Privacy Risks in Event-Based Cameras: The Role of Sensor Configuration

How Different Sensor Configurations Affect Face Identification

Sofyan Ali Ali[1]

Supervisor(s): Nergis Tömen, Tunahan Parlayıcı

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: Sofyan Ali Ali
Final project course: CSE3000 Research Project
Thesis committee: Nergis Tömen, Tunahan Parlayıcı, Ricardo Marroquim

Abstract

Event-based cameras—sensors that asynchronously record pixel-level brightness changes rather than full image frames—are often assumed to be privacy-preserving due to their sparse visual output. This thesis investigates how physical sensor configurations, including temporal bandwidth, contrast thresholds, leak noise, and spatial resolution, affect the trade-off between data utility and biometric privacy risk. We introduce a cross-domain evaluation framework that measures identity leakage using reconstructed event streams and a frozen pre-trained face recognition model. Our results show that event streams retain sufficient facial structure for accurate identification. We further identify a privacy paradox in which reducing temporal bandwidth increases attacker performance by denoising the signal, while background leak noise effectively disrupts reconstruction and lowers identification accuracy. Privacy effectiveness also varies substantially with subject motion, and the apparent benefits of resolution scaling are largely explained by domain mismatch. Overall, the findings suggest that static sensor configurations cannot guarantee anonymity, highlighting the need for threat-aware sensor design and adaptive privacy safeguards. The code used in this research is available at: https://github.com/Stunner070/research-project-bsc-cse-tudelft

1 Introduction

Event-based cameras are neuromorphic sensors that asynchronously report pixel-level brightness changes instead of capturing full image frames at fixed intervals [1]. This event-driven sampling provides low latency and high dynamic range while reducing redundant information in static regions. These properties make event cameras attractive for high-speed robotics, AR/VR, and smart surveillance. At the same time, because raw event streams often appear as sparse edge patterns rather than textured RGB images, they are frequently considered more privacy-preserving than conventional video. Figure 1 shows the difference between an RGB and Event-camera.

This perceived anonymity is particularly important in sensitive environments. Event cameras are increasingly considered for always-on systems such as in-home robots, wearable augmented-reality glasses, and smart surveillance devices, where continuous RGB recording may be legally restricted or socially unacceptable. However, if recognizable faces can be reconstructed or matched from these event streams, the sensor no longer provides the assumed privacy protection. Instead, it may allow identity information to leak in settings where conventional cameras would be considered too invasive.

Recent work has challenged the assumption that event data is naturally anonymous. Event-based representations have been shown to support recognition, reconstruction, and localization tasks [1]. Ahmad et al. introduce event-driven person re-identification (ReID) benchmarks and models, showing that event-frames or voxel grids can be used to match individuals across camera views, despite the absence of full intensity images [2]. Follow-up work on event anonymization further shows that privacy protection must be designed explicitly rather than assumed from the sensing modality alone [3]. Event cameras have also been shown to support face recognition directly, suggesting that facial identity can be recovered from event streams even without full intensity images [4]. More recent work has extended event-based ReID with cross-modality and temporal collaboration approaches, further demonstrating the potential for identity inference from event streams [5; 6]. These results underline that event data can carry rich biometric information, but they do not address how sensitive such information is to low-level sensor choices such as contrast thresholds, leak rate, or temporal bandwidth.

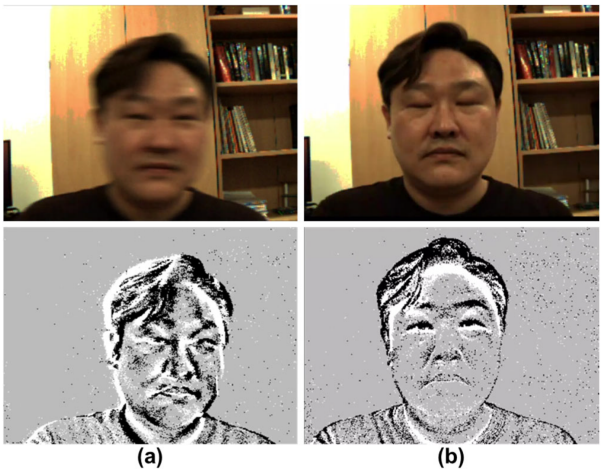

Figure 1: Top Row: RGB Camera, Bottom Row: Event-based Camera. Source: https://www.mdpi.com/2076-3417/13/18/10357.

Similarly. Kim et al. demonstrate that event cameras can support visual localization in real-world environments while also emphasizing the need for explicit privacy safeguards [7]. Together, these studies suggest that event cameras may retain more identity-related information than is commonly assumed.

In parallel, the event-vision community has developed realistic video-to-event simulators that provide fine-grained control over virtual sensor behavior. Simulator surveys cover tools such as ESIM, v2e, DVS-Voltmeter, and related datasets, highlighting the maturity of this research ecosystem [8]. In particular, v2e converts standard RGB video into synthetic event streams while exposing key sensor parameters, including spatial resolution, contrast thresholds, threshold variability, and leak/noise rates [9]. These simulators make it possible to keep scene content fixed while varying the sensor configuration, which is essential for studying how much identity information is encoded under different hardware settings.

Despite these developments, privacy protection in event-based vision remains underexplored compared with traditional RGB pipelines. Existing event-based ReID work often assumes a single physical sensor or simulator configuration and focuses on representations, models, or anonymization mechanisms [2; 3; 5]. Privacy-aware localization work

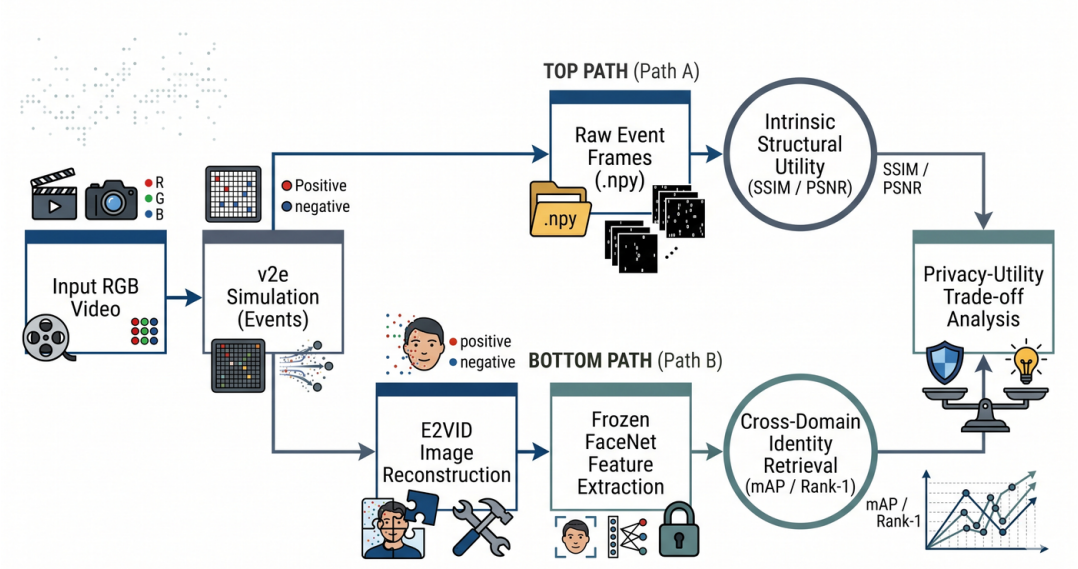

Figure 2: Overview of the methodology, from RGB video input and event-camera simulation to event-domain representations, reconstruction, and identity-risk evaluation.

focuses mainly on architectural safeguards rather than on the information content of the event signal itself [7]. Unlike these works, we treat the sensing hardware itself as the variable under investigation. Rather than proposing a new representation or model, we keep the reconstruction and recognition pipelines fixed and systematically vary physical sensor parameters to study how much identity information remains available to an attacker under different configurations.

A key open question is therefore how much identity information is present in event-based representations and how this depends on the underlying sensor configuration. In particular, it remains unclear how parameters such as resolution, contrast thresholds, and noise levels affect an attacker's ability to infer identity when the underlying scene content is held fixed.

This thesis addresses that gap by studying the relationship between event-camera sensor configurations and identity leakage in event-based representations. We focus on event-derived representations called event-frames and model the attacker as a strong person re-identification system operating directly on these representations. The main research question is:

> **How does the ability of an attacker to infer identity from event-based representations vary across different event-camera sensor configurations, when the underlying scene content is fixed?**

To answer this question, we define an attacker model and a qualitative notion of privacy risk for event-based ReID. We then propose an evaluation protocol in which video-to-event simulation generates multiple event datasets from the same RGB sequences under systematically varied sensor parameters. This setup allows us to empirically characterize how identity inference performance changes across configurations and provides a first step toward understanding how sensor-level choices affect privacy risks in event-based vision systems.

The remainder of this thesis is structured as follows. Chapter 2 describes the methodology, including the attacker threat model and representation formats. Chapter 3 presents the experimental setup, including dataset construction, sensor configurations, and evaluation metrics. Chapter 4 reports the empirical results of the privacy–utility evaluation. Chapter 5 discusses the physical interpretation of the findings, and Chapter 6 concludes with directions for future work.

2 Methodology

2.1 Overall Approach

This work evaluates how event-camera sensor settings affect identity leakage when scene content is fixed. The method follows a simulation–attack pipeline. Identity-disjoint RGB face videos are converted into synthetic events under controlled sensor configurations and encoded as event-frame sequences. A fixed attacker pipeline then reconstructs and matches facial identities using pre-trained models. Changes in attacker performance are interpreted as changes in biometric privacy risk. Figure 2 illustrates the complete methodology pipeline.

The methodology isolates sensor effects from data effects. Each configuration is applied to the same VoxCeleb2-derived clips, so differences in recognition accuracy arise from simulated sensor behaviour rather than from identity, pose, or illumination changes in the source data.

2.2 Attacker Model and Privacy Notion

Let x be an event-derived representation and $y \in Y$ its corresponding identity label. We model the attacker as a high-capability adversary who intercepts event streams and attempts to recover user identity. The attack pipeline consists

of a reconstruction network (E2VID [10]) that converts asynchronous events into grayscale frames, followed by a frozen, pre-trained RGB face recognition backbone $f_\theta : \mathcal{X} \to \mathbb{R}^d$.

Biometric privacy risk is evaluated through a cross-domain retrieval framework. The gallery contains baseline facial embeddings, while the queries are embeddings extracted from reconstructed event streams generated under different sensor configurations. Privacy risk is quantified by retrieval performance: higher identification rates indicate that identity-relevant information remains recoverable despite the applied sensor configuration.

2.3 Sensor Configuration Space and Simulation

Event-camera output depends on resolution, ON/OFF contrast thresholds, threshold mismatch, photoreceptor bandwidth, refractory period, leak events, and polarity asymmetry [1; 8]. Real devices expose these effects through coupled bias parameters, so an exhaustive sweep is impractical. We therefore use one baseline configuration and one-parameter perturbations that represent realistic operating regimes.

Synthetic events are generated with v2e [9], which models contrast-triggered events, photoreceptor filtering, pixel mismatch, shot noise, and leak events. The controlled parameters are:

- spatial resolution of the virtual sensor array;
- positive and negative contrast thresholds (C^+, C^-);
- threshold variability σ_{th};
- photoreceptor cut-off frequency;
- background leak rate;
- polarity asymmetry through unequal ON/OFF thresholds.

Following bias-tuning work, we focus on parameters that dominate event rate, signal-to-noise ratio, and polarity balance [11; 12]. The evaluated variants are: 346×260 versus 640×480 resolution; baseline versus high-sensitivity thresholds; high versus reduced photoreceptor bandwidth; baseline versus high leak rate; and symmetric versus ON-dominant thresholds. Section 3.3 gives the numerical settings.

2.4 Event-Derived Representations

An event stream is an asynchronous sequence of events $e_i = (x_i, y_i, t_i, p_i)$, where each event records the location, timestamp, and polarity of a brightness change. To support both utility evaluation and adversarial reconstruction, we use two representation formats.

For the Intrinsic Structural Evaluation (Path A), events are accumulated over a fixed integration window ΔT into two-channel event-frames, yielding a tensor $X \in \mathbb{R}^{H \times W \times 2}$ containing ON and OFF event counts. This representation directly reflects the simulated sensor output and is used to evaluate structural utility.

For the Adversarial Reconstruction Attack (Path B), the event stream is retained in its native HDF5 format (`events.h5`). E2VID processes the continuous event packets sequentially using a recurrent ConvLSTM-based architecture to reconstruct grayscale image sequences, preserving the temporal information required for identity inference.

2.5 Face Re-Identification Backbone

The face recognition backbone used by the attacker is a deep embedding network (FaceNet [13]) pre-trained on large RGB face datasets (e.g., VGGFace2). Crucially, the model is kept frozen throughout all experiments, and its weights are not fine-tuned or adapted to the event domain.

This constraint is central to the methodology: the face recognition network is not treated as a variable system under test, but rather as a standardized measurement tool. Because the embedding space is optimized strictly to differentiate human identities based on clean facial geometry, its downstream retrieval performance serves as an indicator of information leakage in the sensor data. If a physical parameter change alters the data stream such that E2VID can only reconstruct strongly degraded features, the frozen backbone will fail to compute a matching embedding vector, indicating effective hardware-level privacy protection.

2.6 Evaluation Scope and Threat Model

The evaluation framework assumes an identity-labeled video dataset organized specifically for cross-domain biometric retrieval. Instead of conventional training, validation, and test splits, the data is partitioned into a clean baseline *Gallery* and a parameter-altered *Query* set. The Gallery represents a pre-existing enrollment or watchlist database, while the Query set contains samples generated under different sensor configurations. This setup directly models a real-world re-identification attack scenario, where an adversary attempts to match event-derived observations against a high-quality reference database using a fixed biometric recognition system.

3 Experimental Setup

3.1 Datasets and Identities

Experiments are conducted using a curated, identity-disjoint subset of the VoxCeleb2 dataset [14; 15]. To ensure a controlled evaluation setting, the dataset is instantiated on a fixed subset of 300 video clips representing 51 unique celebrity identities.

The same 300 source clips are reused for every experimental configuration to ensure a controlled comparison. As a result, observed differences in the evaluation metrics can be attributed solely to the manipulated sensor parameters rather than to variations in identity, pose, lighting conditions, or scene content. Furthermore, both clip-level and identity-level annotations are preserved, allowing multi-clip evaluation while capturing variability within the same identity.

To evaluate adversarial privacy risk, standard machine learning partitions are replaced by a strict cross-domain biometric retrieval protocol. Within the 300-clip subset, one clip per identity is designated as the search probe, yielding 51 *Queries* composed of parameter-altered event reconstructions. The remaining 249 pristine baseline video tracks populate the attacker's reference *Gallery*. This configuration models a realistic privacy threat, directly measuring whether the altered sensor stream retains enough biometric structure to link a subject to a pre-existing, high-fidelity watchlist.

3.2 Data Preprocessing Pipeline

Prior to event simulation and feature extraction, raw video clips undergo a standardized preprocessing pipeline to ensure compatibility with the evaluation framework and neural architectures:

- **Bounding-Box Face Cropping:** Bounding box metadata is used to tightly crop faces within each frame, removing irrelevant background context so downstream networks evaluate pure facial geometry rather than peripheral artifacts.
- **Temporal Discretization:** Continuous, asynchronous event streams are sampled and discretized into memory-mapped NumPy arrays, creating discrete event frames compatible with standard deep learning frameworks.
- **Channel Stacking Adaptation:** Because FaceNet expects 3-channel RGB inputs while E2VID outputs single-channel grayscale images, the reconstructions are duplicated across three color channels to bridge this architectural gap.

3.3 Event-Camera Simulation and Sensor Configurations

Synthetic events are generated using the `v2e` simulator [9], starting from a standard DAVIS346 baseline configuration [16]. Secondary characteristics—such as threshold variability ($\sigma_{th} = 0.02$) and the refractory period (0.5 ms)—remain fixed throughout all experiments to isolate the effects of the primary hardware parameters [12; 8]. The physical parameters are systematically altered to test specific hypotheses regarding biometric leakage [11; 17].

For each parameter, we select a single perturbation that is both realistic for current hardware and large enough to induce a measurable change in event statistics. The baseline settings are chosen to match typical DAVIS346 operating points reported in bias-tuning work [12; 17] and manufacturer documentation [16]. The adjusted values (for example, halving the photoreceptor cut-off frequency or increasing the leak rate by roughly one order of magnitude) fall within the bias ranges used in these studies and in practical tuning guides, while keeping the computational load manageable for a multi-configuration evaluation. Table 1 summarizes the parameters and their perturbation.

The controlled parameters are:

- **Spatial Resolution:** Compares the baseline to a modern VGA-resolution configuration (e.g., DVXplorer [16]) to test whether denser spatial detail inherently increases identity leakage [8].
- **Contrast Thresholds:** Evaluates a high-sensitivity condition to determine if preserving smaller facial brightness changes aids the attacker or simply introduces redundant activity [12; 18].
- **Photoreceptor Bandwidth:** Applies a low-pass filter to test whether slower front-end dynamics successfully suppress identity-relevant micro-motions or inadvertently act as a hardware denoiser [1; 11].
- **Background Leak Rate:** Injects additional background noise to measure its efficacy as a privacy mask [19; 8].
- **Polarity Asymmetry:** Introduces an ON-dominant threshold imbalance to reflect practical hardware bias-tuning constraints [11; 17].

Sensor Parameter	Baseline	Adjusted
Spatial Resolution	346×260	640×480
Contrast Thresholds (C^+, C^-)	$(0.2,\ 0.2)$	$(0.1,\ 0.1)$
Photoreceptor Bandwidth	200 Hz	100 Hz
Background Leak Rate	0.1 Hz	5 Hz
Polarity Asymmetry (C^+, C^-)	$(0.2,\ 0.2)$	ON-Dominant $(0.15,\ 0.25)$

Table 1: Sensor parameter configurations used in the controlled simulation study.

3.4 Model Architecture

The attacker's reconstruction stage uses a pre-trained E2VID model [10], with in-memory storage (store_to_ram=True) to track sequential event volumes. The facial recognition layer uses a frozen FaceNet [13] backbone pre-trained on VGGFace2. The embedding extraction pipeline uses the same temporal sampling strategies (e.g., uniform multi-frame averaging with $n = 5$) and spatial normalization layers across both the baseline and all parameter-altered datasets, ensuring that model capacity remains fixed across the parameter sweep. Figure 3 illustrates the visual transformation from the original RGB input to the E2VID-based grayscale reconstruction used by the attacker.

Figure 3: Comparison between original RGB input and E2VID-based event reconstruction used by the adversarial retrieval pipeline.

3.5 Evaluation Metrics

To assess the impact of scaling event-camera sensor parameters, the evaluation framework splits into two evaluation paths: *Intrinsic Structural Utility* (Path A) and *Adversarial Biometric Retrieval* (Path B).

Intrinsic Structural Metrics (Signal Domain)

Path A treats the raw event-frame representations as a physical data stream and quantifies structural degradation caused by sensor-parameter changes relative to the baseline control.

- **Structural Similarity Index (SSIM):** SSIM measures the preservation of structural information between the baseline frame x and the parameter-adjusted frame y by comparing local luminance, contrast, and spatial structure [20; 21]. Unlike pixel-wise error metrics, SSIM emphasizes perceptually relevant structural differences and is therefore well suited for evaluating the preservation of facial geometry. The metric is computed as:

$$\text{SSIM}(x, y) = \frac{(2\mu_x\mu_y + c_1)(2\sigma_{xy} + c_2)}{(\mu_x^2 + \mu_y^2 + c_1)(\sigma_x^2 + \sigma_y^2 + c_2)} \quad (1)$$

where μ_x and μ_y represent the local pixel means, σ_x^2 and σ_y^2 represent the local variances, σ_{xy} is the localized cross-covariance, and c_1, c_2 are stabilizing constants used to prevent division-by-zero errors.

- **Peak Signal-to-Noise Ratio (PSNR):** PSNR assesses the logarithmic ratio between the maximum possible signal power and the Mean Squared Error (MSE) introduced by parameter scaling [20; 21]:

$$\text{PSNR} = 10 \cdot \log_{10} \left(\frac{\text{MAX}^2}{\text{MSE}} \right) \quad (2)$$

where MAX represents the maximum potential floating-point value of the normalized event tensor grid. In scenarios where a parameter configuration leaves a static video track unchanged, the baseline and altered frames are mathematically identical, yielding $\text{MSE} = 0$ and making the standard PSNR calculation infinite (∞). To preserve stable statistical aggregation and avoid skewing the dataset mean, identical pairs are assigned a standard ceiling fallback value:

$$\text{PSNR} = 60.0 \text{ dB} \quad \text{if} \quad \text{MSE} = 0 \quad (3)$$

Adversarial Retrieval Metrics (Threat Domain)
Path B evaluates biometric privacy risk through cross-domain retrieval. Retrieval performance serves as a proxy for identity leakage, with higher matching accuracy indicating that more identity-related information remains recoverable from the event-derived representations.

- **Rank-1 Accuracy:** Quantifies the probability that the closest match retrieved by the attacker's embedding comparison belongs to the true identity of the query subject, as in standard face identification and privacy-protection evaluations [21]:

$$\text{Rank-1} = \frac{1}{Q} \sum_{q=1}^{Q} \mathbb{I}(\text{Rank}(q) = 1) \quad (4)$$

where Q is the total number of query clips and $\mathbb{I}$ is an indicator function returning 1 if the top match is correct, and 0 otherwise.

- **Mean Average Precision (mAP):** Evaluates the distribution and quality of the attacker's entire ranked search space across all identities. It is computed as the mean of the individual Average Precision (AP) scores across the query set [21]:

$$\text{mAP} = \frac{1}{Q} \sum_{q=1}^{Q} \text{AP}_q \quad (5)$$

Given that the evaluation dataset configuration establishes an identity-disjoint protocol where each query clip maps to one correct corresponding video track in the gallery, the individual clip precision simplifies to a direct inverse function of the true match's position:

$$\text{AP} = \frac{1}{\text{Rank of True Match}} \quad (6)$$

- **Attack Success Rate (ASR):** ASR typically measures how often an attacker successfully bypasses a predefined verification threshold in face recognition systems [21]. However, because this evaluation uses a threshold-free, closed-set identification task, an attack is considered successful when the correct identity is retrieved as the top match. Under this definition, ASR is equivalent to Rank-1 Accuracy. To avoid redundancy, we report Rank-1 Accuracy as the primary measure of identification risk and omit ASR from subsequent analyses.

3.6 Implementation Procedure

Algorithm 1, provided in Appendix A, summarizes the end-to-end implementation used for each sensor configuration. The procedure keeps the source identities, gallery construction, reconstruction network, and biometric backbone fixed while varying only the event-camera parameters under evaluation.

4 Results

This chapter presents the empirical findings of the cross-domain biometric retrieval experiments. By evaluating the 300-clip VoxCeleb2 subset across the established hardware configurations, we quantify the precise impact of physical sensor scaling on both intrinsic data utility and adversarial privacy risk. The results are structured progressively: Section 4.1 details the macro-level metric averages, Section 4.2 maps these configurations within the theoretical privacy-utility trade-off space, Section 4.3 isolates the asymmetric effects of individual parameter changes, and Section 4.4 examines the statistical variance across individual video tracks. Figures shown in this section are also visible in the Appendix.

4.1 Macro-Level Evaluation Results

Config.	SSIM μ	SSIM σ	PSNR μ	PSNR σ	Rank-1	mAP
baseline	1.0000	0.0000	60.0000	0.0000	0.5098	0.4743
contrast_1	0.6825	0.3455	21.5556	12.9986	0.2941	0.2483
cutoff_100	0.8073	0.3062	27.9284	14.7623	0.5882	0.4922
leak_5	0.7337	0.3617	26.4514	14.9348	0.2157	0.1730
on_symmetry	0.5999	0.4062	20.1695	14.2704	0.5294	0.4656
resolution_640	0.5560	0.3988	19.7021	15.8493	0.0980	0.1279

Table 2: Macro-Level Evaluation Metrics showing structural utility and biometric retrieval risk across tested sensor configurations.

To establish a high-level understanding of sensor configuration impacts, this section aggregates the performance metrics across the entire query dataset. Table 2 presents the global mean structural utility (SSIM, PSNR) alongside the cross-domain retrieval risk (Rank-1 Accuracy and mAP) for the baseline and each parameter-altered configuration. In the table, μ denotes the mean value across the evaluated query clips, while σ denotes the standard deviation, which captures how much the metric varies between clips.

As shown in Table 2, the baseline configuration establishes a Rank-1 recognition rate of 50.98% alongside perfect structural utility. Adjusting the spatial resolution to 640×480 yields the most significant reduction in privacy risk, dropping

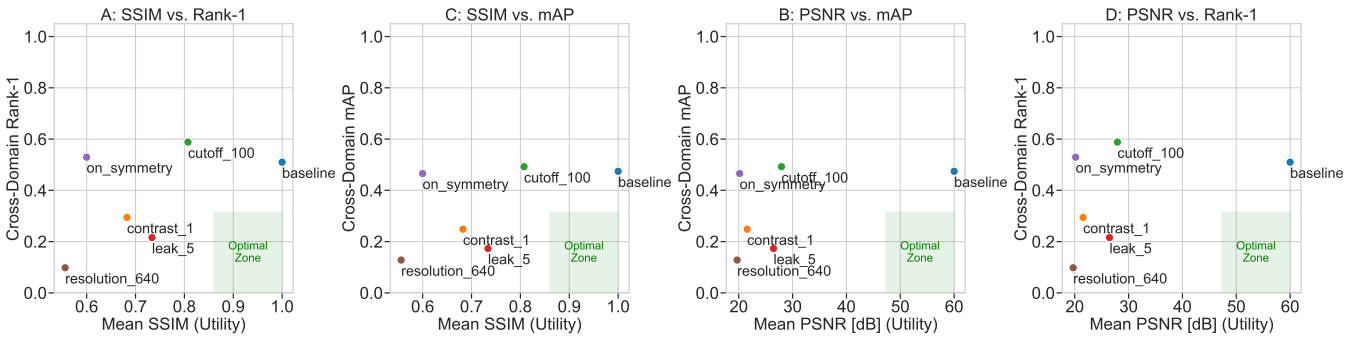

Figure 4: Privacy–utility trade-off overview across the evaluated sensor configurations.

Rank-1 accuracy to 9.80%, though this is accompanied by a severe drop in mean SSIM (0.5560). Conversely, reducing the photoreceptor bandwidth (`cutoff_100`) results in an unexpected increase in attacker success, raising Rank-1 accuracy to 58.82%. Finally, injecting background leak noise (`leak_5`) effectively reduces Rank-1 accuracy to 21.57% while maintaining a relatively moderate mean SSIM of 0.7337.

4.2 Privacy-Utility Trade-off Space

Evaluating a privacy-preserving sensor requires analyzing data utility and biometric risk simultaneously rather than in isolation. To visualize this relationship, the privacy–utility trade-off plot maps each hardware configuration as a discrete operating point across four pairings of utility and privacy metrics. The optimal hardware configuration theoretically occupies the high-utility, low-risk quadrant of these coordinate spaces.

Figure 4 maps the hardware configurations across distinct metric pairings. The `baseline` configuration consistently anchors the high-utility, high-risk regions of the plots. The `resolution_640` parameter appears clearly in the bottom-left quadrant across all charts, indicating severe structural degradation alongside minimized attacker success. Meanwhile, the `leak_5` parameter positions itself centrally, achieving a strong reduction in retrieval metrics while retaining a PSNR above 26 dB. The `cutoff_100` configuration uniquely shifts upward in risk relative to the baseline, highlighting its paradoxical effect. Conversely, `contrast_1` shifts diagonally downward, reflecting a steep, coupled loss in both utility and privacy domains. Finally, the `on_symmetry` configuration drops sharply in structural utility but maintains a high privacy risk profile, indicating a highly inefficient privacy trade-off.

4.3 Asymmetric Parameter Effects

While the absolute metrics provide an overall view of the privacy–utility trade-off, they do not clearly show the effect of individual sensor parameters. To address this, we perform a metric-delta analysis, where each metric is compared against the baseline configuration. This analysis reveals how each parameter change affects privacy and utility, helping identify configurations that achieve meaningful privacy improvements with only limited reductions in utility.

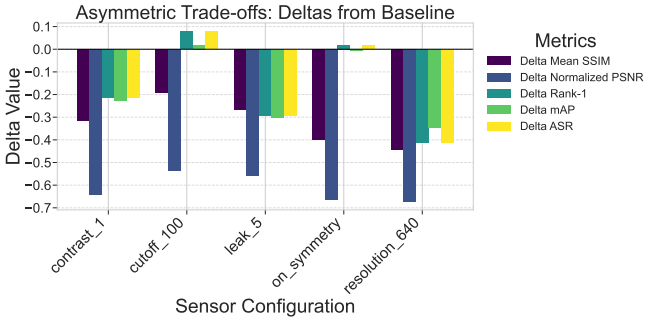

Figure 5: Metric deltas relative to the baseline sensor configuration.

The delta visualization in Figure 5 highlights the trade-offs between privacy and utility across sensor configurations. The `cutoff_100` setting increases both Rank-1 Accuracy and mAP, indicating greater identity leakage despite a small decrease in SSIM. In contrast, the `contrast_1` and `resolution_640` configurations reduce privacy risk but also cause substantial losses in utility. The `leak_5` configuration provides the most favorable trade-off, achieving a large reduction in privacy risk while only moderately affecting utility. Finally, the `on_symmetry` configuration significantly degrades utility but offers little improvement in privacy.

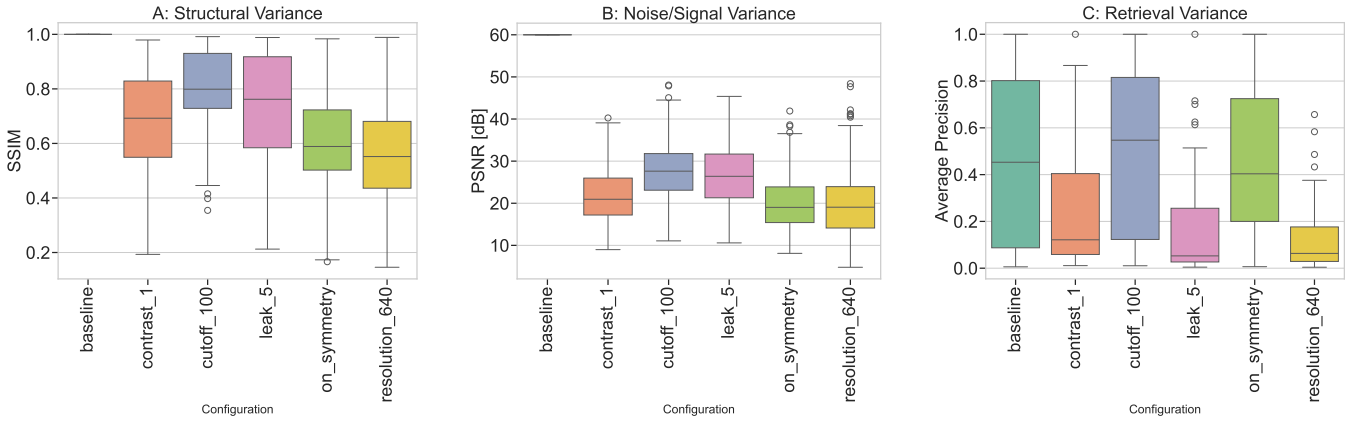

Figure 6: Clip-level metric distributions across the evaluated sensor configurations.

4.4 Clip-Level Variance and Distribution

Global averages provide an overall view of performance, but they can hide differences between individual clips caused by factors such as head movement and lighting conditions. To assess the consistency of each sensor configuration, the clip-level analysis examines the distribution, variability, and outliers of both the structural and retrieval metrics across the 51 query clips.

Figure 6 shows the distribution of structural and retrieval performance across the query dataset. The SSIM distributions for `contrast_1` and `on_symmetry` exhibit large interquartile ranges and long whiskers, indicating that their impact on structural preservation varies considerably between clips. In the retrieval distributions (Average Precision), configurations such as `leak_5` successfully lower the median performance, but several high-performing outliers remain. This suggests that some identities retain enough facial information to be correctly retrieved despite the added noise.

5 Discussion

The results of our experiments suggest that event cameras should not be assumed to be inherently private. In the baseline setting, the Rank-1 recognition rate remains relatively high, indicating that in our reconstruction and retrieval pipeline raw event streams preserve enough facial structure for a standard RGB face recognition model to identify subjects after reconstruction. However, the results also show that privacy and utility change substantially across sensor configurations, sometimes in counter-intuitive ways. The limited privacy improvement of the `on_symmetry` setting suggests that although polarity imbalance reduces signal quality, enough facial geometry remains for identity information to be recovered.

5.1 The Privacy Paradox of Temporal Bandwidth

The most unexpected finding is the behavior of the `cutoff_100` configuration. A straightforward privacy assumption would be that reducing sensor fidelity, such as capping the temporal bandwidth at 100 Hz, should weaken facial features and improve privacy. Instead, this setting increased the attacker's success rate.

This effect can be explained by the denoising behavior of temporal low-pass filtering. High-frequency background noise is suppressed, while slower facial motions, such as blinking or speaking, remain visible in the event stream. As a result, the reconstruction network (E2VID) receives a cleaner signal and produces smoother, more recognizable facial reconstructions. These observations indicate that sensor degradation does not necessarily improve privacy in our setup; in some cases, removing noise can make the attack easier.

5.2 Background Noise as an Effective Privacy Filter

In contrast to temporal bandwidth reduction, adding background leak noise through the `leak_5` configuration provides a more effective privacy intervention. Uniform background events sharply reduce attacker accuracy, while the structural utility of the signal is only moderately degraded.

This effect is linked to the way the reconstruction attacker operates. E2VID relies on temporally coherent event patterns to estimate image intensity over time. Random leak events disrupt this continuity and interfere with the recurrent memory used during reconstruction. The resulting facial images become smeared and less useful for FaceNet matching. In our experiments, controlled hardware-level noise appears to be a practical way to reduce facial recognition risk without fully destroying the scene structure.

5.3 Resolution Mismatch and the Cross-Domain Artifact

The `resolution_640` configuration produces the largest drop in attacker success. However, this result should not be interpreted as evidence that higher resolution is inherently more private. Instead, it reflects the sensitivity of the attacker model to spatial mismatch between query and gallery images.

In the evaluation setup, high-resolution query reconstructions are compared against a lower-resolution baseline gallery. Changing the sensor resolution changes the scale and alignment of reconstructed facial features. Because FaceNet

is sensitive to facial alignment, the embeddings no longer match reliably across the two domains. The observed drop in accuracy is therefore mainly an artifact of comparing mismatched resolutions, rather than a confirmed loss of biometric information. A stronger attacker with a matching high-resolution gallery could likely reduce or remove this apparent phenomenon. Given our cross-domain evaluation, resolution scaling appears less reliable as a privacy filter than background noise injection.

5.4 The Impact of Subject Movement

The clip-level analysis shows that sensor-level privacy protection is not uniform across subjects. The wide variation in the results indicates that privacy depends strongly on the motion dynamics of the captured person.

Event cameras respond to brightness changes. A subject with rapid head movement generates many events, producing a stronger signal that can remain identifiable even under stricter sensor settings. In contrast, a nearly static subject generates fewer events, making the facial structure easier to suppress. This highlights a potential limitation of fixed hardware settings in our evaluation: a configuration that protects a mostly stationary subject may still leak identity information when the subject moves more actively.

5.5 Limitations

Although the `v2e` simulator provides detailed control over sensor parameters, it cannot fully reproduce all physical effects present in real hardware, such as lens distortions, temperature-dependent bias changes, or manufacturing-level pixel variation [8]. The evaluation also relies on a frozen FaceNet model as a standardized privacy measurement tool [13]; a stronger attacker could train a custom model directly on noisy or parameter-shifted event data and potentially recover additional identity features.

A further limitation is that we only evaluate a single perturbed setting per sensor parameter, rather than a dense sweep across the full bias range. Our findings therefore map local trends between a baseline configuration and one realistic perturbation per parameter, but do not fully characterize how privacy and utility evolve across all possible values. The reported privacy–utility trade-offs should be interpreted as properties of this particular simulation and retrieval setup, rather than as definitive bounds for all event-based vision systems.

6 Conclusions and Future Work

This thesis investigated whether event-based cameras are inherently privacy-preserving by examining how identity inference varies across different sensor configurations when scene content is fixed. In our simulated setup, the baseline reconstruction and face-recognition pipeline achieved a Rank-1 identification rate of approximately 51% (Table 2), demonstrating that event streams can retain substantial biometric information. These results suggest that event cameras should not be assumed to provide inherent anonymity.

We further found that privacy risk varies considerably across sensor configurations. Reducing temporal bandwidth unexpectedly increased attacker success by acting as a hardware denoiser, whereas injecting background leak events consistently reduced identification performance while preserving moderate structural utility. Privacy effectiveness also depended strongly on subject motion, indicating that static sensor configurations cannot guarantee uniform protection. Finally, the apparent privacy gains from resolution scaling were largely attributable to cross-domain spatial mismatch rather than a true loss of identity information.

Future Work

Several directions remain for future research. First, these findings should be validated on real hardware, such as DAVIS or DVXplorer sensors. Second, privacy leakage should be evaluated against stronger event-native re-identification models rather than RGB-based attackers. Finally, future systems should investigate adaptive privacy mechanisms that dynamically adjust sensor parameters and combine hardware-level filtering with software-based anonymization techniques.

Declaration of AI Usage

During this project, artificial intelligence tools (Google Gemini, Claude, and Perplexity) were used in a supportive role for literature search, code debugging, and language polishing. All AI-generated suggestions were critically reviewed, and any included material was verified against primary sources. The research questions, experimental design, implementation decisions, and scientific interpretation of the results are my own. AI tools were not used to generate novel results or write sections without my supervision.

Responsible Research

This research studies privacy risks using datasets containing identifiable subjects. To ensure ethical compliance, evaluations strictly adhere to dataset licenses, and no raw biometric data is redistributed; only code, configuration files, and aggregate statistics are released. The methodology prioritizes reproducibility by fixing all simulation parameters (e.g., v2e version, random seeds) and freezing the attacker model prior to evaluation, preventing uncontrolled model tuning.

A core ethical challenge lies in balancing privacy with sensor utility, as extreme privacy filters can destroy the structural data required for a camera's intended tasks. While these methods could theoretically be misused to improve identity inference, our explicit objective is to identify safer hardware configurations and advocate for privacy-aware sensor design combined with explicit software anonymization.

References

[1] 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, July 2020.

[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, 2022.

[3] 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*, pages 11132–11141, 2023.

[4] Qingguo Meng, Xingbo Dong, Zhe Jin, and Massimo Tistarelli. Eventface: Event-based face recognition via structure-driven spatiotemporal modeling. *arXiv preprint arXiv:2604.06782*, 2026.

[5] Renkai Li, Xin Yuan, Wei Liu, and Xin Xu. Event-based video person re-identification via cross-modality and temporal collaboration. *arXiv preprint arXiv:2501.07296*, 2025.

[6] Xiao Wang, Qian Zhu, Shujuan Wu, Bo Jiang, and Shiliang Zhang. When person re-identification meets event camera: a benchmark dataset and an attribute-guided re-identification framework. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 40, pages 10172–10180, 2026.

[7] 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.

[8] Oisín Hageman. *A Survey on Event Camera Simulators and Datasets for Optical Flow Estimation*. Bachelor thesis, Delft University of Technology, 2024.

[9] 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*, pages 1312–1321, 2021.

[10] Henri Rebecq, Rané 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, 2019.

[11] David El-Chai Ben-Ezra, Daniel Brisk, and Adar Tal. On the theory of bias tuning in event cameras. *Available at SSRN 6520048*, 2026.

[12] Rui Graca, Brian McReynolds, and Tobi Delbruck. Optimal biasing and physical limits of dvs event noise. *arXiv preprint arXiv:2304.04019*, 2023.

[13] 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*, pages 815–823, 2015.

[14] Joon Son Chung, Arsha Nagrani, and Andrew Zisserman. Voxceleb2: Deep speaker recognition. *INTERSPEECH*, pages 1086–1090, August 2018.

[15] Arsha Nagrani, Joon Son Chung, Weidi Xie, and Andrew Zisserman. Voxceleb: Large-scale speaker verification in the wild. *Computer Speech & Language*, 60:101027, October 2019.

[16] iniVation AG. DAVIS346, DVXplorer, and DVXplorer Micro hardware documentation. Manufacturer documentation, 2025. Accessed 2026-05-18.

[17] Rui Graça, Brian McReynolds, and Tobi Delbruck. Shining light on the dvs pixel: A tutorial and discussion about biasing and optimization. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 4045–4053, 2023.

[18] Rui Graca, Sheng Zhou, Brian McReynolds, and Tobi Delbruck. Scidvs: A scientific event camera with 1.7% temporal contrast sensitivity at 0.7 lux. In *2024 IEEE European Solid-State Electronics Research Conference (ESSERC)*, pages 205–208. IEEE, 2024.

[19] Evgeny V Votyakov and Alessandro Artusi. Quantifying noise of dynamic vision sensor. *arXiv preprint arXiv:2404.01948*, 2024.

[20] 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.

[21] Xiao Yang, Yinpeng Dong, Tianyu Pang, Hang Su, Jun Zhu, Yuefeng Chen, and Hui Xue. Towards face encryption by generating adversarial identity masks. In *Proceedings of the IEEE/CVF international conference on computer vision*, pages 3897–3907, 2021.

A Implementation Procedure

Algorithm 1 Cross-domain biometric retrieval evaluation pipeline

Require: Source clips V, identity labels Y, sensor configuration set S, E2VID reconstructor R, frozen FaceNet encoder F
Ensure: Structural metrics and retrieval metrics for each sensor configuration

Select an identity-disjoint subset of 300 clips spanning 51 identities
Assign one clip per identity to the query set Q and all remaining clips to the pristine gallery G
Extract gallery embeddings $E_G \leftarrow \{F(g) : g \in G\}$
for all sensor configurations $s \in S$ **do**
    for all query clips $q \in Q$ **do**
        Simulate event stream e_q^s from q using configuration s
        Reconstruct image sequence $\hat{x}_q^s \leftarrow R(e_q^s)$
        Compute structural scores between $\hat{x}_q^s$ and the baseline representation
        Extract query embedding $z_q^s \leftarrow F(\hat{x}_q^s)$
        Rank all gallery embeddings in E_G by embedding similarity to z_q^s
        Record Rank-1 correctness and average precision for q
    end for
    Aggregate SSIM, PSNR, Rank-1 accuracy, and mAP across all queries
end for

B Enlarged Result Figures

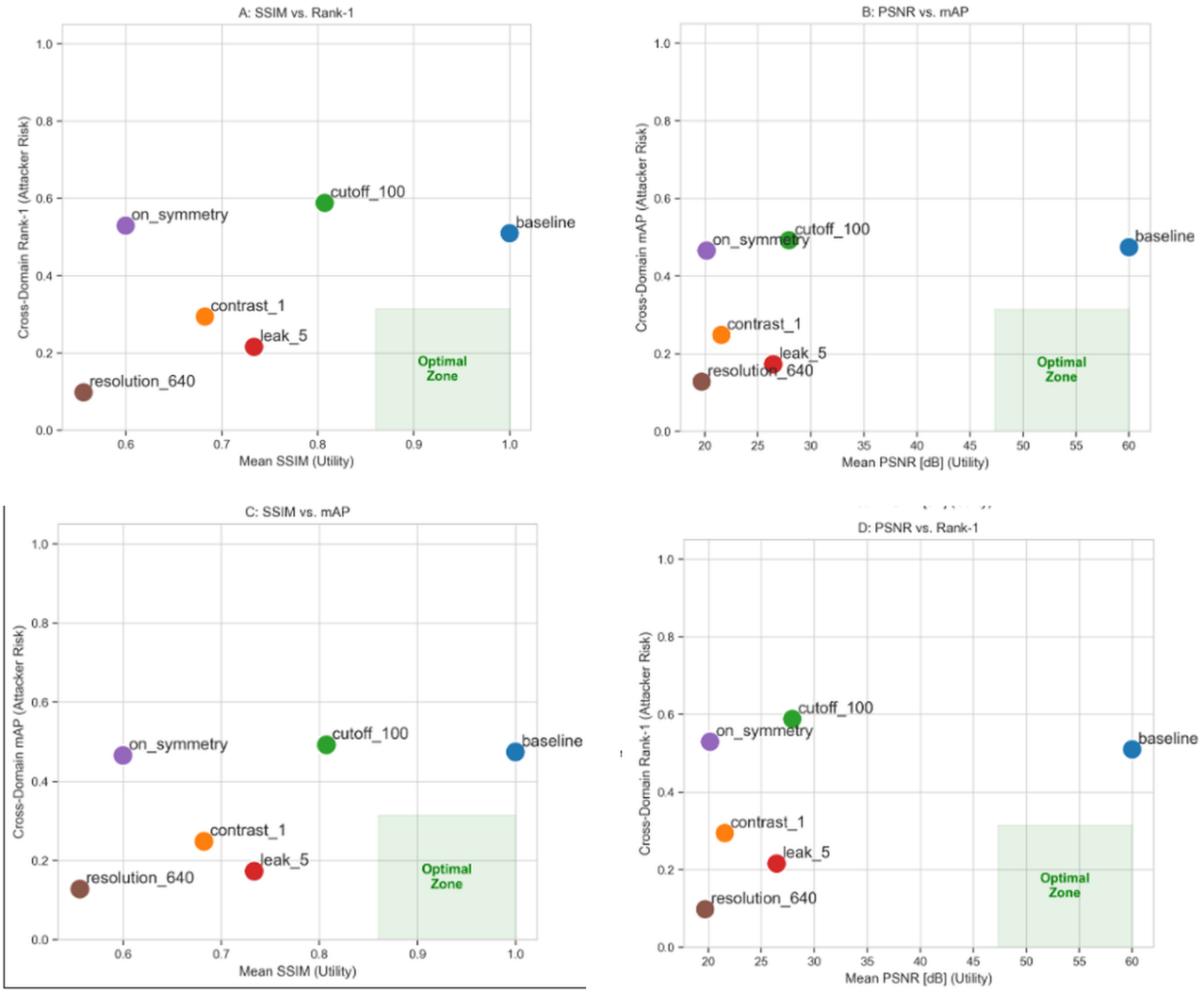

Figure 7: Enlarged privacy–utility trade-off overview across the evaluated sensor configurations.

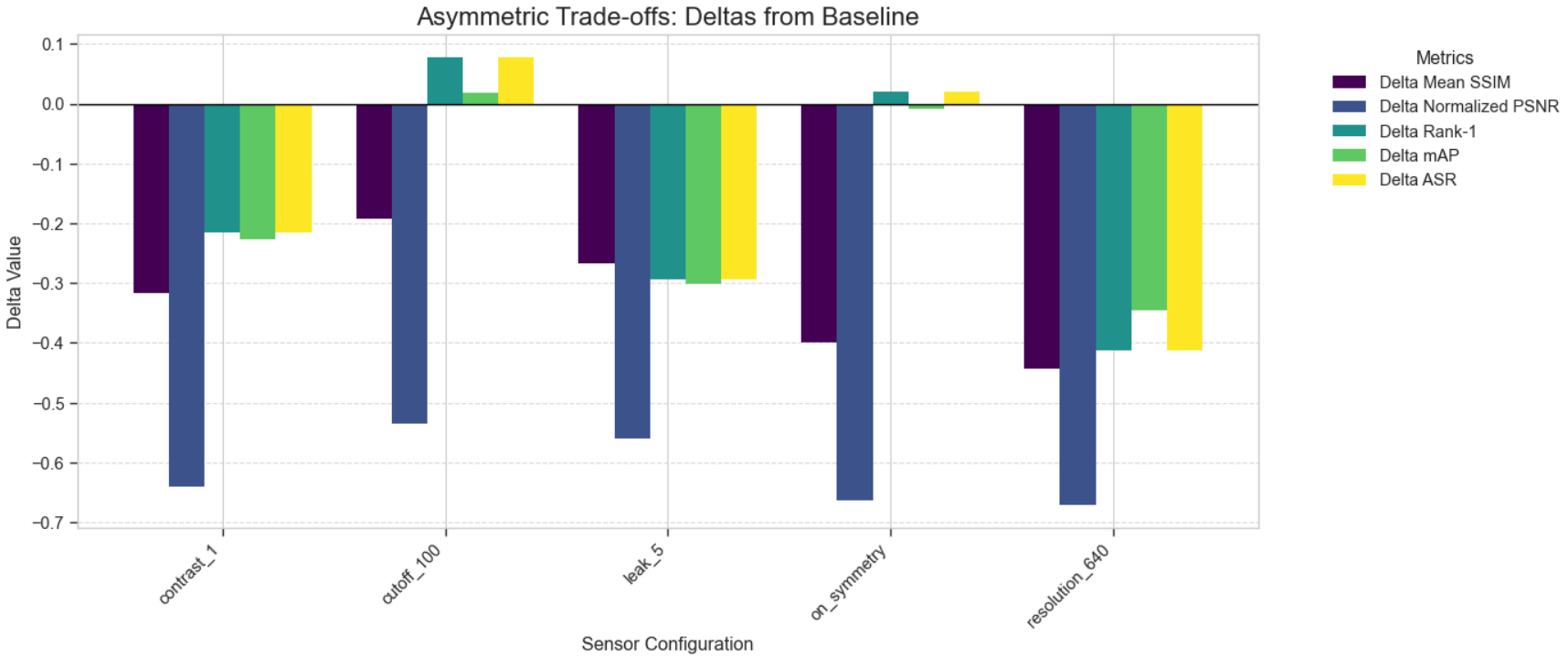

Figure 8: Enlarged metric deltas relative to the baseline sensor configuration.

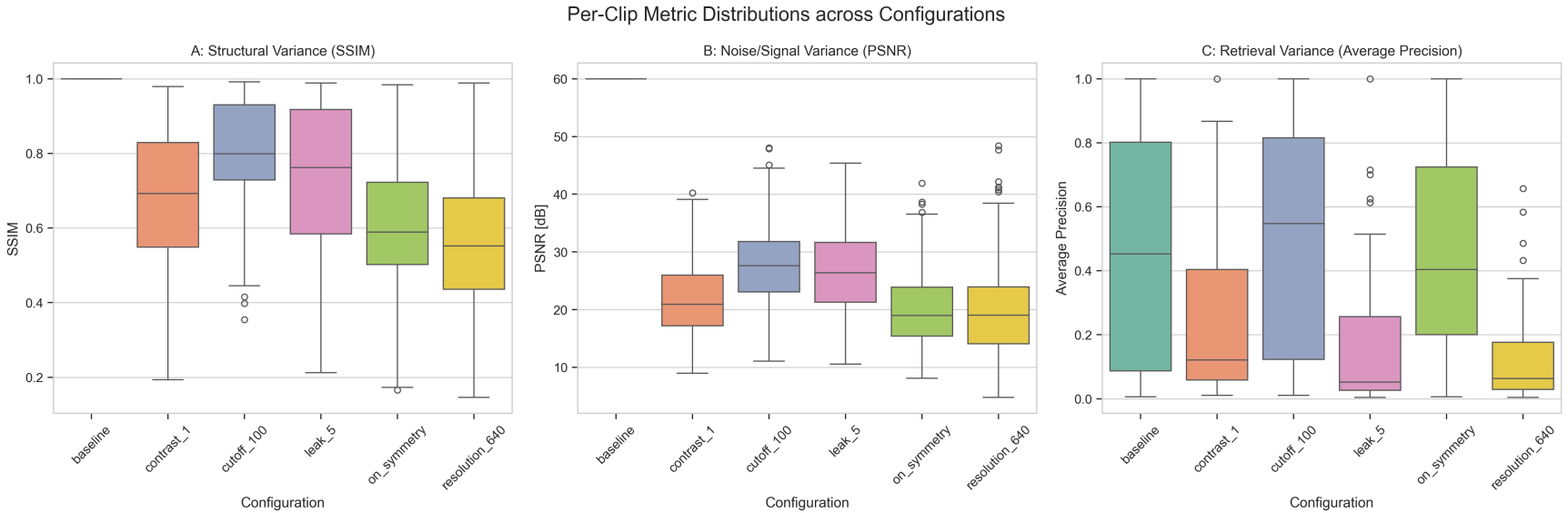

Figure 9: Enlarged clip-level metric distributions across the evaluated sensor configurations.