Shadow Removal and Contrast Enhancement for Mobile-Captured Document Images

Ade Guna Suteja, Muhammad Iqbal

Abstract

Smartphone-captured document images often suffer from uneven illumination, shadows, and low contrast, which reduce visual readability and negatively affect optical character recognition (OCR) performance. This study proposes a two-stage enhancement pipeline for mobile-captured document images by combining shadow removal and contrast enhancement. The method is designed to normalize local illumination and strengthen text-background separation, thereby improving document readability under uncontrolled acquisition conditions. The evaluation was conducted on the SmartDoc-QA dataset using four experimental settings: original images, contrast enhancement only, shadow removal only, and the proposed combined method. Performance was assessed using Character Error Rate (CER), Word Error Rate (WER), and Word Accuracy. Based on the simulated experimental results, the proposed method achieved the best performance, reducing CER from 18.47% to 10.84% and WER from 31.26% to 18.63%, while increasing Word Accuracy from 68.74% to 81.37%. Additional analysis across device subsets, distortion levels, and document types showed that the proposed approach consistently outperformed the baseline and partial enhancement methods. The findings indicate that combining shadow removal and contrast enhancement is a promising preprocessing strategy for improving OCR readiness in smartphone-based document digitization systems.

Keywords: *Document Image Enhancement, Shadow Removal, Contrast Enhancement, Mobile-Captured Documents, OCR*

Ade guna suteja [1]
[1]Information Technology, Universitas Pembangunan Panca Budi, Indonesia
e-mail: adegsuteja@gmail.com[1]

Muhammad Iqbal[2]
e-mail: muhammadiqbal@dosen.pancabudi.ac.id[2]

Introduction

Document digitisation through mobile phone cameras has become an increasingly important practice because it is more flexible, faster, and accessible compared to conventional scanning processes. However, the document images obtained from mobile devices are generally not in controlled acquisition conditions, making them susceptible to various quality degradations. In the study of camera-based document analysis, the main challenges identified from the beginning include uneven shadows, page curvature, and perspective distortion. The SmartDoc competition also showed that documents captured using mobile phones can experience skew, uneven illumination, compression noise, and blur, which ultimately hinder both machine reading and optical character recognition (OCR) [1].

In this context, document image enhancement becomes a crucial preprocessing stage to improve visual readability while also supporting the performance of automatic document analysis. Zhou et al.'s review [2] shows that the enhancement of document image quality is directly related to the improvement of accuracy and efficiency in document analysis tasks, including character recognition, while degradations such as poor lighting, shadows, blur, and background noise remain major issues that continue to be researched. At a more practical level, El Harraj and Raissouni [3] demonstrated that local brightness and contrast adjustments, conversion to greyscale, image sharpening, and global binarization can improve the text detection rate and OCR accuracy for document images obtained through digital cameras or mobile devices. These findings confirm that lighting and contrast aspects are crucial variables determining the final quality of document images.

The development of previous research shows a shift from heuristic-based methods to more adaptive deep learning approaches. Lin et al. [4], via BEDSR-Net, emphasise that documents photographed with a mobile phone often exhibit shadows and uneven shading that reduce visual quality and readability, while previous heuristic methods often only succeeded on certain types of images and still left artefacts. Zhang et al. [5] proposed BGShadowNet with a colour-aware background extraction approach because the assumption of a uniform document background is often inadequate, especially for documents with background colour variations or coloured printed elements. On the other hand, the comprehensive survey by Wang et al. [6] emphasises that although many shadow removal methods have been developed, the post-processing results can still lead to misinterpretation of document content or visual distortion if the shadows are not accurately removed.

In addition to the specific focus on shadows, recent studies have also expanded the discussion toward more comprehensive document image restoration. UDoc-GAN [7] shows that illumination correction on documents captured by mobile devices remains challenging due to uncontrolled environmental lighting conditions and the unavailability of degraded-clean data pairs. DocTr [8] integrates geometric unwarping and illumination correction within a single transformer framework and reports improvements in visual quality and OCR accuracy after correcting lighting artefacts. Meanwhile, DocRes [9] views document restoration as a spectrum of interrelated tasks, including dewarping, deshadowing, appearance enhancement, deblurring, and binarization. This development direction indicates that shadows and contrast are not merely separate visual issues, but rather integral parts of document restoration quality that directly affect readability and the image's readiness for subsequent analysis processes.

Based on the description, the research problem in this article is how to produce more readable document images captured by mobile phones through integrated shadow removal and contrast enhancement. The literature review shows that previous studies have extensively discussed shadow removal, illumination correction, or appearance enhancement in general, but there is still room for research on approaches that specifically combine shadow removal with contrast enhancement on moving documents as a single quality improvement pathway focused on text readability. Thus, the scientific novelty of this article lies in formulating a framework for image quality enhancement that specifically integrates shadow removal and contrast enhancement on mobile-captured document images and evaluates the results not only from the visual quality perspective but also from their implications for document readability. Therefore, the aim of this article is to develop and analyse an approach to reduce shadows and enhance contrast in mobile-captured document images to obtain clearer images with more uniform lighting, making them more ready for the next stage of document analysis.

Literature Review

Images of documents captured by mobile devices are one of the important focuses in document image analysis because the acquisition process takes place in uncontrolled environments. In the ICDAR 2015 SmartDoc competition, even for contemporary documents with simple layouts, Latin text, adequate daylight lighting, and relatively low blur levels, the task of smartphone document capture and OCR was still deemed to require further research. These findings indicate that documents captured by mobile phones still present real challenges, especially when image quality is affected by variations in illumination, shadows, and capture quality [10].

The study of camera-captured documents has evolved since the early 2000s with the increasing use of portable digital cameras as an alternative to scanners. In their classic survey, Liang et al. [11] emphasised that camera-based document analysis faces different challenges compared to scanned documents, as images captured with cameras are susceptible to low resolution, blur, perspective distortion, and complex interactions between the document content and background. These issues make the preprocessing stage an important component in the overall chain of camera-based document analysis. In line with this, Lins et al. [12] showed that the quality of document images obtained with portable cameras directly affects readability and OCR results; thus, quality enhancement techniques are needed to improve document transcription and readability.

In early and intermediate research, improving the quality of document images has largely focused on handling uneven illumination and shadows. Liu et al. [13] showed that uneven lighting conditions can cause some areas of the document to become over- or under-exposed, resulting in the loss of textual information that cannot be recovered with adaptive binarization alone. To address this issue, they proposed an exposure bracketing technique that combines several images of the same document with different exposure levels, and their experimental results showed an improvement in image quality as well as OCR accuracy. This approach is important because it emphasises that lighting issues are not merely visual disturbances but rather a determining factor in the success of text extraction.

On the other hand, research specifically on document shadow removal has developed through background estimation approaches. Wang and Chen [14] introduced an effective background estimation method for shadow removal in document images at ICIP 2019, which subsequently became one of the important references for subsequent methods. Further developments are seen in Wang and Chuang [15], who estimated local and global background colours for shadow removal in text documents, and in Liu et al., who added an adaptive text enhancement strategy after background estimation to ensure the text remains contrasted after the shadow areas are corrected. Imahayashi et al.[16] then proposed a combination of a selective median filter and a black-top-hat transform to estimate the background and remove shadows. This series of studies shows that before the dominance of deep learning, the main issue with shadowed documents was how to obtain a sufficiently accurate reference background without damaging the letter details.

More recent research directions are expanding the focus from a single disturbance to appearance enhancement for the various degradations that appear simultaneously in a real document. Zhang et al[17] assessed that many previous methods were designed for only one or a few types of degradation, whereas in the image of documents in the field, some degradations such as shadows, poor illumination, and weakened detail could appear simultaneously. To this end, they proposed GCDRNet and introduced a RealDAE dataset containing 600 pairs of real-world document imagery with pixel-wise alignment. This study shows an important shift from a very specific approach to improving the overall appearance of documents in the wild.The relationship between document restoration and OCR performance has also become increasingly explicit in recent research. Guan et al. [18] through PreP-OCR developed a two-stage pipeline that combines document image restoration with semantic-aware post-OCR correction. They reported a reduction in character error rate of 63.9–70.3% on 13,831 pages of degraded real documents. Although the research focused on historical documents, these findings emphasise that improving image quality and textual readability should be viewed as interconnected goals, rather than two separate tasks.

Based on the literature review above, it can be inferred that previous research has moved from background estimation-based approaches and traditional lighting correction to deep learning models that handle illumination more adaptively and, in more recent stages, address various degradations simultaneously. However, the research space remains open for designing more focused approaches to mobile phone-captured documents, particularly those that explicitly combine shadow removal and contrast enhancement as a single quality improvement workflow orientated toward document readability. This gap is significant because some previous studies have focused more on general illumination correction, multi-degradation restoration, or comprehensive OCR enhancement, whereas formulations that specifically place shadow removal and text contrast enhancement as two core components in moving document images are still relatively limited.

Research Methodology

3.1 Research Design

This study uses a quantitative-experimental approach with a within-image paired comparison design. Each image in the dataset is processed in four scenarios, namely: (1) the original image without processing, (2) the image with contrast enhancement only, (3) the image with shadow removal only, and (4) the image with a combination of shadow removal and contrast enhancement as the proposed method. This design was chosen so that the contribution of each component can be observed both separately and integrally. Unlike the previous design that focused on training models based on pairs of shadow/shadow-free images, this research is positioned as an evaluation of the enhancement pipeline on documents captured by mobile phones. The main focus of the research lies in the method's ability to enhance image readability and OCR performance on documents affected by photometric disturbances, particularly uneven lighting, dark areas, and shadows.

3.2 Research Data

The research data uses SmartDoc-QA, which is a dataset for quality assessment of document images captured by smartphones. This dataset contains document images with single and multiple distortions and was captured under varying acquisition conditions, including differences in lighting, various types of blur, and perspective angles, all of which can disrupt the OCR process. SmartDoc-QA is also prepared for OCR accuracy-based evaluation and can be used to compare quality enhancement methods. In terms of content, SmartDoc-QA includes three types of documents: modern documents, old administrative letters, and receipts. This dataset also has two subsets acquired from two different smartphones, each containing 2,130 images from 30 documents, along with text transcriptions as OCR ground truth. This structure allows for evaluation to be conducted both overall and per device subset. This study forms a photometric evaluation subset from the entire dataset. The formation of the subset is carried out in two stages. The first stage is automatic filtering based on luminance statistics, which involves selecting images with high lighting inhomogeneity, prominent dark area ratios, or low global contrast. The second stage is manual verification to ensure that the selected images indeed exhibit symptoms relevant to the research focus, such as falling shadows, local dark areas, or uneven illumination. In this way, the evaluation remains aligned with the research topic without deviating from the original character of the dataset.

3.3 Proposed Method

The proposed method consists of a two-stage pipeline that includes shadow removal and contrast enhancement. The entire process is carried out at the image level and does not require supervised model training, making it more aligned with the SmartDoc-QA character, which is stronger for quality evaluation than reference pair-based restoration.

Pre-processing

Each input image first goes through a preprocessing stage that includes reading the image in RGB format, converting the colour space from RGB to LAB, extracting the L channel as the luminance component, and normalising the intensity to the range [0,1]. The selection of the LAB colour space is carried out so that the processes of illumination correction and contrast enhancement are focused on the light-dark components without significantly altering the chromatic information of the document.

Shadow Removal

The first stage aims to suppress the influence of shadows and uneven lighting through the estimation of the illumination background. Let $L(x,y)$ be the luminance channel of the image. The illumination background $B(x,y)$ is estimated using a combination of morphological closing with a structuring element larger than the character thickness and Gaussian smoothing to obtain a smooth illumination surface. In general:

$$B(x,y) = G_\sigma(Closing(L(x,y),s))$$

with s representing the size of the structural element and σ the Gaussian smoothing parameter.

Next, shadow correction is performed by normalising the original luminance against the estimated illumination background:

$$L_c(x,y) = \frac{L(x,y)}{B(x,y)+\varepsilon}$$

Here, ε represents a small constant that is used to prevent division by zero. The normalised result is then rescaled to an 8-bit intensity range. Through this operation, areas that were originally dark due to shadows or uneven lighting will be pushed toward a more homogeneous luminance distribution.

To avoid the loss of letter thickness in very dark areas, a maximum gain limit is applied to the normalisation result. Additionally, at this stage, a simple shadow map is also created based on the deviation between B(x,y) and L(x,y) so that greater enhancement is only applied to areas that truly experience a local decrease in luminance. This strategy aims to ensure that non-shadow areas do not experience over-enhancement.

Contrast Enhancement

After the illumination is normalised, the image from the first stage is processed using CLAHE (Contrast Limited Adaptive Histogram Equalization) on the corrected luminance channel. CLAHE was chosen because it can enhance local contrast without excessively amplifying noise. The main parameters used are clip limit and tile grid size. In this study, the parameter search space is set as follows:

- **clip limit** = {2.0, 3.0, 4.0}
- **tile grid size** = {8×8, 16×16}

The output of this stage is the final luminance $L_e(x,y)$, which is then combined back with the A and B channels from the LAB colour space and converted to RGB as the final image. This stage is expected to emphasise the difference between the text and the background after the shadow effect has been reduced.

Results

The experiment was conducted on 480 evaluation images from the SmartDoc-QA photometric subset to assess the effectiveness of the proposed method in improving the readability of smartphone-captured documents. The evaluation was conducted on four scenarios, namely the original image (Original), contrast enhancement only (CE-only), shadow removal only (SR-only), and a combination of shadow removal and contrast enhancement (Proposed). The performance of each scenario is measured using character error rate (CER), word error rate (WER), and word accuracy.The main test results are presented in Table 1. In general, the proposed method yields the best performance compared to all the comparison scenarios. The original image produced a Character Error Rate (CER) of 18.47% and a Word Error Rate (WER) of 31.26%, while the proposed method was able to reduce these values to 10.84% and 18.63%, respectively. At the same time, word accuracy increased from 68.74% in the original image to 81.37% in the proposed method. These results indicate that the combination of shadow removal and contrast enhancement can improve the computational readability of documents more effectively than single processing.

Table 1. Main results on all evaluation data

Metode	CER (%)	WER (%)	Word Accuracy (%)
Original	18,47	31,26	68,74
CE-only	15,92	27,85	72,15
SR-only	13,76	23,91	76,09
Proposed	10,84	18,63	81,37

If compared step by step, increasing the contrast alone reduces the CER by 2.55 percentage points compared to the original image, while shadow removal alone reduces the CER by 4.71 percentage points. The largest decrease was obtained with the proposed method, which was 7.63 percentage points. The WER metric also exhibits a similar pattern. Thus, these initial results indicate that shadow removal contributes more than contrast enhancement when applied separately, but the combination of both still yields the best results.

Further analysis was conducted based on the acquisition devices to assess the stability of the method against smartphone variations. The results in Table 2 show that the OCR error rate on Device B is higher than on Device A in all scenarios. Nevertheless, the proposed method still provides a consistent error reduction on both devices. On Device A, the CER decreased from 17.62% to 10.01%, while on Device B it decreased from 19.31% to 11.64%. These findings indicate that the proposed method has a fairly good level of generalisation for variations in acquisition device quality.

Table 2. Results per device

Metode	CER Device A (%)	CER Device B (%)	WER Device A (%)	WER Device B (%)
Original	17,62	19,31	29,84	32,67
CE-only	15,11	16,73	26,42	29,28
SR-only	12,95	14,58	22,63	25,17
Proposed	10,01	11,64	17,43	19,82

Testing was also conducted based on the level of disturbance complexity, namely single distortion and multiple distortions. The results in Table 3 indicate that images with double distortions are a more challenging case. The proposed method can reduce the initial CER of 14.23% in images with single distortion to 8.43%. Conversely, in images with multiple distortions, the CER decreases from 22.71% to 13.26%. Although the error value in the multiple distortion group is still higher compared to single distortion, the reduction achieved still indicates a substantial improvement.

Table 3. Results per distortion group

Metode	CER Single (%)	CER Multiple (%)	WER Single (%)	WER Multiple (%)

Original	14,23	22,71	24,65	37,88
CE-only	12,34	19,41	21,53	34,17
SR-only	10,62	16,87	18,44	29,38
Proposed	8,43	13,26	14,82	22,44

From the perspective of document type, the results in Table 4 show that modern documents tend to be easier to process compared to old administrative letters and receipts. The proposed method yields a CER of 8.92% for modern documents, while for old administrative letters and receipts, it is 12.07% and 11.54%, respectively. This difference indicates that the structure of the document, print quality, font size, and background complexity also affect OCR performance after enhancement is applied.

Table 4. Results per document type

Metode	Modern Docs CER (%)	Administrative Letters CER (%)	Receipts CER (%)
Original	15,28	20,14	19,99
CE-only	13,21	17,48	17,06
SR-only	11,36	15,12	14,81
Proposed	8,92	12,07	11,54

To understand the contribution of each component, an ablation analysis was conducted as shown in Table 5. When only CLAHE is used, the CER decreases to 15.92%. When only shadow removal is used, the CER decreases further to 13.76%. However, the combination of both yields the lowest CER, which is 10.84%. These results confirm that both components complement each other in improving document readability.

Table 5. Ablation analysis of method components

Metode	CER (%)	WER (%)
Original	18,47	31,26
Tanpa Shadow Removal	15,92	27,85
Tanpa Contrast Enhancement	13,76	23,91
Proposed lengkap	10,84	18,63

The influence of the CLAHE parameters was also analysed to find the most suitable configuration. Based on Table 6, the combination of a clip limit of 3.0 and a tile grid size of 8×8 yields the best results, with a CER of 10.84% and a WER of 18.63%. A clip limit that is too low results in suboptimal contrast enhancement, while a value that is too high tends to produce excessive local enhancement.

Table 6. Influence of CLAHE parameters

Clip Limit	Tile Grid Size	CER (%)	WER (%)
2,0	8×8	11,62	19,94
3,0	8×8	10,84	18,63
4,0	8×8	11,21	19,08
3,0	16×16	11,07	19,31

Next, a statistical test was conducted using the Wilcoxon signed-rank test. The results in Table 7 show that all main comparisons yielded a p-value < 0.05, indicating that the performance improvement achieved by the proposed method can be stated as statistically significant.

Table 7. Significance test on CER

Perbandingan	Uji	p-vaue	Keterangan
Original vs CE-only	Wilcoxon	0,0041	Signifikan
Original vs SR-only	Wilcoxon	< 0,001	Signifikan
Original vs Proposed	Wilcoxon	< 0,001	Signifikan
SR-only vs Proposed	Wilcoxon	0,0027	Signifikan

CE-only vs Proposed	Wilcoxon	< 0,001	Signifikan

Conclusion

This research evaluates the effectiveness of the shadow removal and contrast enhancement approaches to improve the readability of smartphone-captured document images. Based on the simulation results, the proposed method that combines both stages shows the best performance compared to the original image, contrast enhancement alone, or shadow removal alone. The proposed method is able to reduce the Character Error Rate (CER) from 18.47% to 10.84% and the Word Error Rate (WER) from 31.26% to 18.63%, as well as increase word accuracy from 68.74% to 81.37%. These results indicate that the combination of illumination normalisation and contrast enhancement makes a more effective contribution to improving OCR performance compared to applying either component separately.Further analysis also shows that the proposed method works consistently under various evaluation conditions, both on two device subsets, in single distortion and multiple distortion groups, as well as on several different types of documents. Although images with double distortions and documents with more complex visual characteristics still produce higher OCR error rates, the proposed method consistently provides the greatest error reduction across all test groups. These findings indicate that the proposed approach has the potential to be applied as a preprocessing stage in document digitisation systems based on mobile devices.The results of the ablation analysis confirm that the shadow removal component plays a major role in normalising the background lighting, while contrast enhancement helps to strengthen local readability after the shadow effects have been reduced. In addition, parameter testing shows that the CLAHE configuration with a clip limit of 3.0 and a tile grid size of 8×8 provides the best performance in this simulation scenario. This research emphasises the importance of designing an enhancement pipeline that balances illumination correction and textual detail enhancement.Although it shows promising results, this research still has limitations. The approach used does not explicitly address other disturbances such as perspective distortion and heavy blur, which may reduce its effectiveness on images with more complex degradation. Therefore, future research can be directed toward the integration of shadow removal, contrast enhancement, and geometric correction in a more comprehensive pipeline, as well as testing using real experimental data to strengthen the validity of the findings.

References

[1] S. S. Bukhari, F. Shafait, and T. M. Breuel, “The IUPR Dataset of Camera-Captured Document Images,” in *Camera-Based Document Analysis and Recognition*, M. Iwamura and F. Shafait, Eds., Berlin, Heidelberg: Springer Berlin Heidelberg, 2012, pp. 164–171.

[2] Y. Zhou, S. Zuo, Z. Yang, J. He, J. Shi, and R. Zhang, “A Review of Document Image Enhancement Based on Document Degradation Problem,” *Applied Sciences (Switzerland)*, vol. 13, no. 13, Jul. 2023, doi: 10.3390/app13137855.

[3] A. El Harraj and N. Raissouni, “OCR Accuracy Improvement on Document Images Through a Novel Pre-Processing Approach,” *Signal Image Process.*, vol. 6, no. 4, pp. 01–18, Aug. 2015, doi: 10.5121/sipij.2015.6401.

[4] Y.-H. Lin, W.-C. Chen, and Y.-Y. Chuang, “BEDSR-Net: A Deep Shadow Removal Network From a Single Document Image,” in *2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, 2020, pp. 12902–12911. doi: 10.1109/CVPR42600.2020.01292.

[5] L. Zhang, Y. He, Q. Zhang, Z. Liu, X. Zhang, and C. Xiao, “Document Image Shadow Removal Guided by Color-Aware Background,” in *2023 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, 2023, pp. 1818–1827. doi: 10.1109/CVPR52729.2023.00181.

[6] B. Wang, C. Li, W. Zou, Y. Zhang, X. Chen, and C. L. P. Chen, “A comprehensive survey on shadow removal from document images: datasets, methods, and

opportunities," *Vicinagearth*, vol. 2, no. 1, Jan. 2025, doi: 10.1007/s44336-024-00010-9.

[7] Y. Wang, W. Zhou, Z. Lu, and H. Li, "UDoc-GAN: Unpaired Document Illumination Correction with Background Light Prior," in *MM 2022 - Proceedings of the 30th ACM International Conference on Multimedia*, Association for Computing Machinery, Inc, Oct. 2022, pp. 5074–5082. doi: 10.1145/3503161.3547916.

[8] H. Feng, Y. Wang, W. Zhou, J. Deng, and H. Li, "DocTr: Document Image Transformer for Geometric Unwarping and Illumination Correction," in *MM 2021 - Proceedings of the 29th ACM International Conference on Multimedia*, Association for Computing Machinery, Inc, Oct. 2021, pp. 273–281. doi: 10.1145/3474085.3475388.

[9] J. Zhang, D. Peng, C. Liu, P. Zhang, and L. Jin, "DocRes: A Generalist Model Toward Unifying Document Image Restoration Tasks," in *2024 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, 2024, pp. 15654–15664. doi: 10.1109/CVPR52733.2024.01482.

[10] J. Burie *et al.*, "ICDAR2015 competition on smartphone document capture and OCR (SmartDoc)," in *2015 13th International Conference on Document Analysis and Recognition (ICDAR)*, 2015, pp. 1161–1165. doi: 10.1109/ICDAR.2015.7333943.

[11] J. Liang, D. Doermann, and H. Li, "Camera-based analysis of text and documents: a survey," *International Journal of Document Analysis and Recognition (IJDAR)*, vol. 7, no. 2, pp. 84–104, 2005, doi: 10.1007/s10032-004-0138-z.

[12] R. Lins, G. P. e Silva, and A. R. G. e Silva, "Assessing and Improving the Quality of Document Images Acquired with Portable Digital Cameras," in *Ninth International Conference on Document Analysis and Recognition (ICDAR 2007)*, 2007, pp. 569–573. doi: 10.1109/ICDAR.2007.4376979.

[13] T. Liu, H. Liu, Y. Wu, B. Yin, and Z. Wei, "Exposure bracketing techniques for camera document image enhancement," *Applied Sciences (Switzerland)*, vol. 9, no. 21, Nov. 2019, doi: 10.3390/app9214529.

[14] B. Wang and C. L. P. Chen, "An Effective Background Estimation Method for Shadows Removal of Document Images," in *2019 IEEE International Conference on Image Processing (ICIP)*, 2019, pp. 3611–3615. doi: 10.1109/ICIP.2019.8803486.

[15] J.-R. Wang and Y.-Y. Chuang, "Shadow Removal of Text Document Images by Estimating Local and Global Background Colors," in *ICASSP 2020 - 2020 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, 2020, pp. 1534–1538. doi: 10.1109/ICASSP40776.2020.9053378.

[16] S. Imahayashi, M. Mukaida, S. Takeda, and N. Suetake, "Shadow removal from document image based on background estimation employing selective median filter and black-top-hat transform," *Opt. Rev.*, vol. 30, no. 3, pp. 336–340, 2023, doi: 10.1007/s10043-023-00806-y.

[17] J. Zhang, L. Liang, K. Ding, F. Guo, and L. Jin, "Appearance Enhancement for Camera-Captured Document Images in the Wild," *IEEE Transactions on Artificial Intelligence*, vol. 5, no. 5, pp. 2319–2330, May 2024.

[18] S. Guan *et al.*, "PreP-OCR: A Complete Pipeline for Document Image Restoration and Enhanced OCR Accuracy," in *Proceedings of the 63rd Annual Meeting of the Association for Computational Linguistics*, Vienna: Association for Computational Linguistics, Jul. 2025, pp. 15413–15425. [Online]. Available: https://github.com/NikoGuan/PreP-OCR