REAL VS FAKE

By Prasun Bhattarai

MOTIVATION & OBJECTIVE

Why this problem?

- Deepfakes and synthetic faces are increasingly realistic — detection is important for media integrity, forensics, and platform moderation.

Objective

- Train a CNN-based classifier using a balanced, curated subset (5,000 train / 1,000 test).
- Measure classification accuracy and analyze training dynamics.

DATASET

Sources

- Real faces: FFHQ subset (originally 70k @ 256×256). Source referenced from Kaggle discussion.
- Synthetic faces: Diffusion-model generated (originally 64k @ 1024×1024). Source referenced from Kaggle dataset.

This project — data used

- Selected a subset and re-organized manually:
  - Training: 5,000 images
  - Test: 1,000 images
  - Validation: 10% of training used as validation (i.e., 500 images)

DATASET PREVIEW

PREPROCESSING & DATA PIPELINE

Preprocessing steps (as implemented)

- Resizing / center-cropping images to model input size (consistent across sources).
- Normalization (per-channel mean/std).
- Data augmentation on train : random horizontal flip, random crop, small color jitter — to reduce overfitting from domain differences.
- Labeling: binary labels — real vs synthetic.

Train / Val / Test split

- Train: 5,000 images (of which 10% used for validation)
- Validation: 500 images
- Test: 1,000 images (held out, only used for final eval)

MODEL ARCHITECTURE

High-level architecture

- Convolutional Neural Network (CNN) backbone
  - Several conv layers + ReLU activations
  - Batch Normalization
  - Max-pooling
- Global average pooling / flatten
- Fully connected (dense) layers
- Regularization: dropout.

TRAINING SETUP & HYPERPARAMETERS

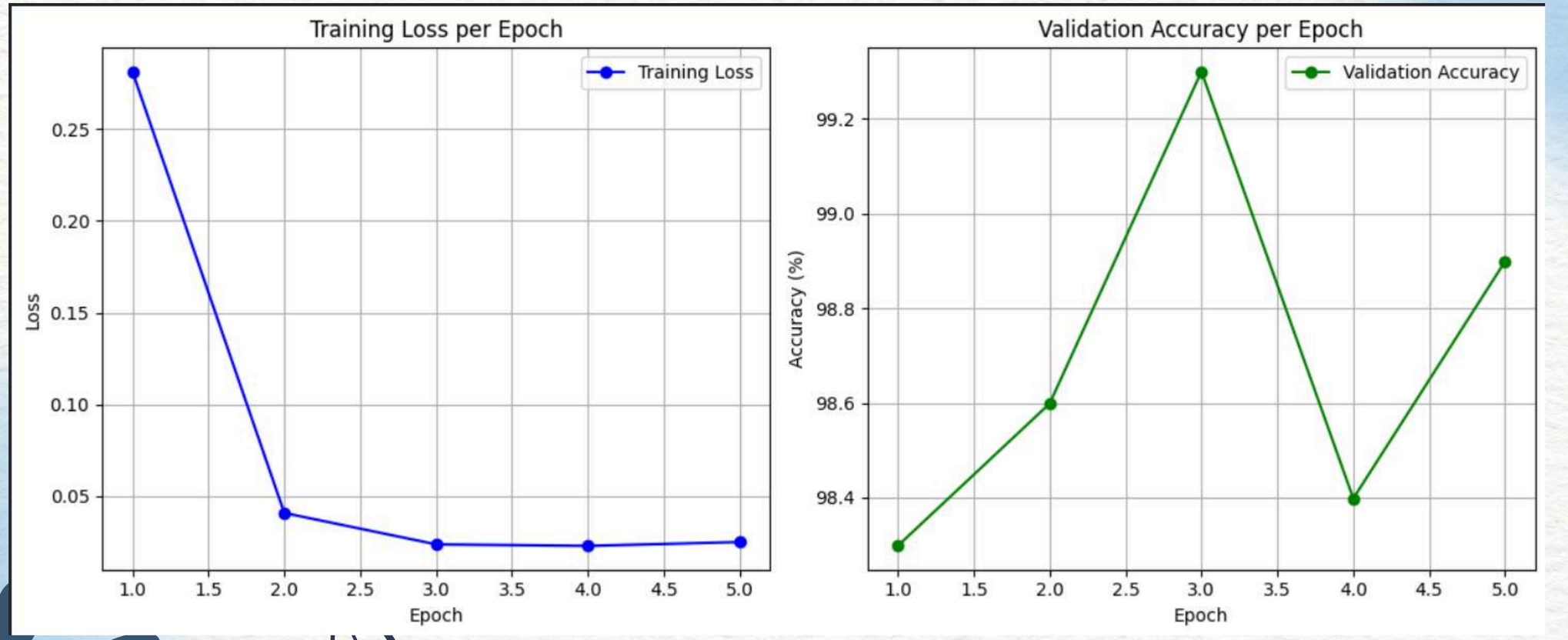
Training details

- Loss function: Cross-Entropy (binary classification)
- Optimizer: (Adam)
- Learning rate: (0.001)
- Epochs: 5
- Batch size: (32)
- Validation: 10% of training data used each epoch

TRAINING LOGS (EPOCH-WISE)

```
0.2811378862771556
Epoch 0, Loss: 79.28088393015787, Test Accuracy: 98.30%
0.04099371107683206 4
Epoch 1, Loss: 11.560226523666643, Test Accuracy: 98.60%
0.023928462637615396
Epoch 2, Loss: 6.747826463807542, Test Accuracy: 99.30%
0.023085641401055 6
Epoch 3, Loss: 6.510150875097679, Test Accuracy: 98.40%
0.025184681252725476
Epoch 4, Loss: 7.102080113268585, Test Accuracy: 98.90%
```

TRAINING & VALIDATION CURVES



CONCLUSION

- Built and trained a face authenticity classifier on a curated subset (5,000 train, 1,000 test).
- Achieved final test accuracy = 99.15% with stable training after a few epochs.
- Results are promising but require further validation to claim general-purpose synthetic face detection capability.

HASTA LA VISTA