NBFIELD: NEURAL BURST FIELD FOR FULL-RESOLUTION BLIND DEBLURRING IN THE WILD

Anonymous authors
Paper under double-blind review

ABSTRACT

Burst deblurring seeks to recover a single sharp photograph from multiple camera-shaken frames. Yet many conventional methods rely on *translation-based, spatially invariant* blur models, whereas real camera motion—from handheld photography to imaging on moving platforms—is often *rotation-dominated* and *strongly spatially varying* at high resolution. In this work, we reinterpret burst reconstruction in the domain of camera rotation, representing exposure motion in $\mathrm{SO}(3)$ and sharing a single 2D implicit neural representation across all frames. A differentiable forward model integrates the neural image along rotational homographies, while alternating optimization jointly refines the image and gyroscope-initialized motion without the need for supervised training. Across 447 in-the-wild GoPro bursts spanning severe shake blur and both natural- and low-light scenes, our method reconstructs native *12 megapixels (12MP)* images and improves PSNR by an average of 5.75 dB over the best blurry input frame, and 3.87 dB over the best baseline. Each burst is reconstructed in 2.4 minutes on a single GPU, demonstrating practical full-resolution deblurring under everyday capture conditions.

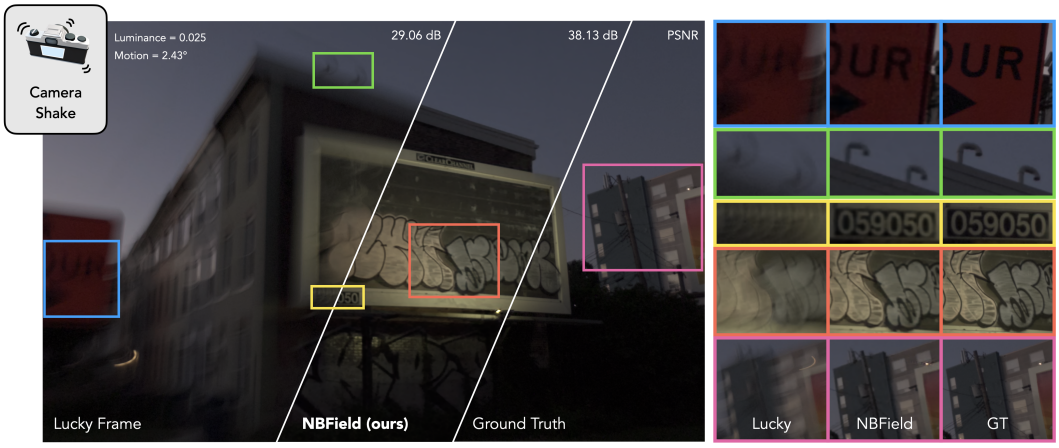

Figure 1: NBField enables full-resolution 12MP burst deblurring under challenging real-world camera shake. Even under severe motion and low light, as shown here, NBField recovers sharp structures and fine details that are heavily blurred in the burst frames. The zoomed regions highlight recovery across diverse scene content and throughout the field of view. Additional visual results are provided in Appendix F.

1 INTRODUCTION

Modern computational photography uses burst capture to reconstruct one high-quality photograph from several exposures (Hasinoff et al., 2016; Wronski et al., 2019). When camera shake blurs successive frames, however, useful scene detail is distributed unevenly across the burst rather than preserved in any single observation. We study this burst deblurring problem at native 12MP resolution (4000×3000 pixels), comparable to the resolution of modern camera systems (Apple, n.d.a; Samsung, n.d.; McHugh-Johnson, 2024)

Reviewers: please read the Reviewer Guidelines (iclr.cc/Conferences/2027/ReviewerGuidelines) and the AI Policy for Reviewers (iclr.cc/Conferences/2027/AIPolicyForReviewers). If you used AI to expand, edit, or polish your review, please provide the input text to the LLM. Better still, consider skipping the LLM and submitting your original text: we, and the authors, are much more interested in your unedited thoughts than in what an LLM has to say. AI-assisted or not, you are putting your name and reputation behind your review: LLM-generated falsehoods, hallucinations or misrepresentations are subject to disciplinary action, which may include desk-rejecting all papers you have authored.

The key inverse problem is to jointly estimate the underlying sharp image and the motion that formed each observation (Levin et al., 2009). If exposure motion is fixed or unmodeled, motion errors can be absorbed as residual blur or duplicated structure in the reconstruction, while an inaccurate image in turn makes motion estimation unreliable. Methods that estimate the image often utilize complementary information extracted from differently degraded frames (Delbracio & Sapiro, 2015a; Aittala & Durand, 2018). These methods focus on inter-frame consistency within the burst, but do not explain how motion blur forms during an individual frame. On the other hand, conventional deblurring methods that estimate motion often approximate image formation with a *spatially invariant* convolution kernel (Fergus et al., 2006; Cho & Lee, 2009; Krishnan et al., 2011). However, this blur model does not fully reflect the motion encountered in in-the-wild smartphone photography, where camera shake is generally dominated by three-dimensional camera *rotation* rather than translation, producing *spatially varying* blur across the image (Whyte et al., 2010; Hirsch et al., 2011). It is therefore crucial to combine a rotation-dominated blur model with burst reconstruction that exploits complementary information across frames.

We propose *Neural Burst Field (NBField)*, a framework for native-12MP burst deblurring that couples a rotational exposure model with a shared 2D implicit neural representation (INR). The INR optimizes a differentiable representation of the shared image, enabling a faithful characterization of the burst under SO(3) motion. This formulation makes burst representation and blur formation complementary components of the inverse problem. Specifically, our contributions are as follows:

- **We reformulate neural burst deblurring as a $2D/\mathrm{SO}(3)$ inverse problem** that combines rotational exposure formation with a 2D INR shared across the burst. The rotation-dominated forward model captures spatially varying camera-shake blur, while the INR provides a differentiable representation of the sharp image shared between frames. Because rotational warps are depth-independent, this formulation operates directly in the image plane without introducing expensive volumetric rendering or scene-depth estimation.
- **We design a reconstruction algorithm that jointly optimizes a shared sharp image and the motion during the burst.** We initialize the motion trajectory from gyroscope measurements and refine it by composing a Catmull–Rom spline correction on $\mathrm{SO}(3)$. The reconstruction requires neither pretraining nor explicit image or motion priors.
- **We develop a practical 12MP implementation for challenging motion and illumination conditions.** We evaluate NBField on 447 real handheld GoPro bursts, covering severe motion and low-light scenarios. Reconstruction achieves a mean PSNR improvement of 5.75 dB over the best blurry input frame and 3.87 dB over the best baseline, with an average reconstruction time of 2.4 minutes on a single GPU. These results demonstrate accurate and practical deblurring at native 12MP resolution, bridging the gap between computational deblurring methods and modern commercial camera systems.

2 Related Work

Burst reconstruction. Burst photography improves restoration by distributing complementary information across several observations. HDR+ established the practicality of aligning and merging raw bursts on mobile devices (Hasinoff et al., 2016), while Fourier Burst Accumulation (FBA) combines registered frames by favoring frequencies that are least attenuated in each observation (Delbracio & Sapiro, 2015a); local Fourier aggregation applies the same idea blockwise to accommodate spatially varying motion (Delbracio & Sapiro, 2015b). Learned burst fusion similarly exploits complementary measurements: the permutation-invariant CNN of Aittala & Durand (2018) uses shared frame encoders and order-independent pooling over an approximately aligned burst. These methods are effective at combining information across frames, but their alignment and fusion stages do not recover the physical camera trajectory integrated within each exposure. We instead explicitly model within-exposure camera motion in the SO(3) space and use readily available gyroscope measurements to initialize the motion trajectory, jointly reconstructing the sharp image and camera motion.

Gyro-assisted deblurring. Gyroscope measurements provide a direct, though imperfect, observation of camera motion, and are easily accessible in consumer imaging devices (Apple, n.d.b; Google, n.d.; Sony, n.d.; GoPro, n.d.). GyroMID integrate gyro measurements to recover rotational motion, convert that motion into spatially varying collections of locally uniform 2D blur kernels, and hold

those kernels fixed during non-blind multi-image deconvolution (Park & Levoy, 2014). More recent learning-based methods use image evidence to compensate for errors or missing components in the inertial motion estimate. GyroDeblurNet refines erroneous gyro-derived motion within a supervised single-image deblurring network (Yang et al., 2025), while GyroDVD combines gyro-derived rotational motion with image-estimated translational motion in a pretrained video-deblurring framework (Rim et al., 2026). Thus, existing gyro-assisted methods either treat sensor-derived blur as fixed or learn a restoration prior from external training data. NBField differs from these methods by treating gyro as an initialization and jointly optimizing the continuous $\mathrm{SO}(3)$ exposure trajectory and shared sharp image from the observed burst. This avoids both supervised pretraining and fixing the blur to the gyro-derived motion.

INRs for joint reconstruction. INRs are widely used for joint reconstruction of the target field and unknown acquisition parameters (Kang et al., 2024; Feng et al., 2023; Shao et al., 2024; Al-Haj Hemidi et al., 2024; Guo et al., 2026). In computer vision, BARF jointly refines a radiance field and inaccurate camera poses (Lin et al., 2021), but its measurement images are still modeled as sharp observations captured at single camera poses. BAD-NeRF, on the other hand, addresses blur explicitly by integrating NeRF renderings along estimated $\mathrm{SE}(3)$ exposure trajectories (Wang et al., 2023). However, its 3D formulation requires repeated multi-view rendering, making a direct extension to native 12MP reconstruction computationally nontrivial. Moreover, reconstructing 3D geometry introduces depth as an additional unknown, while translational parallax is often small and camera-shake blur is dominated by rotation in the wild (Whyte et al., 2010). Other INR-based methods, such as Nam et al. (2022) and SuperF (Jyhne et al., 2026), avoid 3D geometry but assume sharp input frames and perform only image fusion. Single-frame methods such as Deblur-INR jointly estimate a sharp image and a spatially invariant blur kernel for self-supervised blind deconvolution (Zhang et al., 2024). Our work significantly extends this line of research by combining INR-based joint reconstruction with rotational motion modeling, enabling reconstruction at native 12MP burst resolution.

3 Method: Neural Burst Field

3.1 Problem setup and overview

Given a burst of N blurry images $\{B_i\}_{i=0}^{N-1}$, our goal is to recover one sharp image while jointly refining the camera motion that formed the observations. We assume known frame-center timestamps t_i, exposure durations τ_i, and synchronized gyroscope angular-velocity measurements. Following the 2D/SO(3) formulation introduced in § 1, we model each blurry frame as the exposure-time integration of one shared 2D image under continuous rotational camera motion. NBField uses two coupled components: a 2D INR that represents the shared sharp image and a differentiable SO(3) exposure model that explains each blurry frame from continuous camera motion.

Real-world camera shake, particularly at standard and telephoto focal lengths, produces blur dominated by small pitch and yaw rotations (Whyte et al., 2010). Compared to camera translations whose effect scales inversely with scene depth, angular motion can induce significantly larger and nonuniform pixel displacement. Consider a rotation-only lens model, where Π^{-1} unprojects an observed pixel $\mathbf{x} \in \mathbb{R}^2$ into a unit camera ray, and Π projects a ray into reference-image coordinates. We denote the effective camera-to-reference rotation by $R_{i,\phi}(t, \mathbf{x}) \in \mathrm{SO}(3)$, parameterized by motion variables ϕ (defined in the next subsection). The corresponding sampling location is

$$\mathbf{q}_{i,\phi}(t, \mathbf{x}) = \Pi\big(R_{i,\phi}(t, \mathbf{x})\Pi^{-1}(\mathbf{x})\big)\,. \tag{1}$$

A single rotation therefore produces spatially-varying pixel displacement. Because the mapping in Eq. 1 is depth-independent, the forward model remains entirely in 2D while still representing nonuniform blur. This mapping reduces to a rotational homography for an undistorted pinhole camera with spatially constant rotation. We can also accommodate lens distortion and rolling shutter with Eq. 1. During exposure, the predicted blurry RGB value $\widehat{B}_i(\mathbf{x}) \in \mathbb{R}^3_{\geq 0}$ is obtained by averaging the underlying sharp image I_θ along the motion trajectory:

$$\widehat{B}_i(\mathbf{x}) = \frac{1}{\tau_i}\int_{\mathcal{E}_i(\mathbf{x})} I_\theta(\mathbf{q}_{i,\phi}(t, \mathbf{x}))\; dt \approx \frac{1}{Q}\sum_{q=0}^{Q-1} I_\theta(\mathbf{q}_{i,\phi}(t_{iq}(\mathbf{x}), \mathbf{x}))\,. \tag{2}$$

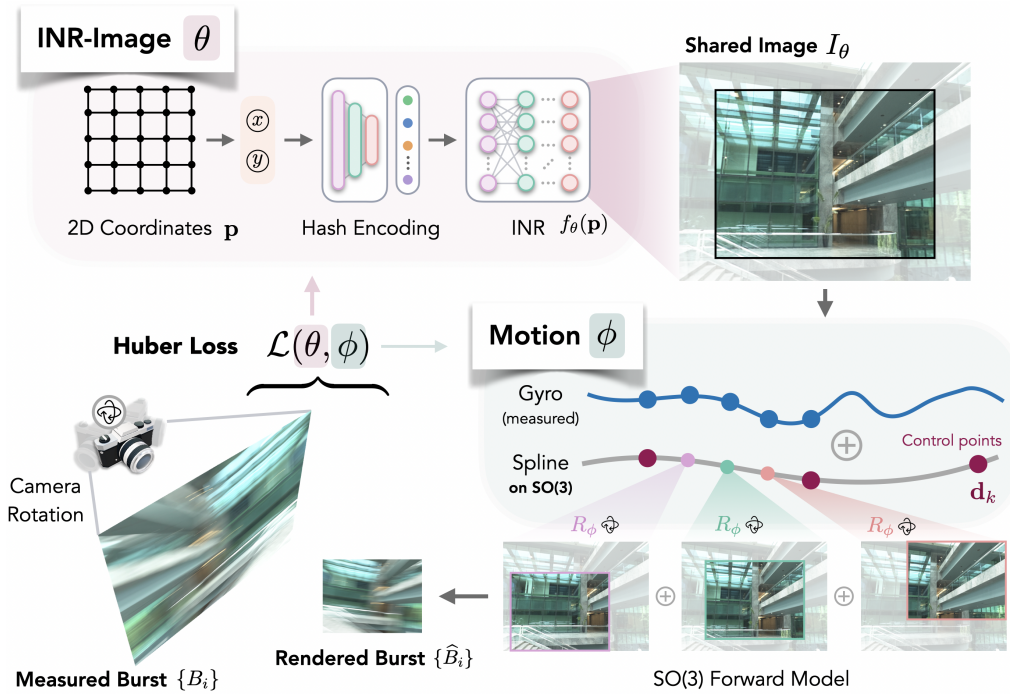

Figure 2: Conceptual overview of how *NBField* jointly reconstructs a shared sharp image and camera motion in $\mathrm{SO}(3)$ from a measured burst. The shared image I_θ is represented by a 2D INR with parameters θ using hash encoding *(INR-Image panel)*. Camera motion is initialized from gyroscope measurements and refined by a spline with parameters ϕ *(Motion panel)*. The forward model integrates I_θ along the estimated motion to render the burst $\{\hat{B}_i\}$. A Huber loss $\mathcal{L}(\theta, \phi)$ between the rendered and measured bursts jointly refines the image and motion *(Loss panel)*.

Here $\widehat{B}_i(\mathbf{x})$ depends on both θ and ϕ, $\mathcal{E}_i(\mathbf{x})$ is the calibrated exposure interval of duration τ_i, and $t_{iq}(\mathbf{x})$ are Q uniformly spaced midpoint samples within that interval. The image values and exposure average are evaluated in linear RGB. This 2D/SO(3) exposure model is the central link between the shared sharp image and the measured burst: unlike frame-level alignment, it explicitly explains how image content moves and accumulates within each observation. To optimize the shared image in this forward model, NBField parameterizes it with a continuous INR

$$I_\theta(\mathbf{p}) = f_\theta(\mathbf{p}), \qquad \mathbf{p} \in \Omega_c \subset \mathbb{R}^2, \tag{3}$$

where Ω_c is the reconstruction canvas, and $f_\theta : \Omega_c \to \mathbb{R}^3_{\geq 0}$ outputs linear RGB. INRs provide high-capacity continuous image representations and can be queried directly at arbitrary subpixel coordinates, making them a natural fit for Eq. 2 (Mildenhall et al., 2020; Sitzmann et al., 2020; Tancik et al., 2020). We use a multiresolution hash encoding with a small MLP to query subpixel locations generated by Eq. 2 because they provide efficient continuous sampling at 12MP resolution (Müller et al., 2022). The hash encoding and MLP parameters θ are optimized at test time and shared across all burst frames.

3.2 Spline-based motion refinement with gyro initialization

The exposure model above is differentiable once a continuous orientation $R_{i,\phi}(t, \mathbf{x})$ is specified. While the gyroscope provides a high-rate estimate of camera motion from angular velocities, treating it as an exact, fixed blur operator would prevent image evidence from correcting small trajectory errors that become visible at native resolution. At the other extreme, independently optimizing a rotation at every temporal sample would introduce many weakly coupled motion variables. NBField therefore uses the measured gyro trajectory for initialization and refines it with a smooth, low-dimensional spline correction. The spline connects the burst frames while maintaining continuous motion within and across exposures.

Let $G(t) \in \mathrm{SO}(3)$ denote the gyro-derived rotation that maps camera rays at time t into the reference view. We introduce rotation-vector controls $\mathbf{d}_k \in \mathbb{R}^3$ at knot times s_k and interpolate them in $\mathbb{R}^3$ with a Catmull–Rom spline (Catmull & Rom, 1974; Yuksel et al., 2011):

$$\begin{aligned} \mathbf{c}_\phi(t) &= \text{Catmull-Rom}\big(\{(s_k, \mathbf{d}_k)\}_{k=0}^{J-1}; t\big)\,, \\ A_\phi(t) &= G(t)\,\mathrm{Exp}(\mathbf{c}_\phi(t))\,, \end{aligned} \tag{4}$$

where $\mathbf{c}_\phi(t) \in \mathbb{R}^3$ is the interpolated rotation-vector correction, and $\mathrm{Exp} : \mathbb{R}^3 \to \mathrm{SO}(3)$ maps it to a correction rotation. Composing this correction with $G(t)$ gives the corrected camera-to-reference rotation $A_\phi(t) \in \mathrm{SO}(3)$. We fix $\mathbf{d}_0 = \mathbf{0}$ at the first measurement timestamp, $s_0 = t_0$, to anchor the residual motion. The remaining controls form the learnable vector $\phi = (\mathbf{d}_1^\top, \ldots, \mathbf{d}_{J-1}^\top)^\top \in \mathbb{R}^{3(J-1)}$ where J is the number of spline control vectors. The spline serves two purposes. First, it converts J corrections into a continuous residual motion that can be queried at every time sample in Eq. 2. Second, because the spline is shared between all burst frames, it couples motion refinement within and across all frames instead of estimating motion independently for each frame. We initialize all spline controls to zero, giving $A_\phi(t) = G(t)$, and keep the gyro-derived trajectory fixed while optimizing ϕ. For a global shutter, we can simplify the rotation in Eq. 1 to $R_{i,\phi}(t, \mathbf{x}) = A_\phi(t)$. NBField also supports row-dependent exposure timing with frame-center compensation, which can be found in Appendix A.

3.3 Joint image-INR and motion-spline optimization

With both image and motion parameterized, the blurry burst then provides supervision for reconstruction. NBField jointly fits the shared sharp image and motion correction by matching the rendered exposures to the measured burst. Let $B_{ic}(\mathbf{x})$ and $\widehat{B}_{ic}(\mathbf{x})$ denote channel $c \in \{1, 2, 3\}$ of the observed and predicted linear RGB values, respectively. For a set $\mathcal{S}$ of valid sampled frame–pixel pairs, we minimize the brightness-weighted Huber loss

$$\mathcal{L}(\theta, \phi; \mathcal{S}) = \frac{1}{3|\mathcal{S}|} \sum_{(i,\mathbf{x})\in\mathcal{S}} w_i(\mathbf{x}) \sum_{c=1}^{3} \rho_\delta\Big(\widehat{B}_{ic}(\mathbf{x}) - B_{ic}(\mathbf{x})\Big)\,, \tag{5}$$

where ρ_δ is the Huber penalty (Huber, 1964) with threshold δ (see Appendix A), and

$$w_i(\mathbf{x}) = \frac{1}{\max(\overline{B}_i(\mathbf{x}), 0) + \epsilon}, \qquad \overline{B}_i(\mathbf{x}) = \frac{1}{3}\sum_{c=1}^{3} B_{ic}(\mathbf{x}).$$

This weighting reduces the dominance of bright pixels in linear-light optimization and mitigates saturation artifacts, while $\epsilon > 0$ bounds the weights near black.

Because the shared image and residual motion are strongly coupled, we alternate between updating θ with motion fixed and updating ϕ with the image fixed, using the same objective in Eq. 5 for both phases. Importantly, optimization does not require explicit construction of the 12MP warped images in Eq. 2. Each NBField update instead samples only a small subset of pixels from the observed frames and queries the INR at the corresponding warped coordinates on the latent canvas.[1] Computation therefore scales with the number of sampled pixel queries rather than with a stack of full-resolution warped images.

3.4 Practical design choices

We use ten burst frames to balance fast capture with sufficient measurement information. During optimization, we progressively activate finer hash-grid levels, densify the motion-correction spline, and increase the number of exposure samples. This schedule couples finer image detail with greater motion capacity and more accurate temporal integration, while keeping all measurements at native resolution. After the final image-optimization phase, we obtain the sharp reconstruction by querying I_θ at the reference-view pixel centers. We also use a safeguard to prevent unfavorable motion updates. After each motion-update block, we evaluate the updated spline controls on a fixed monitoring

[1]The latent canvas can extend beyond the 12MP output region of interest (ROI), provided that the capture device records burst frames at a larger resolution.

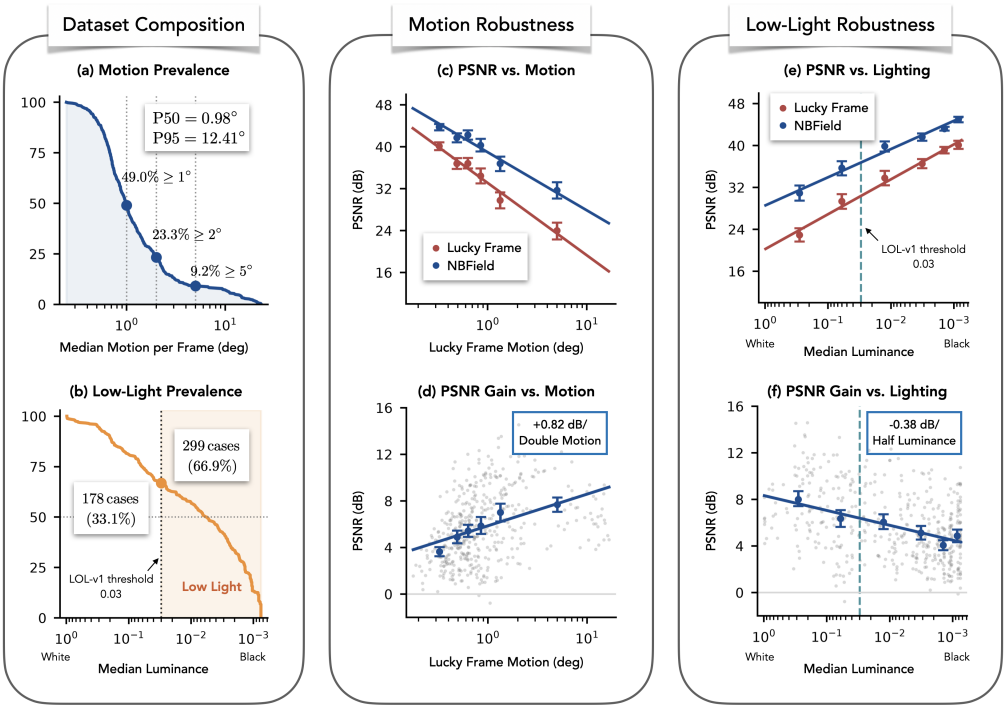

Figure 3: Analysis of dataset composition and reconstruction robustness across all 447 captures. (a–b) Distributions of within-exposure camera motion and scene luminance. (c–d) Reconstruction quality versus motion severity. (e–f) Reconstruction quality versus scene luminance. Note the prevalence of severe motion and adverse lighting cases in the dataset, as well as NBField's strong robustness across both regimes.

sample. The proposed controls are accepted only if they decrease the monitoring loss; otherwise, the previous controls are restored. Appendix A provides the shared hyperparameters, support handling, and evaluation protocol.

4 Experimental Results and Analysis

4.1 Experiment Setup

Data collection protocol. To evaluate our method at full resolution under realistic conditions, we collected an in-the-wild dataset consisting of 12MP image bursts with synchronized gyroscope measurements. The dataset contains 447 sequences collected by a handheld *GoPro Mission 1*, with ten consecutive frames per capture and 4,470 blurry frames in total. Each frame uses a 33.14 millisecond (ms) shutter. We use $4{,}000\times3{,}000$ crops from $7{,}680\times4{,}320$ captures, with a $100.0^\circ\times67.6^\circ$ field of view and a median gyroscope rate of 197.5 Hz. A full summary of the data collection configurations is provided in Appendix Table 3.

Fig. 3a–b characterize the dataset in terms of camera motion and illumination. Fig. 3a shows that the captures span a broad range of *camera motion*. The median frame accumulates 0.98° of $\mathrm{SO}(3)$ motion during the 33.14 ms exposure, while the 95th percentile reaches 12.41°, indicating a substantial tail of severe motion. At native 12MP resolution, such within-exposure motion produces pronounced spatially varying blur and presents a challenging reconstruction regime. Fig. 3b characterizes the corresponding *illumination* conditions. Of the 447 captures, 299 are classified as low light, including a substantial number of near-black scenes. The low-light threshold is set to a median luminance of 0.03 based on the luminance distribution of LOL-v1 (Wei et al., 2018). Experiments on this dataset therefore highlight the algorithmic ability to handle a broad range of camera motion and lighting conditions at high resolution.



Figure 4: Visual comparison of NBField and baselines on a daytime capture. Ground-truth residuals are shown in the bottom-left corner of each reconstruction, with enlarged regions highlighting fine-detail recovery. Note how NBField sharply resolves both the thin power lines and billboard text while leaving substantially less structured error in the residual.

Evaluation protocol. We compare NBField reconstruction with Fourier Burst Accumulation (FBA; Delbracio & Sapiro, 2015a), permutation-invariant burst CNN (PI-CNN; Aittala & Durand, 2018), gyro-based multi-image deconvolution (GyroMID; Park & Levoy, 2014), GyroDVD (Rim et al., 2026), and Deblur-INR (Zhang et al., 2024). These baselines cover classical burst fusion, supervised burst restoration, two gyro-assisted approaches, and blind single-image INR fitting. We denote the best blurry frame as Lucky Frame, selected by the best PSNR after a projective registration to the ground truth (see details in Appendix B). Reconstruction performance is evaluated using PSNR, SSIM (Wang et al., 2004), and LPIPS (Zhang et al., 2018) to assess deblurring quality, together with reconstruction time and peak GPU memory to assess computational efficiency. All experiments are run on a server equipped with Nvidia RTX PRO 6000 GPUs.

4.2 12MP Reconstruction Performance

Overall reconstruction performance. Table 1 (left panel) summarizes the numerical results of NBField and all baselines over the entire dataset. NBField achieves the best reconstruction quality across all three image-quality metrics. Notably, it reaches 39.41 dB PSNR, outperforming the best baseline by 3.87 dB and the lucky frame by 5.75 dB. The perceptual metrics follow the same trend, with NBField attaining the highest SSIM (0.95) and lowest LPIPS (0.29). While operating directly at 12MP, NBField requires only 2.4 minutes per reconstruction with 5.5 GB of peak GPU memory, which substantially narrows the computational gap between INR-based optimization and feed-forward networks.

Fig. 4 provides a visual comparison on a daytime capture. Residuals from the ground truth are shown in the bottom-left corner of each reconstruction, while enlarged regions highlight visual differences. Note how NBField preserves fine billboard text, irregular foliage, brick textures, distant power lines, large-scale building contours, and details near the image boundary. The residuals provide further evidence, with NBField leaving little recognizable scene structure while the baselines retain structured errors. Note also how shift-invariant methods degrade toward the image periphery, consistent with the spatially varying blur induced by camera rotation.

Severe-motion results. Table 1 (middle column) summarizes the numerical results under severe motion. The severe-motion sub-dataset contains $n = 112$ captures from the top quartile of lucky

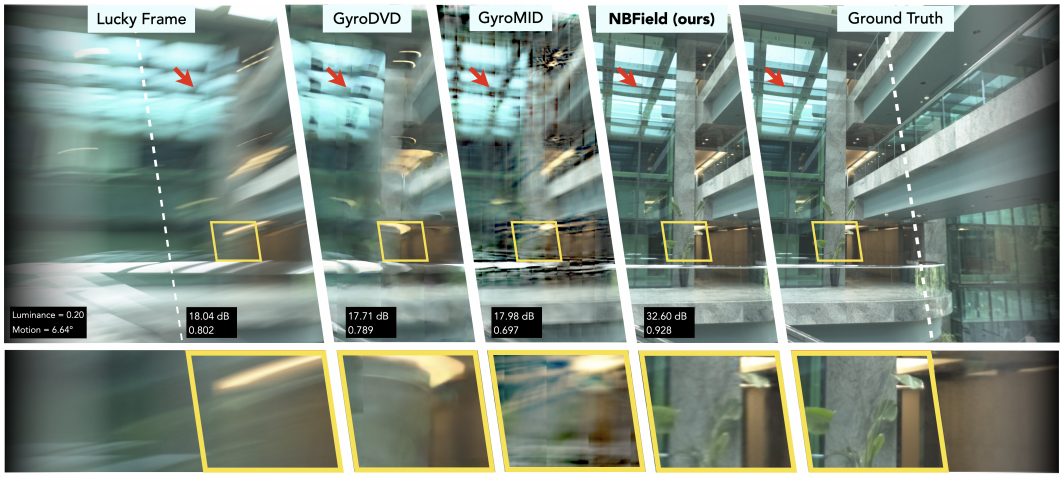

Figure 5: Visual comparison of NBField and baselines under severe camera motion. The lucky frame exhibits $6.64°$ of within-exposure motion, producing severe spatially variant blur across the image. Note how NBField recovers the plant geometry in the enlarged region while preserving reflective glass boundaries and thin railings across the field of view.

Table 1: Quantitative 12MP deblurring results across all 447 GoPro captures. PSNR is reported in dB; runtime is the mean minutes per capture, and memory is the maximum peak GPU usage in GiB. The best result is shown in **green**, and the second best is in orange.

	Overall ($n = 447$)					Severe motion ($n = 112$)			Low light ($n = 299$)		
Method	PSNR↑	SSIM↑	LPIPS↓	Time↓	Mem.↓	PSNR↑	SSIM↑	LPIPS↓	PSNR↑	SSIM↑	LPIPS↓
Lucky Frame	33.66	0.91	0.33	–	–	24.47	0.85	0.39	37.41	0.94	0.32
FBA	31.97	0.92	0.37	1.2	7.0	21.94	0.83	0.42	35.80	0.95	0.36
PI-CNN	27.02	0.73	0.38	4.6	5.5	16.74	0.66	0.47	31.81	0.73	0.36
GyroMID	31.00	0.89	0.36	**0.5**	18.3	22.77	0.79	0.43	34.71	0.92	0.34
GyroDVD	35.54	0.91	0.33	1.3	**3.0**	25.79	0.86	0.39	38.87	0.93	0.33
Deblur-INR	28.05	0.81	0.46	263.1	5.2	18.68	0.73	0.49	32.68	0.84	0.46
NBField (ours)	**39.41**	**0.95**	**0.29**	2.4	5.5	**32.26**	**0.92**	**0.30**	**42.45**	**0.97**	**0.31**

frame motion, with a median motion of $2.31°$. As motion becomes more severe, the performance gap between NBField and the competing methods widens. NBField achieves 32.26 dB PSNR, outperforming the strongest competing method by 6.47 dB and the lucky frame by 7.79 dB. Fig. 5 shows a representative example with $6.64°$ of within-exposure motion, corresponding to approximately 500 pixels of displacement at the image edge. Despite the long, spatially varying blur trajectories, NBField recovers sharp structures across opaque, reflective, and transparent surfaces. In particular, note how NBField resolves the plant geometry in the zoomed-in region, while the baseline methods smear the structure almost completely. The reflections highlighted by the arrows provide another example, which remain resolved in the NBField reconstruction while being spread into broader streaks by baselines.

Fig. 3c–d quantitatively evaluates reconstruction performance as motion severity increases across the entire dataset. Fig. 3c plots the PSNR of the lucky frame and the NBField reconstruction against best-frame motion, while Fig. 3d shows NBField's PSNR improvement over the lucky frame. Across all 447 captures, each doubling of motion magnitude is associated with an additional 0.82 dB improvement over the lucky frame. This margin increases with motion because the lucky-frame PSNR deteriorates rapidly, whereas NBField reconstruction quality degrades more gradually.

Low-light results. We lastly examine NBField under adverse lighting conditions. The low-light sub-dataset contains $n = 299$ captures, selected by ground-truth median luminance using a threshold

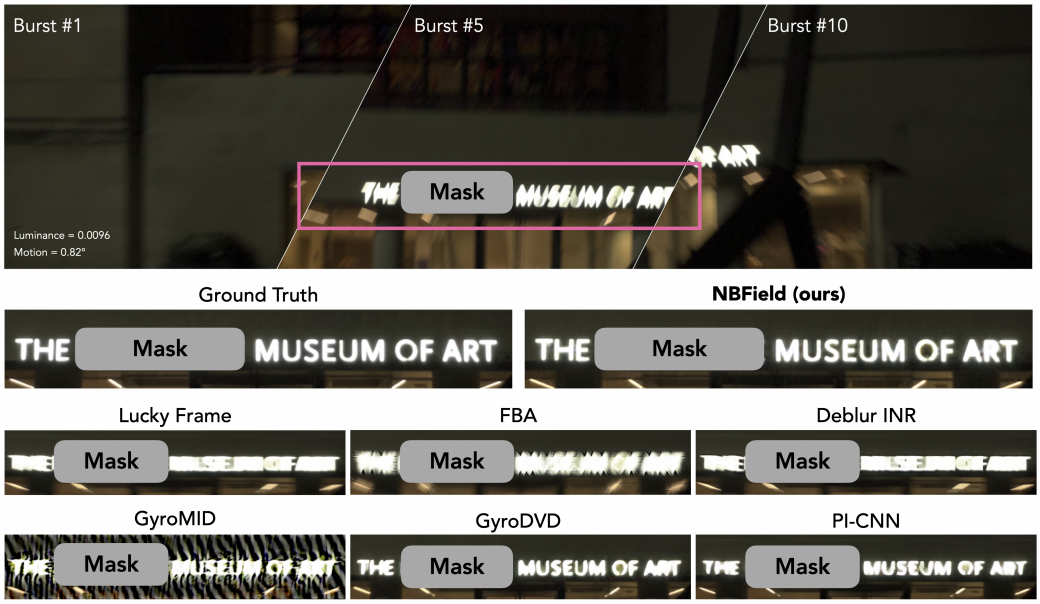

Figure 6: Visual comparison of NBField and baselines on a challenging low-light capture. The top row shows three measured burst frames, and the ground-truth median luminance is 0.0096. The enlarged region compares reconstruction of the distant front-lit sign. Note how NBField recovers sharp, distinct characters despite the limited signal, whereas baselines merge, smear, or distort the text. Sensitive information is masked.

determined from LOL-v1 (Wei et al., 2018). Table 1 (right column) summarizes the numerical results. NBField achieves 42.45 dB PSNR, exceeding the strongest competing baseline by 3.58 dB and the lucky frame by 5.04 dB. It is worth noting that the lucky frame already reaches 37.41 dB PSNR because near-black pixels dominate many captures, making overall PSNR less sensitive to errors in the sparse regions that contain visible scene detail. Fig. 6 illustrates this challenge on a representative capture, where 92.20% of the ground-truth pixels lie in shadow. Despite the limited signal, note how NBField resolves the small, distant front-lit sign into sharp, distinct characters. Baseline methods instead merge characters, diffuse their boundaries, or introduce visible artifacts. This example highlights that NBField can still recover sharp scene detail even when most of the image contains little usable signal.

Fig. 3e–f further examines how illumination affects reconstruction across the entire dataset. Fig. 3e shows that NBField consistently improves upon the lucky frame across the full luminance range, including captures that are almost entirely dark. On the other hand, Fig. 3f shows that the improvement becomes smaller toward the darkest captures. This trend partly reflects the behavior of PSNR on near-black images. As most pixels approach zero, even the lucky frame can reach nearly 45 dB, leaving limited room for further numerical improvement.

5 Conclusion

This work reframes neural burst reconstruction around $\mathrm{SO}(3)$ motion. In contrast to conventional methods that focus on translation, its forward model explicitly captures the rotation-dominated geometry characteristic of real handheld camera shake and addresses spatially varying blur. By jointly recovering a shared 2D neural image and exposure motion, NBField delivers full-resolution restoration across diverse in-the-wild scenes, including low light and severe motion. Its compact 2D formulation scales to 12MP imagery with practical computation, narrowing the gap between research-scale deblurring and contemporary camera resolutions. On the other hand, this compact 2D formulation is best suited to static scenes and rotation-dominated capture, with its assumptions becoming less accurate in near-field settings where parallax is significant. Yet these assumptions align well with practical capture conditions, where rotational shake is dominant and the full burst is captured within only $0.33\,\mathrm{s}$, substantially limiting changes in the scene.

REFERENCES

Miika Aittala and Frédo Durand. Burst image deblurring using permutation invariant convolutional neural networks. In *Proceedings of the European Conference on Computer Vision*, 2018.

Ziad Al-Haj Hemidi, Christian Weihsbach, and Mattias P. Heinrich. IM-MoCo: Self-supervised MRI motion correction using motion-guided implicit neural representations. In *Medical Image Computing and Computer Assisted Intervention 2024*, volume 15007 of *Lecture Notes in Computer Science*, pp. 382–392, 2024.

Apple. Change advanced camera settings on iPhone. Apple Support, n.d.a. URL https://support.apple.com/guide/iphone/change-advanced-camera-settings-iphb362b394e/ios. Accessed: 2026-09-20.

Apple. iPhone 16 – tech specs. Apple Support, n.d.b. URL https://support.apple.com/en-us/121029. Sensors section. Accessed September 24, 2026.

Edwin Catmull and Raphael Rom. A class of local interpolating splines. In Robert E. Barnhill and Richard F. Riesenfeld (eds.), *Computer Aided Geometric Design*, pp. 317–326. Academic Press, 1974. doi: 10.1016/B978-0-12-079050-0.50020-5.

Sunghyun Cho and Seungyong Lee. Fast motion deblurring. *ACM Transactions on Graphics*, 28 (5):145:1–145:8, 2009. doi: 10.1145/1618452.1618491.

Mauricio Delbracio and Guillermo Sapiro. Burst deblurring: Removing camera shake through Fourier burst accumulation. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pp. 2385–2393, 2015a.

Mauricio Delbracio and Guillermo Sapiro. Hand-held video deblurring via efficient Fourier aggregation. *IEEE Transactions on Computational Imaging*, 1(4):270–283, 2015b.

Brandon Y. Feng, Haiyun Guo, Mingyang Xie, Vivek Boominathan, Manoj K. Sharma, Ashok Veeraraghavan, and Christopher A. Metzler. NeuWS: Neural wavefront shaping for guidestar-free imaging through static and dynamic scattering media. *Science Advances*, 9:eadg4671, 2023. doi: 10.1126/sciadv.adg4671.

Rob Fergus, Barun Singh, Aaron Hertzmann, Sam T. Roweis, and William T. Freeman. Removing camera shake from a single photograph. *ACM Transactions on Graphics*, 25(3):787–794, 2006. doi: 10.1145/1141911.1141956.

Google. Pixel phone hardware tech specs. Pixel Phone Help, n.d. URL https://support.google.com/pixelphone/answer/7158570?hl=en. Pixel 9 sensors section. Accessed September 24, 2026.

GoPro. GPMF parser: Parser for GPMF-formatted telemetry data used within GoPro cameras. Official GoPro GitHub repository and documentation, n.d. URL https://github.com/gopro/gpmf-parser. GYRO streams and MP4 Implementation sections. Accessed September 24, 2026.

Wenhan Guo, Yuan Gao, and Yu Sun. Scorefield: Neural inverse scattering with score-based generative priors. *arXiv preprint arXiv:2608.02937*, 2026.

Samuel W Hasinoff, Dillon Sharlet, Ryan Geiss, Andrew Adams, Jonathan T Barron, Florian Kainz, Jiawen Chen, and Marc Levoy. Burst photography for high dynamic range and low-light imaging on mobile cameras. *ACM Transactions on Graphics (ToG)*, 35(6):1–12, 2016.

Michael Hirsch, Christian J. Schuler, Stefan Harmeling, and Bernhard Schölkopf. Fast removal of non-uniform camera shake. In *IEEE International Conference on Computer Vision (ICCV)*, pp. 463–470, 2011. doi: 10.1109/ICCV.2011.6126276.

Peter J. Huber. Robust estimation of a location parameter. *The Annals of Mathematical Statistics*, 35(1):73–101, 1964. doi: 10.1214/aoms/1177703732.

Sander Riisøen Jyhne, Christian Igel, Morten Goodwin, Per-Arne Andersen, Serge Belongie, and Nico Lang. SuperF: Neural implicit fields for multi-image super-resolution. In *International Conference on Learning Representations*, 2026.

Iksung Kang, Qinrong Zhang, Stella X. Yu, and Na Ji. Coordinate-based neural representations for computational adaptive optics in widefield microscopy. *Nature Machine Intelligence*, 6(6): 714–725, 2024. doi: 10.1038/s42256-024-00853-3.

Dilip Krishnan, Terence Tay, and Rob Fergus. Blind deconvolution using a normalized sparsity measure. In *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 233–240, 2011. doi: 10.1109/CVPR.2011.5995521.

Anat Levin, Yair Weiss, Fredo Durand, and William T. Freeman. Understanding and evaluating blind deconvolution algorithms. In *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 1964–1971, 2009. doi: 10.1109/CVPR.2009.5206815.

Chen-Hsuan Lin, Wei-Chiu Ma, Antonio Torralba, and Simon Lucey. BARF: Bundle-adjusting neural radiance fields. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pp. 5741–5751, 2021.

Molly McHugh-Johnson. How to get started with Pixel 8 Pro camera's Pro Controls. The Keyword, Google, April 2024. URL https://blog.google/products/pixel/how-to-use-pixel-8-camera-pro-controls/. Accessed: 2026-09-20.

Ben Mildenhall, Pratul P. Srinivasan, Matthew Tancik, Jonathan T. Barron, Ravi Ramamoorthi, and Ren Ng. NeRF: Representing scenes as neural radiance fields for view synthesis. In *Computer Vision – ECCV 2020*, pp. 405–421. Springer, 2020. doi: 10.1007/978-3-030-58452-8_24.

Thomas Müller, Alex Evans, Christoph Schied, and Alexander Keller. Instant neural graphics primitives with a multiresolution hash encoding. *ACM Transactions on Graphics*, 41(4):102:1–102:15, 2022. doi: 10.1145/3528223.3530127.

Seonghyeon Nam, Marcus A. Brubaker, and Michael S. Brown. Neural image representations for multi-image fusion and layer separation. In *Proceedings of the European Conference on Computer Vision*, pp. 216–232, 2022.

Sung Hee Park and Marc Levoy. Gyro-based multi-image deconvolution for removing handshake blur. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pp. 3366–3373, 2014.

Jaesung Rim, Woohyeok Kim, Haeyun Lee, Heemin Yang, Ke Wang, and Sunghyun Cho. Gyro-based deep video deblurring. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2026.

Samsung. Galaxy S26: Photos save at 12MP instead of 24MP. Samsung Support, n.d. URL https://www.samsung.com/us/support/troubleshoot/TSG10010450/. Accessed: 2026-09-20.

Hua-Chieh Shao, Tielige Mengke, Jie Deng, and You Zhang. 3D cine-magnetic resonance imaging using spatial and temporal implicit neural representation learning (STINR-MR). *Physics in Medicine & Biology*, 69(9):095007, 2024. doi: 10.1088/1361-6560/ad33b7.

Vincent Sitzmann, Julien N. P. Martel, Alexander W. Bergman, David B. Lindell, and Gordon Wetzstein. Implicit neural representations with periodic activation functions. In *Advances in Neural Information Processing Systems*, volume 33, 2020.

Sony. Xperia 1 specifications. Sony Support, n.d. URL https://www.sony.com/electronics/support/mobile-phones-tablets-mobile-phones/xperia-1/specifications. Sensors section. Accessed September 24, 2026.

Matthew Tancik, Pratul P. Srinivasan, Ben Mildenhall, Sara Fridovich-Keil, Nithin Raghavan, Utkarsh Singhal, Ravi Ramamoorthi, Jonathan T. Barron, and Ren Ng. Fourier features let networks learn high frequency functions in low dimensional domains. In *Advances in Neural Information Processing Systems*, volume 33, 2020.

Peng Wang, Lingzhe Zhao, Ruijie Ma, and Peidong Liu. BAD-NeRF: Bundle adjusted deblur neural radiance fields. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 4170–4179, 2023.

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. doi: 10.1109/TIP.2003.819861.

Chen Wei, Wenjing Wang, Wenhan Yang, and Jiaying Liu. Deep Retinex decomposition for low-light enhancement. In *British Machine Vision Conference (BMVC)*, 2018.

Oliver Whyte, Josef Sivic, Andrew Zisserman, and Jean Ponce. Non-uniform deblurring for shaken images. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pp. 491–498, 2010.

Bartlomiej Wronski, Ignacio Garcia-Dorado, Manfred Ernst, Damien Kelly, Michael Krainin, Chia-Kai Liang, Marc Levoy, and Peyman Milanfar. Handheld multi-frame super-resolution. *ACM Transactions on Graphics*, 38(4):28:1–28:18, July 2019. doi: 10.1145/3306346.3323024.

Heemin Yang, Jaesung Rim, Seungyong Lee, Seung-Hwan Baek, and Sunghyun Cho. Gyro-based neural single image deblurring. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 23111–23120, 2025.

Cem Yuksel, Scott Schaefer, and John Keyser. Parameterization and applications of Catmull–Rom curves. *Computer-Aided Design*, 43(7), 2011.

Richard Zhang, Phillip Isola, Alexei A. Efros, Eli Shechtman, and Oliver Wang. The unreasonable effectiveness of deep features as a perceptual metric. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 586–595, 2018.

Tianjing Zhang, Yuhui Quan, and Hui Ji. Cross-scale self-supervised blind image deblurring via implicit neural representation. In *Advances in Neural Information Processing Systems*, volume 37, 2024.