

Supplementary Material: Frame Interpolation for Dynamic Scenes with Implicit Flow Encoding

Pedro Figueirêdo
Texas A&M University
pedrofigueiredo@tamu.edu

Avinash Paliwal
Texas A&M University
avinashpaliwal@tamu.edu

Nima Khademi Kalantari
Texas A&M University
nimak@tamu.edu

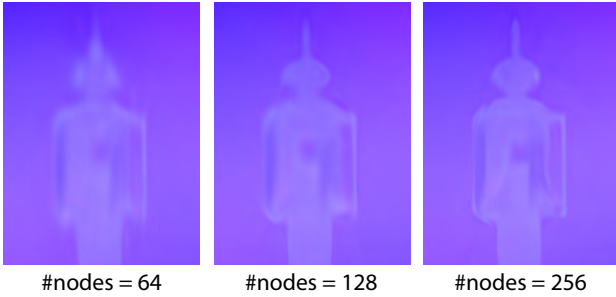


Figure 1: Effect of changing the number of hidden nodes of the hypernetwork on the quality of the interpolated flows.

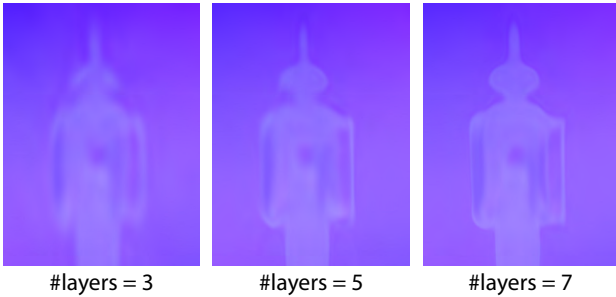


Figure 2: Effect of changing the number of SIREN's hidden layers on the quality of the interpolated flows.

In this supplementary document, we provide additional ablation experiments (Sec. 1) and visual results (Sec. 2). We encourage readers to view the supplementary video which contains visualizations of additional scenes in the targeted use-case of short video interpolations.

1. Additional Ablation Experiments

We evaluate the flow interpolation quality by changing the size of the hypernetwork and SIREN. As shown in Fig. 1, increasing the number of hidden nodes beyond 128, which we use in our implementation, does not significantly impact the quality. Similarly, as shown in Fig. 2, there is no noticeable benefit in going beyond 5 hidden layers in SIREN, which we use in our implementation.

2. Additional Results

Visual Comparisons for Paper's Table 1: Figs. 3, 4, and 5 provide visual comparisons of intermediate flows ($t = 0.5$) generated by our approach and the method by Reda *et al.* [4] (FILM) for a few scenes from the Xiph 2K and 4K [2], and Sintel [1] datasets, respectively. As seen on the figures, our approach can properly interpolate the input RAFT flows and generate intermediate flows that are comparable to the reference.

Additional Intermediate Images and Flows: In Figs. 6, 7, and 8, we provide additional intermediate images and flows for Fig. 4 of the paper. We compare our method with the approaches by Park *et al.* [3] (ABME) and Reda *et al.* [4] (FILM) in terms of image and flow interpolation. For each scene, we show the intermediate images and flows at $t = 0.125, 0.25, 0.5, 0.75$, and 0.875 . Note that we only show intermediate flow comparisons with FILM, since ABME does not explicitly estimate flows. The extended visualizations further demonstrates that our approach significantly outperforms the other techniques.

References

- [1] Joel Janai, Fatma Guney, Jonas Wulff, Michael J Black, and Andreas Geiger. Slow flow: Exploiting high-speed cameras for accurate and diverse optical flow reference data. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 3597–3607, 2017.
- [2] Simon Niklaus and Feng Liu. Softmax splatting for video frame interpolation. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, June 2020.
- [3] Junheum Park, Chul Lee, and Chang-Su Kim. Asymmetric bilateral motion estimation for video frame interpolation. In *Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV)*, pages 14539–14548, October 2021.
- [4] Fitsum Reda, Janne Kontkanen, Eric Tabellion, Deqing Sun, Caroline Pantofaru, and Brian Curless. Film: Frame interpolation for large motion. *arXiv preprint arXiv:2202.04901*, 2022.



Figure 3: Intermediate flow estimations for the Xiph 2K [2] dataset. On the left, we show the perturbed (using various photometric augmentations) input images (see paper Sec. 4.1) and the normalized bidirectional RAFT flows. We show our interpolated middle flow as well as estimated flow by Reda *et al.*'s method [4]. Note that here we use the RAFT flow, computed on the unperturbed middle image and the input as reference.

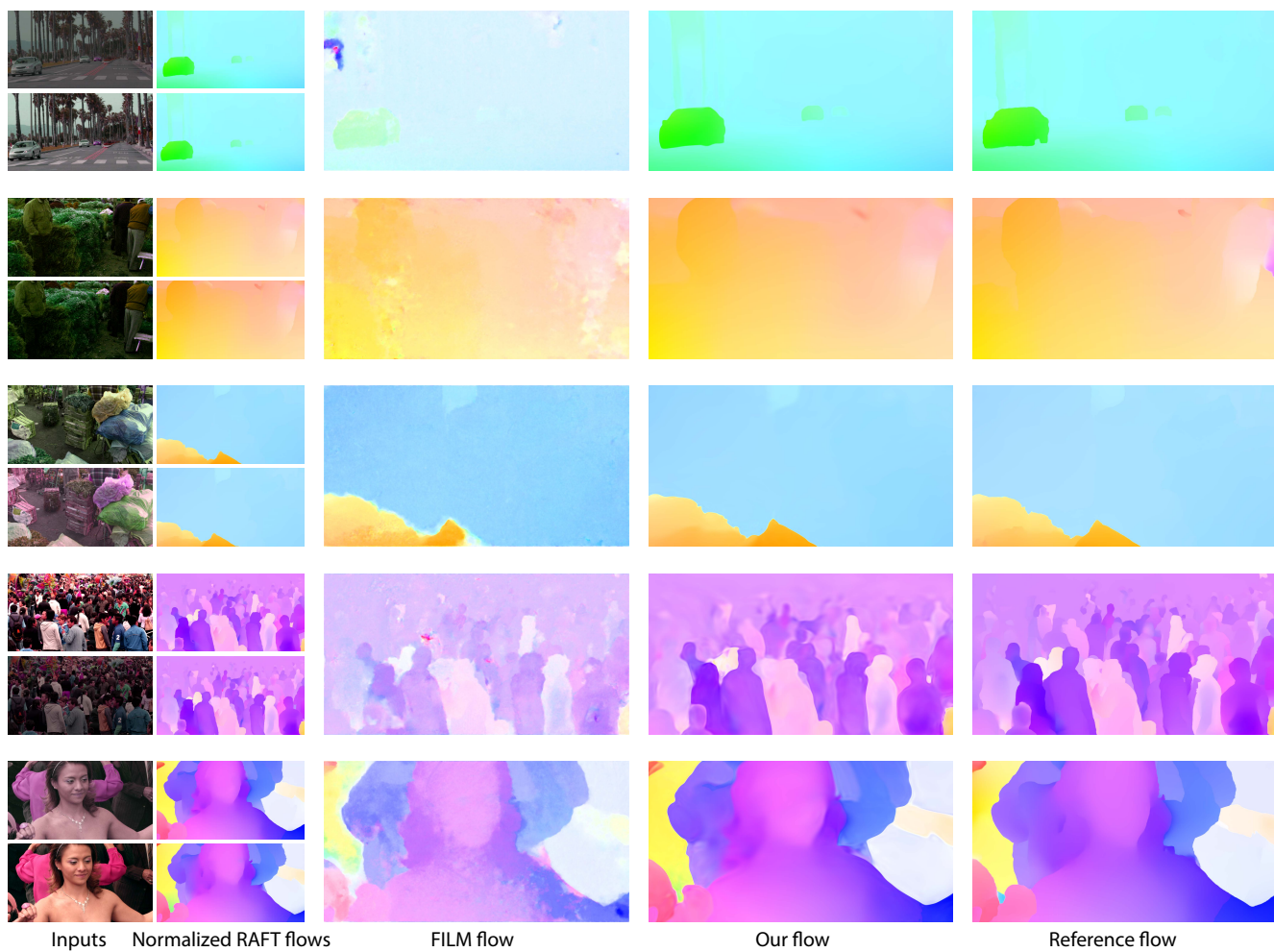


Figure 4: Intermediate flow estimations for the Xiph 4K [2] dataset.

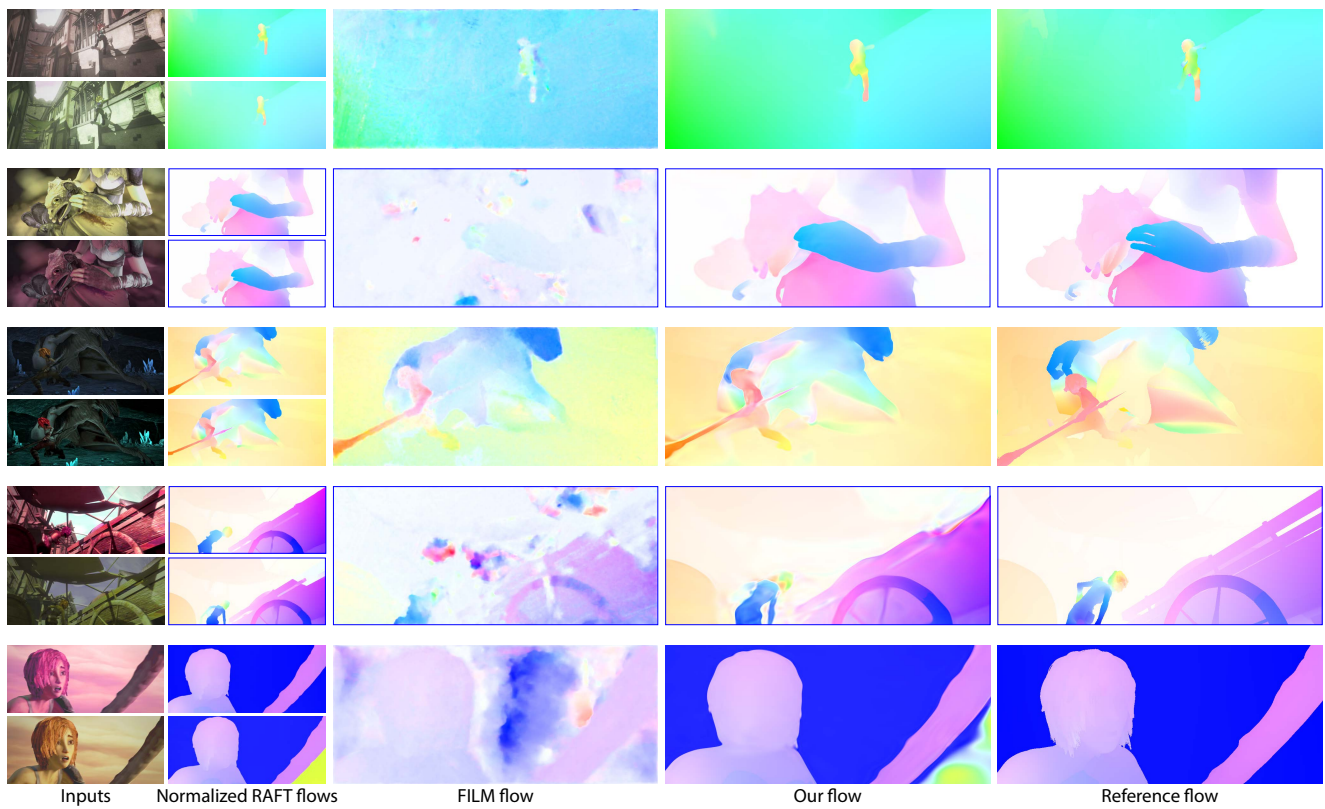


Figure 5: Intermediate flow estimations for the Sintel [1] dataset. Here we use the Sintel’s provided ground truth flow as reference.

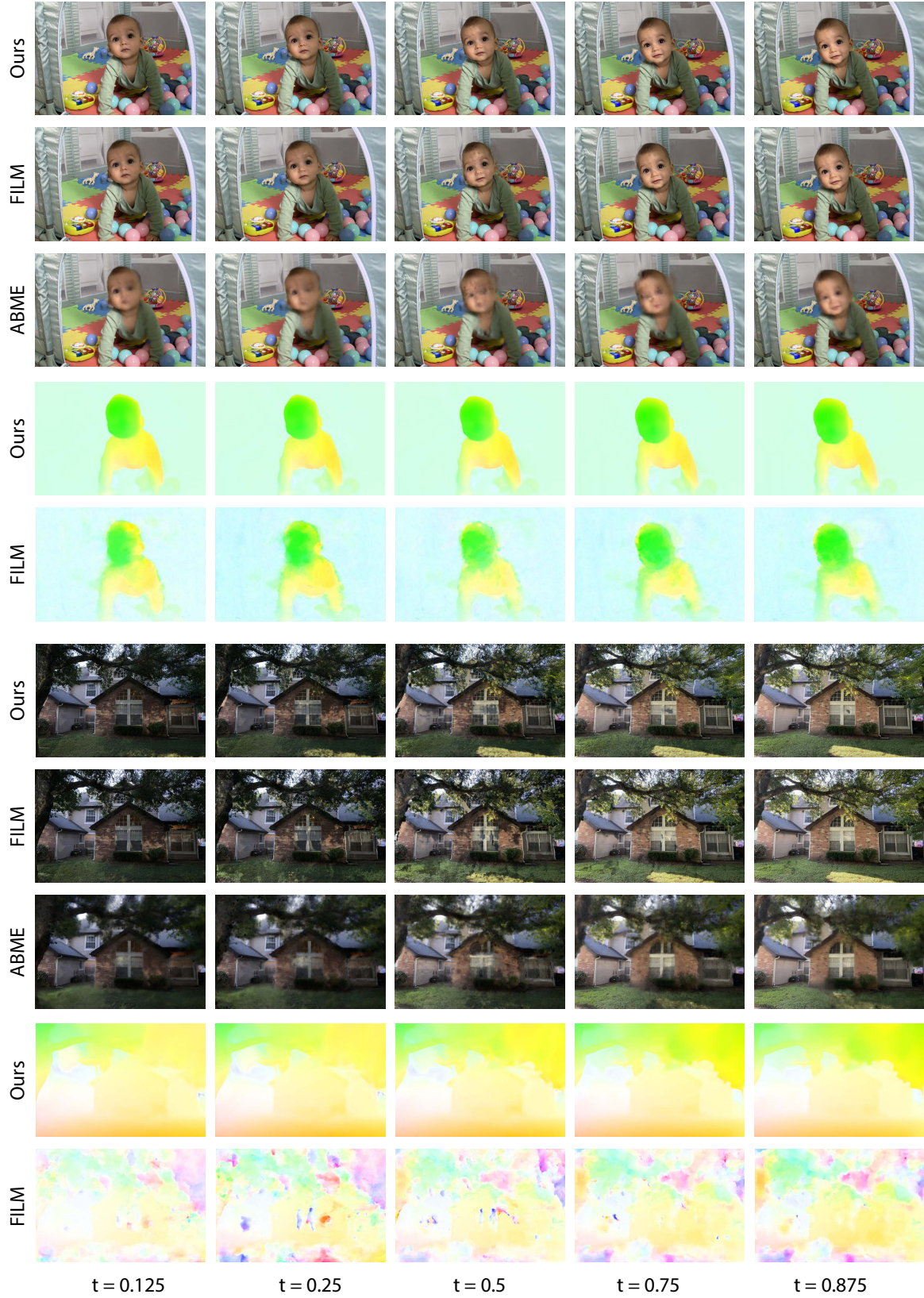


Figure 6: We show detailed comparisons of the BABY and HOUSE scenes against the state-of-the-art methods by Park *et al.* [3] (ABME) and Reda *et al.* [4] (FILM). We only show intermediate flow comparisons with FILM, since ABME does not explicitly estimate flows.

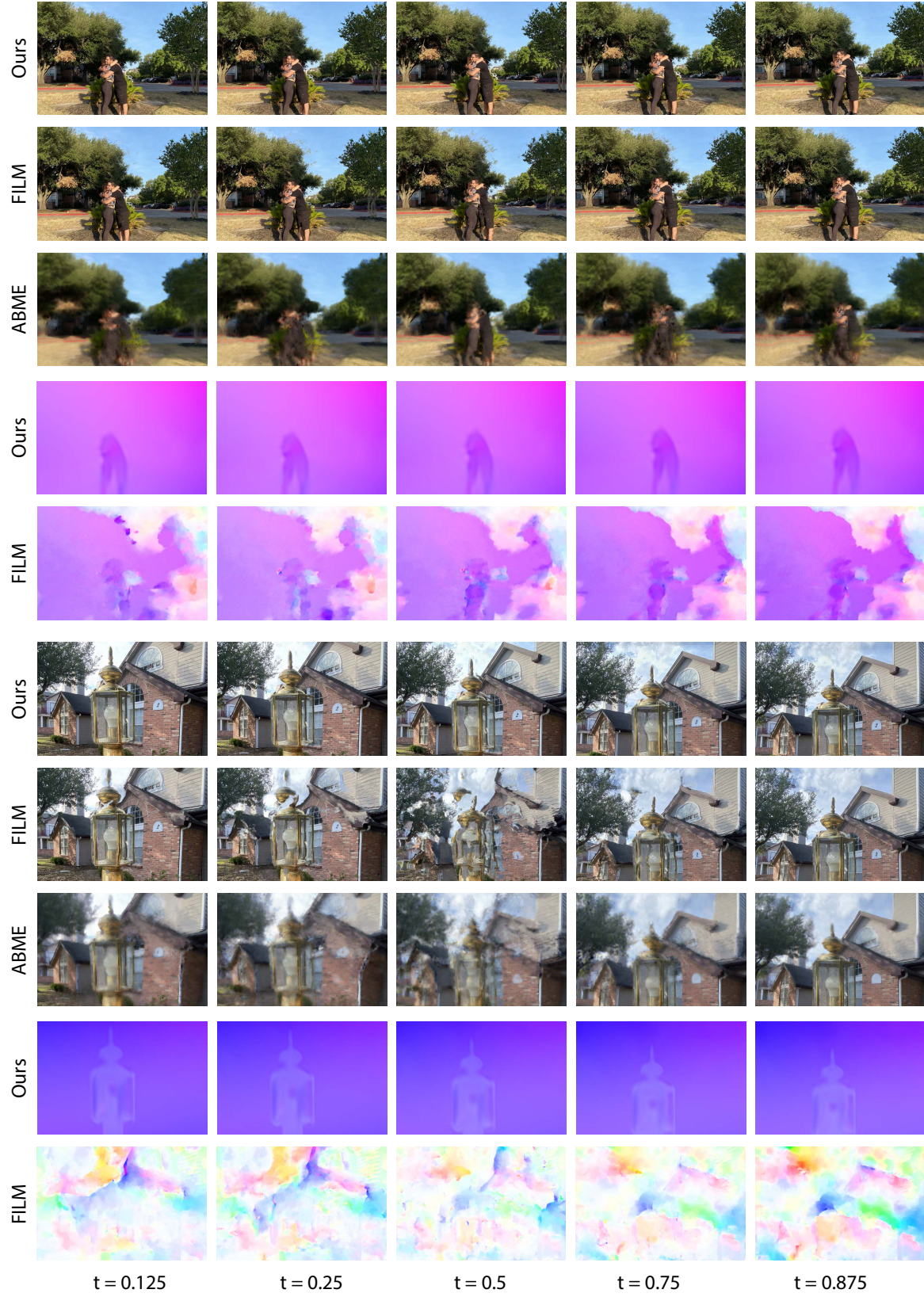


Figure 7: We show detailed comparisons of the HUG and LAMP scenes against the state-of-the-art methods by Park *et al.* [3] (ABME) and Reda *et al.* [4] (FILM).

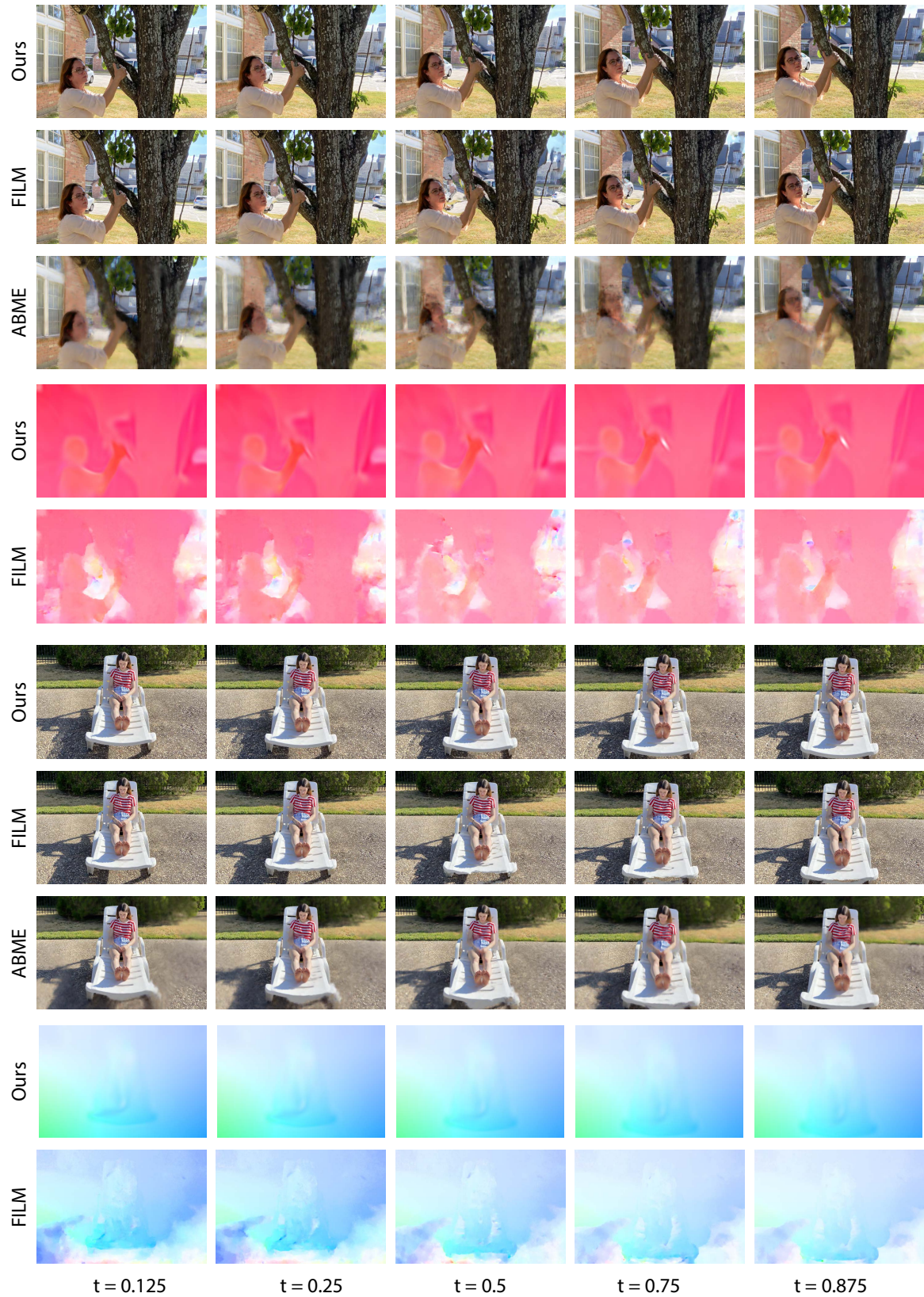


Figure 8: We show detailed comparisons of the TREE and LADY scenes against the state-of-the-art methods by Park *et al.* [3] (ABME) and Reda *et al.* [4] (FILM).