Multimodal AI, Image Analysis, and the Illustrated Periodical Press

The Victorian periodical press inaugurated the mass image. With steam power, industrialized printing, and durable wood engraving techniques, periodicals like the *Penny Magazine* and the *Illustrated London News* accelerated the fidelity and distribution of printed image to create the "first mass-media era" (Hughes 2014). The periodical press achieved enormous volumes of production that practically overwhelmed readers then and continues to challenge researchers now. While the digitization of periodicals has allowed for text-based searching and analysis at increasing scales, dealing with the visual materials of the periodical press is much harder. Thomas (2017) notes that 'while tools for analysing and data mining nineteenth-century texts are becoming ever more sophisticated, those for analysing nineteenth-century images are falling behind.' Existing digital collections do not adequately separate or make searchable their illustrated content. Nor can researchers explore that content visually. Our project offers new approaches to both of these problems: access to a digitized corpus of nineteenth-century periodical illustrations, and analysis of those illustrations through keyword and image-based searches.

Efforts have been made to extract, tag, and categorize nineteenth-century historical illustrations as in projects like the Database of Mid-Victorian Illustration and the Illustration Archive. But these projects are limited by the formidable labor of extracting and tagging images by hand, as it were. Our project introduces machine learning techniques to automate the extraction and labeling of illustrations from digital collections of periodicals, resulting in a dataset researchers can explore with AI-assisted techniques. Our initial efforts will focus on the weekly periodical *Illustrated London News* (*ILN*), though we will also share with other researchers how to adapt these methods to other accessible collections. Building on the project's technical methods, we will also demonstrate how an interpretive methodology of "distant viewing" may help reorient scholarship to the scale of Victorian periodical illustration (Arnold and Tilton, 2019). While existing scholarship necessarily bases its claims about a small set of representative images, our project can facilitate the gathering and analysis of illustrations at close as well as distant scales.

Ultimately, this project aims to (1) produce a new open-access dataset of the *ILN* (1842-1900), where both the text and ca. 90,000 illustrations of the periodical are fully-searchable with textual and visual prompts; (2) show researchers how to produce similar searchable image collections for other digitized illustrated periodicals; and (3) collaboratively develop the experimental methods and case studies with which scholars can newly confront the massive visual archive of the Victorian press. To achieve this, we will publish all our datasets, technical pipelines, and instructions online in stable repositories, and use a substantial part of the funds of the Field Development Grant to organize a hybrid workshop, introducing scholars from various backgrounds to the *ILN* dataset and the computational methods to access and study it. As part of that workshop, we will ourselves undertake a case study in the analysis of the *ILN*, with the goal of publishing in *Victorian Periodicals Review*. Finally, we will also produce an accompanying white paper for other researchers to replicate and/or expand on our work.

This project offers "field development" in a number of ways. First, by producing new data sets and digital collections of periodical illustrations. Second, by helping other scholars develop other collections based on digitally-accessible periodicals. Third, by developing new methods in AI analysis. Fourth, by inviting other scholars to apply those methods to scholarship on periodicals, illustrations, and visual culture.

Work Plan

The project will proceed through four stages of a work plan, including: (1) generating the **data**, (2) developing the distant viewing **methods,** (3) producing the **pipeline**, and (4) **analysis** of the dataset

(pilot study and workshop). The stages are divided into several work packages, described below. An overview of the different stages, including the responsible team member and deliverables, is provided in Table 1.

WP1: Data

Currently, researchers can find a few digitized text-searchable versions of the *ILN* online. In 2010, Gale released its *Illustrated London News* Historical Archive. Since 2016, the publication can also be accessed via the British Newspaper Archive (Fyfe, 2016; Smits, 2016). While Gale mostly focuses on the academic market, the British Newspaper Archive sells individual subscriptions at relatively modest prices. While the paywalls present their own problems (especially for materials in the public domain), these collections also cannot provide separate access to the *ILN*'s key asset: its illustrations. Nor do these collections allow distant reading/viewing methodologies, as researchers cannot access the underlying data.

Our project relies on a digitized version of the *ILN* in the Internet Archive's Serials in Microfilm Collection. This collection, donated to the Internet Archive in 2018, contains over 14,000 titles with an estimated 500 million pages. Along with the *ILN*, the collection includes many other notable titles including *Punch, Leslie's Weekly,* and *Harper's Weekly*. Unlike HathiTrust Digital Library, the Internet Archive allows for bulk downloads of its collections. Thus, we will harvest data from the Internet Archive, and build a pipeline (described below) that can similarly deliver datasets from the other illustrated periodicals it contains.

We start by downloading all 3192 issues of the *ILN* that were published between 14 May 1842 and 1900. These files include PDFs of scanned microfilms and the full text from the Internet Archive's own OCR. In the next step, we extract all the illustrations from the scanned pages. To do this, we have to first train and test an object detection model. We randomly divide the ca. 1151 pages of 58 issues (one per year) into a standard 80/10/10 train, test, and validation set. This allows us to retrain the Newspaper Navigator model, developed by Lee (2021) for the Library of Congress's Chronicling America collection, to detect only illustrations and titles/captions from the *ILN*.

At the end of WP1, we release an initial version of the dataset on Zenodo. This dataset will contain fields for the filename, page number, dimensions (pixels) of the illustration, OCR caption, and a link to the full-text and PDF versions of the issue on the Internet Archive. Using Jupyter Notebooks--a web-based interactive computing environment that allows researchers to create and share documents (notebooks) that contain code and explanatory text--we also release the first part of the pipeline, which enables researchers to (bulk) download other titles from the Internet Archive and extract illustrations, images, or photographs from them.

WP2: Methods

In the second part of the project, we apply recently-developed multimodal AI techniques to the illustrations we harvested in stage one. In the last five years, scholars have used visual AI techniques (computer vision) to analyze extensive collections of digitized images (Arnold and Tilton, 2019; Wevers and Smits, 2020). However, most computer vision techniques were trained on modern high-definition photographs, resulting in what has been called the 'historical bias' of computer vision (Smits and Wevers, 2021). Trained on the visual world of the early twenty-first century, computer vision algorithms find it difficult to detect patterns in historical visual cultures, including the lineated engravings characteristic of many nineteenth-century illustrated newspapers (Fyfe and Ge, 2018).

Recently introduced multimodal machine learning models, such as the Contrastive Language-Image Pre-Training (CLIP) neural network, may provide a new approach to the problem of historically-biased AI. In contrast to traditional computer vision techniques, which learn to recognize a limited set of pre-determined classes (person, dog, car, desk, etc.), multimodal models learn to

connect text to images and vice versa. Trained on 400 million image-text combinations, CLIP is optimized to arrange images and captions by similarity in a mathematical feature space. As a result, it can identify textual aspects that are good predictors of visual elements.. Furthermore, because CLIP's 400 million original image-text pairs included all sorts of visual media, CLIP also works effectively with non-photographic images.

WP2 begins by using CLIP to extract embeddings from all the illustrations (roughly 90,000) of the *ILN* in the period 1842-1900. Researchers can use these embeddings for three retrieval tasks: employing textual queries to retrieve illustrations (text-to-image), visual queries to retrieve illustrations (image-to-image), and connecting illustrations to textual descriptions or metadata categories (image-to-text) (Smits and Wevers, forthcoming). For example, the *text-to-image* task lets researchers use textual prompts to search for illustrations. These prompts can be relatively generic, such as 'an illustration showing an animal,' or more specific, such as 'an illustration showing a zebra.' We can retrieve images of specific scenes – for example, a church service, a horse race, or a mass demonstration, specific places, such as landmarks, like Big Ben, types of persons (priests, politicians, or soldiers) and even specific persons, like Queen Victoria. We can also retrieve illustrations of more abstract concepts, such as 'an illustration of war' or 'an illustration showing a couple in love'. Without having to (manually) add metadata, CLIP allows us to search a large collection of illustrations in the same way that OCR allows keyword searches in large collections of texts.

Visual prompts can also be used to explore the *ILN* dataset (*image-to-image*). For example, we can start with a modern photograph of London Bridge and retrieve the historical illustrations deemed most similar. Thus, CLIP can help with several visual similarity and clustering tasks, including identifying duplicate or reprinted images in this corpus. This possibility becomes especially interesting if we connect embeddings from the *ILN* dataset to those of other visual collections, such as the DMVI or the Illustration Archive. It may allow researchers to explore the movement of visual content across publications or media forms.

Finally, CLIP allows us to efficiently add all sorts of metadata to the illustration in our corpus (*image-to-text*). This method could potentially supplement the manual tagging of images by researchers or crowd workers. To explore this possibility, we will start with a limited number of prompts--for example, identifying images by setting (outside/inside) or medium (photograph/illustration) or content (map/not a map)--and ask CLIP to help classify similar images as belonging to that class.

At the end of the second phase, we will upload an expanded version of the *ILN* dataset to Zenodo, including an embedding of each illustration. We will also produce new Jupyter Notebooks, which should make it possible to use CLIP to extract embeddings from any other newspapers in the Internet Archive's collection, and enable users to try the three retrieval tasks described above (text-to-image, image-to-image, image-to-text).

WP3: Pipeline

Our full code pipeline combines the Jupyter Notebooks produced in stages one and two. The pipeline allows users to bulk download PDFs and OCR text of periodicals in the Internet Archive's Serials in Microfilm Collection, extract illustrations from these digitized periodicals, and apply CLIP for the three distant viewing retrieval tasks. To clarify, we are not producing a website, but publishing data and the code to use it. A Jupyter Notebook (accessed in Google's Collab environment) is a web-based interactive computing platform. Users do not have to know how to code to use it. The notebook explains its code and allows the user to load it, tweak the prompts as needed, and preview the results, which are also downloaded locally to the user's machine.

WP4: Analysis and Dissemination
The project's final stage moves from technical development to the analysis, interpretation, and dissemination of its results. This includes evaluating the effectiveness of CLIP and its implications as a methodology for historical visual studies. In WP4, we undertake our own such reflections in a research case study and white paper. We will also organize a pre-conference workshop at RSVP 2024 to teach other researchers about the methods and possibilities for continued expansion of the field.

WP4.1 Pilot Study and white paper
Using the *ILN* dataset and the methods that the pipeline makes available, the applicants will undertake a pilot study on the 'imperial cartography' of the *ILN* (Braun 2015). As perhaps the first visual mass medium, the *ILN* is a perfect case to study the connection between the constitution of the British Empire and modern regimes of visuality. We examine this relationship by distant viewing the proliferation in the *ILN* of a particular visual 'tool of empire' (Headrick, 1981): the map. Our dataset and searchable embeddings make it possible to view the development of imperial cartography at scale. Using multimodal machine learning to extract all maps from the dataset, then search other illustrations named for these areas, we can chart which parts of the world the *ILN* visualized and study its rhetorics. Alongside this case study, the team will also develop a white paper about the computational techniques and methods.

WP4.2 Workshop
We will also host a hybrid workshop to introduce scholars of different backgrounds and career stages to the *ILN* dataset and our computational methods, seeking to expand scholars' access to image collections, their analytical toolkits, and perspectives on Victorian periodical studies. We propose to use a substantial part of the funds of the Field Development Grant to organize such a workshop, open to applicants alongside invited scholars. We hope to create a diverse, interdisciplinary conversation about the possibilities of these methods, and demonstrate step-by-step how to use them. Ideally, we would like to host this as a pre-conference workshop during RSVP 2024. If that is not possible, we will organize the conference at one of our institutions.

Outcomes
- A dataset of ca. 90,000 illustrations published in the *ILN* between 1842 and 1900.
- A Jupyter Notebook which applies multimodal AI models to the dataset, powering the three search and retrieval tasks described above.
- A pipeline that allows users to (bulk) download other titles from the Internet Archive's Serials in Microfilm Collection, extract illustrations, apply CLIP, and connect results to the *ILN* dataset.
- A hybrid workshop that introduces scholars to the *ILN* dataset and computational methods.
- An article on imperial cartography submitted to *Victorian Periodicals Review*.
- A white paper with complete instructions about the dataset, pipeline, and how to use them.

Digital sustainability
We approach digital sustainability in several ways. First, by not relying on graphic user interfaces, such as a website, which requires continual updating. Jupyter Notebooks and the underlying Python code are industry-standard tools, and have already proven durable. Second, the project works in conjunction with the Internet Archive, whose materials--unlike those provided by commercial publishers--are as accessible and as relatively stable as anything on the web. Third, we produce datasets in a clearly structured tabular format (TSV), which ensures usability across applications.

Finally, we post our datasets, models, and code to GitHub (a web-based platform that allows users to store, manage, and share their software code with others) and Zenodo (a free, open-access digital repository that allows scholars to share and preserve their datasets in a secure and long-term manner). These materials will all be released without restrictions at the end of the project.

Our Team

The collaborators on this project all bring substantial experience on digital projects as well as research expertise on the digital implications for illustration and periodical studies. Please see the accompanying two-page CVs for additional information.

- Thomas Smits (Principal Investigator), Postdoctoral Fellow, University of Antwerp
- Ben Lee, PhD Candidate in Computer Science, University of Washington
- Julia Thomas, Professor in the School of English, Communication and Philosophy, Cardiff University
- Paul Fyfe, Associate Professor of English, North Carolina State University
- The project hires two research interns (see budget). At Cardiff this will be Irene Testini, Research Software Engineer, Cardiff University

Work package	**Team member(s)**	**Deliverable**	**Sep. 2023 - Sep. 2024**
WP1 Data	Lee & Smits	-ILN dataset (V1)	
		-First part pipeline	
WP2 Methods	Smits & Thomas	-ILN dataset (V2)	
		-Second part pipeline	
WP3 Pipeline	Thomas & Lee	Full pipeline	
WP4.1 Pilot	Smits & Fyfe	-*VPR* article	
		-White paper	
WP4.2 Workshop	Fyfe & Smits	Workshop	

Table 1: Responsible team member, deliverables, and planning of the work packages. The research assistants will assist in WP1 - WP 3.

Bibliography

Arnold T and Tilton L (2019) Distant viewing: analyzing large visual corpora. *Digital Scholarship in the Humanities* 34(Supplement_1). Oxford Academic: i3–i16. DOI: 10.1093/llc/fqz013.

Braun L (2015) Colonial and Imperial Cartography. In: Monmonier M (ed.) *The History of Cartography, Volume 6: Cartography in the Twentieth Century*. University of Chicago Press, pp. 251–255.

Fyfe P (2016) An Archaeology of Victorian Newspapers. *Victorian Periodicals Review* 49(4): 546–577. DOI: 10.1353/vpr.2016.0039.

Fyfe P and Ge Q (2018) Image Analytics and the Nineteenth-Century Illustrated Newspaper. *Journal of Cultural Analytics*. DOI: 10.22148/16.026.

Headrick DR (1981) *The Tools of Empire: Technology and European Imperialism in the Nineteenth Century*. Oxford University Press.

Hughes, L (2014) SIDEWAYS!: Navigating the Material(ity) of Print Culture. *Victorian Periodicals Review* 47(1): 1–30.

Lee B (2021) Compounded Mediation: A Data Archaeology of the Newspaper Navigator Dataset. *Digital Humanities Quarterly* 15(4).

Smits T (2016) Making the News National: Using Digitized Newspapers to Study the Distribution of the Queen's Speech by W. H. Smith & Son, 1846–1858. *Victorian Periodicals Review* 49(4): 598–625. DOI: 10.1353/vpr.2016.0041.

Smits T and Wevers M (2021) The Agency of Computer Vision Models as Optical Instruments. *Visual Communication* OnlineFirst. DOI: 10.1177/1470357221992097.

Smits T and Wevers M (forthcoming) A multimodal turn in Digital Humanities? Using contrastive machine learning models to explore, enrich, and analyze digital visual historical collections. *Digital Scholarship in the Humanities*.

Thomas J (2017) *Nineteenth-Century Illustration and the Digital*. Cham: Springer International Publishing. DOI: 10.1007/978-3-319-58148-4.

Wevers M and Smits T (2020) The Visual Digital Turn. Using Neural Networks to Study Historical Images. *Digital Scholarship in the Humanities* 35(1): 194–207. DOI: https://doi.org/10.1093/llc/fqy085.