CS60010 - Deep Learning

Midterm Evaluation Report

Group - 12

Group Members: Sachin Vashistha (22CS91R02), Sidharth Vishwakarma (20CS10082), Tarun Mohanty (22CS60R28), Saras Umakant Pantulwar (20CS30046)

Project Title: Kaggle Photo Reconstruction Challenge

1. Description Of the Problem Statement

Images may get corrupted or destroyed/ degraded due to many reasons. The aim of this project is to build models capable of reconstructing missing parts of images that have been damaged or lost.

2. Progress Made

2.1 Dataset Description: The dataset consists of two folders: "Training_Data" and "Testing_Data". The Training_Data consists of images of four classes of animals, namely the Cat, Dog, Tiger and Elephant. Each of the animal class folders will, in turn, consist of 2 folders - "Masked_Train" and "Unmasked_Train".

- Masked_Train - Consists of images which have holes. This is the primary input to the model. Within this folder, there is a CSV file called the "masked_info.csv", containing the data about where the hole is present.
- Unmasked_Train - This folder consists of the original image of the respective damaged image found in the Masked_Train folder.

Each image in the dataset is a 256 x 256 image, and every image will have precisely two holes. So the "masked_info.csv" will have two info columns for each of these holes.
The information on the masks of the Test Data is available in the corresponding "masked_info.csv".

2.2 Dataset Preprocessing:

1. The images have not been resized during training and testing. The original size of 256 x 256 is used. Resizing will be experimented (as described in the Section "Scope of Improvement").
2. The pixel values of images are normalized before using them for training and testing.

2.3 Baseline Model: UNET model. The details about hyper-parameters is as follows:

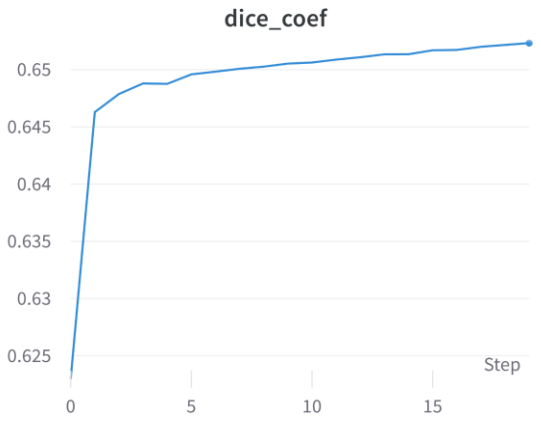
- Batch Size: 32
- Number of Epochs: 20
- Input Image size = Output Image size: (256, 256, 3)
- Optimizer: Adam

2.4 Experiments Carried Out:

- **Choice of metrics:** In tasks like Image reconstruction, two commonly used metrics are IoU (Intersection over Union) and Dice Coefficient. In this project, Dice Coefficient is used whose goal is to maximize the area of overlap between the predicted pixel and the ground truth pixel divided by their union.
- **Loss Function:** Loss function "Mean Absolute Error"(MAE) is used in the experiment. This loss function is used instead of "Mean Squared Error" (MSE) as MSE is more sensitive to outliers than MAE. Here, as the model has to predict the missing hole of a given image, the team made the assumption that outliers may not be rare in this task. But we also plan to compile and train the model using MSE loss than MAE loss.

2.5 Results and Analysis

- **Plot of Dice Coefficient**



Plot of Loss Function

Dice Coeffiecient Value at the end of 20 epochs: 0.6523
Loss Value at the end of 20 epochs: 0.0234

3. Future Direction: Scope of Improvement

The team is going to work in two directions:

- The 1st direction is to resize the given images to 32 x 32 dimension and combine them with the CIFAR-10 dataset which consists of 60000 32 x32 colour images in 10 classes, with 6000 images per class. There are 50000 training images and 10000 test images. Then train the UNET model on the combined dataset.
- The 2nd direction is to use more powerful models. For this, few models have been selected:
    1. CR-Fill
    2. Generative Multi-column Convolutional Neural Networks
    3. Generative Image Inpainting with Contextual Attention

4. Score

The best score Group - 12 has achieved till now is **0.21662**.