DISTINGUISHING AUTHENTIC FROM AI-GENERATED EXPLOSIONS USING SPATIOTEMPORAL DYNAMICS

Sarah Barrington, Hany Farid
University of California, Berkeley
{sbarrington}{hfarid}@berkeley.edu

Motivations

Understanding and responding to world events during military conflicts has always been challenging. This has been heightened in the age of deepfakes, which are now a key enabler of **warzone disinformation**.

As global conflicts intensify, there is an **increasing amount of fake attack and explosion footage** circulating online (Fig. 1).

It can be difficult to authenticate this kind of content because even state-of-the-art **AI-based deepfake detectors fail.**

These AI-based detectors fail because **footage of real explosions is semantically rare.**

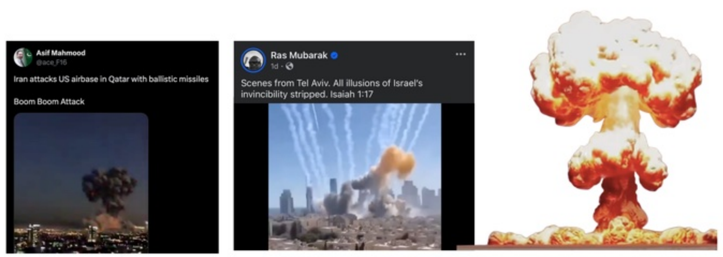


Fig. 1: examples of AI-generated conflict zone videos depicting fabricated explosions.

Research Questions

1. How well do state-of-the-art deepfake detectors work on semantically-rare combat explosion footage?
2. Given that AI detectors fail, can we build a forensic detector that captures something about physical plausibility?
3. Is there a way to combine AI and forensic approaches to get increased performance?

Methodology

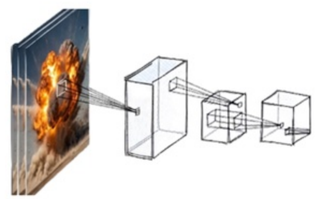
DATA

Cleaned and preprocessed real explosions footage

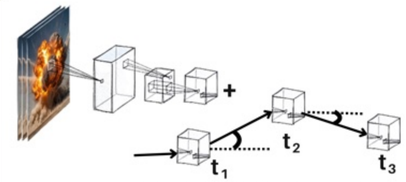
Real video

Matching fake I2V video

FEATURES

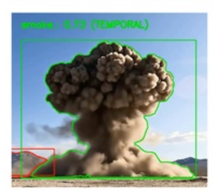


Approach 1: Spatial Embeddings [1]



Approach 2: Spatial + Temporal Embeddings [2]

Approach 3: Spatiotemporal Dynamics



Segment
Grounding Dino ZSOD + Meta SAM

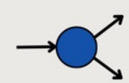
Optical Flow
$\vec{v} = (v_x \quad v_y)$

Fluid dynamics-inspired features
Δ hue and brightness

Divergence
Curl
Gradient

CLASSIFIERS

~6700 frames, 22 videos
Leave-one-out cross validation
Random forest models



Individual models

Ensemble of models

Results

- Baseline models achieve ~**95% accuracy on original datasets** but **near-chance performance (52–55%)** on the custom explosions dataset
- **Retraining** a random forest classifier on the custom dataset **increases accuracy to 78%** (spatial) and **92%** (spatiotemporal)
- **Hand-crafted dynamics** model alone achieves **83%** accuracy (F1 0.80)
- **Ensemble of spatiotemporal + dynamics model is the best performing model** (**94.4%**, F1 0.93)
- **Temporal features outperform spatial ones**, likely reflecting complex and physically constrained temporal properties

Qualitative Analysis

We identified a number of qualitative cues that frequently -- though not always reliably – distinguished real from synthetic explosions:

- Mushroom-cloud-shaped plumes
- Increased and unrealistic rise speed of plume (Fig. 2)
- Visual-acoustic offset

Fig. 2: rise heights of plume 2 seconds into synthetic and real videos

Discussion & Limitations

Authentic explosion videos are **semantically rare**, and so **general purpose deepfake detectors fail to generalize**. While our dataset is limited, we demomstrate the promise of temporal and physics-informed approaches for specific detection. **Bespoke detectors can complement general-purpose tools** by focusing on class-specific artifacts.

References

[1] Maxime Oquab, Timoth´ee Darcet, Th´eo Moutakanni, Huy Vo, Marc Szafraniec, Vasil Khalidov, Pierre Fernandez, Daniel Haziza, Francisco Massa, Alaaeldin El-Nouby, et al. Dinov2: Learning robust visual features without supervision. arXiv:2304.07193, 2023. 3
[2] Christian Intern`o, Robert Geirhos, Markus Olhofer, Sunny Liu, Barbara Hammer, and David Klindt. AI-generated video detection via perceptual straightening. arXiv:2507.00583, 2025. 3

With thanks to Dr. Kirk Yeager, Chief Explosives Scientist at the FBI's TEDAC, Dr. Tom Teschner, and Dr. Matt Williams. This work was funded by the UC Noyce Initiative.