Project Title: Multi-Modal Image Classification: Distinguishing between 2D, 3D, and Realistic Imagery

Submitted by: Orel Cohen, Hen, Shiraz

YouTube short link: https://youtu.be/oQf3URrl71o

1. Problem Description

In the current digital landscape, the boundaries between Computer-Generated Imagery (CGI) and realistic photography are becoming increasingly blurred. Modern rendering engines (such as Unreal Engine 5).

The core problem this project addresses is the automated classification of visual content into three distinct categories:

- **2D Imagery:** Flat illustrations, sketches, cartoons, and line art.
- **3D Imagery:** Computer-generated objects and characters that possess depth, lighting, and rendering artifacts but are not "real" (e.g., video game assets, animated movie characters).
- **Realistic Imagery:** Authentic photographs capturing real-world objects and people.

The main technical challenge lies in training a Deep Learning model (likely a Convolutional Neural Network) to extract subtle visual features—such as texture patterns, lighting physics, and edge characteristics—that differentiate a high-quality 3D render from a photograph, or a stylized 2D drawing from a 3D model.

[Insert Image Here: A comparison collage showing a cartoon object (2D), a rendered object (3D), and a photo of the same object (Realistic)]

2. Business and Technology Use Cases

An automated system capable of distinguishing between these visual modes has significant applications in industry:

- **Content Moderation & Tagging:** Stock photo platforms (like Shutterstock or Getty Images) require accurate metadata. An automated classifier can tag uploaded content as "Illustration," "3D Render," or "Photo" to improve search relevance for users.
- **Augmented Reality (AR):** AR systems need to understand the nature of the environment. Distinguishing between a flat poster (2D) and a physical object (3D/Real) enables the system to overlay virtual elements correctly.
- **Copyright Protection & Media Forensics:** As "Deepfakes" and AI-generated content become prevalent, tools that can flag synthetic media (3D/Generated) versus authentic camera footage are essential for verifying information integrity.

3. Why it is Important

The volume of visual data generated daily makes manual sorting impossible. Developing a robust Deep Learning classifier for this task is important for several reasons:

1. **Automation:** It streamlines archival and database management processes without human intervention.
2. **Semantic Understanding:** It helps computer vision systems understand the *context* of an image—knowing that a photo of a cat implies a real animal, while a 3D model of a cat implies a virtual asset.

3. **Foundation for Generative Models:** Understanding the distinction between these domains is crucial for training future generative models (GANs/Diffusion models) to transfer styles between domains (e.g., turning a photo into a 3D character).