

# PPR10K: A Large-Scale Portrait Photo Retouching Dataset with Human-Region Mask and Group-Level Consistency

Jie Liang, Hui Zeng, Miaomiao Cui, Xuansong Xie, Lei Zhang

Dataset and code: <https://github.com/cslliang/PPR10K> Email: liang27jie@gmail.com



## Introduction



Special and practical requirements of portrait photo retouching (PPR):

- 1) **Human-Region Priority (HRP)**: more attention be paid to human regions
- 2) **Group-Level Consistency (GLC)**: a group of portrait photos be retouched to a consistent tone

Existing datasets (FiveK) are for general purpose without the HRP and GLC consideration.

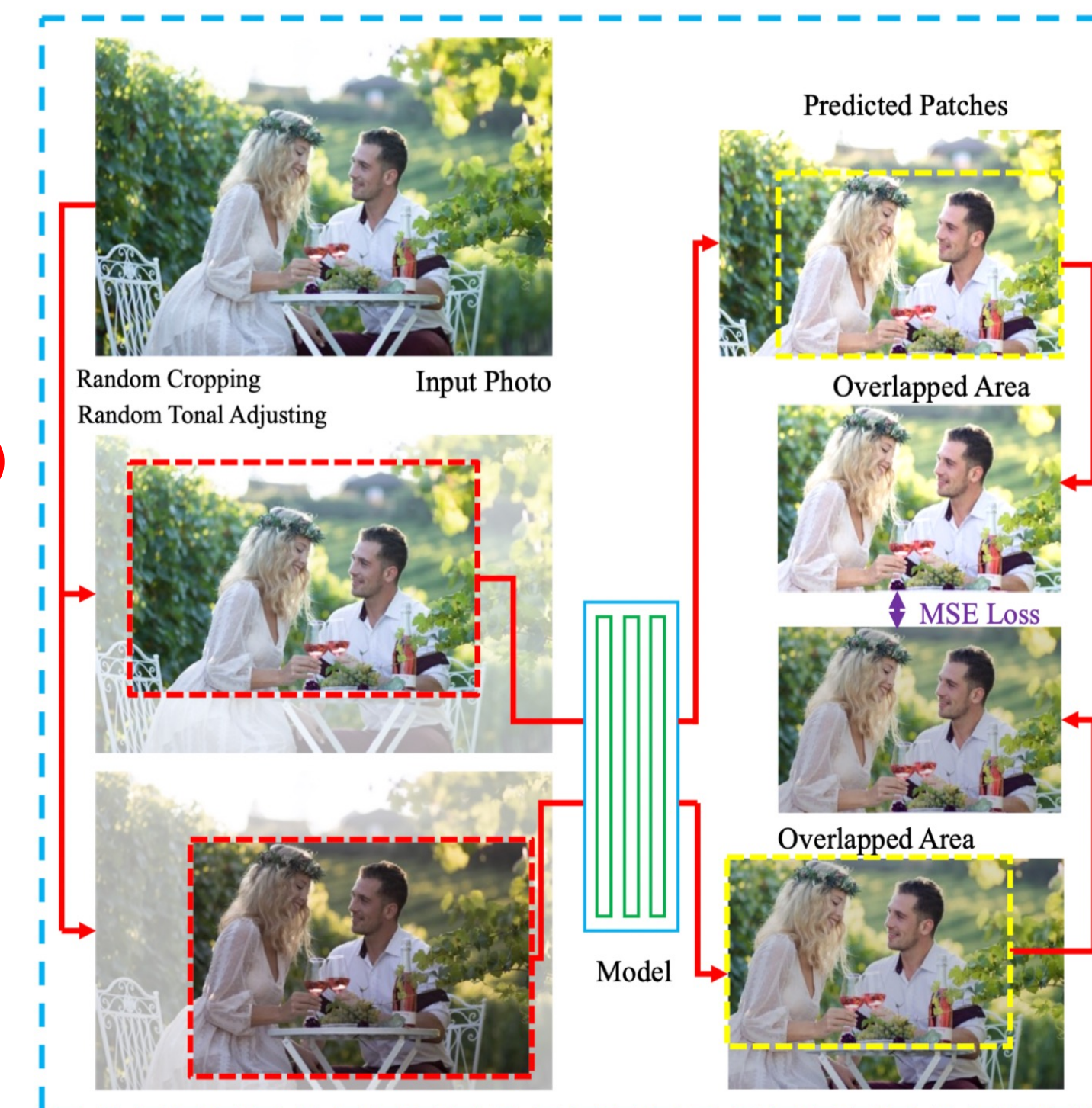
## Measures and Learning Strategies

**Measures:**

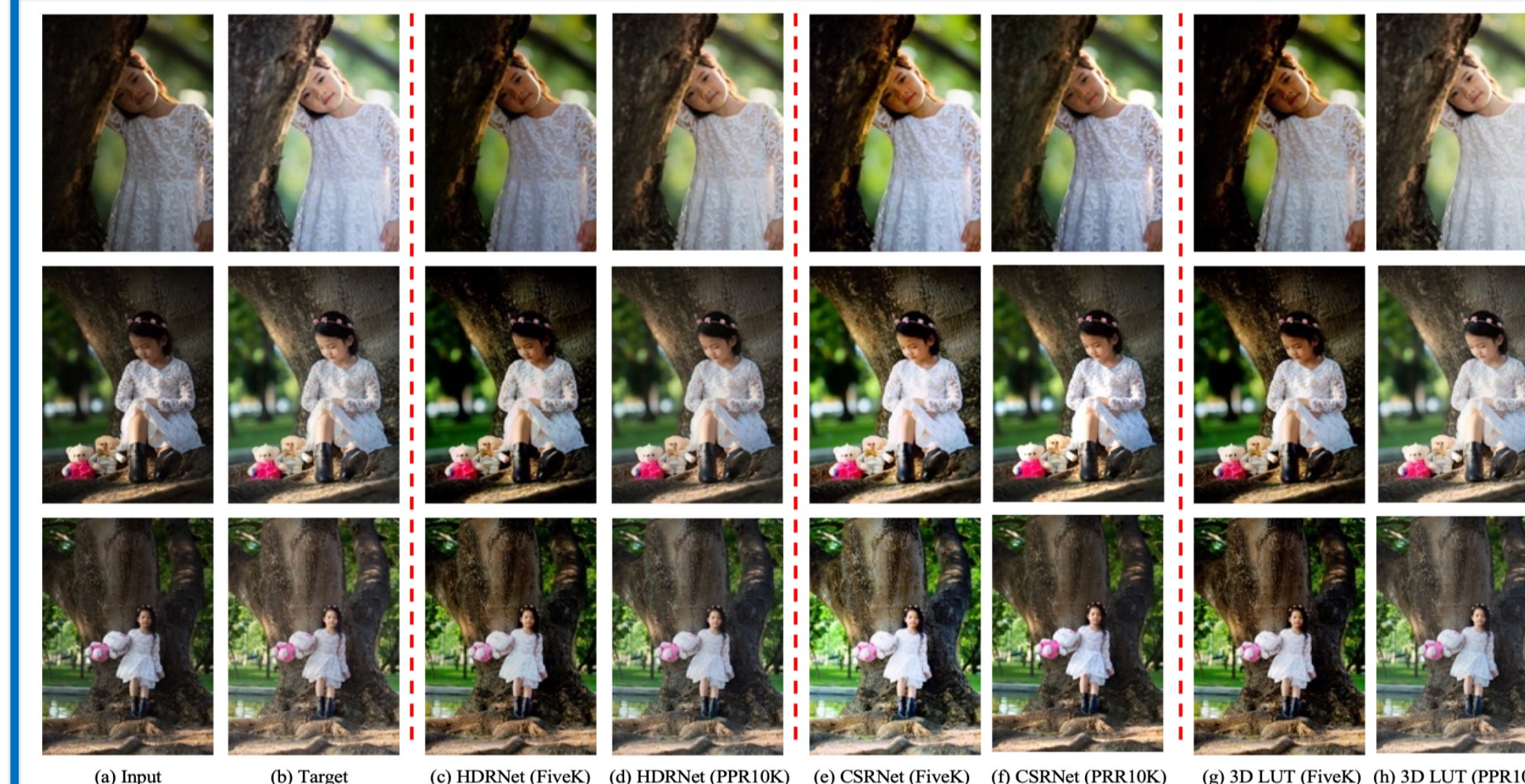
- 1) **Basic**:  $PSNR$  and  $\Delta E_{ab} = \|\hat{I}^{Lab} - Y^{Lab}\|_2$ ,  $\hat{I}^{Lab}$  denotes the predicted photo in Lab color space,  $Y$  denotes the target.
  - 2) **Human-Centered** Measures:  $PSNR^{HC}$  and  $\Delta E_{ab}^{HC}$ , with weight 0.5 in background regions
  - 3) **Group-Level Consistency** Measure: The variance of mean color components.  $\sum_c Var(\mu_{i_1^c}, \mu_{i_2^c}, \dots, \mu_{i_m^c})$
- Most stable: the combination of  $a$  and  $b$  channels

**Learning Strategies:**

- 1) **HRP**: MSE loss with weight 5 in human regions
- 2) **GLC**: As shown in the figure, crop two partially overlapped patches and distort to mimic images in a group. Constraining overlapped region using MSE.



## Experiments



Models training on FiveK obtain much worse performance than trained on the PPR10K dataset: **large domain gap!**

#	Method	Dataset	$PSNR \uparrow$	$\Delta E_{ab} \downarrow$	$PSNR^{HC} \uparrow$	$\Delta E_{ab}^{HC} \downarrow$	$M_{GLC} \downarrow$
1	HDRNet [7]	PPR10K-a	23.93	23.06	8.70	9.13	27.21
2	CSRNet [10]	PPR10K-a	22.72	22.01	9.75	10.20	25.90
3	3D LUT [32]	PPR10K-a	25.64	25.15	6.97	7.25	28.89
4	3D LUT+HRP	PPR10K-a	25.99	25.55	6.76	7.02	28.29
5	3D LUT+GLC	PPR10K-a	25.06	24.39	7.39	7.81	28.34
6	3D LUT+HRP+GLC	PPR10K-a	25.31	24.60	7.30	7.75	28.56
7	HDRNet [7]	PPR10K-b	23.96	23.51	8.84	9.13	27.21
8	CSRNet [10]	PPR10K-b	23.76	23.29	8.77	9.28	27.01
9	3D LUT [32]	PPR10K-b	24.70	24.30	7.71	7.97	27.99
10	3D LUT+HRP	PPR10K-b	25.06	24.66	7.51	7.73	28.36
11	3D LUT+GLC	PPR10K-b	24.16	23.39	8.15	8.70	27.48
12	3D LUT+HRP+GLC	PPR10K-b	24.52	23.81	7.93	8.42	27.82
13	HDRNet [7]	PPR10K-c	24.08	23.66	8.87	9.05	27.32
14	CSRNet [10]	PPR10K-c	23.17	22.85	9.45	9.87	26.47
15	3D LUT [32]	PPR10K-c	25.18	24.78	7.58	7.85	28.49
16	3D LUT+HRP	PPR10K-c	25.46	25.05	7.43	7.69	28.80
17	3D LUT+GLC	PPR10K-c	24.53	23.94	8.10	8.49	27.87
18	3D LUT+HRP+GLC	PPR10K-c	24.59	24.01	8.02	8.39	27.92

Overall comparison. **3D LUT** performs better than HDRNet and CSRNet. **HRP loss** brings better results on most individual measures. **GLC loss** improves the GLC measure yet slightly deteriorates the others. **Combining** them further improves GLC.

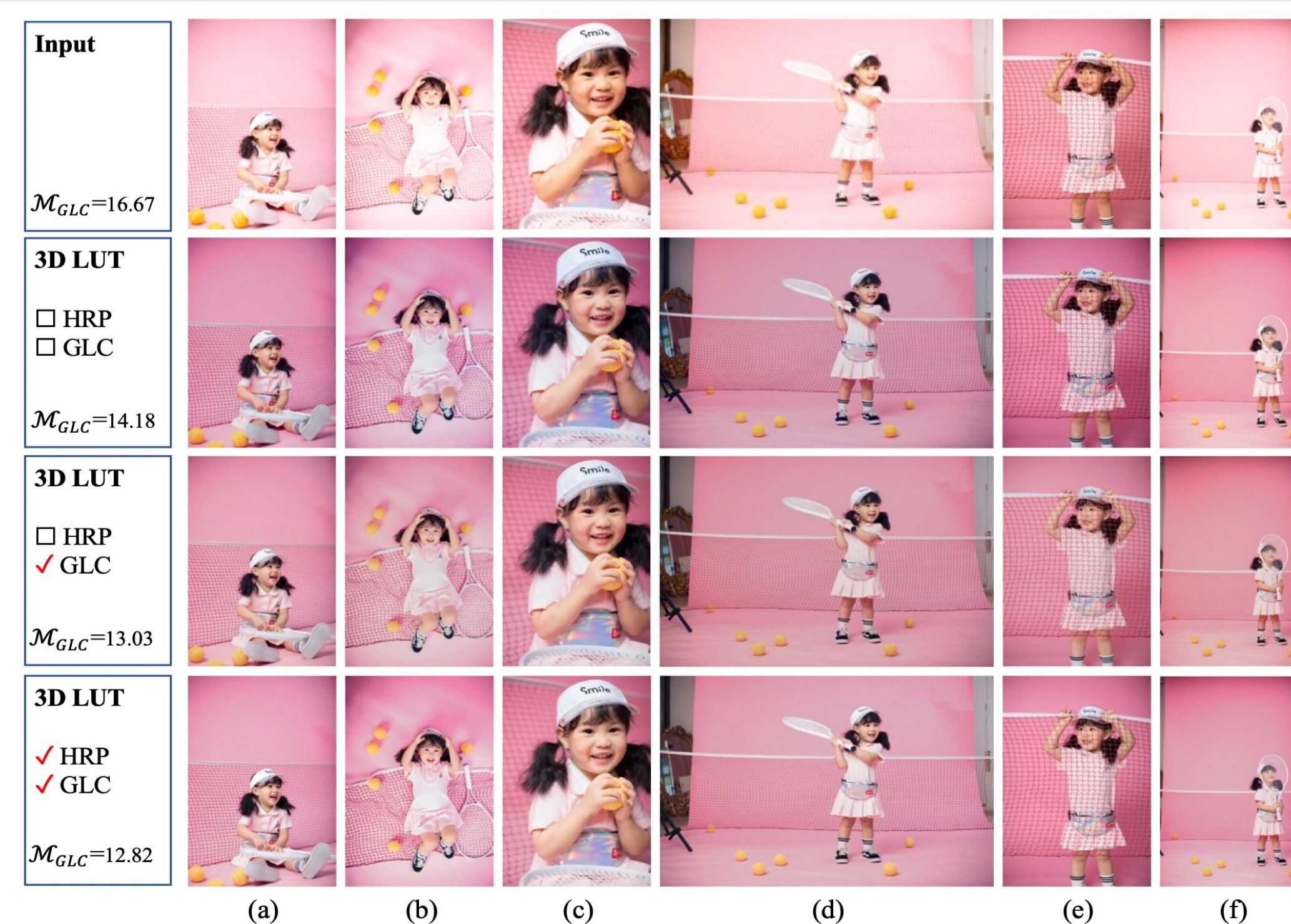
## Dataset Construction

**Challenges:**

- 1) photos should in raw format with high-quality, which is difficult to obtain
- 2) dataset should be large-scale and cover a wide range of real cases
- 3) high-quality retouched results with both good visual quality and group-level consistency and human-region masks should be provided

**Information:**

- 1) **11161 raw-format** portraits in **1681** groups (each group contains 3-18 photos)
- 2) **3 versions of GTs** for each photo manually retouched by 3 experts
- 3) Resolution ranging from **4K-8K**, DSLR **cameras** covering Canon, Nikon and Sony
- 4) Full-resolution human-region **masks** with high-quality



Visually more consistent in a group with **GLC** strategy!

Method	Dataset	$PSNR$	$\Delta E_{ab}$	$PSNR^{HC}$	$\Delta E_{ab}^{HC}$	$M_{GLC}$
HDRNet	a	18.20	17.22	21.44	11.27	20.76
CSRNet	a	19.86	14.07	23.06	9.15	13.97
3D LUT	a	19.92	13.75	23.79	8.90	13.85
HDRNet	b	18.74	16.31	22.00	10.63	20.76
CSRNet	b	19.65	14.47	22.83	9.40	13.97
3D LUT	b	19.74	14.08	23.42	9.22	13.85
HDRNet	c	19.71	14.81	22.96	9.65	20.76
CSRNet	c	19.81	14.57	23.06	9.46	13.97
3D LUT	c	20.03	13.90	23.01	8.85	13.85

Poor performance when training on FiveK and evaluating on PPR tasks!



Brighter faces and more natural temperature with **HRP** strategy!

## Conclusion

A large-scale PPR dataset is constructed, with 11161 diverse raw portraits in 1681 groups. Human-region priority and group-level consistency measures and learning strategies are theoretically demonstrated and experimentally evaluated.