Computer vision and object recognition
Assignment 3

Imane SI SALAH

Abstract

This assignment involves training a baseline Convolutional Neural Network (CNN) on the classifysketch dataset for image classification. We analyzed the initial results, and performed subsequent efforts aiming to enhance the model's performance through iterative improvements and fine-tuning strategies.

1. The Dataset:

The dataset employed for this assignment is a subset of the classifysketch dataset, encompassing 12,000 training images and 5,750 validation images and test images. These images are distributed across 250 classes without noticeable imbalance. The original dimensions of the images are 1111x1111 pixels..

2. Preprocessing:

During the preprocessing stage, we apply various transformations to the images before feeding them into the model. Given that the images in our dataset are sketches without any color information, the features we can extract primarily rely on edges. To enhance the dataset's variability, we do flipping and rotation, along with central cropping. Additionally, we resize the images to 256x256, aligning them with the characteristics of images in pretrained models.

3. Training on CNN:

We started with a model that included five convolutional layers along with ReLU activation and maxpooling layers and three linear fully connected layers with some dropout layers. Our goal was to use hierarchical features to extract complex patterns from the sketch data. The weights of the initial layers were frozen and unfrozen repeatedly during the training phase. By freezing these layers during certain epochs, the model was able to grasp low-level representations that are necessary for differentiating fundamental sketch elements. Further unfreezing contributed to the refinement of higher-level features. We included a learning rate scheduler to improve the effectiveness of the training. During training, this scheduler dynamically changed the learning rate, enabling a more controlled convergence to the optimal solution. We employed the cross-entropy loss function being a suitable choice for multiclass classification as the optimization criterion and used the Adam optimizer for the parameter updates. This adaptive learning rate was essential for optimizing the model's performance on a variety of sketch styles and complexities.
Ther results obtained were not good enough, as shown in the table below:

Model	training	validation
Our CNN	70%	39%

4. Pre-trained model

To address the very low accuracy during the initial training of a convolutional network, we adopted a transfer learning approach by fine-tuning pretrained models—specifically, ResNet50 and ResNet152. These models came pretrained on the ImageNet dataset, showcasing their ability to learn rich hierarchical features from a vast array of images. The essence of transfer learning lies in utilizing the knowledge gained by a model on a source domain (ImageNet, in this case) and applying it to a target domain (our sketch dataset). The pretrained weights serve as a starting point, capturing generalizable features like edges, textures, and shapes. By fine-tuning the model on our specific sketch dataset, we enable it to adapt and refine these features to better align with the intricacies of sketch patterns. This process typically leads to improved convergence and higher accuracy compared to training a model from scratch.
The results are illustrated as follows:

Model	Training	Validation
Resnet50	99%	75%
Resnet152	100%	65%