JUNE 18-22, 2023
CVPR
VANCOUVER, CANADA

SMAE: Few-shot Learning for HDR Deghosting with Saturation-Aware Masked Autoencoders

TUE-PM-157

Qingsen Yan[1], Song Zhang[2], Weiye Chen[2], Hao Tang[3], Yu Zhu[1],

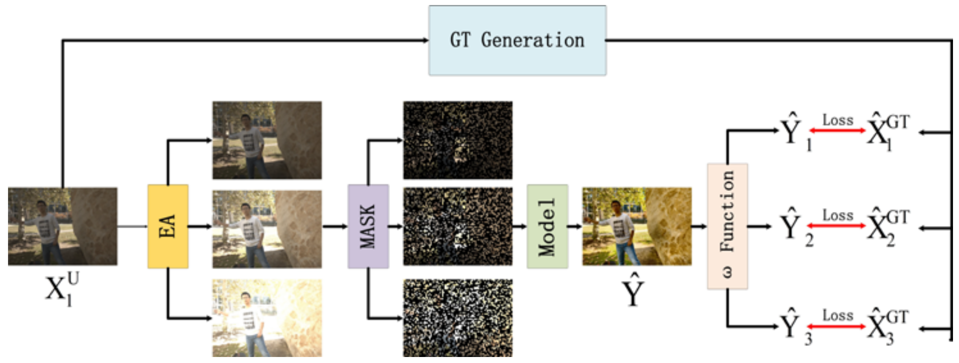
Jinqiu Sun[1], Luc Van Gool[3], Yanning Zhang[1]

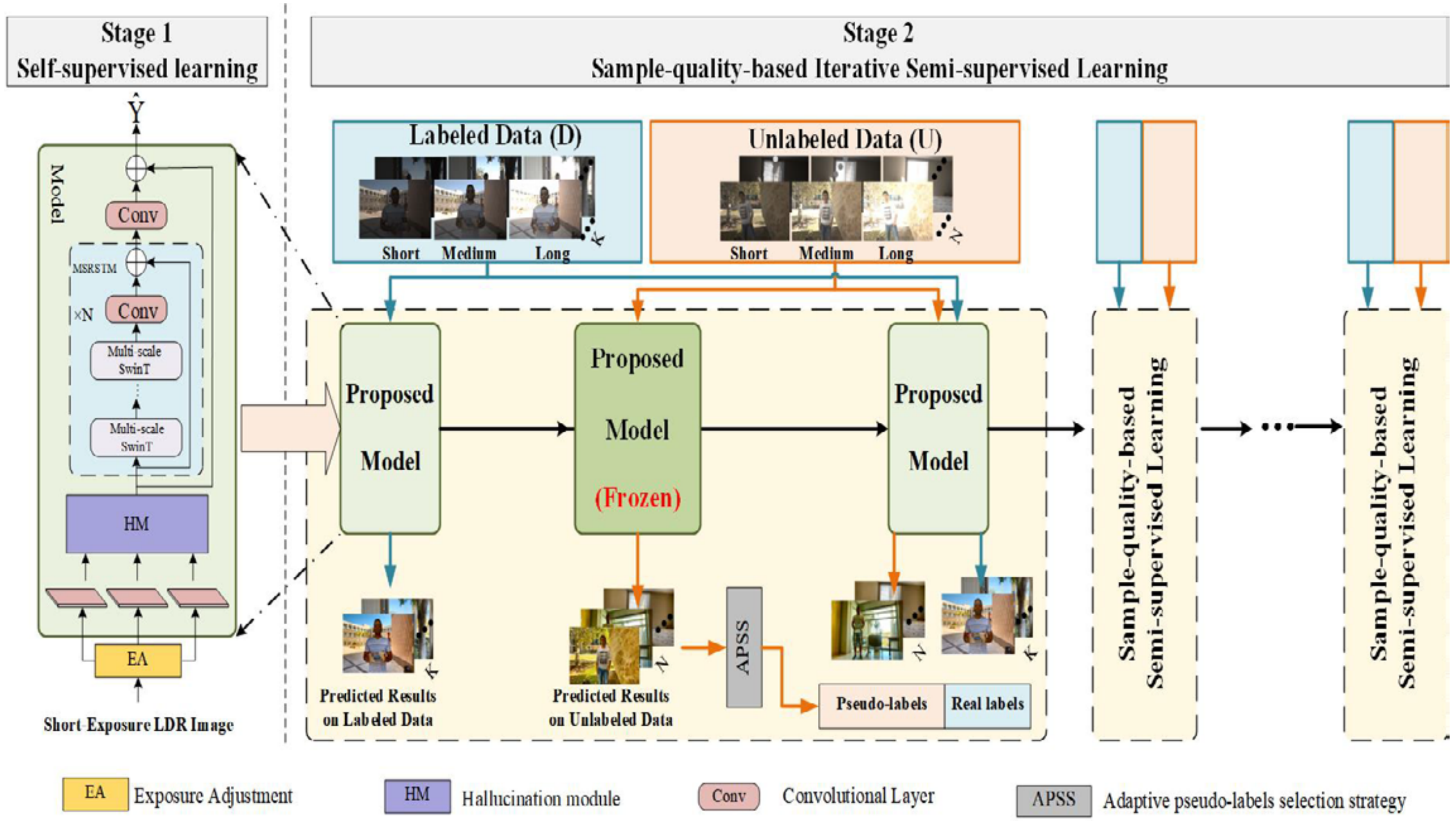
[1]Northwestern Polytechnical University, China, [2]Xidian University, China, [3]ETH Zurich, Switzerland

Summary

Problems

- CNN-based HDR deghosting methods require collecting large datasets with ground truth, which is a tedious and time-consuming process.
- Generating an HDR image on dynamic scenes with only few labels is challenging.





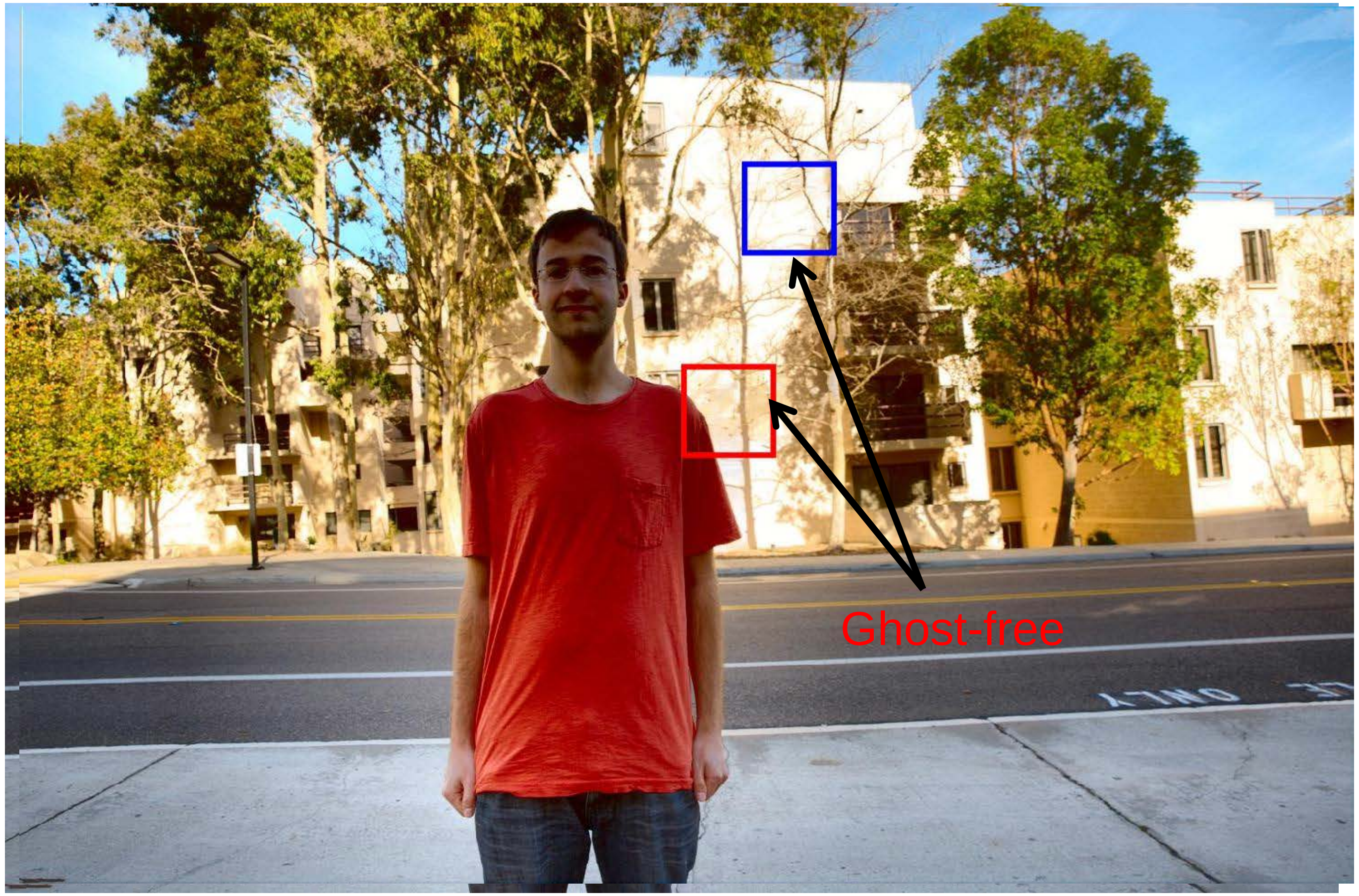
- In the first stage, we propose a multi-scale Transformer model based on self-supervised learning with a saturated-masked autoencoder to make it capable of recovering saturated regions.
- In the second stage, we propose a sample-quality-based iterative semi-supervised learning approach that learns to address ghosting problems.

Background

- **Few-shot HDR Reconstruction**

LDR Images

☑ Few Labeled samples ☑ Saturation ☑ Motion

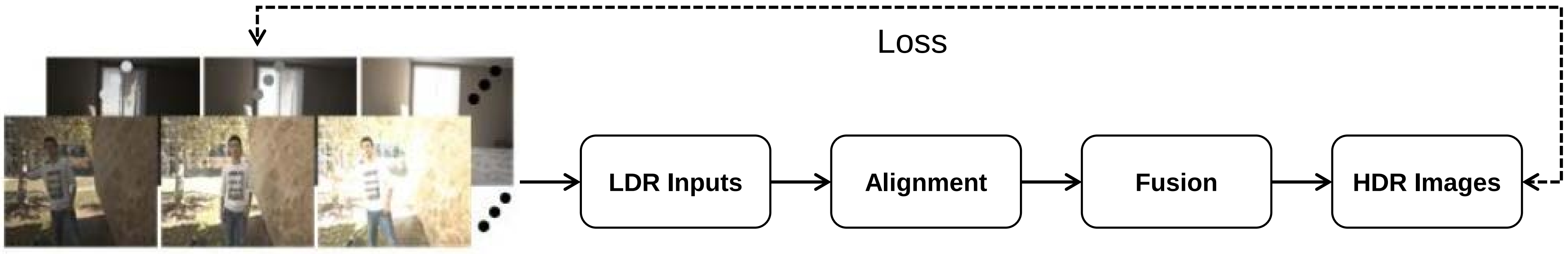


HDR Images with ghosts

Ghost-free HDR Images

Background

- **Existing Methods**



a large amount of labeled samples

It is challenging to collect a large amount of HDR-labeled samples

Reasons:

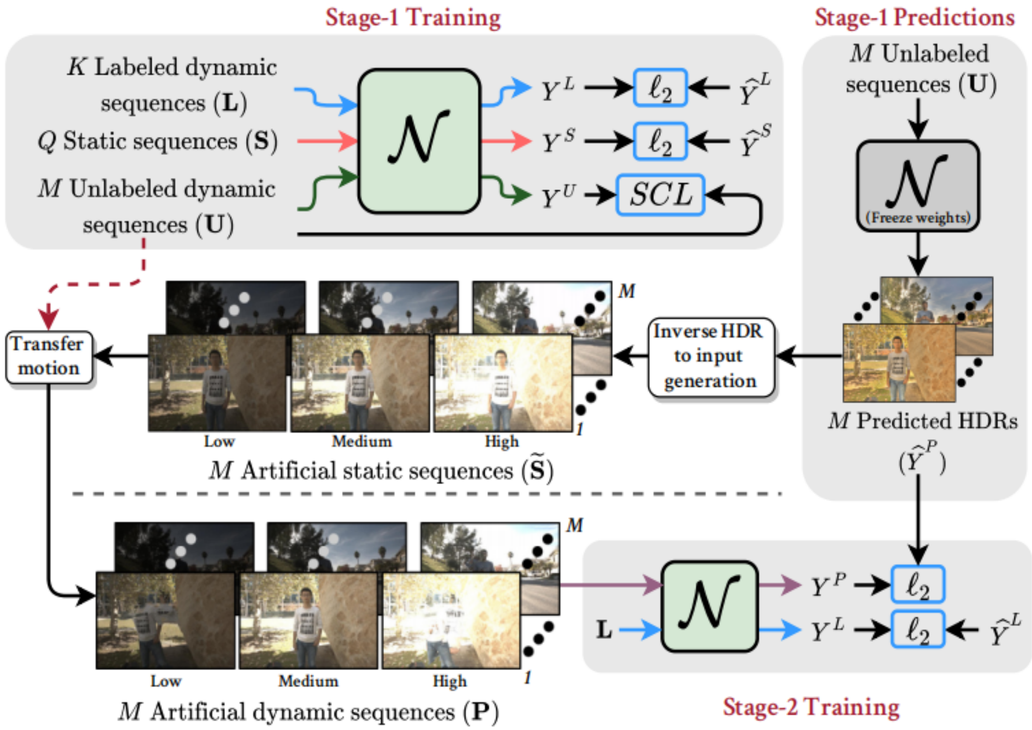
Generating a ghost-free HDR ground truth sample requires an absolute static background

It is time-consuming and requires considerable manpower to do manual post-examination

Background

- **Existing Methods**
  - **FSHDR**[1]

1. Train a preliminary model 2. Generate HDR pseudo-labels 3. Synthesize artificial dynamic LDR inputs



Hard to achieve under the condition of few labeled data

Especially in motion and saturation regions

☑ Handle both the saturation and the ghosting problems simultaneously

Optical flow is error-prone

Distribution shift between LDR training and testing data

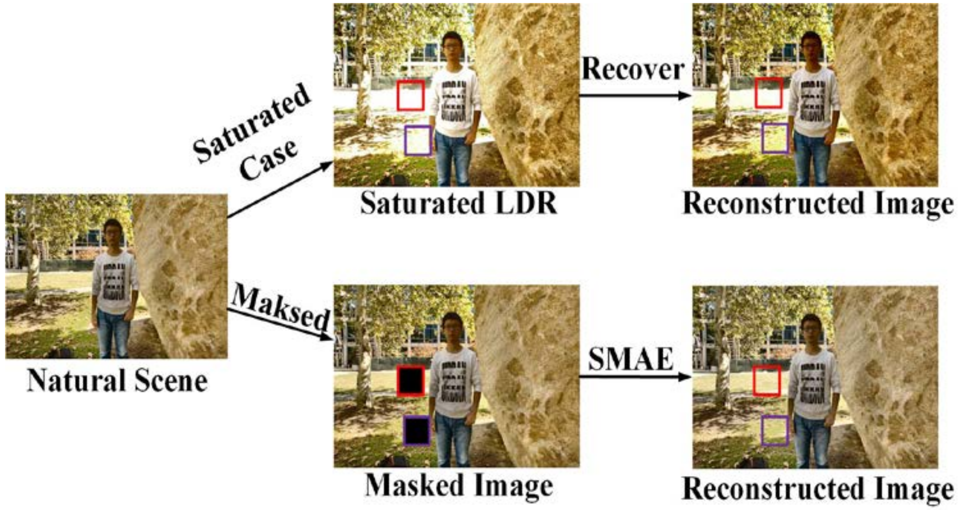
☑ Use optical flow to forcibly synthesize dynamic LDR inputs from poorly generated HDR pseudo-labels

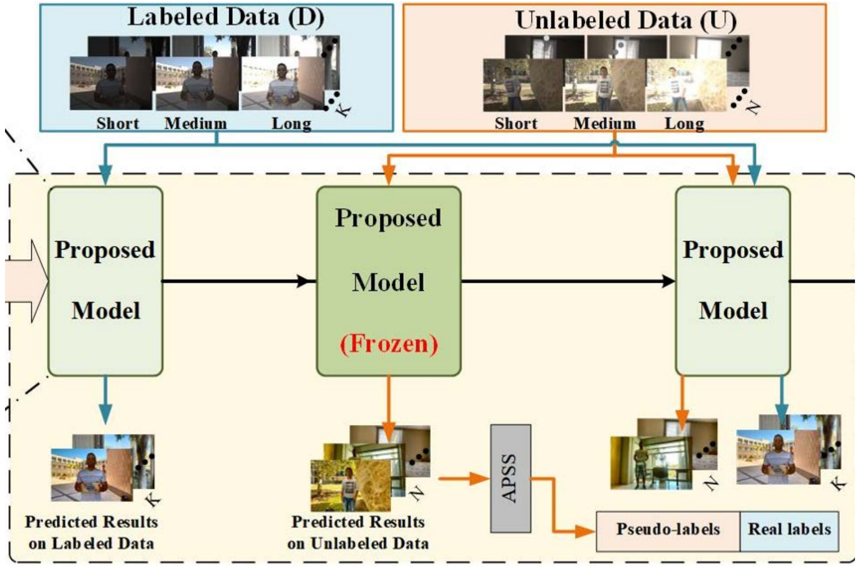
[1] *K Ram Prabhakar, et.al. Labeled from unlabeled: Exploiting unlabeled data for few-shot deep hdr deghosting. In CVPR, 2021.*

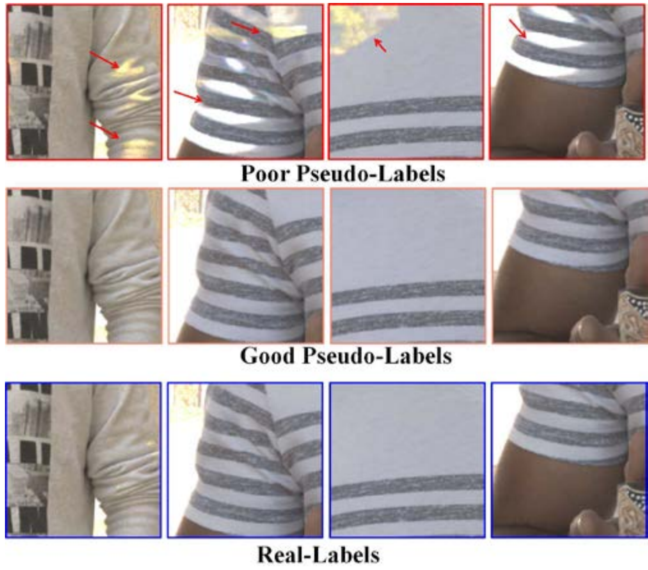
Movtivation

A reasonable way is to address the saturation problems first and then cope with the ghosting problems with a few labeled samples.

- Generate content of saturated regions by self-supervised learning
- Effectively use label data and unlabeled data.
- Select high-quality of pseudo-labels.



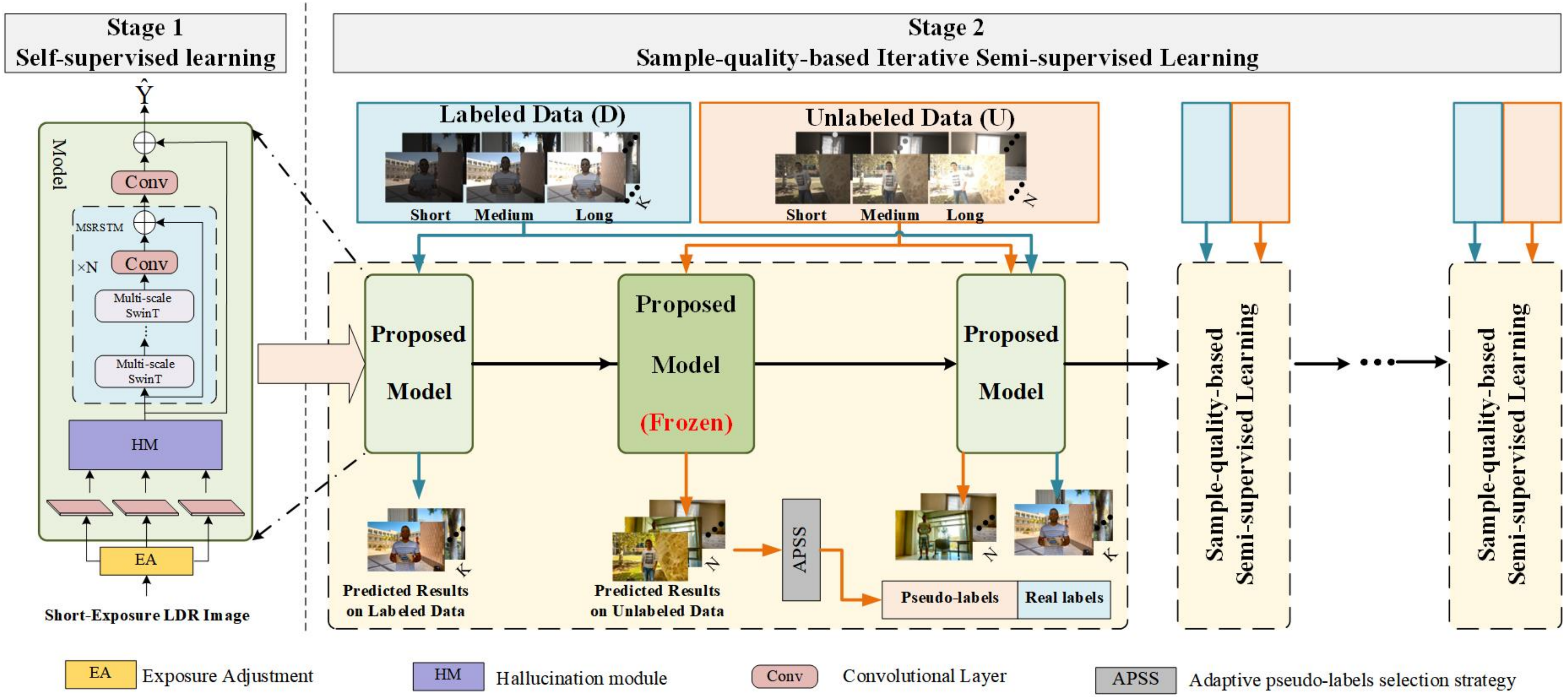




[2] *Kaiming He, et.al. Masked autoencoders are scalable vision learners. In CVPR, 2022.*

Our Method

- **Overview**

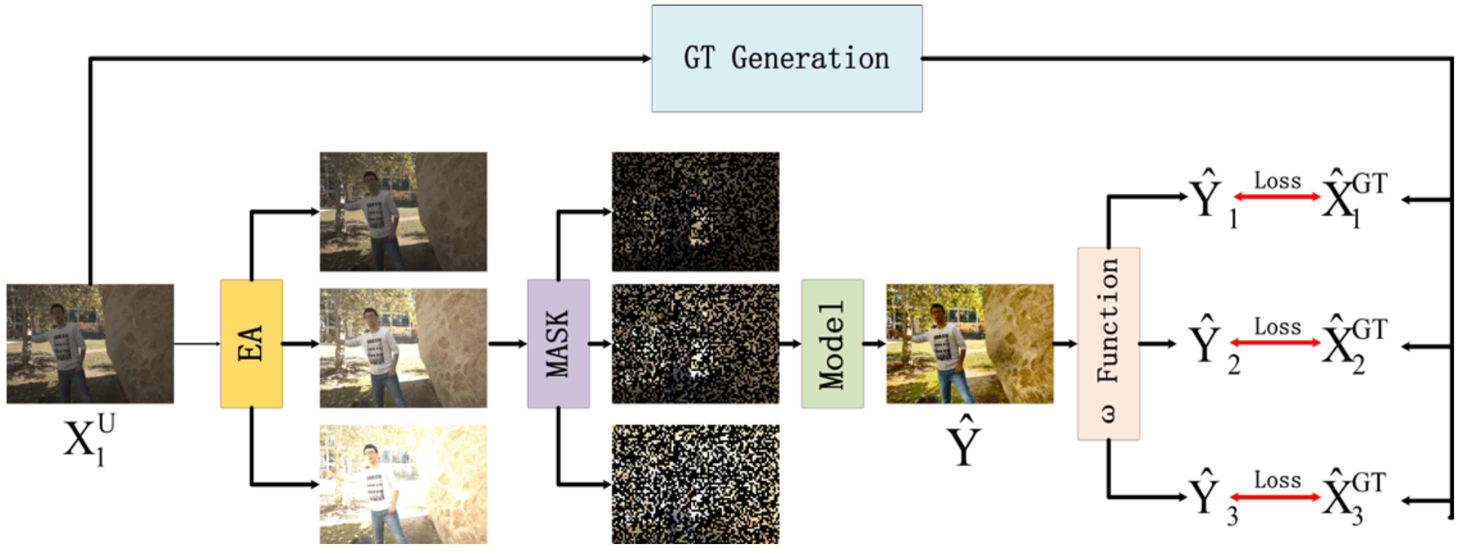


- In the first stage, we propose a multi-scale Transformer model based on self-supervised learning with a saturated-masked autoencoder to make it capable of recovering saturated regions.
- In the second stage, we propose a sample-quality-based iterative semi-supervised learning approach that learns to address ghosting problems.

Our Method

- ## Self-supervised Learning Stage

  - Use mask strategy to reconstruct an HDR image and addresses saturated problems from one LDR image.



EA Module.

$$X_i^U = clip\left(\left(\frac{(X_1^U)^\gamma \times t_i}{t_1}\right)^{\frac{1}{\gamma}}\right), i = 2,3. \ (1)$$

ω Function.

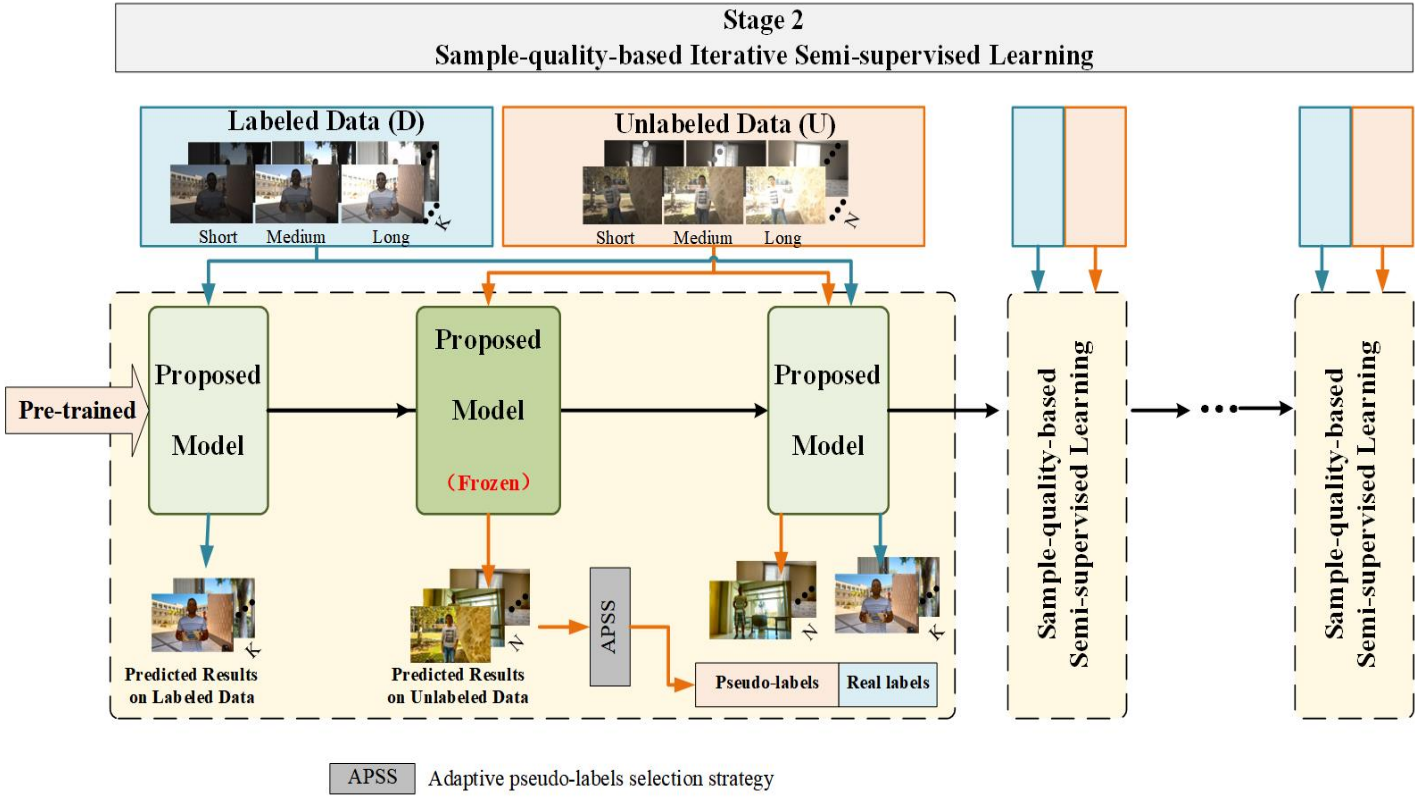
$$\hat{Y}_i = \omega(\hat{Y}) = (\hat{Y} \times t_i)^{\frac{1}{\gamma}}. \ (2)$$

GT Generation.

$$X_i^{GT} = \left(\frac{(X_1^U)^\gamma \times t_i}{t_1}\right)^{\frac{1}{\gamma}}, i = 1,2,3. \ (3)$$

Our Method

- **Semi-supervised Learning Stage**
  - We propose a sample-quality-based iterative semi-supervised learning approach that learns to address ghosting problems.



Finetune.

Iteration.

$$L_{Iteration} = L^{D}_{recon,t+1} + L^{S}_{recon,t+1} + \sum_{i=1}^{N} W^{U_i}_{t+1} L^{U_i}_{recon,t+1}$$

$$+\lambda(L^{D}_{percep,t+1} + L^{S}_{percep,t+1} + \sum_{i=1}^{N} W^{U_i}_{t+1} L^{U_i}_{percep,t+1}), (4)$$

APSS.

$$\tau_t = \sigma(L^{D\cup S}_{select,t}, \beta). (5)$$

$$L^{U}_{select,t} = ||mask(\omega(\hat{Y}^{U}_{t})) - mask(\omega(X^{U}_{2,t}))||_1, (6)$$

$$m^{U}_{t} = max(L^{U}_{select,t}), (7)$$

$$W^{U_i}_{t+1} = \begin{cases} 1 & L^{U_i}_{select,t} \leq \tau_t \\ \dfrac{m^{U}_{t} - L^{U_i}_{select,t}}{m^{U}_{t} - \tau_t} & L^{U_i}_{select,t} > \tau_t \end{cases} \quad (8)$$

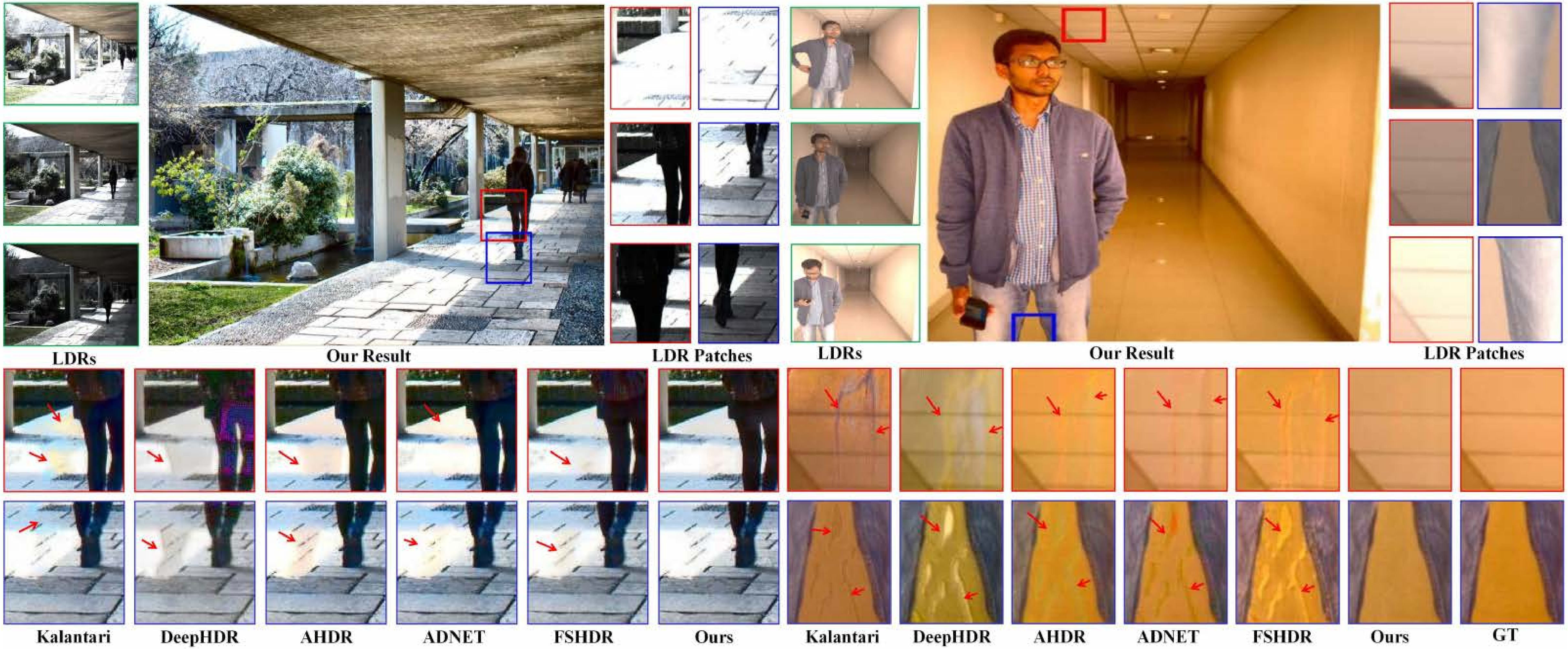
Experiment

- **Individual Datasets.**

(a) Kalantari's dataset (b) Hu's dataset

Experiment

- **Cross Datasets.**



(c) Tursun's dataset (d) Prabhakar's dataset

Experiment

Quantitative Result

Table 1. The evaluation results on Kalantari's[3] and Hu's[4] datasets. The best and the second best results are highlighted in **Bold** and <u>Underline</u>, respectively.

Dataset	Metric	Setting	Kalantari	DeepHDR	AHDRNet	ADNet	FSHDR	Ours
Kalantari	PSNR-l	5way-5shot	39.37±0.12	38.25±0.29	40.61±0.10	40.78±0.15	<u>41.39</u>±0.12	**41.54**±0.10
	PSNR-μ		39.86±0.19	38.62±0.27	41.05±0.32	40.93±0.38	<u>41.40</u>±0.13	**41.61**±0.08
	PSNR-l	5way-1shot	36.94±0.44	36.67±0.67	38.83±0.39	38.96±0.35	<u>41.04</u>±0.11	**41.14**±0.11
	PSNR-μ		37.33±1.21	37.01±1.68	39.15±1.04	39.08±1.06	<u>41.13</u>±0.07	**41.25**±0.05
Hu	PSNR-l	5way-5shot	41.36±0.25	40.73±0.66	46.37±0.76	46.88±0.81	<u>47.13</u>±0.13	**47.41**±0.12
	PSNR-μ		38.95±0.14	39.92±0.22	43.42±0.44	43.79±0.48	<u>43.98</u>±0.27	**44.24**±0.17
	PSNR-l	5way-1shot	38.67±0.43	37.82±0.86	44.64±0.80	44.75±0.84	<u>44.94</u>±0.23	**45.04**±0.16
	PSNR-μ		36.83±0.62	38.49±1.07	42.37±1.42	42.41±1.20	<u>42.50</u>±0.87	**42.55**±0.44

[3] *N. K. Kalantari, et.al. Deep high dynamic range imaging of dynamic scenes. In ACM TOG, 2017.*

[4] *Jinhan Hu, et.al. Sensor-realistic synthetic data engine for multi-frame high dynamic range photography. In CVPRW, 2020.*

Experiment

- **Quantitative Result**

		Kalantari					Hu				
		PSNR-l	PSNR-μ	SSIM-l	SSIM-μ	HV2	PSNR-l	PSNR-μ	SSIM-l	SSIM-μ	HV2
S_1	Sen	38.57	40.94	0.9711	0.9780	64.71	33.58	31.48	0.9634	0.9531	66.39
	Hu	30.84	32.19	0.9408	0.9632	62.05	36.94	36.56	0.9877	0.9824	67.58
	FSHDR	40.97	41.11	0.9864	0.9827	67.08	42.15	41.14	0.9904	0.9891	71.35
	Ours (K=0)	**41.12**	**41.20**	**0.9866**	**0.9868**	**67.16**	**42.99**	**41.30**	**0.9912**	**0.9903**	**72.18**
S_2	Ours (K=1)	41.14	41.25	0.9866	0.9869	67.20	45.04	42.55	0.9938	0.9928	73.23
	Ours (K=5)	41.54	41.61	0.9879	0.9880	67.33	47.41	44.24	0.9974	0.9936	74.49
S_3	Kalantari	41.22	41.85	0.9848	0.9872	66.23	43.76	41.60	0.9938	0.9914	72.94
	DeepHDR	40.91	41.64	0.9863	0.9857	67.42	41.20	41.13	0.9941	0.9870	70.82
	AHDRNet	41.23	41.87	0.9868	0.9889	67.50	49.22	45.76	0.9980	0.9956	75.04
	ADNET	41.31	41.80	0.9871	0.9883	67.57	**50.38**	46.79	0.9987	0.9948	**76.32**
	FSHDR	**41.79**	41.92	0.9876	0.9851	67.70	49.56	45.90	0.9984	0.9945	75.25
	Ours	41.68	**41.97**	**0.9889**	**0.9895**	**67.77**	50.31	**46.88**	**0.9988**	**0.9957**	76.21
S_4	Kalantari	25.87	21.44	0.8610	0.9176	60.00	10.23	16.95	0.6903	0.8346	49.10
	DeepHDR	25.92	21.43	0.8597	0.9170	60.02	25.48	20.86	0.9215	0.8354	66.83
	AHDRNet	26.62	22.08	0.8737	0.9238	58.89	11.44	17.84	0.6732	0.8389	52.79
	ADNET	25.76	21.39	0.8686	0.8217	60.36	10.86	18.09	0.6915	0.8399	49.28
	FSHDR	**28.03**	22.01	0.8751	0.9203	60.53	12.82	19.37	0.7442	0.8347	55.34
	Ours	27.91	**22.45**	**0.8764**	**0.9252**	**61.02**	**30.29**	**21.56**	**0.9440**	**0.8456**	**67.07**
S_5	Kalantari	31.24	33.10	0.9527	0.9593	63.99	19.82	18.63	0.7679	0.8742	59.50
	DeepHDR	30.75	29.01	0.9244	0.9223	63.26	19.84	18.70	0.7698	0.8752	59.48
	AHDRNet	31.84	33.49	**0.9588**	0.9606	64.40	**20.80**	20.51	0.8259	0.9136	**59.79**
	ADNET	31.08	33.50	0.9536	0.9636	63.88	20.78	20.80	**0.8268**	0.9173	59.71
	FSHDR	32.70	32.24	0.9553	0.9465	64.37	20.23	19.71	0.7929	0.9026	59.63
	Ours	**32.72**	**34.49**	0.9586	**0.9713**	**64.45**	20.69	**21.96**	0.8257	**0.9207**	59.76

Experiment

- **Ablation study**

Table 3. Ablation study of 5 shot scenario on Kalantari's dataset.

#	Model	PSNR-l	PSNR-μ	HDR-VDP-2
B1	SSHDR	**41.54**	**41.61**	**67.33**
B2	Stage2Net	41.31	41.43	67.21
B3	w/o APSS	41.49	41.45	67.29
B4	AHDR*	41.48	41.51	67.30
B5	FSHDR*	41.41	41.43	67.26
B6	Vanilla-AHDR	40.61	41.05	66.95
B7	Vanilla-FSHDR	41.39	41.40	67.25

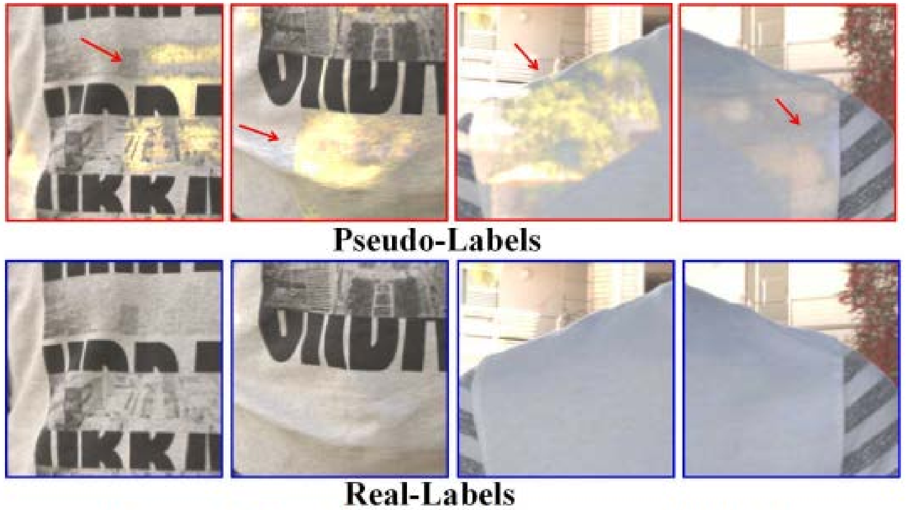


Figure 5. Visual results of poor pseudo-labels.