



浙江大學
ZHEJIANG UNIVERSITY



上海財經大學
SHANGHAI UNIVERSITY OF FINANCE AND ECONOMICS

EchoFence: Non-Intrusive Forgery Detection in Video Conferencing Via Ultrasonic Sensing

Leqi Zhao, Luxin Shi, Jianwei Liu, Rui Xiao, Jinsong Han

INFOCOM 2026

Video Conferencing Forgery



CNN World Sign in

WORLD > ASIA • 2 MIN READ

Finance worker pays out \$25 million after video call with deepfake 'chief financial officer'

FEB 4, 2024

A finance worker at a multinational firm was tricked into paying out \$25 million to fraudsters using deepfake technology to pose as the company's chief financial officer in a video conference call, according to Hong Kong police.

Can we truly trust the person we are talking to?

Existing Solutions



Passive Detection



Spatiotemporal anomalies



Lip-audio mismatch

- ✗ Vulnerable to advanced synthesis
- ✗ Ineffective against replay attacks

Active Detection



Challenge-response prompts



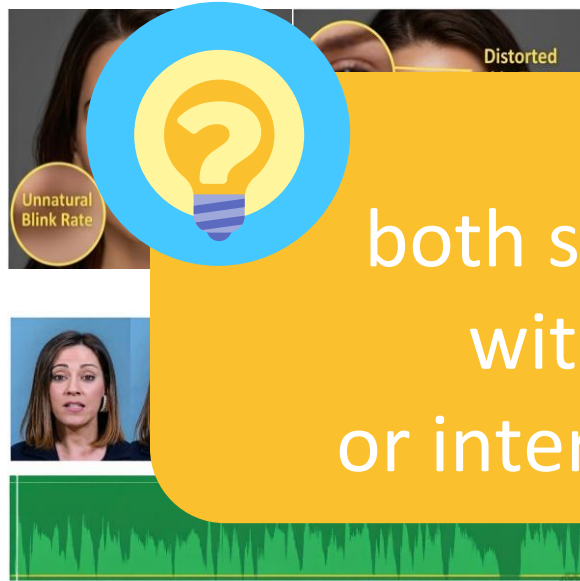
Live visual patterns

- ✗ Intrusive
- ✗ Sensitive to occlusion and behavior variation

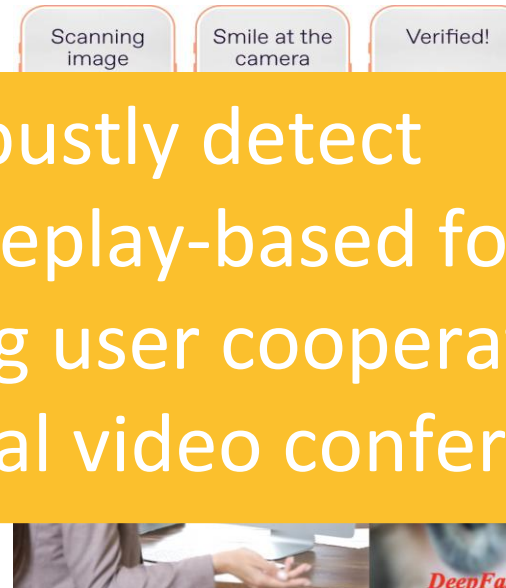
Existing Solutions



Passive Detection



Active Detection

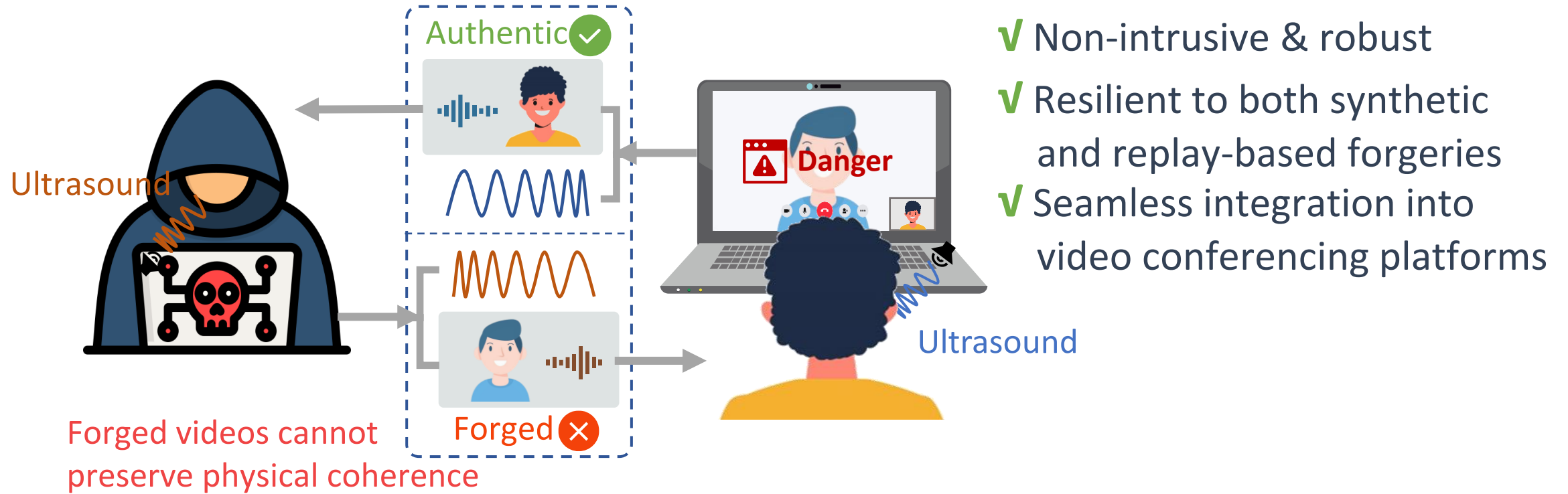


Can we robustly detect both synthetic and replay-based forgeries without requiring user cooperation or interrupting natural video conferencing?

- ✗ Vulnerable to advanced synthesis
- ✗ Ineffective against replay attacks

- ✗ Intrusive
- ✗ Sensitive to occlusion and behavior variation

Our Solution: EchoFence



- ✓ Non-intrusive & robust
- ✓ Resilient to both synthetic and replay-based forgeries
- ✓ Seamless integration into video conferencing platforms

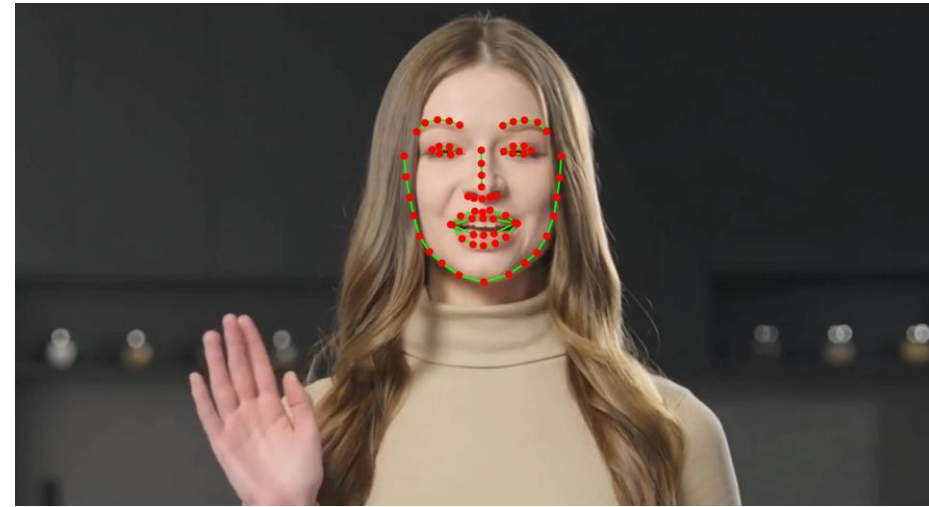
- Physical world probe using imperceptible **ultrasonic signals**
- **Motion coherence verification** between acoustic and visual modalities

Challenge 1



Reliable motion estimation under non-cooperative setting

Subtle motions & unconstrained motions

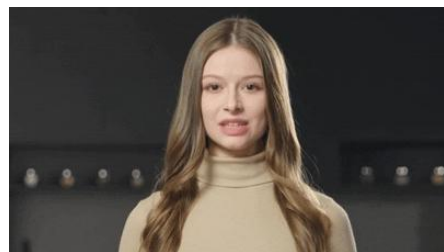


 Capture only coarse motion at predefined facial points

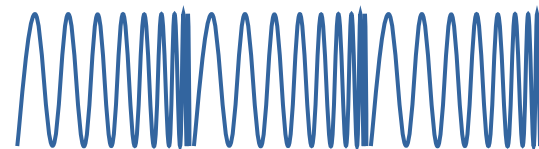
Challenge 1



 Reliable motion estimation under non-cooperative setting



Video frames



Frequency-modulated ultrasonic chirps

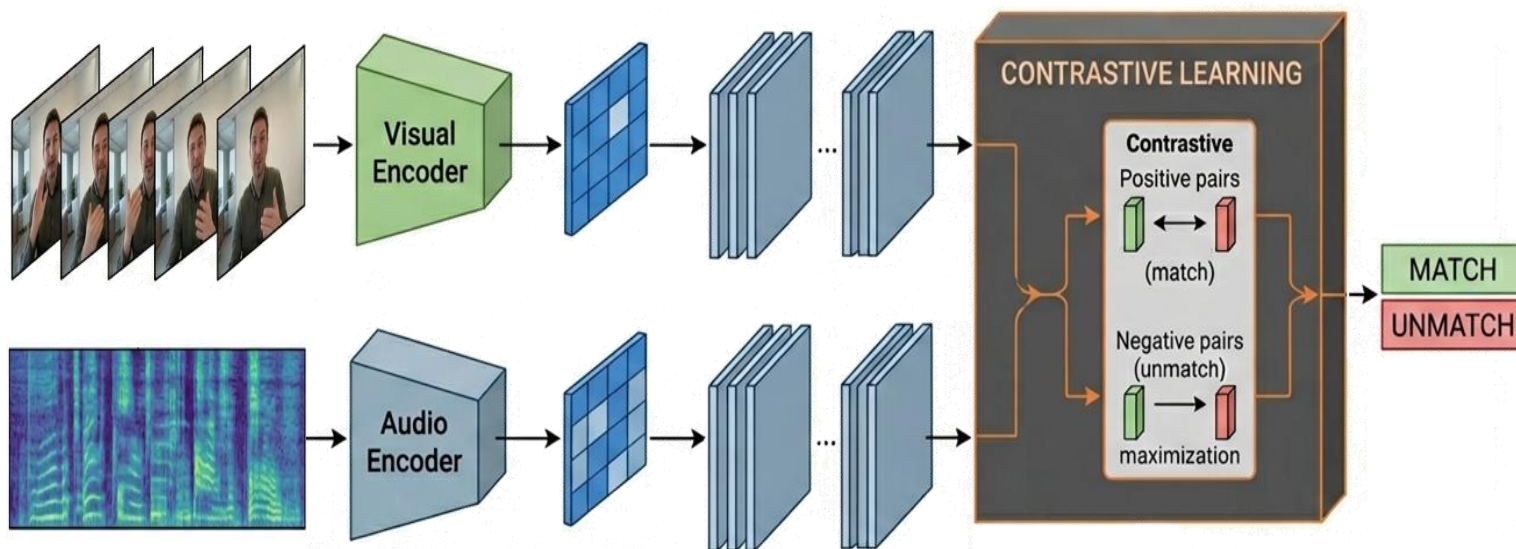


Capture fine-grained facial & upper-body motion

Challenge 2



Cross-modal verification without modality-specific training

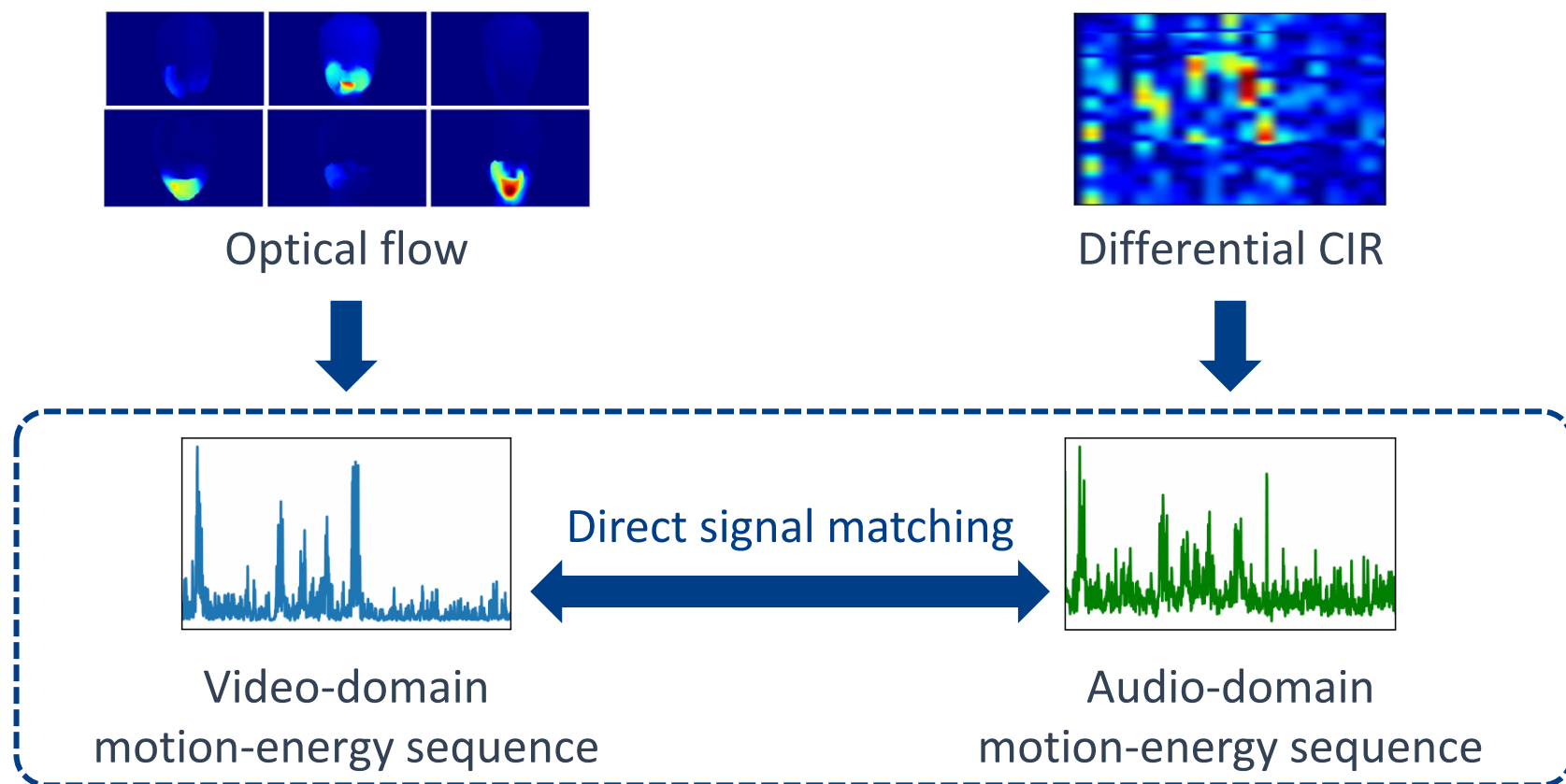


🙄 Extensive training

Challenge 2



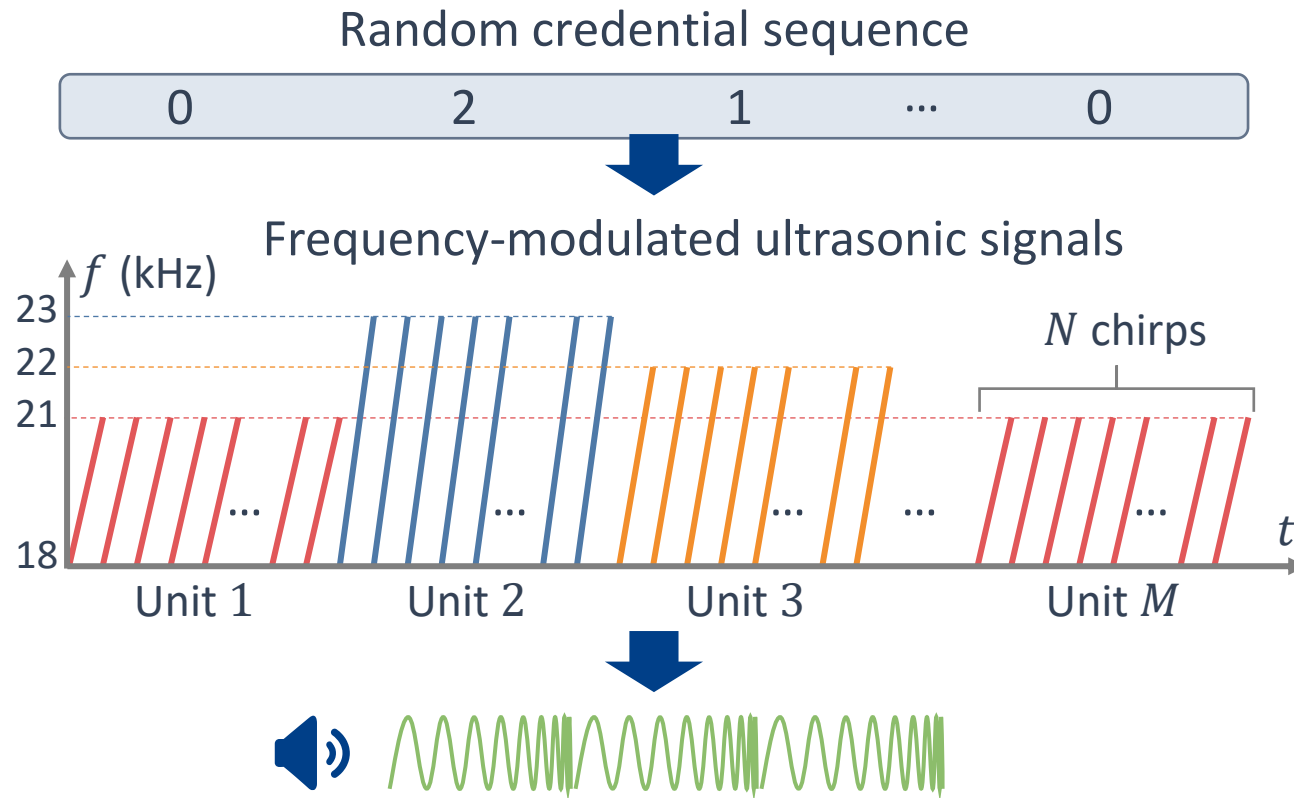
 Cross-modal verification without modality-specific training



Challenge 3



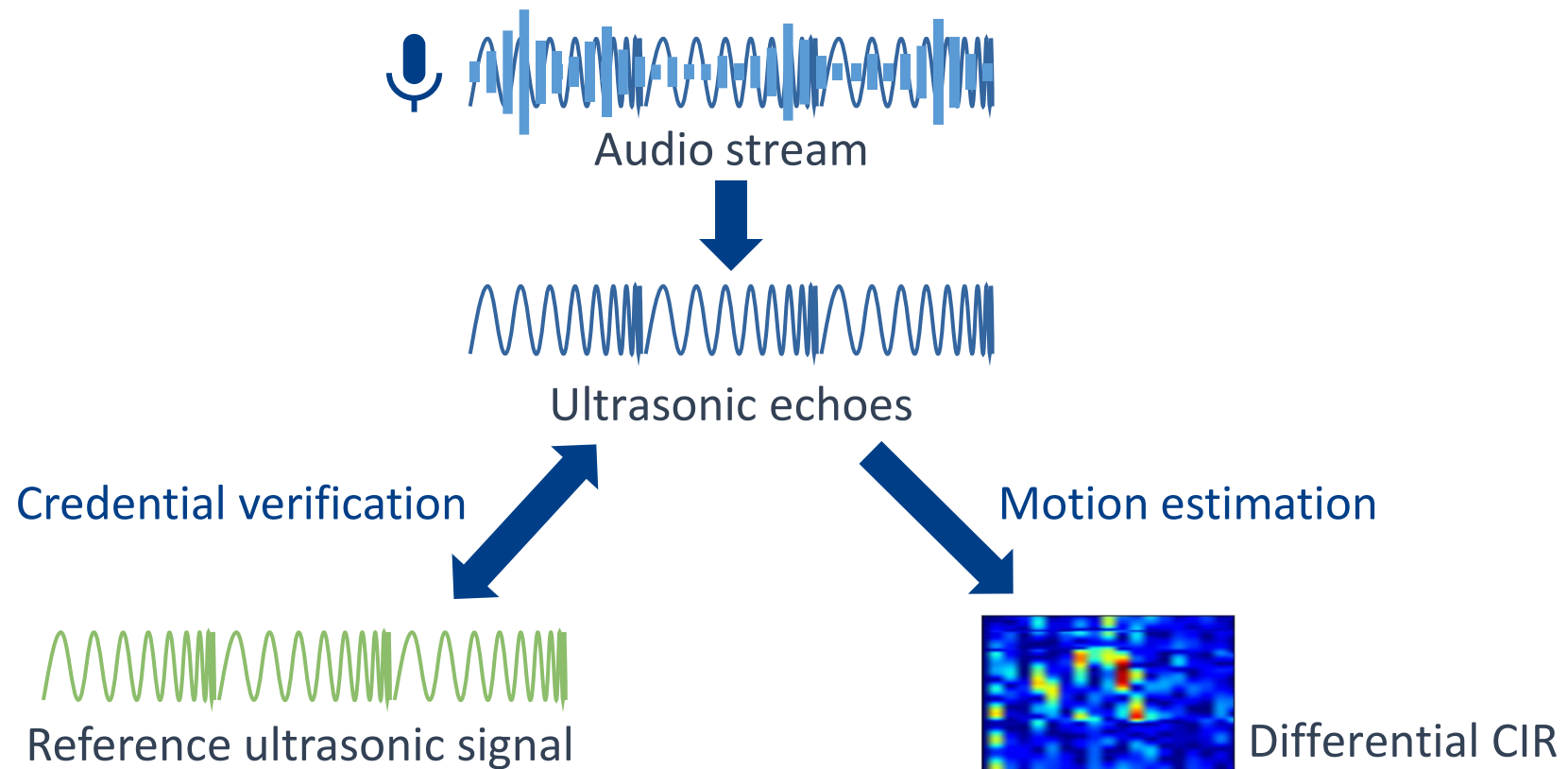
Ensuring replay resilience while maintaining sensing utility



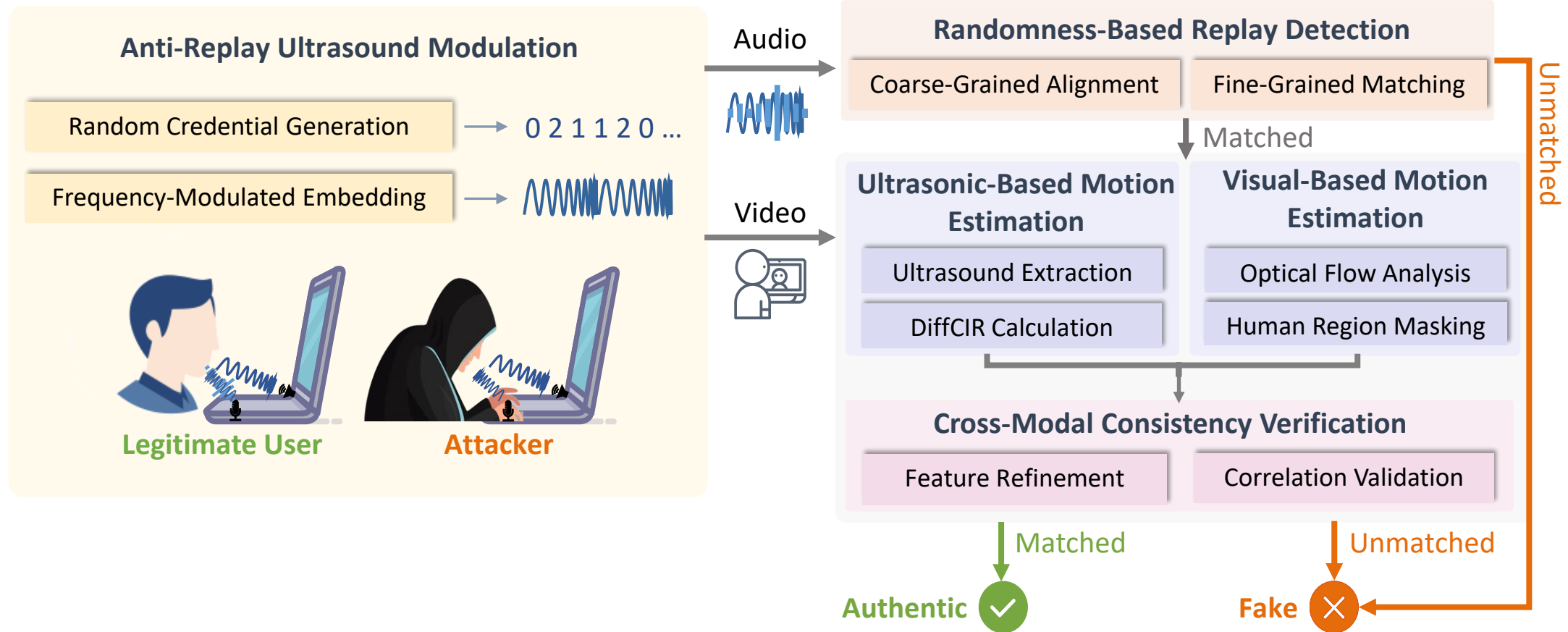
Challenge 3



 Ensuring replay resilience while maintaining sensing utility








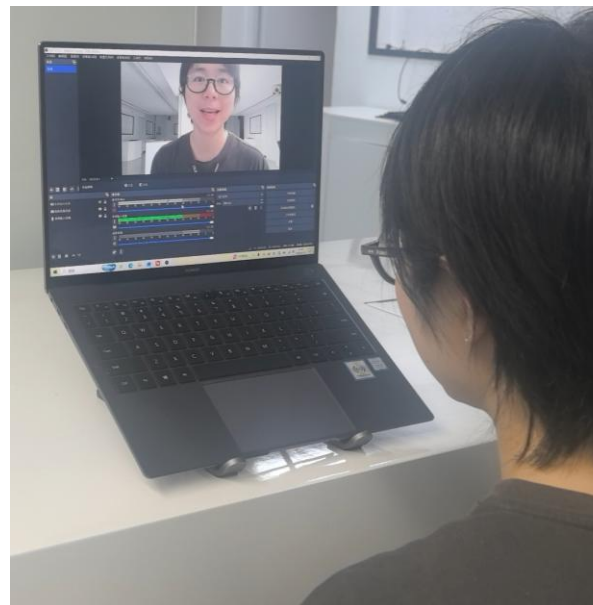
System Overview



Experimental Setup

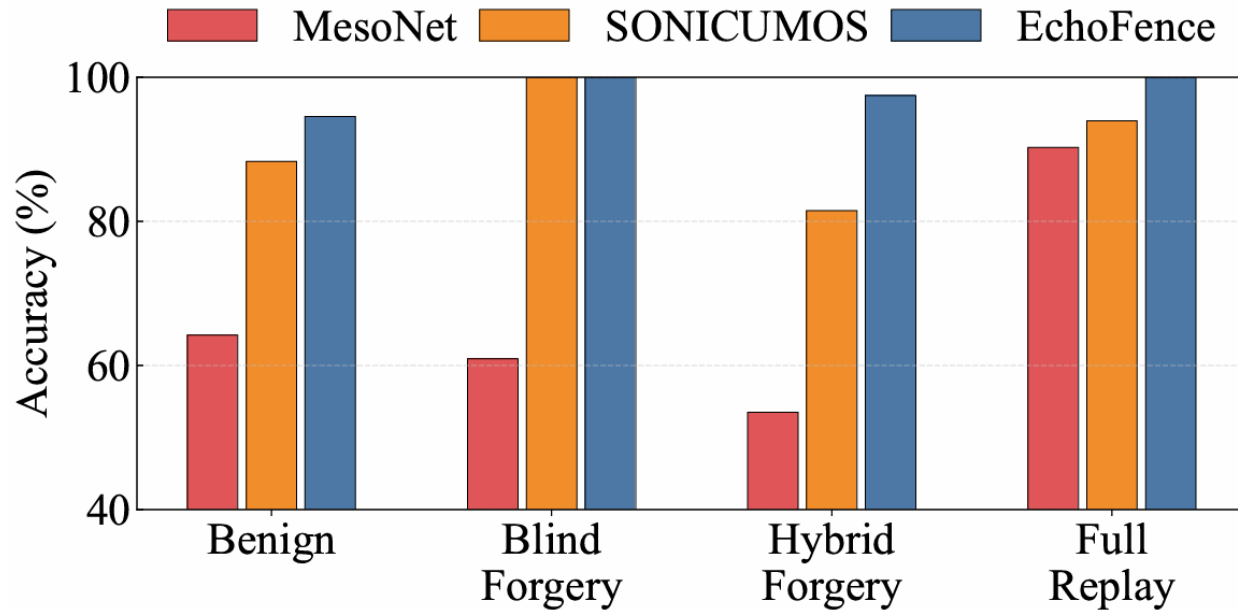


-  Commodity laptop
-  20 FPS video & 48 kHz audio
-  Five volunteers
-  12+ hours benign samples
-  Three forgery attack scenarios



	Benign	Blind forgery	Hybrid forgery	Full replay
Video frames	Live-recorded	Synthetic / Replayed	Synthetic / Replayed	Replayed
Ultrasonic echoes	Live-recorded	None	Live-recorded	Replayed

Overall Performance



■ MesoNet

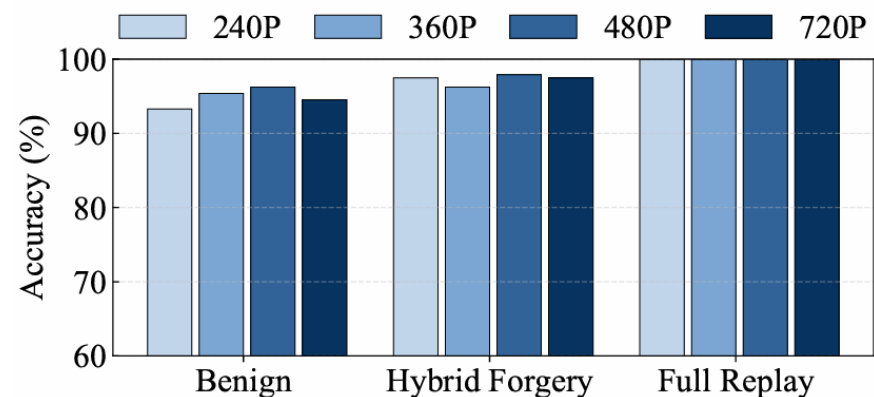
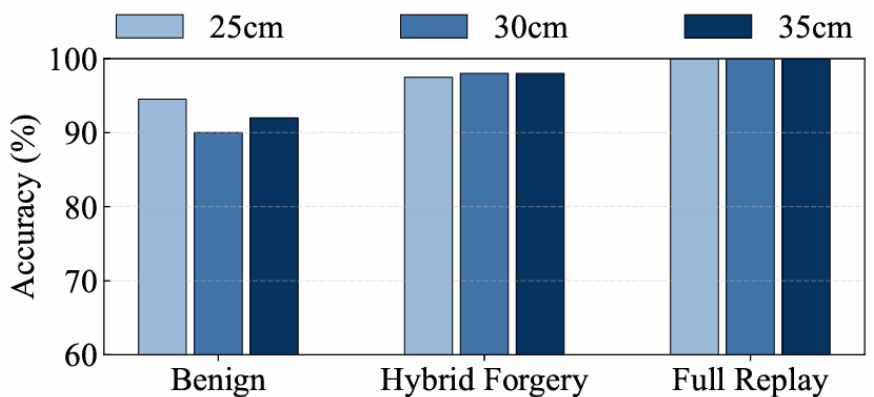
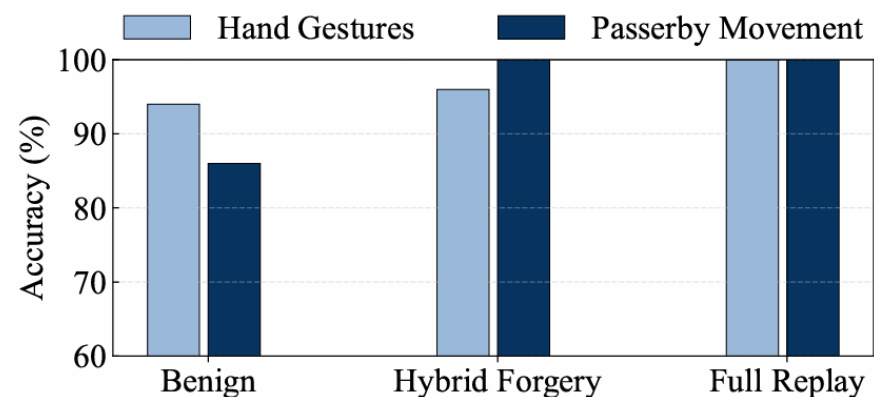
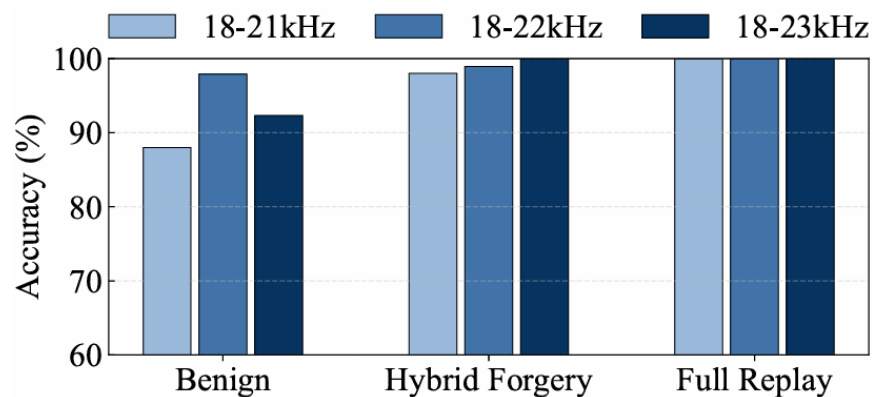
■ Representative passive method

■ SONICUMOS

■ State-of-the-art active approach

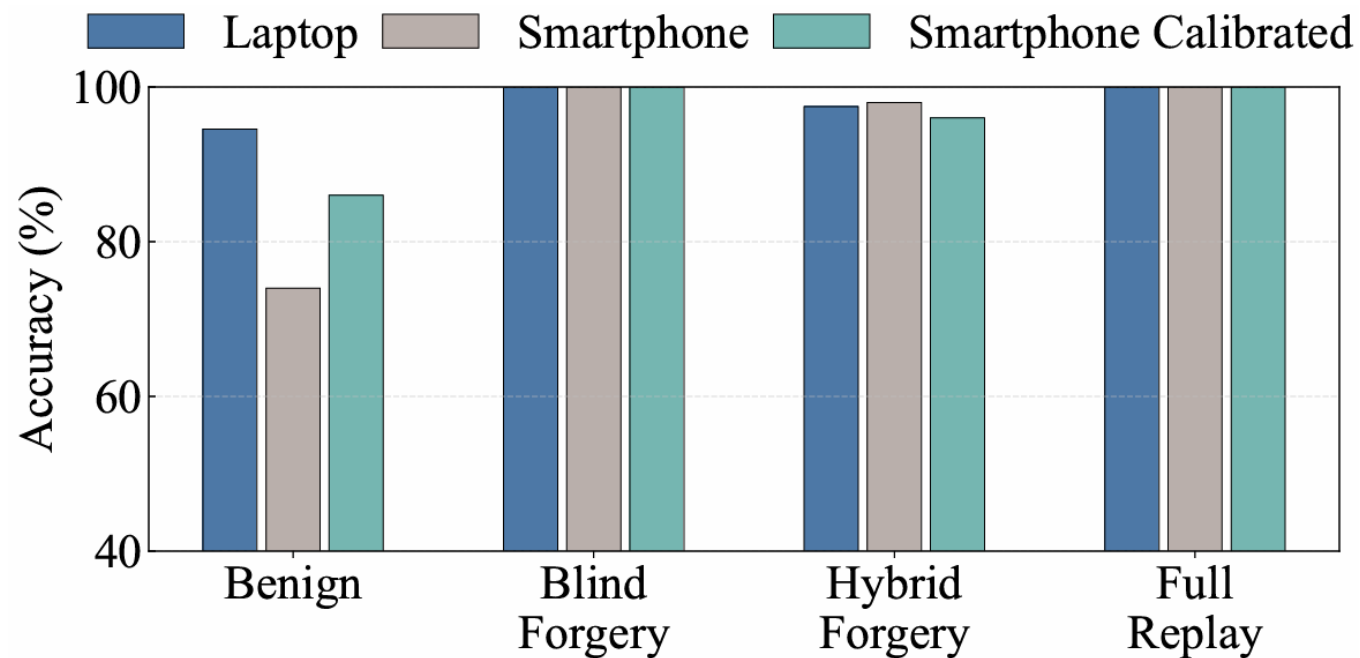
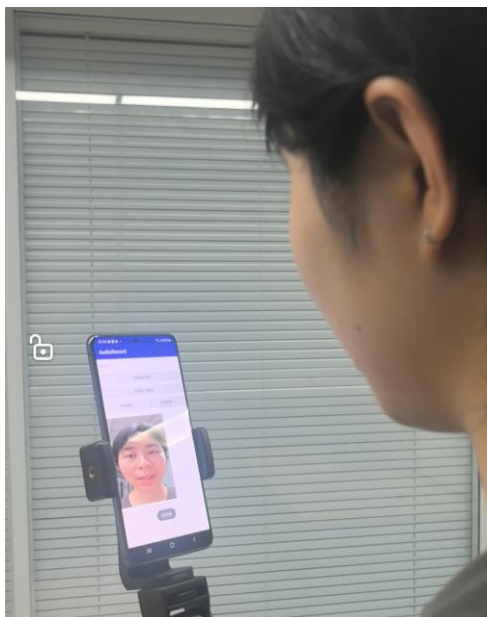
EchoFence consistently achieves over 94% detection accuracy across both benign and attack scenarios, outperforming both baselines

Robustness Evaluation



EchoFence maintains reliable performance across diverse conditions

Device Scalability



EchoFence achieves effective detection with a minor threshold calibration

Conclusion



Toward trustworthy video conferencing

- ✓ Non-intrusive & robust forgery detection framework
- ✓ Replay-resistant ultrasonic sensing
- ✓ Cross-modal physical consistency verification





浙江大學
ZHEJIANG UNIVERSITY



上海財經大學
SHANGHAI UNIVERSITY OF FINANCE AND ECONOMICS

Thank you!

