Detecting Synthetic Multimedia: Imagery vs. Video

EXECUTIVE SUMMARY

Traditional multimedia forensics involves examining imagery, video, or other media to assess origin and likelihood of manipulation. It relies on identifying subtle visual indicators that may or may not be noticeable to the human eye. GenAI synthetic content adds an element of novelty to the field of multimedia forensics. GenAI-derived synthetic media can create incredibly sophisticated depictions of false situations that resemble authentic photos and videos. Fortunately, synthetic content often contains subtle visual clues of its own that can be uncovered with careful human analysis or with the assistance of tools that reveal artifacts imperceptible to the human eye.

The previous *Emerging Tech* report explained key principles of image forensic analysis and its relevance to identifying tampered imagery. This report discusses how principles of multimedia forensics may be applied to detect synthetic imagery and video. Certain anomalies in synthetic content can be detected through careful human observation. In other cases, GenAI detection tools can identify visual irregularities in synthetic media that are too subtle for the human eye to notice. We will discuss three broad categories of algorithms that can detect the presence of synthetic media:

- **Body motion analysis** leverages deep learning to detect subtle anomalies in how a person in a synthetic video moves or blinks.[1]
- **Frame transition analysis** couples elements of pixel analysis with deep learning to predict the logical behavior of a sequence of frames in a video. Deviations from the predicted logic can indicate AI manipulation.[2]
- **Biometric analysis** uses deep learning to enhance the detection of biological indicators like blood flow in the faces of authentic video subjects. The absence of such biometric cues can indicate possible AI manipulation.[3]

References to individual companies are listed in this report only to serve as examples of the existing technology market; they do not entail Department of State GEC endorsement or preference of any private company.

TECHNOLOGY OVERVIEW

Like traditional digital forensics, GenAI synthetic content detection hinges on the identification of subtle visual indicators that may or may not be noticeable to the naked eye. It is important to first explain the common variables of synthetic imagery and synthetic video. As elaborated below, synthetic videos (particularly those which impersonate an individual) are often produced with different tools than synthetic imagery.[4] As such, synthetic image and video can necessitate different angles of analysis.

Synthetic imagery is frequently produced through commercial text-to-image tools. These tools typically utilize GAN and/or diffusion-based technology to generate an image from a text prompt.[5] They pull from a vast trove of internet data to construct the images. The vastness of this dataset can result in stylistic or functional inconsistencies to the output image.[6] For example, the synthetic image of French President Emmanuel Macron (at right) displays multiple clues of inauthentic origin: the eyes are misaligned with a vacant stare and the hands are clasped in an unrealistic way. As a

***Figure 1**: Many synthetic images contain visible errors or clues (example generated from deepai.org).*

further contextual data point, the flag in the background does not resemble the French flag in any way.

Synthetic video is often produced with some combination of Variational Autoencoder (VAE), diffusion, or GAN-based technology.[7] A content creator "trains" a tool on video and/or images of a specific person to generate a lifelike impersonation of that person. Unlike text-to-image tools, the content creator can make the output synthetic video more lifelike by providing the tool with highly specific sample data on the video subject. A skilled content artist can use GenAI tools, sometimes combining with traditional editing techniques, to produce an impersonation video that is more difficult to distinguish from an authentic video recording.[8]

Figure 2: *A text-to-image generation of Ukrainian President Volodymir Zelenskyy (L) beside a still frame from a synthetic video (R). Each example contains visible (yet distinct) flaws.*

These factors are subject to change, particularly as text-to-video technology also becomes more readily available.[9] In the meantime, however, they are key to understanding why certain types of synthetic content require different methods of detection.

VISUAL FACTORS

The last *Emerging Tech* report on *Image Forensic Analysis and Applications to Synthetic Content Detection* introduced some common methodologies on assessing content for possible GenAI manipulation. These methodologies encourage a viewer to examine the context surrounding an image to evaluate for possible inconsistencies or errors.[10] A 2024 study by Northwestern University outlines many of the common errors and inconsistencies that frequently arise in synthetic imagery:

- **Anatomical and biometric irregularities**: Extra fingers or thumbs, limbs at odd angles, etc.[11]
- **Stylistic or functional clues**: Strange colors, shapes, or lighting in the image.[12]
- **Violations of physics**: Scientifically incorrect shadows, reflections, movements, etc.[13]
- **Sociocultural implausibility**: For instance, an alligator crawling on the streets of New York City.[14]

Similar principles apply when analyzing potential synthetic video. As illustrated above, the context and origin of synthetic videos (particularly impersonations) can differ considerably from the most common synthetic images. A study from MIT analyzed certain irregularities that commonly arise in synthetic videos.[15] The visual cues are often distinct from synthetic imagery, but there are certain similarities:

- **Anatomy of the subject's face**: Unnatural appearance of skin, blemishes, hair, etc.[16]
- **Lighting and physics**: Unrealistic pattern of light and shadows when a video subject moves.[17]
- **Blinking**: Excessive, minimal, or inconsistent blinking can indicate GenAI manipulation.[18]
- **Lips**: Unnatural movements that do not match the words being pronounced.[19]

DEEP LEARNING IDENTIFICATION

GenAI algorithms that detect synthetic video also frequently rely on "visual" cues, but they can identify patterns that are practically invisible to the human eye.[20] This is similar to the concept of image forensics, as discussed in the previous report. Just as multimedia forensics can identify subtle manipulations to the pixels of an image, specialized artificial neural networks can help identify telltale patterns of GenAI production in a video. There is a rapidly expanding market of GenAI tools that can detect synthetic videos with varying degrees of accuracy.[21]

Such tools employ a range of deep learning algorithms that compare generalized data from authentic videos as a baseline against possible synthetic videos.[22] Based on this comparison, the algorithm can then return an assessment on the likelihood of GenAI content. For convenience, we have categorized these techniques into three broad categories: body motion analysis, frame transition analysis, and biometric signals. This is not an exhaustive list, and it will inevitably change as technology continues to evolve.

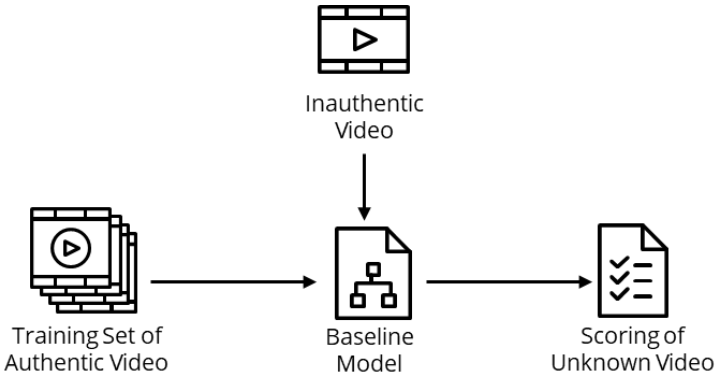

Figure 3: *Deep learning detection tools first train a baseline model on a set of data. In this case, the models are trained on authentic videos to understand what is expected. Once the model is built, the user can supply authentic and inauthentic videos, and the model can determine a probability score of authenticity.*

BODY MOTION ANALYSIS

Specialized algorithms can analyze videos for artificial body movements that are difficult for the human eye to detect alone. Artificial blinking is one such body motion that can leave significant digital clues. Authentic human blinking patterns typically differ by age, time of day, and lighting in a room, amongst other factors.[23] A single blink usually lasts no longer than a fraction of a second, yielding minimal training data for a GenAI content generator to replicate.[24] Certain GenAI video tools may be able to create artificial blinking that appears genuine to a casual observer, but analysis of genuine data with the integration of machine learning techniques can identify many of the inconsistencies.[25] In a related manner, synthetic video "lip syncs" can manipulate a subject's lips to more closely resemble a false or recontextualized video clip.[26] A capable algorithm can spot the inconsistencies in GenAI lip syncs, many of which pass undetected by the human eye.[27] In each of these examples, a deep learning tool generalizes the typical behavior of body motion in authentic video recordings and employs it as a baseline when assessing a video of unknown origin.

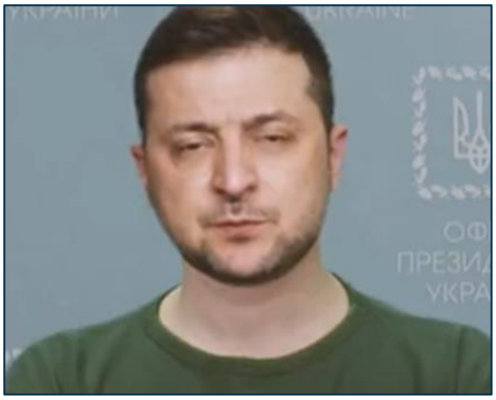

Figure 4: *Abnormally slow blinking was one characteristic of the Zelenskyy impersonation.*

FRAME TRANSITION

Frame transition analysis can predict the logical sequence of frames in a video, thereby identifying disruptions to the logical sequence caused by GenAI manipulation. Face-swap and other types of impersonation videos often leave subtle clues of tampering in the affected regions of a frame.[28] These clues can affect individual video frames, as well as the transition between frames.[29] Convolutional Long Short-Term Memory (ConvLSTM) is one type of recurring neural network that can predict the future of certain cells in a grid based upon the past states of neighboring cells.[30] In the context of synthetic video, unexpected variance in the frame structure and sequence can be a strong indicator of synthetic content.[31] Predictive modeling like ConvLSTM may be considered a close corollary to body motion analysis, but it is more holistic and not limited to facial aspects of the video frame.[32]

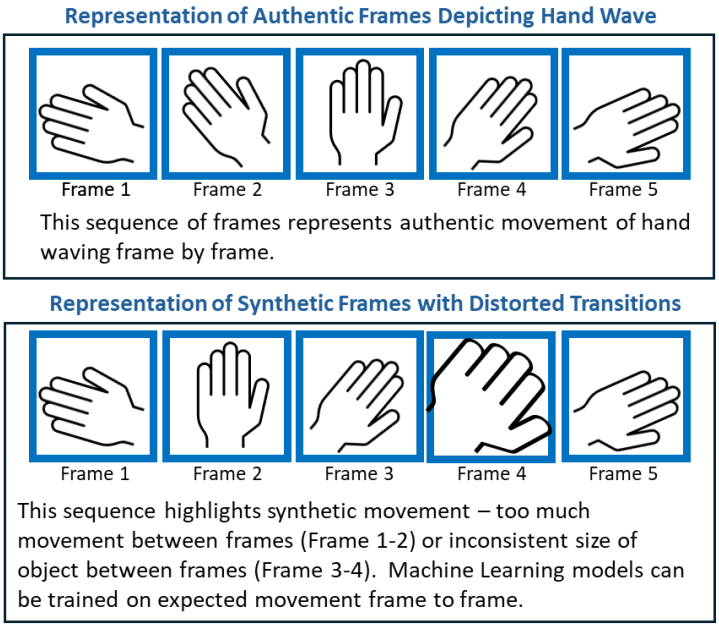

Figure 5: *Highlighting the process of frame algorithms can detect synthetic content.*

VISUAL BIOMETRICS

Biometric signals are an advanced form of pixel analysis that can be used to determine whether a given video is a true recording or synthetically generated. Visual cues associated with heart rate and blood flow are often indistinguishable to the naked eye, but they are nevertheless detectable in many video recordings.[33] Remote photoplethysmography (rPPG) can measure color changes in the face of a video subject based upon heart rate.[34] This methodology assesses the consistency of PPG signals in different regions of the subject's face within a video.[35] A neural network structure then classifies the PPG data, enabling the model to assess whether the facial behavior matches that of a live person or if it is likely to be a synthetic creation.[36] In 2022, Intel became the first major technology firm to market a synthetic video detector tool incorporating this technology.[37] Similar to the other two methods, biometric analysis uses common characteristics of authentic videos as a baseline for evaluating potential synthetic content.

TECHNOLOGY MARKET

Commercially available synthetic content detector tools typically employ one or more of these methods in customized algorithms to detect synthetic media. Many of them employ a multilayered framework of analysis, incorporating body motion, frame transition, and biometric analysis.[38] Many such tools offer only vague descriptions of their underlying technology, likely for proprietary reasons.[39] ***There is no single detection method that returns an accuracy rate of 100% when evaluating synthetic imagery or video.*** Even the most cutting-edge methods return correct results only 90-95% percent of the time on curated datasets.[40] The critical takeaway is that technological detection of synthetic content is essentially an exercise in probability, rather than a definitive result of "fake" or "real." Detectors often reflect this reality by providing results that simply score a likelihood of whether a given image or video is synthetic.

Future advances in synthetic content generation have potential to render current detection methods obsolete. Synthetic media detectors will inevitably need to refine and update their methodology to keep pace with technological advances. In this evolving environment, human digital literacy and critical analysis of online multimedia remains of paramount importance.

CONCLUSION

Visual anomalies play a central role in assessing whether a given video or image is synthetic in nature. An astute human analyst can spot many of these cues through careful examination. Nevertheless, certain visual cues are imperceptible to a human observer. In these instances, detection algorithms are critical in analyzing and flagging the anomalies associated with synthetic media. While the technologies associated with synthetic video and their detection are new, the underlying principles of multimedia forensics remain as relevant as ever. Government agencies and the private sector can train their personnel to critically evaluate digital content and use the relevant tools to analyze it.

We would appreciate your feedback by completing a short survey here.

ACKNOWLEDGEMENTS

The Department of State GEC gratefully acknowledges contributions of the following organizations to this report:

- Artificial Intelligence Security Center and Research Directorate, National Security Agency

REFERENCES

[1] https://arxiv.org/abs/2007.08517v1
[2] https://www.sciencedirect.com/science/article/abs/pii/S0020025522003504
[3] https://ieeexplore.ieee.org/document/9141516
[4] GEC Technology Engagement Technology Report, Synthetic Video Creation: Synthesis, Face-swaps, and Avatars, GEC2024-TET-007.
[5] GEC Technology Engagement Technology Report, Synthetic Image Creation: GANs, Autoencoders, and Diffusion Models, GEC2024-TET-006.
[6] Ibid.
[7] GEC Technology Engagement Technology Report, Synthetic Video Creation: Synthesis, Face-swaps, and Avatars, GEC2024-TET-007.
[8] GEC Technology Engagement Technology Report, Synthetic Video Creation: Synthesis, Face-swaps, and Avatars, GEC2024-TET-007.
[9] https://arxiv.org/abs/2311.15127
[10] GEC Technology Engagement Technology Report, Introduction to Image Forensic Analysis and Applications to Synthetic Content Detection, GEC2024-TET-010.
[11] https://arxiv.org/abs/2406.08651
[12] Ibid.
[13] Ibid.
[14] Ibid.
[15] https://www.media.mit.edu/projects/detect-fakes/overview/
[16] Ibid.
[17] Ibid.
[18] Ibid.
[19] Ibid.
[20] https://link.springer.com/article/10.1007/s10462-024-10810-6#:~:text=Rana%20et%20al.,literature%20on%20detecting%20deepfake%20videos.
[21] https://www.globalmarketestimates.com/market-report/deepfake-detection-software-market-4420
[22] https://gijn.org/stories/spotting-deepfakes-election-year/
[23]https://www.researchgate.net/publication/340813598_DeepVision_Deepfakes_Detection_Using_Human_Eye_Blinking_Pattern
[24] https://arxiv.org/abs/2007.08517
[25] Ibid.
[26] https://arxiv.org/html/2401.15668v1
[27] Ibid.
[28] https://www.sciencedirect.com/science/article/abs/pii/S0020025522003504
[29] Ibid.
[30] https://paperswithcode.com/method/convlstm
[31] https://arxiv.org/abs/2007.08517
[32] https://www.sciencedirect.com/science/article/abs/pii/S0020025522003504
[33] https://ieeexplore.ieee.org/document/9141516
[34] https://link.springer.com/article/10.1007/s00371-023-02833-x#Sec25
[35] https://arxiv.org/abs/1901.02212
[36] Ibid.

[37] https://www.intel.com/content/www/us/en/newsroom/news/intel-introduces-real-time-deepfake-detector.html#gs.csmlyn
[38] https://aimresearch.co/market-industry/5-ai-deepfake-detector-tools-for-2024
[39] https://www.vlinkinfo.com/blog/top-ai-deepfake-detector-tools/
[40] https://arxiv.org/abs/2007.08517v1