Reviews in Digital Humanities • Vol. 2, No. 4

Review: Distant Viewing Toolkit

Patrick Sui¹

¹University of Texas at Austin

Published on: Apr 12, 2021

License: [Creative Commons Attribution 4.0 International License \(CC-BY 4.0\)](https://creativecommons.org/licenses/by/4.0/)

Project

The Distant Viewing Toolkit

Project Directors

[Taylor Arnold](#), University of Richmond

[Lauren Tilton](#), University of Richmond

Project URL

<https://distant-viewing.github.io/dvt/dvt/>

Project Reviewer

[Patrick Sui](#), University of Texas at Austin

Project Overview**Taylor Arnold & Lauren Tilton**

The Distant Viewing Toolkit (DVT) is a software library designed to facilitate the computational analysis of visual culture. The toolkit applies state-of-the-art computer vision algorithms to digitized collections of still and moving images in order to automatically create structured data. Features represented in the structured data include dominant colors found in the images, cut-detection, object detection and recognition, face detection and recognition, and image embeddings. The features map onto concepts from fields such as film, media, and visual culture studies. For example, face detection can be used to analyze shot type. The structured features can be aggregated and analyzed to find patterns across a corpus or used to increase search and discovery within a large archive. The toolkit is provided as a module for the open-source programming language Python, a popular language for research in both computer vision and machine learning.

The design and motivation for the toolkit also comes from the authors' theoretical work on "distant viewing," the concept that the challenges of a computational analysis of large visual corpora can be understood in terms of visual semiotics and communication theory.¹ In conversation with critical work on computer vision, the toolkit provides access to and control over how and what a user decides to view through computer vision algorithms. The toolkit allows users to apply a sequence of algorithms to a corpus with various degrees of customization. The default pipeline works well for many analyses. For more control, and to improve performance on some

input types, the toolkit makes it straightforward to modify additional tuning parameters; it is possible, even, to create entirely new annotation algorithms within the framework of the software.² There is also a module included in the toolkit for visualizing and checking the automated annotations through an interactive website. The toolkit, therefore, is designed to give users the flexibility for customization to pose the questions that animate scholars who work with digital humanities methods, while understanding the assumptions being made by the software.

The primary audience in mind for the toolkit includes people working with collections of still and/or moving images who have some experience writing code but no specialized knowledge with computer vision or machine learning algorithms. In order to make the toolkit accessible, we have developed numerous resources and workshops for introducing the method to audiences from different backgrounds. The current project website has a direct link to a Google Collaboratory environment, which facilitates users from any background to get started with the toolkit.³ Another link provides a demo of the possibilities of visualizing extracted data from several feature-length films. In June 2019, we ran a week-long workshop through the Humanities Intensive Learning and Teaching (HILT) program. These notes, including Python notebooks, are available for a more extensive introduction to computer vision and its applications in the humanities.

DVT was developed with attention to best-practices for software development in the digital humanities and open-source software communities. We follow the guidelines established by *the Journal of Open Source Software*, include an open-source license (GPL-2), have extensive unit testing, and use the Contributor Covenant Code of Conduct (v1.4). The project has numerous forms of documentation, including in-line Python comments, a full API documentation, and several online tutorials. Additionally, a peer-reviewed technical paper outlines the major design decisions of the toolkit.⁴ In order to be consistently mindful of both our technical and humanities-oriented audience, every stage of development was done in direct collaboration between the project directors who have a background in statistics and data science (Arnold) and visual culture studies (Tilton).

More information about the toolkit is available on the project website, in our technical design paper,⁵ and in published research papers outlining our applications of the toolkit to the large-scale analysis of television series⁶ and historic photography.⁷ The toolkit is built and maintained by the Distant Viewing Lab (www.distantviewing.org).

Development of the toolkit has been supported by the National Endowment for the Humanities Office of Digital Humanities (HAA-261239-18).

Project Review

Patrick Sui

The Distant Viewing Toolkit (DVT) is an open-source project that gives humanities scholars access to the latest computer vision and machine learning algorithms to interrogate visual culture. Driven by the theory of distant viewing, it aims to lay bare “the interpretive nature of extracting semantic metadata from images” by unpacking “a representation of elements contained within the visual material” with digital methods.⁸ The project engages in meaning making for large visual corpora by using algorithms that interrogate how raw materials are converted into semiotic symbols, a cognitive process that happens unconsciously while we view pictures and videos. While the title of the tool might imply objectivity, DVT actually asserts that all visual content and the technologies that are used to produce and analyze them are socially and culturally constructed. It echoes much of what American studies, visual studies, and cultural studies more generally have been arguing for decades about photographs, moving images, and other forms of visual representation. From detecting narrative arcs within television and film to identifying stylistic features in photographs, DVT offers humanists another avenue to analyze the plethora of visual evidence in digital and born-digital forms.

DVT aims to make its complex technologies accessible and easy to use. Its powerful Python package can be used to extract data from large visual datasets like TV shows, photography archives, and news footage. Its source code is available on Github and can easily run on Python 3.7 with the guidance of helpful in-line comments. The installation process is very straightforward, requiring only a single command line in Anaconda Prompt. The usage then divides into two interfaces: the high-level command-line, which offers the more accessible parts of its functionality, like running data extractions (allowing input videos and images to be extracted as semantic metadata and converted into callable Python objects) and pre-bundled annotators and aggregators (pipelines) as objects, without requiring much familiarity with Python; and the low-level Python API, which allows more experienced programmers to customize their annotators and aggregators. The former alone will be immensely useful for most scholarly purposes, as the pre-bundled pipelines, like the difference annotator and the cut aggregator, are already sufficiently informative for the analysis of most visual

corpora, be it a video file or a collection of static images. The powerfully coded `DataExtraction` class in the high-level command-line draws attention to the distant viewing framework's focus on the interpretative nature of "viewing" images. Additionally, the package contains a JavaScript visualization engine that presents the metadata gathered on each frame, including the time frame, number of detected faces and people, shot number, shot length, and the category of detected objects.

While DVT provides technical tutorials and an API to users in addition to the Github repository, future developments for the project might consider broadening its appeal to teachers. The development of pedagogical materials for teaching purposes as well as sample datasets might extend the reach of the tool into digital humanities classrooms.

Footnotes

1. T. Arnold and L. Tilton. Distant Viewing: Analyzing Large Visual Corpora. *Digital Scholarship in the Humanities*. 34: i3-i16. DOI: 10.1093/llc/fqz013. [↵](#)
2. For more information on how the design of the software as well as how the design enacts the theory of distant viewing, see JOSS article. For more information on how the design of the software as well as how the design enacts the theory of distant viewing, see JOSS article. [↵](#)
3. Computer vision algorithms require a number of external libraries that are extremely efficient but can be difficult to install. Running on Google's Collab allows users to test out the toolkit with minimal setup. [↵](#)
4. Arnold, T. and L. Tilton, (2020). Distant Viewing Toolkit: A Python Package for the Analysis of Visual Culture. *Journal of Open Source Software*, 5(45), 1800. DOI: 10.21105/joss.01800. [↵](#)
5. Arnold, T. and L. Tilton, (2020). Distant Viewing Toolkit: A Python Package for the Analysis of Visual Culture. *Journal of Open Source Software*, 5(45), 1800. DOI: 10.21105/joss.01800. [↵](#)
6. Arnold, T., A. Berke, and L. Tilton (2019). Visual Style in Two Network Era Sitcoms. *Journal of Cultural Analytics*. DOI: 10.22148/16.043 [↵](#)
7. Arnold, T. and L. Tilton (2020). Enriching Historic Photography with Structured Data using Image Region Segmentation. Proceedings of the 1st International Workshop on Artificial Intelligence for Historical Image Enrichment and Access (AI4HI-2020): 1-10. [↵](#)

8. T. Arnold and L. Tilton. Distant Viewing: Analyzing Large Visual Corpora. *Digital Scholarship in the Humanities*. 34: i3-i16. DOI: 10.1093/llc/fqz013. [↗](#)