CLASSIFICATION OF OIL PAINTING USING MACHINE LEARNING WITH VISUALIZED DEPTH INFORMATION

Jihoon Kim[1], Ji Young Jun[1], Minki Hong[2], Hyeseung Shim[1], Jaehong Ahn[1]*

[1] Graduate School of Culture Technology
Korea Advanced Institute of Science Technology (KAIST)
Daejeon, Republic of Korea-(kjih0314, hs.shim, jiyoungjun, ahnjh)@kaist.ac.kr
[2] Culture Technology Research Institute
Korea Advanced Institute of Science Technology (KAIST)
Daejeon, Republic of Korea-minki.hong@kaist.ac.kr

Commission II, WG II/8

KEY WORDS: Machine Learning, Visualized Depth Information, RTI, Painting Analysis, Artist Classification

ABSTRACT:

In the past few decades, a number of scholars studied painting classification based on image processing or computer vision technologies. Further, as the machine learning technology rapidly developed, painting classification using machine learning has been carried out. However, due to the lack of information about brushstrokes in the photograph, typical models cannot use more precise information of the painters painting style. We hypothesized that the visualized depth information of brushstroke is effective to improve the accuracy of the machine learning model for painting classification. This study proposes a new data utilization approach in machine learning with Reflectance Transformation Imaging (RTI) images, which maximizes the visualization of a three-dimensional shape of brushstrokes. Certain artist's unique brushstrokes can be revealed in RTI images, which are difficult to obtain with regular photographs. If these new types of images are applied as data to train in with the machine learning model, classification would be conducted including not only the shape of the color but also the depth information. We used the Convolution Neural Network (CNN), a model optimized for image classification, using the VGG-16, ResNet-50, and DenseNet-121 architectures. We conducted a two-stage experiment using the works of two Korean artists. In the first experiment, we obtained a key part of the painting from RTI data and photographic data. In the second experiment on the second artists work, a larger quantity of data are acquired, and the whole part of the artwork was captured. The result showed that RTI-trained model brought higher accuracy than Non-RTI trained model. In this paper, we propose a method which uses machine learning and RTI technology to analyze and classify paintings more precisely to verify our hypothesis.

1. INTRODUCTION

In recent years, as the art database expands rapidly, automatic painting classification based on color and morphological features have been gaining much attention (Berns, 2001) (Barni et al., 2005). Various computational methods are used to allow experts to analyze and evaluate different characteristics of paintings in a quantitative way which are difficult to be judged by the naked eye (Berezhnoy et al., 2005) (Berezhnoy et al., 2007). In order to classify paintings, the Image processing technique has been studied to extract the characteristics of paintings such as the shape, directions, and the pattern of brushstrokes (Li et al., 2011). In addition to image processing, machine learning techniques have been actively studied. Painting classification studies are proceeding from the classical Support Vector Machine (SVM) method to Convolution Neural Network (CNN), which is optimized for image learning (Cortes, Vapnik, 1995) (Krizhevsky et al., 2012).

Artists' brushstroke is one of the characteristics which reflects their unique painting styles (Li et al., 2011). It contains a combination of color, pattern, and texture, and the collaboration of each characteristic element shows much information in oil paintings with pigments (Berezhnoy et al., 2009) (Johnson et al., 2008). Nevertheless, the shape and thickness of the brushstrokes are not represented clearly in typical photographs. It

is a significant drawback to the CNN model, which learns the particular properties from the training image set (Krizhevsky et al., 2012) (Zeiler, Fergus, 2014). In order to overcome the limitation, the size of brushstrokes can be measured and used for analysis with technologies which can capture three-dimensional (3D) geometric data such as 3D scanning or photogrammetry (Elkhuizen et al., 2014). However, previous studies show that it is difficult to extract the depth information of the individual brushstrokes (Breuckmann, 2011) (Abate et al., 2014). We also confirmed through experiments that it is time-consuming and hard to acquire 3D data of brushstrokes with 3D scanning or photogrammetry. Our hypothesis is that if the depth information of brushstroke is visualized and used as a training image set, it is effective to improve the accuracy of the machine learning model.

We investigate three machine learning architecture for the painting classification task, which uses the RTI images as a training image set. We also present the result of the accuracy and visualize learning process. The goal of this paper is to verify our hypothesis by comparing the results when RTI and Non-RTI images are used for painting classification with three different machine learning architectures.

In order to capture and visualize the depth of brushstroke, we use Reflectance Transformation Imaging (RTI) technology. RTI is a computational technique that captures the surface shape and

---

*Corresponding author

color of the subject and allows to re-illuminate the subject in any direction (Cultural Heritage Imaging, 2019). It captures the painting's surface with very high resolution and extracts depth information from the captured images (Cultural Heritage Imaging, 2019). The enhancement functions of RTI can reveal each brushstroke that is not disclosed under a direct empirical examination of the physical object (Cultural Heritage Imaging, 2019). This study proposes a new approach to use visualized depth information of paintings as additional input data for machine learning algorithms. It is possible to use an existing machine learning architecture that uses two-dimensional images as the input data.

2. RELATED WORKS

As the technology of computer vision advanced, high-resolution images of paintings have been used for analysis. There are several studies to extract the features of fine art using digital image processing and apply them to classification (Barni et al., 2005). In the early stages of painting classification using image processing, color and texture processing techniques such as complementary colors and Gabor filtering based on RGB values of a picture were studied (Berezhnoy et al., 2007) (Berezhnoy et al., 2005). In the case of the RGB-value-based research, there is a limit that the result is different according to the quality of the input data for the image processing. Berezhnoy et al. extract the artist's brushstroke from the paintings automatically (Berezhnoy et al., 2009). Johnson et al. classify the pictures using wavelets and brush strokes to extract the features of the artist (Johnson et al., 2008).Those researches use only the plane shape information of the brushstrokes, not the three-dimensional information.

With the rapid development of computing power, machine learning has been applied to the classification of artwork. Support Vector Machine (SVM), which has been extensively used in the field of machine learning, was used (Cortes, Vapnik, 1995). Arora et al. classified paintings into 7 genres- Renaissance, Baroque, Impressionism, Cubism, Abstract, Expressionism, and Pop art, using SVM (Arora, Elgammal, 2012). Khan et al. classified artists and styles of works in digitized databases to help manage them focus on both artist and style categorization problems (Khan et al., 2014). These painting classification studies using SVM have attempted to improve the classification accuracy of various characteristics of paintings. In the case of machine learning, much data is necessary. In response to the demand, in 2014, Khan et al. released 'Painting-91' data set consisting of 4,266 pictures of 91 different painters (Khan et al., 2014). Using the 'Painting-91' dataset, a deep-learning model Convolutional Neural Networks (CNN) which specializes in image classification, was attempted to apply artwork classification (Khan et al., 2014) (Krizhevsky et al., 2012). Folego et al. find the way of using the patch with the highest confidence scores turned out to be better than the traditional voting method (Folego et al., 2016). Nanni et al. studied cases in which artistic style, artist, and architectural style classification are performed using various features in several layers of CNN instead of using features extracted from only the top layer of CNN (Nanni et al., 2017). Peng et al. performed the figure analysis using plural CNNs in order to extract multi-scale features (Peng, Chen, 2015). Numerous studies have transformed the CNN model to be optimized for artwork data. However, there is no research on how to enhance classification accuracy by acquiring and using rich data that has more information about the painter's style.

RTI has been applied to many useful applications on a wide range of cultural heritage domain such as condition monitoring, treatment documenting, and surface analyzing. With its mathematical enhancement function, observing features interactively that are difficult to see with naked eyes is possible (Manrique Tamayo et al., 2013) (Giachetti et al., 2017) (Clarricoates, Kotoula, 2019). In recent research, Pamart, A., et al. developed an integrated tool to cross-reference qualitative depth information of RTI and quantitative depth information of photogrammetry (Pamart et al., 2019). They suggest that qualitative depth information of RTI should be supplemented when the purpose is a precise analysis. Ponchio et al. propose that qualitative depth information can be used to classification problems based on machine learning (Ponchio et al., 2018). They studied the automatic classification of cuneiform using images captured by the RTI in the machine learning algorithm. They mainly use partial information, especially the surface normal vector at rapidly changing regions for the analysis of cuneiform. Unlike cuneiform, oil paintings have depth variance distributed in the whole part. This research considered the whole region of the RTI image in order to improve classification accuracy.

3. METHODS

As seen in Figure 1, it is necessary to build an RTI image for extracting data from oil paintings which depicts the brushstrokes cleary with a proper rendering mode. The image set is used as input data to the Residual Network 50(ResNet) architecture of Convolutional Neural Network (CNN) (He et al., 2016) (Krizhevsky et al., 2012). The optimizer uses Adam(Adaptive Moment Estimation), and the ResNet architecture is customized to 224x224 pixel size, which is the size of our data (Kingma, Ba, 2014).

3.1 Dataset Acquisition by RTI from Paintings

In the first stage, to acquire the image dataset, RTI images are built for each part of the painting. When the white light LEDs of the dome illuminate one by one, photos are taken. Then a single Polynomial Texture Mapping (PTM) file is built by combining 80 photos using the RTI Builder (Cultural Heritage Imaging Inc., USA). It can depict clear brushstrokes with a proper rendering mode. The PTM was rendered with static multi-light, which applies optimal light direction on each tile of the image to maximize the contrast of shading. Unlike other rendering methods that manually control the light direction, static multi-light rendering enhances detail information without manipulating controller. It makes a combination of racking light images with only controlling sharpness of images.

3.2 Applying Machine Learning

In the second stage, the two different image datasets from RTI are applied to CNN architectures. We implemented the model in Keras, and an open-source deep-learning package based on Tensorflow (Chollet et al., 2015). All the CNN architectures that we used received images of 224x224 pixels as input data. Therefore, we crop the RTI data and Non-RTI data obtained from the first stage respect of the input size. 20% of all datasets except the test set are randomly extracted and used as a validation set, and 80% is used as a train set. The number of CNN model layers, the fully connected layer, the output layer, the optimizer, and the loss function is set equal for both data sets. Parameters for fine-tuning the CNN model such as batch size, learning rate, and epoch are optimized for each data set because
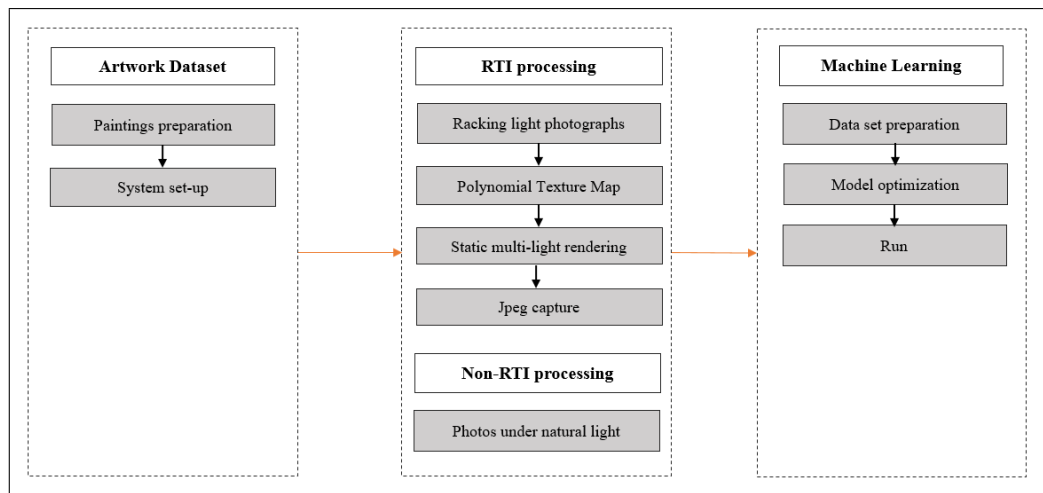
Figure 1. Pipeline of this research for oil painting classification using depth information.

there is a difference between RTI data and Non-RTI data containing depth information. Finally, CNN layer visualization is conducted to investigate the learning process depending on the characteristics of the data set.

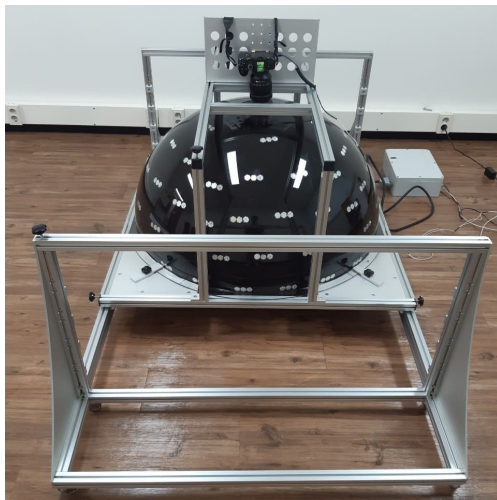## 4. EXPERIMENTS OF PAINTING ANALYSIS USING DEPTH INFORMATION

Figure 2. RTI Dome Setting with the Railed Installation

We used a dome RTI system for the experiments as shown in Figure 2. The dome is a plastic hemisphere with 1m in diameter and an external frame with rail and wheels for easy transportation of the system over a target object. A camera is fixed at the top of the dome, facing the center of the dome. With the 50mm lens, the camera captures 30x50cm of the paintings at once. It contains three different types of LED lights, which are white light, UV and IR, and 80 lights per each type. In this research, we only used white light LEDs. The position of LEDs was calibrated to set up the light position file before capturing the paintings. Two experiments were designed to verify our hypothesis: Experiment 1 is a pilot study to see proof that the depth-visualized images improve the result of painting classification with machine learning, and in experiment 2, the number of training data was increased and the classification classes were enlarged. The results were also compared when three different CNN architectures and were used to see the accuracy of classification according to CNN's layer depth.

4.1 Experiment 1

As shown in Figure 3, we tested eight oil paintings by Jiho Lee as our datasets. Paintings used in the experiment were grouped into two groups according to her painting period. Group A shows a landscape view that was created in 2013, while Group B shows a more abstract environmental view using dark and thin brush strokes, which were her recent works in 2017.
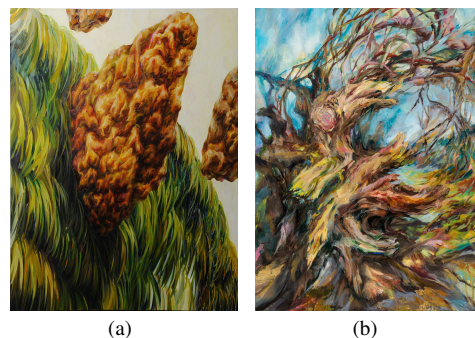

(a) (b)

Figure 3. (a) Group A (b) Group B
The different groups of paintings by Jiho Lee.

4.1.1 Dataset of Experiment 1
PTM files were built by using a dome-RTI system with the Nikon D850 DSLR camera. Due to the limitation of the domes size, we captured several different parts with 30x50cm size. We choose the parts where the brushstrokes were well revealed. RTI image set was captured in jpeg format of 6192 x 4128 resolution from the PTM files under static multi light rendering. The Non-RTI images were captured at the same spot with the same resolution under the natural light.

4.1.2 Categorization the dataset
Among various classifications, binary classification was proceeded because the style of paintings was divided into two groups. We focused on the areas where the brushstrokes were prominent because th e goal was to investigate the difference in the classification accuracy between RTI data and Non-RTI data, and the learning process in CNN layer. About 6.8 partial artwork
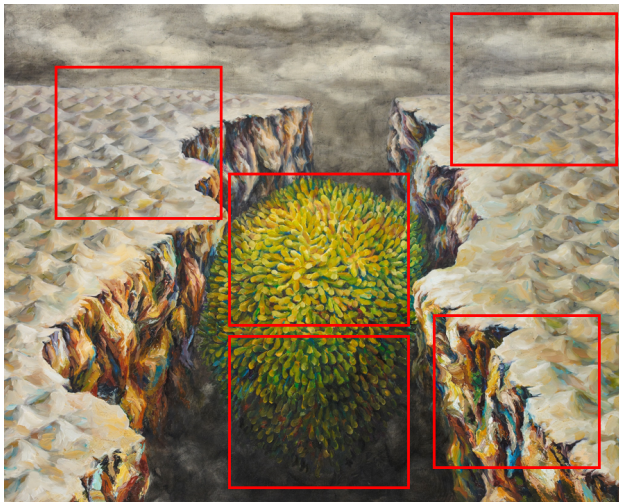
Figure 4. A painting by Jiho Lee. In Experiment 1, we captured key parts of each painting, not the whole part.

images per work in average were acquired from both datasets [Figure 4].

One image from each group was separated for the use as a test set. The CNN architecture used in Experiment 1 was ResNet-50 (He et al., 2016). It took a 224x224 pixel size image as input. When resizing the full size of the artwork directly, a large number of features that can be learned in CNN, such as stroke, overall shape, and RGB value can be distorted or disappear. Due to the reason, both RTI data and Non-RTI data were cropped into 224x224 pixels, and 750 to 850 cropped images per image were acquired. 80% of the cropped data was used as a train set to learn the CNN model, and 20% was used as a data set for validation of the model accuracy. The data images were randomly separated when they were divided into a test set and validation set. For RTI data, 18,434 data were chosen for the train set, 4,608 data for the validation set, and 1,673 data for the test set. For Non-RTI data, 17,667 data sets were chosen for the train set, 4,416 data for the validation set, and 1,512 data for the test set.

The pre-trained model uses ImageNet, which divides 1.2 million pieces of data into 1000 categories (Deng et al., 2009). Since the machine learning model used in Experiment 1 should perform binary classification, we used sigmoid as an activation function and binary cross-entropy as a loss function in the output layer. The optimizer used Adam, and the fully connected layer was customized for ResNet-50 (Kingma, Ba, 2014). Layer visualization was conducted by activating the first, the second, and the third layers of the convolution layer. As the convolution layer moves to the upper layer, the information about the visual content of the image gradually decreases, and the information about the image class gradually increases (Zeiler, Fergus, 2014) (Yosinski et al., 2015). Besides, it becomes increasingly more abstract and more difficult to understand from a human perspective as it goes to the upper layers of the layer (Zeiler, Fergus, 2014) (Yosinski et al., 2015). Therefore, in this paper, only two layers that can be understood by human perception are shown as a result.

4.1.3 Results & Discussion
The model that was trained with the RTI data (RTI-data model) showed the accuracy of 87.43% and the model trained with the Non-RTI data (Non-RTI-data model) showed 82.95%. The
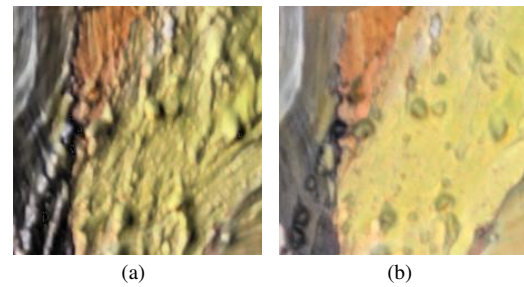

(a)   (b)

Figure 5. (a) Cropped data of RTI (b) Cropped data of Non-RTI

result shows that the accuracy of the model which learned the depth-visualized data is 4.48% higher.

(a) Visualization of first convolution layer using (a) in Figure 5.

(b) Visualization of first convolution layer using (b) in Figure 5.

(c) Visualization of second convolution layer using (a) in Figure 5.

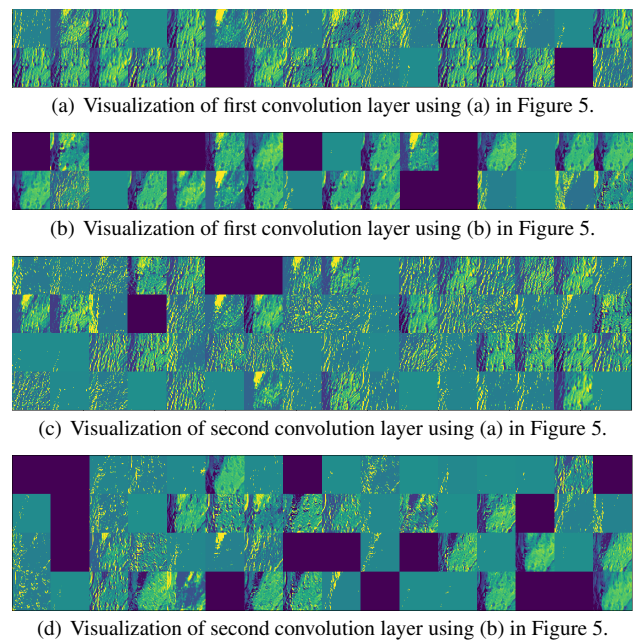(d) Visualization of second convolution layer using (b) in Figure 5.

Figure 6. The results of visualization from convolution layers

When the convolution layer is activated, there is a clear difference between the data, as shown in Figure 6 using the data [Figure 5]. The Non-RTI-data model learned the shape such as line or color, on the other hand, the RTI-data model learned the depth information in addition to the shape. Moreover, the number of deactivated filters in the CNN model was less in the Non-RTI-data model. There is an empty filter represented in Fig.6 as black. This filter appears in both RTI data and Non-RTI data. It shows that the deeper the layer, the more deactivate the filter appeared. This indicates that the pattern encoded in the filter does not appear in the input image. However, when comparing the same layers, the RTI data is less deactivated than the Non-RTI data. In other words, this means that RTI data is rich in information, and CNN is learning the properties appropriately. When the learning process is visualized by layer, it is well observed that RTI data actively learns various data more than Non-RTI data.

4.2 Experiment 2

In experiment 2, we tested 14 paintings by Wi Hyeok Son. There are three different styles painted in the same period (2019). We categorized into three groups by different brush strokes styles for the analysis as seen in Figure 7. Group A' includes paintings with wet and clear brushstrokes, which are more vivid and bold

	VGG16_NonRTI	VGG16_RTI	ResNet50_Non-RTI	ResNet50_RTI	DenseNet121_Non-RTI	DenseNet121_RTI
Group A'	**65.79**	53.80	**55.53**	46.76	**59.55**	40.22
Group B'	10.99	**30.35**	15.90	**25.57**	9.67	**30.62**
Group C'	42.92	**52.82**	45.40	**53.16**	**50.97**	39.61

Table 1. The table shows each accuracy between Non-RTI and RTI datasets for classifying with Group A, B, and C.

compared to the paintings in other groups. Group C' shows dry brush strokes using dried pigments, which is difficult to recognize the outline of each brushstroke. Group B' includes the paintings in which different styles of brush strokes are mixed. Figure 8 shows the details of the style difference of each group.
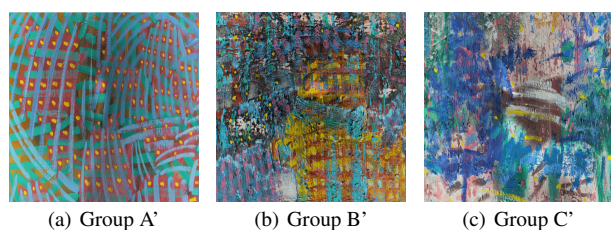

(a) Group A' (b) Group B' (c) Group C'

Figure 7. RTI data examples of three different groups.

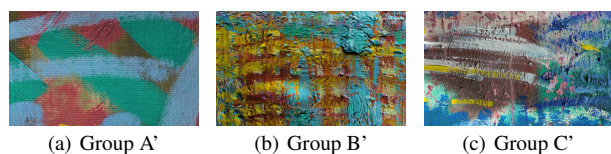(a) Group A' (b) Group B' (c) Group C'

Figure 8. Cropped parts of RTI data from Figure 7 of each group.

4.2.1 Dataset of Experiment 2
The RTI and Non-RTI data were acquired at 7952 x 5304 resolution with Sony a7R III camera. Other processes were the same as Experiment 1.

4.2.2 Categorization the dataset
We tested three different architectures of the CNN model - VGG-16 (16 Layers), ResNet-50 (50 Layers) and DenseNet-121 (121 Layers) to see whether the simple CNN model or the complex model is efficient by increasing the layer from a relatively low-layered architecture (Simonyan, Zisserman, 2014) (He et al., 2016) (Huang et al., 2017). Since the paintings in experiment 2 were grouped into three classes, categorical was used as the loss function of the outlayer, and softmax as the activation function. The fully-connected layer has also been modified for the three classes. Six artworks in Group A', three in Group B', and two in Group C' were chosen as a train set and a validation set. One artwork of each group was separated to be used as a test set. 90% of the dataset except the test set was used as a train set to learn the CNN model and 10% was used as a data set for validation of the model accuracy. The data were randomly separated when they were divided into a test set and a validation set. As a result, 65,930 data were chosen as a train set, 7,325 data as a validation set, and 13,685 data as a test set for Non-RTI data. For RTI data, 65,006 data were chosen as a train set, 7,222 data as a validation set, and 13,457 data as a test set. VGG-16, ResNet-50, DenseNet-121 models were pre-trained with ImageNet (Deng et al., 2009).

4.2.3 Results & Discussion
The results in table 1 shows that RTI-data model is more accurate than Non-RTI-data model except Group A'. The group

A' with wet and clear brushstrokes does not show thick brushstrokes compared to group B' and group C' (Fig 7). That is, the difference between the RTI data and the Non-RTI data in Group A' is not so apparent. Other groups explicitly show the difference in brushstrokes between RTI and Non-RTI data. Group B' with mixed brush strokes has the lowest accuracy among the three groups of RTI trained models. The total cropped image from RTI data of group B' used as a test set in VGG-16 model was 4,830. The prediction results as groups A', B' and C' were 1,862 (38.5 %), 1,466 (30.4%), 1,502 (31.1%) each (rounded to one decimal places). We consider that accuracy was low because both types of brush strokes are observed at the same time in a picture. The results also show that accuracy does not always increase as the layer becomes deeper through experiments. In the case of group A, VGG-16 with Non-RTI dataset, group B, DenseNet-121 with RTI dataset, and group C', ResNet-50 with RTI dataset showed the best performance. This result suggests that there is a need for future research to develop more suitable machine learning architectures and optimizers for utilizing RTI data. In Experiment 2, more data was acquired than Experiment 1, but most data were classified into Group A'. This imbalance should be compensated to obtain better results. The disadvantage of machine learning is that it requires many data, and in most cases, the data must be labeled. The developer of the Generative Adversarial Network (GAN), Lan Goodfellow, said that at least 5,000 learning data per category is required for acceptable performance and at least one million learning examples are needed to match or exceed human performance (Goodfellow et al., 2016). In the field of machine learning, data issues have always existed. This study suggests the possibility of using RTI data that visualizes depth information which can increase accuracy in painting classification where dataset acquisition is limited.

5. CONCLUSION

In this paper, a new type of dataset for automatic oil painting style classification using machine learning was presented and evaluated. We hypothesized that the visualized depth information of brushstroke would be useful to improve the accuracy of the machine learning model for painting classification. RTI technology was applied to extract and visualize the depth information of brushstrokes in paintings. In Experiment 1, the performance of binary classification obtained by learning RTI data was compared with Non-RTI data. In Experiment 2, classifying three classes using more artworks. Since this study only performed data validation, we used a basic CNN model and an optimizer. According to the results, when performing machine learning using the RTI dataset, improved classification accuracy can be obtained compared to the general photography dataset. The results of experiments validated our hypothesis.

However, we did not attempt to categorize various shapes, including sizes, angles, or area of brush strokes, as shown in each canvas. Indeed, the results of irregular accuracy in Experiment

2 allows us to rebuild in terms of setting dataset for the on-going project in the future work, and also it means the importance of classification dependant on which and how much data will be given for us is important to analyze images for the next step. For improving this gap between different styles of paintings, it is necessary for testing a variety of paintings to compare and classify them, not only in the number of paintings or styles but also in various artists. Besides, an optimized model for machine learning is needed, especially in a developed architecture applying RTI technology for future work. Through this approach, it provides us to widen the area of art classification in machine learning, and further, it can provide a sophisticated tool to aid painting classification and appraisal.

In the following research, we will expand the dataset with different styles of artists and paintings, and divide them into the brush strokes of different objects included in the artwork. It is necessary to develop a model and optimizers suitable for the RTI data of the artwork to improve the performance of the painting classification.

REFERENCES

Abate, D., Menna, F., Remondino, F., Gattari, M., 2014. 3D painting documentation: Evaluation of conservation conditions with 3d imaging and ranging techniques. *International Archives of the Photogrammetry, Remote Sensing & Spatial Information Sciences*, 45.

Arora, R. S., Elgammal, A., 2012. Towards automated classification of fine-art painting style: A comparative study. *Proceedings of the 21st International Conference on Pattern Recognition (ICPR2012)*, IEEE, pp. 3541–3544.

Barni, M., Pelagotti, A., Piva, A., 2005. Image processing for the analysis and conservation of paintings: Opportunities and challenges. *IEEE Signal Processing Magazine*, 22(5), pp. 141–144.

Berezhnoy, I. E., Postma, E. O., van den Herik, H. J., 2009. Automatic extraction of brushstroke orientation from paintings. *Machine Vision and Applications*, 20(1), pp. 1–9.

Berezhnoy, I. E., Postma, E. O., van den Herik, J., 2005. Computerized visual analysis of paintings. *Int. Conf. Association for History and Computing*, pp. 28–32.

Berezhnoy, I., Postma, E., van den Herik, J., 2007. Computer analysis of Van Goghs complementary colours. *Pattern Recognition Letters*, 28(6), pp. 703–709.

Berns, R. S., 2001. The science of digitizing paintings for color-accurate image archives: A review. *Journal of imaging science and technology*, 45(4), pp. 305–325.

Breuckmann, B., 2011. 3-Dimensional digital fingerprint of paintings. *2011 19th European Signal Processing Conference*, IEEE, pp. 1249–1253.

Chollet, F. et al., 2015. Keras. https://github.com/fchollet/keras.

Clarricoates, R., Kotoula, E., 2019. The potential of reflectance transformation imaging in architectural paint research and the study of historic interiors: A case study from Stowe House, England. *Journal of the Institute of Conservation*.

Cortes, C., Vapnik, V., 1995. Support-Vector Networks. *Machine Learning*, 20(3), pp. 273–297.

Cultural Heritage Imaging, 2019. Reflectance transformation imaging (RTI).

Deng, J., Dong, W., Socher, R., Li, L.-J., Li, K., Fei-Fei, L., 2009. Imagenet: A large-scale hierarchical image database. *2009 IEEE Conference on Computer Vision and Pattern Recognition*, IEEE, pp. 248–255.

Elkhuizen, W. S., Zaman, T., Verhofstad, W., Jonker, P. P., Dik, J., Geraedts, J. M., 2014. Topographical scanning and reproduction of near-planar surfaces of paintings. *Measuring, Modeling, and Reproducing Material Appearance*, 9018, International Society for Optics and Photonics, pp. 901809.

Folego, G., Gomes, O., Rocha, A., 2016. From impressionism to expressionism: Automatically identifying van gogh's paintings. *2016 IEEE International Conference on Image Processing (ICIP)*, IEEE, pp. 141–145.

Giachetti, A., Ciortan, I., Daffara, C., Pintus, R., Gobbetti, E. et al., 2017. Multispectral RTI analysis of heterogeneous artworks.

Goodfellow, I., Bengio, Y., Courville, A., 2016. *Deep Learning*. MIT press.

He, K., Zhang, X., Ren, S., Sun, J., 2016. Deep residual learning for image recognition. *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pp. 770–778.

Huang, G., Liu, Z., Van Der Maaten, L., Weinberger, K. Q., 2017. Densely connected convolutional networks. *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pp. 4700–4708.

Johnson, C. R., Hendriks, E., Berezhnoy, I. J., Brevdo, E., Hughes, S. M., Daubechies, I., Li, J., Postma, E., Wang, J. Z., 2008. Image processing for artist identification. *IEEE Signal Processing Magazine*, 25(4), pp. 37–48.

Khan, F. S., Beigpour, S., Van de Weijer, J., Felsberg, M., 2014. Painting-91: A large scale database for computational painting categorization. *Machine Vision and Applications*, 25(6), pp. 1385–1397.

Kingma, D. P., Ba, J., 2014. Adam: A method for stochastic optimization. *arXiv preprint arXiv:1412.6980*.

Krizhevsky, A., Sutskever, I., Hinton, G. E., 2012. Imagenet classification with deep convolutional neural networks. *Advances in Neural Information Processing Systems*, pp. 1097–1105.

Li, J., Yao, L., Hendriks, E., Wang, J. Z., 2011. Rhythmic brush-strokes distinguish Van Gogh from his contemporaries: Findings via automated brushstroke extraction. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 34(6), pp. 1159–1176.

Manrique Tamayo, S. N., Andrés, V., Cayetano, J., Osca Pons, M., 2013. Applications of reflectance transformation imaging for documentation and surface analysis in conservation. *International Journal of Conservation Science*, 4, pp. 535–548.

Nanni, L., Ghidoni, S., Brahnam, S., 2017. Handcrafted vs. non-handcrafted features for computer vision classification. *Pattern Recognition*, 71, pp. 158–172.

Pamart, A., Ponchio, F., Abergel, V., Alaoui M'Darhri, A., Corsini, M., Dellepiane, M., Morlet, F., Scopigno, R., De Luca, L., 2019. A complete framework operating spatially-oriented RTI in a 3d/2d cultural heritage documentation and analysis tool. *ISPRS-International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, 422, pp. 573–580.

Peng, K.-C., Chen, T., 2015. A framework of extracting multi-scale features using multiple convolutional neural networks. *2015 IEEE International Conference on Multimedia and Expo (ICME)*, IEEE, pp. 1–6.

Ponchio, F., Lamé, M., Scopigno, R., Robertson, B., 2018. Visualizing and transcribing complex writings through rti. *2018 IEEE 5th International Congress on Information Science and Technology (CiSt)*, IEEE, pp. 227–231.

Simonyan, K., Zisserman, A., 2014. Very deep convolutional networks for large-scale image recognition. *ArXiv Preprint ArXiv:1409.1556*.

Yosinski, J., Clune, J., Nguyen, A., Fuchs, T., Lipson, H., 2015. Understanding neural networks through deep visualization. *ArXiv Preprint ArXiv:1506.06579*.

Zeiler, M. D., Fergus, R., 2014. Visualizing and understanding convolutional networks. *European Conference on Computer Vision*, Springer, pp. 818–833.

Revised May 2019