DKDS: A Benchmark Dataset of Degraded Kuzushiji Documents with Seals for Detection and Binarization

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

[1]Graduate School of Informatics, Kyoto University, Sakyo-ku, Kyoto, 606-8501, Japan.
[2]Academic Center for Computing and Media Studies, Kyoto University, Sakyo-ku, Kyoto, 606-8501, Japan.

*Corresponding author(s). E-mail(s): forest@i.kyoto-u.ac.jp;
Contributing authors: jryjry1094791442@gmail.com; yamashita.kohei.8j@kyoto-u.ac.jp; kameko@i.kyoto-u.ac.jp;

Abstract

Kuzushiji, a pre-modern Japanese cursive script, can currently be read and understood by only a few thousand trained experts in Japan. With the rapid development of deep learning, researchers have begun applying Optical Character Recognition (OCR) techniques to transcribe Kuzushiji into modern Japanese. Although existing OCR methods perform well on clean pre-modern Japanese documents written in Kuzushiji, they often fail to consider various types of noise, such as document degradation and seals, which significantly affect recognition accuracy. To the best of our knowledge, no existing dataset specifically addresses these challenges. To address this gap, we introduce the Degraded Kuzushiji Documents with Seals (DKDS) dataset as a new benchmark for related tasks. We describe the dataset construction process, which required the assistance of a trained Kuzushiji expert, and define two benchmark tracks: (1) text and seal detection and (2) document binarization. For the text and seal detection track, we provide baseline results using multiple versions of the You Only Look Once (YOLO) models for detecting Kuzushiji characters and seals. For the document binarization track, we present baseline results from traditional binarization algorithms, traditional algorithms combined with K-means clustering, and Generative Adversarial Network (GAN)–based methods. The DKDS dataset and the implementation code for baseline methods are available at https://ruiyangju.github.io/DKDS.

Keywords: Kuzushiji, Seals, Pre-modern Japanese Documents, Benchmark Dataset, Object Detection, Document Binarization

1 Introduction

Pre-modern Japanese documents were mainly written in Kuzushiji[1], a pre-modern Japanese cursive script widely used in official records, personal letters, and literary works. Kuzushiji is characterized by complex, flowing strokes and highly variable forms, which differ markedly from modern Japanese characters [1]. Educational reforms in the early 20th century introduced standardized textbooks and led to reduced instruction in Kuzushiji. Following the adoption of modern kana orthography in 1946, most contemporary native Japanese speakers are unable to read Kuzushiji texts without specialized training [2].

[1] くずし字, a form of Japanese cursive script (草書体).

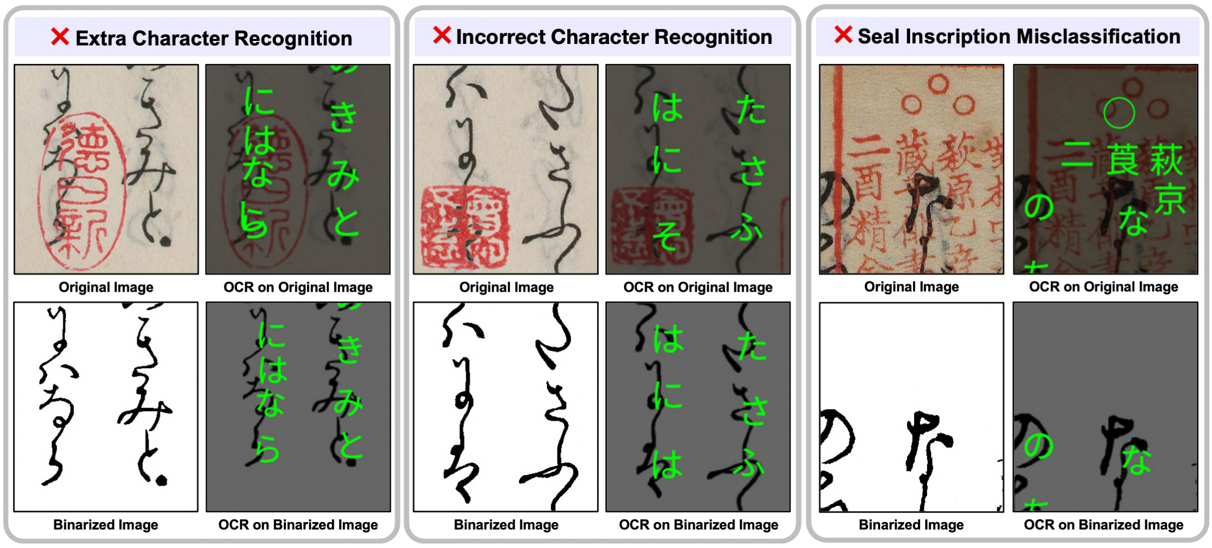

Fig. 1 Comparison of Optical Character Recognition (OCR) results on Kuzushiji characters overlapping with seals between the original and binarized images. OCR was conducted using the "miwo" app [3]. From left to right, the observed OCR errors include recognition of extra character, recognition of incorrect character, and misclassification of seal inscriptions as text.

With the remarkable advances in deep learning for computer vision and natural language processing, researchers have increasingly applied neural network models to Kuzushiji character recognition [4, 5]. While trained models typically achieve satisfactory performance in recognizing Kuzushiji characters seen during training, their performance often drops significantly when tested on unseen (zero-shot) Kuzushiji characters. For instance, the framework proposed by Ishikawa *et al.* [6] achieved an accuracy of only approximately 48% in such zero-shot scenarios. This result presents the critical need to construct high-quality Kuzushiji datasets, including special cases, to improve model generalization and capabilities in complex recognition tasks.

The primary Kuzushiji datasets include Kuzushiji-MNIST, Kuzushiji-49, and Kuzushiji-Kanji[2] [7]. The Kaggle competitions organized using these datasets has further advanced both research and practical applications [8, 9]. Together, these datasets and competitions cover a wide range of Kuzushiji characters and provide a standardized evaluation platform for researchers. With these resources, researchers can perform not only Kuzushiji character classification but also related tasks such as Kuzushiji detection and generation [10], thereby advancing the development of Kuzushiji character recognition techniques and enhancing model performance.

Several methods have been proposed for the Kuzushiji character recognition task. Fujita *et al.* [11] employed Transformer [12] models to analyze and recognize Kuzushiji characters, leveraging attention mechanisms to capture relationships between adjacent characters and accurately separate character regions and boundaries. Ngo *et al.* [13] introduced RNN-Transducer [14] models for handwritten Kuzushiji character recognition, effectively exploiting both visual and text information from the input images and achieving satisfactory performance.

In addition, KuroNet [15, 16] is an end-to-end model based on a residual U-Net [17] architecture, designed to capture long-range context, process extensive vocabularies, and accommodate non-standard character layouts, thereby enabling the recognition of pre-modern Japanese documents. On the Kuzushiji dataset provided by the Kaggle competitions [8, 9], KuroNet achieved an accuracy exceeding 90%. To facilitate the deployment of OCR models on mobile devices, Clanuwat *et al.* [3] developed the "miwo" application, which allows

[2]Kanji refers to Chinese-origin characters used in Japanese.

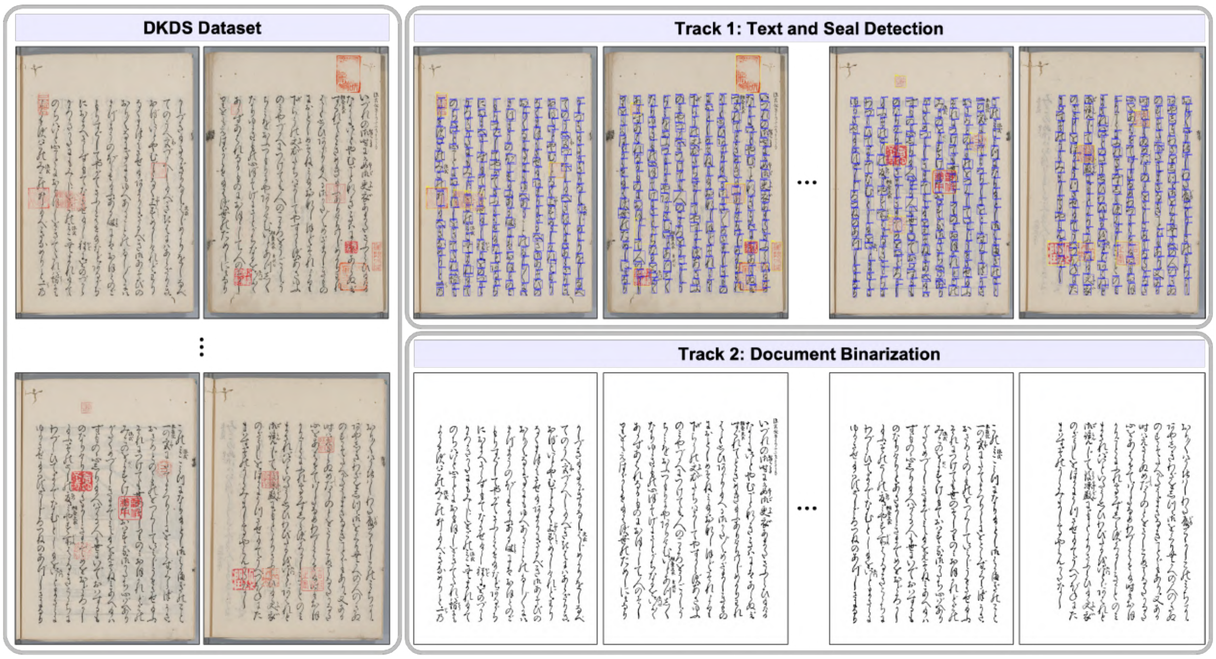

Fig. 2 **DKDS** dataset is the first collection of degraded pre-modern Japanese document images specifically designed to address the challenge of Kuzushiji characters overlapping with seals. Based on the dataset, we define two benchmark tracks: (1) Text and Seal Detection, and (2) Document Binarization.

users to upload or photograph Kuzushiji documents and automatically transcribe the characters into modern Japanese.

The aforementioned OCR methods and applications perform well on clean pre-modern Japanese documents written in Kuzushiji; however, they do not sufficiently consider the effects of noise, such as document degradation and seals, on recognition accuracy. As shown in Fig. 1, such noise can cause errors that significantly reduce recognition accuracy. To address this issue, document binarization, which preserves the foreground (black) while removing background and noise (white), has been widely adopted as a preprocessing step, as it can effectively improve OCR performance.

Among the various types of noise, the overlap between Kuzushiji characters and seals poses a major challenge. Serving as symbols of ownership and collector identity, seals were widely used in ancient and pre-modern Asian document collections [18]. They are typically inscribed with stylized ancient scripts and may contain not only the collector's name but also phrases expressing personal wishes, interests, or social status. In ancient and pre-modern documents, seals usually appear as red marks that frequently overlap with characters, which can blur the characters and lead to OCR errors.

To address this challenge, we introduce the Degraded Kuzushiji Documents with Seals (DKDS) dataset, developed in collaboration with a trained Kuzushiji expert. This novel benchmark dataset comprises degraded pre-modern Japanese documents written in Kuzushiji, with various types of seals randomly added to simulate common types of interference. As shown in Fig. 2, we define two task tracks for the DKDS dataset: (1) text and seal detection, and (2) document binarization. The former track aims to accurately detect text (i.e., Kuzushiji characters) and seal regions within the documents, while the latter focuses on preserving the primary textual content while removing noise, including stains and seals, thereby providing clean input for subsequent OCR tasks.

Overall, the contributions of this work are summarized as follows:

(a) We introduce a novel benchmark dataset, named Degraded Kuzushiji Documents with

Seals (DKDS), developed in collaboration with a trained Kuzushiji expert. To the best of our knowledge, this is the first publicly available dataset specifically addressing the challenge of Kuzushiji characters overlapped with seals.

(b) We provide a detailed description of the dataset construction process, including the generation of binarization ground-truth, its verification and manual correction, and the random addition of seals to the raw Kuzushiji documents.

(c) We define the text and seal detection task, which serves as a preliminary step for recognizing Kuzushiji characters and transcribing them into modern Japanese. Baseline results for this task are also presented using multiple versions of the YOLO models.

(d) We define the document binarization task specifically for pre-modern Japanese document images. Compared to existing binarization datasets, the DKDS dataset presents additional challenges due to the overlap of Kuzushiji characters with seals. Baseline results for this task are provided, including traditional binarization algorithms, traditional algorithms combined with K-means clustering, and Generative Adversarial Network (GAN)–based methods.

The rest of this paper is organized as follows: Section 2 reviews related datasets and previous work relevant to the two task tracks defined here, highlighting their limitations. Section 3 describes the dataset construction process, including raw data collection, generation of binarization ground-truth, and the procedure for adding seals. Section 4 defines the two task tracks and their corresponding evaluation metrics, and introduces the baseline methods used in this work. Section 5 presents the experimental setup and implementation details, followed by a comparative analysis of quantitative results for the two tracks. Finally, Section 6 summarizes this work and discusses the direction of future research.

2 Related Works

2.1 Related Dataset

The datasets relevant to this work can be categorized into three domains: (1) Kuzushiji document datasets, (2) degraded pre-modern document binarization datasets, and (3) seal datasets.

Existing research on pre-modern Japanese documents written in Kuzushiji [4, 7] lacks a dataset that specifically addresses the challenge of overlaps between characters and seals, which can significantly reduce recognition accuracy.

Most publicly available datasets for degraded ancient document binarization focus on documents written in Latin-based alphabetic systems. For instance, the documents in the Document Image Binarization Contest (DIBCO) series datasets [19–28] are primarily written in English, while the Bickley Diary (BD) dataset [29], Synchromedia Multispectral Ancient Document Images (SMADI) [30], and Persian Heritage Image Binarization Dataset (PHIBD) [31] all contain documents written in European languages. In contrast, no publicly available document binarization datasets exist for pre-modern Japanese documents, particularly those written in Kuzushiji.

Furthermore, no existing degraded document datasets include seals as the primary source of interference. Therefore, we aim to randomly add seals to the collected pre-modern Japanese documents written in Kuzushiji. However, existing publicly available seal datasets are all unsuitable, as the seals they contain are either not ancient Asian seals, suffer from poor image quality, or have complex backgrounds. Specifically, the Chinese-Seal Dataset (CSD) [32] focuses on modern Chinese seals; DDI-100 [33] contains Western (English-language) seal samples; and SPODS [34] comprises seal images from contemporary office settings. Although MiikeMineStamps [35] includes some Japanese seal images, they are generally low-resolution and feature non-uniform backgrounds, making background removal difficult.

2.2 Text and Seal Detection

The diverse shapes of ancient seals, their curved inscriptions, and the frequent occurrence of text overlap pose significant challenges for seal processing and text analysis and recognition. As a critical preliminary step for Kuzushiji character recognition, text detection is particularly important [36]. Similarly, seal detection is important for subsequent seal processing.

However, the scarcity of ancient seal samples and the prevalence of low-quality or degraded images severely limit available training data and annotations for related tasks [18]. For instance,

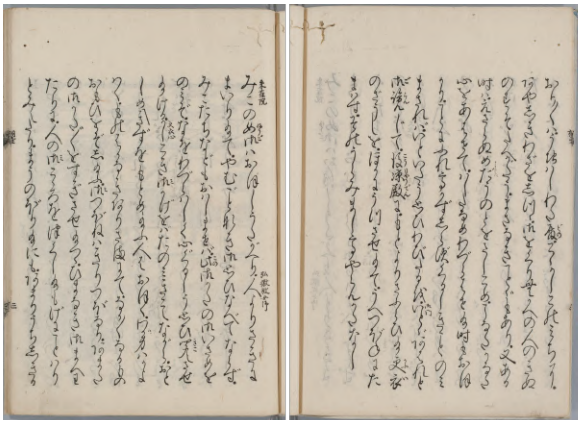

Fig. 3 Examples of raw Kuzushiji document data from the book *Genji Monogatari* (*The Tale of Genji*) [50].

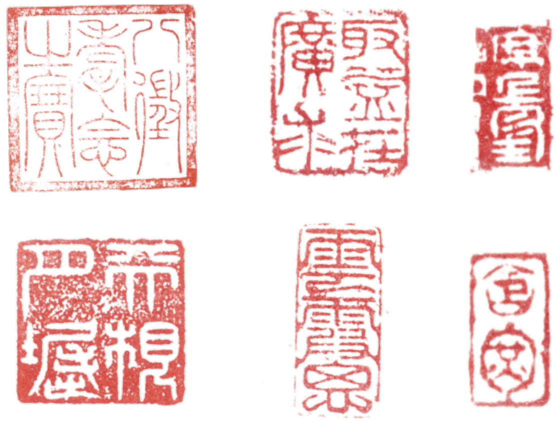

Fig. 4 Examples of our collected imperial seals from the Qing dynasty, with backgrounds removed.

Micenkov *et al.* [37] proposed an automatic seal segmentation system for insurance documents that achieved 84% accuracy. Yu *et al.* [38] developed a modern seal dataset for official and financial contexts, introducing two tasks: seal title text detection and end-to-end seal title recognition. To the best of our knowledge, no dataset currently exists for text and seal detection in degraded pre-modern Japanese Kuzushiji documents.

The YOLO series of models [39–42] achieve an excellent balance between detection accuracy and model size, demonstrating outstanding performance across a wide range of object detection tasks [43–47]. Researchers have explored the application of the YOLO series of models in Kuzushiji character detection [36] and seal detection [48]. For instance, Bento *et al.* [48] merged and adapted the StaVer [37] and DDI-100 [33] datasets to evaluate the performance of YOLOv8, YOLOv9, YOLOv10, and YOLO11 on their newly constructed dataset for seal detection. Similarly, Sun *et al.* [49] enhanced the YOLOv8 architecture and proposed RA-YOLOv8, achieving higher detection accuracy. Nevertheless, due to the lack of seal data based on Kuzushiji documents, these improved YOLO models cannot be directly applied to this specific scenario.

2.3 Document Binarization

Document binarization is an important preprocessing step for OCR and layout analysis, effectively removing various types of noise from document images, such as paper yellowing, text fading, ink bleeding, stains, and seals. The Document Image Binarization Contest (DIBCO), one of the most popular competitions in this field, has provided benchmark datasets from 2009 to 2019 [19–28]. Researchers have proposed a range of neural network models evaluated on these datasets, including those based on GANs [51–53], Transformers [54], and diffusion models [55, 56], which have significantly advanced the field. However, most existing datasets contain ancient and pre-modern document images based on Latin alphabets, while few include Kuzushiji documents, particularly those containing overlapping seals and text.

3 Dataset Construction

3.1 Raw Data Collection

The raw Kuzushiji document data were obtained from the Center for Open Data in the Humanities (CODH) [50]. Specifically, we selected the dataset corresponding to book ID “200003803”, titled *Genji Monogatari* (*The Tale of Genji*)[3]. Examples from this book are shown in Fig 3. This dataset was chosen for two main reasons: (1) its historical significance and widespread popularity; and (2) the availability of detailed OCR annotations provided by CODH, including 237 unique character types and a total of 11,132 characters. These data can be used for training and evaluating OCR models in future research.

[3] https://codh.rois.ac.jp/char-shape/book/200003803/

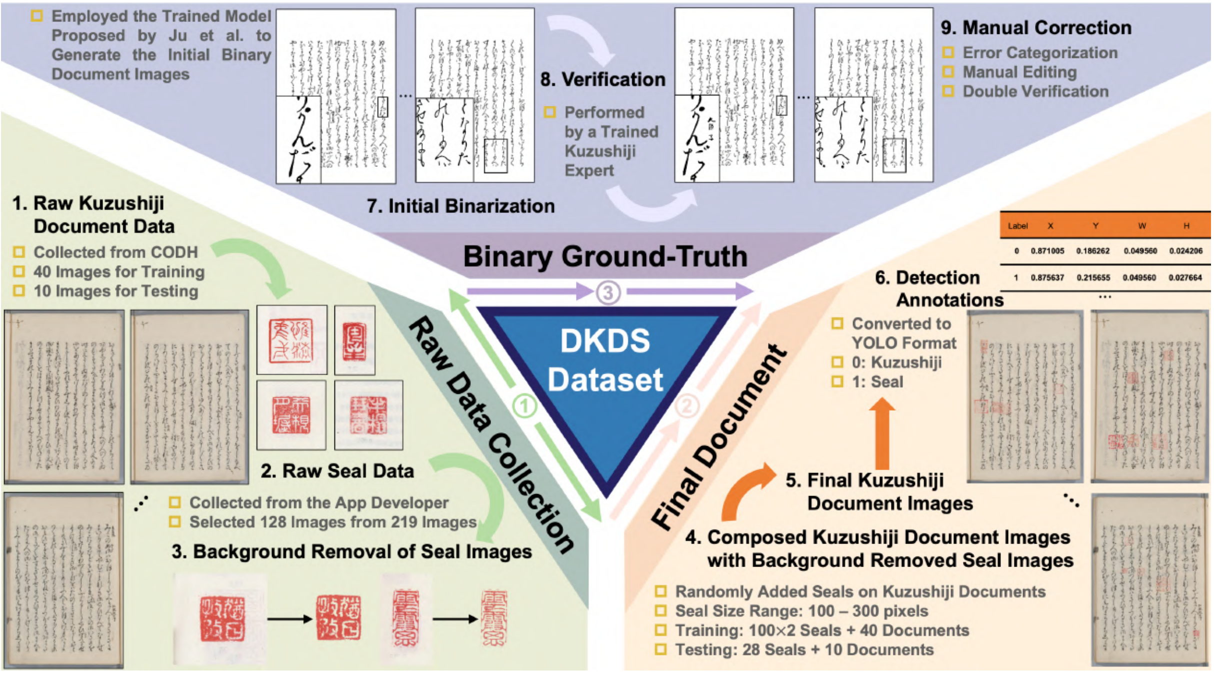

Fig. 5 The overall workflow of the proposed **DKDS** dataset construction includes raw data collection, detection annotations, initial binarization ground-truth generation, verification, and manual correction. The initial binarization ground-truth was generated using the model proposed by Ju *et al.* [57], while the verification was conducted by a trained Kuzushiji expert.

Table 1 Statistics of the proposed DKDS dataset.

Category	Total	Training	Testing
Kuzushiji Documents	50	40	10
Seals	128	100×2	28
Seals per Document	-	4 – 6	2 – 3

Furthermore, to date, there is no publicly available high-resolution dataset of ancient Japanese seals. To address this gap, we collected 219 seals from three emperors of the Qing dynasty (i.e., Kangxi (1661–1722), Yongzheng (1722–1735), and Qianlong (1735–1796)) sourced from the developer of an ancient seal app[4]. The primary reason for selecting ancient Chinese seals is that both ancient Japanese and Chinese seals are mainly inscribed with kanji characters.

Considering factors such as image resolution and seal patterns (e.g., whether the inscriptions are written in kanji), we selected a final set of 128 seal images from the collection. As shown in Fig. 4, these seals are high-resolution, and their backgrounds can be easily removed.

[4]https://www.daizigege.com/seal/

3.2 Construction Workflow

We constructed our dataset by producing color document images and their corresponding binarized ground-truth images from the raw data, and the overall workflow is shown in Fig. 5. To create the final document images, we combined 128 background removed seal images with 50 Kuzushiji document images. Specifically, 40 Kuzushiji documents were selected for training, and the remaining 10 for testing. As detailed in Table 1, the 128 seal images were divided into 100 for the 40 training Kuzushiji documents, and 28 for the 10 test Kuzushiji documents. To increase the cases of overlap between seals and Kuzushiji characters in the training set, we duplicated the 100 training seal images to create 200 seals, which were then randomly added onto the 40 training Kuzushiji documents. As a result, each Kuzushiji document image in the training set contains approximately 4–6 seals, while those in the test set contain about 2–3 seals.

To generate the binarization ground-truth, we employed a combined approach of neural network generation followed by manual verification and

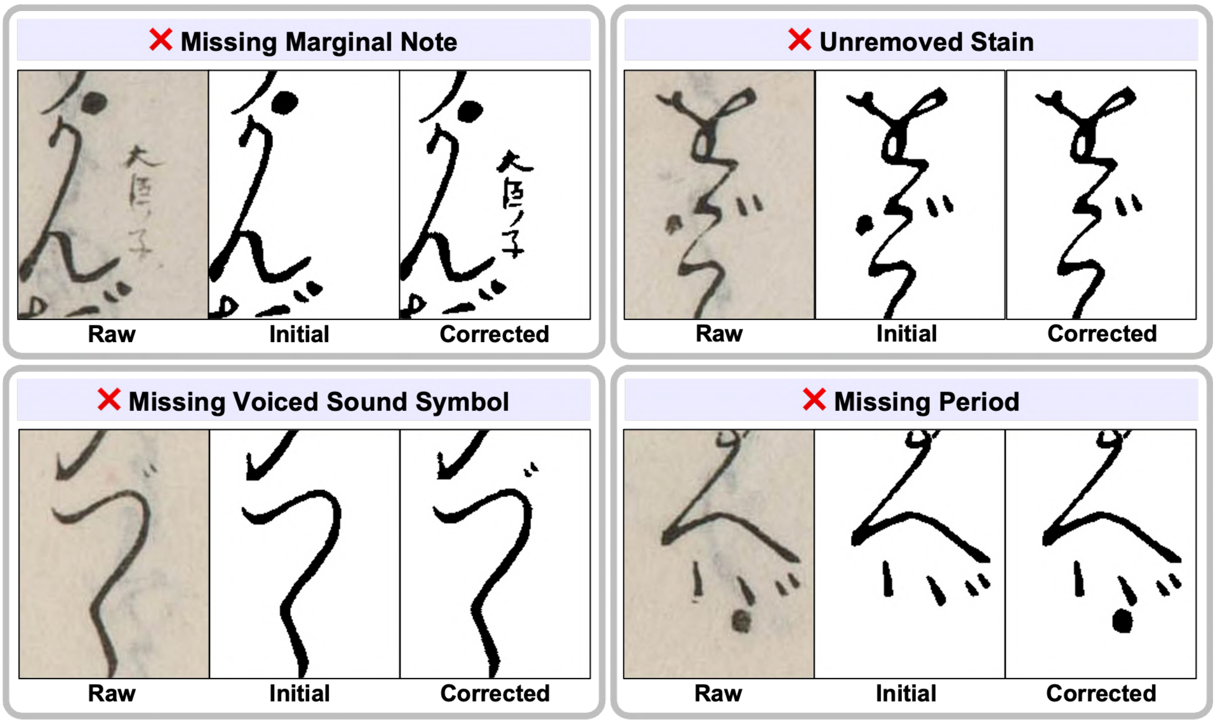

Fig. 6 Examples of verification and manual correction in the dataset construction workflow: "Raw" refers to the input images, "Initial" denotes the initial binarization produced by Ju *et al.* [57], and "Corrected" represents the final ground-truth after verification and manual correction. The trained Kuzushiji expert performed verification to determine which elements should be preserved (e.g., marginal notes, voiced sound symbols, periods) and which should be removed (e.g., stains).

correction. As shown in Fig. 5, we first applied the model proposed by Ju *et al.* [57] to binarize the raw Kuzushiji document images (without added seals), producing the initial binarization ground-truth. Subsequently, a trained Kuzushiji expert manually verified these initial results. Even in the absence of seal interference, the initial binarization often contained errors. As detailed in Fig. 6, the expert identified issues such as missing marginal notes, voiced sound symbols, periods, and unremoved stains. These errors were then manually corrected to create the final binarization ground-truth. Notably, this verification step was critical and required a trained Kuzushiji expert to ensure the validity of the process.

3.3 Detection Annotations

The detection annotations include bounding boxes for both Kuzushiji characters and seals. The bounding box information for Kuzushiji characters was obtained from the OCR annotations provided by CODH [50], while the bounding boxes for the seals were recorded during the process of randomly adding them to the Kuzushiji document images. Notably, the OCR annotations from CODH [50] include only the main Kuzushiji text, excluding marginal notes. Considering practical applications, we followed the same strategy.

All annotations are provided in the YOLO format, enabling researchers to directly download and use the dataset. In addition, detailed information for both Kuzushiji characters and seals, including their coordinates and sizes, is available on our GitHub repository, allowing researchers to convert the data into other object detection formats for model training and evaluation.

4 Task Definition

4.1 Evaluation Tracks

We define two task tracks in the proposed dataset: (1) text and seal detection, and (2) document binarization. As shown in Fig. 7, seals may overlap with Kuzushiji characters or other seals, and

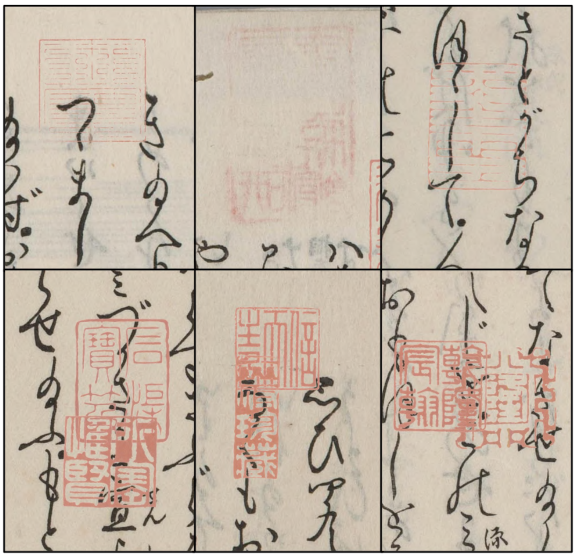

Fig. 7 Challenging examples for the text and seal detection task: The first row presents seals suffering from ink fading, while the second row shows seals overlapping with Kuzushiji characters or other seals.

suffer from ink fading, all of which can significantly reduce detection accuracy, making this task highly challenging. In the document binarization task, the goal is to remove seals while preserving, or even restoring, the underlying Kuzushiji characters. This task becomes particularly difficult when the characters and seals overlap, as shown in Fig. 8.

Furthermore, both tasks hold important research significance. Text and seal detection serves as a critical preliminary step for downstream applications, including Kuzushiji character recognition and seal processing, while document binarization aims to improve the accuracy of subsequent OCR systems. By addressing these challenges, the proposed dataset encourages the development of robust models capable of processing real-world scenarios in Kuzushiji document analysis.

4.2 Evaluation Metrics

For the text and seal detection task, we adopted standard object detection evaluation metrics [58], including model parameters (Params), floating-point operations (FLOPs), average precision at a 50% intersection over union (IoU) threshold

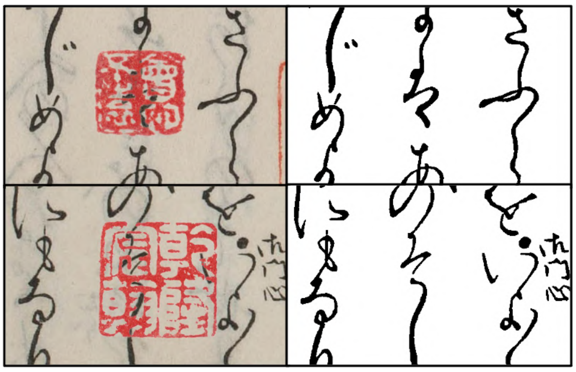

Fig. 8 Challenging examples for the document binarization task: The left column shows the input Kuzushiji document images, while the right column presents the corresponding binarized images.

(AP_{50}), and average precision across IoU thresholds from 50% to 95% ($AP_{50:95}$). Model parameters reflect the model's size and complexity, FLOPs quantify computational cost, and these average precision (AP) metrics provide a comprehensive measure of detection accuracy.

For the document binarization task, we employed classical evaluation metrics [19] for quantitative comparison, including the F-measure (FM), pseudo F-measure (p-FM), peak signal-to-noise ratio (PSNR), and distance reciprocal distortion (DRD). To provide a more comprehensive assessment, we also adopted the Average-Score Metric (ASM) proposed by Ju *et al.* [59], which is calculated as follows:

$$ASM = \frac{FM + p\text{-}FM + PSNR + (100 - DRD)}{4}. \quad (1)$$

A higher ASM value indicates better overall performance, providing a single unified metric for model ranking.

4.3 Baselines

For the text and seal detection task, we employed the YOLO series of object detection models, including YOLOv8 [39], YOLOv9 [40], YOLOv10 [41], and YOLO11 [42]. As one-stage detectors, these models reduce computational complexity and model parameters while maintaining high detection accuracy, achieving a good balance between inference speed and performance. In addition, they support end-to-end training,

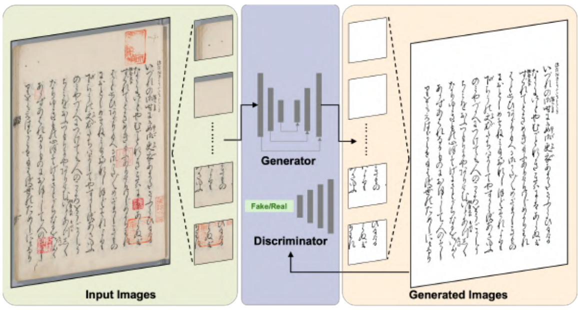

Fig. 9 The architecture of the GAN-based method used.

provide fast inference, and present strong generalization capabilities, making them widely adopted in object detection tasks.

For the document binarization task, we first applied several traditional binarization algorithms to the Kuzushiji documents, including Niblack [60], Otsu [61], and Sauvola [62], which can be directly employed without any training.

Considering that seals in the pre-modern Japanese documents are typically red, while Kuzushiji characters are black and the paper background is pale yellow, we introduced a preprocessing step to remove seal regions based on K-Means clustering [63] before applying the binarization algorithms. This preprocessing effectively reduces the interference caused by red seals and enhances the quality of subsequent binarization results. Specifically, we designed and implemented three combined approaches: K-Means + Otsu, K-Means + Niblack, and K-Means + Sauvola.

In addition, to provide a deep learning-based baseline, we employed a GAN [64]-based model to generate binarized Kuzushiji document images. The network architecture for this model is shown in Fig. 9. Since directly inputting the entire image (approximately 2100×3200 pixels) would require excessive GPU memory, we divided each document image into multiple 512×512 patches; the details of this division are described in Section 5.1.2. These patches were then fed into a U-Net [17] architecture with an EfficientNet-B5 [65] backbone, which served as the generator. A simplified PatchGAN [66, 67] discriminator was adopted to distinguish between real and fake (generated) images. Furthermore, following Suh *et al.* [51] and Ju *et al.* [57], we employed a loss function that combines the Wasserstein GAN with Gradient Penalty (WGAN-GP) loss [68] and an additional Binary Cross-Entropy (BCE) loss.

5 Experiments

5.1 Implementation Details

To ensure a fair performance comparison on the DKDS dataset, all models were trained and evaluated on the same hardware (i.e., NVIDIA RTX 3090 GPU). All experiments were conducted using the Ubuntu operating system, Python 3.9, and the PyTorch framework.

5.1.1 Text and Seal Detection

For the text and seal detection task, the YOLO models were trained for 100 epochs with a batch size of 16 and an input resolution of 640×640 pixels. Model training was performed using the SGD optimizer [69] with an initial learning rate of 0.01. All YOLO models were initialized with weights pre-trained on the COCO dataset [58].

5.1.2 Document Binarization

For the document binarization task, traditional binarization algorithms (i.e., Niblack [60], Otsu [61], and Sauvola [62]) first convert the input image into a grayscale image as a preprocessing step. The Niblack algorithm uses a 25×25 sliding window with $k = 0.8$, while the Sauvola algorithm employs the same window size. For the K-Means clustering method [63], the number of clusters is set to $K = 3$. After clustering, median filtering with a kernel size of 5 is applied to smooth the results and remove isolated noise points.

For the GAN-based method [64], the training data were preprocessed by dividing each original image into 512×512 overlapping patches with a 30% overlap ratio. To further expand the training set, we applied data augmentation, including scaling (factors 0.75, 1.0, 1.25, and 1.5) and rotations (0°and 270°). After preprocessing and augmentation, the 40 original training images were expanded into a total of 71,458 patches of size 512×512 pixels. Both the generator and discriminator were optimized using the Adam [70] optimizer with a learning rate of 2×10^{-4} and β coefficients of $(0.5, 0.999)$. The model was trained for 10 epochs with a batch size of 16. At inference,

Table 2 Quantitative comparison of various YOLO models for text and seal detection on the DKDS dataset.

Model	Params	FLOPs	$AP_{50}^{\text{Kuzushiji}} \uparrow$	$AP_{50:95}^{\text{Kuzushiji}} \uparrow$	$AP_{50}^{\text{Seal}} \uparrow$	$AP_{50:95}^{\text{Seal}} \uparrow$
YOLOv8m [39]	25.86M	79.1G	96.4%	71.2%	**99.1%**	**86.2%**
YOLOv9m [40]	20.16M	77.5G	96.3%	71.7%	97.2%	81.4%
YOLOv10m [41]	**16.49M**	**64.0G**	96.2%	71.4%	**99.1%**	85.7%
YOLO11m [42]	20.05M	68.2G	**97.8%**	**74.1%**	98.5%	85.7%

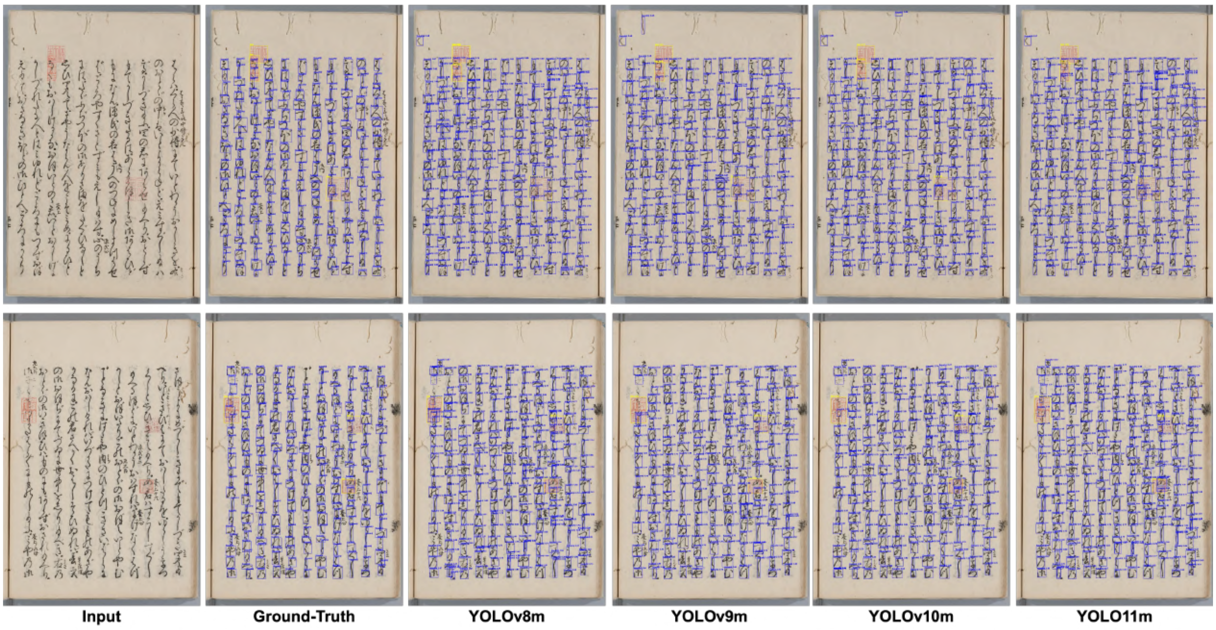

Fig. 10 Visual comparison of text and seal detection results predicted by different models with the corresponding ground-truth annotations. The input images are "200003803_00028_1" (bottom) and "200003803_00028_2" (top).

the predicted outputs were binarized using a 0.5 threshold.

5.2 Track 1: Text and Seal Detection

5.2.1 Quantitative Comparison

We trained and evaluated various YOLO series models on our proposed DKDS dataset, using the medium (M) size variant for all models. Quantitative comparison results are summarized in Table 2.

For the text (i.e., Kuzushiji character) detection, all models achieved over 96% in AP_{50}, indicating that they can accurately detect most Kuzushiji characters. Under stricter IoU thresholds (i.e., $AP_{50:95}$), YOLO11m achieved the highest score of 74.1%, outperforming the other models by about 2–3%. This demonstrates the superior accuracy of YOLO11m in localizing Kuzushiji character bounding boxes.

For the seal detection, both YOLOv8m and YOLOv10m obtained the highest AP_{50} of 99.1%, indicating near-perfect detection performance under the loose IoU criterion. In addition, YOLOv8m reached the highest $AP_{50:95}$ of 86.2%, demonstrating strong localization accuracy under stricter IoU thresholds.

In terms of model size, YOLOv10m presents the lowest Params of 16.49M and FLOPs of 64.0G, making it the most lightweight model.

Meanwhile, YOLO11m achieves the best balance between accuracy and efficiency, with 20.05M parameters, 68.2G FLOPs, $AP_{50}^{Kuzushiji}$ of 97.8%, and AP_{50}^{Seal} of 98.5%. Therefore, YOLO11m can be considered an effective model for the DKDS dataset.

Table 3 Quantitative comparison of various methods for document binarization on the DKDS dataset.

Method	FM↑	p-FM↑	PSNR↑	DRD↓	ASM↑
Niblack [60]	39.13	41.14	8.44dB	79.70	27.25
Otsu [61]	63.01	63.31	11.76dB	37.69	50.10
Sauvola [62]	87.87	90.99	18.34dB	7.01	72.55
K-means [63] + Niblack [60]	39.99	42.03	8.61dB	76.67	28.49
K-means [63] + Otsu [61]	84.76	86.28	17.14dB	9.90	69.57
K-means [63] + Sauvola [62]	88.59	91.48	18.65dB	6.37	73.09
GAN [64]	**98.11**	**98.14**	**26.53dB**	**0.82**	**80.49**

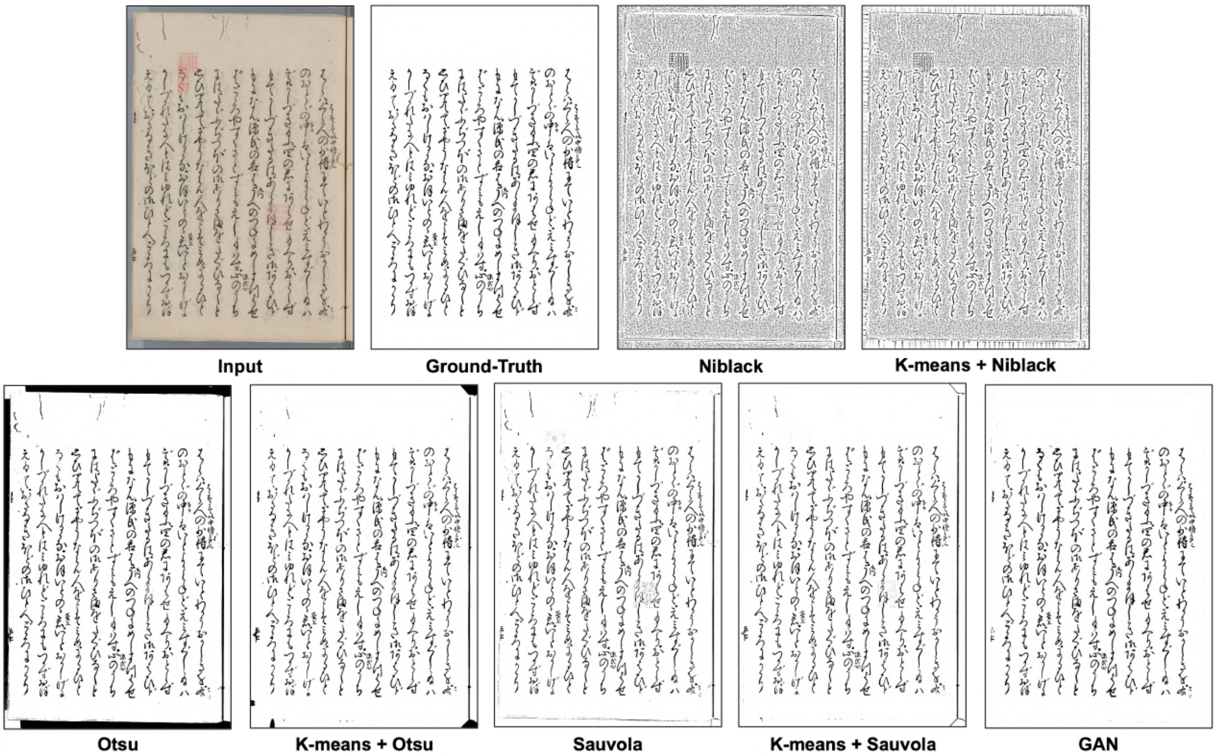

Fig. 11 Visual comparison of results generated by different methods on Kuzushiji document "200003803_00028_2" from the DKDS test set. The input image, ground-truth, and method outputs are shown for comparison.

5.2.2 Visualization

We present the prediction results of various YOLO models for text and seal detection, along with the corresponding input and ground-truth images. As shown in Fig. 10, seals are marked with yellow bounding boxes, while Kuzushiji characters are marked with blue bounding boxes.

In the first row, YOLOv10m failed to detect a Kuzushiji character at the bottom of the third row from the left, while the other models detected all characters correctly. However, all models incorrectly predicted stains in the upper-left corner as Kuzushiji characters, which is the main factor affecting model performance. Furthermore, although our ground-truth annotations do not include positional information for marginal notes, both YOLOv8m and YOLO11m successfully detected several marginal notes, as shown on the far right.

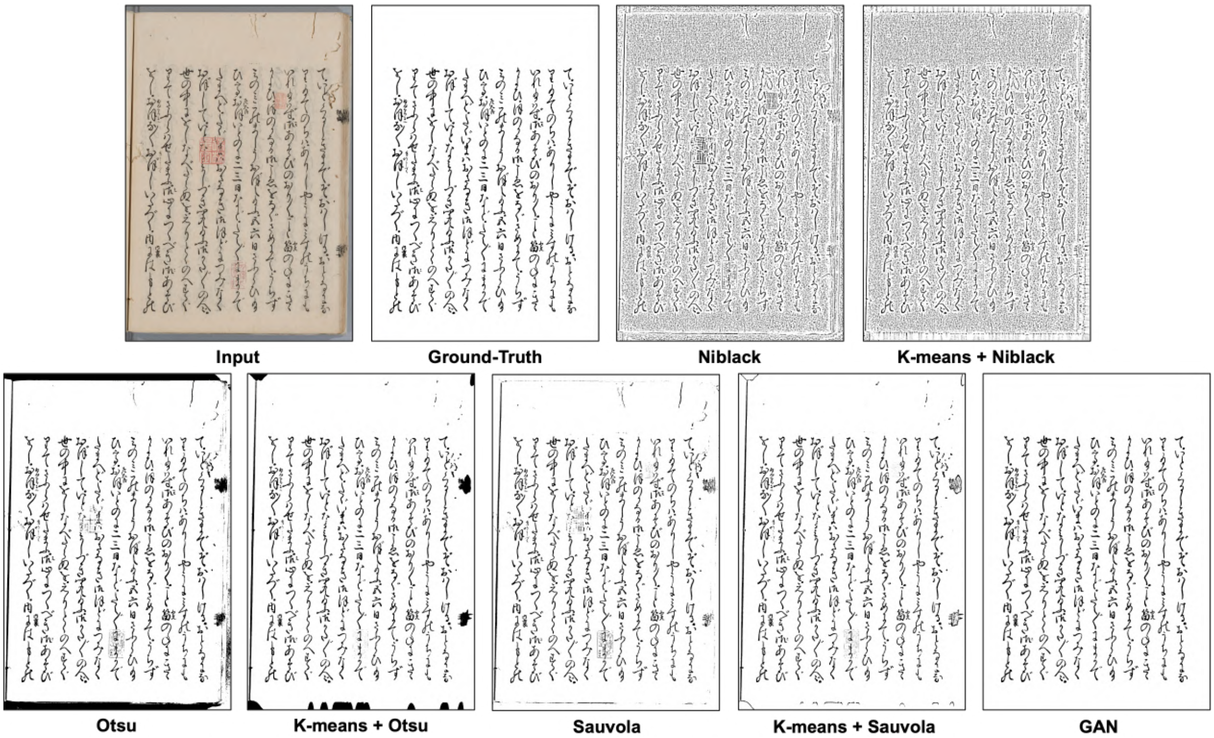

Fig. 12 Visual comparison of results generated by different methods on Kuzushiji document "200003803_00029_1" from the DKDS test set. The input image, ground-truth, and method outputs are shown for comparison.

In the second row, the ink on the Kuzushiji character in the upper-left corner is fading, which poses a challenge for detection. None of the YOLO models except YOLO11m correctly detected this character, demonstrating YOLO11m's superior capability in Kuzushiji character detection.

5.3 Track 2: Document Binarization

5.3.1 Quantitative Comparison

We employed traditional binarization algorithms, traditional algorithms combined with K-means clustering, and deep learning-based (i.e., GAN) methods as baselines for the document binarization task. The experimental results are summarized in Table 3.

Overall, the traditional binarization algorithms presented poor performance. Specifically, Niblack's method [60] performed the worst, as indicated by its high DRD value, reflecting strong sensitivity to noise in the DKDS dataset. Otsu's method [61], a simple global thresholding method, achieved moderate results, while Sauvola's method [62] significantly outperformed both by leveraging locally adaptive thresholding to process complex backgrounds more effectively.

Furthermore, applying K-means clustering [63] as a preprocessing step to remove red interference (e.g., seals) further improved performance. For instance, the ASM increased from 27.25 to 28.49, from 50.10 to 69.57, and from 72.55 to 73.09 for K-means + Niblack, K-means + Otsu, and K-means + Sauvola, respectively.

In contrast, the deep learning–based method (i.e., GAN) demonstrated superior performance, achieving the FM of 98.11, p-FM of 98.14, PSNR of 26.53 dB, DRD of 0.82, and an ASM of 80.49. The high FM, p-FM, and low DRD values indicate that the generated binary images are highly readable. However, the relatively lower PSNR suggests that, although the Kuzushiji character outlines are clear, some noise or imperfections remain along the edges.

5.3.2 Visualization

We present two groups of result images generated by different methods, along with their corresponding input and ground-truth images, as shown in Figs. 11 and 12. The input images contain various forms of degradation, including red seals, paper damage, stains, and ink fading, while the ground-truth images remove all such interference, retaining only the black Kuzushiji characters.

Among the traditional methods, Niblack's algorithm [60] produced the poorest results, with much noise remaining in the background. Otsu's method [61] achieved better separation between text and background but still retained some noise. Sauvola's method [62] performed relatively better but failed to completely remove seal interference.

When combined with K-means clustering [63] preprocessing, all three traditional methods achieved improved performance, further reducing seal marks, though minor residues remained.

In contrast, the GAN-based method [64] obtained the best results, effectively removing most noise and seal interference while maintaining excellent text legibility.

6 Conclusion

Binarizing degraded Kuzushiji document images presents significant challenges, particularly when seals overlap with the characters. Effectively removing seal interference while preserving or restoring character details remains a critical problem. Notably, no publicly available dataset currently contains Kuzushiji documents with seals, limiting the development of methods for this task. To address this gap, we constructed the Degraded Kuzushiji Documents with Seals (DKDS) dataset by collecting and integrating pre-modern Kuzushiji document images with high-resolution images of ancient seals. We detail the construction process of the DKDS dataset and define two task tracks based on it: (1) text and seal detection, and (2) document binarization. Furthermore, we provide baseline evaluation results for several classical methods to facilitate future research.

Notably, the selected Kuzushiji document dataset includes corresponding Unicode character mappings and OCR annotations, providing valuable resources for training models on both document binarization and OCR tasks. In future work, we plan to jointly design and train models for these tasks, aiming to develop an end-to-end system capable of efficiently and accurately converting Kuzushiji characters into modern Japanese in degraded pre-modern Japanese documents.

Acknowledgments

We sincerely appreciate Jie Chen for kindly providing and supporting the collection of the raw seal data used in this work.

Funding

This work was supported by JSPS KAKENHI Grant Numbers 25H01242.

Author Contributions

Rui-Yang Ju: Conceptualization, Data Curation, Formal Analysis, Methodology, Writing – Original Draft Preparation, Writing – Review & Editing; **Kohei Yamashita**: Data Curation, Investigation, Writing – Review & Editing; **Hirotaka Kameko**: Project Administration, Resources, Writing – Review & Editing; **Shinsuke Mori**: Funding Acquisition, Resources, Supervision, Writing – Review & Editing.

Data availability

The proposed dataset and implementation code are publicly available on GitHub at https://github.com/RuiyangJu/DKDS.

Declarations

Competing Interests

The authors declare that they have no conflict of interest.

Ethics approval

This research does not involve human participants and/or animals.

References

[1] Ueki, K., Kojima, T.: Survey on deep learning-based kuzushiji recognition. In: International Conference on Pattern Recognition, pp. 97–111 (2021)

[2] 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)

[3] Clanuwat, T., Kitamoto, A.: 'miwo' ai kuzushiji recognition application for document examination. In: IPSJ Humanities and Computer Symposium (2021)

[4] Le, A.D., Clanuwat, T., Kitamoto, A.: A human-inspired recognition system for pre-modern japanese historical documents. IEEE Access **7**, 84163–84169 (2019)

[5] Le, A.D.: Automated transcription for pre-modern japanese kuzushiji documents by random lines erasure and curriculum training. In: International Workshop on Document Analysis Systems, pp. 371–382 (2020)

[6] Ishikawa, T., Miyazaki, T., Omachi, S.: Japanese historical character recognition by focusing on character parts. Pattern Recognition **148**, 110181 (2024)

[7] Clanuwat, T., Bober-Irizar, M., Kitamoto, A., Lamb, A., Yamamoto, K., Ha, D.: Deep learning for classical japanese literature. In: NeurIPS Workshop on Machine Learning for Creativity and Design (2018)

[8] Kitamoto, A., Clanuwat, T., Lamb, A., Bober-Irizar, M.: Progress and results of kaggle machine learning competition for kuzushiji recognition. In: IPSJ SIG Computers and the Humanities Symposium, vol. 2019, pp. 223–230 (2019). (in Japanese)

[9] Kitamoto, A., Clanuwat, T., Bober-Irizar, M.: Kaggle kuzushiji recognition competition: Challenges of hosting a world-wide competition in the digital humanities. In: Proceedings of the Japanese Society for Artificial Intelligence, vol. 35, pp. 366–376 (2020). (in Japanese)

[10] Center for Open Data in the Humanities (CODH): Soan: Library for rendering modern Japanese using old movable type. http://codh.rois.ac.jp/software/soan/. MIT License (2023)

[11] Fujita, S., Oyama, K.: Learning a sequence of cursive-style japanese characters in classical literary works. In: Asia Pacific Signal and Information Processing Association Annual Summit and Conference, pp. 1–6 (2024)

[12] Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A.N., Kaiser, Ł., Polosukhin, I.: Attention is all you need. Advances in Neural Information Processing Systems **30** (2017)

[13] Ngo, T.T., Nguyen, H.T., Ly, N.T., Nakagawa, M.: Recurrent neural network transducer for japanese and chinese offline handwritten text recognition. In: International Conference on Document Analysis and Recognition, pp. 364–376 (2021)

[14] Graves, A.: Sequence transduction with recurrent neural networks. In: ICML Workshop on Representation Learning (2012)

[15] 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)

[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] 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)

[18] Li, K., Batjargal, B., Maeda, A.: Character segmentation in asian collector's seal

imprints: an attempt to retrieval based on ancient character typeface. Journal of Data Mining & Digital Humanities (HistoInformatics) (2021)

[19] Gatos, B., Ntirogiannis, K., Pratikakis, I.: Icdar 2009 document image binarization contest (dibco 2009). In: International Conference on Document Analysis and Recognition, pp. 1375–1382 (2009)

[20] Pratikakis, I., Gatos, B., Ntirogiannis, K.: H-dibco 2010-handwritten document image binarization competition. In: International Conference on Frontiers in Handwriting Recognition, pp. 727–732 (2010)

[21] Pratikakis, I., Gatos, B., Ntirogiannis, K.: Icdar 2011 document image binarization contest (dibco 2011). In: International Conference on Document Analysis and Recognition, pp. 1506–1510 (2011)

[22] Pratikakis, I., Gatos, B., Ntirogiannis, K.: Icdar 2013 document image binarization contest (dibco 2013). In: International Conference on Document Analysis and Recognition, pp. 1471–1476 (2013)

[23] Ntirogiannis, K., Gatos, B., Pratikakis, I.: Icfhr2014 competition on handwritten document image binarization (h-dibco 2014). In: International Conference on Frontiers in Handwriting Recognition, pp. 809–813 (2014)

[24] Pratikakis, I., Gatos, B., Ntirogiannis, K.: Icfhr 2012 competition on handwritten document image binarization (h-dibco 2012). In: International Conference on Frontiers in Handwriting Recognition, pp. 817–822 (2012)

[25] Pratikakis, I., Zagoris, K., Barlas, G., Gatos, B.: Icfhr2016 handwritten document image binarization contest (h-dibco 2016). In: International Conference on Frontiers in Handwriting Recognition, pp. 619–623 (2016)

[26] Pratikakis, I., Zagoris, K., Barlas, G., Gatos, B.: Icdar2017 competition on document image binarization (dibco 2017). In: International Conference on Document Analysis and Recognition, vol. 1, pp. 1395–1403 (2017)

[27] Pratikakis, I., Zagori, K., Kaddas, P., Gatos, B.: Icfhr 2018 competition on handwritten document image binarization (h-dibco 2018). In: International Conference on Frontiers in Handwriting Recognition, pp. 489–493 (2018)

[28] Pratikakis, I., Zagoris, K., Karagiannis, X., Tsochatzidis, L., Mondal, T., Marthot-Santaniello, I.: Icdar 2019 competition on document image binarization (dibco 2019). In: International Conference on Document Analysis and Recognition, pp. 1547–1556 (2019)

[29] Deng, F., Wu, Z., Lu, Z., Brown, M.S.: Binarizationshop: a user-assisted software suite for converting old documents to black-and-white. In: Joint Conference on Digital Libraries, pp. 255–258 (2010)

[30] Hedjam, R., Cheriet, M.: Historical document image restoration using multispectral imaging system. Pattern Recognition **46**(8), 2297–2312 (2013)

[31] Ayatollahi, S.M., Nafchi, H.Z.: Persian heritage image binarization competition (phibc 2012). In: International Conference on Pattern Recognition and Image Analysis, pp. 1–4 (2013)

[32] Xu, Y., Wang, X., Guo, S., Yu, J.: Chinese-seal dataset (csd): Diverse attributes for seal detection, segmentation and recognition. Segmentation and Recognition (2025)

[33] Zharikov, I., Nikitin, P., Vasiliev, I., Dokholyan, V.: Ddi-100: Dataset for text detection and recognition. In: International Symposium on Computer Science and Intelligent Control, pp. 1–5 (2020)

[34] Nandedkar, A.V., Mukherjee, J., Sural, S.: Spods: A dataset of color-official documents and detection of logo, stamp, and signature. In: International Conference on Computer Vision, Graphics and Image Processing, pp.

219–230 (2016)

[35] Buitrago, P.A., Toropov, E., Prabha, R., Uran, J., Adal, R.: Miikeminestamps: a long-tailed dataset of japanese stamps via active learning. In: International Conference on Document Analysis and Recognition, pp. 3–19 (2021)

[36] Li, M., Yue, X., Meng, L.: Yolo v4 for kuzushiji recognition with synthetic training data generated by gan. In: International Symposium on Advanced Technologies and Applications in the Internet of Things, pp. 37–47 (2023)

[37] Micenkov, B., Beusekom, J.: Stamp detection in color document images. In: International Conference on Document Analysis and Recognition, pp. 1125–1129 (2011)

[38] Yu, W., Liu, M., Chen, M., Lu, N., Wen, Y., Liu, Y., Karatzas, D., Bai, X.: Icdar 2023 competition on reading the seal title. In: International Conference on Document Analysis and Recognition, pp. 522–535 (2023)

[39] Jocher, G., Chaurasia, A., Qiu, J.: Ultralytics YOLOv8 (2023). https://github.com/ultralytics/ultralytics

[40] 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)

[41] Wang, A., Chen, H., Liu, L., Chen, K., Lin, Z., Han, J., *et al.*: Yolov10: Real-time end-to-end object detection. Advances in Neural Information Processing Systems **37**, 107984–108011 (2024)

[42] Jocher, G., Qiu, J.: Ultralytics YOLO11 (2024). https://github.com/ultralytics/ultralytics

[43] Ju, R.-Y., Cai, W.: Fracture detection in pediatric wrist trauma x-ray images using yolov8 algorithm. Scientific Reports **13**(1), 20077 (2023)

[44] Chien, C.-T., Ju, R.-Y., Chou, K.-Y., Xieerke, E., Chiang, J.-S.: Yolov8-am: Yolov8 based on effective attention mechanisms for pediatric wrist fracture detection. IEEE Access (2025)

[45] Chien, C.-T., Ju, R.-Y., Chou, K.-Y., Chiang, J.-S.: Yolov9 for fracture detection in pediatric wrist trauma x-ray images. Electronics Letters **60**(11), 13248 (2024)

[46] Ju, R.-Y., Chien, C.-T., Chiang, J.-S.: Yolov8-rescbam: Yolov8 based on an effective attention module for pediatric wrist fracture detection. In: International Conference on Neural Information Processing, pp. 403–416 (2024)

[47] Ju, R.-Y., Chien, C.-T., Lin, C.-M., Chiang, J.-S.: Global context modeling in yolov8 for pediatric wrist fracture detection. In: International Symposium on Intelligent Signal Processing and Communication Systems, pp. 1–5 (2024)

[48] Bento, J., Paixão, T., Alvarez, A.B.: Performance evaluation of yolov8, yolov9, yolov10, and yolov11 for stamp detection in scanned documents. Applied Sciences **15**(6), 3154 (2025)

[49] Sun, H., Tan, C., Pang, S., Wang, H., Huang, B.: Ra-yolov8: an improved yolov8 seal text detection method. Electronics **13**(15), 3001 (2024)

[50] 国文学研究資料館: 日本古典籍くずし字データセット. https://doi.org/10.20676/00000340

[51] Suh, S., Kim, J., Lukowicz, P., Lee, Y.O.: Two-stage generative adversarial networks for binarization of color document images. Pattern Recognition **130**, 108810 (2022)

[52] Ju, R.-Y., Lin, Y.-S., Chiang, J.-S., Chen, C.-C., Chen, W.-H., Chien, C.-T.: Ccdwt-gan: Generative adversarial networks based on color channel using discrete wavelet transform for document image binarization. In: Pacific Rim International Conference on Artificial Intelligence, pp. 186–198 (2023)

[53] Ju, R.-Y., Wong, K., Chiang, J.-S.: Mfe-gan: Efficient gan-based framework for document image enhancement and binarization with multi-scale feature extraction. arXiv preprint arXiv:2407.04231 (2024)

[54] Souibgui, M.A., Biswas, S., Jemni, S.K., Kessentini, Y., Fornés, A., Lladós, J., Pal, U.: Docentr: An end-to-end document image enhancement transformer. In: International Conference on Pattern Recognition, pp. 1699–1705 (2022)

[55] Saifullah, S., Agne, S., Dengel, A., Ahmed, S.: Coldbin: Cold diffusion for document image binarization. In: International Conference on Document Analysis and Recognition, pp. 207–226 (2023)

[56] 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)

[57] Ju, R.-Y., Lin, Y.-S., Jin, Y., Chen, C.-C., Chien, C.-T., Chiang, J.-S.: Three-stage binarization of color document images based on discrete wavelet transform and generative adversarial networks. Knowledge-Based Systems **304**, 112542 (2024)

[58] 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)

[59] Ju, R.-Y., Wong, K., Chiang, J.-S.: Efficient generative adversarial networks for color document image enhancement and binarization using multi-scale feature extraction. Asia Pacific Signal and Information Processing Association Annual Summit and Conference (2025)

[60] Niblack, W.: An Introduction to Digital Image Processing. Strandberg Publishing Company, Denmark (1985)

[61] Otsu, N.: A threshold selection method from gray-level histograms. IEEE Transactions on Systems, Man, and Cybernetics **9**(1), 62–66 (1979)

[62] Sauvola, J., Pietikäinen, M.: Adaptive document image binarization. Pattern recognition **33**(2), 225–236 (2000)

[63] McQueen, J.B.: Some methods of classification and analysis of multivariate observations. In: Proceedings of Berkeley Symposium on Mathematical Statistics and Probability, pp. 281–297 (1967)

[64] Goodfellow, I., Pouget-Abadie, J., Mirza, M., Xu, B., Warde-Farley, D., Ozair, S., Courville, A., Bengio, Y.: Generative adversarial networks. Communications of the ACM **63**(11), 139–144 (2020)

[65] Tan, M., Le, Q.E., *et al.*: Rethinking model scaling for convolutional neural networks. In: International Conference on Machine Learning, vol. 15 (2019)

[66] Isola, P., Zhu, J.-Y., Zhou, T., Efros, A.A.: Image-to-image translation with conditional adversarial networks. In: IEEE Conference on Computer Vision and Pattern Recognition, pp. 1125–1134 (2017)

[67] Zhu, J.-Y., Park, T., Isola, P., Efros, A.A.: Unpaired image-to-image translation using cycle-consistent adversarial networks. In: IEEE International Conference on Computer Vision, pp. 2223–2232 (2017)

[68] Gulrajani, I., Ahmed, F., Arjovsky, M., Dumoulin, V., Courville, A.C.: Improved training of wasserstein gans. Advances in Neural Information Processing Systems **30** (2017)

[69] Robbins, H., Monro, S.: A stochastic approximation method. Annals of Mathematical Statistics, 400–407 (1951)

[70] Kingma, D., Ba, J.: Adam: A method for stochastic optimization. In: International Conference on Learning Representations (2015)