AI-Generated Image Detection System

Deep Learning-Based Image Classification

Woohyeok Kang/20211004

Course: Deep Learning (CSE403)

Submission Date: 2025.10.24

Background: The Rapid Growth of Generative AI

Generative AI models (e.g., Stable Diffusion, Midjourney, DALL·E) allow for easy creation of high-quality images. However, this techn
ology also presents significant societal challenges.

Problem Definition: Why is a solution needed?

Technical Issues

Difficulty distinguishing real vs. AI im ages.

Social Issues

Increased misinformation from fake i mages.

Ethical Issues

Growing legal and ethical concerns (c opyright, defamation).

Key Research Question

Baseline comparison between CNN and Vision Transformer architectures to analyze representational differences

> "Can deep learning accurately estimate the probability that a given image was generated by AI?"

A positive answer to this question will play a crucial role in ensuring information trustworthiness in the digital age.

Application Scenarios and Expected Effects

News Verification

Enhancing journalistic credibility and filtering misinformation

Platform Security

Automated content verification and filtering for social media and communities

Copyright Protection

Strengthening the protection of digital assets and creative works

Identity Verification

Reinforcing profile and personal verification systems

Dataset Overview

Dataset: Kaggle "Realifake - AI vs Real"

- **Source:** Kaggle Public Dataset
- **Size:** ~174,000 total images (92,000 real photographs, 82,400 AI-generated images)

- **Format:** JPG/PNG
- **Labels:** Binary Classification (Real / AI Generated)
- **Generative Models:** Stable Diffusion, Midjourney, etc. from multiple sources
- **Subject Diversity:** Includes landscapes, portraits, artistic renderings, and surreal sce
  nes

Dataset Structure Analysis

This dataset is designed with a clear binary classification structure and includes images of various resolutions and styles.

Class	Number of Samples	Format	Characteristics
Real Images	**~92,000**	JPG/PNG	Natural camera noise, wide topical diversity
AI Generated	**~82,400**	JPG/PNG	Model-specific artifacts, sur real composition tendencies

Reasons for Dataset Selection

1

Public Accessibility

Ensures reproducibility and academic v erification, fostering open research.

2

Clear Labels

Structure optimized for binary classifica tion, simplifying model development an d evaluation.

3

Unprecedented Scale

With over 174,000 images, it enables ro bust deep learning training and compre hensive model evaluation.

4

Style Diversity

A wide variety in artistic styles, subjects, and visual complexity ensures improved model generalization and real-world applicability.

Conclusion: This dataset is optimal for meeting both academic rigor and practical applicability.

Necessity of Deep Learning Application

1

Subtle Patterns

AI-generated images possess high-dimensional visual features that are difficult to define by human rules.

2

Automatic Feature Learning

Deep learning automatically discovers and learns these features through layer-by-layer abstraction.

3

Generalization Capability

Achieves robust classification performance for images from various generative models.

Potential Limitations

Model Overfitting

Risk of overfitting to specific genera tive models, reducing performance on new AI generators

Domain Gap

Performance degradation on compr essed images from social media pla tforms

Adversarial Robustness

Vulnerability to adversarial attacks d esigned to fool detection systems

Conclusion

- The rapid growth of generative AI has introduced new technical, social, and ethical challenges.
- While acknowledging potential limitations, this project aims to build a deep learning-based detection system to contribute to safer and more trustworthy digital environments.
- Thank you for your attention. We look forward to presenting the results of our research in the final presentation.