Real vs Fake Images Classification Report

Arvind Raghavendran
21f1005301

April 17, 2024

Contents

1 Introduction **3**

2 Preprocessing **3**
2.1 Data Augmentation 3
2.2 Normalization 3

3 Exploratory Data Analysis (EDA) **3**
3.1 Class Distribution 3
3.2 Visual Inspection 3

4 Model Exploration **4**
4.1 Simple CNNs 4
4.2 Transfer Learning 4

5 Model Training and Evaluation **5**
5.1 Metrics 5
5.2 Confusion Matrix 5

6 Observations **5**

7 Conclusions **5**

8 Future Work **6**

9 Acknowledgement **6**

Abstract

This report documents the process and findings of developing a machine learning model to classify real and fake images. The objective is to combat misinformation and maintain the integrity of visual content in the digital age. The report details the preprocessing steps, exploratory data analysis (EDA), model selection, training, and evaluation, along with conclusions and potential areas for further improvement.

1 Introduction

The proliferation of manipulated images in various media platforms has raised concerns about the authenticity of visual content. To address this issue, this project aims to develop a machine learning model capable of distinguishing between real and fake images. The model's performance is evaluated using metrics such as accuracy, precision, and recall. This report provides a detailed overview of the methodology, experimentation, and results obtained during the project.

2 Preprocessing

The preprocessing stage involves preparing the image data for training and testing the model. It includes steps such as data augmentation and normalization.

2.1 Data Augmentation

Data augmentation is essential to increase the diversity of the training data and improve the model's generalization ability. Techniques such as rotation, rescaling, and flipping are applied to augment the dataset.

2.2 Normalization

Normalization is performed to scale the pixel values of the images to a standard range, typically [0, 1]. This ensures that the model learns effectively without being biased by the input data's varying scales.

3 Exploratory Data Analysis (EDA)

EDA is conducted to gain insights into the distribution and characteristics of the dataset. It helps identify potential challenges and biases that may affect model performance.

3.1 Class Distribution

The class distribution of the dataset is analyzed to understand the imbalance between real and fake images. Strategies such as class weighting may be employed to address this imbalance during model training.

3.2 Visual Inspection

Sample images from the dataset are visually inspected to understand the diversity and quality of the data. This helps identify any anomalies or artifacts that may need to be addressed before model training.

4 Model Exploration

Various machine learning models were explored to identify the most suitable architecture for the real vs fake image classification task. The models considered included simple convolutional neural networks (CNNs) as well as transfer learning-based approaches.

4.1 Simple CNNs

Initially, several simple CNN architectures were experimented with, consisting of multiple convolutional layers followed by batch normalization and dropout layers. The configurations tested included:

- Model 1: 3 convolutional layers with 64, 128, and 256 filters respectively, each followed by batch normalization, max-pooling and dropout (dropout rate of 0.5).
- Model 2: Similar to Model 1, but with an additional convolutional layer with 512 filters.

However, these models struggled to effectively distinguish between real and fake images, achieving poor performance. Despite efforts to optimize hyperparameters such as learning rate and dropout rates, the models failed to generalize well, likely due to the limited diversity of the dataset with only 500 samples provided for training and testing.

4.2 Transfer Learning

Given the limitations of the dataset, transfer learning emerged as a promising approach. Pre-trained models such as VGG16 were fine-tuned for the classification task by leveraging their learned features. Various configurations were explored to optimize model performance, including:

- Model 3: VGG16 with the final dense layers replaced by two fully connected layers with 2048 neurons each, followed by ReLU activation and a sigmoid output layer.
- Model 4: Similar to Model 3, but with different numbers of neurons in the fully connected layers (e.g., 4096 neurons each).
- Model 5: VGG16 with different activation functions in the fully connected layers, such as Leaky ReLU.

These transfer learning-based models showed significant improvement compared to the simple CNNs. Model 3, with two fully connected layers with 2048 neurons each, achieved the best performance, with an F1 score of approximately 56%. However, further experimentation and optimization could potentially improve the model's performance given more time and resources.

5 Model Training and Evaluation

The selected model architecture is trained using the preprocessed data, and its performance is evaluated using standard evaluation metrics.

5.1 Metrics

The model's performance is assessed using metrics such as accuracy, precision, recall, and F1 score. These metrics provide insights into the model's ability to correctly classify real and fake images and its balance between precision and recall.

5.2 Confusion Matrix

A confusion matrix is generated to visualize the model's predictions and identify any patterns or areas of improvement. It helps understand the distribution of true positives, true negatives, false positives, and false negatives.

6 Observations

1. Transfer learning is definitely a much-needed boost over models constructed from scratch, due to their already extensive training on huge, diverse datasets.

2. The metric graphs observed while training look promising, yet when we observe the confusion matrix, we realize that the model doesn't perform very well for the minor class, despite using class weights.

3. This discrepancy is mainly due to how misleading these metrics could be, especially when the classes are imbalanced and, for example, $N \ll P$. In such cases, we may still have a high recall just because we do not have enough false negatives proportional to true positives. The same goes for precision and accuracy as well.

4. The graphs also show erratic trends in learning, likely due to the lesser number of samples and hence truly random inputs of data. Fortunately, we make sure to save the best model using `ModelCheckpoint`.

I would like to assume that model performance would be better if we had more data, since even data augmentation is not helping us here.

7 Conclusions

In conclusion, the project demonstrates the effectiveness of transfer learning in addressing the challenges of real vs fake image classification. The chosen model

architecture, fine-tuned VGG16, achieves significantly better performance compared to simple CNNs. However, further improvements are possible with additional experimentation and optimization. The project highlights the importance of robust preprocessing, thorough EDA, and systematic model exploration in developing effective machine learning solutions for image classification tasks.

8 Future Work

Potential areas for future work include:

- Experimenting with different pre-trained models and architectures.
- Incorporating more advanced data augmentation techniques.
- Investigating ensemble methods to combine multiple models for improved performance.
- Collecting and annotating larger datasets to enhance model generalization.

9 Acknowledgement

I'd like to thank IIT Madras and the CV team for this opportunity. This project solidified my understanding of computer vision and how important transfer learning is. It has also improved my confidence in working with Deep Neural Networks. This project is a fantastic addition to my resume and I hope to work on more projects.