CVPR #38625

AdaGaR: Adaptive Gabor Representation for Dynamic Scene Reconstruction

Supplementary Material

Contents

A Activation for Gabor Coefficients 1
A.1. Straight-Through Hard Sigmoid for Frequency Weights 1
B Proof of Adaptive Degradation to Gaussian 1
B.1. Mathematical Formulation 1
B.2. Degradation to Gaussian 1
B.3. Implication for Opacity 1
B.4. Adaptive Gabor Formulation. 2
B.5. Conclusion 2
C Additional Visual Comparisons and Results 2
D Additional Applications 2

A. Activation for Gabor Coefficients

A.1. Straight-Through Hard Sigmoid for Frequency Weights

In Gabor primitives, the frequency coefficients ω_i must satisfy two requirements: (1) values must be constrained within a learnable range, and (2) gradients must flow back through the activation to enable end-to-end optimization.

To achieve this, we employ a Straight-Through Estimator (STE) [1] with a hard sigmoid activation. During the forward pass, we apply hard sigmoid to clip ω_i into the range $[0, 1]$:

$$\hat{\omega} = \text{clip}\left(\frac{\omega + 1}{2}, 0, 1\right). \tag{1}$$

This ensures that the Gabor kernel's frequency modulation remains bounded, preventing unbounded growth that could destabilize energy balance.

However, since the hard clipping operation is non-differentiable, we cannot directly backpropagate through it. Instead, during the backward pass, we use the gradient of the sigmoid function as a surrogate:

$$\frac{\partial L}{\partial \omega} = \frac{\partial L}{\partial \hat{\omega}} \cdot \sigma(\omega)(1 - \sigma(\omega)), \tag{2}$$

where $\sigma(\cdot)$ is the standard sigmoid function. This provides a smooth, bounded gradient signal.

The combination of bounded forward pass and smooth backward pass achieves stable training: the forward pass prevents artifacts by constraining frequency weights, while the backward pass enables effective gradient-based optimization. This approach avoids exploding gradients that can arise from unbounded activations.

B. Proof of Adaptive Degradation to Gaussian

We prove that our Adaptive Gabor representation naturally degrades to a traditional Gaussian when all frequency weights vanish, demonstrating its adaptive capability between Gaussian and Gabor modes.

B.1. Mathematical Formulation

Recall from Eq. (4) and Eq. (5) in the main paper, the adaptive modulation function is defined as:

$$S_{\text{adap}}(\mathbf{x}) = b + \frac{1}{N}\sum_{i=1}^{N} \omega_i \cos(f_i \langle \mathbf{d}_i, \mathbf{x} \rangle), \tag{3}$$

where the compensation term b is given by:

$$b = \gamma + (1 - \gamma)\left(1 - \frac{1}{N}\sum_{i=1}^{N} \omega_i\right), \tag{4}$$

with $\gamma \in [0, 1]$ as a fixed hyperparameter controlling degradation smoothness, and $1/N$ normalizing the weighted average of multiple waves.

B.2. Degradation to Gaussian

Consider the limiting case where all frequency weights approach zero: $\omega_i \to 0$ for all $i \in \{1, \ldots, N\}$.

In this case:

$$\sum_{i=1}^{N} \omega_i \to 0. \tag{5}$$

Substituting into the compensation term:

$$b \to \gamma + (1 - \gamma)\left(1 - \frac{1}{N} \cdot 0\right) = \gamma + (1 - \gamma) \cdot 1 = 1. \tag{6}$$

And the modulation term becomes:

$$\frac{1}{N}\sum_{i=1}^{N} \omega_i \cos(f_i \langle \mathbf{d}_i, \mathbf{x} \rangle) \to 0. \tag{7}$$

Therefore:

$$S_{\text{adap}}(\mathbf{x}) \to 1 + 0 = 1. \tag{8}$$

B.3. Implication for Opacity

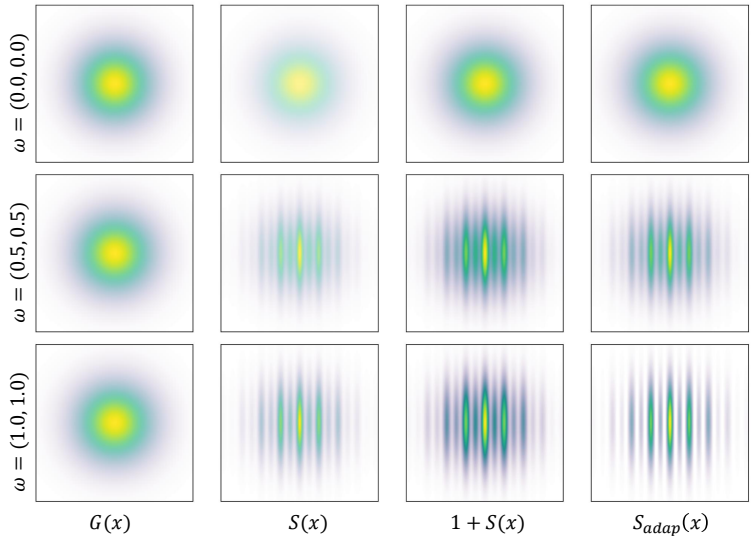
Since the Gabor-modulated opacity is defined as:

$$\alpha_{\text{Gabor}}(\mathbf{x}) = \mathcal{G}(\mathbf{x}) \cdot S_{\text{adap}}(\mathbf{x}), \tag{9}$$

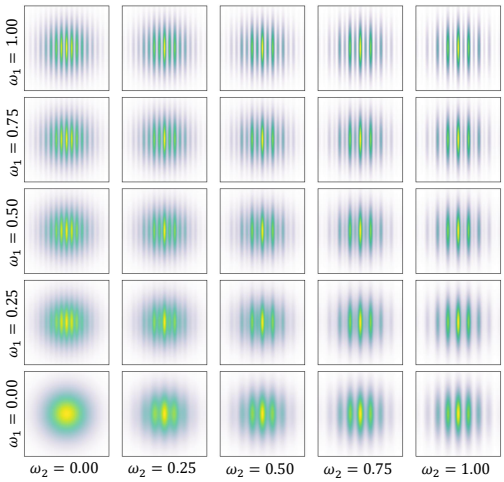
when $S_{\text{adap}}(\mathbf{x}) = 1$, we recover:

$$\alpha_{\text{Gabor}}(\mathbf{x}) = \mathcal{G}(\mathbf{x}) \cdot 1 = \mathcal{G}(\mathbf{x}), \tag{10}$$

which is exactly the traditional Gaussian primitive without frequency modulation.



(a) $G(x)$,Gaussian, $S(x)$, sinusoid part



(b) The effect of different combinations of Gabor wave coefficients (ω_0 and ω_1) on spatial frequency textures.

Figure 1. **Adaptive Gabor formulation (a) Smooth transition between Gaussian and Gabor kernels.** Our method (rightmost column, $S_{\text{adap}}(x)$) uses a compensation term b to maintain energy stability while transitioning from pure Gaussian ($\omega = 0$, top) to frequency-modulated Gabor ($\omega = 1$, bottom). Naive combination $1 + S(x)$ (third column) suffers from intensity artifacts. **(b) Frequency weight combinations.** Different (ω_0, ω_1) pairs generate diverse spatial patterns, from smooth (low ω) to high-frequency textures (high ω), enabling adaptive detail capture in different scene regions.

B.4. Adaptive Gabor Formulation.

We provide additional visualization of the proposed Adaptive Gabor design in Fig. 1. Figure 1(a) shows that a naive Gaussian-to-Gabor replacement can lead to intensity changes and visible artifacts, since the modulation term alters the kernel energy as frequency content increases. In contrast, our compensation term b stabilizes this transition, allowing the primitive to move smoothly from Gaussian-like behavior to frequency-enhanced Gabor modeling. Figure 1(b) further illustrates that different combinations of wave coefficients produce diverse spatial-frequency patterns, ranging from smooth low-frequency responses to richer high-frequency textures. These visualizations help explain why the proposed formulation can adapt its frequency behavior to different scene regions while maintaining stable rendering quality.

B.5. Conclusion

This proof demonstrates that our Adaptive Gabor representation gracefully degrades to a standard Gaussian when frequency content is not needed ($\omega_i \to 0$), while smoothly transitioning to frequency-enhanced Gabor modes when high-frequency details are required ($\omega_i > 0$). This adaptive behavior is crucial for maintaining energy stability across diverse scene regions with varying frequency characteristics.

C. Additional Visual Comparisons and Results

For comprehensive visual comparisons with baseline methods across various dynamic scenes, please refer to Figs. 2, 5 and 6. These figures demonstrate our method's superior performance in preserving high-frequency texture details and maintaining temporal consistency across challenging scenarios including fast motion, occlusions, and complex deformations.

For interactive visualization of downstream application results, including frame interpolation, video editing, and stereo view synthesis, please refer to the supplementary HTML page (`index.html`). The interactive viewer allows frame-by-frame inspection and video playback to better appreciate the temporal coherence and visual quality of our method.

D. Additional Applications

Depth Consistency. We further evaluate whether the reconstructed representation preserves temporally stable geometry in depth space. As shown in Fig. 3, our method maintains consistent depth estimates for static scene elements across frames, while the per-frame baseline exhibits noticeable temporal flickering and boundary inconsistency. This improvement comes from our explicit 3D primitive representation together with smooth temporal motion modeling, which encourages coherent geometry over time rather than frame-wise independent predictions. Such temporal stability is particularly important for downstream applications that rely on consistent depth, such as video editing, view synthesis, and geometry-aware video processing.

Stereo View Synthesis. Our representation also supports stereo view synthesis from monocular video. As shown

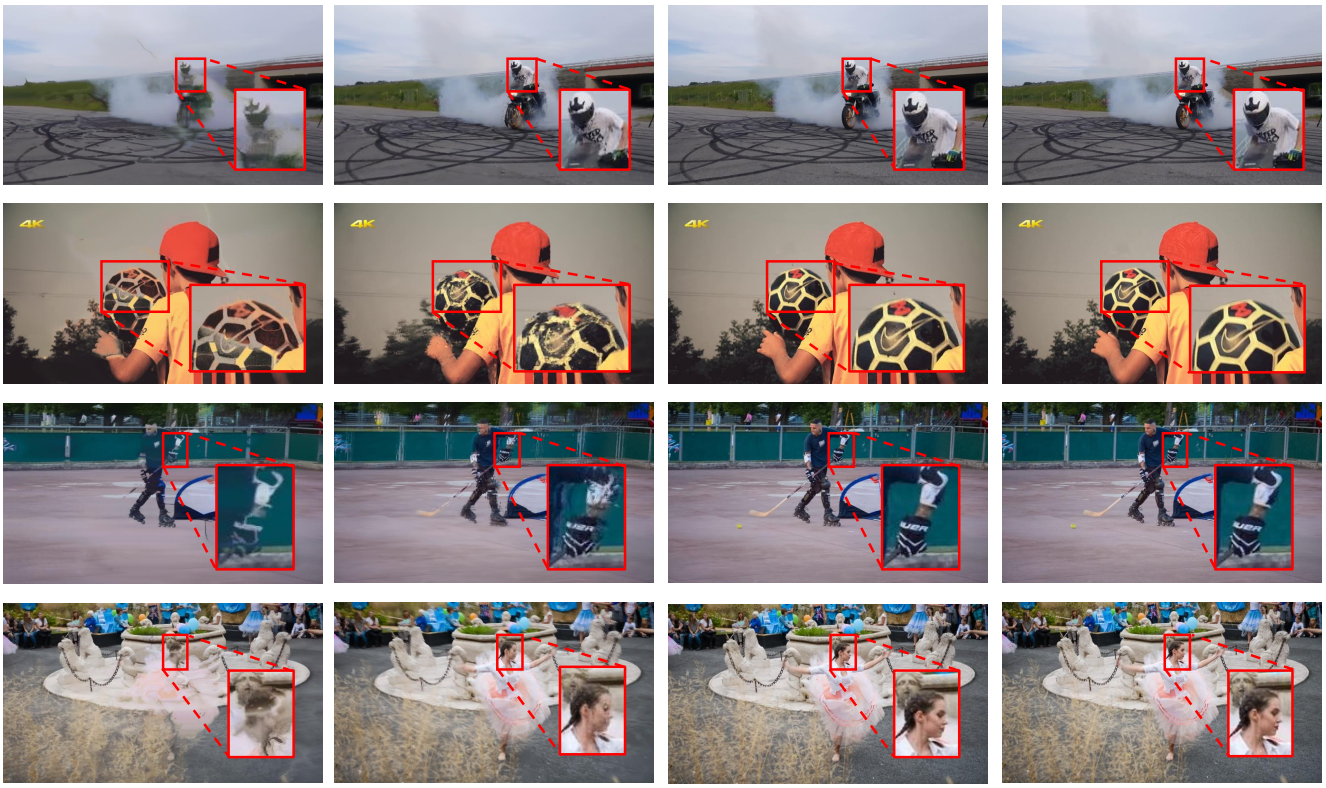


Figure 2. **Visual comparison on the DAVIS dataset.**

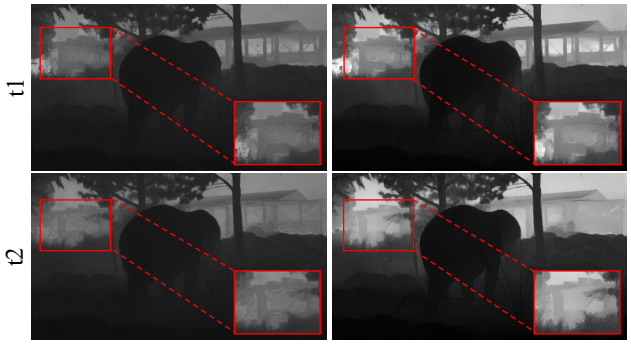


Figure 3. **Depth consistency across time.** *(Left)* Our 3D primitive representation maintains consistent depth for static elements across frames. *(Right)* Per-frame estimation (Marigold [2]) shows temporal flickering (red boxes). Explicit 3D geometry with smooth motion modeling ensures temporal coherence essential for depth-based video applications.

in Fig. 4, the learned 3D structure enables plausible viewpoint variation and consistent scene geometry, producing a convincing stereo effect from a single input video. This result suggests that Adaptive Gabor primitives in the orthographic camera coordinate space capture not only appearance details but also sufficiently stable spatial structure for downstream

Figure 4. **Stereo view synthesis.** Our 3D representation enables novel view synthesis for stereo visualization from monocular video. This demonstrates that Adaptive Gabor primitives in orthographic camera coordinate space capture accurate 3D geometry, enabling immersive applications.

view synthesis. The ability to generate stereo views further demonstrates the practical utility of our explicit dynamic representation beyond standard reconstruction.

References

[1] Yoshua Bengio, Nicholas Léonard, and Aaron Courville. Estimating or propagating gradients through stochastic neurons for conditional computation. *arXiv preprint arXiv:1308.3432*, 2013. 1

[2] Bingxin Ke, Anton Obukhov, Shengyu Huang, Nando Metzger, Rodrigo Caye Daudt, and Konrad Schindler. Repurposing diffusion-based image generators for monocular depth estima-

CVPR
#38625

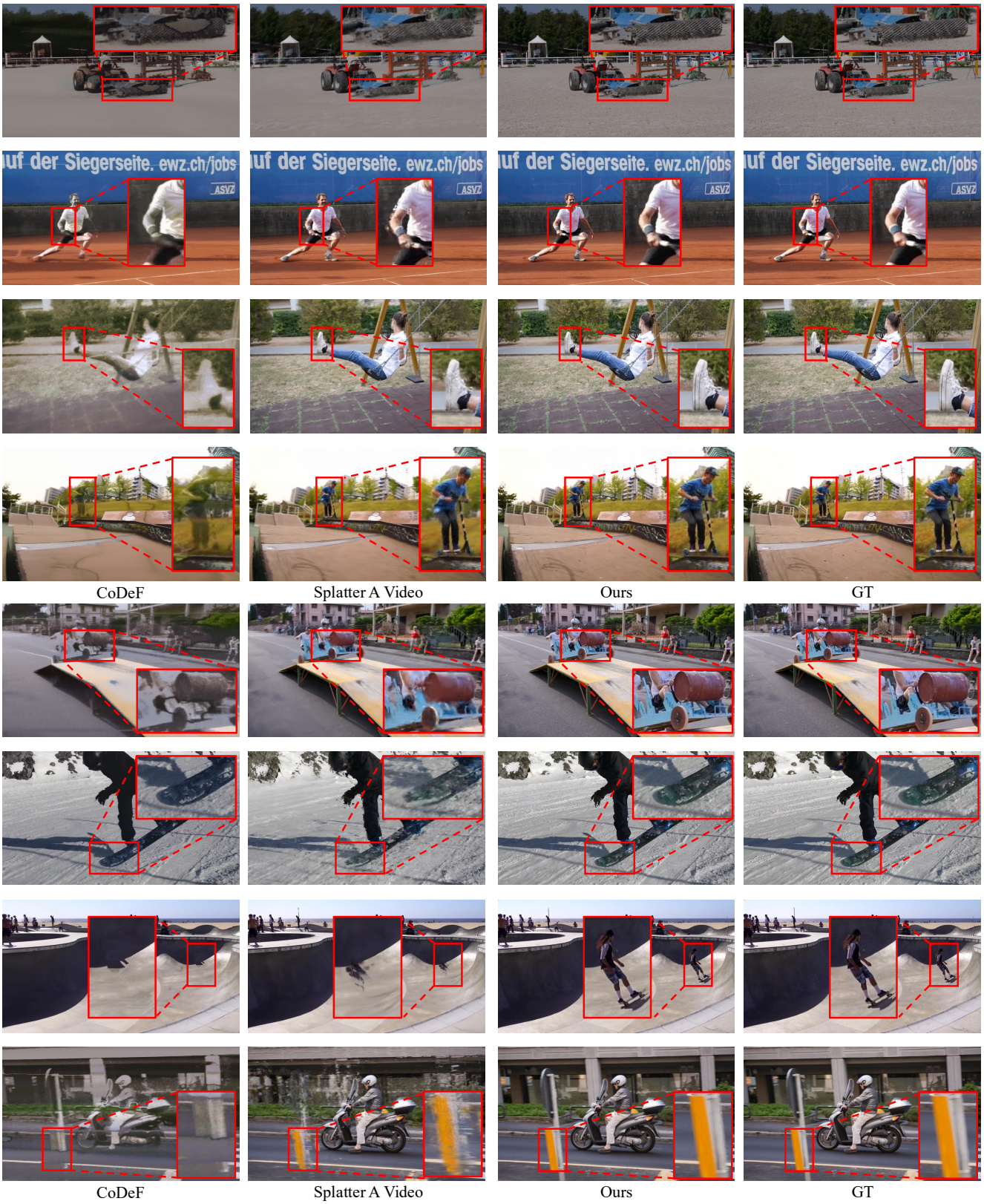


Figure 5. **Visual comparison on the DAVIS dataset.**

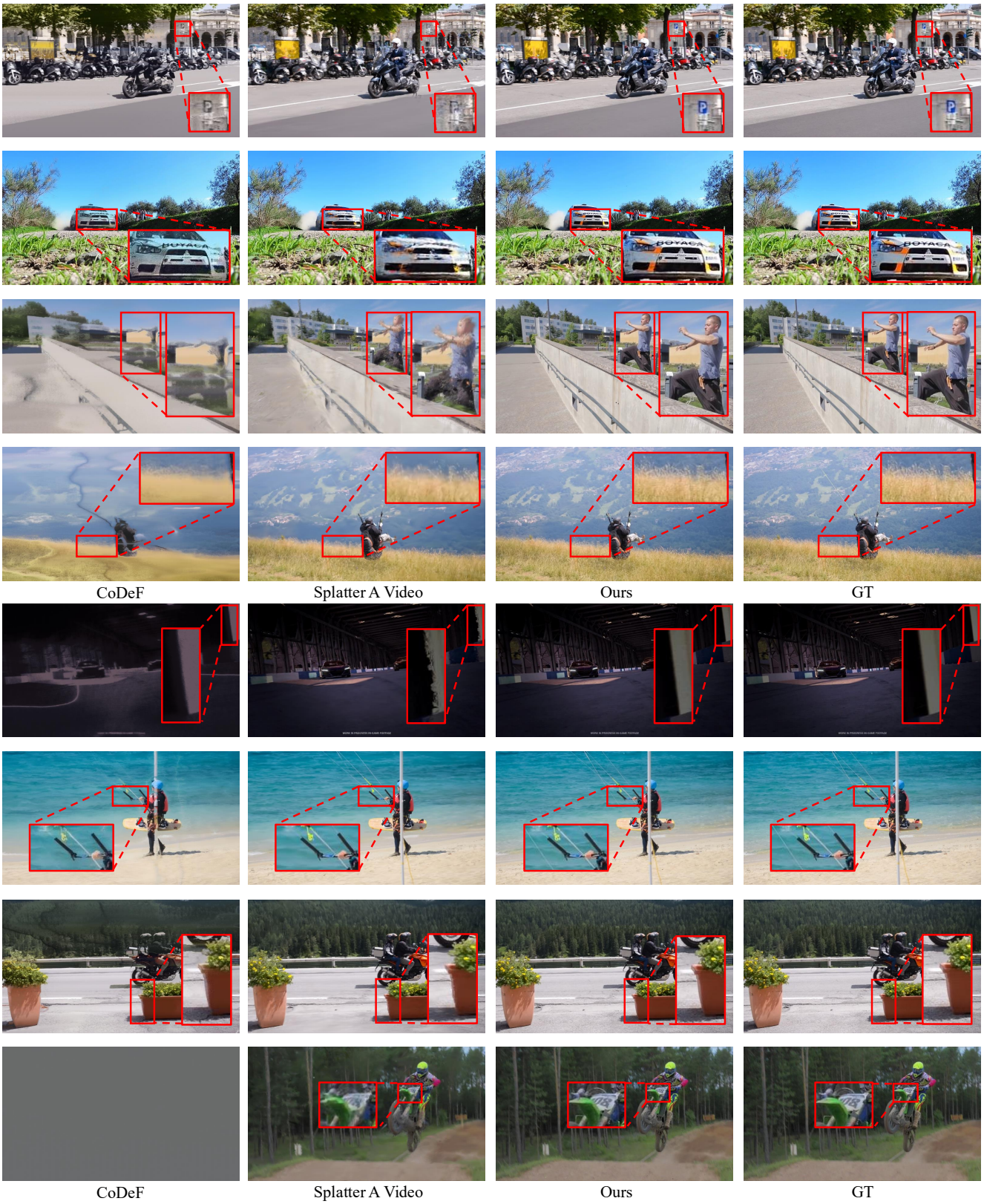


Figure 6. **Visual comparison on the DAVIS dataset.**

tion. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pages 9492–9502, 2024. 3