

1 CNN implementation

1.1 Basic architecture

Training accuracy increases over epochs, resulting in 83% accuracy after 20 epochs of fitting. Validation accuracy also increases, but at a lower rate, leading to 65% accuracy after 20 epochs of fitting. This is indicative of overfitting on the training set, as the training accuracy is higher than the validation accuracy. This divergence, which increases the more fitting is done (also indicative of overfitting), can be seen in figure 1.

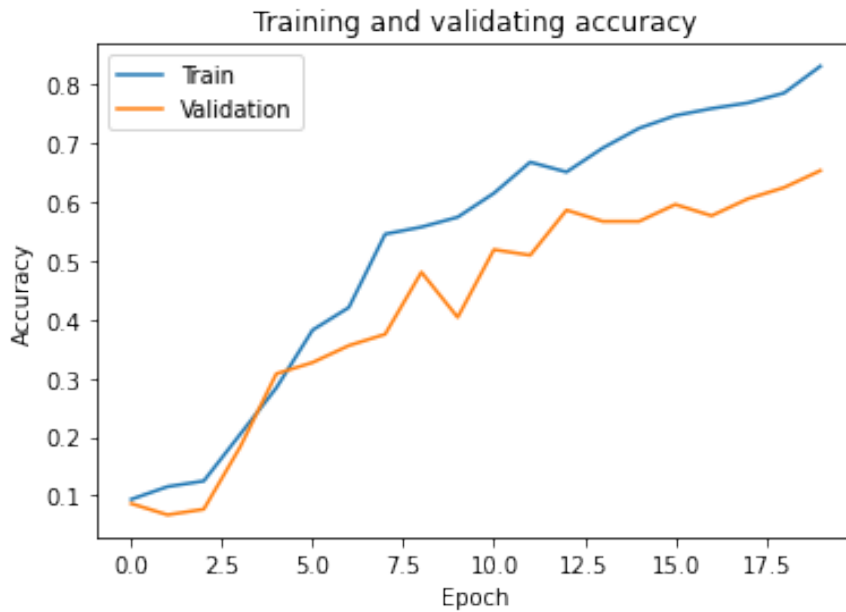


Figure 1: Divergence of training and validation accuracies on basic CNN

1.2 Regularisation and data augmentation

Regularisation was added via dropout after convolution layers, with a rate of 10%. Data augmentation was performed by adding new layers to horizontally flip, rotate, and zoom each training image with a value of 10% for both rotation and zoom. These values were chosen after ad-hoc non-exhaustive grid-search, for which the results can be seen in table 1.

Dropout rate	Rotation value	Zoom value	Training accuracy	Validation accuracy
0.1	0.1	0.1	51%	52%
0.2	0.1	0.1	52%	48%
0.3	0.1	0.1	43%	47%
0.1	0.2	0.1	33%	31%
0.1	0.3	0.1	20%	15%
0.1	0.1	0.2	50%	52%
0.1	0.1	0.3	46%	50%

Table 1: Parameter tuning results for CNN

While both training accuracy and validation accuracy has dropped in the new CNN from 83% to 51% and 65% to 52% respectively, the difference between training and validation accuracies has disappeared. This suggests overfitting is less present in the new model. This can be seen in figure 2 by how closely the training and validation accuracies track each other through the epochs. Note the difference to figure 1 where without these steps, the accuracies diverge. The overall decrease in validation accuracy is likely explained by the increase in sample size due to the data augmentation layer.

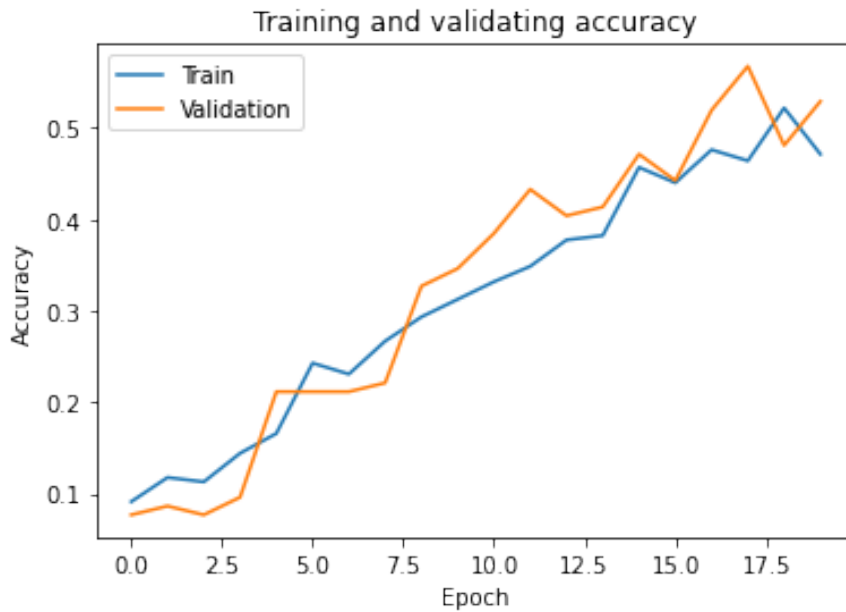


Figure 2: Divergence of training and validation accuracies on new CNN

2 Error analysis

The overall classification accuracy of the CNN on the test set was 54%. The individual accuracies for each class is presented in table 2.

Class	Accuracy
bridge	29%
childs	43%
downwarddog	57%
mountain	100%
plank	43%
seatedforwardbend	14%
tree	29%
trianglepose	71%
warrior1	71%
warrior2	86%

Table 2: Individual class accuracies on CNN

In particular, the "seatedforwardbend" class was difficult to classify, with 4/6 false negatives and 3/6 false positives attributed to the "bridge", "childs", and "downwarddog" classes. An example of the misclassification can be seen in figure 3, where it appears the model has difficulty differentiating between various horizontal-body positions.



Figure 3: Example "seatedforwardbend" classifications

Conversely, the "mountain" class was the easiest to classify, with only 2 false positives from the "tree" class. This leads more credence to the above theory, where the model seems to be strongly drawing predictions

from the overall vertical/horizontal position of the body. Examples of the "mountain" classifications can be seen in figure 4.

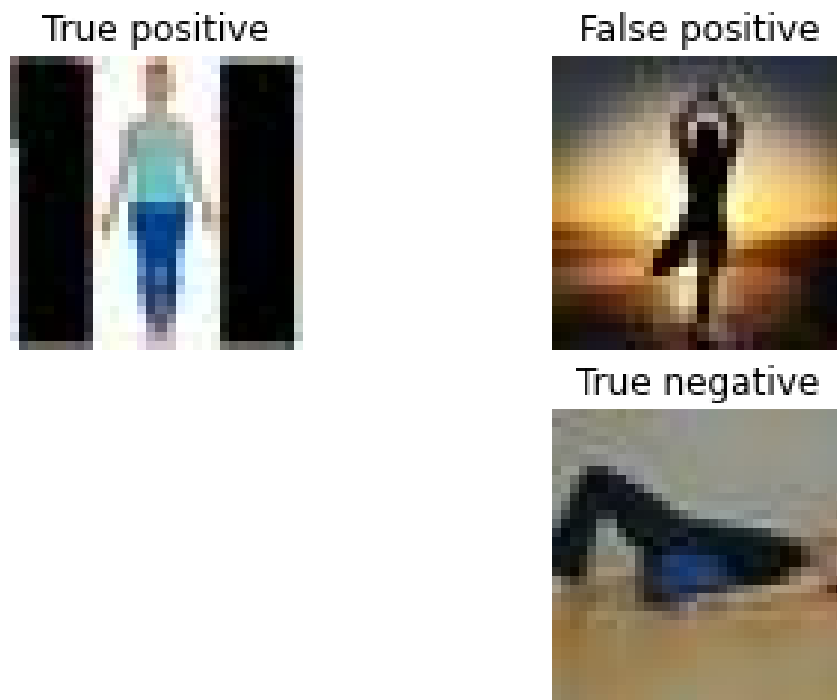


Figure 4: Example "mountain" classifications

3 Visualisation

The nearest neighbours for two test "bridge" images are shown below in figure 5 and figure 6.



Figure 5: Nearest neighbours for "bridge" test image



Figure 6: Nearest neighbours for "bridge" test image

These figures show some limitations of the CNN model - namely, over-reliance on vertical/horizontal position of the body, and inclusion of color as a factor in the predictions (most prevalent in figure 6).

The nearest neighbours for a test "tree" image is shown below in figure 7. Similarly, the model is overestimating the similarities between images with similar coloured backgrounds.



Figure 7: Nearest neighbours for "tree" test image

A complete visualisation of all test images can be found in the jupyter notebook.

Overall, the CNN model is performing much better than random guesswork, at a accuracy of 54% compared to an estimated 10% accuracy when guessing amongst 10 classes. However, as seen above, it is clearly overvaluing non-pose features such as background colours and body size relative to image. It is also not sufficiently separating poses with similar vertical/horizontal components, suggesting a lack of features focusing on limb placement. Likely an edge-detection layer would be beneficial to this CNN to strip out color biases and strengthen limb detection features.