Model Card and Datasheet

- ***Purpose:** Binary classification of images as **real** or **fake** for digital forensics.
- **Architecture Type:** Lightweight Convolutional Neural Network (CNN)
    - Other models are also in use in the code (CNN, SVM and Random Forest)

Model Architecture:

1. **Input Layer**
    1. Accepts 256x256x3 images.
2. **Convolutional Blocks**
   • Four blocks with the following structure:
   • Convolutional layers (kernel sizes: 3x3, 5x5) for localized feature extraction.
   • Batch Normalization for stability and faster convergence.
   • ReLU activation for non-linearity.
   • MaxPooling2D for dimensionality reduction (pool sizes: 2x2, 4x4).
3. **Feature Aggregation**
   • **GlobalAveragePooling2D:** Reduces the spatial dimensions while retaining meaningful feature maps.
   • Two dense layers:
   • First dense layer with LeakyReLU activation and 32 neurons.
   • Dropout (50%) added to reduce overfitting.
   • Final dense layer with sigmoid activation for binary output.
4. **Output Layer**

    - Outputs a single probability value for binary classification (real/fake).

   Training Configuration:

• **Optimizer:** Adam (learning rate: 0.0001)
• **Loss Function:** Binary Cross-Entropy
• **Metrics:** Accuracy
• **Epochs:** 5
• **Batch Size:** 32
• **Class Weights:** Adjusted to handle class imbalance (real:fake = 1.0:1.5).
• **Callbacks:**

```
• ModelCheckpoint: Saves the best model based on validation accuracy.
• ReduceLROnPlateau: Reduces learning rate when validation loss plateaus.
• EarlyStopping: Stops training when validation loss does not improve.
```

Data Augmentation:

• Applied using the following transformations:
• Rescaling: Pixel values normalized to [0, 1].
• Random rotations (up to 20°).
• Shifts: Horizontal and vertical (up to 20%).
• Brightness adjustments (range: 0.8–1.2).
• Horizontal flips.
• Zooming (up to 20%).

Can also be used as part of the search space to investigate the effect of augmentation on classification

Evaluation:

Test Metrics

Accuracy: 0.7800
Precision: 0.9419
Recall: 0.5980
F1: 0.7315

Validation Metrics

Accuracy: 0.9215
Precision: 0.9825
Recall: 0.8578
F1: 0.9159

Confusion Matrix:
Confusion Matrix for Mesonet

		Predicted	
		Fake	Real
Actual	Fake	1817	28
	Real	261	1574

Use Cases

Primary Application: Digital image forensics (e.g., detecting deepfakes).
Secondary Applications: Authentication of image content for journalism, law enforcement, etc.

Primary Users: This should be used by researchers or digital forensic practitioners for detecting potentially altered images

Intended Use Cases: Training models to detect forged images in an unknown dataset

Ethical Considerations

• **Bias and Fairness**: see datasheet

• **Potential Risks**: unknown consent for included persons in dataset

6. **Limitations**

• **Known Limitations**: Computationally expensive and takes a long time to train/optimise

- **Future Improvements**:
- Combinations of the models used to speed up processing by pre-categorising images prior to classification by more computationally expensive methods to speed up workflow.
- better exploration of hyperparameter correlation to remove dimensions from optimisation search space.

7. **Citations and References**

• https://www.kaggle.com/datasets/shivamardeshna/real-and-fake-images-dataset-for-image-forensics/data

==========

Datasheet for Dataset: [Real & Fake Images Dataset-For image forensics]

Dataset Overview
1. **Dataset Name** - **Dataset Name**: Real & Fake Images Dataset for Image Forensics
2. **Dataset Description** - This dataset comprises images categorized as real or fake, intended to support research in image forensics, particularly in detecting and analyzing manipulated images.

Provenance

- ***Source**: The dataset is hosted on Kaggle and was uploaded by the user Shivam Ardeshna. (link: https://www.kaggle.com/datasets/shivamardeshna/real-and-fake-images-dataset-for-image-forensics/data)

Collection Process: unknown

***Licensing Information**: Apache 2.0

3. **Data Characteristics**

Dataset Structure

- ***Directories (four):** Organized into train, test, and validation.
- Each directory contains two subcategories: real and fake.

- Folder layout will be read and labelled directly by TensorFlow
- **Contents**
- ***Train:**
  - Real: Authentic, unaltered images for model training.
  - Fake: Manipulated/generated images representing digital forgeries.
- **Test:**
  - Real: Separate genuine images for model evaluation.
  - Fake: Altered/synthetic images for testing forgery detection.
- **Validation:**
  - Real: Genuine images for fine-tuning and preventing overfitting.
  - Fake: Forged images to validate model generalization.

Utility

- Ensures rigorous evaluation, reliable performance, and applicability to real-world image forensics challenges.
- **Size**: (256, 256, 3)
- **Target Variable**: Binary classifier - Real / Fake

4. **Data Splits**

- ***Train/Test/Validation Split**: Ratios or number of samples in each split.
  - Data Set 1 (as used in project)
    Test: Fake (2,623) / Real (2,604)
    Train: Fake (20,001) / Real (20,001)
    Validation: Fake (6,161) / Real (6,199)

• **Splitting Criteria**: unknown

5. **Data Quality**

• **Missing Data**: No missing data, but class weighting used

• **Noise and Errors**: Known issues in the dataset.

• **Preprocessing Steps**: Cleaning is not required as sizing is standardised. Augmentation optional (contrast, rotations, etc.)

6. **Use Cases**

• **Intended Uses**: Testing of image processing classification methods for determining if an image is real or a fake.

• **Out-of-Scope Uses**: Unknown

7. **Ethical Considerations**

• **Bias**: No data is given concerning the types of images or their content. So caution should be taken. Much of the imagery is of people and no information is given on collection regarding ethnicity, for example.

• **Privacy Concerns**: There is no explicit consent for inclusion shown in the dataset so publishing face should be treated with caution

8. **Citations and References**

• https://www.kaggle.com/datasets/shivamardeshna/real-and-fake-images-dataset-for-image-forensics/data