

UDC 004.932.2+621.397

Multi-Criteria Evaluation of Video Deblurring Algorithms for Deployment in Autonomous Robotic Systems

Liaudanskyi D. D. , Prykhodko I. O. 

National Technical University of Ukraine “Igor Sikorsky Kyiv Polytechnic Institute”, Kyiv, Ukraine

E-mail: liaudanskyidanylo@gmail.com

This paper presents a methodology for multi-criteria evaluation of video deblurring algorithms for deployment in perception pipelines of autonomous robotic systems, with a primary focus on object detection as the target downstream task. Motion blur, caused by platform vibration, rapid manoeuvres, and limited illumination, substantially degrades the accuracy of object detectors and other perception modules; a deblurring stage therefore acts as mandatory pre-processing in deployment-critical systems. A systematic comparison of six algorithms – RVRT, Restormer, PVDNet, ESTRNN, DA, and STFAN – is conducted on two datasets (GoPro and DVD) using an extended set of eight metrics that includes traditional restoration quality measures (PSNR, SSIM, LPIPS), temporal stability metrics (tOF ratio, Warping Error), and computational efficiency indicators (FPS, number of parameters). A Pareto analysis methodology in the space of objectives relevant to the deployment scenario and a weighted utility score with explicitly defined, scenario-dependent weight coefficients are proposed. RVRT achieves the highest frame-level quality on both datasets (30.42 dB on GoPro, 36.08 dB on DVD) and leads on all four quality and temporal stability metrics. For the autonomous detection scenario, PVDNet offers the best quality-throughput-compactness balance on GoPro (29.44 dB at 8.46 FPS, 5.13M parameters), while STFAN joins the Pareto front on DVD (33.11 dB at 8.18 FPS). Cross-dataset analysis reveals significant changes in the relative ranking of algorithms between datasets: STFAN, the bottom-ranked algorithm on GoPro, gains 5.57 dB on DVD and becomes a Pareto-optimal alternative, while RVRT maintains absolute leadership on both datasets. The tOF ratio is systematically lower on DVD for all algorithms, indicating more complex temporal dynamics compared to GoPro. Practical recommendations are provided for two deployment scenarios: autonomous object detection, where restoration quality and throughput are prioritised, and operator-in-the-loop control, where per-frame latency is the decisive constraint.

Keywords: video deblurring; benchmark; autonomous systems; Pareto analysis; temporal stability; cross-dataset validation; robotics

DOI: [10.64915/RADAP.2026.105.52-61](https://doi.org/10.64915/RADAP.2026.105.52-61)

Introduction

The video camera is the primary sensor for a large class of autonomous robotic systems, including unmanned ground vehicles, aerial drones, and mobile monitoring platforms. Real-world deployment of such systems occurs under conditions of platform motion, vibration, limited illumination, and rapid manoeuvres, all of which cause spatiotemporal blur in the video stream. Blur directly degrades the accuracy of the object detector – the central component of the perception pipeline – and also impairs visual odometry, segmentation, and localisation and navigation modules.

Video deblurring is therefore a natural pre-processing step to mitigate these effects; however, inserting an additional module into the pipeline introduces a trade-off between restored frame quality and computational cost. Most existing works on video

deblurring evaluate quality using only PSNR (*Peak Signal-to-Noise Ratio*) and SSIM (*Structural Similarity Index*), without accounting for temporal consistency, per-frame latency, or throughput – parameters that are critical for real deployment. A multi-criteria evaluation that reflects the technical constraints of object detection in autonomous systems is therefore necessary.

This paper proposes a methodology for selecting a video deblurring algorithm for perception pipelines of autonomous systems and validates it on two datasets – GoPro and DVD – to ensure cross-dataset generalisability of the conclusions. The main contributions are: 1) an extended set of eight metrics is employed, including tOF ratio (*temporal Optical Flow ratio*) and WarpErr (*Warping Error*); 2) a Pareto analysis is performed in the space of quality, throughput, and model

size; 3) the primary scenario is autonomous object detection in a frame stream; 4) a cross-dataset analysis is conducted, revealing a significant dependence of algorithm ranking on dataset characteristics; 5) practical algorithm-selection guidelines are provided for latency and quality constraints.

1 Problem Formulation

Let a set of video deblurring algorithms be given: $\mathcal{A} = \{a_1, \dots, a_n\}$, with a metric vector $\mathbf{m}(a)$ defined for each algorithm:

$$\mathbf{m}(a) = \{\text{PSNR}, \text{SSIM}, \text{LPIPS}, \text{tOF}, \text{tOF}_{\text{ratio}}, \text{WarpErr}, \text{FPS}, \text{Params}\},$$

where LPIPS (*Learned Perceptual Image Patch Similarity*) is a deep-feature-based perceptual similarity metric; tOF (*temporal Optical Flow*) is the mean dense-flow magnitude between consecutive frames within a sequence; FPS (*Frames Per Second*) is the number of processed frames per second; Params is the number of model parameters (millions). The goal is to select the algorithm a^* most suitable for a specific deployment scenario within given technical constraints, and to verify the robustness of this choice across different datasets.

The primary scenario considered is **Scenario A (autonomous object detection)**: video deblurring acts as a pre-processing stage in the pipeline camera $\rightarrow$ deblurring $\rightarrow$ detector $\rightarrow$ system decision. Because the decision is made by an algorithm rather than a human, the admissible per-frame latency is broader: 100...500 ms; restoration quality and throughput are the priority criteria. An additional **Scenario B (operator-in-the-loop control)** requires minimal latency (no more than 50 ms per frame) and is therefore more restrictive than Scenario A. The problem is therefore multi-criteria and cannot be reduced to optimising PSNR alone.

The general formulation is:

$$a^* = \arg \max_{a \in \mathcal{A}} U_s(a), \quad \text{s.t. } \mathbf{c}_s(a) \leq \mathbf{c}_s^{\text{max}}, \quad (1)$$

where $U_s(a)$ is the utility function for scenario s , and $\mathbf{c}_s$ is the vector of scenario constraints (primarily latency, throughput, and resource capacity).

2 Review of Related Works

Modern video deblurring architectures can be broadly categorised into three main groups. Recurrent approaches include ESTRNN [1] and PVDNet [2], which exploit the temporal context of neighbouring frames. Transformer-based approaches are represented by RVRT [3] and Restormer [4], which achieve high restoration quality. A separate direction consists of architectures with explicit motion-compensation strategies, notably STFAN [5]. This work also includes the

DA mode (*Domain Adaptation*, based on DeblurGAN-v2) [6] as a representative of batch-inference approaches whose throughput characteristics differ substantially from streaming-mode algorithms.

Traditionally, quality evaluation is based on PSNR, SSIM [7], and, less frequently, LPIPS [8]. For video sequences, however, temporal stability is equally critical. The literature applies optical-flow derivatives and warping errors for this purpose [9, 10]. In practice this matters for the stable operation of detectors, feature trackers, and localisation modules.

A separate line of work addresses deployment-oriented real-time solutions: Real-Time Video Deblurring via Lightweight Motion Compensation [11], RT-Focuser [12], and FPGA implementations for UAVs [13]. These works confirm the relevance of the applied problem but do not provide a unified multi-criteria comparison framework that simultaneously accounts for quality, temporal consistency, and computational constraints. Alongside GoPro [14], an important benchmark is DVD (*Deep Video Deblurring*) [15], which contains real motion blur from combined camera and object motion; yet a systematic cross-dataset comparison on GoPro and DVD with multi-criteria Pareto analysis has not previously been conducted.

2.1 Architectures of the Evaluated Algorithms

To facilitate the interpretation of the benchmark results, the six evaluated algorithms are grouped below into four architectural families, matching the categorisation used in Section 2: recurrent CNN, transformer, motion-compensated (filter-adaptive), and GAN. A full structural comparison — including architectural family, the temporal-modelling mechanism, core building block, inference mode, and parameter count of each algorithm — is given in Table 1.

RVRT [3] (*Recurrent Video Restoration Transformer*) combines clip-wise recurrence with guided deformable attention: overlapping temporal clips of frames are processed recurrently, and within each clip the guided deformable multi-head attention aggregates aligned spatial-temporal features. The architecture achieves state-of-the-art restoration quality at the cost of high memory footprint and low streaming throughput.

Restormer [4] is a single-frame image restoration transformer that processes each frame independently without exploiting temporal context, using Multi-Dconv-head Transposed Attention (MDTA) operating along the channel dimension rather than spatially. The absence of inter-frame information exchange results in the largest model and the lowest throughput among the evaluated algorithms.

PVDNet [2] (*Pixel Volume Deblurring Network*) is a recurrent architecture that models temporal context via pixel volumes: for each target frame, a volume of

spatially aligned candidate pixels from neighbouring frames is constructed using blur-invariant motion estimation, then refined to produce the restored frame, achieving a strong quality-throughput trade-off.

ESTRNN [1] (*Efficient Spatio-Temporal Recurrent Neural Network*) exploits a hidden-state recurrence to propagate temporal information across frames while applying a global spatio-temporal attention fusion module to select the most relevant spatial regions. Its compact architecture makes it the smallest Pareto-optimal algorithm in the evaluated set.

DA [6] (*Domain Adaptation mode of DeblurGAN-v2*) is a GAN-based deblurring algorithm built on a feature-pyramid generator with a PatchGAN discriminator used only during training. Unlike the other

algorithms in this study, DA is run in a batch inference mode at test time, processing multiple frames simultaneously, which substantially inflates its measured FPS and makes direct streaming-mode comparisons methodologically unreliable — this is a deployment-time choice, not an architectural property of the generator.

STFAN [5] (*Spatio-Temporal Filter Adaptive Network*) generates per-pixel spatially adaptive filters for two tasks: frame alignment (removing inter-frame motion) and deblurring (removing within-frame blur). Both filter sets are conditioned on the current blurry frame and the previously restored frame, providing lightweight temporal context without an explicit recurrent state.

Table 1 Structural comparison of the six evaluated video deblurring algorithms, grouped by architectural family

Alg.	Family	Temporal modelling	Core mechanism	Inference	Params, M
ESTRNN	Recurrent CNN	Hidden-state recurrence	Global spatio-temporal attention (GSA)	Streaming	2.47
PVDNet	Recurrent CNN	Pixel volume from $t-1$	Blur-invariant motion estimation (BIMNet)	Streaming	5.13
RVRT	Transformer	Clip-wise recurrence	Guided deformable attention	Streaming	13.57
Restormer	Transformer	None (per-frame)	Multi-Dconv-head transposed attn. (MDTA)	Streaming	26.13
STFAN	Motion-comp.	Filter-adaptive (uses $t-1$)	Align + deblur dynamic filters (FAC)	Streaming	5.37
DA	GAN	None (per-frame)	Feature-pyramid generator	Batch*	2.47

*Batch inference is a deployment-time throughput optimisation, not an architectural requirement of DeblurGAN-v2 (see Section 7.7)

3 Identification of Unresolved Issues

Based on the literature review, four key unresolved aspects are identified.

First, comparative studies are often focused on frame-level quality, while temporal stability metrics (tOF ratio, WarpErr) are reported sporadically or omitted entirely. Second, there is no systematic Pareto analysis in the space {quality, throughput, model size}, which complicates objective selection for real deployment. Third, results for top-quality models and real-time models are rarely presented in a common coordinate system with explicit scenario constraints. Fourth, conclusions on algorithm ranking are typically based on a single dataset (predominantly GoPro), without cross-dataset validation, which calls their transferability to other blur conditions into question.

4 Article Objectives

The aim of this paper is to develop and validate a multi-criteria methodology for selecting video deblurring algorithms for object detection pipelines in autonomous robotic systems.

To achieve this goal, the following tasks are set:

1. Construct an extended metric set combining restoration quality, temporal stability, and computational efficiency indicators, with emphasis on characteristics relevant to the downstream detection task.
2. Conduct a systematic comparison of six algorithms (RVRT, Restormer, PVDNet, ESTRNN, DA, STFAN) on two datasets — GoPro and DVD — in the context of the autonomous object detection scenario.
3. Perform a weighted utility score and Pareto analysis to reveal trade-offs between restoration quality, throughput, and model size.
4. Conduct a cross-dataset analysis of algorithm ranking stability and identify dataset-dependent patterns.
5. Provide practical recommendations for the autonomous detection scenario (Scenario A) and additionally for the operator-in-the-loop scenario (Scenario B).

5 Research Methodology

5.1 Metric Set and Justification

The evaluation framework employs eight metrics, grouped into three categories. Restoration quality is characterised by PSNR, SSIM [7], and LPIPS [8]. Temporal stability is assessed via tOF, tOF ratio, and WarpErr. Computational efficiency is expressed through FPS and the number of model parameters (Params, millions).

All optical-flow-based metrics are computed with the Farnebäck dense optical flow estimator [16] applied to grayscale frames (`cv2.calcOpticalFlowFarneback`, parameters: `pyr_scale=0.5`, `levels=3`, `winsize=15`, `iterations=3`, `poly_n=5`, `poly_sigma=1.2`). The scalar tOF for a sequence $S = \{I_1, \dots, I_T\}$ is the mean per-pixel flow magnitude averaged over all consecutive frame pairs:

$$\text{tOF}(S) = \frac{1}{T-1} \sum_{t=1}^{T-1} \frac{1}{HW} \sum_{h,w} \|\mathbf{v}_t(h, w)\|_2, \quad (2)$$

where $\mathbf{v}_t(h, w) \in \mathbb{R}^2$ is the dense-flow vector at pixel (h, w) between frames I_t and I_{t+1} , and $H \times W$ is the frame resolution.

The normalised temporal stability indicator tOF ratio is then defined as:

$$\text{tOF}_{\text{ratio}} = \frac{\text{tOF}(\hat{I})}{\text{tOF}(I^{gt})}, \quad (3)$$

where $\hat{I}$ are the restored frames and I^{gt} the ground-truth frames. A value of $\text{tOF}_{\text{ratio}} \approx 1.0$ indicates that the motion energy of the restored sequence matches that of the ground truth; values below 1.0 indicate over-smoothing (temporal flattening); values above 1.0 indicate excessive inter-frame variation (flicker).

WarpErr is the Mean Absolute Error (MAE) between the current restored frame and a prediction obtained by warping the previous restored frame forward using the dense optical flow estimated on the corresponding *ground-truth* pair (I_{t-1}^{gt}, I_t^{gt}) [17]:

$$\text{WarpErr} = \frac{1}{T-1} \sum_{t=1}^{T-1} |\hat{I}_t - \text{warp}(\hat{I}_{t-1}, \mathbf{v}_t^{gt})|_{\text{mean}}, \quad (4)$$

where warping uses bilinear interpolation (`cv2.remap`), pixel values are in the $[0, 255]$ range, and no occlusion masking is applied. This metric reflects the geometric consistency of the restored sequence.

5.2 Deployment Scenario Definitions

Two application scenarios are considered, with Scenario A being primary.

Scenario A (autonomous object detection) describes a typical autonomous system perception

pipeline: camera $\rightarrow$ buffer $\rightarrow$ deblurring $\rightarrow$ detector $\rightarrow$ system decision. Deblurring serves as mandatory pre-processing that improves the input frame quality for the detector. Because the decision is made algorithmically rather than by a human, the admissible per-frame latency is broader: 100...500 ms. The key algorithm selection criteria are restoration quality (PSNR, SSIM, LPIPS), temporal consistency of the sequence, and acceptable throughput.

Scenario B (operator-in-the-loop control) is an auxiliary scenario: camera $\rightarrow$ deblurring $\rightarrow$ video stream $\rightarrow$ operator $\rightarrow$ control command. Here, minimal latency (no more than 50 ms) is decisive to avoid operator disorientation. The latency threshold values for this scenario are given in Table 2.

Table 2 Latency threshold intervals for the operator-in-the-loop scenario

Latency, ms	Effect on control
< 50	Virtually imperceptible, natural control
50...100	Perceptible but acceptable
100...200	Noticeable degradation of operator action accuracy
> 200	Disorientation, high risk of erroneous decisions

5.3 Multi-Criteria Evaluation

For Scenario A the utility score accounts for restoration quality (a proxy for detector accuracy), throughput, and model compactness. The weighted utility function is:

$$U(a) = w_1 \text{PSNR}_{\text{norm}}(a) + w_2 \text{FPS}_{\text{norm}}(a) + w_3 (1 - \text{Params}_{\text{norm}}(a)), \quad (5)$$

where $w_1, w_2, w_3 \geq 0$ are scenario- and hardware-dependent weight coefficients reflecting the relative importance of restoration quality, throughput, and model compactness respectively; $\text{PSNR}_{\text{norm}}(a)$, $\text{FPS}_{\text{norm}}(a)$, $\text{Params}_{\text{norm}}(a)$ are the metric values normalised to $[0, 1]$; and $(1 - \text{Params}_{\text{norm}}(a))$ inverts the Params metric since a smaller value is preferable. The results reported in this paper correspond to equal weighting $w_1 = w_2 = w_3 = 1$; the sensitivity of the ranking to alternative weight profiles is analysed in Section 7.7. Normalisation follows the min-max scheme:

$$x_{\text{norm}}(a) = \frac{x(a) - x_{\min}}{x_{\max} - x_{\min}}. \quad (6)$$

Here $x(a)$ is the value of an arbitrary metric (PSNR, FPS, or Params) for algorithm a ; $x_{\min} = \min_{a \in \mathcal{A}} x(a)$ and $x_{\max} = \max_{a \in \mathcal{A}} x(a)$ are the minimum and maximum values of that metric over the set of algorithms $\mathcal{A}$; and $x_{\text{norm}}(a) \in [0, 1]$ is the normalised value.

Pareto dominance is defined in the standard way [18, 19]:

$$a_i \succ a_j \Leftrightarrow (\forall k : f_k(a_i) \geq f_k(a_j)) \wedge (\exists r : f_r(a_i) > f_r(a_j)), \quad (7)$$

where $a_i \succ a_j$ reads “ a_i dominates a_j ” and f_k are the objective functions stated in maximisation form: PSNR $\uparrow$, FPS $\uparrow$, $-\text{Params}$ $\uparrow$ (i.e. fewer parameters is better). The *Pareto front* is the subset of non-dominated algorithms. The “Pareto” column in Tables 3 and 4 indicates Pareto-front membership (“yes”) or non-membership (“no”) in this three-objective space.

The Pareto analysis serves as a *weight-free pre-filter*: it eliminates algorithms that are strictly worse on every objective simultaneously, without requiring the analyst to commit to specific weight values $\mathbf{w}$. On GoPro, four algorithms are Pareto-optimal (RVRT, PVDNet, ESTRNN, DA), reducing six candidates to four; on DVD, five are optimal. In both cases Restormer is eliminated as dominated by RVRT on all three criteria. Within the Pareto front, the scenario constraint vector $\mathbf{c}_s$ from Eq. (1) (latency budget, memory limit) narrows the admissible set further, and $U(a)$ with the chosen weights then provides a final ranking among the remaining candidates. This two-stage approach — Pareto filtering followed by weighted scoring — makes the selection process objective in the first stage and scenario-adaptable in the second.

5.4 Experimental Setup

Experiments were conducted on two datasets.

GoPro [14]: test split — 11 sequences, 1111 frames, spatial resolution 960×540. The dataset consists predominantly of camera-shake blur captured with a GoPro high-speed camera.

DVD (*Deep Video Deblurring*) [15]: test_std split — 10 sequences, 1000 frames, spatial resolution 720×1280. The dataset contains blur from combined camera and object motion in real video recordings; frames have a larger area than GoPro (921 600 vs 518 400 pixels).

An identical metric computation protocol was applied to all algorithms on each dataset. To reduce measurement variance, the first 10 frames of each sequence were excluded from FPS measurements (warm-up period). The DA inference mode — batch processing — was recorded separately as it affects FPS interpretation. On DVD, RVRT and PVDNet were evaluated with in-distribution weights, while the remaining four algorithms used weights trained on GoPro — a discrepancy

that constitutes a key methodological limitation of the cross-dataset comparison, discussed further in Section 7.6.

6 Results

6.1 Restoration Quality Metrics (GoPro)

According to Table 3, RVRT achieves the highest quality in terms of PSNR, SSIM, and LPIPS. Restormer ranks second in quality but has the lowest throughput and the largest model size. PVDNet delivers a balanced result: PSNR of 29.44 dB at 8.46 FPS.

6.2 Temporal Stability Metrics (GoPro)

The best tOF ratio is achieved by RVRT (0.9771), while STFAN has the lowest (0.9013). STFAN simultaneously exhibits the smallest WarpErr. This combination (low PSNR together with the smallest WarpErr) is interpreted as an over-smoothing effect: geometric consistency increases, but useful frame detail is lost. Consequently, tOF ratio and WarpErr should be used together with frame-level quality metrics.

6.3 Computational Efficiency (GoPro)

Throughput ranges from 1.17 FPS (Restormer) to 22.85 FPS (DA). Per-frame latency τ is estimated as $\tau \approx 1000/\text{FPS}$ ms: PVDNet $\tau \approx 118$ ms, ESTRNN $\tau \approx 155$ ms, STFAN $\tau \approx 65$ ms, RVRT $\tau \approx 537$ ms. These values directly determine the suitability of each algorithm for a given deployment scenario.

6.4 Utility Score and Pareto Front (GoPro)

Using Equation (5), the following ranking is obtained: DA (2.299), PVDNet (1.884), RVRT (1.562), ESTRNN (1.558), STFAN (1.535), Restormer (0.777). However, the high score of DA is driven by its batch inference mode, so this result is treated as a methodological artefact for streaming systems.

Table 3 Summary metric matrix for video deblurring algorithms on GoPro

Alg.	PSNR	SSIM	LPIPS	tOF r.	WarpErr	FPS	Params	Pareto
RVRT	30.42	0.9672	0.0660	0.9771	31.9247	1.86	13.57	yes
Restormer	29.78	0.9513	0.0910	0.9410	31.3522	1.17	26.13	no
PVDNet	29.44	0.9376	0.1279	0.9257	31.5020	8.46	5.13	yes
ESTRNN	28.44	0.9285	0.1431	0.9591	31.5004	6.47	2.47	yes
DA	28.40	0.9278	0.1415	0.9638	31.6417	22.85	2.47	yes
STFAN	27.54	0.8971	0.1651	0.9013	30.1733	15.43	5.37	no

The Pareto front for the triple (PSNR $\uparrow$, FPS $\uparrow$, Params $\downarrow$) is formed by RVRT, PVDNet, ESTRNN, and DA (Fig. 1). Restormer is dominated by RVRT on all three criteria. STFAN is dominated by ESTRNN on PSNR and Params. Fig. 2 shows clearly that PVDNet exhibits the most balanced profile across quality, throughput, and model compactness.

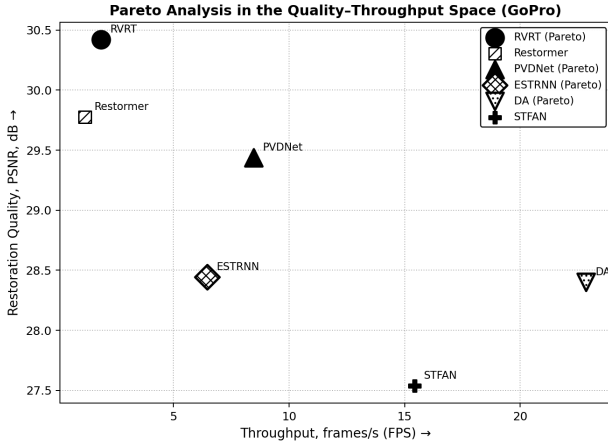


Fig. 1. PSNR–FPS diagram with highlighted Pareto-optimal algorithms (GoPro)

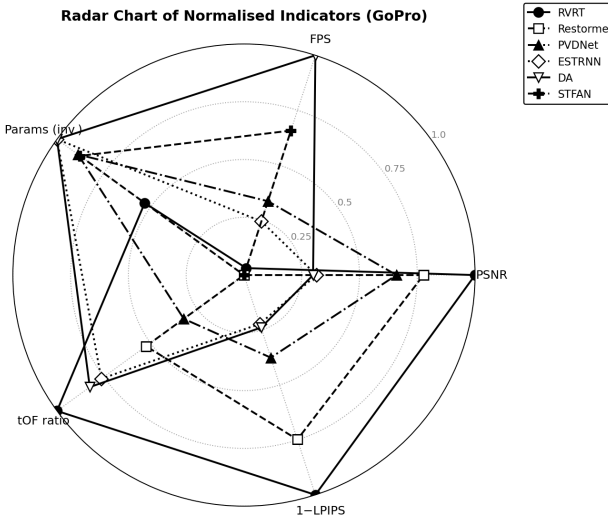


Fig. 2. Radar chart of normalised quality, throughput, and compactness indicators (GoPro)

6.5 Results on DVD

According to Table 4, RVRT retains absolute leadership on DVD (36.08 dB), exceeding the nearest competitor by 3.0 dB. A key difference from GoPro is the position of STFAN: on GoPro it was the bottom-ranked algorithm (27.54 dB, last place), whereas on DVD it reaches 33.11 dB (second place), a gain of +5.57 dB. Restormer has the lowest PSNR (30.27 dB) but exhibits the smallest WarpErr (21.13), analogous to the over-smoothing effect observed for STFAN on GoPro.

6.6 Cross-Dataset Comparison GoPro–DVD

Table 5 reveals two key patterns. First, all algorithms achieve higher PSNR on DVD, with the gain ranging from +0.49 dB (Restormer) to +5.66 dB (RVRT); this is linked to the blur characteristics of DVD, which appear better suited to certain architectures, and additionally to the use of in-distribution weights for RVRT and PVDNet. Second, the tOF ratio is *systematically lower* on DVD for all algorithms: the drop ranges from -0.013 (RVRT) to -0.183 (Restormer), indicating that DVD presents more complex temporal dynamics than GoPro, posing a greater challenge for inter-frame consistency restoration.

6.7 Utility Score and Pareto Front (DVD)

Using Equation (5) for DVD: DA (2.164), STFAN (1.991), PVDNet (1.720), RVRT (1.582), ESTRNN (1.430), Restormer (0.000). Restormer scores zero on all normalised components simultaneously (worst PSNR, lowest FPS, largest model size).

Pareto front on DVD: RVRT, STFAN, PVDNet, ESTRNN, and DA (5 algorithms; DA FPS reflects batch-inference mode). Restormer is dominated by RVRT on all three criteria. Compared with GoPro, the main change is the entry of STFAN onto the Pareto front (on GoPro it was dominated by ESTRNN on PSNR and Params). PVDNet remains on the front thanks to its smaller model size ($5.13\text{M} < 5.37\text{M}$ of STFAN), making it Pareto-incomparable with STFAN.

Table 4 Summary metric matrix for video deblurring algorithms on DVD (test_std)

Alg.	PSNR	SSIM	LPIPS	tOF r.	WarpErr	FPS	Params	Pareto
RVRT	36.08	0.9816	0.0903	0.9640	22.24	1.28	13.57	yes
STFAN	33.11	0.9641	0.1214	0.887	21.91	8.18	5.37	yes
PVDNet	33.07	0.9625	0.1559	0.845	21.83	4.89	5.13	yes
ESTRNN	31.32	0.9459	0.1521	0.898	22.47	3.67	2.47	yes
DA	31.22	0.9457	0.1494	0.905	22.60	12.70	2.47	yes
Restormer	30.27	0.9283	0.1853	0.758	21.13	0.66	26.13	no

Table 5 Cross-dataset comparison of PSNR (dB) and tOF ratio: GoPro vs DVD

Alg.	PSNR GoPro	PSNR DVD	Δ PSNR	tOF r. GoPro	tOF r. DVD
RVRT	30.42	36.08	+5.66	0.977	0.964
STFAN	27.54	33.11	+5.57	0.901	0.887
PVDNet	29.44	33.07	+3.63	0.926	0.845
DA	28.40	31.22	+2.82	0.964	0.905
ESTRNN	28.44	31.32	+2.88	0.959	0.898
Restormer	29.78	30.27	+0.49	0.941	0.758

7 Discussion

7.1 Scenario A: Autonomous Object Detection

Scenario A is the primary subject of this study: video deblurring acts as pre-processing in a detection pipeline, and the quality of the restored frame directly affects detector accuracy. Blur degrades object boundary localisation, causes false negatives, and lowers classifier confidence, especially for small or fast-moving objects. Consequently, improvements in PSNR and SSIM from deblurring are functionally significant for the downstream detection task.

On GoPro the best trade-off is offered by PVDNet (PSNR 29.44 dB, 8.46 FPS, 5.13M parameters): it provides acceptable frame quality at sufficient throughput and is Pareto-optimal. On DVD, STFAN joins the Pareto front (33.11 dB, 8.18 FPS), offering an analogous quality-throughput balance. When the hardware platform is resource-constrained, ESTRNN (2.47M parameters) is the most compact Pareto-optimal alternative, appearing on the front for both datasets. When maximum frame detail for the detector is the priority, RVRT is preferred (30.42 dB on GoPro, 36.08 dB on DVD); its low throughput is acceptable in buffered or asynchronous pipeline configurations.

7.2 Scenario B: Operator-in-the-Loop Control

For operator control, the target latency is below 50 ms (Table 2); this scenario is treated as auxiliary. Among the evaluated algorithms, none simultaneously satisfies both conditions: PSNR of at least 29 dB and latency below 50 ms. This scenario requires additional optimisations (TensorRT, half-precision, resolution reduction, or new real-time architectures).

7.3 Cross-Dataset Ranking Stability

Comparing results on GoPro and DVD reveals three important patterns.

STFAN ranking inversion. STFAN is the bottom-ranked algorithm on GoPro (PSNR 27.54 dB, last place) and is dominated by ESTRNN; on DVD it jumps to 33.11 dB (+5.57 dB) and enters the Pareto front. This may be attributed to the alignment between

STFAN’s spatially adaptive filtering mechanism and the blur characteristics of DVD (combined camera and object motion at 720p resolution). This result indicates that ranking conclusions obtained on one dataset are not automatically transferable to other conditions.

Systematic drop in tOF ratio on DVD. All six algorithms exhibit a lower tOF ratio on DVD compared to GoPro (Table 5). The smallest drop is observed for RVRT (−0.013), the largest for Restormer (−0.183). Probable causes are: (1) the more complex temporal structure of DVD (ground-truth frames obtained by averaging consecutive high-frame-rate captures at 240 fps); (2) domain shift for algorithms trained on GoPro. For detection tasks this means that temporal stability of the sequence may degrade substantially when transitioning to a different blur characteristic.

RVRT as cross-dataset leader. RVRT maintains absolute leadership in PSNR, SSIM, LPIPS, and tOF ratio on both datasets, confirming the robustness of the transformer approach to video restoration. On DVD RVRT achieves 36.08 dB — 3.0 dB above the nearest competitor (STFAN, 33.11 dB).

7.4 Over-Smoothing as a Deployment Anti-Pattern

On GoPro, STFAN exhibits a typical over-smoothing anti-pattern: the worst detail (PSNR 27.54 dB) paired with the very low WarpErr. On DVD this effect shifts to Restormer: lowest PSNR (30.27 dB), lowest tOF ratio (0.758), but smallest WarpErr (21.13). Restormer, as a frame-by-frame model without temporal context, generates geometrically accurate but temporally incoherent frames — a classic signature of sequence-level over-smoothing. Thus, algorithm suitability cannot be concluded from a single temporal metric, and over-smoothing may manifest in different algorithms depending on the dataset.

7.5 FPS and Resolution: Linear Relationship

FPS for all algorithms on DVD is lower than on GoPro proportionally to the increase in frame area. For ESTRNN, PVDNet, Restormer, and STFAN the coefficient $\text{FPS}_{\text{GoPro}}/\text{FPS}_{\text{DVD}} \approx 1.77$, which corresponds to the frame area ratio $921\,600/518\,400 \approx 1.78$. For RVRT

the deviation is larger (≈ 1.43), likely due to the non-linear dependence of the window-attention mechanism on spatial resolution. This regularity allows prediction of an algorithm's FPS when the deployment resolution changes.

7.6 Methodological Limitations

The cross-dataset comparison has the following limitations: (1) RVRT and PVDNet on DVD use in-distribution weights, while the remaining algorithms use weights trained on GoPro, which may inflate their scores; (2) DVD test_std (10 sequences) is a subset of the full DVD test (71 sequences), so results are not directly comparable with reference values from the literature; (3) FPS measurements were averaged over three repeated runs per algorithm on a single hardware configuration.

7.7 Sensitivity of the Ranking to Weight Selection

Table 6 presents the $U(a)$ ranking produced by four representative weight profiles applied to the same normalised metric values as Eq. (5). The profiles cover four practically motivated scenarios: balanced evaluation; quality-priority (GPU-equipped, buffered pipeline); throughput-priority (CPU/embedded, no accelerator); and compactness-priority (on-chip deployment with limited memory, e.g. NPU/TPU).

Two key observations follow from Table 6. First, the Pareto front (RVRT, PVDNet, ESTRNN, DA on GoPro; plus STFAN on DVD) remains unchanged across all weight profiles, confirming that Pareto dominance is a weight-free property: no choice of $\mathbf{w}$ can make Restormer non-dominated or move a Pareto-optimal algorithm off the front. Second, the $U(a)$ ranking within the front is highly sensitive to weights: RVRT leads only under quality-priority, whereas DA leads in all other profiles — primarily due to its batch-mode FPS inflating FPS_{norm} . This confirms that $U(a)$ should be reported with explicit weights and interpreted in conjunction with the Pareto front, not as a standalone score.

7.8 Limitations of the Multi-Criteria Framework

The following limitations of the proposed multi-criteria framework should be considered when applying it to autonomous robotic systems.

Sensor domain gap. Both GoPro and DVD were captured with consumer-electronics colour cameras at 8-bit sRGB dynamic range. Robotic platforms frequently employ high-sensitivity global-shutter sensors, monochrome cameras, or SWIR/thermal imagers with a substantially higher signal-to-noise ratio and 12–16 bit dynamic range. Data-driven deblurring algorithms internalise a noise model and tone-response curve tied to the training sensor, so rankings obtained on GoPro and DVD may not transfer to cameras with different sensor characteristics. Evaluation on domain-appropriate datasets is therefore an important direction for future work.

Optical-electronic channel not parameterised. The metric vector $\mathbf{m}(a)$ and the constraint vector $\mathbf{c}_s$ in Eq. (1) do not include any descriptor of the optical-electronic channel. In practice, the point-spread function (PSF), the modulation transfer function (MTF), sampling relative to the diffraction limit, and the channel SNR all define the nature and magnitude of blur [20]. Extending the framework with channel-level descriptors (e.g. MTF at the Nyquist frequency, PSF support width, signal-to-noise ratio) would allow selection recommendations to account for the specific optics of the deployed sensor, and constitutes a recommended direction for future work.

Absence of an optimal-filter reference. Deblurring algorithms are not benchmarked against an optimal linear restoration baseline such as the Wiener filter [21], which minimises the mean-square restoration error under known signal and noise spectra. Such a baseline would serve as a principled reference point for evaluating the benefit of learned, non-linear methods. Note that the Wiener transfer function depends on the channel MTF and the power spectral densities of signal and noise — parameters that are camera- and scene-specific — so the reference is not dataset-independent. Including this baseline is left as a practical extension, pending characterisation of the deployment camera.

Weight-selection subjectivity. As shown in Section 7.7, the ranked position of an algorithm inside the Pareto front can change substantially with the choice of (w_1, w_2, w_3) . In particular, the weight w_2 assigned to computational efficiency is meaningful only relative to the available hardware accelerator (GPU, NPU, or TPU with specific INT8 support, memory bandwidth, and on-chip capacity). Practitioners should recompute the sensitivity table with weights calibrated to their hardware configuration, and treat the weight-free Pareto front as the primary algorithmic selection criterion.

Table 6 Utility ranking under four weight profiles (w_1, w_2, w_3) for GoPro and DVD. The Pareto front (Pf) is invariant across all profiles

Profile	Weights	GoPro ranking	DVD ranking	Pf?
Balanced	(1, 1, 1)	DA, PVDNet, RVRT, ESTRNN, STFAN, Rst	DA, STFAN, PVDNet, RVRT, ESTRNN, Rst	unchanged
Quality	(2, 0.5, 0.5)	RVRT, PVDNet, DA, Rst, ESTRNN, STFAN	RVRT, STFAN, PVDNet, DA, ESTRNN, Rst	unchanged
Throughput	(0.5, 2, 0.5)	DA, STFAN, PVDNet, ESTRNN, RVRT, Rst	DA, STFAN, PVDNet, ESTRNN, RVRT, Rst	unchanged
Compactness	(0.5, 0.5, 2)	DA, ESTRNN, PVDNet, STFAN, RVRT, Rst	DA, STFAN, ESTRNN, PVDNet, RVRT, Rst	unchanged

7.9 Comparison with Existing Research

The obtained results are consistent with the trend in the literature: highest-quality models typically exhibit lower throughput. This work refines that statement through a scenario-driven approach, Pareto analysis, and cross-dataset validation. Compared with the real-time systems direction [11–13], this study reveals the gap between top-quality academic methods and the practical requirements of operator-in-the-loop scenarios. The cross-dataset analysis confirms the instability of single-dataset conclusions: the ranking of mid-tier algorithms (STFAN, PVDNet, ESTRNN) changes substantially between GoPro and DVD, while the leader (RVRT) and the bottom-ranked model (Restormer) retain their positions. It is also shown that batch inference mode can substantially distort FPS-based ranking, which should be explicitly documented in future benchmarking studies.

Conclusions

This paper proposes and validates a multi-criteria evaluation methodology for video deblurring algorithms for object detection pipelines in autonomous robotic systems. The methodology combines restoration quality, temporal stability, throughput, and model resource footprint indicators; its validity is confirmed on two datasets — GoPro and DVD.

The main conclusions are as follows:

1. RVRT is the unconditional leader in frame-level quality on both datasets: 30.42 dB (GoPro) and 36.08 dB (DVD). Cross-dataset leadership across PSNR, SSIM, LPIPS, and tOF ratio confirms the robustness of the transformer architecture for video restoration.
2. For the autonomous detection scenario on GoPro, the most balanced choice is PVDNet (29.44 dB, 8.46 FPS, 5.13M parameters); on DVD, STFAN joins the Pareto front (33.11 dB, 8.18 FPS) and offers an analogous balance.
3. Under platform resource constraints, the most compact Pareto-optimal alternative is ESTRNN (2.47M parameters), which appears on the front for both datasets.
4. Cross-dataset analysis reveals a ranking inversion for STFAN: the outlier on GoPro (27.54 dB, last place) rises to 33.11 dB on DVD (+5.57 dB) and enters the Pareto front, indicating dataset-dependent ranking behaviour.
5. The tOF ratio is systematically lower on DVD for all algorithms (ranging from -0.013 for RVRT to -0.183 for Restormer), indicating the more complex temporal dynamics of DVD and the

necessity of cross-dataset validation of temporal stability.

6. Restormer is dominated by RVRT in the criteria space on both datasets and does not form its own Pareto niche. On DVD, Restormer additionally exhibits critical over-smoothing (tOF ratio 0.758).
7. FPS scales linearly with frame area ($\text{FPS} \propto 1/S$) for most algorithms, enabling throughput prediction when deployment resolution changes.
8. For the operator-in-the-loop scenario (Scenario B), none of the tested models simultaneously achieves latency < 50 ms and quality $\text{PSNR} \geq 29$ dB.
9. The DA FPS result requires careful interpretation, as it is heavily dependent on the batch inference mode.
10. Methodological limitations include the use of in-distribution weights for RVRT and PVDNet on DVD, and testing on test_std (10 sequences) rather than the full test (71 sequences).
11. The multi-criteria framework has identified limitations: both evaluation datasets originate from consumer-electronics colour cameras, which may not represent the sensor characteristics of robotic deployments; the optical-electronic channel (PSF, MTF, SNR) is not parameterised; a Wiener-filter baseline is absent; and the utility-score ranking is sensitive to the choice of weight coefficients. Future extensions should address these points.
12. Promising future directions include direct evaluation of deblurring impact on detector mAP, addition of real-time models (RT-Focuser), extension to the full DVD test split, and application of the proposed multi-criteria methodology to adjacent perception tasks such as visual odometry and SLAM.

References

- [1] Zhong Z. *et al.* (2020). Efficient Spatio-Temporal Recurrent Neural Network for Video Deblurring. In: Vedaldi, A., Bischof, H., Brox, T., Frahm, JM. (eds) Computer Vision – ECCV 2020. *ECCV 2020*, Lecture Notes in Computer Science, Vol. 12351. Springer, Cham. doi: 10.1007/978-3-030-58539-6_12.
- [2] Son H. *et al.* (2021). Recurrent Video Deblurring with Blur-Invariant Motion Estimation and Pixel Volumes. *ACM Trans. Graphics (TOG)*, Vol. 40, Iss. 5, Art. No.: 185, pp. 1–18. doi: 10.1145/3453720.
- [3] Liang J. *et al.* (2022). Recurrent Video Restoration Transformer with Guided Deformable Attention. *Proc. NeurIPS*. doi: 10.48550/arXiv.2206.02146.

- [4] Zamir S. W. *et al.* (2022). Restormer: Efficient Transformer for High-Resolution Image Restoration. *Proc. CVPR*. doi: 10.48550/arXiv.2111.09881.
- [5] Zhou S. *et al.* (2019). Spatio-Temporal Filter Adaptive Network for Video Deblurring. *Proc. ICCV*. doi: 10.48550/arXiv.1904.12257.
- [6] Kupyn O., Martyniuk T., Wu J., Wang Z. (2019). DeblurGAN-v2: Deblurring (Orders-of-Magnitude) Faster and Better. *Proc. ICCV*. DOI: 10.1109/ICCV.2019.00897.
- [7] Wang Z., Bovik A. C., Sheikh H. R., Simoncelli E. P. (2004). Image Quality Assessment: From Error Visibility to Structural Similarity. *IEEE Trans. Image Processing*, Vol. 13, Iss. 4, pp. 600–612. DOI: 10.1109/TIP.2003.819861.
- [8] Zhang R., Isola P., Efros A. A., *et al.* (2018). The Unreasonable Effectiveness of Deep Features as a Perceptual Metric. *Proc. CVPR*. DOI: 10.1109/CVPR.2018.00068.
- [9] Sajjadi M. S. M., Vemulapalli R., Brown M. (2018). Frame-Recurrent Video Super-Resolution. *Proc. CVPR*. DOI: 10.1109/CVPR.2018.00693.
- [10] Lai W.-S., Huang J.-B., Ahuja N., Yang M.-H. (2016). A Comparative Study for Single Image Blind Deblurring. *Proc. CVPR*. DOI: 10.1109/CVPR.2016.188.
- [11] Son H., Lee J., Cho S., Lee S. (2022). Real-Time Video Deblurring via Lightweight Motion Compensation. *Computer Graphics Forum (Proc. Pacific Graphics)*, Vol. 41, Iss. 7, pp. 177–188. doi: 10.1111/cgf.14667.
- [12] Wu S. *et al.* (2025). RT-Focuser: A Real-Time Lightweight Model for Edge-side Image Deblurring. *Proc. ICTA*. DOI: 10.1109/ICTA68203.2025.11329854.
- [13] Cheng J. *et al.* (2025). FPGA-Based Real-Time Deblurring and Enhancement for UAV-Captured Infrared Imagery. *Remote Sensing*, Vol. 17, Iss. 20, 3446. doi: 10.3390/rs17203446.
- [14] Nah S., Kim T. H., Lee K. M. (2017). Deep Multi-Scale CNN for Dynamic Scene Deblurring. *Proc. CVPR*. DOI: 10.1109/CVPR.2017.35.
- [15] Su S., Delbracio M., Wang J., *et al.* (2017). Deep Video Deblurring for Hand-held Cameras. *Proc. CVPR*. DOI: 10.1109/CVPR.2017.33.
- [16] Farnebäck G. (2003). Two-Frame Motion Estimation Based on Polynomial Expansion. In: Bigun, J., Gustavsson, T. (eds) *Image Analysis: Proc. 13th Scandinavian Conference (SCIA)*, LNCS 2749, Vol. 2749, pp. 363–370, Springer, Berlin, Heidelberg. doi: 10.1007/3-540-45103-X_50.
- [17] Lai W.-S., Huang J.-B., Wang O., *et al.* (2018). Learning Blind Video Temporal Consistency. *Proc. ECCV*. doi: 10.48550/arXiv.1808.00449.
- [18] Miettinen K. (1999). *Nonlinear Multiobjective Optimization*. Kluwer Academic Publishers. Def. 2.2.1.
- [19] Ehrgott M. (2005). *Multicriteria Optimization*, 2nd ed. Springer. Ch. 2.
- [20] Boreman G. D. (2001). *Modulation Transfer Function in Optical and Electro-Optical Systems*. SPIE Press. PDF ISBN: 9780819480453, Print ISBN: 9780819441430. doi: 10.1117/3.419857.
- [21] Gonzalez R. C., Woods R. E. (2018). *Digital Image Processing*, 4th ed. Pearson. Ch. 5.

Мультикритеріальна оцінка алгоритмів усунення розмиття відеосигналів для розгортання в автономних роботизованих системах

Ляуданський Д. Д., Приходько І. О.

У роботі представлено методологію мультикритеріального оцінювання алгоритмів усунення розмиття відеосигналів для розгортання в перцептивних конвеєрах автономних роботизованих систем із акцентом на детекцію об'єктів як цільову підпорядковану задачу.

Розмиття зображень, спричинене вібрацією платформи, швидкими маневрами та недостатнім освітленням, суттєво знижує точність детекторів об'єктів та інших модулів сприйняття; тому усунення розмиття є обов'язковим етапом попередньої обробки у критичних для розгортання системах. Проведено систематичне порівняння шести алгоритмів – RVRT, Restormer, PVDNet, ESTRNN, DA та STFAN – на двох датасетах (GoPro та DVD) за розширеним набором з восьми метрик, що включає як традиційні показники якості відновлення (PSNR, SSIM, LPIPS), так і метрики часової стабільності (tOF ratio, Warping Error), а також показники обчислювальної ефективності (FPS, кількість параметрів).

Запропоновано методику Парето-аналізу в просторі цільових функцій, релевантних до сценарію розгортання, та зважену інтегральну оцінку з явно заданими ваговими коефіцієнтами. RVRT демонструє найвищу якість відновлення на обох датасетах (30,42 дБ на GoPro, 36,08 дБ на DVD) та лідирує за всіма чотирма показниками якості й часової стабільності. Для сценарію автономної детекції PVDNet забезпечує найкращий баланс якості–продуктивності–компактності на GoPro (29,44 дБ при 8,46 FPS, 5,13M параметрів), тоді як на DVD до Парето-фронті долучається STFAN (33,11 дБ при 8,18 FPS).

Крос-датасетний аналіз виявив суттєву зміну відносного ранжування алгоритмів між датасетами: STFAN, що є аутсайдером на GoPro, набирає 5,57 дБ на DVD та стає конкурентоспроможною Парето-оптимальною альтернативою, тоді як RVRT зберігає абсолютне лідерство на обох наборах.

Значення tOF ratio систематично нижчі на DVD для всіх алгоритмів, що свідчить про складнішу часову динаміку цього датасету порівняно з GoPro. Надано практичні рекомендації для двох сценаріїв розгортання: автономної детекції об'єктів, де пріоритетами є якість відновлення та пропускну здатність, і операторського управління, де вирішальним обмеженням є затримка обробки кадру.

Ключові слова: усунення розмиття відеосигналів; порівняльний аналіз; автономні системи; Парето-аналіз; часова стабільність; крос-датасетна валідація; робототехніка