CS337 Course Project
Face Recognition

Pranjal Kushwaha, Shashwat Garg, Vedang Asgaonkar, Virendra Kabra

Autumn 2022

Contents

1 Dataset

We use the Labelled Faces in the Wild (LFW) dataset. It has been taken from Kaggle. There is an uneven distribution of images, with only 10 people having at least 53 images. Thus, we train the model to classify these 10 people, with our dataset containing 53 images for each. The train-test split is 80-20, and the train set is further split to create the validation set.

2 Model Architecture

We use a Convolutional Neural Network (CNN) to create a multi-class image classifier. The network is inspired from FaceNet [**?**].

Layer	**In**	**Out**	**Kernel**	**Params**
conv1	$250 \times 250 \times 3$	$123 \times 123 \times 64$	$7 \times 7 \times 3, 2$	9K
batchnorm1	$123 \times 123 \times 64$	$123 \times 123 \times 64$		128
relu1	$123 \times 123 \times 64$	$123 \times 123 \times 64$		0
maxpool1	$123 \times 123 \times 64$	$61 \times 61 \times 64$	$2 \times 2 \times 64, 2$	0
dropout1	$61 \times 61 \times 64$	$61 \times 61 \times 64$		0
conv2	$61 \times 61 \times 64$	$61 \times 61 \times 128$	$3 \times 3 \times 64, 1$	74K
batchnorm2	$61 \times 61 \times 128$	$61 \times 61 \times 128$		256
relu2	$61 \times 61 \times 128$	$61 \times 61 \times 128$		0
maxpool2	$61 \times 61 \times 128$	$30 \times 30 \times 128$	$2 \times 2 \times 128, 2$	0
dropout2	$30 \times 30 \times 128$	$30 \times 30 \times 128$		0
conv3	$30 \times 30 \times 128$	$30 \times 30 \times 256$	$3 \times 3 \times 128, 1$	295K
batchnorm3	$30 \times 30 \times 256$	$30 \times 30 \times 256$		512
relu3	$30 \times 30 \times 256$	$30 \times 30 \times 256$		0
maxpool3	$30 \times 30 \times 256$	$15 \times 15 \times 256$	$2 \times 2 \times 128, 2$	0
dropout3	$15 \times 15 \times 256$	$15 \times 15 \times 256$		0
conv4	$15 \times 15 \times 256$	$15 \times 15 \times 64$	$3 \times 3 \times 256, 1$	148K
batchnorm4	$15 \times 15 \times 64$	$15 \times 15 \times 64$		128
relu4	$15 \times 15 \times 64$	$15 \times 15 \times 64$		0
dropout4	$15 \times 15 \times 64$	$15 \times 15 \times 64$		0
flatten	$15 \times 15 \times 64$	14400		0
fc1	14400	1024		14M
relu5	1024	1024		0
dropout5	1024	1024		0
fc2	1024	64		66K
relu5	64	64		0
dropout5	64	64		0
fc3	64	10		650
Total				15M

Table 1: Model Architecture

- Convolutions: To learn hierarchical representations of the input data, we use several convolutional layers.
- Batch Normalization: Adding these layers lead to faster convergence.
- ReLU: These are added for non-linearity, which is necessary for the universal approximation theorem to hold.
- Max Pooling: This helps make the representation become approximately invariant to small translations of the input.
- Dropout: This acts as a regularizer. We use dropout with $p = 0.2$ after the maxpool layers [**?**], and with $p = 0.5$ after the fully-connected layers [**?**].
- Optimizer: Stochastic Gradient Descent (SGD) with a learning rate of 10^{-3} and weight decay of 10^{-3}.
- Loss: We use two losses in addition: