Report

Part 1: Spoof Detection

Introduction

Spoof detection is an essential task in biometric security to differentiate between real and fake facial images. In this task, a Vision Transformer (ViT) model was used for classifying real and spoofed images. The dataset used was CelebA-Spoof.

Methodology

1. **Dataset Preparation**:

   o The dataset was loaded from Hugging Face.

   o It was reduced to 20% of its original size to optimize computation.

   o The dataset was split (80-20%) into training and testing sets.

2. **Model Selection & Training**:

   o Used ViT-Base Patch16-224 from Hugging Face.

   o Applied AdamW optimizer with cross-entropy loss.

   o Training was done for 10 epochs with a batch size of 16.

3. **Results**:
   After training, the model achieved the following metrics:

   o **Accuracy: 99.63%**

   o **Precision: 99.63%**

   o **Recall: 99.84%**

   o **F1-Score: 99.74%**

4. **Prediction Example:**

   o Real Image Prediction: Real (Confidence: 0.92)

   o Spoof Image Prediction: Spoof (Confidence: 1.00)

Real Image - Prediction: Real   Spoof Image - Prediction: Spoof

Key Takeaways:

- The ViT model was highly effective in detecting spoofed images.

- The high recall score suggests minimal false negatives.

- Corrupted or missing images were handled using an exception-based approach.

Part 2: AI-Powered Visual Search (CLIP Model)

Introduction

CLIP (Contrastive Language-Image Pretraining) is an advanced AI model that connects text and images to perform semantic search. This task aimed to implement AI-powered visual search using OpenAI's CLIP model.

Methodology

1. **Model Setup**:
   - Installed ftfy, regex, tqdm, and OpenAI's CLIP.
   - Loaded pre-trained CLIP ViT-B/32 model.
2. **Implementation Steps**:
   - Extracted image embeddings and text embeddings.
   - Compared embeddings using cosine similarity to find the best matches.
3. **Results**:
   - Tested with sample images and text queries.

- o CLIP successfully matched text descriptions with relevant images**.**
- o The model was able to distinguish similar objects based on contextual differences.

Top 5 images for query: 'A person riding a horse'

Sim: 0.3108

Sim: 0.3155   Sim: 0.3098   Sim: 0.3074   Sim: 0.3033

Part 3: Stable Diffusion - Parameters & Their Effects

Introduction

Stable Diffusion is a text-to-image generative AI model that can generate high-quality images from prompts. This task focused on analyzing how different parameters affect image generation.

Parameters Tested & Effects

Parameter	Definition	Values Tested	Effects
Strength	Controls how much the original image influences the output.	0.5, 0.7, 0.9	Higher values lead to stronger adherence to text prompt.
Guidance Scale	Controls strictness of following text prompt.	7, 10, 12	Higher values enforce text prompt more strictly.
Steps	Number of diffusion steps during image generation.	50, 75, 100	Higher steps produce finer details but take longer to process.

Results & Observations

1. **Prompts Used**:

   - o "Watercolor painting of a lion."

   - o "A pixel art lion."

   - o "A lion in the style of Salvador Dalí."

   - o "A lion in the style of Van Gogh."

   - o "Futuristic cyberpunk lion."

2. **Key Findings:**

   o Van Gogh style had strong texture variations when guidance scale was high.

   o Pixel art performed better with lower strength values to retain structure.

   o Cyberpunk images improved with a medium-strength setting for balanced results.

watercolor painting of a lion
Strength: 0.5, Scale: 7

watercolor painting of a lion
Strength: 0.7, Scale: 10

watercolor painting of a lion
Strength: 0.9, Scale: 12

a pixel art lion
Strength: 0.5, Scale: 7

a pixel art lion
Strength: 0.7, Scale: 10

a pixel art lion
Strength: 0.9, Scale: 12

a lion in the style of Salvador Dalí
Strength: 0.5, Scale: 7

a lion in the style of Salvador Dalí
Strength: 0.7, Scale: 10

a lion in the style of Salvador Dalí
Strength: 0.9, Scale: 12

a lion in the style of Van Gogh
Strength: 0.5, Scale: 7

a lion in the style of Van Gogh
Strength: 0.7, Scale: 10

a lion in the style of Van Gogh
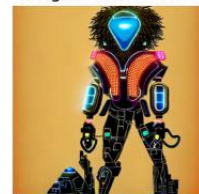Strength: 0.9, Scale: 12

futuristic cyberpunk lion
Strength: 0.5, Scale: 7

futuristic cyberpunk lion
Strength: 0.7, Scale: 10

futuristic cyberpunk lion
Strength: 0.9, Scale: 12

Conclusion

This assignment explored three key Generative AI tasks:

1. **Spoof Detection**: Achieved 99.63% accuracy using ViT for image classification.

2. **AI-Powered Visual Search**: Used CLIP to match images and text with high precision.

3. **Stable Diffusion Analysis**: Experimented with parameters like strength, steps, and guidance scale to optimize image generation.