ndt

Article

Non-Destructive Estimation of Paper Fiber Using Macro Images: A Comparative Evaluation of Network Architectures and Patch Sizes for Patch-Based Classification

Naoki Kamiya [1,*], Kosuke Ashino [2,*], Yasuhiro Sakai [1], Yexin Zhou [3], Yoichi Ohyanagi [4] and Koji Shibazaki [4]

[1] School of Information Science and Technology, Aichi Prefectural University, Nagakute 480-1198, Japan; koinobori8282@gmail.com
[2] Graduate School of Information Science and Technology, Aichi Prefectural University, Nagakute 480-1198, Japan
[3] Life and Culture Department, Seirei Women's Junior College, Akita 011-0937, Japan; vincentchou0307@gmail.com
[4] Faculty of Fine Arts, Aichi University of the Arts, Nagakute 480-1194, Japan; yoichi_ohyanagi@mail.aichi-fam-u.ac.jp (Y.O.); ksibazaki@mail.aichi-fam-u.ac.jp (K.S.)
* Correspondence: n-kamiya@ist.aichi-pu.ac.jp (N.K.); im243001@cis.aichi-pu.ac.jp (K.A.)

Abstract: Over the years, research in the field of cultural heritage preservation and document analysis has exponentially grown. In this study, we propose an advanced approach for non-destructive estimation of paper fibers using macro images. Expanding on studies that implemented EfficientNet-B0, we explore the effectiveness of six other deep learning networks, including DenseNet-201, DarkNet-53, Inception-v3, Xception, Inception-ResNet-v2, and NASNet-Large, in conjunction with enlarged patch sizes. We experimentally classified three types of paper fibers, namely, kozo, mitsumata, and gampi. During the experiments, patch sizes of 500, 750, and 1000 pixels were evaluated and their impact on classification accuracy was analyzed. The experiments demonstrated that Inception-ResNet-v2 with 1000-pixel patches achieved the highest patch classification accuracy of 82.7%, whereas Xception with 750-pixel patches exhibited the best macro-image-based fiber estimation performance at 84.9%. Additionally, we assessed the efficacy of the method for images containing text, observing consistent improvements in the case of larger patch sizes. However, limitations exist in background patch availability for text-heavy images. This comprehensive evaluation of network architectures and patch sizes can significantly advance the field of non-destructive paper analysis, offering valuable insights into future developments in historical document examination and conservation science.

Keywords: paper fiber estimation; non-destructive analysis; patch-based image classification; deep learning

check for updates

Citation: Kamiya, N.; Ashino, K.; Sakai, Y.; Zhou, Y.; Ohyanagi, Y.; Shibazaki, K. Non-Destructive Estimation of Paper Fiber Using Macro Images: A Comparative Evaluation of Network Architectures and Patch Sizes for Patch-Based Classification. *NDT* **2024**, 2, 487–503. https://doi.org/10.3390/ndt2040030

Academic Editor: Fabio Tosti

Received: 31 July 2024
Revised: 24 October 2024
Accepted: 6 November 2024
Published: 7 November 2024

1. Introduction

The analysis of raw materials of paper and the corresponding manufacturing methods plays a vital role in historical research, cultural heritage preservation, and industrial quality control [1]. Non-destructive paper analysis has been instrumental in examining historical documents, authenticating artworks, and evaluating materials without causing physical harm. For example, techniques like support vector machine (SVM)-based image classification have been used to analyze paper characteristics such as fiber composition and production processes, including beating cycles and additives [2]. Fourier image analysis using digital microscopes has also been applied to reveal fiber orientation in historical papers [3]. In the broader context of non-destructive testing (NDT), techniques like digital X-radiography have been used to classify tempering materials in pottery [4] and combined non-destructive and destructive testing has been employed to assess the structural integrity of ancient timber beams [5].

While these methods offer valuable insights, particularly for other materials, traditional paper analysis methods often face challenges, such as subjectivity in sensory evaluations [6] and the potential for damage when destructive testing [7,8] is required. Non-destructive techniques, though promising, sometimes require specialized equipment, limiting their accessibility. To address these challenges, this study focuses on developing an objective and accessible non-destructive method specifically tailored for paper fiber analysis, utilizing macro images and deep learning techniques.

Recent advancements in image processing technology and machine learning techniques have presented new possibilities for non-destructive paper analysis. The application of deep learning techniques to historical document analysis has significantly expanded in recent years [9], with diverse applications such as manuscript dating [10] and the decipherment of historical manuscripts [11]. Previously, our research group has contributed to this field by proposing a method for estimating paper fibers based on patch classification using macro images captured via consumer digital cameras [12–14]. This approach does not require specialized imaging equipment and can be applied to a wide range of paper materials. In our initial research, we achieved a high accuracy of 94.2% for classifying three types of fibers using the VGG-16 architecture [12]. Subsequent studies have been reported using EfficientNet-B0, achieving 87.5% accuracy in classifying the origin of kozo paper [13] and 86.0% accuracy in fiber estimation of macro images containing text [14]. Our previous works on non-destructive paper fiber estimation using macro images have contributed to the growing field of historical document analysis by focusing on the analysis of paper materials rather than textual content. This aligns with the broader trend of applying sophisticated machine learning techniques to solve complex problems in cultural heritage preservation and historical research [15].

The use of patch-based classification is supported by successful applications in other fields, such as medical imaging [16]. This technique facilitates the detailed analysis of specific regions within larger images, which is particularly useful for heterogeneous materials such as paper. Based on these studies, we have previously established a fundamental procedure for paper fiber estimation using patch classification. However, several challenges remain in the pursuit of more advanced analyses and applications for diverse paper materials.

Although our previous study focused on EfficientNet for network architecture evaluation, the performance of the latest architectures, such as Inception-ResNet-v2 [17] and Xception [18], was not sufficiently assessed. Khan et al. [19] conducted a comprehensive survey of recent convolutional neural network (CNN) architectures and highlighted the potential benefits of exploring various network designs for specific tasks. Additionally, although previous studies have reported that a 500-pixel square patch size yields the highest accuracy with respect to patch size optimization [20], the effect of larger patch sizes has not been verified. Furthermore, improving the applicability of paper materials that contain text remains challenging. Although previous studies [14] have reported some success, there exists a need to enhance the robustness of the method for more complex documents and various typefaces.

In this study, we aim to re-evaluate the effectiveness of widely accepted network architectures and patch sizes for non-destructive paper fiber estimation. Expanding on our previous research that implemented EfficientNet-B0, we investigate the performance of six additional architectures (DenseNet-201, DarkNet-53, Inception-v3, Xception, Inception-ResNet-v2, NASNet-Large) and examine the impact of larger patch sizes (up to 1000 pixels). Additionally, we assess the applicability of the method to paper materials containing text, which remains a significant challenge in the field. By comprehensively evaluating these architectures and patch sizes, we aim to provide insights that will advance non-destructive paper analysis methods and contribute to the broader fields of cultural heritage preservation and historical research.

2. Materials and Methods

The proposed method for paper fiber estimation comprises four stages: input image acquisition, patch generation, deep-learning-based classification, and result aggregation. Figure 1 illustrates the overall process of the proposed approach.

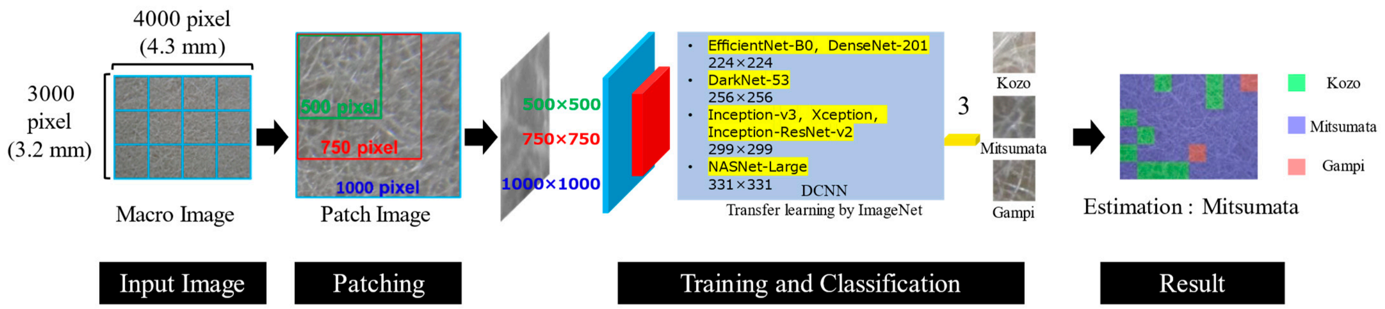


Figure 1. Overview of the proposed paper fiber estimation method. The process begins with a macro image (4000 × 3000 pixels) of paper, which is divided into patches of various sizes (500 × 500, 750 × 750, and 1000 × 1000 pixels). These patches are fed into different pretrained deep convolutional neural networks (DCNNs) for classification. Finally, the results are aggregated to estimate the fiber type of the entire macro image.

2.1. Dataset

2.1.1. Paper Sample Preparation

In this study, we focused on three types of paper fibers commonly used in traditional Japanese papermaking: kozo, mitsumata, and gampi. We prepared 1080 macro images, comprising 360 images for each fiber type. These samples were obtained from our paper database [21], which contains various traditional Japanese papers. The fiber composition of these samples was predetermined and verified via optical microscope fiber composition tests conducted at the Kochi Prefectural Paper Industry Technology Center, following the JIS P 8120 "Paper, board and pulps—Fiber furnish analysis" standard [7,8].

2.1.2. Macro Image Acquisition

We used the Olympus Tough TG-5 digital camera to capture macro images based on the method described in our previous study [20]. The camera was set to microscope mode with autofocus, an aperture value of f/6.3, 4× magnification, and a focal length of 18 mm. An LED light guide (LG-1; Olympus, Tokyo, Japan) was used to ensure uniform illumination. Images were saved in JPEG format with a resolution of 4000 × 3000 pixels, covering an area of approximately 4.3 mm × 3.2 mm, with each pixel corresponding to nearly 1.08 µm. We placed the camera lens directly on the paper sample and even performed imaging using a thin acrylic plate whenever necessary to maintain a consistent focal distance.

2.1.3. Patch Image Generation

Each macro image was used to generate patch images of three different sizes: 500 × 500, 750 × 750, and 1000 × 1000 pixels. These sizes were selected based on our previous studies [14,20], wherein we determined 500 × 500 pixels to be optimal among smaller sizes with respect to the divisibility of the original 4000 × 3000 pixels macro images. A non-overlapping sliding window approach was employed to generate 51,840, 21,600, and 12,960 patches for each size, respectively. The use of larger patch sizes enables the capture of more detailed structural features of the fibers, which could be valuable for future studies involving more complex or mixed fiber compositions. In this study, the patch sizes were tested using seven deep learning models: EfficientNet-B0, DenseNet-201, DarkNet-53, Inception-v3, Xception, Inception-ResNet-v2, and NASNet-Large, to assess their impact on classification accuracy.

2.2. Network Architectures

The selection of appropriate network architectures is crucial for historical document analysis tasks [15]. In this study, we evaluated seven pretrained CNN architectures available in the MATLAB R2022b deep network designer [22], including the EfficientNet-B0 (baseline) [23], DenseNet-201 [24], DarkNet-53 [25], Inception-v3 [26], Xception [18], Inception-ResNet-v2 [17], and NASNet-Large [27]. We selected these networks based on their theoretical performance compared with EfficientNet-B0, our baseline from previous studies [14,20]. The selection was guided by a performance comparison provided by MathWorks [22], which clearly demonstrated the relationship between EfficientNet-B0 and other networks in terms of accuracy and computational efficiency.

The baseline model EfficientNet-B0 [23] features a compound scaling method that uniformly scales the network width, depth, and resolution, offering a balance between efficiency and performance with an input size of 224 × 224 and a depth of 132 layers. By contrast, DenseNet-201 [24] introduces a dense connectivity pattern, wherein each layer is directly connected to every other layer in a feed-forward fashion, promoting feature reuse and substantially reducing the number of parameters in its 201-layer structure. Originally designed for YOLOv3, DarkNet-53 [25] employs residual blocks in its 53-layer architecture with an input size of 256 × 256, optimizing both the accuracy and speed in object detection tasks.

In the case of larger input sizes, Inception-v3 [26] uses factorized convolutions and auxiliary classifiers in its 48-layer structure with a 299 × 299 input, aiming to reduce computational costs while maintaining high accuracy. Xception [18] expands this concept and replaces the Inception modules with depthwise separable convolutions in its 71-layer architecture, offering improved performance with the same number of parameters as Inception-v3. Inception-ResNet-v2 [17] further develops this concept by combining the Inception architecture with residual connections in its 164-layer network, thereby accelerating the training of very deep convolutional networks. Finally, NASNet-Large [27], developed using the Neural Architecture Search technique, represents an automatically optimized network structure with 270 layers and an input size of 331 × 331, designed to maximize both accuracy and efficiency.

Table 1 summarizes the input size and depth of each network architecture used in this study.

Table 1. Input size and depth of the evaluated network architectures.

Network Architecture	Input Size	Depth
EfficientNet-B0 [23]	224 × 224	132
DenseNet-201 [24]	224 × 224	201
DarkNet-53 [25]	256 × 256	53
Inception-v3 [26]	299 × 299	48
Xception [18]	299 × 299	71
Inception-ResNet-v2 [17]	299 × 299	164
NASNet-Large [27]	331 × 331	270

2.3. Experimental Setup

2.3.1. Three-Fold Cross-Validation

We employed three-fold cross-validation for the performance evaluation. The dataset was divided into three subsets at the macro-image level to ensure that patches from the same macro image were not split between the training and testing sets. Each subset contained 360 macro images (120 for each fiber type). We used two subsets (720 macro images) for training and one subset (360 macro images) for testing in each iteration. This process was repeated three times, with each subset serving as the test set once.

2.3.2. Hyperparameter Settings

We empirically explored the hyperparameters for each network, starting with the default settings of MATLAB. The number of epochs was set to 100 for all experiments. We evaluated batch sizes of 32 and 64 considering two optimization functions, namely Stochastic Gradient Descent with Momentum (SGDM) and Adam. The learning rates were set to 0.01 and 0.001 for SGDM and Adam optimizers, respectively. These parameters were selected based on the best performance achieved in our exploratory experiments.

2.3.3. Hardware and Software Environment

We performed the experiments using a system equipped with dual AMD Rome7742 CPUs and an NVIDIA A100 GPU with 80 GB of memory; Ubuntu 20.04.6 was used as the operating system. All implementations and training procedures were performed using MATLAB R2022b.

2.4. Evaluation Method

The performance of the networks was assessed using the classification accuracy obtained from the three-fold cross-validation. Two primary metrics were calculated for each combination of patch size and network architecture. The first metric was the patch classification accuracy of the image, which represented the percentage of correctly classified individual patch images. The second metric was the macro image fiber estimation accuracy, which was determined by applying a majority voting scheme to the classification results of patch images derived from each macro image. In the cases where the majority vote resulted in a tie, we considered the macro image unclassifiable and deemed it a misclassification. This dual evaluation approach enabled the assessment of both the fine-grained performance at the patch level and the holistic accuracy in determining the fiber composition of complete paper samples.

3. Results

We evaluated seven distinct network architectures across three patch sizes for paper fiber classification and estimation. The findings were classified into two categories: patch classification accuracy and macro image fiber estimation accuracy.

3.1. Patch Classification Accuracy

Table 2 summarizes the patch classification accuracy for each network architecture across three patch sizes (500 × 500, 750 × 750, and 1000 × 1000 pixels).

Table 2. Patch classification accuracy (%).

Network	500 × 500	750 × 750	1000 × 1000
EfficientNet-B0	78.5	79.6	77.9
DenseNet-201	76.1	77.5	76.9
DarkNet-53	74.1	72.7	72.0
Inception-v3	77.6	80.1	81.5
Xception	77.3	80.3	82.2
Inception-ResNet-v2	75.7	80.8	82.7
NASNet-Large	74.0	76.1	78.7
Average	76.2	78.2	78.8

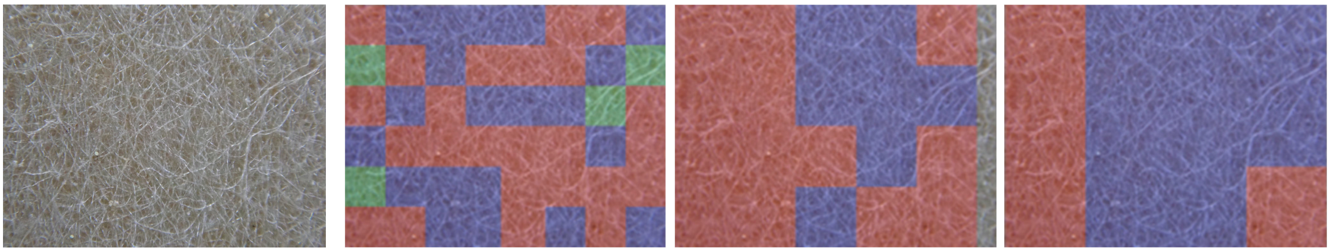
Our analysis revealed significant variations in performance across different network architectures and patch sizes. Inception-ResNet-v2 achieved the highest classification accuracy of 82.7% with 1000 × 1000 pixel patches, representing a 7% improvement over its performance with 500 × 500 pixel patches. This substantial improvement highlighted the potential benefits of using larger patch sizes in some network architectures. Conversely, EfficientNet-B0 demonstrated the highest accuracy (78.5%) for the smallest patch size (500 × 500 pixels); however, its performance deteriorated slightly for larger patches.

This contrasting behavior underscores the varying impacts of patch size across different network architectures and emphasizes the need for a careful selection of patch size for optimal performance.

The effect of increasing the patch size varied significantly among networks, revealing diverse responses to changes in the input dimensions. While Inception-ResNet-v2, Xception, and Inception-v3 exhibited substantial improvements with larger patches, DarkNet-53 exhibited a decline in performance as the patch size increased. This diversity in the responses indicates that the effectiveness of larger patches significantly relies on the specific network architecture employed, implying that the optimal patch size should be determined based on each architecture.

In general, we observed a trend of improved accuracy with increased patch size, validated by the mean accuracies of 76.2%, 78.2%, and 78.8% for the 500 × 500, 750 × 750, and 1000 × 1000 pixel patches, respectively. However, this trend was inconsistent across networks, with some networks exhibiting peak performances at different patch sizes. This variability further emphasizes the complex relationship between the network architecture and patch size in the context of paper fiber classification.

To provide a concrete example of the impact of patch size on the classification performance, Figure 2 illustrates the results of Inception-ResNet-v2 on a single macro image of the mitsumata fiber. This visual representation clearly demonstrates an improvement in classification accuracy and consistency as the patch size increases, offering insights into the potential benefits of larger patch sizes in capturing more comprehensive features for classification.



Macro image
Correct fiber: Mitsumata

500 [pixel]
Estimated fiber: Gampi

750 [pixel]
Estimated fiber: Gampi

1000 [pixel]
Estimated fiber: Mitsumata

Figure 2. Macro image fiber estimation results based on patch classification using Inception-ResNet-v2 across three patch sizes (500 × 500, 750 × 750, and 1000 × 1000 pixels) for a macro image of mitsumata fiber. The colored areas represent the classification results, with green, blue, and red corresponding to kozo, mitsumata, and gampi, respectively. The gray areas in the 750 × 750 patch size image indicate regions at the edges of the original macro image where patches could not be extracted completely, and thus, no classification was performed. This visualization demonstrates the impact of increasing patch size on classification accuracy and consistency. As the patch size increases from left to right, a clear improvement is observed in the correct classification of mitsumata fibers (blue areas). The larger patch sizes (750 × 750 and 1000 × 1000) exhibit significantly fewer misclassifications, particularly reducing the instances of kozo (green) misclassification. This improvement suggests that larger patches capture more comprehensive fiber structures and patterns, resulting in more accurate and stable classifications. The progression also highlights the trade-off between classification detail and the number of available patches for voting in macro image estimation.

The performance variations observed across the various network architectures and patch sizes underscore the complex interplay between these factors in paper fiber classification. Networks with larger input sizes, such as Inception-ResNet-v2 and Xception (both with 299 × 299 inputs), generally demonstrate better performance with larger patch sizes. This indicates that these architectures may be better equipped to process and learn from the additional contextual information provided by larger patches. By contrast, networks with smaller input sizes, such as EfficientNet-B0 and DenseNet-201 (both with 224 × 224 inputs), exhibit less consistent improvements or even slight declines with increasing patch sizes. This implies that the relationship between a native input size of the network and its ability to effectively utilize larger patches is not straightforward and may rely on other architectural features.

These findings highlight that careful consideration of both network architecture and patch size is critical when designing a paper fiber classification system. Although larger patch sizes generally lead to improved classification accuracy, the optimal patch size may vary depending on the selected network architecture. Furthermore, trade-offs between computational complexity and accuracy improvements must be considered because larger patch sizes inevitably increase the processing requirements. This complex relationship between network architecture, patch size, and classification performance underscores the need for a nuanced approach to developing effective paper fiber classification systems.

3.2. Macro Image Fiber Estimation Accuracy

Table 3 presents the fiber estimation accuracy of macro images across various network architectures and patch sizes. This macro-level estimation was determined by applying a majority voting scheme to the classification results of the patch images derived from each macro image.

Table 3. Macro image fiber estimation accuracy (%).

Network	500 × 500	750 × 750	1000 × 1000
EfficientNet-B0	84.8	84.2	80.6
DenseNet-201	83.1	81.2	80.1
DarkNet-53	82.9	78.7	74.6
Inception-v3	83.1	83.4	83.1
Xception	83.8	84.9	83.1
Inception-ResNet-v2	80.2	84.3	84.6
NASNet-Large	79.8	80.1	82.6
Average	82.5	82.4	81.2

Our analysis of the macro image fiber estimation accuracy revealed interesting trends that differed from those observed in the patch-level classification. Xception achieved the highest estimation accuracy of 84.9% using 750 × 750 pixel patches, slightly surpassing the performance of EfficientNet-B0 (84.8%) with 500 × 500 pixel patches. Contrary to the patch classification results, the macro image estimation accuracy did not consistently improve with larger patch sizes across all networks. This discrepancy highlights the complex relationship between patch-level classifications and macro-level estimations in the context of paper fiber analysis.

The impact of patch size on the estimation accuracy varied significantly among the networks, revealing diverse patterns of performance. For most networks, the peak performance was observed with either 500 × 500 or 750 × 750 pixel patches, with a general decline in accuracy for 1000 × 1000 pixel patches. This trend was reflected in the average accuracy, which exhibited a slight decline from 82.5% for the 500 × 500 patches to 81.2% for the 1000 × 1000 patches. This suggests that an optimal patch size range may exist for macro image estimation, which balances the trade-off between feature richness and the number of patches available for majority voting.

Individual network performance provides further insights into the interplay between architecture and patch size. EfficientNet-B0 exhibited the best performance with the smallest patch size (500 × 500), achieving 84.8% accuracy; however, its performance deteriorated significantly with larger patches. This indicates that the architecture of EfficientNet-B0 might be well suited for extracting relevant features from smaller patches in the context of macro image estimation. By contrast, Inception-ResNet-v2, which performed best in patch classification for larger patch sizes, exhibited an interesting trend in macro image estimation. Its performance improved significantly from 80.2% with 500 × 500 patches to 84.6% with 1000 × 1000 patches, demonstrating the best performance among all the networks for the largest patch size.

The performance of DarkNet-53 consistently deteriorated with the increase in patch size, with the accuracy decreasing from 82.9% for 500 × 500 patches to 74.6% for 1000 × 1000 patches. This consistent decline suggests that the architecture of DarkNet-53 may not be optimal for processing larger patches in this specific task. Conversely, Xception and Inception-v3 maintained relatively stable performances across different patch sizes, with Xception achieving the highest overall accuracy of 84.9% with 750 × 750 patches. This stability across patch sizes indicates that these architectures may be more robust to changes in input dimensions for macro-level estimation tasks.

The aforementioned results collectively highlight the complex relationship between the network architecture, patch size, and estimation accuracy in the context of macro image fiber estimation. Additionally, the findings emphasize the importance of carefully selecting both the network architecture and patch size for optimal performance as the best combination may vary depending on the specific characteristics of the task and network architecture. These findings underscore the need for a nuanced approach that considers the interplay between patch size and network architecture when designing and implementing deep learning models for paper fiber estimation to achieve the highest accuracy in both patch-level classification and macro-level estimation tasks.

Furthermore, the discrepancy between the patch-level classification and macro-level estimation accuracies for some networks suggests that the process of aggregating patch-level results to perform macro-level predictions is not straightforward. This observation opens avenues for future research into more sophisticated aggregation methods that could potentially improve the overall system performance.

4. Discussion

4.1. Impact of Network Architecture and Patch Size on Classification Performance

Our comprehensive investigation of the impact of various network architectures and patch sizes on paper fiber classification performance yielded several significant insights, demonstrating the complex interplay between these factors and classification accuracy. Both Inception-ResNet-v2 and Xception exhibited superior performances, particularly for larger patch sizes. Inception-ResNet-v2 achieved a remarkable classification accuracy of 82.7% with 1000 × 1000 pixel patches, representing a substantial improvement of 7% over its performance with 500 × 500 pixel patches. This enhancement can be attributed to the larger patch size, which enables the capture of more extensive structural features such as fiber orientation and distribution patterns. Conversely, EfficientNet-B0 demonstrated high performance (78.5%) with smaller patch sizes (500 × 500 pixels), suggesting that its architecture is well suited to smaller input sizes. However, EfficientNet-B0 showed limited improvement with the increase in patch size, even exhibiting a slight decrease to 77.9% at 1000 × 1000 pixels.

The varying responses of different networks to the increase in patch size provide valuable insights into the relationship between architecture and input dimensions. Although Inception-v3, Xception, and Inception-ResNet-v2 exhibited significant performance improvements with larger patch sizes, the improvements were limited in the case of EfficientNet-B0 and DenseNet-201. This variability indicates that the optimal patch size significantly relies on network architecture. Furthermore, networks with larger native input sizes (299 × 299) demonstrate excellent performance with larger patch sizes, suggesting that they are inherently more suitable for processing larger fields of view. Conversely, DarkNet-53 consistently exhibited the lowest performance across all patch sizes, with the performance declining as the patch size increased, indicating its unsuitability for this particular task.

These findings collectively demonstrate that the selection of the network architecture and optimization of the patch size are key determinants of performance improvement in paper fiber classification tasks. Networks that can effectively use larger patch sizes, such as Inception-ResNet-v2 and Xception, enable more detailed and accurate fiber classification. However, this must be balanced with the computational requirements and specific constraints of the application environment.

4.2. Comparison of Patch Classification and Macro Image Estimation

Our analysis of patch-level classification and macro-image-level fiber estimation accuracies revealed trends that underscored the complex relationship between these two scales of analysis in paper fiber characterization. Notably, the macro image estimation accuracy tended to be higher than the patch classification accuracy. For instance, the patch classification accuracy was 80.3% when using Xception with 750 $\times$ 750 pixel patches, whereas the corresponding macro image estimation accuracy was 84.9%. This indicates that the majority voting integration process employed in the macro-level estimation partially mitigates the misclassification of individual patches, improving the overall accuracy.

However, the relationship between patch size and estimation accuracy was nonlinear and varied significantly among the networks. Although increasing the patch size from 500 $\times$ 500 pixels to 750 $\times$ 750 pixels generally improved the macro image estimation accuracy for most networks, a further increase to 1000 $\times$ 1000 pixels often reduced the accuracy. This trend was exemplified by EfficientNet-B0, where the macro image estimation accuracies for patch sizes of 500 $\times$ 500, 750 $\times$ 750, and 1000 $\times$ 1000 pixels were 84.8%, 84.2%, and 80.6%, respectively. This implies that although larger patches can capture more comprehensive fiber structure information, they simultaneously reduce the number of patches available for majority voting, thereby destabilizing the estimation process at the macro-image level.

The performance patterns of individual networks provide further insights into the interplay between architecture and patch size in macro-level estimations. The peak performance of EfficientNet-B0 with the smallest patch size (500 $\times$ 500) suggests that its architecture is particularly adept at extracting relevant features from smaller patches for macro image estimation. By contrast, Inception-ResNet-v2 demonstrates a significant performance improvement from 80.2% with 500 $\times$ 500 patches to 84.6% with 1000 $\times$ 1000 patches, validating the effective utilization of larger input sizes. The consistent performance decline observed in DarkNet-53 with the increase in patch size (from 82.9% at 500 $\times$ 500 pixels to 74.6% at 1000 $\times$ 1000 pixels) indicates that its architecture may be suboptimal for processing larger patches in this specific task. Xception and Inception-v3 maintain relatively stable performances across different patch sizes, with Xception achieving the highest overall accuracy of 84.9% with 750 $\times$ 750 patches. This implies that these architectures are more robust to input size variations in macro-level estimation tasks.

The aforementioned findings emphasize that the optimal combination of the network architecture and patch size varies depending on the specific characteristics of the task and the architecture itself, necessitating a nuanced approach for the development and implementation of deep learning models for paper fiber estimation. Furthermore, the discrepancies between patch-level classification and macro-level estimation accuracies for some networks highlight the non-trivial nature of aggregating patch-level results for macro-level predictions. In the future, more sophisticated aggregation methods must be developed to enhance the overall system performance.

4.3. Performance of Text-Containing Images

We evaluated the effectiveness of the proposed method on macro images containing text to address the practical challenges of analyzing actual documents and books. Our dataset comprised 237 text-containing macro images, including 129 kozo, 72 mitsumata, and 36 gampi samples. We applied the seven network architectures and three patch sizes to these images to assess both patch classification and macro image estimation accuracies.

Tables 4 and 5 summarize the overall patch classification and macro image estimation accuracies of text-containing images, respectively.

Table 4. Patch classification accuracy of text-containing images (%).

Network	500 × 500	750 × 750	1000 × 1000
EfficientNet-B0	56.7	64.2	68.0
DenseNet-201	58.8	61.0	62.9
DarkNet-53	60.9	62.0	57.8
Inception-v3	58.7	64.8	60.2
Xception	59.5	66.0	67.3
Inception-ResNet-v2	62.5	66.2	69.1
NASNet-Large	56.9	59.5	64.3
Average	59.1	63.4	64.2

Table 5. Macro image fiber estimation accuracy of text-containing images (%).

Network	500 × 500	750 × 750	1000 × 1000
EfficientNet-B0	84.8	84.2	80.6
DenseNet-201	83.1	81.2	80.1
DarkNet-53	82.9	78.7	74.6
Inception-v3	83.1	83.4	83.1
Xception	83.8	84.9	83.1
Inception-ResNet-v2	80.2	84.3	84.6
NASNet-Large	79.8	80.1	82.6
Average	82.5	82.4	81.2

The results indicated a general decrease in patch classification accuracy for text-containing images in comparison with text-free images. For instance, the accuracy reached 69.1% for text-containing images when Inception-ResNet-v2 was used with 1000 × 1000 pixel patches, which was lower than the accuracy of 82.7% achieved for text-free images. However, the macro image estimation accuracy remained relatively high, with Xception achieving an accuracy of 84.9% for 750 × 750 pixel patches; this was comparable to its performance on text-free images. The disparity between patch-level and macro-level performances indicates that the majority voting process used in macro image estimation can partially compensate for the challenges posed by the presence of text in individual patches.

We conducted a more granular analysis by distinguishing between patches containing text and those without text to gain further insights. Tables 6 and 7 present the patch classification and fiber estimation accuracies of patches containing text, respectively.

Table 6. Patch classification accuracy for text-containing patches (%).

Network	500 × 500	750 × 750	1000 × 1000
EfficientNet-B0	44.8	55.1	65.3
DenseNet-201	45.5	50.3	53.1
DarkNet-53	49.8	51.9	49.6
Inception-v3	45.9	55.9	53.7
Xception	47.4	61.8	62.5
Inception-ResNet-v2	52.3	59.2	64.9
NASNet-Large	48.0	51.4	58.8
Average	47.7	55.1	58.3

Table 7. Fiber estimation accuracy using text-containing patches (%).

Network	500 × 500	750 × 750	1000 × 1000
EfficientNet-B0	46.6	57.0	66.1
DenseNet-201	53.9	54.1	53.2
DarkNet-53	56.7	53.2	47.1
Inception-v3	51.1	63.6	50.0
Xception	56.7	61.9	66.1
Inception-ResNet-v2	60.4	62.0	70.7
NASNet-Large	52.3	59.4	63.5
Average	54.0	58.7	59.5

The classification accuracy of text-containing patches (Table 6) was generally low, averaging 47.7% and 58.3% for the 500 × 500 and 1000 × 1000 pixel patches, respectively.

This reduction in accuracy was attributed to text-obscuring fiber features. Interestingly, the fiber estimation accuracy of text-containing patches (Table 7) was slightly higher than the patch classification accuracy, with Inception-ResNet-v2 achieving 70.7% accuracy for 1000 × 1000 pixel patches. This indicates that useful fiber information can be extracted even in the presence of text, particularly when using larger patch sizes.

Tables 8 and 9 summarize the patch classification and fiber estimation accuracies for text-free patches within text-containing macro images, respectively.

Table 8. Patch classification accuracy of text-free patches within text-containing images (%).

Network	500 × 500	750 × 750	1000 × 1000
EfficientNet-B0	62.3	72.7	72.9
DenseNet-201	66.3	69.2	75.1
DarkNet-53	68.0	70.5	68.2
Inception-v3	66.0	72.3	69.5
Xception	67.7	69.5	75.0
Inception-ResNet-v2	68.6	73.9	75.0
NASNet-Large	62.0	67.3	70.4
Average	65.8	70.8	72.3

Table 9. Fiber estimation accuracy of text-free patches within text-containing images (%).

Network	500 × 500	750 × 750	1000 × 1000
EfficientNet-B0	62.1	72.5	72.0
DenseNet-201	63.0	70.5	75.1
DarkNet-53	74.7	72.0	66.5
Inception-v3	63.8	72.1	66.9
Xception	69.5	69.0	72.3
Inception-ResNet-v2	67.1	70.0	74.5
NASNet-Large	62.0	67.8	70.7
Average	66.0	70.6	71.1

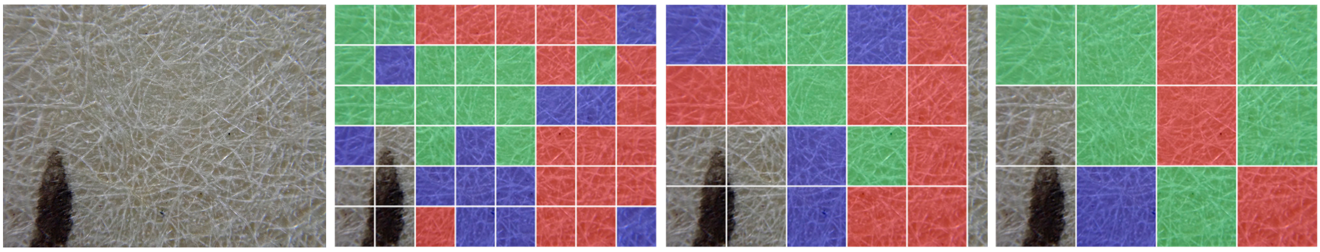
As expected, the classification accuracy for text-free patches within text-containing images (Table 8) was higher, reaching an average of 72.3% for 1000 × 1000 pixel patches; this was closer to the results obtained from entirely text-free images. Similarly, the fiber estimation accuracy of text-free patches (Table 9) was high, exceeding 70% for both the 750 × 750 and 1000 × 1000 pixel patch sizes. These results highlight the importance of identifying and leveraging text-free regions within documents for an accurate fiber estimation.

The following conclusions can be drawn from the aforementioned findings. First, the use of larger patch sizes (750 × 750 and 1000 × 1000 pixels) is particularly effective for text-containing patches. This is because larger patches are likely to include both text and non-text regions, thereby capturing more fiber features. Second, Inception-ResNet-v2 and Xception consistently demonstrate high performance regardless of the presence of text, indicating that these architectures are particularly well suited for this task. Third, fiber estimation at the macro-image level maintains a higher accuracy than patch-level classification, indicating that the majority voting integration process can partially compensate for misclassifications in individual patches.

However, increasing the patch size exhibits some limitations, particularly for images with high text density. In the case of 1000 × 1000 pixel patches, the number of background regions (text-free patches) decrease significantly, rendering fiber estimation impossible for some images. This observation underscores the need for a balanced approach to patch-size selection, particularly when handling text-heavy documents.

Our findings confirm that careful selection of patch size and network architecture are both crucial in the fiber analysis of text-containing paper documents. In the future, the introduction of a text detection and removal preprocessing stage, the development of a multitask learning approach that distinguishes between text and non-text regions, and the adoption of multiscale analysis methods that combine results from different patch sizes should be explored. The implementation of these enhancements can improve the accuracy and robustness of fiber estimation for text-containing paper documents, thereby enabling efficient non-destructive analysis of actual historical documents and books and contributing to the fields of cultural heritage research and conservation science.

Figure 3 illustrates the performance of the proposed approach on text-containing images using Inception-ResNet-v2. The figure depicts the results for a kozo fiber sample with text across different patch sizes. We observed changes in the patch-level classifications as the patch size increased from 500 × 500 pixels to 1000 × 1000 pixels. Although the 500 × 500 pixel patches resulted in an incorrect macro-level estimation (gampi), the larger patch sizes correctly identified the fiber as kozo. This suggests that larger patches may capture more comprehensive fiber characteristics, thereby enabling a more accurate macro-level estimation despite the presence of text.



Macro image Correct fiber: Kozo | 500 [pixel] Estimated fiber: Gampi | 750 [pixel] Estimated fiber: Gampi | 1000 [pixel] Estimated fiber: Kozo

Figure 3. Macro image fiber estimation results based on patch classification using Inception-ResNet-v2 across three patch sizes (500 × 500, 750 × 750, and 1000 × 1000 pixels) for a kozo fiber sample containing text. The colored areas represent patch-level classification results, with green, blue, and red corresponding to kozo, mitsumata, and gampi, respectively. The gray areas in the 750 × 750 patch size image indicate regions at the edges of the original macro image where patches could not be fully extracted, and thus, no classification was performed. Additionally, in all patch sizes, gray areas represent regions where patches contained text, and these were excluded from classification. The estimated fiber type for the entire macro image is specified below each patch-based result.

4.4. Implications for Paper Fiber Analysis

The developed novel approach addresses several key challenges in the field of non-destructive paper fiber analysis. As the proposed method enables the analysis of a wide range of paper materials using macro images captured by consumer-grade digital cameras without requiring expensive specialized equipment, it is particularly valuable for historical documents and artworks where destructive analysis is not permissible. Furthermore, the proposed approach aligns with the growing trend in conservation science towards non-invasive techniques and presents new possibilities for fiber analysis of actual documents and books. The ability to analyze text-containing images complements other non-destructive techniques and facilitates the efficient analysis of large quantities of paper materials, thereby contributing to document authentication and conservation science.

The proposed method offers new perspectives to historical and archaeological research by enabling the statistical analysis of paper characteristics across different eras and regions. This can enhance our understanding of the evolution of papermaking techniques. Moreover, our analysis expands the work on the historical development of papers and contributes to studies on the history of technology and cultural exchange. The accurate understanding of paper fiber composition using the proposed method has important implications for the selection of appropriate conservation and restoration techniques, which is a crucial aspect in the field of cultural heritage preservation. Another potential application of the method is to the quality control processes in the paper industry, particularly for evaluating high-quality handmade paper.

The interdisciplinary nature of this study spans computer vision, machine learning, materials science, history, and conservation science, highlighting the importance of cross-disciplinary approaches in advancing cultural heritage studies. The combination of conventional knowledge on paper characteristics with advanced computational methods presents new avenues for understanding and preserving documentary heritage. However, despite the significant progress in analyzing text-containing documents, further improvement is necessary for the handling of densely written texts and complex layouts. Future research should focus on refining the proposed technique to address these challenges by incorporating more advanced text detection and segmentation methods.

In summary, our study represents a significant step forward in non-destructive paper analysis, with wide-ranging implications for cultural heritage studies, conservation science, and the paper industry. With further refinement and expansion of the proposed technique, we anticipate greater contributions to the understanding and preservation of paper-based cultural heritage. The practical applications of this research extend beyond academic circles, offering valuable tools for conservation, authentication, and quality control in paper production, thereby bridging the gap between theoretical research and practical implementation in the field of paper fiber analysis.

4.5. Comparison with Previous Studies

In this study, we extend the previous analyses on EfficientNet-B0 [14] by investigating the impact of a wider range of network architectures and patch sizes. The performed comparative analysis yields several significant insights, particularly regarding the influence of network architecture on performance. We demonstrated that more complex architectures, such as Inception-ResNet-v2 [17] and Xception [18], exhibit superior performance with larger patch sizes. This revealed the critical importance of network selection in paper fiber classification tasks. Additionally, our research advances the optimization of patch sizes. Although a previous study [20] has suggested that 500 $\times$ 500 pixel patches are optimal, we verified the performance improvements with larger patch sizes of 750 $\times$ 750 and 1000 $\times$ 1000 pixels, enabling the capture of more extensive fiber structures and contributing to improved classification accuracy.

A significant advancement in this study is the improved capability to analyze text-containing images, an aspect that has not been sufficiently addressed in previous studies [14]. This progress has substantially enhanced the potential of applying the proposed method to actual document analysis, presenting new possibilities for studying historical documents and manuscripts. Our multi-network comparison evaluated seven different network architectures under identical conditions, thereby providing valuable insights into the characteristics and strengths of each network and offering guidance for the selection of the most appropriate network for specific tasks. Furthermore, we achieved notable improvements in classification accuracies, increasing the patch classification accuracy from 78.5% in previous studies [14] to 82.7% and slightly improving the macro-image-estimation accuracy from 84.8% to 84.9%.

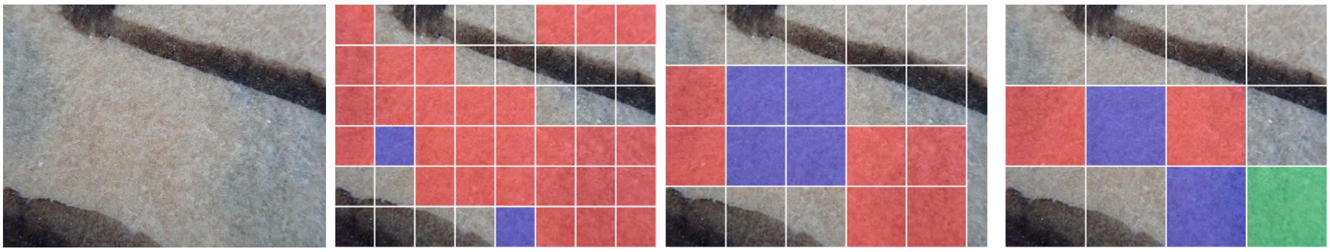
These advancements demonstrate the applicability of the proposed method to actual documents containing text, which can lead to technological innovations in cultural heritage protection and historical research. The implications of the study findings extend beyond the immediate field of paper analysis, contributing to the broader domain of cultural heritage studies and offering new tools and methodologies for the non-destructive examination of historical artifacts.

Finally, our findings on the impact of network architecture and patch size align with observations from other historical document analysis tasks. For instance, Hamid et al. [10] demonstrated the effectiveness of CNNs for manuscript dating, whereas Yin et al. [11] applied deep learning techniques to decipher historical manuscripts. In addition to these studies, our study contributes to the growing body of evidence supporting the efficacy of deep learning in historical document analysis [9,10]. The variability in performance across different architectures and patch sizes highlights the importance of careful parameter selection and experimentation in developing effective solutions for specific tasks in historical document analysis.

4.6. Limitations and Future Work

Although our research on non-destructive paper fiber analysis demonstrated significant advancements, several limitations exist that warrant consideration. Primarily, we focused on three specific fiber types, including kozo, mitsumata, and gampi, which represent only a fraction of the diverse range of paper types used throughout history and across different regions. This limitation of the dataset scope restricts the broad applicability of the proposed method. Additionally, the analysis of text-containing images presents ongoing challenges, particularly for documents with high text density and complex layouts, where the accuracy and robustness of the developed method can be further improved.

A significant limitation of our approach is apparent when handling text-heavy images, particularly when larger patches are used. Figure 4 demonstrates this issue using a gampi fiber sample containing text.



Macro image
Correct fiber: Gampi

500 [pixel]
Estimated fiber: Gampi

750 [pixel]
Estimated fiber: Gampi

1000 [pixel]
Estimated fiber:
No majority

Figure 4. Macro image fiber estimation results based on patch classification using Inception-ResNet-v2 across three patch sizes (500 × 500, 750 × 750, and 1000 × 1000 pixels) for a gampi fiber sample containing text. The colored areas represent patch-level classification results, with green, blue, and red corresponding to kozo, mitsumata, and gampi, respectively. The gray areas in the 750 × 750 patch size image indicate regions at the edges of the original macro image where patches could not be fully extracted, and thus, no classification was performed. Additionally, in all patch sizes, gray areas represent regions where patches contained text, and these were excluded from classification. The estimated fiber type for the entire macro image is mentioned below each patch-based result, with “No majority” indicating the case where the majority voting method could not determine the predominant fiber type because of the equal distribution of classifications.

As indicated in Figure 4, while the 500 × 500 and 750 × 750 pixel patches result in accurate macro-level estimation, the 1000 × 1000 pixel patches lead to a “No majority” state. This occurs when the majority voting process cannot determine the predominant fiber type because of the equal distribution of different patch-level classifications. This limitation highlights a key trade-off in the proposed approach: although larger patch sizes can potentially capture more comprehensive fiber structures, they may reduce the number of patches available for majority voting. In text-heavy documents, this reduction may result in the failure of definitive macro-level classification owing to insufficient or equally distributed voting patches. Furthermore, the computational costs associated with large patch sizes and complex network architectures pose potential barriers to their practical implementation, particularly in resource-constrained environments. A comprehensive comparison of the developed method with other non-destructive analysis techniques, such as spectroscopic methods, should be performed to better understand its relative strengths and weaknesses.

The transition from laboratory results to real-world applications in historical documents and artworks presents its own set of challenges, such as variations in document conditions, environmental factors, and the practicalities of on-site analysis. These factors can potentially affect the performance and reliability of the proposed method in real-world scenarios.

These limitations should be addressed in future works to advance the field of non-destructive paper analysis. For instance, the dataset should be expanded to encompass a wider variety of paper types, geographical origins, and historical periods to broaden the applicability of the method. Additionally, as demonstrated in our previous study [14], our two-stage network approach successfully achieved over 99% accuracy in recognizing and extracting text-free patches, enabling selective analysis of background patches in text-heavy images. However, more sophisticated text detection and separation algorithms that can potentially incorporate advanced natural language processing techniques must be developed to further enhance the analysis of text-containing images. Moreover, the development of multimodal analysis techniques that combine image analysis with other non-destructive methods, such as spectroscopy, could provide a more comprehensive understanding of paper characteristics.

Other potential avenues for future research include the application of transfer learning and domain adaptation techniques, which could improve learning efficiency with limited datasets and enhance the method's applicability to a wider range of paper types. Additionally, exploring time-series analysis to evaluate paper aging and preservation states may offer valuable insights into degradation processes and conservation strategies.

5. Conclusions

In this study, we comprehensively evaluated the impact of multiple deep learning network architectures and patch sizes on the non-destructive estimation of paper fibers using macro images. We explored six deep learning networks, including DenseNet-201, DarkNet-53, Inception-v3, Xception, Inception-ResNet-v2, and NASNet-Large, in conjunction with enlarged patch sizes of 500, 750, and 1000 pixels. Our experiments demonstrated that Inception-ResNet-v2 with 1000-pixel patches achieved the highest patch classification accuracy of 82.7%, while Xception with 750-pixel patches yielded the best macro-image-based fiber estimation performance at 84.9%. These findings highlight the importance of network architecture selection and patch size optimization for improving classification performance. Additionally, we demonstrated the method's capability to handle text-containing images through selective analysis of background patches. While the proposed method has proven effective in laboratory settings, future implementations of this research could benefit from leveraging edge computing technologies, as discussed in Ashino (2024) [28], enabling more scalable and accessible applications beyond laboratory environments. Open dissemination of these research findings will also be critical for the advancement of non-destructive paper analysis. Future research should focus on expanding the dataset to cover more diverse paper types and periods, improving text detection algorithms, and integrating multimodal techniques such as spectroscopy for a more comprehensive analysis. This work significantly contributes to the advancement of non-destructive paper analysis, with promising implications for cultural heritage preservation and historical document examination.

Author Contributions: Conceptualization, N.K.; methodology, N.K.; software, Y.S. and K.A.; validation, N.K., Y.S. and K.A.; formal analysis, K.A.; investigation, N.K., Y.O. and K.S.; resources, N.K.; data curation, Y.Z., Y.O., Y.S. and K.A.; writing—original draft preparation, N.K. and K.A.; writing—review and editing, N.K.; visualization, K.A.; supervision, N.K. and K.S.; project administration, K.S. and N.K.; funding acquisition, K.S. and N.K. All authors have read and agreed to the published version of the manuscript.

Funding: This research was funded by JSPS KAKENHI, grant number JP22H00003, and a research grant from the Naito Science & Engineering Foundation.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: The macro images of the paper samples used in this study are partially available on the Paper Propagation website [21]. Although not fully publicly available, these images can be accessed for collaborative research. Researchers interested in using this dataset for academic studies are encouraged to contact the authors for potential research collaborations. Access to the complete dataset may be granted, subject to agreement on research cooperation and adherence to data usage policies. This approach ensures the integrity of the data while promoting further research in the field of paper fiber analysis.

Conflicts of Interest: The authors declare no conflicts of interest.

References

1. Shibazaki, K. Propagation of paper in the world and Samarkand paper. In Proceedings of the International Seminar on HandMade Paper and Artistic Expression “Research into Ancient Samarkand Paper, Islamic Manuscripts, and Miniatures, Aichi, Japan, 16 November 2019; pp. 6–9.
2. Ishikawa, T.; Yata, N.; Nagano, T. Automatic classification of paper using combinational optimization of image features. *Jpn. Tappi J.* **2011**, *65*, 585–594. [CrossRef]
3. Han, Y.H.; Enomae, T.; Isogai, A.; Yamamoto, H.; Hasegawa, S.; Song, J.J.; Jang, S.W. Traditional Papermaking Techniques revealed by Fibre Orientation in Historical Papers. *Stud. Conserv.* **2006**, *51*, 267–276. [CrossRef]
4. Nagaya, A.; de Lucio, O.G.; Ortiz Ruiz, S.; Uc González, E.; Peraza Lope, C.; Cruz Alvarado, W. Description and Classification of Tempering Materials Present in Pottery Using Digital X-Radiography. *ndt* **2024**, *2*, 456–473. [CrossRef]
5. Marzo, A.; Carpani, B.; Marghella, G.; Tripepi, C. A Methodology to Manage and Correlate Results of Non-Destructive and Destructive Tests on Ancient Timber Beams: The Case of Montorio Tower. *ndt* **2024**, *2*, 311–329. [CrossRef]
6. Masuda, K. Examination for reproduction of antique paper. *J. Soc. Instrum. Control Eng.* **1989**, *28*, 689–692.
7. ISO. *Standard 9184-1*; International Organization for Standardization: Geneva, Switzerland, 2022.
8. JISC. *Standard P8120*; Japanese Industrial Standards Committee: Tokyo, Japan, 1994.
9. Lombardi, F.; Marinai, S. Deep learning for historical document analysis and recognition—A survey. *J. Imaging* **2020**, *6*, 110. [CrossRef] [PubMed]
10. Hamid, A.; Bibi, M.; Moetesum, M.; Siddiqi, I. Deep learning based approach for historical manuscript dating. In Proceedings of the 2019 International Conference on Document Analysis and Recognition, ICDAR 2019, Sydney, Australia, 20–25 September 2019; pp. 967–972. [CrossRef]
11. Yin, X.; Aldarrab, N.; Megyesi, B.; Knight, K. Decipherment of historical manuscript images. In Proceedings of the 2019 International Conference on Document Analysis and Recognition, ICDAR 2019, Sydney, Australia, 20–25 September 2019; pp. 78–85. [CrossRef]
12. Ikuta, A.; Oshima, A.; Urano, Y.; Shibazaki, K.; Kamiya, N. Automatic classification of fiber composition in digital macro images using VGG-16 for nondestructive paper analysis. *J. Inst. Image Electron. Eng. Jpn.* **2020**, *49*, 41–46.
13. Kitamura, T.; Iwata, A.; Urano, Y.; Zhou, Y.; Shibazaki, K.; Kamiya, N. Initial study on classification of Japanese paper by kozo name using EfficientNet with digital camera. In Proceedings of the IEEE 9th Global Conference on Consumer Electronics (GCCE), Kobe, Japan, 13–16 October 2020; pp. 844–847. [CrossRef]
14. Kamiya, N.; Yoshizato, Y.; Zhou, Y.; Ohyanagi, Y.; Shibazaki, K. Fiber estimation from paper macro images via EfficientNet-based patch classification. *IEEE Access* **2024**, *12*, 12271–12278. [CrossRef]
15. Liwicki, F.S.; Liwicki, M. Deep learning for historical document analysis. In *Handbook of Pattern Recognition and Computer Vision*; Chen, C.H., Ed.; World Scientific: Singapore, 2020; pp. 287–303. [CrossRef]
16. Hou, L.; Samaras, D.; Kurc, T.M.; Gao, Y.; Davis, J.E.; Saltz, J.H. Patch-based convolutional neural network for whole slide tissue image classification. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Las Vegas, NV, USA, 27–30 June 2016; pp. 2424–2433.
17. Szegedy, C.; Ioffe, S.; Vanhoucke, V.; Alemi, A. Inception-v4, Inception-ResNet and the impact of residual connections on learning. *Proc. AAAI Conf. Artif. Intell.* **2017**, *38*, 4278–4284. [CrossRef]
18. Chollet, F. Xception: Deep learning with depthwise separable convolutions. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Honolulu, HI, USA, 21–26 July 2017; pp. 1251–1258.
19. Khan, A.; Sohail, A.; Zahoora, U.; Qureshi, A.S. A survey of the recent architectures of deep convolutional neural networks. *Artif. Intell. Rev.* **2020**, *53*, 5455–5516. [CrossRef]
20. Okada, K.; Iwata, A.; Urano, Y.; Zhou, Y.; Shibazaki, K.; Kamiya, N. Initial study on multi-scale patch-based classification of paper fibers based on EfficientNet using consumer digital camera. In Proceedings of the IEEE 10th Global Conference on Consumer Electronics (GCCE), Kyoto, Japan, 12–15 October 2021; pp. 275–277. [CrossRef]
21. Paper Propagation. Available online: https://paper.a-mz.com/paper (accessed on 30 July 2024).
22. MathWorks, Pretrained Deep Neural Networks. Available online: https://jp.mathworks.com/help/deeplearning/ug/pretrained-convolutional-neural-networks.html (accessed on 8 July 2024).
23. Tan, M.; Le, Q. EfficientNet: Rethinking model scaling for convolutional neural networks. In Proceedings of the 36th International Conference on Machine Learning, Long Beach, CA, USA, 9–15 June2019; pp. 6105–6114.

24. Huang, G.; Liu, Z.; Van Der Maaten, L.; Weinberger, K.Q. Densely connected convolutional networks. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Honolulu, HI, USA, 21–26 July 2017; pp. 4700–4708.
25. Redmon, J.; Farhadi, A. YOLOv3: An incremental improvement. *arXiv* **2018**, arXiv:1804.02767. [CrossRef]
26. Szegedy, C.; Vanhoucke, V.; Ioffe, S.; Shlens, J.; Wojna, Z. Rethinking the inception architecture for computer vision. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Las Vegas, NV, USA, 27–30 June 2016; pp. 2818–2826.
27. Zoph, B.; Vasudevan, V.; Shlens, J.; Le, Q.V. Learning transferable architectures for scalable image recognition. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Salt Lake City, UT, USA, 18–22 June 2018; pp. 8697–8710.
28. Ashino, K.; Kamiya, N.; Zhou, Y.; Ohyanagi, Y.; Shibazaki, K. End-To-End Paper Fiber Classification Using Consumer Digital Camera: A Practical Approach for Non-Destructive Analysis. In Proceedings of the IEEE 13th Global Conference on Consumer Electronics (GCCE), Kitakyushu, Japan, 29 October–1 November 2024; pp. 264–265.