





MiikeMineStamps: A Long-Tailed Dataset of Japanese Stamps via Active Learning

Paola A. Buitrago^{1,2}, Evgeny Toropov³, Rajanie Prabha^{1,2}, Julian Uran^{1,2}, and Raja Adal⁴

¹ Pittsburgh Supercomputing Center, Pittsburgh, PA 15203, USA

² Carnegie Mellon University, Pittsburgh, PA 15203, USA

³ DeepMap Inc., East Palo Alto, CA 94303, USA

etoropov@nvidia.com

⁴ University of Pittsburgh, Pittsburgh, PA 15260, USA

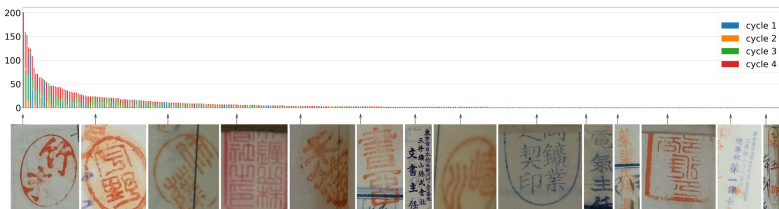


Fig. 1. Long tail distribution of stamps in the MiikeMineStamps dataset and stamp samples from selected classes. The labeling was performed in cycles across documents from different time periods.

Abstract. Mining existing image datasets with rich information can help advance knowledge across domains in the humanities and social sciences. In the past, the extraction of this information was often prohibitively expensive and labor-intensive. AI can provide an alternative, making it possible to speed up the labeling and mining of large and specialized datasets via a human-in-the-loop method of active learning (AL). Although AL methods are helpful for certain scenarios, they present limitations when the set of classes is not known before labeling (i.e. an open-ended set) and the distribution of objects across classes is highly unbalanced (i.e. a long-tailed distribution). To address these limitations in object detection scenarios we propose a multi-step approach consisting of 1) object detection of a generic “object” class, and 2) image classification with an open class set and a long tail distribution. We apply our approach to recognizing stamps in a large compendium of historical documents from the Japanese company Mitsui Mi’ike Mine, one of the largest business archives in modern Japan that spans half a century, includes tens of thousands of documents, and has been widely used by labor historians, business historians, and others. To test our approach we produce and make publicly available the novel and expert-curated MiikeMineStamps dataset. This unique dataset consists of 5056 images

of 405 different Japanese stamps, which to the best of our knowledge is the first published dataset of historical Japanese stamps. We hope that the MiikeMineStamps dataset will become a useful tool to further explore the application of AI methods to the study of historical documents in Japan and throughout the world of Chinese characters, as well as serve as a benchmark for image classification algorithms with an open-ended and highly unbalanced class set.

Keywords: Active learning · Object detection · Long tail · Open set · Stamp · Japanese · Historical · Dataset

1 Introduction

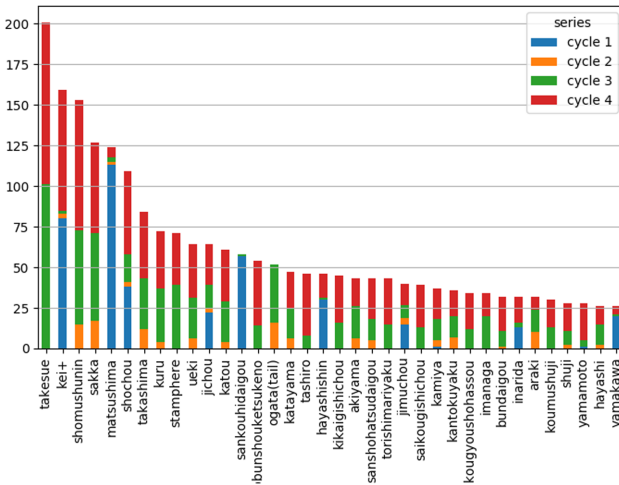


Fig. 2. The most frequent stamps by class as collected across active learning cycles.

Mining existing image datasets with rich information can help advance knowledge across various domains in the humanities, social sciences, and beyond. In East Asia, stamps often take the place of signatures. When opening a bank account or completing a contract, instead of signing one’s name, it is common to stamp it onto the account application or contract. Stamps are therefore the primary instrument for verifying one’s identity, but they have also been used for a number of other purposes. It is not uncommon for businesses and government offices to stamp the date onto a document, along with the name of the company or branch office, or the status of the document, such as “approved” or “top secret”. Documents emanating from East Asian government bureaucracies or businesses often feature multiple stamps on a single page. Mining these stamps opens unprecedented scenarios, making it possible to transform a document archive into a rich dataset that can reveal individual names, information flows, and interpersonal networks.

The Mitsui Mi'ike Mine archive is probably the most complete business archive for the study of modern Japan available today. Its uniqueness lies in its size, more than 30,000 pages, and its span, half a century ranging from 1889 to 1940. Without the aid of machine learning, mining the tens of thousands of stamps in this archive would require an expensive team of research assistants trained in reading the frequently stylized and hard-to-read Chinese characters that are used in East Asian stamps. The research assistants would need to open a photograph of each document, input a document identifier in a spreadsheet, and then work on recognizing the stamps that appear on that document. Since every document has, on average, several stamps, this would have to be repeated tens of thousands of times, requiring thousands of hours of work. However, the cost of such work would be secondary to the real challenge of finding, hiring, and training such a team of expert assistants.

During the past decade, machine learning has been widely used and applied to discovering and automating such tasks. The most promising type of algorithms falls under the supervised learning category [31]. These algorithms depend on the availability of large volumes of labeled data, making it possible to learn “by example”. Producing the much-needed labeled data traditionally requires an expensive and heavily involved process which can be prohibitive. The labeling challenge is particularly significant in domains that involve specialized knowledge. Active learning (AL), which is concerned with optimally selecting the next data samples to label based on feedback from prior iterations, has become a useful approach to making labeling possible while making a reasonable investment in time and effort. Until now, most AL research has been applied to classification rather than object detection. For AL in detection, however, the main area of focus is defining ideal criteria that make it possible to select ideal next candidates for labeling.

A significant limitation of the existing AL approaches for detection is that they do not consider open-class (i.e. undefined number of classes), long-tail data distributions (i.e. a large number of classes and few samples for a significant portion of them). Datasets exhibiting these characteristics are common and particularly challenging.

To facilitate the labeling work in scenarios like the one described here, we propose a method that leverages active learning concepts and popular algorithms in the area tuned to the application. The method relies on the following elements:

1. Break the task into two parts: detect generic “objects” and classify them.
2. Use a classification model to manage open-class, long-tail distributions.

We illustrate this method by applying it to the Mitsui Mi'ike Mine catalogue of historical documents, whose characteristics make it ideal for this type of work:

- Stamps share similar visual features. Previously unseen classes of stamps can still be identified by a generic stamp detector.
- The stamp class set is not known in advance.
- The long-tail distribution limits the accuracy of off-the-shelf object detectors.

In this work, we use AL to crop out and annotate stamps from the historical documents and produce the resulting MiikeMineStamps dataset. This unique

dataset contains 5056 images of 405 different Japanese stamps enriched with relevant domain-specific metadata. Figure 1 presents examples of the stamp images and the distribution of the stamp classes.

The contribution of this paper is therefore twofold. First, we introduce MiikeMineStamps, a unique dataset of stamps from Japanese historical documents, and second, we also present the application of a known AL approach for the object detection of datasets with open class and long-tail distributions. We trust that this dataset will become a useful tool to further study Japanese historical documents, as well as serve as a benchmark for image classification algorithms with highly unbalanced class sets.

2 Related Work

2.1 Kuzushiji and Stamps in Japanese Historic Documents

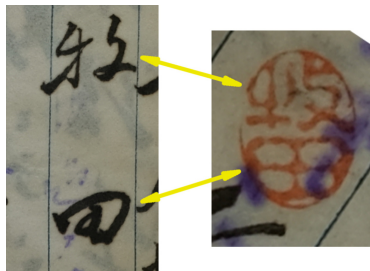


Fig. 3. Handwritten and stamped Chinese characters. The last name Makita looks very different when handwritten in cursive (left) and imprinted as a stamp (right).

Much of the leading-edge research in the recognition of the Chinese characters, which are used in Japan, China, Korea, and a few other parts of East Asia, has focused on the recognition of handwritten cursive script. Today, most Japanese is printed or handwritten in easy-to-read block or semi-cursive characters. Until the beginning of the twentieth century, however, most documents were either printed with woodblocks or handwritten with a brush using a cursive script known as *kuzushiji*. Not only does the kuzushiji cursive script link multiple characters, making it difficult to know where a character begins and where it ends, but there was no standard way for writing each character. Learning how to read a character meant learning three, four, or more ways in which it could be written. Since reading the kuzushiji cursive script requires special training, only trained archivists and historians are able to read it. Recently, however, the Center for Open Data in the Humanities (CODH) in Tokyo published a revolutionary machine-learning model known as Kuronet [4], which made it possible to read cursive kuzushiji script with an F-score in the range of 80% to 90% for most woodblock-printed books and with lower and sufficient accuracy for handwritten documents. This model has elicited enormous interest from archivists and historians in Japan and internationally.

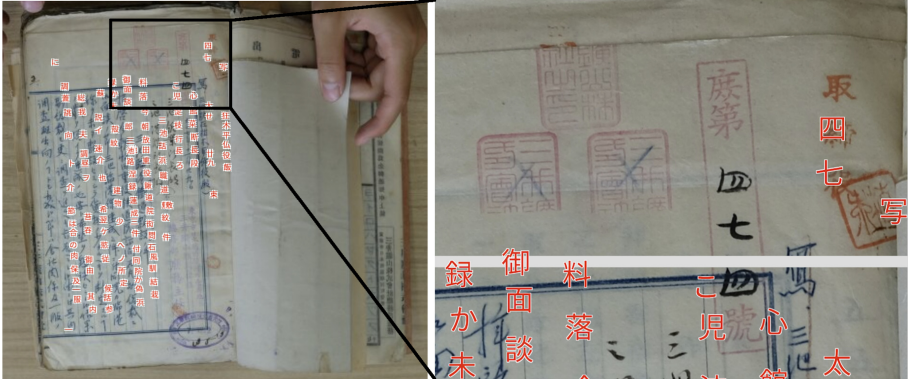


Fig. 4. The result of using Kuronet Kuzushiji Recognition Service on a document with stamps. Kuronet [4] successfully recognizes semi-cursive, but stamps are not detected at all and when they are, they are incorrectly recognized.

Machine learning models for reading kuzushiji cursive characters such as Kuronet, however, are not capable of recognizing the innumerable stamps that populate Japanese bureaucratic documents, as well as documents from China, Korea, and other parts of East Asia, from several thousand years ago to today. The scripts used to make stamps are stylized in ways that are very different from the kuzushiji cursive writing and use multiple, often archaic, fonts (Fig. 3). They can also combine multiple characters on a single stamp or can be used to simultaneously stamp two conforming copies of a document, such as a letter or a contract, so that the top half of the stamp appears on one document and the bottom half of the stamp on another. Figure 4 shows how Kuronet, a model created to recognize cursive handwritten or woodblock printed documents, is incapable of recognizing stamps. Although Kuronet successfully recognized portions of the semi-cursive writing, it did not recognize any of the stamps or even detect most of them. This is not surprising if we consider that Kuronet was never trained to recognize stamps.

As a result, a different model and a different dataset are needed for recognizing stamps in historical documents. The labeled dataset of stamps MiikeMineStamps together with the AI model to distinguish them fills this gap.

2.2 Active Learning

With the wide adoption of data-hungry deep learning methods, the need for large labeled datasets encouraged the development of Active Learning (AL) methods. AL aims to efficiently label large datasets in order to reduce the annotation cost. In AL, a small subset of data is annotated first, then an acquisition function selects the next batch to be annotated. A machine learning model trained on previously collected data helps the annotator by producing machine-generated labels, which the annotator verifies or corrects. The process repeats until the whole dataset has been labeled.

Until recently, AL research in computer vision focused primarily on image classification [2, 8, 14, 25, 26, 33, 37]. Only a few recent works explore AL in context of object detection [1, 12, 21, 22]. All these works focus on the optimal selection of the acquisition function.

A few works consider the class imbalance when applying AL. In [21], the authors address the class imbalance for the task of object detection in aerial images. The long tail distribution of classes is also explored in [9] in the image classification scenario. Most other AL approaches consider datasets with a small number of well-balanced classes, such as DOTA [35] with 15 classes, CityPersons [38] with 30 classes, PASCAL2007 [6] with 7 classes, BDD100K [36] with 10 classes, or CIFAR100 [15] with 100 classes.

In our dataset, the number of classes is not predefined during the cycles of AL (i.e. the open class set problem) reaching 405 by the final iteration. On every cycle, most classes contain only a handful of instances, making the dataset highly unbalanced (Fig. 2) and meaning that the object detectors used in existing AL work simply can not be bootstrapped.

2.3 Image Classification with Unbalanced Data

Over the past few years, convolutional neural networks (CNNs) have excelled on image classification tasks. These classic CNN architectures, however, only perform well on well-balanced academic datasets, such as ImageNet [5], CIFAR-100 [15], COCO [17], Caltech-256 [10], CelebA [18], VisualGenome [13], and others. Most of these datasets rarely capture the state of the real world in which highly skewed, unbalanced data prevails.

As a result, multiple few-shot learning algorithms [34] were introduced. Matching Networks [32], Prototypical Networks [27], and Model-Agnostic Meta-Learning [7] are some cutting-edge research papers that aim at solving the image classification problem with very few images or instances per class. In 2019, “Large-Scale Long-Tailed Recognition in an Open World” (OLTR) [19] was presented. It addressed the long-tail and open-set nature of real-world datasets. We compare three of the aforementioned models for the task of classifying stamps in our MiikeMineStamps dataset.

3 Methodology

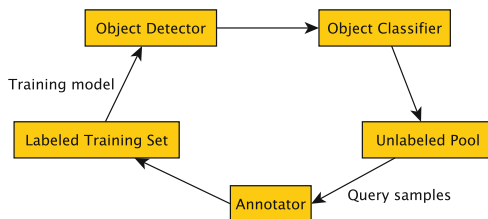


Fig. 5. The proposed AL pipeline. Arrows indicate the flow of information.

The proposed method (Fig. 5) follows the general principle of active learning. On every cycle, a machine learning model first predicts bounding boxes and object class for all unlabeled images in the dataset. Then we pick a subset of images based on an adjustable criterion. In our experiments, we favor images with a large number of objects that have high uncertainty. The images and the predictions are passed over to human experts to verify the labels and correct them if necessary. The ML model is then retrained on all the verified data available, and the cycle is considered complete.

In the case of the open class set, we do not know object classes beforehand and cannot train an object detector model that looks for a specific set of classes. Instead, we propose a two-step approach. First, an object detector model finds instances of the generic “stamp” class, then an image classification model is used to recognize a specific class in cropped out images of “stamps”. This approach provides the advantage of transferring the difficulty of dealing with open class sets and long-tail distribution from the detection to the classification setup, where there are more tools to manage it.

We apply the detection algorithm on the images to extract stamps and resize these stamps to 80×80 pixels. Then, the cropped stamps are individually passed to the image classifier. While any off-the-shelf object detector architecture can be taken for the “stamp” detection step, the image classification model must be able to handle the open class set and the long tail challenges. We assume the number of instances per class varies from one to several hundred. Furthermore, we assume the existence of previously unseen classes. We compared three image classification models: FaceNet [24], Prototypical Networks [27], and OLTR [19]. While FaceNet and Prototypical Networks produce reasonable results, these models do not address the long-tail class distribution, and their performance falls behind OLTR. We direct the reader to the respective work for the details of the architecture.

4 Experimental Results

4.1 Mitsui Mi’ike Mine Documents

We use the presented two-step active learning approach on a compendium of historical documents from the Mitsui Mi’ike Mine company. In this company, like in many other Japanese companies from this era, when a letter or other document crossed someone’s desk, it usually incurred a stamp, either to inscribe the name of the manager who approved it or to label it in some other way. Recognizing stamps across this archive will make it possible to trace the circulation of documents within this company. Considering that the full archive consists of more than thirty-two thousand pages and each document usually features multiple stamps, the automatic detection and classification of stamps is of considerable advantage.

Figure 6 (left) shows examples of the ground truth annotations in a page of a document. Documents were photographed as colored images with resolution 6000×4000 pixels. We collected annotations for the total of 677 images that have 5056 stamps. In Sect. 6, we present a dataset that consists of images of stamps, cropped from the original archive, annotated with stamp names and other metadata. The original image archive is not published in order to preserve the privacy of employees.

The active learning workflow follows a pipeline proposed in Sect. 3. Below, we describe the detection and classification components in detail.

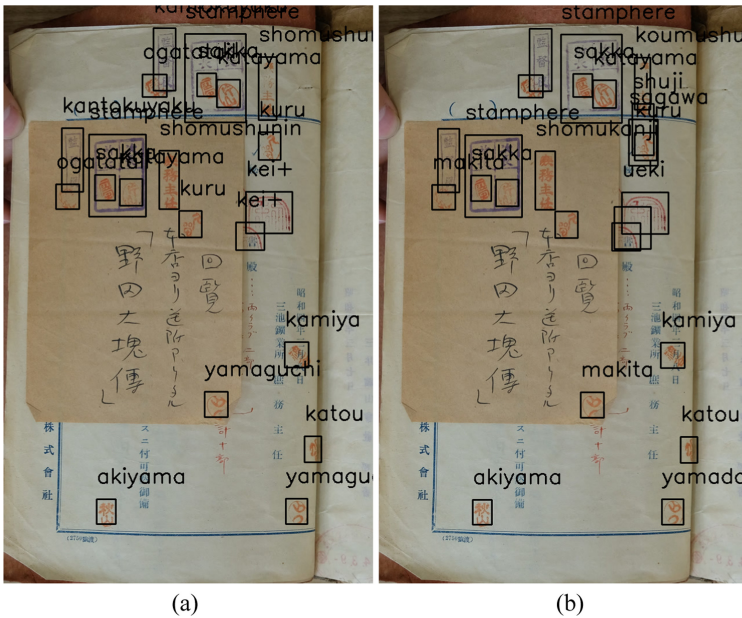


Fig. 6. (a): an example image with ground truth labels; (b): predictions of detector + classifier.

4.2 Detection

Given that the method is agnostic to the specific choice of a detector, we used a well-known RetinaNet [16] detector with ResNet-50 backbone, pretrained on COCO.

Hyperparameters were chosen via 5-fold cross-validation separately for every cycle. For the last cycle, the learning rate was set to $lr = 0.0001$ and batch size to $batch = 4$.

Table 1. Object detection average precision (@IoU = 0.5) across active learning cycles (%). Numbers on the main diagonal are the average “test” result in 5-fold cross-validation. Models trained on data from the 1st, 2nd, 3rd, and 4th cycle produce exceedingly better results when evaluated on the 4th cycle (in bold).

Trained on	Tested on			
	Cycle 1	Cycle 2	Cycle 3	Cycle 4
Cycle 1	89.3		31.6	44.7
Cycles 1–2		86.8	63.2	55.3
Cycles 1–3			83.5	75.0
Cycles 1–4				84.3

Table 1 tracks the performance of the object detector across AL cycles. The numbers on the main diagonal, i.e. trained and tested on the same cycle, are obtained from training with cross-validation. The three models trained on the first three cycles (in bold) perform increasingly better when trained on more data, showing that the active learning is gathering useful training data.

Figure 7 presents the precision-recall curves of detectors trained on cycles 1, 1–2 and 1–3, and evaluated on the last cycle 4. The detector performance can be seen to be steadily improving.

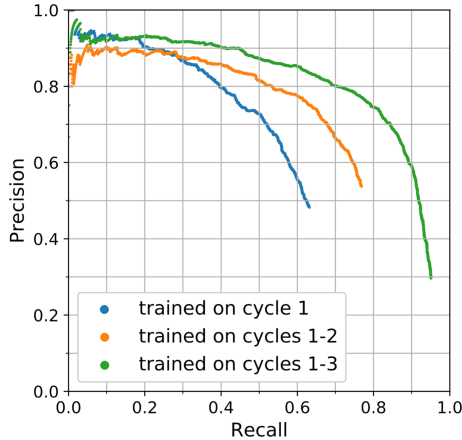


Fig. 7. Object detector trained on different cycles and evaluated on cycle 4.

4.3 Classification

Once stamps are detected via an object detector, the next task is to classify them. Figure 1 shows the high imbalance across classes. In fact, many classes have only a handful of examples. To overcome that, we picked classification models that are capable of working with long-tailed and open-set datasets. We

evaluate (1) FaceNet [24], (2) Prototypical Networks [27], and (3) OLTR [19] image classification models. We now describe the experiments with each of them.

FaceNet is a popular architecture designed to work with a high number of classes but few instances per class. We split all data from cycle 1 with classes having more than 2 instances into the train, validation, and test sets. The remaining classes with 2 or fewer instances were combined into a class called “other”, which was added to the test set. This gave us a total of 29 classes with 507 images in the train-val set and 30 classes and 135 images in the test set. We used Inception ResNet v1 as the backbone model for the FaceNet model with the softmax loss. We trained the model for 200 epochs on images of stamps resized 160×160 and generated embeddings in the 512-dimensional space. SVM was chosen as the last layer of FaceNet owing to its popularity [24,28]. It proved to be a better choice over the Random Forest classifier as per our experiments. After applying the SVM classifier in this embedding space, we achieved the test accuracy of 63% with RBF (radial basis function) kernel. The triplet loss failed to work because of the bias in the selection of the triplets. Randomly selected triplets do not lead to model convergence, and using the hardest triplets results in the model getting stuck in local minima. Additionally, adjusting class weights proved to have no effect on such a long-tailed dataset.

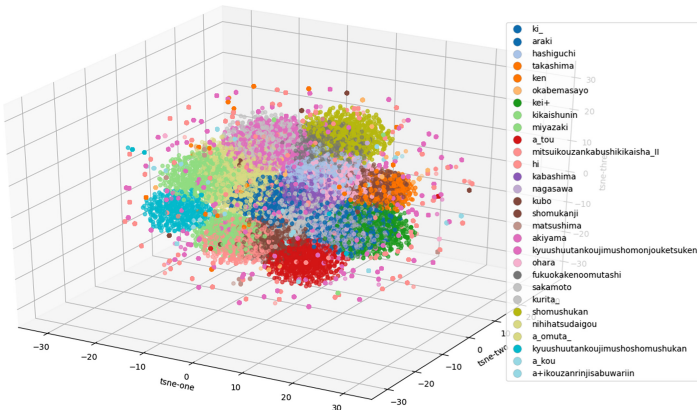


Fig. 8. Prototypical network: t-SNE on the test set (29 classes)

Owing to these limitations of the FaceNet model, we explored a few-shot learning architecture, specifically Prototypical Networks that is additionally tolerant to the long-tail data distribution. We trained this model with 5-shot, 5-query examples per class, and achieved the test accuracy of 69% for cycle 1 and 76% on the combined cycles 1 and 2 respectively. The t-SNE plot for the cycle 1 test set is shown in Fig. 8. Through this figure, we aim to illustrate that some classes form well-defined clusters in the t-SNE space, while other classes are highly diffuse. It graphically represents the class imbalance in the dataset.

The motivation behind exploring OLTR model was because of its ability to handle the open-set property of the dataset. This model promised successful results based on similarly distributed datasets, and henceforth will be used for future cycles of our dataset. All classes from cycle 1–3 with less than 3 image instances were moved to the open set (novel set). We used ResNet-10 [11] as our backbone and trained it with feature dimension 512 on 200 classes (many-shot, median-shot, low-shot combined). One important aspect of our work is reducing the labeling effort for subsequent cycles. To this end, subsequent cycles are automatically annotated with top-3 class predictions, given these predictions are above a certain confidence threshold. The expert can either choose from them or input their own class. Accordingly, we report top-3 and top-5 accuracy of 71.15% and 78.30% on the test set respectively. As expected, classes from the “many-shot” set perform better than classes from the “median-shot” set by 15%, which in turn perform better than classes from the “low-shot” set by another 15%. For the open-set (novel classes), we achieved 64% accuracy with the confidence threshold of 0.4. In order to assert robustness, we did 5-fold cross-validation for all models. A few examples of correct predictions by the OLTR model are presented in the top row of Fig. 10. The first stamp in the bottom row was incorrectly predicted to belong to either “kodama” or “sakka” classes, instead of the correct “kurihara” class. The last four stamps in the second row of the same figure illustrate the visual similarity between these three classes, which led to the incorrect prediction.

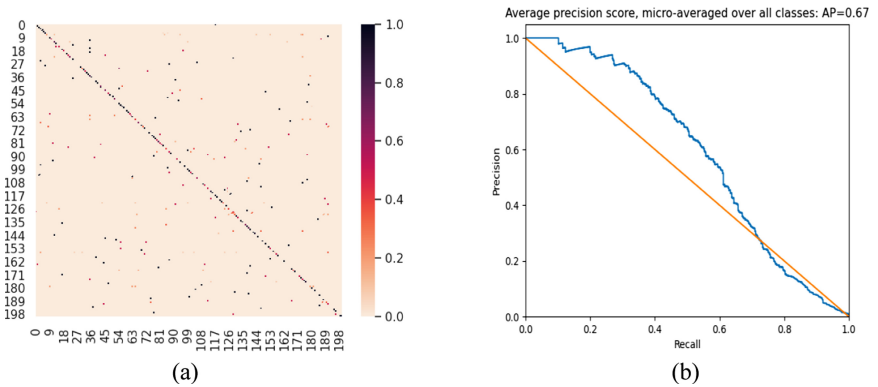


Fig. 9. OLTR model. Classification results for 200 random classes from cycle 3. (a) confusion matrix (b) precision-recall curve.

To sum up, OLTR performs best on our dataset as compared to FaceNet or Prototypical Networks, but the challenge of achieving high classification confidence ($> 50\%$) still exists. Figure 9(a) shows the confusion matrix for the test set of 200 classes and Fig. 9(b) shows the precision-recall curve showing that the threshold lies close to 0.4.

The OLTR model trained on all the available stamps is released together with the MiikeMineStamps dataset.

5 Technical Details

Each cycle of AL includes the manual labeling process that requires a domain expert to inspect and annotate labels for hundreds of images. A labeling tool was required for streamlining the annotation process and making it as fast and accurate as possible. Our research identified critical requirements for a labeling tool: web-based, open-source, and/or free of charge for the relevant volume of data, the ability to specify labels dynamically in the interface as opposed to choosing from a given set, the compatibility of the label files format, the ability to export, the support for uploading new or modified labels, and the usability of the interface.

As a result of comparing 17 different tools, the well-known LabelMe Annotation Tool [23] was chosen for this project. The comparison is released together with the code. The authors hope that it will be useful for future AL researchers.

Furthermore, active learning with thousands of objects presents the challenge of tracking changes in the datasets. As one example, the manual cleaning step after each labeling cycle included (1) expanding the bounding box around each stamp, (2) tiling stamps of the same class into one “collage” image, (3) exporting



Fig. 10. 1st row: Examples of correct classification. 2nd row: The first stamp “kurihara” was incorrectly classified as “kodama” or “sakka”. The last four stamps belong to the “kodama” and “sakka” classes. The visual similarity between these three classes explains the model’s incorrect prediction of the first stamp.

to LabelMe format, (4) importing the cleaned results from LabelMe, (5) back-projecting stamps from collages back into their original images, and (6) shrinking bounding boxes back to their original size. This pipeline as well as other work on managing datasets, including filtering, splitting and merging, visualization and querying, was performed using the Shuffler toolbox [29].

The project code is available at <https://github.com/pscedu/ml4docs>.

6 MiikeMineStamps Dataset

In this section, we describe the published MiikeMineStamps dataset.

Once the annotation process via AL was completed, the annotated stamps were cropped out of the original documents, resulting in 5056 images from 405 stamp classes. The average dimensions of a stamp are 167×257 pixels, but both width and height vary significantly from 27 pixels to 1200 pixels (Fig. 11b).

The distribution of the number of stamps by class is very unbalanced. The most common class, “takesue”, has 201 images, at the same time, 158 classes are represented by a single instance. Two stamps with the same letters but different shapes belong to the same class. The published dataset contains 14 such classes.

Additionally, the date on each original document was transferred to the stamps, which allowed us to track the flow of individual stamps over decades. Figure 12 illustrates this distribution for a small subset of stamps, while the full information is available in the published dataset.

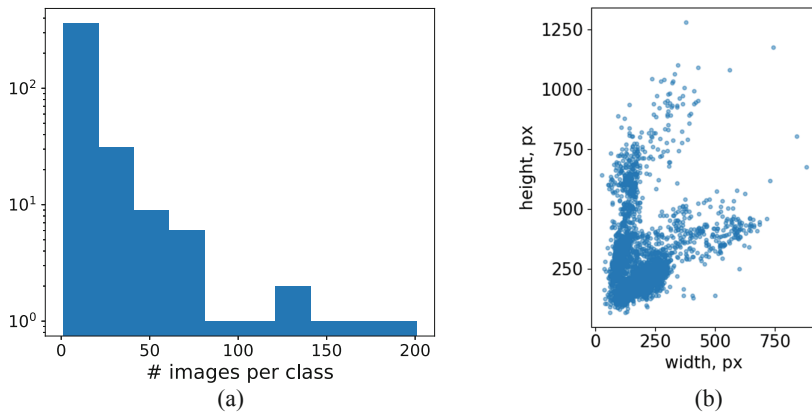


Fig. 11. (a) histogram of the number of stamp images per class; (b) distribution of stamps sizes.

The dataset introduced in this paper is publicly available under a Creative Commons Attribution 4.0 International license. The data is available for free to researchers for non-commercial use. This dataset includes the stamp images and labels. Additionally, we attach the information about the position of each stamp

relative to its page, and other useful details, such as the year of the source document. The dataset DOI is <https://doi.org/10.1184/R1/14604768>. More information on the dataset and how to retrieve it can be found at <https://kukuruza.github.io/MiikeMineStamps/>. The original images of the historic documents are not publicly available as they may contain sensitive and personally identifiable information.

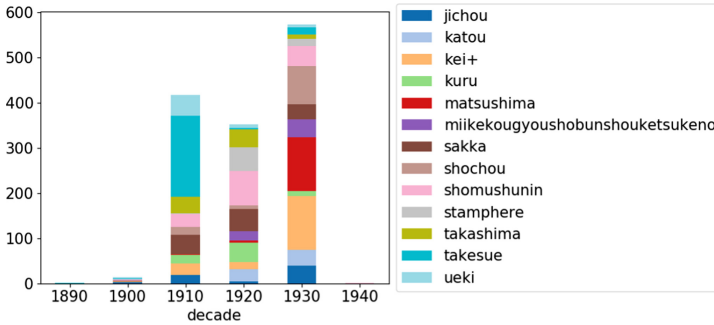


Fig. 12. Distribution of the most frequent classes across decades.

7 Discussion and Conclusion

The dataset of cropped stamps is interesting as it provides a completely different perspective on an archive. It instantly shows, for example, which stamps are most commonly used, providing clues as to who might be the gatekeepers of the organization. The benefits of this dataset increases considerably when stamps are classified into a series of classes and matched to the individual document(s) on which they appear. This will make it possible to identify all of the documents that came across the desk of an individual. Even more interesting is that stamps can show the way in which documents circulate in a company or government office. A memorandum will often circulate across the desk of multiple individuals, departments, or branches. At each location, it will usually incur a stamp that attests that someone has seen and approved it. Mining stamps on a large scale opens the door to tracing the circulation not only of one such document but of thousands of them. It helps to answer numerous questions in archives that feature a large number of stamps, not only in this archive of the Mitsui Mi'ike Mine but in most institutional archives in East Asia. For example, how does the circulation of documents change when a family-owned company becomes a joint-stock company? How does the circulation of documents in a ministry of foreign affairs change during wartime? Do the gatekeepers change? How is censorship implemented? What is the decision-making process in times of crisis? And more broadly, how do different bureaucratic decision-making processes lead to different outcomes? The answers to these questions are of interest to historians, political scientists, sociologists, anthropologists, media scholars, and researchers interested in the study of business management, among other fields.

The recent Large-Scale Long-Tailed Recognition in an Open World paper [19] presents long-tailed versions of three well-known datasets: ImageNet-LT, Places-LT, and MS1M-LT. In this work, we collected a naturally long-tailed dataset in the domain of documents, that we called MiikeMineStamps, which can serve as a benchmark for OLTR problems.

Acknowledgements. This work used the Extreme Science and Engineering Discovery Environment (XSEDE) which is supported by National Science Foundation grant number ACI-1548562. Specifically, it used the Bridges and Bridges-2 systems, which is supported by NSF award number ACI-1445606 and ACI-1928147, at the Pittsburgh Supercomputing Center (PSC) [3, 20, 30]. The work was made possible through the XSEDE Extended Collaborative Support Service (ECSS) program.

We are grateful to the Mitsui Archives for giving us permission to reproduce their documents and publish the stamps.

Finally, this work would not have been possible without the expert labeling and assistance of Ms. Mieko Ueda.

References

1. Aghdam, H.H., González-García, A., van de Weijer, J., López, A.M.: Active learning for deep detection neural networks. In: ICCV, pp. 3671–3679 (2019)
2. Beluch, W.H., Genewein, T., Nurnberger, A., Kohler, J.M.: The power of ensembles for active learning in image classification. In: CVPR, pp. 9368–9377 (2018). <https://doi.org/10.1109/CVPR.2018.00976>
3. Buitrago, P.A., Nystrom, N.A.: Neocortex and bridges-2: a high performance AI+HPC ecosystem for science, discovery, and societal good. In: Nesmachnow, S., Castro, H., Tchernykh, A. (eds.) High Performance Computing, pp. 205–219. Springer International Publishing, Cham (2021)
4. Clanuwat, T., Lamb, A., Kitamoto, A.: KuroNet: pre-modern Japanese Kuzushiji character recognition with deep learning. In: ICDAR, pp. 607–614 (2019)
5. Deng, J., Dong, W., Socher, R., Li, L.J., Li, K., Fei-Fei, L.: ImageNet: a large-scale hierarchical image database. In: CVPR (2009)
6. Everingham, M., Van Gool, L., Williams, C.K.I., Winn, J., Zisserman, A.: The PASCAL visual object classes (VOC) challenge. IJCV **88**(2), 303–338 (2010)
7. Finn, C., Abbeel, P., Levine, S.: Model-agnostic meta-learning for fast adaptation of deep networks. In: Precup, D., Teh, Y.W. (eds.) ICML, vol. 70, pp. 1126–1135 (2017)
8. Gal, Y., Islam, R., Ghahramani, Z.: Deep Bayesian active learning with image data. ICML **70**, 1183–1192 (2017)
9. Geifman, Y., El-Yaniv, R.: Deep active learning over the long tail (2017)
10. Griffin, G., Holub, A., Perona, P.: Caltech-256 object category dataset. CalTech Report, March 2007
11. He, K., Zhang, X., Ren, S., Sun, J.: Deep residual learning for image recognition. In: 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), pp. 770–778 (2016). <https://doi.org/10.1109/CVPR.2016.90>
12. Kao, C.-C., Lee, T.-Y., Sen, P., Liu, M.-Y.: Localization-aware active learning for object detection. In: Jawahar, C.V., Li, H., Mori, G., Schindler, K. (eds.) ACCV 2018. LNCS, vol. 11366, pp. 506–522. Springer, Cham (2019). https://doi.org/10.1007/978-3-030-20876-9_32

13. Krishna, R., et al.: The visual genome dataset v1.0 + v1.2 images. <https://visualgenome.org/>
14. Krishnamurthy, A., Agarwal, A., Huang, T.K., Daume, H., III, Langford, J.: Active learning for cost-sensitive classification. *JMLR* **20**(65), 1–50 (2019)
15. Krizhevsky, A., Nair, V., Hinton, G.: CIFAR-100 (Canadian Institute for Advanced Research)
16. Lin, T., Goyal, P., Girshick, R., He, K., Dollár, P.: Focal loss for dense object detection. In: *ICCV*, pp. 2999–3007 (2017)
17. 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: Fleet, D., Pajdla, T., Schiele, B., Tuytelaars, T. (eds.) *ECCV 2014*. LNCS, vol. 8693, pp. 740–755. Springer, Cham (2014). https://doi.org/10.1007/978-3-319-10602-1_48
18. Liu, Z., Luo, P., Wang, X., Tang, X.: Deep learning face attributes in the wild. In: *ICCV* (2015)
19. Liu, Z., Miao, Z., Zhan, X., Wang, J., Gong, B., Yu, S.X.: Large-scale long-tailed recognition in an open world. In: *CVPR* (2019)
20. Nystrom, N.A., Levine, M.J., Roskies, R.Z., Scott, J.R.: Bridges: a uniquely flexible HPC resource for new communities and data analytics. In: *XSEDE 2015: Scientific Advancements Enabled by Enhanced Cyberinfrastructure* (2015). <https://doi.org/10.1145/2792745.2792775>
21. Qu, Z., Du, J., Cao, Y., Guan, Q., Zhao, P.: Deep active learning for remote sensing object detection (2020)
22. Roy, S., Unmesh, A., Namboodiri, V.: Deep active learning for object detection. In: *BMVC* (2019)
23. Russell, B., Torralba, A., Murphy, K., Freeman, W.: LabelMe: a database and web-based tool for image annotation. *Int. J. Comput. Vis.* **77**, 157–173 (2008)
24. Schroff, F., Kalenichenko, D., Philbin, J.: FaceNet: a unified embedding for face recognition and clustering. *CoRR abs/1503.03832* (2015)
25. Sener, O., Savarese, S.: Active learning for convolutional neural networks: a core-set approach. In: *ICLR* (2018)
26. Sinha, S., Ebrahimi, S., Darrell, T.: Variational adversarial active learning. In: *ICCV*, pp. 5971–5980 (2019). <https://doi.org/10.1109/ICCV.2019.00607>
27. Snell, J., Swersky, K., Zemel, R.: Prototypical networks for few-shot learning. *NIPS* **30**, 4077–4087 (2017)
28. Taigman, Y., Yang, M., Ranzato, M., Wolf, L.: Deepface: closing the gap to human-level performance in face verification. In: *CVPR*, pp. 1701–1708 (2014). <https://doi.org/10.1109/CVPR.2014.220>
29. Toropov, E., Buitrago, P.A., Moura, J.M.F.: Shuffler: A large scale data management tool for machine learning in computer vision. In: *PEARC* (2019)
30. Towns, J., Cockerill, T., Dahan, M., Foster, I., Gafter, K., Grimshaw, A., Hazelwood, V., Lathrop, S., Lifka, D., Peterson, G.D., Roskies, R., Scott, J., Wilkins-Diehr, N.: XSEDE: accelerating scientific discovery. *Comput. Sci. Eng.* **16**(05), 62–74 (2014). <https://doi.org/10.1109/MCSE.2014.80>
31. Villalonga, G., Lopez, A.M.: Co-training for on-board deep object detection (2020)
32. Vinyals, O., Blundell, C., Lillicrap, T., Kavukcuoglu, k., Wierstra, D.: Matching networks for one shot learning. In: Lee, D., Sugiyama, M., Luxburg, U., Guyon, I., Garnett, R. (eds.) *NIPS*, vol. 29, pp. 3630–3638 (2016)
33. Wang, K., Zhang, D., Li, Y., Zhang, R., Lin, L.: Cost-effective active learning for deep image classification. *IEEE Trans. Circ. Syst. Video Technol.* **27**(12), 2591–2600 (2017). <https://doi.org/10.1109/TCSVT.2016.2589879>

34. Wang, Y., Yao, Q., Kwok, J., Ni, L.: Few-shot learning: a survey. arXiv preprint [arXiv:1904.05046](https://arxiv.org/abs/1904.05046) (2019)
35. Xia, G., et al.: DOTA: a large-scale dataset for object detection in aerial images. In: CVPR, pp. 3974–3983 (2018). <https://doi.org/10.1109/CVPR.2018.00418>
36. Xu, H., Gao, Y., Yu, F., Darrell, T.: End-to-end learning of driving models from large-scale video datasets. In: CVPR, pp. 3530–3538 (2017)
37. Yoo, D., Kweon, I.S.: Learning loss for active learning. In: CVPR, pp. 93–102 (2019). <https://doi.org/10.1109/CVPR.2019.00018>
38. Zhang, S., Benenson, R., Schiele, B.: CityPersons: a diverse dataset for pedestrian detection. In: CVPR, pp. 4457–4465 (2017). <https://doi.org/10.1109/CVPR.2017.474>