INTER IIT
TECH MEET 13.0

MID
PREP

Adobe

IMAGE CLASSIFICATION AND ARTIFIACT DETECTION

About Adobe Research

Adobe Research combines the latest in cutting-edge technology with creative digital experiences, pioneering advancements in artificial intelligence, machine learning, computer vision, and more. Our mission is to drive the future of digital creativity and user experiences, enabling professionals and creators worldwide to unlock their full potential through innovative tools and solutions.

Background

The rapid evolution of artificial intelligence has significantly impacted the creation of digital content. Advanced generative models, such as Adobe Firefly, Stable Diffusion, and Midjourney, have made it possible to produce highly realistic synthetic content.

These diffusion-based models open new creative possibilities and provide businesses with efficient solutions for automating content creation. However, this surge in synthetic media also raises pressing concerns about content authenticity and reliability.

For Adobe, ensuring content authenticity is crucial to maintaining user trust and the integrity of digital experiences. As AI-generated content becomes widespread, the ability to detect inauthentic or manipulated images has become essential for content verification, brand reputation, and the protection of digital spaces from misuse. This challenge aims to address these concerns by advancing technologies that differentiate real from AI-generated images.

Problem Description

We invite participants to develop a model that not only detects AI-generated images but also explains the basis of its classifications. This challenge comprises two primary tasks:

Task 1: Detection of AI-Generated Images

Objective:
Build a model that accurately identifies AI-generated images, focusing on those generated by diffusion and other advanced techniques. With the growing prevalence of AI-generated imagery, challenges such as Deepfakes and fabricated visuals are increasingly common. Effective detection is essential for applications in news verification, e-commerce, and personal identification, ensuring integrity and trust in digital media.

Task 2: Artifact Identification & Explanation Generation

Objective:
Develop a model that not only identifies distinguishing artifacts within images but also provides interpretable explanations for its classifications of real vs. AI-generated images.
Many AI models are opaque "black boxes," making it difficult to trace the reasoning behind their outputs. To address this, participants should design a model that interprets and highlights identifiable features in real and synthetic images.

Examples of these artifacts might include: subtle irregularities in texture, inconsistencies in lighting, or unnatural patterns in fine details, all of which contribute to differentiating synthetic images from genuine ones. By generating plausible explanations for the detected artifacts, this model will help prevent hallucinations and build user confidence in the AI's assessments.

Recommended Reading

To better understand the problem, consider reviewing these key papers: [1], [2], [3]

INTER IIT TECH MEET 13.0

Dataset

The CIFAKE dataset provides 60000 labelled samples of real and synthetically generated images, which can be used to develop the image detection model
Dataset Link

Problem Statement:

Task 1: Detection of AI-Generated Images

- Goal: Classify images as either AI-generated or real
- Data sample is provided below. The dataset consists of labelled examples of real and fake images

REAL: containing real images
FAKE: containing synthetically generated images

Fake **Real**

- Size of training data: 60000 samples
- The results will be evaluated on a custom test set which will be shared 2 days prior to final submission.

Task 2: Artifact Identification and Explanation Generation

- For images classified in Task 1, develop a model to detect distinguishing artifacts and provide clear, interpretable explanations supporting the classification
- Example generation is provided below –

- Unnatural Blending: Soft, artificial transitions between horse and background
- Anatomical Inconsistencies: Distorted leg proportions and joint angles
- Color Treatment: Artificially uniform coat coloring without the complex variations seen in real horses
- Resolution and Detail: Missing fine details like individual hairs, replaced by smooth texture
- Motion Effects: Unrealistic motion blur pattern

A comprehensive set of potential artifacts, along with a structured framework for generating explanations, will be provided.
NOTE: The test set would be generated using different generative techniques and model variants incorporating adversarial perturbations to prevent detection.

Evaluation Guidelines

Submission Deadline:
Testing Dataset Availability: 2 days before final submission
Evaluation Metrics:
Task 1:

- Classification Accuracy
- Precision/Recall, and F1 Score

Task 2:

- Content similarity

Evaluation Criteria:

- Solution Accuracy (50%): Performance metrics as specified for each task.
- Approach (35%): Innovative techniques, efficiency, and generalizability of the solution.
- Presentation (15%): Final presentation, Quality of documentation, clarity of code

Submission Guidelines

- **Final Code files need to be submitted.**
- **Technical Report**: A concise technical report (maximum 10 pages LaTeX format, excluding appendices and references) outlining the approach, model architecture, performance, and any assumptions or inferences.