# Studies on Intelligent Curation for the Korean Traditional Cultural Heritage

Jae-Ho Lee, Hee-Kwon Kim, Chan-Woo Park
*Creative Content Research Division*
*Electronics and Telecommunications Research Institute*
Daejeon, Korea
jhlee3@etri.re.kr, hkkim79@etri.re.kr, gamer@etri.re.kr

***Abstract*—In this paper, we introduce the necessary technologies to use the Korean traditional cultural heritage in immersive content by applying artificial intelligence technology. In fact, the data stored in museums has already been expanded to a huge amount. Recently, there are increasing efforts to convert such vast amounts of traditional cultural heritage image or text data into usable content through the analysis and connection of information using them. The main purpose of this study is to support the response to various content demands such as meta-verse and virtual reality for traditional cultural heritage of Korea. Representative technologies used in Korean traditional cultural heritage introduced in this study can be classified into artificial intelligence-based object detection and high-resolution conversion, and text analysis suitable for the characteristics of Korean traditional cultural heritage.**

***Index Terms*—traditional heritage, digital heritage, Korean heritage, super-resolution, Named Entity Recognition**

## I. Introduction

Recently, a lot of technologies for applying cultural heritage to immersive content are introduced and in progress under the influence of the spread of meta-verse, virtual reality, mixed reality, etc. Access to these new markets is an important factor in expanding the new role of museums and galleries in historical culture and art. However, to effectively compose cultural heritage-based immersive content, high-quality digitization of cultural heritage must be preceded. Additionally, it is necessary to develop AI visual search and search-based curation support technology and platform to manage vast digital heritage data.

With the development of artificial intelligence technology, researches to use the extensive cultural heritage data that have been continuously built until now as actual content and to activate it as a multifaceted approaches are gradually increasing, mainly in developed countries [1][2][3][4]. Meaning-based image search technologies that identify and compare major meanings between visual data, rather than simple keyword searches for vast cultural heritage data, are expanding to related application fields [5][6]. Artificial intelligence-based high-quality data conversion technologies are continuously being introduced [7][8].

Despite the continuous development of technology, the practical application of traditional cultural heritage in museums and exhibition halls has not yet been properly implemented. The reason for this inadequate is that most of the staff working at the museum are mainly composed of studies based on archaeology, and the main task of the museum until now has actually been the preservation and management of these relics. Due to the increase in user-experienced exhibitions around the world, a change of times that requires new contents fused with technology as a new task of museums is rapidly being pursued. However, this new approach must be promoted based on digital transformation to ensure its effectiveness and continuity.

Digital transformation of traditional cultural heritage includes simply digitizing data, and it is necessary to consider new data generation methods and standards according to technology and equipment. In addition, it is necessary to define methods for changing and improving existing data according to the use of content. These digitized data can be easily retrieved and, if necessary, the relationship between each relic's must be established to be reborn as information necessary for practical use. When the data analysis is completed in this way, the traditional cultural heritage management platform that provides an intuitive interface that can be easily accessed and used by curators working in the museum and continuously updates the development of technologies must be completed.

For this digital transformation, we are currently conducting research to develop a platform, including interfaces for creation, enhancement, analysis, search, relationship definition, and practical use of traditional cultural heritage data. In this paper, we present applied artificial intelligence technologies in our development of an effective Korean traditional cultural heritage management platform. Among the research in progress in our work, object detection technology such as animals, plants, and people in traditional Korean painting, high-resolution conversion technology of the previously photographed low-resolution Korean traditional cultural heritage image data, and recognition of related information in texts related to the Korean traditional cultural heritage. We also introduce the application contents related to this research and development in the following sections.

## II. Object Detection in Korean Traditional Paintings

The development of object detection technology in images is currently showing excellent results based on natural images.

However, to apply this object detection to Korean traditional painting, other works are needed. In particular, in the case of Korean painting, there is not much development of object detection technologies in painting using artificial intelligence. Therefore, in this study, from the preparation of the dataset to the ground truth dataset, the research was conducted in parallel with the preparations to complete the experimental environment.

### A. Preparation of Object Detection

For the compositing of the experiment, in this study, a GT data generation tool for learning about cultural heritage image data was first developed. This GT data generation is a basic function to verify the effectiveness of research results, and it was conducted in parallel with the study because there is no defined data set for the information of individual objects in Korean traditional painting.

Due to the characteristics of Korean traditional painting, experts with an understanding of traditional painting were employed as image annotators. In the tagging tool, various functions were added to make it easier for workers based on the traditional culture major to increase the ease of work. By supporting the progress bar that allows you to check the current progress and before and after images of the image to be worked on, the user can make tagging easier. Additionally, a shortcut function was supported to facilitate annotation work on image extraction properties. It supports the property status window that shows the image information (file name, image size, etc.) currently being worked on to the worker to check the work progress. Simple image processing and image editing functions (image resize, image filtering, etc.) are also provided for accurate bounding box selection. Figure 1 shows examples of traditional Korean paintings that were tagged using these tools.

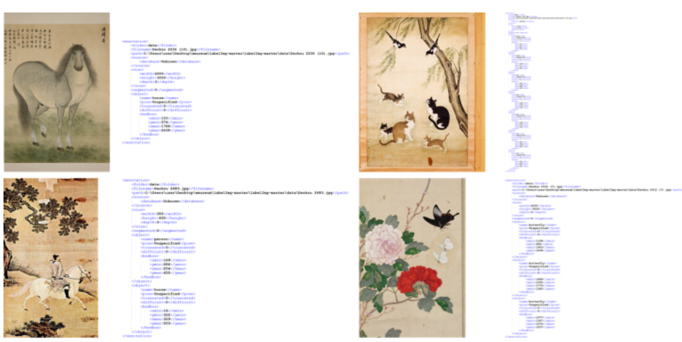

Fig. 1. Examples of annotated Korean traditional paintings.

### B. Examples of Object Detection

First, in this study, object detection was applied to images of Korean cultural heritage based on the previously learned deep learning model and the results were reviewed. The deep learning framework used was Tensor-Flow, and the COCO data-set was used as the training data, and the applied network was Faster-rcnn-inception-v2. Figure 2 shows the object detection results in Korean traditional painting using the existing model.

Fig. 2. Object detection based on the conventional model.

After attempting object detection using the existing learned model, Korean painting cultural properties are additionally learned, and additionally, transfer learning technology is applied to the object detection model to make better results. The deep learning framework used in this study is Tensor-Flow 1 and Keras, the training data are additionally learning the painting cultural heritage data-set to the existing COCO data-set, and the applied network is using RetinaNet.

RetinaNet is a structure that combines Focal Loss and Feature Pyramid Network and has an advantage in fast detection time and improves object detection performance in degradation problem of one stage detector. Using RetinaNet, we are confirming results that are advantageously applied to the detection of small objects included in painting cultural properties. Figure 3 is an overview of the deep learning model used in this study, and Figure 4 is the research result in the current situation, and it can be confirmed that the detection performance is improving little by little compared to the previous one.

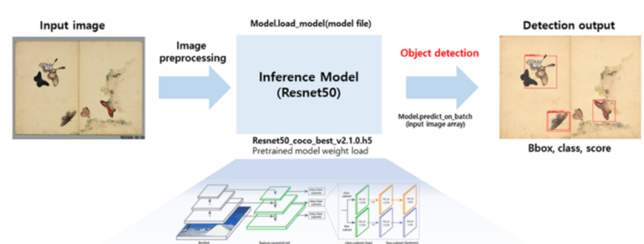


Fig. 3. Overview of deep learning model.

## III. Super-Resolution of the Old Korean Cultural Heritage Image

Super-resolution is an image processing technology that converts a low-resolution image into a high-resolution image. It can be used in various fields by allowing images that have

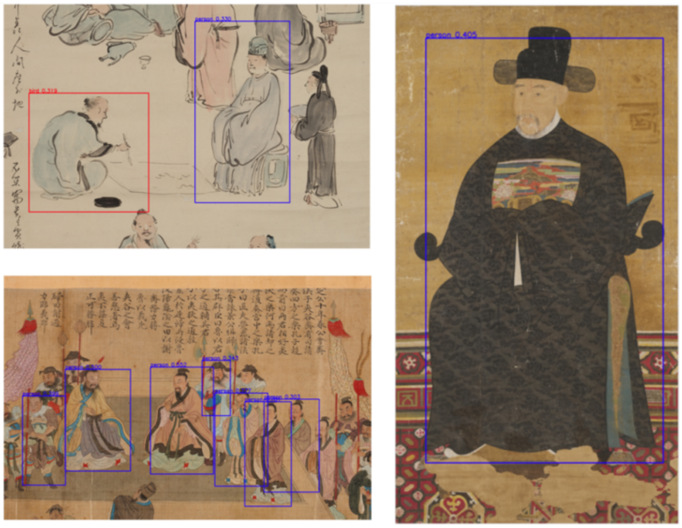

Fig. 4. Current Object detection Results.

been lost or stored in a low resolution due to transmission or storage to be viewed as clear, high-definition images. This image high-resolution technique started with simple interpolation methods, such as bilinear interpolation and bicubic interpolation and has developed into a form using machine learning. Before the advent of super-resolution technology using deep learning, non-deep learning machine learning methods such as SelfExSR [9] and SI [10] were typical. SI is a method for learning linear mapping from low-resolution to high-resolution images with a small amount of data. However, in most super-resolution research fields, since a large amount of data can be easily obtained, the performance is inevitably inferior to that of the deep learning method that uses a large amount of data. While deep learning is a convolutional neural network that can learn various features through non-linear mapping, the non-deep learning method has limitations because it learns linear mapping. Additionally, there is a clear limitation in that performance changes significantly even if hyper-parameters such as patch size change even slightly.

The super-resolution method using deep learning started with SRCNN ((Super-Resolution Convolutional Neural Network) [11], and many technologies such as VDSR (Accurate Image Super-Resolution Using Very Deep Convolutional Networks) [12], ESPCN ( Efficient Sub-Pixel Convolutional Neural Network) [13], SRResnet [14], and EDSR (Enhanced Deep Residual Networks for Single Image Super-Resolution) [15] have been introduced. Models suitable for super-resolution have been rapidly developed and have shown steady performance improvement. In particular, RCAN (Residual Channel Attention Networks) [16] has an uncomplicated structure and consistently shows high performance in benchmark tests, so it is an excellent model for natural images. In anticipation that it will show excellent results in cultural property images, this study was conducted by modifying the model structure of the previous studies.

Existing super-resolution research has been conducted in various ways, but mostly it has been done only on unspecified general natural images, and has not been verified in cultural heritage image data so far. In this respect, there is a distinct difference from natural images, so a specialized method is needed. This paper proposes a method for super-resolution 4x and 8x images of cultural assets using deep learning. The model structure inspired by RCAN We propose a patch extraction method that uses the characteristics of cultural assets images and a deep learning method that uses cultural heritage image data-sets in various ways. Therefore, it is expected that it will be helpful in research related to cultural heritage in the future.

### A. *Learning Model of super-resolution of the Korean cultural heritage images*

In this paper, the study was conducted by referring to the structure of the Residual in Residual Network. The network is composed of two ResGroups in a reduced form than in the paper. Each group has 10 ResBlocks, and each ResBlock is the type referenced in [16]. After adding the input image to the result after passing through the ResGroup, the 2x magnification module consisting of convolution, ReLU activation function, and pixel shuffle [13] is used to increase the image size to the desired size. At this time, in the case of a 4x magnification network, the above module was repeatedly configured 2 times, and in the case of an 8x magnification network, 3 times.

For the learning of cultural data properties, it was conducted in parallel with learning using general images. For general image learning, the DIV2K [17] dataset given in NTIRE2017 was used, and 2K quality images consist of 800 images for training and 100 images for testing. It is composed of general natural images not specific to either side, so it is easy to learn the characteristics of general images necessary for super-resolution.

We used the weight learned above to further learn by using cultural assets images for transfer learning. In this way, with the model learning the characteristics of natural images, additional learning was conducted using the cultural assets image dataset of Korean traditional cultural heritage. The advantage of this method is that it allows additional learning from cultural heritage image data while having a filter that can extract significant features for super-resolution from natural images.

### B. *Results of super-resolution*

In all experiments, L1 loss was used, optimization was performed using the Adam optimizer, and the initial learning rate was set to 1e-4. The batch size was 8, and patches cut to 32x32 were used. In the case of training with only one dataset, the test was conducted with the weight with the lowest valid loss while training up to 65 K iterations. In the case of transfer learning, the test was conducted with the weight with the lowest effective loss by additionally learning as much as 10 K repetitions from the cultural heritage image data with the best weight obtained from learning with DIV2K.

All the result images of the deep learning model show better results both numerically and visually than the results of bicubic interpolation. In particular, the average PSNR value for the resulting image is at least 1.1 dB at 4x super-resolution and at least 1.09 dB at 8x super-resolution, showing a significant performance improvement. Among the deep learning methods, numerically, the result image of the learning model through transfer learning is the best, but only subtle differences can be seen with the naked eye. Figure 5 shows the results of the proposed method, the linear interpolation method, and the ground thumb image.

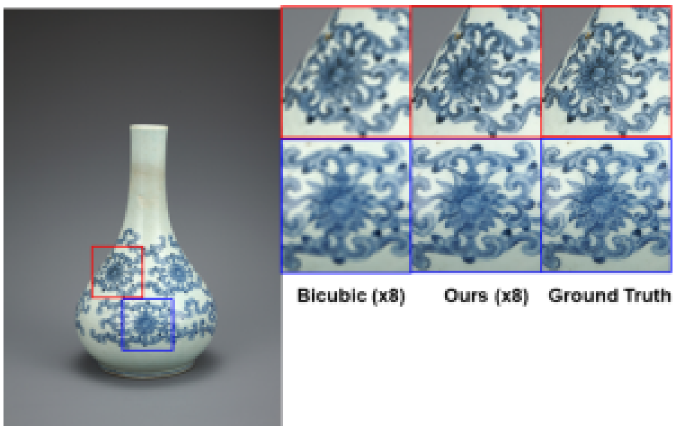


Fig. 5. Comparison of super-resolution result.

Most cultural heritage image data have a monotonous background and an object is located in the middle of the image. To this end, we newly construct a cultural property dataset, apply a method for extracting patches only from the central part, and propose a learning method using DIV2K, a natural image dataset, and cultural property data appropriately. As a result, compared to bicubic interpolation, about 1.25 dB at 4 times super-resolution and about 1.26 dB at 8 times super-resolution increased. Compared to the simple DIV2K learning method, performance increased by 0.06 dB at 4x magnification and 0.17 dB at 8x magnification. Figure 6 shows one of the super-resolution images of the proposed method.

## IV. Text Analysis in Korean Heritage Description Data

To extract formal and meaningful information from atypical traditional cultural heritage text data, this study applied a deep learning-based language model to learn semantic information and structural information of Korean sentences to develop an entity name recognition model and relationship extraction model.

### *A. Named Entity Recognition*

Named Entity Recognition (NER) technology is one of the basic technologies, and. it is also an important research area in terms of application. The process of learning with a language model optimized for traditional cultural heritage through post-training and fine-tuning based on the Korean


Fig. 6. Super-resolution result of the Korean cultural heritage image.

language model can further improve the model's performance. To recognize proper nouns in traditional culture, the above method is required, and to better understand traditional culture individual information in sentences, research and development is needed accordingly.

To understand the stylistic characteristics of domain-specific sentences and the complex understanding of the language, it is necessary to construct learning data, and fine-tuning of learning is required for more accurate proper noun entity recognition. NER is a technology for extracting individual names, such as person names (PS), place names (LC), and organization names (OG), which have unique meanings from a document and recognizing the extracted entity names.

Recently, NER in the field of natural language processing proceeds by pre-learning the LM through the encoder of the Bi-directional Transformer that can consider the context from a large corpus, and then apply it to the NER task. For the Pre-trained Language Model (PLM) to more effectively handle semantic and structural information of language from the text, model structures such as BERT, RoBERTa, and ELECTRA should be used [18][19][20].

This approach is because these models consider self-supervised learning objectives of language models such as Masked Language Modeling (MLM) and Next Sentence Prediction (NSP). For this purpose, the NSP technique suggested by BERT was excluded. In this study, the RoBERTa model, which adopts the masking pattern of the MLM technique as a dynamic method, and the Korean language model of ELECTRA, which introduces the replaced token detection technique that learns each sample data more effectively than the MLM technique, were applied.

Compared to the multilingual language model, the Korean version of the language model, pre-learned with a large-capacity Korean Wiki corpus, can better capture the features of the Korean language and is built as a lexicon that expresses the Korean language better in terms of vocabulary. For this study, we tried to enable the language model to capture the semantic and structural meaning of tokens by using various

Korean-based language models considering Korean, which has agglutinative language characteristics. In this study, not only the existing traditional vocabulary-based tokenization method, but also a study was conducted to extract objects according to the characteristics of the language model by using the Korean language model that can consider the Korean semantic form.

The proposed model represents the structure of entering Korean language models such as KoBERT, RoBERTa, and KoELECTRA through the tokenization process based on the segmentation method by receiving sentences as input. Classifies input tokens through a language model that extends the transformer encoder structure. In the fine-tuning process, we tried to improve the model performance by adding MLP Layers to consider in more detail the context that not be considered only with the language model. The used MLP Layers structure can be divided into Bi-LSTM, Bi-LSTM-CRF, and CRF. In addition to KoBERT, RoBERTa, and KoELECTRA, the language models used in this study include KorBERT, KorBERT-morph, and HanBERT.

### *B. Results of NER in the Korean cultural heritage description*

Using the published NER dataset, preprocessing for performance evaluation of each language model and performance evaluation in several models were performed. The corpus used in this study is the below [21].

- NIKL – The NER corpus distributed by NIKL with data of 3 million words (2 million written, 1 million spoken) includes 15 analysis markers to recognize the entity name boundary.
- AIR and NAVER NER Challenge – The data designed based on the CoNLL-2003 data format are for competition data, and the test set is not disclosed. Therefore, the verification data set will be used for testing in the evaluation of actual learning.
- KMOU-NER Corpus – The data constructed by Korea Maritime University is constructed by dividing approximately 24 K of ignition data into 10 classes.
- KLUE – data built to perform 8 Korean Natural Language Understanding (NLU) tasks, including 6 classes for Korean NER

The actual experimental result is a simulation in progress based on the recognition data of all corpus objects, and the entire dataset is divided into train, dev, and test 8:1:1, respectively, and early stopping is applied to prevent overfitting problems. The experimental model was KoBERT, KoELECTRA base v3 model of the Korean model, and in the case of the multilingual model, the experiment was conducted in xlm-roberta-based. So far, although the multilingual model has the largest model size, the KoBERT model pre-trained in Korean has recorded the best performance.

In this study, research is in progress to properly prove the effectiveness of constructing NER data in the traditional culture domain through the developed Korean model. The goal is to visualize the knowledge relationship graph through the NER model and relationship extraction results developed by understanding the linguistic characteristics of the Korean language from traditional cultural heritage text data. By using this knowledge, it is expected to provide a service to researchers and learners who want to use traditional cultural content to understand the subject matter they want by looking at the graph connected according to the conditions such as genre and era.

## V. Conclusions and Further Works

Recently, as the performance of computers is improved and the capacity of memory is increased, digitalization around the world has become a realistically feasible state. This phenomenon shows the possibility of digitization even for vast relics in museums and exhibition halls. Whereas digitization in the past was to create simple digital data, current digitization is applying it to the digital world or reality, such as virtual reality or digital twin. It is being changed for be used in the world.

In line with this trend, the digitization of relics in museums and exhibition halls around the world is changing for use for other applications such as virtual exhibition halls beyond the purpose of preserving the information on existing relics. To use it for other purposes, information must be put into the data stored. Such information is made through data analysis, but there is a clear limitation in the fact that a person performs analysis on a large amount of information and informatics it.

For this reason, many technologies for data processing using artificial intelligence technology are being studied. Technologies for analyzing information from numerous images or text and processing it into the desired form have been developed to an astonishing level. Unfortunately, however, these technologies are concentrated on universal images and natural language, so it is difficult to use them directly for analysis of old museums or ancient documents. For this reason, the actual development of artificial intelligence-based analysis and transformation of Korean traditional cultural heritage is still insufficient.

In this situation, this study studied the analysis and transformation technology of artificial intelligence-based cultural heritage data that can be used more by using the data of the actual museum and presenting its direction. In this ongoing study, the main purpose is to create an optimal usage model for the data of Korean traditional cultural heritage in the actual museum.

## Acknowledgment

This research is supported by the Ministry of Culture, Sports and Tourism and Korea Creative Content Agency (Project Number: R2020040045)

## References

[1] www.wikiart.org.
[2] www.europeana.eu.
[3] Gjorgji Srezoski, et al., "Omniart: Multi-task Deep Learning for Artistic Data Analysis," arXiv preprint arXiv:1708.00684, CoRR, Aug. 2017.
[4] www.museumnext.com.

[5] Nam Vo, et al., “Composing Text and Image for Image Retrieval – An Empirical Odyssey,” IEEE Conference on Computer Vision and Pattern Recognition (CVPR), pp. 6439-6448, California, America, June 2019.
[6] Marvin Teichmann et al., “Detect-to-Retrieve: Efficient Regional Aggregation for Image Search,” IEEE Conference on Computer Vision and Pattern Recognition (CVPR), pp. 5109-5118, California, America, June 2019
[7] [7] Tao Dai, et el., “Second-order Attention Network for Single Image Super-Resolution,” IEEE Conference on Computer Vision and Pattern Recognition (CVPR), pp. 11065-11074, California, America, June 2019.
[8] ] Bee Lim, et al., “Enhanced Deep Residual Networks for Single Image Super-Resolution,” NTIRE 2017 Workshop. pp.136-144, Hawaii, America, July 2017.
[9] J. B. Huang, A. Singh, and N. Ahuja, “Single image super-resolution from transformed self-exemplars,” IEEE Conference on Computer Vision and Pattern Recognition (CVPR), pages 5197–5206, 2015.
[10] J. Choi and M. Kim, ”Super-Interpolation With Edge-Orientation-Based Mapping Kernels for Low Complex 2× Upscaling,” IEEE Transactions on Image Processing, vol. 25, no. 1, pp. 469-483, January 2016, doi: 10.1109/TIP.2015.2507402.
[11] C. Dong, C. C. Loy, K.He, and X. Tang, “Learning a deep convolutional network for image super-resolution,” European Conference on Computer Vision (ECCV), pages 184–199. Springer, 2014.
[12] J. Kim, J. K. Lee and K. M. Lee, “Accurate image super-resolution using very deep convolutional networks,” IEEE Conference on Computer Vision and Pattern Recognition, pp. 1646-1654, June 2016.
[13] W. Shi, J. Caballero, F. Huszar, J. Totz, A. P. Aitken, R. Bishop, D. Rueckert, and Z. Wang, “Real-Time Single Image and Video Super-Resolution Using an Efficient Sub-Pixel Convolutional Neural Network,” IEEE Conference on Computer Vision and Pattern Recognition (CVPR), pages 1874–1883, 2016.
[14] C. Ledig, L. Theis, F. Huszar, J. Caballero, A. Cunningham, A. Acosta, A. Aitken, A. Tejani, J. Totz, Z. Wang, et al., “Photo-realistic single image super-resolution using a generative adversarial network,” arXiv preprint arXiv:1609.04802, 2016.
[15] B. Lim, S. Son, H. Kim, S. Nah, and K.M.Lee, “Enhanced Deep Residual Networks for Single Image Super-Resolution,” In CVPRW, 2017.
[16] Y. Zhang, K. Li, K. Li, L. Wang, B. Zhong, and Y. Fu, “Image Super-Resolution Using Very Deep Residual Channel Attention Networks,” In ECCV, 2018.
[17] E. Agustsson and R. Timofte, “Ntire 2017 challenge on single image su-per-resolution: Dataset and study,” IEEE Conference on Computer Vision and Pattern Recognition Workshop, pp. 1122-1131, July 2017.
[18] Lample, G., Ballesteros, M., Subramanian, S., Kawakami, K., and Dyer, C., ”Neural architectures for named entity recognition,” arXiv preprint arXiv:1603.01360, 2016.
[19] Nie, B., Ding, R., Xie, P., Huang, F., Qian, C., and Si, L., ”Knowledge-aware Named Entity Recognition with Alleviating Heterogeneity,” AAAI Conference on Artificial Intelligence, Vol. 35, No. 15, pp. 13595-13603, May 2021.
[20] Tran, Q., MacKinlay, A., and Yepes, A. J., ”Named entity recognition with stack residual lstm and trainable bias decoding,” arXiv preprint arXiv:1706.07598, 2017.
[21] Park, S., Moon, J., Kim, S., Cho, W. I., Han, J., Park, J., et al. ”KLUE: Korean Language Understanding Evaluation,” arXiv preprint arXiv:2105.09680, 2021.