

A. Additional Qualitative Results

In this section we show more qualitative results on illumination enhancement and photo retouching to compare STAR and CNN backbones. The codes and full results will be released in the future.

A.1. Illumination Enhancement Results

Fig. 1 give additional illumination results on FiveK [1] datasets. The same to Section 4.1, DCE-Net and DCE-Net_D here denote original pixel-wise implementation of [2] and its downsampled form (predicting token-wise curve like STAR). As Fig. 1 shows, compared to CNN backbone, STAR can yield high quality results which are much closer to the references, with only 13% float operations.

A.2. Photo Retouching Results

In Fig. 2 we show more photo retouching results. As shown, compared to [4], semantic methods deliver both higher quantitative results and better image perceptual quality. Note that CNN_{semantic} methods can achieve comparable results to the proposed STAR, but at cost of much more computations (0.29G FLOPS vs. 0.04G FLOPS).

B. Typo Correction

We are sorry that there's a typo in Figure.6 of the submitted main text. The Complexity of CNN_{semantic} is 0.29G FLOPS (the same to Table.5) rather than 0.21G GFLOPS. We will correct it in future version.

References

- [1] Vladimir Bychkovsky, Sylvain Paris, Eric Chan, and Frédo Durand. Learning photographic global tonal adjustment with a database of input/output image pairs. In *CVPR 2011*, pages 97–104. IEEE, 2011. 1, 2
- [2] Chunle Guo, Chongyi Li, Jichang Guo, Chen Change Loy, Junhui Hou, Sam Kwong, and Runmin Cong. Zero-reference deep curve estimation for low-light image enhancement. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 1780–1789, 2020. 1
- [3] Samuel W Hasinoff, Dillon Sharlet, Ryan Geiss, Andrew Adams, Jonathan T Barron, Florian Kainz, Jiawen Chen, and Marc Levoy. Burst photography for high dynamic range and low-light imaging on mobile cameras. *ACM Transactions on Graphics (TOG)*, 35(6):1–12, 2016. 3
- [4] Hui Zeng, Jianrui Cai, Lida Li, Zisheng Cao, and Lei Zhang. Learning image-adaptive 3d lookup tables for high performance photo enhancement in real-time. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2020. 1

108
109
110
111
112
113
114
115
116
117
118
119
120
121
122
123
124
125
126
127
128
129
130
131
132
133
134
135
136
137
138
139
140
141
142
143
144
145
146
147
148
149
150
151
152
153
154
155
156
157
158
159
160
161

162
163
164
165
166
167
168
169
170
171
172
173
174
175
176
177
178
179
180
181
182
183
184
185
186
187
188
189
190
191
192
193
194
195
196
197
198
199
200
201
202
203
204
205
206
207
208
209
210
211
212
213
214
215

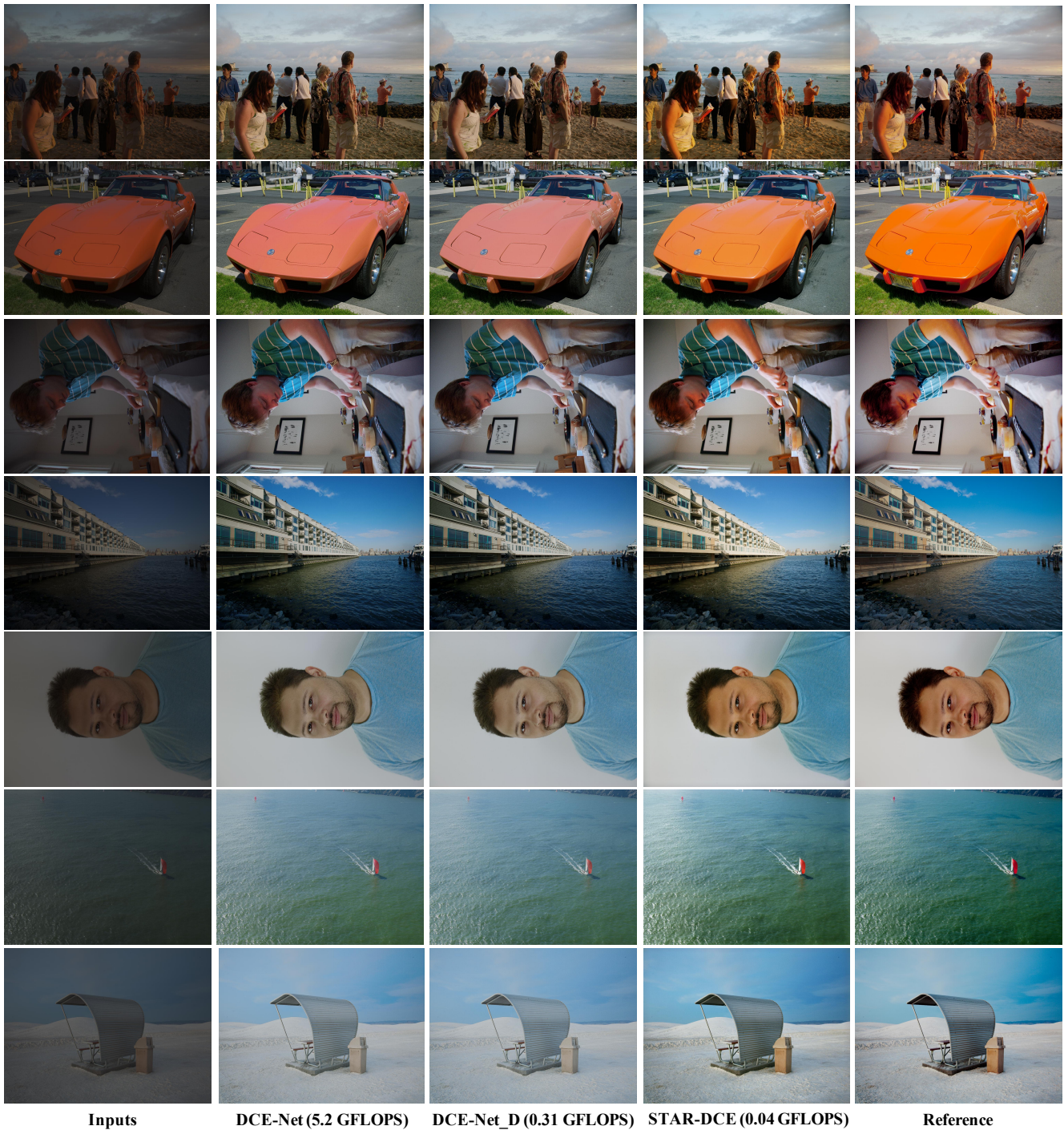


Figure 1. Visual comparison of illumination enhancement on MIT-Adobe FiveK [1].

216
217
218
219
220
221
222
223
224
225
226
227
228
229
230
231
232
233
234
235
236
237
238
239
240
241
242
243
244
245
246
247
248
249
250
251
252
253
254
255
256
257
258
259
260
261
262
263
264
265
266
267
268
269

270
271
272
273
274
275
276
277
278
279
280
281
282
283
284
285
286
287
288
289
290
291
292
293
294
295
296
297
298
299
300
301
302
303
304
305
306
307
308
309
310
311
312
313
314
315
316
317
318
319
320
321
322
323

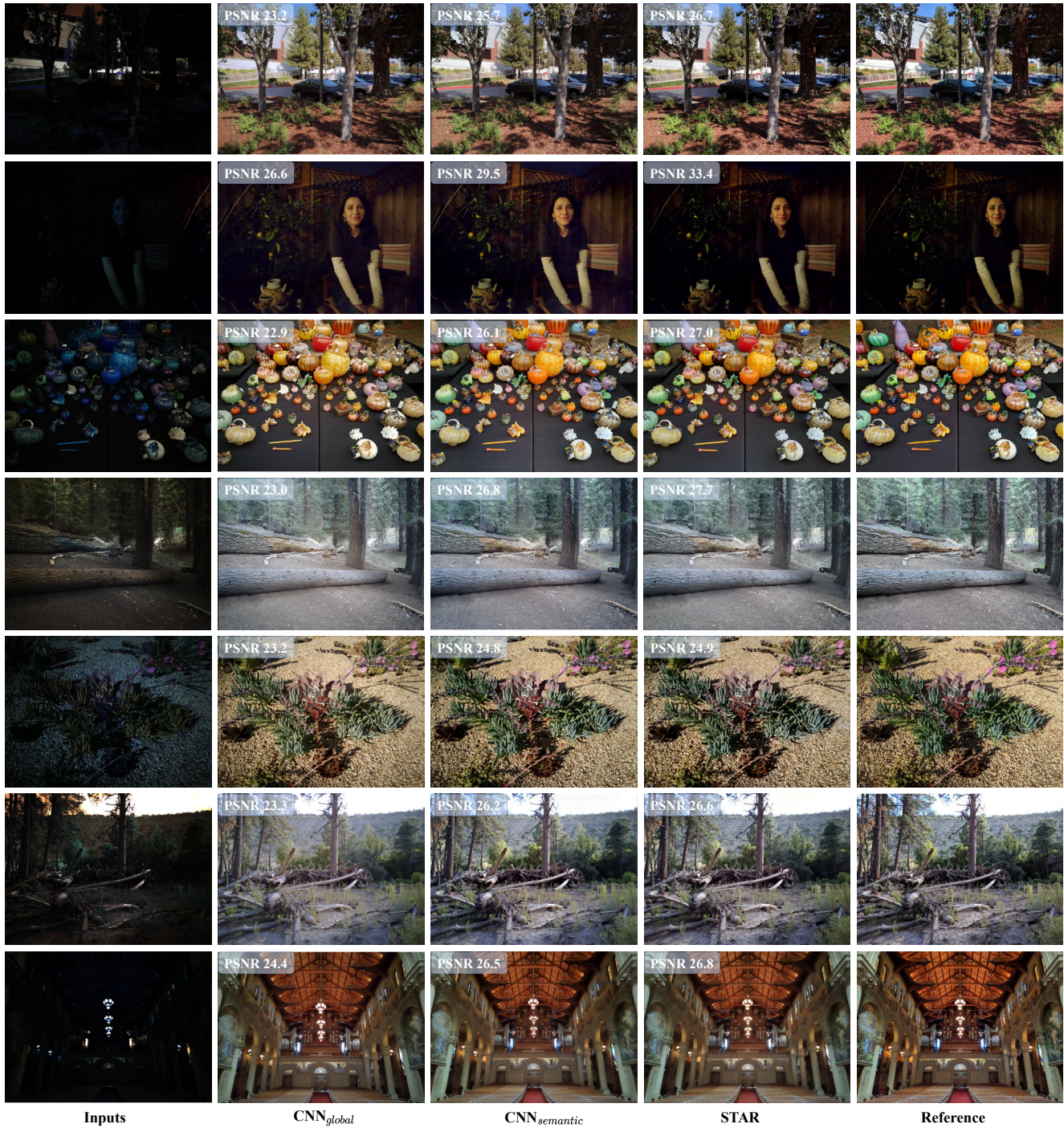


Figure 2. Visual comparison of photo retouching on HDR+ [3]. We mark PSNR on all results.