Restoration-Guided Kuzushiji Character Recognition Framework under Seal Interference

Rui-Yang Ju[1], Kohei Yamashita[2], Hirotaka Kameko[2], and Shinsuke Mori[2]

Graduate School of Informatics, Kyoto University, Sakyo-ku, Kyoto, 606-8501, Japan
jryjry1094791442@gmail.com
Academic Center for Computing and Media Studies, Kyoto University, Sakyo-ku, Kyoto, 606-8501, Japan
yamashita.kohei.8j@kyoto-u.ac.jp, {kameko,forest}@i.kyoto-u.ac.jp

Abstract. Kuzushiji was one of the most popular writing styles in pre-modern Japan and was widely used in both personal letters and official documents. However, due to its highly cursive forms and extensive glyph variations, most modern Japanese readers cannot directly interpret Kuzushiji characters. Therefore, recent research has focused on developing automated Kuzushiji character recognition methods, which have achieved satisfactory performance on relatively clean Kuzushiji document images. However, existing methods struggle to maintain recognition accuracy under seal interference (e.g., when seals overlap characters), despite the frequent occurrence of seals in pre-modern Japanese documents. To address this challenge, we propose a three-stage restoration-guided Kuzushiji character recognition (RG-KCR) framework specifically designed to mitigate seal interference. We construct datasets for evaluating Kuzushiji character detection (Stage 1) and classification (Stage 3). Experimental results show that the YOLOv12-medium model achieves a precision of 98.0% and a recall of 93.3% on the constructed test set. We quantitatively evaluate the restoration performance of Stage 2 using PSNR and SSIM. In addition, we conduct an ablation study to demonstrate that Stage 2 improves the Top-1 accuracy of Metom, a Vision Transformer (ViT)-based Kuzushiji classifier employed in Stage 3, from 93.45% to 95.33%. The implementation code of this work is available at `https://ruiyangju.github.io/RG-KCR` .

Keywords: Kuzushiji characters · pre-modern Japanese documents · object detection · document restoration · character classification · seals

1 Introduction

Kuzushiji[1] is a traditional cursive writing style widely found in pre-modern Japanese materials, including historical records, narrative texts, and classical po-

[1] Kuzushiji (くずし字) refers to historical cursive character forms used in pre-modern Japanese (before the Meiji Restoration) documents.

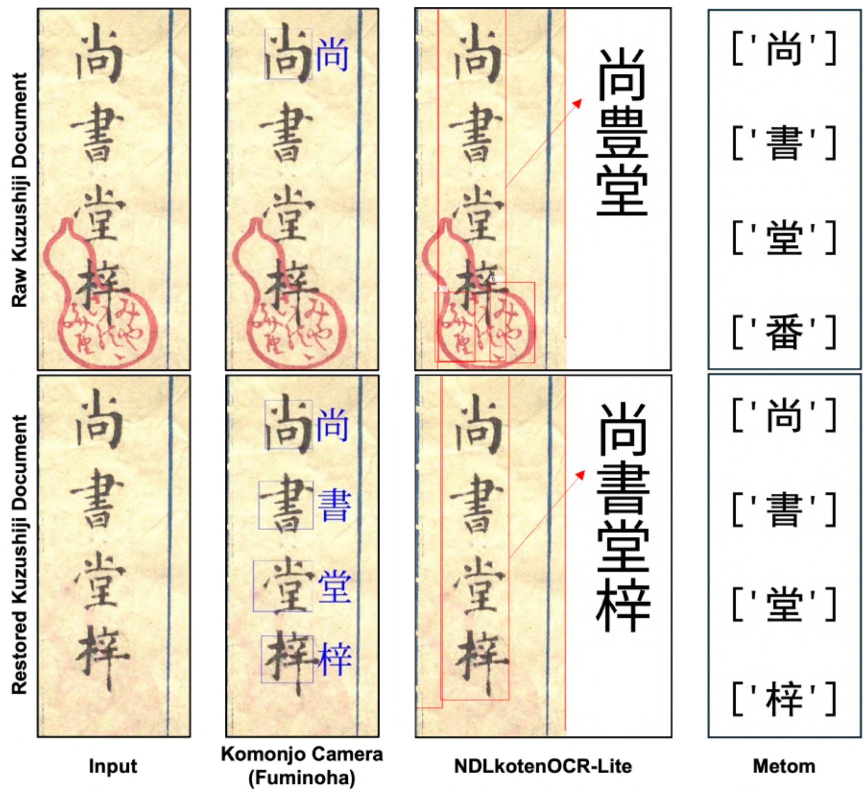

Fig. 1. Example recognition results under seal interference, where seals overlap with Kuzushiji characters. The left column shows the input document region containing the Kuzushiji characters "尚書堂梓", and the right column shows the corresponding recognition outputs produced by three systems: Fuminoha [31], NDLkotenOCR-Lite [1], and Metom [13]. The top row uses the raw document as input, while the bottom row reports the recognition results after applying our document restoration method.

etry such as waka[2], as well as prose written in literary styles such as sōrōbun[3] [25]. Although Kuzushiji is based on the same underlying character system as modern Japanese (i.e., kanji[4] and kana[5]), its glyphs have undergone extensive simplification and stylistic transformation, resulting in highly fluid and compact forms that are difficult to recognize. Moreover, Kuzushiji covers diverse cursive forms of kanji and kana (i.e., hiragana and katakana), as well as numerous variant forms. Before the Meiji Restoration, Kuzushiji was widely used in Japanese society.

[2] Waka (和歌) refers to a type of classical Japanese poetry.
[3] Sōrōbun (候文) refers to a classical Japanese prose style widely used in letters and official documents from the medieval period through the early Meiji era.
[4] Kanji (漢字) refers to Chinese-origin characters used in Japanese.
[5] Kana (仮名) refers to the Japanese syllabaries used to represent phonological units.

However, with the spread of modern printing and the establishment of modern schooling, its use gradually declined in daily life [10]. As a result, most contemporary Japanese readers are unable to directly interpret pre-modern documents written in Kuzushiji, and only a limited number of trained specialists can read these characters accurately.

Recent advances in machine learning (ML) have enabled neural networks to recognize Kuzushiji characters by being trained on expert-annotated data. For instance, TOPPAN Inc. introduced the "Fuminoha" program to support the interpretation and practical use of historical materials. As part of this program, the "Komonjo Camera" [31] applies high-precision AI-based OCR to recognize Kuzushiji characters in documents captured with mobile devices. Furthermore, the Center for Open Data in the Humanities (CODH) proposed Metom [13], a Vision Transformer (ViT) [8]-based Kuzushiji character classifier capable of recognizing more than one million character classes. In addition, the National Diet Library (NDL) released NDLkotenOCR-Lite [1], a lightweight OCR system designed to run without GPU acceleration. These systems represent some of the most practical solutions currently available for Kuzushiji character recognition.

Pre-modern Japanese documents frequently contain seals. As symbols of ownership and identity, seals were often stamped not only by the original creators but also by later owners and collectors. Typically rendered in red ink, they may include stylized characters indicating the owner's name, social status, or personal aspirations. From the perspective of Kuzushiji character recognition, seals introduce substantial interference, especially when they occlude or overlap Kuzushiji characters, thereby severely degrading recognition accuracy. As shown in Fig. 1, the three representative Kuzushiji recognition systems struggle to correctly recognize characters under such seal interference.

To address this challenge, we propose a restoration-guided Kuzushiji character recognition (RG-KCR) framework that mitigates seal interference. The proposed framework consists of three stages: Kuzushiji character detection (Stage 1), Kuzushiji document restoration (Stage 2), and Kuzushiji character classification (Stage 3). The main contributions of this work are summarized as follows:

(a) We propose a novel restoration-guided Kuzushiji character recognition (RG-KCR) framework that improves recognition performance when seals overlap with Kuzushiji characters.
(b) To mitigate red-seal interference, we introduce a training-free document restoration algorithm that is computationally efficient and reduces seal artifacts in occluded regions.
(c) We conduct an ablation study to demonstrate the impact of the proposed document restoration stage on Kuzushiji character classification performance.
(d) We construct a Kuzushiji character detection dataset comprising 1,000 document images by synthetically overlaying pre-modern real seal images onto Kuzushiji documents sourced from the CODH, and we manually review and correct the annotations to ensure high reliability.
(e) We further construct a Kuzushiji character classification test set consisting of 100 Kuzushiji document images with synthetically overlaid seals, containing 17,982 character instances to evaluate recognition performance.

2 Related Work

2.1 Kuzushiji Character Recognition

Kuzushiji character recognition has attracted increasing attention in Japan, and several institutions have released practical systems for reading pre-modern Japanese documents. Among them, the CODH developed the mobile application "miwo" [5] and the web-based platform "KuroNet" [7, 16] for full-page Kuzushiji document recognition. Both systems first perform pixel-wise character segmentation using U-Net [27]-based semantic segmentation [6], and then classify the segmented character regions to obtain recognition results. In addition, the CODH introduced "Metom", a ViT [8]-based model for single-character Kuzushiji recognition that is accessible via a web interface.

TOPPAN Inc. has been actively developing Kuzushiji recognition technologies since 2015. In 2023, the "Fuminoha" project released a mobile application, "Komonjo Camera" [31], which integrates two recognition engines: "Komonjo AI" and "Kotenseki AI". "Komonjo AI" targets documents containing many kanji characters with variable stroke widths, while "Kotenseki AI" is optimized for documents that are mainly kana with relatively uniform stroke width. These engines are trained on diverse handwriting styles and character variants to improve recognition robustness.

The NDL developed NDLkotenOCR [22], an OCR system for historical materials (e.g., Japanese documents from the Edo period), in 2022 and subsequently released updated versions in 2023 and 2024. NDLkotenOCR employs Cascade Mask R-CNN [3] for layout analysis and builds upon the TrOCR framework [17] with a RoBERTa-small decoder [20] to recognize Kuzushiji character sequences. Because NDLkotenOCR requires GPU acceleration at inference time, the NDL further introduced a lightweight variant, NDLkotenOCR-Lite [1], to enable efficient processing in CPU-only environments (e.g., web-based platforms).

Despite their strong performance on relatively clean Kuzushiji documents and non-occluded characters, these methods often have to process severely degraded and occluded documents (e.g., when seals overlap Kuzushiji characters), which leads to unsatisfactory recognition accuracy.

2.2 Object Detection

Object detection has been widely studied in computer vision (CV) and is applicable to many object categories, including character instances in historical documents. The rise of convolutional neural networks (CNNs) facilitated the development of two-stage detectors, such as R-CNN [9], Mask R-CNN [11], and Cascade R-CNN [3], which established strong baselines in practice. Subsequently, one-stage detectors, including You Only Look Once (YOLO) [26] and the Single Shot Detector (SSD) [19], substantially reduced computational cost and enabled real-time detection. With the success of Transformer-based architectures in CV, the DEtection TRansformer (DETR) [4] was introduced, formulating object detection as a set prediction problem and attracting considerable research interest.

Despite continuous improvements in state-of-the-art (SOTA) performance on the COCO benchmark [18], existing Kuzushiji character detection methods still mainly rely on conventional approaches. For instance, NDLkotenOCR [22] is based on Cascade Mask R-CNN [3], while more recent one-stage and Transformer-based detectors have yet to be systematically evaluated for this task.

2.3 Document Restoration

Document restoration aims to improve the readability of degraded document images and to support downstream document analysis tasks such as optical character recognition (OCR). It encompasses several complementary subtasks, including: (a) geometric restoration (e.g., document dewarping) to correct distortions caused by page curvature, camera viewpoint, or scanning; (b) illumination and shadow correction (e.g., deshadowing) to mitigate uneven illumination, cast shadows, and background intensity variations; (c) document enhancement (e.g., denoising and deblurring) to suppress noise and recover clear text structures; (d) document binarization to separate foreground text from background regions and produce binary images for OCR and layout analysis; and (e) inpainting-based restoration (e.g., seal removal) to reconstruct occluded or damaged regions.

With the rise of large-scale neural networks, document restoration has increasingly shifted from task-specific pipelines toward unified multi-task frameworks. Recent methods [34, 35] have reported strong performance across multiple restoration tasks by jointly modeling diverse degradation types. However, these approaches typically rely on large-scale neural networks at inference time, leading to substantially higher inference cost and latency than conventional image processing techniques. Such efficiency limitations can hinder deployment in latency-sensitive and resource-constrained applications, including practical Kuzushiji character recognition systems.

2.4 Image Classification

Image classification is a fundamental task in CV that assigns an input image to a predefined category. CNNs have achieved remarkable success in image classification, with representative architectures such as ResNet [12] and EfficientNet [28] remaining widely used in practical systems. More recently, the ViT [8] introduced Transformer-based modeling to image classification and demonstrated strong global representations. Based on this, hybrid designs that combine the local inductive biases of CNNs with the long-range dependency modeling of Transformers have been actively explored. Among them, the Swin Transformer [21] has achieved SOTA performance across multiple benchmarks. In Kuzushiji character recognition, character classification is a subtask of image classification: each character instance is assigned to a character class, which can be mapped to a Unicode code point when applicable, enabling automated transcription into modern Japanese.

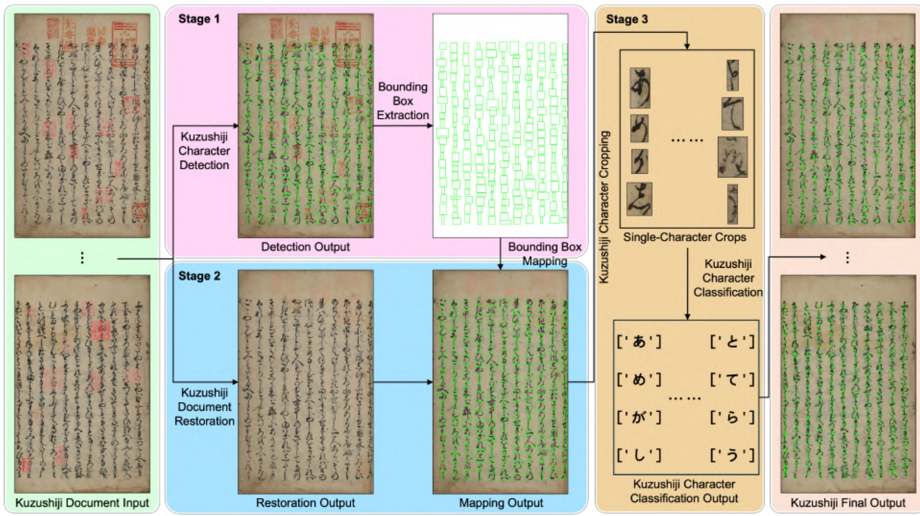

Fig. 2. The pipeline of the proposed RG-KCR framework, consisting of three stages: Kuzushiji character detection (Stage 1), Kuzushiji document restoration (Stage 2), and Kuzushiji character classification (Stage 3).

3 Proposed Method

3.1 Overall Framework

To mitigate seal interference in Kuzushiji character recognition, we propose a restoration-guided framework consisting of three stages: Kuzushiji character detection (Stage 1), Kuzushiji document restoration (Stage 2), and Kuzushiji character classification (Stage 3). The pipeline of the proposed RG-KCR framework is shown in Fig. 2. As reported in Section 4.4, our employed detector (i.e., YOLOv12-medium [30]) achieves 98% precision even under seal interference; therefore, we do not apply document restoration prior to detection stage (Stage 1). In contrast, Kuzushiji character classification is considerably more sensitive to seal interference, with recognition performance decreasing substantially, as shown in Section 4.5 Based on this, we apply document restoration (Stage 2) before the classification stage (Stage 3) to mitigate seal interference during recognition. Finally, the predicted label (i.e., a Unicode code point) for each Kuzushiji character is mapped to its corresponding modern Japanese character. These characters are then overlaid onto the restored document image at the corresponding bounding box locations, thereby producing the final output of the proposed RG-KCR framework. By overlaying the recognized modern Japanese characters back onto the original document layout, the framework enables intuitive reading and interpretation of the document content. Furthermore, our implementation allows users to adjust the color and size of the modern Japanese characters, as well as the visibility of bounding boxes, to improve readability and support interactive visualization.

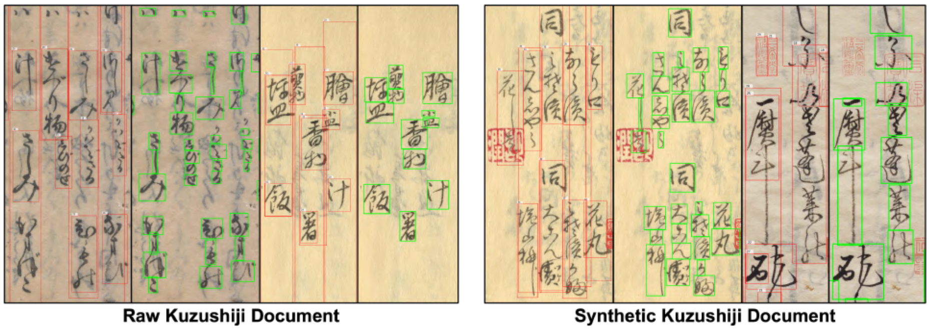

Fig. 3. Qualitative comparison of line-level and character-level Kuzushiji detection. Red bounding boxes show line-level detections produced by NDLkotenOCR-Lite [1], while green bounding boxes show character-level detections produced by the YOLOv12-medium [30] detector adopted in this work.

3.2 Kuzushiji Character Detection

Kuzushiji character detection can be categorized into character-level and line-level approaches. The CODH dataset [24] provides character-level annotations, while the NDL dataset [23] adopts line-level annotations. As shown in Fig. 3, for both raw and seal-overlaid synthetic documents, line-level systems (e.g., NDLkotenOCR-Lite) can produce duplicate bounding boxes on images with complex layouts, especially when the reading order differs from the conventional top-to-bottom, right-to-left order. This leads to multiple red bounding boxes covering the same character regions. Therefore, we adopt character-level detection as Stage 1 of the proposed RG-KCR framework and employ a SOTA detector to produce accurate bounding boxes for each Kuzushiji character.

3.3 Kuzushiji Document Restoration

To mitigate interference from red seals, we propose a training-free and efficient color-based seal removal method. Given an input document image $I \in \mathbb{R}^{H \times W \times 3}$, let R, G, and B denote the RGB channel intensities of a pixel. Because red seals typically present a higher red-channel intensity than the green and blue channels, we detect seal regions using a thresholding rule based on channel ratios. Specifically, a pixel is classified as a red-seal candidate if it satisfies:

$$(R \geq \tau_r) \land (R \geq \tau_{rg} \cdot G) \land (R \geq \tau_{rb} \cdot B), \tag{1}$$

where τ_r is the minimum red-channel intensity threshold, and (τ_{rg}, τ_{rb}) control the dominance of the red channel relative to the green and blue channels, respectively. This criterion reduces false positives from non-seal reddish regions while retaining the majority of seal pixels. The resulting binary mask M can be optionally refined using morphological dilation to expand the mask boundaries

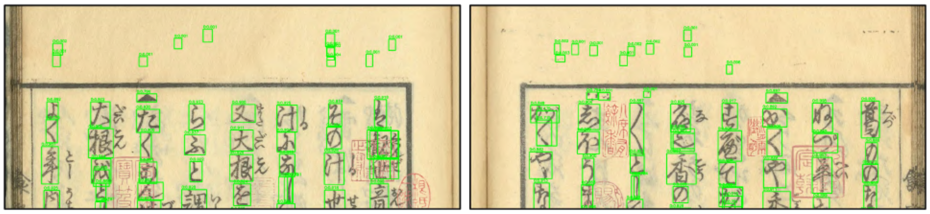

Fig. 4. Examples of low-confidence bounding boxes produced by the detection model. Stains in pre-modern documents can cause false positives, where background artifacts are mistakenly detected as Kuzushiji characters with confidence scores < 0.01.

and compensate for color bleeding from seal ink into surrounding regions. We then remove the detected seal areas via image inpainting:

$$I_{\text{restored}} = \text{Inpaint}(I, M, \rho), \tag{2}$$

where ρ denotes the inpainting radius. We use OpenCV implementations of Telea's fast marching method [29] and the Navier-Stokes-based approach [2], which propagate surrounding texture and structural cues into masked regions to produce visually plausible reconstructions.

3.4 Kuzushiji Character Cropping

To extract individual Kuzushiji character instances for the subsequent single-character classification using Metom [13], we crop character regions based on the bounding boxes predicted in Stage 1. Specifically, we obtain the predicted bounding box parameters from Stage 1 outputs, including the top-left coordinates (x, y) and the width w and height h, and use them to crop patches from the restored document image produced in Stage 2. Because Stage 2 preserves the image geometry, the restored image shares the same coordinate system as the input, allowing the same bounding boxes to be used for cropping. We then crop the corresponding character patches from the restored document. In documents with background stains or noise, the detector may produce false positives with extremely low confidence (e.g., confidence scores < 0.01), as shown in Fig. 4. To ensure reliable cropping, we filter detections using a confidence threshold of 0.5 (i.e., we keep only bounding boxes with confidence scores > 0.5).

3.5 Kuzushiji Character Classification

Kuzushiji character classification can be categorized into two approaches corresponding to the detection strategies introduced in Stage 1. For line-level detection methods, the outputs typically consist of character sequences, which require sequence recognition models for transcription. For instance, NDLkotenOCR [22] adopts the TrOCR [17] to perform end-to-end recognition of entire text lines. In contrast, character-level detection methods produce cropped images of individual characters. This setting can be formulated as an image classification task,

Table 1. Details of the constructed dataset from 13 distinct Kuzushiji books, where #Images denotes the number of page images for each book.

Index	NIJL ID	Book Title	#Images
1	100241706	Usonarubeshi (虚南留別志)	67
2	100249376	Gozenkashi Hiden-shou (御前菓子秘伝抄)	104
3	100249416	Mochigashi Sokuseki Teseishuu (餅菓子即席手製集)	58
4	100249476	Meshi Hyakuchin Den (飯百珍伝)	46
5	200006663	Diguchi (ぢぐち)	8
6	200015843	Nippon Eitaigura (日本永代蔵)	180
7	200017458	Soga Monogatari (曾我物語)	78
8	200020019	Chikusai (竹斎)	146
9	200021086	Isoho Monogatari (伊曾保物語)	60
10	200021763	Zenbu Ryouri-shou (膳部料理抄)	94
11	200021802	Ryouri Monogatari (料理物語)	105
12	200021869	Ryourikata Kokoroenokoto (料理方心得之事)	30
13	200022050	Ryouri Hiden-shou (料理秘伝抄)	24

Table 2. Matched bounding box pairs between the CODH [24] ground-truth annotations and YOLOv12-medium [30] predictions. A pair is considered matched if the Intersection-over-Union (IoU) is at least 0.5 (IoU $\geq$ 0.5).

Test Set #Images	Total Ground-Truth Bounding Boxes	Total Predicted Bounding Boxes	Total Matched Pairs (IoU $\geq$ 0.5)
100	19,035	18,656	17,982

where the class labels are Unicode code points, and each character image is assigned to its corresponding label. The predicted labels can then be mapped to modern Japanese characters when applicable. For instance, Metom [13] follows this approach to achieve single-character Kuzushiji recognition.

4 Experiments

4.1 Datasets

The raw Kuzushiji document dataset [24] is provided by the CODH, with source materials collected and preserved by the National Institute of Japanese Literature (NIJL). Considering the significant variations in writing styles, page layouts, and degrees of degradation across different books, we select 13 representative books from the CODH collection, whose bibliographic information is summarized in Table 1. We further exclude page images that contain no Kuzushiji characters, resulting in a dataset of 1,000 page images.

During annotation verification for Kuzushiji character detection, we found that 267 of the 1,000 images had incomplete bounding box annotations. To improve dataset quality and reliability, we review all images and manually add the missing bounding boxes for Kuzushiji characters. As shown in Fig. 5, the red bounding boxes denote our supplementary annotations, while the green bounding boxes indicate the original annotations.

Fig. 5. Examples of incomplete annotations in the raw dataset. Red boxes indicate our corrected annotations, and green boxes indicate the original annotations.

Fig. 6. Examples comparing raw (real) Kuzushiji documents and seal-overlaid synthetic documents. Seals in the first row are present in the original documents, while seals in the second row are synthetically overlaid.

Because the raw 1,000 images contain relatively few seals and thus do not sufficiently capture real-world seal interference, we follow prior work [15, 34] and incorporate high-quality seal images for synthetic data generation. As shown in Fig. 6, we adopt the synthesis procedure of Ju *et al.* [15], using the same parameter settings, to randomly overlay real seal images onto document images. Specifically, we overlay 10 seal instances per image, allowing at most two seals to overlap, to simulate seal interference patterns observed in real documents.

For Kuzushiji character classification, the CODH annotations [24] provide a Unicode code point for each character instance. We use the Unicode code point as the ground-truth label for classification, and map it to a modern Japanese character for visualization when applicable. In total, the test split contains 100 document images with 19,035 character instances. To decouple Stage 3 evaluation from Stage 1 localization errors, we match YOLOv12-medium predictions (as shown in Section 4.4) to the ground-truth boxes using an Intersection-over-Union (IoU) threshold of 0.5 (i.e., IoU $\geq$ 0.5). As reported in Table 2, this process yields 17,982 matched instances with associated labels, which are used as the test set for Stage 3 (Kuzushiji character classification).

Table 3. Quantitative comparison of medium variants of different detection models for Kuzushiji character detection on the constructed test set.

Method	#Params	FLOPs	Precision	Recall	AP_{50}	$AP_{50:95}$
RT-DETR (CVPR'24) [36]	31.99M	103.4G	86.2%	79.2%	84.1%	57.0%
YOLOv9 (ECCV'24) [33]	20.01M	76.5G	97.7%	93.0%	96.6%	80.9%
YOLOv10 (NeurIPS'24) [32]	**15.31M**	**58.9G**	97.7%	92.3%	96.5%	80.5%
YOLO11 (Ultralytics'24) [14]	20.03M	67.6G	**98.1%**	93.3%	96.7%	81.7%
YOLOv12 (NeurIPS'25) [30]	20.11M	67.1G	98.0%	**93.9%**	**97.0%**	**82.3%**

Table 4. Parameter sensitivity analysis of τ_r and (τ_{rg}, τ_{rb}) for Kuzushiji document restoration, evaluated using PSNR and SSIM on the constructed dataset.

τ_r	(τ_{rg}, τ_{rb})	$PSNR_{Valid}$	$SSIM_{Valid}$	$PSNR_{Test}$	$SSIM_{Test}$
–	–	29.15dB	0.9655	28.71dB	0.9639
80	(1.2, 1.2)	29.76dB	0.9470	29.61dB	0.9465
80	(1.3, 1.3)	33.64dB	0.9736	33.73dB	0.9731
80	(1.4, 1.4)	33.87dB	0.9756	33.77dB	0.9745
80	(1.5, 1.5)	31.97dB	0.9717	31.68dB	0.9706
90	(1.2, 1.2)	30.37dB	0.9522	30.19dB	0.9619
90	(1.3, 1.3)	**34.09dB**	0.9757	**34.13dB**	0.9750
90	(1.4, 1.4)	34.05dB	**0.9763**	33.94dB	**0.9753**
90	(1.5, 1.5)	32.03dB	0.9721	31.74dB	0.9710

4.2 Evaluation Metrics

For Stage 1 (Kuzushiji character detection), we employ standard object detection metrics [14, 30, 32, 33], including the number of model parameters (#Params), floating-point operations (FLOPs), precision, recall, average precision at IoU= 0.50 (AP_{50}), and mean average precision averaged over IoU thresholds from 0.50 to 0.95 ($AP_{50:95}$). For Stage 2 (Kuzushiji document restoration), we quantitatively assess restoration quality using image fidelity metrics commonly adopted in document restoration research [34, 35], namely the peak signal-to-noise ratio (PSNR, in dB) and the structural similarity index (SSIM). For Stage 3 (Kuzushiji character classification), we adopt standard image classification metrics [12, 28], including Top-1 and Top-5 accuracy (reported as percentages).

4.3 Implementation Details

All experiments are conducted on Ubuntu, using an Intel Core i7-14700K CPU and an NVIDIA RTX A6000 GPU. For Stage 1, we train all detection models using their medium variants and initialize them with COCO-pretrained weights [18]. All detectors are implemented using the Ultralytics [14] framework, with the SGD optimizer and an initial learning rate of 0.01. All models are trained for 100 epochs on the training set with a batch size of 16 and an input size of 640×640 pixels. For Stage 2, we evaluate various combinations of the restoration hyperparameters τ_r and (τ_{rg}, τ_{rb}) and report the parameter sensitivity results in

Table 5. Ablation study results on the impact of Stage 2 (Kuzushiji document restoration) on Stage 3 (Kuzushiji character classification).

Method	Stage 2 (Restoration)	Restoration Time	Top-1 Accuracy	Top-5 Accuracy
Metom [13]	–	–	93.45%	97.46%
Metom [13]	✓	0.51 s/image	95.33%	98.62%

Fig. 7. Qualitative detection results produced by YOLOv12-medium [30] on Kuzushiji document images from the constructed dataset.

Section 4.4. Throughout our experiments, we set $\tau_{rg} = \tau_{rb}$ in our experiments. For seal-mask refinement, we apply morphological dilation to the binary mask using a square structuring element of size $k \times k$ with t iterations to compensate for color bleeding around seal boundaries. By default, we set $k = 3$ and $t = 1$ for all experiments. The inpainting radius ρ, which controls the spatial propagation range of surrounding textures, is fixed at $\rho = 3$ pixels to balance artifact removal and stroke preservation. For Stage 3, we directly use the pretrained Metom model released on Hugging Face [13] to classify and evaluate the cropped Kuzushiji character instances.

4.4 Quantitative Results

Kuzushiji Character Detection Considering computational constraints in mobile and web deployment scenarios, we evaluate several recent YOLO variants, including YOLOv9 [33], YOLOv10 [32], YOLO11 [14], and YOLOv12 [30], together with the lightweight Detection Transformer variant RT-DETR [36], on the constructed dataset. All models are trained, validated, and tested under

Fig. 8. Qualitative comparison of restoration results between raw and restored Kuzushiji documents with hyperparameters $\tau_r = 90$ and $(\tau_{rg}, \tau_{rb}) = 1.3$.

identical settings, and the quantitative results are reported in Table 3. Although YOLOv10 has the smallest parameter count and the lowest FLOPs, its overall detection performance is slightly lower than the other YOLO variants on this dataset. In contrast, YOLOv12 achieves the best overall performance, attaining 98.0% precision (0.1 percentage point below the highest value), 93.9% recall (the highest value), 97.0% AP_{50} (the highest value), and 82.3% $\mathrm{AP}_{50:95}$ (the highest value). Therefore, we adopt YOLOv12-medium as the detector for Stage 1 of the proposed framework.

Kuzushiji Document Restoration We conduct a parameter sensitivity analysis of the key hyperparameters τ_r and (τ_{rg}, τ_{rb}) in Stage 2 (Kuzushiji document restoration). The results are summarized in Table 4. The first row reports a no-restoration baseline, where Stage 2 is not applied and the evaluation metrics are computed between the raw document images and their corresponding synthetic ones. Across both the validation and test sets, $\tau_r = 90$ consistently yields better restoration quality than $\tau_r = 80$. Moreover, when $(\tau_{rg}, \tau_{rb}) = (1.3, 1.3)$, PSNR reaches its highest values of 34.09 dB and 34.13 dB on the validation and test sets, respectively. The corresponding SSIM scores (0.9757 and 0.9750) are the second-highest, only slightly below the best results of 0.9763 and 0.9753. Considering overall performance across all metrics, we select $\tau_r = 90$ and $\tau_{rg} = \tau_{rb} = 1.3$ as the final hyperparameter configuration for Stage 2 of the proposed framework.

4.5 Ablation Study

To quantify the contribution of Stage 2 (Kuzushiji document restoration) to Stage 3 (Kuzushiji character classification) in the proposed RG-KCR framework,

Fig. 9. Examples of seal-overlaid synthetic Kuzushiji document images and the corresponding final outputs produced by the proposed RG-KCR framework.

we conduct an ablation study. We adopt Metom [13] as the character classifier given its strong single-character recognition capability across more than one million character classes. The ablation results are summarized in Table 5. On 17,982 cropped character instances, direct classification without restoration (baseline) achieves 93.45% Top-1 accuracy and 97.46% Top-5 accuracy. After incorporating Stage 2, performance improves to 95.33% (Top-1 Accuracy) and 98.62% (Top-5 Accuracy). Stage 2 introduces an additional runtime of 0.51 s per document image on average, which remains acceptable given the accuracy gains.

4.6 Qualitative Results

Kuzushiji Character Detection Fig. 7 shows qualitative detection results produced by YOLOv12-medium on the constructed test set. The evaluated documents present diverse degradations due to long-term preservation processes and storage conditions, including paper yellowing, stains, and uneven ink intensity. In addition, seals vary in size and appear at different locations across pages. Despite these challenges, YOLOv12-medium [30] accurately detects Kuzushiji characters even when seals overlap with them, demonstrating robustness to seal interference.

Kuzushiji Document Restoration Fig. 8 presents representative examples comparing input document images and the corresponding restoration results obtained with $\tau_r = 90$ and $(\tau_{rg}, \tau_{rb}) = (1.3, 1.3)$ in Stage 2. The examples are drawn from ten different books and cover diverse handwriting styles, page layouts, and degradation levels. Most obvious seal artifacts are effectively removed, with only faint pink residues remaining in some cases. As shown in the upper-right corner of the fourth image in the top row, even large seal regions can be substantially removed. After restoration, stain-induced visual artifacts are reduced, and the overall contrast and structural clarity are improved, which benefits subsequent character classification. Nevertheless, in regions with dense seal coverage, the restoration process may introduce locally overexposed areas or overly smoothed textures, potentially losing fine paper details. Although such artifacts may remain visually acceptable, they can adversely affect pixel-level fidelity metrics (e.g., PSNR and SSIM) and may also impact the preservation of local character stroke structures.

5 Discussion

Fig. 9 shows representative Kuzushiji document images and their corresponding final outputs produced by the proposed RG-KCR framework. For visualization, green bounding boxes and the recognized modern Japanese characters are overlaid on the documents with a uniform font size of 64 pixels to improve clarity. After applying the RG-KCR framework, readers can interpret the content of the original Kuzushiji documents more intuitively. Notably, the current framework facilitates comprehension by overlaying recognized modern Japanese characters on the document images, but it does not yet generate continuous, ordered Japanese text. This limitation mainly stems from the complexity of reading-order recovery in Kuzushiji documents: substantial layout variations across pages make reliable character ordering and text-line reconstruction challenging. In future work, we plan to incorporate character ordering and layout analysis to enable end-to-end text reconstruction and continuous textual output.

6 Conclusion

Kuzushiji character recognition has attracted increasing attention, as it enables contemporary Japanese readers to access and interpret pre-modern Japanese documents. However, seals frequently overlap with Kuzushiji characters and significantly decrease recognition accuracy. To the best of our knowledge, existing Kuzushiji recognition systems do not explicitly address seal interference.

To address this challenge, we propose a restoration-guided Kuzushiji character recognition (RG-KCR) framework. The framework first employs YOLOv12-medium for character-level detection. It then applies a document restoration stage to mitigate seal interference and reduce seal-induced artifacts in occluded regions. Next, the restored character crops are classified using Metom to recognize individual Kuzushiji characters. Finally, the recognized modern Japanese characters are overlaid on the restored documents for intuitive interpretation.

Acknowledgments

This work was supported by JSPS KAKENHI Grant Number 25H01242, and JST SPRING Grant Number JPMJSP2110.

References

1. Aoike, T.: Development of ndlkotenocr-lite, a lightweight ocr that runs at high speed in a cpu environment. In: IPSJ SIG Computers and the Humanities Symposium. pp. 181–186 (2024)
2. Bertalmio, M., Bertozzi, A.L., Sapiro, G.: Navier-stokes, fluid dynamics, and image and video inpainting. In: IEEE Computer Society Conference on Computer Vision and Pattern Recognition. vol. 1, pp. I–I (2001)
3. Cai, Z., Vasconcelos, N.: Cascade r-cnn: High quality object detection and instance segmentation. IEEE Transactions on Pattern Analysis and Machine Intelligence **43**(5), 1483–1498 (2019)
4. Carion, N., Massa, F., Synnaeve, G., Usunier, N., Kirillov, A., Zagoruyko, S.: End-to-end object detection with transformers. In: European Conference on Computer Vision. pp. 213–229 (2020)
5. Clanuwat, T., Kitamoto, A.: "miwo" ai kuzushiji recognition application for document examination. In: IPSJ SIG Computers and the Humanities Symposium. pp. 302–309 (2021)
6. Clanuwat, T., Lamb, A., Kitamoto, A.: End-to-end pre-modern japanese character (kuzushiji) spotting with deep learning. In: IPSJ SIG Computers and the Humanities Symposium. pp. 15–20 (2018)
7. Clanuwat, T., Lamb, A., Kitamoto, A.: Kuronet: Pre-modern japanese kuzushiji character recognition with deep learning. In: International Conference on Document Analysis and Recognition. pp. 607–614 (2019)
8. Dosovitskiy, A., Beyer, L., Kolesnikov, A., Weissenborn, D., Zhai, X., Unterthiner, T., Dehghani, M., Minderer, M., Heigold, G., Gelly, S., Uszkoreit, J., Houlsby, N.: An image is worth 16x16 words: Transformers for image recognition at scale. In: International Conference on Learning Representations (2021)
9. Girshick, R., Donahue, J., Darrell, T., Malik, J.: Rich feature hierarchies for accurate object detection and semantic segmentation. In: IEEE Conference on Computer Vision and Pattern Recognition. pp. 580–587 (2014)
10. Hashimoto, Y., Iikura, Y., Hisada, Y., Kang, S., Arisawa, T., et al.: The kuzushiji project: Developing a mobile learning application for reading early modern japanese texts. Digital Humanities Quarterly **11**(1) (2017)
11. He, K., Gkioxari, G., Dollár, P., Girshick, R.: Mask r-cnn. In: IEEE International Conference on Computer Vision. pp. 2961–2969 (2017)
12. He, K., Zhang, X., Ren, S., Sun, J.: Deep residual learning for image recognition. In: IEEE Conference on Computer Vision and Pattern Recognition. pp. 770–778 (2016)
13. Imajuku, Y., Clanuwat, T.: Metom, `https://huggingface.co/SakanaAI/Metom`
14. Jocher, G., Qiu, J., Chaurasia, A.: Ultralytics, `https://www.ultralytics.com`
15. Ju, R.Y., Yamashita, K., Kameko, H., Mori, S.: Dkds: A benchmark dataset of degraded kuzushiji documents with seals for detection and binarization. arXiv preprint arXiv:2511.09117 (2025)
16. Lamb, A., Clanuwat, T., Kitamoto, A.: Kuronet: Regularized residual u-nets for end-to-end kuzushiji character recognition. SN Computer Science **1**(3), 1–15 (2020)
17. Li, M., Lv, T., Chen, J., Cui, L., Lu, Y., Florencio, D., Zhang, C., Li, Z., Wei, F.: Trocr: Transformer-based optical character recognition with pre-trained models. In: AAAI Conference on Artificial Intelligence. vol. 37, pp. 13094–13102 (2023)
18. Lin, T.Y., Maire, M., Belongie, S., Hays, J., Perona, P., Ramanan, D., Dollár, P., Zitnick, C.L.: Microsoft coco: Common objects in context. In: European Conference on Computer Vision. pp. 740–755 (2014)

19. Liu, W., Anguelov, D., Erhan, D., Szegedy, C., Reed, S., Fu, C.Y., Berg, A.C.: Ssd: Single shot multibox detector. In: European Conference on Computer Vision. pp. 21–37 (2016)
20. Liu, Y., Ott, M., Goyal, N., Du, J., Joshi, M., Chen, D., Levy, O., Lewis, M., Zettlemoyer, L., Stoyanov, V.: Roberta: A robustly optimized bert pretraining approach. arXiv preprint arXiv:1907.11692 (2019)
21. Liu, Z., Lin, Y., Cao, Y., Hu, H., Wei, Y., Zhang, Z., Lin, S., Guo, B.: Swin transformer: Hierarchical vision transformer using shifted windows. In: IEEE/CVF International Conference on Computer Vision. pp. 10012–10022 (2021)
22. Nagasaki, K., Ohmukai, I., Shimoda, M.: Enhancing digital scholarly editions through advanced ocr accuracy: A new methodological development. In: IPSJ SIG Computers and the Humanities Symposium. pp. 177–182 (2023)
23. National Diet Library (国立国会図書館): NDL Kotenseki OCR Training Dataset (Minna de Honkoku Processed Data) (NDL古典籍OCR学習用データセット (みんなで翻刻加工データ)) (2023), `https://github.com/ndl-lab/ndl-minhon-ocrdataset`
24. National Institute of Japanese Literature (国文学研究資料館): Japanese Classical Cursive Script Kuzushiji Dataset（日本古典籍くずし字データセット） (2016). https://doi.org/10.20676/00000340
25. National Institute of Japanese Literature (国文学研究資料館): What is Kuzushiji (くずし字って何) (2018), `https://www.nijl.ac.jp/koten/kuzushiji/post.html`
26. Redmon, J., Divvala, S., Girshick, R., Farhadi, A.: You only look once: Unified, real-time object detection. In: IEEE Conference on Computer Vision and Pattern Recognition. pp. 779–788 (2016)
27. Ronneberger, O., Fischer, P., Brox, T.: U-net: Convolutional networks for biomedical image segmentation. In: International Conference on Medical Image Computing and Computer-Assisted Intervention. pp. 234–241 (2015)
28. Tan, M., Le, Q.: Efficientnet: Rethinking model scaling for convolutional neural networks. In: International Conference on Machine Learning. pp. 6105–6114 (2019)
29. Telea, A.: An image inpainting technique based on the fast marching method. Journal of graphics tools **9**(1), 23–34 (2004)
30. Tian, Y., Ye, Q., Doermann, D.: Yolov12: Attention-centric real-time object detectors. In: Advances in Neural Information Processing Systems (2025)
31. TOPPAN Inc.: ふみのは (Fuminoha): 古文書カメラ (Komonjo Camera) (2023), `https://camera.fuminoha.jp`
32. Wang, A., Chen, H., Liu, L., Chen, K., Lin, Z., Han, J., et al.: Yolov10: Real-time end-to-end object detection. In: Advances in Neural Information Processing Systems. vol. 37, pp. 107984–108011 (2024)
33. Wang, C.Y., Yeh, I.H., Mark Liao, H.Y.: Yolov9: Learning what you want to learn using programmable gradient information. In: European Conference on Computer Vision. pp. 1–21 (2024)
34. Yang, Z., Liu, B., Xxiong, Y., Yi, L., Wu, G., Tang, X., Liu, Z., Zhou, J., Zhang, X.: Docdiff: Document enhancement via residual diffusion models. In: ACM International Conference on Multimedia. pp. 2795–2806 (2023)
35. Zhang, J., Peng, D., Liu, C., Zhang, P., Jin, L.: Docres: A generalist model toward unifying document image restoration tasks. In: IEEE/CVF Conference on Computer Vision and Pattern Recognition. pp. 15654–15664 (2024)
36. Zhao, Y., Lv, W., Xu, S., Wei, J., Wang, G., Dang, Q., Liu, Y., Chen, J.: Detrs beat yolos on real-time object detection. In: IEEE/CVF Conference on Computer Vision and Pattern Recognition. pp. 16965–16974 (2024)