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 several recent 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, two state-of-the-art (SOTA) Generative Adversarial Network (GAN) methods, as well as our Conditional GAN (cGAN) baseline. The DKDS dataset and the implementation code for baseline methods are available at https://ruiyangju.github.io/DKDS.

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

1 Introduction

Pre-modern Japanese documents were mainly written in Kuzushiji[1], which is a 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 [2]. 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

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

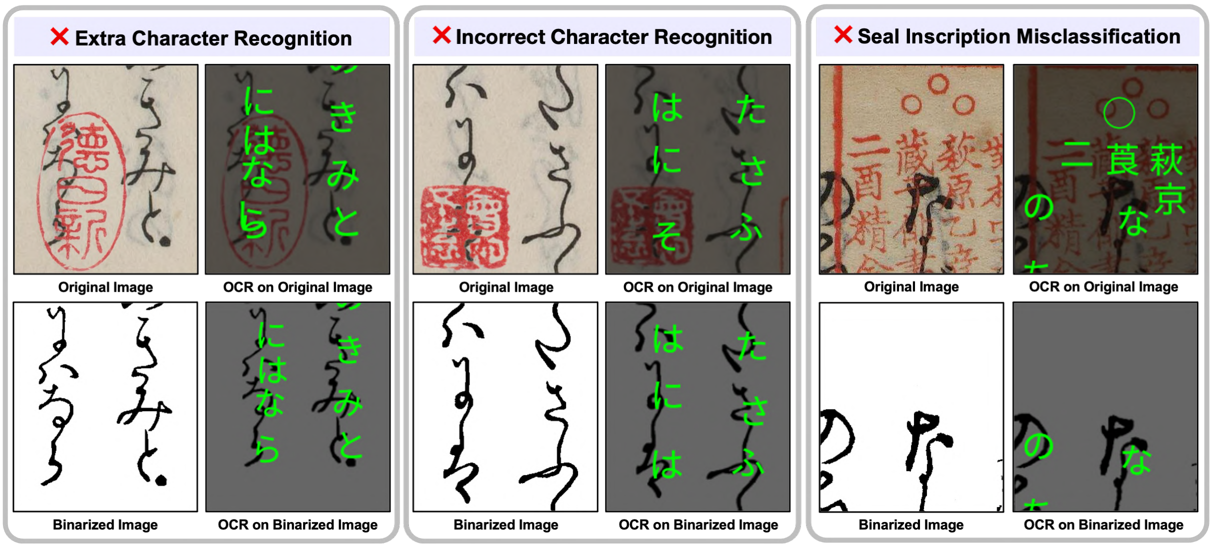

Fig. 1 Comparison of Optical Character Recognition (OCR) results on Kuzushiji characters overlapping with seals between the original and binarized images. OCR was conducted using the "miwo" app [1]. From left to right, the observed OCR errors include recognition of extra character, recognition of incorrect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

Japanese speakers are unable to read Kuzushiji texts without specialized training [3].

With the remarkable advances in deep learning for computer vision, researchers have increasingly applied neural network models to Kuzushiji character recognition. The primary datasets for Kuzushiji character recognition include Kuzushiji-MNIST, Kuzushiji-49, and Kuzushiji-Kanji[2] [4]. The Kaggle competitions organized using these datasets have further advanced both research and practical applications [5, 6]. 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 [7], thereby advancing the development of Kuzushiji character recognition and enhancing model performance.

Meanwhile, the Center for Open Data in the Humanities (CODH) has developed several Kuzushiji OCR systems that address both character-level and page-level recognition scenarios. In 2021, CODH introduced the mobile application "miwo" [1], which enables users to capture or upload images of Kuzushiji documents and automatically transcribe them into modern Japanese. CODH also released KuroNet [8, 9], an end-to-end OCR model based on a Residual U-Net architecture [10]. Designed to capture long-range context, process extensive vocabularies, and accommodate non-standard character layouts, KuroNet supports robust recognition of pre-modern Japanese texts. On Kaggle's Kuzushiji datasets [5, 6], KuroNet achieved classification accuracies exceeding 90%. More recently, SakanaAI introduced Metom [11], a Vision Transformer [12]-based Kuzushiji classifier specialized for single-character recognition. Metom was trained on 2,703 types of Kuzushiji characters that appear at least five times in the large-scale Kuzushiji document provided by CODH [13], offering a powerful and extensible foundation for character-level analysis.

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

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

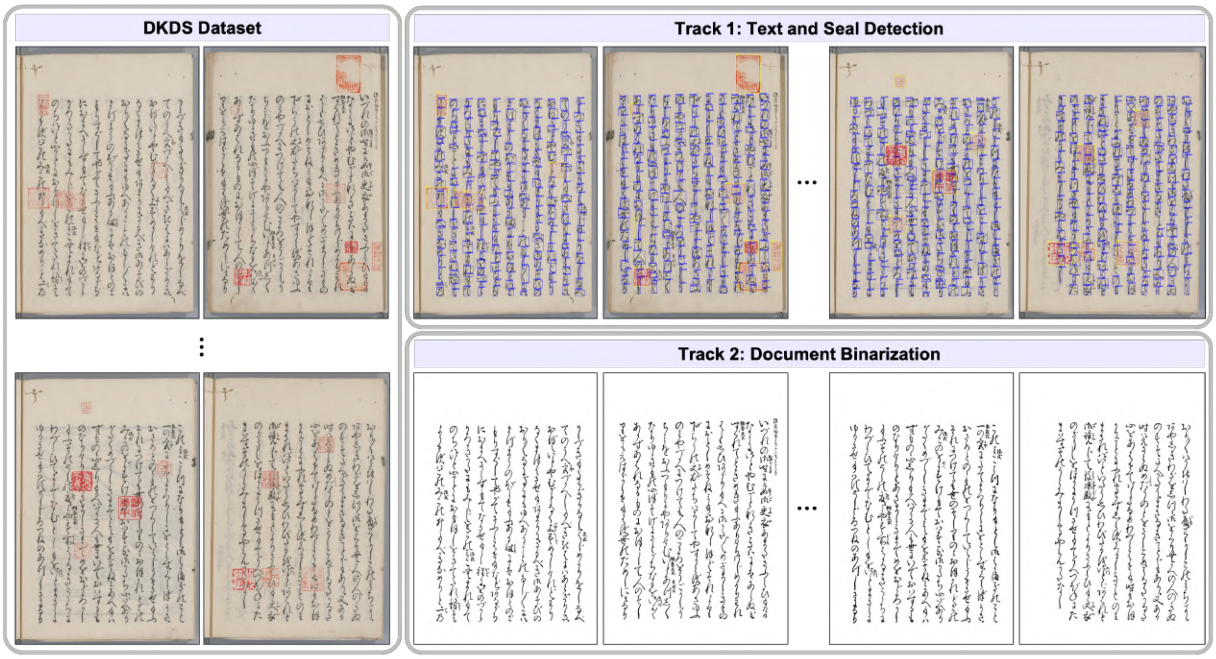

Fig. 2 **DKDS** dataset is the first collection of degraded pre-modern Japanese document images specifically designed to address the challenge of Kuzushiji characters overlapping with seals. Based on the dataset, we define two benchmark tracks: (1) Text and Seal Detection, and (2) Document Binarization.

(white), has been widely adopted as a preprocessing step, as it can effectively improve OCR performance.

Seals, serving as symbols of ownership and collector identity, are widely used in ancient and pre-modern Asian documents [14]. In these documents, seals usually appear as red marks inscribed with stylized ancient scripts, and may contain the collector's name as well as phrases reflecting personal wishes, interests, or social status. Among the various types of noise, the overlap between Kuzushiji characters and seals poses a major challenge for text and seal detection as well as document binarization.

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 the subsequent OCR task.

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 overlapping 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 several recent 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 two state-of-the-art (SOTA) Generative Adversarial Network (GAN) methods, as well as our Conditional GAN (cGAN) baseline.

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 document binarization datasets, and (3) seal datasets.

Existing research [4] on pre-modern Japanese documents written in Kuzushiji 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 document binarization focus on documents written in Latin-based alphabetic systems. For instance, the documents in the Document Image Binarization Contest (DIBCO) series datasets [15–24] are primarily written in English, while the Bickley Diary (BD) [25], Synchromedia Multispectral Ancient Document Images (SMADI) [26], and Persian Heritage Image Binarization Dataset (PHIBD) [27] all contain documents written in alphabetic scripts. 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. Yang *et al.* [28] synthesized paired data by overlaying seal images onto clean documents; however, their synthesized documents that primarily contain modern Chinese characters, which differ significantly from the pre-modern Japanese Kuzushiji characters targeted in this work. Following a similar strategy, we also simulate seal-induced degradation by randomly adding seal images to collected pre-modern Japanese documents. Regarding the source of seal images, existing publicly available seal datasets are unsuitable for this purpose, as they either do not contain ancient Asian seals, suffer from poor image quality, or exhibit complex backgrounds. For instance, MiikeMineStamps [29] includes some Japanese seal images; however, 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 as well as text analysis and recognition. As a critical preliminary step in Kuzushiji character recognition, text detection is particularly important. For instance, DBNet [30], a widely used text detection method in OCR systems, achieved an F-measure of 85.4% at 26 frames per second on the ICDAR 2015 [31] dataset, demonstrating real-time performance with high detection accuracy.

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 [14]. Micenkov *et al.* [32] proposed an automatic seal segmentation system for insurance documents. When evaluated on the collected dataset of 400 document images, named StaVer, their system achieved a recall of 83% and a precision of 84%. Yu *et al.* [33] 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 documents written in Kuzushiji.

The YOLO series of models [34–37] achieve an excellent balance between detection accuracy and model size, demonstrating outstanding performance across a wide range of object detection tasks. Researchers have explored the application of the YOLO series of models in text detection and seal detection. Nevertheless, due to the lack of ancient seal data from pre-modern Japanese documents, the recent versions of YOLO models cannot be directly applied to this specific scenario.

2.3 Document Binarization

Document binarization is an important preprocessing step for OCR, as it effectively removes various types of noise, such as paper yellowing, text fading, ink bleeding, stains, and seals. Early work in this area primarily relied on threshold-based methods. Specifically, Otsu [38] introduced the global thresholding algorithm in 1979. Subsequently, Niblack [39] and Sauvola [40] proposed local adaptive thresholding techniques in 1985 and 2000, respectively, enabling more robust binarization of documents with non-uniform illumination and complex background variations.

To facilitate systematic evaluation, the Document Image Binarization Contest (DIBCO), one of the most popular competitions in this field, released benchmark datasets from 2009 to 2019 [15–24]. These datasets have played an important role in advancing both traditional and deep learning-based binarization methods by providing standardized document image evaluation benchmarks.

The introduction of GANs [41] has enabled the generation of high-quality binary document images. Suh *et al.* [42] proposed a two-stage GAN framework that employed six enhanced CycleGANs [43] to produce binarized document images. Building on this line of work, Ju *et al.* [44, 45] further introduced a three-stage GAN architecture, was also composed of six improved CycleGAN

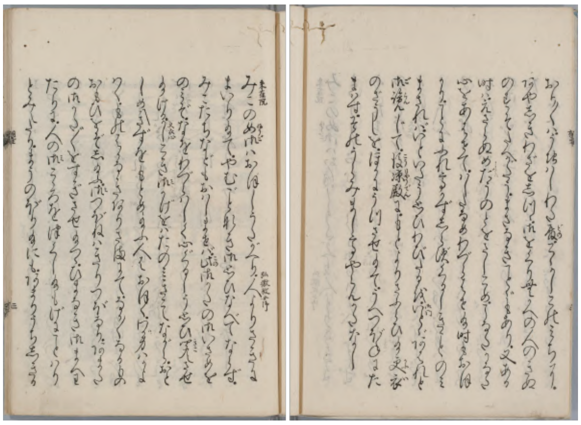

Fig. 3 Examples of raw Kuzushiji document data from the book *Genji Monogatari* (*The Tale of Genji*) [13].

modules, to advance document binarization performance. Both methods consistently achieved SOTA results on the DIBCO benchmarks.

However, neither traditional binarization methods nor SOTA GAN-based methods have been evaluated on degraded pre-modern Japanese documents written in Kuzushiji, particularly those containing seals, overlapping text, and complex degradations. This gap highlights the need to develop datasets and models that are tailored to the unique characteristics of pre-modern Japanese documents.

3 Dataset Construction

3.1 Raw Data Collection

The raw Kuzushiji document data were obtained from the CODH [13]. 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 also be used for training and evaluating OCR models in future research.

Furthermore, to date, there is no publicly available high-resolution dataset of ancient

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

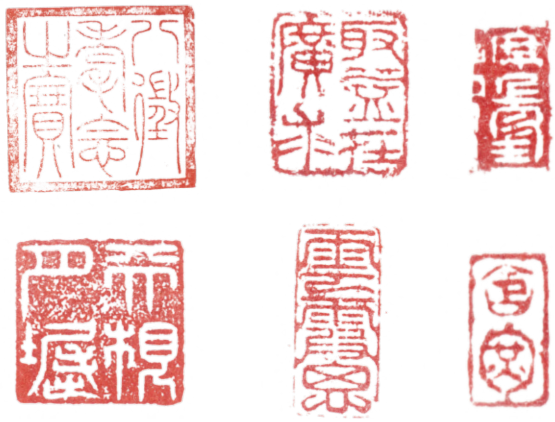

Fig. 4 Examples of our collected imperial seals from the Qing dynasty, with backgrounds removed, which are used to simulate seal interference.

Japanese seals. To address this gap, we collected 219 seals belonging to three Chinese ancient 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, and no high-resolution images of ancient Japanese seals are currently publicly available.

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.

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,

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

Table 1 Statistics of the proposed DKDS dataset.

Category	Total	Training	Testing-E	Testing-D
Document	50	40	10	10
Seal	128	100×2	28	28×5
S/D	-	$4-6$	$2-3$	14

S/D denotes the number of seals per document.

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 to the 40 training Kuzushiji documents. As a result, each Kuzushiji document image in the training set contains approximately 4–6 seals. In addition, we constructed easy-level (Testing-E) and difficult-level (Testing-D) versions of the test set according to the number of seals per document. Each document image in Testing-E contains approximately 2–3 seals, while each document image in Testing-D contains 14 seals.

To generate the binarization ground-truth, we employed a combined approach of neural network generation followed by manual verification and correction. As shown in Fig. 5, we first applied a pre-trained binarization model trained on the DIBCO benchmarks 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 shown 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 [13], while the bounding boxes for the seals were recorded during the process

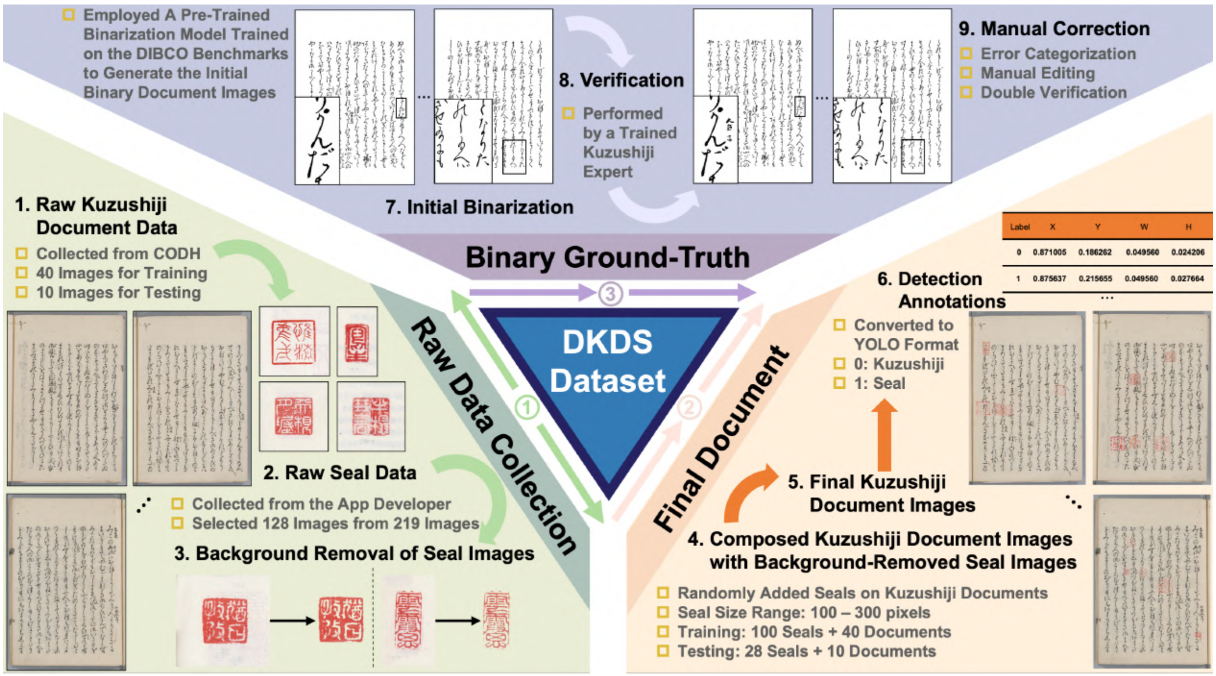

Fig. 5 The overall workflow of the proposed **DKDS** dataset construction includes raw data collection, text and seal detection annotations, initial binarization ground-truth generation, verification, and manual correction. The initial binarization ground-truth was generated using a pre-trained binarization model trained on the DIBCO benchmarks, while the verification was conducted by a trained Kuzushiji expert.

of randomly adding them to the Kuzushiji document images. Notably, the OCR annotations from CODH [13] 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 with other seals, and suffer from ink fading. These degradations can lead to incomplete seal detection or significantly reduce Kuzushiji character detection accuracy, making the detection 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 DKDS 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 (Track 1), we adopt standard evaluation metrics commonly used in text detection [30, 31], including the number of

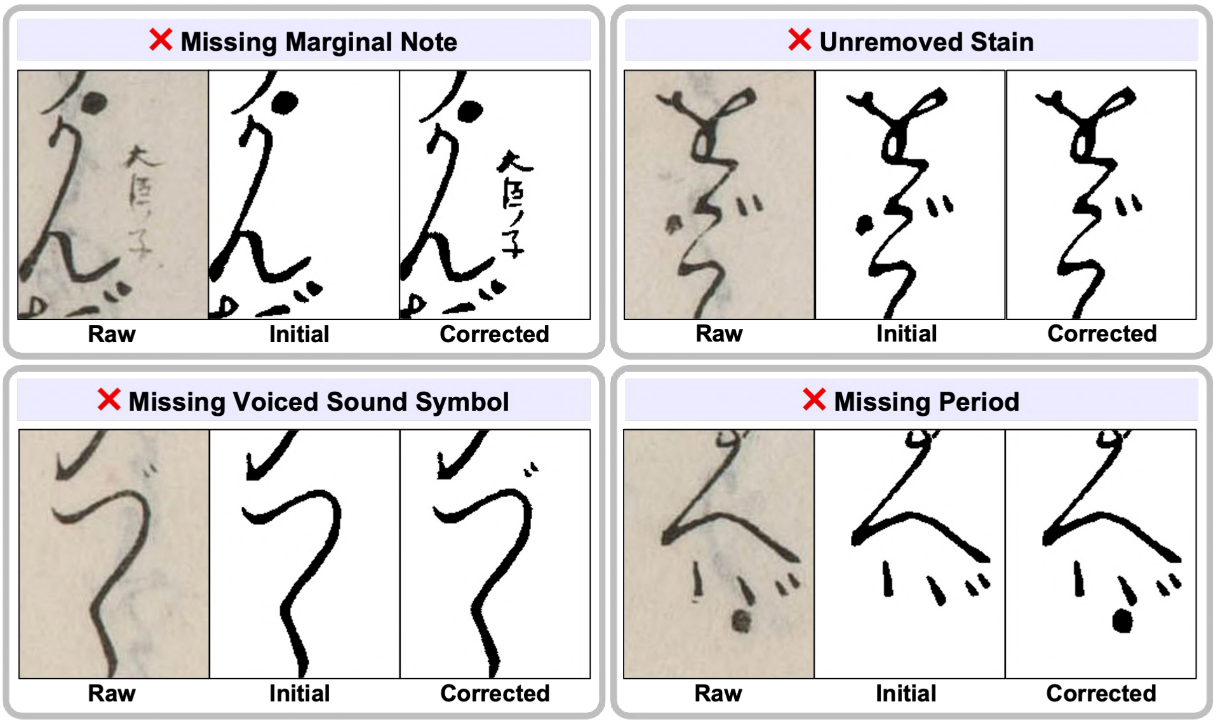

Fig. 6 Examples of verification and manual correction in the dataset construction: "Raw" refers to the input images, "Initial" denotes the initial binarized images produced by a pre-trained binarization model, and "Corrected" represents the final ground-truth images after verification and manual correction.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

model parameters (Params), floating-point operations (FLOPs), precision (P, in %), recall (R, in %), and F-measure (F, in %).

For the document binarization task (Track 2), we employ classical evaluation metrics [15] 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 adopt the Average-Score Metric (ASM) proposed by Ju *et al.* [46, 47], which is calculated as follows:

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

where a higher ASM value indicates better overall performance, providing a single unified metric.

4.3 Baselines

For the text and seal detection task (Track 1), we employ the YOLO series of object detection models, including YOLOv8 [34], YOLOv9 [35], YOLOv10 [36], and YOLO11 [37]. 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, provide fast inference, and present strong generalization capabilities, making them widely adopted in object detection tasks.

For the document binarization task (Track 2), we apply several traditional binarization algorithms to the Kuzushiji documents, including Niblack [39], Otsu [38], and Sauvola [40], 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 introduce a pre-processing step to remove seal regions based on

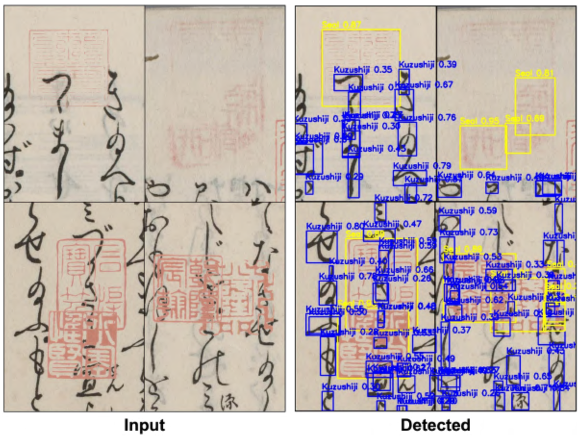

Fig. 7 Challenging examples in text and seal detection. The first row presents seals suffering from ink fading, while the second row shows seals overlapping with Kuzushiji characters or other seals.

K-Means clustering [48] 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 implement three combined approaches: K-Means + Otsu, K-Means + Niblack, and K-Means + Sauvola.

To provide deep learning-based baselines, we employ two SOTA GAN-based methods (i.e., Suh *et al.* [42] and Ju *et al.* [45]) to evaluate the proposed DKDS dataset.

In addition, we employ a cGAN [49]-based method 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 $\times$ 3200 pixels) would require excessive GPU memory, we divide each document image into multiple 512 $\times$ 512 patches; the details of this division are described in Section 5.1.2. These patches are then fed into a U-Net [10] architecture with an EfficientNet-B5 [50] backbone, which serves as the generator. A simplified PatchGAN [43, 49] discriminator is adopted to distinguish between real and fake (generated) images. Furthermore, following Suh *et al.* [42] and Ju *et al.* [45], we employ a loss function that combines the Wasserstein GAN with Gradient Penalty (WGAN-GP) loss [51] and an additional Binary Cross-Entropy (BCE) loss.

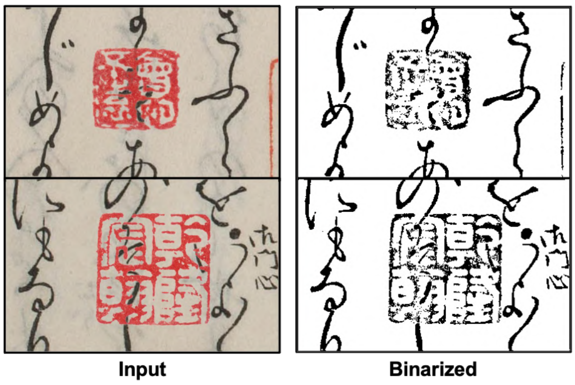

Fig. 8 Challenging examples in document binarization, where Kuzushiji characters overlap with seals.

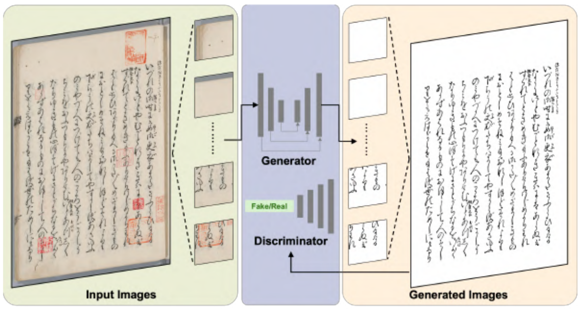

Fig. 9 The architecture of the cGAN-based method used.

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 GeForce 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 (Track 1), the YOLO models were trained for 100 epochs with a batch size of 16 and an input resolution of 640 $\times$ 640 pixels. Model training was performed using the SGD optimizer [52] with an initial learning rate of 0.01. All YOLO models were initialized with weights pre-trained on the COCO dataset [53].

Table 2 Quantitative comparison of medium-sized YOLO models for text and seal detection on the DKDS Testing-E set.

Model	Params	FLOPs	Kuzushiji			Seal		
			P	**R**	**F**	**P**	**R**	**F**
YOLOv8 (Ultralytics) [34]	25.86M	79.1G	95.5	89.1	92.2	95.6	96.4	96.0
YOLOv9 (ECCV2024) [35]	20.16M	77.5G	95.9	89.9	92.8	99.9	96.4	98.1
YOLOv10 (NeurIPS2024) [36]	16.49M	64.0G	93.5	89.9	91.7	99.6	96.4	98.0
YOLO11 (Ultralytics) [37]	20.05M	68.2G	97.1	92.4	94.7	97.0	92.9	94.9

Table 3 Quantitative comparison of medium-sized YOLO models for text and seal detection on the DKDS Testing-D set.

Model	Params	FLOPs	Kuzushiji			Seal		
			P	**R**	**F**	**P**	**R**	**F**
YOLOv8 (Ultralytics) [34]	25.86M	79.1G	90.5	87.9	89.2	92.9	81.4	86.8
YOLOv9 (ECCV2024) [35]	20.16M	77.5G	94.0	87.1	90.4	88.4	81.4	84.8
YOLOv10 (NeurIPS2024) [36]	16.49M	64.0G	90.4	87.1	88.7	96.4	77.1	85.7
YOLO11 (Ultralytics) [37]	20.05M	68.2G	94.7	89.1	91.8	98.5	84.3	90.8

5.1.2 Document Binarization

For the document binarization task (Track 2), traditional binarization algorithms (i.e., Niblack [39], Otsu [38], and Sauvola [40]) first convert the input images into grayscale 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 [48], the number of clusters is set to $K = 3$. After clustering, median filtering with a kernel size of 5×5 is applied to smooth the results and remove isolated noise points.

For the two SOTA GAN-based methods, we follow their original training data preprocessing strategies, resulting in 325,608 patches of size 256×256 pixels and 240 downscaled full-page images of 512×512 pixels. In terms of model architecture, we adopt a U-Net [10] generator for the method of Suh *et al.* [42], and employ a U-Net++ [54] generator for the method of Ju *et al.* [45]. Both methods use EfficientNet-B5 [50] as the encoder backbone.

For our cGAN [49]-based method, the training data are preprocessed by dividing each original image into 512×512 overlapping patches with a 30% overlap ratio. To further expand the training set, we apply 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 are expanded into a total of 71,458 patches of size 512×512 pixels. Both the generator and discriminator are optimized using the Adam [55] optimizer with a learning rate of 2×10^{-4} and β coefficients of $(0.5, 0.999)$. The model is trained for 10 epochs with a batch size of 16. At inference, the predicted outputs are binarized using a threshold of 0.5.

5.2 Track 1: Text and Seal Detection

5.2.1 Quantitative Comparison

We train and evaluate various YOLO series models on the proposed DKDS dataset, using the medium (M) size variant for all models, and the quantitative comparison results are shown in Tables 2 and 3.

On the Testing-E set, all models demonstrate strong overall performance. For Kuzushiji text detection, YOLOv11 achieves the best results with an F-measure of 94.7, maintaining an excellent balance between a precision of 97.1 and a recall of 92.4, and significantly outperforms YOLOv8, YOLOv9, and YOLOv10. In contrast, for seal detection, YOLOv9 achieves the highest performance with an F-measure of 98.1, closely followed by YOLOv10 with an F-measure of 98.0 and YOLOv8 with an F-measure of 96.0, while YOLOv11 exhibits comparatively lower performance with an F-measure of 94.9.

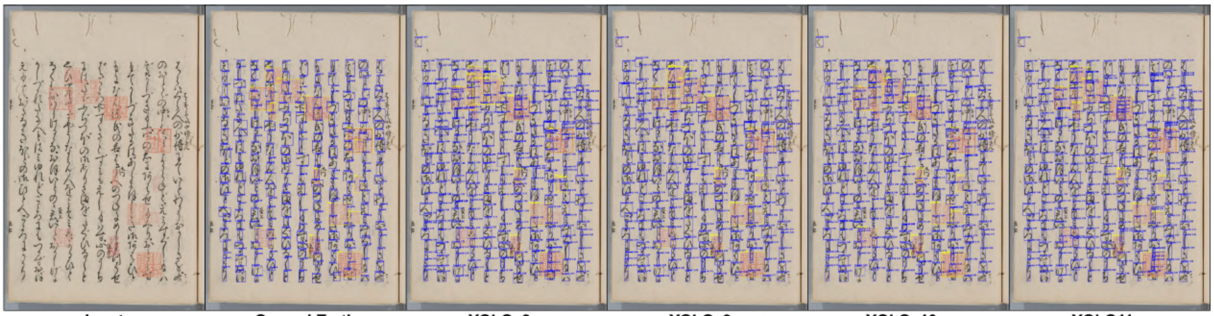

Fig. 10 Visual comparison of text and seal detection results predicted by different models, together with the corresponding ground-truth annotations. The input image is "200003803_00028_2" from the DKDS Testing-D set.

The Testing-D set contains a larger number of seals per document, which markedly increases the overall detection difficulty. For Kuzushiji text detection, YOLOv11 continues to achieve the highest F-measure of 91.8, demonstrating robust text modeling and detection capability even under complex degradation conditions. For seal detection, YOLOv11 shows an even more pronounced advantage, attaining an F-measure of 90.8 and clearly outperforming other models, including YOLOv8 with an F-measure of 86.8, YOLOv9 with 84.8, and YOLOv10 with 85.7.

Compared with Testing-E, the recall of seal detection decreases markedly for all models on Testing-D, indicating that incomplete seal detection is the primary source of the increased difficulty observed in Testing-D.

Regarding model efficiency, YOLOv10 has the smallest parameter count at 16.49M and the lowest computational cost at 64.0G FLOPs. Meanwhile, YOLOv11 demonstrates an excellent balance between efficiency and performance while maintaining high detection accuracy.

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.

It can be seen that YOLOv10m fails to detect a Kuzushiji character located at the bottom of the third column from the left, whereas the other models successfully detect all characters. In addition, all YOLO models incorrectly predicted stains in the upper-left corner as Kuzushiji characters, which constitutes a primary factor affecting detection performance. Furthermore, although the ground-truth annotations do not include positional information for marginal notes, all YOLO models detect several marginal notes to varying extents, as shown on the far right of the figure.

5.3 Track 2: Document Binarization

5.3.1 Quantitative Comparison

We employ 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 Tables 4 and 5.

On the Testing-E set, the traditional binarization algorithms exhibit generally poor performance. Specifically, Niblack's method [39] performs the worst, as indicated by its high DRD value, reflecting strong sensitivity to noise in the DKDS dataset. Otsu's method [38], a simple global thresholding method, achieves moderate results, while Sauvola's method [40] significantly outperforms both by leveraging locally adaptive thresholding to process complex backgrounds more effectively. Furthermore, applying K-means clustering [48] as a preprocessing step to remove red interference (e.g., seals) further improves performance. For instance, the ASM increases 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. These results indicate that when the number of seals

Table 4 Quantitative comparison of various methods for document binarization on the DKDS Testing-E set.

Method	Params	FM↑	p-FM↑	PSNR↑	DRD↓	ASM↑
Otsu [38]	–	63.01	63.31	11.76dB	37.69	50.10
Niblack [39]	–	39.13	41.14	8.44dB	79.70	27.25
Sauvola [40]	–	87.87	90.99	18.34dB	7.01	72.55
K-means [48] + Otsu [38]	–	84.76	86.28	17.14dB	9.90	69.57
K-means [48] + Niblack [39]	–	39.99	42.03	8.61dB	76.67	28.49
K-means [48] + Sauvola [40]	–	88.59	91.48	18.65dB	6.37	73.09
Suh *et al.* (PR2022) [42]	187.30M	93.09	93.09	20.99dB	3.12	76.01
Ju *et al.* (KBS2024) [45]	191.46M	95.80	95.81	23.14dB	1.76	78.25
Ours	31.22M	98.11	98.14	26.53dB	0.82	80.49

Table 5 Quantitative comparison of various methods for document binarization on the DKDS Testing-D set.

Method	Params	FM↑	p-FM↑	PSNR↑	DRD↓	ASM↑
Otsu [38]	–	60.51	60.80	11.39dB	40.89	47.95
Niblack [39]	–	37.52	39.41	8.28dB	82.29	25.73
Sauvola [40]	–	84.00	87.04	17.08dB	9.60	69.63
K-means [48] + Otsu [38]	–	68.10	69.04	13.37dB	33.32	54.30
K-means [48] + Niblack [39]	–	35.98	37.04	8.27dB	82.66	24.66
K-means [48] + Sauvola [40]	–	70.50	69.98	15.80dB	14.69	60.40
Suh *et al.* (PR2022) [42]	187.30M	91.95	91.96	20.10dB	3.85	75.04
Ju *et al.* (KBS2024) [45]	191.46M	94.86	94.90	22.11dB	2.27	77.40
Ours	31.22M	97.08	97.13	24.58dB	1.38	79.35

is limited, color-based clustering effectively mitigates seal interference for traditional thresholding methods.

However, when traditional methods are applied to the Testing-D set, which contains a larger number of seals, the limitations of combining K-means clustering with thresholding methods become more evident. Although K-means preprocessing still provides some improvement for simple thresholding approaches, overall performance deteriorates significantly. In particular, the FM, p-FM, and DRD metrics for both K-means + Otsu and K-means + Sauvola exhibit noticeable degradation. This suggests that in more complex scenarios, clustering-based binarization strategies struggle to robustly process the effects of severe degradation.

In contrast, deep learning-based methods consistently demonstrate superior performance over traditional methods. On the Testing-E set, the method of Suh *et al.* [42] achieves both FM and p-FM values of 93.09 with a DRD of 3.12, while the method of Ju *et al.* [45] further improves these results to FM and p-FM values of 95.80 and 95.81, and reduces the DRD to 1.76. Our method achieves the best overall performance across all evaluation metrics, attaining FM and p-FM values of 98.11 and 98.14, a PSNR of 26.53 dB, a DRD of 0.82, and the highest ASM of 80.49.

When evaluated on the more challenging Testing-D set, deep learning-based methods maintain strong robustness. The FM score of the method of Suh *et al.* [42] decreases from 93.09 to 91.95, while the method of Ju *et al.* [45] exhibits a reduction from 95.80 to 94.86. In comparison, our method experiences only a slight decrease from 98.11 to 97.08, while still achieving the lowest DRD value of 1.38. Notably, compared with the method of Ju *et al.* [45], our method further reduces the DRD on Testing-D by approximately 39%, decreasing it from 2.27 to 1.38, while consistently outperforming competing methods in FM, p-FM, and PSNR.

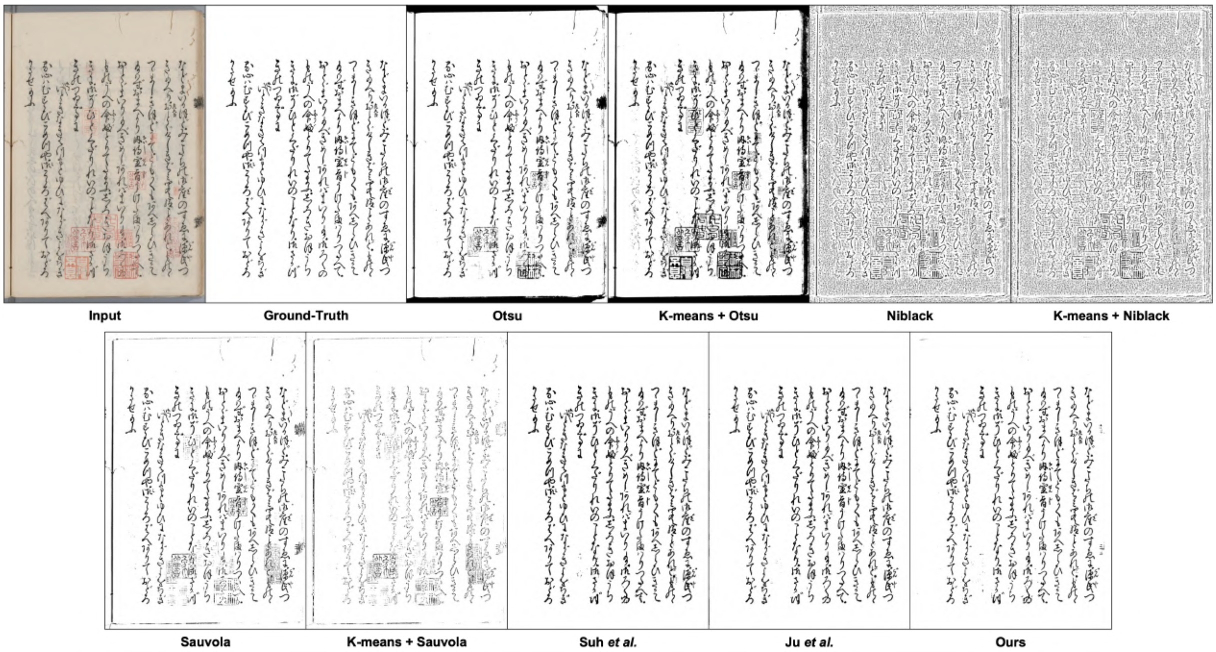

Fig. 11 Visual comparison of results generated by different methods on Kuzushiji document "200003803_00027_1" from the DKDS Testing-D set. The input image, ground-truth image, and various method outputs are shown for comparison.

Finally, in terms of model complexity, the methods of Suh *et al.* [42] and Ju *et al.* [45] contain 187.30M and 191.46M parameters, respectively, whereas our model requires only 31.22M parameters, achieving superior performance with a substantially reduced model size.

5.3.2 Visualization

We present visualized results generated by different methods, along with their corresponding input and ground-truth images, as shown in Fig. 11. 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 [39] produces the poorest results, with much noise remaining in the background. Otsu's method [38] achieves better separation between text and background but still retains some noise. Sauvola's method [40] performs relatively better but failed to completely remove seal interference.

Furthermore, as shown in Table 5, the incorporation of K-means clustering does not lead to a substantial performance improvement when documents contain a large number of seals.

In contrast, deep learning-based methods, including those proposed by Suh *et al.* [42], Ju *et al.* [45], and our method, achieve markedly better results by effectively suppressing noise and seal interference while preserving high text legibility.

5.4 OCR Evaluation

To quantitatively assess OCR performance on degraded and binarized documents, we employ the "miwo" app to recognize Kuzushiji characters. OCR is conducted on combined sets consisting of the training set and each version of the test set, with each evaluation involving a total of 50 document images.

The experimental results show that binarization leads to a substantial reduction in character error rate (CER). On the combined set of 50 images from the training set and the Testing-E set, the CER decreases by 27.6% in relative terms after binarization. Similarly, on the combined set of 50 images from the training set and the Testing-D set, the CER is reduced by 34.9%. Specifically, for the training set and Testing-E set, the CER of the binarized documents is 1.497%, compared with 2.067% for the original degraded documents. For the training set and Testing-D set, the CER of the original degraded documents reaches 2.301%, which is also markedly reduced after binarization (CER = 1.497%).

Notably, both the raw Kuzushiji document images [13] and the "miwo" OCR app [1] are provided by the Center for Open Data in the Humanities (CODH), which may partially account for the strong recognition performance observed on these documents.

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 and deep learning-based 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 Number 25H01242, and JST SPRING Grant Number JPMJSP2110.

Author Contributions

Rui-Yang Ju: Conceptualization, Data Curation, Formal Analysis, Methodology, Writing – Original Draft Preparation, Writing – Review & Editing; **Kohei Yamashita**: Methodology, 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] Clanuwat, T., Kitamoto, A.: 'miwo' ai kuzushiji recognition application for document examination. In: IPSJ Humanities and Computer Symposium (2021)

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

[3] 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)

[4] 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)

[5] 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)

[6] 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)

[7] 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/ (2023)

[8] 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)

[9] 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)

[10] 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)

[11] Imajuku, Y., Clanuwat, T.: Metom (2024). https://huggingface.co/SakanaAI/Metom

[12] 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)

[13] 国文学研究資料館: 日本古典籍くずし字データセット (2016). https://doi.org/10.20676/00000340

[14] 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)

[15] 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)

[16] 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)

[17] 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)

[18] 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)

[19] 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)

[20] 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)

[21] 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)

[22] 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)

[23] 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)

[24] 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)

[25] 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)

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

[27] 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)

[28] 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)

[29] 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)

[30] Liao, M., Wan, Z., Yao, C., Chen, K., Bai, X.: Real-time scene text detection with differentiable binarization. In: AAAI Conference on Artificial Intelligence, vol. 34, pp. 11474–11481 (2020)

[31] Karatzas, D., Gomez-Bigorda, L., Nicolaou, A., Ghosh, S., Bagdanov, A., Iwamura, M., Matas, J., Neumann, L., Chandrasekhar, V.R., Lu, S., *et al.*: Icdar 2015 competition on robust reading. In: International Conference on Document Analysis and Recognition, pp. 1156–1160 (2015)

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

[33] 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)

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

[35] 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)

[36] 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)

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

ultralytics

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

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

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

[41] 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)

[42] 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)

[43] 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)

[44] 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)

[45] 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)

[46] Ju, R.-Y., Wong, K., Chiang, J.-S.: Efficient generative adversarial networks for color document image enhancement and binarization using multi-scale feature extraction. In: Asia Pacific Signal and Information Processing Association Annual Summit and Conference, pp. 1898–1903 (2025)

[47] Ju, R.-Y., Wong, K., Jin, Y., Chiang, J.-S.: Mfe-gan: Efficient gan-based framework for document image enhancement and binarization with multi-scale feature extraction. arXiv preprint arXiv:2512.14114 (2025)

[48] 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)

[49] 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)

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

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

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

[53] 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)

[54] Zhou, Z., Siddiquee, M.M.R., Tajbakhsh, N., Liang, J.: Unet++: Redesigning skip connections to exploit multiscale features in image segmentation. IEEE Transactions on Medical Imaging **39**(6), 1856–1867 (2019)

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