Event-based Motion Deblurring with Unpaired Data

Supplementary Material

In this supplementary material, we provide more details that are not included in the paper due to space limitations. This includes the details of the EDI-based deblurring architecture (Sec. S1), experiments on the cross-domain unpaired setting (Sec. S2), details of the confidence map (Sec. S3), effectiveness of frequency module (Sec. S4), more qualitative results (Sec. S5), and further analysis experiments (Sec. S6) respectively.

S1. Details of the EDI-based Architecture

As mentioned in Sec. 3.1, we incorporate the inverse double integral G into the blurring process (BL) so that the model receives guidance for deblurring during unsupervised learning and continues to use it at inference time. The effectiveness of this approach is demonstrated in Table 2 of the main paper. We apply this strategy with slight variations to the backbone models GEM [12], EFNet [8], and REDNet [10], and we describe these details below.

For GEM and REDNet, the overall architectures are relatively simple, and their designs are not specifically centered around fusion. Therefore, we incorporate the inverse double integral, concatenating it with the original inputs, and increasing the input channel size of the encoder accordingly. In contrast, EFNet performs fusion in a much more tightly coupled manner. Simply injecting the blurred image into the existing pipeline results in the model barely utilizing it. To address this, we keep the original architecture largely unchanged but attach an additional encoder dedicated to processing the inverse double integral. The encoded features are then fused with the encoded event features.

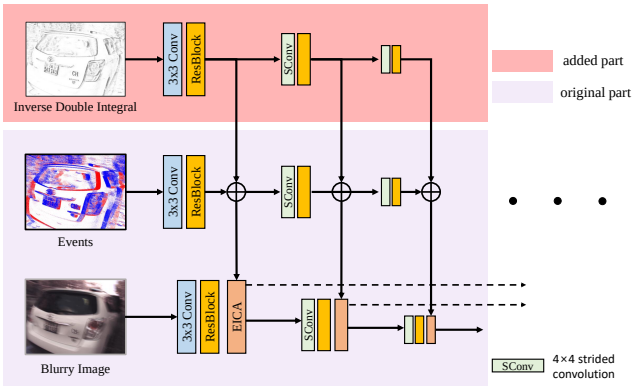

Figure S1. Details of the EDI-Based EFNet Architecture. We adopt only the encoder component and modify it for our framework, while the decoder remains identical to the original implementation. "Added part" denotes the components that we additionally incorporated into the model.

Table S1. Comparison on HighREV [9] and EventAid [2]. An experiment using datasets from different domains in an unpaired setting. **H** denotes HighREV [9], and **A** denotes EventAid [2]. † denotes cross-domain invariance via unsupervised learning or external domain knowledge. **Ev** denotes the event-based method.

		Blur: H / **Sharp:** A		**Blur:** A / **Sharp:** H	
Method	**Ev**	**PSNR** ↑	**SSIM** ↑	**PSNR** ↑	**LPIPS** ↓
CycleGAN [13]		16.08	0.440	17.16	0.619
CUT [6]		17.00	0.471	15.95	0.536
Blur2Blur† [7]		27.77	0.880	21.37	**0.135**
EDI† [5]	✓	28.49	0.839	16.67	0.395
EVDI† [11]	✓	27.36	0.834	31.05	0.293
EMP (Ours)	✓	**31.38**	**0.947**	**32.94**	0.155

As shown in Fig. S1, we illustrate the original EFNet architecture and indicate our modifications. The key difference is that we introduce a separate encoder for the inverse double integral and add its features to the event stream before fusion. We observe that this approach yields better performance than directly concatenating the inverse double integral with the blur input and feeding it to EFNet without any architectural modifications.

S2. Unpaired Setting across Different Domains

In the main paper, following prior work [1, 7], we assumed an unpaired setting in which blurred and sharp images are sampled from the *same* dataset. To explore a more practical and realistic scenario, where one may wish to reuse an existing real blur dataset, we further consider an unpaired *cross-domain* setting, in which blur and sharp data are taken from different datasets. As shown in Table S1, when the blurred images are taken from HighREV, the corresponding sharp images for training are sampled from EventAid, and evaluation is performed on the HighREV test set. Conversely, when the blurred images come from EventAid, the sharp images are drawn from HighREV, and evaluation is carried out on the EventAid test set. Even under this cross-domain unpaired configuration, the proposed method consistently achieves the best performance among unpaired learning approaches and also outperforms event-based unsupervised baselines, demonstrating that our framework is both practical and robust in heterogeneous domain settings.

S3. Details of the Confidence Map

S3.1. Motivation of the confidence map

As shown in Table 2 of the main paper, removing the self-supervised deblurring loss $\mathcal{L}_{self}$ leads to a noticeable performance drop. We attribute this to the well-known instabil-

ity of unpaired learning with generators, where the model can easily drift toward undesirable solutions or suffer from mode collapse. To prevent such a collapse, it is crucial to provide a meaningful guidance signal that encourages the generator to move in a favorable direction. In our case, we introduce an event-based self-supervised deblurring loss as this additional guidance. As shown in Fig. S4, real event streams inevitably contain noise and saturation effects, which introduce noticeable artifacts in the deblurring results from EDI. Consequently, applying a reconstruction loss uniformly over the entire image is not ideal, since unreliable regions dominated by such artifacts can adversely affect training. This motivates our confidence-map design, which selectively emphasizes reliable regions when computing the self-supervised deblurring loss $\mathcal{L}_{self}$.

S3.2. Principles of the confidence map

The goal is to show that our confidence measure is not just a heuristic, but an event-based model-driven consistency metric that evaluates how well the brightness evolution predicted by EDI agrees with the actual event firing behavior of the sensor. This confidence computation is grounded in three principles about the physics of event cameras and the structure of the EDI model.

Principle 1: Event cameras follow a strict triggering model

Let $\mathcal{S}(\mathbf{x}, t)$ denote the latent sharp-image intensity at pixel $\mathbf{x}$ and time t. By design, an event at pixel $\mathbf{x}$ is triggered when the log-intensity change crosses the contrast threshold θ:

$$\log \mathcal{S}(\mathbf{x}, t_k) - \log \mathcal{S}(\mathbf{x}, t_{k-1}) = q_k\, \theta, \tag{S1}$$

with polarity $q_k \in \{-1, +1\}$. Using the accumulated event signal $\mathcal{G}(\mathbf{x}, t, \Delta)$,

$$\mathcal{G}(\mathbf{x}, t, \Delta) = \int_t^{t+\Delta} q(\tau)\, \delta(\tau)\, d\tau, \tag{S2}$$

the log-intensity change over an interval satisfies

$$\log \mathcal{S}(\mathbf{x}, t + \Delta) - \log \mathcal{S}(\mathbf{x}, t) = \theta\, \mathcal{G}(\mathbf{x}, t, \Delta), \tag{S3}$$

and the corresponding intensity relation becomes

$$\mathcal{S}(\mathbf{x}, t + \Delta) = \mathcal{S}(\mathbf{x}, t)\, \exp\big(\theta\, \mathcal{G}(\mathbf{x}, t, \Delta)\big). \tag{S4}$$

Hence, once EDI reconstructs the latent brightness trajectory $\mathcal{S}(\mathbf{x}, t)$, the *expected number and polarity of events in each time interval* are implicitly determined through $\mathcal{G}(\mathbf{x}, t, \Delta)$. EDI thus serves as a generative model for event statistics. If the reconstructed trajectory is accurate, then the *observed event counts and polarities should agree* with these predictions. This provides a principled way to measure reconstruction reliability via model–sensor consistency.

Principle 2: Consistency must be evaluated locally in time

Event rates are highly non-uniform in time due to motion, edges, and texture. Evaluating consistency over the entire exposure would obscure local temporal mismatches. Instead, as in the main text, we divide the exposure into K bins $\{[t_{k-1}, t_k]\}_{k=1}^{K}$ and work with the per-bin event integrals, where we set $K = 17$ in our implementation:

$$\mathcal{G}_k(\mathbf{x}) = \int_{t_{k-1}}^{t_k} q(\tau)\, \delta(\tau)\, d\tau, \tag{S5}$$

$$\log \mathcal{S}(\mathbf{x}, t_k) - \log \mathcal{S}(\mathbf{x}, t_{k-1}) = \theta\, \mathcal{G}_k(\mathbf{x}). \tag{S6}$$

This bin-wise design captures (i) *when* events occur, (ii) *how many* are triggered, and (iii) with *which polarity*, all of which reflect the local temporal correctness of the reconstruction.

Principle 3: Event count and polarity are complementary

Event count and polarity capture different aspects of reconstruction errors:

- **Count consistency** measures whether the *magnitude* of the predicted brightness change is correct (*i.e.*, whether EDI predicts the right number of threshold crossings).
- **Polarity consistency** measures whether the *direction* of the brightness change is correct (*i.e.*, whether the scene is predicted to become brighter or darker in agreement with the observed event signs).

A reconstruction may predict the correct polarity but over- or underestimate the magnitude, or vice versa. Both terms are therefore necessary, and in our implementation they are combined multiplicatively into a single per-bin score (with no additional weighting, *i.e.*, both count and polarity are treated equally important).

S3.3. Formulating the Confidence Map

We now restate the confidence formulation using the notation introduced in the main EDI subsection.

S3.3.1. Predicted event statistics from EDI

Given the per-bin event accumulation $\mathcal{G}_k(\mathbf{x})$, the EDI model implies a log-intensity change

$$\log \mathcal{S}(\mathbf{x}, t_k) - \log \mathcal{S}(\mathbf{x}, t_{k-1}) = \theta\, \mathcal{G}_k(\mathbf{x}). \tag{S7}$$

Since each increment of $\pm\theta$ in log-intensity corresponds to a threshold crossing, we interpret the magnitude and sign of $\mathcal{G}_k(\mathbf{x})$ as the predicted event count and dominant polarity in bin k:

$$n_k^{\text{pred}}(\mathbf{x}) \approx \big|\mathcal{G}_k(\mathbf{x})\big|, \qquad \pi_k^{\text{pred}}(\mathbf{x}) = \operatorname{sign}\big(\mathcal{G}_k(\mathbf{x})\big). \tag{S8}$$

S3.3.2. Observed event statistics

Let $n_k^+(\mathbf{x})$ and $n_k^-(\mathbf{x})$ denote the number of observed positive and negative events at pixel $\mathbf{x}$ within bin k, respectively. We denote the total observed count by

$$n_k(\mathbf{x}) = n_k^+(\mathbf{x}) + n_k^-(\mathbf{x}), \tag{S9}$$

and define the observed positive fraction as

$$p_k^+(\mathbf{x}) = \frac{n_k^+(\mathbf{x})}{\max\big(1,\, n_k(\mathbf{x})\big)}. \tag{S10}$$

Here, $p_k^+(\mathbf{x}) \in [0, 1]$ describes the empirical polarity distribution in bin k.

S3.3.3. Per-bin consistency measures

Count consistency. We first evaluate how well the EDI-predicted event count agrees with the observed count using an exponential penalty on their absolute difference:

$$s_{\mathrm{cnt}}(k, \mathbf{x}) = \exp\big(-|\, n_k(\mathbf{x}) - n_k^{\mathrm{pred}}(\mathbf{x})\,|\big). \tag{S11}$$

This score is close to 1 when the predicted and observed counts match and decays rapidly as the discrepancy increases.

Polarity consistency. Next, we measure the agreement between the predicted polarity (from $\mathcal{G}_k(\mathbf{x})$) and the observed polarity distribution:

$$s_{\mathrm{pol}}(k, \mathbf{x}) = \begin{cases} p_k^+(\mathbf{x}), & \mathcal{G}_k(\mathbf{x}) > 0, \\ 1 - p_k^+(\mathbf{x}), & \mathcal{G}_k(\mathbf{x}) < 0, \\ \frac{1}{2}, & \mathcal{G}_k(\mathbf{x}) = 0. \end{cases} \tag{S12}$$

If the EDI model predicts positive polarity (*i.e.*, $\mathcal{G}_k > 0$), a large fraction of observed events should be positive, and similarly for negative polarity. When $\mathcal{G}_k(\mathbf{x}) = 0$ (no net predicted change), we assign a neutral value $1/2$.

S3.3.4. Aggregated per-pixel confidence

We combine the per-bin count and polarity consistencies into a single per-pixel confidence by averaging their product over all bins:

$$\mathrm{Conf}(\mathbf{x}) = \frac{1}{K} \sum_{k=1}^{K} s_{\mathrm{cnt}}(k, \mathbf{x})\, s_{\mathrm{pol}}(k, \mathbf{x}). \tag{S13}$$

Finally, if no events are observed at pixel $\mathbf{x}$ over the entire exposure, i.e.,

$$\sum_{k=1}^{K} n_k(\mathbf{x}) = 0, \tag{S14}$$

we set

$$\mathrm{Conf}(\mathbf{x}) = 1, \tag{S15}$$

to avoid penalizing static or low-contrast regions where the absence of events is fully consistent with the reconstructed brightness.

Algorithm S1 Per-pixel confidence estimation from EDI

Require: $\{\mathcal{G}_k(\mathbf{x})\}_{k=1}^{K}$ ▷ per-bin event integrals implied by $\mathcal{S}(\mathbf{x}, t)$
Require: $\{n_k^+(\mathbf{x})\}_{k=1}^{K}$, $\{n_k^-(\mathbf{x})\}_{k=1}^{K}$
Ensure: $\mathrm{Conf}(\mathbf{x})$

```
Initialize C_sum ← 0
Initialize N_tot ← 0
for k = 1 to K do
    n_k ← n_k^+(x) + n_k^-(x)
    N_tot ← N_tot + n_k
    if n_k > 0 then
        p_k^+ ← n_k^+(x) / n_k
    else
        p_k^+ ← 0.5
    end if
    n_k^pred ← |G_k(x)|
    π_k^pred ← sign(G_k(x))
    s_cnt ← exp(−|n_k − n_k^pred|)
    if π_k^pred > 0 then
        s_pol ← p_k^+
    else if π_k^pred < 0 then
        s_pol ← 1 − p_k^+
    else
        s_pol ← 0.5
    end if
    C_k ← s_cnt · s_pol
    C_sum ← C_sum + C_k
end for
if N_tot = 0 then
    return Conf(x) ← 1
else
    return Conf(x) ← (1/K) C_sum
end if
```

S3.4. Interpretation of the confidence map

The value $\mathrm{Conf}(\mathbf{x}) \in [0, 1]$ quantifies the *physical consistency* between (i) the brightness trajectory implied by the EDI model and (ii) the actual event firing behavior in terms of counts and polarities.

High confidence indicates that the EDI-predicted event statistics agree with the measured event stream at pixel $\mathbf{x}$, suggesting a reliable local reconstruction. Low confidence signals potential temporal misalignment, incorrect contrast estimation, motion boundary artifacts, or other modeling errors. In the main text, this confidence is used as a loss mask to down-weight unreliable regions during training.

S3.5. Pseudo-code Implementation

For completeness, we summarize the confidence computation in pseudo-code in Algorithm S1.

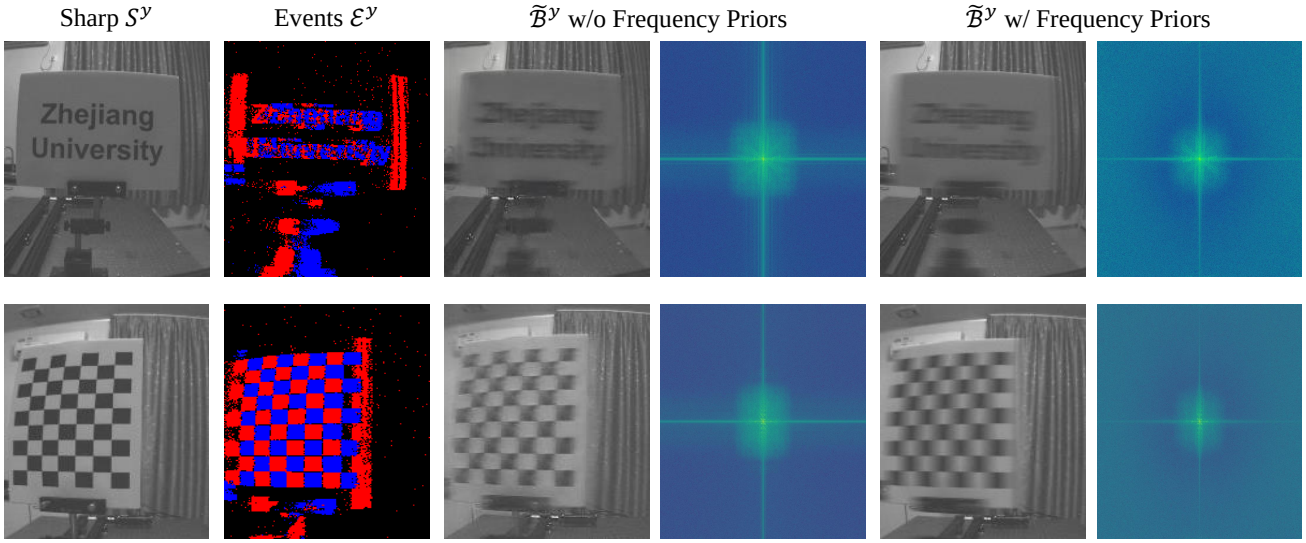

Figure S2. Visualization of FFT spectra of blur images generated by the blurring process (BL) with and without frequency priors. When the priors are applied, the spectra exhibit stronger low-frequency concentration, clearer motion-aligned energy patterns, and a more structured distribution overall, while high-frequency components are substantially suppressed. These differences highlight that the frequency prior guides the BL module to produce more physically consistent and realistic motion blur.

S4. Effectiveness of Frequency Domain

To analyze how the proposed frequency prior influences the blur images generated by the blurring process (BL), we visualize the FFT magnitude spectra of the blur images with and without the frequency prior in Figure S2. It not only reveals a noticeable difference in the quality of the generated blur images, but also a clear discrepancy in the frequency domain: the blur images produced with the proposed frequency prior exhibit stronger low-frequency concentration and more structured, motion-aligned spectral patterns, while high-frequency components are significantly suppressed compared to those generated without the prior. This demonstrates that the frequency prior guides the blurring process toward a more controlled and physically plausible blur, rather than inducing arbitrary smoothing or residual high-frequency artifacts.

S5. More Qualitative Results

The qualitative results presented in Figures 5 and 6 of the main paper correspond to applying the proposed EMP to the GEM [12] backbone. In this supplementary material, we primarily provide results based on REDNet [10] and EFNet [8] to further illustrate the effectiveness of our approach. We provide additional qualitative results in Figs. S5 and S6, where our method shows clearly superior visual quality compared to other unpaired approaches, especially those that do not use events, and also produces results that are competitive with, or even better than, existing event-based unsupervised learning methods.

S6. Additional Analysis Experiments

Sensitivity to EDI threshold θ. Following prior works [3, 4], we set $\theta = 0.2$ and analyze sensitivity by varying this value. Performance remains stable due to confidence weighting and the unpaired generation pathway.

θ	0.1	0.2	0.3	0.4	0.5
PSNR	35.32	35.49	35.43	35.16	34.96

Benefits of training at scale. As described in REBlur [8], additional data without sharp GT are provided beyond the standard splits. Leveraging this data, we perform the requested scale analysis by training with 10%, 25%, and 50% of the original training set we used, as well as expanded settings of 125% and 150%, where extra unpaired data are included. Results show consistent performance gains with more data, supporting the scalability.

Ratio (%)	10	25	50	100 (Original)	125	150
PSNR	34.42	34.87	35.01	35.49	35.61	35.84

Frequency transfer and synthetic blur analysis. As shown in Fig. S3, we fix the target sharp image $\mathcal{S}^{\mathcal{Y}}$ and vary unpaired blur references $\mathcal{B}^{\mathcal{X}}$, observing changing blur characteristics while structural content remains stable. In addition, we fix the target sharp data and vary the unpaired reference, including multiple blurry samples and a sharp sanity reference. Notably, samples (1) and (2) show different blur tendencies. While blur appearance varies with the reference, the scene structure and motion remain aligned with the target sharp data, indicating that the synthesized blur is

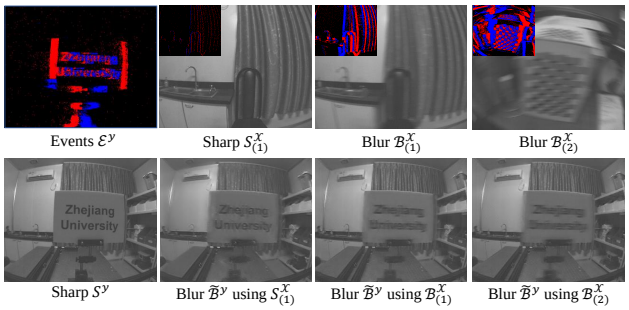

Figure S3. Qualitative analysis with a fixed target $\mathcal{S}^{\mathcal{Y}}$ and varying unpaired references $\mathcal{B}^{\mathcal{X}}_{(1)}$ and $\mathcal{B}^{\mathcal{X}}_{(2)}$, including a sharp reference $\mathcal{S}^{\mathcal{X}}_{(1)}$. For $\mathcal{S}^{\mathcal{X}}_{(1)}$, events are accumulated with a shorter exposure time. Blur appearance varies with the reference, while structural content of $\mathcal{S}^{\mathcal{Y}}$ remains stable.

compatible with the sharp-domain events rather than inheriting motion patterns from the reference. When a sharp image is used as the reference, blur synthesis becomes weaker due to the lack of strong blur cues, confirming that the model conditions on blur-related characteristics from the reference.

References

[1] J. Cheng, W.-T. Chen, X. Lu, and M.-H. Yang. Unpaired deblurring via decoupled diffusion model. *arXiv preprint arXiv:2502.01522*, 2025. 1

[2] P. Duan, B. Li, Y. Yang, H. Lou, M. Teng, Y. Ma, and B. Shi. Eventaid: Benchmarking event-aided image/video enhancement algorithms with real-captured hybrid dataset. *arXiv preprint arXiv:2312.08220*, 2023. 1

[3] S. Lee and G. H. Lee. Diet-gs: Diffusion prior and event stream-assisted motion deblurring 3d gaussian splatting. In *Proceedings of the Computer Vision and Pattern Recognition Conference*, pages 21739–21749, 2025. 4

[4] L. Pan, C. Scheerlinck, X. Yu, R. Hartley, M. Liu, and Y. Dai. Bringing a blurry frame alive at high frame-rate with an event camera. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pages 6820–6829, 2019. 4

[5] L. Pan, C. Scheerlinck, X. Yu, R. Hartley, M. Liu, and Y. Dai. Bringing a blurry frame alive at high frame-rate with an event camera. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pages 6820–6829, 2019. 1

[6] T. Park, A. A. Efros, R. Zhang, and J.-Y. Zhu. Contrastive learning for unpaired image-to-image translation. In *European conference on computer vision*, pages 319–345. Springer, 2020. 1

[7] B.-D. Pham, P. Tran, A. Tran, C. Pham, R. Nguyen, and M. Hoai. Blur2blur: Blur conversion for unsupervised image deblurring on unknown domains. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 2804–2813, 2024. 1

[8] L. Sun, C. Sakaridis, J. Liang, Q. Jiang, K. Yang, P. Sun, Y. Ye, K. Wang, and L. V. Gool. Event-based fusion for motion deblurring with cross-modal attention. In *European conference on computer vision*, pages 412–428. Springer, 2022. 1, 4

[9] L. Sun, C. Sakaridis, J. Liang, P. Sun, J. Cao, K. Zhang, Q. Jiang, K. Wang, and L. Van Gool. Event-based frame interpolation with ad-hoc deblurring. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 18043–18052, 2023. 1

[10] F. Xu, L. Yu, B. Wang, W. Yang, G.-S. Xia, X. Jia, Z. Qiao, and J. Liu. Motion deblurring with real events. In *Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV)*, pages 2583–2592, October 2021. 1, 4

[11] X. Zhang and L. Yu. Unifying motion deblurring and frame interpolation with events. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 17765–17774, 2022. 1

[12] X. Zhang, L. Yu, W. Yang, J. Liu, and G.-S. Xia. Generalizing event-based motion deblurring in real-world scenarios. In *ICCV*, 2023. 1, 4

[13] J.-Y. Zhu, T. Park, P. Isola, and A. A. Efros. Unpaired image-to-image translation using cycle-consistent adversarial networks. In *Proceedings of the IEEE international conference on computer vision*, pages 2223–2232, 2017. 1

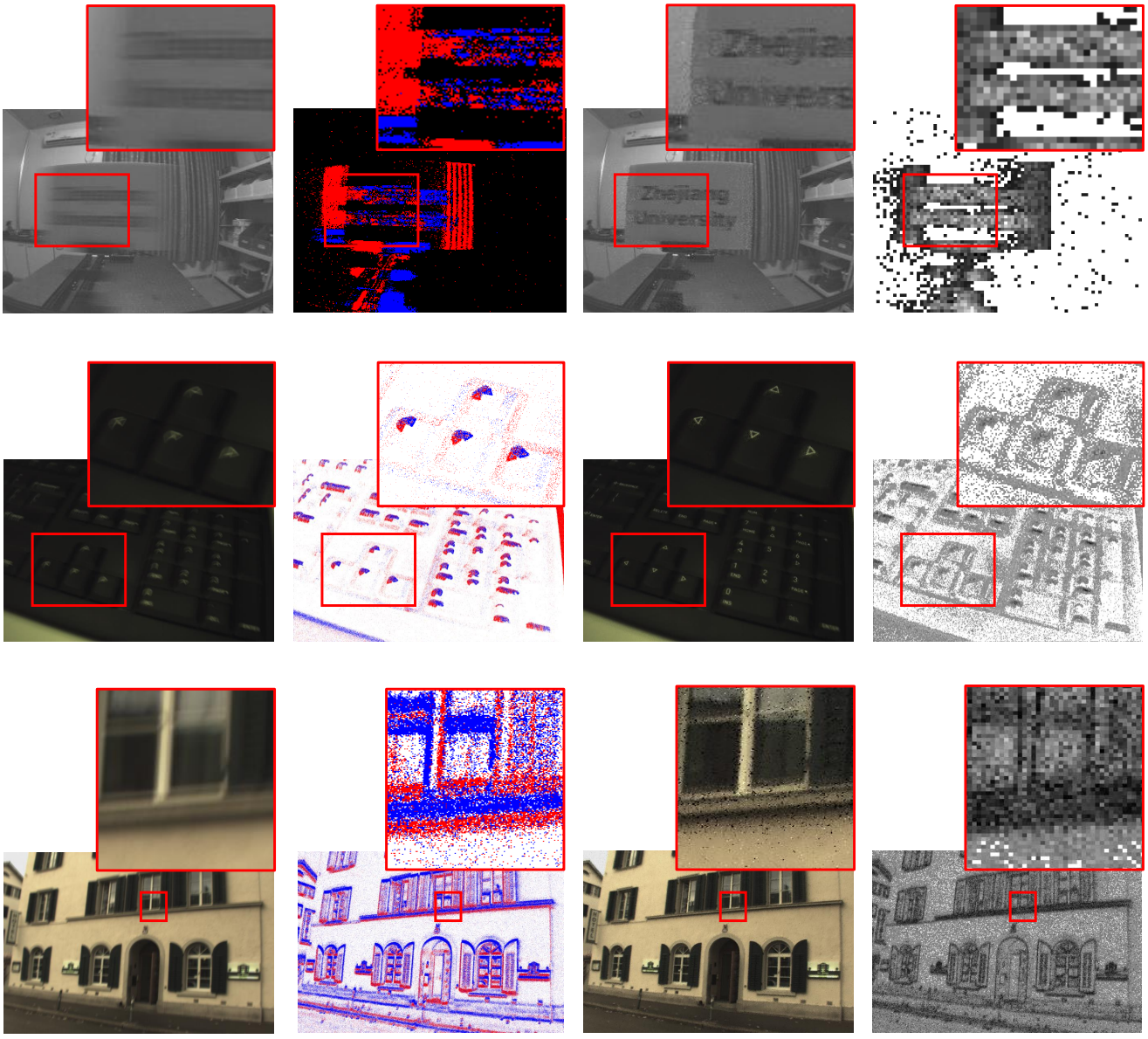

Figure S4. Visualization of the EDI deblurring results and the proposed confidence map. In the confidence map, pixels closer to white have values approaching 1, while pixels closer to black have values approaching 0. EDI deblurring is highly sensitive to noise in the event stream, which can significantly degrade the deblurring quality. In particular, noisy events often introduce salt-and-pepper–like artifacts into the reconstruction. The proposed confidence map assigns high confidence to well-reconstructed regions, whereas regions affected by noise or poor reconstruction receive low confidence.

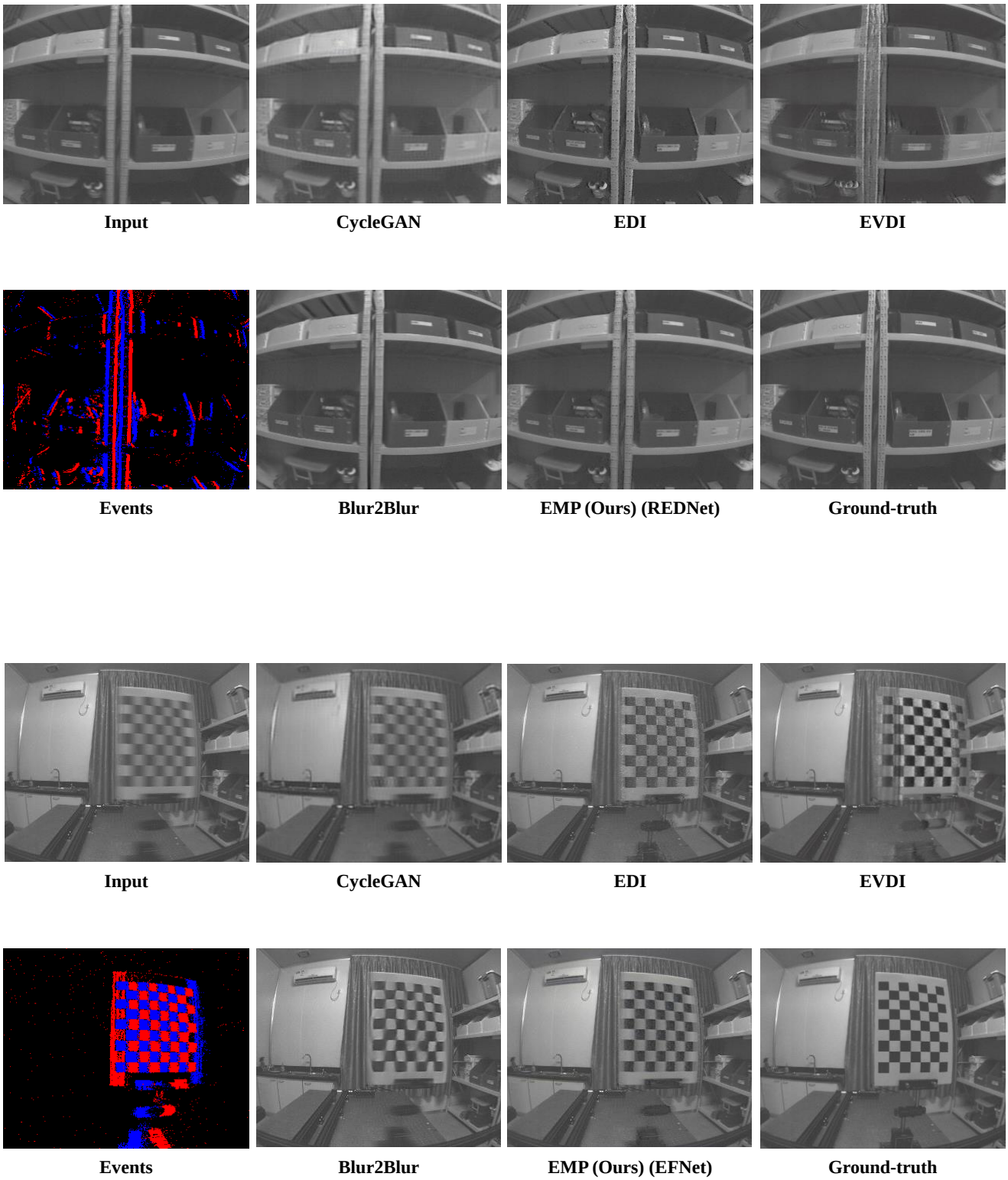

Figure S5. Qualitative comparisons with other unpaired deblurring methods on the REBlur dataset.

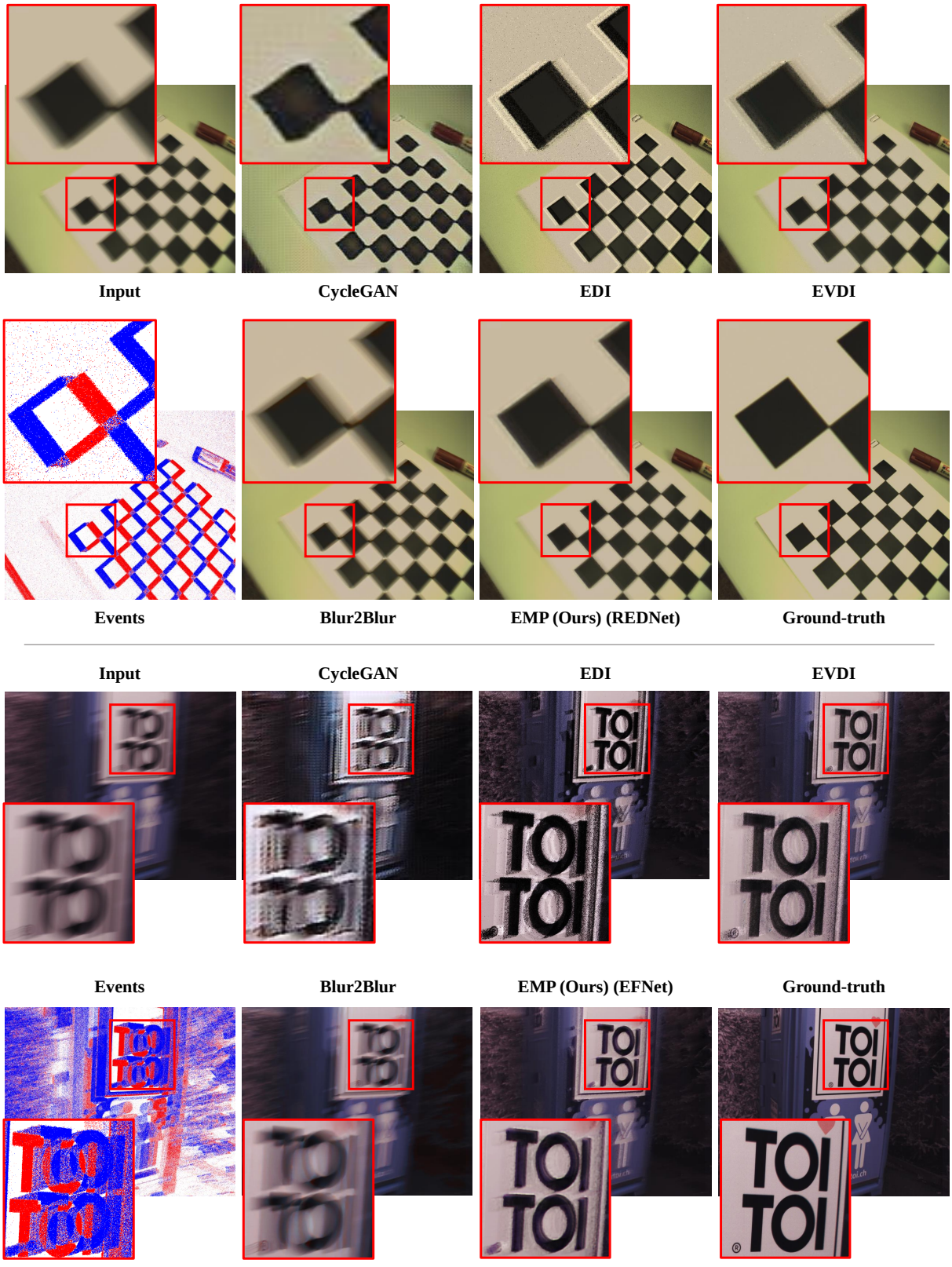

Figure S6. Qualitative comparisons with other unpaired deblurring methods on the HighREV dataset.