Detection of Facial Deepfakes Using Interactive Liveness Tests

"Interaction exposes the fake."

Bc. David Drtil • supervisor Ing. Anton Firc, Ph.D.

Faculty of Information Technology, Brno University of Technology • 2026

1. The Challenge

Every day we meet our colleagues and family over **Microsoft Teams**, **Google Meet**, **Messenger** and **WhatsApp** video calls. A **single profile photo**, such as the one already public on Facebook or LinkedIn, is enough for tools like *FaceFusion* or *DeepFaceLive* to wear that face live on a webcam.

The result: impersonation scams that feel exactly like the real person, extortion built on fake intimate footage, and corporate attackers who walk into internal meetings with a stolen identity and ask for money, credentials, or a quick favour.

Passive detection alone is losing the arms race. Xception trained on one generator drops from **99.4 %** AUC on the training manipulation to **49.1 %** on FaceSwap (Li et al., CVPR 2020).

2. Our Solution

A **hybrid framework** that combines passive video analysis with an interactive **challenge–response protocol**. The system asks the user to perform **random actions** in ten-second windows.

The deepfake generator cannot smoothly reproduce these motions in real time, which **surfaces artifacts a passive detector can finally exploit**.

Randomised sequences defeat replay attacks, and the per-action peak passive score is fused with the binary success of each challenge before the session is allowed to pass. An early-exit flag ends the session the moment a frame is flagged as fake, so later interactions cannot fix it.

3. Threat vs. Defence

Figure 1 — the threat

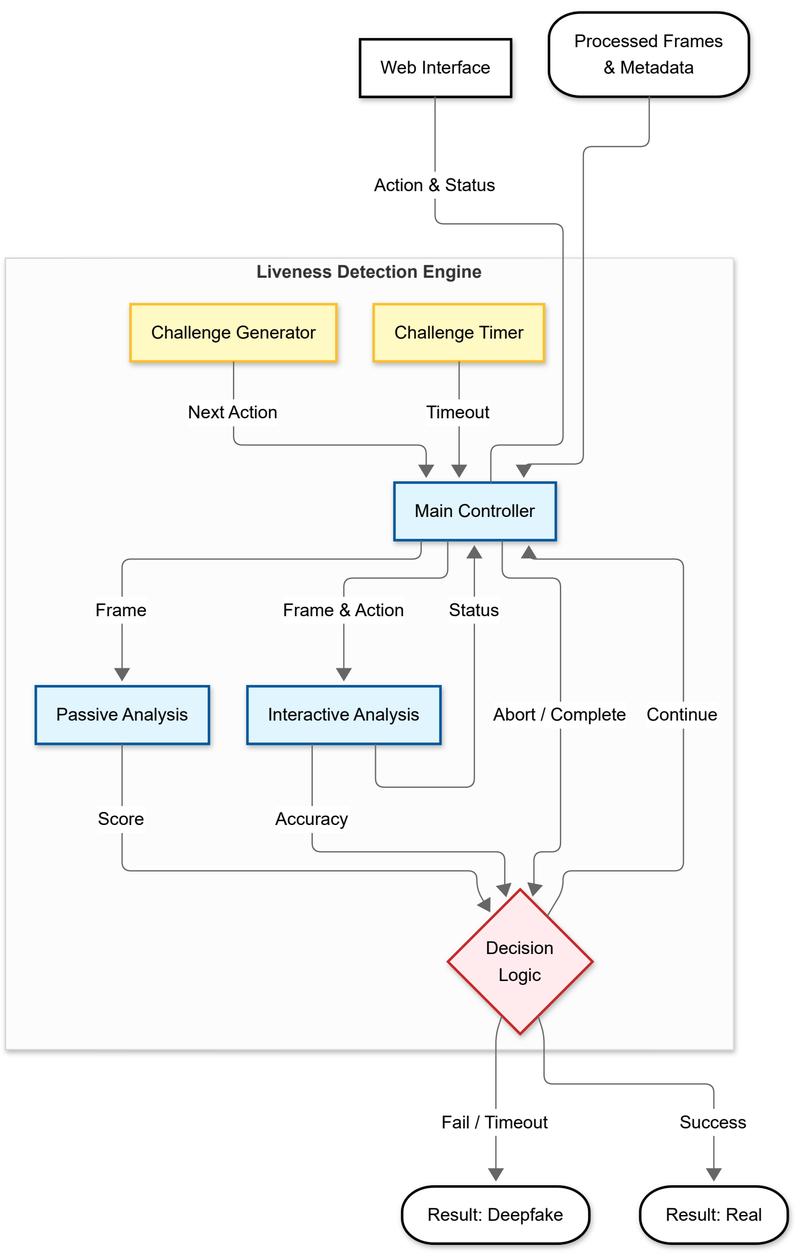
Live face-swap of the author onto Tom Cruise. The webcam feed is intercepted by *DeepFaceLive*, passive verifier sees a realistic face in a realistic setting.

Figure 2 — our detection in action

Our web app during a hand-occlusion challenge: MediaPipe face mesh, hand skeleton, *passive score*, pose angles and the prompted action are all fused in one frame.

4. System Pipeline

Figure 3 — end-to-end detection engine



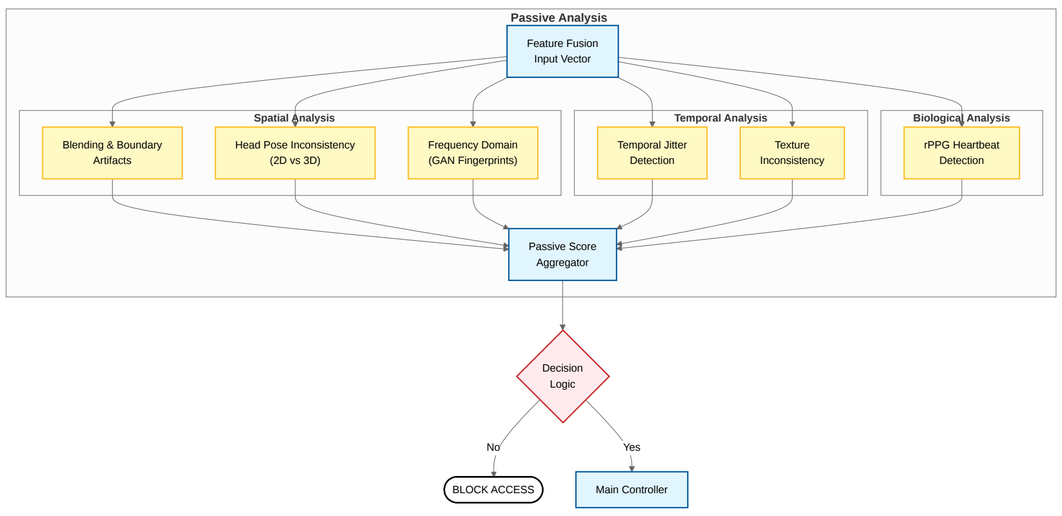
Webcam frames run through **Pre-processing** (One Euro smoothing, face alignment and resize). **Passive** and **Interactive** modules work in parallel. **Decision Logic** fuses per-action peak passive scores with the binary success of all challenges.

5. Passive Branch

Three parallel detectors score every frame:

- **Spatial** – UCF on an Xception backbone, boundary and blending artifacts.
- **Frequency** – SPSL on the image spectrum, GAN upsampling fingerprints.
- **Temporal** – CViT-v2 as a *video* model over a 15-frame sliding window. It also catches **replay attacks** (pre-recorded loops) by flagging large frame-to-frame jumps and unnatural temporal transitions.

Figure 4



A weighted aggregator fuses the three scores and applies a 10-frame moving average:

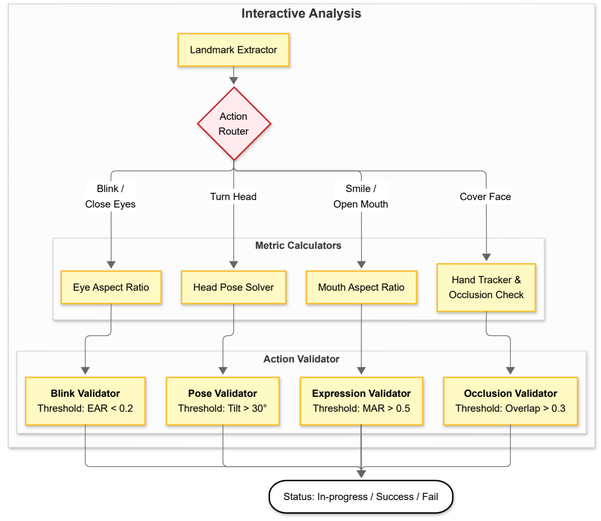
$$S_{\text{passive}} = \text{ma}_{10}\left(w_s s_s + w_t s_t + w_f s_f \right)$$

6. Interactive Branch

MediaPipe emits **468 face** and **21 hand** landmarks per frame. An *Action Router* activates only the metrics needed for the current challenge and enforces hard geometric thresholds:

- **Head Pose Solver** — pitch/yaw/roll from the PnP problem on 5 landmarks.
- **EAR / MAR** — eye and mouth aspect ratios for blink and expression.
- **Hand Tracker** — intersection of the hand mask and the face bounding box.

Figure 5



Each action has a 10-second timeout. A *randomised sequence* of 8 actions is selected for each session, so an attacker cannot prepare a pre-rendered clip that happens to match.

7. Key Contributions

Hybrid passive + interactive framework with a per-action peak score aggregation that lifts the signal exactly when artifacts are most likely.

Real-time web server built on MediaPipe, UCF, SPSL and CViT-v2, streaming over WebSocket to a thin-client web browser.

InsightFace ArcFace identity embeddings filter source-target pairs by cosine similarity.

Replay-safe protocol with randomised action sequences, per-action timeouts, and CViT-v2 as a temporal watchdog for recorded loops.

A system tested against live deepfake attacks generated by DeepFaceLive and FaceFusion.

8. Try It

GitHub repository

`daviddrtil/DeepFaceID`

`xdrtil03@stud.fit.vutbr.cz`

Selected references. [1] Firc, Malinka, Hanáček. *Deepfakes as a threat to speaker and facial recognition.* Heliyon 9(4), 2023. [2] Šalko, Firc, Malinka. *Security Implications of Deepfakes in Face Authentication.* ACM SAC, 2024. [3] Wang, Luo, Lau. *Living a Lie: Security Analysis of Facial Liveness Detection in Mobile Apps.* ACNS, 2024. [4] Li et al. *Seeing is Living? Rethinking Facial Liveness in the Deepfake Era.* USENIX Sec., 2022. [5] Shiohara, Yamasaki. *Self-Blended Images.* CVPR, 2022. [6] Yan et al. *UCF: Uncovering Common Features for Generalizable Deepfake Detection.* ICCV, 2023. [7] Wodajo et al. *CViT-v2: Deepfake Video Detection.* 2024. [8] Lugaresi et al. *MediaPipe: A Framework for Perceiving & Processing Reality.* CVPR Wkshp, 2019.