Unsupervised Intrinsic Image Decomposition with LiDAR Intensity Enhanced Training (Supplementary Material)

Contents

A LiDAR intensity 1

B USI^{3}D, IID-LI, and LIET 1
- B.1. Relationship between USI^{3}D, IID-LI, and LIET 2
- B.2. IID-LI 2
- B.3. Ablation study for model architecture 3

C Effect of smoothing loss 4

D Loss effectiveness 4

E. Implementation details 5
- E.1. Dataset 5
- E.2. Network implementation 6
- E.3. IQA metrics 6
- E.4. IID metrics 6

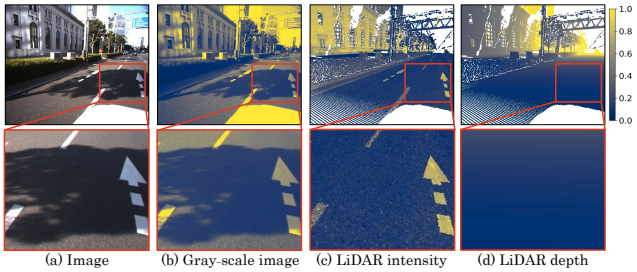
F. Shades inference 7

A. LiDAR intensity

LiDAR intensity is a crucial aspect of the data collected by LiDAR systems, offering insights into the surface characteristics and material composition of scanned objects. This intensity measurement reflects the amount of light that returns to the sensor after being emitted and interacting with a target. Several factors influence LiDAR intensity, including the infrared wavelength reflectance of the surface ρ, which depends on the material's ability to reflect light; the incidence angle θ, with perpendicular angles typically resulting in stronger intensity readings; and the distance between the LiDAR sensor and the target D_{t}, where greater distances lead to diminished intensity due to dispersion and absorption. Thus, LiDAR intensity L is defined as Eq. (1).

$$L = \frac{\eta_{\mathrm{apa}}^2 \eta_{\mathrm{sys}} \eta_{\mathrm{atm}}}{4 D_{\mathrm{t}}^2} \rho \cos\theta, \tag{1}$$

where, η_{apa}, η_{sys}, and η_{atm} denote attenuation by aperture size, camera response function, and atmospheric transmittance accounting. The NTT-IID dataset prepares the LiDAR intensity L, corrected by the object distance and surface angle. The LiDAR intensity represents the intrinsic ability of a surface to reflect light, which is determined by the material composition of the target. Surfaces with high reflectance, such as white concrete, reflect more light back to the sensor compared to those with low reflectance, like asphalt as Fig. 7.



(a) Image (b) Gray-scale image (c) LiDAR intensity (d) LiDAR depth

Figure 7. Examples of (a) input image, (b) gray-scale image, and its corresponding (c) LiDAR intensity and (d) LiDAR depth. The whole image (top) and cropped image (bottom) are illustrated. The shadow and the white arrow are visible in gray-scale image. LiDAR intensity has no cast shadows while maintaining white arrows, since LiDAR intensity is calculated from the intensity ratio of irradiated and reflected lights, equivalent to an albedo at infrared wavelength. Conversely, LiDAR depth represents the distance to objects, resulting in the absence of cast shadows and white arrows.

B. USI^{3}D, IID-LI, and LIET

This section describes the relations between the three models; USI^{3}D, IID-LI, and LIET. Appendix B.1 describes the relationship between these models in terms of model structure and loss including intermediate models. Appendix B.2 discusses in detail IID-LI, which was omitted in the main paper. We evaluate the contribution of each structure and loss by evaluating USI3D, IID-LI, LIET and their intermediate models in Appendix B.3.

B.1. Relationship between USI^{3}D, IID-LI, and LIET

Initially, we describe the relations between USI^{3}D [18], IID-LI [20], and LIET. IID-LI is the model of USI^{3}D incorporating the intensity-consistency (IC) loss, LiDAR-intensity input, and the LiDAR intensity densification (LID) module. The IC loss aligns the luminance value of the albedo inferred from an input image to its corresponding LiDAR intensity for reducing cast shadows. LiDAR-intensity input refers to the input of four channels data consisting of image and LiDAR intensity in IID-LI instead of image input in USI3D to make effective use of LiDAR intensity. The LID module densifies the LiDAR intensity since its sparsity may deteriorate the IID quality. On the other hand, LIET is also the model of USI^{3}D incorporating LiDAR-encoder path, image-LiDAR conversion (ILC) paths, and albedo-alignment loss. To clearly the relationships among these models, we summarize the implementation process beginning from USI3D, along with the intermediary models as illustrated in Fig. 8. Initially, USI^{3}D is implemented for the baseline of IID-LI and LIET. Subsequently, we extended USI^{3}D by incorporating an IC loss that aligns the albedo inferred from an input image into its corresponding LiDAR intensity as USI3D+IC. IID-LI is implemented by incorporating LiDAR-intensity input and the LID module to USI3D+IC. Since the distribution of LiDAR intensity varies across samples, features are well trained by applying instance normalization [24] rather than by scaling uniformly across all samples. Thus, we also define USI^{3}D+ICIN as the model in which the scaling of the IC loss is changed to instance normalization. In addition, a LiDAR-encoder path and ILC path is add to USI^{3}D+ICIN, and LIET+IC is defined to effectively extract the feature of LiDAR intensity. Finally, LIET is implemented by changing IC loss into albedo aligning loss. Since LiDAR and cameras typically operate in distinct wavelength bands, the luminance values of an inferred albedo may be unnatural on the chromatic material surface due to the direct comparison in IC loss. Thus, we introduce albedo aligning loss that aligns the gray-scale albedo inferred from an image to that from its corresponding LiDAR intensity. The LiDAR-intensity utilization and model architecture are summarized in Tab. 4.

B.2. IID-LI

IID-LI [20] is an unsupervised learning model that utilizes LiDAR intensity during training and inference. IID-LI is characterized by IC loss, LiDAR-intensity input, and LID module. IC loss aligns the albedo inferred from an image x_{ri} to the LiDAR intensity x_l in gray scale to enhance the IID quality. In addition, IID-LI is a completely-shared model that accepts both an image and LiDAR intensity as input and processes them simultaneously for both the training and inference processes as shown in Fig. 9 (b). The LID

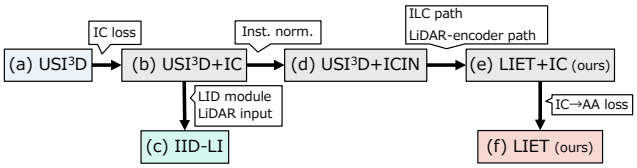


Figure 8. Relations between USI^{3}D, IID-LI, and LIET. (a) USI^{3}D is the baseline model of IID-LI and LIET. (b) USI^{3}D+IC is defined by adding the intensity consistency (IC) loss to USI^{3}D. (c) IID-LI is implemented by adding LiDAR intensity densified (LID) module and LiDAR-intensity input. (d) USI^{3}D+ICIN is the model in which the scaling of the IC loss is changed to instance normalization. (e) LIET+IC is defined for adding the LiDAR-encoder path and ILC path to USI^{3}D+ICIN. (f) LIET is implemented by changing IC loss into albedo aligning (AA) loss. IID-LI is defined for adding LID module and four-channel input into USI^{3}D+IC.

Model	LiDAR utilization	Architecture	Loss for LiDAR
USI^{3}D [18]	–	completely-shared	–
USI^{3}D+IC	train	completely-shared	IC
IID-LI [20]	train&infer	completely-shared	IC
USI^{3}D+ICIN	train&infer	completely-shared	ICIN
LIET+IC	train	partially-shared	ICIN
LIET	train	partially-shared	AA

Table 4. Summary of LiDAR-intensity utilization, model architecture, and LiDAR-intensity related loss for intermediate models including USI^{3}D, USI^{3}D+IC, IID-LI, USI^{3}D+ICIN, and LIET+IC, and LIET in Fig. 8.

module densifies the LiDAR intensity based on deep image prior [25] since the sparsity of LiDAR intensity deteriorates the IID quality. IID-LI has notably enhanced IID quality for the real-world data by utilizing LiDAR intensity, however, IID-LI has restricted applicability due to the need for LiDAR intensity even during inference.

In some cases, IID-LI degrades the image quality of inferred albedo. In the NTT-IID dataset [20], a part of LiDAR intensity is misaligned with the image as shown in Fig. 10 (a) and (b), due to the difficulty of precisely calibrating between the camera and LiDAR. IID-LI processes the image and LiDAR intensity simultaneously by one encoder in such a misaligned sample during inference process. As shown in Fig. 10 (middle), the precisely aligned region of the image quality is maintained in the inferred albedo. Since the misalignment is limited to a portion of the entire training data, the image quality is not degraded for use in training. Conversely, the image quality of the misaligned region is degraded in the inferred albedo as shown in Fig. 10 (bottom), when the data for inference includes the misaligned region. Consequently, the architecture of the completely-shared model may result in image quality degradation for misalignment data. To address this issue, LIET is a partially-shared model that accepts an image and

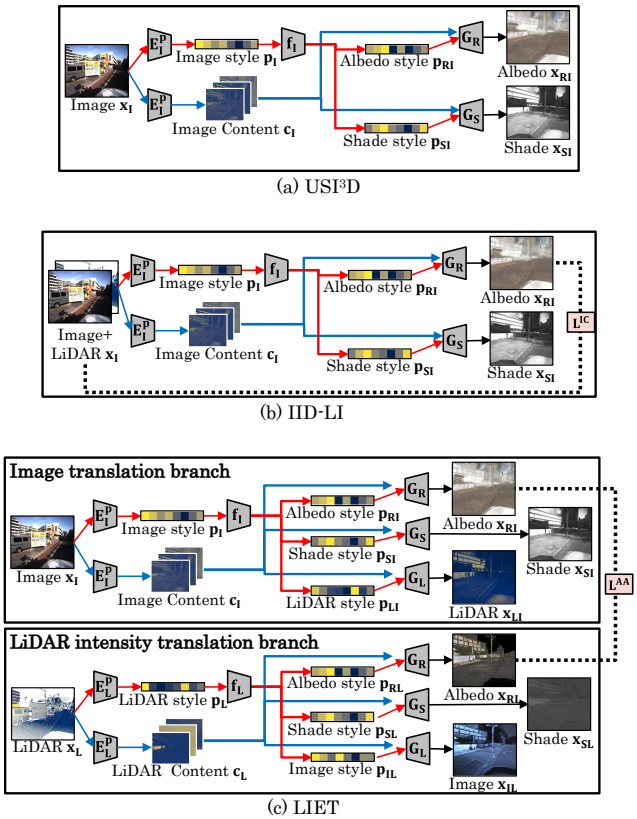


Figure 9. Cross domain translation model in (a) USI^{3}D, (b) IID-LI, and (c) LIET. (a) USI^{3}D infer albedo x_{RI} and shade x_{SI} from an input image x_{I}, through content encoder E_{I}^{c} and style encoder E_{I}^{p}. The albedo style p_{RI} and shade style p_{SI} are derived from image style p_{I} and mapping function f_{I}. (b) IID-LI substitutes the input of USI^{3}D from an image into four-channel data including both an image and its corresponding LiDAR intensity. (c) LIET utilized an image for image-translation branch and LiDAR intensity for LiDAR intensity translation branch. This two-branch architecture realizes the single image inference while an image plus LiDAR intensity training.

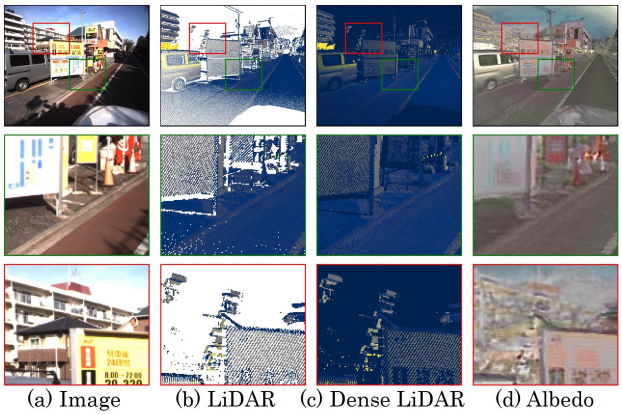


Figure 10. Example of (a) input image x_i, (b) LiDAR intensity x_l, (c) densified LiDAR intensity by LID module, (d) albedo x_{ri} inferred from x_i by IID-LI. The whole images (top row), cropped images for the precisely aligned region (middle), and cropped images for the misaligned region (bottom). In the precisely aligned region between the input image x_i and its corresponding LiDAR intensity x_l, the albedo x_{ri} maintains the same image quality as the input image x_i. However, the image quality of the misaligned region is degraded due to the four-channel inputs.

Method	F-score(↑)	WHDR(↓)	Precision(↑)	Recall(↑)
USI^{3}D [18]	0.454	0.422	0.539	0.500
USI^{3}D+IC	0.516	0.423	0.561	0.511
IID-LI [20]	0.602	0.353	0.625	0.596
USI^{3}D+ICIN	0.567	0.408	0.569	0.569
LIET+IC	0.602	0.377	0.606	0.606
LIET+MMF	0.583	0.383	0.592	0.578
LIET	0.607	0.340	0.649	0.601

Table 5. Ablation study for USI^{3}D, USI^{3}D+IC, IID-LI, USI^{3}D+ICIN, LIET+IC, LIET+MMF, and LIET in Fig. 8, in terms of IID quality.

LiDAR intensity individually using a different specific encoder but processes them together in specific components to learn shared representations.

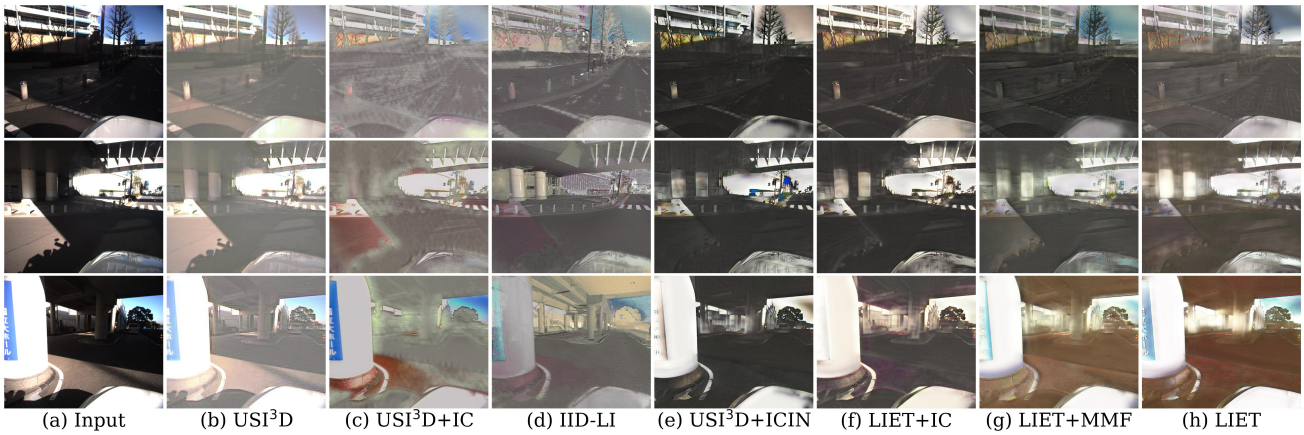
B.3. Ablation study for model architecture

Initially, we perform an ablation study for the model architecture depicted in Fig. 8. The evaluation results for IID quality is described in Tab. 5 and Fig. 11.

USI^{3}D+IC. USI^{3}D+IC is evaluated to verify the contribution of IC loss. The inference process in USI^{3}D+IC only employs a single image, while the LiDAR intensity is used for loss calculation during training. Furthermore, utilizing LiDAR intensity during loss calculation produces better IID quality than USI^{3}D. Consequently, though USI^{3}D+IC achieves single-image input and improves the inferred image quality, the IID quality of USI^{3}D+IC is inferior to that of LIET due to the lack of effective LiDAR-intensity utilization.

USI^{3}D+ICIN. Since the distribution of LiDAR intensity varies across samples, features are well-trained by applying instance normalization rather than by scaling uniformly across all samples. Instance normalization improves the IID quality while keeping the inferred image quality. However, the IID quality still has not reached the level of LIET.

LIET+IC. To extract features of LiDAR intensity more effectively, the LiDAR-encoder path and ILC path are implemented to USI^{3}D+ICIN, and LIET+IC is defined. Consequently, the IID quality is improved and almost reached the level of LIET. However, the inferred albedo luminance may be unnatural on the chromatic material surface with a direct comparison in IC loss. Thus, we introduce an albedo



(a) Input (b) USI^{3}D (c) USI^{3}D+IC (d) IID-LI (e) USI^{3}D+ICIN (f) LIET+IC (g) LIET+MMF (h) LIET

Figure 11. Visual results for an ablation study. The compared models include USI^{3}D, USI^{3}D+IC, IID-LI, USI^{3}D+ICIN, LIET+IC, LIET+MMF, and LIET in Fig. 8.

Method	F-score(↑)	WHDR(↓)	Precision(↑)	Recall(↑)
LIET (Ours)	0.607	0.340	0.649	0.601
w/o $\mathcal{L}^{\text{phy}}$	0.465	0.524	0.489	0.490
w/o $\mathcal{L}^{\text{AA}}$	0.437	0.473	0.497	0.476
w/o $\mathcal{L}^{\text{VGG}}$	0.558	0.398	0.583	0.551
w/o $\mathcal{L}^{\text{KLD}}$	0.579	0.379	0.599	0.573
heavy $\mathcal{L}^{\text{phy}}$	0.581	0.380	0.598	0.576
heavy $\mathcal{L}^{\text{AA}}$	0.529	0.458	0.528	0.539
heavy $\mathcal{L}^{\text{KLD}}$	0.609	0.354	0.626	0.600
heavy $\mathcal{L}^{\text{VGG}}$	0.489	0.488	0.489	0.490

Table 6. Ablation study for each loss in terms of IID quality including $\mathcal{L}^{\text{adv}}$ [8], $\mathcal{L}^{\text{phy}}$ [18], $\mathcal{L}^{\text{AA}}$, $\mathcal{L}^{\text{VGG}}$ [11], and $\mathcal{L}^{\text{KLD}}$.

aligning loss that aligns the gray-scale albedo inferred from a single image to that from its corresponding LiDAR intensity.

LIET+MMF. (Q5) Following the base model USI^{3}D, LIET employs a single mapping function f_{I}. However, we verify the effect of multiple mapping functions by replacing f_{I} into f_{RI}, f_{SI}, f_{LI}.

C. Effect of smoothing loss

Smoothing loss is not implemented in LIET, despite USI^{3}D [18] and IID-LI [20] utilizing it to improve IID quality. Thus, this section describes the effect of smoothing loss for IID quality and IQA, by removing and adding the smooth loss for USI^{3}D [18], IID-LI [20], and LIET [1]. As a result, for both USI^{3}D and IID-LI, eliminating smoothing loss improves the image quality of the inferred albedos as shown in Tab. 7, since the blurring by smoothing loss is reduced. On the other hand, variation caused by cast shadows in the albedo was reduced by smoothing, hence, IID quality degrades without smoothing loss. Conversely, the albedo inferred from LiDAR intensity has less variability derived from cast shadows, hence the variation in the albedo inferred from the image has already reduced due to albedo-alignment loss in LIET. Thus, LIET achieves albedo local flatness without using smoothing loss, which keeps the IQA closely aligned with the input image. Adding further smoothing only degrades IQA and does not improve IID quality.

[1] USI^{3}D with smooth loss, IID-LI with smooth loss, and LIET without smooth loss are marked with a double asterisk (**) since these models are original

D. Loss effectiveness

In LIET, we calculate eight losses including $\mathcal{L}^{\text{img}}$, $\mathcal{L}^{\text{sty}}$, $\mathcal{L}^{\text{cnt}}$ [10], $\mathcal{L}^{\text{adv}}$ [8], $\mathcal{L}^{\text{phy}}$ [18], $\mathcal{L}^{\text{AA}}$, $\mathcal{L}^{\text{VGG}}$ [11], and $\mathcal{L}^{\text{KLD}}$. These losses are divided into two types, essential losses and extra losses. The essential losses including $\mathcal{L}^{\text{img}}$, $\mathcal{L}^{\text{sty}}$, $\mathcal{L}^{\text{cnt}}$, and $\mathcal{L}^{\text{adv}}$ are necessary for stable model training. Additionally, the extra losses including $\mathcal{L}^{\text{phy}}$, $\mathcal{L}^{\text{AA}}$, $\mathcal{L}^{\text{VGG}}$, and $\mathcal{L}^{\text{KLD}}$ are used for improving IID quality. For the training process to proceed well, it is helpful to initially set essential losses to a certain appropriate value. Subsequently, $\mathcal{L}^{\text{phy}}$ is set to reach consistency between the inferred albedo and shade. Next, $\mathcal{L}^{\text{AA}}$ is added to enhance the IID quality. To keep the color of the input image in the inferred albedo, $\mathcal{L}^{\text{VGG}}$ should be added after $\mathcal{L}^{\text{AA}}$. Finally, $\mathcal{L}^{\text{KLD}}$ is used to improve IID quality and image quality. This section describes details of the effect of each loss. Especially, the effect is experimentally verified for extra loss as shown in Tab. 6 and Fig. 12.

Image reconstruction loss $\mathcal{L}^{\text{img}}$. Initially, the input images should be reconstructed after passing through the within-domain reconstruction process, hence image reconstruction loss $\mathcal{L}^{\text{img}}$ is utilized in LIET to extract features. When the weight of the image reconstruction loss is low, feature ex-

Model	$\mathcal{L}^{\text{smooth}}$	IID quality metrics				IQA metrics		
		F-score(↑)	WHDR(↓)	Precision(↑)	Recall(↑)	MANIQA(↑)	TReS(↑)	MUSIQ(↑)
USI^{3}D [18]		0.444	0.427	0.540	0.502	0.629	67.2	52.6
USI^{3}D** [18]	✓	0.454	0.422	0.539	0.500	0.492	53.2	40.1
IID-LI [20]		0.570	0.389	0.581	0.565	0.522	80.9	56.4
IID-LI** [20]	✓	0.602	0.353	0.625	0.596	0.461	55.5	43.5
LIET** (Ours)		0.607	0.340	0.649	0.601	0.570	75.1	56.3
LIET (Ours)	✓	0.593	0.368	0.634	0.583	0.556	69.5	50.1

Table 7. Effect of smoothing loss $\mathcal{L}^{\text{smooth}}$ in IID quality and IQA with the NTT-IID dataset [20]. Smoothing loss improves the IID quality for USI^{3}D and IID-LI, while degrading image quality due to its blurring effect. LIET has already achieved albedo local flatness, hence adding smoothing loss will not improve the IID quality.

traction becomes less effective, making the inferred albedo and shade tend to be blurred. On the other hand, increasing the weight up to 10 times does not have much effect.

Style reconstruction loss $\mathcal{L}^{\text{sty}}$ and content code reconstruction loss $\mathcal{L}^{\text{cnt}}$. Since the reconstructed images should maintain their styles and contents, style reconstruction loss $\mathcal{L}^{\text{sty}}$ and content code reconstruction loss $\mathcal{L}^{\text{cnt}}$ are utilized. These losses are relatively sensitive to parameter weights, and training does not proceed well with inappropriate values.

Adversarial loss $\mathcal{L}^{\text{adv}}$. the adversarial loss $\mathcal{L}^{\text{adv}}$ [8] is defined to ensure that the image inferred through cross-domain translation aligns with the distribution of the target domain. When the weight of adversarial loss is increased, the inferred albedo is more likely to be influenced by the dataset trend in the albedo domain. In the free supervision from video games (FSVG) dataset [13] used in this study, most of the roads in albedo are black, hence the roads in the inferred albedo are more likely to be black. On the other hand, when the weight of adversarial loss is decreased, x_{ri} and x_{si} tend to value 1.0 in all pixel and x_i, respectively, due to the losses such as $\mathcal{L}^{\text{phy}}$ and $\mathcal{L}^{\text{AA}}$. **Physical loss $\mathcal{L}^{\text{phy}}$.** Physical loss is calculated to maintain consistency between the inferred albedo and shade, and is the distance between the input image and the product of the inferred albedo and shade. The inferred albedo tends to be dark and blurred without the physical loss as shown in Fig. 12-(top) since the dark and blurred inference shows the lower loss value in $\mathcal{L}^{\text{img}}$, $\mathcal{L}^{\text{sty}}$ and $\mathcal{L}^{\text{cnt}}$. On the other hand, shadows are more likely to remain in inferred albedo with heavy physical loss weight to strictly match the input image with the product of the inferred albedo and shade as shown in Fig. 12-(bottom).

Albedo aligning loss $\mathcal{L}^{\text{AA}}$. To improve the IID quality, we propose albedo-alignment loss $\mathcal{L}^{\text{AA}}$, aligning the gray-scale albedo from an image to that inferred from its corresponding LiDAR intensity. Cast shadows tend to remain without albedo aligning loss. On the other hand, increasing the weight up to 10 times blurs the inferred albedo as shown in Fig. 12.

VGG loss $\mathcal{L}^{\text{VGG}}$. To preserve the object edges and colors of the input image, the distance between the input image and the inferred albedo within the VGG feature space is computed [4, 22, 26] for the VGG loss $\mathcal{L}^{\text{VGG}}$ [11]. Consequently, the object color in translated image is not consistent without VGG loss, while the inferred albedo is strongly blurred with large weight as shown in Fig. 12.

KLD loss $\mathcal{L}^{\text{KLD}}$. The Kullback-Leibler divergence (KLD) loss $\mathcal{L}^{\text{KLD}}$ aligns the distribution of style codes from within-domain reconstruction and from the mapping module. As the KLD loss helps to extract style code features, KLD loss should be adjusted at the end to improve IID quality or inferred-albedo quality.

E. Implementation details

E.1. Dataset

In this paper, we utilize the NTT-IID dataset [20] for RGB images, LiDAR intensity, and annotations. Additionally, we also utilize FSVG dataset [13] for albedo and shade domains.

NTT-IID dataset. NTT-IID dataset was collected by a mobile mapping system (MMS) with cameras and LiDAR in Japan. To measure the precise position and direction, global navigation satellite system (GNSS) and an inertial measurement unit (IMU) were used. ZF profiler was utilized as a LiDAR, which has 0.0009 degree angular resolution and a maximum range of 120 meters. For IID annotation, they extracted 110 samples and performed Poisson-disk sampling [2]. These sample points were annotated for relative intensity in albedo based on human judgements, and a total number of the annotation is 12,626 in NTT-IID dataset.

FSVG dataset. FSVG dataset prepares synthetic rendered images and its corresponding albedos in outdoor scene. Thus, shades were calculated by the pixel quotient of each image and albedo due to the lack of shade data. Since unsupervised learning was targeted, 10000 samples each of albedo and shade were extracted from the FSVG dataset to eliminate duplication of albedo and shade. In this paper, we employ the same albedo and shade samples as those used in

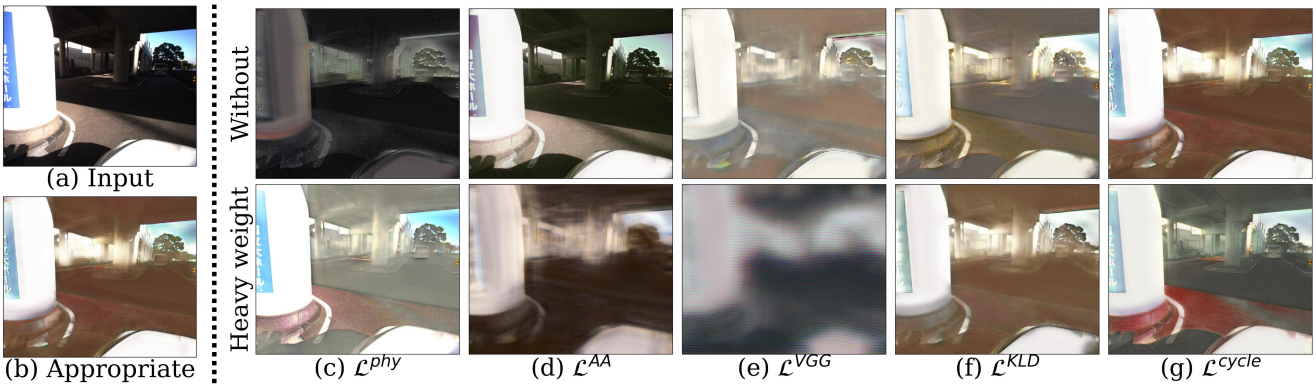


Figure 12. Visual results for each loss effectiveness. Examples of (a) input image and (b) albedo from input image by LIET are shown. The verified loss includes (c) $\mathcal{L}^{\text{phy}}$ [18], (d) $\mathcal{L}^{\text{AA}}$, (e) $\mathcal{L}^{\text{VGG}}$ [11], and (f) $\mathcal{L}^{\text{KLD}}$. The inferred albedos without each loss (top), and with ten times the weight according to each loss (bottom).

Dataset	PLCC	SROCC
LIVE [21]	MANIQA [27]	MANIQA [27]
CSIQ [14]	MANIQA [27]	MANIQA [27]
TID2013 [19]	MANIQA [27]	MANIQA [27]
KADID [17]	MANIQA [27]	MANIQA [27]
CLIVE [6]	HyperIQA [23]	DBCNN [29]
KonIQ [9]	TReS [7]	MUSIQ [12]
LIVEFB [28]	TReS [7]	TReS [7]

Table 8. List of major IQA datasets and the best evaluation models for each dataset in PLCC and SROCC.

IID-LI [20] for fair comparison.

E.2. Network implementation

LIET is implemented based on USI^{3}D [18][2]. The style code encoder E^s_{X}, content code encoder E^c_{X}, generator G_{X}, and discriminator D_{X}, $(X \in I, L, R, S)$, are implemented with the same model structures and parameters as USI^{3}D. Initially, we added generator for LiDAR intensity, and discriminators for image and LiDAR intensity with the same architecture of above models. Subsequently, the channel of multi-layer perceptron in mapping module was expanded from two channels to three channels for inferring LiDAR intensity as well as albedo and shade from an image. Next, the LiDAR intensity inferring path from an image was implemented. Additionally, we prepared a model that infers albedo, shade, and image from LiDAR intensity with exactly the same structure as the model that infers albedo, shade, and LiDAR intensity from image. The albedo aligning loss is implemented with gray scale and instance normalization. For stop gradient, "detach" function was used for the inferred albedo from LiDAR intensity. Finally, VGG loss was implemented following MUNIT [10][3].

E.3. IQA metrics

In this section, we describe the decision process of image quality assessment (IQA) metrics. Initially, we conducted a comprehensive survey across seven prominent IQA datasets: LIVE [21], CSIQ [14], TID2013 [19], KADID [17], CLIVE [6], KonIQ [9], LIVEFB [28]. Subsequently, we identified the best performing IQA models for each dataset based on Pearson's linear correlation coefficient (PLCC) and Spearman's rankorder correlation coefficient (SROCC) as shown in Tab. 8. Consequently, we employed the following IQA models by using pyIQA [3]: MANIQA [27], TReS [7], MUSIQ [12], HyperIQA [23], DBCNN [29].

E.4. IID metrics

This section describes the IID quality assessment metrics. Typically, IID quality for actual measurement datasets is evaluated by the degree of agreement with human-judged annotation. As annotations, sparse points are initially extracted from the test image by Poisson sampling. Then, two neighboring points are selected from the sampled points, and the annotators judge the point to be darker in case the test image is albedo. The decision is based on the majority vote, and the confidence level of the decision is calculated based on the agreement rate. This operation is repeated for all neighboring points. The total number of annotations is 12,626 in NTT-IID dataset. To quantify the IID quality, for albedo inferred from the image, two neighboring points that have already been annotated are picked up, and the lumi-

[2]GitHub page for USI^{3}D (https://github.com/DreamtaleCore/USI3D.git)

[3]GitHub page for MUNIT (https://github.com/NVlabs/MUNIT.git)

Figure 13. Examples of inferring shades obtained from various existing models and LIET (Ours) with NTT-IID dataset [20]. The compared models include Revisiting [5], IIDWW [16], UidSequence [15], USI^{3}D [18], and IID-LI [20].

nance between the two points is compared on a gray scale.

$$(\hat{J}) = \begin{cases} D & \text{if } \mathrm{R_L}/\mathrm{R_D} > 1+\delta \\ L & \text{if } \mathrm{R_D}/\mathrm{R_L} > 1+\delta \\ E & \text{else.} \end{cases} \quad (2)$$

R_L and R_D denote lighter and darker points of annotated points. The threshold is set to $\delta = 0.1$, as the existing papers [1, 20]. Finally, IID quality is quantitatively evaluated by comparing the estimation results with human-judged annotation.

We computed the macro average of the three-class classification with confidence-level weighting. Annotation bias towards "same intensity" in NTT-IID dataset favors flat images in accuracy and precision, but the bias is less effective in F-score, recall, and random sampling assessments.

F. Shades inference

In this section, we describe the shadow inference results obtained from various existing models and LIET with the NTT-IID dataset [20]. The compared models include Revisiting [5], IIDWW [16], UidSequence [15], USI^{3}D [18], and IID-LI [20]. As shown in Fig. 13, the shadow inference results from each model are almost the same. On the other hand, the shadow inferred by USI^{3}D is slightly whitish.

References

[1] Sean Bell, Kavita Bala, and Noah Snavely. Intrinsic Images in the Wild. *ACM TOG*, 33(4):1–12, 2014. 7

[2] Robert Bridson. Fast Poisson disk sampling in arbitrary dimensions. *SIGGRAPH*, 10(1):1, 2007. 5

[3] Chaofeng Chen and Jiadi Mo. IQA-PyTorch: PyTorch Toolbox for Image Quality Assessment. [Online]. Available: https://github.com/chaofengc/IQA-PyTorch, 2022. 6

[4] Qifeng Chen and Vladlen Koltun. Photographic Image Synthesis with Cascaded Refinement Networks. In *ICCV*, pages 1511–1520, 2017. 5

[5] Qingnan Fan, Jiaolong Yang, Gang Hua, Baoquan Chen, and David Wipf. Revisiting Deep Intrinsic Image Decompositions. In *CVPR*, pages 8944–8952, 2018. 7

[6] Deepti Ghadiyaram and Alan C Bovik. Massive Online Crowdsourced Study of Subjective and Objective Picture Quality. *IEEE TIP*, 25(1):372–387, 2015. 6

[7] S Alireza Golestaneh, Saba Dadsetan, and Kris M Kitani. No-Reference Image Quality Assessment via Transformers, Relative Ranking, and Self-Consistency. In *WACV*, pages 1220–1230, 2022. 6

[8] Ian Goodfellow, Jean Pouget-Abadie, Mehdi Mirza, Bing Xu, David Warde-Farley, Sherjil Ozair, Aaron Courville, and Yoshua Bengio. Generative Adversarial Nets. *NeurIPS*, 27:139–144, 2014. 4, 5

[9] Vlad Hosu, Hanhe Lin, Tamas Sziranyi, and Dietmar Saupe. KonIQ-10k: An ecologically valid database for deep learning of blind image quality assessment. *IEEE TIP*, 29:4041–4056, 2020. 6

[10] Xun Huang, Ming-Yu Liu, Serge Belongie, and Jan Kautz. Multimodal Unsupervised Image-to-Image Translation. In *ECCV*, pages 172–189, 2018. 4, 6

[11] Justin Johnson, Alexandre Alahi, and Li Fei-Fei. Perceptual Losses for Real-Time Style Transfer and Super-Resolution. In *ECCV*, pages 694–711, 2016. 4, 5, 6

[12] Junjie Ke, Qifei Wang, Yilin Wang, Peyman Milanfar, and Feng Yang. MUSIQ: Multi-scale Image Quality Transformer. In *ICCV*, pages 5148–5157, 2021. 6

[13] Philipp Krähenbühl. Free Supervision From Video Games. In *CVPR*, pages 2955–2964, 2018. 5

[14] Eric C Larson and Damon M Chandler. Most apparent distortion: full-reference image quality assessment and the role of strategy. *Journal of electronic imaging*, 19(1):011006–011006, 2010. 6

[15] Louis Lettry, Kenneth Vanhoey, and Luc Van Gool. Unsupervised Deep Single Image Intrinsic Decomposition using Illumination Varying Image Sequences. *Computer Graphics Forum*, 37, 2018. 7

[16] Zhengqi Li and Noah Snavely. Learning Intrinsic Image Decomposition from Watching the World. In *CVPR*, pages 9039–9048, 2018. 7

[17] Hanhe Lin, Vlad Hosu, and Dietmar Saupe. KADID-10k: A large-scale artificially distorted IQA database. In *QoMEX*, pages 1–3, 2019. 6

[18] Yunfei Liu, Yu Li, Shaodi You, and Feng Lu. Unsupervised Learning for Intrinsic Image Decomposition from a Single Image. In *CVPR*, pages 3248–3257, 2020. 2, 3, 4, 5, 6, 7

[19] Nikolay Ponomarenko, Oleg Ieremeiev, Vladimir Lukin, Karen Egiazarian, Lina Jin, Jaakko Astola, Benoit Vozel, Kacem Chehdi, Marco Carli, Federica Battisti, et al. Color image database TID2013: Peculiarities and preliminary results. In *EUVIP*, pages 106–111, 2013. 6

[20] Shogo Sato, Yasuhiro Yao, Taiga Yoshida, Takuhiro Kaneko, Shingo Ando, and Jun Shimamura. Unsupervised Intrinsic Image Decomposition With LiDAR Intensity. In *CVPR*, pages 13466–13475, 2023. 2, 3, 4, 5, 6, 7

[21] Hamid R Sheikh, Muhammad F Sabir, and Alan C Bovik. A Statistical Evaluation of Recent Full Reference Image Quality Assessment Algorithms. *IEEE TIP*, 15(11):3440–3451, 2006. 6

[22] Karen Simonyan and Andrew Zisserman. Very Deep Convolutional Networks for Large-Scale Image Recognition. In *ICLR*, pages 1–14, 2015. 5

[23] Shaolin Su, Qingsen Yan, Yu Zhu, Cheng Zhang, Xin Ge, Jinqiu Sun, and Yanning Zhang. Blindly Assess Image Quality in the Wild Guided by a Self-Adaptive Hyper Networkk. In *CVPR*, pages 3667–3676, 2020. 6

[24] Dmitry Ulyanov, Andrea Vedaldi, and Victor Lempitsky. Instance Normalization: The Missing Ingredient for Fast Stylization. *arXiv preprint arXiv:1607.08022*, 2016. 2

[25] Dmitry Ulyanov, Andrea Vedaldi, and Victor Lempitsky. Deep Image Prior. In *CVPR*, pages 9446–9454, 2018. 2

[26] Ting-Chun Wang, Ming-Yu Liu, Jun-Yan Zhu, Andrew Tao, Jan Kautz, and Bryan Catanzaro. High-Resolution Image Synthesis and Semantic Manipulation with Conditional GANs. In *CVPR*, pages 8798–8807, 2018. 5

[27] Sidi Yang, Tianhe Wu, Shuwei Shi, Shanshan Lao, Yuan Gong, Mingdeng Cao, Jiahao Wang, and Yujiu Yang. MANIQA: Multi-dimension Attention Network for No-Reference Image Quality Assessment. In *CVPRW*, pages 1191–1200, 2022. 6

[28] Zhenqiang Ying, Haoran Niu, Praful Gupta, Dhruv Mahajan, Deepti Ghadiyaram, and Alan Bovik. From patches to pictures (PaQ-2-PiQ): Mapping the perceptual space of picture quality. In *CVPR*, pages 3575–3585, 2020. 6

[29] Weixia Zhang, Kede Ma, Jia Yan, Dexiang Deng, and Zhou Wang. Blind Image Quality Assessment Using A Deep Bilinear Convolutional Neural Network. *IEEE TCSVT*, 30(1):36–47, 2018. 6