CBDAR2021, 06 September 2021

Inscription Segmentation Using Synthetic Inscription Images for Text Detection at Stone Monuments

Naoto Morita, Ryunosuke Inoue, Masashi Yamada*,Takatoshi Naka,
Atsuko Kanematsu, Shinya Miyazaki, Junichi Hasegawa
Chukyo University, Japan

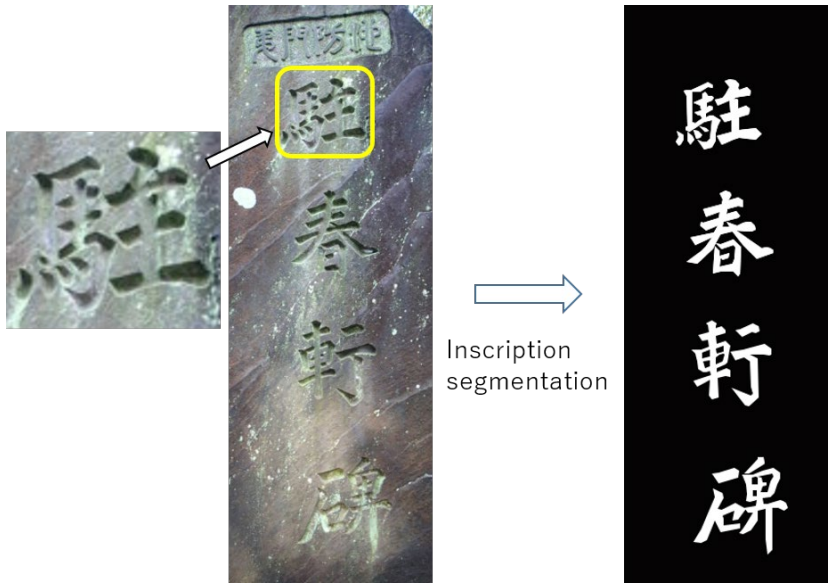
Introduction

- Stone monuments have an important place in history research.
- Photography has been used as an easy way to record these inscriptions.
- However, the light present at the time of photography, the resulting shadows created, and the basic texture of the stone can make the text in the photographs indistinct and difficult to recognize.

Stone monument

Objective

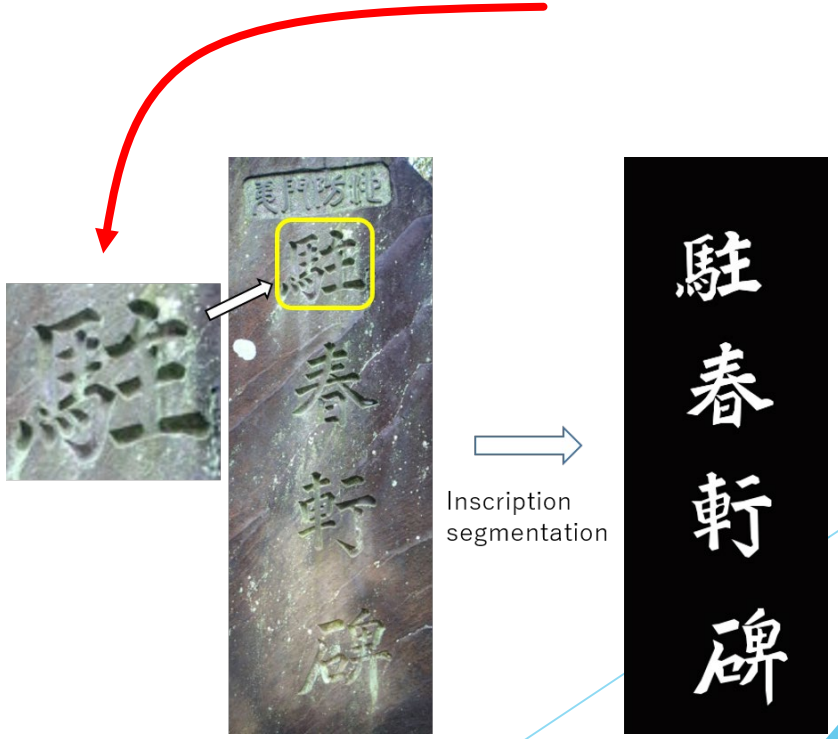
- This paper proposes a method for inferring pixel-wise text areas from the inscription images of stone monuments using deep learning that would help to decipher the script.



The text area corresponds to the area on the image that has been engraved according to the shape of the text characters.

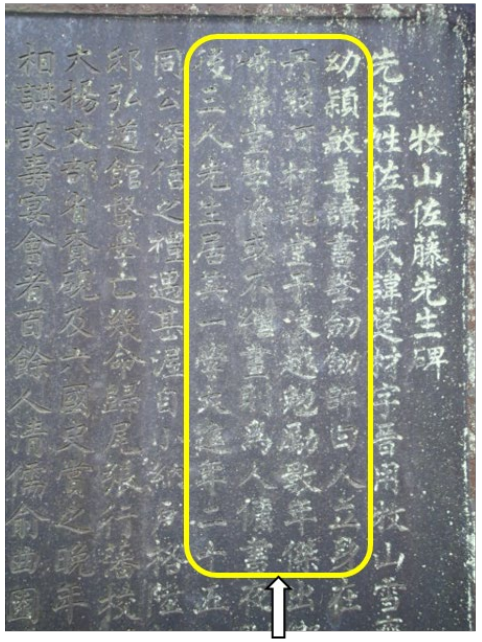
Difficulties of detecting the text area

- Depending on how the light hits the engraved area, the brightness of the reflected light varies.
- In other words, the shading in the text area is not uniform.

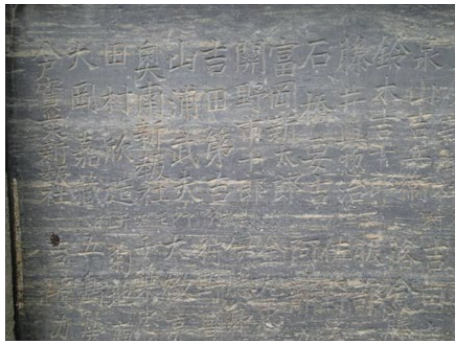


Difficulties of detecting the text area

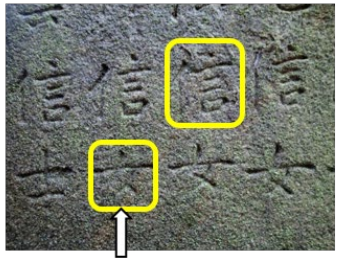
- In actual stone monuments, the factors that obscure the text are not only the stone texture but also other factors that include dirt and weathering.



Obscure text due to dirt



Obscure text due to stone texture



Obscure text due to weathering

Difficulties of detecting the text area

- Our target is stone monuments in Japan that contain Chinese characters, `kanji'.
- There are more than 2,000 classes of kanji characters currently used in Japan, and more than 5,000 if we include characters used in older eras.
- However, deep learning requires sufficient training data and datasets of stone monuments containing kanji are not yet available.

Approach

- Our method uses pseudo-inscription images for training a deep neural network.
- Pseudo images are generated by synthesizing a shaded image representing the engraved text and stone texture image.
- Through experiments using a network model, we confirm that the network model achieves high accuracy in the task of inscription segmentation and that training with pseudo-inscription images is effective in detecting inscriptions on real stone monuments.

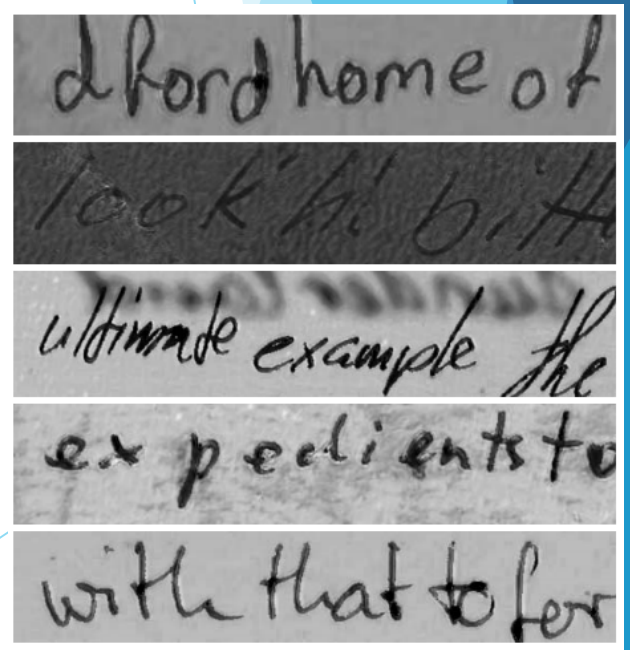
To the best of our knowledge, this is the first study that involves training a network using pseudo-inscription images and detecting text engraved on stone monuments.

Related Work

- Text detection on inscription images
  - Liu et al. proposed a method for obtaining the bounding box of characters engraved on oracle bones using a conventional model of object detection [12].
  - Qin et al. proposed another method for obtaining the bounding box of characters from a scene with stone monuments [17]
- Binarization of document images.
  - Kitadai et al. proposed a deciphering support system using four basic binarization methods [10].
  - Peng et al. proposed a binarization method incorporating multi-resolution attention that achieved better accuracy than the best model in the ICDAR2017 competition [16][15].

Related Work

- Pseudo text image generation for training networks
  - Jaderberg et al. generated text in various fonts, styles, and arrangements, and created a dataset of automatically generated text images that are difficult to recognize[7] .
  - Tensmeyer et al. proposed a DGT-CycleGAN that generated highly realistic synthetic text data [22].



Proposed Method:
Creating Images of Pseudo-Inscription

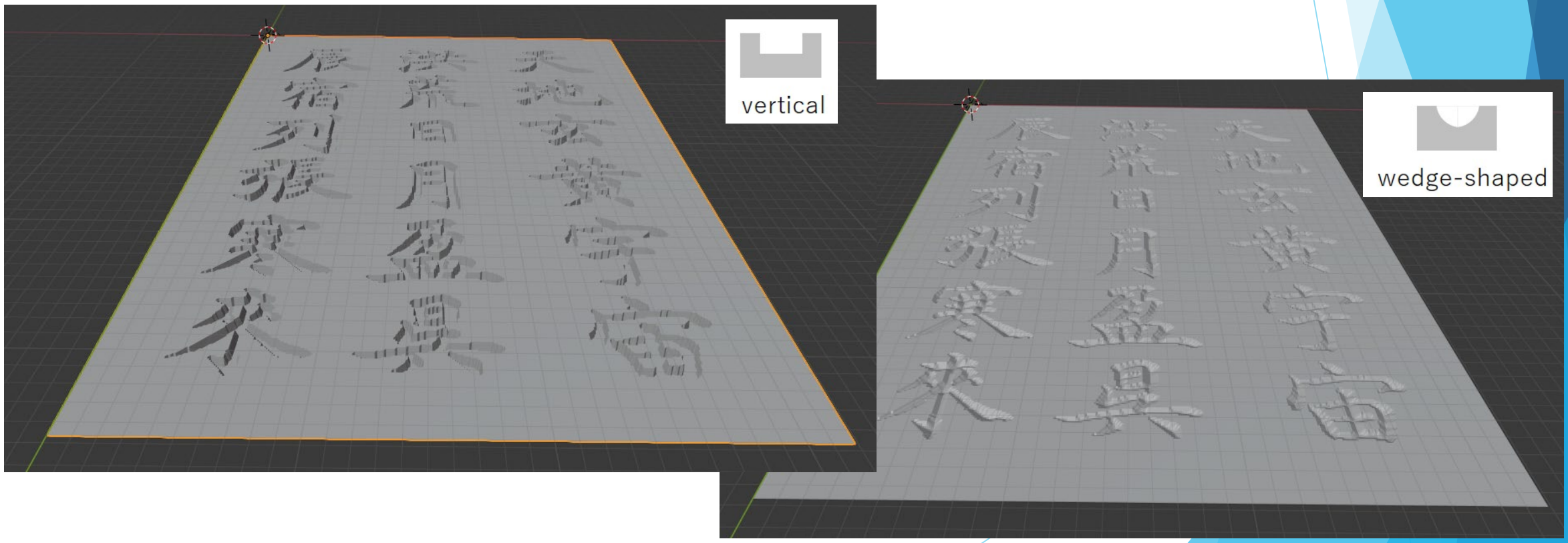
1. Generate **the text image**.
2. Generate **the three-dimensional mesh data** with the text engraved into it.
3. Generate **a shaded text image** by setting a light source and rendering it.
4. Blend with **a stone texture image**.

1. Generating text images

- The text image is composed of the handwritten kanji characters found in a calligraphy instruction book [14] containing 1,000 different classes of characters
- Of these, 50x18 characters are extracted, 18 at a time, and transferred to a single image, making a total of 50 images. The text image is a binary image with a spatial resolution of 320x448.

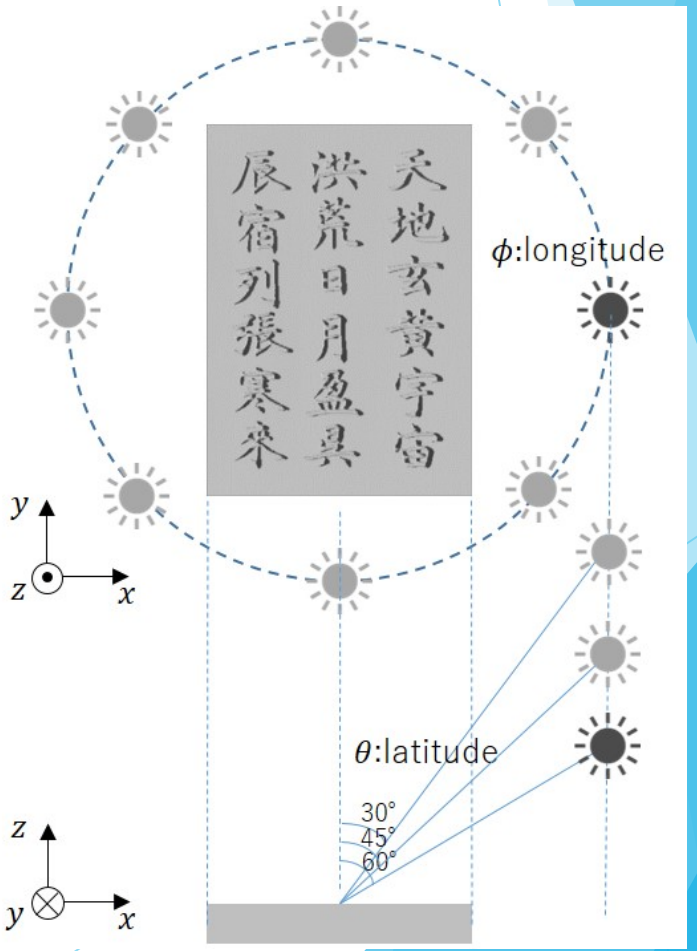
2. Generating the three-dimensional mesh data with the text engraved into it

- Engraving type: vertical and wedge-shaped

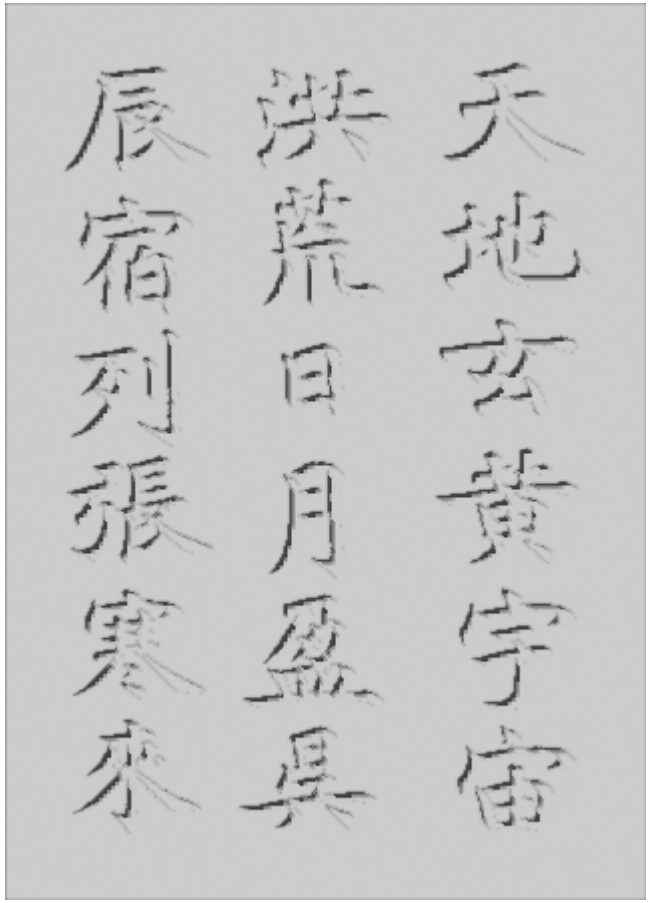


3. Generating a shaded text image by setting a light source and rendering it.

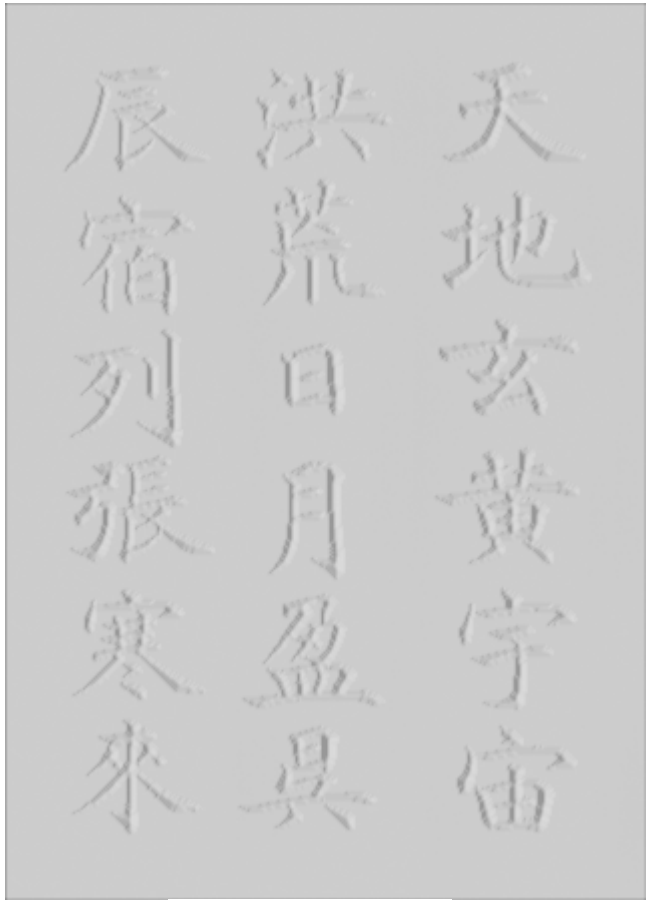
- Setting a parallel light source at the polar coordinate $(\mathrm{r}, \theta, \phi)$
  where $\phi = 0°, 45°, \cdots, 315°, \theta = 30°, 45°, 60°$
- The total of possible light positions is 8x3=24



Examples of shaded text images



vertical



wedge-shaped

In total, 2400 shaded images are generated;
(50 text images × 2 engraving types
× 24 light position)

4. Blending with a stone texture image

- 18 stone texture images

4. Blending with a stone texture image

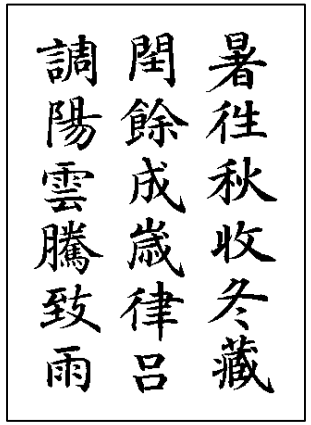
- Pseudo inscription image p

$\odot$: Hadamard product

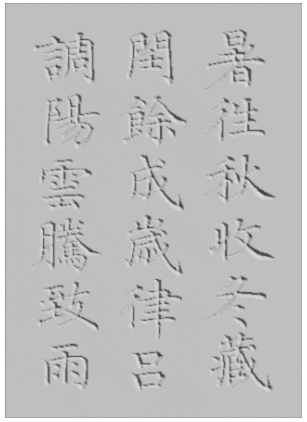
α: blend ratio

$$p = \bar{m} \odot s \odot t + \alpha m + (1 - \alpha) m \odot s \odot t$$

m: text image



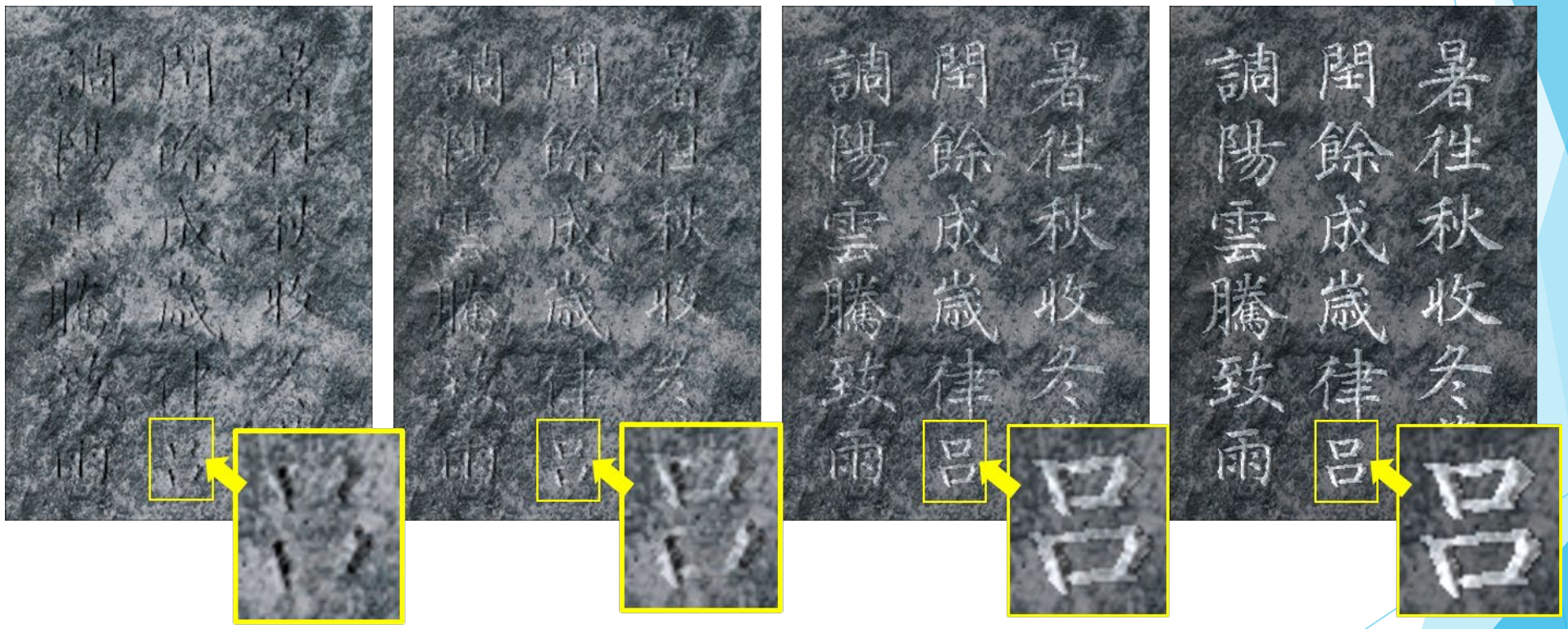
$\bar{m}$: inverse of m



s: shaded image

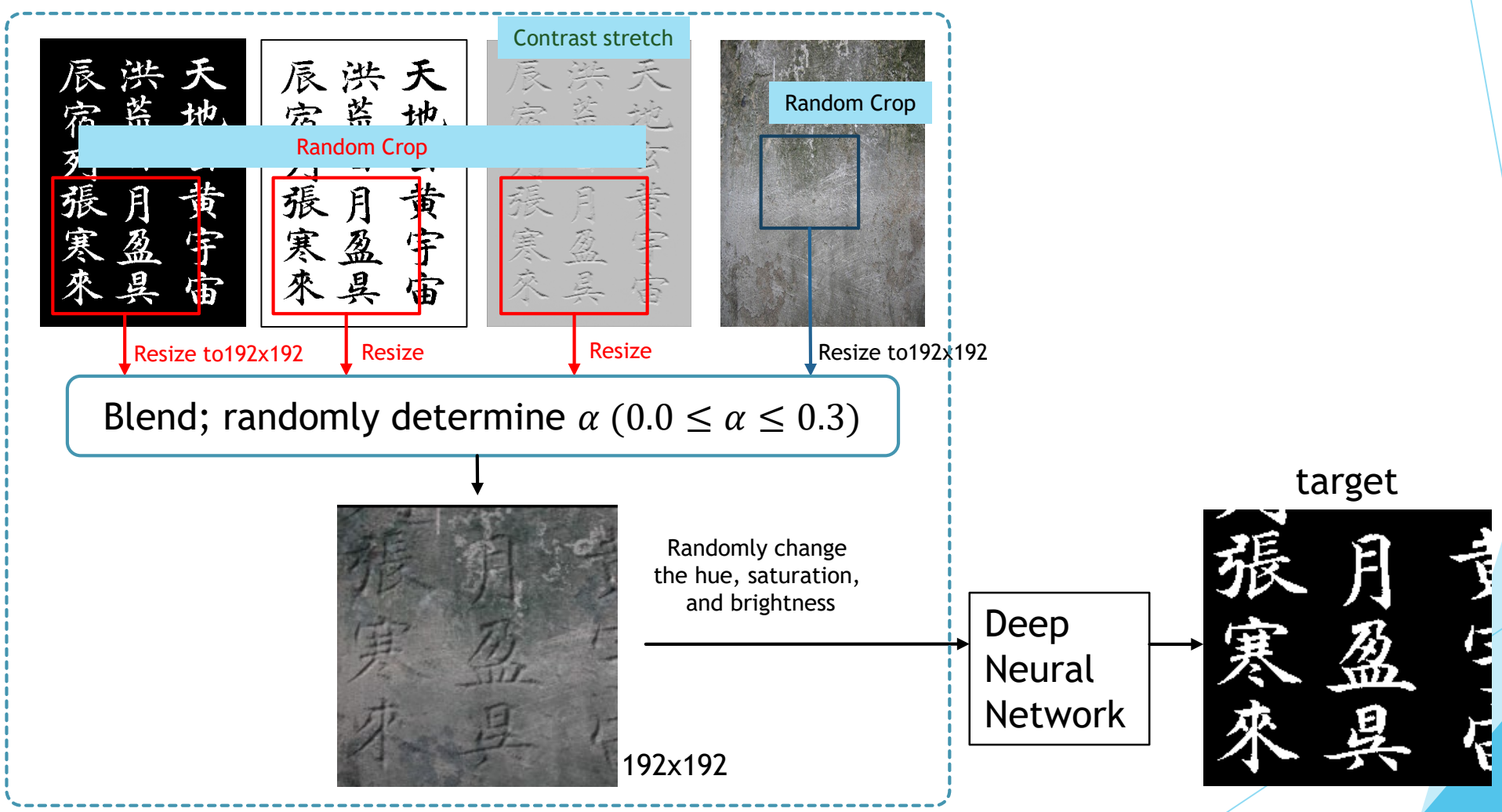
t: stone texture

Examples of pseudo-inscription image



$\alpha = 0.0$ $\alpha = 0.1$ $\alpha = 0.2$ $\alpha = 0.3$

Data Augmentation

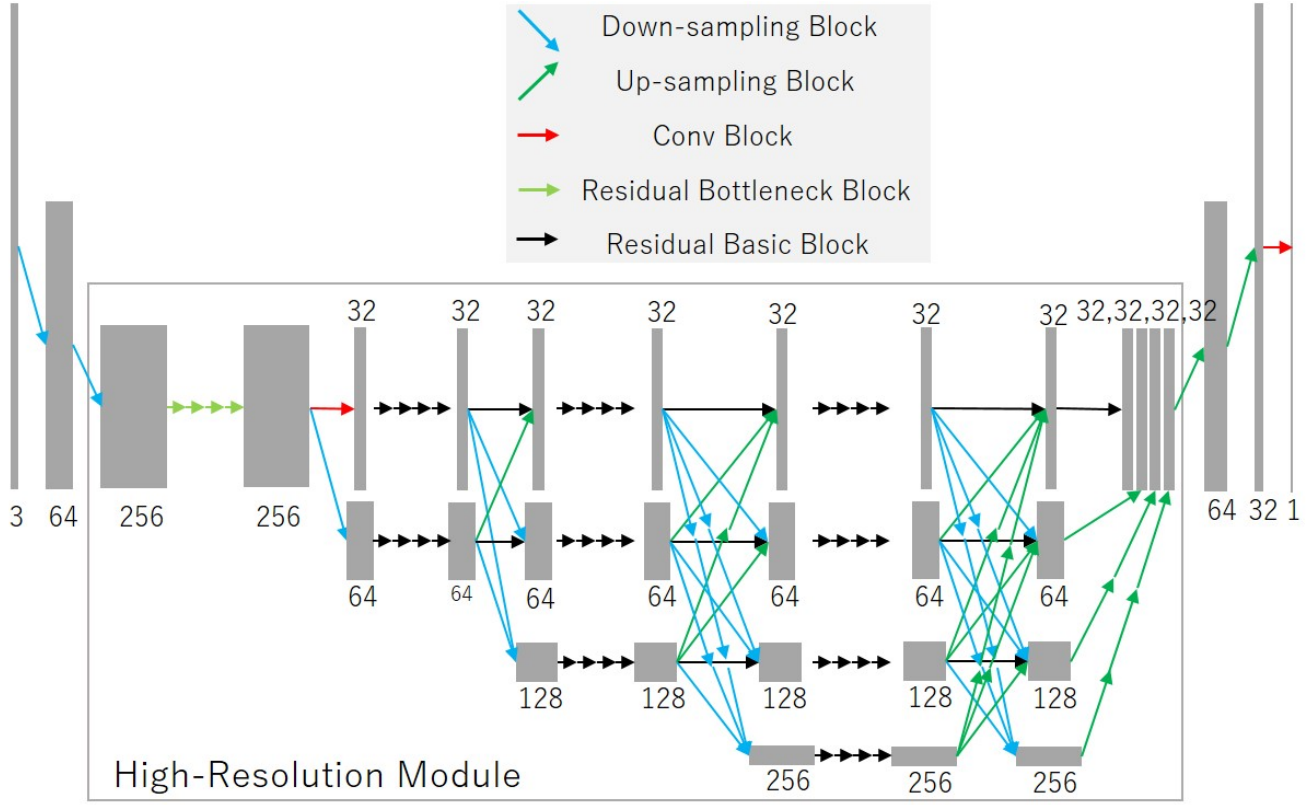


Pseudo-inscription images generated by data augmentation processing

Network Model

one of the state of the art models for semantic segmentation

- High Resolution Network (HRNet) [21]



Experiments

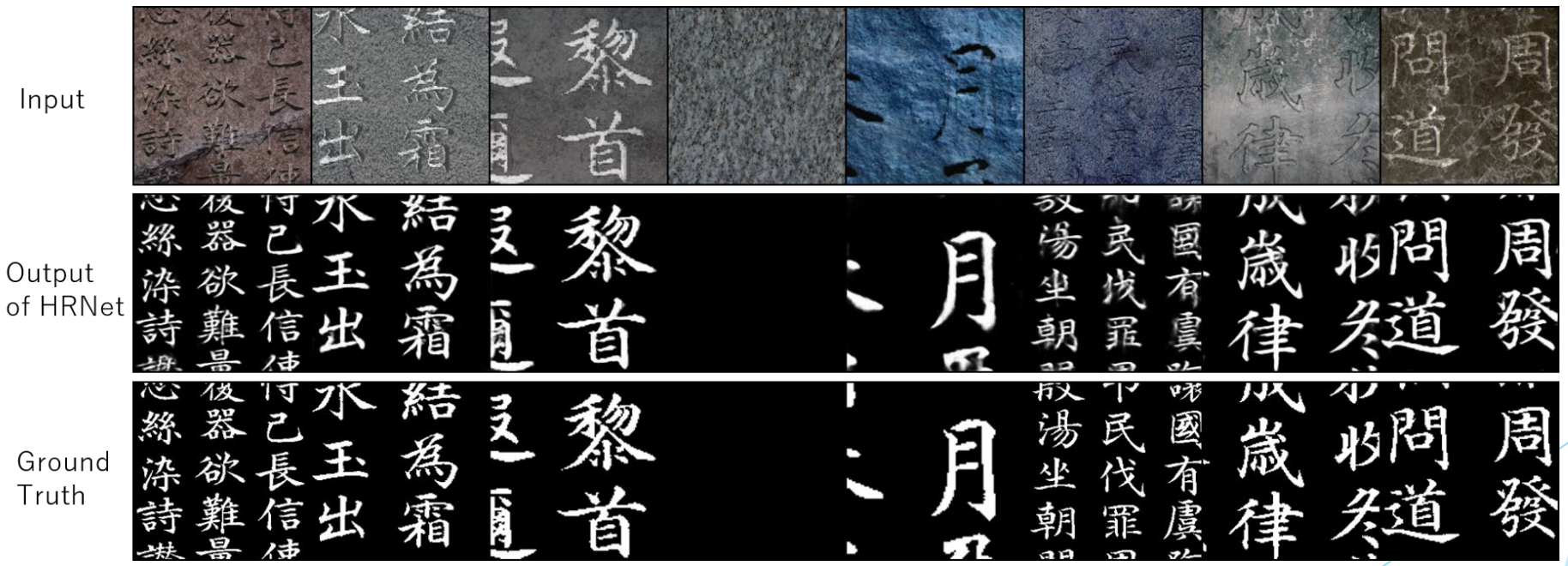
- Training and Validation data
  - Training : Validation = 4 : 1
    - Text images 40 : 10 (included characters 720: 180)
    - Shaded images 1,920 : 480
  - The same 18 stone textures are used for both training and validation phases.
  - In the training phase, the network is trained on negative samples in addition to the training data.
    - Negative samples are texture images without text.

Experiments

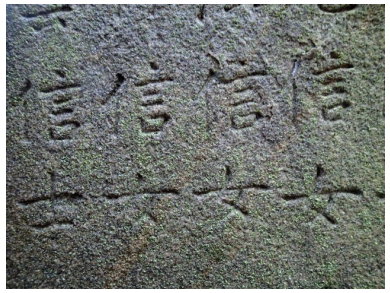
- Implementation Details
  - Loss function: mean square error
  - Learning rate: 10^{-4}
  - Optimization algorithm: adam
  - Batch size: 48
  - DL framework: Pytorch
  - GPU: GeForce RTX 2080 Ti

Experiments

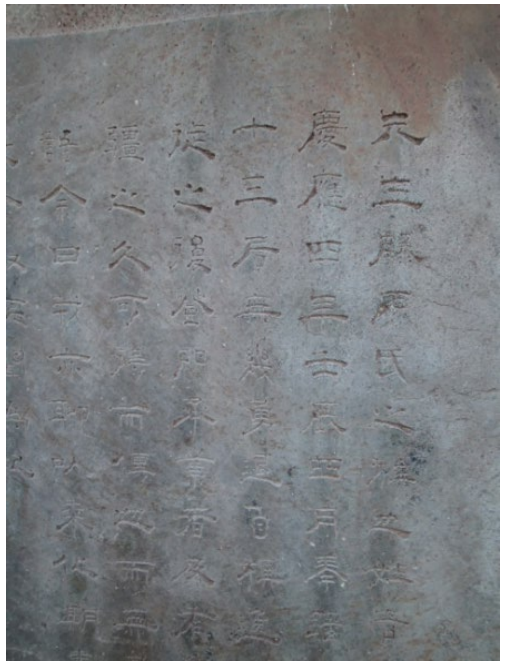
- Results for validation data (after 500 epochs)
  - Precision 0.92, recall 0.86, f-measure 0.88

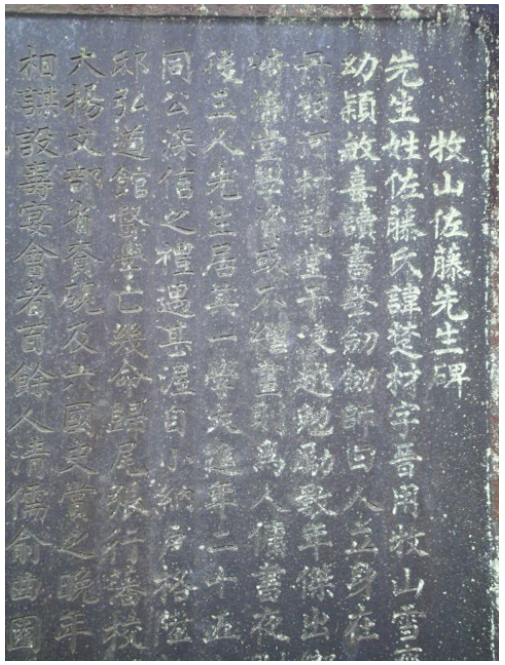


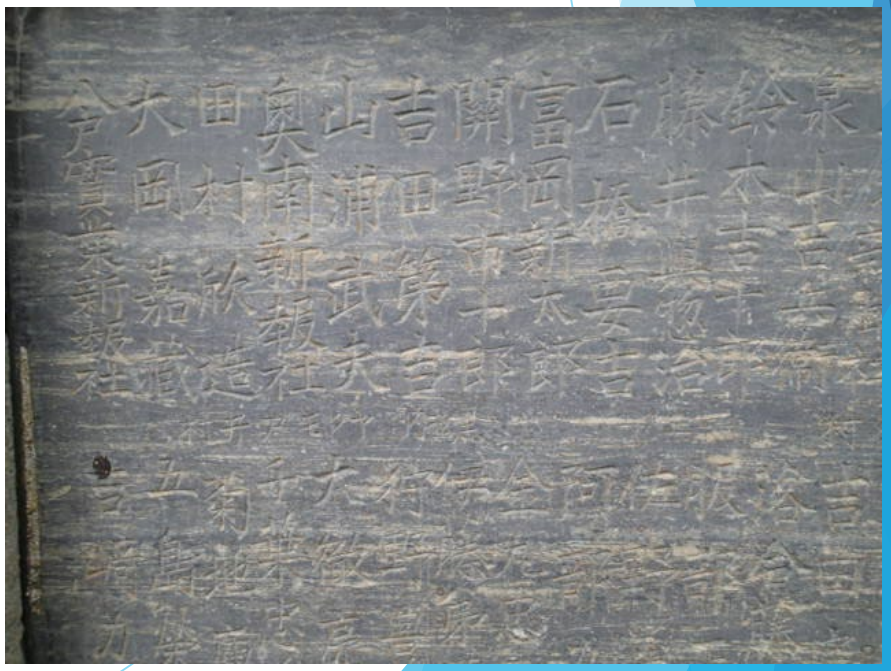
Experiments



- Experiment on real inscription images
  - The images of actual stone monuments were taken at Hakodate Park and Koushoji Temple in Japan.



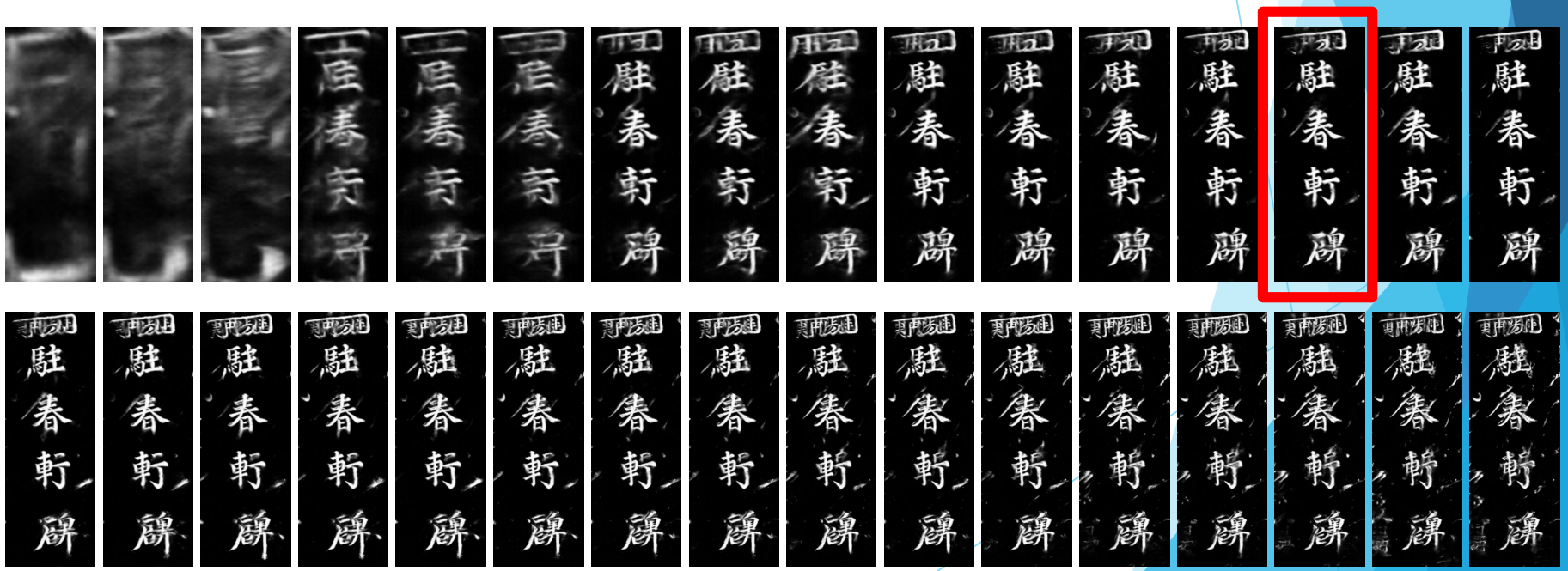




Experiments

- Experiment on real inscription images
  - Each image of the actual stone monuments is resized at 0.05 intervals from 0.1 to 1.8, and 32 images with different resolutions are provided.
  - The results with the highest value of f-measure is used for evaluation.

Resize to 32 resolutions and input them to the network



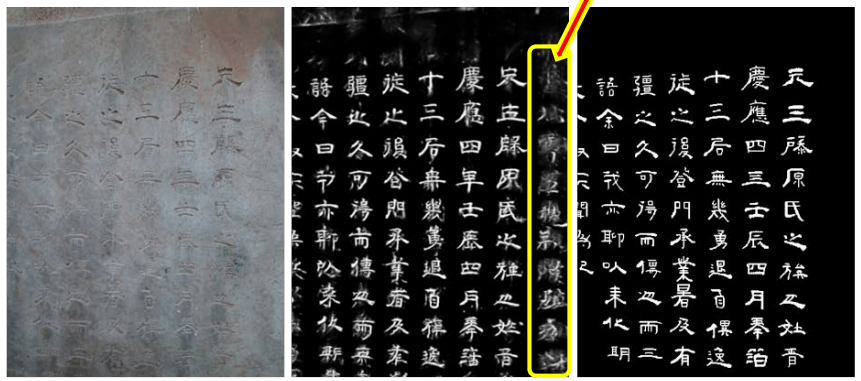
32 output images are resized to the original resolution

Experiments

- Results for real inscription images

Input | Output of HRNet (a) | Ground Truth

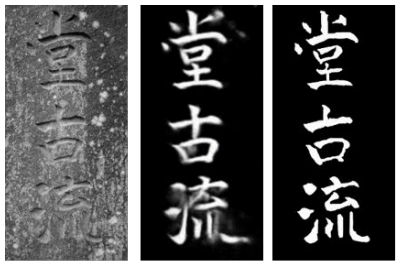
Over detection



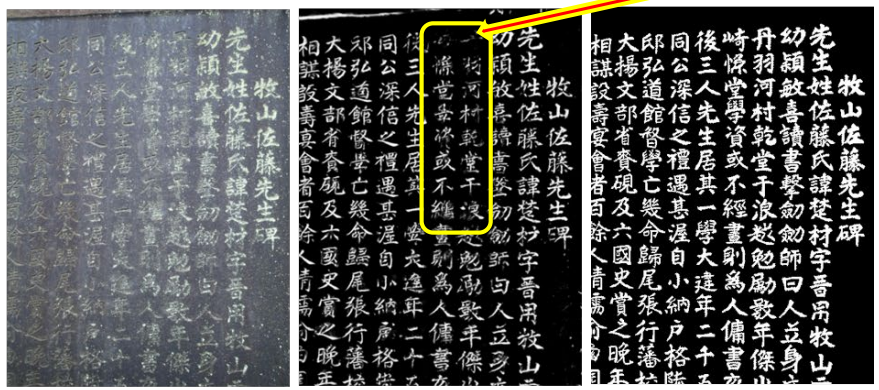
Input | Output of HRNet (c) | Ground Truth

Over detection

Insufficient detection



Input | Output of HRNet (b) | Ground Truth



Input | Output of HRNet (d) | Ground Truth

Partial over-detection or insufficient detection is observed.

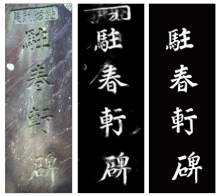
Experiments

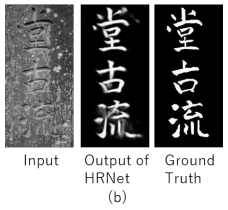
- Results for real inscription images

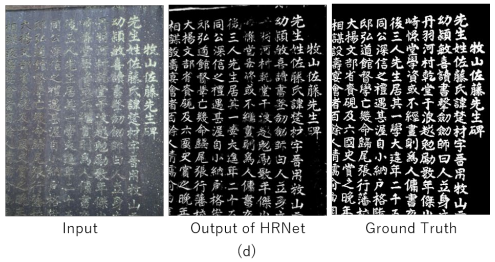
The best result among four basic binarization methods proposed in [10]:

Input		HRNet			binarization proposed in [10]		
image	size	*precision*	*recall*	*f-measure*	*precision*	*recall*	*f-measure*
a	160×480	0.72	0.85	0.78	0.14	0.62	0.23
b	96×256	0.74	0.82	0.78	0.16	0.55	0.25
c	576×800	0.61	0.73	0.66	0.10	0.44	0.17
d	572×768	0.91	0.50	0.65	0.70	0.55	0.62

The f-measure values are less than those for validation data.
This implies the over-fitting to pseudo inscription images.

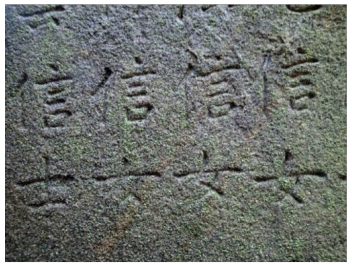




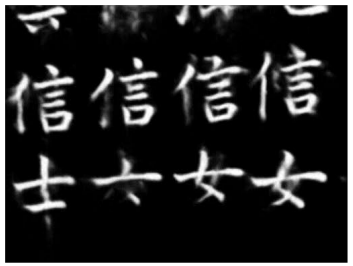


Experiments

- Results for other real inscription images



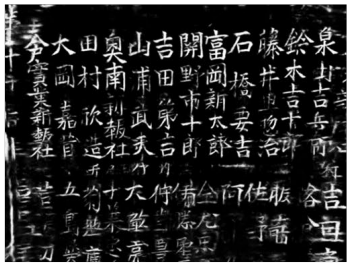
Input



(a) Output of HRNet

The strokes in the weathered area were not detected

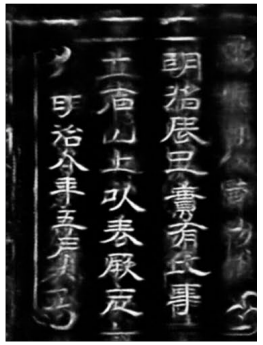
Input



(b) Output of HRNet

Many text areas were not detected because of the high contrast texture.

Input



(c) Output of HRNet

This is written in a peculiar calligraphy style. Our pseudo-inscription may need to adapt to a variety of calligraphy style.

Conclusion

- We proposed a low-cost technique for generating pseudo-inscription images. It is useful for training the network for inscription segmentation.
- Experimental results implied the over-fitting to pseudo inscription images.
- In order to improve the accuracy for real stone monuments, it is necessary to have pseudo inscription images with more variety of character placement, calligraphy style and stone texture.
- HRNet is not sufficient to various resolution. We need to investigate other network models.

Thank you for your attention.