# EfficientNet-B6 Model

**EfficientNet-B6** is a high-performance convolutional neural network architecture that is part of the **EfficientNet** family, which was introduced by **Google AI** in 2019. The EfficientNet models use a novel **compound scaling** method to systematically scale up the network's depth, width, and resolution in a balanced way, leading to improved accuracy and efficiency compared to traditional CNN architectures.

**EfficientNet-B6** specifically offers **state-of-the-art accuracy** while maintaining a relatively **low computational cost** compared to other large models like **ResNet** or **Inception**. It is designed for high-resolution image inputs (528x528) and has **approximately 43 million parameters**. The model is well-suited for tasks requiring fine-grained image classification and is widely used in industrial and academic applications.

Appendices

**C.7 Efficient Net Model Architecture**

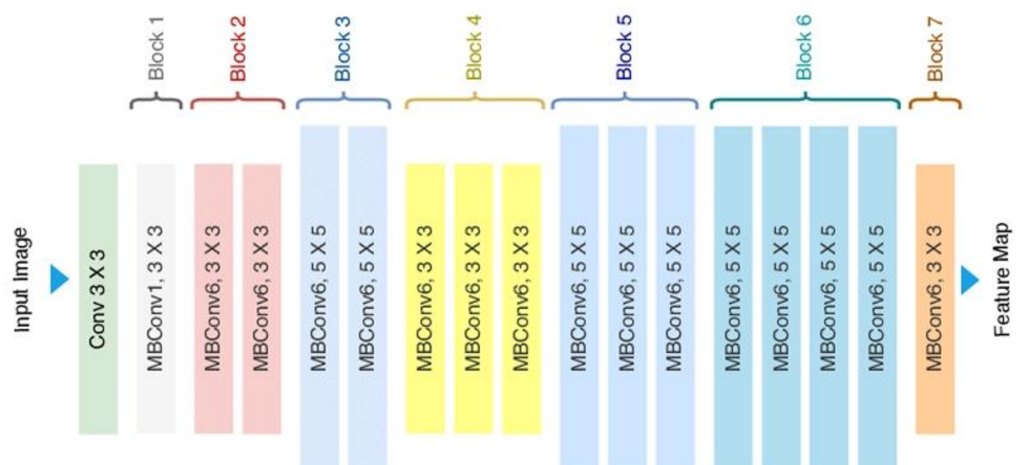



Figure 0C.7 Efficient Net Model Architecture

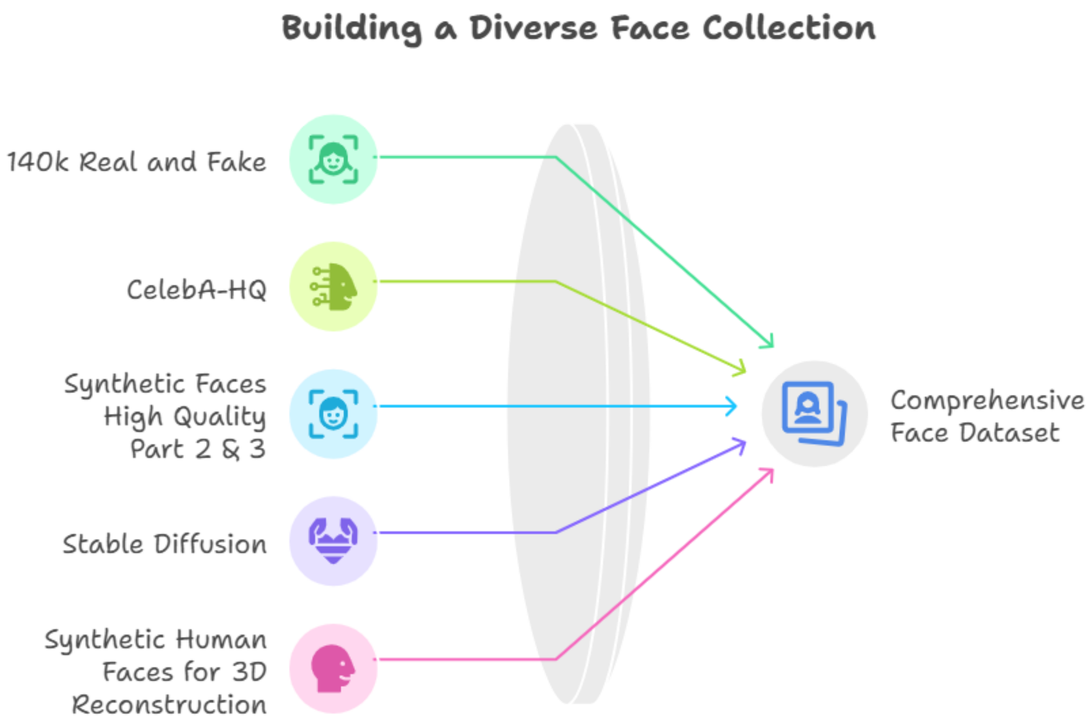
# Task Objective:

## AI Image Generation Detection:

Our system leverages advanced deep learning techniques to **accurately detect AI-generated images** (such as those produced by GANs, diffusion models, or other generative techniques) and distinguish them from authentic, human-created images. This involves:

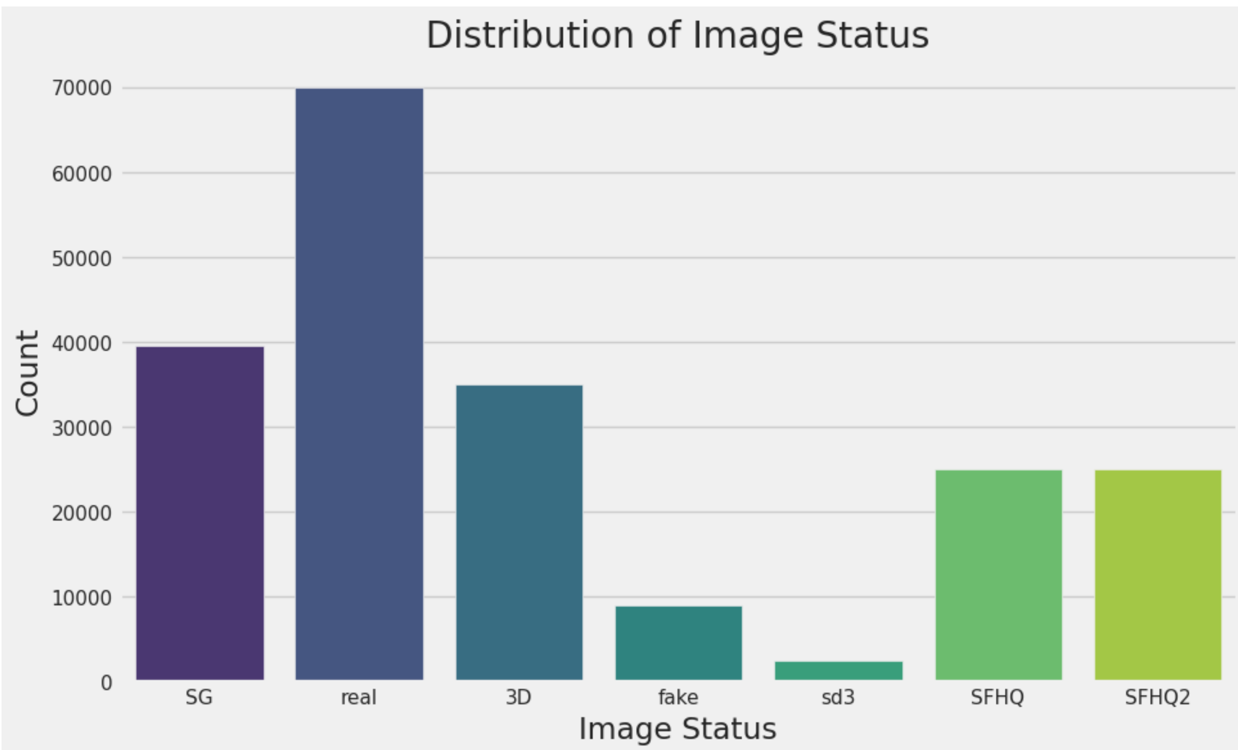
- **Feature Extraction**: Utilizing high-capacity convolutional neural networks (e.g., **EfficientNet**, **ResNet**) to extract subtle statistical patterns and artifacts commonly found in AI-generated images.
- **Multi-domain Training**: Training on diverse datasets that include synthetic images from popular generative models like StyleGAN, StyleGAN2, EF3D and Stable Diffusion to ensure generalization across different sources.
- **Fine-grained Classification**: Implementing classifiers that can detect minute inconsistencies in textures, lighting, edges, and noise patterns that are often imperceptible to the human eye.
- **Robust Performance**: Evaluated across multiple benchmarks to ensure high precision and recall, even when images are post-processed, resized, or compressed.

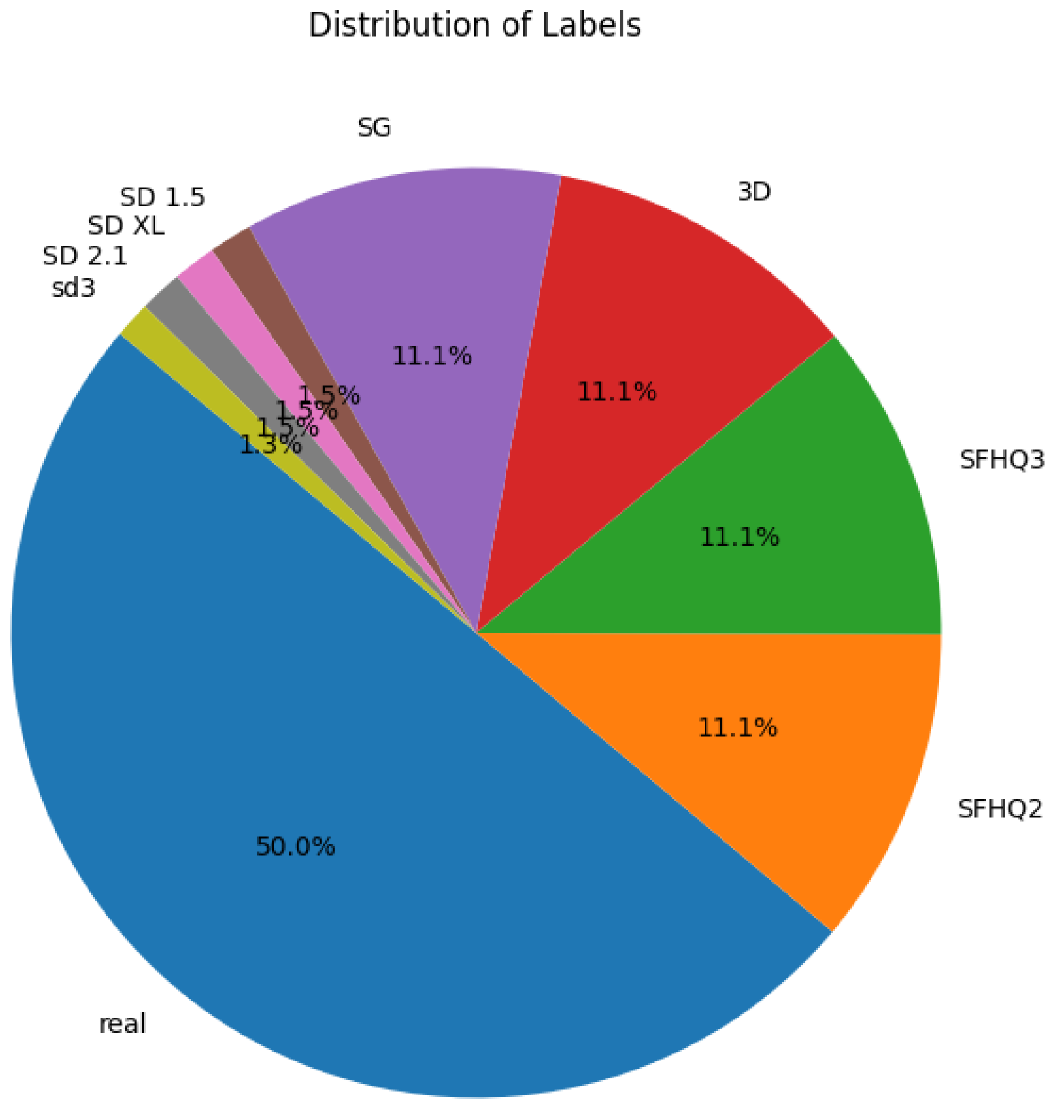
# Used Datasets:



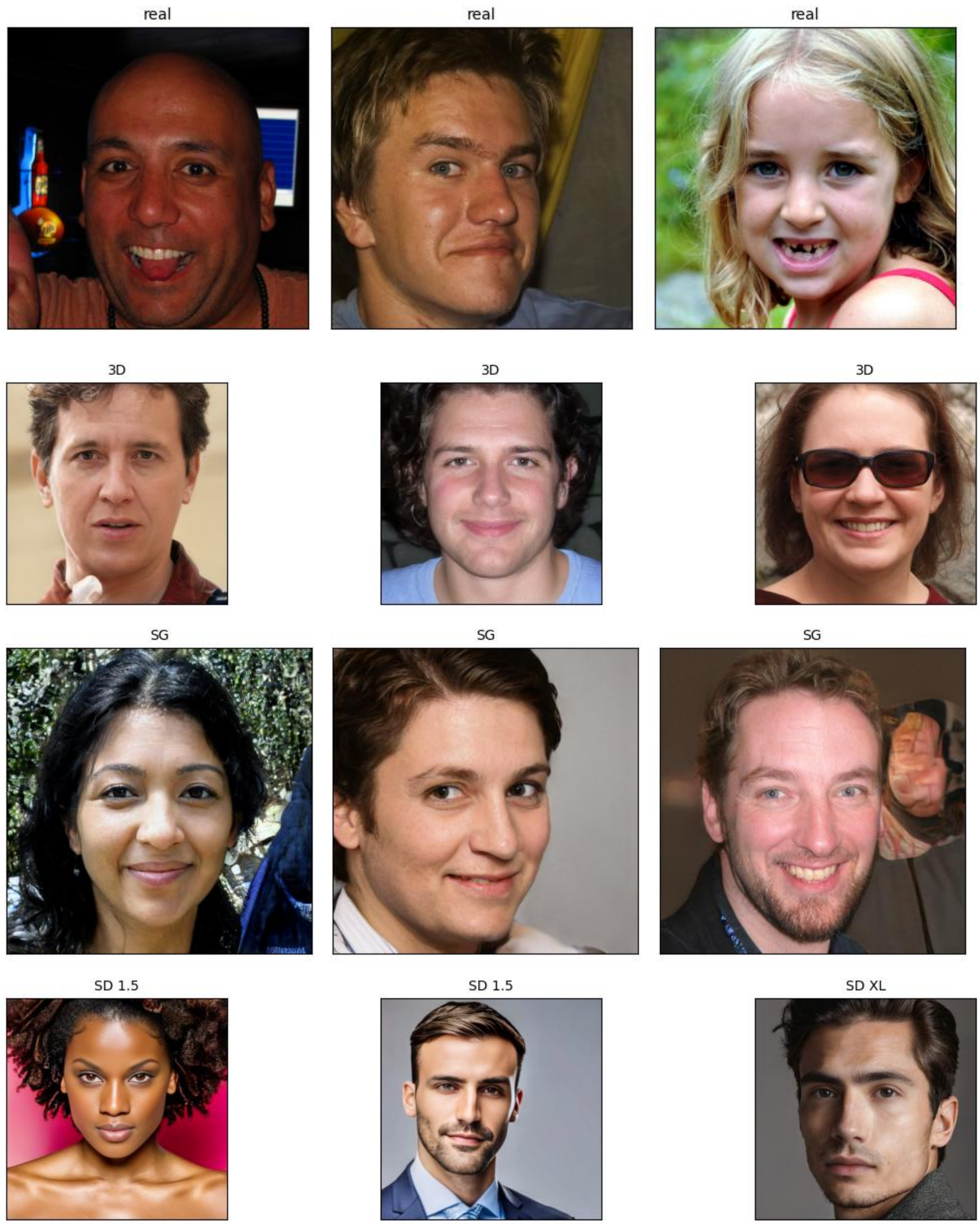

- **140k Real and Fake Faces**: Combines 70,000 real faces from the Flickr dataset with 70,000 fake faces generated by StyleGAN, resized to 256px.
- **CelebA-HQ resized (256x256)**: Contains 30,000 high-quality celebrity faces, suitable for training and evaluating generative models**.**
- **Synthetic Faces High Quality (SFHQ) Part 2**: Includes 91,361 high-quality 1024x1024 curated face images, enhanced using StyleGAN2 techniques.
- **Face Dataset Using Stable Diffusion v1.4**: Comprises real faces from the Flickr dataset and fake faces generated by Stable Diffusion models, resized to 256px.
- **Stable Diffusion Face Dataset**: AI-generated human faces using Stable Diffusion 1.5, 2.1, and SDXL 1.0 checkpoints, covering resolutions of 512x512, 768x768, and 1024x1024.
- **Synthetic Faces High Quality (SFHQ) Part 3**: Contains 118,358 high-quality 1024x1024 face images generated by StyleGAN2, utilizing advanced truncation tricks.
- **Synthetic Human Faces for 3D Reconstruction:** Generated by drawing samples from the EG3D model, resulting in high-quality 512x512 synthetic face images.

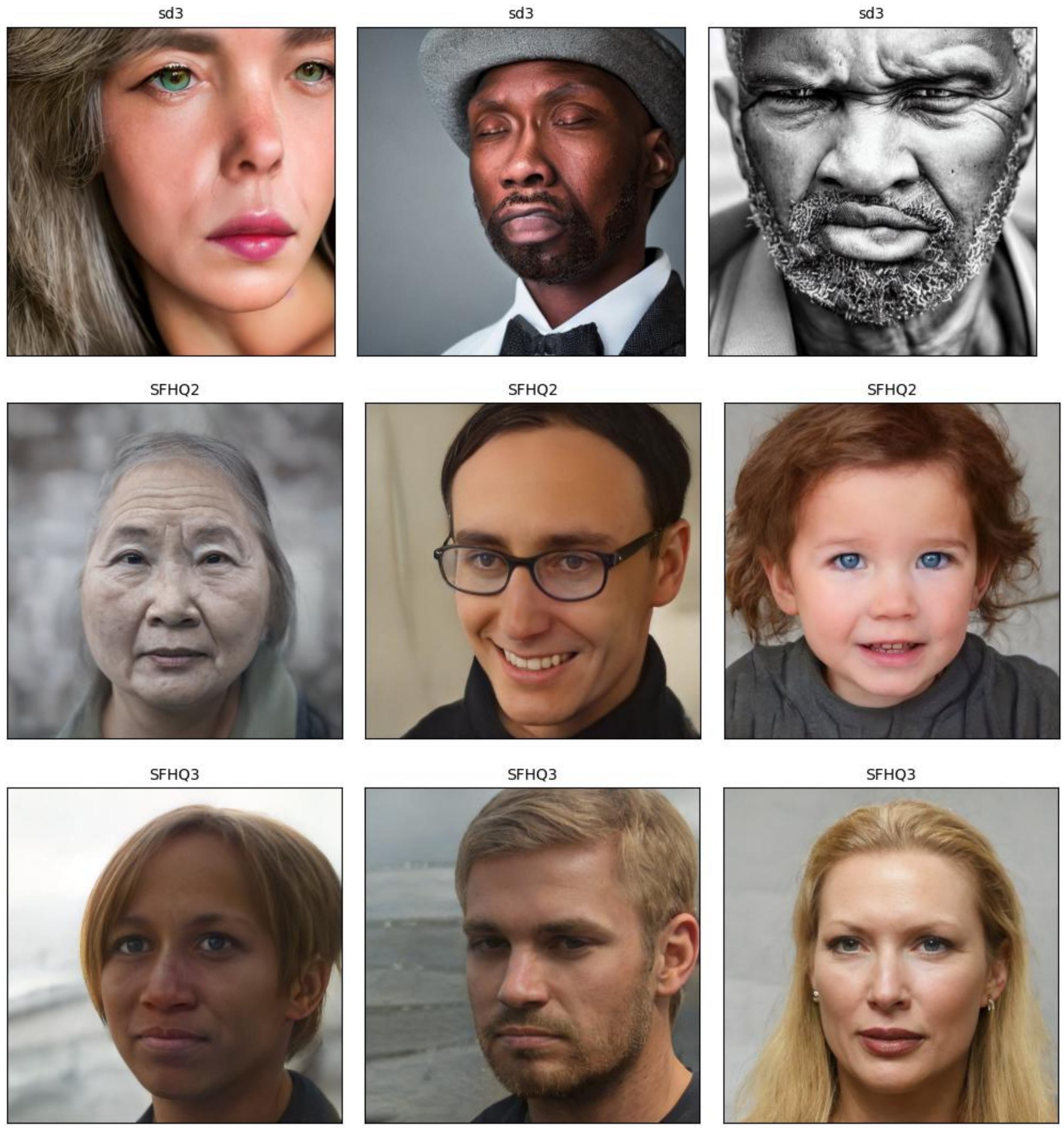
| 200K images | | | | | |
|---|---|---|---|---|---|
| 100 k Real<br>100 k Fake | | | | | |
| Train | | Validate | | Test | |
| Real: 70 k | Fake: 70 k | Real: 15 k | Fake: 15 k | Real: 15 k | Fake: 15 k |
| 70k Flickr | 3k SD V2.1 | 15 k CelebA HQ | 3k SDXL 1.0 | 15 k CelebA HQ | 2536 SD 1.4 |
| | 3k SD V1.5 | | 3k SG1 | | 3116 SG1 |
| | 16k SG1 | | 3k EG3D | | 3116 EG3D |
| | 16k EG3d | | 16k S SG2 | | 3116 S SG2 |
| | 16k S SG2 | | 16k S SD1.4 | | 3116 S SD1.4 |
| | 16k S SD1.4 | | | | |
| 70% Data | | 15% Data | | 15% Data | |




Distribution of Labels
SG
SD 1.5
SD XL
SD 2.1
sd3
3D
SFHQ3
SFHQ2
real
11.1%
11.1%
11.1%
11.1%
1.5%
1.5%
1.5%
1.3%
50.0%

## Dataset Overview:




sd3
sd3
sd3
SFHQ2
SFHQ2
SFHQ2
SFHQ3
SFHQ3
SFHQ3

# Model Selection:

## Overview of Initial Model Candidates:

We initially evaluated several models for image detection, and we chose one based on the reported performance for our dataset characteristics.

- **EfficientNet**: Chosen for its balance between accuracy and computational efficiency.
- **ResNet**: Evaluated for its strong performance in image classification.
- **Xception**: Considered for its capability to detect fine-grained image features.

## Evaluation Criteria:

Models were evaluated based on the following criteria

- **Accuracy**: Overall classification performance.
- **Precision and Recall**: Balance between false positives and false negatives.
- **F1 Score**: Combined measure of precision and recall.
- **ROC-AUC**: Trade-off between true positive and false positive rates.

## Experimental Setup:

Each model was trained using a standardized dataset split of 70% training, 15% validation, and 15% testing, and models were trained on high-performance GPUs.

## Results of Initial Models:

- **EfficientNet**: Balanced accuracy and computational efficiency, with an accuracy of 99%.
- **ResNet**: Achieved an accuracy of 99% but required more computational resources.
- **Xception**: Offered detailed feature extraction but was less efficient compared to EfficientNet.

| Metrics | Accuracy | Precision | Recall | F1-Score |
| --- | --- | --- | --- | --- |
| **EfficientNetB2** | 94.3% | 94.3% | 94.2% | 94.2% |
| **EfficientNetB6** | 99.5% | 99.5% | 99.5% | 99.5% |
| **Xception** | 97% | 98% | 97% | 98% |
| **ResNet-50** | 98.2% | 98.2% | 98.2% | 98.2% |
| **VIT-224** | 96.7% | 96.6% | 96.7% | 96.6% |

## Final Model Selection:

Based on the initial evaluations, the following refinements were made:

**EfficientNet** was selected for its optimal balance between accuracy and computational efficiency, with additional data augmentation applied to improve generalization, Preferred for its efficiency and high accuracy in distinguishing Real from AI-generated images.

## Training Setup:

**EfficientNet-B6** Model was trained with:

- Batch Size:16
- Learning rate: 5e-5
- Epochs: 2.5

## Performance Metrics:

### Classification Report:

```
Classification report:

              precision    recall  f1-score   support

        Real     0.9978    0.9915    0.9946     15000
        Fake     0.9916    0.9978    0.9947     15000

    accuracy                         0.9947     30000
   macro avg     0.9947    0.9947    0.9947     30000
weighted avg     0.9947    0.9947    0.9947     30000
```

### Confusion Matrix: