© FEB 2026 | IRE Journals | Volume 9 Issue 8 | ISSN: 2456-8880
DOI: https://doi.org/10.64388/IREV9I8-1714721

A Multi-Stage Deep Learning Framework for Document Image Restoration

VEESAM VENKATA SRINIVAS[1], INUKOLLU ANANTHA PRAKASH REDDY[2], GOSU MADHU[3], DHARMAVARAPU JAYARAJU[4], GADIPUDI KRISHNA VAMSI[5]
[1]*Assistant Professor, Department of IT, R.V.R & J.C.C.E, Guntur, India*
[2, 3, 4, 5]*Final Year Students, Department of IT, R.V.R & J.C.C.E, Guntur, India*

Abstract - Real-world camera-captured document images often exhibit complex degradations, including cast shadows, non-uniform illumination, and contrast distortion, which severely degrade visual quality and prevent robust document analysis. In this paper, we present an illumination estimation multi-stage deep learning framework for restoring document images that explicitly separates shadow suppression from illumination normalization. The proposed pipeline involves an initial deep network estimating and mitigating shadow-induced intensity variations before a refinement network corrects global illumination consistency while maintaining textual structure and fine document details. By decomposing the enhancement task into complementary stages, the framework effectively copes with both local shadow artifacts and global lighting imbalance in unconstrained document imaging scenarios. Extensive experiments on real-world camera-captured document images reveal that the proposed method provides visually coherent enhancement with more readable results compared to conventional image processing techniques and existing deep learning-based methods. Standard image quality metrics have been quantitatively evaluated, showing notable gains. The results indicate that the proposed framework offers a robust and practical preprocessing solution for analyzing camera-based document images.

Keywords: Document image enhancement, Shadows and Illumination, Multi-Stage Deep Learning, Camera-Captured Documents

I. INTRODUCTION

Document digitization has become an integral part of information management systems in recent times. The digitization process allows for access, sharing, and analysis of large volumes of textual data with ease. With the massive proliferation of smartphones and portable imaging devices, camera-captured document images are extensively employed in applications pertaining to mobile scanning, e-governance, digital archiving, and OCR [1], [10], [16]. However, unlike flatbed-scanner-captured documents, images captured under unconstrained environments often exhibit severe appearance degradation due to non-uniform illumination, cast shadows, and complex lighting conditions [1], [5], [6]. These detrimental effects severely reduce readability and affect the performance of downstream document analysis systems.

Traditional document image enhancement approaches, relying on handcrafted assumptions, include histogram equalization, homomorphic filtering, and methods leveraging Retinex [6], [12]. While these methods work well in controlled conditions, they inevitably fail to generalize under challenging real-world conditions where shadows and various illumination variations manifest themselves in highly complex spatial patterns [1], [6]. Recently developed deep models have indeed achieved very good results in enhancing degraded images; most existing deep models for image enhancement make use of a single-stage architecture that tries to jointly handle the factors contributing to degradation, often at the expense of incomplete shadow suppression or a loss of fine textual information.

Fig. 1. Examples of camera-captured document images under real-world conditions showing shadows, non-uniform illumination, and contrast degradation [2]–[4].

To overcome these limitations, this paper proposes an Illumination estimation multi-stage deep learning framework for document image restoration. The proposed approach breaks down the enhancement process into complementary stages: first mitigating shadow-induced artifacts, then normalizing illumination and refining appearance. By explicitly decoupling these tasks, the network can effectively handle both the local shadow regions and global illumination imbalance while maintaining the document structure and textual clarity. Extensive experiments on real-world camera-captured

© FEB 2026 | IRE Journals | Volume 9 Issue 8 | ISSN: 2456-8880
DOI: https://doi.org/10.64388/IREV9I8-1714721

document images demonstrate that the proposed approach accomplishes superior visual coherence and better readability compared to traditional enhancement techniques and existing deep learning-based methods.

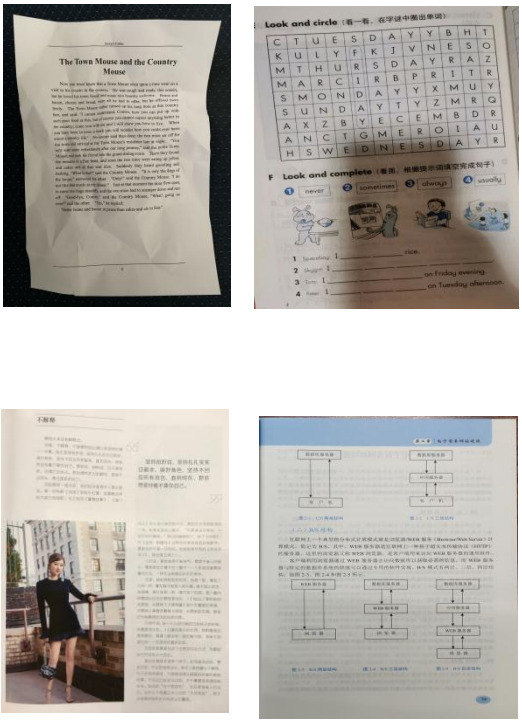


Fig. 1 illustrates representative examples of camera-captured document images exhibiting common degradations such as cast shadows, uneven illumination, and contrast distortion. [2], [3]

We address the challenges above from an algorithmic perspective by proposing an illumination estimation multi-stage deep learning approach to document image restoration [1]. The proposed approach follows a cascaded design where the enhancement task is decomposed into complementary stages. In the first stage, we employ a deep network that models global illumination characteristics and suppresses shadow-induced intensity variations by leveraging contextual information from the whole document image. This global enhancement stage focuses on learning large-scale illumination patterns, which are critical for preventing misinterpretation of shadows as foreground content. However, relying solely on global illumination correction may result in the loss of fine textual details and the introduction of local artifacts. So we address this limitation by including a refinement stage in the second part to recover local details, improve contrast consistency, and preserve structural information. The integrated multi-stage architecture is able to effectively handle global illumination imbalance and local shadow artifacts of document images captured under challenging real-world conditions. Extensive qualitative and quantitative studies on camera-captured document images demonstrate that our proposed framework enhances visual clarity and document readability compared to conventional enhancement techniques and existing deep learning-based methods while acting as an effective preprocessing module for downstream document analysis tasks.

The main contributions of this work can be summarized as follows:

1) In this paper, an illumination- and shadow-degradation-aware multi-stage deep learning framework is proposed for document images taken by cameras.
2) It breaks the enhancement task down into global illumination correction and local detail refinement stages, allowing it to perform robust restoration under unconstrained lighting conditions.
3) Experimental results show improved visual quality and readability of documents, therefore proving the effectiveness of the proposed framework as a preprocessing solution to real-world document analysis.

II. RELATED WORK

Document image enhancement has been an active research area owing to its importance in document digitization, optical character recognition, and content understanding. Broadly, the approaches taken so far may be divided into classical image processing methods and deep learning-based techniques. Recently, more and more multi-stage and cascaded learning frameworks were explored for addressing complex real-world degradations.

A. Algorithms

1) Shadow Handling Algorithms

Most of the document enhancement algorithms for shadow removal emphasize detecting and compensating for illumination attenuation due to occlusion [5], [6]. A commonly adopted strategy for these methods involves estimations of shadows followed by normalization of intensity. These algorithms first model the background illumination surface and suppress the shadow regions by rescaling the pixel intensities. Conventionally, shadows are detected based on local intensity ratios or gradient

discontinuities under the assumption that shadows reduce brightness while preserving chromatic consistency [6], [12]. In methods based on learning, shadow regions are implicitly modeled by leveraging global contextual information, hence enabling distinguishing between large shadowed areas and foreground text [1], [15]. Then, for global illumination correction, the estimated representation of shadows may serve as guidance, preventing the shadow from being mistakenly treated as document content.

2) Illumination Correction Algorithms

Illumination correction algorithms focus on the restoration of uniform lighting across the document surface. Various classical methods approximate the illumination component either by smoothing it or fitting a low-frequency background surface and then subtracting it from the original [6], [12]. Adaptive thresholding and Retinex-inspired methods try to separate illumination and reflectance components for brightness normalization [6]. More recent deep learning–based approaches conduct illumination normalization in a purely data-driven way, learning the nonlinear mapping between degraded and enhanced images. They emphasize local contrast and textual structure preservation in conjunction with correcting large-scale illumination imbalance and often benefit from multi-scale or cascaded processing strategies [1], [15].

B. Formulas

1) Shadow Modeling: Document images as captured by a camera can be modeled as interaction between illumination and reflectance components [5], [6]. Consider the observed image to be

$$I(x, y) = R(x, y) \cdot L(x, y)$$

where I(x,y) denotes the observed pixel intensity at location (x,y), R(x,y) represents the reflectance corresponding to document content, and L(x,y) denotes the illumination component. In shadowed regions, the illumination term L(x,y) exhibits significant attenuation. Shadow removal methods aim to estimate an illumination-normalized image by compensating for the spatial variation in L(x,y), thereby reducing shadow influence while preserving reflectance information.

2) Illumination Normalization: For correction of illumination, the estimated component of illumination is generally normalized to a reference level. A corrected image can be represented as

$$I_{norm}(x, y) = \frac{I(x, y)}{L(x, y) + \epsilon}$$

where I_{norm}(x,y) is the illumination-corrected image and ϵ is a small constant introduced for numerical stability. In learning-based frameworks, the illumination function L(x,y) is implicitly learned through network parameters, allowing the model to adaptively correct both global illumination imbalance and local shadow artifacts [1], [15].

Connection between the Theoretical Model and Proposed Method

The illumination-reflectance formulation above provides a conceptual basis for the shadow and illumination degradations in camera-captured document images. This decomposition is not done explicitly in the proposed multi-stage deep learning framework; instead, the network learns implicitly to suppress illumination-induced variations through data-driven feature representations. First, the enhancement stage focuses on capturing global contextual information corresponding to large-scale illumination patterns and shadow distributions, while subsequently, the refinement stage emphasizes local structure preservation and detail restoration. By applying pretrained networks sequentially these stages in an end-to-end manner, the proposed approach effectively approximates the process of illumination normalization and reflectance Preservation without explicit estimation of the illumination component, thus performing robust document image restoration under unconstrained lighting conditions.

Table 1: Image Quality and Readability Metrics for Classical Enhancement Methods

Technique / Method	Classical Deshadow	Illumination Correction	Images
Entropy	3.26384	2.32	10
RMS Contrast	29.09	2.83	10
Edge Sharpness	290.96	10.63	10
BRISQUE	73.53	68.92	10

© FEB 2026 | IRE Journals | Volume 9 Issue 8 | ISSN: 2456-8880
DOI: https://doi.org/10.64388/IREV9I8-1714721

III. PROPOSED METHOD

Within this research, we also introduce a two-stage deep learning scheme to improve document images captured by cameras that are compromised by shadows and illumination irregularities. The proposed scheme is inspired by the fact that document degradation exists at different spatial scales: a global illumination imbalance that comes along with shadows and illumination, as well as degradation on a local level that impacts text clarity and detail. To overcome these, our proposed scheme aims to leverage a two-stage structure composed of a Global Context Network (GC-Net) and a Detail Restoration Network (DR-Net) to complete global illumination correction and local detail restoration, respectively [1].

As depicted in Fig. 2, the proposed framework adopts a sequential inference pipeline. Given a camera-captured document image, the input first passes through GC-Net to suppress large-scale illumination variations and dominant shadows from global contextual information. Then the output of GC-Net is a globally enhanced document image, which is further refined by DR-Net to restore local details, improve text clarity, and enhance contrast consistency. Both networks run, and no additional training is performed during inference based on the pretraining checkpoints without any implicit shadow suppression or handcrafted illumination modeling.

With the help of this insight, the method proposes a two-phase enrichment technique that breaks down the document image restoration task into illumination correction at the global scale and refining at the local level. The initial stage has been formulated with the intention of eliminating the major effects of shadows and lighting by utilizing global information available in the document image. Yet, at the global level, there could be slight artifacts or insufficient enrichment at the text detail level [1].

Based on this design rationale, the proposed GCDRNet framework is formed by cascading a Global Context Network (GC-Net) and a Detail Restoration Network (DR-Net). This cascaded representation enables the two pretrained networks to concentrate on different degradation aspects at their respective spatial scales. For instance, the GC-Net mainly handles large-scale illumination degradation, while the DR-Net mainly focuses on local structure and detail restoration. Therefore, a balanced document restoration result can be obtained. The system architecture is shown in Fig. 2.

In summary, the GCDRNet framework is designed to handle the complementary degradation aspects in camera-captured document images. By separating the global illumination normalization process from the local detail restoration process in a cascaded inference framework, the proposed method overcomes the drawbacks of existing single-stage enhancement.

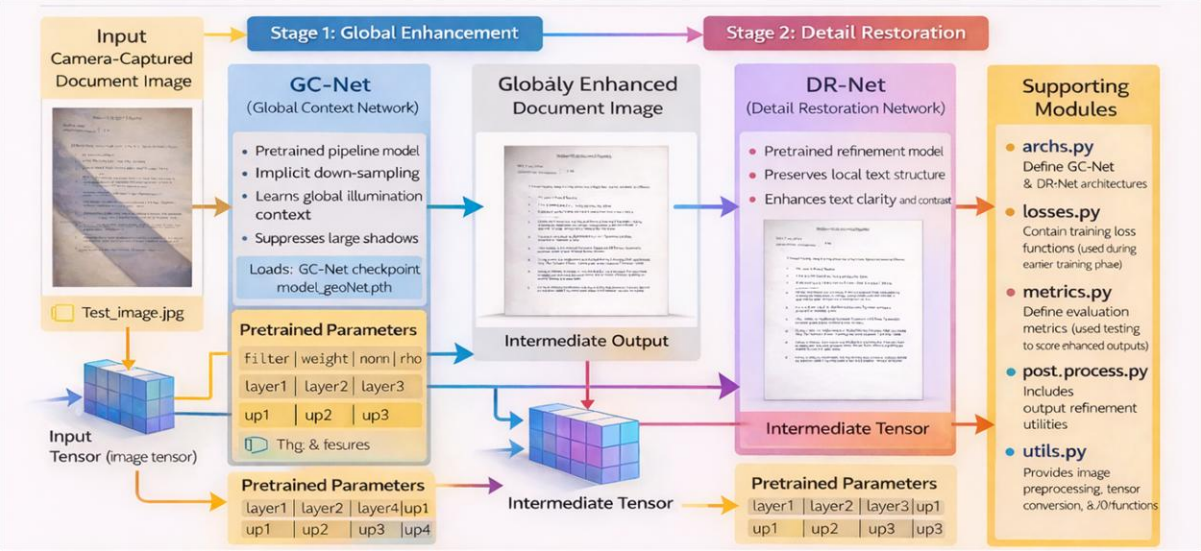


Fig. 2. Overview of the proposed GCDRNet inference pipeline. The input camera-captured document image is first processed by GC-Net for global illumination correction via shadow suppression, followed by DR-Net for local detail restoration to produce the final enhanced output.

© FEB 2026 | IRE Journals | Volume 9 Issue 8 | ISSN: 2456-8880
DOI: https://doi.org/10.64388/IREV9I8-1714721

A. Global Illumination Enhancement using GC-Net:
Global context information is very important for document image enhancement, and large shadow areas and lighting variations cannot be properly addressed by using the proposed approach to focus solely on local imaging of an input document. Normally, captured images of documents contain shadows with large areas, which may also include text, so such images cannot be properly addressed by focusing solely on the local aspects by using the proposed approach. The proposed approach uses the first stage of the GC-Net technique [1].

We load a camera-captured document image and prepare it for the network:

$$I \in \mathbb{R}^{H \times W \times 3}$$

Where H & W are used to represent the spatial dimensions in the image, while the other three represent the color channels of the RGB color model.

i. Image Formation Assumption

The proposed scheme utilizes an assumption in the reflectance-illumination imaging model that has been commonly acknowledged for document image enhancement. The document image obtained by capturing the document can be represented by the equation [5], [6]:

$$I = R \odot S$$

Where R represents the reflectance component of the clean content of the documents, S denotes the lighting or dark portion, and $\odot$ represents element-wise multiplication. This definition leads to the design of the global enhancement stage, which strives to estimate and compensate for the illumination effect.

ii. Shadow / Illumination Estimation

The Global Context Network (GC-Net) is used to simulate large-scale illumination changes and shadow areas. For an image, I, GC-Net make a prediction of an illumination (shadow).

$$\hat{S} = G_{\theta_g}(I)$$

Where $G_{\theta_g}(\cdot)$ denotes the GC-Net with pretrained parameters θg, and $S \in \mathbb{R}^{H \times W \times 3}$ is the estimated illumination map. GC-Net relies on the global receptive field and the downsampling process to capture large-scale shadows [1], [15].

iii. Global Illumination Correction

The globally enhanced image is then achieved by element-wise division using the predicted illumination map:

$$I_{GC} = \text{clip}\left(\frac{I}{\hat{S}}, 0, 1\right)$$

Where I_{GC} represents the illumination-corrected image and clip(·) ensures that pixel values remain within a valid intensity range. This division-based correction explicitly removes shadow effects, thereby retaining the original document structure.

iv. Intermediate Representation

After the first stage, the intermediate image $I_{GC} \in \mathrm{R}^{H \times W \times 3}$ exhibits reduced shadow dominance and improved global illumination consistency, providing a suitable input for subsequent refinement.

B. Detail Restoration Network (DR-Net)

Even with the GC-Net, there may still be a problem with the final output, such that the textual details may not have been sufficiently enhanced, while minor artifacts might have been added to the document while correcting the global illumination effect. This problem is then resolved with the Detail Restoration Net DR-Net [1], [15].

DR-Net is based on the globally enhanced image obtained from the GC-Net and aims at the restoration of the high-frequency channels like text strokes, edges, and local contrast. Contrary to the global enhancement process, the restoration process demands a subtle attention to the fine patterns in the spatial domain, which assume paramount importance in the identification of characters and the document layout.

i. Input Fusion

To maintain the input content and the globally enhanced info, the DR-Net network uses a combined representation, which is achieved through:

$$X_{DR} = \text{Concat}(I, I_{GC})$$

where $X_{DR} \in \mathrm{R}^{H \times W \times 6}$.

ii. Detail Refinement

The Detail Restoration Network refines the local structures and improves the details by applying the following nonlinear transformation:

$$\hat{I} = D_{\theta_d}(X_{DR})$$

© FEB 2026 | IRE Journals | Volume 9 Issue 8 | ISSN: 2456-8880
DOI: https://doi.org/10.64388/IREV9I8-1714721

Where $D_{\theta_d}(\cdot)$ denotes DR-Net with pretrained parameters θ_d, and $\hat{\ }$ is the final restored document image [13], [14].

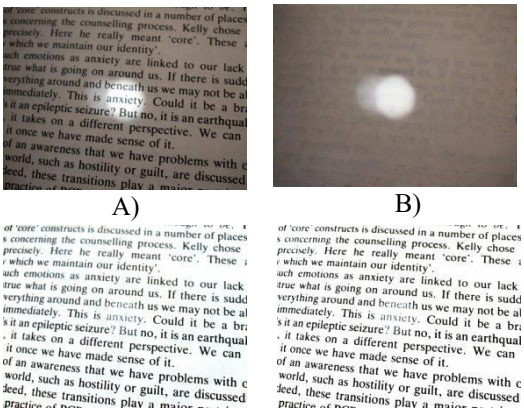
End-to-End Formulation:

Combining both stages, the complete proposed GCDRNet pipeline can be expressed as

$$\hat{I} = D_{\theta_d}\left(\text{Concat}\left(I, \text{clip}\left(\frac{I}{G_{\theta_g}(I)}, 0, 1\right)\right)\right)$$

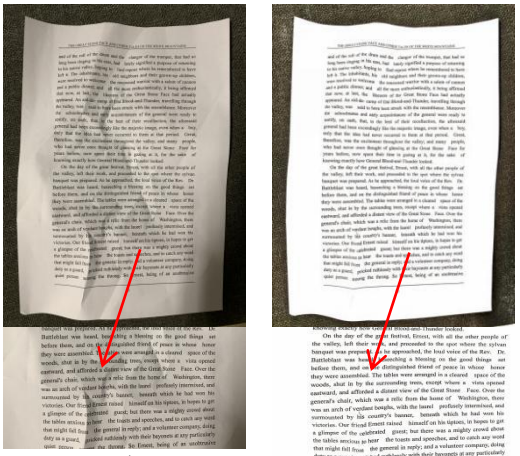
This formulation compactly represents the two-stage global-to-local enhancement strategy employed in the proposed framework.

By breaking down document image enhancement into two sub-tasks: illumination correction and local detail restoration, the proposed method is able to cope with drawbacks existing in single-stage models. The cascaded architecture helps each network focus on their specific task and therefore enhances the visualization quality and text readability for complex lighting conditions. At the same time, this structure is lightweight and aligned with implementation.



A) B) C) D)

Fig. 3. Visual comparison of document enhancement results. (A) Input camera-captured document image with shadows and uneven lighting conditions. (B) Predicted shadow/illumination map generated by GC-Net. (C) Illumination-corrected and then globally enhanced image. (D) Final restored document image produced by the proposed GCDRNet.



A) B)

Fig. 4. Some example image pairs of the enhancement outcome on the captured document images by the suggested framework. (A) Degraded source images with strong shadow and lightness inconsistency. (B) Enhanced images obtained by the suggested GCDRNet. Compared with the degraded source images, the enhanced images have better lightness uniformity, less shadow dominance, and more details in the highlighted areas.

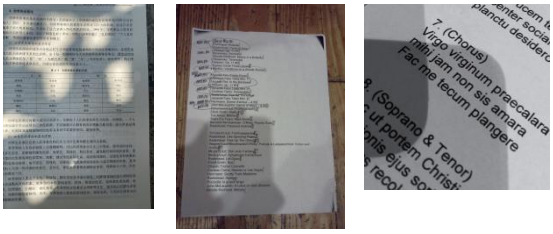
A) B) C)

Color Adjustment

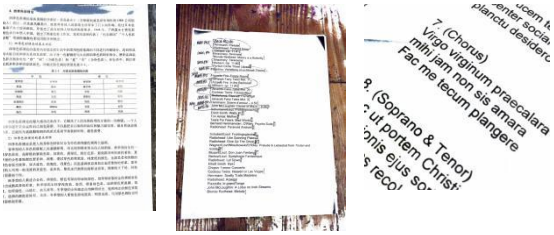
© FEB 2026 | IRE Journals | Volume 9 Issue 8 | ISSN: 2456-8880
DOI: https://doi.org/10.64388/IREV9I8-1714721

D)

Bleed-through Removal



E)



F)

Shadow Removal

Fig. 5. Image pairs illustrating the effectiveness of the proposed GCDRNet approach for various document degradations. (A) (C) (E) Camera-captured document images with various degradations due to the presence of shadows, bleed-through effects, and color distortions. (B) (D) (F) Restored document images obtained through the proposed approach, which clearly present improved illumination distribution, eliminated interference effects, and clear text contents despite document degradations.

Table 2. Visual Quality and Readability Assessment of GCDRNet Outputs

Output Type	GCDRNet (final)	GC-Net	Shadow map	Images
Entropy	3.59232	4.44394	6.31801	12
RMS Contrast	71.2573	55.1817	30.4636	12
Edge Sharpness	1211.2	800.16	4.69538	12
BRISQUE	39.9925	24.4147	61.3031	12

Table 3 shows a quantitative comparison of character accuracy using OCR on the original input document images and their corresponding enhanced outputs produced by the proposed GC-Net + DR-Net pipeline. The accuracy of the OCR process is used as a proxy measure to assess the readability of the documents, as it is a direct measure of the effectiveness of enhancement methods in improving text recognition tasks. The experiment was conducted on a sample set of 30 document images with varied illumination distortions, bleed-through, and text content [8], [9], [10], [16].

Table 3. OCR-Based Character Accuracy Comparison for Document Readability Evaluation

Image Type	Average OCR accuracy (%)	Images
Input (RAW)	45.32	30
DR-Net Output	62.78	30
Improvement	+17.46	30

As evident from Table 3, the proposed enhancement pipeline brings about a significant improvement in the OCR-based accuracy of characters. The raw input images resulted in an average accuracy of 45.32%, whereas the enhanced images resulted in an improved accuracy of 62.78%, thus bringing about an average improvement of 17.46%. This result clearly shows that the proposed GC-Net + DR-Net system is successful in enhancing the readability of the documents by removing the inconsistencies in illumination and the degradation artifacts.

Though the degree of improvement is different for different samples, the accuracy improvement is greater for samples with moderate to severe illumination degradation, where the role of enhancement is more important for improving the readability of the text. For some samples, the accuracy is as high as 95.60%, thus indicating a complete recovery of the text. The low accuracy for some challenging samples is mainly due to the presence of handwritten text and severe degradation of the document, which is always difficult to read even for OCR systems.

Qualitative Analysis & Observations

The above qualitative results clearly indicate the effectiveness of the proposed GCDRNet model on

various real-world document degradation types. From the stage-wise comparison graph, the document quality is gradually enhanced from the degraded input to the final output. The shadow or illumination maps generated by GC-Net help to effectively rectify the large-scale illuminations, which results in the effective removal of dominant shadow areas and the uneven illuminations present in the degraded document. This is quite effective in correcting the uneven illuminations on the document surface [1], [5].

After this global illumination enhancement, Detail Restoration Network (DR-Net) is employed for further improvement on enhancing details of the mapped image. The outputs show more defined strokes of texts, clear boundary lines of characters, and even contrast, unlike the input and intermediate results from GC-Net. In addition, this two-step method is also capable of avoiding over-smoothing and losing textual details, a typical drawback of a single-step enhancement technique [1], [15].

The visual results of the comparison on varied degradation types such as shadow images, artifacts, and color contrast clearly reveal the effectiveness of the proposed model across different environmental conditions of the document. It has been found in all cases that the improved output images possess better readability and possess minimal background interference effects, which clearly justify the design concept of the model based on decomposing the task of enhancing a document image by a two-step process involving global and local operations.

IV. CONCLUSION

In this paper, a two-stage deep learning framework, namely GCDRNet, is proposed to improve camera-captured document images degraded by complex real-world degradations. For the first time, the proposed approach decomposes the restoration process into global illumination correction and local detail refinement, which effectively tackles challenging defects such as shadow artifacts, uneven lighting conditions, bleed-through interference, and color imbalance. The large-scale illumination variations are suppressed by the explicit shadow compensation of the Global Context Network (GC-Net), while the Detail Restoration Network (DR-Net) refines the local structures to improve text clarity and consistency in contrast [1], [15].

Extensive qualitative analysis shows that the proposed framework gradually enhances the appearance of documents for various types of degradation, and the visual clarity and readability of the document images are better than those of intermediate enhancement stages. Such observations are further established through quantitative evaluation using no-reference image quality metrics, which ensure that the final enhanced outputs have better contrast, edge sharpness, and perceptual quality. Stage-wise analysis underlines the complementary role of GC-Net and DR-Net and justifies the effectiveness of the cascaded global-to-local enhancement strategy [1].

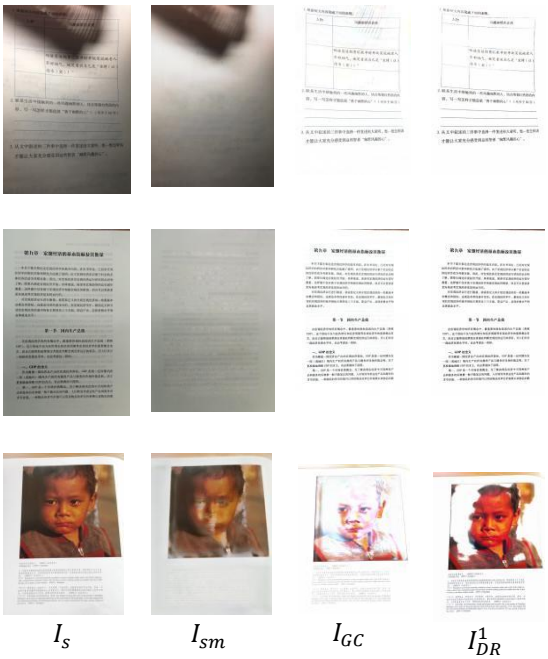


Fig. 6. Stage-wise enhancement results of the proposed GCDRNet. From left to right: input image I_s, predicted illumination/shadow map I_{sm}, GC-Net output I_{GC}, and final enhanced image I_{DR}^1.

In all, the GCDRNet framework proposed here is lightweight and implementation aligned, which makes it suitable for real-world document image enhancement with no dependency on pixel-aligned ground truth data [2]-[4]. Further extensions could be made in introducing other document artifacts within this framework, integrating task-driven objectives such as OCR-aware optimization, and improving its robustness across a wider variation of capture conditions and document types [9], [10].

REFERENCES

[1] J. Zhang, C. Liu, J. Yu, and Z. Guo, "Appearance Enhancement for Camera-Captured Document Images in the Wild," *IEEE Transactions on Artificial Intelligence*, vol. 5, no. 5, pp. 2314–2328, May 2024.

[2] A. Das, H. Ma, and S. Kar, "DocProj: A Projective Distortion Dataset for Document Image Enhancement," *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 1474–1483, 2020.

[3] X. Yi, Y. Zhou, and L. He, "Doc3D: A Large-Scale Synthetic Dataset for Document Image Deformation," *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 2252–2260, 2016.

[4] X. Yi, L. Zhang, and J. Yu, "Doc3DShade: A Synthetic Dataset for Shadow Removal in Document Images," *Proceedings of the IEEE International Conference on Computer Vision (ICCV)*, pp. 157–165, 2017.

[5] J. Jung, S. Cho, and N. I. Cho, "Illumination Correction for Document Images Using Surface Reconstruction," *IEEE Transactions on Image Processing*, vol. 26, no. 9, pp. 4362–4376, Sept. 2017.

[6] R. Hedjam, R. F. Moghaddam, and M. Cheriet, "A Statistical Approach for Illumination Compensation of Historical Documents," *IEEE Transactions on Image Processing*, vol. 21, no. 12, pp. 5111–5121, Dec. 2012.

[7] Z. Liu et al., "UNeXt: MLP-Based Rapid Medical Image Segmentation Network," *Proceedings of the International Conference on Medical Image Computing and Computer-Assisted Intervention (MICCAI)*, 2022.

[8] A. Mittal, A. K. Moorthy, and A. C. Bovik, "No-Reference Image Quality Assessment in the Spatial Domain," *IEEE Transactions on Image Processing*, vol. 21, no. 12, pp. 4695–4708, Dec. 2012.

[9] R. Smith, "An Overview of the Tesseract OCR Engine," *Proceedings of the International Conference on Document Analysis and Recognition (ICDAR)*, pp. 629–633, 2007.

[10] A. Bissacco, M. Cummins, Y. Netzer, and H. Neven, "Photo OCR: Reading Text in Uncontrolled Conditions," *Proceedings of the IEEE International Conference on Computer Vision (ICCV)*, pp. 785–792, 2013.

[11] L. Kang, Y. Li, and D. Doermann, "Orientation Robust Text Line Detection in Natural Images," *IEEE Transactions on Image Processing*, vol. 25, no. 9, pp. 4182–4195, Sept. 2016.

[12] S. Lu, B. Su, and C. L. Tan, "Document Image Binarization Using Background Estimation and Stroke Edges," *International Journal on Document Analysis and Recognition (IJDAR)*, vol. 13, no. 4, pp. 303–314, 2010.

[13] S. Xie, Z. Tu, "Holistically-Nested Edge Detection," *Proceedings of the IEEE International Conference on Computer Vision (ICCV)*, pp. 1395–1403, 2015.

[14] C. Tensmeyer and T. Martinez, "Document Image Binarization with Fully Convolutional Neural Networks," *ICDAR*, pp. 99–104, 2017.

[15] Y. Wang, J. Zhang, and Z. Guo, "Document Image Appearance Enhancement via Deep Learning," *Pattern Recognition*, vol. 95, pp. 292–304, 2019.

[16] D. Karatzas et al., "ICDAR 2015 Competition on Robust Reading," *ICDAR*, pp. 1156–1160, 2015.