The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences, Volume XLVIII-M-9-2025
30th CIPA Symposium "Heritage Conservation from Bits:
From Digital Documentation to Data-driven Heritage Conservation", 25–29 August 2025, Seoul, Republic of Korea

Classification of Trees in <Donggwoldo (東闕圖)> Using CNN Deep Learning - Focusing on Tree Representation Techniques-

GaYeon Lee [1], SunYong Sung [1]

[1] Dept. of Traditional Landscape Architecture, Korea National University of Heritage, Republic of Korea – (20221016, sysung85)@knuh.ac.kr

Keywords: CNN Deep Learning, Donggwoldo, Tree representation techniques, Multiclass classification, MobileNetV2.

Abstract

This study explores the potential of applying Convolutional Neural Network (CNN)-based deep learning techniques to the analysis of vegetation landscapes depicted in traditional pictorial records. The research focuses on <Donggwoldo (東闕圖)>, a 19th-century Korean court painting that visually documents the architectural and natural landscape of the Joseon royal palaces. Drawing upon prior studies and the Chinese painting manual『Jieziyuan Huazhuan (芥子園畵傳)』, six distinct types of tree representation were defined to serve as classification categories for the analysis. To construct a training dataset, a total of 580 tree images were manually cropped from two versions of <Donggwoldo>, held respectively by the Korea University Museum and the other by the Dong-A University Museum. These images were then augmented using techniques such as rotation, flipping, and zooming to enhance dataset diversity and reduce overfitting. Seven pre-trained CNN models (including ResNet50V2, InceptionV3, and DenseNet121) were tested in a CPU-based Google Colab environment. Among them, MobileNetV2 was selected as the most suitable model based on its balance of high accuracy, low computational demand, and relatively fast training time. The selected model achieved 100% validation accuracy and a minimal validation loss of 0.3% by epoch 65. Final performance on the test set reached 95.5% accuracy, with precision, recall, and F1 score each measuring 86.6%. These findings demonstrate that deep learning can effectively classify tree types in historical paintings and offer valuable tools for digital heritage preservation, landscape interpretation, and cultural informatics. The approach also underscores the potential for machine learning to assist in the systematic study of historical visual culture.

1. Introduction

1.1 Background and Objective

Documentary paintings are visual records that realistically depict historically significant events, figures, and places, offering spatial and temporal information about the past much like modern photography. These visual materials serve as crucial references in the restoration and conservation of cultural heritage, assisting in determining the form, location, and scale of historical objects.

Gyehwa(界畵), a type of documentary painting, refers to a painting or painting method that accurately depicts buildings such as palaces, pavilions, and houses using a ruler, and has been emphasized in the interest of the court since King Jeongjo's reign.

Figure 1. Donggwoldo (Dong-A University Museum version).

https://doi.org/10.5194/isprs-archives-XLVIII-M-9-2025-785-2025

At the same time, the Dohwa-Won(圖畫署), a government office established to handle painting-related affairs, selected Chabidae-Ryeong-Hwawon(差備待令畫員, court painters), and the field of court painting at that time made significant progress based on their contributions.

<Donggwoldo> employs a parallel diagonal composition to depict the Eastern Palaces specifically Changdeokgung(昌德宮) and Changgyeonggung(昌慶宮) in detailed accordance with actual topography. Unlike many other documentary paintings, it gives significant visual weight to the surrounding natural landscapes, making it a valuable source for studies in landscape architecture (Lee Miyeon, 2024).

This study categorizes the tree illustrations in <Donggwoldo> based on artistic representation techniques and develops a CNN-based classification model accordingly. Through this approach, this study seeks to explore the potential of applying deep learning techniques to vegetation landscape analysis using documentary paintings as historical visual data.

1.2 Research Subject

This study focuses on <Donggwoldo>, a representative example of Gyehwa from the Joseon Dynasty (Figure 1). Donggwoldo is a detailed and realistic depiction of the Eastern Palaces (present-day Changdeokgung and Changgyeonggung Palaces) during the 19th century. As a documentary painting, it provides valuable visual information for understanding the architecture, various facilities, vegetation, and topography of the palace grounds at that time.

The techniques used to depict trees in <Donggwoldo> are based on realistic representation, aiming to reflect the morphological characteristics of actual trees. Although these tree illustrations do not allow for precise identification of specific species, analyzing their representational characteristics offers meaningful insights into the general trends of the historical vegetation landscape of the Eastern Palaces.

2. Methodology

This study applied a Convolutional Neural Network (CNN)-based image classification technique to categorize tree illustrations depicted in <Donggwoldo>. The research was carried out in the following stages (Figure 2).

First, a classification framework was established to determine how the tree images would be categorized. Based on classification criteria proposed in previous studies and considering the unique characteristics of tree representations found in <Donggwoldo>, a modified classification system was developed.

Second, a dataset was constructed for training the CNN model. Tree regions were cropped from the <Donggwoldo> images, and each image was labeled according to the newly defined classification criteria. Image augmentation techniques were then applied to increase the volume of training data. The dataset was divided into training, validation, and test sets for model development and evaluation.

Third, a CNN-based classification model was built using the constructed dataset. To achieve high performance with relatively limited data, a pre-trained MobileNetV2 model was adopted, and transfer learning was applied to enhance training efficiency.

The model was developed using the free version of Google Colab, implemented in a CPU-based environment with TensorFlow and Keras (Google, 2017). Google Colab offers a convenient cloud-based environment that enables deep learning experiments directly in a web browser without requiring any local installation. It also provides access to pre-installed libraries and supports easy sharing and storage of experimental results. However, the free version of Google Colab has limitations such as restricted access to GPU/TPU and automatic session termination after a certain period of inactivity. Additionally, the CPU environment performs slower than GPU-based systems, making it less suitable for large-scale datasets or complex model training. Despite these constraints, the relatively small size of the image dataset in this study allowed for stable experimentation in a CPU-based environment.

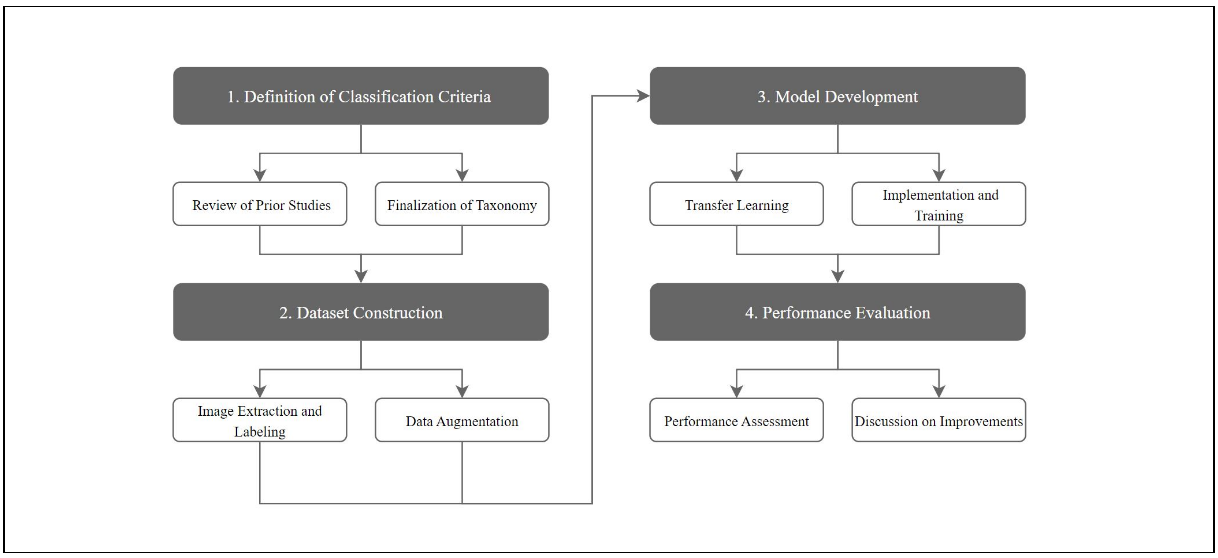

Figure 2. Research flowchart.

Label	A (0)	B (1)	C (2)	D (3)	E (4)	F (5)	Sum
Type	‘sohonjeom’ (小混點)	‘gaejajeom’ (介字點)	long oval 'guyeob' (勾葉)	willow tree	‘ang-yeobjeom’ (仰葉點)	flowering trees	-
				-		-	-
Img							-
Ct.	102	93	91	93	98	103	580

Table 1. Tree representation techniques and data count (Cultural Heritage Administration., 2016).

2.1 Classification Framework & Data Collection

A study on the classification and distribution of tree representation types in <Donggwoldo> was previously conducted in 2007. However, that study lacked clearly defined classification criteria, which led to inconsistencies in results depending on the individual conducting the classification.

In response to this limitation, a subsequent study in 2016 focused on the historical context in which <Donggwoldo> was produced. It identified that during the 19th century the late Joseon period there was an influx of the Southern School painting style (Namzong Hua-feng, 南宗畵風) from China, along with the dissemination of painting manuals such as the 『Jieziyuan Huazhuan (芥子園畫傳)』, which were used for learning tree illustration techniques at the time. Based on this context, a total of 13 tree depiction types used in <Donggwoldo> were derived (Cultural Heritage Administration, 2016).

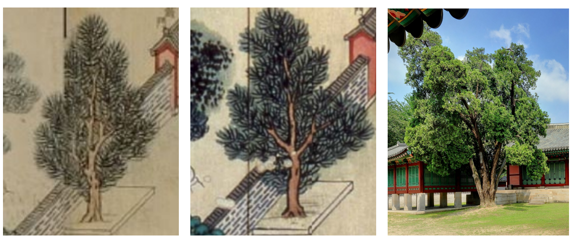

Figure 3. Trees in the Same Location Depicted in the <Donggwoldo>[1] (A: Dong-A University Museum, B: Korea University Museum, C: Real object).

The <Donggwoldo> currently exists in only two versions: one held by the Korea University Museum and the other by the Dong-A University Museum[2]. Both versions were created based on the same preliminary drawing but exhibit slight stylistic differences due to their nature as pictorial records (Figure 3). Therefore, data were collected from both versions of the <Donggwoldo>.

The Dong-A University version collected data from <Donggwoldo> (the 19th Century Dong-A University Museum) provided by Google Art & Culture, and the Korea University version from the <Donggwoldo> PDF file[3] (Cultural Heritage Administration, 1991), provided by the Korea Heritage Service. Additionally, images where tree representation techniques were not visually distinguishable due to poor image quality were excluded from the dataset to prevent noise during model training. Ultimately, this study focused on six tree representation types that had a relatively enough data (Table 1). A total of 580 images were collected and used to construct the final dataset for classification (Table 1).

2.2 Image Augmentation

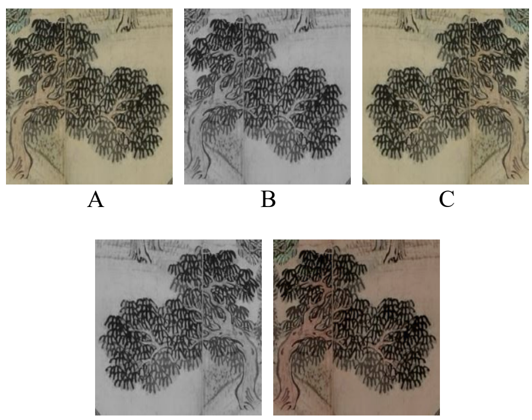

Figure 4. Examples of Image Augmentation (A: Original, B: Gray, C: Flipping, D: Gray + Flipping, E: Color Adjusting).

(1) There is a slight difference between the locations of trees shown in <Donggwoldo> and the actual locations of trees. It is presumed that this is due to changes in the locations of buildings or fences after the production of <Donggwoldo>.

(2) The Korea University Museum version bears the phrase ”Donggwoldo-in (東闕圖人)” on each panel. Based on painting conventions, it is likely that companion panels titled ”Donggwoldo-cheon (東闕圖天)” and ”Donggwoldo-ji (東闕圖地)” also existed, suggesting that a total of three scrolls were originally produced.

(3) The book used as research data was selected because it explained the collection of Korea University Museum.

One way to improve the performance of CNN models and overcome overfitting is to increase the amount of training image data. However, the amount of data that an individual can realistically obtain is limited, and the quality of the collected data is often inconsistent. To address these limitations, data augmentation techniques are employed.

In image classification, common data augmentation methods include rotation, translation, resizing, cropping, brightness adjustment, noise removal or addition, color adjustment, and gaussian blur. In this study, data augmentation was performed using techniques such as grayscale conversion, horizontal flipping, random zooming, and random color adjustment (Figure 4).

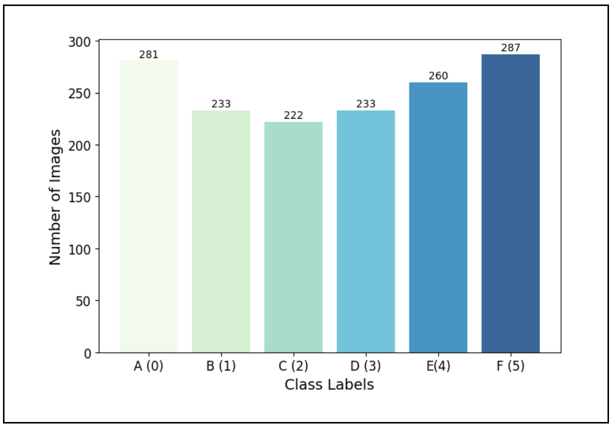

Figure 5. Train Data Distribution.

At this stage, unused data was required for an accurate evaluation of the trained model. To address this, 25 images were randomly selected from each label before image augmentation to create a separate test set. The remaining data were then augmented fivefold using data augmentation techniques, resulting in a total of 1,680 training images (Figure 5), 10% of which were allocated for validation. As a result, the final dataset was organized into training, validation, and test sets in an approximate ratio of 8:1:1.

2.3 Selection of Transfer Learning Model

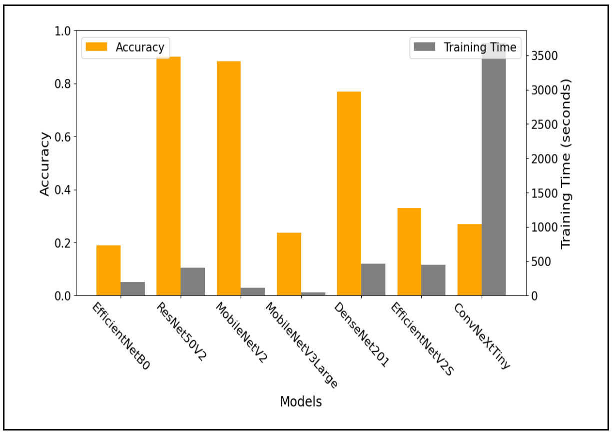

Figure 6. Accuracy and learning time by model.

To build a high-performance CNN model, not only is a large volume of training data required, but also a high-spec computational system capable of efficiently processing such data, as well as considerable training time. However, individual researchers may not always have access to these resources. One solution to overcome these limitations is transfer learning. Representative pretraining models include VGGNet, ResNet, SENet, GoogleNet, and EfficientNet.

For the evaluation of the transfer learning model to be used in this study, a total of seven models were selected including EfficientNetB0, ResNet50V2, MobileNetV2, DenseNet201, and EfficientNetV2S among the well-run models in the Google Colab CUP environment.

Subsequently, the training time and accuracy were evaluated for each model, and MobileNetV2, which recorded high accuracy and fast learning speed with approximately 88% accuracy and 113.3 seconds of training time, was selected as the final transfer learning model (Figure 6).

3. Results and Discussion

3.1 Training Preparation

Based on the pre-trained MobileNetV2 model, a custom classification model was constructed to suit the objectives of this study. The top classification layer of MobileNetV2 was removed so that the model functioned solely as a feature extractor, and all layers were frozen to prevent their weights from being updated during training.

Global Average Pooling was applied to the extracted feature maps to reduce spatial dimensions, followed by two fully connected (Dense) layers with 128 and 64 neurons, respectively. Each Dense layer used the ReLU activation function, and dropout was applied between them to prevent overfitting, with dropout rates set to 0.4 and 0.3, respectively. The final output layer consisted of 6 neurons corresponding to the six tree classes, using the Softmax activation function to output class probabilities.

The model was optimized using the Categorical Crossentropy loss function and the Adam optimizer with a learning rate of 0.0002. Training was conducted with a batch size of 16 for up to 100 epochs, and Early stopping was implemented to halt training if validation performance did not improve over a certain number of epochs. At each epoch, training accuracy and loss, as well as validation accuracy and loss, were monitored. The model weights from the epoch with the best validation performance were saved.

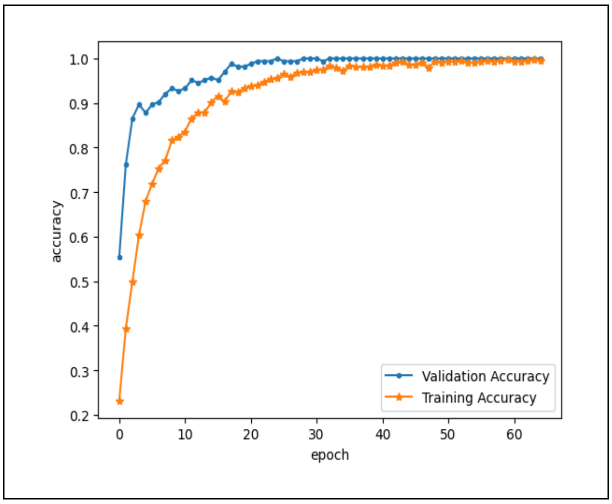

Figure 7. Training and Validation Accuracy.

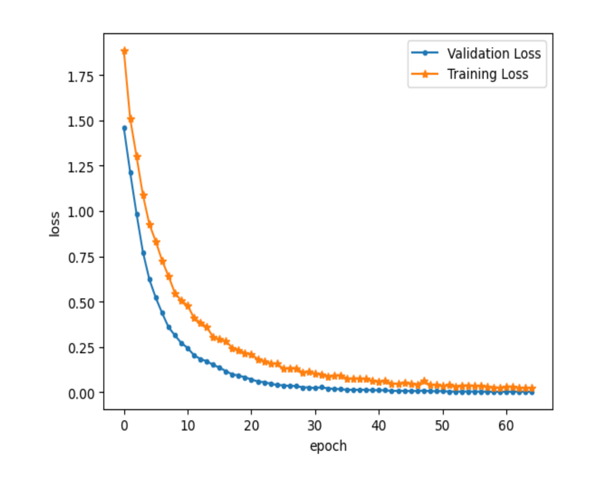

Figure 8. Training and Validation Loss

As a result, early stopping was triggered at epoch 65, with a training accuracy of 99.21% and a validation accuracy of 100.00% (Figure 7). The training loss was 2.3%, while the validation loss was 0.3% (Figure 8). These results indicate that the model achieved high generalization performance without overfitting, demonstrating its effectiveness in classifying tree types depicted in <Donggwoldo>.

3.2 Performance Evaluation

To evaluate how the model trained on the training and validation datasets performs on new data, an unused test dataset was used. The model's classification performance during the testing process was assessed using commonly used metrics such as accuracy, precision, recall, and F1 score. In addition, a confusion matrix was used to check how each class was classified.

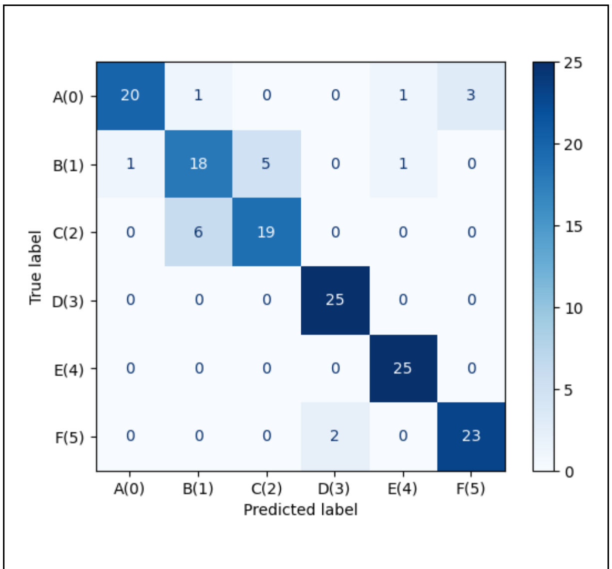

Figure 9. Confusion Matrix.

As a result of classifying 150 test images, 5 out of 25 actual images of type A were misclassified, and 7 out of 25 actual images of type B were misclassified. In addition, 6 out of 25 actual images of type C were misclassified, while all 25 images of types D and E were correctly classified. Finally, 2 out of 25 actual images of type F were misclassified (Figure 9).

		Predicted		
		Negative	Positive	
Actual	True	**TN** 730	**FP** 20	**Specificity** 97.3%
	False	**FN** 20	**TP** 130	**Recall** 86.6%
		Accuracy 95.5%	**Precision** 86.6%	**F1-Score** 86.6%

Table 2. Evaluation Metrics.

Based on the results derived from the confusion matrix, the model achieved an accuracy of 95.5%, precision of 86.6%, recall of 86.6%, and an F1 score of 86.6% (Table 2). These results indicate that the model correctly classified a high proportion of the overall test data. However, it is also notable that, despite the high accuracy, the values of the other evaluation metrics are relatively lower in comparison.

Figure 10. Misclassified Image Data.

As previously mentioned, the model misclassified 20 out of 150 test images after training and classification on actual image data. Although the overall accuracy was high, other evaluation metrics were relatively lower. This discrepancy is primarily due to image blurring caused by low resolution (Dodge & Karam., 2016), interference from background facilities included in the images (Rosenfeld et al., 2018), and tree regions that were not perfectly separated during preprocessing (Lee et al., 2018) (Figure 10).

4. Conclusion

4.1 Limitations

This study focused on tree illustrations depicted in <Donggwoldo>, a court documentary painting from the late 19th-century Joseon Dynasty. As such, the tree classification system and deep learning model developed in this research are optimized for the iconographic style of that specific period. Therefore, while the proposed classification criteria and model may be highly applicable to other pictorial records that share a similar historical background and painting style, their applicability to records from different eras or with different artistic techniques may be limited.

In addition, the image data used in this study were sourced from digital materials provided by Google Arts & Culture and the Korea Heritage Service. Some of these images were of low resolution, with blurred lines and colors in the depiction of trees. Such low-quality images may fail to accurately reflect the characteristics of tree rendering techniques, potentially having a negative impact on the learning process and classification accuracy of the deep learning model.

Future research should aim to include a more diverse dataset that covers various historical periods and artistic styles and develop more refined training models based on high-resolution imagery.

4.2 Implications

This study applied CNN deep learning to analyze and classify tree illustrations in <Donggwoldo>, achieving a high classification accuracy of approximately 95.5%. This result demonstrates the potential for using traditional painting-based tree representation typologies in vegetation and landscape analysis through historical documentary paintings.

Moreover, the study presents a novel approach by extracting visual information embedded in documentary paintings and utilizing it as digital data for AI-based analysis. This highlights the applicability of artificial intelligence in the interpretation and digital archiving of cultural heritage. As such, the research serves as a meaningful foundation for future AI-based studies in the field of cultural heritage.

Acknowledgements

This paper is the result of the R&D support project "Development of a Climate Change Risk Assessment Framework for Natural Heritage" of the Korea National University of Heritage.

References

Cultural Heritage Administration, 1991: Donggwoldo. Cultural Heritage Administration

Cultural Heritage Administration, 2016: Analysis of Major Vegetation and Restoration of Partitions in Donggwol (Changdeokgung and Changgyeonggung Palaces), Cultural Heritage Administration

Dong-A University Museum, 19th century. Donggwoldo (Eastern Palace Map). *Google Arts and Culture*, https://artsandculture.google.com/asset/the-eastern-palace-unknown/hQEsB-GVjRR09A?hl=ko (28 February 2025).

Google, 2017. Google Colaboratory (Google Colab), Version 1.0.0. Google. colab.research.google.com (13 March 2025).

Lee, M.Y., 2024. A Study on the Painting Characteristics of Gyehwa (界畫) in Donggwoldo (東闕圖). *PhD dissertation*, Kyonggi University, Suwon.

Lee, S.J., Lee, K.D., Lee, S.W., Ko, J.G., & Yoo, W.Y., 2018. Technology trends and analysis of deep learning-based object classification and detection. *Electronics and Telecommunications Trends*, 33(4), pp. 33-42.

Rosenfeld, A., Zemel, R., & Tsotsos, J. K., 2018. The elephant in the room: Examining the importance of context for object detection. *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 876-885. doi.org/10.48550/arXiv.1808.03305

S. Dodge and L. Karam, 2016. Understanding how image quality affects deep neural networks, *Proceedings of the 8th International Conference on Quality of Multimedia Experience (QoMEX)*, pp. 1-6. doi.org/10.48550/arXiv.1604.04004