Semantic Segmentation with GLCM Images

Akira Nakajima[1] [a] and Hiroyuki Kobayashi[2] [b]
[1] *Graduate School of Robotics and Design, Osaka Institute of Technology, Osaka, Japan*
[2] *Department of System Design, Osaka Institute of Technology, Osaka, Japan*
{m1m23r25, hirokyuki.kobayashi}@oit.ac.jp

Keywords: GLCM Images, Semantic Segmentation, U-Net, Ready-Mixed Concrete.

Abstract: At construction sites, there is a problem of excess ready-mixed concrete due to ordering errors being disposed of as industrial waste, and there is a need to introduce image recognition technology as an indicator to determine the appropriate amount to order. In this study, we attempted to detect ready-mixed concrete using a machine learning technique called semantic segmentation. We believe that texture analysis can solve the problem that raw concrete is difficult to recognize accurately because its texture is similar to that of other building materials and backgrounds and its texture fluctuates depending on the amount of moisture and mixing conditions. In this study, we proposed to perform texture analysis using GLCM (Gray Level Co-occurrence Matrix) and use the resulting image dataset. the results using GLCM images show that, compared to conventional segmentation, the GLCM images can be used to identify a variety of raw The results using the GLCM images provided highly accurate predictions for a wide variety of raw concrete placement conditions at construction sites, compared to conventional segmentation methods.

1 INTRODUCTION

The problem of excess ready-mixed concrete due to over-ordering at construction sites is becoming a serious issue as it is disposed of as industrial waste. In 2023, more than 2 million cubic meters of ready-mixed concrete were discarded annually in Japan, not only posing environmental challenges but also placing a financial burden on concrete manufacturers for disposal costs. Therefore, determining the appropriate order quantity has become a pressing issue. To address this, there is a growing demand for the introduction of image recognition technology to provide accurate order volume estimations, especially in order to accommodate the various concrete pouring conditions on construction sites. In this study, we propose using semantic segmentation, a machine learning technique, to detect ready-mixed concrete from construction site images. However, the texture of fresh concrete is similar to that of other construction materials and the background, and the texture fluctuates depending on the amount of moisture and mixing conditions, making accurate recognition difficult. Therefore, we believe that by introducing texture analysis, it will be possible to extract the texture and detailed surface features of raw concrete and enable recognition that can cope with texture similarity and variation, which has been difficult with conventional segmentation methods. In this study, we propose to use images with texture analysis added using Gray Level Co-occurrence Matrix (GLCM) as a dataset. Semantic segmentation using texture analysis has demonstrated its effectiveness in various fields. For example, in garment segmentation, combining texture and semantic decoding modules has been shown to improve accuracy.(Liu et al., 2023) In the classification of herbal plants, hybrid methods using GLCM with CNN or SVM have achieved high classification accuracy (Purnawansyah et al., 2023). Additionally, for 3D urban scene mesh data, the introduction of a texture convolution module significantly improved segmentation accuracy compared to traditional methods (Yang et al., 2023). Furthermore, for SAR images, a new method based on texture complexity analysis and key superpixels has been proposed, enhancing noise resistance and distinguishing different landforms (Shang et al., 2020).

[a] https://orcid.org/0009-0002-1142-9470
[b] https://orcid.org/0000-0002-4110-3570

Nakajima, A. and Kobayashi, H.
Semantic Segmentation with GLCM Images.
DOI: 10.5220/0013072200003822

In *Proceedings of the 21st International Conference on Informatics in Control, Automation and Robotics (ICINCO 2024) - Volume 1*, pages 527-531
ISBN: 978-989-758-717-7; ISSN: 2184-2809

2 METHOD

2.1 GLCM Images

Gray-level co-occurrence matrix (GLCM) is one of the methods used in image texture analysis. It is a matrix that represents the frequency with which a particular gray level combination occurs between adjacent pixels in an image.

GLCM defines pairs of pixels based on a given direction (e.g., 0, 45, 90, or 135 degrees) and distance. In this case, the direction is 0 degrees and the distance is 1 pixel. Based on the direction and distance, count how often a given gray level pixel i and its adjacent pixel gray level j occur simultaneously and reflect it in the matrix $P(i,j)$. For example, if one gray level is

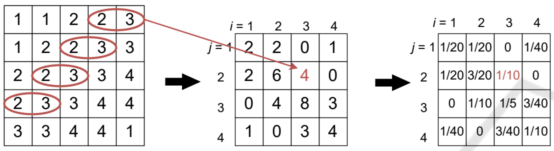


Figure 1: Procedure for calculating the GLCM.

"2" and the gray level of an adjacent pixel is "3" as in Figure1, there are four combinations, so we assign "4" to the position$P(i,j)$ in the GLCM.
Furthermore, by dividing the value of each element of the GLCM by the total number of pixel pairs, we can express the co-occurrence frequency as a probability. A scalar value is obtained by applying a function to the GLCM matrix. This is the GLCM feature. In this study, six GLCM features were obtained by applying the functions shown in Equation 1 to Equation 6 to the GLCM matrix.

$$\text{Contrast} = \sum_{i=0}^{N-1}\sum_{j=0}^{N-1}(i-j)^2 P_{ij} \tag{1}$$

$$\text{Homogeneity} = \sum_{i=0}^{N-1}\sum_{j=0}^{N-1}\frac{P_{ij}}{1+|i-j|} \tag{2}$$

$$\text{Energy} = \sum_{i=0}^{N-1}\sum_{j=0}^{N-1}(P_{ij})^2 \tag{3}$$

$$\text{Dissimilarity} = \sum_{i=0}^{N-1}\sum_{j=0}^{N-1}|i-j|P_{ij} \tag{4}$$

$$\text{Correlation} = \frac{\sum_{i=0}^{N-1}\sum_{j=0}^{N-1}(i-\mu_i)(j-\mu_j)P_{ij}}{\sigma_i\sigma_j} \tag{5}$$

$$\text{Entropy} = -\sum_{i=0}^{N-1}\sum_{j=0}^{N-1}P_{ij}\log(P_{ij}) \tag{6}$$

Contrast represents the difference in intensity of edges and gray levels in an image. Homogeneity indicates the uniformity of the image texture. Energy indicates the regularity or consistency of repeated patterns in the texture of an image. Dissimilarity represents the difference in gray levels in the image and is used to detect changes in the texture. Correlation represents regularity in the texture of an image. Entropy indicates randomness or complexity in an image.
GLCM features are obtained as a single scalar value for a single image, but to use them as a segmentation dataset, a GLCM feature is needed for each pixel. Therefore, we divided the image into smaller regions and calculated GLCM features for each region as Figure 2.

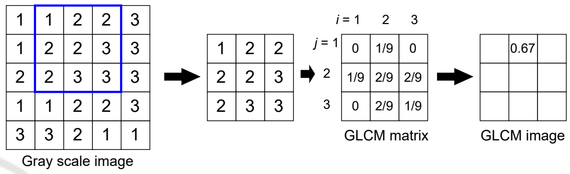


Figure 2: Truth A.

Specifically, for a 512 x 512 pixel RGB image, we generated a GLCM matrix using a 7 x 7 pixel kernel and applied six different functions to obtain GLCM features for each pixel. Figure 3 shows the original image and examples of GLCM images are shown in Figures 4 through 9.

Figure 3: Sample.

Figure 4: Contrast.

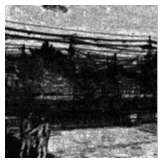


Figure 5: Homogeneity.

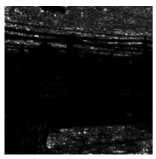


Figure 6: Energy.

Figure 7: Dissimilarity.

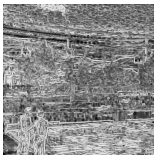


Figure 8: Correlation.

Figure 9: Entropy.

2.2 U-Net++

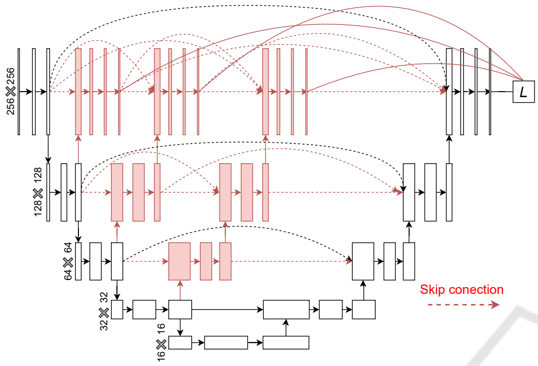


Figure 10: U-Net++

U-Net is a convolutional neural network (CNN) specialized for image segmentation.(Ronneberger et al., 2015) It has a structure with an encoder to extract image features and a decoder to reconstruct images, and can retain spatial information through skip connections. Its advantage is that it can perform highly accurate segmentation even with small amounts of data. In this study, we employed U-Net++ (Figure 10), which has been improved to extract more detailed features by increasing the number of skip connections.(Zhou et al., 2018)

2.3 Transfer Learning

The technique of pre-training another dataset to improve segmentation accuracy is called “Transfer Learning”. ResNet is a very deep convolutional model, and the skip connection solves the problem of “learning does not proceed well with deeper layers of the model, even though gradient loss does not occur.(He et al., 2015) VGG11 was used as the encoder in the TernausNet(Iglovikov and Shvets, 2018) paper that showed the effectiveness of U-Net transition learning, but we chose ResNet, which has deeper layers and better performance. I used ResNet, a model that excels at image recognition, which has already been trained on a large dataset called ImageNet, as the encoder for the training model. I applied learned weights to the encoder and random weights to the decoder. These weights are updated during the training process.

When training the new 9-channel dataset, the weights for the 6 channels corresponding to the additional GLCM features were set randomly, since ImageNet is a 3-channel image dataset. To increase the effectiveness of the transition learning, we also normalized the distribution of the new dataset to match the distribution of the pre-training data. The mean and standard deviation of each channel of the pre-training data were used for normalization; the mean and standard deviation corresponding to the 3-channel images were taken from ImageNet, and for the additional 6 channels, the mean was set to 0 and the standard deviation to 1, based on batch normalization.(Bjorck et al., 2018)

2.4 Augmentation

Augmentation is a method of extending data by performing various transformations on the original image. In this case, the data was theoretically augmented by a factor of 4 by applying a horizontal inversion and a viewpoint change in 3D space at each epoch with a probability of 50 % each.

2.5 Loss Functions and Evaluation Metrics

The Dice coefficient was used as the loss function and the IoU score as the evaluation index. Both are widely used in segmentation tasks, with the Dice coefficient emphasizing the degree of overlap between predicted and true regions and the IoU score being a direct measure of agreement between predicted and true regions.

$$\text{Dice} = \frac{2|A \cap B|}{|A| + |B|} \tag{7}$$

$$\text{IoU} = \frac{|A \cap B|}{|A \cup B|} \tag{8}$$

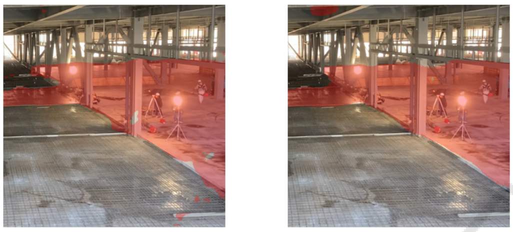
3 RESULTS

The dataset consisted of 104 images of 9 channels with the addition of the GLCM image, and the number of epochs was set to 120 for the comparison experiment with the 3-channel image. The Dice coefficient was used as the loss function, and the loss function was optimized using Adam. Evaluation was performed on two images A and B. The IoU score, which indicates overall recognition accuracy, is shown in Tab.1, and the segmentation prediction results are shown in Fig.11 to Fig.14.

Table 1: Accuracy.

	3 channels	9 channels
IoU score	0.8888	0.8979

Figure 11: Truth A.



(a) 3 channels (b) 9 channels

Figure 12: Prediction A.

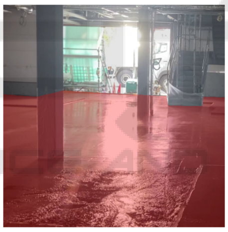
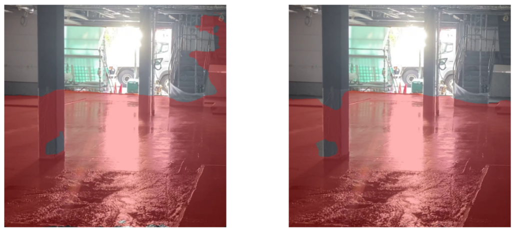


Figure 13: Truth B.

From Table 1, the 9-channel image with the GLCM features was more accurate than the 3-channel image using the conventional method. In Fig. 12, the 3-channel image had a detection error from the center to the lower right, whereas the 9-channel image showed a detection error in the upper left. The 3-channel image had a false positive in the upper left corner, but there was no significant difference in overall prediction accuracy. Furthermore, in Fig. 14, the prediction result for the 3-channel image showed a false positive in the upper right corner, while the prediction result for the 9-channel image had no false positive and was accurate.

4 DISCUSSIONS

From Fig.12, the reason for the higher number of false positives in the prediction results for the 9-channel



(a) 3 channel (b) 9 channel

Figure 14: Prediction B.

images compared to the 3-channel images can be attributed to overfitting of the model. Factors contributing to overfitting include the fact that the model became too complex due to the increased dimensionality of the features, and that the training data set was small.

On the other hand, from Fig.14, the reason why the prediction results for the 9-channel image were more accurate than for the 3-channel image is that the model was able to capture the boundaries between the raw concrete surface and the interior walls. The features that were effective in detecting edges and boundaries were Contrast, Correlation, and Entropy.

5 CONCLUSIONS

The addition of GLCM images for texture analysis to the segmentation dataset resulted in more accurate recognition of raw concrete at construction sites, with texture and boundaries being effectively extracted as features. This improvement is expected to facilitate the determination of appropriate order quantities, thereby reducing the amount of raw concrete that becomes industrial waste.

6 PERSPECTIVES

The approach is to reduce the dimensionality of the features in the dataset to mitigate model overfitting. Specifically, we will consider using 3-channel images as the dataset and computing GLCM features within the model.

REFERENCES

Bjorck, J., Gomes, C. P., and Selman, B. (2018). Understanding batch normalization. *CoRR*, abs/1806.02375.

He, K., Zhang, X., Ren, S., and Sun, J. (2015). Deep residual learning for image recognition. *CoRR*, abs/1512.03385.

Iglovikov, V. and Shvets, A. (2018). Ternausnet: U-net with VGG11 encoder pre-trained on imagenet for image segmentation. *CoRR*, abs/1801.05746.

Liu, Z., Li, H., and Zhu, P. (2023). Garment segmentation network based on texture and semantic decoding module. In *Journal of Physics: Conference Series*, volume 2562, page 012028. IOP Publishing.

Purnawansyah, P., Wibawa, A. P., Widyaningtyas, T., Haviluddin, H., Hasihi, C. E., Teng, M. F., and Darwis, H. (2023). Comparative study of herbal leaves classification using hybrid of glcm-svm and glcm-cnn. *ILKOM Jurnal Ilmiah*, 15(2):382–389.

Ronneberger, O., Fischer, P., and Brox, T. (2015). U-net: Convolutional networks for biomedical image segmentation. *CoRR*, abs/1505.04597.

Shang, R., Peng, P., Shang, F., Jiao, L., Shen, Y., and Stolkin, R. (2020). Semantic segmentation for sar image based on texture complexity analysis and key superpixels. *Remote Sensing*, 12(13):2141.

Yang, Y., Tang, R., Xia, M., and Zhang, C. (2023). A texture integrated deep neural network for semantic segmentation of urban meshes. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, 16:4670–4684.

Zhou, Z., Siddiquee, M. M. R., Tajbakhsh, N., and Liang, J. (2018). Unet++: A nested u-net architecture for medical image segmentation. *CoRR*, abs/1807.10165.