DATA MINING

Assignment 2

NOVEMBER 28, 2022

Bastiaansen, Thijs

Jacobs, Iris

van Pinxteren, Milo

Rijkers, Niek

Thijssen, Aisja

Link to the GitHub repository: https://github.com/milovanpinxteren/DataMiningAssignment2

PART 1: CONVOLUTIONAL NEURAL NETWORKS

1.1: DESIGNING A CONVNET

For this exercise, two functions to build the convolutional neural networks were created; one for a dataset that contains pictures of the hand gestures made during the game "rock paper scissors" and one for a dataset that contains pictures of flowers. Below, the design choices and the performance of both models are be discussed.

ROCK PAPER SCISSORS (RPS) MODEL

The first model that was created tries to predict whether a player plays rock paper or scissors based on an image of the hand. Figure 1 contains a visual representation of this final model.

To create the model, first, some simple models with a very minimal number of convolutional layers were tested. With these simple models, it became clear that with the rock paper scissors dataset training and validation accuracy reached 100% after only a few epochs. However, the test set was only able to reach about 65 to 70% accuracy with these simple models. Meaning that it was very easy to overfit with this data, which was likely due to a large number of parameters and the simplicity of the images used to train the data (e.g., no background, high contrast, etc.). After trying multiple models with different numbers of convolutional layers, it was decided to use 3 convolutional layers. This yielded better performance than models with 2, 4, or 5 layers and did not make the model too complicated. The first layer includes 32 filters, and the deeper layers include 64. We experimented with including more filters in the last 2 layers. We did this because patterns get more complex in further layers, hence, there are a larger number of patterns to capture, but this decreased the training accuracy slightly so it was decided to keep 64 filters in the final layers.

Furthermore, to minimize the risk of overfitting and increase translation invariance, Max Pooling and dropout were added in every convolutional layer. The dropout rate increases by 0.1 every time you go deeper into the network and ranges from 0.2 to 0.5. Moreover, only one dense layer with 64 nodes was included since it was assumed that the test accuracy of the model would decrease when more dense layers would be added. More dense layers would mean more parameters which in turn leads to a higher chance of overfitting. In turn, this was also tested by adding extra dense layers to the model. The addition of more dense layers indeed lowered the test accuracy. Finally, to reduce the chance of overfitting a low batch size (32) was chosen (a large learning rate would have worked for this as well).

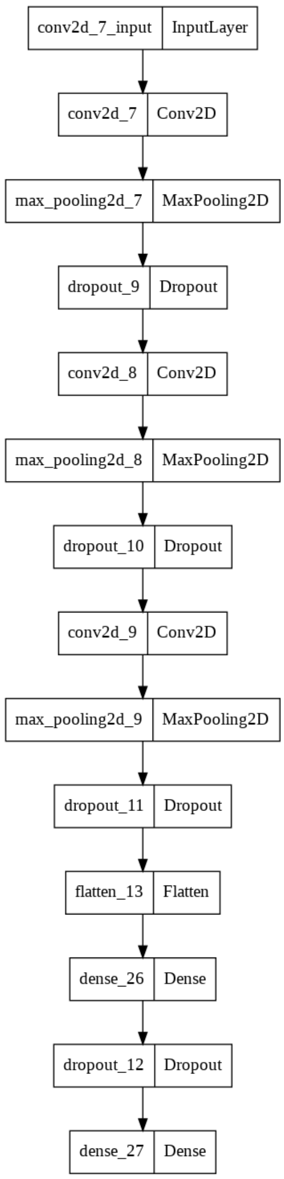


FIGURE 1: RPS MODEL

Figure 2 shows that this final model resulted in 100% training and validation accuracy in the 3rd epoch. It was considered to use early stopping since the validation accuracy was stable in later epochs. However, in the assignment, it was mentioned to use at least 5 epochs, so this is what was used. This resulted in a test accuracy of around 84%.

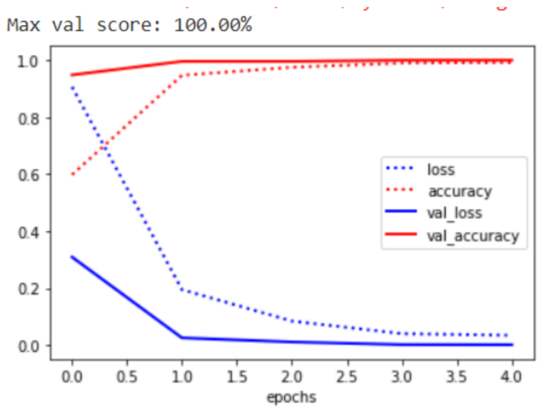


FIGURE 2: ACCURACY OF RPS MODEL PLOTTED

FLOWER MODEL

The second model aims to classify the type of flower in an image based on that image. Just like the RPS model, due to a large number of trainable parameters, the flower model is likely to overfit. For this reason, a similar approach as for the rock paper scissors dataset was used, though, there are a few key differences between the models.

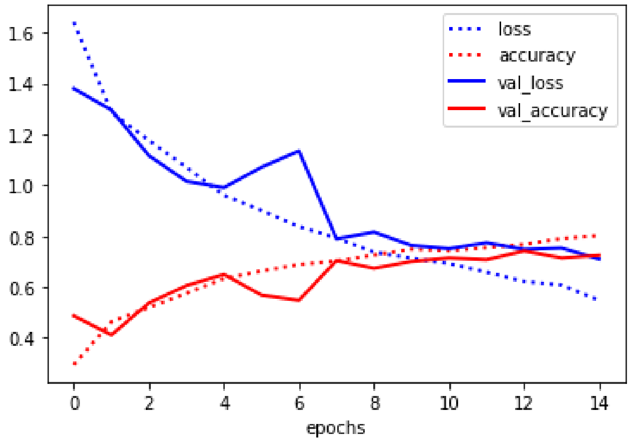


FIGURE 3: ACCURACY OF FLOWER MODEL PLOTTED

As the flower classification task considers five different classes compared to the three of the RPS task and as the images of the flower classification task are much noisier and more crowded, we expected to need a more complex model. This expectation was tested by trying both simple and more complicated models. The performance measures indicated that, indeed, a somewhat more complex model than the RPS model was needed. This led to the flower model having a 4th convolutional layer where the final two layers have 128 filters instead of 64. As mentioned before, patterns get more complex in further layers, hence, there are a larger number of patterns to capture, and therefore more filters in the two final layers increase the training and validation accuracy for this dataset. Likewise, the number of nodes in the dense layer was increased to 128. Also, for this model, 15 epochs were used instead of 5 because here the curve for the validation accuracy needed longer to flatten (see figure 3). This model had 80% accuracy on the training set, 74% accuracy on the validation set, and 74% on the test set. Hence, the flower model performs worse than the RPS model but does not seem to overfit as much. The task is more complicated and the images are much noisier with crowded backgrounds. Therefore, it is not surprising that the performance of the model is lower.

Finally, for both models, it was considered to use batch normalization to decrease overfitting. However, this lowered accuracy significantly and therefore it was discarded. This might be due to the small batch size of 32. A batch normalization layer must calculate the mean and variance to normalize the previous outputs across the batch. This will be accurate if the batch size is large, but this will be less accurate for smaller batches.

1.2: DATA AUGMENTATION

RPS MODEL

Before we started experimenting with different ways to augment the data, we hypothesized which augmentations could work for both datasets and which ones would not. For the RPS dataset, we hypothesized that rotation is not likely to work well because the rotation of the hand is not crucial for determining which gesture it represents. And as experiments showed, rotation indeed did not improve performance. However, we did think that zooming would work well because a zoomed-in paper gesture might appear as rock. When we train our model with more cases where rock and paper are difficult to distinguish, the model will become better at this; indeed this improved performance, and therefore zooming was added to the augmentation function. We also hypothesized that color augmentation would not work for the RPS model, because the dataset already included hands with different skin colors. This indeed did not work. Finally, we also tried some additional augmentations, some of which seemed to work well. Namely, width and height shift, and horizontal flip increased performance, so these were added to the augmentation function. With the augmented data, the accuracy curves took a bit longer to flatten, and therefore the number of epochs was increased to 10. Also, the training data and validation accuracy did not reach 100% like before but both reached 98%. The test accuracy reached 93% which is about 9 percentage points more than before. Concluding, for the RPS model, the data augmentation worked well.

FLOWER MODEL

For the flower model, we also constructed some hypotheses. First of all, flowers are very colorful, and the current model might use these colors to make distinctions between flower species. So, when the training data is augmented on color, the model can no longer train on the differences in color which would likely decrease performance. Indeed, performance decreased significantly with color augmentation. We also tried rotation and shear, but these did not improve the accuracy of the model. Finally, width and height shift and horizontal flip were also added to the augmentation function but these also do not improve performance as the training, validation and test accuracy are all around 70%. A reason for this could be that the current dataset already has enough variety and therefore these forms of augmentation did not improve the training.

PART 2: MODEL EVALUATION

2.1: ACCURACY ON THE TEST SET

RPS MODEL

The performance of the model was already discussed before, but in this section, we will summarize it for a better overview. The performance of the model can be observed by calculating the accuracy. This accuracy is computed with the “sklearn.metrics” package. For the RPS data, this resulted in a 98% accuracy on the train and test set, and a 93% accuracy on the validation set.

The test accuracy is a bit lower than the train and validation set, thus it can be concluded that the model is overfitting a little bit.

Flower model

Again, the performance of the flower model with the augmented data was already briefly discussed, but we will summarize it here. The flower model described above resulted in a training accuracy of 75% and a validation accuracy of 73%, whilst reaching an accuracy of 72% on the test set.

As can be seen, the accuracy is lower than the RPS model. This, however, is most likely because the dataset contains more classes and is more difficult to accurately predict. The accuracy is more or less the same for the train, test, and validation set. Thus it can be concluded that the model is neither underfitting nor overfitting.

2.2: ANALYZATION OF THE ERRORS

RPS model

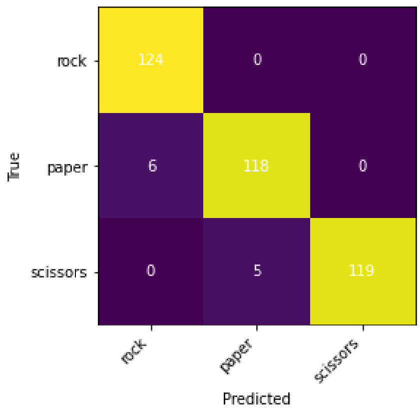


Figure 4: Confusion matrix RPS model

As can be seen in the confusion matrix in figure 4, most of the cases are classified correctly. However, 11 cases are not classified correctly. 5 of these are classified as paper, whilst the actual class is scissors, and 6 are classified as rock whilst the actual class is paper. The rock class has yielded a perfect classifying score. Zero instances of the rock class have been classified incorrectly by the model. To learn more about why some cases are misclassified, and how the cases look, 5 misclassified images are plotted for each class. However, since the rock class has no misclassified images, no images were generated for this class.

These are 5 out of the 6 misclassified images. All of them look extremely similar, with minor differences. They have the same background. Therefore, one can conclude that the background color does not influence the classification of the images. This can also be concluded since altering the colors of the images did not improve the accuracy of the model, as said before. The images contain relatively small hands, with not a lot of space between the fingers. This could be a reason for the model to think that these images are part of the rock class.

The misclassified images in the scissors class are all classified as paper. These images all have a relatively small portion of the image covered by hand. The hands are relatively small, with the thumbs being clearly visible. It could be that, since the hands are small and there is a whitespace between the thumb and hand, the model has more difficulty identifying two fingers. Therefore, the model will 'think' that the images are part of the paper class. In addition to this, the nails of the hands are polished with yellow, pink, and blue. This might influence the model.

FLOWER MODEL

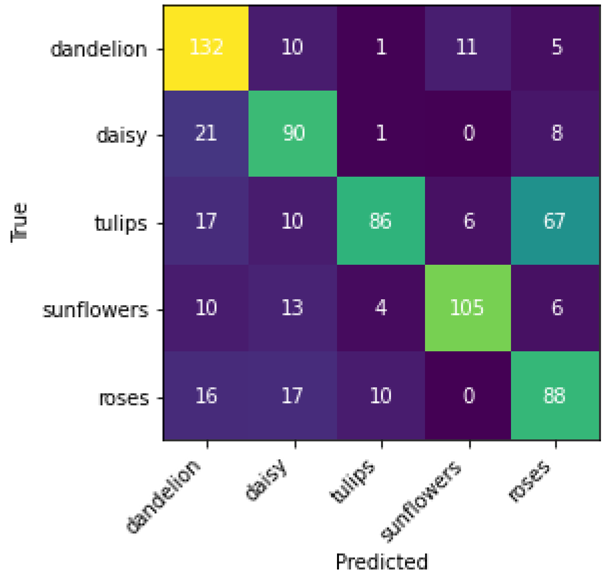


FIGURE 5: CONFUSION MATRIX FLOWER MODEL

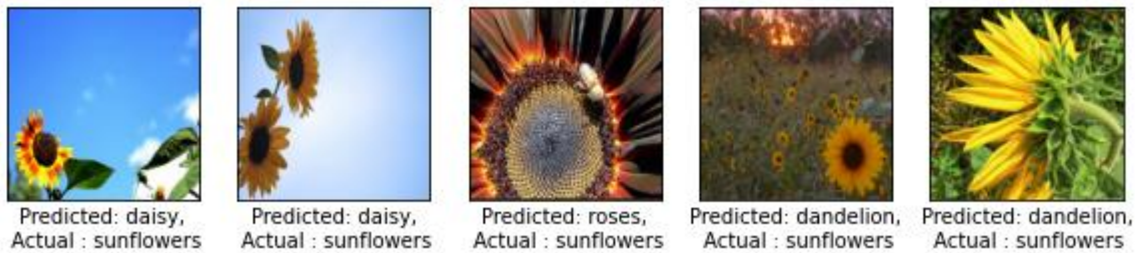
As can be seen, most cases are classified correctly and there is quite a lot of spread in the misclassified cases. It stands out that the images of tulips are very often classified as roses, compared to the other images. When the model classifies an image as a sunflower, it is right almost all the time. The flowers that are most often misclassified are tulips and roses.

Plotted below are 5 misclassified images for each class. There is a lot more color and difference in the images compared to the RPS dataset. This would make it more difficult to identify structural flaws in the classification of the images. However, some interesting insights can be retrieved when inspecting the images.

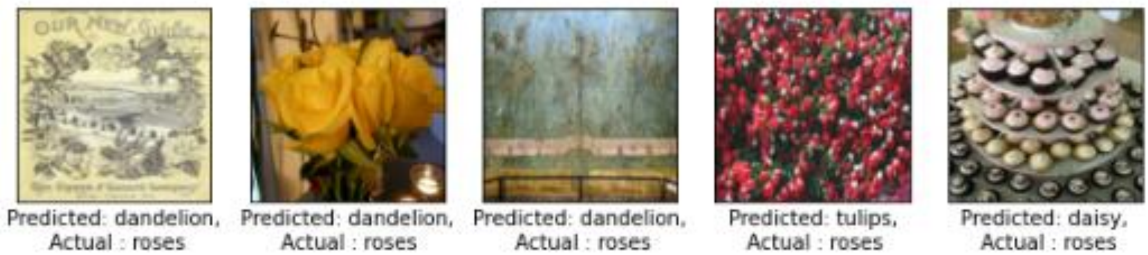
The model predicts that the dandelion images are of sunflowers if the dandelion is in the flower stadium and has a yellow color. Some images, like the middle image, are vague and show colors different from what one would expect from a dandelion, the background overpowers the image. This would make it very hard to predict.

Most of the misclassified cases of the daisy class are classified as dandelion (⅗). This could be due to the white nature of the flower.

All of these images are (mis)classified as roses whilst being tulips. Even though the two flower classes are not in the same family or subclass, the two are alike. Even humans have trouble distinguishing tulips and roses (Plantsheaven. (n.d.)).



Sunflowers are correctly classified the most out of all the flowers, as can be seen in the confusion matrix. The pictures with a lot of blue sky, or where the sunflowers are photographed from a distance are often classified as a daisy. This is most likely since daisies are quite small and often photographed in fields from below. The middle image is an unusual representation of a sunflower since the flower lost most of its characteristics and color. This would make it very hard to classify anyway for a model.



Roses come in all colors, shapes, and sizes. The rose has over 300 subspecies (Mabberley, 1997). This makes it difficult to classify. In addition to this, the pictures shown here are not typical representations of roses. One is modified to look grey, one seems to be a postcard of some sort. One looks like a bridal cake. This would make it very difficult to distinguish a rose from another flower. Even most humans would have trouble classifying these as roses.

PART 3: TRANSFER LEARNING

3.1: TRANSFER LEARNING FROM MOBILENET

In this section, the design choices for the RPS model and flower model that use the convolutional base of the MobileNetV2 model are discussed. Also, the performance of the other pre-trained models will be evaluated.

RPS MODEL

First, a model was constructed that imported the convolutional base from MobileNetV2 with GlobalAveragePooling and one dense hidden layer, and one output layer. The model uses the augmented data with 10 epochs and reaches 98% training and 100% validation accuracy and has 85% accuracy in the test set, which is a worse score than before (the test score was 93% on augmented data) so it is also overfitting more. To fine-tune this model first it was tried to unfreeze some of the last convolutional layers, however, this decreased validation accuracy significantly while the training accuracy stayed the same. So, the model started to overfit more when more layers were unfrozen and therefore the convolutional base was kept frozen. Finally, to minimize overfitting, dropout was added after the convolutional base and after the dense hidden layer. This resulted in the model depicted in figure 4, which had 100% accuracy in the training, validation, and test set, which is an improvement to the first model used.

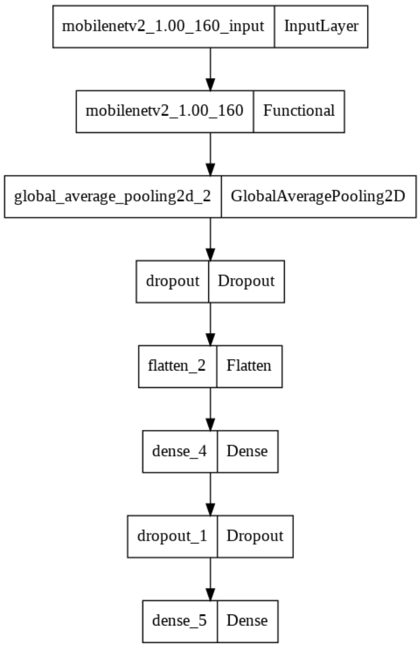


FIGURE 6: FINE-TUNED RPS MODEL

In addition to MobileNetV2, two other pre-trained models that were listed on the Keras website were explored. The first pre-trained model that was explored was DenseNet121. Just like the other RPS models, this model scored 100% on training and validation accuracy after only a few epochs. The test set accuracy reached about 87% accuracy, hence it was overfitting. The second pre-trained model that was explored was MobileNet. This model reaches 100% training and validation accuracy, but only 86% on the test set. So, the DenseNet121 and the MobileNet model perform about the same as the MobileNetV2 model on the RPS dataset without fine-tuning.

FLOWER MODEL

For the flower dataset, the model with the convolutional base from MobileNetV2 had 69% accuracy in the training and validation set, and 66% accuracy on the test set. To fine-tune this model, the final layer of the convolutional base was unfrozen. This actually improved the training accuracy to 87%, while the validation accuracy stayed roughly the same around 70%, thus this model was now overfitting. Because of this, dropout was added after the dense layer which increased the validation and test accuracy to 85%. Unfreezing more layers caused even more overfitting so for that reason the rest of the convolutional base was kept frozen.

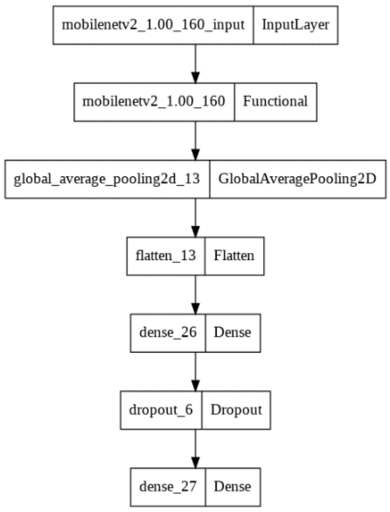


FIGURE 7: FINE-TUNED FLOWER MODEL

Just like for the RPS dataset, the pre-trained models DenseNet121 and MobileNet were explored. With the convolutional base of DenseNet121, the model scores 87% on the training set and 90% on the validation set, so it does not seem to overfit. The accuracy on the test set is also 90%, so this model performs better than the model that uses the convolutional base of MobileNetV2. With the convolutional base of MobileNet, the model reaches just below 90% validation accuracy, but the training accuracy starts getting higher at 5 epochs and thus starts overfitting. This model also had 90% accuracy on the test set and therefore also performs better than the model with MobileNetV2.

3.2: VISUALIZING THE LEARNED EMBEDDINGS WITH TSNE

RPS MODEL

The following plots show the 2D TSNE map for the rock paper scissors model (figure 8) and the flower model (figure 9)

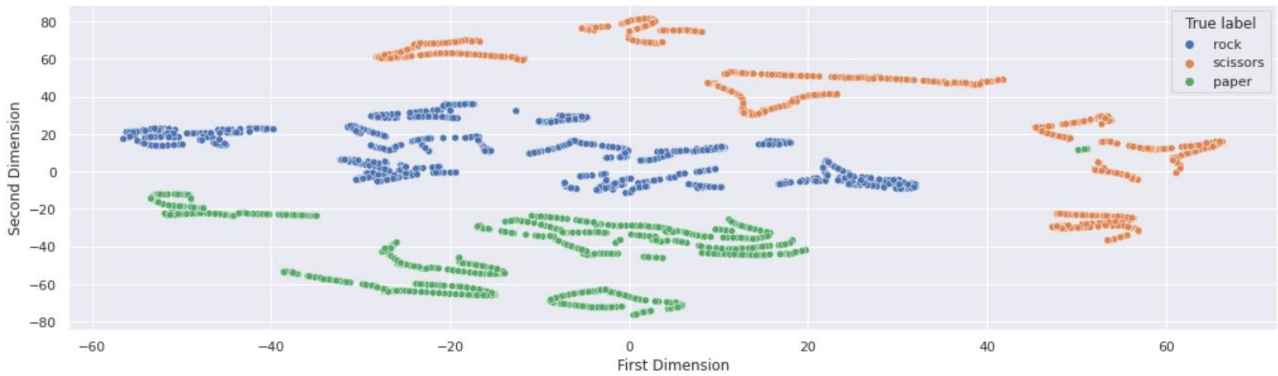


FIGURE 8: 2D TSNE MAP FOR THE RPS MODEL

As can be read earlier in the report, the model predicting the class of the rock paper scissors dataset has a very high score. The reason for this is that very few pictures are misclassified. When looking at figure 8, we can see that classes are clustered together. This makes it easier to predict and therefore classify. In the cluster of the scissors class, there are a few green dots. These are misclassified. There seems to be a small subset of very similar pictures which are misclassified. Besides this, clear separations can be made between the clusters, which is in accordance with the learned embedding.

FLOWER MODEL

FIGURE 9: 2D TSNE MAP FOR THE FLOWER MODEL

For the flower dataset, the number of classes is higher. Nevertheless, there are clusters, all be it with overlap among these clusters. When looking at the model performance, this makes sense. The model is not as good as the model for the rock paper scissors dataset, however, the model is still quite good.

BONUS

Due to time constraints, successful implementation of the bonus is not reached.

REFERENCES

Plantsheaven. (n.d.). *Tulip Vs Rose: What's The Difference Between Both? | Plants Heaven*. Retrieved November 28, 2022, from https://plantsheaven.com/tulip-vs-rose/

Mabberley, D. J. (1997). *The Plant-Book: A Portable Dictionary Of The Vascular Plants* (2 Rev ed). Cambridge University Press.