Facial Invisible Watermarking for Deepfake Sexual Crime Defense

Jang, Namwoo (School: Cheonan Upsung High School)
Choi, Sion (School: Seolhwa High School)
Han, Jiwoo (School: Gongju Meister High School)

Deepfake technology has rapidly evolved into a pervasive tool for digital exploitation. By the end of 2025, over 8 million deepfakes had been shared online, a dramatic increase from about 500,000 in 2023, with roughly 98% consisting of non-consensual pornographic content targeting private individuals and minors. This surge has caused serious societal harm, including lasting emotional and psychological damage to victims. However, existing countermeasures remain largely reactive, relying on post-hoc detection trained on public-figure datasets with limited protection for ordinary users, especially minors. To address these limitations, this research proposes OnMark, a personalized pre-emptive deepfake defense framework. OnMark embeds invisible identity-linked watermarks into facial images before manipulation. A unique pseudo-random (PN) sequence from facial features is embedded into the Level-2 LH/HL subbands of the blue channel using DWT. An oval-region masking strategy selectively encodes central facial areas preserved in face-swap pipelines. The method maintains good visual fidelity with PSNR 31.73 dB and SSIM 0.9558. Evaluated on 529 facial images under attacks including Delaunay face swapping, Poisson blending, JPEG (Q=75), and Gaussian blur, it achieved TPR 71.3%, FPR 12.9%, and AUC 0.8206. Oval masking notably improved sample separability by doubling the separation between watermarked and non-watermarked samples while reducing FPR by over 25 percentage points. Unlike conventional post-hoc detection, OnMark enables persistent identity linkage after manipulation. This shifts the paradigm from detection to proactive prevention and traceability, offering scalable protection for vulnerable users, especially minors.

Awards Won:
Association for the Advancement of Artificial Intelligence: AAAI Student Memberships for each finalist that is part of the 1st, 2nd, and 3rd Prize Winning projects and 5 Honorable Mention winning projects (up to 3 students per project) (in-kind award / part of the 1st-3rd prize)