Small Patterns Detection in Historical Digitised Manuscripts Using Very Few Annotated Examples

Hussein Mohammed[a] and Mahdi Jampour[b]

Cluster of Excellence, Understanding Written Artefacts, Universität Hamburg, Hamburg, Germany

Keywords: Pattern Detection, Deep Learning, Historical Manuscripts, Datasets.

Abstract: Historical manuscripts can be challenging for computer vision tasks such as writer identification, style classification and layout analysis due to the degradation of the artefacts themselves and the poor quality of digitization, thereby limiting the scope of analysis. However, recent advances in machine learning have shown promising results in enabling the analysis of vast amounts of data from digitised manuscripts. Nevertheless, the task of detecting patterns in these manuscripts is further complicated by the lack of annotations and the small size of many patterns, which can be smaller than 0.1% of the image size. In this study, we propose to explore the possibility of detecting small patterns in digitised manuscripts using only a few annotated examples. We also propose three detection datasets featuring three types of patterns commonly found in manuscripts: words, seals, and drawings. Furthermore, we employed two state-of-the-art deep learning models on these novel datasets: the FASTER ResNet and the EfficientDet, along with our general approach for standard evaluations as a baseline for these datasets.

1 INTRODUCTION

Object detection is a task in computer vision that involves locating and identifying objects within an image or video. There have been several advances in object detection in recent years. One of the main areas of progress has been in the development of deep learning-based approaches, which have achieved state-of-the-art results on a number of benchmarks.

The use of visual-pattern detection in manuscript research is crucial for addressing various research queries. This technology enables scholars to efficiently explore digitised manuscripts, locating relevant images through specific patterns. It enhances searchability for textual content and visual elements like seals and drawings. Even when Handwriting Recognition (HTR) is viable, the patterns tied to research questions may pertain to specific visual styles within the handwriting itself, such as that of a particular scribe.

Detecting visual patterns in historical manuscripts presents challenges distinct from object detection tasks, where objects typically have clear boundaries. Unlike standard benchmarks with well-defined objects like animals or vehicles, patterns in manuscripts may lack distinct boundaries, making detection more challenging. The annotation process for such datasets is often more time-consuming due to unclear pattern boundaries, posing an additional challenge in pattern detection for historical manuscripts.

Applying deep learning models to object detection demands extensive training on labelled examples, specifying object locations and classes in each image. This requirement, vital yet costly, poses challenges, particularly for datasets needing specialized annotation. Researchers employ data augmentation to maximize annotated data use, but substantial examples per class remain essential, particularly for manuscript patterns differing significantly from those in standard benchmarks.

Digitised manuscript annotation typically requires expert supervision, often from relevant research fields, yet even with this, some annotations are subjective. Obtaining annotations for more than a few examples per pattern is challenging and sometimes impossible. Manuscript images often feature scripts understood by only a few humanities experts, making context-dependent patterns challenging for non-experts. Additionally, images may suffer degradation due to poor manuscript preservation or writing support nature, further complicating the annotation process.

[a] https://orcid.org/0000-0001-5020-3592
[b] https://orcid.org/0000-0002-1559-1865

Mohammed, H. and Jampour, M.
Small Patterns Detection in Historical Digitised Manuscripts Using Very Few Annotated Examples.
DOI: 10.5220/0012269500003654

In *Proceedings of the 13th International Conference on Pattern Recognition Applications and Methods (ICPRAM 2024)*, pages 605-612
ISBN: 978-989-758-684-2; ISSN: 2184-4313

Figure 1: Examples of detection by the proposed approach using three novel datasets of digitised historical manuscripts are shown. The detected patterns represent three common types found in manuscripts: seals, drawings, and words. The word examples have been enlarged for improved visibility.

Finally, patterns in digitised manuscripts often occupy a small image area, as illustrated in Fig. 1. This poses a known challenge in computer vision, where small objects may appear blurry or pixelated due to limited model input resolution, hindering accurate detection. Additionally, small objects may lack distinctive features, making them harder to identify, complicating the task of accurately distinguishing them from their surroundings for detection algorithms.

In this research, we present three fully annotated detection datasets featuring three types of patterns commonly found in manuscripts: words, seals, and drawings. We employed two state-of-the-art deep learning models on these novel datasets: a two-stage detector and a single-stage detector. Finally, we propose a general approach to improve detection performance on all datasets and evaluate it using standard object detection metrics to serve as a baseline for future studies.

2 RELATED WORK

There are two main approaches for object detection: two-stage and single-stage (Liu et al., 2020; Zaidi et al., 2022; Jiao et al., 2019). Two-stage approaches first identify regions of the image that are likely to contain objects, and then classify those objects and refine their locations. Examples include the Faster R-CNN (Ren et al., 2017) and the Region-based Fully Convolutional Network (R-FCN) (Dai et al., 2016). Single-stage approaches, on the other hand, aim to identify and classify objects in a single step, without first identifying regions that are likely to contain objects (Ren et al., 2015). Examples include the You Only Look Once (YOLO) (Redmon et al., 2016) and EfficientDet (Tan et al., 2020). Single-stage algorithms tend to be faster than two-stage algorithms, but may have lower accuracy.

The concept of using machine learning to automatically detect patterns in manuscript images has been around for at least a decade (Yarlagadda et al., 2011), but progress has been limited due to the lack of standard and publicly available datasets with ground-truth annotations. Additionally, the reliance of state-of-the-art methods on annotated training data has hindered the progress.

However, several pattern detection methods have been proposed to detect symbols, logos, and other types of patterns found in documents (Mohammed et al., 2021; Le et al., 2014; Wiggers et al., 2019). Some of these methods have been specifically designed to detect patterns in historical documents and manuscripts (Úbeda et al., 2020; En et al., 2016b), and optimized for certain types of patterns and manuscripts. More recently, a general training-

(a) (b)

Figure 2: (a) Example images and patterns from the SAM dataset. (b) An example from each of the selected seals. The most complete and clear instances are selected in this figure for better visibility.

free approach has been proposed by Mohammed *et al*.(Mohammed et al., 2021) for detecting patterns in manuscript images. The authors of this work argued that using a training-free approach can eliminate the problem of annotations availability and provide state-of-the-art results. While this approach may be useful for many scholars in manuscript research, it has two major drawbacks: first, performance can only be slightly enhanced by adding more examples per pattern, due to the lack of a training phase. Second, the hand-crafted features used in this research may not be useful in detecting some types of visual patterns.

There are several benchmark datasets that are commonly used for evaluating object detection algorithms. One widely used dataset is the PASCAL Visual Object Classes (VOC) dataset (Everingham et al., 2010), which consists of images annotated with bounding boxes around objects of 20 different classes. Another popular dataset for object detection is the Microsoft Common Objects in Context (COCO) dataset (Lin et al., 2014), which consists of images annotated with bounding boxes around objects of 90 different classes.

These datasets are typically not relevant for pattern detection in manuscript research, as the annotated objects are everyday items such as cars, planes, and animals. On the other hand, two challenging datasets for pattern detection in manuscripts have been published in the past few years: the *AMADI_LontarSet* dataset (Burie et al., 2016), which consists of handwriting on palm leaves for word spotting, and the *DocExplore* dataset (En et al., 2016a), which consists of medieval manuscripts for pattern detection. Despite being valuable contributions, the first dataset is highly unbalanced, very specific, and some queries are merely letters or other visual marks. In addition, the annotation is provided as part of the file name. The second dataset does not include any annotation. Therefore, there is a significant demand for datasets, especially historical data, intended for pattern detection.

3 DETECTION DATASETS

The primary motivation for creating these three datasets is to investigate the possibility of detecting medium to small patterns in digitised manuscripts using a small number of annotated examples while both small patterns and a low number of annotated data are open challenges. To this end, the datasets were chosen to represent different types of typical patterns found in manuscripts. All of the datasets are annotated using the Pascal VOC format and saved as XML files. All datasets are split into training, validation, and test sets; however, one can alter the splits based on the requirements of individual experiments.

The distribution of pattern instances per class in the training subset is kept balanced as much as possible in order to focus on the main research question and to make interpretation of results easier. Furthermore, the resolution of images in all datasets is kept high enough to preserve the visual features of small patterns. Finally, all images are saved in ".jpg" format to standardise any required image processing.

The main challenges in all of the datasets presented in this work are the extremely limited number of training samples (down to only three examples) per pattern and the small size of many instances compared to the image size. In addition, each dataset poses a different set of challenges, such as fading, low contrast, arbitrary orientation, interclass similarities, and etc.

3.1 Dataset of Seals in Arabic Manuscripts (SAM)

A dataset of seals in Arabic manuscripts has been created from the publicly available images of the "Staatsbibliothek zu Berlin" in (van Lit, 2020). Sample images of different seals are presented in Fig. 2a. Only seals with a minimum of 4 occurrences in different images have been selected, resulting in 8 different seals and 77 images in total. The complete statis-

(a) Example images from DMM dataset with various challenges such as fading, orientation.

(b) One example from each of the selected drawings in the DMM dataset.
Figure 3: Example images and patterns in the DMM dataset.

tics are provided in Table 1. One example from each of the selected seals is presented in Fig. 2b. The SAM dataset is made publicly available in a research data repository (Mohammed, 2023b) under the Creative Commons license. As can be seen from the presented examples, the main challenges for patterns in this dataset include their small size, fading, low contrast, arbitrary orientation, and interclass similarities.

3.2 Dataset of Drawings in Medieval Manuscripts (DMM)

A subset of 124 images has been selected from the *DocExplore* images (En et al., 2016a) in order to create a detection dataset of drawings in medieval manuscripts. Since the original dataset has been published without providing any annotations, we selected and annotated 8 different patterns in the subset, which resulted in a total of 268 annotated instances. The complete statistics are provided in Table 2. Sample images of different drawing are presented in Fig. 3a, and one example from each of the selected patterns is presented in Fig. 3b. The DMM dataset is made publicly available (Mohammed, 2023a) under the Creative Commons license. Some of the main challenges for patterns in this dataset include their small size, as well as the colour and scale variance of different instances for the same pattern.

(a) Example images.

(b) One example from each of the selected words. Complete and clear instances are selected in this figure for better visibility.
Figure 4: Example images and patterns from the WPM dataset.

3.3 Dataset of Words in Palm-Leaf Manuscripts (WPM)

A dataset of words from colophons found in palm-leaf manuscripts hailing from Tamil Nadu (a state in India) has been created from images provided by Centre for the Study of Manuscript Cultures (CSMC) for the manuscripts belonging to the Staats- und Universitätsbibliothek (SUB) Hamburg, and images provided by the Bibliothèque nationale de France (BnF), the library of the École française d'Extrême Orient (EFEO) in Pondicherry and the Cambridge University Library for their manuscript collections. All images and patterns are selected and annotated by Giovanni Ciotti from the CSMC within the scope of the activities of the Palm-Leaf Manuscript Profiling Initiative (PLMPI). A total of 10 words have been selected and annotated in 69 images. The complete statistics are provided in Table 3. Sample images from the WPM dataset are presented in Fig. 4a, and one example from each of the selected patterns is presented in Fig. 4b. The WPM dataset is made publicly available in a research data repository (Mohammed and Ciotti, 2023) under the Creative Commons license.

Table 1: Number of pattern instances in each subset within the SAM dataset.

Pattern / No. of	Train	Validate	Test
Seal 1	3	1	7
Seal 2	3	1	2
Seal 3	3	1	30
Seal 4	3	1	1
Seal 5	3	1	7
Seal 6	4	3	3
Seal 7	3	1	0
Seal 8	3	1	0
Total		85	

Table 2: Number of pattern instances in each subset within the DMM dataset.

Pattern / No. of	Train	Validate	Test
Corner Diamond	16	2	96
Letter A	4	1	9
Letter BP	4	1	7
Pine cone	16	1	12
Letter T	4	9	26
Letter D	4	1	34
Statue	8	2	2
Coat Shield	4	1	1
Total		115	

Table 3: Number of pattern instances in each subset within the WPM dataset.

Pattern / No. of	Train	Validate	Test
Varsa	8	1	2
Samvatsa	4	3	3
Yeluti	8	1	1
Srikosa	10	1	2
Karakrta	8	1	1
Svahasta-likhita	5	3	2
Naksatra	6	2	3
Subhadin-attil	5	1	5
Kutu	11	1	3
Pillai	6	2	6
Total		265	

The WPM dataset presents a unique set of challenges in addition to all the aforementioned issues mentioned in the other two datasets. The patterns in this dataset are extremely small compared to image size. If we scale down the images to fit the input size of the detection models, most of the annotated patterns will be represented by only few pixels with no meaningful visual features.

Furthermore, the annotated patterns themselves have no distinctive visual features to define them as objects with clear boundaries. Most of the visual features in each of these patterns exist also in other parts of the images (e.g. other words). In addition, the boundaries of these patterns can only be accurately detected after correctly classifying the patterns, because they are merely defined by the spacial relations between the visual features of these patterns (e.g. letters sequence). Moreover, the patterns in this dataset are handwritten words by different scribes on palm-leaves. Therefore, the handwriting style can differ greatly between different instances of the same pattern (word), and the texture of the writing support (leaf) can differ significantly as well.

4 PROPOSED APPROACH

Two state-of-the-art models are employed to execute and evaluate the suggested approach on the three datasets introduced in this study. The initial model is Faster R-CNN (Ren et al., 2017), which exemplifies the two-stage methodology. In its first stage, Faster R-CNN employs a region proposal network (RPN) to produce a collection of region proposals, i.e., potential object locations. In the second stage, these region proposals are passed through a classifier to determine the class and location of the objects within the image. The use of a two-stage approach allows for greater accuracy and efficiency in object detection compared to single-stage approaches. Faster R-CNN also incorporates a ResNet (He et al., 2016) architecture, which utilizes skip connections and batch normalization to improve the accuracy and efficiency of the model. the ResNet50 variant will be used for the rest of this work, as incremental gains are not the focus of this research.

The second model is EfficientDet (Tan et al., 2020), which represents the single-stage approach. This model performs object detection in a single stage using a single neural network. This allows for faster inference times and a simpler overall architecture. Additionally, EfficientDet utilizes a weighted bi-directional feature pyramid network (BiFPN) to efficiently combine multi-scale feature maps, leading to improved performance on small objects.

Table 4: Detection results of Faster R-CNN and Efficient-Det models using transfer learning, fine tuning and data augmentation on the SAM, DMM and WPM datasets.

Model	Metric	SAM	DMM	WPM
Faster R-CNN ResNet	COCO mAP	0.84	0.56	≈ 0.0
	mAP@0.5	0.99	0.97	≈ 0.0
	Recall@1	0.79	0.47	≈ 0.0
Efficient-DetD1	COCO mAP	0.77	0.53	≈ 0.0
	mAP@0.5	0.97	0.86	≈ 0.0
	Recall@1	0.75	0.42	≈ 0.0

The standard parameter values mentioned in the original publications of the corresponding models are used in all our experiments. However, the number of training steps is fixed at 10 thousand in order to make the results of different experiments comparable and to speed up the training phase for all experiments. The details of all used parameters and configurations for both models are published in (Mohammed, 2023c) as public research data.

As an evaluation metric, we used the coco mAP metric which is the average value of the calculated mAPs at IoU thresholds ranging from 0.5 to 0.95 with a step of 0.05. In addition, we provided other metrics in our base results such as mAP at 0.5 and 0.7, and recall rate.

4.1 Learning from Few Examples

Transfer learning, a valuable technique for limited annotated datasets (Li et al., 2020), enhances object detection performance, particularly with few annotated images (Talukdar et al., 2018). This study employs transfer learning to leverage insights from a larger dataset, improving pattern detection across three datasets. Fine-tuning of pre-trained models is necessary due to dissimilarities between patterns in these datasets and standard benchmarks. The models, initially trained on the COCO 2017 dataset with 200,000+ images and 250,000+ annotated objects (Lin et al., 2014), were trained on 640x640 pixel resolution images. During fine-tuning, images in smaller datasets were resized to 640 pixels on the smaller dimension while maintaining the aspect ratio.

Data augmentation enhances model performance by increasing training data. This study employs basic augmentations—random jpeg quality, contrast and brightness adjustment, and random black patches. For the SAM dataset, 90-degree rotation and vertical flip augmentations are included due to variant pattern orientations. Table 4 displays performance metrics for both models across the three datasets, incorporating the mentioned techniques. Results indicate the FASTER R-CNN model's superior performance on the SAM and DMM datasets. Subsequent experiments focus solely on the FASTER R-CNN model.

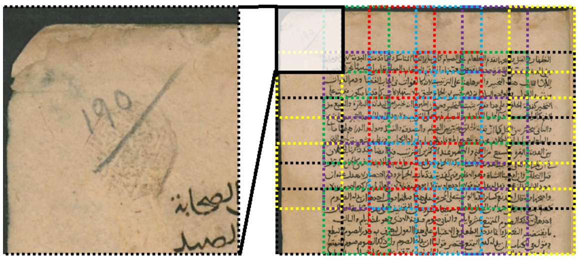

Figure 5: An illustration of the proposed image tiling. The used example in this illustration is the upper part of an image from the SAM dataset. Each image is split into 640x640 sub-images, and then the corresponding annotations are mapped properly into their new position within each tile. The tiles are overlapped by 25% in order to avoid missing the patterns located at the borders between the tiles.

However, both models struggle with the WPM dataset, primarily due to handwritten words lacking distinct visual features against a background of similar words. These words also lack clear boundaries and distinctiveness, making object recognition challenging. Additionally, there's significant intra-class variation between instances of the same pattern, stemming from differences in handwriting styles among different scribes. Furthermore, images in the WPM dataset are large, with selected patterns occupying less than 0.1% of the image. Scaling down during training results in annotated patterns represented by only a few pixels. While a larger model input could address this issue, it comes with a substantial increase in computational cost.

4.2 Detecting Small Patterns

Detecting small objects in images poses challenges for deep learning models (Tian et al., 2018) due to several factors. Small objects often have fewer pixels, providing less visual information for the model to extract useful features. Furthermore, these objects may be easily occluded or concealed by other elements in the scene, complicating detection. The complexity of shapes and features in small objects adds another challenge for the model to accurately recognize and classify them.

Image tiling technique is one of the approaches used to help improving the performance of detecting small objects in which an input image is divided into a grid of smaller tiles or patches (Ozge Unel et al., 2019). Each tile is then processed independently by a machine learning model, which generates a prediction for the presence or absence of small objects within the tile. The predictions from the individual tiles can

Table 5: The impact of image tiling on the detection performance of the FASTER ResNet50 model.

Faster R-CNN ResNet	Metric	SAM	DMM	WPM
Without image tiling	COCO mAP	0.84	0.56	≈ 0.0
	mAP@0.5	0.99	0.97	≈ 0.0
	Recall@1	0.79	0.47	≈ 0.0
With image tiling	COCO mAP	0.84	0.66	≈ 0.10
	mAP@0.5	1.00	1.00	≈ 0.15
	Recall@1	0.86	0.61	≈ 0.11

then be combined to generate a final prediction for the entire image.

The key benefit of the image tiling technique is its ability to enable a machine learning model to concentrate on smaller, more manageable regions of an image instead of processing the entire image simultaneously. This is particularly advantageous for small object detection, enabling the model to focus on areas more likely to contain small objects and make more accurate predictions. As part of this approach, an image tiling mechanism has been implemented to enhance the performance of small pattern detection across all three datasets.

Every image in the datasets undergoes division into sub-images sized 640x640. These tiles have a 25% overlap with adjacent tiles to incorporate features from pattern borders effectively. To accommodate this, annotations of the original image are recalculated by shifting coordinates, ensuring accurate placement in relevant tiles. Refer to Fig. 5 for an illustration of the proposed image tiling. Proper resizing of images precedes the tiling process to include image borders in the tiles.

At the end of each row and column, tiles extend beyond the image boundary due to the fixed tile size. Therefore, we shift these tiles inwards so that they are contained within the image boundary. As a result, these tiles have a larger overlap area with their preceding tiles. The overlap increment is directly proportional to the number of shifted pixels.

Concerning annotated patterns on the borders, inclusion occurs only when the overlap between the annotation and the tile is 50% or more. This ensures that all annotated patterns are incorporated into at least one tile, given their size is not comparable to the tile size—a condition met in all three datasets. The impact of image tiling on detection performance is evident in Table 5. The proposed image tiling technique markedly improves detection performance across all datasets, including the WPM dataset.

Improved detection results on the WPM dataset are anticipated by incorporating advanced augmentations, a larger model-input size, more training steps, and applying the dropout technique. However, achieving high performance on handwritten patterns requires further research. Consequently, we encourage researchers in the community to explore these possibilities. The base results on both validation and test sets are outlined in Table 6.

Table 6: The base results for the SAM, DMM and WPM datasets obtained using the FASTER R-CNN ResNet50 model.

Faster R-CNN ResNet	SAM val/test	DMM val/test	WPM val/test
COCO mAP	0.84 / 0.78	0.66 / 0.64	0.10 / 0.09
mAP@0.5	1.00 / 0.87	1.00 / 0.89	0.15 / 0.13
Recall@1	0.86 / 0.77	0.61 / 0.56	0.11 / 0.10

5 CONCLUSIONS

We explored detecting small patterns in digitised manuscripts with limited annotated examples per pattern. Three detection datasets were created and annotated, featuring words, seals, and drawings commonly found in manuscripts. Challenges included limited training samples, small instance size, fading, and interclass similarities. Two deep learning models were tested, namely the FASTER ResNet and the EfficientDet, and detection performance was reported using COCO metrics. A general approach was proposed to serve as a baseline for these datasets, utilizing standard techniques and image tiling. While improvements were made, performance on the WPM dataset remained poor due to factors such as lack of saliency and intra-class variations. Therefore, further research is required to enhance the performance on such types patterns.

ACKNOWLEDGEMENTS

The research for this work was funded by the Deutsche Forschungsgemeinschaft (DFG, German Research Foundation) under Germany's Excellence Strategy – EXC 2176 'Understanding Written Artefacts: Material, Interaction and Transmission in Manuscript Cultures', project no. 390893796. The research was conducted within the scope of the Centre for the Study of Manuscript Cultures (CSMC) at Universität Hamburg.

In addition, we would like to thank Giovanni Ciotti for providing, selecting, and annotating all the images and patterns in the WPM dataset, and Aneta

Yotova for annotating the SAM and DMM datasets and preparing the "Number of instances" tables.

REFERENCES

Burie, J.-C., Coustaty, M., Hadi, S., Kesiman, M. W. A., Ogier, J.-M., Paulus, E., Sok, K., Sunarya, I. M. G., and Valy, D. (2016). ICFHR competition on the analysis of handwritten text in images of balinese palm leaf manuscripts. In *15th International Conference on Frontiers in Handwriting Recognition*, pages 596–601.

Dai, J., Li, Y., He, K., and Sun, J. (2016). R-fcn: Object detection via region-based fully convolutional networks. *Advances in neural information processing systems*, 29.

En, S., Nicolas, S., Petitjean, C., Jurie, F., and Heutte, L. (2016a). New public dataset for spotting patterns in medieval document images. *Journal of Electronic Imaging*, 26(1):1 – 15.

En, S., Petitjean, C., Nicolas, S., and Heutte, L. (2016b). A scalable pattern spotting system for historical documents. *Pattern Recognition*, 54:149 – 161.

Everingham, M., Van Gool, L., Williams, C. K., Winn, J., and Zisserman, A. (2010). The pascal visual object classes (voc) challenge. In *International Conference on Computer Vision*, pages 404–417.

He, K., Zhang, X., Ren, S., and Sun, J. (2016). Deep residual learning for image recognition. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 770–778.

Jiao, L., Zhang, F., Liu, F., Yang, S., Li, L., Feng, Z., and Qu, R. (2019). A survey of deep learning-based object detection. *IEEE access*, 7:128837–128868.

Le, V. P., Nayef, N., Visani, M., Ogier, J.-M., and De Tran, C. (2014). Document retrieval based on logo spotting using key-point matching. In *2014 22nd international conference on pattern recognition*, pages 3056–3061. IEEE.

Li, X., Grandvalet, Y., Davoine, F., Cheng, J., Cui, Y., Zhang, H., Belongie, S., Tsai, Y.-H., and Yang, M.-H. (2020). Transfer learning in computer vision tasks: Remember where you come from. *Image and Vision Computing*, 93:103853.

Lin, T.-Y., Maire, M., Belongie, S., Hays, J., Perona, P., Ramanan, D., Dollár, P., and Zitnick, C. L. (2014). Microsoft coco: Common objects in context. In *European Conference on Computer Vision*, pages 740–755. Springer.

Liu, L., Ouyang, W., Wang, X., Fieguth, P., Chen, J., Liu, X., and Pietikäinen, M. (2020). Deep learning for generic object detection: A survey. *Int. journal of computer vision*, 128(2):261–318.

Mohammed, H. (2023a). Dataset of drawings in medieval manuscripts (dmm).

Mohammed, H. (2023b). Dataset of seals in arabic manuscripts (sam).

Mohammed, H. (2023c). Model Parameters of FASTER ResNet and EfficientDet Model Parameters of FASTER ResNet and EfficientDet.

Mohammed, H. and Ciotti, G. (2023). Dataset of words in palm-leaf manuscripts (wpm).

Mohammed, H., Märgner, V., and Ciotti, G. (2021). Learning-free pattern detection for manuscript research. *International Journal on Document Analysis and Recognition (IJDAR)*, 24(3):167–179.

Ozge Unel, F., Ozkalayci, B. O., and Cigla, C. (2019). The power of tiling for small object detection. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops*, pages 0–0.

Redmon, J., Divvala, S., Girshick, R., and Farhadi, A. (2016). You only look once: Unified, real-time object detection. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 779–788.

Ren, S., He, K., Girshick, R., and Sun, J. (2017). Faster r-cnn: Towards real-time object detection with region proposal networks. *IEEE Transactions on Pattern Analysis & Machine Intelligence*, 39(06):1137–1149.

Talukdar, J., Gupta, S., Rajpura, P. S., and Hegde, R. S. (2018). Transfer learning for object detection using state-of-the-art deep neural networks. In *2018 5th International Conference on Signal Processing and Integrated Networks (SPIN)*, pages 78–83.

Tan, M., Pang, R., and Le, Q. V. (2020). Efficientdet: Scalable and efficient object detection. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pages 10778–10787.

Tian, Y., Li, B., Chen, C., Fu, Y., and Huang, Q. (2018). Tiny object detection in dense crowds. In *Proceedings of the European Conference on Computer Vision (ECCV)*, pages 497–513.

van Lit, L. C. (2020). Seals from the staatsbibliothek zu berlin and their automated detection.

Wiggers, K. L., Britto, A. S., Heutte, L., Koerich, A. L., and Oliveira, L. S. (2019). Image retrieval and pattern spotting using siamese neural network. In *2019 International Joint Conference on Neural Networks (IJCNN)*, pages 1–8. IEEE.

Yarlagadda, P., Monroy, A., Carque, B., and Ommer, B. (2011). Recognition and analysis of objects in medieval images. In Koch, R. and Huang, F., editors, *Computer Vision – ACCV 2010 Workshops*, pages 296–305, Berlin, Heidelberg. Springer Berlin Heidelberg.

Zaidi, S. S. A., Ansari, M. S., Aslam, A., Kanwal, N., Asghar, M., and Lee, B. (2022). A survey of modern deep learning based object detection models. *Digital Signal Processing*, page 103514.

Úbeda, I., Saavedra, J. M., Nicolas, S., Petitjean, C., and Heutte, L. (2020). Improving pattern spotting in historical documents using feature pyramid networks. *Pattern Recognition Letters*, 131:398 – 404.