The Paleographer's Eye *ex machina*: Using Computer Vision to Assist Humanists in Scribal Hand Identification

Samuel Grieggs[1] C. E. M. Henderson[2] Sebastian Sobecki[2] Alexandra Gillespie[2] Walter Scheirer[3]
[1]Indiana University of Pennsylvania
[2]University of Toronto
[3]University of Notre Dame
sgrieggs@iup.edu, cai.henderson@mail.utoronto.ca, sebastian.sobecki@utoronto.ca, alexandra.gillespie@utoronto.ca, walter.scheirer@nd.edu

Abstract

The steady digitization of medieval manuscripts is rapidly changing the field of paleography, challenging existing assumptions about handwriting and book production. This development has identified historically important centers for the production of scribal texts, and even individual scribes themselves. For example, scholars of late medieval English literature have identified the copyists of a number of literary manuscripts, and the important role of London government clerks in shaping literary culture. However, traditional paleography has no agreed-upon methodology or fixed criteria for the attribution of handwriting to a particular community, period, or scribe. The approach taken by paleographers is inherently qualitative and subject to personal bias. Even those wielding the mighty "paleographer's eye" cannot claim objectivity. Computer vision offers solutions with spectacular performance on writer identification and retrieval benchmarks, but these have not been widely adopted by the paleography community because they tend not to hold up in practice. In this work, we attempt to bridge the divide with a software package designed not to automate paleography, but to augment the paleographer's eye. We introduce automated handwriting identification tools for which the results can be quickly visually understood and assessed, and used as one feature among many by expert paleographers when attributing previously unknown scribal hands. We also demonstrate a use case for our software by analyzing several items believed to be written by Thomas Hoccleve, a highly productive clerk of the Privy Seal who is also an important fifteenth-century English poet.

1. Introduction

The steady digitization of medieval manuscripts and other historical records is rapidly changing the field of paleography — the study and identification of historical script — challenging existing assumptions about handwriting and book production. This development has identified historically important centers for the production of scribal texts, and even individual scribes themselves. Several fascinating areas of research exist where computer vision scientists and paleographers can collaborate to make progress on humanistic questions that are important to fields that have not traditionally used algorithmic analysis.

Figure 1. Our tool analyzing an example of Thomas Hoccleve's hand from our Privy Seal examples. This software looks at image patches and compares them to a template of a specific hand. Green patches have the highest similarity to the template, yellow patches are moderately close, and red patches have a low similarity.

A number of researchers, for example, have focused on identifying the copyists of medieval English literary manuscripts and on the important role of government clerks working in London institutions in shaping literary culture [4, 17, 18, 26, 30, 36–38, 47]. However, paleography has no agreed-on methodology or fixed criteria for the attribution of handwriting to a particular community, period, or scribe. Experts in the field make attributions primarily based on their familiarity, developed over long experience, with a variety of scripts and individual scribal "hands." In particular, paleographers focus on the "aspect," the overall appearance of a scribe's hand. This approach is inherently qualitative and subject to personal bias. In addition, this approach makes it difficult for paleographers to describe their results for publication: "It just really looks like Hoccleve" may be true, but does not sound very convincing. Paleographers therefore take pains to describe, as thoroughly as possible, particular features of a hand that seem most diagnostic. However, a lack of controlled vocabulary for this practice makes it difficult to compare one hand to another using these descriptions alone.

This is problematic not only for paleographers, but also for scholars whose arguments depend on paleographic attributions. When a non-paleographer reads Linne R. Mooney's hugely influential 2006 article attributing the work of a scribe previously known only as "Scribe B" to the London scrivener Adam Pinkhurst [26], how convinced should they be? *Is* his "distinctive formation of *g*, in which the strokes often cross to create a projecting spike on the right of the lower lobe but occasionally fail to complete it" really all that distinctive? Is the "double-slash, dot, double-slash, dot decoration" indeed "virtually a signature"? A reader used to critically analyzing written arguments and their sources will have no real idea. They must take the paleographer's word for it — and the paleographer, for all their training, is nevertheless usually going by vibes. It wasn't until 2011 that an objection to Mooney's attribution appeared in print [30]; this objection did not prevent literature scholars from continuing to take the Pinkhurst attribution as settled historical fact.

Ironically, computer-vision-based approaches for writer identification and retrieval often come with some of the same problems as paleographers' attempts. It is most common to end up with just a number that represents a sample's similarity to other documents. While this can lead to excellent scores on benchmark datasets such as the CVL and ICDAR challenge datasets [7, 12, 22], we can expect paleographers to be understandably hesitant about treating these numbers as truth values. Therefore, our objective in this work was to create a software package that instead provides a visual representation of a questioned document's similarity to reference examples of a particular hand, so that paleographers can use that information to fortify their argument as to whether or not a certain document was written by a particular known scribe. Not only does our tool offer a visual representation of the document's similarity in individual regions, it provides a convenient Graphical User Interface (GUI) to make it accessible for paleographers who do not have access to an on-staff machine learning engineer. We have taken steps to make sure that this software can be run on Windows, Mac OS, and Linux, with pre-built executables for each. The software is even designed to be run on machines without a GPU, further reducing the barrier to entry.

In summary, this work introduces a software tool that is distributed, ready to run with a GUI designed to augment the Paleographer's Eye by producing a visualization showing which portions of the image are most similar to the reference hand. Furthermore, we have collected a unique dataset of text images from a closely associated group of scribes who deliberately wrote very similarly in the same type of script. Finally, we give an interesting case study by using the tool to identify some "in the wild" images that may have been transcribed by Thomas Hoccleve.

2. Related Work

Related work in this area consists of approaches from the document analysis community on writer identification and retrieval, and work from the digital humanities community on automated paleography. We review the references most related to our work below.

2.1. Writer Identification and Retrieval

At its highest level, the problem of handwriting or writer identification is a well-explored area of human biometrics within document analysis. An early example of this includes the work of Bulacu et al., who developed a methodology that utilizes text-independent statistical features to measure attributes like letter slant and spacing to make distinctions between writers [2]. Interestingly, a great deal of even the recent work in this area does not use deep learning. The International Conference on Document Analysis and Recognition (ICDAR) held challenges on very similar tasks in 2017 [12] and 2019 [7]. Both challenges were framed as an Image Retrieval task, where the objective was to retrieve all the images from a specific writer. The 2017 ICDAR challenge had five entries that used traditional handcrafted image features, and one that used a Convolutional Neural Network (CNN), which generally performed worse than the other entries. In 2019, none of the teams used artificial neural networks as a main approach, although one team used a neural network for the auxiliary task of binarizing color images.

When deep learning is used in handwriting recognition, it tends to be used as a type of feature descriptor for areas of interest which have been identified by classical feature matching algorithms like SURF [1] or SIFT [24]. Christlein et al. [6] and Fiel et al. [13] both used the last hidden layer of a convolutional neural network trained on small patches of handwriting drawn from a closed set image classification problem as a feature descriptor. They then used a local feature extractor to identify features, and their CNN outputs as the descriptor. Tang et. al [44] took a similar approach, but instead of using small patches, built global representations by training their network on larger synthetic text documents created by stitching cropped words together. The current state of the art on the ICDAR challenge datasets are held by a similar method that uses a ResNet backbone with an improved encoding layer and additional analysis of page embeddings [29]

With the great success that margin losses have achieved in facial recognition (*e.g.,* Facenet [34] and ArcFace [9]), other application areas have made use of such losses when considering deep learning. For example, in writer identification, work has emerged that uses margin losses in tandem with a self-supervised representation to classify authorship of Vatican documents [23].

These projects focus on solving the computer vision problem itself, rather than creating a tool that can be easily used by paleographers to do their work. Our main contribution is this attempt to bridge the gap by making a tool that exists beyond the realm of research code and has features for accessibility such as a simple GUI and a prepackaged executable that can be run on multiple platforms. Our tool uses similar approaches to many of the other methodologies mentioned above, but we have attempted to package it in a way that is consistent with the expectations of paleographers. We have taken inspiration from other tools for historical document processing, such as READ-COOP's[1] Transkribus [19] for text transcription, which have been very successful in bringing computer vision capabilities to a broader audience.

2.2. Automated Paleography

As this work represents a collaboration between computer vision scientists and paleographers, it is also important to consider its position in the literature from a paleographer's perspective. A great deal of the good work done by computer scientists is not yet in a form that can be utilized by paleographers. When code is distributed, it is usually set up to reproduce experiments, and not set up to be run in a way that is useful for paleographers and needs to be adapted. The increasing availability of digitized manuscripts has led to major advances in digital paleography in recent years. Simply being able to refer to high-quality images of manuscripts in different libraries at the same time, from any location, makes the work of paleographers considerably easier. For example, the Late Medieval English Scribes (LMES) website[2], developed by Linne Mooney, Estelle Stubbs, and Simon Horobin, allows users to compare characteristics of certain scribes, especially their use of particular letterforms, across multiple manuscripts. Similarly, the DigiPal project[3], led by Peter Stokes, has made significant progress in mapping letterforms in insular scripts from the period 1000-1100 [42,43].

However, most work in digital paleography, such as that of the aforementioned READ-COOP, focus on the problem of transcribing difficult and frequently highly abbreviated medieval scripts, not on identifying the scribes responsible. Other work by historians, such as the DEEDS project,[4] has been primarily interested in dating charters. The Medieval Paleographic Scale (MPS) project, led by Jan Burgers with Lambert Schomaker and Sheng He, has developed a paleographic scale, using image processing and pattern recognition to date medieval charters [15]. Schomaker also led the NWO HIMANIS project (HIstorical MANuscript Indexing for User-controlled Search) on documents produced by the French royal chancery (14th-15th c.), tackling problems in image processing, word segmentation, and allographic variation [5, 16, 33].

3. Datasets of Handwritten Documents

Digital Humanities problems like this one offer unique challenges when using machine learning. While just about anyone can, say, classify a photo, write a caption for it, or otherwise provide useful ground truth for more general computer vision problems, most people cannot accurately identify scribal hands. Furthermore, digitized manuscripts are often an intellectual property minefield. Many libraries and archives charge for non-personal use or publication of their images, and permission must be acquired before redistributing even personal photos of their collections.

Generating ground truth for this problem involves negotiating with archives and consulting some of the world's leading paleographers, making it a much more challenging and expensive process than contracting out to services such as Amazon Mechanical Turk. This is compounded by the fact that in most cases there is no recorded evidence of scribal identity, so even uncontested identifications are often the result of a modern paleographer's intuition rather than explicit written evidence. And not infrequently, as with Adam Pinkhurst or Thomas Hoccleve, paleographers disagree! This may be acceptable for training data but makes it difficult to evaluate that data.

With these considerations in mind, we have collected a new evaluation dataset of 14 scribes who were working in the English central government's Privy Seal Office in the late 14th and early 15th centuries (referred to as the "Privy Seal Dataset" below). These scribes include both senior and junior clerks of the Privy Seal, all of whom were based in Westminster, now part of London. We selected them because they executed their handwriting professionally, in a standardized script, using all three languages of the central government: Latin, French, and English. Since these clerks formed a tightly-knit and rigorous community of practice, writing the same script, called Secretary, an example of which is shown in Figure 2, in a form specific to the Privy Seal Office [39], these 14 hands are highly similar and are not easily distinguishable. Additionally, because

[1] https://readcoop.eu/transkribus/
[2] https://www.medievalscribes.com
[3] https://www.digipal.eu/
[4] https://deeds.library.utoronto.ca/

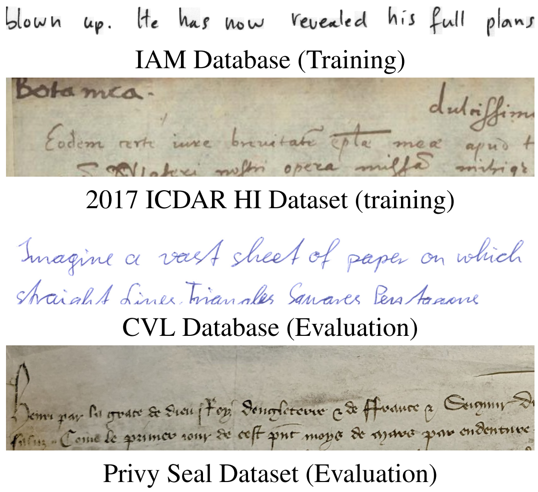


Figure 2. Examples of pages from the various datasets used in training and evaluation. An additional dataset used in training contains data we have not obtained permission to share.

this dataset includes examples in all three languages, letterforms and variations in aspect that complicate scribal attribution between one language to another are captured in the dataset. This dataset contains up to 38 images per scribe, totaling in 146.

The majority of these examples come from the C 81 document series (Chancery warrants for the Great Seal) in the National Archives of the UK (TNA), with additional records drawn from the following series: E 28 (Exchequer, treasury of the receipt: Council and Privy Seal records), E 30 (Exchequer, treasury of the receipt: diplomatic documents), E 101 (King's Remembrancer: accounts various), and E 404 (Exchequer of Receipt: warrants for issues). We have received permission from TNA to share them with the computer vision community.

Among these scribes is Thomas Hoccleve, a Privy Seal clerk and poet whose handwriting is the best-known and most widely studied example from medieval England. Unlike many medieval authors, whose work does not survive in their own handwriting, four Hoccleve holographs are known to exist. He also copied literary work by other authors, and wrote hundreds of records during his time as a clerk: Mooney identified over a hundred [27], to which Helen Killick added a further 900+ [20], and Sobecki still more [40]. Hoccleve's handwriting features in our figures and our case study for this paper.

This is not enough data by itself to provide a rigorous evaluation, so we have also utilized some existing standard benchmark datasets. The IAM Handwriting Database [25] is a dataset of 9862 images of lines of modern English handwritten text from 500 different writers that is most commonly used for handwritten text transcription but also has writer-level annotations. We have used this dataset as a portion of our training dataset. Another commonly used dataset in handwriting / writer identification is the CVL database [22], which contains 1604 page images of writing, produced by 310 different writers, mostly in English but also with some German. We use this to evaluate our tool on modern text. We have also drawn on the entirety of the data from the 2017 ICDAR Handwriting Identification Challenge as training data [12], as per the protocol prescribed in the 2019 ICDAR Image Retrieval challenge [7]. This gives us an additional 4782 images from 1114 writers that come from the electronic holdings of the Universitätsbibliothek Basel. These documents represent the domain of interest more accurately than the other two sets from the literature. An example of each of these datasets can be seen in Figure 2.

We have also collected some additional training images that are not from open data sources. Many of these come from the "Subscriptions to the oath" in the Common Paper of the Company of Scriveners of the City of London, a group of scribes — among them Pinkhurst — trained in executing notarial and official documents. Importantly for our purposes, unlike most examples of medieval writing, each oath is signed by its scribe, giving us gold-standard ground truth. From the Common Paper we have 337 images of texts of varying length from 164 Scriveners. The paleographers on our team also sourced and identified the scribes for some additional signed or otherwise certain ground-truth examples from personal photo collections and various online repositories. The online examples are publicly available; we will provide instructions for obtaining them. From these we collected an additional 24 identities with 121 images in total, including a "null" identity which has examples of parchment and paper containing no handwritten text, which is intended to make the model less likely to match blank paper or parchment to a specific hand.

4. Methodology for Stylistic Analysis

Paleographers often find that computer vision models do not work nearly as well on real world data as their reported performance would suggest. Therefore, when building our tool, we decided it was less important to achieve high overall accuracy on benchmark datasets than it was to create a model that could generalize across a variety of script types.

In our software, the stylistic analysis begins by building templates from a directory of samples taken from a known identity. Our model considers 64×64 patches from an image. In order to build an effective template, we take the 100 ORB [31] features with the highest activation. We chose 100 locations to sample the image, since we found that increasing the number of samples per image did not significantly improve performance, and minimizing the number of samples is important to improve inference speed on CPUs. After sampling, we pass each patch through the model, and taking the sum of the model output for all of the known sam-

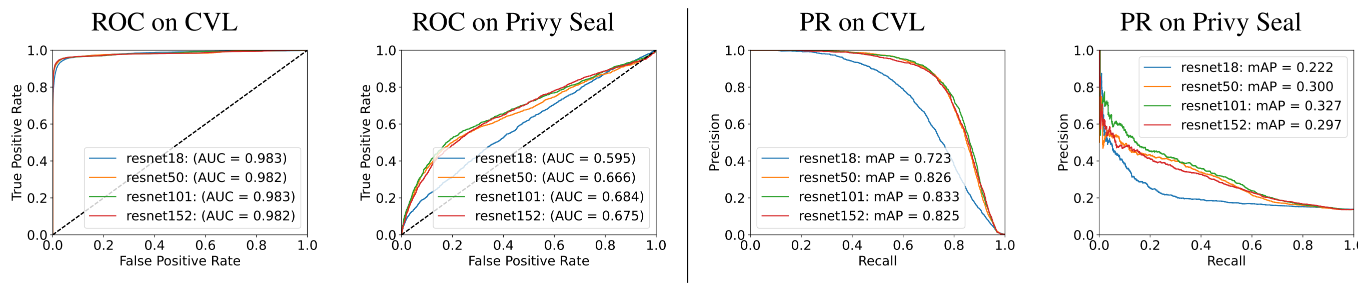


Figure 3. ImageNet pretrained ResNet Performance on The CVL dataset and our Privy Seal evaluation dataset.

ples for said identity. After a template is built, the user can compare arbitrary samples against it. The algorithm takes the mean of all of the image patches for that writer, and takes the cosine distance between that vector and the representation generated by passing each tile in the query image through the network. The output module overlays a color on top of the original image, depending on the similarity score of each patch. This allows the user to make a visual assessment of the similarity of the two images.

4.1. Model Ablation Study

To begin, we show the results on several "off the shelf" models. We looked at different sizes of ResNet [14], specifically, ResNet-18, ResNet-50, ResNet-101, and ResNet-152. These models were each initialized with their default ImageNet pretrained weights, and their penultimate layer was replaced with a randomly initialized fully connected layer with an output size of 512, which gives us our model output. Each model took 64×64 pixel patches of the text as input. Our evaluation procedure for these models is slightly different than how the software itself works. Building templates for each writer would require class labels from the test set. Therefore, for a preliminary examination of each model's performance we looked used a more standard image retrieval procedure, and took each image in the evaluation dataset and ran the model over each 64×64 patch of the image, and then took the mean of the model output for each patch across the whole image. Other than cropping the image into tiles, no other pre-processing steps, such as binarization, were taken during evaluation. We then compared each sample to every other sample in the dataset using cosine distance, to give us a similarity score for each sample. We used those scores to calculate Receiver Operating Characteristic (ROC) and Precision Recall Curves, which are shown in Figure 3. We found that the performance of the ResNets larger than ResNet 18 were more or less interchangeable, so to make the software easier to run, we chose to use a ResNet50 for further experiments.

4.1.1 Training

We trained two ResNets, both ResNet-50s. One model was trained completely end-to-end, and the other only trained the weights that were randomly initialized in the previous experiment. Each model was trained on the datasets de-

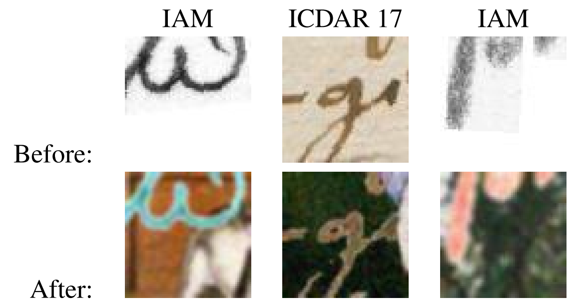


Figure 4. Examples of our augmentations from the IAM [25] and ICDAR 2017 [12] datasets. The top row shows the cropped images before augmentation, and the bottom row with augmentation.

scribed in Section 3. For each image shown to the model, we used the following data augmentation procedure: (1) randomly rotate the image up to 15 degrees, (2) randomly crop a 64×64 patch, (3) binarize the image to get a mask of the ink, (4) use that mask to replace the background of the image with a random crop from the ImageNet [8] training set, (5) apply a random color jitter only to the text portion of the image, and, finally, (6) add Gaussian Noise and Blur. Examples of images from this augmentation process can be seen in Figure 4. These augmentations were chosen to force the model to focus on the shape of the handwriting rather than extraneous features such as the color of ink or the type of paper, which could be useful for classification but are less likely to generalize. This training regime offers a number of benefits for this task. In particular, by training on patches rather than whole images, we can prevent over-fitting, since many of the full-page images are quite large and most contain more than 1600 unique 64×64 patches. With the aggressive augmentation scheme, it is unlikely that the model will ever see the same image twice during training.

The model was trained using an angular margin loss similar to ArcFace [9]. However, the angular margin loss encourages class-clustering in the last hidden layer and in theory improves the feature representation using a distance learning strategy. The models were trained for 150 epochs, using an Adam Optimizer [21] starting with a learning rate of $1e^{-3}$ and a weight decay of $1e^{-5}$. The code used to train the model will be released with publication.

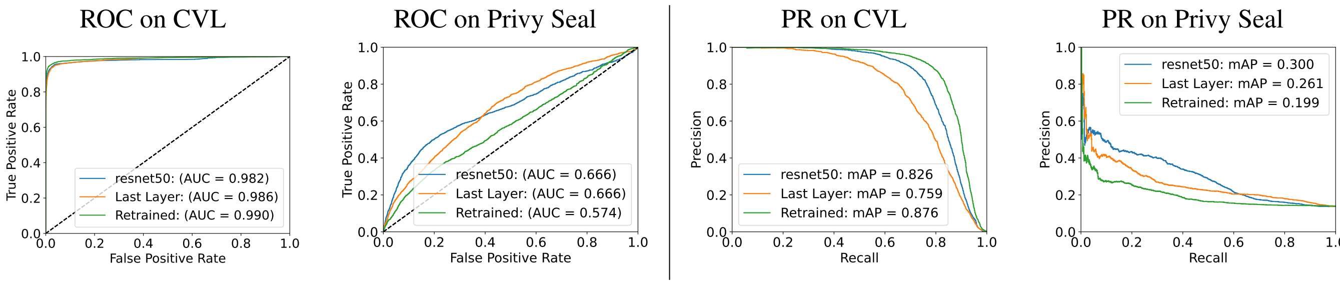


Figure 5. Our trained ResNets' Performance on the CVL dataset and our Privy Seal evaluation dataset as well as that of a ResNet-50 only trained on ImageNet

4.1.2 Evaluation of trained models

To evaluate our trained models' performance, we used the same evaluation procedure as the on the untrained models. Each query image is broken up into 64×64 tiles and we use our model to generate a feature vector for each tile containing text. We then average the vector generated by our model for each input patch to give us a feature vector that is representative of the totality of the image. The cosine similarity is then compared between every image to give us a similarity score. These results are shown in Figure 5. The performance on CVL does not approach state of the art, but this was intended more as a sanity check.

Interestingly, the trained models perform only marginally better on CVL, and significantly worse on the Privy Seal dataset. This is despite being trained on more "Scribal" data than modern handwriting, as the ICDAR 17 dataset is larger than the IAM dataset of modern text. Upon further investigation, we found that this does not appear to hold for individual patches, and the trained models worked significantly better in that case. We used the "Fully Trained" model for our qualitative examples.

5. Software: The Paleographer's AI

This work's most important contribution is the software package that we have developed to streamline the process of using automated handwriting identification features in paleographic attribution. Here we will discuss the steps we've taken to make the software easy to use and encourage adoption. All of the code is open source and available for use with under a MIT License to maximize accessibly. We have also built the tool such that the previously described methodology for comparing patches to a template can be replaced with another model or method of comparison.

5.1. Graphical User Interface

In order to make the software sufficiently usable, we developed a GUI using PyWebView[5]. When starting the software the user is presented with a splash screen as shown in Figure 6. From that screen, the user can choose an algorithm to use to evaluate the image, and then build a template for a specific identity. The software then displays a file chooser screen, and the user is prompted to select a folder containing the images that will build the template.

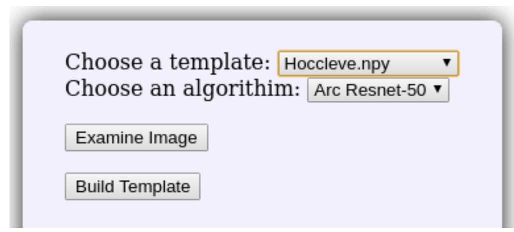


Figure 6. The initial Splash Screen of our software.

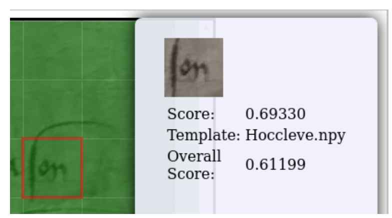


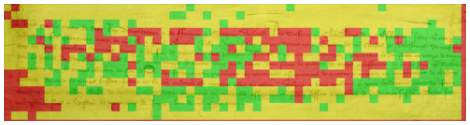
Figure 7. The user can mouse over portions of the document to see more details about each tile.

After building as many templates as needed, the user can choose a template from the drop-down list and use it to examine an image. This brings the user to a new "Examine Image" screen. From here a user can choose an image to load, and then the model will process it, giving similarity scores for each tile. In the output, tiles with a low similarity are shaded red, tiles that are on the lower end of the accept threshold, which were determined empirically from the Privy Seal dataset, and can also be adjusted by the user, are shaded yellow, and very high similarity tiles are shaded green. Users can mouse over areas of interest, and the tiles will be displayed on a side panel along with their raw similarity scores. The "Examine Image" screen is shown in Figure 1, a closer view of some of the elements can be seen in Figure 7, and a zoomed out view of a positive and negative match are shown in Figure 8.

5.2. Distribution

Even the most technical among us has struggled with installing the correct dependencies, and paleographers are

[5] https://pywebview.flowrl.com/



Hoccleve Example Compared to Hoccleve Template

Hoccleve Example Compared to Langeport Template

Figure 8. An example showing a document written by Hoccleve compared to a Hoccleve template (top) and the same document compared to a template of another Privy Seal scribe, Richard Langeport (bottom).

not typically the most technical among us. Therefore, we have packaged our tool for Windows, MacOS, and (Ubuntu) Linux using PyInstaller[6] to make standalone executable files with all of the dependencies included. This makes getting started with our tool much easier than it would be otherwise, completely removing the need to set up a virtual environment and install dependencies. To maximize compatibility the "prebuilt" version of our software is designed to be run on the CPU, but inference can also be run on a GPU when the environment is set up manually.

6. Case Study: Hoccleve or not?

To demonstrate one way in which our software could be used, we go back to Thomas Hoccleve, a highly productive and leading clerk of the Privy Seal who also happens to be one of the most important fifteenth-century English poets. His handwriting is the most extensively studied of any medieval English writer, and Hoccleve has left behind more than 1000 government records written in his hand and several holograph manuscripts as well as books to which he contributed his handwriting [3, 10, 11, 27, 35, 40]. This makes identifying his handwriting a very interesting use case for our software. To demonstrate how our tool might be used, we took some instances of his handwriting and examples not written by him, and ran these through our fully trained model. We generated a biometric template for Hoccleve's handwriting using the patches surrounding the top 100 ORB [31] keypoints in the Hoccleve images from our Privy Seal Dataset.

Our first example is a sanity test. The image is from Bodmer 48, a later 15th-century manuscript of Chaucer's *Canterbury Tales* copied by a scribe 40-50 years after Hoccleve's death. ("Bodmer" in Figure 9) Like the other examples in this case study, it is written in an English Secretary script. This script bears many similarities to our target but is immediately visually distinguishable from Hoccleve's handwriting to anyone familiar with his hand; no human paleographer would identify this example as Hoccleve. The model agrees: almost all of the text in our example image has been colored red. (Blank expanses of paper or parchment tend to result in false positives, as shown here.)

The second image is from a copy of Hoccleve's *Regiment of Princes*, British Library MS Royal 17 D. xviii. ("Royal" in Figure 9) In 2011, Linne R. Mooney advanced an argument that it was in fact a holograph — the original manuscript itself [28]. However, this attribution was quickly met with skepticism: Daniel Wakelin first noted that some letterforms in the manuscript are uncharacteristic for Hoccleve; Lawrence Warner offered a more thorough analysis of the differences, concluding that the manuscript is not in Hoccleve's hand; and Sebastian Sobecki further observed that the script is dissimilar to that used by scribes of Hoccleve's office, the Privy Seal [40,45,46]. As an example that caused disagreement amongst experts, it represents an interesting benchmark. One would expect the model to evaluate it as significantly more similar to Hoccleve than the Bodmer manuscript used as a control. Our model's evaluation of it is shown in Figure 9. While there is a not insignificant amount of green on the paper parts of the image, we see the majority of the text itself is yellow or red. Accordingly, we align ourselves with the critics: the Royal manuscript is unlikely to be Hoccleve's.

British Library MS Harley 219 contains many different texts and was written by several scribes, one of whom was recently identified as Hoccleve by Misty Schieberle [32]. ("Harley" in Figure 9) Here we compare folio 134r, a portion of Christine de Pizan's Middle French *Epistre Othea*, to our Hoccleve template, and find significantly positive results. So far, no one has contested Schieberle's attribution. Our results do not compel us to do so, either.

The scribe of our final example, Cambridge, Trinity College MS O.7.43, has never been identified. ("TCC" in Figure 9) The manuscript is a small miscellany of only 43 folios, written by a single scribe in brown ink, with some blue initials with red detailing. The sparse description by cataloguer M. R. James says only: "Cent. xv, very well written." Very well written... by Hoccleve? Plausibly so, according to our model, and the paleographers on our team agree [41].

The paleographers on our team (whose own *oculi palaeographici* accept the Harley and TCC attributions and reject that of the Royal) are highly encouraged by these results, because the Hoccleve items we used to generate this comparison template are all from his work at the Privy Seal. They are, therefore, examples of how Hoccleve wrote when he was writing documents, not literary works. This shows that the tool can give useful results across similar scripts, rather than only the exact same form of script. As shown in Figure 8, it can also clearly distinguish between different scribes writing precisely the same script.

[6] https://pyinstaller.org/

Figure 9. The result of comparing known Hoccleve examples against folios from four different manuscripts: Cologny, Fondation Martin Bodmer, Cod. Bodmer 48; London, British Library, MS Royal 17 D.xviii; London, British Library MS Harley 219; Cambridge, Trinity College MS O.7.43.

7. Conclusion and Future Work

Despite excellent results on benchmark datasets in the literature, previous work by computer vision scientists has not significantly impacted the field of medieval paleography. Paleographers have found no turnkey methodology for the stylistic analysis of handwriting that can help identify different scribes. While our tool does not entirely solve this problem, we believe that it takes some important first steps. Firstly, our approach has included paleographers in every step of the design process. With the tool already in the hands of paleographers, development of our software will continue with their feedback as a guiding force. Our case study on manuscripts attributed to Thomas Hoccleve is an example of how paleographers are already able to use this tool in support of their scribal attributions. We offer solid foundations for important future research lines that are central to our understanding of literary and intellectual culture, such as identifying scribal collaboration in manuscripts and locating spatial clusters of scribes affiliated with various civic, national, and religious institutions in medieval London and elsewhere.

This tool is very much a work in progress. We plan to continue supporting and developing it as an open source project. We do hope to include a wider variety of algorithms in future versions of the software. Including some of the pre-machine learning methods that are still quite competitive, different types of metric learning techniques such as siamese networks, and transformer based models that can take in additional context. Additionally, we hope to add more sophisticated layout analysis tools that can give the software more consistent alignment to the text.

A tool that can automatically identify script types and scribal hands would mark a watershed in the field of paleography. Software that succeeds at this task could be run over large and previously unanalyzed corpora, offering insight in days that may have taken an entire career's worth of work before. Even a moderately accurate tool of this sort would help paleographers identify interesting digitized manuscripts available online.

However, paleographers are likely to be skeptical of new findings that are supported only by a high similarity score. We expect that the identification of TCC MS O.7.43 as the hand of Hoccleve on the basis of our model alone is unconvincing without further explanation of the paleographic features of this manuscript, which will follow in a separate paper. Computer vision scientists may be able to create tools that assist paleographers with this descriptive step as well, by calculating some explicit measurements relevant to scribal aspect. For example, Sobecki used an orthopedic goniometer to measure — by hand — the angles of long letters in order to distinguish individual Privy Seal scribes from one another [40]. A computer application would be more convenient. While our software does not currently do this, it is something that we would like to work on in future versions. This kind of measurement would not only make comparison of hands and collaboration between paleographers easier, it would also make the confidence level of paleographers' results more apparent to non-specialists.

8. Acknowledgements

We gratefully acknowledge The National Archives; The Martin Bodmer Foundation, Cologny (Geneva); the British Library; and Trinity College, Cambridge, for use of their collections. The original images from the British Library are © British Library Board. The original image from Cambridge, Trinity College is Copyright the Master and Fellows of Trinity College, Cambridge, CC-BY-NC 4.0 (https://mss-cat.trin.cam.ac.uk/Manuscript/O.7.43). We also want to thank Euan Rodgers from the British Library and Jessica Lockhart from the Old Books New Science Lab at the University of Toronto.

References

[1] Herbert Bay, Andreas Ess, Tinne Tuytelaars, and Luc Van Gool. Speeded-up robust features (surf). *Computer vision and image understanding*, 110(3):346–359, 2008. 2

[2] Marius Bulacu and Lambert Schomaker. Automatic handwriting identification on medieval documents. In *14th International Conference on Image Analysis and Processing (ICIAP 2007)*, pages 279–284. IEEE, 2007. 2

[3] J. A. Burrow and A. I. Doyle, editors. *Thomas Hoccleve: A facsimile of the autograph verse manuscripts: Henry E. Huntington Library, San Marino (California), MSS HM III and HM 744; University Library, Durham (England), MS Cosin V. III. 9*. Early English Text Society. Published for the Early English Text Society by Oxford University Press, Oxford, 2002. 7

[4] Martha Carlin. Thomas spencer, southwark scrivener (d. 1428): Owner of a copy of chaucer's troilus in 1394? *the chaucer review*, 49(4):387–401, 2015. 1

[5] Sukalpa Chanda, Emmanuel Okafor, Sebastien Hamel, Dominique Stutzmann, and Lambert Schomaker. Deep learning for classification and as tapped-feature generator in medieval word-image recognition. In *2018 13th IAPR International Workshop on Document Analysis Systems (DAS)*, pages 217–222. IEEE, 2018. 3

[6] Vincent Christlein, David Bernecker, Andreas Maier, and Elli Angelopoulou. Offline writer identification using convolutional neural network activation features. In *Pattern Recognition: 37th German Conference, GCPR 2015, Aachen, Germany, October 7-10, 2015, Proceedings 37*, pages 540–552. Springer, 2015. 2

[7] Vincent Christlein, Anguelos Nicolaou, Mathias Seuret, Dominique Stutzmann, and Andreas Maier. Icdar 2019 competition on image retrieval for historical handwritten documents. In *2019 International Conference on Document Analysis and Recognition (ICDAR)*, pages 1505–1509. IEEE, 2019. 2, 4

[8] Jia Deng, Wei Dong, Richard Socher, Li-Jia Li, Kai Li, and Li Fei-Fei. Imagenet: A large-scale hierarchical image database. In *2009 IEEE conference on computer vision and pattern recognition*, pages 248–255. Ieee, 2009. 5

[9] Jiankang Deng, Jia Guo, Niannan Xue, and Stefanos Zafeiriou. Arcface: Additive angular margin loss for deep face recognition. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pages 4690–4699, 2019. 3, 5

[10] A Ian Doyle and Malcolm Parkes. The production of copies of the canterbury tales and the confessio amantis in the early fifteenth century. In V. J. Scattergood and Andrew G. Watson, editors, *Medieval Scribes, Manuscripts, and Libraries: Essays Presented to N. R. Ker*, pages 163–210. Scolar, 1978. 7

[11] A. I. Doyle and M. B. Parkes. Palaeographical introduction. In Paul G. Ruggiers, editor, *The Canterbury Tales: A Facsimile and Transcription of the Hengwrt Manuscript with Variants from the Ellesmere Manuscript*, page xix–xlix. University of Oklahoma Press, Norman, OK, 1979. 7

[12] Stefan Fiel, Florian Kleber, Markus Diem, Vincent Christlein, Georgios Louloudis, Stamatopoulos Nikos, and Basilis Gatos. Icdar2017 competition on historical document writer identification (historical-wi). In *2017 14th IAPR International Conference on Document Analysis and Recognition (ICDAR)*, volume 1, pages 1377–1382. IEEE, 2017. 2, 4, 5

[13] Stefan Fiel and Robert Sablatnig. Writer identification and retrieval using a convolutional neural network. In *Computer Analysis of Images and Patterns: 16th International Conference, CAIP 2015, Valletta, Malta, September 2-4, 2015, Proceedings, Part II 16*, pages 26–37. Springer, 2015. 2

[14] Kaiming He, Xiangyu Zhang, Shaoqing Ren, and Jian Sun. Deep residual learning for image recognition. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 770–778, 2016. 5

[15] Sheng He, Petros Sammara, Jan Burgers, and Lambert Schomaker. Towards style-based dating of historical documents. In *2014 14th International Conference on Frontiers in Handwriting Recognition*, pages 265–270. IEEE, 2014. 3

[16] Sheng He and Lambert Schomaker. Beyond ocr: Multifaceted understanding of handwritten document characteristics. *Pattern Recognition*, 63:321–333, 2017. 3

[17] Simon Horobin. Adam pinkhurst, geoffrey chaucer, and the hengwrt manuscript of the canterbury tales. *Chaucer Review: A Journal of Medieval Studies and Literary Criticism, The*, 44(4):351–367, 2010. 1

[18] Simon Horobin. Thomas hoccleve: Chaucer's first editor? *The Chaucer Review*, 50(3-4):228–250, 2015. 1

[19] Philip Kahle, Sebastian Colutto, Günter Hackl, and Günter Mühlberger. Transkribus - a service platform for transcription, recognition and retrieval of historical documents. In *2017 14th IAPR International Conference on Document Analysis and Recognition (ICDAR)*, volume 04, pages 19–24, 2017. 3

[20] Helen K. S. Killick. *Thomas Hoccleve as poet and clerk*. phd, University of York, Nov 2010. 4

[21] Diederik P Kingma and Jimmy Ba. Adam: A method for stochastic optimization. *arXiv preprint arXiv:1412.6980*, 2014. 5

[22] Florian Kleber, Stefan Fiel, Markus Diem, and Robert Sablatnig. Cvl-database: An off-line database for writer retrieval, writer identification and word spotting. In *2013 12th international conference on document analysis and recognition*, pages 560–564. IEEE, 2013. 2, 4

[23] Lorenzo Lastilla, Serena Ammirati, Donatella Firmani, Nikos Komodakis, Paolo Merialdo, and Simone Scardapane. Self-supervised learning for medieval handwriting identification: A case study from the vatican apostolic library. *Information Processing & Management*, 59(3):102875, 2022. 3

[24] David G Lowe. Distinctive image features from scale-invariant keypoints. *International journal of computer vision*, 60:91–110, 2004. 2

[25] U-V Marti and Horst Bunke. The iam-database: an english sentence database for offline handwriting recognition. *International Journal on Document Analysis and Recognition*, 5:39–46, 2002. 4, 5

[26] Linne R Mooney. Chaucer's scribe. *Speculum*, 81(1):97–138, 2006. 1, 2

[27] Linne R. Mooney. Some new light on thomas hoccleve. *Studies in the Age of Chaucer*, 29(1):293–340, 2007. 4, 7

[28] Linne R. Mooney. A holograph copy of thomas hoccleve's regiment of princes. *Studies in the Age of Chaucer*, 33(1):263–296, 2011. 7

[29] Marco Peer, Florian Kleber, and Robert Sablatnig. Towards writer retrieval for historical datasets. In *International Conference on Document Analysis and Recognition*, pages 411–427. Springer, 2023. 3

[30] Jane Roberts. On giving scribe b a name and a clutch of london manuscripts from c. 1400. *Medium Aevum*, 80(2):247–270, 2011. 1, 2

[31] Ethan Rublee, Vincent Rabaud, Kurt Konolige, and Gary Bradski. Orb: An efficient alternative to sift or surf. In *2011 International conference on computer vision*, pages 2564–2571. Ieee, 2011. 4, 7

[32] Misty Schieberle. A new hoccleve literary manuscript: The trilingual miscellany in london, british library, ms harley 219. *The Review of English Studies*, 70(297):799–822, Nov 2019. 7

[33] Lambert Schomaker. How deep is deep and what's next in computational intelligence? In *2016 15th International Conference on Frontiers in Handwriting Recognition (ICFHR)*, page xx–xxii, Shenzhen, China, Oct 2016. IEEE. 3

[34] Florian Schroff, Dmitry Kalenichenko, and James Philbin. Facenet: A unified embedding for face recognition and clustering. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, June 2015. 3

[35] H. C. Schulz. Thomas hoccleve, scribe. *Speculum*, 12(1):71–81, Jan 1937. 7

[36] Sebastian Sobecki. " the writyng of this tretys": Margery kempe's son and the authorship of her book. *Studies in the Age of Chaucer*, 37(1):257–283, 2015. 1

[37] Sebastian Sobecki. Ecce patet tensus: The trentham manuscript, in praise of peace, and john gower's autograph hand. *Speculum*, 90(4):925–959, 2015. 1

[38] Sebastian Sobecki. *Last words: The public self and the social author in late medieval England*. Oxford University Press, 2019. 1

[39] Sebastian Sobecki. Communities of practice: Thomas hoccleve, london clerks, and literary production. *Journal of the Early Book Society for the Study of Manuscripts and Printing History*, 24:51–403, 2021. 3

[40] Sebastian Sobecki. The handwriting of fifteenth-century privy seal and council clerks. *The Review of English Studies*, 72(304):253–279, 2021. 4, 7, 8

[41] Sebastian Sobecki. Gens sans argent: A new holograph manuscript by thomas hoccleve. *The Library*, 25, 2024. 7

[42] Peter A Stokes. Digital approaches to paleography and book history: some challenges, present and future. *Frontiers in Digital Humanities*, 2:5, 2015. 3

[43] Peter A Stokes. Scribal attribution across multiple scripts: A digitally aided approach. *Speculum*, 92(S1):S65–S85, 2017. 3

[44] Youbao Tang and Xiangqian Wu. Text-independent writer identification via cnn features and joint bayesian. In *2016 15th International Conference on Frontiers in Handwriting Recognition (ICFHR)*, pages 566–571. IEEE, 2016. 2

[45] Daniel Wakelin. *Scribal correction and literary craft: English manuscripts 1375-1510*. Cambridge studies in medieval literature. Cambridge University Press, Cambridge, U.K, 2014. 7

[46] Lawrence Warner. Scribes, misattributed: Hoccleve and pinkhurst. *Studies in the Age of Chaucer*, 37(1):55–100, 2015. 7

[47] Lawrence Warner. *Chaucer's Scribes: London Textual Production, 1384–1432*. Cambridge University Press, 2018. 1