Supplementary materials for "Event-based Video Reconstruction via Potential-assisted Spiking Neural Network"

Table 1. EVSNN Architecture Details.

Layer	Description	Input	Output	Spike Rate
Head	Conv2D: kernel size = 5×5, stride = 1, padding = 2 LIF spiking neuron (V_{th} = 1, τ = 2, $V_{reset} = 0$)	$\mathbf{x}$: 1×H×W	$\mathbf{d}_0$: 32×H×W	0.2479
Down1	Conv2D: kernel size = 5×5, stride = 2, padding = 2 LIF spiking neuron (V_{th} = 1, τ = 2, $V_{reset} = 0$)	$\mathbf{d}_0$: 32×H×W	$\mathbf{d}_1$: 64×$\frac{1}{2}$H×$\frac{1}{2}$W	0.2459
Down2	Conv2D: kernel size = 5×5, stride = 2, padding = 2 LIF spiking neuron (V_{th} = 1, τ = 2, $V_{reset} = 0$)	$\mathbf{d}_1$: 64×$\frac{1}{2}$H×$\frac{1}{2}$W	$\mathbf{d}_2$: 128×$\frac{1}{4}$H×$\frac{1}{4}$W	0.1352
Down3	Conv2D: kernel size = 5×5, stride = 2, padding = 2 LIF spiking neuron (V_{th} = 1, τ = 2, $V_{reset} = 0$)	$\mathbf{d}_2$: 128$\frac{1}{4}$H×$\frac{1}{4}$W	$\mathbf{d}_3$: 256×$\frac{1}{8}$H×$\frac{1}{8}$W	0.1174
Res1-1	Conv2D: kernel size = 3×3, stride = 1, padding = 1 LIF spiking neuron (V_{th} = 1, τ = 2, $V_{reset} = 0$)	$\mathbf{d}_3$: 256×$\frac{1}{8}$H×$\frac{1}{8}$W	$\mathbf{r}_1$: 256×$\frac{1}{8}$H×$\frac{1}{8}$W	0.1241
Res1-2	Conv2D: kernel size = 3×3, stride = 1, padding = 1 LIF spiking neuron (V_{th} = 1, τ = 2, $V_{reset} = 0$)	$\mathbf{r}_1$: 256×$\frac{1}{8}$H×$\frac{1}{8}$W	$\mathbf{r}_2$: 256×$\frac{1}{8}$H×$\frac{1}{8}$W	0.1308
Up1	Interp2D: upsampling-factor = 2 Conv2D: kernel size = 5×5, stride = 1, padding = 2 LIF spiking neuron (V_{th} = 1, τ = 2, $V_{reset} = 0$)	$\mathbf{r}_2 \odot \mathbf{d}_3$: 512×$\frac{1}{8}$H×$\frac{1}{8}$W	$\mathbf{u}_1$: 128×$\frac{1}{4}$H×$\frac{1}{4}$W	0.1905
Up2	Interp2D: upsampling-factor = 2 Conv2D: kernel size = 5×5, stride = 1, padding = 2 LIF spiking neuron (V_{th} = 1, τ = 2, $V_{reset} = 0$)	$\mathbf{u}_1 \odot \mathbf{d}_2$: 256×$\frac{1}{4}$H×$\frac{1}{4}$W	$\mathbf{u}_2$: 64×$\frac{1}{2}$H×$\frac{1}{2}$W	0.3338
Up3	Interp2D: upsampling-factor = 2 Conv2D: kernel size = 5×5, stride = 1, padding = 2 LIF spiking neuron (V_{th} = 1, τ = 2, $V_{reset} = 0$)	$\mathbf{u}_2 \odot \mathbf{d}_1$: 128×H×W	$\mathbf{u}_3$: 32×H×W	0.3580
Prediction	Conv2D: kernel size = 5×5, stride = 1, padding = 1 MP_LIF neuron (V_{th} = Inf, τ = 2, V_{reset} = None)	$\mathbf{u}_3 \odot \mathbf{d}_0$: 64×H×W	$\mathbf{p}$: 1×H×W	-

1. SNN Architecture Details

The detailed architecture of our SNNs are shown in Table 1 and 2. For each spiking layer, we count the spike rate on IJRR dataset. The average spike rates of EVSNN and PA-EVSNN are 0.2642 and 0.2511 respectively, which indicate low energy consumption.

Our EVSNN is a variant of the U-shaped model. All neurons in head, encoder, and decoder layers are spiking neuron. Based on EVSNN, we further propose a potential-assisted EVSNN model, the details are shown in Table 2. MP neuron is introduced in each encoder and decoder layer to help extract the temporal information hidden in the spikes.

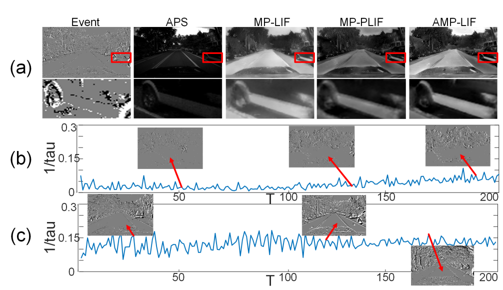

Figure 1. **Ablation of AMP neuron.** (a) Qualitative results of AMP neurons. By introducing AMP neurons, the results are clearer and the contrast is better. (b)(c) are the membrane time constant τ of AMP neuron under **slow** and **fast** light changing scenes, respectively.

Table 2. PA-EVSNN Architecture Details.

Layer	Description	Input	Output	Spike Rate
Head	Conv2D: kernel size = 5×5, stride = 1, padding = 2 LIF spiking neuron (V_{th} = 1, τ = 2, $V_{reset} = 0$)	$\mathbf{x}$: 1×H×W	$\mathbf{d}_0$: 32×H×W	0.2479
Head_MP	Conv2D: kernel size = 1×1, stride = 1, padding = 0 AMP_LIF neuron (V_{th} = Inf, $V_{reset} = 0$)	$\mathbf{d}_0$: 32×H×W	$\mathbf{d}_0^{\mathbf{MP}}$:32×H×W	-
Down1	Conv2D: kernel size = 5×5, stride = 2, padding = 2 LIF spiking neuron (V_{th} = 1, τ = 2, $V_{reset} = 0$)	$\mathbf{d}_0$: 32×H×W	$\mathbf{d}_1$: $64\times\frac{1}{2}$H$\times\frac{1}{2}$W	0.2459
Down1_MP	Conv2D: kernel size = 1×1, stride = 1, padding = 0 AMP_LIF neuron (V_{th} = Inf, $V_{reset} = 0$)	$\mathbf{d}_1$: $64\times\frac{1}{2}$H$\times\frac{1}{2}$W	$\mathbf{d}_1^{\mathbf{MP}}$: $32\times\frac{1}{2}$H$\times\frac{1}{2}$W	-
Down2	Conv2D: kernel size = 5×5, stride = 2, padding = 2 LIF spiking neuron (V_{th} = 1, τ = 2, $V_{reset} = 0$)	$\mathbf{d}_1$: $64\times\frac{1}{2}$H$\times\frac{1}{2}$W	$\mathbf{d}_2$: $128\times\frac{1}{4}$H$\times\frac{1}{4}$W	0.1352
Down2_MP	Conv2D: kernel size = 1×1, stride = 1, padding = 0 AMP_LIF neuron (V_{th} = Inf, $V_{reset} = 0$)	$\mathbf{d}_2$: $128\times\frac{1}{4}$H$\times\frac{1}{4}$W	$\mathbf{d}_2^{\mathbf{MP}}$: $32\times\frac{1}{4}$H$\times\frac{1}{4}$W	-
Down3	Conv2D: kernel size = 5×5, stride = 2, padding = 2 LIF spiking neuron (V_{th} = 1, τ = 2, $V_{reset} = 0$)	$\mathbf{d}_2$: $128\frac{1}{4}$H$\times\frac{1}{4}$W	$\mathbf{d}_3$: $256\times\frac{1}{8}$H$\times\frac{1}{8}$W	0.1174
Down3_MP	Conv2D: kernel size = 1×1, stride = 1, padding = 0 AMP_LIF neuron (V_{th} = Inf, $V_{reset} = 0$)	$\mathbf{d}_3$: $256\times\frac{1}{8}$H$\times\frac{1}{8}$W	$\mathbf{d}_3^{\mathbf{MP}}$: $32\times\frac{1}{8}$H$\times\frac{1}{8}$W	-
Res1-1	Conv2D: kernel size = 3×3, stride = 1, padding = 1 LIF spiking neuron (V_{th} = 1, τ = 2, $V_{reset} = 0$)	$\mathbf{d}_3$: $256\times\frac{1}{8}$H$\times\frac{1}{8}$W	$\mathbf{r}_1$: $256\times\frac{1}{8}$H$\times\frac{1}{8}$W	0.1241
Res1-2	Conv2D: kernel size = 3×3, stride = 1, padding = 1 LIF spiking neuron (V_{th} = 1, τ = 2, $V_{reset} = 0$)	$\mathbf{r}_1$: $256\times\frac{1}{8}$H$\times\frac{1}{8}$W	$\mathbf{r}_2$: $256\times\frac{1}{8}$H$\times\frac{1}{8}$W	0.1308
Res1_MP	Conv2D: kernel size = 1×1, stride = 1, padding = 0 AMP_LIF neuron (V_{th} = Inf, $V_{reset} = 0$)	$\mathbf{r}_2$: $256\times\frac{1}{8}$H$\times\frac{1}{8}$W	$\mathbf{r}_2^{\mathbf{MP}}$: $32\times\frac{1}{8}$H$\times\frac{1}{8}$W	-
Up1	Interp2D: upsampling-factor = 2 Conv2D: kernel size = 5×5, stride = 1, padding = 2 LIF spiking neuron (V_{th} = 1, τ = 2, $V_{reset} = 0$)	$\mathbf{r}_2 \odot \mathbf{d}_3$: $512\times\frac{1}{8}$H$\times\frac{1}{8}$W	$\mathbf{u}_1$: $128\times\frac{1}{4}$H$\times\frac{1}{4}$W	0.1905
Up1_MP	Conv2D: kernel size = 1×1, stride = 1, padding = 0 AMP_LIF neuron (V_{th} = Inf, $V_{reset} = 0$)	$\mathbf{u}_1$: $128\times\frac{1}{4}$H$\times\frac{1}{4}$W	$\mathbf{u}_1^{\mathbf{MP}}$: $32\times\frac{1}{4}$H$\times\frac{1}{4}$W	-
Up2	Interp2D: upsampling-factor = 2 Conv2D: kernel size = 5×5, stride = 1, padding = 2 LIF spiking neuron (V_{th} = 1, τ = 2, $V_{reset} = 0$)	$\mathbf{u}_1 \odot \mathbf{d}_2$: $256\times\frac{1}{4}$H$\times\frac{1}{4}$W	$\mathbf{u}_2$: $64\times\frac{1}{2}$H$\times\frac{1}{2}$W	0.3338
Up2_MP	Conv2D: kernel size = 1×1, stride = 1, padding = 0 AMP_LIF neuron (V_{th} = Inf, $V_{reset} = 0$)	$\mathbf{u}_2$: $64\times\frac{1}{2}$H$\times\frac{1}{2}$W	$\mathbf{u}_2^{\mathbf{MP}}$: $32\times\frac{1}{2}$H$\times\frac{1}{2}$W	-
Up3	Interp2D: upsampling-factor = 2 Conv2D: kernel size = 5×5, stride = 1, padding = 2 LIF spiking neuron (V_{th} = 1, τ = 2, $V_{reset} = 0$)	$\mathbf{u}_2 \odot \mathbf{d}_1$: 128×H×W	$\mathbf{u}_3$: 32×H×W	0.3580
Agg1_MP	Interp2D: upsampling-factor = 2 Conv2D: kernel size = 5×5, stride = 1, padding = 2	$\mathbf{d}_3^{\mathbf{MP}} \oplus \mathbf{r}_2^{\mathbf{MP}}$: $32\times\frac{1}{4}$H$\times\frac{1}{4}$W	$\mathbf{a}_1$: $32\times\frac{1}{4}$H$\times\frac{1}{4}$W	-
Agg2_MP	Interp2D: upsampling-factor = 2 Conv2D: kernel size = 5×5, stride = 1, padding = 2	$\mathbf{a}_1 \oplus \mathbf{d}_2^{\mathbf{MP}} \oplus \mathbf{u}_1^{\mathbf{MP}}$: $32\times\frac{1}{4}$H$\times\frac{1}{4}$W	$\mathbf{a}_2$: $32\times\frac{1}{2}$H$\times\frac{1}{2}$W	-
Agg3_MP	Interp2D: upsampling-factor = 2 Conv2D: kernel size = 5×5, stride = 1, padding = 2	$\mathbf{a}_2 \oplus \mathbf{d}_1^{\mathbf{MP}} \oplus \mathbf{u}_2^{\mathbf{MP}}$: $32\times\frac{1}{4}$H$\times\frac{1}{4}$W	$\mathbf{a}_3$: 32×H×W	-
Prediction_spike	Conv2D: kernel size = 5×5, stride = 1, padding = 1 MP_LIF neuron (V_{th} = Inf, τ = 2, V_{reset} = None)	$\mathbf{u}_3 \odot \mathbf{d}_0$: 64×H×W	$\mathbf{p}_s$: 1×H×W	-
Prediction_final	Conv2D: kernel size = 1×1, stride = 1, padding = 0	$\mathbf{p}_s \oplus \mathbf{a}_3 \oplus \mathbf{d}_0^{\mathbf{MP}}$: 1×H×W	$\mathbf{p}$: 1×H×W	-

2. Ablation of AMP neuron

Fig. 2(a) shows the qualitative results of AMP neurons. In Fig. 2, we plot the corresponding event frames to visualize the light changing of the scene (the more events there are, the faster the scene changes). $1/\tau$ controls balance between remembering X_t and forgetting V_{t-1} (see Eq. 2). In the scene of fast light change, the neuron adaptively chooses a large $1/\tau$ to remember more new information while forgetting more memory, thus improves the performance.

3. Additional Quantitative Results

Table 3 shows the breakdown of the quantitative results of EVSNN, PA-EVSNN, E2VID [2], FireNet [3], and SPADE-E2VID [1] on the HQF, IJRR and MVSEC respectively. Our models are based on SNN architecture, while the other three methods are ANN-based models. These five models are trained on the same dataset provided by [2]. All the results are generated by the original code and pretrained weight provided by the authors. The quantitative results show that our SNN models achieve comparable performance on these public datasets, while the energy consumption is much lower.

Since the frames accompanying the events in the commonly used MVSEC and IJRR datasets are of low quality, as recommend in [4], we only evaluate on select sequences and for select cuts of those sequences. In Table 4, we show the cut time for each sequence of the IJRR and MVSEC datasets respectively.

4. Additional Qualitative Results

Additional qualitative results on IJRR, HQF, and MVSEC datasets are shown in Fig. 2, 3, and 4, respectively. Our models perform well on complex scenes.

5. Discussion on the Input Data

EVSNN is a fully spiking neural network, but when the input data contains floating-point values, such as voxels, the first convolution layer in 'head layer' will bring some MAC operations (e.g., 0.038 GMACs with 180×240 resolution input). This motivated us to explore a binary event representation for SNN input in the future.

6. Inference Speed on GPU

We use an NVIDIA Titan Xp GPU for all experiments, with 180×240 input size, the inference speeds of EVSNN and PA-EVSNN are 250+ FPS and 70+ FPS, respectively.

References

[1] Pablo Rodrigo Gantier Cadena, Yeqiang Qian, Chunxiang Wang, and Ming Yang. Spade-e2vid: Spatially-adaptive denormalization for event-based video reconstruction. *IEEE Transactions on Image Processing*, 30:2488–2500, 2021. 2

[2] Henri Rebecq, René Ranftl, Vladlen Koltun, and Davide Scaramuzza. High speed and high dynamic range video with an event camera. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2019. 2

[3] Cedric Scheerlinck, Henri Rebecq, Daniel Gehrig, Nick Barnes, Robert Mahony, and Davide Scaramuzza. Fast image reconstruction with an event camera. In *IEEE Winter Conference on Applications of Computer Vision (WACV)*, pages 156–163, 2020. 2

[4] Timo Stoffregen, Cedric Scheerlinck, Davide Scaramuzza, Tom Drummond, Nick Barnes, Lindsay Kleeman, and Robert Mahony. Reducing the sim-to-real gap for event cameras. In *European Conference on Computer Vision (ECCV)*, pages 534–549. Springer, 2020. 2

Table 3. MSE/SSIM/LPIPS of our models and ANN-based methods on IJRR, HQF, and MVSEC Datasets.

	Sequence	E2VID			FireNet			SPADE -E2VID			EVSNN			PA-EVSNN		
IJRR	boxes_6dof_cut	0.0365	0.6602	0.2926	0.0410	0.6470	0.2741	0.0495	0.5709	0.3294	0.0446	0.6074	0.3062	0.0429	0.6052	0.3145
	calibration_cut	0.0324	0.5692	0.4090	0.0446	0.5394	0.3842	0.0302	0.5718	0.3714	0.0483	0.5031	0.4080	0.0405	0.5197	0.4262
	dynamic_6dof_cut	0.0423	0.7181	0.3238	0.0643	0.6286	0.3906	0.0484	0.6536	0.3436	0.0683	0.6251	0.3670	0.0363	0.7335	0.3347
	office_zigzag_cut	0.0400	0.6985	0.2966	0.0502	0.6575	0.3311	0.0482	0.6244	0.3715	0.0469	0.6044	0.3509	0.0352	0.7185	0.2995
	poster_6dof_cut	0.0464	0.6906	0.2226	0.0268	0.7301	0.1792	0.0667	0.5678	0.3021	0.0325	0.6771	0.2494	0.0373	0.6605	0.2508
	shapes_6dof_cut	0.1774	0.4888	0.4886	0.1445	0.3785	0.5067	0.1373	0.4151	0.4904	0.1270	0.4378	0.4665	0.0879	0.5187	0.5723
	slider_depth_cut	0.0396	0.6754	0.3328	0.0517	0.6366	0.3170	0.0607	0.6033	0.3470	0.0547	0.5703	0.3768	0.0414	0.6253	0.3742
	Mean	0.0592	**0.6429**	0.3380	0.0604	0.6025	0.3404	0.0630	0.5724	0.3651	0.0603	0.5750	0.3607	**0.0459**	0.6259	0.3675
HQF	bike_bay_hdr	0.0938	0.4748	0.4280	0.0873	0.4939	0.4118	0.1007	0.4425	0.4542	0.1033	0.4345	0.4421	0.0820	0.4551	0.4577
	boxes	0.0260	0.7019	0.2937	0.0285	0.7056	0.2747	0.0454	0.6276	0.3346	0.0681	0.5882	0.3511	0.0377	0.6584	0.3141
	desk_6k	0.0444	0.6882	0.2783	0.0446	0.6666	0.2909	0.0409	0.6761	0.2811	0.0716	0.5848	0.3577	0.0424	0.6575	0.3205
	desk_fast	0.0624	0.6193	0.3248	0.0580	0.6135	0.3298	0.0616	0.5814	0.3700	0.0808	0.5328	0.3870	0.0523	0.6170	0.3575
	desk_hand_only	0.1185	0.4738	0.5713	0.1010	0.3595	0.6443	0.0902	0.5410	0.5504	0.1196	0.4299	0.6109	0.0844	0.5255	0.5946
	desk_slow	0.0660	0.5988	0.4346	0.0763	0.3883	0.5349	0.0500	0.6013	0.3683	0.0861	0.5346	0.4288	0.0534	0.5915	0.4300
	engineering_posters	0.0648	0.5901	0.4072	0.0372	0.6340	0.3423	0.0696	0.5450	0.4034	0.0715	0.5105	0.4592	0.0421	0.5770	0.3825
	high_texture_plants	0.0989	0.4881	0.3360	0.0498	0.5767	0.2603	0.0656	0.4968	0.3039	0.0716	0.4436	0.3752	0.0560	0.4610	0.3242
	poster_pillar_1	0.0949	0.4187	0.4803	0.0658	0.4666	0.4047	0.0784	0.3991	0.4504	0.0850	0.3843	0.4611	0.0611	0.3981	0.4737
	poster_pillar_2	0.1750	0.2756	0.5633	0.1559	0.2897	0.5179	0.1433	0.2611	0.5624	0.1411	0.2393	0.5403	0.1061	0.2591	0.5706
	reflective_materials	0.0781	0.5232	0.3878	0.0705	0.5343	0.3771	0.1116	0.4493	0.4611	0.1054	0.4333	0.4177	0.0798	0.4860	0.4111
	slow_and_fast_desk	0.0722	0.6047	0.3961	0.0450	0.6371	0.3718	0.0643	0.5846	0.4086	0.0639	0.5522	0.4068	0.0470	0.6076	0.3886
	slow_hand	0.1085	0.4884	0.5167	0.0689	0.5215	0.4771	0.1130	0.4604	0.5603	0.0855	0.4514	0.5166	0.0571	0.5340	0.4941
	still_life	0.0354	0.6796	0.2708	0.0247	0.6978	0.2386	0.0888	0.5031	0.4233	0.0440	0.6242	0.3087	0.0471	0.6234	0.3092
	Mean	0.0813	**0.5446**	0.4063	0.0653	0.5418	**0.3912**	0.0802	0.5121	0.4237	0.0855	0.4817	0.4331	**0.0606**	0.5322	0.4163
MVSEC	indoor_flying1_cut	0.1201	0.3958	0.6304	0.0870	0.4078	0.5539	0.0766	0.4911	0.4913	0.0871	0.4287	0.4925	0.0883	0.4404	0.5184
	indoor_flying2_cut	0.1369	0.3683	0.6775	0.1071	0.3661	0.5839	0.0990	0.4473	0.5248	0.1002	0.4080	0.5086	0.1041	0.4130	0.5436
	indoor_flying3_cut	0.1296	0.3912	0.6363	0.0910	0.4006	0.5652	0.0865	0.4844	0.4940	0.0936	0.4359	0.4912	0.0932	0.4430	0.5234
	indoor_flying4_cut	0.1422	0.2913	0.7181	0.1154	0.2747	0.6531	0.1228	0.3085	0.6454	0.1153	0.3093	0.5962	0.1177	0.3112	0.6300
	outdoor_day1_cut	0.1458	0.3760	0.6570	0.1173	0.3451	0.6844	0.0927	0.4310	0.6545	0.0954	0.3680	0.6179	0.1040	0.3884	0.6482
	outdoor_day2_cut	0.1557	0.4404	0.5850	0.1126	0.3723	0.5598	0.0931	0.4965	0.5270	0.1320	0.3847	0.5188	0.1319	0.4218	0.5295
	Mean	0.1384	0.3772	0.6507	0.1051	0.3611	0.6000	**0.0951**	**0.4431**	0.5562	0.1039	0.3891	**0.5375**	0.1065	0.4030	0.5655

Table 4. Start and end times for sequences in IJRR and MVSEC.

IJRR			MVSEC		
Sequence	Start [s]	End [s]	Sequence	Start [s]	End [s]
boxes_6dof	5.0	20.0	indoor_flying1	10.0	70.0
calibration	5.0	20.0	indoor_flying2	10.0	70.0
dynamic_6dof	5.0	20.0	indoor_flying3	10.0	70.0
office_zigzag	5.0	12.0	indoor_flying4	10.0	19.8
poster_6dof	5.0	20.0	outdoor_day1	0.0	60.0
shapes_6dof	5.0	20.0	outdoor_day2	100.0	160.0
slider_depth	1.0	2.5			

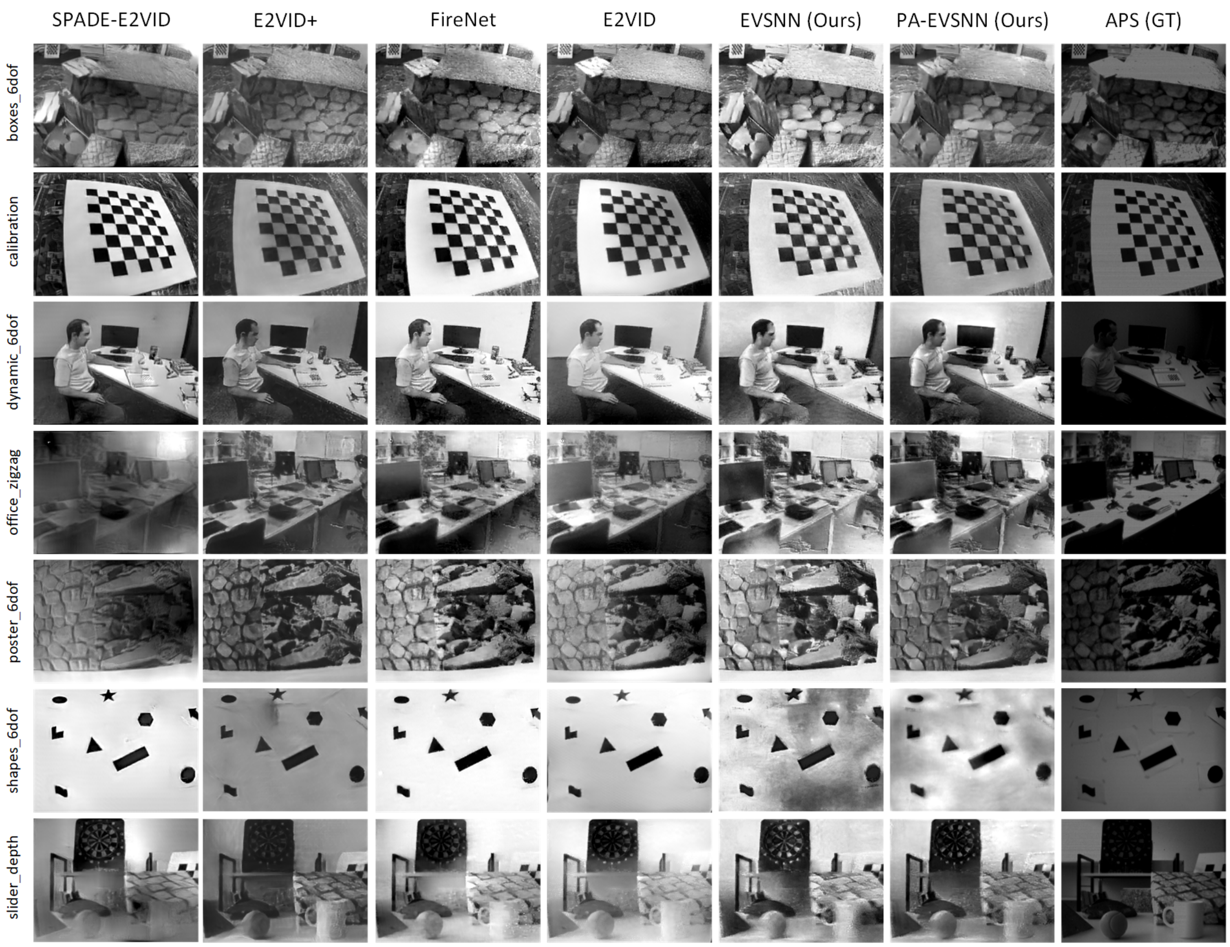

Figure 2. **Qualitative results of our models and ANN-based methods on IJRR dataset.**

Figure 3. **Qualitative results of our models and ANN-based methods on HQF dataset.**

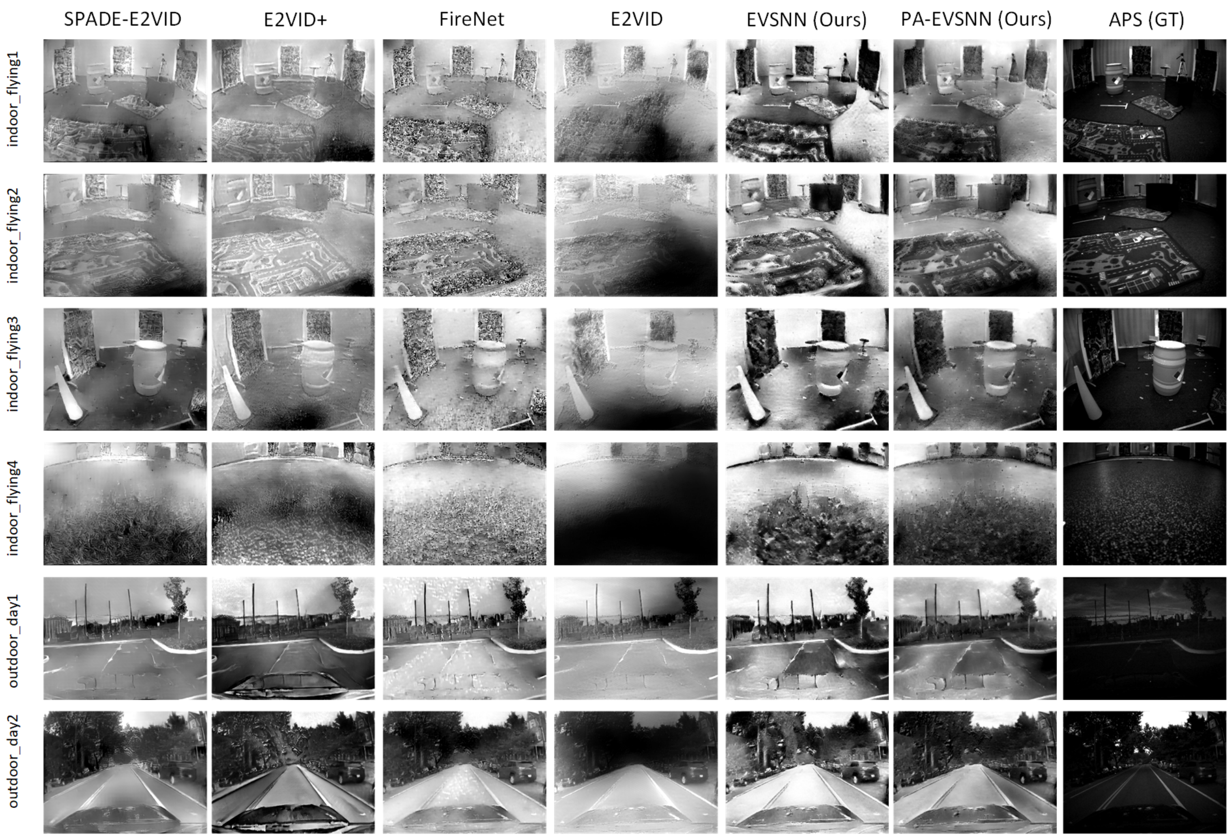

Figure 4. **Qualitative results of our models and ANN-based methods on MVSEC dataset.**