CSCI-SHU 360: Machine Learning

Final Competition Report - Fall 2024

VIT and Binary Model

Zhuolin Li
zl4241@nyu.edu

Abstract

1. *I train a binary classifier for AI-generated images and perform a cleaning. (based on ResNet 50 framework)*
2. *I train the VIT model to make predictions based on the cleaned data set.*
3. *I perform another cleaning based on my previous VIT model to correct the mislabeled AI images.*
4. *Re-Train the VIT model based on the refined dataset.*
5. *Find that the Ming and Qing images are hard to classify within the VIT model from the confusion matrix, train another binary model for the Ming and Qing images. Also apply fine-tuning threshold according to the confusion matrix and the report.*

My method of training performs a score of 0.7660 after training the VIT model on the refined data set. My Avg-AI is not that good compared to the other at the same level. After tuning the threshold, my model performance increases.

1 Introduction

The data set consists of 3600 images while 3200 of them are training data sets. Among those images of training dataset, we have around **3% Tang dynasty images, 18.6% Song images, 8.7% Yuan images, 18.6% Ming images, 31.3% Qing images and 12.5% AI-generated images**. There are also **400 mislabeled AI-generated images** within the dataset.
The task is to make prediction for the images' label in the test dataset.
Metrics are consisting of overall acuracy, f1 scores for AI and non AI images; the **confusion matrix** is also a great tool.
Below are my findings for features that is helpful for identifying the images' dynasty.

1. Tang Dynasty: Many paintings features human, they look **fat** and individual subjects take up more of the picture, large figure paintings. Often the whole painting is more colorful (high saturation).
2. Song Dynasty: If its is a painting contains figures: often have multiple figures in the picture. Lots of landscape paintings.There are paintings of flowers and use more white in the picture compared to other dynasty.
3. Yuan Dynasty: Overall painting style is coarse, brush strokes are more rough (the paper seems to be more damaged).Horses appear in large numbers. Landscape paintings have very distinctive brushstrokes.
4. Ming Dynasty: Staining with ink became more common (brush strokes).
5. Qing Dynasty: The strokes are clear and realistic (brushwork).

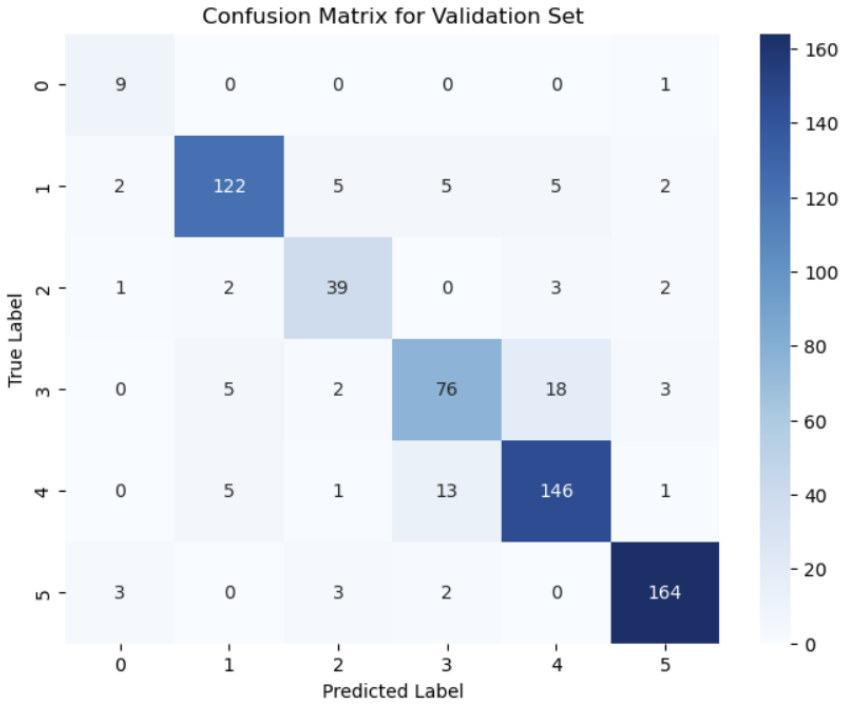
2 Method

From the file I included,
vit_model file is for training the original vit_model.
The confusion matrix file is for checking the confusion matrix and adjust the threshold for label 5 and 3 and 4. I also get the final output from that file.
Binary classifier file is for training a more robust binary classifier for AI and non AI dataset (which also provide a csv file for my future training). I also tried to train a Ming Qing classifier there.

Lastly, although I tried to train Ming and Qing binary dataset on both confusion matrix file and binary classifier file, I failed.
Below is a detailed introduction.

1. I train a binary classifier for AI-generated images and perform a cleaning. (based on ResNet 50 framework)
2. I train the VIT model to make predictions based on the cleaned data set. (Label Smoothing entropy, AdamW optimizer, CosineAnnealingLR scheduler, mix up function)
3. I perform another cleaning based on my previous VIT model to correct the mislabeled AI images. For example, if model predicts AI with probability > 0.9 but label is not AI, relabel as AI.
4. Re-Train the VIT model based on the refined dataset. (fine tuning learning rate)
5. Find that the Ming and Qing images are hard to classify within the VIT model from the confusion matrix (the cleaned dataset are got from the binary classifier file and predicted by binary_model.pth), decide to train another binary model for the Ming and Qing images with Focal loss, but failed.

   Also, find that when making certain classification, it is better to adjust the threshold for better performance, try on different threshold. (Adjust the threshold of AI and nonAI dataset to be 0.53, adjust the threshold for Song dynasty and Ming and Qing dynasty)



3 Experiment

Why my model is effective:

1. Training VIT model with strong augmentation and regulation. (Mix up augmentation to help model be more robust within the noise, use label smoothing for regulation)
2. Use weightedRandomSampler to handle the class imbalance.
3. Only use randomcrop in transform to avoid potential risk of messing up the features in the images.
4. Use effective techniques which have been proved by my previous trials like AdamW optimizer, CosineAnnealingLR scheduler, and mixup and scaler for this new trial within VIT model.
5. Run about 40 epochs in total for training the VIT model
6. Further implementation within the report from the spread sheet (Non-AI F1, etc.)
7. Look at the confusion matrix and make more trials on fine-tuning the threshold
8. The design of my training process is suitable for such dataset.

Some of my previous trials:

Use SVM, logistic and XGBoost for training the dynasty classifier: Even within feature extraction, the model architecture is too simple for handling such a task.

Train 6 different binary classifier for each dynasty's model because the AI binary classifier perform great (does not perform well), it fails on certain dynasties images even within augmentation, it will easily go overfitting or underfitting.

Try different model architecture like DenseNet, ResNet 18, ResNet 50, etc.

4 Conclusion

Summarize your findings on how to tackle a noisy and unbalanced dataset.

1. Firstly, we should look and examine the dataset and identify the type of the task (for instance, image classification or sound classification might need different type of model framework)
2. Find a suitable model framework to handle certain project, firstly design the process of such task.
3. Perform the dataset cleaning for a noisy set for better performance later. Explore certain method to deal with the unbalance issue (I use the mix up function to handle the class imbalance)
4. Handle the transform method carefully in case of messing up important features. For instance, after flipping, the textural features might be distorted.
5. Evaluate the performance and adjust the method of training. Start from the base model and learn different techniques for tuning hyperparameter and understand its benefits and flaws.
6. **Start early!**