Motivation

- Problem: Distinguishing between computer-generated (CG) images and natural photographic (PG) images is crucial for verifying authenticity in the digital era, with important applications in social media verification, journalism, and forensics.
- Challenge: Modern AI generative models such as GANs, diffusion models, and autoregressive models create images that are almost indistinguishable from real photographs. Traditional detection methods struggle in this area.

Proposed Method – MDTL-NET

The authors introduce MDTL-NET (Multi-scale Deep Texture Learning Network), a deep learning framework designed to capture subtle texture differences between CG and PG images.

Key modules inside MDTL-NET:

1. Global Texture Representation Module (GTRM):
    - Built on top of ResNet.
    - Captures multi-scale texture patterns across different image levels.
    - Uses Gram matrix-based activation to represent global textures more effectively.
2. Deep Texture Enhancement Module (DTEM):
    - Uses semantic segmentation map-guided affine transformations.
    - Amplifies discriminative traces, the small and hard-to-see differences between CG and PG textures.
    - Enhances high-frequency components such as edges and fine patterns that are often lost in CG images.
3. Attention Mechanisms (Channel and Spatial):
    - Makes the model focus on informative regions and features instead of irrelevant content.
    - Channel attention learns the importance of feature channels.
    - Spatial attention highlights where in the image the useful features are.
4. Feature Fusion via Low-rank Tensor Representation (LTR):
    - Instead of simple concatenation, this method uses tensor representation and low-rank decomposition.
    - Allows the network to capture interactions between features while keeping computation efficient.

Main Contributions

1. A new detection network (MDTL-NET) that combines GTRM, DTEM, attention mechanisms, and LTR for robust CG detection.
2. Simultaneous enhancement of global texture patterns and local discriminative traces, unlike traditional CNNs.
3. A novel deep texture enhancement module that magnifies subtle differences between CG and PG images, guided by semantic segmentation maps.
4. A new dataset (DSGCG) with 84,000 images (42,000 CG and 42,000 PG). It includes traditional CG, GAN-based, and diffusion-model-generated images, covering diverse scenarios such as indoor, outdoor, objects, people, grayscale, and more.

Experimental Results

- Datasets used: DSGCG (new, proposed by authors), DSRah (2017), DSTok (2013), DSMan (2022).
- MDTL-NET consistently outperformed 10 state-of-the-art methods, including CNN-based and handcrafted-feature-based detectors.
- Achieved accuracy of 96.28 percent on DSGCG, higher than competing methods that often stayed below 90 percent.
- Showed strong generalization to unseen types of CG images, such as those generated by diffusion models.
- Demonstrated robustness to post-processing operations like compression, resizing, and filtering.

Why This Matters

This paper advances CG image detection beyond simple deep CNN classifiers. By enhancing and focusing on textures, MDTL-NET leverages the fact that CG and PG images differ not in obvious content but in subtle texture statistics, especially in high-frequency analysis. The new dataset also addresses the lack of large, diverse, real-world image collections for this research area.