

# ASCII Art Classification Model by Transfer Learning and Data Augmentation

Akira Fujisawa<sup>a,1</sup>, Kazuyuki Matsumoto<sup>b</sup>, Kazuki Ohta<sup>c</sup>, Minoru Yoshida<sup>b</sup>,  
Kenji Kita<sup>b</sup>

<sup>a</sup>*Aomori University, Japan*

<sup>b</sup>*Graduate School of Technology, Industrial and Social Sciences, Tokushima University, Japan*

<sup>c</sup>*Faculty of Engineering, Tokushima University, Japan*

**Abstract.** In this study, we propose an ASCII art category classification method based on transfer learning and data augmentation. ASCII art is a form of nonverbal expression that visually expresses emotions and intentions. While there are similar expressions such as emoticons and pictograms, most are either represented by a single character or are embedded in the statement as an inline expression. ASCII art is expressed in various styles, including dot art illustration and line art illustration. Basically, ASCII art can represent almost any object, and therefore the category of ASCII art is very diverse. Many existing image classification algorithms use color information; however, since most ASCII art is written in character sets, there is no color information available for categorization. We created an ASCII art category classifier using the grayscale edge image and the ASCII art image transformed from the image as a training image set. We also used VGG16, ResNet-50, Inception v3, and Xception's pre-trained networks to fine-tune our categorization. As a result of the experiment of fine tuning by VGG16 and data augmentation, an accuracy rate of 80% or more was obtained in the "human" category.

**Keywords.** ASCII art, transfer learning, fine tuning, data augmentation

## 1. Introduction

In recent years, so-called net framing, whereby critical comments have increased explosively due to comments and behaviors on the Web, has become a problem. From the viewpoint of risk management, the use of social media is often prohibited or censored by persons belonging to organizations such as companies. On the Web, on the other hand, electronic bulletin boards have existed as an online culture that flourished before the development of social media. In particular, anonymous bulletin boards allow users to avoid the risk of spreading personal information about themselves and they are easy to use because they allow users to write their intentions easily. As a result, the number of anonymous bulletin board users has exploded with the spread of the Internet.

The disadvantage of being anonymous, however, is that it leaves users prone to crime notifications, online bullying, and profanity. For example, bulletin boards often have users who engage in trolling on the Internet and they are often confusing. On the

---

<sup>1</sup> Corresponding Author: Akira Fujisawa, Faculty of Software and Information Technology, Aomori University. E-mail: fujisawa@aomori-u.ac.jp

operation side, it is costly to delete harmful write monitors such as net patrol settings. Therefore, although it may be possible to deal with this issue by filtering writing using stop words, doing so is often delayed until the problem becomes serious.

Some online trolls use ASCII art copy and paste. Since ASCII art contains visual information rather than regular text, it has a strong impact and can be said to have a powerful effect on first-time users.

In this research, we grasp the visual representation of ASCII art as an image rather than text information, and classify objects drawn in ASCII art into categories. Most ASCII art is unrestricted in its use and it rarely occurs in typical real-world scenes; therefore, it is rarely treated as a target for image processing. For this reason, there is no large-scale data set of ASCII art, and it is difficult to secure a sufficient amount of data for machine learning such as deep learning. Therefore, this study uses a deep learning model trained on various objects and tasks for transfer learning and the learning data are enriched. Instead of simply processing and padding the training data, it is recommended to use different kinds of image corpus and use more diverse images as training data.

By this research, the ASCII Art classification model by using DCNN would become easier to generate. These methods become to help to judge that text which appeared in Internet are trolls or not.

## **2. Related Research**

### *2.1. Studies on ASCII arts*

ASCII art research includes text extraction, automatic generation, and content classification. Suzuki [1] proposed a method that uses the REL (Run Length Encoding) algorithm to regard ASCII art, which comprises text data, as data of a specific length, and to extract ASCII art from text.

Fujisawa et al. [2] proposed a method of using image features to recognize pictograms or ASCII art that has more information than pictograms. Matsumoto et al. [3] proposed a method to classify input ASCII art by work category by focusing on the types and numbers of characters that appear in ASCII art and extracting each as a feature.

In a previous study, Fujisawa et al. [4] also proposed how to process a regular photo image and treat it as training data for ASCII art classification to create a deep learning classifier for ASCII art.

### *2.2. Transfer Learning*

Transfer learning is a method of applying a model trained to solve a specific task to solve a task that differs from the original target. If we do not have sufficient training data to perform deep learning, our goal is to supplement the missing training data by applying a model created on the basis of sufficient training data for various tasks. Nakayama [5] evaluated the effectiveness of transfer learning aimed at compensating for the lack of data, and in particular the diversity of deep learning models created using ImageNet [6].

In transfer learning, the weights of the original model are used as they are by adjusting only the final output layer of the pre-trained model; however, by adding another training data and retraining, the learning model can be expanded. There is also a method called fine adjustment of fine tuning, which allows us to obtain more accurate output for unknown input data that were not supported by the original model by using the data that correspond to the new task for training.

Hetsugi et al. [7] created a classifier using CNN (convolutional neural network) to classify excavated stone tools. To extend the learning model, they performed transfer learning based on an existing learning model created using ImageNet. This is a fine-tuning using existing high-performance classification models, similar to the classification problem built into this study. As in this study, we are studying data expansion at the same time as transfer learning, and are studying the methods necessary to create an efficient classifier for classification objects that do not have a large data set.

Another study that uses fine-tuned transfer learning is Yamada et al., [8] which aims to identify artifact areas from landscape images. In their research, the model that is the basis of the fine adjustment uses ImageNet data, while the model created based on ImageNet in the field of image processing using CNN is a reliable and effective way to recognize an object.

### *2.3. Domain Adaptation*

Domain adaptation is a kind of transfer learning method. When we create a classification model in CNN, we need training and test data, which are also called source and target domains, respectively. In a real classification task, the source and target domains rarely have the same environmental data, and the classification performance varies from domain to domain. Domain adaptation aims to improve the classification performance of the generated model by processing the source domains and reducing the differences between domains.

Ishii et al. [9] realized domain adaptation by executing conversion processing when extracting features from the source domain. Suzuki et al. [10] proposed a method for domain adaptation by simplifying the domain and reducing the image in the source domain. This study also proposes a way to treat the source domain to mimic the ASCII art of the target domain when creating a classifier based on different source domains for adaptive learning.

## **3. Proposed Method**

### *3.1. Data Augmentation*

In this research, data expansion is performed from the following two viewpoints:

1. Add processed data by adding rotation etc. to the original training data.
2. From the material that differs from the original training data, add the processed data like the original training data.

In both cases, the original image set is required. This study classifies ASCII art using a dataset that collects ASCII art by category. In step 1, the training data auto-expansion feature is used to increase the number of images by three for each image. The transformation parameters (such as rotation angle) are random, and each experiment performs different expansions.

In step 2, since data are added from image datasets with different characteristics, learning efficiency may be reduced depending on the data collection method, which may adversely affect accuracy. To keep the selection criteria consistent, we used the same search engine (such as Google search) to collect data for deployment and the same operator selected the image.

### 3.2. Fine Tuning with the Pre-trained Model

This study used four pre-trained deep learning models: VGG16 [11], ResNet-50 [12], Inception-v3 [13], and Xception [14].

VGG16 is a kind of deep convolutional neural network consisting of 13 layers of convolution and 3 layers of total connections. We achieved high performance in the 2014 ILSVRC (ImageNet Large Scale Visual Recognition Challenge) image classification task. ResNet-50 is a network consisting of 50 layers, which has a structure called shortcut connection (residual module) and which uses the output of residuals to avoid the problem that the gradient disappears when the layers become deeper.

Inception v3 is a convolutional neural network trained using over 1 million images in the ImageNet database. It has a structure called Inception module, which is a stack of small networks with a depth of 48 layers. Xception outperformed Inception v3 in the ImageNet image recognition task. A structure called separable convolution (spatial and channel convolution to reduce the number of parameters) is used.

By using the weights of the learned network as they are, it is possible to create a classifier with a certain level of performance, without additional learning of a large amount of data from the beginning, by additional learning of a small amount of data. For this reason, in this study, performing a transfer learning using a trained model and performing fine adjustment, which is a method of re-learning the layers near the output layer, cause overlearning and a decrease in accuracy due to lack of training data.

## 4. Evaluation experiment

### 4.1. Experimental Data

The image data used in the experiment comprised a web search engine for each of the five categories of “car,” “cat,” “fish,” “flower,” and “human.” Data were collected manually and there are four types of image data: “ASCII art (AA),” “photo/illustration (image),” and “images converted from pictures or illustrations to ASCII art using the automatic conversion tool (ImgToAA).” Table 1 shows an example. Of these, only ASCII art is classified in this study, while others are used as training data.

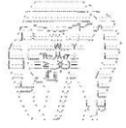
Table 2 shows a breakdown of the collected data. For learning, preprocessing involves cropping into squares with short sides, and for photos/illustrations, grayscale conversion and conversion to edge images.

### 4.2. Transfer Learning and Fine Tuning

In this study, the weights of the pre-trained network are used as they are, but the transfer characteristics are learned and fine-tuned because it is necessary to match the property and category type of data to the task of ASCII art classification.

Instead of removing the output layer of the trained network, we added a 2D Global Average Pooling (GAP) Layer, 256 units of fully connected layer, and a fully connected layer (using Softmax function as an activated function) into five categories as a new output layer. Here, in the case of transfer learning, the weights just before the GAP layer were not updated in learning. For fine tuning, the layer immediately before the GAP layer was updated.

**Table 1.** Example of Dataset

Category	Car	Cat	Fish	Flower	Human
AA					
Image					
Image to AA					

**Table 2** Breakdown of Collected Image Dataset

Category	AA	Image	ImgToAA	Total
Car	200	200	200	600
Cat	200	200	200	600
Fish	200	200	200	600
Flower	200	200	200	600
Human	200	200	200	600
Total	1000	1000	1000	3000

We used TensorFlow’s Keras module (version 2.2.0) to train and evaluate the model. We implemented RMSprop as an optimization method, set the learning rate to 0.001, and implemented early stop, which stops learning when the loss value of the loss function becomes less than 0.1.

### 4.3. Evaluation Method

In this experiment, four types of pre-training network were used to train the ASCII art category classification model by combining the seven types of training data shown in

Table 3 to evaluate performance. In addition, the data were randomly sorted by category and created in five parts. Cross-validation of the five parts calculates and evaluates Accuracy, Recall, Precision, and mean F-score. The equations for Accuracy, Recall, Precision, and F-score are shown in Equations 1 - 4. In this formula,  $M$  is the number of categories,  $C_i$  is the number of successful classifications of ASCII art for categories  $i$ ,  $N_i$  is the number of ASCII arts in category  $I$ , and  $O_i$  is the number of ASCII arts classified as category  $i$  by the classifier.

$$\text{Accuracy(\%)} = \frac{\text{Number of Correct AA}}{\text{Total number of AA}} \times 100 \tag{1}$$

$$\text{Recall(\%)} = 100 \times \frac{1}{M} \sum_{i=1}^M \frac{C_i}{N_i} \tag{2}$$

$$\text{Precision(\%)} = 100 \times \frac{1}{M} \sum_{i=1}^M \frac{C_i}{O_i} \tag{3}$$

$$\text{F-score} = \frac{\text{Recall} \times \text{Precision} \times 2}{\text{Recall} + \text{Precision}} \tag{4}$$

**Table 3** Combination of Training Data

No.	Combination of Training Data
1	AA
2	AA, Image, ImgToAA
3	AA, Image
4	AA, ImgToAA
5	Image
6	ImgToAA
7	Image, ImgToAA

## 5. Results and Discussion

### 5.1. Experimental Results

The experimental results (correct answer rate, F value) by transfer learning are shown in Tables 4 -7, and the experimental results (correct answer rate, F value) by fine adjustment are shown in Tables 8 - 11. Bold number means a best score of accuracy and f-measure in each category.

**Table 4** Experimental Result (Transfer Learning: VGG16)

Combination	1	2	3	4	5	6	7	
Accuracy	78.0	76.1	<b>79.6</b>	75.8	36.3	31.5	35.0	
F	Car	<b>89.6</b>	88.6	87.1	88.6	30.1	11.4	12.1
	Cat	62.3	58.4	<b>66.7</b>	59.5	2.9	25.4	5.4
	Fish	72.6	71.5	<b>75.5</b>	70.3	21.7	1.0	11.2
	Flower	72.7	72.5	<b>76.6</b>	69.7	51.8	36.0	53.0
	Human	<b>91.9</b>	88.2	91.0	90.2	46.0	46.7	47.1

**Table 5** Experimental Result (Transfer Learning: ResNet-50)

Combination	1	2	3	4	5	6	7	
Accuracy	<b>59.2</b>	52.1	53.8	57.9	28.7	23.2	31.5	
F	Car	76.9	73.4	<b>77.6</b>	76.8	40.3	26.3	48.7
	Cat	<b>42.6</b>	29.1	15.7	28.5	8.1	2.8	11.7
	Fish	<b>53.5</b>	42.2	42.3	45.8	21.8	2.0	9.5
	Flower	47.3	45.3	52.3	<b>55.3</b>	33.0	34.3	34.5
	Human	<b>71.0</b>	65.4	62.3	67.3	34.4	26.9	38.1

**Table 6** Experimental Result (Transfer Learning: Inception v3)

Combination	1	2	3	4	5	6	7	
Accuracy	<b>71.7</b>	68.6	69.7	69.6	41.8	29.2	42.4	
F	Car	81.2	81.3	80.8	<b>81.4</b>	52.6	19.7	42.9
	Cat	<b>54.2</b>	47.8	51.7	48.8	14.4	5.2	9.9
	Fish	<b>67.8</b>	62.7	67.0	65.4	32.0	19.0	39.2
	Flower	<b>69.6</b>	67.0	66.1	68.0	43.4	42.2	50.4
	Human	<b>84.9</b>	80.6	81.2	81.9	51.5	36.2	52.8

**Table 7** Experimental Result (Transfer Learning: Xception)

Combination	1	2	3	4	5	6	7	
Accuracy	<b>73</b>	68.6	70.2	69.4	44.1	31.6	42.8	
F	Car	<b>84.9</b>	80.1	82.0	80.7	53.9	21.3	46.5
	Cat	<b>55.2</b>	48.9	48.9	50.4	15.7	4.0	5.4
	Fish	<b>66.7</b>	63.9	65.4	66.3	36.7	19.1	39.0
	Flower	<b>72.4</b>	68.0	69.4	66.3	46.7	43.5	50.1
	Human	<b>84.4</b>	79.1	82.4	80.4	53.7	42.7	53.5

**Table 8** Experimental Result (Fine Tuning: VGG16)

Combination	1	2	3	4	5	6	7	
Accuracy	73.8	59.4	74.8	<b>75.2</b>	42.3	31.4	46.1	
F	Car	<b>71.5</b>	69.3	67.4	68.0	69.6	65.4	66.4
	Cat	<b>46.6</b>	44.8	43.5	43.3	44.1	41.7	41.7
	Fish	<b>55.1</b>	53.2	51.5	52.1	52.9	49.9	50.2
	Flower	<b>57.6</b>	56.2	55.0	55.5	56.1	53.3	54.5
	Human	<b>71.8</b>	69.7	68.4	68.8	68.8	66.8	67.5

**Table 9** Experimental Result (Fine Tuning: ResNet-50)

Combination	1	2	3	4	5	6	7	
Accuracy	40.1	40.2	<b>49.4</b>	42.8	32.1	24.9	35.1	
F	Car	<b>56.7</b>	55.4	54.1	54.1	55.6	52.3	53.3
	Cat	<b>23.7</b>	23.6	21.7	22.1	22.4	21.1	21.1
	Fish	<b>37.6</b>	36.9	35.3	35.9	36.1	34.5	35.2
	Flower	<b>42.7</b>	41.8	41.9	40.9	42.3	41.4	41.2
	Human	38.9	<b>41.6</b>	38.7	39.6	38.4	37.3	39.4

**Table 10** Experimental Result (Fine Tuning: Inception v3)

Combination	1	2	3	4	5	6	7	
Accuracy	57.4	47.1	57.5	<b>59.7</b>	38.4	23.4	35.9	
F	Car	<b>58.0</b>	56.3	55.2	55.1	56.6	53.7	53.2
	Cat	<b>34.4</b>	33.3	32.6	32.7	33.1	31.6	31.6
	Fish	<b>47.3</b>	46.3	45.3	45.4	46.6	44.1	44.5
	Flower	<b>47.5</b>	46.8	45.9	46.3	46.7	45.0	45.6
	Human	<b>53.9</b>	52.3	50.8	51.2	52.4	49.2	50.2

**Table 11** Experimental Result (Fine Tuning: Xception)

Combination	1	2	3	4	5	6	7
Accuracy	<b>58.9</b>	45.4	57.3	53.4	33.6	29.9	30.1
Car	<b>57.5</b>	54.7	54.8	54.2	55.9	52.6	52.6
Cat	<b>31.6</b>	30.3	29.3	28.5	29.9	28.3	28.0
F	<b>40.2</b>	37.9	37.9	37.0	38.4	35.8	36.4
Flower	<b>43.8</b>	43.4	43.1	43.4	42.4	43.0	42.9
Human	<b>54.5</b>	52.6	52.7	51.8	52.7	51.6	51.2

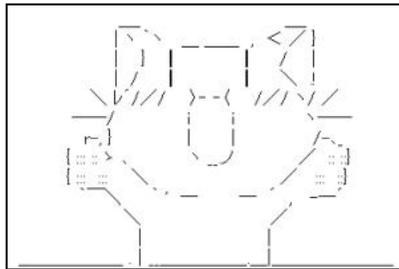
In the transfer learning results, the maximum correct answer rate was 79.6% when the combination of VGG16 and learning data (3:AA, Image) was obtained. As a result of fine tuning, the correct answer rate and F value were lower than those of transfer learning.

When fine tuning was performed, the combination of learning data with VGG16 (4:AA, ImgToAA) gave the highest correct answer rate. In both transfer learning and fine tuning, the combination of learning data (7: Image, ImgToAA) shows the lowest correct answer rate, while the correct answer rate is the same when using Image and ImgToAA individually as the learning data.

Furthermore, the F-values of “human” and “cars” tended to be high. One cause of this may be that the dataset contained many similar pieces of ASCII art in the automobile category.

## 5.2. Discussion

Comparing all the results, regardless of the model used, the F value is the lowest value for cat ASCII art in all five categories. This is because most of the AA used as learning data were deformed. When we expanded the data, it was difficult to find common points in the learning data and feature extraction failed. Figure 1 is an example of a deformed “cat” AA.

**Figure. 1** Deformed cat’s ASCII art

This time, the image was converted to ASCII art for domain adaptation. Many of the original images included the background, and when converted to ASCII art, the background information was also included in the conversion.

Figure 2 shows an example of an image before and after conversion of the category “flower.” In this example, the background of the original image (pot clay and shadow) is also represented as a component of ASCII art, so the converted ASCII art itself has a large amount of information.



**Figure 2** Photo of flower and Image, which converted into AA from the photo of a flower

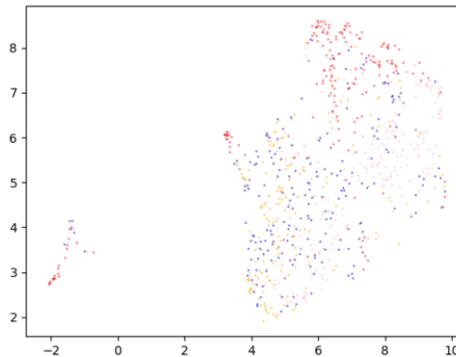
Figure 3 shows an example of the image before and after the conversion of the category “car.” In this example, the original image has a solid background, so converting it to ASCII art will only properly convert the target object to ASCII art. As a result of the experiment, many images with a composition with little background were seen even among “human” with high F values on average. Therefore, when converting an image to ASCII art, it is considered that the classification performance is greatly affected by the influence of the background. Therefore, when transforming the original image for domain adaptation, in addition to monochromaticization by grayscale conversion, preprocessing and region extraction are performed, and only the target object is extracted to determine the matching rate between domains.

### 5.3. Visualizations

To explore how the learning network used in the experiment extracts features, we used UMAP to visualize in two dimensions [15]. Figure 4 shows the result of inputting ASCII art into VGG16 and visualizing the output of the GAP layer. Categories are displayed in color. From this figure, it can be seen that the ASCII art features are not grouped into categories and have a wide distribution.

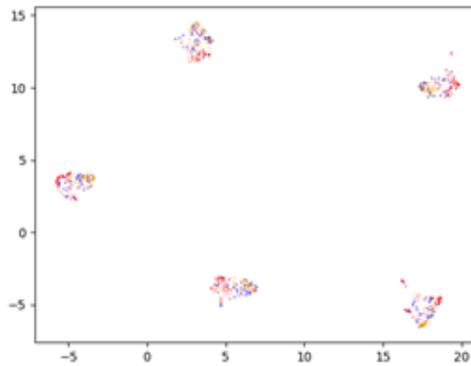


**Figure 3** Photo of a car and Image which converted into AA from the photo of a car



**Figure 4** Visualization by UMAP (AA, VGG16)

After tweaking and learning only the ASCII art, these features can be visualized in UMAP as shown in Figure 5.



**Figure 5** Visualization by UMAP (Fine Tuned)

From this result, it can be seen that a highly accurate classifier cannot be learned. This is because the ASCII ART data set collected in this study may have been over-learned due to lack of diversity and lack of data only by data expansion. In fact, there was a problem that the value of verification loss in cross validation did not decrease even if learning progressed.

For these reasons, expansion of the amount of learning data is vital, while the edge feature is important for ASCII art that does not have color information. Therefore, both training and evaluation perform edge enhancement and preprocessing. It may be necessary to clarify the shade to improve accuracy.

## 6. Conclusion

This study (1) data expansion and (2) trained models to compensate for the lack of training data for ASCII art categorization. In data expansion, the volume of ASCII art data collected is small, so in addition to conventional data expansion methods such as rotation, the same processing as ASCII art is performed from image corpora (photos, illustrations) that have different characteristics from ASCII. We also proposed a method to increase learning data by applying art.

We used multiple neural network architectures as trained models. As a result of the evaluation experiment, transfer learning and fine adjustment was performed. When transfer learning was performed by including ASCII art in the learning data, a high F value was obtained as a whole. The highest correct answer rate (79.6%) was obtained when using the learning images obtained by converting images and illustrations into edge images in addition to ASCII art.

Furthermore, most of the ASCII art in the dataset created this time contained many modified images. However, the illustrations other than the photos that were added when the data were developed were almost untransformed and realistic. This difference in data quality may have affected the loss of accuracy. To improve the classification accuracy, it is necessary not only to prepare all kinds of image data but also to select an image according to the tendency of the ASCII art to be classified.

In future work, we will perform object detection to the training image as preliminary preparation. By this, we try to remove the evil effect of background of test data. By using the test image that were removed background as training data, we would become able to generate the good performance ASCII Art classification model.

## Acknowledgments

This work was supported by JSPS KAKENHI Grant Numbers JP20K12027, JP18K11549.

## References

- [1] Tetsuya Suzuki: Introduction of N-gram into a run-length encoding based ASCII art extraction method, Proceedings of 15th International Conference on Current Trends in Web Engineering, ICWE 2015 Workshops NLPIT, PEWET, SoWEMine, Lecture Notes in Computer Science, 9396, pp. 28-39, 2015.
- [2] Akira Fujisawa, Kazuyuki Matsumoto, Noriyuki Okumura, Minoru Yoshida, and Kenji Kita : Challenge to ASCII Art – An Image Feature-Based Approach –, Journal of JSAI, 32(3), pp.364-371, 2017.
- [3] Kazuyuki Matsumoto, Akira Fujisawa, Minoru Yoshida and Kenji Kita: ASCII Art Classification based on Deep Neural Networks Using Image Feature of Characters, Journal of Software, Vol.13, No.10, pp. 559-572, 2018.
- [4] Akira Fujisawa, Kazuyuki Matsumoto, Kazuki Ohta, Minoru Yoshida and Kenji Kita: ASCII Art Category Classification based on Deep Convolutional Neural Networks, Proceedings of the 5th IEEE International Conference on Cloud Computing and Intelligence Systems (CCIS), pp.345-349, 2018.
- [5] Hideki Nakayama, Image Feature Extraction and Transfer Learning Using Deep Convolutional Neural Networks, IEICE Technical Report, vol. 115, no. 146, SP2015-45, pp. 55-59, 2015.
- [6] ImageNet: <http://www.image-net.org/> (accessed 2020-07-13).
- [7] Michihiro Hetsugi, and Fumihiko Yamagushi, Excavator Classification model by CNN and Data Augmentation about training model., Proc.82nd National Convention of IPSJ, 2020, 1, pp.553-554, 2020.
- [8] Sohei Yamada, and Yuko Osana, Improving Accuracy of Artifact Region Identification using Convolutional Neural Networks with Fine Turing, Proc.82nd National Convention of IPSJ, 2020, 1, pp.295-296, 2020.
- [9] Masato Ishii, and Atsushi Sato, Joint Optimization of Feature Transform and Instance Weighting for Domain Adaptation, The 30th Annual Conference of the Japanese Society for Artificial Intelligence, 2016, 4K1-1in2, pp.1-4, 2016.
- [10] Ryosuke Suzuki, Tadachika Ozono, and Toramatsu Shintani, On a Domain Adaptation Method Based on Image Simplification for Machine Learning with Synthetic Data, The 34th Annual Conference of the Japanese Society for Artificial Intelligence, 2020 , 2H5-GS-13-03, pp.1-2, 2020.
- [11] Karen Simonyan and Andrew Zisserman: Very Deep Convolutional Networks for Large-Scale Image Recognition, Proceedings of ICLR 2015.
- [12] Kaiming He, Xiangyu Zhang, Shaoqing Ren and Jian Sun: Deep Residual Learning for Image Recognition, Proceedings of 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2016.
- [13] Christian Szegedy, Vincent Vanhoucke, Sergey Ioffe, Jonathon Shlens, Zbigniew Wojna: Rethinking the Inception Architecture for Computer Vision, Proceedings of IEEE Conference on Computer Vision and Pattern Recognition, 2016.
- [14] François Chollet: Xception: Deep Learning with Depthwise Separable Convolutions, Proceedings of 2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2017.
- [15] Leland McInnes, John Healy, James Melville: UMAP: Uniform Manifold Approximation and Projection for Dimension Reduction, The Journal of Open Source Software Vol. 3, No. 29, 861, 2018.