Effect of Privacy Preservation Strategies on Event-to-Image Reconstruction

A Comparative Study of Raw-Event Perturbation Strategies

Arda Tamgaç[1]
Supervisor(s): Nergis Tömen[1]**, Tunahan Parlayıcı**[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: Arda Tamgaç
Final project course: CSE3000 Research Project
Thesis committee: Nergis Tömen, Tunahan Parlayıcı, Ricardo Marroquim

Abstract

Event cameras are increasingly deployed in privacy-sensitive applications such as surveillance, autonomous vehicles, and human-computer interaction. Unlike conventional cameras, they record only per-pixel brightness changes as a sparse, asynchronous stream of events, making them efficient and potentially privacy-preserving. However, reconstruction models such as E2VID can recover recognizable facial images from event streams, undermining this assumption. This paper investigates whether simple perturbations applied directly to raw event streams can reduce face identifiability while preserving reconstruction quality. Three perturbation methods are compared: polarity flipping, spatial jitter, event insertion and deletion. Each method is evaluated across multiple strength levels on 300 face video clips from the CelebV-HQ dataset, converted to synthetic events using v2e. Reconstruction quality is measured using PSNR, SSIM, and LPIPS, and face identifiability is measured using FaceNet re-identification. The implementation is available at https://github.com/ardatamgac/event-to-image-privacy.

1 Introduction

The widespread deployment of surveillance systems has made privacy preservation an increasingly important concern in computer vision research. As cameras become cheaper and more pervasive, the volume of visual data collected in public spaces has grown substantially, raising both legal and ethical questions about the right to privacy [3].

Event cameras have emerged as a promising alternative to traditional RGB cameras in this context. They are bio-inspired sensors that, rather than capturing full intensity frames, asynchronously detect per-pixel changes in brightness and represent them as a sparse stream of events encoding the time, location, and polarity of each change [11; 6]. Because they discard absolute intensity and record only sparse changes, event streams have often been assumed to be inherently privacy-preserving, and this assumption has motivated their use in privacy-sensitive settings.

This assumption has, however, been challenged. Reconstruction networks such as E2VID, [12] can recover high-quality intensity video directly from event streams [1], showing that the discarded appearance information can be inferred rather than lost. Ahmad et al. [2] made this concrete for identity, introducing the first event-based person re-identification benchmark and showing that reconstruction can recover identifying detail from event data. Event streams are therefore not private by default.

A growing body of work has responded by trying to make event data private by design. These approaches act at different stages of the pipeline. At the sensor and network level, Kim et al. [9] protect event-based visual localization while preserving task performance. At the representation level, AnonyNoise [3] learns a noise pattern on event histograms that reduces re-identification while retaining utility. Other methods operate on the raw stream or the reconstruction model itself. These defences are effective but typically require training or are tied to a specific downstream model, and each is studied in isolation, usually not on face data.

This leaves a gap: no prior work systematically compares simple, training-free raw-event perturbation strategies head-to-head on face data using consistent reconstruction-quality and identifiability metrics. This project addresses that gap by applying simple perturbations directly to raw event streams and measuring their effect on both reconstruction quality and identifiability. The main research question is: What is the effect of applying simple privacy-preserving perturbations to event camera data in event-to-image reconstruction pipelines, in terms of reconstruction quality and identifiability? The study is divided into four sub-questions:

1. How does polarity flipping affect reconstruction quality and face identifiability across varying flip probabilities?
2. How does spatial jitter affect reconstruction quality and face identifiability across varying noise strengths?
3. How does event insertion and deletion affect reconstruction quality and face identifiability across varying rates?
4. Which perturbation strategy provides the best privacy-utility tradeoff across the tested strength levels?

The hypothesis is that lightweight alterations of the event stream can reduce the identifiability of reconstructed images while preserving enough structure for downstream tasks.

The rest of this paper is organized as follows. Section 2 describes the methodology, detailing the CelebV-HQ dataset, synthetic event generation, the perturbation methods, the E2VID reconstruction pipeline, and the face evaluation metrics. Section 3 outlines the experimental setup, including the specific strength parameters used across the video clips. Section 4 presents the results. Section 5 discusses the privacy-utility tradeoffs and the limitations of the evaluation. Section 6 reflects on the responsible research aspects of using facial video data, and Section 7 concludes with a summary of the findings and possible future work.

2 Methodology

2.1 Dataset and Event Generation

For this study the CelebV-HQ [18] dataset was chosen. CelebV-HQ is a large-scale RGB video dataset which consists of 35,666 clips involving 15,653 identities, with a resolution of at least 512x512 pixels. The clip durations range from 3 to 20 seconds. It was found suitable for this study as it provides high-quality face videos making it suitable for both event simulation and face re-identification evaluation.

RGB video clips are converted into synthetic event streams using v2e [7]. To approximate the high temporal resolution required for event generation, v2e temporally upsamples the input video using the pretrained Super-SloMo frame interpolation network [8], before applying a DVS event camera model to generate per-pixel brightness change events. We chose to generate synthetic events as there are few public event camera datasets available, and access to ground-truth frames is

Figure 1: Top row shows the ground truth frames and the clean reconstruction. Remaining rows show increasing corruption strength from left to right, set by p for polarity flip, σ for spatial jitter and ρ for event insertion & deletion.

needed to compute baseline reconstruction quality metrics. The full pipeline, from RGB video through event generation, perturbation, and reconstruction to evaluation, is shown in Figure 2.

Each output event is represented as a tuple (t, x, y, p), where t is the timestamp in seconds, x and y are the pixel coordinates, and $p \in \{0, 1\}$ is the polarity indicating a brightness increase or decrease.

Input videos are resized to 346×260 pixels during simulation to match the resolution of the DAVIS346 event camera. The DAVIS architecture combines an asynchronous event stream with a conventional frame-based readout, providing both brightness-change events and intensity frames [4].

2.2 Event-Stream Perturbation Methods

Three event-stream perturbation methods were selected, each targeting a different property of an event: polarity flipping, spatial jitter, event insertion and deletion. Each was applied directly to the raw event stream at varying strengths to assess its effect on reconstruction quality and face identifiability.

Polarity Flipping

Polarity flipping randomly inverts the polarity of events in the event stream. Each event's polarity is flipped with probability p, where $p \in [0, 1]$ controls the strength of the perturbation. A higher value of p results in more events being flipped, resulting in a greater disruption to the event stream. At $p = 0$ no events are modified, and at $p = 0.5$ half of all events are flipped on average.

Spatial Jitter

Spatial jitter randomly displaces the pixel coordinates of each event by a small Gaussian offset. For each event, independent noise is sampled from $\mathcal{N}(0, \sigma^2)$ and added to both the x and y coordinates, where σ controls the strength of the perturbation. Coordinates are clamped to the sensor boundaries after displacement to ensure all events remain within the valid pixel range of $[0, 345] \times [0, 259]$. A higher value of σ results in larger displacements, introducing greater spatial disruption to the event stream.

Event Insertion & Deletion

Event insertion and deletion modifies the event stream by randomly adding synthetic events and removing real events. For insertion, fake events are generated at random pixel coordinates (x, y) drawn uniformly across the sensor and with random polarity, with timestamps t drawn from the existing real events, and appended to the stream. For deletion, each real event is independently removed with probability ρ. The number of inserted events is set to match the expected number deleted, so a single rate parameter $\rho \in [0, 1]$ controls both operations: a higher ρ deletes more real events and inserts a correspondingly larger number of fake ones, while the total event count stays roughly constant.

2.3 Reconstruction

Event streams are reconstructed into grayscale images using E2VID, a pretrained recurrent U-Net [13] model. No retraining or fine-tuning is performed, the publicly available `E2VID_lightweight.pth` checkpoint is used directly. Each event stream is processed using fixed-duration temporal windows of $\lfloor 1000/\text{fps} \rfloor$ ms, matching the frame rate of the source video, so that one reconstruction is produced per source frame. E2VID outputs a sequence of grayscale 346×260 images alongside a `timestamps.txt` file recording the centre timestamp of each output window.

Both the unperturbed and perturbed event streams are processed through the same E2VID pipeline, ensuring that any

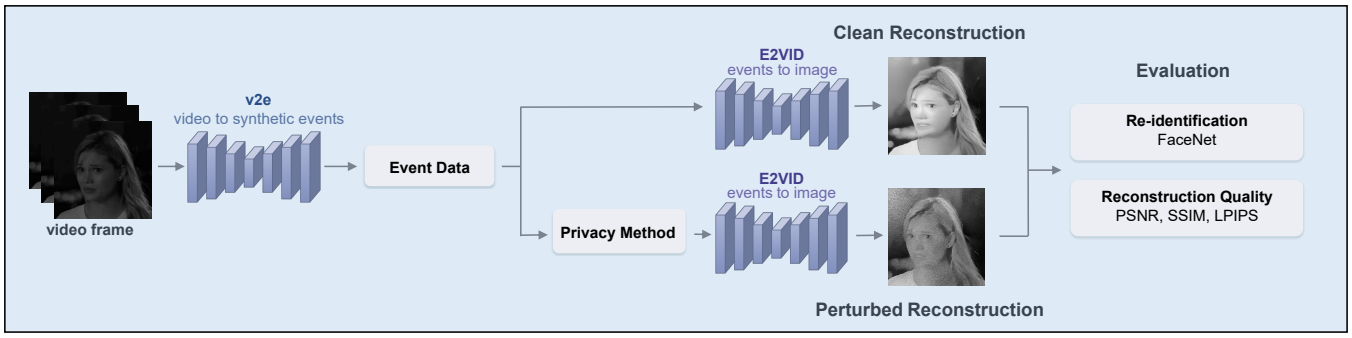

Figure 2: The evaluation pipeline: CelebV-HQ clips are converted to events with v2e, perturbed at the raw-event level, reconstructed with E2VID, and evaluated for reconstruction quality and face identifiability.

differences in reconstruction quality are attributable solely to the applied perturbation. Figure 1 shows representative reconstructions across perturbation types and strengths.

2.4 Evaluation

For each reconstructed frame, the source video (already resized to 346×260) is queried at the exact timestamp reported by E2VID, and the corresponding frame is extracted and converted to grayscale. This ensures that every reconstructed frame has a temporally matched ground-truth counterpart for metric computation.

The unperturbed event stream produces a clean baseline reconstruction against which perturbed reconstructions are compared. Evaluation metrics are computed frame-by-frame between each reconstructed image and its matched ground-truth frame.

Reconstruction Quality

We measure reconstruction quality using three complementary metrics: Peak Signal-to-Noise Ratio (PSNR), Structural Similarity Index (SSIM), Learned Perceptual Image Patch Similarity (LPIPS). All three are computed between each reconstructed frame $\hat{I}$ and its temporally matched ground-truth frame I, both of size $M \times N$ with pixel values in $[0, 255]$.

PSNR measures the ratio between the maximum possible pixel value and the mean squared error between the reconstructed and ground-truth frames,

$$\mathrm{MSE} = \frac{1}{MN} \sum_{i=1}^{M} \sum_{j=1}^{N} \big(I(i,j) - \hat{I}(i,j)\big)^2, \tag{1}$$

$$\mathrm{PSNR} = 10 \cdot \log_{10}\left(\frac{L^2}{\mathrm{MSE}}\right), \tag{2}$$

where $L = 255$ is the maximum pixel value. Higher PSNR indicates greater pixel-level similarity.

SSIM compares images across luminance, contrast, and structure [15], and is one of the most widely used perception-based image quality metrics. It has also been employed in prior event-camera privacy evaluations [1]. It is computed locally over an 11×11 Gaussian window as

$$\mathrm{SSIM}(x,y) = \frac{(2\mu_x\mu_y + c_1)(2\sigma_{xy} + c_2)}{(\mu_x^2 + \mu_y^2 + c_1)(\sigma_x^2 + \sigma_y^2 + c_2)}, \tag{3}$$

where μ_x, μ_y are local means, σ_x^2, σ_y^2 local variances, and σ_{xy} the local covariance. The stabilising constants are $c_1 = (k_1 L)^2$ and $c_2 = (k_2 L)^2$ with $k_1 = 0.01$, $k_2 = 0.03$, and $L = 255$. SSIM produces a score in $[-1, 1]$ where higher values indicate greater perceptual similarity between images. Lower PSNR and SSIM values indicate greater disruption to the reconstructed image.

LPIPS measures perceptual similarity using deep features extracted from pretrained convolutional networks [17],

$$\mathrm{LPIPS}(x,y) = \sum_{l} \frac{1}{H_l W_l} \sum_{h,w} \Big\| w_l \odot \big(\hat{\phi}^l_{hw}(x) - \hat{\phi}^l_{hw}(y)\big) \Big\|_2^2, \tag{4}$$

producing a distance score where lower values indicate higher perceptual similarity, where $\hat{\phi}^l$ are the channel-normalised activations of layer l of spatial size $H_l \times W_l$, w_l is a learned per-channel weight, and $\odot$ denotes element-wise multiplication. In this work, we use an AlexNet backbone as the feature extractor [10].

LPIPS is included because it correlates more closely with human visual perception than PSNR or SSIM [17]. While PSNR and SSIM operate on raw pixel values, LPIPS compares high-level feature representations learned from large-scale image data, capturing perceptual differences that low-level metrics may miss.

Face Identifiability

Face identifiability is evaluated using a two-stage pipeline consisting of face detection with MTCNN [16] followed by face re-identification with FaceNet [14]. For each clip, the best-detected reconstructed frame serves as a probe, whose embedding is compared against a gallery of ground-truth embeddings, one matching identity (positive) and several others (negatives), as shown in Figure 3. The exact detection, embedding, and gallery parameters are given in Section 3.

Similarity between embeddings is measured using cosine similarity, the cosine of the angle between two embedding vectors, and from these similarities three identifiability metrics are computed: Rank-1 accuracy, identification rate, and verification AUC.

Rank-1 accuracy is the fraction of probes for which the positive entry achieves the highest similarity among all six candidates, and therefore reflects whether the correct identity

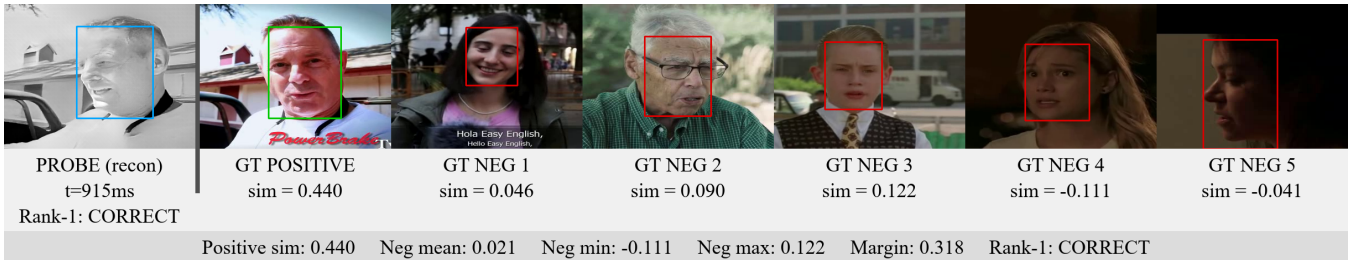

Figure 3: Baseline re-identification evaluation for a single clip. The probe frame at $t = 915$ms (blue) is correctly matched to the positive gallery entry (green) with cosine similarity $= 0.440$, over five negative entries (red) with mean cosine similarity $= 0.021$. The margin between the positive and the highest-scoring negative is 0.318.

can be picked out from a set of candidates. It is computed only over clips in which a face is detected. The identification rate instead counts clips with no detected face as failed identifications, and so measures success over all clips rather than only over detected faces.

These metrics capture two distinct ways in which identifiability can fall. A perturbation may cause matcher confusion, in which a face is still detected but is degraded enough that it no longer matches the correct identity; this lowers Rank-1 accuracy. Alternatively it may cause detection failure, in which the reconstruction is so degraded that no face is detected at all; this leaves Rank-1 unaffected, since Rank-1 is computed only over detected faces, but lowers the identification rate.

The gap between these two metrics therefore indicates how much of a method's apparent privacy gain comes from detection failure rather than genuine matcher confusion. This distinction matters because privacy obtained through detection failure is inseparable from severe quality loss, whereas matcher confusion may leave a usable reconstruction intact.

Finally, verification AUC is the area under the ROC curve formed by treating each probe's similarity to its positive entry as a genuine score and the average of its similarities to the five negative entries as a single impostor score, and is equal to the probability that a genuine pair is scored higher than an impostor pair, where a value of 1.0 indicates perfectly separable identities and 0.5 indicates chance. Unlike Rank-1, which depends on a single matching decision, AUC measures how separable identities remain across all decision thresholds.

3 Experimental Setup

Experiments are conducted on 300 video clips drawn from the CelebV-HQ dataset, spanning 291 unique YouTube identities. The full pipeline was executed on Google Colab's GPU runtime, with code and pretrained weights stored on Google Drive. For each clip, synthetic events are generated using v2e with output resolution 346×260 pixels and default threshold parameters, and reconstructed using E2VID with fixed-duration temporal windows of $\lfloor 1000/\text{fps} \rfloor$ ms as described in Section 2. The full fixed configuration of the pipeline, including all v2e, E2VID, and detection parameters held constant across experiments, is given in Table 1.

Reconstruction quality is measured using PSNR and SSIM via `scikit-image` with `data_range=255`, and LPIPS via the `lpips` library using a pretrained AlexNet network with inputs normalized to $[-1, 1]$.

Table 1: Fixed configuration of the experimental pipeline, held constant across all perturbation methods and strength levels so that observed differences are attributable to the perturbations rather than to the setup.

Stage	Parameter	Value
v2e	sensor model	DAVIS346 (346×260)
	pos/neg threshold	0.2 (log intensity)
	threshold variation σ	0.03
	cutoff frequency	300 Hz
	leak rate	0.01 Hz/px
	shot-noise rate	0.001 Hz
	leak jitter fraction	0.1
	noise CV	0.1
	refractory period	0.5 ms
	frame interpolation	Super-SloMo
	RNG seed	not fixed
E2VID	checkpoint	E2VID_lightweight.pth
	window mode	fixed-duration
	window duration	$\lfloor 1000/\text{fps} \rfloor$ ms
	auto-HDR	enabled
	event normalisation	enabled
	recurrent state	enabled
	retraining	none
MTCNN	crop size	160×160
	margin	20
	keep_all	False
	grayscale handling	channel replicated ×3
FaceNet	architecture	Inception-ResNet-V1
	weights	VGGFace2
	embedding	512-d, L2-normalised
	similarity	cosine

Face identifiability is evaluated per clip: the reconstructed frame with the highest MTCNN detection confidence serves as the probe, and a 512-dimensional L2-normalised FaceNet embedding is extracted using an Inception-ResNet-V1 network pretrained on VGGFace2 [5]. Each probe is compared against a gallery as mentioned in Section 2.4. Clips sharing the same YouTube base identity are excluded from the negative set, as multiple clips from the same source video may depict the same person and would therefore not constitute true negatives.

Rank-1 accuracy, identification rate, and verification AUC are reported across all clips for each perturbation method

and strength level, with all metrics computed from cosine similarities between embeddings as defined in Section 2.4.

For polarity flipping, the flip probability p is varied across $0.05, 0.1, 0.2, 0.3$. For spatial jitter, the standard deviation σ is varied across $1.0, 2.0, 3.0, 4.0$ pixels. For event insertion and deletion, the rate ρ is varied across $0.1, 0.2, 0.3, 0.4, 0.5$. An unperturbed baseline is also evaluated for each clip.

4 Results

4.1 Baseline Reconstruction and Identifiability

The unperturbed event streams establish the reference point against which all perturbations are measured. Reconstruction quality at baseline yields the following results: PSNR is 9.78 dB, SSIM 0.372, and LPIPS 0.459. These values reflect the difficulty of the reconstruction task itself, as E2VID recovers grayscale intensity from sparse brightness-change events alone, without colour or absolute-intensity information.

The baseline reaches a Rank-1 accuracy of 97.5%, and an identification rate of 92.5% once clips with no detected face are taken into account. How identifiability behaves as the reconstruction is perturbed is examined in the following sections. Finally, even at baseline 5.1% of clips yield no detectable face, most likely due to side-profile or non-frontal source frames rather than reconstruction failure. This baseline rate is the reference for the perturbed conditions: any no-face rate above 5.1% can be attributed to the perturbation, not to the dataset.

4.2 General Trends

Across all three perturbation methods, increasing the perturbation strength reduces face identifiability and degrades reconstruction quality, as reported in Table 2. The privacy-utility tradeoff is therefore present in every case: identity becomes harder to recover, but only at the cost of a worse reconstruction. The methods differ, however, in how steeply each quantity falls and in which aspect of quality is most affected, and these differences are the focus of the following sections.

At the highest strength of every method the no-face rate rises sharply, well above the 5.1% baseline, indicating that strong perturbation eventually reduces identifiability through detection failure rather than by anonymising recognisable faces. AUC, by contrast, remains high for almost all conditions and falls clearly only for the strongest spatial-jitter settings, as shown in Figure 4 and examined in Section 5.2.

The following subsections examine each method in these terms, using the matcher confusion and detection failure mechanisms defined in Section 2.4 to interpret how identifiability falls. Polarity flipping, spatial jitter, event insertion and deletion are treated in turn, each answering one of the study's sub-questions, before the methods are compared on the overall tradeoff.

4.3 Polarity Flipping

Polarity flipping has little effect on identifiability at low strength and a growing effect as the flip probability increases. At p=0.05, Rank-1 accuracy is 97.8% and the identification rate is 92.2%, both effectively unchanged from baseline. Identifiability then declines steadily: the identification rate falls to

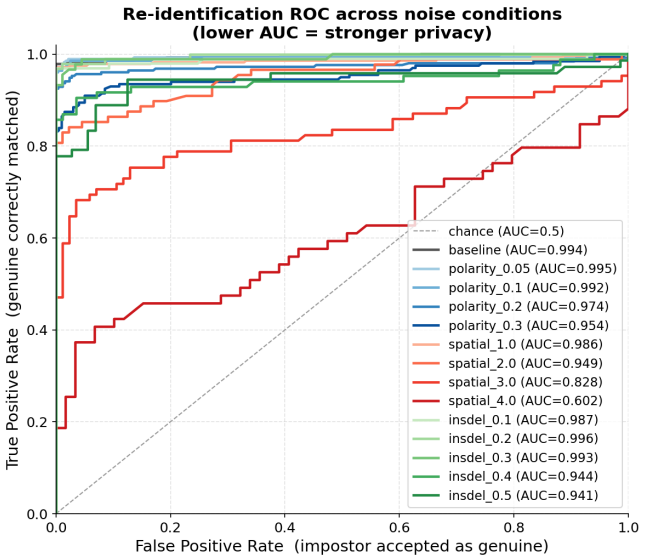

Figure 4: Re-identification ROC curves across all conditions. Almost all conditions stay near the top-left despite large drops in Rank-1 accuracy, showing identity remains separable under verification. Only the two strongest spatial-jitter settings bend toward chance.

87.8% at p=0.1, to 79.0% at p=0.2, and to 53.8% at p=0.3. The reconstruction model is therefore largely insensitive to a small fraction of flipped events, and only begins to lose identity information once a substantial share of polarities is inverted.

At higher strength, much of this reduction is driven by detection failure rather than matcher confusion. The no-face rate rises from 6.1% at p=0.1 to 14.2% at p=0.2 and 33.4% at p=0.3, the last far above the 5.1% baseline. Over the faces that are still detected, Rank-1 accuracy remains comparatively high at 80.9% at p=0.3, so the gap between this value and the 53.8% identification rate shows that the apparent privacy gain at high strength comes largely from reconstructions degraded past the point of detection, not from detected faces being mismatched. At p=0.05 and p=0.1, where the no-face rate stays near baseline, the two measures track each other and the modest identifiability loss reflects genuine matcher confusion. Verification AUC stays high throughout, from 0.995 at p=0.05 to 0.954 at p=0.3, indicating that identity remains separable even where the identification rate falls.

The accompanying quality loss is predominantly perceptual rather than pixel-level. LPIPS rises sharply, from 0.497 at p=0.05 to 0.762 at p=0.2, while PSNR and SSIM decline more gradually over the same range. The reason this loss is largely perceptual, and why the model is robust at low p, is examined in Section 5.1.

In answer to the first sub-question, polarity flipping reduces face identifiability only at high flip probabilities, and does so largely by causing detection failure rather than by anonymising recognisable faces. Because this reduction comes with substantial perceptual quality loss, polarity flipping does not provide a favourable privacy–utility tradeoff.

Table 2: Face identifiability and reconstruction quality for the baseline and all perturbation conditions, with metrics as defined in Section 2.4. Lower identifiability indicates stronger privacy. AUC stays high (> 0.94) for most conditions, declining only for spatial jitter at σ=3.0 (0.828) and collapsing at σ=4.0 (0.602), where privacy comes only from collapsed reconstructions, as the high No-face(%) and low SSIM values show.

Condition	Rank-1 (%)	Identification Rate (%)	AUC	No-face (%)	PSNR (dB) ↑	SSIM ↑	LPIPS ↓
Baseline	97.5	92.5	0.994	5.10	9.78	0.372	0.459
Polarity Flipping p=0.05	97.8	92.2	0.995	5.76	9.36	0.356	0.497
Polarity Flipping p=0.1	93.5	87.8	0.992	6.10	8.97	0.340	0.550
Polarity Flipping p=0.2	92.1	79.0	0.974	14.2	8.18	0.293	0.762
Polarity Flipping p=0.3	80.9	53.8	0.954	33.4	8.32	0.302	0.756
Spatial Jitter σ=1.0	91.8	85.9	0.986	6.35	10.1	0.357	0.574
Spatial Jitter σ=2.0	78.4	69.0	0.949	12.0	10.1	0.347	0.682
Spatial Jitter σ=3.0	**54.1**	**46.0**	**0.828**	**15.0**	**9.95**	**0.334**	**0.748**
Spatial Jitter σ=4.0	22.0	13.0	0.602	41.0	9.92	0.317	0.801
Event Insertion & Deletion ρ=0.1	96.9	94.0	0.987	3.00	9.86	0.375	0.451
Event Insertion & Deletion ρ=0.2	95.7	90.0	0.996	6.00	10.9	0.297	0.598
Event Insertion & Deletion ρ=0.3	88.8	79.0	0.993	11.0	11.0	0.258	0.677
Event Insertion & Deletion ρ=0.4	79.8	67.0	0.944	16.0	11.2	0.230	0.743
Event Insertion & Deletion ρ=0.5	80.6	58.0	0.941	28.0	11.3	0.207	0.800

4.4 Spatial Jitter

Spatial jitter produces the steepest reduction in identifiability of the three methods, and it does so while leaving pixel-level quality almost unchanged. As the standard deviation increases, Rank-1 accuracy falls from 91.8% at σ=1.0 to 78.4% at σ=2.0, 54.1% at σ=3.0, and 22.0% at σ=4.0. The identification rate follows the same trajectory, from 85.9% to 13.0% across the sweep. By σ=3.0, identifiability is roughly halved relative to baseline, and by σ=4.0 it approaches the level expected from chance.

The mechanism behind this reduction changes with strength. Up to σ=3.0, the no-face rate stays modest, at 6.4%, 12.0%, and 15.0% for σ=1.0, 2.0, and 3.0 respectively, only slightly above the 5.1% baseline. The identifiability loss in this range therefore comes predominantly from matcher confusion: faces are still detected, but the displaced events yield reconstructions that no longer match the correct identity. At σ=4.0 the behaviour changes sharply, with the no-face rate jumping to 41.0%, indicating that the displacement has become large enough to prevent detection altogether. Spatial jitter thus transitions from anonymising recognisable faces at moderate strength to destroying reconstructions at high strength. This shift is mirrored in the verification AUC, which stays high at 0.949 for σ=2.0 but falls to 0.828 at σ=3.0 and 0.602 at σ=4.0, the only condition in the study to approach chance. Spatial jitter is therefore the only method where the drop in Rank-1 accuracy is accompanied by a real loss of identity separability, rather than the identity remaining recoverable despite the lower Rank-1 score.

The most notable feature of this method is the decoupling of identifiability from pixel-level quality. PSNR remains essentially constant across the entire sweep, at 10.1, 10.1, 9.95, and 9.92 dB for σ=1.0 to 4.0, despite Rank-1 accuracy falling by almost 70 percentage points over the same range (Figure 5).

SSIM declines only mildly, from 0.357 to 0.317. Spatial displacement therefore degrades identity far more than it degrades pixel-error metrics: the same jitter that collapses recognizability leaves PSNR and SSIM nearly unchanged. The mechanism behind this decoupling is examined in Section 5.1.

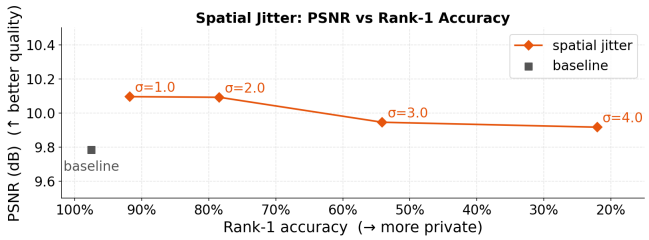

Figure 5: PSNR vs Rank-1 accuracy for different jitter strengths σ.

This decoupling makes σ=3.0 the most favourable operating point observed for spatial jitter. At this strength, identifiability is reduced substantially, to 54.1% Rank-1 and a comparable identification rate, while the no-face rate remains low and PSNR and SSIM are close to their baseline values. The privacy gain at σ=3.0 is therefore obtained primarily through matcher confusion on reconstructions that remain detectable and pixel-wise faithful, rather than through the reconstruction collapse that characterises σ=4.0.

In answer to the second sub-question, spatial jitter reduces face identifiability more effectively than the other methods and, at moderate strength, does so while preserving pixel-level reconstruction quality. It offers the clearest separation between privacy and reconstruction quality among the methods tested.

4.5 Event Insertion and Deletion

Event insertion and deletion reduces identifiability gradually with strength, and less steeply than spatial jitter. Rank-1 accuracy stays near baseline at low rates, at 96.9% for ρ=0.1 and 95.7% for ρ=0.2, then declines to 88.8% at ρ=0.3 and to around 80% at ρ=0.4 and ρ=0.5. The identification rate falls more clearly over the same range, from 94.0% at ρ=0.1 to 58.0% at ρ=0.5. Identity is therefore retained well for ρ values between 0.1 and 0.3.

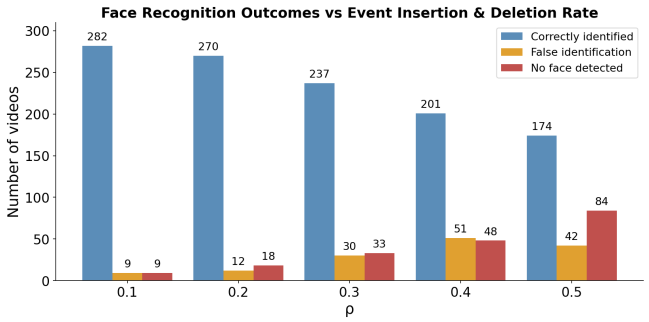

Figure 6: Face identification at different strength levels ρ

The two identifiability measures diverge in an informative way at high strength. Rank-1 accuracy is slightly higher at ρ=0.5 (80.6%) than at ρ=0.4 (79.8%), yet the identification rate continues to fall, from 67.0% to 58.0%, over the same step. Figure 6 makes the cause explicit: between the two conditions the no-face count rises sharply from 48 to 84 (from 16% to 28%) while false identifications actually decrease (from 51 to 42). In other words, the additional perturbation at ρ=0.5 does not confuse the matcher further, it degrades the reconstruction so severely that the detector can no longer find a face at all, converting would be matches (correct or incorrect) into detection failures. Matcher accuracy on the faces that remain detectable has effectively saturated, so the additional privacy at ρ=0.5 comes entirely from detection failure rather than from further matcher confusion. This is the clearest case in the study where the identification rate, rather than Rank-1 accuracy, reflects the true change in privacy. Verification AUC remains above 0.94 across the entire sweep, confirming that the reduction in identification comes from detection failure rather than from any loss of identity separability.

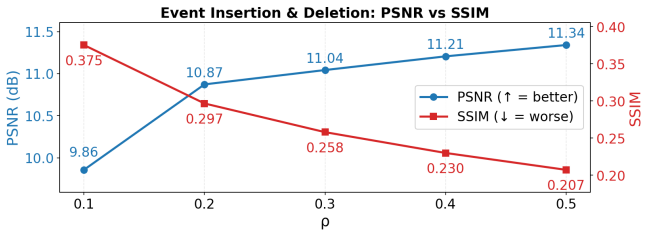

Figure 7: PSNR and SSIM as a function of event insertion & deletion rate ρ. Despite an increase in PSNR, SSIM decreases, suggesting that PSNR alone may not accurately reflect perceptual image quality in reconstructed event data.

Reconstruction quality behaves unusually under this method. As ρ increases, SSIM falls steeply (from 0.375 to 0.207) and LPIPS worsens (from 0.451 to 0.800), yet PSNR moves the opposite way, rising from 9.86 to 11.3dB (Figure7). We explain why PSNR can rise as the reconstruction degrades in Section 5.2. In answer to the third sub-question, event insertion and deletion reduces face identifiability only at high rates, and its privacy gain at the highest strength is driven mainly by detection failure. This privacy is not genuine anonymisation but a by-product of degrading the reconstruction until no face can be detected, so event insertion and deletion does not offer a usable privacy utility tradeoff.

5 Discussion

5.1 Effectiveness of Perturbations

We have seen in the results that the three methods reduce identifiability at very different rates relative to the quality they sacrifice, with spatial jitter reducing identity far more steeply than polarity flipping or event insertion and deletion at comparable levels of structural quality. We attribute this difference to which part of the reconstruction each perturbation disturbs, relative to the information a face recognition model actually uses.

With spatial jitter identifiability falls sharply while PSNR stays almost constant. Since jitter displaces the coordinates of events but leaves their overall distribution intact, the global intensity statistics that govern pixel-error are largely preserved, while the precise spatial relationships between facial features are scrambled. We therefore reason that jitter removes close to only the information the recogniser relies on, which is why it can suppress identity while leaving pixel-level quality nearly untouched. This targeted disruption of geometry is, in our interpretation, the reason it preserves the tradeoff better than the other two methods.

Polarity flipping shows a contrasting pattern, with identifiability staying close to baseline at low strength and the cost appearing mostly in perceptual quality. Flipping inverts an event's polarity but leaves it in place, and since E2VID integrates many events recurrently within each reconstruction window, we reason that a minority of inverted polarities is averaged out, so the facial geometry survives largely intact. The flipped events that remain, however, reverse the direction of the brightness change they encode, and we reason that the resulting contrast artefacts are penalised heavily by a perceptual metric such as LPIPS while having comparatively little effect on mean-squared error. This is why the cost of polarity flipping falls almost entirely on perceptual quality.

Event insertion and deletion is similarly inefficient, with identity persisting well into its sweep. Because this perturbation adds and removes events across the whole frame rather than acting on facial structure specifically, we reason that it degrades the image globally while leaving the geometry of the detectable faces broadly recognisable. Its eventual reduction in identifiability, as the results show, comes less from confusing the matcher than from corrupting the reconstruction until no face is detected at all.

Finally, we reason that a raw-event perturbation can reduce identifiability while preserving reconstruction quality only when it disrupts the structural features the recogniser uses; otherwise it lowers identifiability only by degrading the reconstruction as a whole. This is, in our interpretation, why spatial jitter provides the most favourable privacy–utility tradeoff of the three methods: it is the only one that shows a clear decoupling of identifiability from pixel-level quality, reducing identifiability while keeping reconstruction quality within the low-degradation band shown in Figure 8 in the Appendix. This answers the fourth sub-question: spatial jitter offers the best

tradeoff among the methods tested, with σ=3.0 as its most favourable operating point.

5.2 Observed Anomalies

A number of results stood out as counterintuitive, and we discuss them here. The first is the behaviour of PSNR under event insertion and deletion. We have seen that PSNR rises as the rate increases, even though SSIM and LPIPS fall and identifiability declines over the same range. We did not expect a quality metric to improve as the reconstruction visibly degrades. We attribute this to the way the perturbation reshapes the image: inserting random events and removing real ones pushes the reconstruction toward a flatter, lower-contrast output, and since PSNR is governed by mean-squared error against an already low-contrast grayscale face, a flatter image can reduce that error even as facial structure is destroyed. This show that no single metric captures reconstruction quality on its own: had we relied on PSNR alone, this condition would have appeared to improve, while SSIM and LPIPS reveal the degradation that PSNR misses. This result justifies our use of three complementary metrics rather than one.

The second concerns the difference between retrieval and verification. We have seen that AUC in Table 2 stays above 0.94 for almost every condition, falling meaningfully only for spatial jitter at 4.0. This was unexpected, because retrieval accuracy drops substantially for several conditions where AUC barely moves. We interpret this as a sign that most perturbations do not remove identity from the embedding space; they only make the correct identity harder to single out from a gallery. Since AUC measures how separable a genuine match is from an impostor across all thresholds, its persistence means that an attacker asking only whether a reconstruction matches a specific person would still succeed where retrieval appears to fail. We therefore read the retrieval-based privacy gains as real but shallow, and we note that even our most effective operating point leaves identity largely separable under verification.

The third is the non-monotonic behaviour of event insertion and deletion at high rates. We have seen that Rank-1 accuracy is slightly higher at ρ=0.5 than at ρ=0.4, even though the identification rate continues to fall. We initially found this contradictory, but the breakdown in Figure 6 resolves it: between these two strengths the number of undetected faces rises sharply while false matches decline, so the added perturbation no longer misleads the matcher but instead degrades marginal detections into complete detection failures. We take this as a demonstration in the study that Rank-1 accuracy alone can misrepresent privacy, since it is computed only over detected faces.

5.3 Limitations

Several factors bound the conclusions of this study. The events are synthesised from RGB video using v2e rather than captured with a real event camera, and while this is necessary to obtain ground-truth frames for the quality metrics, synthetic events may not reproduce the noise characteristics of real sensors. The observed trends could therefore differ on real hardware.

Identifiability is measured with a single detector and recogniser pair, MTCNN for detection and FaceNet for re-identification. A different recognition model may rely on the facial information in different proportions and could prove more or less robust to each perturbation, so the relative ordering of the methods is specific to this pipeline rather than guaranteed to hold in general.

The attacker model is also limited. The perturbations are fixed and applied without any knowledge of the recogniser, so the evaluation assumes a non-adaptive attacker. An attacker who knew which perturbation had been applied could retrain a recognition model to be robust to it, which means the identifiability we report is best understood as a lower bound on what a determined attacker could recover.

Finally, identity is assessed only through face re-identification. Other biometric cues that may survive reconstruction, such as gait or body shape, are not considered, so the perturbations cannot be claimed to anonymise a subject completely even where they reduce face identifiability.

6 Responsible Research

This section addresses the ethical implications and data usage boundaries of our work, ensuring that our technical evaluations align with responsible academic research practices.

6.1 Ethical Considerations

CelebV-HQ is used strictly for non-commercial research purposes, in accordance with its license. The dataset consists of publicly available YouTube videos of celebrities. While the subjects are real, identifiable individuals, their identities are used solely as labels for evaluating face re-identification metrics. It should also be noted that this study uses synthetic event streams generated from RGB video rather than real event camera recordings. No new data is collected from real individuals, and explicit consent was not necessary as the study relies entirely on publicly available datasets that preserve privacy. This work is intended solely for academic research purposes.

6.2 Reproducibility of Methods

Reproducibility is a core requirement of reliable research. To support replicability, the experimental pipeline uses only publicly available pretrained models (E2VID, v2e, MTCNN, and FaceNet), and all exact parameter values are reported in Section 3, with the code released at the repository linked in the abstract. All mathematical formulations, algorithmic steps, and hyperparameter selections for the evaluated perturbation methods are explicitly documented to ensure the pipeline is fully open.

6.3 Use of LLMs

Large language models were used in a supporting role only. For writing, they were used to correct grammar and spelling, improve readability and conciseness, to refine text, and assist with LaTeX formatting, including the creation of figures and graphs. For code, they were used to improve code readability, writing comments and as a debugging aid.

7 Conclusions and Future Work

Event cameras are often assumed to be privacy-preserving because they record only sparse brightness changes rather than

full images, yet reconstruction models can recover recognisable faces from these events. This work asked what effect simple, training-free perturbations applied directly to the raw event stream have on the resulting reconstruction quality and on how identifiable the reconstructed faces are, and it provides the first systematic head-to-head comparison of three such perturbations, polarity flipping, spatial jitter, and event insertion and deletion, on face data under consistent metrics.

We find that the three methods behave very differently. Polarity flipping reduces identifiability only at high strength, and largely by degrading the image until no face can be detected, at a substantial perceptual cost. Spatial jitter reduces identifiability most steeply and, at moderate strength, does so while leaving pixel-level quality almost unchanged, making it the best privacy–utility tradeoff among the methods tested. Event insertion and deletion is the least effective, retaining identity until high rates and reducing it only by corrupting the reconstruction. The broader conclusion is that simple raw-event perturbations can reduce identifiability, but always at a cost to reconstruction quality, and the methods differ mainly in how favourable that exchange is. Spatial jitter offers the most favourable exchange of the three, though even it reduces identity only partially.

Two important nuances affect how these reductions in identifiability should be read, detailed in Section 5.2: no single image-quality metric is sufficient, since PSNR can stay flat or even improve as reconstructions degrade; and the privacy obtained is real but shallow, since identity remains separable under verification (AUC) even where retrieval accuracy falls. Simple raw-event perturbations should therefore be seen as a lightweight first layer applied directly to the event stream, rather than a complete anonymisation method.

Several directions remain for future work. The most important is evaluation on real event-camera recordings rather than the synthetic events used here, to confirm the trends hold under genuine sensor noise. The perturbations should also be tested against an adaptive attacker who knows the applied method and can retrain a recogniser to counter it, giving a tighter bound on the protection they offer. Finally, combining complementary perturbations may achieve stronger anonymisation than any single one, and extending the evaluation to other biometric cues such as gait or body shape would test whether identity is protected beyond the face alone.

References

[1] Shafiq Ahmad, Pietro Morerio, and Alessio Del Bue. Person re-identification without identification via event anonymization. In *Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV)*, pages 11132–11141, October 2023.

[2] 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. In *Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision (WACV) Workshops*, pages 459–468, January 2022.

[3] Katharina Bendig, René Schuster, Nicole Thiemer, Karen Joisten, and Didier Stricker. Anonynoise: Anonymizing event data with smart noise to outsmart re-identification and preserve privacy. In *Proceedings of the Winter Conference on Applications of Computer Vision (WACV)*, pages 3159–3161, February 2025.

[4] Christian Brandli, Raphael Berner, Minhao Yang, Shih-Chii Liu, and Tobi Delbruck. A 240 × 180 130 db 3 µs latency global shutter spatiotemporal vision sensor. *IEEE Journal of Solid-State Circuits*, 49(10):2333–2341, 2014.

[5] Qiong Cao, Li Shen, Weidi Xie, Omkar M. Parkhi, and Andrew Zisserman. Vggface2: A dataset for recognising faces across pose and age. In *2018 13th IEEE International Conference on Automatic Face & Gesture Recognition (FG 2018)*, pages 67–74, 2018.

[6] Guillermo Gallego, Tobi Delbrück, Garrick Orchard, Chiara Bartolozzi, Brian Taba, Andrea Censi, Stefan Leutenegger, Andrew J. Davison, Jörg Conradt, Kostas Daniilidis, and Davide Scaramuzza. Event-based vision: A survey. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 44(1):154–180, 2022.

[7] Yuhuang Hu, Shih-Chii Liu, and Tobi Delbruck. v2e: From video frames to realistic dvs events. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR) Workshops*, pages 1312–1321, June 2021.

[8] Huaizu Jiang, Deqing Sun, Varun Jampani, Ming-Hsuan Yang, Erik Learned-Miller, and Jan Kautz. Super slomo: High quality estimation of multiple intermediate frames for video interpolation. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, June 2018.

[9] Junho Kim, Young Min Kim, Ramzi Zahreddine, Weston A. Welge, Gurunandan Krishnan, Sizhuo Ma, and Jian Wang. Privacy-preserving visual localization with event cameras. *IEEE Transactions on Image Processing*, 34:6215–6230, 2025.

[10] Alex Krizhevsky, Ilya Sutskever, and Geoffrey E Hinton. Imagenet classification with deep convolutional neural networks. In F. Pereira, C.J. Burges, L. Bottou, and K. Weinberger, editors, *Advances in Neural Information Processing Systems*, volume 25. Curran Associates, Inc., 2012.

[11] Patrick Lichtsteiner, Christoph Posch, and Tobi Delbruck. A 128× 128 120 db 15 μs latency asynchronous temporal contrast vision sensor. *IEEE Journal of Solid-State Circuits*, 43(2):566–576, 2008.

[12] Henri Rebecq, René Ranftl, Vladlen Koltun, and Davide Scaramuzza. High speed and high dynamic range video with an event camera. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 43(6):1964–1980, 2021.

[13] Olaf Ronneberger, Philipp Fischer, and Thomas Brox. U-net: Convolutional networks for biomedical image segmentation. In Nassir Navab, Joachim Hornegger, William M. Wells, and Alejandro F. Frangi, editors,

Medical Image Computing and Computer-Assisted Intervention – MICCAI 2015, pages 234–241, Cham, 2015. Springer International Publishing.

[14] Florian Schroff, Dmitry Kalenichenko, and James Philbin. Facenet: A unified embedding for face recognition and clustering. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, June 2015.

[15] Zhou Wang, Alan C Bovik, Hamid R Sheikh, and Eero P Simoncelli. Image quality assessment: from error visibility to structural similarity. *IEEE Transactions on Image Processing*, 13(4):600–612, 2004.

[16] Kaipeng Zhang, Zhanpeng Zhang, Zhifeng Li, and Yu Qiao. Joint face detection and alignment using multitask cascaded convolutional networks. *IEEE Signal Processing Letters*, 23(10):1499–1503, 2016.

[17] Richard Zhang, Phillip Isola, Alexei A. Efros, Eli Shechtman, and Oliver Wang. The unreasonable effectiveness of deep features as a perceptual metric. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, June 2018.

[18] 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. In Shai Avidan, Gabriel Brostow, Moustapha Cissé, Giovanni Maria Farinella, and Tal Hassner, editors, *Computer Vision – ECCV 2022*, pages 650–667, Cham, 2022. Springer Nature Switzerland.

Appendix

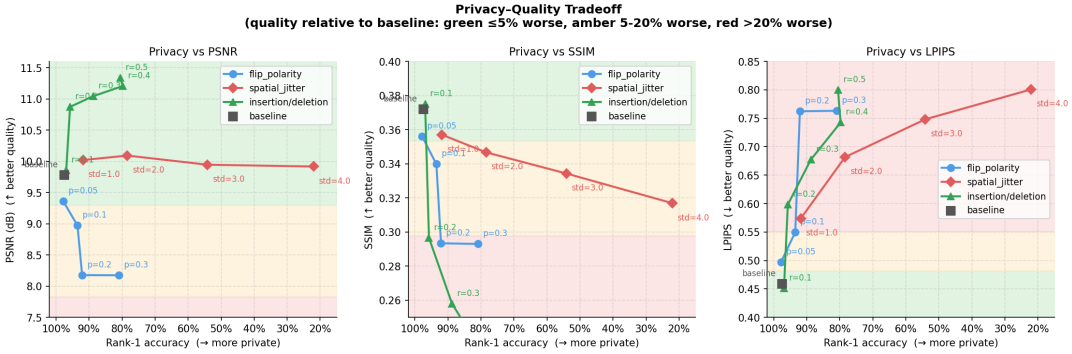

Figure 8: Privacy–utility tradeoff for all three methods across the three quality metrics, plotted against Rank-1 accuracy. Shaded bands mark quality loss relative to baseline (green $\leq$5%, amber 5–20%, red $>$20%). Spatial jitter (red) stays in or near the low-degradation bands while its Rank-1 accuracy falls furthest, making it the most favourable of the three; polarity flipping and event insertion and deletion leave the favourable bands at smaller privacy gains.