Distant Viewing

Computational Exploration of Digital Images

By: Taylor Arnold, Lauren Tilton

Citation:

Distant Viewing: Computational Exploration of Digital Images
By: Taylor Arnold, Lauren Tilton
DOI: 10.7551/mitpress/14046.001.0001
ISBN (electronic): 9780262375160
Publisher: The MIT Press
Published: 2023

The open access edition of this book was made possible by generous funding and support from the author

The MIT Press

Notes

Introduction

1. Wu et al., "Detectron2"; Abadi et al., "TensorFlow," 265–283; Chollet, *Deep Learning with Python*; Paszke et al., "PyTorch," 8026–8037; Redmon and Farhadi, "YOLO9000," 7263–7271.

2. Berger, *Ways of Seeing*; Bordwell, *Making Meaning*; Cartwright, *Moral Spectatorship*; Hall, "Encoding/Decoding," 128–138; Manovich, *Language of New Media*; Nakamura, *Digitizing Race*; Raiford, *Imprisoned in a Luminous Glare*; Sturken and Cartwright, *Practices of Looking*; Wexler, *Tender Violence*.

3. Underwood, "A Genealogy of Distant Reading."

4. Jockers, *Macroanalysis*.

5. Piper, *Enumerations*.

6. Salt, *Moving into Pictures*; Tsivian and Civjans, *Cinemetrics*; Butler, "Statistical Analysis of Television Style."

7. Burghardt et al., "Film and Video Analysis in the Digital Humanities"; Flueckiger, "Methods and Advanced Tools"; Masson et al., "Exploring Digised Moving Image Collections."

8. Arnold et al., "Introduction."

9. We were unaware of Bender's contemporaneous use of the term when we also started using it in 2015 in relationship to our early work on the Distant Viewing Toolkit. Bender, "Distant Viewing in Art History."

10. Van Noord, Hendriks, and Postma, "Toward Discovery of the Artist's Style"; Thompson and Mimno, "Computational Cut-Ups."

11. Additional technical details and references are included throughout in notes.

Chapter 1

1. For work on documentary and truth claims, see Nichols, *Representing Reality;* Nichols, *Speaking Truths with Film;* Kahana and Musser, *The Documentary Film Reader*.

2. Abel, *Signs of the Times*; Berger, *Ways of Seeing*; Gray, *Watching Race*; Wexler, *Tender Violence*.

3. For a small sample of scholars studying the visual methods of a set of images in relationship to a larger whole, see Campt, *Image Matters*; Hirsch, *Family Frames;* Raiford, *Imprisoned in a Luminous Glare*; Smith, *Photography on the Color Line;* Hartman, *Wayward Lives*.

4. Neuendorf, *The Content Analysis Guidebook*.

5. Wevers and Smits, "The Visual Digital Turn," 207.

6. Manovich, *Cultural Analytics*, 9.

7. For example, see Barthes, *Image—Music—Text*; Metz, *Film Language*.

8. Saussure, *Cours de Linguistique Générale,* 159.

9. The concept and terminology of a *symbol* originates for Charles Sanders Peirce's first trichotomy of signs into *icons, symbols*, and *indices*. See Peirce, "Of Reasoning in General," 11–26. His later work would further divide semiotic signs into ten categories. See Peirce, "Nomenclature and Divisions of Triadic Relations," 289–299.

10. Barthes, *Image—Music—Text*.

11. In Peirce's trichotomy, this kind of sign is called an *icon*. The concept is so closely tied to visuality that Peirce occasionally uses the alternative name *image* in place of *icon*. Peirce, "Of Reasoning in General," 13, 19.

12. Throughout, we use the term *photograph* to broadly include manual and digital still photography, film, video, and any other methods for recording a measurement of light to replicate the human visual system.

13. Scott, *The Spoken Image*.

14. Barthes, *Image—Music—Text*, 17.

15. Other semioticians—including Umberto Eco, Jean Baudrillard, and Group µ—have extended the ideas of Pierce, Saussure, and Barthes presented here. The continued scholarship on visual semiotics offers important new ways that visual materials make meaning; we encourage reading and engaging with this literature. However, these arguments largely consist of nuances and expansions that do not affect our core points here regarding the lack of an explicit linguistic indexical system contained in visual media. For further reading, see Groupe µ, *Rhétorique*

générale; Groupe μ, *Rhétorique de la poésie*; Groupe μ, *Traité du signe visual: pour une rhétorique de l'image*; Eco, "Sémiologie des messages visuels," 11–51; Baudrillard, *Simulacra and Simulation*.

16. It is possible to have an image that itself depicts another set of annotations, such as the scanned image of a textual document. The analyses in this book generally focus on image collections that do not consist of such self-contained annotations.

17. Scott, *Spoken Image*, 20.

18. "Regarder, c'est-à-dire oublier le nom des choses que l'on voit." Translation by the authors. Paul Valéry, *Degas, Danse, Dessin*, 178.

19. Barthes, *Image—Music—Text*, 19; Barthes, *Camera Lucida.*

20. Hall, "Encoding/Decoding," 128–138.

21. Hall, "The Whites of Their Eyes: Racist Ideologies and the Media."

22. Bonilla-Silva, "The Invisible Weight of Whiteness," 173–194; St. Clair, "Social Scripts," 171–183; Taylor, *The Archive and the Repertoire*.

23. *Oxford English Dictionary*, s.v. "pixel."

24. The first known use of the term comes from a 1965 edition of the *Proceedings Society Photo-optical Instrumentation Engineers*. *Oxford English Dictionary*, s.v. "pixel."

25. There is a complex relationship between the color perceived by a digital sensor, the color stored in a digital file, and the color displayed on a digital screen. These differences are not crucial for our analysis; we simplify the discussion by focusing on the way the image is stored. For more information, see Poynton, *Digital Video and HD*.

26. Lyon, "A Brief History of 'Pixel,'" 2–16.

27. For example, in UTF-8 encoding, the sequence (01110100) indicates the lowercase letter "t," and the sequence (11001111 10000000) indicates a lowercase Greek letter "π" (pi).

28. A study by NASA showed that the JPEG algorithm was able to store an image in a file that was ten-times smaller than the original pixel intensities with no detectable difference in the end product. See Haines and Chuang, "The Effects of Video Compression."

29. Saussure, *Cours de Linguistique Générale*, 23. Some preprocessing must first be applied, as the text typically needs to be split apart into words by the process known as *tokenization*. For written text, however, this process can usually be accomplished unambiguously through a simple deterministic algorithm. Languages that make use of a syllabary, such as Japanese and Cherokee, require additional work.

30. The call to create structured annotations to describe the messages conveyed by digital images may seem reminiscent of the theory and practice of structuralism. While we are advocating for the need to approximate visual messages using a structured system, our application is driven by the limitations of existing computational techniques. One of the defining features of our methodology is the post-structural critique that these structured annotations are at best an incomplete and imperfect approximation of the visual messages.

31. While there is no single agreed-upon definition of the term *content analysis*, nearly all frame it in the language of the inferential analyses and the scientific method. See, for example, Neuendorf, *Content Analysis Guidebook*; Mayring, *Qualitative Content Analysis;* Riffe, Lacy, Watson, and Fico, *Analyzing Media Messages*.

32. While the terms *annotations* and *annotation* are also often used for labels created directly by human input, through the remainder of the text we will use these terms to describe the automatically produced annotations created by computer vision algorithms. When referring to labels that are not produced algorithmically, we will explicitly mention *manually* constructed annotations.

33. Mitchell and Gillies, "Model-Based Computer Vision System," 231–243.

34. Jin et al., "Methodology for Potato Defects Detection," 346–351.

35. Zhang, Li, and Zhang, "Boosting Image Orientation Detection," 95–99.

36. Ahmed, *The Cultural Politics of Emotion*.

37. Dhall et al., "Emotion Recognition," 509–516.

38. Berger, *Ways of Seeing*; Sturken and Cartwright, *Practices of Looking*.

39. Hall, "Encoding/Decoding."

40. Barthes, *Image—Music—Text*; Hill and Helmers, *Defining Visual Rhetorics*; Kress and van Leeuwen, *Reading Images*.

41. Shannon, "A Mathematical Theory of Communication," 379–423.

42. As further evidence of the duality between textual and visual messages, the guiding example throughout Shannon's article is the transmittance of a short passage of text.

43. Porter, *Trust in Numbers*; Wernimont, *Numbered Lives*.

44. There is extensive scholarship in visual culture studies on the topic. Works include Berger, *Ways of Seeing*, Mirzoeff, *The Visual Culture Reader*; Sturken and Cartwright, *Practices of Looking*.

45. Hall, "The Whites of Their Eyes," 89–93.

46. Sturken and Cartwright, *Practices of Looking*.

47. Berger, *Ways of Seeing*, 10.

48. Hall, "The Whites of Their Eyes"; Nakamura, *Digitizing Race*. See also Chun, *Updating to Remain the Same;* Joanna Drucker, *Graphesis*.

49. Smith, "Visual Culture Studies," 1–16; Rogoff, "Studying Visual Culture," 24–36.

50. Zylinksa, *Nonhuman Photography*.

51. Marblestone, Wayne, and Kording, "Deep Learning and Neuroscience."

52. Research in this area is quickly expanding. For example, see Rettberg et al., "Mapping Cultural Representations of Machine Vision," 97–101.

53. For examples, see Ruth Kim, "Top 4 Beauty Apps,"; Smith, *Photography on the Color Line*; Mulvey, *Visual and Other Pleasures*, 14–26.

54. Selbst et al., "Fairness and Abstraction," 59–68; Noble, *Algorithms of Oppression*; Eubanks, *Automating Inequality*; O'Neil, *Weapons of Math Destruction*; Broussard, *Artificial Unintelligence*.

55. Noble, *Algorithms of Oppression*.

56. O'Neil, *Weapons of Math Destruction*.

57. Anderson, *Technologies of Vision*.

Chapter 2

1. Donoho, "50 Years of Data Science," 745–766.

2. Wickham and Grolemund, *R for Data Science,* ix.

3. Wickham, "Tidy Data," 1–23.

4. While our conceptual pipeline begins with the digitized visual material, we echo Lisa Gitelman's caution that this starting point is not outside of our analysis; rather, "the collection and management of data may be said to presuppose interpretation." Gitelman, *"Raw Data" Is an Oxymoron*, 3.

5. The inclusion of an explicit step showing how the annotations are created through computer vision systems draws on a "feminist strategy for considering context" by acknowledging the "cooking process that produces 'raw' data." D'Ignazio and Klein, *Data Feminism*, 160.

6. Arnold and Tilton, "Enriching Historic Photograph with Structured Data."

7. Brightness and saturation are given here on a scale from 0 to 1. We provide a detailed description and application of these measurements of color in chapter 3.

8. Trachtenberg, "From Image to Story," 43–73.

9. On medical imaging, see Hecht-Neilsen, "Neural Networks," 449–460; on medical imaging, see Ramos et al., "Detecting Unexpected Obstacles," 1025–1032, and Bar et al., "Chest Pathology," 294–297; on militarism, see Zuboff, "Big Other," 75–89; on surveillance capitalism, see Zuboff, *The Age of Surveillance Capitalism*.

10. Ramos et al., "Detecting Unexpected Obstacles," 1025–1032.

11. Redmon et al., "You Only Look Once," 779–788.

12. King, "dlib-ML: A Machine Learning Toolkit," 1755–1758.

13. "Panoptic" is short for *panopticon*, a surveillance system designed by English philosopher Jeremy Bentham. Bentham, *The Panopticon Papers*. It also references the Greek word *panoptes*, "all seeing," which is derived from the many-eyed Greek giant Argus Panoptes. The names of the algorithms further indicate entanglement with practices of looking through computer vision.

14. *Oxford English Dictionary*, s.v. "metadata."

15. Hirsch, *Family Frames*, 74.

16. McAuley et al., "Image-Based Recommendations," 43–52.

17. For example, see Arnold and Tilton, "Distant Viewing: Analyzing Large Visual Corpora."

18. For example, see Arnold et al., "Uncovering Latent Metadata."

19. Ni, Li, and McAuley, "Justifying Recommendations Using Distantly-Labeled Reviews and Fine-Grained Aspects," 188–197.

20. Mayer, "Linked Open Data," 2–14.

21. Dijkshoorn et al., "The Rijksmuseum Collection as Linked Data," 221–230.

22. Fear, "User Understanding of Metadata" 26–60.

23. Riley and Shepherd, "A Brave New World," 91–112.

24. Dorman, "Technically Speaking," 116–17.

25. Library of Congress Names, accessed April 15, 2020, http://id.loc.gov/authorities/names.html.

26. Wickham, "Tidy Data."

27. Arnold, Leonard, and Tilton, "Knowledge Creation through Recommender Systems."

28. Donoho, "50 Years of Data Science," 755.

29. Tukey, *Exploratory Data Analysis*.

30. Arnold and Tilton, "New Data?," 293–299.

31. Arnold et al., *Photogrammar*.

32. Hanson, *Mass Communication*.

33. See, for example, Beebe et al., *Interpersonal Communication*; Croteau and Hoynes, *Media/Society*; Martin and Nakayama, *Intercultural Communication*.

34. Whitelaw, "Generous Interfaces for Digital Cultural Collections."

35. Gibbs and Owens, "Building Better Digital Humanities Tools."

36. Wickham and Grolemund, *R for Data Science*, 423.

37. boyd and Crawford, "Critical Questions for Big Data," 662–679.

38. D'Ignazio and Klein, *Data Feminism*, 110.

39. Benjamin, *Race after Technology*.

40. D'Ignazio and Klein, *Data Feminism*.

41. There are also plenty of images that would be traumatizing for people to view and are better analyzed using a computational process. Newton, "The Trauma Floor."

Chapter 3

1. Sargent, *Picture Theatre Advertising*, 59.

2. Smith, *Selling the Movie*, 68.

3. Street and Yumibe, *Chromatic Modernity*, 123.

4. "Six Secrets of Movie Posters," BBC Bitesize, accessed November 1, 2021, https://www.bbc.co.uk/bitesize/articles/zr9s6v4/.

5. Verdesoto, "Movie Poster Color Schemes Explained."

6. Qiang, "A Study on the Metaphor of Red in Chinese Culture," 100–105.

7. Flueckiger, "A Digital Humanities Approach to Film Colors," 71–94.

8. Heftberger, *Digital Humanities and Film Studies*.

9. Vane, "Making 'Dive into Color.'"

10. Palmer, *Vision Science*, 29.

11. Palmer, 338.

12. Palmer, 254.

13. Polarity, the direction(s) that the light oscillates, is an additional property of light waves, but it does not have a significant effect on human perception. However, it is a critical aspect of the visual systems of certain insects, cephalopods, and other animals, where it aids in navigation and communication.

14. Palmer, *Vision Science*, 112.

15. Palmer, 102.

16. Palmer, 113.

17. Palmer, 137.

18. Palmer, 115.

19. Newhall, *The History of Photography*.

20. Hirsch, *Seizing the Light*, 180.

21. Hirsch, 28.

22. Palmer, *Vision Science*, 181.

23. The early speculation on this fact is credited to the trichromatic theory of Thomas Young and Hermann von Helmholtz. The core idea of their theory has turned out to be remarkably accurate, as medical research subsequently discovered the physiological processes behind their observations. Palmer, 180.

24. Palmer, 185–187.

25. Palmer, 187.

26. Palmer, 96.

27. Palmer, 99.

28. Palmer, 97.

29. There are many other color models that could also be argued for. Options such as CIELAB give a more accurate spacing of colors relative to the sensitivity of the human visual system. However, they are more challenging to perform the kinds of aggregative analysis that follow. Likewise, more popular options such as using saturation in place of chroma and value in place of intensity are slightly more difficult to describe mathematically and offer no benefit in the context of our analysis. See Levkowitz and Herman, "Generalized Lightness, Hue, and Saturation," 271–285.

30. Palmer, *Vision Science*, 97.

31. Palmer, 98.

32. While *chroma* is not commonly used in colloquial English and may appear to be a more technical term than *saturation*, it is easier to both define and visualize. For this reason, many online and popular press sources show color space diagrams using the chroma value but mistakenly label it with the term *saturation*.

33. Palmer, *Vision Science*, 98.

34. Rhodes, "Origin and Development," 228–246.

35. Fuller-Seeley, "Storefront Theatre Advertising," 398–419.

36. There are several extensive histories of the Motion Picture Association of America and early cinema. See, for example, Doherty, *Pre-Code Hollywood*; Jeff and Simmons, *Dame in the Kimono;* Browser, *Transformation of Cinema*.

37. Fuller-Seeley, "Storefront Theatre Advertising," 406.

38. *Moving Picture World: The Film Exhibitor's Guide* (1916), accessed November 1, 2021, https://archive.org/stream/mowor29chal#page/n421/mode/1up/.

39. Fuller-Seeley, "Storefront Theatre Advertising," 417.

40. Staiger, "Announcing Wares, Winning Patrons, Voicing Ideals," 3–31.

41. Cohen, *A Consumers' Republic*, 281.

42. Fuller-Seeley, "Storefront Theatre Advertising," 400.

43. Sargent, *Picture Theatre Advertising*, 68, 250.

44. Franklin, *Motion Picture Theater Management*, 258.

45. Sargent, *Picture Theatre Advertising*, 64; Fuller-Seeley, "Storefront Theatre Advertising," 412.

46. Fuller-Seeley, "Storefront Theatre Advertising," 264.

47. Smith, *Selling the Movie*, 17.

48. Rhodes, "Origin and Development," 232.

49. Rhodes, "Origin and Development," 242.

50. "21 Poster Exchanges Added to Centralized Accessories System," *Motion Picture Herald*, February 17, 1940, 17, accessed November 1, 2021, http://lantern.mediahist.org/catalog/motionpictureher138unse_0635/.

51. Lawlor v. National Screen Service Corp., 349 US 322 (1955).

52. For more on the studio system, see Schatz, *The Genius of the System*. For information on the impact of television, see Anderson, *Hollywood TV*.

53. Poole and Poole, *Collecting Movie*, 110.

54. Bogart, *Artists, Advertising, and the Borders of Art*, 81–89.

55. Karal Ann Marlin, quoted in the film *The Man Who Drew Bug-Eyed Monsters* (1994), dir. Mel Bucklin.

56. Fuller, "The Arts Desk."

57. Bordwell, "It's the 80s, Stupid."

58. Handel, “Hollywood Market Research,” 304–310.

59. Street and Yumibe, *Chromatic Modernity*, 27.

60. Handel, “Hollywood Market Research,” 308.

61. Films are associated with the year in which they were first released, but box office sales were computed over all time, as of February 2020.

62. For more details about the role of exploratory data analysis, see the discussion in chapter 2.

63. Fisk, *Television Culture*.

64. As a thought experiment, we can take the percentages of white and black in each half-decade and assume that all posters are three-tone: white, black, and a shade of gray with an intensity of 50. This would yield an average intensity of 60.3 for the half-decade between 1970 and 1974 and an average intensity of 34.2 for the years 1995 to 1999. The actual median intensities are very close to this, with values of 60.9 and 34.0, respectively. This further provides evidence that our chosen cutoff values offer reasonable summary statistics for the data.

65. Rhodes, “Origin and Development of the American Moving Picture Poster,” 230.

66. We could also account for this difference by using more complex multivariate regression models. However, these require more explanation and are more prone to detect spurious correlations that arise from the multitude of different ways of specifying a multivariate model. We will attempt to use the most easy-to-understand and straightforward approaches to addressing each task throughout this text.

67. Verdesoto, “Movie Poster Color Schemes Explained.”

68. There are other relatively rare tags that exist on IMBd but are not present in our data. For example, “Short” appears as a genre for *Serta Perfect Sleeper* (1972), one of the earlier films in our larger fifty-year dataset.

69. Throughout this chapter, we will use capitalized terms to refer to the genre categories in the dataset. Lowercase forms are used to refer to a genre in a general or abstract sense that is not necessarily tied to this specific corpus.

70. Before looking at the results, our intuition was that pairs such as Biography and History, or Thriller and Horror, might have a very high degree of overlap and thus require removing one or the other from the analysis. However, on closer examination, each of these genre tags has a specific meaning that is not mutually synonymous with the other. This is an illustration of the utility that computational analysis provides to challenge seemingly obvious assumptions about a corpus. Even when the analysis confirms our assumption, it is still a valuable exercise to rule out cases such as the example here.

71. As an alternative, we could study the ways in which the genre tags intersect with one another by investigating the full set of many-to-many mappings between the films and tags. This approach would require the use of more involved multivariate methods.

72. Auster and Mansbach, "The Gender Marketing of Toys," 375–388.

73. Deleyto, *The Secret Life of Romantic Comedy*, 24.

74. Cooper, "Escapist Adventure, Timely Irreverence," 25–26.

75. There are techniques for taking means and distances in meaningful ways for circular values such as hue; these are common when working with data that includes observations at different times of the day. These can be useful for a clustering analysis of hue, but do not solve the general problem of combining hues of very different values across a corpus.

76. Roberson et al., "Color Categories," 378–411.

77. Berlin and Kay's classic study of this phenomenon proposes a hierarchical view of how languages evolve terms for color types. See Berlin and Kay, *Basic Color Terms*. More recent work has challenged this hypothesis through the use of additional field methods and experimental research. See Wierzbicka, "The Semantics of Colour: A New Paradigm," 1–24.

78. Winawer et al., "Russian Blues," 7780–7785.

79. Pemberton et al., "CSS Color Module Level 3."

80. Eden, Grizzard, and Lewis, "Disposition Development in Drama," 33–47.

81. Smuts, "The Desire-Frustration Theory of Suspense," 281–290.

82. While outside the scope of this chapter, we could use another set of more advanced computer vision algorithms to detect people, which would be a potential avenue for a future study.

83. Yoon, "Color Symbolisms of Diseases," 1–4.

84. Byrnes, "Color Associations of Children," 247–250.

85. It may seem that the color percentages of each genre in table 3.3 should match the percentages in table 3.5. They have a similar pattern but are measured in different ways. Table 3.3 uses medians to be consistent with the other metrics on the same table; in contrast, table 3.5 uses averages to be consistent with its other measurements. Also, table 3.5 does not include posters that have less than 3 percent of their area covered with colorful pixels. This is the reason that the total count columns are slightly lower in this table.

Chapter 4

1. For a classic overview of the project, see Fleischhauer and Brannan, eds., *Documenting America*.

2. Howells and Negreiros, *Visual Culture*.

3. Sabatelli et al., "Advances in Digital Music Iconography."

4. Wevers and Smits, "The Visual Digital Turn," 194–207.

5. Bermeitinger et al., "Deep Watching."

6. The photo was taken in Cimarron County, Oklahoma, in April 1936.

7. Neural networks, which are currently the most popular approach to computer vision research, were largely dismissed and ignored for over fifty years based on the initial difficulty in getting them to "beat" other techniques.

8. Elgendy, *Deep Learning for Vision Systems*, 263.

9. Elgendy, 264.

10. Elgendy, 310.

11. Elgendy, 266.

12. Finnegan, "What Is This a Picture of?," 116–123.

13. Caesar, Uijlings, and Ferrari, "COCO-stuff," 1209–1218.

14. The original paper uses the terminology of *stuff* and *things*, but we will continue to refer to the latter category as *objects* to coincide with the terminology of other methods.

15. The indoor stuff categories consist of all human-made materials that are not roads or railroads, including regions such as windows and walls, which can of course be photographed from the outside just as easily as they are from the inside.

16. We were, in fact, very skeptical that the stuff algorithm would work well on black-and-white photography, as the algorithm could easily have learned to identify some categories, such as sky, through the detection of color. The positive results were an exciting surprise.

17. These are what are known as false positive rates and are based on whether the tags were reasonable given the image. Not all categories are easy to distinguish properly; for example, what is the exact cutoff in rock size between a dirt road and a gravel road? For a similar reason, it is tough to accurately get false negative rates, which we did not attempt to do outside of the detection of people.

18. For examples of work on the evolution of state power during the New Deal, see Cebul, Geismer, and Williams, eds., *Shaped by the State*; Fraser and Gerstle, eds., *The Rise and Fall of the New Deal*.

19. Finnegan, *Picturing Poverty*.

20. Finnegan, *Picturing Poverty*; Fleischhauer and Brannan, *Documenting America*; Tagg, *Disciplinary Frame*.

21. Stryker and Wood, *In This Proud Land*, 14.

22. Fleischhauer and Brannan, *Documenting America*, 331.

23. Arnold et al., "Uncovering Latent Metadata."

24. Fleischhauer and Brannan, *Documenting America*, 335.

25. Wexler, private communication (2010).

26. The API can be found here: https://libraryofcongress.github.io/data-exploration/.

27. Alpers, *Walker Evans*.

28. Gordon, *Dorothea*.

29. Fleischhauer and Brannan, *Documenting America*.

30. Allred, *American Modernism*, 5.

31. Strange, *Symbols of Ideal Life*.

32. The first level contains only thirteen categories and the third is far too large, containing 1,153 unique values. Using the middle set of categories allows us to give a wide range of topics while still fitting all the results on a single table.

33. To explore the photos in each Vanderbilt system category, visit https://photogrammar.org.

34. We might even expect larger proportion for some categories, such as the 99 percent and greater percentages of outdoor photographs corresponding with *The Land* categories. Manually looking at a sample of one hundred images from the *Children* category reveals that some of these misclassifications are the result of an algorithmic error—a person was present but not detected—but just as many are the result of images that seem to be given a category either in error or because no other better tag existed.

35. In our experiments with the FSA-OWI data, "building" is most usually used to tag a building framed in an outdoor scene. A building from the inside is generally classified with more specific categories, such as a ceiling or wall.

36. Hall, "Encoding/Decoding," 128–138.

37. For example, Russell Lee and Walker Evans photographed small towns and communities. They often focus on one building or a few buildings, taking the photo as if the viewer were looking straight at the building. Evans's compositional approach would become known as "straight documentary" for its seemingly impartial point of view, due to the lack of angles, shadows, and related compositional features. Curtis and Grannen, "Let Us Now Appraise Famous Photographs," 1–23.

38. Berman and Cronin, "Project DOCUMERICA," 186–197.

39. Caesar et al., "COCO-stuff," 1209–1218.

40. Our previous analysis indicated that there was a difference in the proportion of images with people among the photographers, which may at first seem to contradict the observation here. The apparent contradiction is explained by the difference in what each metric is measuring. Before, we looked at the specific probabilities by photographer; here, we measure how spread out the photographs with people are over the photographers. The distribution is changing (it focuses more on OWI photographers and less on FSA photographers), but the overall spread is still large.

41. The images were in the form of a cartoon painting on the side of a circus wagon in Alger, Montana. This can be found by slightly decreasing the cutoff confidence score.

42. For a powerful example of how a small set of photographs can open histories, see Grace Elizabeth Hale's study using a set of FSA-OWI photographs to argue how the rise of consumer culture challenged existing race relations in the American South. Hale, *Making Whiteness*.

43. Buolamwini and Gebru, "Gender Shades," 77–91.

44. Trachtenberg, "From Image to Story," 50.

45. Finnegan, "What Is This a Picture of?," 116.

46. Trachtenberg, "From Image to Story," 45.

47. Gordon, *Dorothea Lange*, 264.

Chapter 5

1. Cameron and Jeffery, "The Universal Hitchcock," 271.

2. Ott and Keeling, "Cinema and Choric Connection," 370.

3. Sharits, "Red, Blue, Godard," 24–29.

4. Dyer, "Lighting for Whiteness," 282; DuVernay, "Lighting the Black Body."

5. Everett, "The Other Pleasures," 26–38.

6. Some recent television series, such as HBO's *Game of Thrones*, have been produced and funded more in line with feature-length films than a typical television series. However, we focus on the far more "typical" television series that dominate both synchronous and asynchronous television. See Mittell, *Complex TV*.

7. This relationship seems to be described well by Ernie Kovacs's (likely misattributed) joke that "television is a medium. So-called because it is neither rare nor well done."

8. An earlier version of the analysis in this chapter appeared in the *Journal of Cultural Analytics*; see Arnold, Tilton, and Berke, “Visual Style in Two Network Era Sitcoms.” It was co-written with our collaborator, film and TV scholar Dr. Annie Berke, who has graciously supported our extension of the analysis here; see Berke, *Their Own Best Creations*.

9. For a sense of scale, consider the collected works of William Shakespeare. The digitized text of all his plays, sonnets, and other materials can be saved in a file taking up under 7 megabytes of storage space. In comparison, the standard size of just a single photo on a modern smartphone is currently between 5 and 12 megabytes. Even heavily compressed, a half-hour video file digitized in standard definition requires over 1000 megabytes of storage.

10. Salt, “Statistical Style Analysis of Motion Pictures,” 13.

11. Tsivian and Civjans, *Cinemetrics*; Acland and Hoyt, *The Arclight Guidebook*; Butler, “Statistical Analysis of Television Style,” 25–44.

12. Burghardt, Kao, and Wolff, “Beyond Shot Lengths”; Ferguson, “Digital Surrealism.” Burges, Dimmock, and Romphf, “Collective Reading.”

13. These are standard playback frame rates. Slow-motion or time-lapse effects are created by recording at a different frame rate than the playback frame rate. Julie Turnock provides an excellent summary of the history and theory behind frame rates in her analysis of *The Hobbit*, which was known for being played at the unusually high 48 frames per second. Turnock, “Removing the Pane of Glass,” 30–59.

14. While not used in the analysis shown in this chapter, the distant viewing toolkit includes algorithms for the analysis of both sound and subtitle information.

15. Butler, *Television,* 30.

16. Miller, “Anal Rope,” 119–172.

17. Due to the centrality of this task in video processing, many competing terms are used in the literature, including *shot transition detection*, *shot detection*, and *cut detection*.

18. We became aware of a very accurate, pretrained algorithm for shot boundary detection that was published well after we completed the work presented in this chapter. We would recommend this algorithm for others looking to perform a similar analysis on another dataset. Souček and Lokoč, “TransNet V2.”

19. There are a few other minor technical details in the algorithm, such as how to avoid making shots that are too small. These are described in the implementation included in the supplementary materials. Also, note that the introductory cartoon sequences to both series (shown with the titles) were excluded in this and subsequent analyses.

20. See, respectively, Papageorgiou, Oren, and Poggio, "A General Framework for Object Detection," 555–562; Viola and Jones, "Rapid Object," 1–8; Dalal and Triggs, "Histograms of Oriented Gradients for Human Detection," 886–893.

21. For example, a trainable HOG detector algorithm is provided by the popular OpenCV library, dlib library, and the scikit-image Python package.

22. Sun, Wu, and Hoi, "Face Detection Using Deep Learning," 42–50; Sharma, Shanmugasundaram, and Ramasamy, "CNN Based Efficient Face Recognition Technique Using Dlib,"192–195.

23. We applied the Faster R-CNN to every ten frames in six episodes of each of our two series and labeled where the algorithm had mistakenly identified objects as positive faces or had failed to detect faces when present. For the latter, we only considered a face as being missed if at least one eye of a face was present in the frame, and we excluded any extraneous actors (extras) in the far background and removed from the main action. Compared to the hand-labeled faces, the algorithm performed well, with a *precision* (proportion of detected faces that were faces) of over 98.3 percent and a *recall* (proportion of faces that were correctly identified) of 95.1 percent. For comparison, the HOG detector had an overall recall of only 55.2 percent.

24. Baltrusaitis, Zadeh, Lim, and Morency, "OpenFace 2.0," 59–66; Klontz et al., "Open-Source Biometric Recognition," 42–50.

25. Cao et al., "VGGFace2,"67–74.

26. The algorithm can be applied to faces that are not in the training set by supplying a single *reference* image for every character of interest. The algorithm returns which faces in the dataset appear to be the same as one of the reference images.

27. Defining precision as the proportion of assigned faces that were correctly identified and recalled as the proportion of main characters correctly identified, our choice of cutoff values yields an overall precision of 92.8 percent and recall of 87.0 percent. Similarly, a precision of 99.0 percent can be achieved while maintaining a recall of only 67 percent. Kaiming He et al., "Deep Residual Learning," 770–778.

28. We are aware of only one prior attempt to offer an algorithmic taxonomy of film shots. The presented taxonomy mainly related to camera movement in action films and was not very relevant to the features we are interested in classifying. See Wang and Cheong, "Taxonomy of Directing Semantics," 1529–1542.

29. Close shots were identified with a precision of 99.0 percent and recall of 93.5 percent (F1 Score: 0.962); group shots were classified with a precision of 98.25 percent and recall of 95.73 percent (F1 Score: 0.969); and over-the-shoulder shots had a precision of 88.24 percent and recall of 95.74 percent (F1 Score: 0.918). Much of our analysis focuses on the timing and presence of close shots. The precision is exceptionally high because of the conservatively chosen logic in our algorithm.

30. Edgerton, *The Columbia History of American Television*,178.

31. Edgerton, *255*.

32. Lotz, *The Television Will Be Revolutionized*, 51.

33. Eco, "The Myth of Superman," 17.

34. Lotz, *The Television Will Be Revolutionized*, 101.

35. Metz, *Bewitched*, 14–16.

36. Metz, 17.

37. Stoddard, "Bewitched and Bewildered," 50.

38. Spigel, *Welcome to the Dreamhouse*, 132.

39. Fairfield-Artman, Lippard, and Sansom, 27.

40. As further evidence of the uncertainty about the centrality of characters in these series, translations of the title into other languages often reverse the perspective of the original. For example, in Italian, *Bewitched* is called *Vita da Strega* (The life of a witch). The original working title of the script, in fact, was *The Witch of Westwood*; see IMDb, https://www.imdb.com/title/tt0057733/releaseinfo.

41. The character of Darrin on *Bewitched* was played by two actors: Dick York (seasons 1–5) and Dick Sargent (season 6–8). We created separate face detection algorithms for both actors. However, in this and all other results, we have combined them to their common character.

42. This four-part structure roughly follows the chart provided by Jeremy Butler and very accurately matches the structure of our two series. We were able to automatically extract the narrative parts through the chapter breaks encoding in our DVD materials. Butler, *Television*.

43. A common trope on the show involves fast cuts between Darrin and Samantha as they talk over the telephone between the office and home.

44. Sherif, Taub, and Hovland, "Assimilation and Contrast Effects," 150.

45. Only 1.04 percent of the dataset started with a shot containing multiple faces; in these cases, none of the characters was counted.

46. It is unclear from this analysis whether her centrality is due to the magical abilities of Samantha Stevens or the star power of Elizabeth Montgomery. Likely both contribute in some way.

47. Salt, *Moving into Pictures*.

48. Regression analysis can be used to detect how strongly various factors influence median shot length. Predicting shot length first as a function of the series and then

as a function of shot length reveals that the latter explains two orders of magnitude more variation. Series has an R2-value of only 0.000055, whereas a regression model using shot type provided an R2-value of 0.0534. Following prior work on shot lengths, we assumed that shot length followed a log-normal distribution, and ran each regression of the logarithm of each shot length. Cutting, DeLong, and Nothelfer, "Attention and the Evolution," 432–439.

49. Stiller, Nettle, and Dunbar, "The Small World of Shakespeare's Plays," 397–408; Xanthos et al., "Visualizing the Dynamics of Character Networks," 417–419.

50. Blatt, "Which Friends on *Friends* Were the Closest Friends?"

51. Mulvey, *Visual and Other Pleasures*.

52. Bordwell, Staiger, and Thompson, *Classical Hollywood Cinema*, 13.

53. Butler, *Television*, 369.

54. Bordwell, Staiger, and Thompson, *Classical Hollywood Cinema*, 374.

Chapter 6

1. Brennan, "Public, First," 384–390.

2. Digital Collections, Library of Congress, accessed November 4, 2021, https://www.loc.gov/collections/.

3. Rijks Data, Rijksmuseum, accessed November 4, 2021, https://data.rijksmuseum.nl/object-metadata/api/.

4. Baca, *Introduction to Metadata*.

5. Metropolitan Museum of Art Collection API, updated November 17, 2020, https://metmuseum.github.io/.

6. Smithsonian Open Access, accessed July 17, 2022, https://www.si.edu/openaccess.

7. Louvre Collections database, accessed November 4, 2021, https://collections.louvre.fr/.

8. "The Santa Barbara Statement on Collections as Data, Version 2" (2020), accessed November 4, 2021, https://collectionsasdata.github.io/statement/.

9. Padilla, *Responsible Operations*.

10. Cordell, *Machine Learning + Libraries*.

11. "Digital Strategy for the Library of Congress," accessed November 2, 2021, https://www.loc.gov/digital-strategy/.

12. Smithsonian 2022 Strategic Plan, accessed November 2, 2021, https://www.si.edu/strategicplan.

13. Elgendy, *Deep Learning for Vision Systems,* 6.

14. Elgendy, 406.

15. Elgendy, 402.

16. The terms *deep learning* and *neural networks* are often used interchangeably. Originally, deep learning referred to a general approach to building models through a series of iterative transformations, whereas a neural network was a specific example of a deep learning algorithm. However, neural networks and their variants have remained the only commonly used type of deep learning over time. In this section, we use the term *deep learning* to highlight our focus on the concepts behind the approach rather than the specific details of individual model implementations.

17. There are many other ways of building a classification based on two features; this is just one example. A slightly more complex and general-purpose approach would be to use a logistic regression, which learns to weight the two features together as a model of the probability that the image was taken outdoors.

18. Arnold and Tilton, "Depth in Deep Learning," 309–328.

19. Arnold and Tilton, 311.

20. Montúfar et al., "On the Number of Linear Regions," 1–14.

21. Elgendy, *Deep Learning for Vision Systems*, 213.

22. Elgendy, 253.

23. Bansal et al., "Transfer Learning for Image Classification," 1.

24. Elgendy, *Deep Learning for Vision Systems*, 240.

25. Unlike manually constructed features, it is not possible to assign any meaning to the individual features that are automatically generated. This occurs for two reasons. First, since the final features are themselves generated by many interconnected layers, their relationship to the original image becomes highly obfuscated. Secondly, the model is not incentivized to isolate elements in the image in individual features. For example, rather than a single feature indicating whether a photograph was taken outdoors, this aspect of the image is likely to be captured by a more complex combination of the features.

26. Arnold and Tilton, "Depth in Deep Learning," 317.

27. McInnes, Healy, Saul, and Großberger, "UMAP: Uniform Manifold Approximation and Projection," 861.

28. The formal algorithm is slightly more complicated because it is usually not possible to replicate the neighborhood structure exactly. Instead, an approximation is made that measures how closely the two-dimensional neighbors match the neighbors in the original space. Conceptually, the idea is the same as described in the text.

29. Szegedy et al., “Rethinking the Inception Architecture,” 2818–2826.

30. “About the Met,” Metropolitan Museum of Art, accessed July 17, 2022, https://www.metmuseum.org/about-the-met.

31. Charr, “Decolonize This Place Targets New York Museums,” *MuseumNext*, October 15, 2019, https://www.museumnext.com/article/decolonize-this-place-targets-new-york-museums; Larkin, “The Met Museum Misses the Mark.”

32. The number 375,000 is contradicted elsewhere in the press release as being only 200,000. The latter number more accurately matches the number of digital images in the collection as of 2021. It is unclear whether the first number is a typo, whether the legal policies around some records changed at some point, or whether slightly different things are being counted.

33. Loic Tallon, Chief Digital Officer, “Introducing Open Access at The Met,” February 7, 2017, https://www.metmuseum.org/blogs/digital-underground/2017/open-access-at-the-met.

34. As we have seen in our study of the FSA-OWI, access to physical, organizational information can open many possibilities of analysis that would not otherwise be possible.

35. See, for example, “Decolonizing Museum Collections: A Conversation between Colleagues in the Field (CSAAM),” American Alliance of Museums, September 29, 2021, accessed November 4, 2021, https://www.aam-us.org/2021/09/29/decolonizing-museum-collections-a-conversation-between-colleagues-in-the-field-csaam/.

36. Simon Wakeling et al., “Readers Who Borrowed This.”; Herlocker et al., “Evaluating Collaborative Filtering Recommender Systems,” 5–53.

37. See “Amulet of Baboon in Act of Adoration 525–30 B.C.,” The Metropolitan Museum of Art Collections, accessed September 30, 2022, https://www.metmuseum.org/art/collection/search/570613.

38. “Google Apologises for Photos App’s Racist Blunder,” *BBC News*, July 1, 2015, accessed November 4, 2021.

39. Patterns of color can be difficult to make out in the printed black-and-white image; they can be seen more clearly in the full color version available on the book’s website.

40. The name *Islamic Art* might suggest a cultural or religious logic, but according to the metadata in the Met’s collection, it is better understood as representing the

geographical regions of Northern Africa and the Near East. For example, the twelfth-century BCE stucco fragments from modern-day Iran are housed in the Islamic Art department, despite being made nearly 1,700 years before the start of Islam.

41. Whitelaw, "Generous Interfaces."

42. For readers looking to reproduce the analysis on a similar dataset, the code to produce our interface is included in the text's supplemental material available at https://distantviewing.org/book.

43. Buley, *The User Experience Team of One*; MacKenzie, *Human-Computer Interaction*; Cox and Tilton, "The Digital Public Humanities," 127–146; Smulyan, *Doing Public Humanities*.

44. Caddick and Cable, *Communicating the User Experience*, 28; Warner, "Publics and Counterpublics"; Black, *Transparent Designs*; Emerson, *Reading Writing Interfaces*; Wardrip-Fruin, *Expressive Processing*.

45. The design and code for the interface was adapted from our project *Access and Discovery of Documentary Images* (ADDI), a project designed to adapt and apply computer vision algorithms to aid in the discovery and use of digital collections, specifically documentary photography collections held by the Library of Congress. Funding and support were provided by the Library of Congress through the *Computing Cultural Heritage in the Cloud* project. See https://github.com/distant-viewing/addi.

46. For more on the lifecycle of a digital project and best practices for sustainability, see Visual Media Workshop at the University of Pittsburgh, *The Socio-Technical Sustainability Roadmap*, accessed April 26, 2022, http://sustainingdh.net.

47. Whitelaw, "Generous Interfaces," 3.

48. One of the most important tuning parameters for UMAP controls whether the algorithm tends to create one giant cluster with all the images or many unconnected clusters that split the images into discrete groups. We chose a value that created just a few independent groups because that seemed to be the most interesting on its own. However, there is a lot to be gained by looking at several different embeddings together.

Conclusion

1. Arnold and Tilton, "Distant Viewing Toolkit."

2. Laws regarding the distribution of annotations under the fair use doctrine are largely untested. For an overview, see Lemley and Casey, "Fair Learning," 743–785.

The open access edition of this book was made possible by generous funding from the National Endowment for the Humanities and the University of Richmond.

The MIT Press would like to thank the anonymous peer reviewers who provided comments on drafts of this book. The generous work of academic experts is essential for establishing the authority and quality of our publications. We acknowledge with gratitude the contributions of these otherwise uncredited readers.

This book was set in Stone Serif and Stone Sans by Westchester Publishing Services.

Library of Congress Cataloging-in-Publication Data

Names: Arnold, Taylor, author. | Tilton, Lauren, author.
Title: Distant viewing : computational exploration of digital images / Taylor Arnold and Lauren Tilton.
Description: Cambridge, Massachusetts : The MIT Press, [2023] | Includes bibliographical references and index.
Identifiers: LCCN 2022052202 (print) | LCCN 2022052203 (ebook) | ISBN 9780262546133 (paperback) | ISBN 9780262375177 (epub) |
ISBN 9780262375160 (pdf)
Subjects: LCSH: Computer vision. | Image data mining. | Image processing—Digital techniques. | Visual sociology—Technique.
Classification: LCC TA1634 .A76 2023 (print) | LCC TA1634 (ebook) | DDC 006.4/2—dc23/eng/20230111
LC record available at https://lccn.loc.gov/2022052202
LC ebook record available at https://lccn.loc.gov/2022052203