COMP90086 Computer Vision Assignment 2

Einon McGrory-Perich 992697

September 2022

1 Question One

Before the model can be trained, the data must be preprocessed. By normalising the pixels to have a value between 0 and 1, the gradient descent algorithm will be smoother leading to more accurate results.

To train the convolutional neural network (CNN) models, found in subsections 1.1 and 1.2, the parameters found in table 1 were used.

Parameter	Value
Epochs	30
Batch Size	32

Table 1: CNN Parameters

1.1 Base Model

When training the base model without any data augmentation or regularisation, it performs quite poorly with a validation accuracy of 50%. As seen in the learning curve in figure 1 there is significant over fitting, as the training loss decreases but validation loss increases. The required structure specified was also used for both models.

1.2 Modified Model

To reduce over fitting there a number of methods that can be employed.

First, data augmentation is a useful tool where more training images can be created by performing slight modifications, such as randomly flipping the image and adjusting the contrast ratio. Only horizontal flipping was performed, as upside yoga poses will not be encountered traditionally, and a contrast factor of 0.2 was found to perform well.

The next method is to regularize the model. Regularization is a way to reduce over fitting as it reduces the magnitude of various features, so no dominant components decide the classification. Of this there are 2 forms: L1, which aims to reduce feature values to 0, and L2 which tries to make them small. It was found that L1 generally performed worse than L2, as identified by a lower accuracy. For both methods, it was found that a high scaling factor classified every image as a specific class, indicating only few features were being used. L2 regularization was chosen and

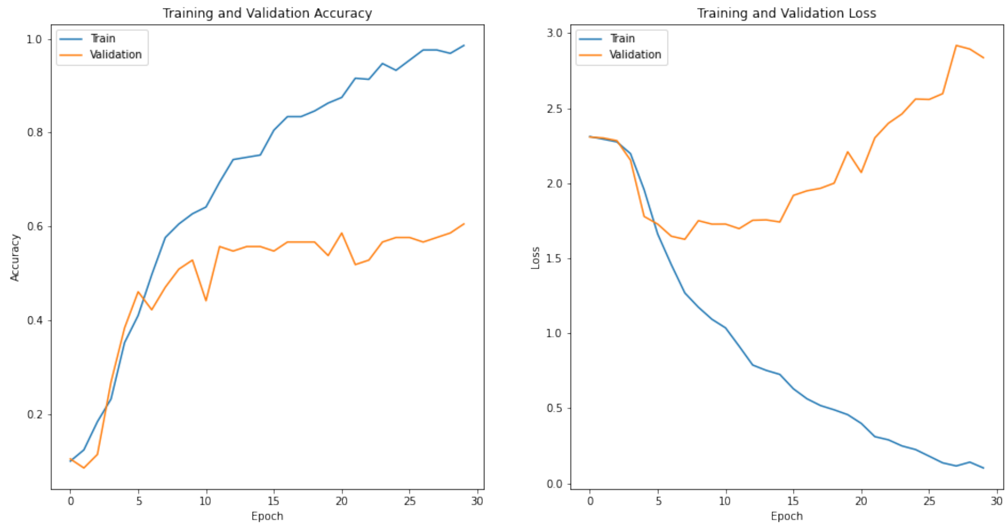

Figure 1: Base CNN Models Learning Curve

applied to the last two layers, as these have lower dimensions and improves the weights chosen for particular features. A default scaling factor of 0.02 was chosen for the second last layer, and a value of 0.01 in the last layer as it resulted in the best results.

As seen in the learning curve of the modified model in figure 2, there is less over fitting as the validation loss does not increase after a certain point, but the training loss continues to decrease.

Figure 2: Modified CNN Models Learning Curve

2 Question Two

Table 2 demonstrates the accuracy of each model on the test images. Considering that there are 10 possible classes, an accuracy of nearly 53% is not too poor, as randomly choosing would only have an accuracy of 10%. The modified model does have noticeably better performance as it has nearly a 10% improvement on accuracy.

Table 2: Accuracy of Models

Model	Accuracy
Base	52.86 %
Modified	61.43 %

When observing the individual class performance of the modified model, it's clear that some images are more successful to classify than others. The best label was found to be the Mountain pose, with an 85.71% accuracy, and the worst was the Childs pose with 14.29% accuracy.

Table 3: Modified Models Accuracy by Class

Label	Accuracy
Bridge	71.43 %
Childs	14.29 %
Downward Dog	71.43 %
Mountain	85.71 %
Plank	42.86 %
Seated Forward Bend	71.43 %
Tree	57.14 %
Triangle Pose	57.14 %
Warrior 1	71.43 %
Warrior 2	71.43 %

To understand why a class such as childs performs so badly it can be useful to extract and observe what it's actually classified as. Figures 3 and 4 show the child image being classified as seatedforwardbend both times. Both images are very similar and therefore it is very likely that most childs are classified as seatedforwardbend. Further to this, it was also found that the image found in figure 9 was found in both the training sets of the child and seatedforwardbend poses.

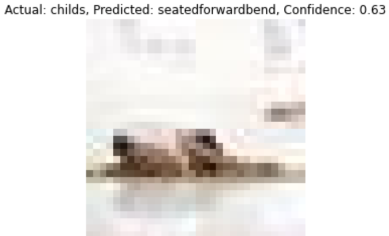

Figure 3: Child

Figure 4: Child 2

Figure 5: Mountain

Figure 5 shows the mountain having 100% confidence in the image being a mountain, demonstrating there is a particular feature in the mountain that it looks for. Another image, shown in figure 6 is classified as a tree, potentially because of the upwards stroke in the centre of the image, and the triangle pose in figure 7 was classified as a downward dog, which could be because of the triangular shape. Finally figure 8 shows an unusual plank that was not correctly classified, most likely due to the extension of limbs.

Figure 6: Downward Dog

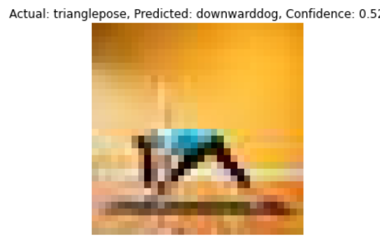

Figure 7: Triangle Pose

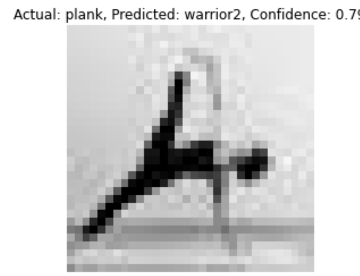

Figure 8: Plank Image

Figure 9: Image Found in Child and SeatedForwardBend Training Set

The modified model appears to perform quite well on most classes, but on images that have more features, it can be prone to incorrect classifications which could be due to the features it searches for.

3 Question Three

When evaluating the model on a specific instance, it's possible to extract the values for the features the model has identified. The most useful layer to visualise for this would be the one right before the final softmax layer, as this is the layer that weights each feature and predicts a class.

By applying the model on each test and train image, the feature values of each can be identified. The euclidean distances between each test and train image can then be calculated, and the smallest 5 distances represent the 5 nearest neighbours. Applying this approach to the model and data sets, figure 10 was produced and shows the 5 nearest neighbours for a sample of test images. The leftmost image is the test image, and the 5 to the right are the nearest neighbours.

It can clearly be seen that the nearest training images are all very similar to the test image. The first row shows 4 other mountains which is quite accurate, indicating it searches for a dark upright shape. The second row has the biggest error, as the test image does not match any of it's nearest neighbours, which explains why it was incorrectly classified as identified above. The third row appears to identify features with a diagonal shape. The last two rows show images that appear to have very similar colour profiles, which might also explain why the test image in the last row was incorrectly classified in the above section.

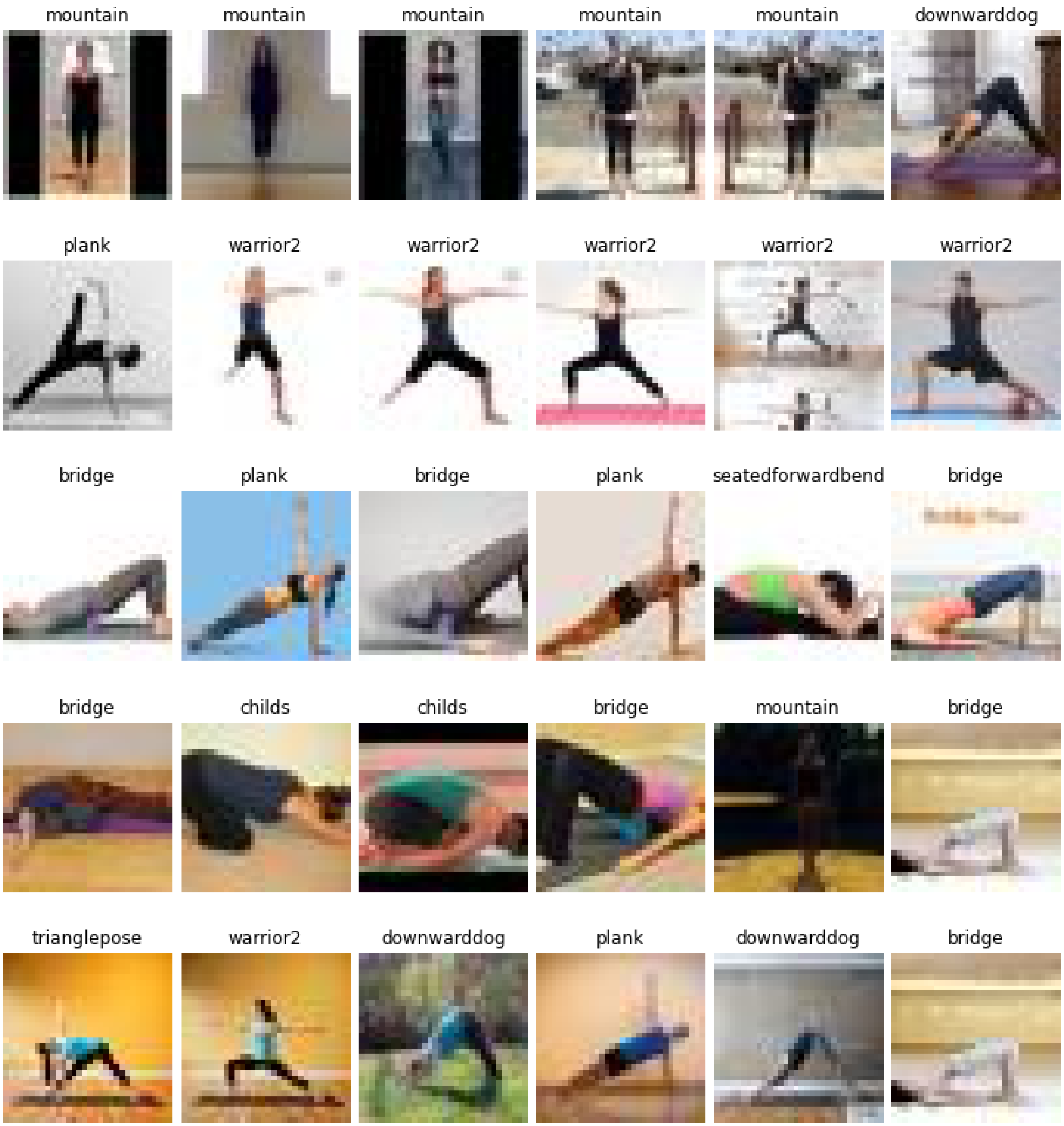

Figure 10: 5 Nearest Neighbours of Test Images

Clearly the model has learned a good feature space, as it can identify similar images based off general shapes and other properties in the image. However, the fact that colour has such an impact on the model could be a reason for poor classifications, and a potential improvement could be to include more augmentations, such as providing grey scale images. It was also seen to classify the child pose incorrectly nearly every time, which is due to it's similarity to other classes. The only difference between the two poses is the position of the legs, and therefore it can become quite hard to distinguish.

To improve the performance of this model a deeper CNN structure could be investigated. Another suitable option would be to expand the number of training images, as well as perform more augmentations to them so that it's more robust to similar images that aren't of the same class.