**Deliverable 2**

1. Problem statement: Restate the initial project that you proposed in deliverable one in 2 - 3 sentences. Be sure to refer back to this problem statement in the following questions.

The project proposed is to develop an art style classification model using the WikiArt Art Movements/Styles dataset from Kaggle. The goal is to create a machine learning model using Convolutional Neural Networks (CNNs), which will be capable of predicting the art style given an input image. The model will be integrated into a web application where users can upload images of artworks to receive genre classifications. The goal of our project is to help users enhance their appreciation of classical artwork.

2. Data Preprocessing: Confirm the dataset you are working with. State any changes from the initial dataset you chose. Discuss the content of the dataset (number of samples, labels, etc.). Describe and justify your data preprocessing methods (did you delete or modify any data? If so, why?).

The dataset we are planning to work with is the WikiArt Art Movements/Styles dataset from Kaggle. There are 1305 files under Academic Art, 3035 under Art Nouveau, 5312 under Baroque, 2607 under Expressionism, 2235 under Japanese Art, 3115 under Neoclassicism, 1324 under Primitivism, 5375 under Realism, 6192 under Renaissance, 2521 under Rococo, 6813 under Romanticism, 1510 under Symbolism, 1158 under Western Medieval. This brings us to a total of 42 502 files classified under 13 different genres. The entire dataset is a whopping size of 29.23 GB. There are no changes to report; we are still working with our initially chosen dataset. We did not modify or delete any data, because the data set is catered very well to our project. However, we will need to normalize the size of the data.

3. Machine learning model: In the first deliverable, you proposed a model for your project. If you decided to change your model, explain why. Restate your chosen model and elaborate on the design decisions. Report the following:
    a. Specify the framework and tools that you used to implement your model. (For instance, did you use any libraries such as PyTorch, Keras, etc. to implement the model? Any other tools? What does the architecture of your model look like? How many layers/modules? etc.)

We will use TensorFlow, Keras as well as opencv to process our data and classify it. TensorFlow and Keras will be used to train our dataset, whereas opencv will be used later to classify the input image.

Note: We spent most of the time this week learning how to implement image classification and did not get to classifying our chosen images. Instead, we classified images in CIFAR-10 as a first step. We will modify the model to classify paintings this week.

We first implemented an image classification, to get familiar with tensorflow and the use of convolutional neural networks since none of us were familiar with tensorflow and keras. We also used a data set provided by the tensorflow.keras.datasets library, to allow us to focus on the implementation of the CNN.

- Justify any decision about training/validation/test splits, regularization techniques, optimization tricks, setting hyper-parameters, etc.
- Description of validation methods: How did you test your model? Is your model overfitting or underfitting?
- Did you face any challenges implementing the model? If so, how did you solve it?

We successfully implemented the image classifier, but noticed some limitations. First, the accuracy of the model was about 79% (see Version 5 on Kaggle). Also, the images from the dataset have to be resized to 32x32 pixels in the implemented model. This was done to ensure a quick implementation of the training model, which resulted in images with very few details. Therefore when processing our actual dataset, we will need to ensure that the resolution is high enough for the CNNs to be able to pick up details. Furthermore, all the images in our dataset might not all have the same size and we would need to normalize this. One option might be truncating parts of the images to fit them into a given frame and another option would be to use padding and extend the images. The main downsides are, for the first option, the loss of potentially important details and, for the second option, the increase in the dataset size which will slow down the process.

Note : the following section does not really apply to our project situation since we didn't train the model on our chosen dataset yet

4. Preliminary results: In this section, you will focus on the performance of your model. Confirm the metric discussed in Deliverable 1. Present a detailed analysis of your results, providing graphs as appropriate. In addition to an evaluation metric, discuss the overall performance of the model and the feasibility of the project with these results. Remember, graphs are beautiful and we love them!

5. Next steps: Discuss your next steps. Describe the pros/cons of your approach and future work. Will you be altering your model? For example, will you be fine-tuning it? At this point, if you think that your model is not performing well and/or does not work, please reach out to your assigned TPM to see what you can do to improve it.