VT-Intrinsic: Physics-Based Decomposition of Reflectance and Shading using a Single Visible-Thermal Image Pair

Supplementary Material

Table 2. Result of si-MSE (↓) on simulated MIT-Intrinsic Dataset.

Method	Average Albedo	Average Shading
Ours	1.9%	0.5%

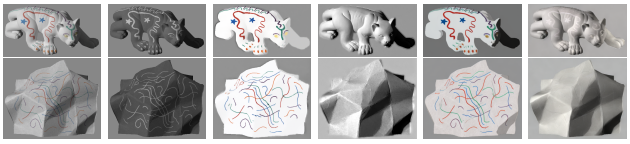

Figure 8. Qualitative comparison with the best baseline on MIT-Intrinsic dataset [19], IntrinsicDiffusion [30] . Our method excels without learning priors by leveraging the simulated thermal image.

1. Simulated MIT-Intrinsic Dataset: Investigating the Informativeness of Ordinality

The effectiveness of our method depends on two orthogonal factors: (1) the correctness of albedo–shading ordinalities derived from visible–thermal pairs, determined by imaging quality and the robustness of our theory to variations in physical properties, and (2) the informativeness of such ordinalities for intrinsic decomposition. With (1) validated in Sec. 6.2.1, we isolate (2) by generating ideal absorbed-light images as thermal input using pseudo ground truth from the MIT-Intrinsics dataset [19] via Eq. 2. Tab. 2 shows that a single ideal thermal image achieves a low average si-MSE on both albedo and shading.

2. Ablation on Loss Terms and DDIP

We ablated the loss functions and Double-DIP parameterization using pseudo ground truth from the *Painted Mask* scene. Tab. 3 shows that combining ordinal, edge, and reconstruction losses with Double-DIP yields the best result.

Table 3. Ablation study on loss terms and Double-DIP parameterization. We report the si-MSE for the *Painted Mask* scene.

$\mathcal{L}_{\text{recon}}$	$\mathcal{L}_{\text{edge}}$	$\mathcal{L}_{\text{ord}}$	DDIP	Albedo ↓	Shading ↓
✓	✓	✓	✓	$\mathbf{1.1 \times 10^{-1}}$	$\mathbf{9.7 \times 10^{-4}}$
✓	✓	✓	✗	1.6×10^{-1}	32×10^{-4}
✓	✓	✗	✓	2.2×10^{-1}	18×10^{-4}
✓	✗	✓	✓	2.0×10^{-1}	13×10^{-4}
✗	✓	✓	✓	4.0×10^{-1}	79×10^{-4}
✓	✗	✗	✓	3.3×10^{-1}	22×10^{-4}

3. Proof for Propositions

Proposition 1. *Given two pixels with visible and heat intensities as in Eq. 6, if $\mathcal{H}(x_i) > \mathcal{H}(x_j)$ and $I_v(x_i) < I_v(x_j)$, then $\rho(x_i) < \rho(x_j)$, and vice versa.*

Proof. Given

$$(1-\rho(x_i))\eta(x_i) > (1-\rho(x_j))\eta(x_j) \tag{21}$$
$$g\rho(x_i)\eta(x_i) < g\rho(x_j)\eta(x_j). \tag{22}$$

Dividing the first eq. by the second and noting that all terms are positive, we get

$$\frac{1-\rho(x_i)}{g\rho(x_i)} > \frac{1-\rho(x_j)}{g\rho(x_j)} \implies \rho(x_i) < \rho(x_j) \tag{23}$$

Proof for the complement is omitted for brevity. □

Proposition 2. *Given two pixels with visible and heat intensities as in Eq. 6, if $I_v(x_i) < I_v(x_j)$ and $\mathcal{H}(x_i) < \mathcal{H}(x_j)$, then $\eta(x_i) < \eta(x_j)$, and vice versa.*

Proof. Since multiplying an inequality by a positive scalar and adding two inequalities of same order preserves the order, we have

$$\frac{I_v(x_i)}{g} + \mathcal{H}(x_i) < \frac{I_v(x_j)}{g} + \mathcal{H}(x_j). \tag{24}$$

From Eq. 1 and Eq. 2, note that $\frac{I_v(x)}{g} + \mathcal{H}(x) = \eta(x)$. Substituting in Eq. 24, we can see that

$$\eta(x_i) < \eta(x_j) \tag{25}$$

Proof for the complement is omitted for brevity. □

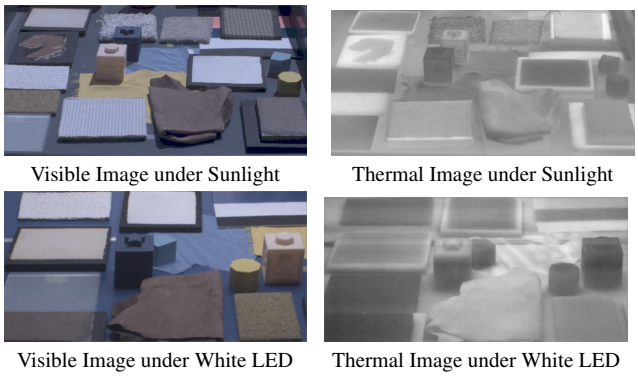

Figure 9. An example scene in expert validation on ordinalities across diverse materials (Sec. 3.5).

Table 4. The 20 materials used in ordinality validation.

1. Terrycloth	11. Orange peel
2. Plaster	12. Wooden block
3. Felt	13. Yellow silk
4. Cork	14. Blue silk
5. Frosted glass	15. Painted aluminum can
6. Sponge	16. Painted metal handcart
7. Carpet	17. Plastic board w/ black paint
8. White leather	18. Plastic board w/ white paint
9. Brick	19. Beige rubber block
10. Suede leather	20. Blue rubber block

4. Ordinality Validation on Diverse Materials

To further examine how material properties affect the validity of our theory and assumptions, we created scenes with 20 common material samples from CUReT dataset [10] and daily objects under sunlight and white-LED, and conducted expert validation on albedo-shading ordinalities. The detailed list of 20 materials used is shown in Tab. 4.

5. Additional Limitation Analysis

The key limitations of our method arise when the relationship between the absorbed heat from light (S) and the thermal image intensity (I_t) is violated, which can be summarized by three categories: external heat generation, non-opaque surfaces, and low signal-to-noise ratio (SNR). Fig. 7 shows representative failure cases.

5.1. Low SNR in Low-Light Conditions

To investigate how thermal image SNR is influenced by low-light condition, we captured visible-thermal image pairs of a color chart under an incandescent light at different distance. We measured illuminance using a light meter and computed si-MSE on the albedo decomposed by our method. As shown in Fig. 10, si-MSE decreases with increasing illuminance, indicating improved thermal SNR and decomposition quality under higher illumination. For reference, direct sunlight reaches about 100,000 lux, while overcast daylight is around 6,000 lux [15].

5.2. Non-Equilibrium Thermal Conditions

Fig. 7 shows a failure case of a truck where a heat source (engine) disturbs the thermal equilibrium. However, our approach can tolerate small deviations from thermal equilibrium since it relies on relative intensities (ordinalities) rather than absolute intensities. Theoretically, thermal transient exponentially converges to equilibrium [34], consequently, the observed thermal contrast follows this trend. Empirically, this tolerance can be evaluated on the JoLHT-Video dataset since all scenes were captured with a strong nearby source switched on at the beginning of videos, introducing a disturbance more severe than most natural settings. Please note that this disturbance is more severe than typical natural variations. We ran our method at 10, 20, and 30 seconds, achieving 72.3%, 79.0%, 89.4%, respectively, of the steady-state performance at 60s.

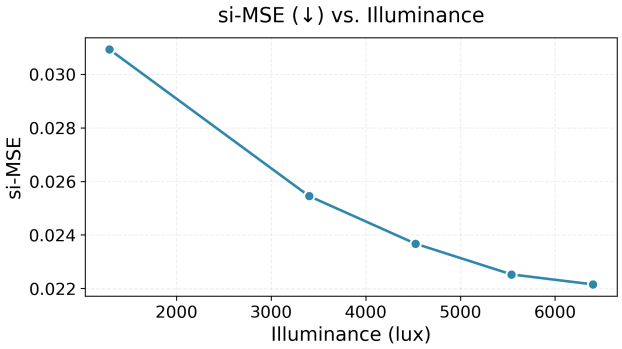

Figure 10. Effect of illuminance on albedo decomposition accuracy. The albedo si-MSE (↓) decreases as illuminance increases (incandescent bulb at varying distances), reflecting improved thermal SNR under stronger illumination. For reference, bright sunlight reaches 111,000 lux, while overcast daylight is typically 1,000–2,000 lux.

6. Additional Method Details

For point-pair sampling, the first point is sampled uniformly over the image. The second point is sampled at a random angle in $[0, 2\pi]$ and a random distance in $[0, 0.2]$ times the image diagonal from the first. Points outside the image are mirrored back.

7. Runtime Analysis

The optimization generally converges within 5000 iterations with Double-DIP parameterization and 500 without it. The average runtime for 5000 vs. 500 iterations is 48.22 s vs. 5.63 s, evaluated on the data from [34]. The measurements were obtained by running five parallel jobs on a single GeForce RTX 4090 GPU and normalizing the runtime per job. In challenging cases, increasing the number of iterations can further improve the results.

8. Additional Qualitative Results

We provide additional qualitative results on VT-Intrinsic dataset in comparison with state-of-the-art baselines. Each case shows visible input with albedo estimations above and thermal with shading below. Images are tonemapped / colormapped for visualization.

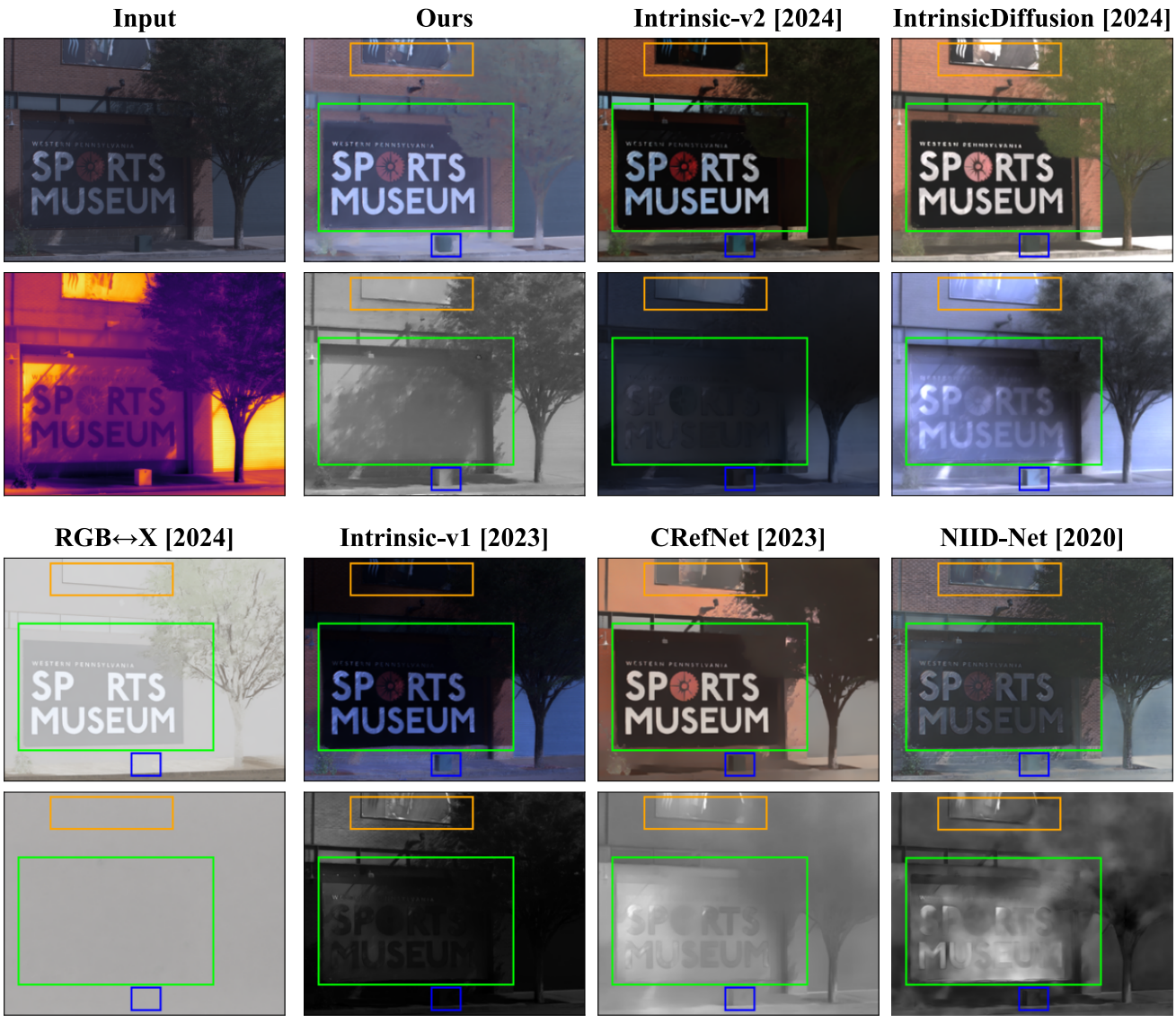

Figure 11. Qualitative comparisons to state-of-the-art baselines. Visible input with albedo estimations are shown above and thermal with shading below.

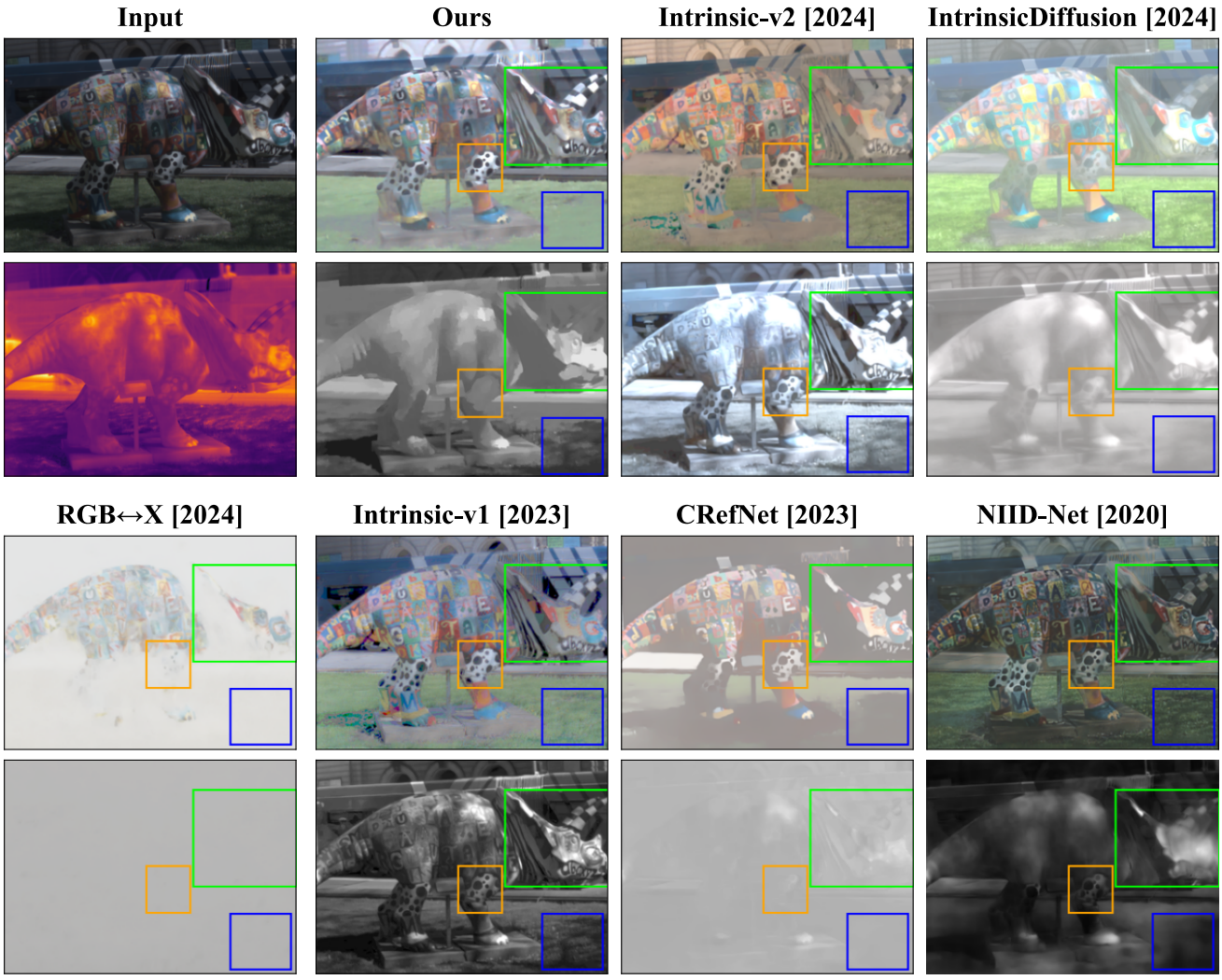

Figure 12. Qualitative comparisons to state-of-the-art baselines.

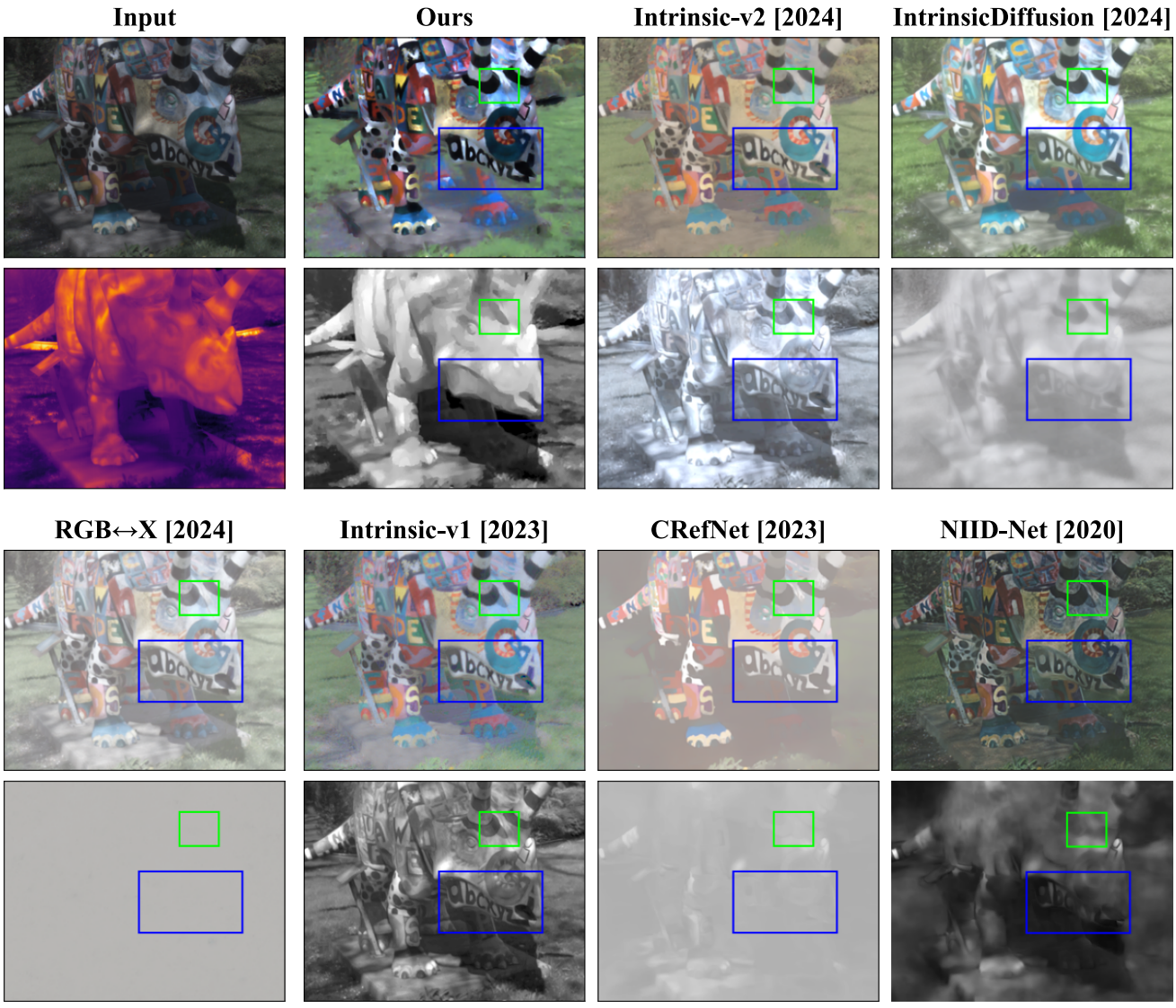

Figure 13. Qualitative comparisons to state-of-the-art baselines.

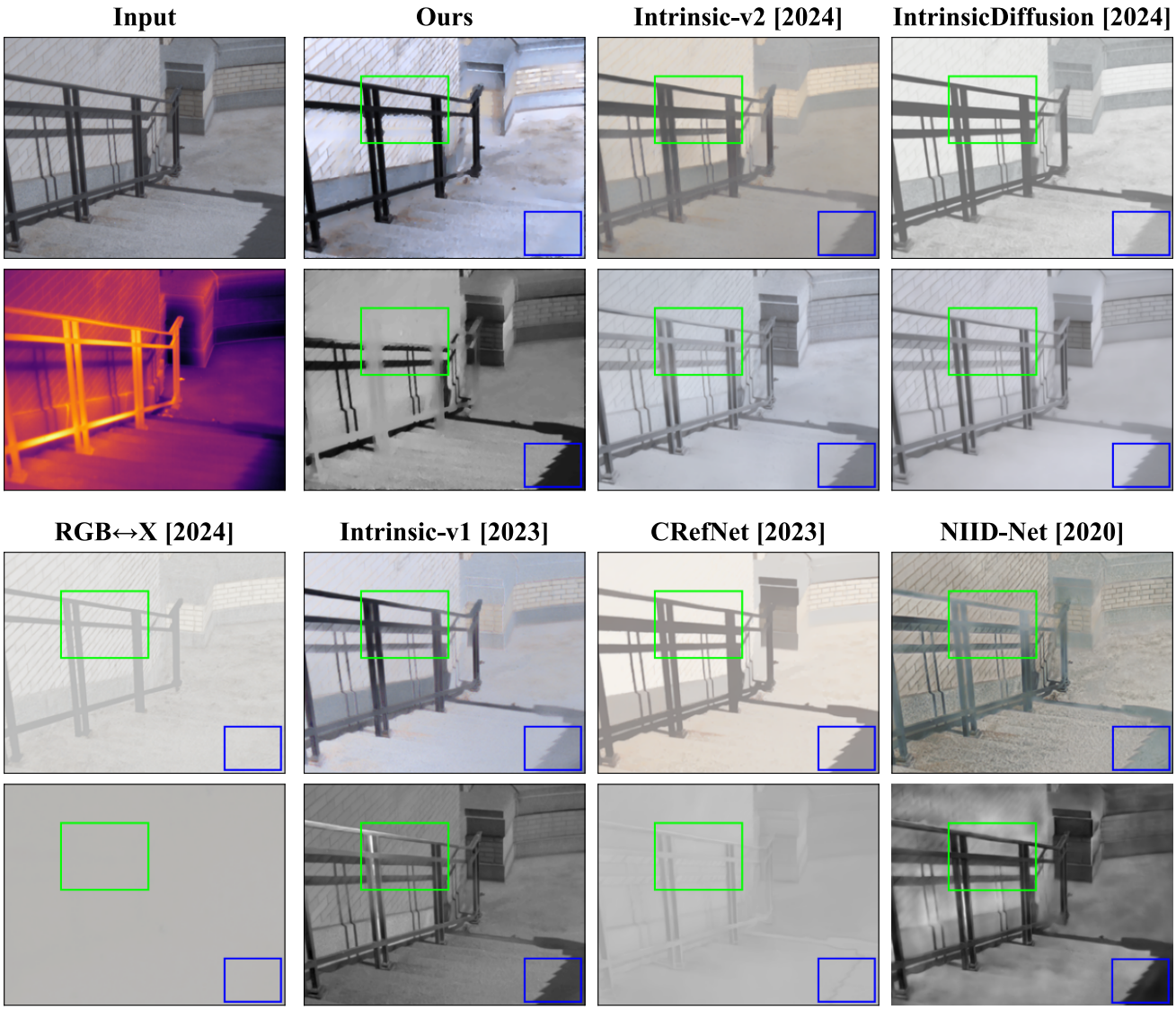

Figure 14. Qualitative comparisons to state-of-the-art baselines.

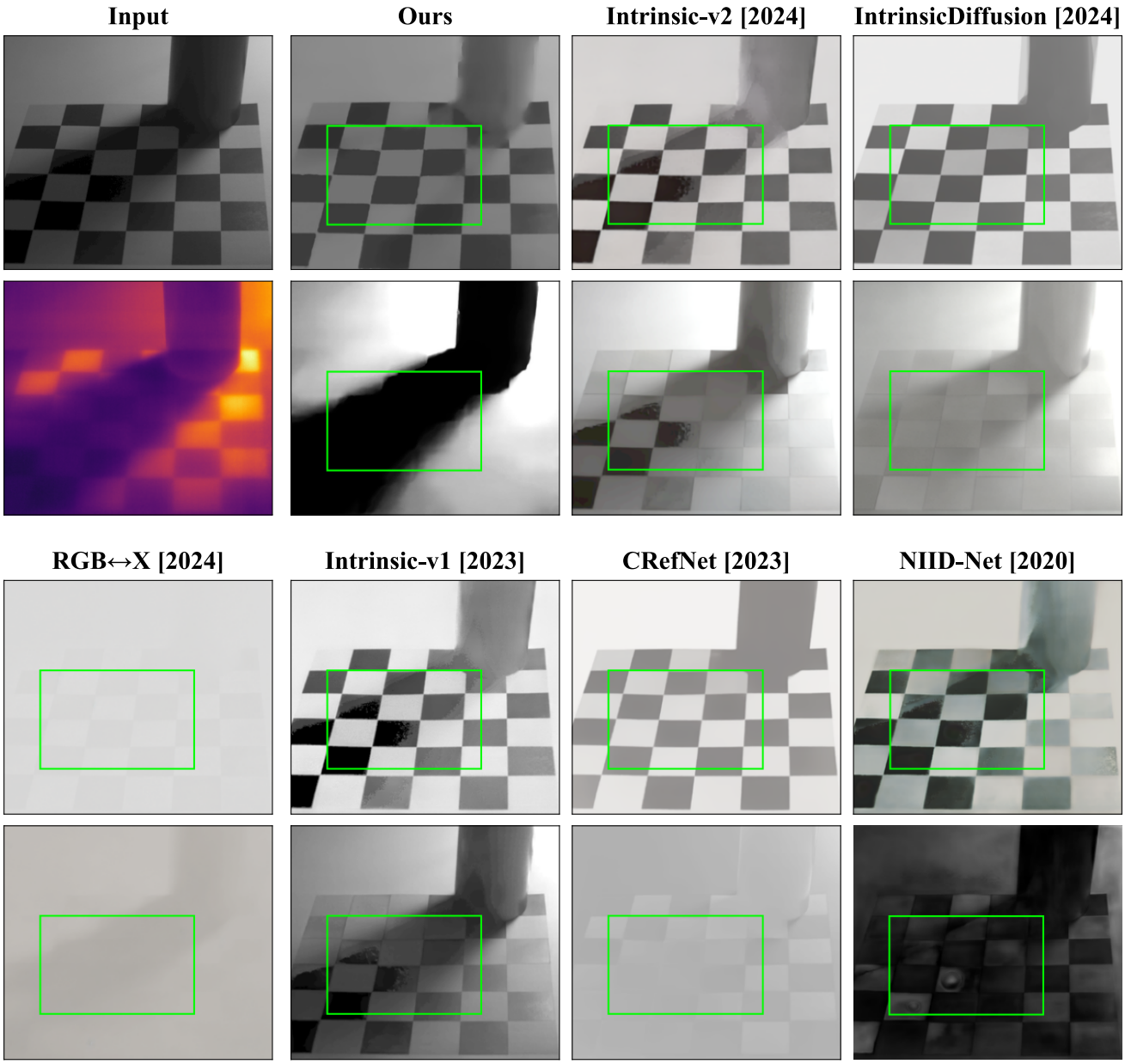

Figure 15. Qualitative comparisons to state-of-the-art baselines.

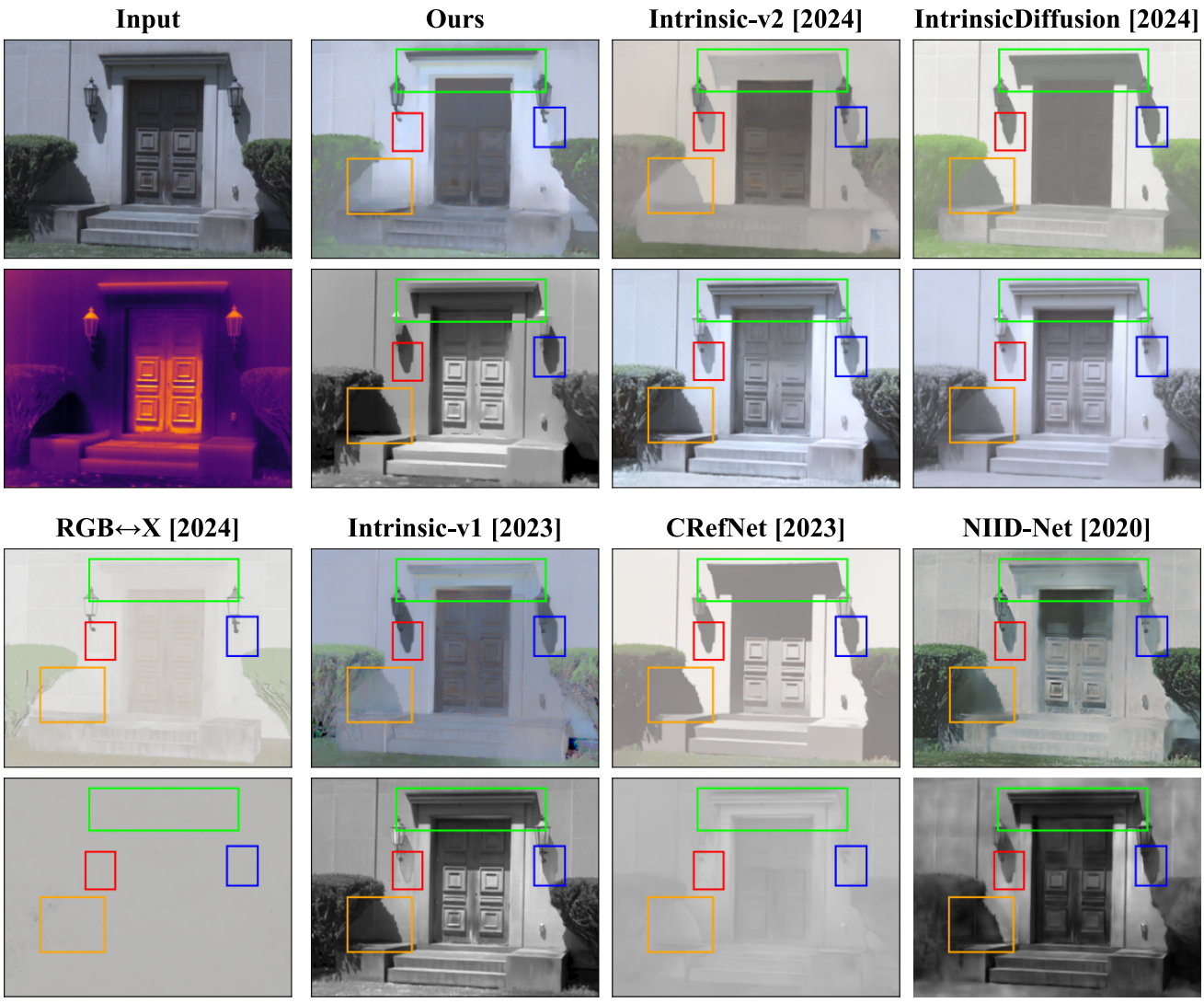

Figure 16. Qualitative comparisons to state-of-the-art baselines.

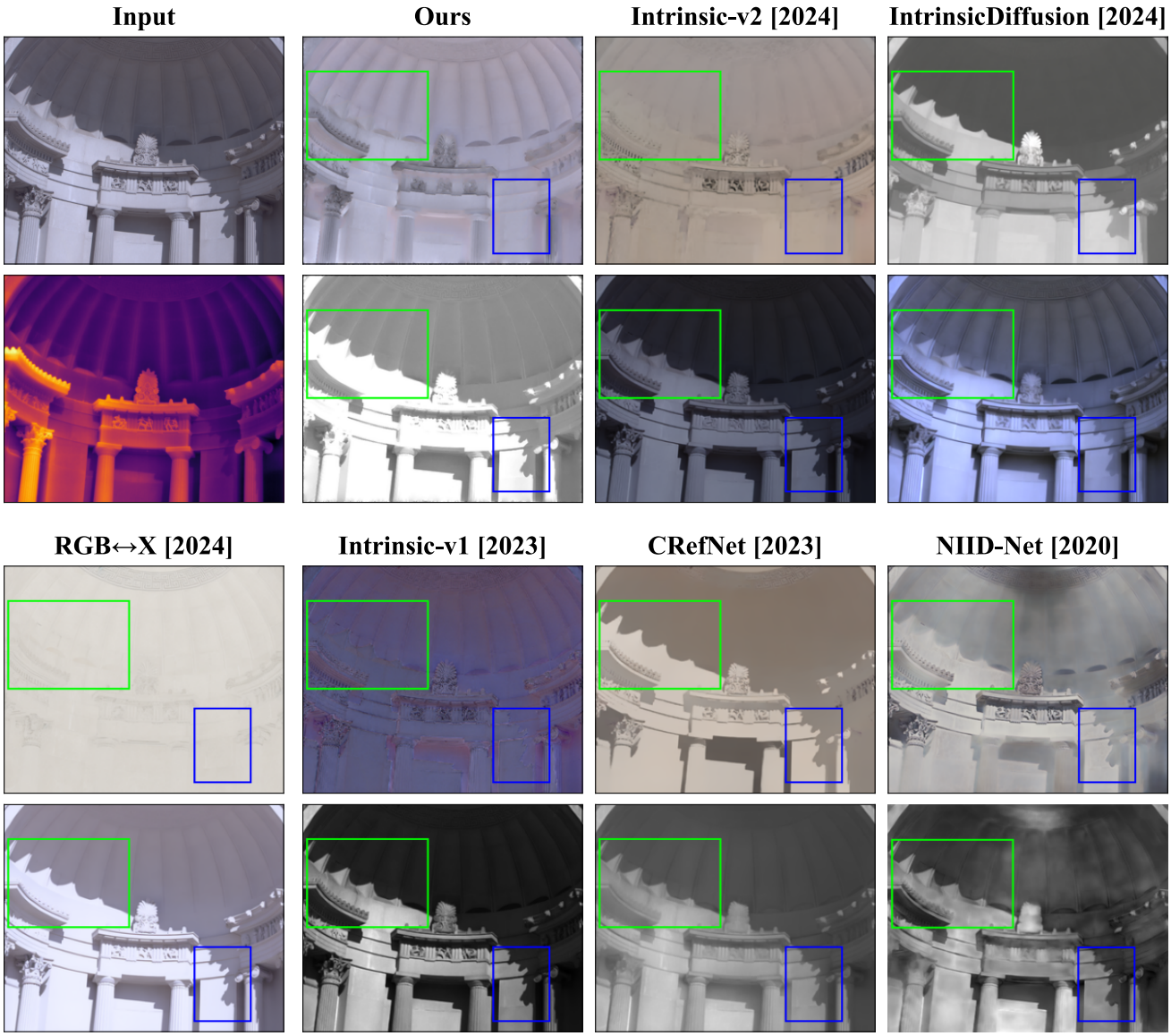

Figure 17. Qualitative comparisons to state-of-the-art baselines.

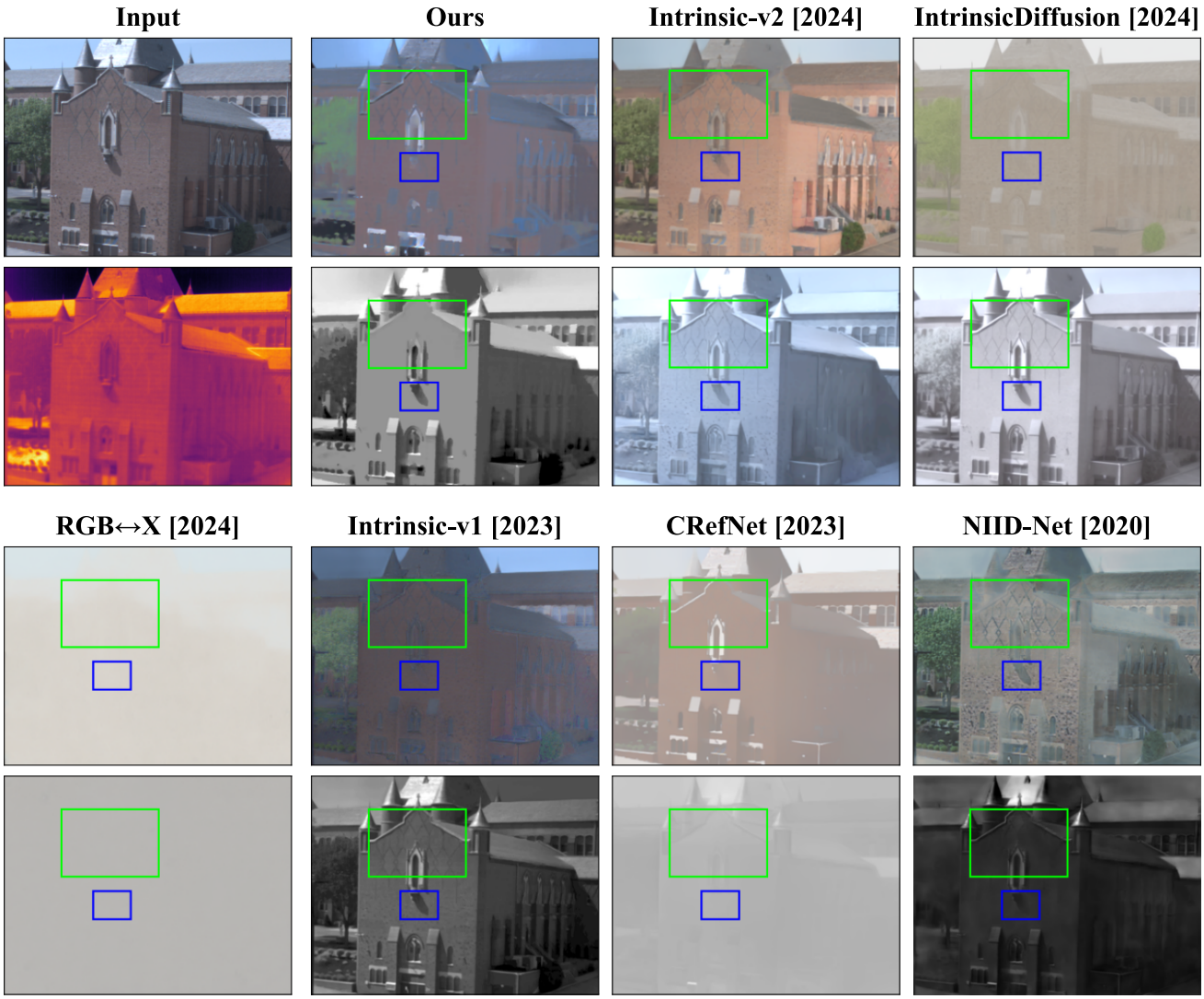

Figure 18. Qualitative comparisons to state-of-the-art baselines.

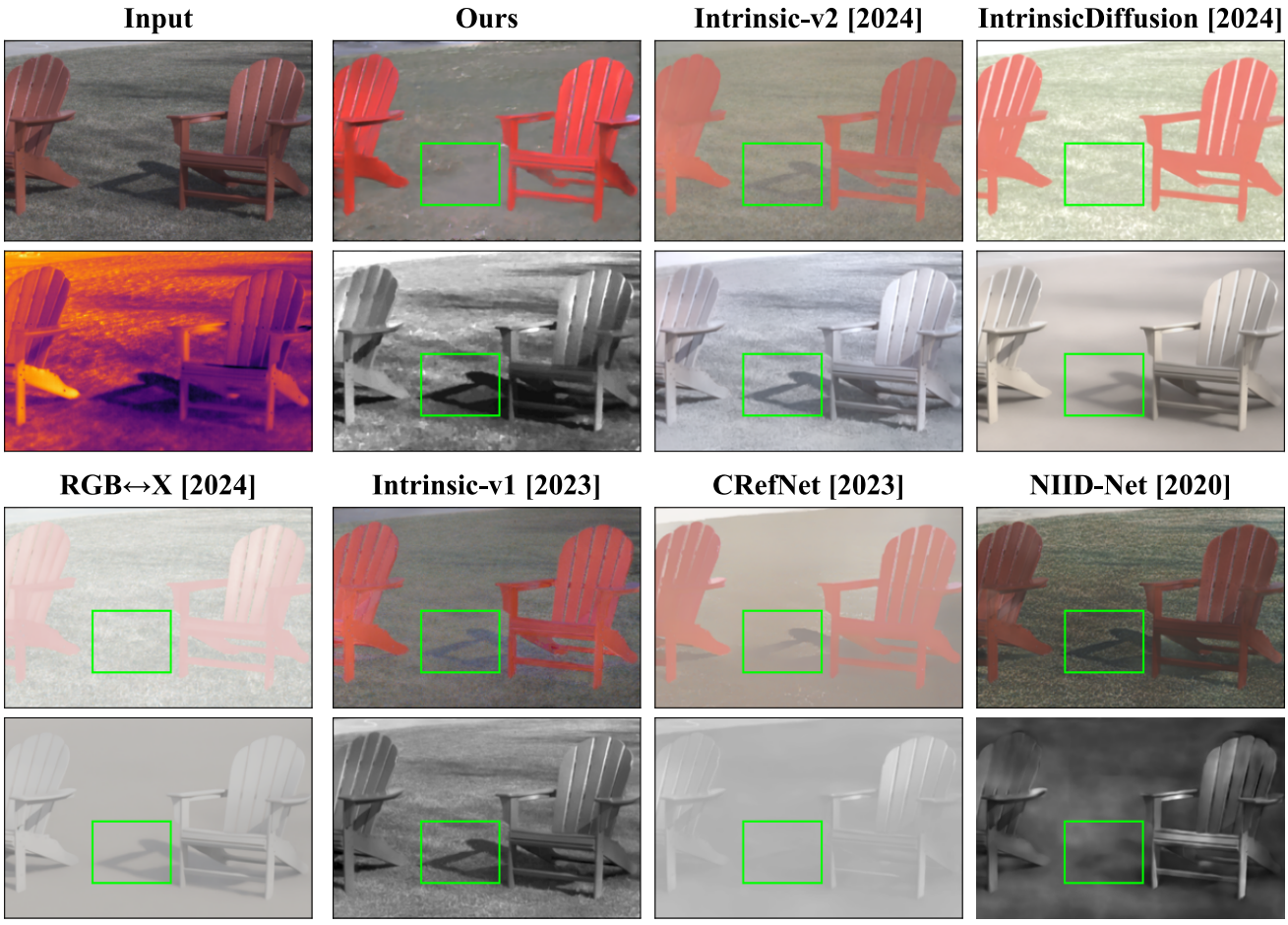

Figure 19. Qualitative comparisons to state-of-the-art baselines.

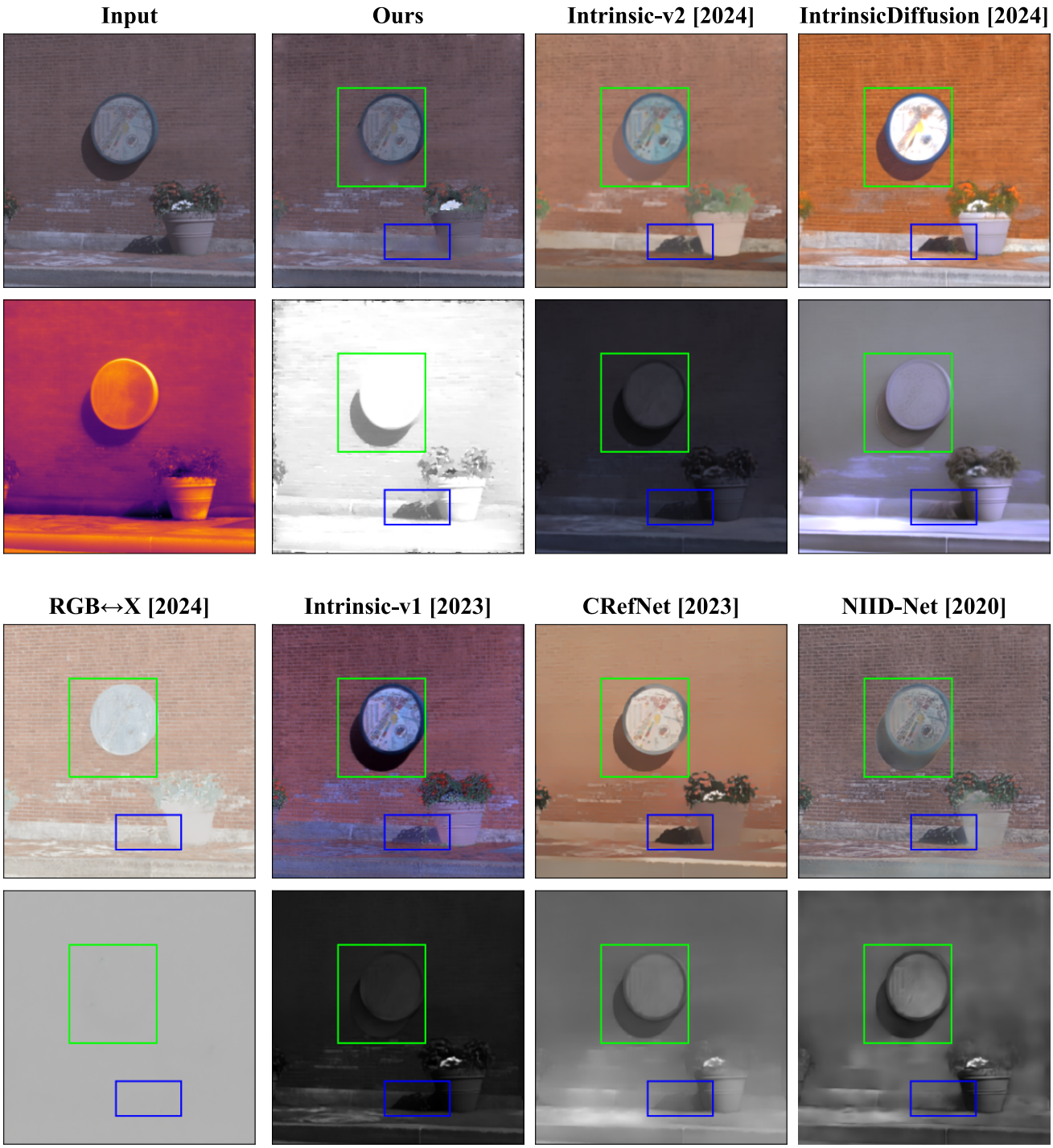

Figure 20. Qualitative comparisons to state-of-the-art baselines.

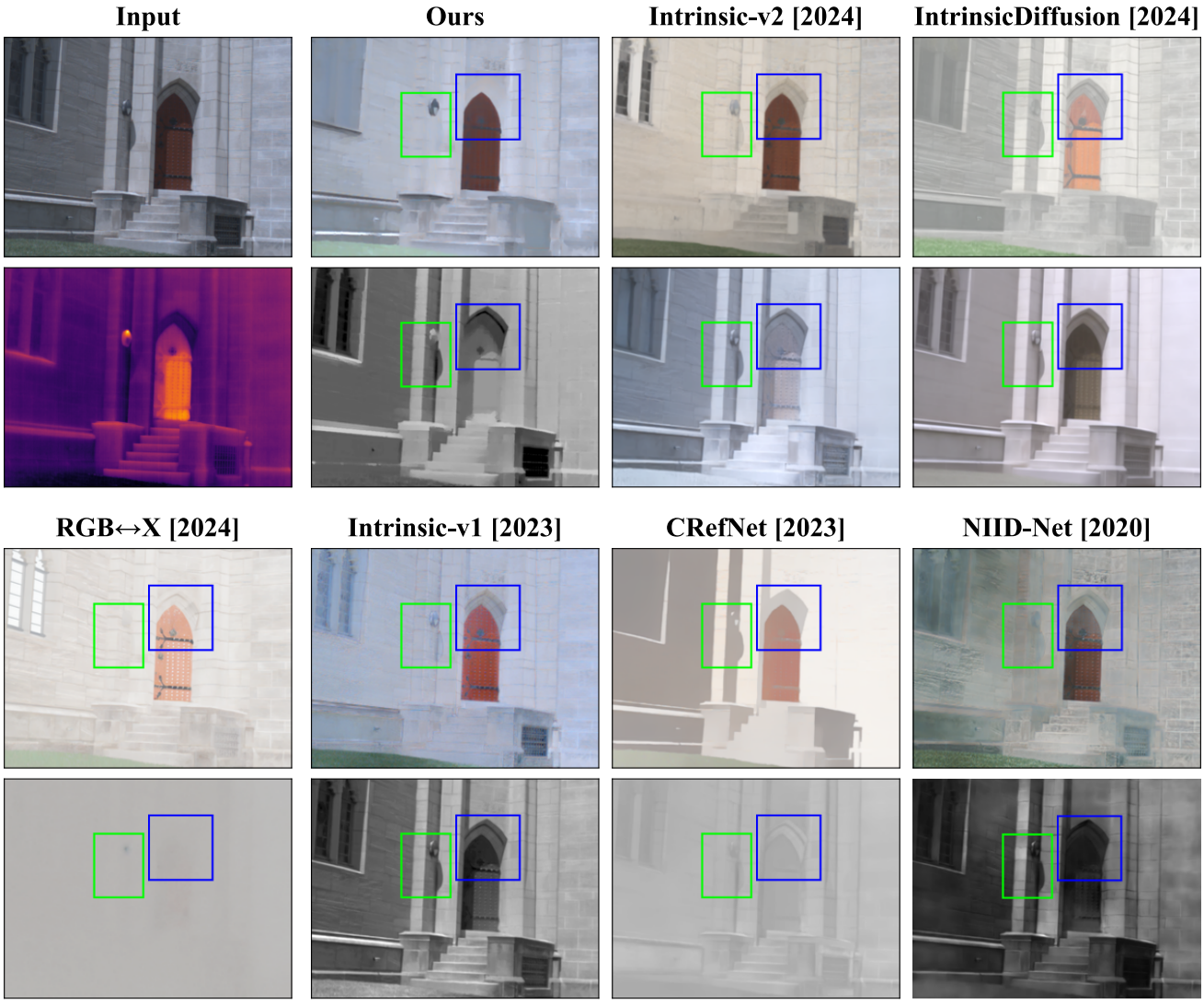

Figure 21. Qualitative comparisons to state-of-the-art baselines.

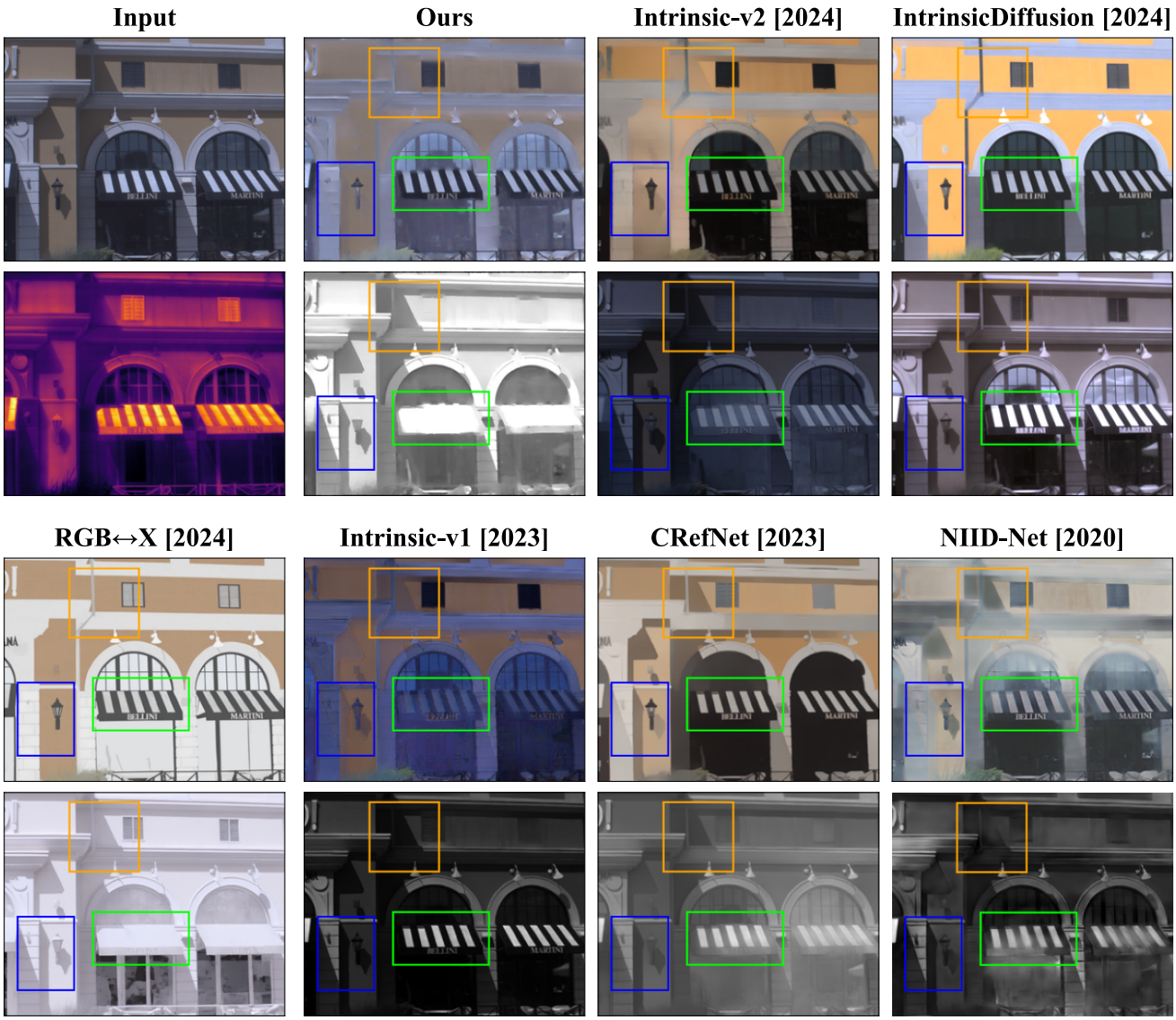

Figure 22. Qualitative comparisons to state-of-the-art baselines.

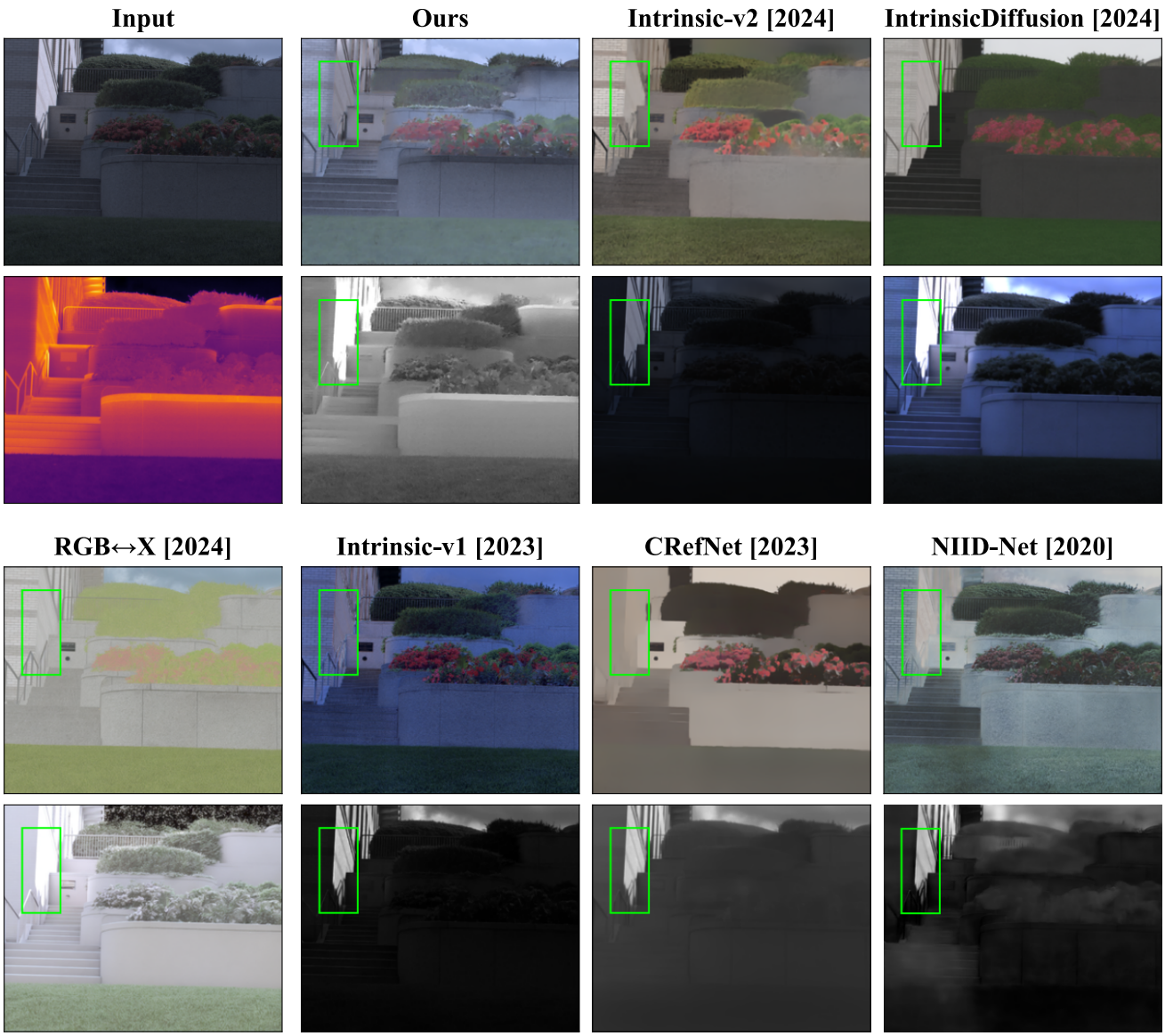

Figure 23. Qualitative comparisons to state-of-the-art baselines.

Figure 24. Qualitative comparisons to state-of-the-art baselines on a printed image.

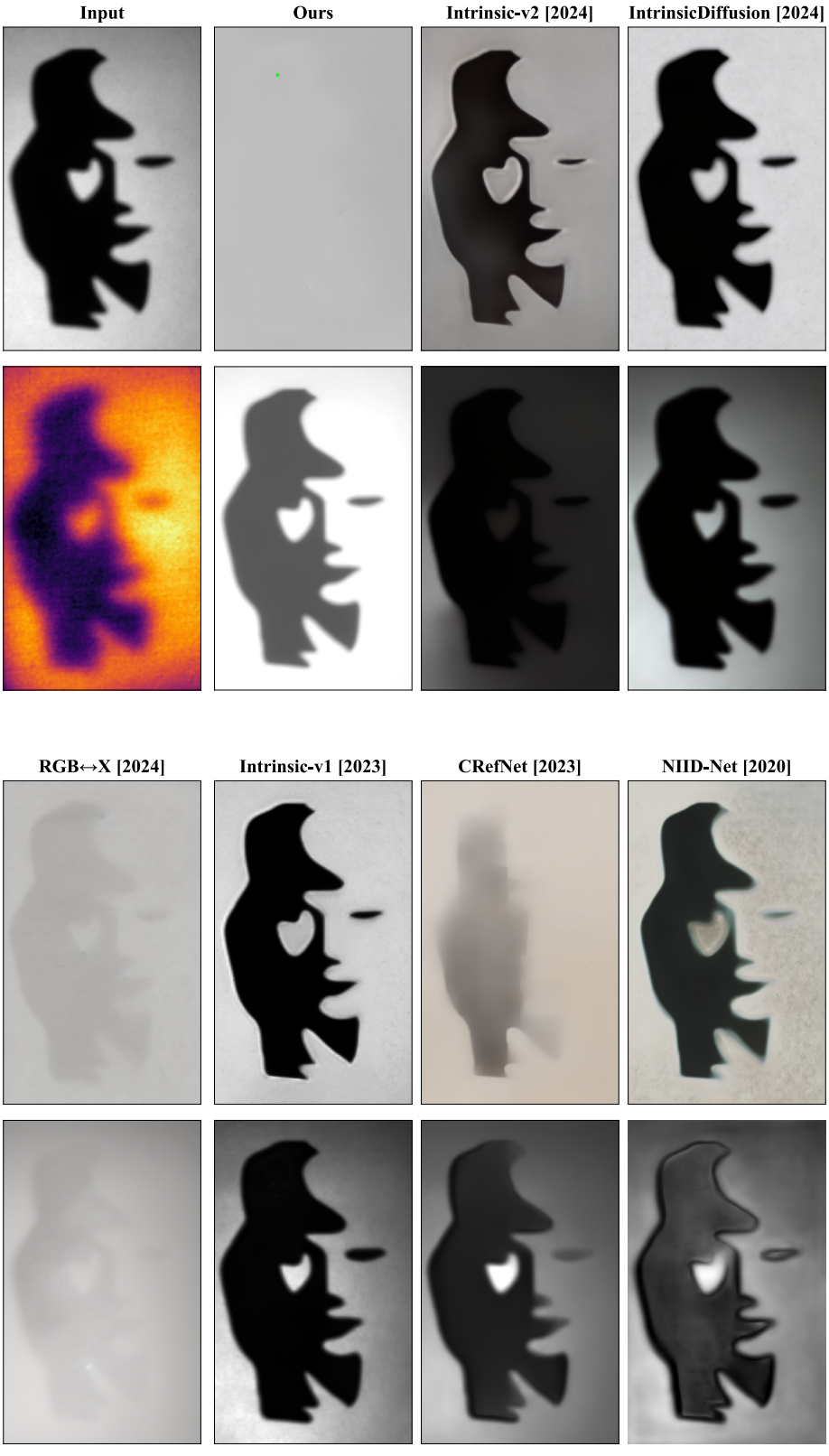

Figure 25. Qualitative comparisons to state-of-the-art baselines on a projected image.