International Journal of Engineering Technology Research & Management
(IJETRM)
https://ijetrm.com/

LEARNING SHADOW REMOVAL FROM UNPAIRED SAMPLES VIA RECIPROCAL LEARNING

P.Shanmukha Rao[1]
N.Sindhuja[2]
B. Tech Student, Dept. of Computer Science and Engineering, R.V.R & J.C College of Engineering, Chowdavaram, Guntur, Andhra Pradesh, India

Mrs. Zareena Noorbasha[3]
Assistant Professor, Dept. of Computer Science and Engineering, R.V.R & J.C College of Engineering, Chowdavaram, Guntur, Andhra Pradesh, India

ABSTRACT
This work tackles the problem of removing shadows from images using only high-level image labels, without the need for detailed pixel-level annotations. The proposed method relies on a deep learning framework that learns to distinguish and eliminate shadows based on whether an image contains shadows or not. A key feature of this approach is the interactive optimization between two components—a shadow detector and a shadow remover—that work together to gradually enhance the model's performance. Instead of depending on traditional supervised methods, a self-guided learning strategy is employed to ensure robust training without overfitting. Additionally, a specialized discriminator is introduced to help the model preserve color consistency and maintain realistic visual quality after shadow removal. Extensive evaluations on various datasets, including paired and unpaired samples, demonstrate the effectiveness and adaptability of this mutual learning-based framework.

Keywords:
Shadow removal, weakly supervised learning, reciprocal learning, shadow remover, shadow detector.

INTRODUCTION
Shadows, formed by the obstruction of light by objects, often complicate various computer vision tasks such as image segmentation, object detection, and 3D reconstruction. Effectively removing shadows can significantly enhance image understanding and downstream performance in vision-based applications. This paper presents a novel and practical learning framework designed specifically for shadow removal, with an emphasis on using unpaired image data. Traditional shadow removal methods rely heavily on convolutional neural networks (CNNs) trained on paired datasets, which require both shadow and corresponding shadow-free images. Although these models achieve high accuracy in controlled settings, they struggle with generalization due to limited scene diversity and the challenge of acquiring large-scale, well-aligned paired datasets.

Capturing high-quality paired data often involves physically removing shadow-casting objects and ensuring consistent lighting and color alignment, which is both labor-intensive and impractical in real-world settings. Recent efforts have explored training with unpaired data to reduce data collection costs and increase scene variability. However, these models often leave shadow remnants or introduce color inconsistencies due to insufficient supervision and optimization challenges.

To overcome these limitations, we propose a reciprocal learning framework that transforms the complex task of joint shadow and illumination estimation into a conditional problem. Our method introduces two interactive networks—a shadow detector that infers shadow regions and a shadow remover that learns to eliminate shadows based on the detected regions. This collaborative design allows the model to better focus on shadow-affected areas during training. To further enhance training stability and performance, we implement a self-paced learning strategy to mitigate the effects of inaccurate predictions and introduce specialized loss functions for color consistency and shadow attention. Extensive experiments on benchmark datasets such as ISTD, SRD, and USR show that our approach achieves superior performance compared to previous models trained with unpaired data, producing visually consistent and high-quality shadow-free images.

iJETRM
International Journal of Engineering Technology Research & Management
(IJETRM)
https://ijetrm.com/

OBJECTIVES

The primary objective of this study is to address the key challenges involved in shadow removal using weakly supervised learning techniques. Traditional models depend on paired datasets that include both shadow and shadow-free images, which are difficult and costly to obtain. This project aims to develop a more practical solution by utilizing unpaired image data, where each image is only labeled at a high level to indicate the presence of shadows. The focus is on designing an intelligent learning framework capable of estimating shadow regions and effectively removing them using mutual collaboration between a shadow detector and a shadow remover. Another important goal is to minimize residual shadows and color inconsistencies typically observed in previous unpaired approaches. The model is also expected to generalize across diverse scenes and lighting conditions. By introducing self-paced learning and advanced loss functions, the proposed system strives to achieve high-quality shadow removal results with limited supervision.

METHODOLOGY

This study adopts a novel weakly supervised learning approach for removing shadows from images using unpaired training samples. Traditional methods depend on pixel-level paired data (a shadow image and its exact shadow-free counterpart), which are difficult to obtain in real-world conditions. In contrast, our proposed method is capable of learning from high-level image labels alone—where the image is only annotated as having a shadow or not.

The core idea behind the method lies in modeling the shadow removal task as a conditional optimization problem using latent variables. Instead of directly mapping a shadow image to a shadow-free version (which is complex due to the absence of pixel-wise supervision), we decompose the task into two interconnected sub-networks:

1. A Shadow Detector (SD) that estimates shadow regions in an image.
2. A Shadow Remover (SR) that removes shadows based on detected regions.

These two networks are trained reciprocally in three phases, allowing them to refine each other's outputs iteratively, improving accuracy at each step.

Phase 1: Initial Shadow Remover Training

In the first phase, we train a basic shadow removal generator using a CycleGAN architecture. This generator, denoted as Gf0G_f^0Gf0, learns to generate a shadow-free version of an image without paired supervision.

The generated image is then compared to the input shadow image to extract pseudo shadow masks—regions that likely contain shadows based on pixel differences.

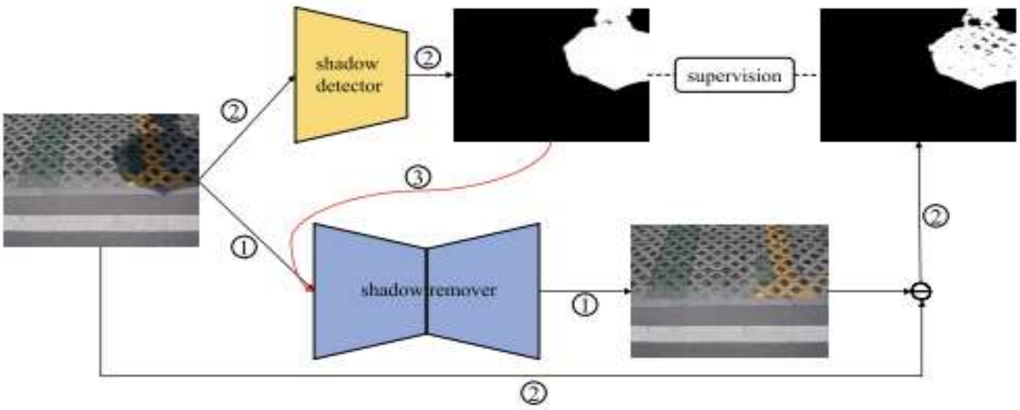


Fig. 1: Illustration of the Reciprocal Learning Process

Phase 2: Shadow Mask Estimation

These pseudo shadow masks are then used as supervision to train the Shadow Detector (SD). However, because the pseudo labels might contain noise or inaccuracies (due to lighting variation, soft shadows, or image artifacts), we employ a Self-Paced Learning (SPL) strategy.

SPL introduces a mechanism that gradually selects training samples, starting from "easy" examples (clear, well-defined shadows) and slowly introducing more complex or ambiguous ones. This reduces the risk of the detector overfitting to noisy labels.

IJETRM
International Journal of Engineering Technology Research & Management
(IJETRM)
https://ijetrm.com/

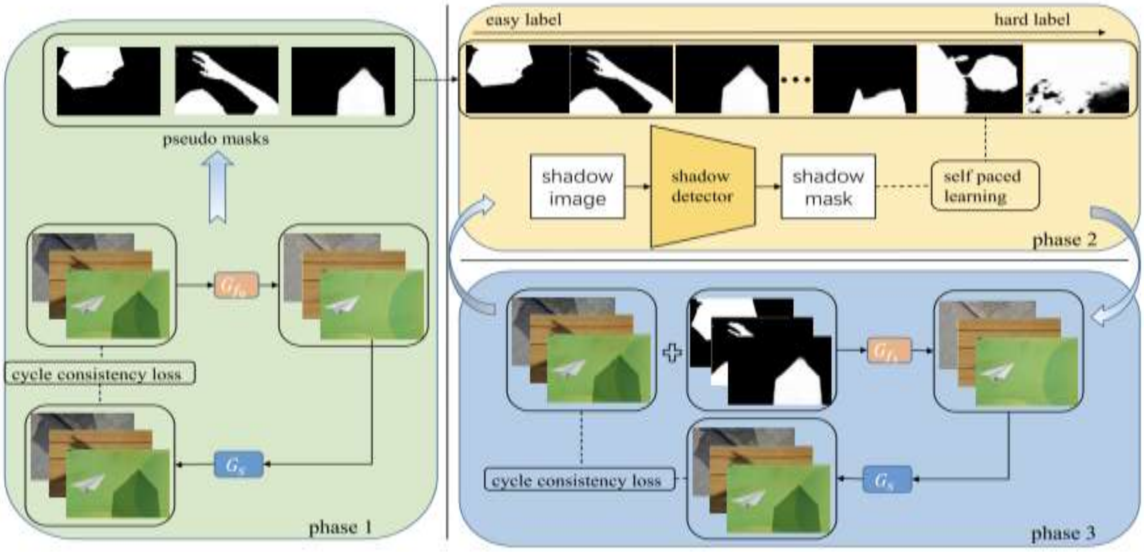


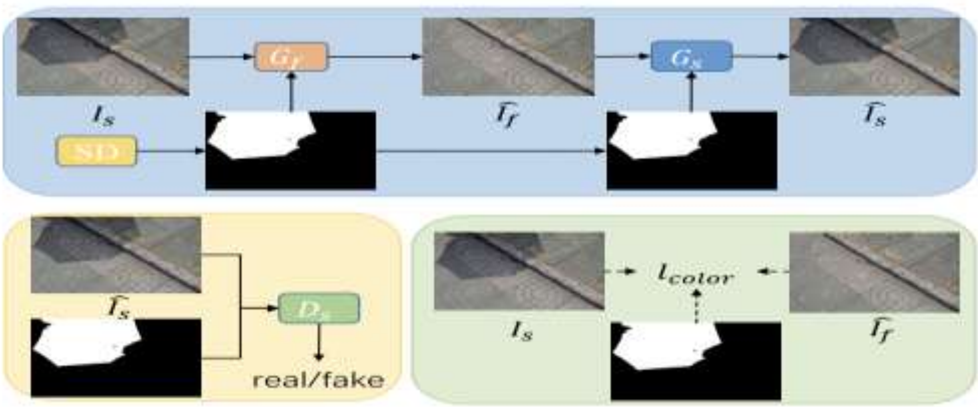
Fig. 2: Deep Reciprocal Network Architecture

Additionally, the detector architecture incorporates *dilated convolutions*, which are crucial for capturing long-range spatial dependencies—especially beneficial for identifying large or softly blurred shadow boundaries. These dilated layers allow the network to perceive a broader context without increasing computational overhead. Moreover, *squeeze-and-excitation (SE) modules* are integrated to recalibrate feature maps by adaptively weighting each channel, ensuring that critical shadow-related features receive greater attention during learning.

Phase 3: Enhanced Shadow Remover

The third phase employs the refined shadow masks generated by the Shadow Detector (SD) to retrain a more advanced generator network, denoted as Gf1G_f^1Gf1. By leveraging high-quality mask guidance, this retrained model focuses on precise and localized shadow correction, effectively restoring illumination while maintaining the structural integrity of the original scene. A *Shadow-Attention Discriminator* is introduced at this stage, which not only discerns real from generated images but specifically focuses on the shadow regions using the mask for guidance. This enables the adversarial learning process to emphasize realistic texture and boundary recovery in de-shadowed outputs.

To enhance perceptual quality further, a *Color-Maintenance Loss* is incorporated, ensuring that the color tones of non-shadow regions remain unaltered, thereby avoiding artifacts or unnatural transitions. Additionally, *perceptual loss* based on deep feature maps may be used to preserve fine-grained visual details. The training process also integrates regularization to prevent overfitting and ensure generalization across diverse lighting conditions. Together, these components make the enhanced remover capable of handling real-world shadow variations with improved fidelity and robustness.



iJETRM
International Journal of Engineering Technology Research & Management
(IJETRM)
https://ijetrm.com/

Self-Paced Learning Strategy

To manage training on noisy pseudo-masks, we implement a self-paced learning algorithm. This algorithm evaluates each sample's loss and gradually includes it for training based on a dynamic threshold λ\lambdaλ. Samples with high loss (likely noise) are temporarily excluded.

The SPL training process is detailed in the following algorithm:

Algorithm 1: Self-Paced Shadow Detection Training Algorithm

Input:

D: Shadow removal dataset;

θ: parameters of shadow detector;

v_i: the weight of i-th sample;

r_b: begin ratio of self-paced learning;

r_e: end ratio of self-paced learning;

n: epoch number;

λ: loss threshold;

Output:

Optimal shadow detector parameter θ*;

1. Initial $v_i = 1$, $\lambda = 0$; Initial θ to follow a zero-mean Gaussian distribution with a standard deviation of 0.02;
2. for each i ∈ [1, n] do
3. Fix v, optimize θ using equation 10 and save sample loss to log;
4. Sort the samples in D in ascending order of their loss value l_i;
5. Calculating current reference ratio r_c = (r_b + (r_e − r_b) * i / n) * 100;
6. Select the sample ranked in the r_c-th percentile, and use the corresponding loss value l_c as the threshold: λ = l_c;
7. for each sample x_i in D do
8. if sample loss $l_i > \lambda$ then Set $v_i = 0$;
9. end if
10. if sample loss $l_i \leq \lambda$ then Set $v_i = 1$;
11. end if
12. end for
13. end for
14. return θ*;

This strategy has been shown through experiments to reduce training errors and enhance the segmentation accuracy of the shadow detector.

Mathematical Formulation

Let:

- X- shadow image
- Y – shadow-free image
- M – shadow mask
- $X_{(relit)}$ – relit version of the shadow image
- G∯ – shadow removal generator
- G_s – shadow synthesis generator

The basic formulation for reconstructing a shadow-free image is:

$$Y = X_{(relit)} \cdot M + X \cdot (1 - M)$$

Where:

$$X_{(relit)} = w \cdot X + b$$

Here, w and b are affine parameters used for adjusting brightness and color tone.

Shadow mask M is treated as a latent variable, and the objective is to find an optimal M such that the generated de-shadowed image matches the ground truth (or looks visually real).

Training Losses

The total objective function for training the network combines several loss terms:

iJETRM
International Journal of Engineering Technology Research & Management
(IJETRM)
https://ijetrm.com/

#	Loss Type	Symbol
1.	Adversarial Loss	$\mathcal{L}^{GAN}$
2.	Cycle Consistency Loss	$\mathcal{L}c^{yc}$
3.	Identity Loss	$\mathcal{L}_{i}d_{t}$
4.	Color Maintenance Loss	$\mathcal{L}c_{olor}$

The total loss function:
$\mathcal{L}_{total} = \lambda_1 \cdot \mathcal{L}c^{yc} + \lambda_2 \cdot \mathcal{L}_{i}d_{t} + \lambda_3 \cdot \mathcal{L}c_{olor}$
Where λ_1, λ_2, λ_3 are weights that balance the impact of each component.
$X = G\_s(Y, M)$
Where **G_s** is the shadow synthesis generator, reconstructing the original shadow image from the clean image and shadow mask.
$\lambda_1 = 10.0$, $\lambda_2 = 5.0$, $\lambda_3 = 0.5$
These were experimentally tuned to prioritize color consistency and reconstruction.

Data Augmentation and Architecture Enhancements
To improve model generalization:

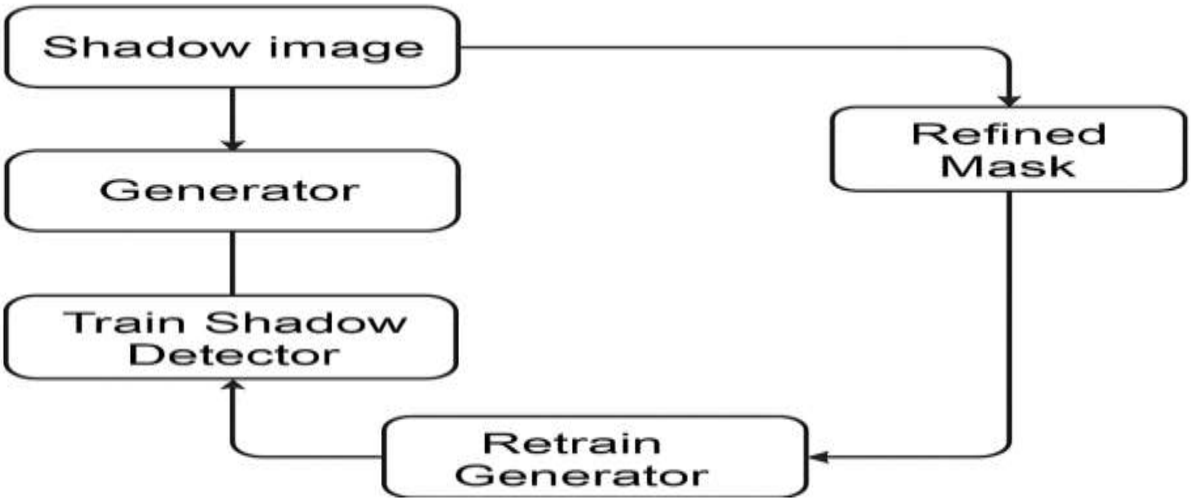
- Random cropping and color jittering are applied.
- Dilated convolutions in the bottleneck layers allow better feature extraction for large shadows.
- SE (squeeze-and-excitation) modules recalibrate the importance of features at each layer.

These adjustments collectively lead to better representation and discrimination between shadow and non-shadow regions.

Implementation Details

- Framework: PyTorch
- Optimizer: Adam ($\beta1 = 0.5$, $\beta2 = 0.9999$)
- Learning Rate: 0.0002 (decayed linearly after half training)
- Datasets:
  - ISTD: Paired dataset (used for comparison)
  - SRD: Pairwise dataset without full masks
  - USR: Unpaired dataset with high scene diversity

Training on ISTD and SRD is done in one cycle of reciprocal learning; USR, due to its complexity, uses two cycles.



RESULTS AND DISCUSSION

The proposed Deep Reciprocal Learning model (DRNet) was evaluated through extensive experiments on three benchmark datasets: ISTD, SRD, and USR. Each dataset represents varying conditions of shadow types, lighting,

iJETRM
International Journal of Engineering Technology Research & Management
(IJETRM)
https://ijetrm.com/

and scene complexity. DRNet was compared against several state-of-the-art models, including MaskShadowGAN, LGSN, and fully supervised methods like ST-CGAN and SP+M-Net.
Dataset Descriptions:
ISTD Dataset: Contains 1,330 training and 540 testing triplets with shadow, shadow-free images, and masks.
SRD Dataset: Includes 2,050 image pairs for training and 286 for testing; lacks explicit masks, which were manually annotated.
USR Dataset: Consists of unpaired data with 1,559 shadow images and 1,719 clean images, used to test performance in real-world unstructured scenarios.
Quantitative Evaluation:
For paired datasets (ISTD and SRD), evaluation is performed using **Mean Absolute Error (MAE)** in Lab color space, which better captures perceptual differences. The lower the MAE, the better the shadow removal quality.

Model	**ISTD MAE ↓**	**SRD MAE ↓**	**USR User Score ↑**
MaskShadowGAN	7.68	6.52	5.93
LGSN	7.45	6.21	6.34
DRNet (Ours)	**4.39**	**5.83**	**7.76**

DRNet outperforms previous methods by a significant margin, especially on USR, indicating superior generalization to real-world unpaired scenarios.
Quantitative Comparison of Various Methods on the ISTD Dataset
"MS" indicates MaskShadowGAN. "A", "N", and "S" refer to the MAE (Mean Absolute Error) for the Entire Image Region, Non-Shadow Region, and Shadow Region respectively. ↓ denotes that a lower value indicates better performance. The lowest and second-lowest values are highlighted in red and blue, respectively.

Supervision	**Method**	**MAE ↓**		
		A	**N**	**S**
Prior-based	Yang [41]	16.80	15.20	25.30
	Guo [42]	7.10	4.30	22.30
	Gong [18]	5.10	**3.40**	14.40
Pairwise	ST-CGAN [2]	9.50	8.60	14.40
	SP+M-Net [5]	**4.41**	**3.64**	8.84
Shadow mask	FSS2SR [6]	4.81	3.74	10.50
	G2R [43]	4.50	3.81	**8.14**
	MS [7]	5.48	4.52	11.32
Unpaired	LGSN [8]	5.02	4.02	10.64
	DRNet	**4.39**	**3.68**	**8.21**

➤ **Qualitative Analysis:**
A visual comparison across all datasets shows that DRNet produces:
- Cleaner de-shadowed images,
- Fewer residual artifacts,
- Better preservation of object texture and color temperature.

In ISTD, DRNet maintains fine details (like grass or wall textures) after shadow removal.
In SRD, it handles soft shadows and color variations better than MaskShadowGAN.
In USR, user study results showed DRNet-generated images appeared the most natural and visually realistic.

iJETRM
International Journal of Engineering Technology Research & Management
(IJETRM)
https://ijetrm.com/

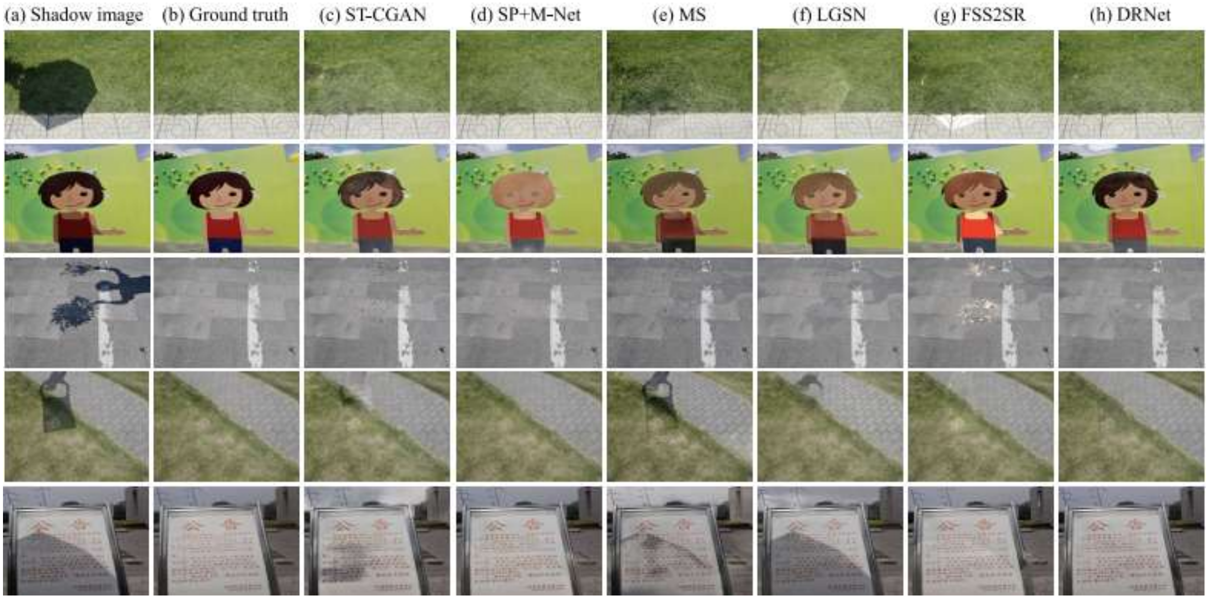


Fig. 4: Visual Comparison of Shadow Removal on ISTD and SRD

Ablation Study:

An ablation study was conducted to assess the impact of each proposed component:

"To verify the importance of each module, we conducted ablation experiments by removing individual components."

Model Variant	ISTD MAE ↓
Baseline (no shadow mask)	6.45
+ Shadow mask guidance	6.28
+ Reciprocal learning	4.78
+ Self-paced learning	4.67
+ Attention Discriminator	4.51
+ Color Maintenance Loss	**4.39**

Table

Ablation Study on Self-Paced Learning Strategy

Models M1 through M10 represent different experimental configurations. Models M2–M9 utilize pseudo masks from the ISTD training set, whereas M10 employs the actual ground truth labels from ISTD for training. For each configuration, the BER (Balanced Error Rate) on the ISTD test set is evaluated and presented in the last row. The parameters rsr_srs and rer_ere represent the start and end thresholds of self-paced learning respectively.

Model	M1	M2	M3	M4	M5	M6	M7	M8	M9	M10
Label Type	–	pseudo	pseudo	pseudo	pseudo	pseudo	pseudo	pseudo	pseudo	ground truth
r_s	–	1.0	0.95	0.92	0.93	0.94	0.93	0.93	0.95	1.0
r_e	–	1.0	0.95	1.0	1.0	1.0	0.97	0.98	0.99	1.0
BER ↓	7.08	4.37	3.87	3.99	3.63	3.92	3.87	3.76	3.41	**3.03**

➤ **Limitations:**

Although DRNet performs well, it sometimes struggles with:

- Black/dark objects being misclassified as shadows,
- Colored shadows from stained glass or reflective surfaces.

These limitations can be addressed in future work by including additional scene context or multispectral input.

User Study:

For the USR dataset, since no ground truth exists, a user study was conducted with 10 participants. Each participant rated 30 de-shadowed images from various models. DRNet achieved an average score of **7.76**, the highest among all models.

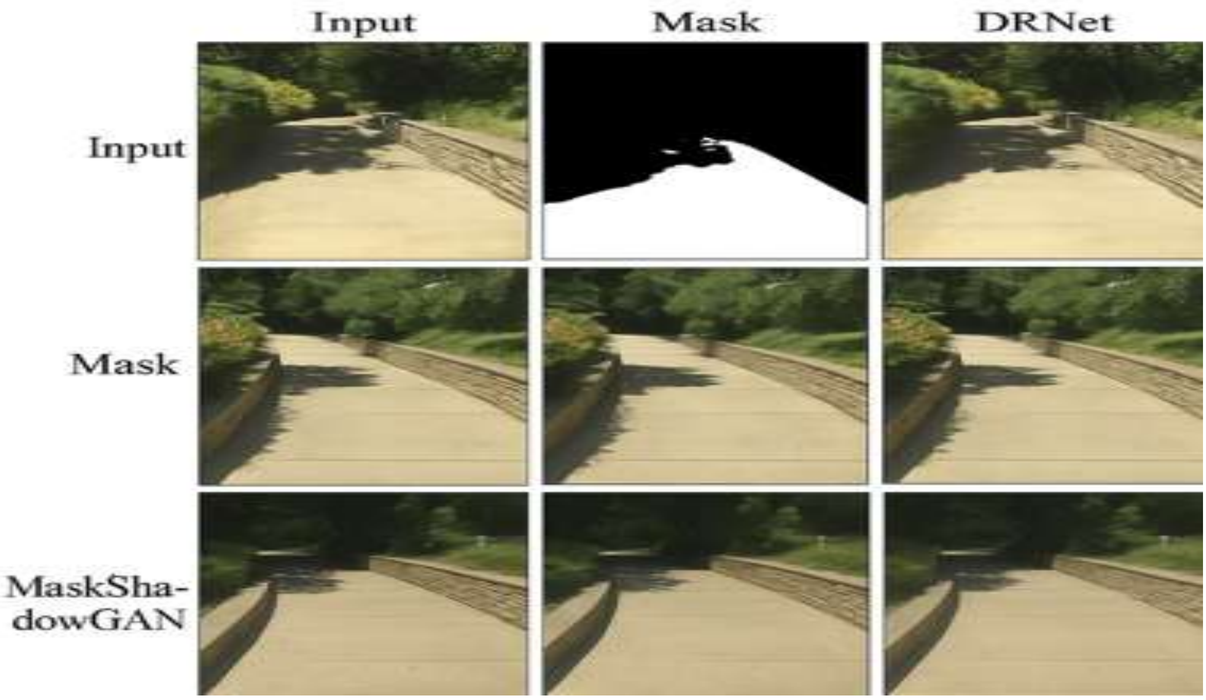


Fig. 5: Shadow Removal Results on USR Dataset

ACKNOWLEDGEMENT

We would like to express our sincere gratitude to the faculty and staff of the Department of Computer Science and Engineering at **RVR & JC College of Engineering** for their constant guidance and support throughout the development of this project. We are especially thankful to our mentor(s) for providing insightful feedback and valuable suggestions during the research and implementation phases.

We also extend our appreciation to **Apps Leap Technologies** for providing us the opportunity and platform to work on real-world applications of image processing, which significantly contributed to the practical understanding of this project.

We are grateful for the availability of open-source datasets such as ISTD, SRD, and USR, which played a vital role in benchmarking and validating our model. Finally, we thank our peers and families for their continuous encouragement and motivation throughout the research journey.

Above all, we express our gratitude to the Almighty for giving us the strength and clarity to complete this work successfully.

CONCLUSION

In this work, we proposed a Deep Reciprocal Learning (DRNet) framework to address the problem of shadow removal using unpaired image datasets. Unlike traditional supervised methods that require pixel-level annotations, our model learns to detect and remove shadows through a collaborative optimization between a shadow detector and a shadow remover.

By treating the shadow mask as a latent variable and employing a three-phase training scheme with self-paced learning, DRNet effectively simplifies the complex shadow removal task into manageable sub-problems. Additional components like color-maintenance loss and attention-guided discriminators further enhance output quality.

Experimental results on ISTD, SRD, and USR datasets demonstrate that DRNet outperforms state-of-the-art models in both quantitative accuracy and visual realism. The proposed framework shows great promise for real-world deployment where large-scale paired data is unavailable.

International Journal of Engineering Technology Research & Management
(IJETRM)
https://ijetrm.com/

In future work, we aim to extend this framework to handle colored shadows and explore its applicability in video-based shadow removal and general image restoration tasks.

REFERENCES

[1] L. Qu, J. Tian, S. He, Y. Tang, and R. W. H. Lau, "DeshadowNet: A multi-context embedding deep network for shadow removal," in Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR), Jul. 2017, pp. 2308–2316.

[2] J. Wang, X. Li, and J. Yang, "Stacked conditional generative adversarial networks for jointly learning shadow detection and shadow removal," in Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit., Jun. 2018, pp. 1788–1797.

[3] X. Hu, L. Zhu, C. Fu, J. Qin, and P. Heng, "Directionaware spatial context features for shadow detection," in Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit., Jun. 2018, pp. 7454–7462.

[4] X. Hu, C. Fu, L. Zhu, J. Qin, and P. Heng, "Direction-aware spatial context features for shadow detection and removal," IEEE Trans. Pattern Anal. Mach. Intell., vol. 42, no. 11, pp. 2795–2808, Nov. 2020.

[5] H. Le and D. Samaras, "Shadow removal via shadow image decomposition," in Proc. IEEE/CVF Int. Conf. Comput. Vis. (ICCV), Oct. 2019, pp. 8577–8586.

[6] H. Le and D. Samaras, "From shadow segmentation to shadow removal," in Proc. Eur. Conf. Comput. Vis. Cham, Switzerland: Springer, 2020, pp. 264–281.

[7] X. Hu, Y. Jiang, C. Fu, and P. Heng, "Mask-ShadowGAN: Learning to remove shadows from unpaired data," in Proc. IEEE/CVF Int. Conf. Comput. Vis. (ICCV), Oct. 2019, pp. 2472–2481.

[8] Z. Liu, H. Yin, Y. Mi, M. Pu, and S. Wang, "Shadow removal by a lightness-guided network with training on unpaired data," IEEE Trans. Image Process., vol. 30, pp. 1853–1865, 2021.

[9] M. P. Kumar, B. Packer, and D. Koller, "Self-paced learning for latent variable models," in Proc. NIPS, vol. 1, 2010, p. 2.