Cow Teat Keratosis level identification using SOTA Image algorithms

Haider Ali, Yeshiva Univesity
Hali4@mail.yu.edu

Abstract

In this paper, I've tried several models to achieve better accuracy than the VGG16 model, which is 66.8 accurate in identifying cow teat thickness /roughness. Classes range from 1 (low roughness/thickness) to 4 (highest roughness/thickness). The health of the cow's teat is essential to identify the cow's health at the early stage and the milk quality, which can help to treat the cow better. I've used the teat images dataset from Zhang's Teat classification dataset [2] to identify teat thickness or roughness (Hyperkeratosis). I started with a basic CNN model with 10 CNN + Batchnorm + Relu layers to understand how the model behaves on an imbalanced dataset. I reached 48 percent accuracy. Then I moved to the VGG19 model, and even after many trying combinations of hyperparameters, it was getting stuck at local minima and gave an accuracy of 50 percent. I then tried ResNet18 and ResNet 34, finding that ResNet34 was overfitting and ResNet18 was 55 percent accurate. My final model consists of the ResNet18 + CNN model + Linear, which leads to 61 percent accuracy.

1. Introduction

Teat Hyperkeratosis is a very complex problem to tackle. Deep learning models can help identify it easily and help prevent end products and help teat the cow better in an earlier stage.

I started with a basic CNN model and found that the model accuracy stayed the same. To improve accuracy, I started with VGG 16 and VGG 19 and found that they were stuck in a local minimum. Then I focused on deeper models, which can learn more than just forward layers, i.e., ResNets, and found that they improve accuracy. Still, as we go deeper, they started to lose information, also known as the vanishing gradients problem. I kept my focus on basic models and tried to preserve the information.

2. Related Work

Previously, using Separable Confident Transductive learning [1], a group of members from Cornell university achieved 77.6 accuracy using GoogleNet, which was

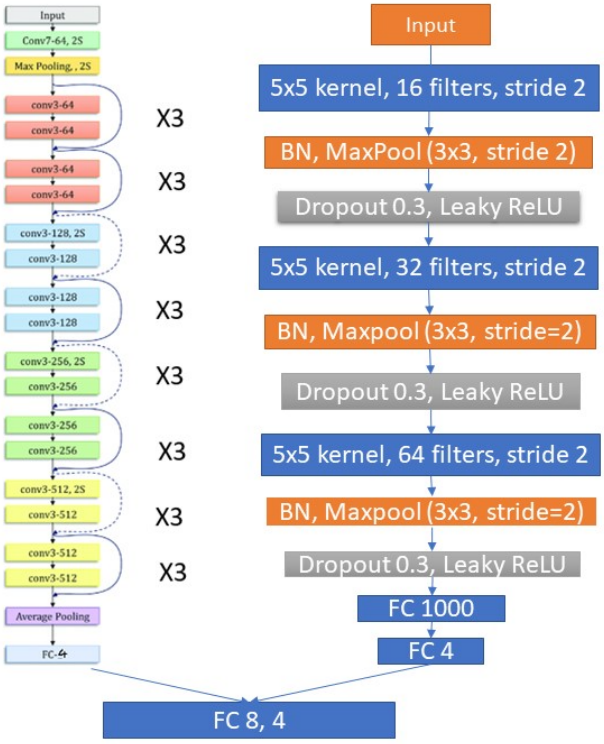

Figure 1. Proposed CNN model: ResNet18 + 3CNN + Linear

13.4 more accurate than the ImageNet dataset. Transductive learning is a type of semi-supervised learning technique that trains using both labeled training data and unlabeled test data. They used pseudo labeling and transfer learning with separable loss to achieve the above accuracy.

3. Methods

I realized that the dataset needs to be balanced using oversampling methods or different strategies to handle an imbalanced dataset. To deal with overfitting, I used OTSU's thresholding method using the mean of each image to help models extract important features, inverse class weights as weightage parameters in the Cross entropy loss, and Focal loss. Using label smoothing, the model gets stuck at a local minimum.

Below are the Cross-Entropy loss and Focal loss functions. For the final model, I've used the combined weighted loss as Focal loss (0.6) and Cross Entropy loss (0.4).

$$\text{Cross Entropy Loss} = -\sum_{c=1}^{M} y_{o,c} \log(p_{o,c}), \qquad (1)$$

$$\text{Focal Loss} = -\sum_{c=1}^{M} y_{o,c} \gamma t (1 - pt) log(pt), \qquad (2)$$

where Y_i is the original value, $\hat{Y}_i$ is the prediction value. In Focal loss, pt is the cross-entropy loss, and gamma is the weightage parameter.

4. Dataset

The dataset consists of 1529 images, where 1149 images were train images and 380 were test images. I've further bifurcated train images into trainset of 919 images and 230 valid images.

The dataset has four classes with class-to-image distribution: Score 1: 0.6083551, Score 2: 0.5726719, Score 3: 0.83724976, Score 4: 0.98172325. It can be seen that the dataset is imbalanced.

5. Results

Here are the losses for the final model

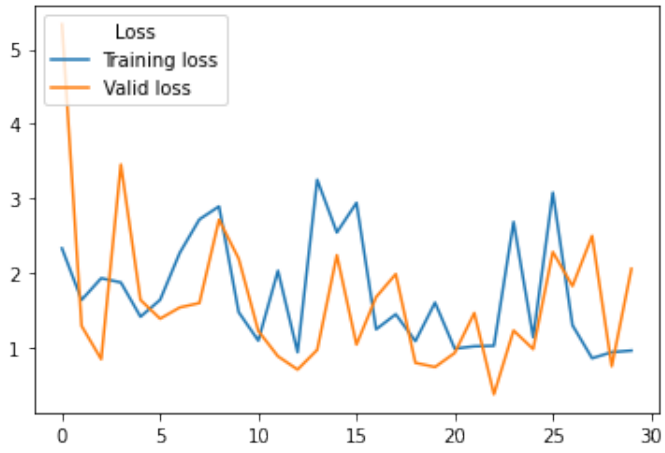

Figure 2. Figure 2. Final model loss plot

Here are the accuracies for the final model

6. Discussion

The result of this study suggests that the basic SOTA models, such as VGG models and ResNet with basic augmentations techniques and loss functions, can only go until 61 accuracy.

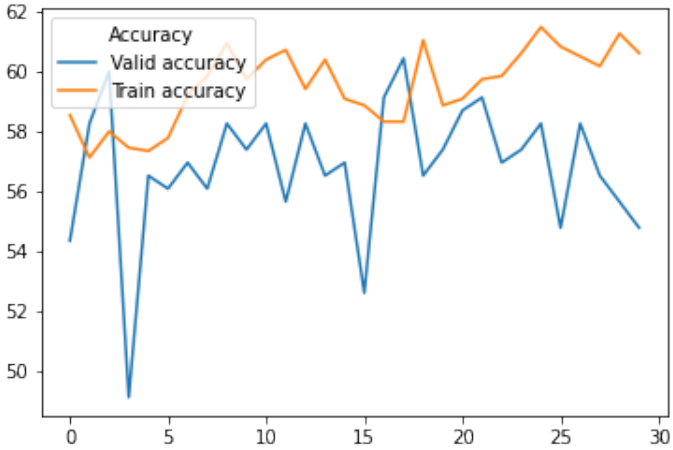

Figure 3. Figure 3: Final model accuracy plot

Table 1. Comparison of models

Model	Train accuracy	Test accuracy
VGG 19	0.48	0.46
ResNet 18	0.52	0.53
ResNet 34	0.60	0.52
ResNet18+CNN+Linear	0.56	0.61

I've reshaped images to a 256x256 shape and applied random rotation with 20 degrees, Random Erasing with an erasing range of 0.02 - 0.1, and finally used OTSU's mean thresholding using the mean of each image.

I've tried VGG19, ResNet18, and ResNet34. I also tried to build a meta-model with VGG19 and ResNet18 model outputs. This also gave me only 55 percent accuracy, as I inferred that both models were competing. So my final model was adding a CNN model to help ResNet18 increase accuracy.

I used Cross Entropy Loss with inverse class weights, NLL loss, and Focal Loss with gamma values of 2 and 2.5. CEL and Focal Loss with gamma values of 2.5 performed similarly. I finalized the Focal Loss because I found the loss slope steeper than CEL.

I've used Adam and SGD as optimizers with different batch sizes and loss functions. I found Adam the best performer, with a batch size of 8, Cross entropy loss, and Focal loss with gamma 2.5.

Per my experimentations, I can infer that loss functions and transductive learning can greatly boost the model's performance as the model is trained using the test dataset.

7. Conclusion

I started the project to get more than 66 percent accuracy greater than VGG 16 model to classify teat images. I've run all the models for 40 epochs with Adam optimizer,

Cross entropy, and Focal loss. I've tried this using multiple state-of-the-art models, but only applying the models or increasing layers won't help increase the accuracy.

There are lots of scope for improvement. I want to use semi-supervised learning and learn how current models work. I want to use CutMix, and CutNet paper and see if VGG 16 and 19 still overfit the dataset. Creating a synthetic dataset for under-labeled classes to improve accuracy. Applying GoogleNet and Inception models to get more accuracy.

I want to research and expand my vision and find more strategies and methods to increase accuracy and reach 90+ accuracy.

References

[1] Youshan Zhang, Parminder S Basran, Ian R Porter, and Matthias Wieland. Dairy cows teat-end condition classification using separable transductive learning. In *61st National Mastitis Council (NMC) Annual Meeting*, 2022. 1

[2] Youshan Zhang, Ian R Porter, Matthias Wieland, and Parminder S Basran. Separable confident transductive learning for dairy cows teat-end condition classification. *Animals*, 12(7):886, 2022. 1