International Information and Engineering Technology Association
Advancing the World of Information and Engineering

Traitement du Signal
Vol. 41, No. 1, February, 2024, pp. 511-517
Journal homepage: http://iieta.org/journals/ts

Enhancing Intrinsic Image Decomposition with Transformer and Laplacian Pyramid Network

Jianxin Liu[1], Yupeng Ma[2], Xi Meng[2], Shan Zhang[1], Zhiguo Liu[2], Yufei Song[2*]

[1] National Computer System Engineering Research Institute of China, China Electronics Corporation, Beijing 100000, China
[2] College of Future Information Technology, Shijiazhuang University, Shijiazhuang 050035, China

Corresponding Author Email: 2010005@sjzc.edu.cn

https://doi.org/10.18280/ts.410146

Received: 3 September 2023
Revised: 26 December 2023
Accepted: 10 January 2024
Available online: 29 February 2024

Keywords:
Intrinsic Image Decomposition (IID), transformer, Laplacian pyramid, reflectance, shading, feature encoding

ABSTRACT

Intrinsic Image Decomposition (IID) remains a pivotal challenge in the domain of computer vision, with applications spanning image editing, color image denoising, and segmentation, among others. Despite notable successes, there exists a significant opportunity for enhancing the feature encoding process to improve the accuracy of predicted outcomes. In response to this, a novel framework, termed Transformer and Laplacian Pyramid Network (TLPNet), is introduced. TLPNet comprises two distinct sub-networks: the Transformer for Reflectance Network (TRNet) and the Laplacian Pyramid for Shading Network (LPSNet). Within this framework, the Transformer module is strategically employed within the reflectance imaging component to effectively address the challenge of inadequate feature information extraction. Comprehensive experiments conducted on the ShapeNet Dataset and MIT Dataset have demonstrated the efficacy of TLPNet in predicting more accurate reflectance and shading images. This study contributes to the field by presenting an innovative approach that leverages the strengths of transformer models and Laplacian pyramid structures for the task of IID, setting a new benchmark for future research in the area.

1. INTRODUCTION

IID has made great progress. In 1971, Land and McCann [1] proposed the Retinex Theory. The Retinex Theory, a portmanteau of “retina” and "cortex," suggests that color perception is not solely determined by the wavelength of light entering the eyes. Instead, it proposes that the brain compares and processes the light reflected from different parts of a scene to comprehend the color and brightness consistently. This process helps in maintaining color constancy under varying shading conditions. The Retinex theory differentiates between shading and reflectance components in an image by analyzing gradients [1, 2]. Barrow et al. [3] explored the idea of decomposing an image into its intrinsic components. Eq. (1) encapsulates this idea, where I represents the observed image:

$$I = R * S \tag{1}$$

R is the reflectance image, indicating the intrinsic color properties of the objects in the scene, independent of the illumination conditions; S is the shading image, which contains the incident illumination and the shape information. Reflectance images and shading images have great significance in other research fields. Reflectance has been found to enhance various image processing and computer vision tasks. In particular, it can greatly improve the performance of semantic segmentation [4], help find attempts to spoof face recognition systems [5], lead to more in-depth facial intrinsic analysis [6], and make photo editing better [7]. On the other hand, shading is particularly useful in extracting shape information from images (shape from shading) [8], assisting in relighting tasks for more realistic lighting effects [9], and playing a critical role in 3D reconstruction processes [10]. Each of these applications leverages the unique properties of reflectance and shading to improve the quality and accuracy of the results. Ma et al. [11] comprehensively examine various approaches and methodologies developed over the years for separating the reflectance and illumination components of an image. It likely covers the theoretical underpinnings of the technique, evaluates different algorithms, and discusses their applications and limitations.

It can be known from Eq. (1) that the IID is ill-posed. Many solutions were proposed by researchers to address this problem. Before deep learning and neural networks became mainstream, IID mainly relied on some traditional computer vision techniques. These methods are usually based on the physical properties of the image, heuristic rules, or statistical models [12-16]. These methods also have problems. Different scenes may require different physical constraints, and selecting or adjusting these constraints can be difficult without prior knowledge about the scene. In dynamic or changing environments, such as outdoor scenes or changing shading conditions, physical constraints may need to be adjusted in real time, which increases the complexity of the algorithm. Deep learning models can better generalize to different images and scenes after being trained on large and diverse data, rather than being limited to specific physical conditions or constraints. Deep learning and neural networks have developed vigorously in recent years [16, 17]. In this field of IID, there are two main methods: unsupervised learning [18, 19] and supervised learning [20-24]. Unsupervised learning does not rely on labeled training data. It learns image decomposition by exploring the intrinsic structure of the data. Supervised learning relies on labeled training data. These labels guide the learning process to ensure that the decomposition results match the expected output.

In this paper, we propose IID using a TLPNet. First, the application of Transformer in this field of image processing is gradually increasing. On the one hand, Transformer is able to capture global dependencies through the self-attention mechanism, and even regions that are far apart in the image can directly influence each other. On the other hand, the Transformer architecture performs well when handling large-scale data sets, especially when large amounts of training data are available. As model size increases, transformers are often better able to utilize the additional data to improve performance. In addition, we also use the Laplacian pyramid to improve the effect of the shading image. Laplacian pyramid by representing images at different scales. This method allows the shading image to extract image features from coarse to fine. The Laplacian pyramid can be used to enhance the details of an image, such as by sharpening edges and textures. In summary, our contributions are as follows:

●We apply Transformer to the encoder of IID, which can better extract features from the image. Using transformers can often make better use of the extra data to improve performance.

●When the Laplacian pyramid is used to process shading images, its multiscale representation is used to improve image features step by step from big to small, especially by sharpening edges and textures. This makes the shading image better overall and better at capturing shading details.

●Multiple loss functions, which include mean square error (MSE) and cosine similarity error (CSE), are applied in our TLPNet. The loss of reflectance image, shading image, and reconstruction image have different weights.

2. RELATED WORKS

2.1 Transformer module

The Transformer model was originally proposed by Vaswani et al. [25]. It is mainly used to solve sequence-to-sequence tasks in natural language processing (NLP), such as machine translation. Its core innovation is the self-attention mechanism, which enables the model to more effectively capture long-distance dependencies when processing sequence data.

The Transformer model is also used in the image field. It can help improve the performance of image classification [26, 27], contribute to generating more realistic pictures [28], assist in object detection [29], and play a critical role in semantic segmentation [30].

2.2 Laplacian Pyramid

The origins of the Laplacian Pyramid date back to 1983, as proposed by Burt and Adelson [31] in their seminal paper. This concept is developed on the basis of the Gaussian pyramid and aims to represent the multiscale information in images more effectively. The Laplacian pyramid is constructed by subtracting the unsampled and smoothed version of the previous layer from each layer of the Gaussian pyramid. This method can effectively capture the detailed information of images at different scales and is suitable for various applications such as image compression, image fusion, and feature extraction. Burt and Adelson [32] introduced the multiresolution spline method using the Laplacian pyramid for image stitching and fusion. Jacobs et al. [33] used the Laplacian Pyramid for fast multiresolution image query and demonstrated its application in image retrieval. Do and Vetterli [34] proposed the contourlet transform, which combines the Laplacian Pyramid and the directional filter to represent the directional information of the image more effectively. Song et al. [35] introduce a novel method that leverages a Laplacian pyramid-based architecture to decode depth residuals, focusing on refining the depth boundaries and global layout of depth maps.

3. IMAGE DECOMPOSITION USING TRANSFORMER AND LAPLACIAN PYRAMIDBASED (IIDTLP)

3.1 Architecture network

Currently, there are two different network structures in deep learning for IID. One is the parallel structure, with the reflectance image and shading image using the same encoder to extract image features but using different decoders to generate the reflectance image and shading image. Another one is serial structure [22], which uses two encoders and decoders to reconstruct the reflectance image and the shading image.

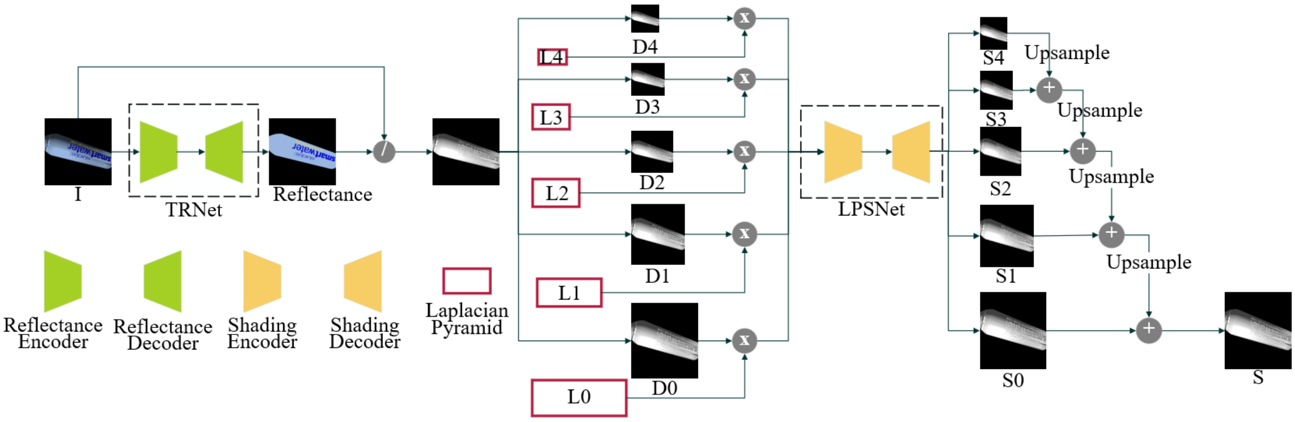

Figure 1. Architecture network of TLPNet

As shown in Figure 1, we propose TLPNet, which adopts a similar structure to CasQNet. The reason we use a serial network structure is that the reflection image generated by the Transformer network is better. First, we use a Transformer encoder to extract image features; it generates a reflectance image from a natural image. Second, we use the outcome of the quotient of the natural image and the reflectance image to splice the natural image. We use the splicing outcome to extract image features to get the shading image.

3.2 Transformer for reflectance

TRNet is designed for constructing reflection images from nature images. The architecture is shown in Figure 2. First of all, we use a transformer encoder to exact the features from the nature image. Transformer encoders can catch more features from nature images than other neural networks. Second, we down-sample the reflectance image. We get the reflectance images of 1/2, 1/4, 1/8, and 1/16 sizes to prepare for obtaining the shading image. Figure 3 has more detail about the Transformer encoder and decoder. The Transformer block in Figure 3 is shown in Figure 4.

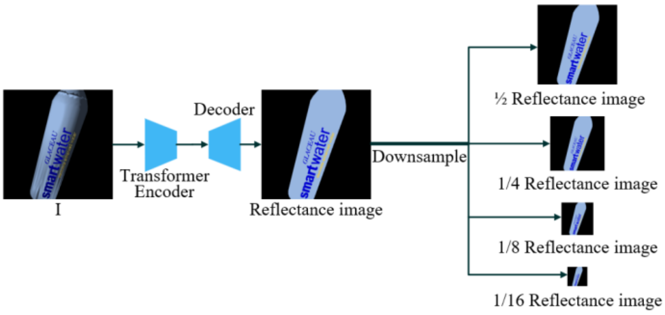

Figure 2. Architecture network of TRNet

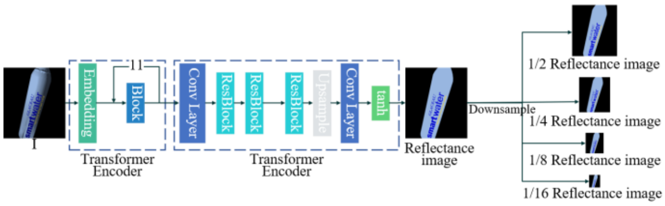

Figure 3. Architecture network of transformer encoder and decoder

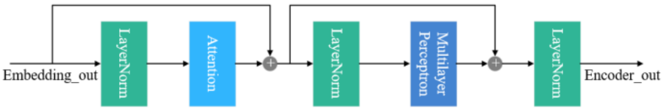

Figure 4. Architecture network of transformer block

3.3 Laplacian pyramid for shading

In the process of predicting shading images, we first use Eq. (2) to perform quotient operations on original images and reflectance images of different sizes. Since the predicted reflection image is not completely accurate, this will cause the edges of the shading map to be unsmooth and the texture to be unclear. We call this result the temporary shading image. In order to solve the problems raised above, we applied the Laplacian Pyramid to our model. Laplacian Pyramid is a technique used in image processing, mainly for multi-scale decomposition of images. It is a further development based on the Gaussian Pyramid and is used to capture details in images in greater detail. Before extracting shading features, we concatenate the temporary shading map and the Laplacian map of corresponding sizes. This can better preserve important information in the image. Figure 5 also shows the process of shading the image from coarse to fine. In Figure 5, Ri (i from 2 to 5) is the different size of reflectance image, R size is the same to the nature image; Di (i from 0 to 4) is the different size of temporary shading image; and L is the result of the different sizes of Laplacian pyramid. The highest level of the shading image is as follows:

$$S = I/R \tag{2}$$

$$S = S_k + Up(S_{k+1}), k = 0,1,2,3 \tag{3}$$

S4 contains the global layout of the shading map at the image pyramid. By iteratively computing Eq. (3) with the order of $k = 3 \rightarrow 2 \rightarrow 1 \rightarrow 0$, S is computed as the final shading image.

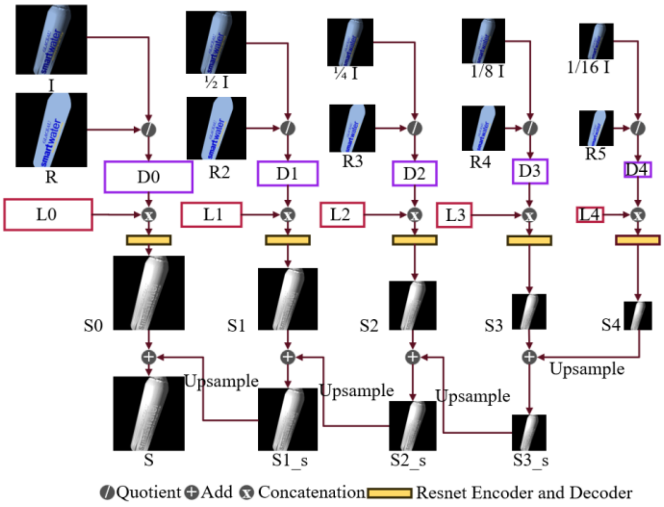

Figure 5. Architecture network of LPSNet

3.4 Loss function

The trainable parameters of the TLIID network are optimized based on three loss functions, which include L_R, L_S, and L_C, as follows:

$$L_t = \begin{cases} L_R & epoch < 20 \\ L_S & 20 \leq epoch < 40 \\ \lambda_R L_R + \lambda_S L_S + \lambda_I L_C & 40 \leq epoch < 60 \end{cases} \tag{4}$$

The loss function of L_R is designed for TRNet, L_S is designed for LPSNet. Also, we design L_C for the product of the predicting R and S. In training process, the parameters are empirically set as: $\lambda_R = 1.0$; $\lambda_S = 1.0$; $\lambda_I = 0.25$; We can conclude from I = R * S, R and S are two parts of I, R and S perform multiplication operations, resulting in $\lambda_I = 0.5 * 0.5$.

The effect of reflection images is an important section on IID. The reflectance loss function is the most important part to impact reflection images. Therefore, we use L_{Rmse} and L_{Rcse} to compose the reflectance loss, which are defined as follows:

$$L_R = L_{Rmse} + L_{Rcse} \tag{5}$$

$$L_{Rmse} = MSE(R,\hat{R}) = \frac{1}{n}\sum_{i=0}^{n} (R_i - \widehat{R_\iota})^2 \tag{6}$$

$$L_{cse}(R,\hat{R}) = \frac{R * \hat{R}}{\|R\|\|\hat{R}\|} = \frac{\sum_{i=0}^{n} (R_i * \widehat{R_\iota})}{\sqrt{\sum_{i=0}^{n} R_i^2} * \sqrt{\sum_{i=0}^{n} \widehat{R_\iota}^2}} \tag{7}$$

$$L_{Rcse} = 1 - \frac{L_{cse}(R,\hat{R}) + 1}{2} \tag{8}$$

where, R_i represents the actual point values and $\widehat{R_\iota}$ represents the predicted point values. For each data point, subtract the predicted value from the actual value. n is the quantity of data points. Cos Similarity Error (CSE)∈[-1, 1], In order to achieve the best effect when L_{Rcse} is 0. We did some processing on the CSE in Eq. (8).

For the LPSNet, the loss function is composed of MSE and CSE for predicting shading images. The shading loss is as follows:

$$L_S = L_{Smse} + L_{Scse} \tag{9}$$

$$L_{Smse} = MSE(S,\hat{S}) = \frac{1}{n}\sum_{i=0}^{n} (S_i - \widehat{S_\iota})^2 \tag{10}$$

$$L_{cse}(S,\hat{S}) = \frac{S * \hat{S}}{\|S\|\|\hat{S}\|} = \frac{\sum_{i=0}^{n} (S_i * \widehat{S_\iota})}{\sqrt{\sum_{i=0}^{n} S_i^2} * \sqrt{\sum_{i=0}^{n} \widehat{S_\iota}^2}} \tag{11}$$

$$L_{Scse} = 1 - \frac{L_{cse}(S,\hat{S}) + 1}{2} \tag{12}$$

where, L_{Smse} is the MSE of shading image, L_{Scse} is the CSE of the shading image. S is the ground-truth image. $\hat{S}$ is the estimated shading image.

According to $\hat{I} = \hat{R} * \hat{S}$, we estimate $\hat{I}$ by using $\hat{R}$ and $\hat{S}$. Reconstruction image loss functions are also composed of MSE and CSE. The reconstruction image loss as follows:

$$L_I = L_{Imse} + L_{Icse} \tag{13}$$

$$L_{Imse} = MSE(I,\hat{I}) = \frac{1}{n}\sum_{i=0}^{n} (I_i - \widehat{I_\iota})^2 \tag{14}$$

$$L_{Icse}(I,\hat{I}) = \frac{I * \hat{I}}{\|I\|\|\hat{I}\|} = \frac{\sum_{i=0}^{n} (I_i * \widehat{I_\iota})}{\sqrt{\sum_{i=0}^{n} I_i^2} * \sqrt{\sum_{i=0}^{n} \widehat{I_\iota}^2}} \tag{15}$$

$$L_{Icse} = 1 - \frac{L_{cse}(I,\hat{I}) + 1}{2} \tag{16}$$

where, L_{Imse} is the MSE of the reconstruction image, L_{Icse} is the CSE of the reconstruction image.

The loss function was designed using MSE and CSE. MSE can compare pixel values directly. It is sensitive to noise and outliers. CSE focuses on directional similarity rather than size. CSE is more in line with human perceptions of image similarity.

4. EXPERIMENTS AND ANALYSIS

In this section, we evaluated the TLPNet on the MIT intrinsic image dataset [36] and the ShapeNet dataset [37] to evaluate its effectiveness.

4.1 Experiments setup

Evaluation Indicators: We use three main evaluation methods: MSE, local mean square error (LMSE) [38], and structural dissimilarity (DSSIM) [39].

Implementation Details: We use the PyTorch machine learning framework to implement TLPNet. In training, the input image size is 256*256. We introduce a transformer in the reflectance image net, and the shading image adopts a 5-level Laplacian Pyramid. The learning rate is 10^{-4}. The epoch is 60. We train TRNet in the first 20 epochs; LPSNet is trained in the middle 20 epochs; And in the last 20 epochs, we train the total net (TLPNet).

4.2 Quantitative comparison

We performed a series of comparative experiments on the ShapeNet dataset and the MIT dataset. In these datasets, our method gets great performance both in terms of quantitative analysis and visual quality. In this section, we compare the one-level or many-level shading image, applying or not applying the Laplacian pyramid, adding or not adding reconstruction image error, and the weight of different reconstruction image errors.

4.2.1 Ablation study

In this subsection, comparative experiments are conducted on a quarter of ShapeNet dataset to verify the effectiveness of the proposed architecture. Laplacian pyramid-based and multi-layer shading images on different weights in the loss function. First, performance variations are evaluated according to the level of shading image. The architecture changes according to different levels of shading image. One-level shading image is shown in Figure 6. The change from concatenating the Laplacian pyramid to a temporary shading image is shown in Figure 3. Table 1 shows the estimated results of different architectures and weights in the loss function. Our final model, TLPNet, achieves the best overall performance on the ShapeNet dataset. Thanks to our proposed Laplacian Pyramid and the special weight of reconstruction loss function.

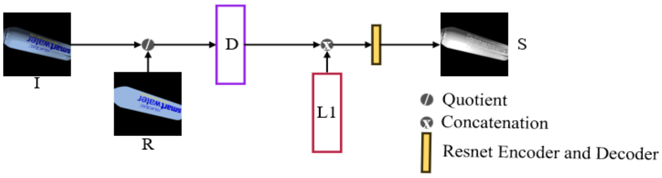

Figure 6. Architecture network of one shading image

4.2.2 Comparison on ShapeNet dataset

In this experiment, the proposed network is compared to other methods. All methods presented here are using the ShapeNet dataset. The results are shown in Table 2. The proposed method generally outperforms, on average, all other methods except for the DSSIM metric. Particularly, the proposed method outperforms other methods for the MSE and LMSE metrics.

Table 1. Performance analysis of the proposed method on the ShapeNet dataset according to the architecture and loss function

Method	MSE		LMSE		DSSIM	
	R	***S***	***R***	***S***	***R***	***S***
one level shading + Laplacian Pyramid + ($L_R + L_S + 0.25L_C$)	0.0060	0.0037	0.0065	0.0041	0.0376	0.0385
multi-level shading + ($L_R + L_S + 0.25L_I$)	0.0061	0.0039	0.0067	0.0044	0.0436	0.0418
multi-level shading + Laplacian Pyramid + ($L_R + L_S + 0.5L_C$)	0.0058	0.0044	0.0064	0.0049	0.0393	0.0547
multi-level shading + Laplacian Pyramid + ($L_R + L_S$)	0.0056	0.0035	0.0063	0.0042	0.0389	0.0451
multi-level shading + Laplacian Pyramid + ($L_R + L_S + 0.25L_C$)	**0.0056**	**0.0035**	**0.0061**	**0.0040**	**0.0376**	**0.0350**

Table 2. Quantitative evaluation on ShapeNet dataset

Method	MSE		LMSE		DSSIM	
	R	***S***	***R***	***S***	***R***	***S***
Baseline	-	-	0.0789	0.0231	0.2273	0.2341
SIRFS [36]	0.0061	0.0039	0.0067	0.0044	0.0436	0.0418
IIW [40]	0.0167	0.0127	0.4810	0.2280	0.1679	0.1367
DI [41]	0.0252	0.0245	0.0711	0.0275	0.1987	0.1454
Han [42]	-	-	0.0101	0.0119	0.0490	0.0503
Shi [43]	0.0278	0.0126	0.0353	0.0097	0.0939	0.0622
CascadeQ [22]	0.0047	0.0035	0.0053	0.0053	**0.0060**	**0.0065**
Our method	**0.0037**	**0.0024**	**0.0041**	**0.0027**	0.0243	0.0227

Table 3. Quantitative evaluation on MIT dataset

Method	MSE		LMSE		DSSIM	
	R	***S***	***R***	***S***	***R***	***S***
Retinex [1]	-	-	0.0353	0.1027	0.1825	0.3987
SIRFS [36]	-	-	0.0416	0.0168	0.1238	0.0985
DI [41]	0.0252	0.0245	0.0585	0.0295	0.1526	0.1328
Shi [43]	0.0278	0.0126	0.0503	0.0240	0.1465	0.1200
CasQNet [22]	0.0107	0.0106	0.0206	0.0186	0.0734	0.0705
NCCNet [44]	0.0104	0.0081	0.0137	0.0128	**0.0581**	**0.0580**
Our method (MIT)	**0.0077**	**0.0068**	**0.0087**	**0.0077**	0.0932	0.0628

4.2.3 Comparison on MIT dataset

In this section, we compare the other methods on the MIT Dataset [36]. The results are shown in Table 3. Only the DSSIM metric performs best; the other metric generally outperforms on TLPNet. Based on the training results of the ShapeNet dataset, we input the MIT data set for training for 20 epochs.

5. CONCLUSION

In this work, we present a novel model called TLPNet for IID. TLPNet is a hybrid transformer for reflectance and a Laplacian Pyramid for shading. We train both sub-nets jointly using our proposed special weight loss function. The TLPNet estimated reflectance image and shading image are substantially more consistent than the state-of-the art baselines, both locally and globally. The result image of different algorithms on the ShapeNet dataset is shown in Figure 7. From Figure 7, the first row is the clean image from the ShapeNet dataset. For each sample, R is the estimated reflectance and S is the estimated shading. Ranjit is the algorithm from the study [45]. Carega is the algorithm from the study [46].

Nevertheless, there is still future work to consider. Our model struggles to accurately estimate reflectance and shading in DSSIM. Besides, as the training data mainly consists of indoor scenes, it seems harder to deal with outdoor scenes. Further research could focus on improving high-quality reconstruction in these two aspects. If we have more datasets under water, we would apply the method in the study [47] to get reflectance and shading images under water.

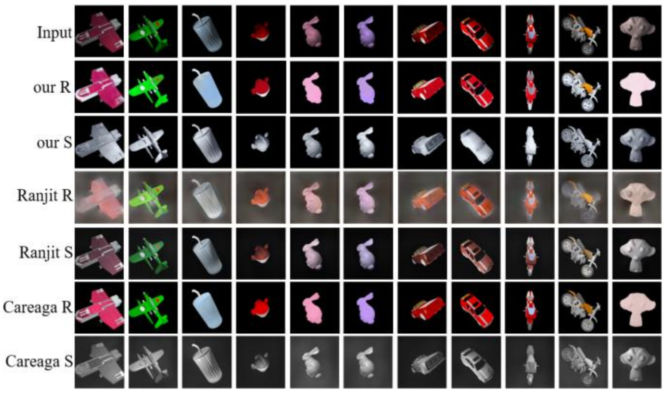

Figure 7. The result image of different algorithms on the ShapeNet dataset

ACKNOWLEDGMENT

We would like to thank Luo et al. [21] for publishing their source code; Ma et al. [22] for creating cascaded quotient networks (CasQNet) for us; Do and Martin [34] and Grosse et al. [38] for creating metrics for us.

FUNDING

This work was supported in part by the National Nature Science Foundation of China (Grant No.: 62201195); and in part by the Doctoral Startup Foundation of Shijiazhuang University (Grant No.: 23BS022 and 21BS017); China

University Industry University Research Innovation Fund (Grant No.: 2021ITA05002 and 2021ITA05004).

REFERENCES

[1] Land, E.H., McCann, J.J. (1971). Lightness and retinex theory. Journal of the Optical Society of America, 61(1): 1-11. https://doi.org/10.1364/JOSA.61.000001

[2] Marr, D. (1974). The computation of lightness by the primate retina. Vision Research, 14(12): 1377-1388. https://doi.org/10.1016/0042-6989(74)90012-1

[3] Barrow, H., Tenenbaum, J., Hanson, A., Riseman, E. (1978). Recovering intrinsic scene characteristics. Computer Vision Systems.

[4] Baslamisli, A.S., Groenestege, T.T., Das, P., Le, H.A., Karaoglu, S., Gevers, T. (2018). Joint learning of intrinsic images and semantic segmentation. In 15th European Conference, Munich, Germany, pp. 286-302. https://doi.org/10.1007/978-3-030-01231-1_18

[5] Li, L., Xia, Z., Jiang, X., Ma, Y., Roli, F., Feng, X. (2020). 3D face mask presentation attack detection based on intrinsic image analysis. Iet Biometrics, 9(3): 100-108. https://doi.org/10.1049/iet-bmt.2019.0155

[6] Liang, L., Jin, L., Xu, Y. (2020). PDE learning of filtering and propagation for task-aware facial intrinsic image analysis. IEEE Transactions on Cybernetics, 52(2): 1021-1034. https://doi.org/10.1109/TCYB.2020.2989610

[7] Shekhar, S., Reimann, M., Mayer, M., Semmo, A., Pasewaldt, S., Döllner, J., Trapp, M. (2021). Interactive photo editing on smartphones via intrinsic decomposition. Computer Graphics Forum, 40(2): 497-510. https://doi.org/10.1111/cgf.142650

[8] Barron, J.T., Malik, J. (2011). High-frequency shape and albedo from shading using natural image statistics. In CVPR 2011, Colorado Springs, CO, USA, pp. 2521-2528. https://doi.org/10.1109/CVPR.2011.5995392

[9] Duchêne, S., Riant, C., Chaurasia, G., et al. (2015). Multi-view intrinsic images of outdoors scenes with an application to relighting. ACM Transactions on Graphics, 34(5): 164. https://doi.org/10.1145/2756549

[10] Wu, S., Rupprecht, C., Vedaldi, A. (2023). Unsupervised learning of probably symmetric deformable 3D objects from images in the wild. IEEE Transactions on Pattern Analysis and Machine Intelligence, 45(4): 5268-5281. https://doi.org/10.1109/TPAMI.2021.3076536

[11] Ma, Y., Feng, X., Jiang, X., Xia, Z., Peng, J. (2017). Intrinsic image decomposition: A comprehensive review. In Image and Graphics: 9th International Conference, ICIG 2017, Shanghai, China, pp. 626-638. https://doi.org/10.1007/978-3-319-71607-7_55

[12] Horn, B.K., Brooks, M.J. (1986). The variational approach to shape from shading. Computer Vision, Graphics, and Image Processing, 33(2): 174-208. https://doi.org/10.1016/0734-189X(86)90114-3

[13] Bousseau, A., Paris, S., Durand, F. (2009). User-assisted intrinsic images. In ACM SIGGRAPH Asia 2009 papers, No. 130. https://doi.org/10.1145/1661412.1618476

[14] Zhang L. M., Zhang D., Wang, Z.Q., Cong Y., Wang X.S., (2023). Constructing a three-dimensional creep model for rocks and soils based on memory-dependent derivatives: A theoretical and experimental study. Computers and Geotechnics, 159: 105366. https://doi.org/10.1016/j.compgeo.2023.105366

[15] Tappen, M., Freeman, W., Adelson, E. (2002). Recovering intrinsic images from a single image. Advances in Neural Information Processing Systems.

[16] Song, Z., Ma, Y., Tan, F., Feng, X. (2022). Hybrid dilated and recursive recurrent convolution network for time-domain speech enhancement. Applied Sciences, 12(7): 3461. https://doi.org/10.3390/app12073461

[17] Huang, D., Xia, Z., Li, L., Ma, Y. (2023). Pain estimation with integrating global-wise and region-wise convolutional networks. IET Image Processing, 17(3): 637-648. https://doi.org/10.1049/ipr2.12639

[18] Zhang, Q., Zhou, J., Zhu, L., Sun, W., Xiao, C., Zheng, W. S. (2021). Unsupervised intrinsic image decomposition using internal self-similarity cues. IEEE Transactions on Pattern Analysis and Machine Intelligence, 44(12): 9669-9686. https://doi.org/10.1109/TPAMI.2021.3129795

[19] Sato, S., Yao, Y., Yoshida, T., Kaneko, T., Ando, S., Shimamura, J. (2023). Unsupervised intrinsic image decomposition with LIDAR intensity. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, Vancouver, BC, Canada, pp. 13466-13475. https://doi.org/10.1109/CVPR52729.2023.01294

[20] Shen, L., Yeo, C. (2011). Intrinsic images decomposition using a local and global sparse representation of reflectance. In CVPR 2011, Colorado Springs, CO, USA, pp. 697-704. https://doi.org/10.1109/CVPR.2011.5995738

[21] Luo, J., Huang, Z., Li, Y., Zhou, X., Zhang, G., Bao, H. (2020). NIID-Net: Adapting surface normal knowledge for intrinsic image decomposition in indoor scenes. IEEE Transactions on Visualization and Computer Graphics, 26(12): 3434-3445. https://doi.org/10.1109/TVCG.2020.3023565

[22] Ma, Y., Jiang, X., Xia, Z., Gabbouj, M., Feng, X. (2020). Casqnet: Intrinsic image decomposition based on cascaded quotient network. IEEE Transactions on Circuits and Systems for Video Technology, 31(7): 2661-2674. https://doi.org/10.1109/TCSVT.2020.3024687

[23] Zhang L. M., Chao W.W., Liu Z.Y., Cong Y., Wang, Z.Q. (2022). Crack propagation characteristics during progressive failure of circular tunnels and the early warning thereof based on multi-sensor data fusion. Geomechanics and Geophysics for Geo-Energy and Geo-Resources, 8: 172. https://doi.org/10.1007/s40948-022-00482-3

[24] Zhang, F., You, S., Li, Y., Fu, Y. (2022). HSI-Guided Intrinsic Image Decomposition for Outdoor Scenes. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, New Orleans, LA, USA, pp. 313-322. https://doi.org/10.1109/CVPRW56347.2022.00046

[25] Vaswani, A., Shazeer, N., Parmar, N., et al. (2017). Advances in Neural Information Processing Systems 30 (NIPS 2017).

[26] Yan, F., Yan, B., Pei, M. (2023). Dual transformer encoder model for medical image classification. In 2023 IEEE International Conference on Image Processing, Kuala Lumpur, Malaysia, pp. 690-694. https://doi.org/10.1109/ICIP49359.2023.10222303

[27] Yang, H., Yu, H., Hong, D., Xu, Z., Wang, Y., Song, M.

(2022). Hyperspectral image classification based on multi-level spectral-spatial transformer network. In 2022 12th Workshop on Hyperspectral Imaging and Signal Processing: Evolution in Remote Sensing, Rome, Italy, pp. 1-4. https://doi.org/10.1109/WHISPERS56178.2022.9955116

[28] Li, S. (2022). Trans-CycleGAN: Image-to-Image Style Transfer with Transformer-based Unsupervised GAN. In 2022 3rd International Conference on Computer Vision, Image and Deep Learning & International Conference on Computer Engineering and Applications, Changchun, China, pp. 1-4. https://doi.org/10.1109/CVIDLICCEA56201.2022.9824311

[29] Masood, A., Naseem, U., Razzak, I. (2023). Multi-scale swin transformer enabled automatic detection and segmentation of lung metastases using CT images. In 2023 IEEE 20th International Symposium on Biomedical Imaging (ISBI), Cartagena, Colombia, pp. 1-5. https://doi.org/10.1109/ISBI53787.2023.10230663

[30] Zhang, Y., Jiang, X., Liu, S., Hu, B., Gao, X. (2022). Boundary-aware bias loss for transformer-based aerial image segmentation model. In ICASSP 2022 - 2022 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), Singapore, pp. 3528-3532. https://doi.org/10.1109/ICASSP43922.2022.9747074

[31] Burt, P.J., Adelson, E.H. (1987). The Laplacian pyramid as a compact image code. In Readings in Computer Vision, pp. 671-679. https://doi.org/10.1016/B978-0-08-051581-6.50065-9

[32] Burt, P.J., Adelson, E.H. (1983). A multiresolution spline with application to image mosaics. ACM Transactions on Graphics (TOG), 2(4): 217-236. https://doi.org/10.1145/245.247

[33] Jacobs, C.E., Finkelstein, A., Salesin, D.H. (1995). Fast multiresolution image querying. In Proceedings of the 22nd Annual Conference on Computer Graphics and Interactive Techniques, Los Angeles, CA, USA, pp. 277-286. https://doi.org/10.1145/218380.218454

[34] Do, M.N., Vetterli, M. (2005). The contourlet transform: An efficient directional multiresolution image representation. IEEE Transactions on Image Processing, 14(12): 2091-2106. https://doi.org/10.1109/TIP.2005.859376

[35] Song, M., Lim, S., Kim, W. (2021). Monocular depth estimation using laplacian pyramid-based depth residuals. IEEE Transactions on Circuits and Systems for Video Technology, 31(11): 4381-4393. https://doi.org/10.1109/TCSVT.2021.3049869

[36] Barron, J.T., Malik, J. (2014). Shape, illumination, and reflectance from shading. IEEE Transactions on Pattern Analysis and Machine Intelligence, 37(8): 1670-1687. https://doi.org/10.1109/TPAMI.2014.2377712

[37] Chang, A.X., Funkhouser, T., Guibas, L., et al. (2015). Shapenet: An information-rich 3D model repository. arXiv preprint arXiv:1512.03012. https://doi.org/10.48550/arXiv.1512.03012

[38] Grosse, R., Johnson, M.K., Adelson, E.H., Freeman, W.T. (2009). Ground truth dataset and baseline evaluations for intrinsic image algorithms. In 2009 IEEE 12th International Conference on Computer Vision, Kyoto, Japan, pp. 2335-2342. https://doi.org/10.1109/ICCV.2009.5459428

[39] Clevert, D.A., Unterthiner, T., Hochreiter, S. (2015). Fast and accurate deep network learning by exponential linear units (ELUs). arXiv preprint arXiv:1511.07289. https://doi.org/10.48550/arXiv.1511.07289

[40] Bell, S., Bala, K., Snavely, N. (2014). Intrinsic images in the wild. ACM Transactions on Graphics (TOG), 33(4): 159. https://doi.org/10.1145/2601097.2601206

[41] Narihira, T., Maire, M., Yu, S.X. (2015). Direct intrinsics: Learning albedo-shading decomposition by convolutional regression. In 2015 IEEE International Conference on Computer Vision, Santiago, Chile, pp. 2992-2992. https://doi.org/10.1109/ICCV.2015.342

[42] Han, G., Xie, X., Lai, J., Zheng, W.S. (2018). Learning an intrinsic image decomposer using synthesized RGB-D dataset. IEEE Signal Processing Letters, 25(6): 753-757. https://doi.org/10.1109/LSP.2018.2820041

[43] Shi, J., Dong, Y., Su, H., Yu, S.X. (2017). Learning non-lambertian object intrinsics across shapenet categories. In 2017 IEEE Conference on Computer Vision and Pattern Recognition, Honolulu, HI, USA, pp. 1685-1694. https://doi.org/10.1109/CVPR.2017.619

[44] Zhang, F., Jiang, X., Xia, Z., Gabbouj, M., Peng, J., Feng, X. (2022). Non-local color compensation network for intrinsic image decomposition. IEEE Transactions on Circuits and Systems for Video Technology, 33(1): 132-145. https://doi.org/10.1109/TCSVT.2022.3199428

[45] Ranjit, S.S., Jaiswal, R.K. (2020). Single image intrinsic decomposition using transfer learning. In Proceedings of the 2020 12th International Conference on Machine Learning and Computing, Shenzhen, China, pp. 418-425. https://doi.org/10.1145/3383972.3384062

[46] Careaga, C., Aksoy, Y. (2023). Intrinsic image decomposition via ordinal shading. ACM Transactions on Graphics, 43(1): 12. https://doi.org/10.1145/3630750

[47] Ma, Y., Feng, X., Chao, L., Huang, D., Xia, Z., Jiang, X. (2018). A new database for evaluating underwater image processing methods. In 2018 Eighth International Conference on Image Processing Theory, Tools and Applications, Xi'an, China, pp. 1-6. https://doi.org/10.1109/IPTA.2018.8608131