Interpreting Intrinsic Image Decomposition using Concept Activations

Supplementary Material: ICVGIP, 2022

Avani Gupta
CVIT, KCIS
IIIT-Hyderabad
avani.gupta@research.iiit.ac.in

Saurabh Saini
CVIT, KCIS
IIIT-Hyderabad
saurabh.saini@research.iiit.ac.in

P J Narayanan
CVIT, KCIS
IIIT-Hyderabad
pjn@iiit.ac.in

We first provide the raw sensitivity scores obtained in our experiments, concept set temperature details followed by detailed comparison of models over existing IID evaluation metrics MSE, LMSE, D-SSIM. We then provide the qualitative results over different temperatures in concept sets and additional results on Δ_a experiments. We finally provide results on all of our four experimental settings, ARAP[2] dataset and real-world concept sets datasets: MIW[6], PS[1] along with MIT Intrinsics[3] in supplementary videos.

1 RAW SENSITIVITY SCORES

We report the raw sensitivity scores of experiments in Table 4. Note: the range of scores is between 0-1, as mentioned in the paper. We observe that these scores are high for both R and S by models in some experiments. The concept captured by CAV vectors is dependent on the model's activations and some models might be affected by that concept for both R and S and hence the raw sensitivity scores by themselves do not provide much information about the importance given by the model for $\hat{R}$ vs $\hat{S}$. Note: TCAV does analysis in classification problems and thus calculates sensitivity over classifiers while we are using it in a decomposition(reconstruction problem) for a multi-branch network. Hence the sensitivity scores are standalone for classifier setting as by Kim et al. [5] but not for our problem of comparing outputs of multi-branch network. We are interested in the R *vs.* S sensitivity scores, hence the ratios of scores matter for us and not the raw scores.

2 CONCEPT SETS TEMPERATURES

We experiment with three temperatures for concept sets which are shown in Figure 1. These temperatures have been inspired from temperature settings of vidit dataset [4]. We report qualitative results over our three temperature settings in Figure 2 where $A_i I_j T_k$ represent a scene having albedo i, illumination j and temperature k. k = 0 for T = 2500, k = 1 for T = 4500 and k = 2 for T = 6500.

We observe that all three models confuse temperature with albedo. We also verify our CSM_S and CSM_R scores (given in paper) for different temperatures from qualitative results. Overall, for Δ_a USI3D>IIWW>CGIID, for Δ_i CGIID>IIWW>USI3D. For T = 2500 and 4500 the trend according to CSM_S is USI3D>IIWW»CGIID while T = 6500 has trend as IIWW>USI3D>CGIID.

3 METRICS COMPARISON

We report the D-SSIM, LMSE and MSE metrics over MIT Intrinsics dataset[3], ARAP[2] and our newly introduced Δ_a and Δ_i concept sets in Table 1, Table 2 and Table 3 respectively. Each of these metrics measures different aspects of closeness to Ground Truth. MSE measures the average squared pixel-wise difference between predicted and GT. LMSE is local-MSE and measures MSE patch-wise, while D-SSIM gives structural dis-similarity between predicted and GT images. MSE measures absolute error, not taking spatial information of pixels into account, whereas LMSE and D-SSIM consider spatially close pixels separately. These metrics are not designed to measure R *vs.* S disentanglement as pointed in the paper. According to these metrics Table 2, USI3D has best $\hat{S}$ while IIWW has best $\hat{R}$ on ARAP dataset[2]. (LMSE $\hat{R}$ for CGIID and IIWW are comparable). On MIT Intrinsics[3] all models have comparable performance Table 1 and there is no common trend of performance established as such.

Model	MSE↓		LMSE↓		DSSIM↓	
	R	S	R	S	R	S
IIWW	**0.0147**	0.0135	0.0341	0.0253	0.1398	**0.1266**
USI3D	0.0156	**0.0102**	0.064	0.0474	**0.1158**	0.131
CGIID	0.0167	0.0127	**0.0319**	**0.0211**	0.1287	0.1376

Table 1: Pixel-wise comparison metrics on MIT Intrinsics datataset[3]

Model	MSE↓		LMSE↓		D-SSIM↓	
	R	S	R	S	R	S
IIWW	**0.056**	0.033	0.066	0.054	**0.448**	0.522
USI3D	0.095	**0.021**	0.072	**0.052**	0.486	**0.347**
CGIID	0.073	0.037	**0.064**	0.054	0.512	0.498

Table 2: Pixel-wise comparison metrics on 42 scenes of ARAP dataset[2] used as test set in paper.

Figure 1: Illumination temperatures: We use three temperatures for illumination which are 2500, 4500, 6500 as shown in from left to right.

Scene type	Albedo type	Model	Δ_a concept set						Δ_i concept set					
			R			S			R			S		
			MSE	**LMSE**	**D-SSIM**	**MSE**	**LMSE**	D-SSIM	**MSE**	**LMSE**	**D-SSIM**	**MSE**	**LMSE**	**D-SSIM**
Simple	RGB	IIWW	0.116	0.027	0.277	0.053	0.006	0.409	0.145	0.031	0.339	0.021	0.006	0.415
		USI3D	**0.06**	**0.025**	**0.213**	0.041	**0.001**	**0.234**	**0.067**	**0.021**	**0.23**	0.035	**0.004**	**0.198**
		CGIID	0.134	0.033	0.344	**0.014**	0.002	0.265	0.183	0.04	0.426	**0.012**	0.005	0.23
	Textured	IIWW	0.057	0.012	0.208	0.075	0.005	0.348	0.076	0.015	0.337	0.071	**0.004**	0.355
		USI3D	**0.026**	**0.008**	**0.138**	0.054	0.005	0.326	**0.041**	**0.008**	**0.284**	0.045	0.007	0.31
		CGIID	0.062	0.016	0.255	**0.018**	**0.002**	**0.287**	0.08	0.019	0.359	**0.016**	0.005	**0.269**
Complex	RGB	IIWW	0.07	0.014	0.246	0.057	0.006	0.382	0.09	0.016	0.306	0.059	0.006	0.398
		USI3D	**0.024**	**0.007**	**0.19**	0.039	**0.003**	**0.268**	**0.034**	**0.006**	**0.228**	0.035	**0.004**	**0.23**
		CGIID	0.082	0.019	0.294	**0.022**	**0.003**	0.308	0.128	0.027	0.396	**0.02**	0.005	0.272
	Textured	IIWW	0.044	0.014	0.25	0.064	0.006	0.383	0.048	0.014	0.322	0.112	**0.005**	**0.414**
		USI3D	**0.019**	**0.009**	**0.184**	0.062	0.007	0.4	**0.031**	**0.01**	**0.31**	0.071	0.01	0.449
		CGIID	0.046	0.014	0.284	**0.02**	**0.005**	**0.372**	0.05	0.013	0.335	**0.024**	0.007	0.421

Table 3: Pixel-wise comparison metrics on scenes of our concept sets: USI3D does best in general in terms of the above metrics.

Temp	Model	Δ_a								Δ_i							
		Textured				**RGB**				**Textured**				**RGB**			
		Simple		**Complex**		**Simple**		**Complex**		**Simple**		**Complex**		**Simple**		**Complex**	
		R_{Δ_a}	S_{Δ_a}	R_{Δ_a}	S_{Δ_a}	R_{Δ_a}	S_{Δ_a}	R_{Δ_a}	S_{Δ_a}	R_{Δ_i}	S_{Δ_i}	R_{Δ_i}	S_{Δ_i}	R_{Δ_i}	S_{Δ_i}	R_{Δ_i}	S_{Δ_i}
2500	IIWW	0.34	0.196	0.309	0.278	0.336	0.376	0.338	0.237	0.326	0.333	0.307	0.337	0.349	0.347	0.346	0.262
	USI3D	0.331	0.08	0.311	0.111	0.337	0.082	0.296	0.112	0.37	0.166	0.36	0.146	0.375	0.103	0.285	0.184
	CGIID	0.53	0.588	0.71	0.515	0.347	0.626	0.613	0.583	0.518	0.604	0.422	0.596	0.419	0.637	0.353	0.586
4500	IIWW	0.385	0.193	0.366	0.237	0.385	0.403	0.383	0.204	0.379	0.305	0.335	0.328	0.387	0.343	0.389	0.3
	USI3D	0.294	0.125	0.335	0.149	0.267	0.099	0.271	0.176	0.325	0.21	0.352	0.175	0.301	0.116	0.304	0.196
	CGIID	0.421	0.565	0.671	0.503	0.445	0.63	0.532	0.591	0.447	0.602	0.468	0.6	0.346	0.643	0.333	0.62
6500	IIWW	0.395	0.158	0.378	0.303	0.386	0.4	0.386	0.196	0.384	0.293	0.34	0.299	0.387	0.36	0.391	0.271
	USI3D	0.263	0.241	0.275	0.214	0.264	0.128	0.277	0.221	0.257	0.242	0.323	0.239	0.285	0.146	0.315	0.229
	CGIID	0.365	0.597	0.648	0.507	0.353	0.633	0.487	0.582	0.375	0.609	0.468	0.606	0.3	0.644	0.327	0.61
Avg	IIWW	0.373	0.182	0.351	0.273	0.369	0.393	0.369	0.213	0.363	0.311	0.327	0.322	0.374	0.35	0.375	0.278
	USI3D	0.296	0.148	0.307	0.158	0.289	0.103	0.282	0.169	0.318	0.206	0.345	0.187	0.321	0.122	0.301	0.203
	CGIID	0.439	0.583	0.676	0.509	0.382	0.63	0.544	0.585	0.448	0.605	0.453	0.601	0.355	0.641	0.338	0.605

Table 4: TCAV sensitivity scores for concepts albedo change and illumination change in our 4 experimental settings.

4 ADDITIONAL RESULTS

We provide additional results on Δ_a concept for each of our 4 experimental settings for T = 6500, 4500 and 2500 in Figures 3, 4 and 5 respectively.

Please refer Supplementary video for more results.

REFERENCES

[1] Neil Alldrin, Todd Zickler, and David Kriegman. 2008. Photometric stereo with non-parametric and spatially-varying reflectance. In *Computer Vision and Pattern Recognition (CVPR)*.

[2] Nicolas Bonneel, Balazs Kovacs, Sylvain Paris, and Kavita Bala. 2017. Intrinsic Decompositions for Image Editing. *Computer Graphics Forum (Eurographics State of The Art Report)* (2017).

[3] Roger Grosse, Micah K Johnson, Edward H Adelson, and William T Freeman. 2009. Ground truth dataset and baseline evaluations for intrinsic image algorithms. In *International Conference on Computer Vision (ICCV)*. IEEE, 2335–2342.

[4] Majed El Helou, Ruofan Zhou, Johan Barthas, and Sabine Süsstrunk. 2020. VIDIT: Virtual image dataset for illumination transfer. *arXiv preprint arXiv:2005.05460* (2020).

[5] Been Kim, Martin Wattenberg, Justin Gilmer, Carrie Cai, James Wexler, Fernanda Viegas, et al. 2018. Interpretability beyond feature attribution: Quantitative testing with concept activation vectors (tcav). In *International Conference on Machine Learning (ICML)*. PMLR.

[6] Lukas Murmann, Michael Gharbi, Miika Aittala, and Fredo Durand. 2019. A Multi-Illumination Dataset of Indoor Object Appearance. In *2019 IEEE International Conference on Computer Vision (ICCV)*.

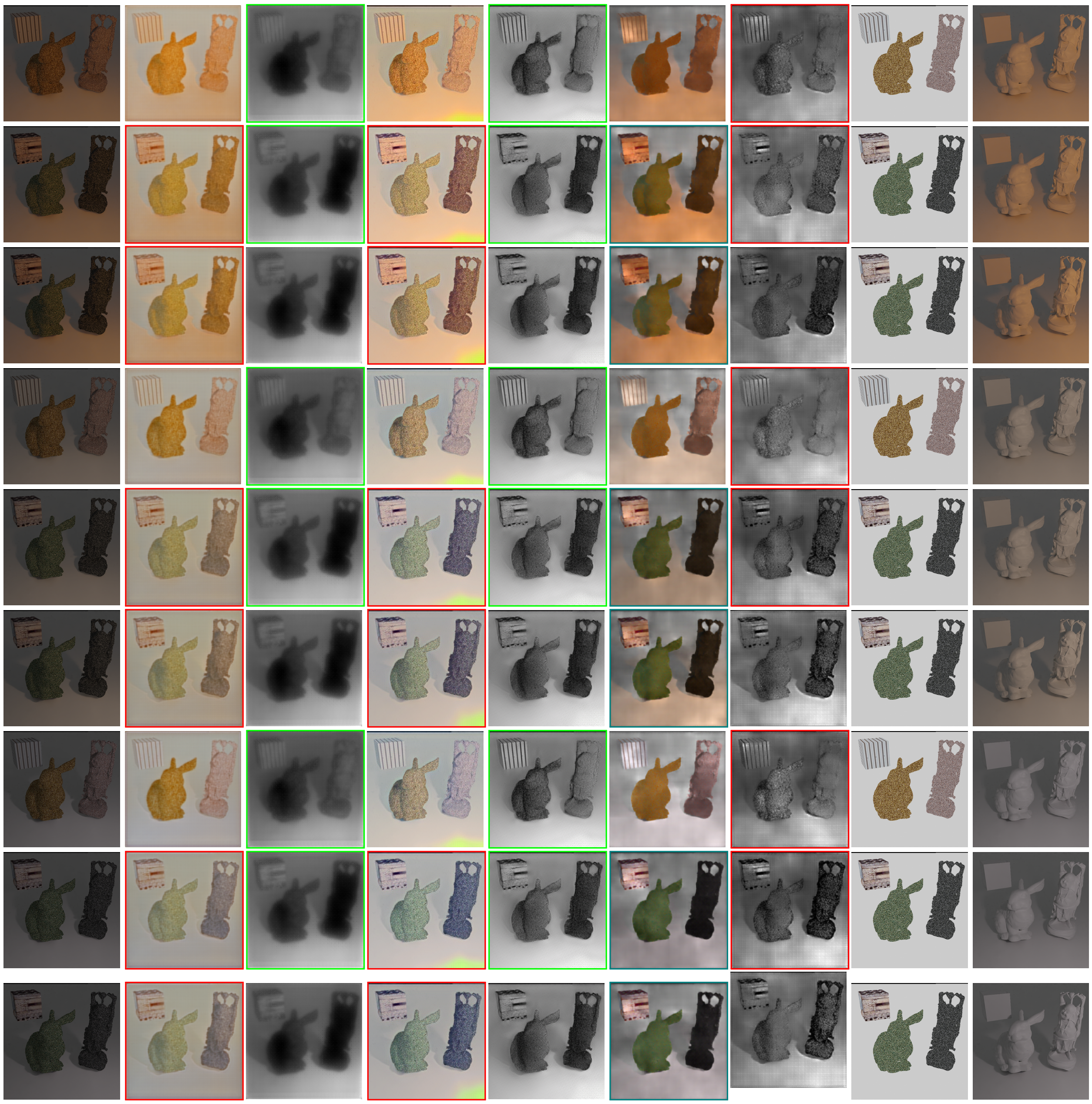
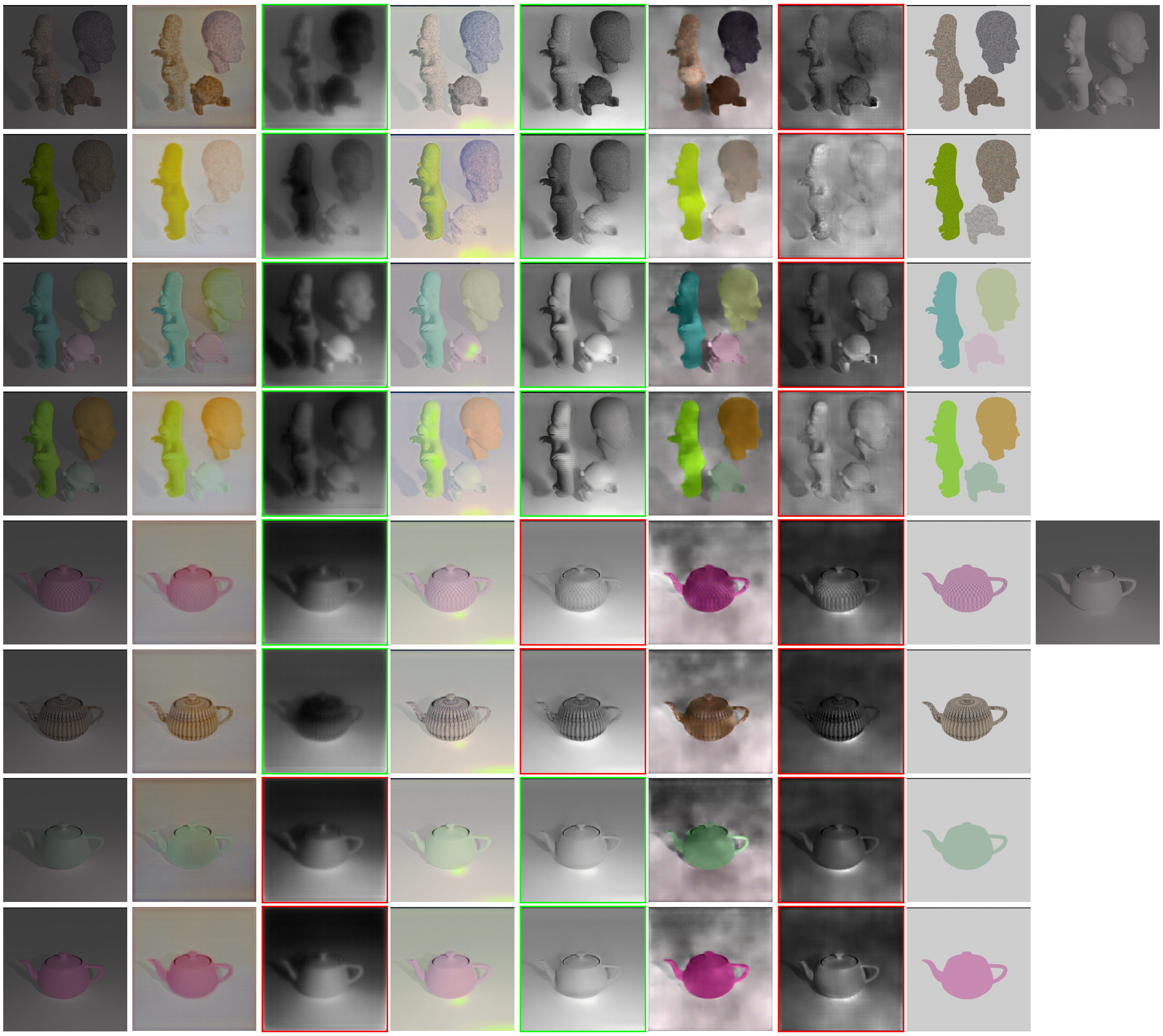
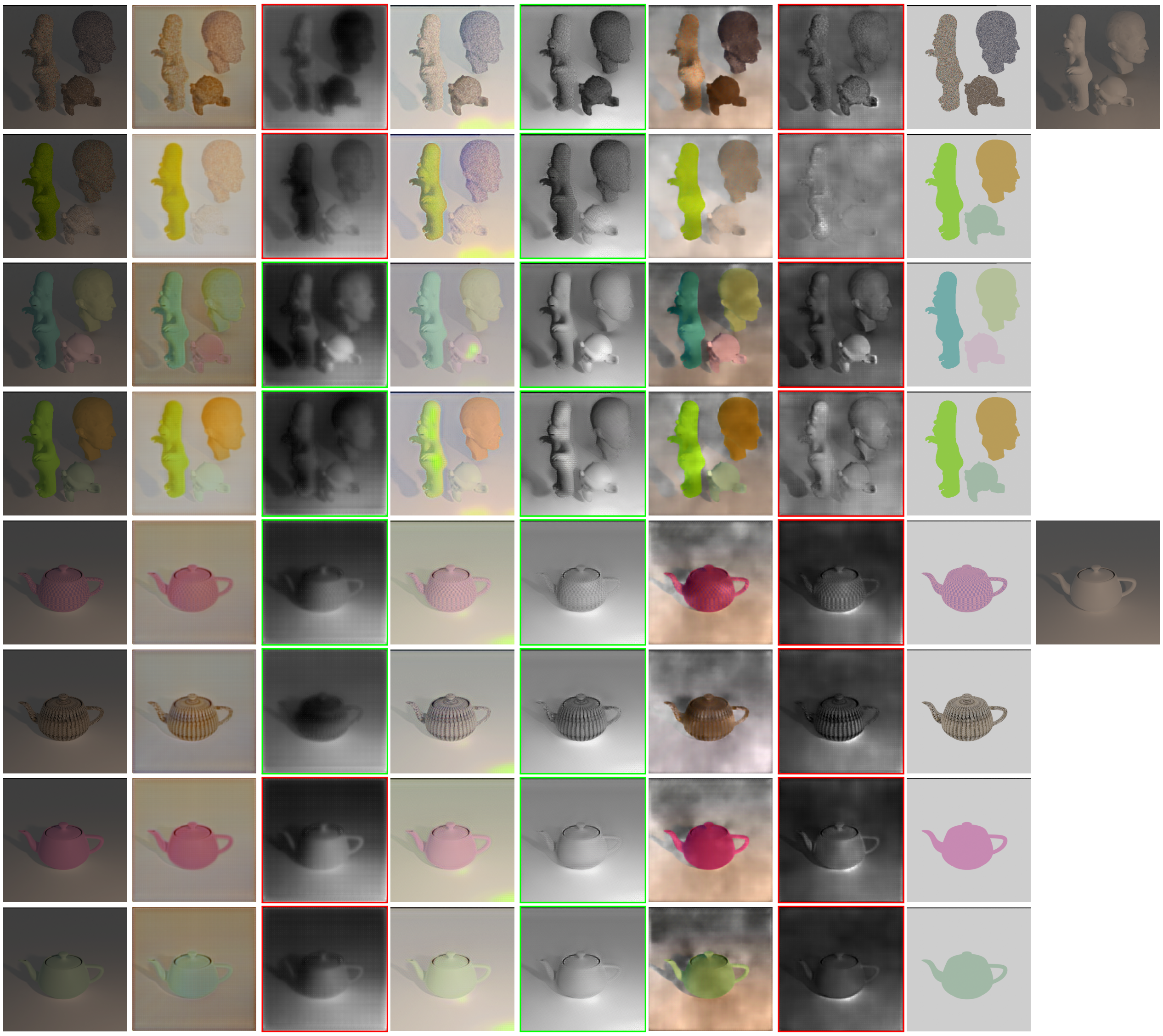


Figure 2: Illustrative qualitative results for albedo and illumination variation experiments in different temperatures. $A_iI_jT_k$ represent scene having albedo i, illumination j and temperature k. For albedo variation $A_0 \longrightarrow A_1$, in temperatures T_0 and T_1 rows first to second $A_0I_0T_0 \longrightarrow A_1I_0T_0$ and fourth to fifth $A_0I_0T_1 \longrightarrow A_1I_0T_1$, USI3D has least $\hat{S}$ changes (green) followed by IIWW and CGIID (red) while for temperature T_2 ($A_0I_0T_2 \longrightarrow A_1I_0T_2$, IIWW has least $\hat{S}$ changes (green) followed by USI3D while CGIID has most $\hat{S}$ changes (red). For illumination variation $I_0 \longrightarrow I_1$, the same trend CGIID>IIWW>USI3D is observed for all the temperatures. CGIID has least $\hat{R}$ changes for Δ_i (teel) followed by IIWW and USI3D which has most $\hat{R}$ changes (magenta). Further, CGIID has lesser illumination leakage in $\hat{R}$ for all three rows while IIWW and USI3D have clear illumination leakages (shadows) in R.



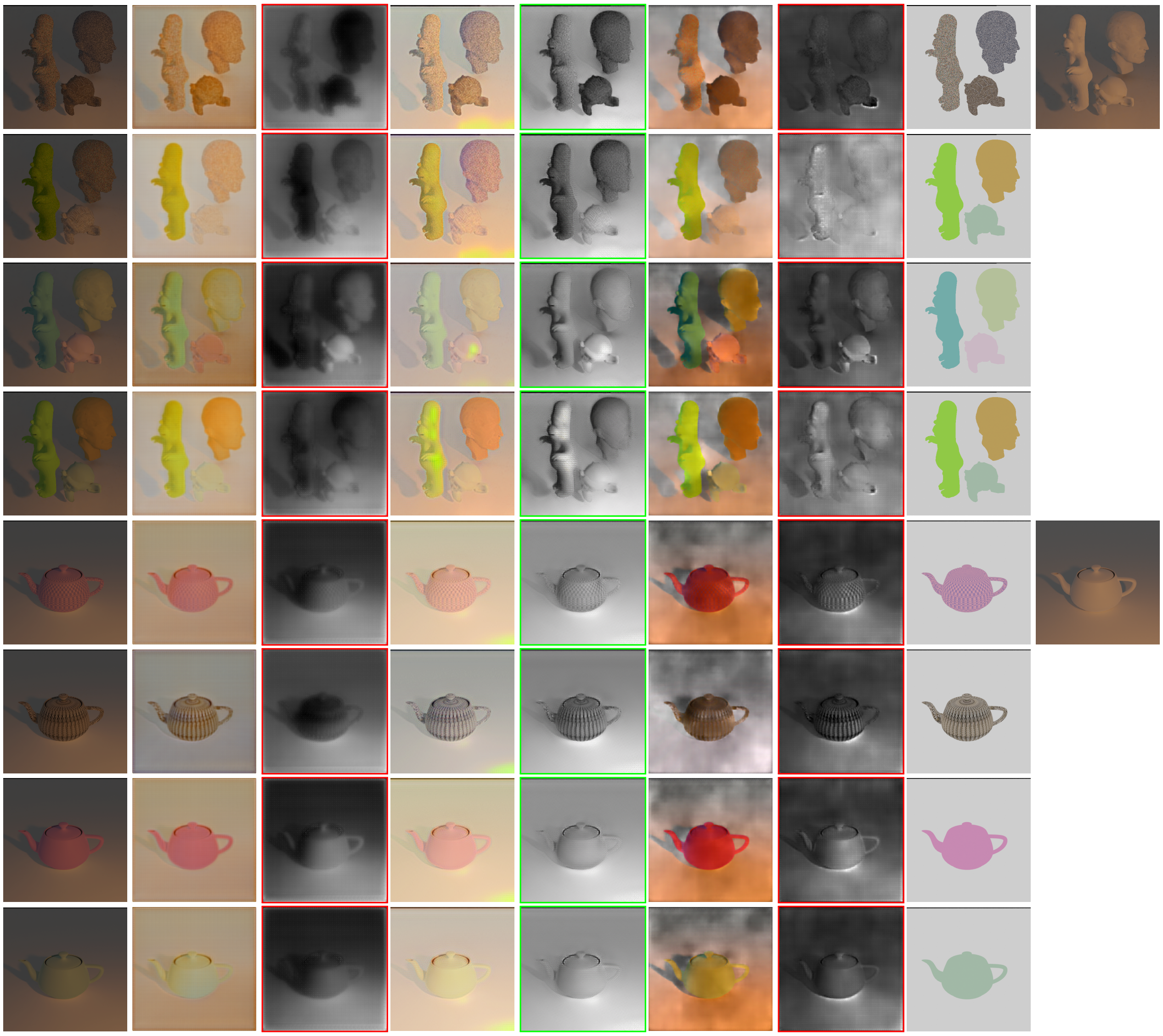
(a) Input (b) IIWW $\hat{R}$ (c) IIWW $\hat{S}$ (d) USI3D $\hat{R}$ (e) USI3D $\hat{R}$ (f) USI3D $\hat{S}$ (g) CGIID $\hat{S}$ (h) GT R (i) GT S

Figure 3: Albedo change experiments for T = 6500: Overall performance order is: USI3D=IIWW»CGIID. Rows 1 and 2 are scenes in Textured-Complex setting: USI3D has least S_{pred} variations followed by IIWW and CGIID which has significant global changes (shading intensity varions from light to dark). Also, CGIID's R_{pred} for second row is very smooth and most texture information is leaked in S_{pred}. Third and fourth rows have RGB-complex Δ_a: IIWW followed by USI3D have less changes in S_{pred} compared to CGIID which observes global changes. Fifth and sixth rows are Textured-Simple Δ_a: IIWW observes least changes in S_{pred} over teapot followed by USI3D. CGIID and IIWW have significant S_{pred} variations in background. For second last and last rows which is RGB-Simple Δ_a setting: USI3D has a nearly constant $\hat{S}$, while IIWW has siginificant variations in background(at top) where intensity changes becomes darker and CGIID has shading intensity variations over teapot).



(a) Input (b) IIWW $\hat{R}$ (c) IIWW $\hat{S}$ (d) USI3D $\hat{R}$ (e) USI3D $\hat{S}$ (f) CGIID $\hat{R}$ (g) CGIID $\hat{S}$ (h) GT R (i) GT R

Figure 4: Albedo change experiments for T = 4500 Order of performance: USI3D> IIWW> CGIID. Note: Since shading is constant we represent shading for rows 1, 2, 3, 4 in row 1 and rows 5, 6, 7, 8 in row 5.



(a) Input (b) IIWW $\hat{R}$ (c) IIWW $\hat{S}$ (d) USI3D $\hat{R}$ (e) USI3D $\hat{S}$ (f) CGIID $\hat{R}$ (g) CGIID $\hat{S}$ (h) GT R (i) GT R

Figure 5: Albedo change experiments for T = 2500 Order of performance: USI3D> IIWW> CGIID. Note: Since shading is constant we represent shading for rows 1, 2, 3, 4 in row 1 and rows 5, 6, 7, 8 in row 5. Note: USI3D has sharper textures in S, hence in texutured setting, textures might seem a bit change, but its light intensity is constant compared to other models which have both texture leakage and light intensity changes. IIWW has smoother S leading to lesser texture leakage but has more light intensity changes(as seen from rows 5 and 6) while CGIID has both sharp texture leakages and light intensity changes. Hence USI3D gets a good CSM_S overall.