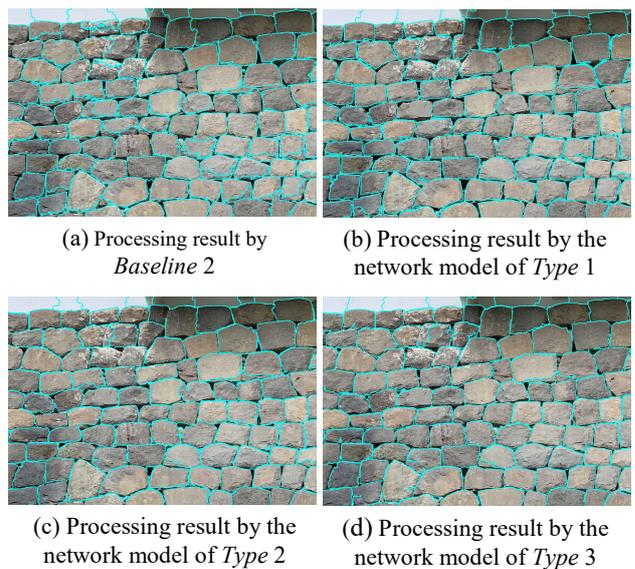


Figure 8. Comparison of wall stone polygon extraction after multiscale image segmentation

Method	Valid wall stones	Under-merged wall stones	Over-merged wall stones
Baseline 1	73	16	36
Baseline 2	83	18	24
Type 1	110	4	11
Type 2	98	6	21
Type 3	117	4	4

Table 1. Comparison of the number of extracted wall stones

Fig. 9 is an example showing the effect of the combined use of the edge image generated by the proposed Stacked cGAN and multiscale image segmentation. In Fig.9 (b), most of the outlines of the extracted wall stone polygons are generated in the region along the edges, but not all the edges are reflected (edges in red circle). Shown as the red circle in Fig. 9 (c), we can observe that the wall stone polygons are appropriately extracted by the multiscale image segmentation processing even in the area where the edge is not generated. This is a typical example of the effectiveness of the proposed method.

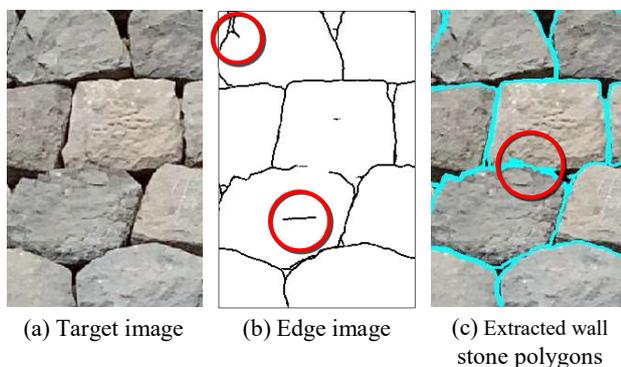


Figure 9. Effects of combinational use of edge image by Stacked cGAN and multiscale image segmentation

#### 4. CONCLUSIONS

In the present study, we have addressed a solution to the problem that the extraction performance of wall stone polygons would be degraded when there were no clear edges or boundaries between wall stones. To solve this problem, we improved the multiscale image segmentation technique used in our previous studies. The first improvement is that in the region growing process, selecting the best combination of a plurality of objects instead of two. The second improvement is that we modified the shape criterion of *convex fill fitness* which was proposed in our previous study.

We proposed a new cGAN based network model, called Stacked cGAN, to deal with the case where no boundaries between wall stones could not be solved by the improved image segmentation technique. This model enabled to generate edges even in areas where the boundaries of wall stones were unclear by using the existing wall stone edges as context.

It was confirmed that the separation performance of individual wall stone polygons was improved by appending the edge image generated by the proposed method to the input channel in baseline image segmentation. Furthermore, the proposed method is highly effective to reduce the difficulty in setting of the scale parameter, which is usually sensitive to segmentation results and requires trial and error.

#### REFERENCES

- Baatz, M. and Schäpe, A., 2000. Multiresolution segmentation: an optimization approach for high quality multi-scale image segmentation. In: *Angewandte Geographische Informations-Verarbeitung*, Vol. XII, pp. 12-23.
- Chen, Z., Zhao, Z.M., Yan, D.M., Chen, R.X., 2005. Multi-scale segmentation of the high resolution remote sensing image. In: *Proceedings of IEEE International Geoscience and Remote Sensing Symposium*, IGARSS '05, pp. 3682-3684.
- Esch, T., Thiel, M., Bock, M., Roth, A., Dech, S., 2008. Improvement of Image Segmentation Accuracy Based on Multiscale Optimization Procedure. *IEEE Geoscience and Remote Sensing Letters*, Vol. 5, No. 3, pp. 463-467.
- Goodfellow, I., Pouget-Abadie, J., Mirza, M., Xu, B., Warde-Farley, D., Ozair, S., Courville, A., Bengio, Y., 2014. Generative adversarial nets. In *Advances in neural information processing systems*, pp. 2672-2680.
- Ibrahim, Y., Nagy, B., Benedek, C., 2019. CNN-Based Watershed Marker Extraction for Brick Segmentation in Masonry Walls. In: Karray F., Campilho A., Yu A. (eds) *Image Analysis and Recognition*, ICIAR 2019. Lecture Notes in Computer Science, Vol. 11662, Springer, Cham.
- Isola, P., Zhu, J.-Y., Zhou, T., Efros, A. A., 2017. Image-to-image translation with conditional adversarial networks. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*.
- Křemen, T., Koska, B., Pospíšil, J., 2011. Laserscanning for Castle Documentation. In: *Proceedings of the 23rd CIPA Symposium*, Prague, pp. 1-8.
- Li, H. T., Gu, H.Y., Han, Y. S., Yang, J. H., Han, S. S., 2009. An Efficient Multi-scale Segmentation for High-resolution Remote Sensing Imagery Based on Statistical Region Merging and Minimum Heterogeneity Rule. *IEEE J. Selected Topics in Appl. Earth Observations and Remote Sensing*, 2(2), pp. 67-73.
- Mirza, M., Osindero, S., Conditional generative adversarial nets., 2014. arXiv preprint arXiv:1411.1784.
- Ronneberger, O., Fischer, P., Brox, T., 2015. U-net: Convolutional networks for biomedical image segmentation. In: *MICCAI*.
- Ryu, H., Yamamoto, H., Nishida, K., Nishigata, T., 2014. Dismantling and Repair Method for Stone Wall, Japan Pat. JP5457116B2 (17, January, 2014).
- Sakamoto, M., Tsuguchi, M., Chhatkuli, S., Satoh, T., 2018. Extended Multiscale Image Segmentation for Castellated Wall Management. *International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, XLII-2, 999-1005.
- Vacca, G., Deidda, M., Dessi A., Marras, M., 2012. Laser Scanner Survey to Cultural Heritage Conservation and Restoration, *International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, Vol. XXXIX-B5, Melbourne, Australia, pp. 589-594.