1. Egg Classifier Model, Matheus Pupp de Araujo Rosa

2. Abstract

Egg image classifiers were attempted on different occasions and were relatively successful in their tasks. With that in mind I started coding a CNN model with VGG19 architecture with all layers frozen and only the final classification layers were trained. I used ChatGPT and Gemini AI tools to help write and troubleshoot the code. Got the base model to work with a weighted average precision of 0.75, a recall of 0.57, and an F1-score of 0.53. The training set got an accuracy of 0.3369 and loss of 1.3489. As for the validation set, it got an accuracy of 0.5867 and a loss of 1.2481, both being better than the training set, so I should be able go forward to run the unfrozen model.

3. Introduction

It is known that we will always have some degree of problems related to egg abnormalities [2], especially when we think about contamination, and for that reason it is made the egg classification, to better deal with each type of chicken egg. Some genetic companies like to separate those into 3 categories, clean, bed clean or nest dirty and bed eggs, but that classification does not account for egg cracks, which can be an entrance point for contamination in the eggs.

Egg quality is a crucial factor in food production and safety. With that in mind, my idea was to account for the egg cracks and their dirtiness using a CNN model to identify and classify them into 4 different categories (Healthy, Dirty, Cracked, Dirty and Cracked) to better help farm workers deal with the handleling of this situation.

The plan is to follow what was done by Moreno et al. [1], and with the help of AI tools, code a VGG19 architecture CNN model, that gets as close as possible to their results, with an accuracy of 93% being the ideal.

The VGG models are a family of deep learning models designed for image classification. They use a series of small 3×3 convolutional filters stacked in deep layers to extract features from images. The most common version, VGG19, has 19 layers and was trained on millions of images (ImageNet). It's popular because it’s simple, deep, and very effective at recognizing image patterns like edges, shapes, and textures.

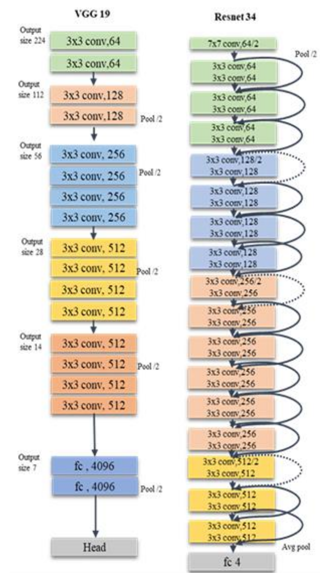

Fig. 1. Structure of the VGG 19 and Resnet 34 models [1]

4. Background

For this project, the work of Moreno and his team on the paper entitled “Poultry Egg Classification System Using Deep Learning” was used as a baseline to what I wanted my model to look like, and for that I asked the paper team for their dataset, which they kindly gave me access to, since it would not be possible for me to gather by myself all that information in that short amount of time.

I will use other papers to better understand different models used for that same task, or similar task, to be able to better wright this report.

Another very important aspect of my project will be my personal experience in the years of the Veterinary Medicine graduation back in Brazil on

UFSM (Federal University of Santa Maria), which gave me the opportunity to gather a lot of experience in the poultry field.

5. Approach

VGG19, proposed by Simonyan and Zisserman (2015) [3], is widely used for image classification tasks and benefits from its deep architecture and simplicity. Transfer learning with pre-trained models has been shown to significantly reduce training time and improve performance, especially in domain-specific problems with limited datasets.

I started by asking ChatGPT to generate the codes for a CNN VGG19 model for classification of eggs into four different categories (Healthy, Dirty, Cracked, Dirty and Cracked), giving as input the details of what I needed the model to be and also asking it to get base information from the paper I am using as a guideline/base for my project. I had to troubleshoot the way to get to the dataset that the authors from the base paper gave me access to, since Collab was not able to directly get it, so I needed to download it in a ZIP file and extract it into my own drive account (which I needed to buy more space because I was running out of it and the dataset is large). I used both ChatGPT and Gemini to troubleshoot any errors.

With what ChatGPT and Gemini suggested, my code will end up with 11 steps, them being: mounting Google Drive, setup and import the necessary Python libraries, data loading and preprocessing (with class table output), base VGG19 model (frozen), train the base model, plot training history, evaluate base model, fine-tune VGG19, plot fine-tuning history & compare, evaluate fine-tuned model, and accuracy improvement summary.

I used a VGG19 model pre-trained on ImageNet. All convolutional layers are initially frozen to use the learned features, and only a new dense classifier is trained. The dataset consists of images divided into train, validation, and test folders. Each image is resized to 224x224 pixels to match VGG19 input requirements. We use categorical crossentropy as the loss function and Adam as the optimizer. Training is done over 10 epochs with a batch size of 64 [1].

The evaluation parameters used followed almost the same ones used in the base paper [1], them being:

- Precision - proportion of correct predictions out of all predicted instances for a class. High precision means few false positives;
- Recall - proportion of correct predictions out of all actual instances for a class. High recall means few false negatives;
- F1-score - harmonic mean of precision and recall. Useful when class imbalance exists.

Image augmentation has not yet been applied. Training and validation accuracies are monitored to detect underfitting or overfitting. All experiments were conducted on Google Colab using GPU acceleration.

5. Preliminary Results

I was able to run and troubleshoot up to the evaluation of the base model step. The model got a weighted average accuracy of 53% on the validation set after 10 epochs, which outperforms the random baseline of 25% for four classes (fig. 2). The classification report showed relatively balanced performance across all categories, and the "support" column confirmed that class distribution is relatively even (fig. 3). Weighted averages for evaluation parameters were, as seen on image 2:

- Precision = 75%;
- Recall = 57%;
- F1-score = 53%.

Training and validation accuracy curves indicated continued learning without clear signs of overfitting (fig. 4).

$$\text{Random Accuracy} = \frac{1}{\text{Number of Classes}} = \frac{1}{4} = 0.25 \text{ or } 25\%$$

Fig. 2. Random accuracy formula

Classification Report (Base VGG19):

	precision	recall	f1-score	support
cracked	0.49	0.95	0.64	41
cracked and dirty	0.57	0.79	0.66	38
dirty	1.00	0.28	0.43	36
healthy	1.00	0.20	0.33	35
accuracy			0.57	150
macro avg	0.76	0.55	0.52	150
weighted avg	0.75	0.57	0.53	150

Fig. 3. Classification report scores

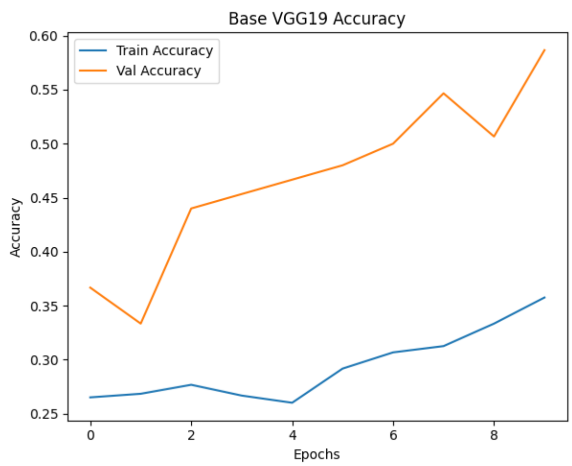

Fig. 4. Base VGG19 accuracy plot

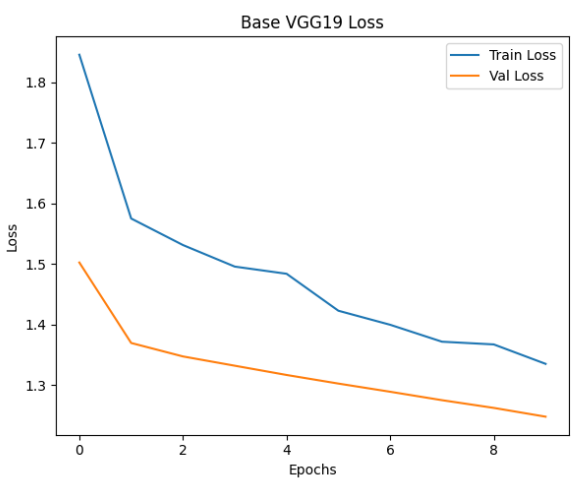

Fig. 5. Base VGG19 loss plot

According to the confusion matrix (fig. 6), I should be more attentive with the class pairs “dirty/cracked” and “dirty/cracked and dirty”, which makes sense since these are harder aspects/classes to distinguish from one another, sometimes even for the human trained eye. This class pairs will probably have be the target for a better feature separation during fine-tuning.

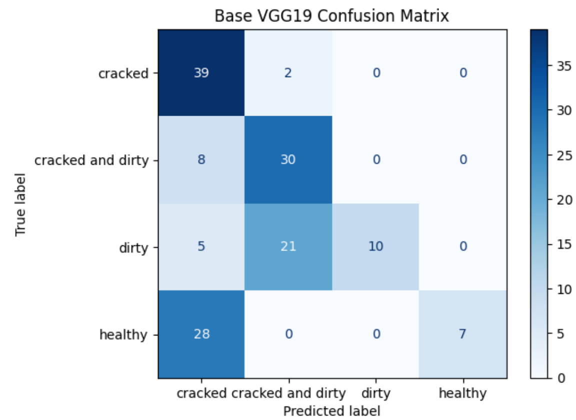

Fig. 6. Confusion Matrix of base model

6. Conclusion

This initial performance supports the hypothesis that the model has learned useful features from the dataset, despite frozen convolutional layers. The next step will involve unfreezing the top convolutional blocks of VGG19 and fine-tuning them to the egg classification task. I expect this to significantly improve accuracy, as indicated in the reference study.

References

[1] Sánchez, J. M. P., Moreno, L. A. O., Rodríguez, J. L. R., & Corro, I. D. M. (2023). Poultry Egg Classification System Using Deep Learning (pp. 1–6). *2023 5th International Congress on Human-Computer Interaction, Optimization and Robotic Applications (HORA)*. IEEE. https://doi.org/10.1109/HORA58378.2023.10156776

[2] Dunkley C. S. A Dozen Egg Abnormalities: HOW THEY AFFECT EGG QUALITY, UGA Cooperative Extension Circular 1255, April 2022. Available at: <https://secure.caes.uga.edu/extension/publications/files/pdf/C%201255_1.PDF>

[3] Simonyan K., Zisserman A. VERY DEEP CONVOLUTIONAL NETWORKS FOR LARGE-SCALE IMAGE RECOGNITION, Published as a conference paper at ICLR, 2015. https://doi.org/10.48550/arXiv.1409.1556.

[4] Victor Massaki Nakaguchi, R.M. Rasika D. Abeyrathna, Tofael Ahamed. Development of a new grading system for quail eggs using a deep learning-based machine vision system, Computers and Electronics in Agriculture, Volume 226, 2024, 109433, ISSN 0168-1699. https://doi.org/10.1016/j.compag.2024.109433.

Çelik A., Tekin El. Classification of Hatchery Eggs Using a Machine Learning Algorithm Based on Image Processing Methods: A Comparative Study, Brazilian Journal of Poultry Science, e-ISSN: 1806-9061 2024 / v.26 / n.2 / 001-012. http://dx.doi.org/10.1590/1806-9061-2023-1882.