Alexandria University
Faculty of Engineering
Computer and Communications program

Spring 2025
CSE357 Pattern Recognition
Assigned: 3rd of May

Assignment 3

Style Transfer with Multi-Input Conditioning:

In this assignment, you will implement a neural style transfer pipeline that takes two input images and a style threshold:

- A content image, such as a photograph of a city or a person
- A style image, such as a painting by Van Gogh or Picasso
- The style threshold is a scalar value that determines the degree to which the style is transferred onto the content image

Your model will generate a new image that preserves the content of the first image while adopting the style of the second, modulated by the threshold value. To understand the underlying principles of neural style transfer, you are expected to read the foundational paper on this topic.

Data Loading and Splitting (15 points)

You are required to find or construct a suitable dataset consisting of not more than 10,000 images. This dataset should primarily include content images, with a smaller portion dedicated to style images. A practical approach would be to sample content images from a subset of a Google-scraped image dataset and use the WikiArt dataset for style images. You must divide your dataset into training, validation, and test splits to ensure proper evaluation and generalization.

Data Processing (15 points)

This step involves resizing and preparing the images, as well as constructing the input format for your model. Each input should consist of a content image, a style image, and a style threshold. The quality and consistency of your preprocessing pipeline will significantly influence your model's performance and behavior.

Creating the Model (20 points)

You are expected to use a convolutional neural network (CNN) that has been pre-trained, preferably a VGG model. You will need to modify this model by removing or adding layers to adapt it to the style transfer task. Your pipeline should be designed to accept the new multi-input format and produce the stylized output image, following the methodology described in the paper. You can add other models in addition to the pre-trained model to generate the final image.

Dr. Nada Osman

Eng. Abdelrahman Wael
Eng. Mohamed Saeed

Fine-Tuning the Model (30 points)

Using your dataset and the modified pre-trained model, you will implement a neural style transfer training pipeline. You will need to adapt the original approach to incorporate the style threshold, allowing for dynamic control over the degree of style transfer. Your evaluation should be based on the various loss functions used in the training process, applied across the different data splits.

Evaluation and Report (20 points)

You will prepare a comprehensive report summarizing your approach, including model architecture, hyperparameters, training process, and evaluation results. You must include at least ten example inputs and their corresponding outputs, and provide a critical analysis of your model's performance. Discuss what worked well, what didn't, and what could be improved in future iterations.

Bonus 1: Use a Vision Transformer (10 points)

Replace the CNN with a Vision Transformer (ViT). You should repeat the previous steps using the ViT architecture, explain how it integrates into your pipeline, and compare its performance with the CNN-based approach.

Bonus 2: Create a Web App (10 points)

Create a web application that allows users to upload a content image and a style image, set the style threshold, and receive the stylized output. The backend implementation of this web app will be the basis for evaluating this component.

Notes:

- You are allowed to work in teams of up to 3.

- Submit a **notebook/file** with your code and a **report** detailing your work.

- Teams may lose credit in the discussion if any member has not contributed.

- Any cases of cheating will result in a score of 0% for all involved teams.

Dr. Nada Osman

Eng. Abdelrahman Wael
Eng. Mohamed Saeed