Privacy Preservation in Event-Based Vision: Risks, Methods, and Trade-offs

The effect of applying perturbations on the privacy and visual naturalness of face images reconstructed from event-based data

Matei-Ioan Oprescu[1]
Supervisor(s): Nergis Tömen[1], **Tunahan Parlayici**[1]
[1]EEMCS, Delft University of Technology, The Netherlands

A Thesis Submitted to EEMCS Faculty Delft University of Technology,
In Partial Fulfilment of the Requirements
For the Bachelor of Computer Science and Engineering
June 21, 2026

Name of the student: Matei-Ioan Oprescu
Final project course: CSE3000 Research Project
Thesis committee: Nergis Tömen, Tunahan Parlayici, Ricardo Marroquim

Abstract

Facial recognition systems pose significant privacy risks, encouraging the development of generative adversarial evasion methods, such as AMT-GAN and Adv-CPG. While effective on clean, high-resolution RGB images, it remains unknown whether facial protection methods are still effective under the reconstruction pipeline of event-based cameras. This research investigates the privacy-naturalness trade-off of applying adversarial makeup to event-reconstructed faces. CelebV-HQ video clips were converted to event streams, reconstructed into grayscale images using E2VID under different thresholds, and evaluated for Attack Success Rate (ASR) and Structural Similarity (SSIM). The results reveal the following: the event-reconstruction process reduced AMT-GAN's protection effectiveness, dropping mean ASR across four white-box models. A contrast-threshold ablation indicated this reduction is a direct result of the event-generation process itself, rather than just data loss from sparse event streams. Furthermore, a qualitative evaluation of Adv-CPG showed a serious identity over-shifting and mode collapse, failing to mentain the structural diversity of the reconstructed face-image inputs. Finally, my research shows that current RGB-based adversarial protections are highly sensitive to domain shifts and fail to provide appropriate privacy for event-reconstructed vision. The scripts and jobs used in this paper can be found in the public repository: https://github.com/MateiOpr/research-project

1 Introduction

Facial recognition (FR) systems are becoming widely deployed in surveillance, social media, and autonomous systems, raising concerns about unauthorized identity tracking done by bad actors and mass surveillance. Protecting facial identity from FR systems while preserving the natural appearance of the images has therefore become an active research problem, further motivated by legal frameworks, such as General Data Protection Regulation (GDPR)[1]. Recent unrestricted adversarial methods focus on learning perturbations within high-level facial attributes, using strategies such as cosmetic makeup transfer to improve black-box transferability against unauthorized FR models[2]

Privacy protection methods applied to RGB images are divided into multiple categories. Simple noise-based methods, such as TIP-IM [3], generate adversarial identity masks through iterative optimization repeatedly calculating and applying small pixel adjustments over multiple steps to maximize the system's error, managing to fool FR models. These have the downside of introducing pixel-level distortions which are visible to the human eye. More recent generative privacy protection approaches embed adversarial perturbations within human-recognizable facial features. Among these, Adv-CPG [2] is a customized portrait generation framework that injects adversarial signals through identity-preserving latent space optimization, while AMT-GAN [4] embeds adversarial perturbations within transferred cosmetic makeup using a Generative Adversarial Network (GAN). Both Adv-CPG and AMT-GAN achieve a high ASR (Attack Success Rate), while also preserving the visual naturalness of images, which is measured by SSIM (Structural Similarity Index). While these protection methods have been evaluated on standard RGB images, their effect on images produced by event-based cameras remains unexplored.

Event-based cameras represent visual input as asynchronous streams of brightness changes rather than frame-based images [5, 6]. While raw event data is often considered privacy-preserving due to its asynchronous and sparse nature, this assumption has been increasingly challenged by advancements in event-to-image deep learning reconstruction techniques[7]. Because networks like E2VID[8] can recover highly detailed intensity images from event streams, hardware-level obfuscation is bypassed, requiring the application of algorithmic adversarial protections[7, 8]. Futhermore, the reconstruction process introduces artifacts and information loss that may interact differently with protection methods. Therefore, whether protection methods designed for clean RGB images remain effective under the reconstruction pipeline, and whether the event reconstruction parameters affect the protection quality are two questions being investigated in this research.

This reserach study focuses primarily on AMT-GAN, an adversarial makeup-transfer framework that embeds perturbations targeting a fixed identity during training. To extend the analysis, this research also incorporates a qualitative comparison for Adv-CPG. Unlike AMT-GAN, Adv-CPG leverages a diffusion-based architecture that accepts arbitrary target identities at inference. Although, its substantial computational overhead constrains it to a secondary comparative role in this research

To study the impact of event-based reconstruction on privacy protection, CelebV-HQ[9] video clips are converted to event streams using v2e[10] and reconstructed into grayscale facial images using E2VID[8]. The reconstructed images are then processed by each protection method and evaluated using Attack Success Rate (ASR) and Structural Similarity Index Measure (SSIM). ASR is computed using AMT-GAN's official evaluation protocol against the four white-box Face Recognition models it was trained on: IR152, IRSE50, FaceNet, and MobileFace.[11]

For AMT-GAN, I run three experiments. The main evaluation applies AMT-GAN with nine reference makeup styles to event-reconstructed inputs and computes ASR against CelebA-HQ identity 047073. Additonally, I create a baseline from applying AMT-GAN directly to the original RGB inputs (skipping the event-reconstruction pipeline) to isolate the effect of reconstruction. Finally, an event-side ablation varies the v2e contrast threshold across five values to investigate how event-simulation density affects

downstream protection effectiveness.

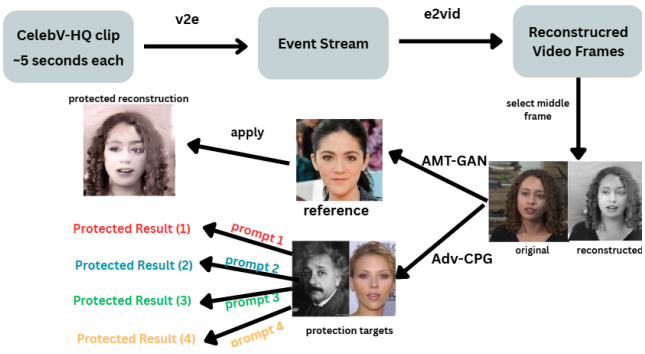
For Adv-CPG, I evaluate a 10-image qualitative subset across varying target identities and text prompts to observe its visual trade-offs on event-reconstructed images.

Research Question: How do event-based reconstruction parameters and the choice of generative adversarial protection method determine the privacy-naturalness trade-off on event-reconstructed face images?

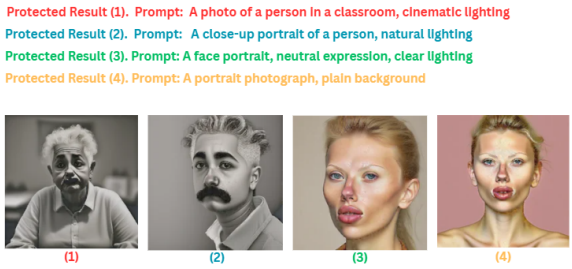
Research Questions:

- **RQ1:** Does AMT-GAN's adversarial signal transfer from clean RGB to event-reconstructed inputs?
- **RQ2:** How do event-simulation parameters (specifically the v2e contrast threshold) affect protection effectiveness?
- **RQ3:** Where does any observed failure originate? In the event reconstruction, the protection method, or the FR evaluator?
- **RQ4:** How does the domain shift of event reconstruction impact the visual naturalness and structural diversity of diffusion-based protections like Adv-CPG?

The remainder of this paper is structured as follows:
Section 2 covers the background and methodology,
Section 3 outlines the contributions of this work,
Section 4 presents the experimental setup and results,
Section 5 addresses the ethical aspects of the research,
Section 6 discusses the results,
Section 7 presents the conclusion and future work.



(a) Reconstruction and Protection Pipeline



(b) Adv-CPG Protection Scenarios

Figure 1: Overview of the research pipeline and protection scenarios. **(a)** CelebV-HQ video clips (5 seconds) are converted to asynchronous event streams using v2e, then reconstructed into grayscale frames using E2VID. The reconstructed baseline is passed through two generative protection approaches: AMT-GAN (applying a reference makeup style) and Adv-CPG (targeting a specific identity via text prompts). **(b)** Visual breakdown of the four distinct Adv-CPG prompt scenarios applied during the qualitative evaluation.

2 Background

Event cameras differ from RGB cameras because they do not produce frames at fixed time intervals. In event-based camera vision, each pixel independently emits an event whenever its log-intensity changes by a fixed threshold. Each event is represented as a tuple (t, x, y, p), where t is the timestamp, (x, y) is the pixel coordinate, and $p \in \{-1, +1\}$ is the polarity of the brightness change [7].

Recently, neural networks such as E2VID[12] have achieved reconstruction of grayscale frames from event streams. When applied to event streams of human faces, E2VID recovers enough structural and textural details that the identity of the subject typically remains recognizable to face recognition systems. This is a privacy risk that further motivates this study.

2.1 Adversarial Face Privacy Protection

Adversarial facial privacy protection methods generally operate under two distinct threat models: data poisoning attacks, which inject perturbations into public photos to disrupt the training phase of unauthorized FR systems [13], and evasion attacks, which modify a face image to cause an already trained FR system to mistake the identity at inference time while maintaining visual naturalness [3]. In this research I focus exclusively on evasion techniques.

Early evasion methods, such as LowKey [14], rely on l_p-bounded noise optimization, which frequently introduces noticeable, high-frequency artifacts that reduce the visual quality of the image. To overcome this limitation, recent unrestricted adversarial attacks embed perturbations into meaningful facial features. Approaches like Adv-Makeup [15], DiffAM [16], and CLIP2Protect [17] integrate natural-looking cosmetic additions, such as eyeshadow or lipstick. These cosmetic additions are guided by either reference images or text prompts, achieving high black-box transferability without compromising the visual naturalness of the image.

2.1.1 AMT-GAN

AMT-GAN [4] is a Generative Adversarial Network (GAN) based protection method introduced in 2022 that generates adversarial face images by transferring cosmetic makeup from a reference image onto the source image. The adversarial perturbation is embedded into the transferred makeup style and is optimized against a fixed target identity during training using four white-box FR models: IR152, IRSE50, FaceNet, and MobileFace. This multi-model training strategy over multiple surrogate white-box models is essential for preventing the adversarial signal from overfitting to a specific neural network, a technique also employed by generative frameworks like CLIP2Protect[17] and DiffAM[16] to guarantee black-box transferability against unknown commercial FR systems.

The adversarial perturbation is fixed toward a specific identity during the model's training phase. While the official AMT-GAN evaluation protocol averages results across four different fixed target identities (such as CelebA-HQ identities 085807 and 047073 found in their released codebase) [11], the model requires separate training for each. Figure 2 illustrates two of these fixed target identities (085807 and 047073) utilized in their evaluation codebase. On clean RGB images (CelebA-HQ), the standard AMT-GAN method achieves a mean ASR of 76.96% on IRSE50 and a mean SSIM of 0.7873 [4].

2.1.2 Adv-CPG

Adv-CPG [2] is a generative protection method which accepts an arbitrary target identity at inference time and builds on Stable Diffusion XL. It uses three modules: a Multi-Modal Image Customizer, an ID Encryptor that injects target-identity features into the diffusion process, and an Encryption Enhancer that adds gradient-based identity guidance at each denoising step. The result is a customised portrait that is visually similar to the original while embedding adversarial signals that misleads FR models. On clean, unreconstructed RGB benchmarks (CelebA-HQ), Adv-CPG achieves an average ASR of 79.65% while maintaining an average SSIM of 0.8978[2].

2.2 Evaluation Metrics

During this research, two metrics will be studied while finding out the trade-off: (1) The Attack Success Rate (ASR), which measures protection effectiveness as the percentage of protected images for which the FR model fails to match the correct identity, ranging from 0% (no protection) to 100% (fully protected). (2) The Structural Similarity Index (SSIM)[18], which measures visual naturalness as a perceptual similarity score between the protected image and its unprotected counterpart, with values ranging from 0 to 1, where 1 means identical images. In order for a protection method to be successful, the protection method should achieve high ASR without significantly lowering SSIM. Therefore, studying this tradeoff is important.

2.3 Methodology

A subset of 200 video clips from CelebV-HQ[9] is processed through v2e[10], an event camera simulator that converts video frames into asynchronous event streams. The resulting event streams are then passed through E2VID[12] to reconstruct grayscale frames. The middle frame of each reconstruction is selected for protection. The reconstructed images are visually grayscale (since E2VID outputs intensity only) but stored as three-channel RGB. These reconstructed images also form the input set common to both protection methods. Because event cameras only record relative brightness changes that exceed a specific contrast threshold, low-intesity, high-frequency adversarial perturbations can sometimes be discarded during the sensing phase[10, 19]. Furthermore, image reconstruction networks like E2VID rely on tracking data over time to build back the visual scene. This acts as a filter that smooths out delicate, pixel-level adversarial noise [8].

2.3.1 AMT-GAN evaluation

AMT-GAN is applied to each reconstructed RGB image with nine different reference makeup styles (ref01–ref10, skipping ref05). To align with the original paper's methodology, ASR is evaluated against CelebA-HQ identity 047073, which is the standard evaluation target in AMT-GAN's released codebase.[11] Two additional experiments are performed to localize the source of the protection failure: AMT-GAN is applied directly to the original RGB inputs (skipping the event-reconstruction pipeline) to isolate the effect of the reconstruction. Additionally, the v2e contrast threshold is varied across five values $\{0.1, 0.15, 0.2, 0.25, 0.3\}$ to investigate how event-simulation density affects downstream protection effectiveness. Lower thresholds produce richer event streams and higher-detail reconstructions, while higher thresholds produce sparser events and lower-detail reconstructions. For this ablation, a single reference makeup style (ref01) is used to isolate the effect of the threshold parameter.

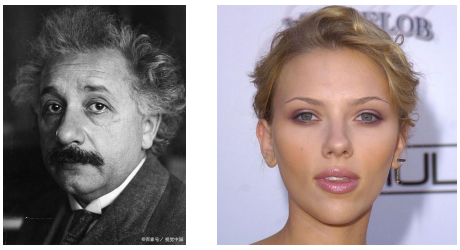
Identity 047073 (Evaluation Target)

Identity 085807 (Training Target)

Figure 2: CelebA-HQ target identities utilized in the AMT-GAN pipeline. Identity 047073 serves as the standard target for downstream evaluation metrics (ASR), while identity 085807 represents the fixed target optimization embedded within the pretrained model weights.

2.3.2 Adv-CPG evaluation

To perform a qualitative analysis of Adv-CPG on event-reconstructed images, a subset of 10 reconstructed source images was evaluated. To assess the impact of the target identity and scene prompt on the privacy-naturalness trade-off, my evaluation combines two target identities (Albert Einstein and Scarlett Johansson) across four distinct text-prompt scenarios (Figure 1): (1) a default classroom prompt ("A photo of a person in a classroom, cinematic lighting") with the Einstein target; (2) a close-up prompt ("A close-up portrait of a person, natural lighting") with the Einstein target; (3) a facial-focus prompt ("A face portrait, natural expression, clear lighting") with the Scarlett Johansson target; and (4) a minimal prompt ("A face portrait, natural expression, clear lighting") with the Scarlett Johansson target. All Adv-CPG generations were initialized with a fixed random seed of 42 to ensure consistency.



Albert Einstein target Scarlett Johansson target

Figure 3: The two distinct target identities utilized for the Adv-CPG qualitative evaluation across different prompt scenarios.

2.3.3 ASR computation

ASR is computed using AMT-GAN's official evaluation protocol against the four white-box Face Recognition models it was trained on: IR152, IRSE50, FaceNet, and MobileFace[4]. A protected image is counted as a successful attack against a given FR model if its cosine similarity to the target identity exceeds that model's specific FAR@0.01 verification threshold (0.167, 0.241, 0.409, and 0.302)[2].

2.3.4 Hardware

The pipeline was initially developed on an NVIDIA GTX 1660 Ti GPU running inside a WSL2 Ubuntu environment. Because Adv-CPG requires more VRAM than the GTX 1660 Ti has, the protection step was migrated to TU Delft's DelftBlue computer[20], where the jobs are ran on NVIDIA A100 GPU with 40 GB of memory. AMT-GAN's GAN is a lighter protection method that runs locally on the GTX 1660 Ti. The complete pipeline (event generation, reconstruction, protection, and evaluation) is implemented using Python with PyTorch.

3 Contributions

I am evaluating generative adversarial face-protection methods on event-reconstructed face images. Prior work on AMT-GAN and Adv-CPG [2] [4] has been evaluated on clean, high-resolution RGB datasets, such as CelebA-HQ.

Event cameras don't capture normal pictures. Their data has to be rebuilt into grayscale, lower-resolution frames. When a network like E2VID [8] does this rebuilding, it focuses on making the video look smooth from one frame to the next. The downside is that it sacrifices sharp spatial details. This means the final images look a bit smoothed out, missing the fine, sharp textures from a standard RGB camera.

I introduce a reconstruction-vs-original baseline comparison. AMT-GAN is run on both event-reconstructed inputs and the corresponding original RGB inputs, with all parameters kept constant. This isolates the effect of event-based reconstruction from the effect of the protection method and the face recognition evaluation. By comparing the performance on original versus reconstructed inputs, this baseline evaluates the event-simulation and image-reconstruction pipeline as input transformation. Just like ordinary image edits, blurring or compression reduce adversarial protection [14], the data loss from event recording and subsequent image reconstruction washes away the exact pixel patterns the protection relies on.

I introduce an event-side ablation that varies the v2e[10] reconstruction contrast threshold across five values to quantify how event-simulation density affects downstream protection effectiveness. This links the privacy-naturalness trade-off to the event-camera parameters that produced the input image.

I provide a reproducible end-to-end pipeline that converts CelebV-HQ face video clips into events (using v2e[10]), reconstructs grayscale frames (using E2VID), applies adversarial protection (using the official implementations of Adv-CPG and AMT-GAN), and evaluates the result using AMT-GAN's official white-box FR models, with ASR and SSIM.

This research analyzes four findings. While AMT-GAN affords partial protection on event-reconstructed inputs, the reconstruction pipeline affects the adversarial signal, lowering the effectiveness by approximately 56% relative to the original RGB baseline. Furthermore, the baseline reproduction on clean RGB inputs result in an ASR of 12.90%, which falls significantly short of the 76.96% reported in the original AMT-GAN literature. This difference suggests that unrestricted adversarial makeup is highly sensitive to distribution shifts, specifically the transition from high-resolution static portraits to upscaled, lower-quality video frames.

Additionally, I contribute a qualitative evaluation of

Adv-CPG across multiple text-prompt scenarios and target identities. This analysis reveals that diffusion-based protections struggle significantly with event-reconstructed inputs. I found that the domain shift affects the generative model, causing it to over-shift toward the target identity, create unnatural caricatures by exaggerating facial features, and trigger severe mode collapse that destroys the original structural diversity of the source images.

4 Setup and Results

4.1 Experimental Setup

Dataset. CelebV-HQ [9] is a large-scale face video dataset collected from YouTube, where each clip is cropped to the face region and labeled with attribute annotations. A target subset of 205 clips was downloaded using `yt-dlp`. However, 5 clips from the dataset (`6GO1ToF0Tnk_12_0.mp4`, `Agw7D-sR-c0_5.mp4`, `i65EoVftrM8_5.mp4`, `o9DsswQ0FhY_1.mp4`, and `pYMxUisY8Oc_0_0.mp4`) failed event simulation because they were unreadable due to missing `moov` atoms. This resulted in a set of 200 clips for the main evaluation. For each valid clip, the middle frame was extracted as the source RGB image. Event streams were then generated from this 200-clip set using v2e's `esim_torch` simulator [10].

Reconstruction. Each event stream is passed through E2VID[12] using the publicly released pretrained checkpoint. The middle reconstructed frame is selected and saved as a three-channel grayscale image for compatibility with the protection methods and FR models, because they only accept RGB input.

AMT-GAN. AMT-GAN is ran using the official implementation [11] with the publicly released pretrained generator checkpoint. Because some the reconstructed inputs are too small for AMT-GAN (e.g. 188×188), each source image is first upscaled to a maximum dimension of 512 pixels. AMT-GAN is applied with nine different reference makeup styles (ref01–ref10, skipping ref05).

Adv-CPG. Due to substantial computational overhead, the Adv-CPG evaluation is limited to a 10-image qualitative subset using the official implementation. The generation process requires significant VRAM, requiring the use of NVIDIA A100 GPUs (40 GB) scheduled on TU Delft's Delft Blue supercomputer. To analyze the privacy-naturalness trade-off, this subset is evaluated across two target identities (Albert Einstein and Scarlett Johansson) and four distinct text-prompt scenarios, with all generations initialized using a fixed random seed of 42 to ensure consistency.

Evaluation. ASR is computed using AMT-GAN's official evaluation script, testing against the four white-box FR models used in its training: IR152, IRSE50, FaceNet, and MobileFace. A protected image is counted as a successful attack if its cosine similarity to the target identity exceeds the model's FAR@0.01 threshold (0.167, 0.241, 0.409, and 0.302). To match the official methodology, the target identity is CelebA-HQ 047073. These evaluation parameters and thresholds are inherited from the official AMT-GAN implementation, allowing for a direct comparison with the original benchmarks. This proves that the lower ASR values represent a real drop in protection rather than a difference in measurement. SSIM is computed in grayscale at the source resolution between each protected image and its corresponding unprotected reconstructed source.

Hardware. Event generation and reconstruction run on an NVIDIA GTX 1660 Ti (6 GB VRAM) on WSL2 Ubuntu. Adv-CPG runs on NVIDIA A100 (40 GB VRAM) on TU Delft's DelftBlue supercomputer [20], scheduled via SLURM. AMT-GAN runs locally on the GTX 1660 Ti.

Face detection limitations. While running AMT-GAN on event-reconstructed inputs, even after upscaling to 512 pixels, the detector fails on some images due to the lower texture quality and smooth E2VID reconstructions. Affected source images are excluded from the AMT-GAN evaluation. Out of the 200 event-reconstructed sources, 139 had a detectable face after upscaling and were processed by AMT-GAN with nine reference makeup styles.

4.2 AMT-GAN main results

Table 1 reports the ASR at three False Acceptance Rate (FAR) thresholds across the four white-box models. At FAR@0.01 (the standard reporting threshold in the AMT-GAN literature), the FR models give ASR values between 4.72% (irse50) and 7.11% (ir152), with a mean of 5.64%. Furthermore, the mean SSIM for reconstructed images is 0.5743.

The protection is non-zero, indicating that some of the adversarial signal survives the event-reconstruction pipeline. However, this effectiveness is significantly below AMT-GAN's published 76.96% on clean images. The visual makeup transfer succeeds in all cases, but the adversarial perturbation is weaker.

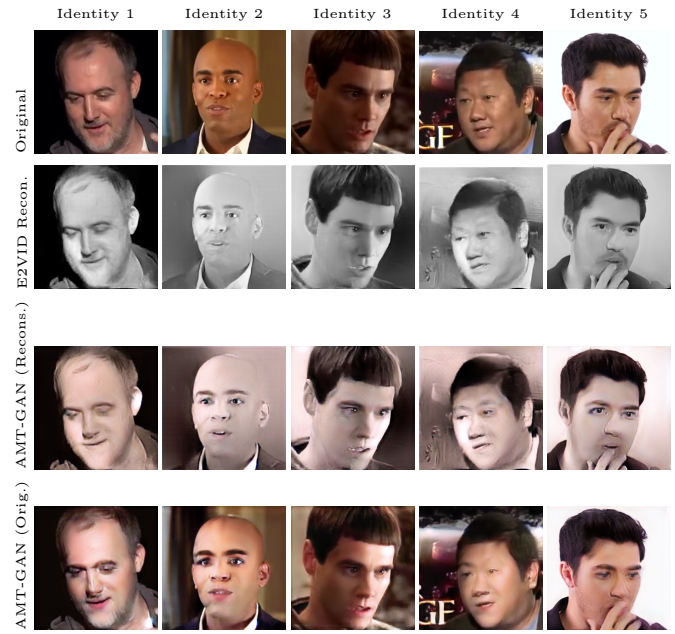


Figure 4: AMT-GAN qualitative output for five CelebV-HQ identities, under the two input conditions evaluated. Top two rows: original CelebV-HQ frames and their E2VID reconstructions. Bottom two rows: AMT-GAN protected outputs with reference makeup style ref01 applied to each input type.

Table 1: AMT-GAN protection results on 139 event-reconstructed CelebV-HQ faces, evaluated using the official AMT-GAN protocol.

FR Model	ASR @ FAR=0.1	ASR @ FAR=0.01	ASR @ FAR=0.001
ir152	27.82%	7.11%	1.92%
irse50	17.91%	4.72%	0.64%
facenet	26.70%	5.04%	0.00%
mobile_face	29.58%	5.68%	1.04%
Mean	**25.50%**	**5.64%**	**0.90%**

4.3 Baseline: AMT-GAN on original RGB

To isolate the source of the protection failure, AMT-GAN is applied directly to the original RGB inputs from CelebV-HQ, skipping the event reconstruction pipeline. Out of the 200 sources, 146 had detectable faces and were processed.

Table 2 reports the results. On the original RGB inputs, the mean ASR at FAR@0.01 rises to 12.90%. Comparing this to the 5.64% achieved on reconstructed inputs confirms that the event-reconstruction pipeline degrades protection effectiveness by approximately 56%.

The mean SSIM for original RGB images drops to 0.4681. This suggests that AMT-GAN's makeup style changes are applied more heavily or are more visible against the original RGB baseline than against the lower-detail reconstructions. Even on this clean baseline, the protection falls shorter than published 76.96%, indicating a decrease caused by the dataset distribution (video frames vs. static portraits) and upscaling artifacts.

Table 2: AMT-GAN baseline on 146 original RGB CelebV-HQ sources (no event reconstruction applied).

FR Model	ASR @ FAR=0.1	ASR @ FAR=0.01	ASR @ FAR=0.001
ir152	36.38%	12.79%	4.49%
irse50	34.09%	10.81%	3.58%
facenet	40.11%	12.56%	0.61%
mobile_face	41.55%	15.45%	2.97%
Mean	**38.03%**	**12.90%**	**2.91%**

4.4 Event-side ablation: contrast threshold

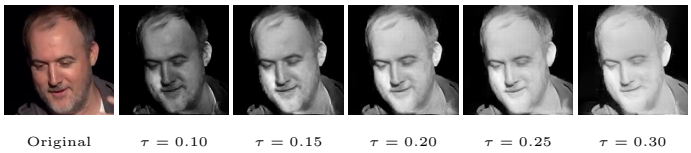


Figure 5: Effect of the v2e contrast threshold τ on E2VID reconstruction quality on a single CelebV-HQ source. Lower thresholds produce denser events and richer reconstructions. Higher thresholds produce sparser events and smoother reconstructions.

To examine whether event-simulation parameters affect protection effectiveness, the v2e contrast threshold is varied across five values: {0.10, 0.15, 0.20, 0.25, 0.30}. Lower thresholds produce denser event streams and richer reconstructions. Higher thresholds produce sparser events and lower-quality reconstructions. For each threshold, the full pipeline is run across the complete 200-clip dataset using reference makeup style ref01 to thoroughly isolate the effect of the threshold parameter.

Table 3: AMT-GAN performance as a function of v2e contrast threshold evaluated on the full 200-clip dataset (ref01 only). ASR is reported at FAR=0.01.

Threshold	n	Mean SSIM	ir152	irse50	facenet	mobile_face	**Mean ASR**
0.10	135	0.499	8.89%	5.93%	5.93%	5.93%	**6.67%**
0.15	138	0.547	7.97%	5.07%	6.52%	5.80%	**6.34%**
0.20	139	0.573	7.19%	3.60%	4.32%	6.47%	**5.40%**
0.25	130	0.578	6.15%	3.85%	7.69%	3.85%	**5.39%**
0.30	131	0.592	6.11%	4.58%	4.58%	3.05%	**4.58%**

SSIM rises with contrast threshold. Visual similarity increases from 0.499 at threshold 0.10 to 0.592 at threshold 0.30. Because higher thresholds produce lower-detail, structurally smoother reconstructions, the AMT-GAN protected outputs deviate less from these low-information inputs, resulting in artificially higher SSIM scores.

Face detection degradation at sparse thresholds: Across the 200-clip input set, successful face detections

varied depending on the threshold. Sparser event streams generally resulted in fewer successful detections ($n = 131$ at $\tau = 0.30$ versus $n = 139$ at $\tau = 0.20$), directly reflecting the loss of facial geometry in low-density event streams.

ASR declines with sparsity: Mean ASR drops slightly as the threshold increases, moving from 6.67% at $\tau = 0.10$ down to 4.58% at $\tau = 0.30$. This indicates that richer reconstructions (lower thresholds) preserve more of the adversarial signal, though the overall protection remains low across all of the conditions.

This ablation isolates the temporal event density, revealing that the degradation of the adversarial signal is not only a result of poor visual reconstruction. Instead, the signal loss persists regardless of how rich or dense the event stream is, suggesting a fundamental incompatibility between event-reconstructed face images and existing adversarial masking techniques.

The fundamental reason behind this protection effectiveness reduction lies in the neuromorphic sensing mechanism itself. Because event cameras only trigger events when log-contrast crosses a fixed threshold, they operate as a strict, non-linear sampler. That kind of discrete sampling throws away exactly what gradient-based adversarial attacks need to work: the smooth, low-amplitude pixel gradients and fine textures that the perturbations are built on, regardless of the overall event density[19].

By looking closely at event density in this ablation test, it can be seen that the loss of protection goes deeper than just low-quality image reconstruction. The protection degrades even when the event stream is packed with data and the resulting images look highly detailed. This highlights an incompatibility between how event cameras capture the world and how current adversarial masks are designed.

4.5 Adv-CPG qualitative evaluation

A qualitative analysis of Adv-CPG on 10 event-reconstructed images (targeting Albert Einstein and Scarlett Johansson) reveals a serious identity over-shift. While this successfully makes the original source unrecognizable, artificially lowering the ASR, the outputs lose visual resemblance to the source, sacrificing naturalness.

In terms of image quality, the protected outputs show obvious unnatural artifacts because Adv-CPG overly emphasizes and artificially exaggerates the most obvious facial features of the target identity. For example, protections targeting Einstein generated overly emphasized, unnatural mustaches, while protections targeting Scarlett Johansson produced unnaturally exaggerated lips. Furthermore, the choice of text prompt and target governs the variance of the generated outputs. For the Einstein target, the default prompt results in a strong over-shift with differing face shapes and expressions, while the close-up prompt introduces high variance among the 10 images, resulting in distinct smiles, eyebrows, and hairstyles. Both prompts targeting Scarlett Johansson triggered a serious mode collapse, where all 10 protected outputs appear almost identical in pose, look, and expression, completely disregarding the diversity of the source images.

While Adv-CPG is designed to guarantee high visual similarity with the original face on clean RGB images[2], this research's findings indicate that it over-shifts towards the target when applied to event-reconstructed inputs. This suggests the smooth, lower-detail textures from the reconstruction pipeline destabilizes Adv-CPG's ID Encryptor and Multi-Modal Image Customizer (MMIC) modules, which are highly tuned to the high-frequency feature distributions of standard, clean RGB datasets [2], causing the protected image to lose resemblance to the original face.

Figure 6: Adv-CPG results across different text prompts.

4.6 Cross-condition summary

Table 4 compares the main result against the original-RGB baseline and the literature claims. Both methods show an important decrease when moving from clean RGB to event-reconstructed inputs. Adv-CPG tends to over-shift toward the target, while AMT-GAN preserves the source identity but suffers a massive drop in ASR from the published 76.96% down to 12.90% (original RGB) and 5.64% (event-reconstructed).

Table 4: Summary of AMT-GAN performance under different conditions (Mean ASR @ FAR=0.01). Literature values are from [4].

Condition	Mean SSIM	Mean ASR
Event-reconstructed (Main result, 9 refs)	0.574	5.64%
Original RGB (Baseline, 9 refs)	0.468	12.90%
Clean RGB on IRSE50 (Literature) [4]	0.787	76.96%

5 Responsible Research

5.1 Reproducibility

The complete pipeline is built using open-source tools: Vid2e [10] for event simulation, E2VID [8] for image reconstruction, and AMT-GAN [4] and Adv-CPG [2] for adversarial protection. Evaluation is conducted using AMT-GAN's official evaluation protocol and its four white-box FR models. All model weights used in this research are publicly available from the original authors' repositories.
The CelebV-HQ metadata [9] that determines which YouTube clips are sampled is also public, although clips can become unavailable over time as YouTube videos are removed (23 of the attempted downloads failed for this reason). Additionally, the pipeline handles corruption at the ingestion level: 5 downloaded files with missing video container `moov` atoms were caught and filtered out automatically by the simulation script, ensuring no corrupted streams disrupted the downstream E2VID or AMT-GAN evaluation.

To ensure the experiment can be replicated, it is worth mentioning that AMT-GAN's dlib-based face detector is sensitive to the upscaling size of the reconstructed images. I tested several dimensions and found that 512×512 pixels was the appropriate size. Anything smaller caused the detector to miss too many faces, while anything larger simply slowed down the processing time without offering any benefit. Furthermore, all Adv-CPG generations were initialized using a fixed random seed of 42.

5.2 Dataset ethics

CelebV-HQ is a dataset of YouTube clips of public figures, with each clip pre-cropped to the face region and labelled with attribute annotations. The dataset is distributed as URLs and metadata, not image files. This requires the download of the YouTube videos. This respects YouTube's terms of service and allows content owners to remove material from the dataset by removing their videos. However, several issues arise:

CelebV-HQ can contain identifiable individuals who did not consent to having their face used as training or evaluation data for face recognition research. The dataset's choice of public figures partially mitigates this. I use CelebV-HQ for evaluation only. No models were trained on these identities, and the analysis is statistical rather than focused on any individual. Furthermore, the usage of these face images highlights the tradeoff between public computer vision datasets and legal frameworks like the GDPR, which requires explicit consent for biometric data used for identification.

Additionally, the target identity used in this work is also a real person. CelebA-HQ identity 047073 is a young woman whose photograph was released by the CelebA-HQ dataset and adopted as the standard evaluation target in AMT-GAN's official codebase. I use 047073 only because it is required to replicate AMT-GAN's published evaluation conditions.

5.3 Dual-use risk

Face privacy protection methods such as AMT-GAN and Adv-CPG are dual-use by design. The same techniques that protect a private individual's photograph from unauthorized face recognition can also be used to evade legitimate identity verification, e.g, by helping a person avoid recognition by law enforcement.

This work does not introduce a new protection method, and the evaluated methods (AMT-GAN, Adv-CPG) are already publicly available from their authors. My finding that AMT-GAN's protection degrades on event-reconstructed inputs is more useful to defenders (who can understand the limitations of privacy-preserving sensors and update their FR systems) than to attackers (who already know the limitations of older protection methods). Additionally, the event-side ablation informs system designers about reconstruction effects on protected face images rather than enabling new attacks.

5.4 Limitations

Dataset size. The main evaluation uses 200 CelebV-HQ clips, resulting in 139 successfully reconstructed faces and 146 original RGB faces. While this sample can be used for identifying performance degradation, it is still smaller than the thousands of images typically used in face-recognition benchmarks.

Unequal evaluation across methods. AMT-GAN is evaluated comprehensively on the full dataset, while the Adv-CPG evaluation is limited to a 10-image qualitative subset due to its heavy computational constraints. Therefore, the two methods cannot be fairly compared at scale based on this paper alone.

Reconstruction model. Only E2VID is used for reconstruction. Other event-to-image reconstruction methods may produce different reconstructions, which could affect the ASR and SSIM results.

Single image per clip. Only the middle frame of each reconstructed clip is used as the source. Choosing different temporal samples may result in different protection outcomes, especially for clips where the subject moves significantly.

5.5 AI Usage

In accordance with the course guidelines on academic integrity, I want to clearly outline how Large Language Models (LLMs) were utilized in this research. I employed LLMs exclusively as a software development aid to troubleshoot, optimize, and write the Python and Bash scripts necessary for automating the v2e event-camera simulations, the E2VID reconstructions, and the ASR and SSIM evaluation pipelines.

The AI was used strictly for code generation and debugging. It did not contribute to the literature review, the formulation of the methodology, the qualitative analysis of the images, or the writing of this manuscript. The theoretical framing and the interpretation of the privacy-naturalness trade-offs presented in this paper are entirely my own original work. For full transparency, the exact prompts used to assist with the coding process are documented in Appendix A.

6 Discussion

6.1 The Discrepancy with Published Literature

Even on the original, unreconstructed RGB baseline, AMT-GAN achieved a mean ASR of 12.90% across the four white-box models at FAR@0.01. This is far below the 76.96% reported in the original paper for the IRSE50 model. Because this evaluation mirrors the authors' exact codebase, target identity, and verified FAR thresholds, this massive drop cannot be attributed to an evaluator or threshold mismatch. This gap likely results from the input data. CelebV-HQ consists of video frames extracted from YouTube clips, which naturally have motion blur, varied lighting, and compression artifacts compared to the high-resolution static portraits of CelebA-HQ used in AMT-GAN's training.

6.2 The Impact of Event-Reconstruction on Adversarial Signals

The primary finding is that the event-reconstruction pipeline reduces protection effectiveness by approximately 56%, dropping mean ASR from 12.90% to 5.64%. The adversarial signal is partially preserved through the simulation and reconstruction pipeline, but it is significantly reduced. Additionally, the contrast-threshold ablation (Section 4.4) shows that varying event-simulation density does not substantially improve the protection effectiveness. The mean ASR remains bounded to low values (between 4.58% and 6.67%) regardless of how dense or sparse the input event stream is. This rules out the hypothesis that the failure is caused simply by sparse events losing the perturbation. Instead, the conversion from RGB to asynchronous brightness changes and back reduces adversarial noise.

6.3 Why does SSIM rise with contrast threshold?

A result from the contrast-threshold ablation is that SSIM increases with threshold: $0.499 \rightarrow 0.547 \rightarrow 0.573 \rightarrow 0.578 \rightarrow 0.592$. This might be counter-intuitive, because it would be expected that higher-quality reconstructions produce more natural-looking protected outputs.

The explanation is that SSIM measures similarity between the protected image and the reconstructed source, not the original. Higher contrast thresholds produce smoother reconstructions. AMT-GAN's makeup color changes are less visible against low-detail inputs, resulting a higher SSIM. The protected output is not actually more natural, it just differs less from an input that contains less information. Therefore, when comparing protection methods on lower-quality inputs, SSIM against the lower-quality source can mislead.

6.4 Implications for event-camera privacy

A common assumption in event-camera privacy literature is that event streams are privacy-preserving by design because they discard absolute brightness. Recent reconstruction methods, including E2VID [8], challenge this by showing that identity is recoverable. This introduces another variable to consider: applying an existing RGB-domain adversarial face-protection method to the reconstructed image does not provide the appropriate privacy, resulting in less than 6% ASR. However, the fact that the protection is non-zero proves that adversarial signals can partially survive the reconstruction pipeline. Because parts of the adversarial noise persist through the reconstruction loop, future protection methods can no longer ignore them. Closing this gap requires a different approach. Researchers will most likely have to create protection methods specifically designed for event data, or optimize standard RGB protection methods to withstand the heavy smoothing that happens during the reconstruction process.

7 Conclusions and Future Work

In this work, I investigated the privacy-naturalness trade-off of adversarial face-protection methods applied to images reconstructed from event-based cameras. By evaluating AMT-GAN across 200 event-reconstructed video clips and comparing the results to both an original RGB baseline and published literature, several important conclusions can be drawn.

The assumption that the event-camera bottleneck entirely reduces adversarial perturbations is incorrect. AMT-GAN provides partial protection on event-reconstructed inputs, resulting in a mean Attack Success Rate (ASR) of 5.64% at FAR=0.01 across four white-box Face Recognition (FR) models. This proves that

adversarial signals can partially survive the reconstruction pipeline.

The event-reconstruction process substantially reduces the protection. Compared to the original RGB baseline (which achieved 12.90% ASR), the reconstruction pipeline degrades protection effectiveness by roughly 56%. An ablation of the contrast threshold demonstrated that this degradation is related to the event-to-pixel reconstruction process, rather than simply a lack of event data.

The event pipeline significantly reduces the protection. While the original RGB baseline manages a 12.90% ASR, running the data through the reconstruction loop reduces that effectiveness by roughly 56%. The contrast threshold ablation proves this drop isn't just a side effect of having sparser event streams. Instead, the drop in protection results from the way event data is converted back into pixels.

In conclusion, the reduced effect of protection applied on the event-reconstructed inputs, which is just a 5.64% ASR (in this research) compared to the 76.96% reported in the AMT-GAN literature shows the strict limitations of these adversarial facial protection methods. While dataset differences motivate for the baseline drop to 12.90%, the E2VID reconstruction process further reduces the protection. This makes clear a weakness in the tested facial-privacy methods: they are not yet effective under the image reconstruction pipeline

Future work should expand the evaluation of generative protection methods to determine if different adversarial embedding techniques are more effective to event-camera reconstruction. Finally, providing reliable privacy for event-based vision may require designing adversarial protections specifically built for the event domain, rather than reusing RGB-based methods.

Appendix A: Generative AI Prompts

In alignment with the academic integrity guidelines for this course, this appendix details the generative AI assistance used along the project. AI tools were used for coding support, debugging specific errors, and writing LaTeX, and not for generating core scientific claims or analysis. Below is a description of some of the queries used during the research process.

Dataset & Simulation Scripts

- "I need to download a subset of 205 specific CelebV-HQ clips. Write a Python script using `yt-dlp` that downloads these videos and automatically extracts the middle frame to serve as my baseline."

- "I'm running the v2e `esim_torch` simulator and keep getting a 'missing `moov` atom' error on mp4 files. Help me write a try-catch block to catch this corrupted file error and safely skip to the next video?"

- "I need to iterate over 200 event stream files, feed them into the E2VID model, and save the reconstructed outputs as 3-channel grayscale images. Help me write a bash script for it."

Generative Protections & DelftBlue SLURM jobs

- "My event-reconstructed frames are too small (some are 188×188) and the dlib face detector in AMT-GAN keeps failing. How can I write a batch processing script in Python to upscale these to 512×512?"

- "I need to run ADV-CPG. My local GTX 1660 Ti doesn't have enough VRAM for running Adv-CPG. Help me write a SLURM job script for TU Delft's DelftBlue supercomputer. I need to run it on a A100 GPU (40GB) with a fixed random seed of 42."

Metrics & LaTeX

- Help me calculate the Attack Success Rate (ASR) in Python for AMT-GAN under 4 different FR models, given that a successful attack means the score must be higher than the FAR@0.01 thresholds (0.167, 0.241, 0.409, and 0.302) for identity 047073?"

- "How do I format a 4×5 image grid in LaTeX. The columns have be the Adv-CPG text prompts, and the rows should be my source images with labels rotated 90 degrees."

- "My LaTeX figure is overlapping into the next section because it's too tall. How do I force it to stay in Section 5.5? I need to make the row heights smaller?"

References

[1] Peter I Gasiokwu, Ufuoma Garvin Oyibodoro, and Michael O Ifeanyi Nwabuoku. GDPR safeguards for Facial Recognition Technology: A critical analysis. volume 06, pages 407–423, 1 2025.

[2] Junying Wang, Hongyuan Zhang, and Yuan Yuan. Adv-Cpg: A Customized Portrait Generation Framework with Facial Adversarial Attacks. pages 21001–21010, 6 2025.

[3] Xiao Yang, Yinpeng Dong, Tianyu Pang, Hang Su, Jun Zhu, Yuefeng Chen, and Hui Xue. Towards face encryption by generating adversarial identity masks. pages 3877–3887, 10 2021.

[4] Shengshan Hu, Xiaogeng Liu, Yechao Zhang, Minghui Li, Leo Yu Zhang, Hai Jin, and Libing Wu. Protecting facial privacy: generating adversarial identity masks via style-robust makeup transfer. *2022 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 14994–15003, 6 2022.

[5] Shafiq Ahmad, Gianluca Scarpellini, Pietro Morerio, and Alessio Del Bue. Event-driven Re-Id: A New Benchmark and Method Towards Privacy-Preserving Person Re-Identification. pages 459–468, 1 2022.

[6] Junho Kim, Young Min Kim, Ramzi Zahreddine, Weston A. Welge, Gurunandan Krishnan, Sizhuo Ma, and Jian Wang. Privacy-Preserving visual localization with event cameras. volume 34, pages 6215–6230, 1 2025.

[7] Mira Adra, Simone Melcarne, Nelida Mirabet-Herranz, and Jean-Luc Dugelay. Event-based solutions for human-centered applications: a comprehensive review. *Frontiers in Signal Processing*, 5:1585242, 2025.

[8] Henri Rebecq, Rene Ranftl, Vladlen Koltun, and Davide Scaramuzza. Events-To-Video: Bringing Modern Computer Vision to Event Cameras. pages 3852–3861, 6 2019.

[9] Hao Zhu, Wayne Wu, Wentao Zhu, Liming Jiang, Siwei Tang, Li Zhang, Ziwei Liu, and Chen Change Loy. Celebv-hq: A large-scale video facial attributes dataset. `https://celebv-hq.github.io/`, 2022.

[10] Yuhuang Hu, Shih-Chii Liu, and Tobi Delbruck. v2e: From Video Frames to Realistic DVS Events. pages 1312–1321, 6 2021.

[11] Shengshan Hu, Xiaogeng Liu, Yechao Zhang, Minghui Li, Leo Yu Zhang, Hai Jin, and Libing Wu. Amt-gan: Protecting facial privacy via style-robust makeup transfer (source code). `https://github.com/CGCL-codes/AMT-GAN`, 2022.

[12] Henri Rebecq and Davide Scaramuzza. E2vid: Event-to-video reconstruction (source code). `https://github.com/uzh-rpg/rpg_e2vid`, 2019.

[13] Shawn Shan, Emily Wenger, Jiayun Zhang, Huiying Li, Hai-Tao Zheng, and Ben Y. Zhao. Fawkes: Protecting Personal Privacy against Unauthorized Deep Learning Models. 2 2020.

[14] Valeriia Cherepanova, Micah Goldblum, Harrison Foley, Shiyuan Duan, John P. Dickerson, Gavin Taylor, and Tom Goldstein. LowKey: Leveraging Adversarial Attacks to Protect Social Media Users from Facial Recognition. 5 2021.

[15] Bangjie Yin, Wenxuan Wang, Taiping Yao, Junfeng Guo, Zelun Kong, Shouhong Ding, Jilin Li, and Cong Liu. Adv-Makeup: A New Imperceptible and Transferable Attack on Face Recognition. pages 1252–1258, 8 2021.

[16] Yuhao Sun, Lingyun Yu, Hongtao Xie, Jiaming Li, and Yongdong Zhang. DiffAM: Diffusion-Based Adversarial Makeup Transfer for Facial Privacy Protection. pages 24584–24594, 6 2024.

[17] Fahad Shamshad, Muzammal Naseer, and Karthik Nandakumar. CLIP2Protect: Protecting Facial Privacy Using Text-Guided Makeup via Adversarial Latent Search. pages 20595–20605, 6 2023.

[18] Zhou Wang, A.C. Bovik, H.R. Sheikh, and E.P. Simoncelli. Image quality assessment: from error visibility to structural similarity. volume 13, pages 600–612, 4 2004.

[19] Guillermo Gallego, Tobi Delbruck, Garrick Orchard, Chiara Bartolozzi, Brian Taba, Andrea Censi, Stefan Leutenegger, Andrew J. Davison, Jorg Conradt, Kostas Daniilidis, and Davide Scaramuzza. Event-Based Vision: a survey. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 44(1):154–180, 7 2020.

[20] Delft High Performance Computing Centre (DHPC). DelftBlue: TU Delft Supercomputer. `https://www.tudelft.nl/dhpc/`, 2024.