Who’s The Pokemon

CSE-472 Project

Nazia Afreen (1605015)
Tasin Hoque (1605029)

Project Description

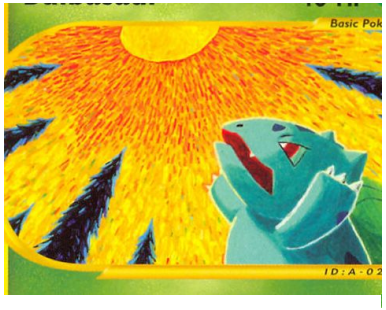
In this project our model is trained to identify pokemons.

For instance, our train dataset has only one image of

Bulbasaur (Pokedex #001) with transparent background

Project Description

Whereas out test dataset has 6 images of Bulbasaur , each with different background and angle.



Our objective is to identify these pokemon images.

Architecture

For our project, we have used meta learning paradigm. We tried to implement ‘One Shot Image Recognition’ from the paper- Siamese Neural Networks for One-shot Image Recognition.

Koch et al’s approach to getting a neural net to do one-shot classification is to give it two images and train it to guess whether they have the same category.

Then while testing, the network compares the test image to each image in the support(training) set, and chooses the closest probable category based on L1 distance.

Architecture

Therefore, our model takes two images as input at a time and the output is the probability of whether they share the same category based on Sigmoid Function.

The loss function is binary cross entropy between the predictions and targets and there is also a L1 weight decay term in the loss.

Architecture

The preprocessing of the dataset:

- Dataset from Kaggle:

  https://www.kaggle.com/datasets/aaronyin/oneshotpokemon

- The support images are loaded as b&w images.
- Images are scaled between 0-255 and reshaped to (128,128,1)
- For data augmentation, images are randomly rotated.
- To avoid overfitting, random noises are added to the background.

Architecture

Input of the model:

- The model is trained in batches where each batch size is 32
- For each batch, a 2 dimensional list (pairs = [[],[]]) of training images and a list of labels go as input. Length of each 1d list equals to batch size.
- The first 32/2 = 16 images of each list in the training 2d list are of same class; therefore, the first 16 entries of the label list is 1.
- The next 16 images of both lists in the 2d list are of different classes; therefore, the last 16 values in the label list is 0.

Architecture

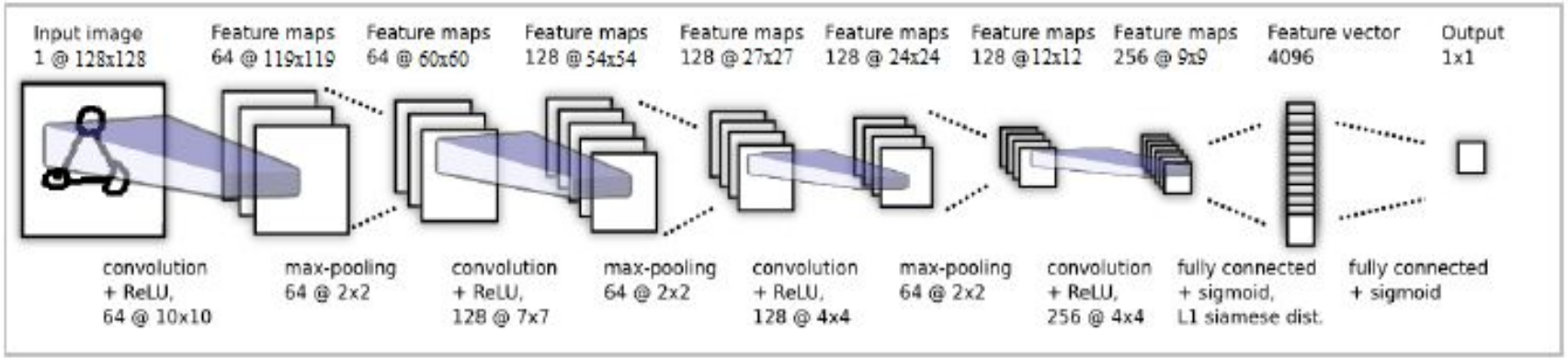
The model architecture:

- In the model we have used several conv2D layers with relu followed by max pooling layer.
- The two inputs are encoded using the convolution network.
- Encoded inputs are merged through L1 distances in between them.
- This merged input is passed through a dense layer.
- The output is squashed into [0,1] with a sigmoid function to make it a probability.

Architecture

The model Architecture:

The image of model architecture is given below:

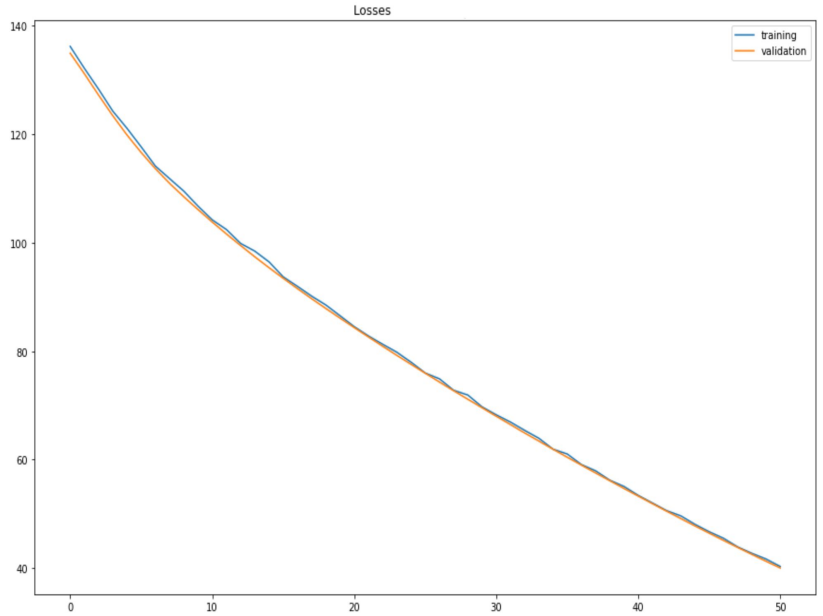


Evaluation Metrics

On training dataset of 150 classes-

Loss graph for 5000 epoch is-

and accuracy is ~10%

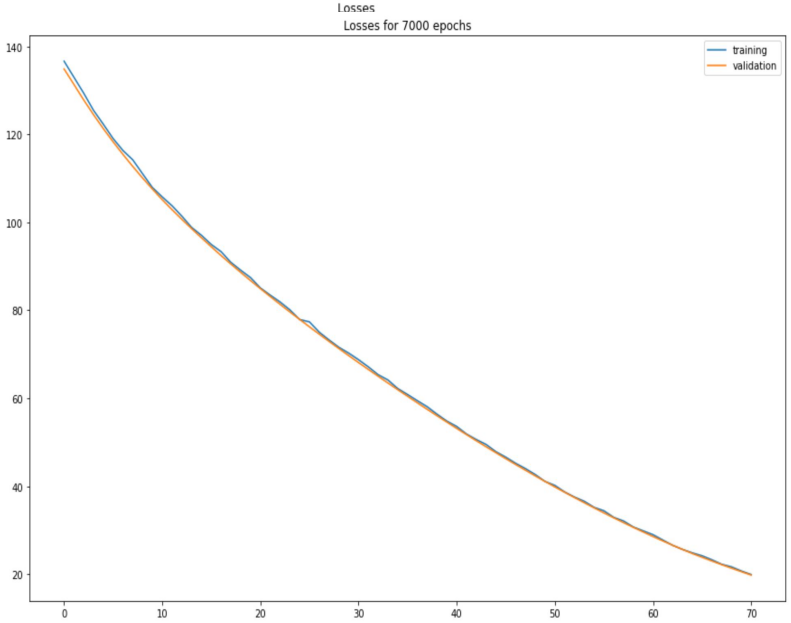


Evaluation Metrics

On training dataset of 150 classes-

Loss graph for 7000 epoch is-

and accuracy is ~21%

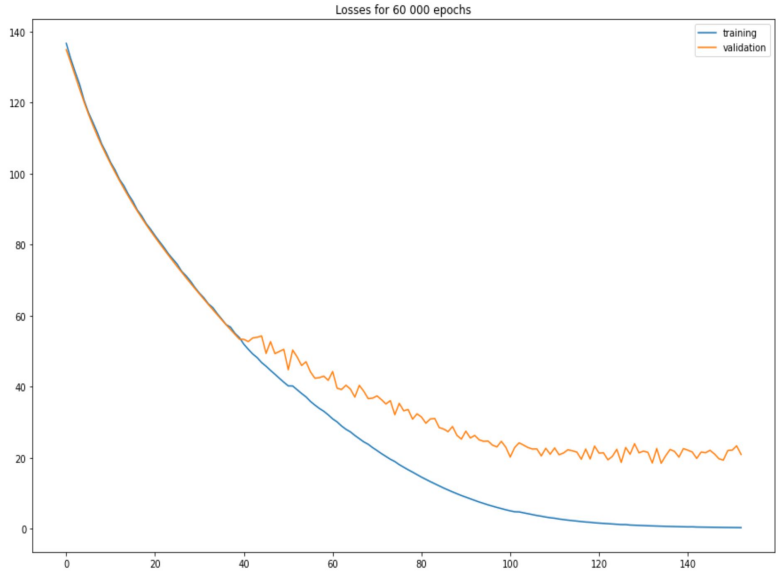


Evaluation Metrics

On training dataset of 150 classes-

Loss graph for 60000 epoch is-

and accuracy is ~51%

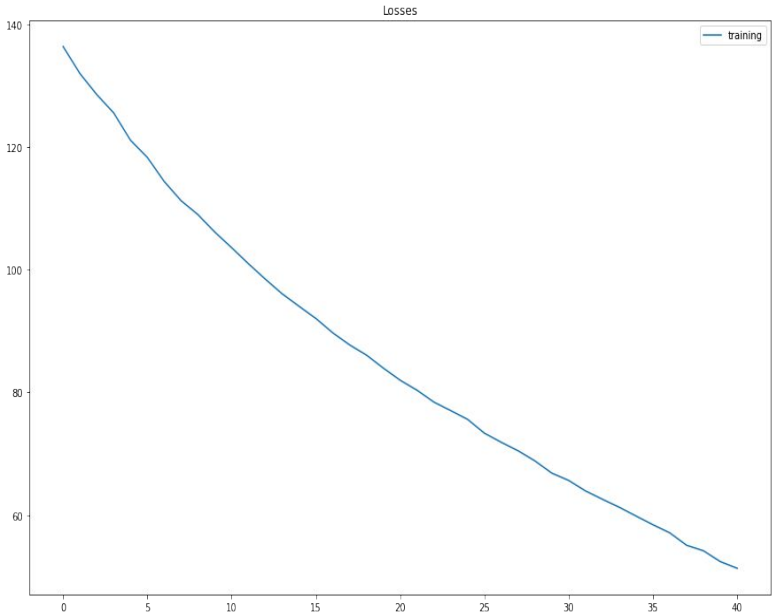


Evaluation Metrics

On training dataset of 50 classes-

Loss graph for 4000 epoch is-

and accuracy is ~13%

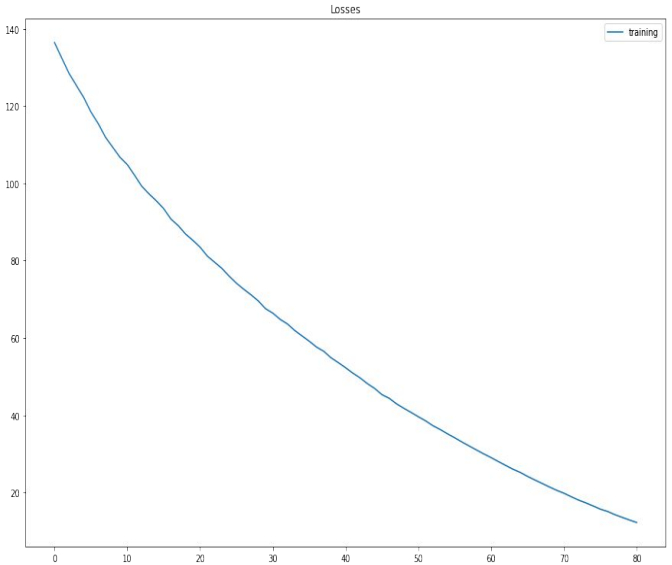


Evaluation Metrics

On training dataset of 50 classes-

Loss graph for 8000 epochs is-

and accuracy is ~35%

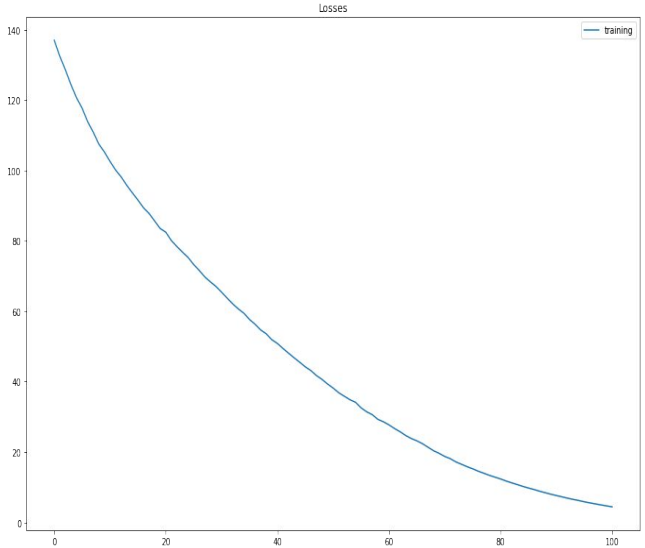


Evaluation Metrics

On training dataset of 50 classes-

Loss graph for 10000 epoch is-

and accuracy is ~52%



Evaluation Metrics

Due to hitting GPU usage limit in Colab, we could not finish running 60000 epochs for 50 classes.

But from the previous 3 graphs it can be seen that, with increasing epochs, the loss decreases and accuracy increases correspondingly.

Challenges

- Each class has only one image of a specific type of pokemon. Therefore, conventional CNN gives a poor accuracy. For this reason, Siamese Neural Network is used.
- Since there is shortage in training data, there is always a chance to overfit. To check this overfitting problem, weights and biases are initialized from a normal distribution, Dropout layers, kernel regularizar, introducing random noises are used according to the paper.