12-15-2024

Distinguishing the Real from the Artificial in Digital Content

Nakul R. Padalkar
Boston University, npadalka@bu.edu

Rajiv Kohli
College of William and Mary, rajiv.kohli@mason.wm.edu

TREO
Technology, Research, Education, Opinion

Distinguishing the Real from the Artificial in Digital Content

A Critical Examination of Image Authenticity with Generative Models

Nakul R. Padalkar npadalka@bu.edu; Rajiv Kohli rajiv.kohli@mason.wm.edu;

With 5.43 billion individuals interconnected via social media, the rapid dissemination of digital content raises the importance of verifying authenticity. When creating content, generative artificial intelligence (GAI) further blurs the lines between what is real and what is synthetic in an ideologically divided society. Digital forensics to authenticate content will be critical to maintaining political stability and public trust (Auxier & Anderson, 2021; DataRePortal et al., 2024).

We leverage Convolutional Neural Networks (CNNs) and Vision Transformers to distinguish between device-captured and AI-generated images. We explore three questions: How effectively are computer vision algorithms distinguishing Real from Artificial? What is the comparative effectiveness of algorithms in feature extraction? How do prompt characteristics influence the perceived authenticity of images?

To address the questions, we enhance the VISION dataset (Shullani et al., 2017) with synthetic images generated from the generative AI models using prompt engineering. We used this enriched dataset to evaluate how the proposed model's performance. We combine prompt evaluation, advanced image recognition algorithms, and visual feature extractions. The prompt analysis provides insights into the effect of textual inputs (features) on generating visual outputs. This analysis helps identify the specific characteristics of AI-generated images and enables us to understand the underlying biases and patterns embedded within the AI models.

Our preliminary findings indicate substantial advancements in detecting synthetic images using ResNet architectures, significantly outperforming traditional image-only detectors. The ResNet18 model achieves a Top 1 Accuracy of 72.87% and a Weighted AUC of 97.95%, with deeper models like ResNet34 and ResNet50 showing consistent enhancements across metrics. The pinnacle, ResNet152, reaches a Top 1 Accuracy of 81.72% and a Weighted AUC of 99.18%. By integrating advanced computer vision techniques, this research aims to strengthen the reliability of digital media content, ensuring that the independent press can continue to uphold its role as a watchdog of democracy. Beyond technical solutions, our findings offer insights into the ethical dimensions of AI and journalism and inform regulatory frameworks needed for tomorrow's media landscape.

References

Auxier, B., & Anderson, M. (2021). *Social Media Use in 2021*. Pew Research Center. https://www.pewresearch.org/internet/2021/04/07/social-media-use-in-2021/#fn-27044-1

DataRePortal, K., Meltwater, I., & WeAreSocial, I. (2024). *Digital 2024: Global Overview Report —*. DataReportal – Global Digital Insights. https://datareportal.com/reports/digital-2024-global-overview-report

Shullani, D., Fontani, M., Iuliani, M., Shaya, O. Al, & Piva, A. (2017). VISION: a video and image dataset for source identification. *EURASIP Journal on Information Security*, *2017*(1), 15. https://doi.org/10.1186/s13635-017-0067-2