Supplementary Material

Anonymous CVPR submission

Paper ID 6829

1. Gradient of the pose estimation loss to pose parameters in 3DGS

The gradient of the loss function $\mathbf{L}$ to $\mathbf{x}$ is used to refine the initial pose T^0 as introduced in Subsection 3.3 of the main paper. It is defined as follows:

$$\frac{\partial \mathcal{L}}{\partial \mathbf{x}} = \left(\frac{\partial \mathcal{L}}{\partial I} \frac{\partial I}{\partial V} + \frac{\partial \mathcal{L}}{\partial R} \frac{\partial R}{\partial V} \right) \frac{\partial V}{\partial \mathbf{x}} \tag{1}$$

The derivatives $\frac{\partial \mathcal{L}}{\partial I}$, $\frac{\partial \mathcal{L}}{\partial R}$, and $\frac{\partial V}{\partial \mathbf{x}}$ are relatively straightforward to compute and can be efficiently implemented using PyTorch's autograd framework. Consequently, the primary challenge lies in the computation of $\frac{\partial I}{\partial V}$ and $\frac{\partial R}{\partial V}$. Given that the rendering processes of I and R are analogous, we focus on I as a representative example to elucidate the procedure for clarity. The gradient calculation formula is as follows:

$$\frac{\partial I}{\partial V} = \frac{\partial I}{\partial \mu'} \frac{\partial \mu'}{\partial V} + \frac{\partial I}{\partial \Sigma'} \frac{\partial \Sigma'}{\partial V} = \left(\frac{\partial I}{\partial V} \right)_1 + \left(\frac{\partial I}{\partial V} \right)_2, \tag{2}$$

where the first and second parts represent the gradient backtracked through the 2D mean μ' and covariance matrix Σ' of the 3D Gaussian. Next, we compute these two parts separately.

1.1. Mean Component

3D Gaussians are projected to 2D Gaussians for rendering a 2D image with the following 2D mean μ':

$$\mu' = \begin{bmatrix} \mu_1 \\ \mu_2 \end{bmatrix} = \begin{bmatrix} \frac{x_{hom}}{w_{hom}} \\ \frac{y_{hom}}{w_{hom}} \end{bmatrix}, \tag{3}$$

where $[x_{hom}, y_{hom}, z_{hom}, w_{hom}]^T = P\mu_c$, P is the perspective matrix and $\mu_c = V\mu$.

Let $F_1 = PV$ represent the composition of the perspective and view transformations, we calculate:

$$\frac{\partial I}{\partial F_1} = \begin{bmatrix} \frac{1}{w_{hom}} \frac{\partial I}{\partial \mu_1} \\ \frac{1}{w_{hom}} \frac{\partial I}{\partial \mu_2} \\ 0 \\ -\lambda \end{bmatrix} \cdot \begin{bmatrix} x & y & z & 1 \end{bmatrix}. \tag{4}$$

$$\lambda = \frac{\mu_1}{w_{hom}} \frac{\partial I}{\partial \mu_1} + \frac{\mu_2}{w_{hom}} \frac{\partial I}{\partial \mu_2}, \tag{5}$$

where $\frac{\partial I}{\partial \mu_1}$ and $\frac{\partial I}{\partial \mu_2}$ can be computed with the CUDA kernels provided by the original 3DGS [1]. So, we compute the first term of $\frac{\partial I}{\partial V}$ as:

$$\left(\frac{\partial I}{\partial V} \right)_1 = \sum_{i=1}^{4} \alpha_i^T \cdot \beta_i, \tag{6}$$

where

$$\begin{cases} \alpha_i = [P_{i1}, P_{i2}, P_{i3}, P_{i4}], \\ \beta_i = \left[\left(\frac{\partial I}{\partial F_1} \right)_{i1}, \left(\frac{\partial I}{\partial F_1} \right)_{i2}, \left(\frac{\partial I}{\partial F_1} \right)_{i3}, \left(\frac{\partial I}{\partial F_1} \right)_{i4} \right]. \end{cases}$$

1.2. Covariance Component

The projected 2D covariance matrix Σ' can be represented as:

$$\Sigma' = JW\Sigma W^T J^T, \tag{7}$$

where W is the 3x3 part on the top left of the view matrix V. J is the Jacobian matrix of perspective transformation at μ_c:

$$J = \begin{bmatrix} 1/\mu_{c,2} & 0 & -\mu_{c,0}/\mu_{c,2}^2 \\ 0 & 1/\mu_{c,2} & -\mu_{c,1}/\mu_{c,2}^2 \\ \mu_{c,0}/l & \mu_{c,1}/l & \mu_{c,2}/l \end{bmatrix}, \tag{8}$$

where $l = \left\| (\mu_{c,0}, \mu_{c,1}, \mu_{c,2})^T \right\|$ [4].

According to Eq. 8, we obtain J in the camera space μ_c. Let $F_2 = JW$, we compute the second term of $\frac{\partial I}{\partial V}$ as:

$$\left(\frac{\partial I}{\partial V} \right)_2 = \begin{bmatrix} \frac{\partial I}{\partial \mu_c} \mu^T + \frac{\partial I}{\partial F_2} J & \frac{\partial I}{\partial \mu_c} \\ O & 0 \end{bmatrix}, \tag{9}$$

where $\frac{\partial I}{\partial \mu_c}$ and $\frac{\partial I}{\partial F_2}$ can be compute with the CUDA kernels provided by the 3DGS. Finally, we can optimize the camera pose of 3DGS. To improve efficiency, our adapted backpropagation module computes gradients for pose estimation, skipping those required for training standard 3D Gaussians.

2. More Results

2.1. Quantitative Results

Quantitative comparisons on synthetic datasets – Tables 1 and 2. The quantitative comparison of the baseline methods and our proposed approach on each object of MAD-Sim [3] and our synthetic dataset are presented in Tables 1 and 2, respectively. Compared to Table 3 in the main paper, these comparisons offer more detailed insights. The experiments demonstrate that our method achieves significantly better performance than the baselines regarding pixel and image-level AUROC.

Objects	Pixels AUROC↑			Images AUROC↑		
	OmniAD	SplatPose	Our	OmniAD	SplatPose	Our
Gorilla	99.5	99.5	99.8	93.6	91.1	97.4
Unicorn	98.2	99.7	99.7	94.0	98.8	99.4
Mallard	97.4	99.7	99.8	94.7	97.7	99.3
Turtle	99.1	99.5	99.4	95.6	97.1	96.8
Whale	98.3	99.5	99.6	92.5	97.9	99.9
Bird	95.7	99.5	99.4	92.4	92.9	98.0
Owl	99.4	99.2	99.6	88.2	88.0	93.9
Sabertooth	98.5	99.4	99.3	95.7	96.6	98.6
Swan	98.8	99.3	99.4	86.5	93.7	97.7
Sheep	97.7	99.6	99.4	90.1	96.5	98.5
Pig	97.7	99.8	99.8	88.3	96.7	99.0
Zalika	99.1	89.5	99.5	88.2	99.3	94.2
Pheonix	99.4	99.5	99.7	82.3	84.6	94.0
Elephant	99.0	99.7	99.6	92.5	95.3	99.3
Parrot	99.5	99.5	99.5	97.0	93.6	99.8
Cat	97.7	99.6	99.5	84.9	86.1	93.1
Scorpion	95.9	99.4	99.2	91.5	99.3	99.7
Obesobeso	98.0	99.5	98.9	97.1	96.1	91.9
Bear	99.3	99.6	99.5	98.8	98.9	99.8
Puppy	98.8	99.1	99.4	93.5	97.1	97.8
MEAN	98.35	99.01	99.50	91.87	94.87	97.41

Table 1. **Anomaly detection; MAD-Sim dataset –** Comparisons of pixel and image-level AUROC. The best results are color-coded.

Objects	Pixels AUROC↑			Images AUROC↑		
	OmniAD	SplatPose	Our	OmniAD	SplatPose	Our
Axletree	98.1	98.1	99.5	77.0	77.3	95.3
Box	98.1	95.8	99.3	78.6	86.8	95.9
Can	99.0	97.8	99.4	99.1	95.5	99.9
Chain	98.9	97.5	99.1	95.6	98.4	100.0
Gear	95.8	95.6	97.1	98.1	88.6	99.7
Keyring	99.3	98.8	99.5	98.4	100.0	100.0
Motor	99.4	95.7	99.0	81.5	77.6	98.4
Parts	-	95.2	99.5	-	54.1	99.4
Picker	98.0	98.7	99.4	96.2	93.3	99.4
Section	-	96.2	99.2	-	82.6	99.5
Shaft	99.2	98.7	99.6	99.1	92.4	100.0
Spray_can	98.8	98.9	99.3	63.1	92.7	100.0
Spring	99.6	99.3	99.5	86.7	82.4	92.4
Sprockets	98.9	98.7	99.6	97.8	96.4	99.2
Amphora	85.2	96.9	97.5	57.6	76.3	79.5
Teapot	88.7	96.4	97.6	59.4	79.5	87.8
MEAN	96.93	97.39	99.01	84.87	85.87	96.65

Table 2. **Anomaly detection; our dataset (synt) –** Comparisons of pixel and image-level AUROC. The best results are color-coded.

Objects	Pixels AUROC↑			Images AUROC↑		
	OmniAD	SplatPose	Our	OmniAD	SplatPose	Our
Valve	97.3	92.9	99.3	91.7	74.1	98.8
Tube	97.2	99.5	99.6	95.7	81.5	94.7
Cup	92.5	98.8	99.5	63.6	83.1	92.5
USB	96.1	99.1	99.4	51.8	41.9	55.8
Joint	94.0	99.6	99.7	57.6	100.0	100.0
PaperCup	91.5	98.7	99.1	62.1	71.4	91.1
Lighter	98.5	99.5	99.8	88.0	90.9	99.9
Cube	97.3	99.0	99.3	89.7	93.5	87.7
Lamp	85.5	94.6	95.8	95.6	73.8	95.4
Bolt	95.6	98.0	98.9	90.3	83.5	99.1
Filter	96.6	99.7	99.9	78.7	81.9	97.0
Wand	92.7	98.1	99.6	39.1	76.0	94.4
Wheel	95.6	96.5	97.1	48.1	77.3	94.8
Bearing	97.6	98.1	99.7	90.8	88.5	95.6
MEAN	94.86	98.01	99.05	73.20	79.82	92.63

Table 3. **Anomaly detection; our dataset (real) –** Comparisons of pixel and image-level AUROC. The best results are color-coded.

Quantitative results on real dataset – Table 3. The detailed performance of the baselines and our method on our real dataset are reported in Table 4 of the main paper and are reproduced in Table 3 here for convenience. The comparisons show that our method significantly outperforms the other two methods in pixel-level and image-level AUROC.

2.2. Qualitative Results - Figure 1

We provide qualitative comparisons for all objects in our real dataset in Figures 1 to 3, as a supplement to Figure 6 in the main paper. For each object, one randomly selected defect type is showcased. For the three objects featured in Figure 6 of the main paper (Filter, Wheel, and Valve), we vary the camera pose, defect size, and defect type to provide a broader comparison. Our method accurately detects anomalies even when the lighting conditions of the query images differ from those of the training images.

3. Ablation Studies

As claimed in lines 401-403 of the main paper, we provide the ablation study results on our complete synthetic and real datasets in Tables 4 to 6. *CL* denotes the data with consistent lighting, and *IL* refers to the data with inconsistent lighting (marked with a gray background).

Pose initialization and optimization – Table 4. Our method utilizes reflectance images for pose initialization and combines them with color images for pose estimation. To validate the effectiveness of this strategy, we evaluate various configurations. We denote the use of color (I) and reflectance (R) images in a module. As shown in Table 4, the selected configuration (R+IR) performs best, especially for data with inconsistent lighting. While there is a slight drop in pixel-level AUROC for lighting-consistent data compared to using only color images (I+I), this is ex-

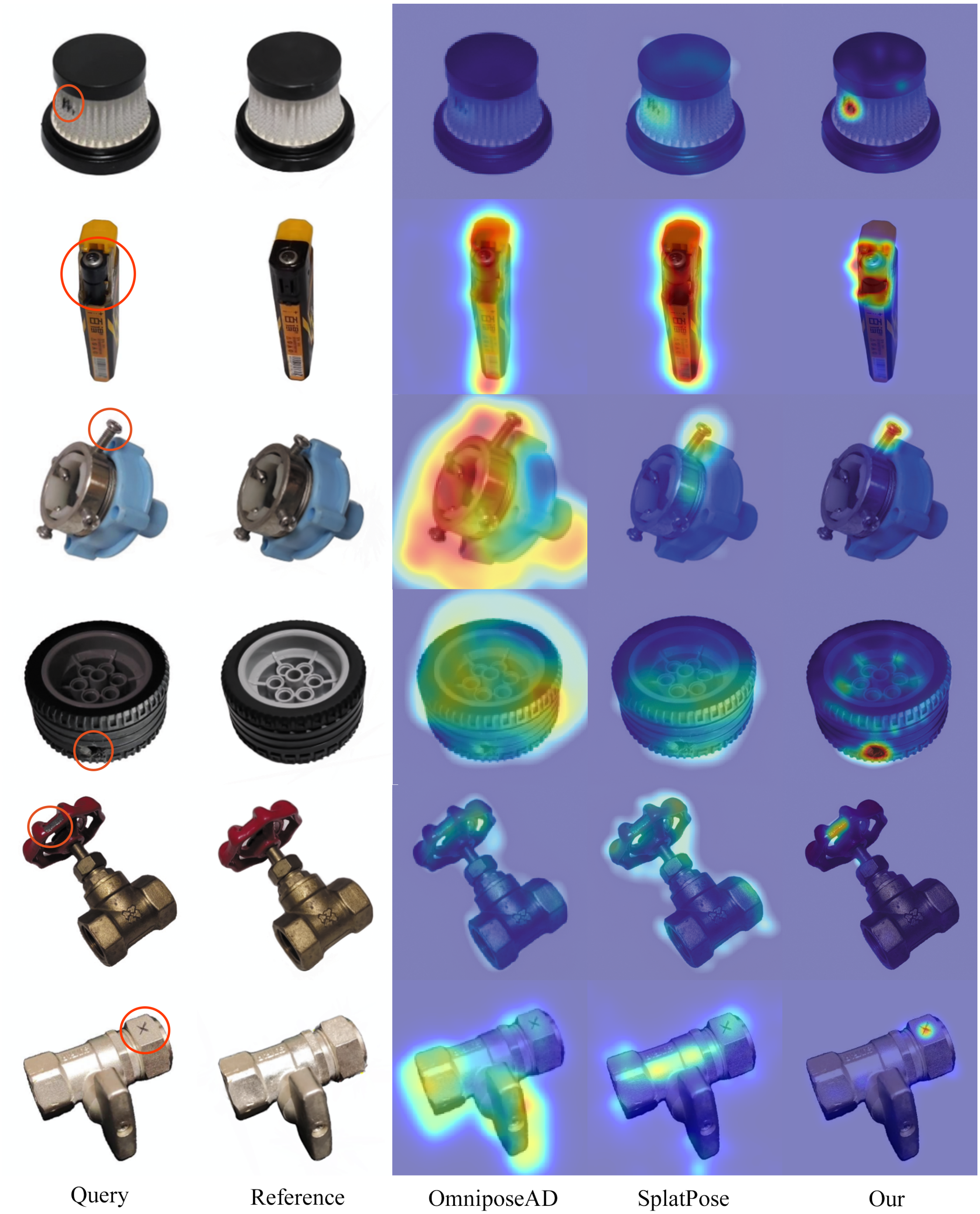

Figure 1. **Qualitative comparisons on anomaly detection (Part 1 of 3).** In the left two columns, we visualize the query images and reference images. The right three columns compare the heatmaps generated by our method and the baselines.

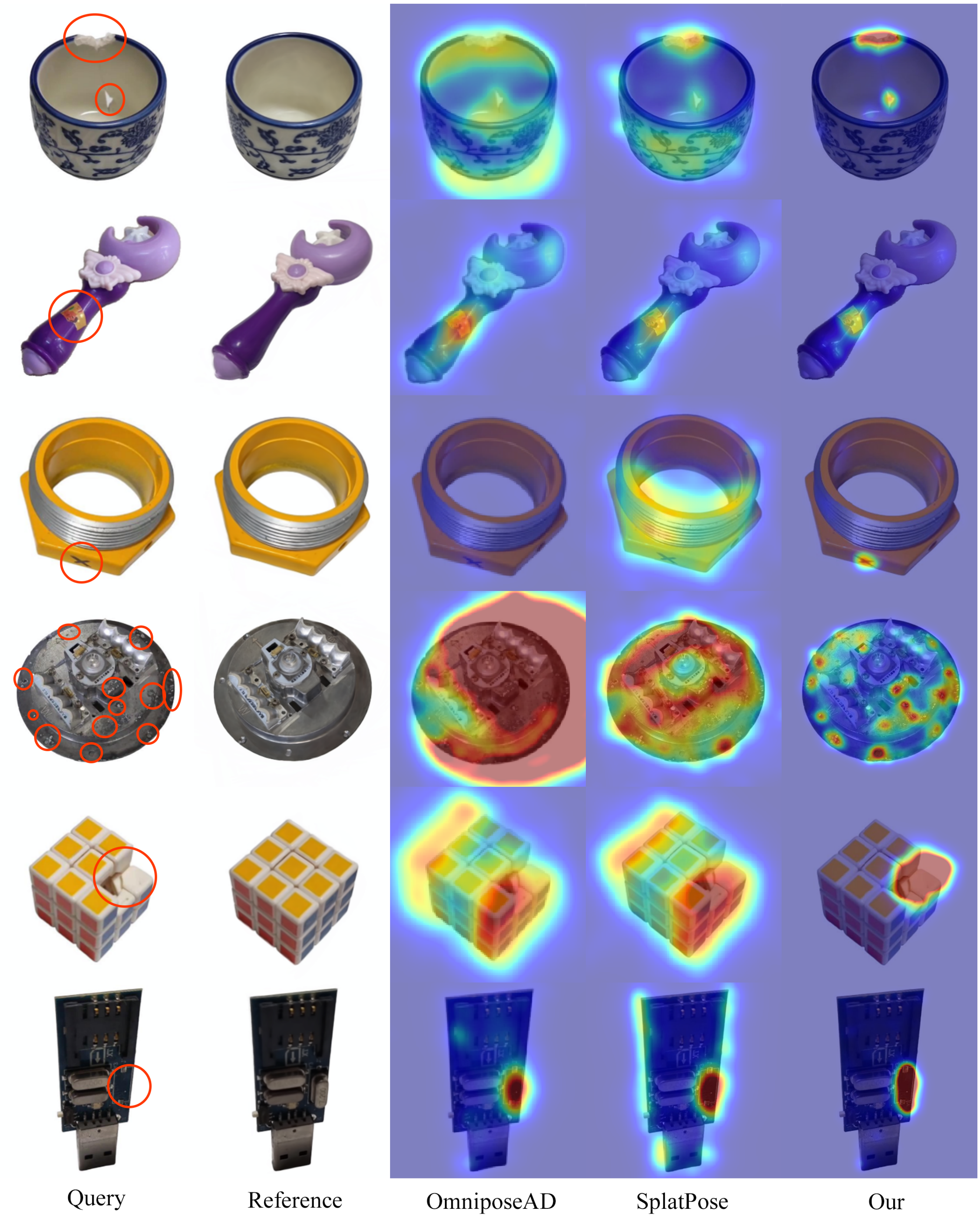

Figure 2. **Qualitative comparisons on anomaly detection (Part 2 of 3).** In the left two columns, we visualize the query images and reference images. The right three columns compare the heatmaps generated by our method and the baselines.

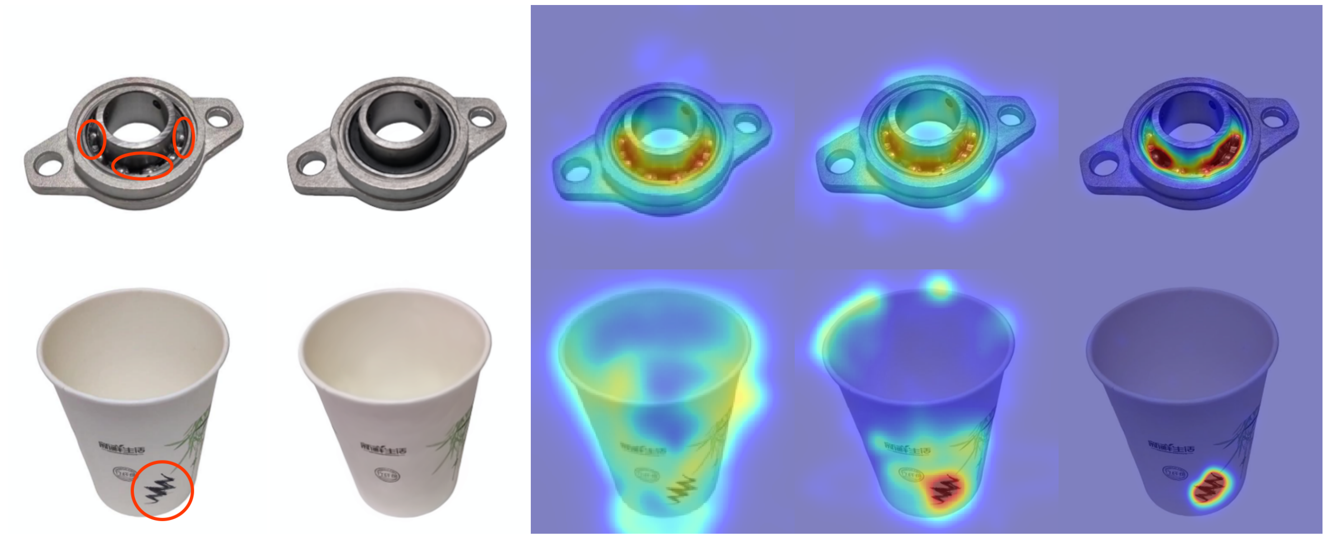

Figure 3. **Qualitative comparisons on anomaly detection (Part 3 of 3).** In the left two columns, we visualize the query images and reference images. The right three columns compare the heatmaps generated by our method and the baselines.

pected. The quality of the rendered reflectance image is limited by the pre-trained RetinexNet [2], which has not been fine-tuned on our dataset. Moreover, the rendered color reference image remains accurate under consistent lighting conditions.

Weights of pose optimization loss – Table 5. We analyze the impact of λ in the pose optimization loss in Table 5. A weight of 0.6 was chosen to achieve the best performance overall. This reveals that the high-frequency gradients of the color channel are still beneficial for fine-grained registration.

Loss components for anomaly detection – Table 6. We conducted three ablation experiments on different combinations of color and reflectance features for anomaly detection, as shown in Table 6. The results demonstrate that the color or reflectance feature alone may be more accurate at detecting differences at a pixel level, while their combination offers better detection performance and yields the best results overall.

References

[1] Bernhard Kerbl, Georgios Kopanas, Thomas Leimkühler, and George Drettakis. 3d gaussian splatting for real-time radiance field rendering. *ACM Transactions on Graphics (ToG)*, 42(4): 1–14, 2023. 1

[2] Wenhui Wu, Jian Weng, Pingping Zhang, Xu Wang, Wenhan Yang, and Jianmin Jiang. Uretinex-net: Retinex-based deep unfolding network for low-light image enhancement. In *CVPR*, pages 5901–5910, 2022. 5

[3] Qiang Zhou, Weize Li, Lihan Jiang, Guoliang Wang, Guyue Zhou, Shanghang Zhang, and Hao Zhao. Pad: A dataset and benchmark for pose-agnostic anomaly detection. In *NeurIPS*, pages 44558–44571. Curran Associates, Inc., 2023. 2

[4] Matthias Zwicker, Hanspeter Pfister, Jeroen Van Baar, and Markus Gross. Ewa splatting. *IEEE Transactions on Visualization and Computer Graphics*, 8(3):223–238, 2002. 1

Class	Objects	Pixels AUROC↑				Images AUROC↑			
		I+I	I+IR	R+R	R+IR	I+I	I+IR	R+R	R+IR
Real	Valve	99.8	99.7	94.5	99.3	97.3	95.1	98.8	98.8
	Tube	99.6	99.6	98.9	99.6	95.8	92.4	89.4	94.7
	Cup	99.2	99.4	96.8	99.5	95.5	95.3	62.0	92.5
	USB	99.4	99.4	94.9	99.4	52.2	52.4	52.0	55.8
	Joint	99.6	99.7	99.7	99.7	100.0	100.0	99.9	100.0
	PaperCup	99.3	99.6	99.1	99.1	85.9	99.0	86.6	91.1
	Lighter	99.8	99.7	99.2	99.8	98.6	100.0	100.0	99.9
	Cube	99.9	99.1	98.0	99.3	100.0	95.2	69.5	87.7
	Lamp	97.4	97.4	93.7	95.8	100.0	100.0	91.6	95.4
	Bolt	99.7	98.6	98.0	98.9	92.8	92.0	99.1	99.1
	Filter	99.9	99.9	98.4	99.9	97.7	97.7	77.1	97.0
	Wand	99.5	99.3	99.4	99.6	93.1	89.2	92.4	94.4
	Wheel	97.1	96.7	89.3	97.1	92.5	92.3	59.3	94.8
	Bearing	99.6	99.6	92.5	99.7	91.7	91.7	49.7	95.6
Synt	Axletree	96.4	96.3	93.0	99.5	77.8	69.4	61.8	95.3
	Box	99.1	99.1	99.2	99.3	90.0	89.9	92.8	95.9
	Can	99.5	99.5	99.2	99.4	99.9	99.9	100.0	99.9
	Chain	99.2	99.2	98.9	99.1	100.0	100.0	100.0	100.0
	Gear	97.3	97.3	97.0	97.1	97.7	99.7	99.7	99.7
	Keyring	99.5	99.5	99.5	99.5	100.0	100.0	100.0	100.0
	Motor	99.1	99.0	98.9	99.0	98.3	98.4	98.5	98.4
	Parts	99.6	99.6	99.4	99.5	99.4	99.4	99.4	99.4
	Picker	99.4	99.4	99.4	99.4	99.4	99.4	99.4	99.4
	Section	99.2	99.2	99.2	99.2	99.5	99.5	99.4	99.5
	Shaft	99.7	99.7	99.4	99.6	100.0	100.0	100.0	100.0
	Spray_can	99.3	99.3	99.3	99.3	100.0	100.0	100.0	100.0
	Spring	99.6	99.6	99.6	99.5	92.2	92.4	92.6	92.4
	Sprockets	99.6	99.6	99.5	99.6	99.2	99.2	99.2	99.2
	Amphora	96.4	97.4	97.5	97.5	81.5	89.4	84.2	79.5
	Teapot	95.6	95.8	97.5	97.6	82.4	83.5	83.9	87.8
MEAN of *CL*		99.18	99.10	98.10	99.14	94.65	94.53	91.32	95.59
MEAN of *IL*		98.02	98.12	95.77	98.57	89.82	90.63	74.43	91.52
MEAN of All		98.94	98.91	97.63	99.03	93.68	93.75	87.94	94.77

Table 4. **Ablation –** on pose initialization and optimization. The selected configuration (R+IR) performs best overall. We denote the use of color (I) and reflectance (R) images in a module.

Class	Objects	Pixels AUROC↑					Images AUROC↑				
		0.0	0.3	0.6	0.9	1.0	0.0	0.3	0.6	0.9	1.0
Real	Valve	98.9	98.9	99.3	95.3	94.6	99.0	98.9	98.8	97.5	96.5
	Tube	99.6	99.6	99.6	99.2	98.9	95.7	95.6	94.7	90.2	89.4
	Cup	99.4	99.3	99.5	98.8	96.8	92.1	92.2	92.5	78.1	64.4
	USB	99.6	99.6	99.4	98.6	95.0	57.9	56.7	55.8	56.4	52.7
	Joint	99.6	99.6	99.7	99.7	99.7	94.1	94.1	100.0	99.9	99.9
	PaperCup	98.8	98.9	99.1	99.1	99.1	82.9	83.3	91.1	95.5	84.8
	Lighter	99.8	99.8	99.8	99.5	99.1	97.3	98.4	99.9	100.0	100.0
	Cube	99.9	99.8	99.3	98.5	98.0	100.0	97.6	87.7	82.6	69.8
	Lamp	95.9	96.3	95.8	96.5	93.4	100.0	100.0	95.4	95.0	91.6
	Bolt	99.6	99.8	98.9	98.3	97.8	98.4	99.1	99.1	98.9	99.0
	Filter	99.9	99.9	99.9	99.8	98.6	97.6	97.5	97.0	96.8	75.9
	Wand	99.6	99.6	99.6	99.5	99.4	94.5	94.4	94.4	93.7	92.3
	Wheel	97.5	97.5	97.1	95.2	90.2	94.9	94.5	94.8	88.1	58.6
	Bearing	99.7	99.7	99.7	97.2	92.7	95.6	95.5	95.6	73.7	46.2
Synt	Axletree	96.4	96.4	99.5	95.0	93.1	77.7	75.3	95.3	61.7	57.0
	Box	99.3	99.3	99.3	99.3	99.2	96.1	95.7	95.9	92.7	92.6
	Can	99.4	99.4	99.4	99.4	99.2	99.9	99.9	99.9	99.9	100.0
	Chain	99.1	99.2	99.1	99.0	98.9	100.0	100.0	100.0	100.0	100.0
	Gear	97.1	97.1	97.1	97.0	97.0	99.7	99.7	99.7	99.7	99.7
	Keyring	99.5	99.5	99.5	99.5	99.5	100.0	100.0	100.0	100.0	100.0
	Motor	99.1	99.1	99.0	99.0	98.9	98.3	98.3	98.4	98.4	98.5
	Parts	99.5	99.5	99.5	99.5	99.3	99.4	99.4	99.4	99.4	99.4
	Picker	99.5	99.4	99.4	99.4	99.4	99.4	99.4	99.4	99.4	99.4
	Section	99.2	99.2	99.2	99.2	99.1	99.5	99.5	99.5	99.5	99.4
	Shaft	99.6	99.6	99.6	99.6	99.4	100.0	100.0	100.0	100.0	100.0
	Spray_can	99.3	99.3	99.3	99.3	99.3	100.0	100.0	100.0	100.0	100.0
	Spring	99.6	99.6	99.5	99.5	99.6	92.2	92.3	92.4	92.6	92.6
	Sprockets	99.6	99.6	99.6	99.6	99.4	99.2	99.2	99.2	99.2	99.2
	Amphora	97.2	97.4	97.5	97.6	97.7	76.8	76.1	79.5	83.4	85.1
	Teapot	97.2	97.6	97.6	97.4	97.6	82.1	84.7	87.8	80.7	84.3
MEAN of CL		99.05	99.08	99.14	98.66	98.07	94.95	94.78	95.59	93.19	91.08
MEAN of IL		98.52	98.62	98.57	97.78	96.03	90.25	90.45	91.52	86.07	73.73
MEAN		98.95	98.98	99.03	98.48	97.66	94.01	93.91	94.77	91.77	87.61

Table 5. **Ablation –** balance between color and reflectance losses. A weight of 0.6 was chosen to achieve the best performance overall.

Class	Objects	Pixels AUROC↑			Images AUROC↑		
		$\mathcal{S}_I^{\mathcal{F}}$	$\mathcal{S}_R^{\mathcal{F}}$	$\mathcal{S}_I^{\mathcal{F}} + \mathcal{S}_R^{\mathcal{F}}$	$\mathcal{S}_I^{\mathcal{F}}$	$\mathcal{S}_R^{\mathcal{F}}$	$\mathcal{S}_I^{\mathcal{F}} + \mathcal{S}_R^{\mathcal{F}}$
Real	Valve	99.3	99.3	99.3	95.7	99.7	98.8
	Tube	99.6	99.6	99.6	93.1	90.0	94.7
	Cup	99.6	99.5	99.5	96.1	93.8	92.5
	USB	99.5	99.4	99.4	51.1	53.6	55.8
	Joint	99.8	99.8	99.7	100.0	99.6	100.0
	PaperCup	99.1	98.9	99.1	91.8	89.9	91.1
	Lighter	99.9	99.8	99.8	99.5	98.7	99.9
	Cube	99.4	99.2	99.3	90.0	86.7	87.7
	Lamp	96.1	94.9	95.8	95.4	88.8	95.4
	Bolt	98.9	98.8	98.9	100.0	96.6	99.1
	Filter	99.9	99.5	99.9	98.8	82.5	97.0
	Wand	99.6	99.5	99.6	93.8	92.9	94.4
	Wheel	96.8	96.8	97.1	81.1	98.3	94.8
	Bearing	99.7	99.7	99.7	96.4	98.6	95.6
Synt	Axletree	99.5	99.6	99.5	93.2	95.1	95.3
	Box	99.3	99.5	99.3	95.5	94.0	95.9
	Can	99.4	99.5	99.4	99.9	97.7	99.9
	Chain	99.1	99.4	99.1	99.8	99.5	100.0
	Gear	97.1	97.9	97.1	99.9	98.1	99.7
	Keyring	99.6	99.7	99.5	100.0	100.0	100.0
	Motor	99.1	99.0	99.0	85.5	97.8	98.4
	Parts	99.6	99.6	99.5	99.3	97.8	99.4
	Picker	99.5	99.5	99.4	98.4	96.3	99.4
	Section	99.3	99.5	99.2	99.2	98.2	99.5
	Shaft	99.5	99.8	99.6	99.5	100.0	100.0
	Spray_can	99.4	99.5	99.3	100.0	99.9	100.0
	Spring	99.6	99.5	99.5	93.5	87.9	92.4
	Sprockets	99.6	99.5	99.6	99.8	88.4	99.2
	Amphora	96.5	98.4	97.5	72.6	82.6	79.5
	Teapot	97.3	98.0	97.6	83.1	86.7	87.8
MEAN of *CL*		99.20	99.20	99.14	94.84	93.67	95.59
MEAN of *IL*		98.30	98.65	98.57	87.63	90.27	91.52
MEAN of All		99.02	99.09	99.03	93.40	92.99	94.77

Table 6. **Ablation –** on loss components for anomaly detection. Our selected configuration ($\mathcal{S}_I^{\mathcal{F}} + \mathcal{S}_R^{\mathcal{F}}$) yields the best results overall. $\mathcal{S}_I^{\mathcal{F}}$ represents the color feature, while $\mathcal{S}_R^{\mathcal{F}}$ denotes the reflectance feature.