

TOWARDS LEARNING LOW-LIGHT INDOOR SEMANTIC SEGMENTATION WITH ILLUMINATION-INVARIANT FEATURES

N. Zhang^{1,*}, F. Nex¹, N. Kerle¹, G. Vosselman¹

¹ Faculty of Geo-Information Science and Earth Observation (ITC), University of Twente, Enschede, the Netherlands
(n.zhang, f.nex, n.kerle, george.vosselman)@utwente.nl

Commission II, WG II/5

KEY WORDS: Semantic Segmentation, Dataset, Low-light, Image Decomposition, Deep Learning, Scene Understanding.

ABSTRACT:

Semantic segmentation models are often affected by illumination changes, and fail to predict correct labels. Although there has been a lot of research on indoor semantic segmentation, it has not been studied in low-light environments. In this paper we propose a new framework, LISU, for Low-light Indoor Scene Understanding. We first decompose the low-light images into reflectance and illumination components, and then jointly learn reflectance restoration and semantic segmentation. To train and evaluate the proposed framework, we propose a new data set, namely LLRGBD, which consists of a large synthetic low-light indoor data set (LLRGBD-synthetic) and a small real data set (LLRGBD-real). The experimental results show that the illumination-invariant features effectively improve the performance of semantic segmentation. Compared with the baseline model, the mIoU of the proposed LISU framework has increased by 11.5%. In addition, pre-training on our synthetic data set increases the mIoU by 7.2%. Our data sets and models are available on our project website.

1. INTRODUCTION

Indoor semantic segmentation is a fundamental computer vision task, which assigns a semantic label to each pixel in an indoor scene image. In recent years, convolutional neural networks (CNNs) have made remarkable achievements in indoor semantic segmentation. However, most of the research focuses on the segmentation of normal-light scenes, while scene understanding in low-light indoor scenes is practical but has not been investigated much. When robots or first responders perform search tasks and use a light source to illuminate a dark room, the lack of illumination will make the colors and textures of the same objects look different when views change. The illumination variances reduce the robustness of CNN-based methods and yield inaccurate segmentation. It is nearly impossible to build a sufficiently large data set to cover all illumination settings and train a robust CNN on it to learn complete representations of illumination changes. To overcome the negative influence of illumination changes on semantic segmentation, some research work in the field of autonomous driving pre-processed RGB images and transform them into illumination-invariant images based on spectral response (Alshammari et al., 2018, Mader et al., 2014, Upcroft et al., 2014). However, these transformation methods are sensitive to the saturation of images, so they are not always effective (Upcroft et al., 2014).

This paper attempts to fill the gap of indoor semantic segmentation under low-light, and explores to take advantage of illumination-invariant features to improve the segmentation accuracy. Inspired by some intrinsic image decomposition methods based on Retinex theory (Land and McCann, 1971), an image I can be factorized as the product of a reflectance map $R(I)$ and an illumination map $S(I)$:

$$I = R(I) \cdot S(I). \quad (1)$$

* Corresponding author

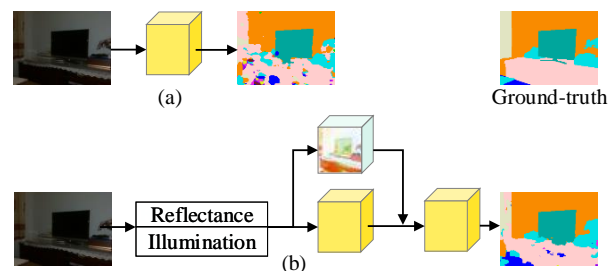


Figure 1. Compared with the direct segmentation of low-light images (a), our approach (b) seeks clues from illumination-invariant features.

The reflectance component is beneficial to segmentation task because it represents the intrinsic property of a scene and shows the original colors of objects that are not affected by illumination. As shown in Figure 1, we propose a novel framework dubbed LISU, which segments low-light indoor scenes by embedding illumination-invariant features into the segmentation branch. We decompose paired low/normal-light images of the same scene into reflectance and illumination components in an unsupervised manner. Then, we feed the decomposed components of the low-light image into a simple encoder-decoder network. This network has two separate decoders, learning reflectance restoration and semantic segmentation, respectively. The features from each task's decoder are linked to the decoder of the other task, so as to enable tighter joint learning of two tasks. Our contributions can be summarized as follows:

- We propose a novel framework which exploits the illumination-invariant features for robust low-light indoor semantic segmentation. As far as we know, we are the first to propose an end-to-end (without any pre/post-processing) trainable framework for semantic segmentation in low-light indoor scenes.

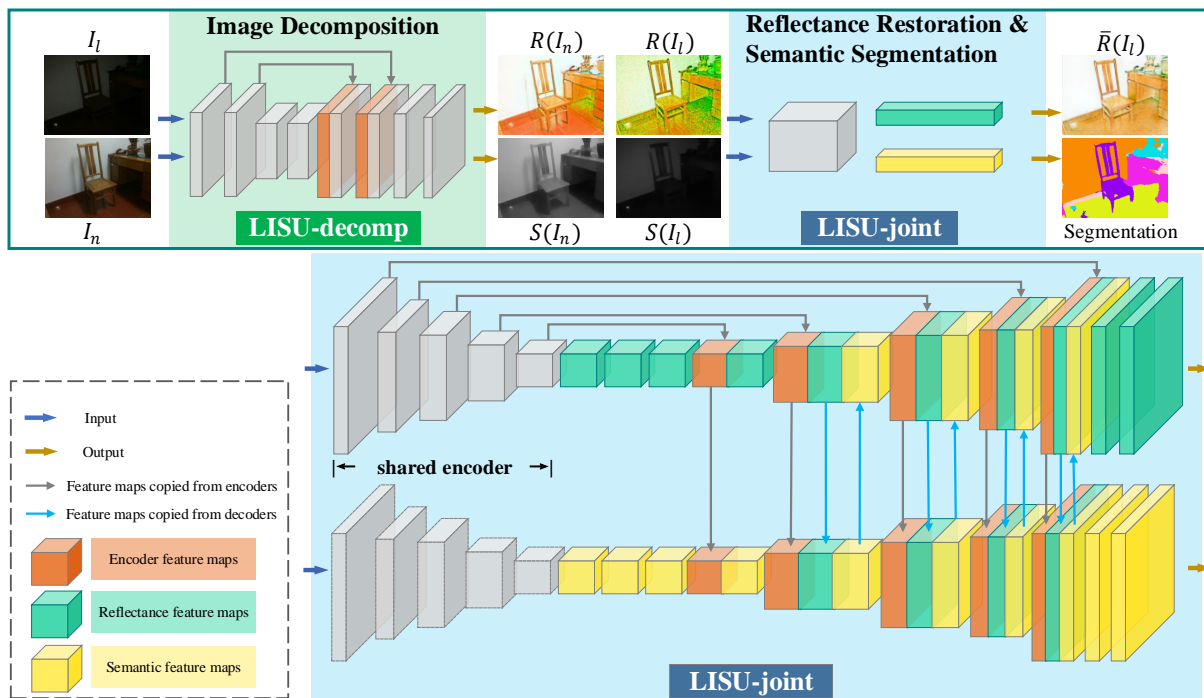


Figure 2. The cascade architecture of our LISU network, which consists of LISU-decomp and LISU-joint. LISU-decomp learns the decomposition of each input image, and LISU-joint jointly learns the reflectance restoration and semantic segmentation.

- Since there is no available data set for low-light indoor scene understanding, we collect and annotate a large synthetic data set and a small real data set to train and evaluate our method. For each scene in the data sets, two different illumination settings are deployed. Besides, the corresponding segmentation maps and depth maps are provided.
- The experimental results show the effectiveness of introducing illumination-invariant features to the low-light semantic segmentation task. Our data sets and models are available on our project website¹.

The remainder of the article is organized as follows. Section 2 introduces the architecture and loss functions of our LISU framework. Section 3 presents the details of the data sets. Section 4 elaborates on the experimental results and discussion. Section 5 concludes the paper.

2. LISU: A FRAMEWORK FOR LOW-LIGHT INDOOR SCENE UNDERSTANDING

The upper part of Figure 2 shows our LISU framework for Low-light Indoor Scene Understanding. It consists of two sub-networks: LISU-decomp and LISU-joint. The former is responsible for decomposing an image into reflectance and illumination components. The second sub-network receives the reflectance map and illumination map of the low-light image output by LISU-decomp and performs joint learning of reflectance restoration and semantic segmentation. Our framework needs paired low/normal-lights for training, while only low-light images are needed at the inference stage. Next, we will introduce the detailed structures and loss functions of the two sub-networks.

¹ <https://noahzn.netlify.app/project/low-light-indoor-scene-understanding/>

2.1 LISU-decomp: Intrinsic Image Decomposition

Our decomposition network is shown in the upper left (green box) of Figure 2. The long skip connection (gray arrows) concatenates the features from the encoder to corresponding decoder layers and enables the network to preserve low-level information and generate sharper result. Similar network structures have been used for intrinsic image decomposition (Dai et al., 2016, Rematas et al., 2016) and semantic segmentation (Ronneberger et al., 2015, Wu et al., 2018). Since the ground-truth of the reflectance map and illumination map of real images are not available, we adopt the unsupervised method proposed in (Zhang et al., 2019) and suppose that we have paired images of the same scene taken under low-light and normal-light conditions $[I_l, I_n]$. We use the same network to decompose these two images into a reflectance map and an illumination map, namely $[R(I_l), S(I_l)]$ and $[R(I_n), S(I_n)]$, respectively. According to Eq. 1, the first part of our reconstruction loss can be defined as:

$$L_{recon1} = \|I_l - R(I_l) \cdot S(I_l)\|_1 + \|I_n - R(I_n) \cdot S(I_n)\|_1, \quad (2)$$

where $\|\cdot\|_1$ denotes the l^1 norm. Ideally, the reflectance maps of two images should be equal. Therefore, we can reconstruct the low-light image using the reflectance map of the normal-light image and the illumination map of the low-light image, and vice versa. We construct another part of reconstruction loss as follows:

$$L_{recon2} = \|I_l - R(I_n) \cdot S(I_l)\|_1 + \|I_n - R(I_l) \cdot S(I_n)\|_1. \quad (3)$$

Following Retinex-based methods (Guo et al., 2016, Handa et al., 2016, Land and McCann, 1971), we use the maximum values of RGB channels of input as an initial estimation of the

illumination map:

$$L_{init} = \|S(I_l) - \max_{c \in \{R, G, B\}} I_l^c\|_1 + \|S(I_n) - \max_{c \in \{R, G, B\}} I_n^c\|_1. \quad (4)$$

We also use a loss to constrain the smoothness of the illumination map. Our structure-aware smoothness loss is defined as:

$$L_{smooth} = \|\nabla S(I_l) \cdot \exp(-\lambda_g \nabla I_l)\|_1 + \|\nabla S(I_n) \cdot \exp(-\lambda_g \nabla I_n)\|_1, \quad (5)$$

where ∇ represents for the first-order derivative in both horizontal and vertical directions. λ_g is a weight term to ensure piece-wise smooth, and it is set to 10 as in (Wei et al., 2018). Unlike (Wei et al., 2018) that used reflectance maps to weight the function, we look for clues from the original low-light images to weight the loss function. The reason is that the quality of the reflectance map generated at this stage is noisy and not reliable to guide the decomposition of illumination map. Our final loss function for the decomposition network is:

$$L_{decomp} = L_{recon1} + \lambda_1 L_{recon2} + \lambda_2 L_{init} + \lambda_3 L_{smooth}, \quad (6)$$

where λ_1 , λ_2 and λ_3 are weight factors.

2.2 LISU-joint: joint learning of reflectance restoration and semantic segmentation

We tend to use the output of LISU-decomp to train our segmentation network as the reflectance maps are not affected by illumination. However, the reflectance maps obtained by decomposing low-light images have serious degradation, and they can be further restored using the corresponding reflectance maps of normal-light images (Zhang et al., 2019). Therefore, we extend the single task segmentation network to jointly learn reflectance restoration.

The proposed joint learning network LISU-joint shown in the lower part (blue box) of Figure 2 is also a U-shaped network similar to our decomposition network, but with deeper convolutional layers for the encoder. The five-layer encoder takes the reflectance map and illumination map of a low-light image output by LISU-decomp as input and learns shared features. Then, two distinct decoders learn reflectance restoration and semantic segmentation, respectively. We enhance the correlation between two tasks by linking the features (light blue arrows) from two decoders together. As the training goes on, the semantic segmentation task benefits from the gradually restored illumination-invariant features. At the same time, the segmentation branch also provides semantic information to the restoration branch, and promotes it to produce better restoration at boundaries. Following (Zhang et al., 2019) the loss function for reflectance restoration is defined as:

$$L_{restore} = \|\bar{R}(I_l) - R(I_n)\|_2^2 - SSIM(R(I_l), R(I_n)) + \|\nabla \bar{R}(I_l) - \nabla R(I_n)\|_2^2, \quad (7)$$

where $\bar{R}(I_l)$ is the restored reflectance map, and $\|\cdot\|_2^2$ denotes the l^2 reconstruction loss (MSE). $SSIM(\cdot, \cdot)$ measures the structural similarity (Wang et al., 2004) of two reflectance maps. The last term makes the restored reflectance map have textures close to the reference. For the semantic segmentation task, we use cross-entropy as loss function:

$$L_{ce} = -\frac{1}{n} \sum_i \sum_{c \in M} \log(p_i^c), \quad (8)$$

where p_i^c denotes the probability of a pixel i being predicted as class c . M is the semantic category. The combined objective function for LISU-joint is:

$$L_{joint} = L_{restore} + L_{ce}. \quad (9)$$

3. NEW DATA SETS OF LOW-LIGHT INDOOR SCENES

In this section we introduce the synthetic and real data sets, namely LLRGBD-synthetic and LLRGBD-real, which are used for understanding low-light indoor scenes. Figure 3 shows some sample images in our data sets.

3.1 LLRGBD-synthetic

We render realistic image pairs using a modified version of Opposite Renderer (McCormac et al., 2017, Pedersen, 2013), which is an open-source renderer based on the Nvidia OptiX ray tracing engine. For each rendering, the engine selects one layout from a total of 57 layouts, and randomly places relevant 3D objects from ShapeNet repository (Chang et al., 2015) into the scene. Then the camera moves in the scene to generate random collision-free trajectories. We maintain a high-quality texture library to realize photorealistic rendering of scenes. In order to simulate an LED light source in the dark, we modify the source code of the renderer and installed a white light source on the camera. The engine uses z-buffer of OpenGL to render the depth map, and we get the 13-class (Couprie et al., 2013) labels by simple object-class matching process.

We render $29K \times 2$ images at 640×480 resolution, and each rendering takes 2-3 minutes on a Nvidia Titan XP GPU. This data set is divided into training set and test set according to the ratio of 90%-10%.

3.2 LLRGBD-real

This data set consists of 515 pairs of real low-light and normal-light images of indoor scenes, including offices, bedrooms, bathrooms, kitchens, and living rooms. These image pairs are taken with a fixed Intel RealSense D435i camera, which is an advanced RGB-D camera usually used as SLAM sensor. Unlike those methods that collected image pairs by changing shutter speed and ISO (Chen et al., 2018a, Guo et al., 2016, Wei et al., 2018), when capturing low-light images, we ensure that the scenes are completely dark and we only use one point light as illumination. The color temperature of this light is $5500 \pm 200K$ and the illuminance is about 800 lm. The purpose of using a light instead of controlling the camera parameters is that we aim to focus on real exploration missions with a light as illumination in dark indoor environments. Then the white lights in the room are turned on and we take these pictures as normal-light images. We capture the images at 640×480 resolution, and we use 13 categories defined in (Couprie et al., 2013) to label our data set. This data set is divided into 415 image pairs for training and 100 for testing. Corresponding depth map of each scene is collected by the infrared sensor of RealSense D435i and aligned to the RGB images. We further refine these depth maps using the colorization method proposed in (Levin et al.,

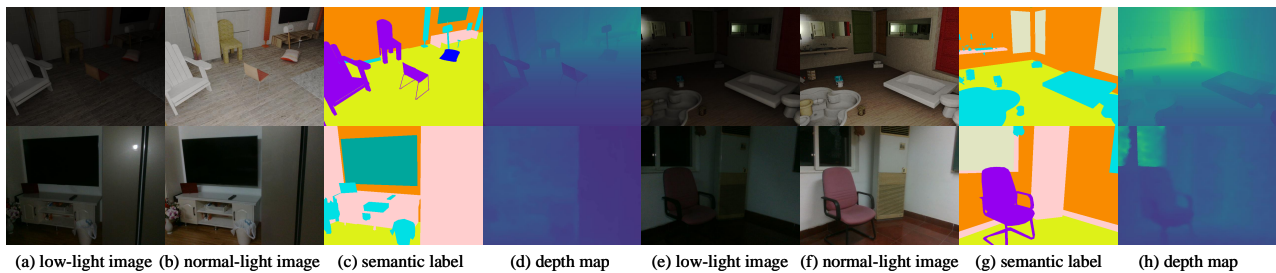


Figure 3. Sample images from the proposed LLRGBD data sets. The images in the first row are synthetic, and the second row shows some real images.

2004). Although depth maps are not used in this paper, we still release them as a part of our data set. Researchers can use this data set to study other scene understanding tasks, such as depth estimation of low-light indoor scenes.

4. EXPERIMENTS AND EVALUATION

4.1 Metrics

We evaluate segmentation with three metrics: 1) **pixel accuracy (PixAcc.)**, which calculates the percentage of correctly classified pixels; 2) **mean accuracy (mAcc.)**, which takes the average of all the pixel accuracy over all the classes; 3) **mean Intersection over Union (mIoU)**. IoU is the intersection between the classified pixels and the true labels divided by the union between the classified pixels and the true labels. mIoU takes the average of all the IoUs over all classes.

4.2 Implementation details

We use PyTorch to implement our network, and train the models on a Nvidia Titan XP GPU with batch size 12. We use Adam solver with $(\beta_1, \beta_2) = (0.95, 0.999)$ as the optimizer. The initial learning rate is 0.001 and it is scheduled with polynomial decay with power $p = 0.9$. When training on LLRGBD-synthetic, the total number of epochs is set to 50, while when training on LLRGBD-real, it is set to 200. All the training images and labels are down-sampled to 320×240 , and no data augmentation is applied. The influence of the weights on the results is studied in the **Experiment II** of section 4.5, and we finally set $\lambda_1 = 0.01$, $\lambda_2 = 0.1$ and $\lambda_3 = 0.5$ because this combination of weights achieves the best segmentation accuracy.

4.3 Evaluation of the baseline model for low-light images segmentation.

We first evaluate the segmentation branch of our LISU-joint, which serves as the baseline model in this paper. It is the bottom part of LISU-joint shown in Figure 2, and contains only the gray encoder features, the yellow decoder features and the red features from the encoder. We call the baseline model **LISU-seg**. In this experiment we directly use the original low-light images to train the segmentation models. Table 1 shows the results of our baseline model LISU-seg, and the other two encoder-decoder models, SegNet (Badrinarayanan et al., 2017) and U-Net (Ronneberger et al., 2015). We report the results trained on LLRGBD-synthetic and LLRGBD-real respectively, and our LISU-seg performs best on both data sets. SegNet does not perform well because the pixel values of low-light images

are very small, and the use of max pooling loses a lot of information. Although U-Net also uses max pooling layers to reduce the sizes of feature maps, the U-shaped structure that uses long-skip connections helps to maintain shallow convolutional features. LISU-seg has a structure similar to U-Net, but it has less encoder features and we avoid to use max pooling.

Data set	Method	PixAcc.	mAcc.	mIoU
LLRGBD-synthetic	SegNet	73.4	33.0	25.6
	U-Net	77.8	43.1	33.2
	LISU-seg	82.3	49.0	39.5
LLRGBD-real	SegNet	42.8	34.2	22.4
	U-Net	54.6	47.4	32.8
	LISU-seg	59.0	50.0	36.1

Table 1. Comparison of accuracy of direct segmentation of low-light images.

4.4 Evaluation of segmentation using degraded reflectance maps

In this experiment we explore if segmenting the reflectance maps of low-light images can achieve better results than using original low-light images. However, the ground-truth of reflectance is not available. Therefore, we first train our decomposition network LISU-decomp with paired low/normal-light images, and then feed the output three-channel reflectance map of the low-light images to the segmentation networks. We still train and evaluate three segmentation networks on the proposed data sets. Table 2 shows the results and we can find that the accuracy of all three segmentation models on both synthetic and real data sets has increased. It is worth noting that the reflectance maps used in the training come from our decomposition network, and these reflectance maps degrade very much. But they are still helpful to semantic segmentation.

Data set	Method	PixAcc.	mAcc.	mIoU
LLRGBD-synthetic	SegNet	80.3	45.1	35.9
	U-Net	78.7	45.5	35.6
	LISU-seg	82.5	49.9	40.3
LLRGBD-real	SegNet	48.2	36.8	26.2
	U-Net	59.3	52.1	38.1
	LISU-seg	60.1	52.6	38.9

Table 2. Comparison of segmentation accuracy of degraded reflectance maps.

4.5 Evaluation of LISU for low-light images segmentation

Experiment I: In this experiment we evaluate the segmentation accuracy of LISU, which contains the joint learning network LISU-joint. The results in Table 3 show that the proposed joint learning network outperforms the baseline model

on two data sets. Especially on LLRGBD-real data set, there is a great improvement (for mIoU 47.6 versus 36.1). The training of segmentation task on the small data set derives benefit from the illumination-invariant features. The effectiveness of LISU is also reflected in LLRGBD-synthetic (for mIoU 43.4 versus 39.5). Since this synthetic data set contains sequential images, the network has a better chance to learn the representations of same objects under illumination variations. The strategy of joint learning further improves the learning ability of the network.

Data set	Method	PixAcc.	mAcc.	mIoU
LLRGBD-synthetic	LISU-seg	82.3	49.0	39.5
	LISU	84.5	52.3	43.4
LLRGBD-real	LISU-seg	59.0	50.0	36.1
	LISU	67.3	61.2	47.6

Table 3. Evaluation of LISU. We copy the results of the baseline model LISU-seg from Table 1 for clear comparison.

Experiment II: In our method, LISU-decomp is very important because it provides the input of LISU-joint. In this experiment we evaluate the influence of weights in Eq. 6 on segmentation accuracy. The results in Table 4 show that when we fix λ_2 and λ_3 , a larger λ_1 generates worse segmentation. The first combination of weights is chosen ($\lambda_1 = 0.01, \lambda_2 = 0.1, \lambda_3 = 0.5$) because it has the highest mIoU and mean accuracy.

No.	Weights of Eq. 6			LISU		
	λ_1	λ_2	λ_3	PixAcc.	mAcc.	mIoU
1	0.01	0.1	0.5	<u>67.3</u>	61.2	47.6
2	0.1	0.1	0.5	65.7	58.9	45.8
3	0.5	0.1	0.5	64.1	55.8	42.3
4	0.01	0.5	0.5	65.0	57.9	44.6
5	0.01	0.01	0.5	68.7	<u>60.0</u>	<u>47.5</u>
6	0.01	0.01	0.1	67.1	59.7	47.1
7	0.01	0.01	0.01	65.2	55.9	43.1

Table 4. Study of the influence of weights of Eq. 6 on segmentation. The best and the second best results are highlighted in **bold** and underlined.

4.6 Evaluation of the effectiveness of pre-training on synthetic data

In this experiment we explore if a model pre-trained on our synthetic data set can further improve the segmentation performance on small-scale real data. We first train LISU for 50 epochs on LLRGBD-synthetic, and then fine-tune the model on LLRGBD-real by freezing the encoders of LISU-decomp and LISU-joint. Compared with LISU without pre-training, the mIoU shown in Table 5 increases by 7.2% (54.8 versus 47.6).

Data set	Method	PixAcc.	mAcc.	mIoU
LLRGBD-real	LISU (pre-trained)	72.3	68.2	54.8

Table 5. Evaluation of the effectiveness of pre-training on LLRGBD-synthetic.

4.7 Segmentation with modified DeepLab v3+

Since our LISU-joint uses simple encoder-decoder structure as segmentation network, in principle it can be replaced by any network with a similar structure. In order to further verify the superiority of joint learning of semantic segmentation and reflectance restoration in low-light indoor scene segmentation,

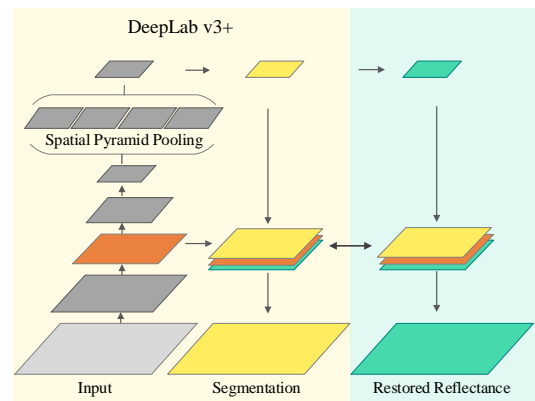


Figure 4. A modified DeepLab v3+ that jointly learns to restore reflectance.

Method	PixAcc.	mAcc.	mIoU
DLv3p (Chen et al., 2018b)	54.9	46.9	33.4
DLv3p-joint	68.4	62.0	49.2

Table 6. Comparison of segmentation accuracy using DeepLab v3+ and its variants.

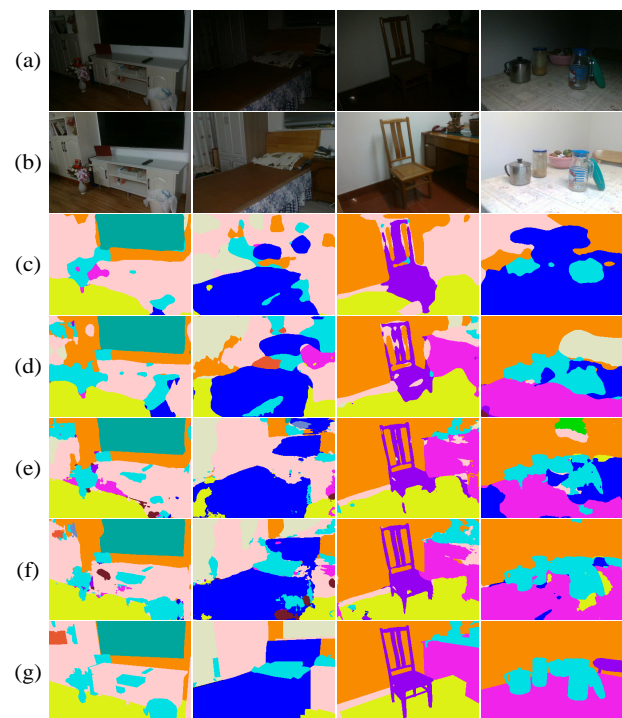


Figure 5. Qualitative results on LLRGBD-real. (a) is the input low-light images, and (b) shows the corresponding normal-light images; (c): DLv3p (Chen et al., 2018b); (d): LISU-seg; (e) LISU; (f) LISU (pre-trained on LLRGBD-synthetic); (g) Ground-truth.

we evaluate our approach with the state-of-the-art DeepLab v3+ (Chen et al., 2018b) (DLv3p) as shown in Figure 4. The structure in the yellow box is the original DLv3p (without green feature maps), and it consists of an encoder (ResNet50 backbone (He et al., 2016)) and a decoder. We modify its original structure by adding an extra decoder to restore the reflectance

(DLv3p-joint), and we remove the dropout layers for this decoder. The features from two decoders are concatenated together (yellow and green feature maps). We use original low-light images from LLRGBD-real to train DLv3p. And for the modified structure, the input comes from the output of LISU-decomp. We encountered overfitting problems when directly training DLv3p on LLRGBD-real. Thus, we adopted early stopping, and the results are shown in the first row of Table 6. The introduction of joint learning makes the training more stable. Note that only one layer in the decoders participates in feature exchange, which once again proves the effectiveness of our approach. We show qualitative results in Figure 5. Although we do not need normal-light images when inferring the model, we still show them here because some low-light images are too dark to be seen clearly.

5. CONCLUSION

In this paper, we present a novel end-to-end trainable CNN framework that takes advantage of the illumination-invariant features for low-light indoor scene segmentation. We also present a new data set for understanding low-light indoor scenes. The experimental results on both synthetic and real data sets show the effectiveness of our approach.

ACKNOWLEDGEMENTS

This project has received funding from the European Union's Horizon 2020 Research and Innovation Programme and the Korean Government under Grant Agreement No 833435. Content reflects only the authors' view and European Commission is not responsible for any use that may be made of the information it contains.

REFERENCES

- Alshammari, N., Akcay, S., Breckon, T. P., 2018. On the impact of illumination-invariant image pre-transformation for contemporary automotive semantic scene understanding. *Intelligent Vehicles Symposium*.
- Badrinarayanan, V., Kendall, A., Cipolla, R., 2017. Segnet: A deep convolutional encoder-decoder architecture for image segmentation. *IEEE TPAMI*, 39(12), 2481–2495.
- Chang, A. X., Funkhouser, T., Guibas, L., Hanrahan, P., Huang, Q., Li, Z., Savarese, S., Savva, M., Song, S., Su, H., Xiao, J., Yi, L., Yu, F., 2015. Shapenet: An information-rich 3d model repository. *arXiv preprint arXiv:1512.03012*.
- Chen, C., Chen, Q., Xu, J., Koltun, V., 2018a. Learning to see in the dark. *CVPR*.
- Chen, L.-C., Zhu, Y., Papandreou, G., Schroff, F., Adam, H., 2018b. Encoder-decoder with atrous separable convolution for semantic image segmentation. *ECCV*.
- Coupric, C., Farabet, C., Najman, L., LeCun, Y., 2013. Indoor semantic segmentation using depth information. *ICLR*.
- Dai, J., He, K., Sun, J., 2016. Instance-aware semantic segmentation via multi-task network cascades. *CVPR*.
- Guo, X., Li, Y., Ling, H., 2016. LIME: Low-light image enhancement via illumination map estimation. *IEEE TIP*, 26(2), 982–993.

- Handa, A., Patraucean, V., Badrinarayanan, V., Stent, S., Cipolla, R., 2016. Understanding real world indoor scenes with synthetic data. *CVPR*.
- He, K., Zhang, X., Ren, S., Sun, J., 2016. Deep residual learning for image recognition. *CVPR*.
- Land, E. H., McCann, J. J., 1971. Lightness and retinex theory. *Josa*, 61(1), 1–11.
- Levin, A., Lischinski, D., Weiss, Y., 2004. Colorization using optimization. *SIGGRAPH*.
- Maddern, W., Stewart, A., McManus, C., Upcroft, B., Churchill, W., Newman, P., 2014. Illumination invariant imaging: Applications in robust vision-based localisation, mapping and classification for autonomous vehicles. *ICRA*.
- McCormac, J., Handa, A., Leutenegger, S., Davison, A. J., 2017. Scenenet rgb-d: Can 5m synthetic images beat generic imagenet pre-training on indoor segmentation? *ICCV*.
- Pedersen, S. A., 2013. Progressive photon mapping on gpus. Master's thesis, Institutt for datateknikk og informasjonsvitenskap.
- Rematas, K., Ritschel, T., Fritz, M., Gavves, E., Tuytelaars, T., 2016. Deep reflectance maps. *CVPR*.
- Ronneberger, O., Fischer, P., Brox, T., 2015. U-net: Convolutional networks for biomedical image segmentation. *MICCAI*.
- Upcroft, B., McManus, C., Churchill, W., Maddern, W., Newman, P., 2014. Lighting invariant urban street classification. *ICRA*.
- Wang, Z., Bovik, A. C., Sheikh, H. R., Simoncelli, E. P., 2004. Image quality assessment: from error visibility to structural similarity. *IEEE TIP*, 13(4), 600–612.
- Wei, C., Wang, W., Yang, W., Liu, J., 2018. Deep retinex decomposition for low-light enhancement. *BMVC*.
- Wu, B., Wan, A., Yue, X., Keutzer, K., 2018. Squeezeseg: Convolutional neural nets with recurrent crf for real-time road-object segmentation from 3d lidar point cloud. *ICRA*.
- Zhang, Y., Zhang, J., Guo, X., 2019. Kindling the darkness: A practical low-light image enhancer. *ACM MM*.