Pizza or Donuts: Food Image Classification with VGG16

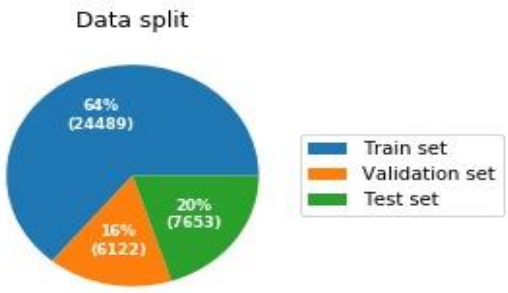
Starring: Ruchella Kock & Nynke van ‘t Hof (Group 18)

Introduction

- Over the last few years, many advancements have been made in the field of Deep Learning and Convolutional Neural Networks. These can be used for various tasks [1].
- Food recognition is such as task that could benefit from Deep Learning and CNN because of the wide variety of foods.
- Food image classification and recognition are crucial steps for dietary assessment[2].
- Several food classification models have been trained, starting out with the pretrained model from ImageNet (transfer learning).
- ImageNet uses more than 14 million hand-annotated images with various objects in them[3].

Methods

Data



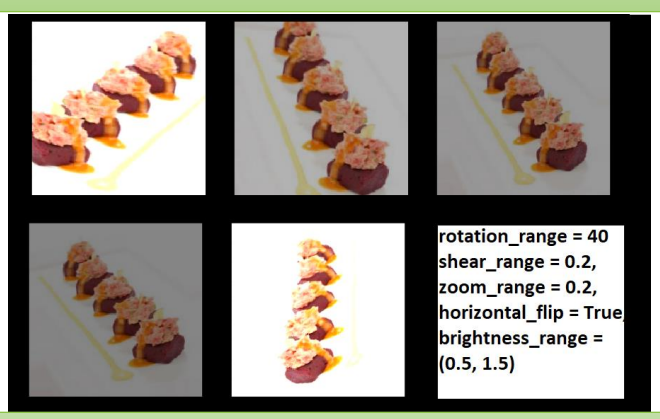
Experimenting

We experimented with different parameters:

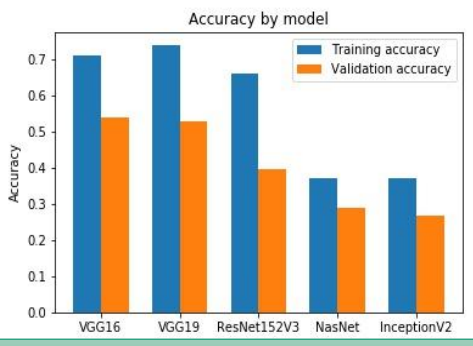
- Multiple Models
- Different frozen and trainable layers
- Different Layers
- Data augmentation
- Different image input sizes
- Different activation functions
- Different Batch sizes & Epochs
- Ensembling

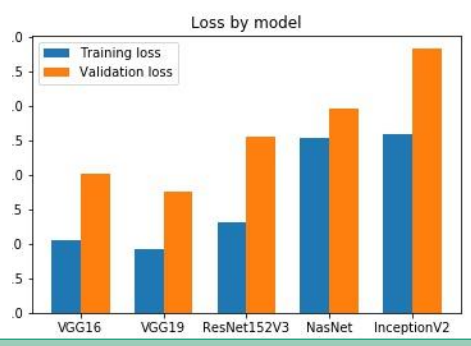
A few of these experiments are visualized hereunder

Data Augmentation



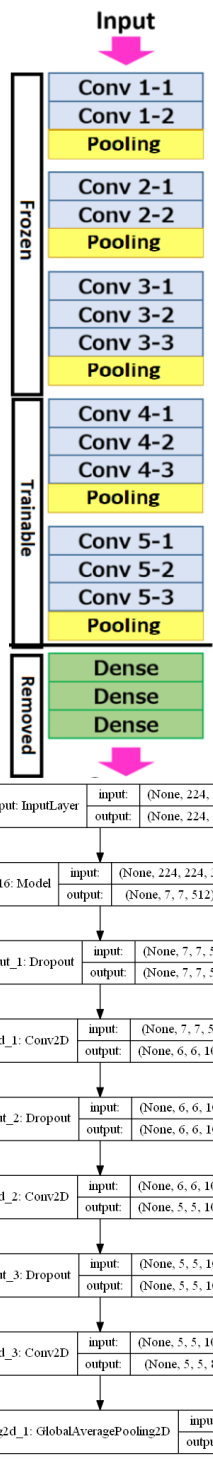
Multiple Models





Best Model

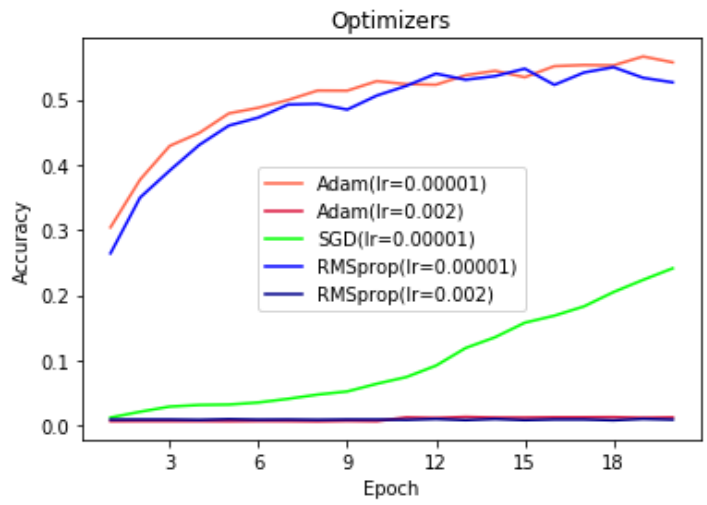
VGG16



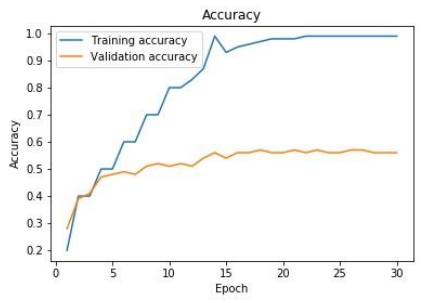
Parameters

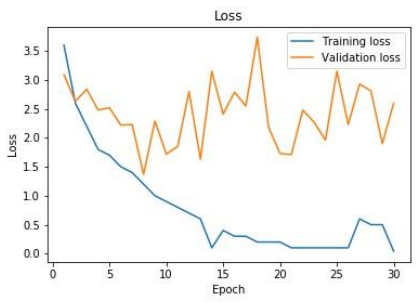
Optimization:
SGD(lr=0.002)
Loss:
Categorical crossentropy
Input size:
(244,244,3)
Batch size:
32
Epochs:
20

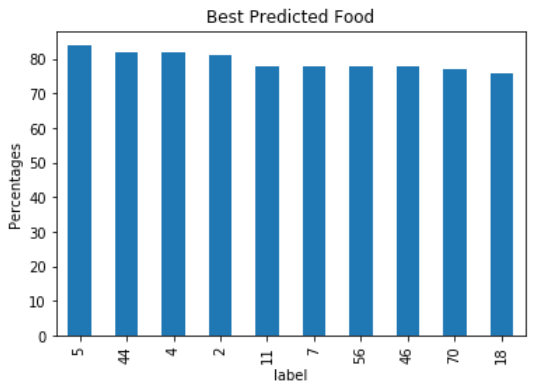
Optimizers



Results VGG16







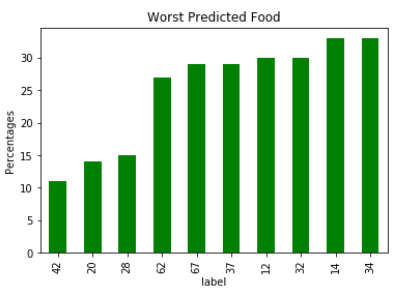
Best Predicted food

5: Quail Eggs
4: Salad
44: Pasta
2: Bread
46: Edamame
56: Gingerbread cookie
7: Chocolate
11: Stuffed bell peppers
70: Guacamole
18: Squid rings

- Most pictures in the categories look alike, have similar color schemes and basic shapes.

Worst predicted food

42: Melted cheese/sauce
20: Cake with wave pattern
28: Banquet
62: Tenderloin steak
37: Crab
67: Scrapple/ Rabbit
32:Hagis
12:Duck
34:Beef
14:Ambrosia salad



- Pictures vary a lot in these categories because of many different ways to prepare the dishes

Limitations

- Majority voting Ensembling possibly not complex enough
- Model computationally heavy
- We couldn’t experiment enough with many hyperparameters and different models in the given time

Future work

- Try other Ensembling method e.g. Boosting
- Explore the weaknesses from the model e.g. by investigating what was often wrongly classified and try to improve upon the model
- More dropouts or regularization to avoid overfitting

Conclusion

- In a short amount of time, it is possible to classify 80 classes with an accuracy of 58%, which is significantly better than randomness.
- However, with more attention to prevention of overfitting and more attentions to hyperparameter tuning the task could be performed even better.

References

[1] Singla, A., Yuan, L. and Ebrahimi, T., ‘Food/Non-food Image Classification and Food Categorization using Pre-Trained GoogLeNet Model’, *Proceedings of the 2nd International Workshop on Multimedia Assisted Dietary Management* (2016) 3-11.
[2] Zhu, F., Bosch, M., Schap, T., Khanna, N., Ebert, D.S., Boushey, C.J. and Delp, E.J., ‘Segmentation Assisted Food Classification for Dietary Assessment’, *spiedigitallibrary* (2011).
[3] http://www.image-net.org/