

Design Flow and Implementation of an AI-Enabled Smart Guidance System for Art Galleries

Alvaro Belloso Esteban^{1,2}, Mario Silvestre Delgado^{1,2}, Xinrui Yu¹, Won-Jae Yi¹ and Jafar Saniie¹

¹*Embedded Computing and Signal Processing Research Laboratory (<http://ecasp.ece.iit.edu/>)*

Department of Electrical and Computer Engineering, Illinois Institute of Technology, Chicago, IL, U.S.A.

²*E.T.S.I. Industriales, Universidad Politécnica de Madrid, Madrid, Spain*

Abstract— Audio guides are commonly utilized to enrich the experience of art gallery visitors and to fully engage them with the artwork by providing background, contexts, and other information related to the corresponding artists. However, this method may be monotonous to the public and cannot automatically change to the appropriate recording when users move to the next artwork. The traditional operation of audio guides is only enabled by the user’s visual recognition of the artwork to activate appropriate commentaries manually. In this paper, we explore the potential enhancement of the art gallery visitors’ experience by applying Artificial Intelligence (AI), machine learning, and computer vision techniques to recognize the artwork’s main features automatically and provide the corresponding artist and other information to the visitors. We train a model using a Deep Convolutional Neural Network (CNN) and Kaggle’s Painter by Numbers dataset, and host the trained model on a low-power AI accelerator equipped with a camera to recognize the artwork in real-time. The results are then transmitted to a database, which is utilized by a custom Android application on the smartphone to provide updated artwork-related information on the display or by audio. Our system targets not only the general public for an interactive experience in the art gallery but also visually impaired people for proper guidance with the smartphone’s camera and its voice feedback capabilities.

I. INTRODUCTION

The most traditional method for guiding the visitor in the art gallery is by providing an audio-enabled guidance system that contains pre-recorded narrations. This mechanism implies that the manual user is responsible for triggering the correct audio guides based on visual confirmation of the artwork, and could result in providing misinformation if the visitor is unaware of the location of the displayed artwork.

Artificial Intelligence (AI), machine learning, and computer vision techniques are expected to play a key role in the field of visual arts [1] [2], where these techniques can be applied to real-time recognition of the artwork such as paintings, noteworthy objects from historical buildings, sculptures, and decor, and to provide valuable knowledge including information of the artist, background, context, and more. AI techniques can be used in many different ways in this environment: to enhance museum facility and supply management systems, to adopt robots to take over time-consuming tasks [2], or to favor a better user engagement with the artwork [1]. In addition, AI techniques can provide greater assistance to people with disabilities, including visually impaired people to correctly identify and provide accurate audio guides to the visitor [3].

An example of applying AI and machine learning to the artwork, specifically in paintings, is the Painter by numbers Kaggle competition. This challenge aims at the detection of two given paintings and having the algorithms to determine their artists’ match [4]. Such an application that adopts AI and machine learning techniques would also be useful to detect the potential forgery of the artwork. This implementation can be achieved by using a specific machine learning algorithm called Support Vector Machine (SVM), which can identify artists from different artworks. Two types of SVM models are trained using the pixel’s information from the obtained images and their color histograms [5].

In this paper, we explore the possibility of adopting AI, machine learning, and computer vision techniques to enhance the user experience of the art gallery visitors by enabling automatic recognition of the artwork and providing contextual information on their Android smartphones. This study focuses on obtaining biographical information about the most probable artist of the artwork that the visitor is stand-facing, where the artist prediction result is generated by a Deep Convolutional Neural Network (CNN) implemented on a low-power AI accelerator system. Also, the obtained result updates the cloud-based database, where the custom Android application synchronizes to provide artwork-related information, such as artist background and contexts, by displaying the information on the screen or reading it out by the smartphone’s audio speaker.

II. SYSTEM DESIGN

Fig. 1 shows the overview of the system design. First, a machine learning model is created using Deep CNN and Kaggle’s Painter by Numbers dataset for recognizing artists from different artworks. Once the model is established, we implemented it on an NVIDIA Jetson Nano [6], a low-power AI application accelerator, to predict the artwork’s most probable author from the image captured by the integrated camera. That will result in updating the obtained result on a cloud-based database, where it synchronizes with our custom Android application. The Android application has two dedicated modes: normal mode, which provides information on the display using texts, and audio mode, which provides information via the audio speaker on the smartphone.

A. Deep Convolutional Neural Network Architecture

We utilize Deep CNN to create a model to extract the relevant features from the input images (artworks) and perform the inference procedures to classify them by their predicted

artists. The initially designed approaches involved lightweight, yet accurate convolutional neural networks such as MobileNet [7], and deeper neural networks such as VGG-16 [8] were also explored to increase the accuracy in validation and testing phases. However, complexity and overfitting issues caused these networks to obtain poor results on the selected Kaggle's Painter by Numbers dataset. In this study, ResNet-18 CNN [9] was chosen to perform image (artwork) classification, with enabled improved accuracy via residual architecture by enabling two different paths for the gradients to propagate during training, as shown in Fig. 2. Furthermore, regularization and data augmentation are also incorporated to significantly reduce overfitting. To increase the robustness of the detection algorithm, images fed into the neural network from the training set were randomly flipped vertically and horizontally. Also, additional dropout layers were included in the model which reduce the number of neurons that activate further layers in the fully connected phase. Consequently, the complexity of the trained neural network is reduced, avoiding optimizing the results for non-relevant features from the training set.

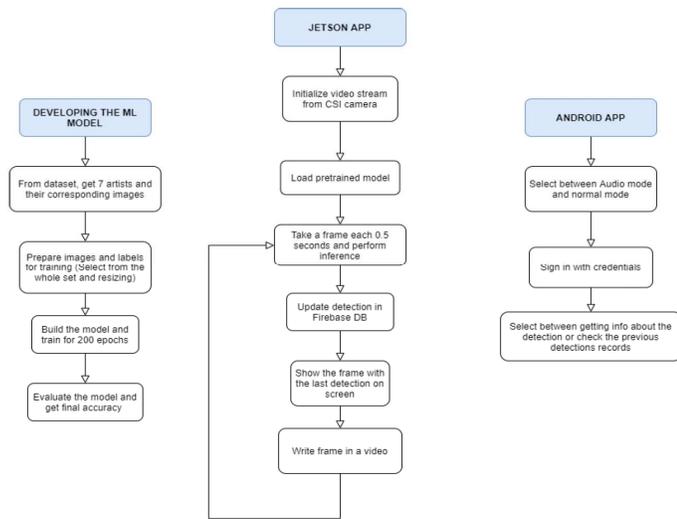


Fig. 1. System Flowchart

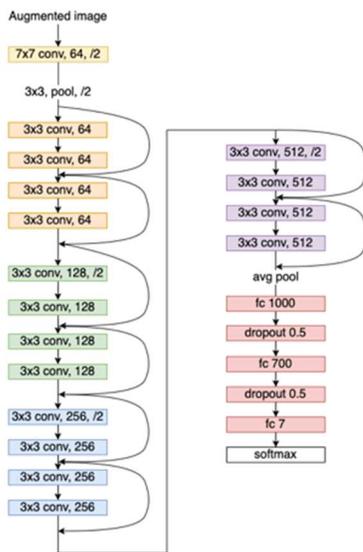


Fig. 2. ResNet-18 Architecture

B. Cloud-based Database and Android Application

As shown in Fig. 1, Google Firebase is used for creating an easy-to-use and tree-based cloud-based database [10]. It serves as the interaction and synchronization platform among the code contained in the NVIDIA Jetson Nano and the Android application. Google Firebase eases its integration with Android applications developed utilizing MIT App Inventor [11]. The customized Android application includes functionality to synchronize with the database where the information regarding the artist's biographical data, previously detected artists' data, and the ability to securely login to the database with username and password is contained. Fig. 3 shows the main screen of the customized Android application, as well as the login screen. As discussed in Section I, our system design goal is to provide artists' information using both normal mode (display textual information on the display) and audio mode (for visually impaired people). *Audio mode* button on the bottom of the screen enables the mode for visually impaired people, where the login process can be resolved by voice inputs. The login allows accessing the data stored in the cloud-based database, being the first common step providing the user's username and password. Later, as shown in Fig. 4, two options are available for the users to select: corresponding artist's information and background of the detected artwork, and displaying the record of artists who were detected previously.

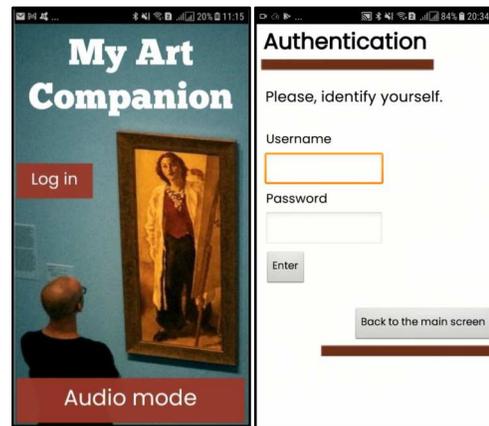


Fig. 3. *My Art Companion* Application's Main Screen (left) and Login Screen (right).

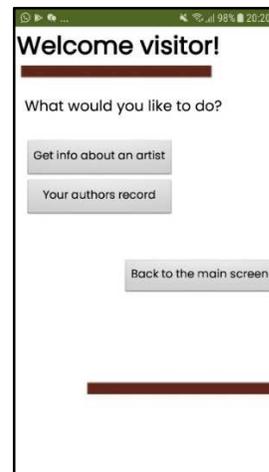


Fig. 4. *My Art Companion* Application's Information Retrieval Screen

III. IMPLEMENTATION AND RESULTS

In this study, we have chosen seven artists from the Kaggle’s Painter by Numbers dataset including Pierre-Auguste Renoir, Gustave Dore, Albrecht Durer, Ivan Aivazovsky, Rembrandt, Zinaida Serebriakova, and William Merritt Chase. Table I shows the number of training and testing images for each set utilized to validate the developed Deep CNN model.

TABLE I. KAGGLE DATASET DEEP CNN TRAINING AND TESTING SETS

Artist	Image Samples	
	Training Sets	Testing Sets
Renoir	388	112
Dore	387	113
Durer	383	117
Aivazovsky	378	122
Rembrandt	376	124
Serebriakova	304	108
Chase	281	82

Python programming language has been chosen for both the training and testing phases of the proposed neural network. We utilized Keras [12] to implement the neural network architecture and to load the trained weights. OpenCV [13] was used for image loading and preprocessing purposes. Once the model is trained for 200 epochs utilizing the Adam Optimization Algorithm [14] and the categorical cross-entropy [15] as the loss function, it is saved in an H5 file format which is ported to the inference platform, the NVIDIA Jetson Nano. NVIDIA Jetson Nano was chosen as the inference platform to take advantage of its dedicated NVIDIA Maxwell GPU with 128 NVIDIA CUDA cores which help to accelerate AI applications, along with the ability to integrate CSI camera into the system. To obtain these video frames, OpenCV is used to open a video feed which loads each frame as a separate image. Then, the model is loaded and takes a frame every 0.5 seconds to process the input from the camera.

As mentioned in Section II, MobileNet and VGG-16 have been explored to increase the accuracy in the validation and testing phases. Due to the overfitting issues caused by the combination of having a relatively small training set and an excessive number of parameters for estimation, poor validation values were obtained in terms of loss and accuracy when using these two networks. Thus, ResNet-18 scheme incorporating data augmentation and regularization procedures was selected, resulting in obtaining a validation accuracy of approximately 75% and performance of the testing set of approximately 70%, as shown in Fig. 5.

Despite the results obtained from the ResNet-18 scheme, our test results demonstrated that the system was able to identify the correct artist of the artwork shown on camera. Fig. 6 illustrates the live camera feed of the NVIDIA Jetson Nano along with the prediction confidence level of 85.35% that the artwork is by Ivan Aivazovsky. Fig. 7 shows the corresponding artist's background information on the custom Android application in real-time.

In addition, Fig. 8 illustrates the live camera feed of the NVIDIA Jetson Nano along with the prediction confidence level of 89.33% that the artwork is by Gustave Dore. Fig. 9 shows the corresponding artist background information on the custom Android application in real-time.

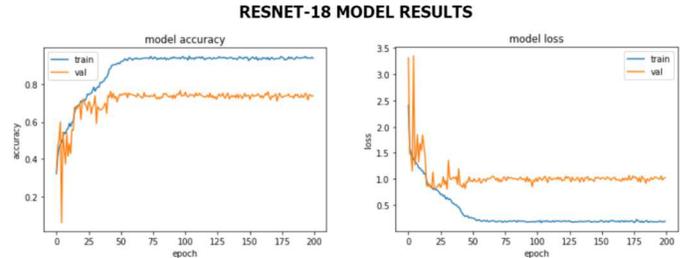


Fig. 5. Training Results using ResNet-18 Scheme

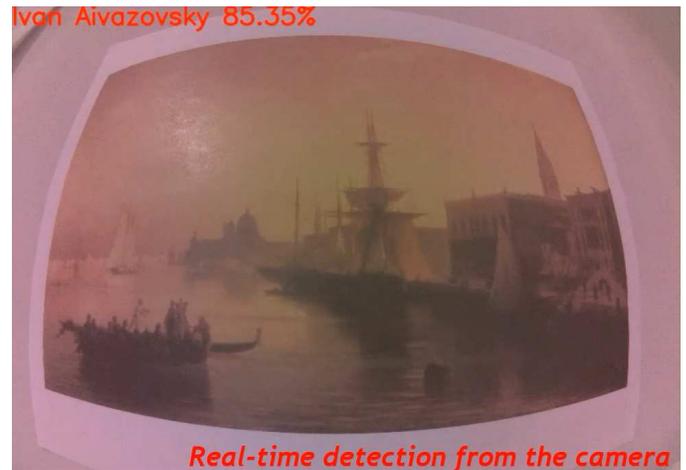


Fig. 6. Detected Artwork from AI Accelerator’s Camera – Aivazovsky

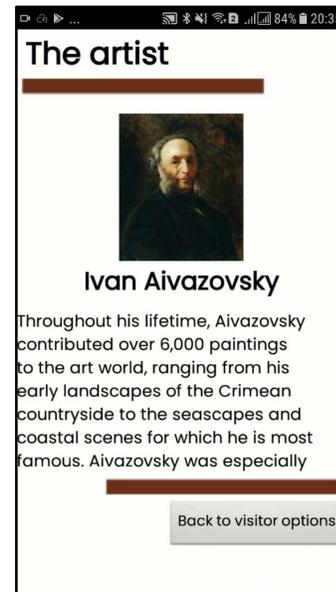


Fig. 7. Artist Information shown on Android Application - Aivazovsky.



Fig. 8. Detected Artwork from AI Accelerator's Camera – Dore

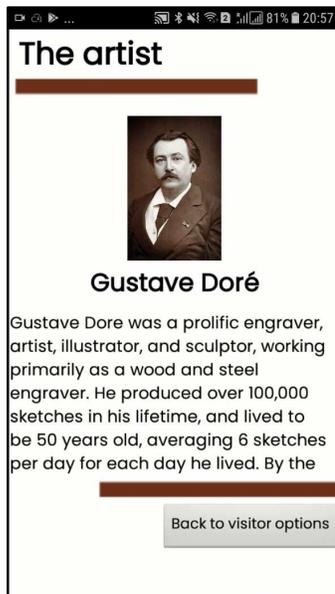


Fig. 7. Artist Information Shown on Android Application – Dore.

IV. CONCLUSION

In this study, the development of a smart assistant utilizing AI, machine learning, and computer vision techniques to enhance art gallery visitors' experience have been explored. Our implementation has included the connection of a CSI camera integrated to an AI accelerator, NVIDIA Jetson Nano, to obtain real-time images from artworks inside an art gallery. Based on the trained model hosted on this device, the most probable artist from the obtained image is predicted by utilizing a Deep CNN model. The result of the inference is then updated to a cloud-based Google Firebase database to enable a custom Android

application to access the detected result and to provide information related to the artist of the artwork. The users, including people with visual disabilities, will be able to acquire information using our customized Android application. Our application can be coupled with an audio tour guide system, and provide corresponding information based on the observed painting. Our system can be expanded to more AI accelerators deployed to different art galleries, to collect more image data, and to improve the datasets to enhance the system's prediction.

REFERENCES

- [1] L. Styx, "How are museums using artificial intelligence, and is AI the future of museums?" Museum Next. Sept. 18, 2020. [Online]. Available: <https://www.museumnext.com/article/artificial-intelligence-and-the-future-of-museums/> [Last Accessed: Jan. 28, 2022].
- [2] O. Murphy, E. Villaespesa, "The Museums + AI network – AI: A Museum Planning Toolkit." Jan. 2020. [Online]. Available: https://themuseumsainetwork.files.wordpress.com/2020/02/20190317_museums-and-ai-toolkit_rl_web.pdf [Last Accessed: Jan. 28, 2022].
- [3] N. Cherukuru, R. Manuel, AJ. Lauer, T. Scheitlin, B. Bhagchandani, "Using Augmented Reality (AR) to Create Immersive and Accessible Museums for People with Vision-Impairments." MW21: MW 2021. Jan. 29, 2021. [Online]. Available: <https://mw21.museweb.net/paper/using-augmented-reality-ar-to-create-immersive-and-accessible-museums-for-people-with-vision-impairments/> [Last Accessed: Jun. 28, 2021].
- [4] Kaggle Competition. "Painter by numbers". [Online]. Available: <https://www.kaggle.com/c/painter-by-numbers/overview/evaluation> [Last Accessed: Jan. 28, 2022].
- [5] A. Blessing, K. Wen, "Using Machine Learning for Identification of Art Paintings". [Online]. Available: <http://cs229.stanford.edu/proj2010/BlessingWen-UsingMachineLearningForIdentificationOfArtPaintings.pdf> [Last Accessed: Mar. 4, 2022].
- [6] NVIDIA Jetson Nano Datasheet. [Online]. Available: <https://developer.nvidia.com/embedded/dlc/jetson-nano-system-module-datasheet> [Last Accessed: Mar. 4, 2022].
- [7] AG. Howard, M. Zhu, B. Chen, D. Kalenichenko, W. Wang, T. Weyand, M. Andreetto, H. Adam, "MobileNets: Efficient Convolutional Neural Networks for Mobile Vision Applications," arXiv:1704.04861, Apr 2017.
- [8] K. Simonyan, A. Zisserman, "Very Deep Convolutional Networks for Large-Scale Image Recognition," arXiv:1409.1556, Apr 2015.
- [9] K. He, X. Zhang, S. Ren, J. Sun, "Deep Residual Learning for Image Recognition," *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 770-778, 2016.
- [10] Google, "Firebase". [Online]. Available: <https://firebase.google.com/> [Last Accessed: Mar. 4, 2022].
- [11] MIT, "MIT App Inventor". [Online]. Available: <https://appinventor.mit.edu/> [Last Accessed: Mar. 4 2022].
- [12] Keras. "Keras: the Python deep learning API". [Online]. Available: <https://keras.io/> [Last Accessed: Mar. 4, 2022].
- [13] OpenCv. [Online]. Available: <https://opencv.org/> [Last Accessed: Mar. 4, 2022].
- [14] D. P. Kingma, J. Ba, "Adam: A Method for Stochastic Optimization", arXiv:1412.6980, 2014.
- [15] J. Brownlee, "A Gentle Introduction to Cross-Entropy for Machine Learning". [Online]. Available: https://www.tensorflow.org/api_docs/python/tf/keras/losses/CategoricalCrossentropy [Last Accessed: Mar. 4, 2022].