What Type of Data Are Images?

Nora Webb Williams

February 2023

Abstract

Images are an intriguing source of data for social scientists. Increasingly, the phrase "images as data" implies big data, digitized images, and quantitative analyses with elements of machine learning. This perspective, however, obscures the many other ways to conceive of what images are as data for social scientists. For example, big data, machine learning image researchers rely on labeled data to build classification algorithms. The process of labeling the data is essentially content analysis, a familiar tool to qualitative researchers. Thus, quantitative analyses are built on qualitative approaches and both are a potentially valid way to use images to better understand the world. This chapter reviews current trends in images-as-data research while considering how the complicated nature of meaning derived from images lends itself to a pluralistic research approach. I propose four guiding questions to use when embarking on or evaluating image research in political science.

1 Introduction

What type of data are images? In this piece, I argue that images are valuable data sources for both qualitative and quantitative approaches. I note that the content of and reaction to images are often highly subjective, which necessitates careful consideration of research procedures. In an attempt to expand the frame of what counts as "images as data" scholarship, I propose four guiding questions to use when embarking on or evaluating image research. The use of images as data is a frontier of political science research and the discipline will benefit from taking a broad perspective when evaluating how we do this work.

In most political science discourse today, the phrase "images as data" (or "image as data"), which appears in many recent publications (Joo et al., 2022; Loken, 2021; Schwemmer et al., 2020; Webb Williams et al., 2020), seems to imply "big data."[1] Big data sets lend themselves to the use of automated and machine learning tools, as processing so much data by hand would be tedious and expensive. Most big data analyses rely on quantitative tools to aggregate the massive amounts of information contained in the data.

Much of the substantive images-as-data work in the big data sphere wrestles with the proliferation of images and image-making/sharing technologies in the 21st century, such as digital cameras and other forms of digitization; the Internet; and social media (Anastasopoulos et al., 2016; Cantú, 2019; Casas et al., 2018; Joo et al., 2019; Peng, 2018; Won et al., 2017). These studies might consider the effects of these new technologies directly, such as work that considers how images shared on social media might mobilize social movements (Casas et al., 2018). Or they might leverage the technologies indirectly, for example by using scanned ballots or pictures of ballot counts to detect electoral fraud (Callen et al., 2015; Cantú, 2019). Images-as-data methods pieces highlight the potential of leveraging the deep learning revolution in computer science by applying convolutional neural nets and other

[1] The threshold for what constitutes big data is fluid but can be thought of as "big to the extent that new technologies had to be created by specialists to collect, store, and analyze it" (Lazer et al., 2017, p. 21).

computer vision frameworks to politically-relevant images (Peng, 2022; Torres, n.d.; Torres et al., 2021; Webb Williams et al., 2020). I provide additional thoughts on recent trends in images as data scholarship in section 6.

Yet images themselves are far from a new development in human history. Image analysis and critique are similarly established. This chapter describes the current state of the subfield (both substantive and methodological) in political science image research while attempting to place these analyses and their challenges within a larger, pluralistic framework for understanding visuals. Evoking the need for quantitative researchers to maintain a qualitative sensibility (Tanweer et al., 2021), my main point is that images are rarely perfectly objective sources of data. A picture may be worth a thousand words, but what those thousand words mean, and to whom, are subject to debate. In other words, measurement error in understanding meaning from images is relevant at all scales and for all methods of image analysis (and for the analysis of other "new" mediums as well, including video and audio). Sometimes that measurement error is correctable – for example, errors that occur due to low quality, blurry images could be improved with higher resolution images – but at other times the differences in how individuals see and react to images is a feature of the image analysis process, not a bug to be corrected.

2 The Meaning of Images

Political scientists care about images as a source of data because they convey meaning. But it is only sometimes the case that there is a universally acknowledged meaning or truth that can be derived from a picture. One can imagine a spectrum of meaning-agreement for images. At one end, the meaning of images might be more objective, or what we might consider purely informational (e.g. does the picture show a dog?). At the other end of the spectrum, the meaning might be more subjective or context dependent (e.g. is this picture scary?). At the more informational/objective end of the scale, for example, satellite pictures of the earth at night contain information about the extent of electrical grids and access to electric lighting (Min et al., 2013). Looking at these pictures, it is obvious where they are bright or not. Adjusting for cloud cover, in this case, helps to produce reliable and valid measurements from images. Other examples that are fairly object include daytime satellite images that show where road and roofing materials change (Jean et al., 2016; Philipp et al., n.d.; Varshney et al., 2015). Or drone or helicopter shots of an inauguration contain information about how many people attended the event. We could group these types of data under the general heading of "remote sensing." They are images taken from a remove, and yet it is relatively easy to confirm their accuracy by seeing whether the aerial imagery matches what is on the ground; hence the terms "ground-truth" or "ground-truthing" to mean verifying that what is in an image reflects the truth when examined at a much closer scale. For example, we might check the validity of the people-count images by having researchers on the ground counting the number of people in a set area. In a recent example of this type of validation, Livny (2021) develops a measure of religiosity by analyzing changes in nighttime lights brightness during Ramadan, which is then confirmed with survey and electoral data.

In these remote-sensing cases, images are a source of objective information for the quantity the researcher actually wants to measure (electrification; religiosity; attendance). This type of data from images is certainly of use to political science researchers and social scientists more broadly. Yet often our work edges in a less objective direction[2] that forces us to address context, interpretation, and subjectivity.

To give a concrete example, image analysis studies often involve extracting information about what is in a picture (this is generally called *object recognition* in computer vision litera-

[2] Although even the remote-sensing data are never completely perfect – satellites and maps lie (the images may be misaligned, for example, or map could have been drawn to suit a particular political purpose, or the satellites may not be sensitive enough to register very faint lights at night), as do respondents and our own eyes as we try to count protesters or otherwise validate data. As most quantitative researchers would likely agree, it is impossible to completely eliminate concerns about measurement error.

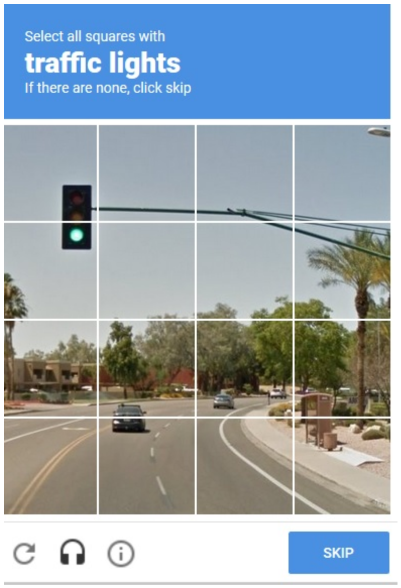
ture). Extracting the information usually starts with a process called "image annotation" or "image labeling," which means that humans look at the images and answer questions about what is pictured. The human-generated labels are also referred to as "manual labels," "gold standard labels" or "ground-truth labels" (with reference to the remote-sensing applications described above). Most Internet users are familiar with image annotation, though they may not be aware of it. Figure 1a, for example, shows a CAPTCHA[3] example where humans are required to perform a task to prove that they are not a computer – here the task is identifying traffic lights. This is an object recognition task, and not only are you proving your humanity, you are also providing labels that might later be used to train a computer vision algorithm in a self-driving car.[4] The CAPTCHA tests also help demonstrate the challenges of image interpretation. How would you label the image in Figure 1b? Is this a traffic light, or not? How you interpret the image (and the task) will vary greatly depending on what instructions you have received (Does the labeling form indicate if traffic lights on signs count, or not?), your own personal judgement (What is your gut sense on if it counts or not?), and the context (If you knew you were providing labels to train a self-driving car, would that change your response? What about if you were labeling when you were tired versus well-rested?).

Returning to the larger point, the idea that images convey information and meaning leads to the question of: "To whom is the meaning conveyed and what meaning do they perceive?" What people read into a image is heavily context dependent. If someone were faced with a traffic light CAPTCHA who had never seen a traffic light before, for example, they would not take the expected meaning from the picture. Or imagine an image that has bands of English-language text running along the top and the bottom (in other words, imagine a

[3]Which stands for: Completely Automated Public Turing test to tell Computers and Humans Apart

[4]Training a computer to recognize a traffic light from a static image is a fairly easy task, and because of this, these CAPTCHAs often do not a great job at sorting the humans from the machines Dzieza (2019). In the near future, the more "human" response may be to incorrectly label the image.

Figure 1: Image annotation CAPTCHAs



(a) A relatively easy annotation task

(b) A trickier task

meme as they currently often appear on the Internet). If you showed that image to someone who reads English, they will get a very different sense of what the image is about compared to sharing it with a non-English reader. The text content of the image, and whether or not you can read it, matters for understanding meaning.

One critique of the current images-as-data literature in political science is that we rarely address the issue of subjectivity or context. We seem to assume that there is a universally true answer about, for example, the gender or race of someone in a picture and that we can get close to that true answer by accounting for common measurement errors (e.g. making sure our labelers are well trained and paying attention). Yet the constructed nature of gender and race is well-established. What an image viewer in the 17th century would say about the race of someone pictured in a Twitter post would likely be quite different from a 21st century response to the same picture. To bridge the gaps in meaning, it behooves us to draw on the valuable insights from our fellow social scientists in communications, anthropology, sociology, and psychology, not to mention the qualitative and interpretive scholars in our own field. The solution to the challenges of using images as data is to approach the possibilities the data afford with flexibility and openness to a variety of methodological approaches.[5]

3 Four Guiding Questions

Approaching images for political science research from a pluralistic perspective can be helpfully guided by four questions. As we think about research with images, past or present, and keep the often subjective nature of images in mind, these questions help clarify our thinking about methodology and method. In deriving these questions, I draw on work that falls under the heading of *visual research methods*, *visual methodologies* or *visual analysis*, catch-all terms used across many social science disciplines (Emmel et al., 2011; Knoblauch et al., 2008; Webb et al., 2017), and also from Critical Visual Theory.[6] The four suggested questions to guide image research are:

- Who produces the images?
- Who gives the images purpose?
- Who interprets the images?
- Are research subjects active or passive in the research process?

Depending on the answers to these questions, a study will use different methods and will draw on different intellectual traditions. Yet the fundamental challenge of assigning a meaningful truth to an image remains regardless of the methodological approach.

[5]This call echoes those from Sarah Shugars (this volume) and Janet M. Box-Steffensmeier (2021 APSA presidential address).

[6]In a Critical Visual Theory tradition, the relevant questions might be phrased slightly differently as being about image "creation, distribution, and reception" (Ludes et al., 2014, p. 203)

3.1 Who Produces?

The question of who produces the research image(s) has two main answers: usually either the research subject produces or the researcher themselves produces. Subject-produced images could include social media posts, drawings made at the behest of the researcher during an interview, street graffiti, photo journals, recruitment posters, maps, and much more. Subject-produced images could come from amateurs (e.g. asking a child to draw a map from their home to their school) or from experts (e.g. a historical map produced by a professional cartographer). Subject-produced images tell us something of the subject's worldview: what they see; what they think is worth photographing; what they think is worth remembering; how they perceive the world. They are potentially very powerful sources of information and understanding. For example, (Wong et al., 2020) use subject-drawn maps as a measure of place, community, and context.

On the other hand, researcher-produced images allow for a greater degree of researcher control. For example, a researcher might create an image to use as an experimental treatment. A hypothetical survey experiment might vary the image accompanying a news story to see if the image affects how respondents perceive the issue. By making or manipulating the image themselves, researchers have control over the differences between images. Researchers might make their images from scratch, or they might use technology to alter existing images – changing the expression on a politician's face from a smile to a frown, for example (Homan et al., n.d.), to see if that affects a subjects willingness to donate to a political campaign. Deepfake technologies are the current leading edge of manipulated or manufactured images and videos. These technologies produce fake images and videos that can be difficult to tell apart from reality – for example, one might make a fake video that shows a politician saying something they have never actually said on camera. While creating deepfakes is tempting for experimental researchers because of the high degree of verisimilitude, it is crucial to debrief with research subjects to emphasize the manufactured nature of the material.[7]

In some cases, the image production could be a hybrid between researcher and subject. Perhaps the researcher provides the tools – a disposable camera or an app, for example – that they give to the subjects to produce the images. In this example, the researcher provides some bounds (or constraints) to the image production.

In other cases, the image producer is a third party – neither the researcher nor the subject. For example, we might want to know how college students respond to propaganda images. The students would be the research subjects. Instead of making our own example images, we could use real-world propaganda images made by, for example, an opposition party. What we give up in researcher control in designing the images, we gain in verisimilitude of the images.

3.2 Who Gives Purpose?

Answering the question of who gives purpose to an image requires an understanding of which actor uses a picture to some end. After the picture has been produced (taken, drawn, etc.), it is put to some use (posted to social media, turned into a protest poster, used as propaganda, etc.). As social scientists, we are not generally interested in the image for the image's sake. Instead, we want images to help us understand something about humans and human society. For example, When someone chooses a selfie for a profile photo instead of a landscape image, what does that tell us about the role social media plays in our lives? Or how does the socialist realism art movement help us understand political life in the Soviet Union? We are not interested simply in the fact that an image exists; we want to understand how it is tied to other human activities.

As with the first question, the answer to the question of who gives purpose to an image

[7] And, given the propensity of humans to accept information that fits their worldview, some respondents may not believe a debrief that contradicts their impression of the deepfake. The ethical issues of using deepfakes are many, and as such they should be used very carefully in political science research.

is usually either the researcher or the research subject. In experimental cases, the researcher is deploying the image by embedding it in a survey or other format. Varying the images in researcher-made social media posts to track differential engagement would also involve researcher purpose.

In observational studies, however, research subjects decide which images to use for their purposes. The "use" might look like posting an image to social media or hanging a favorite piece of artwork on the wall. Here the subject is attempting to do something by using an image. The researcher has less control over what is posted and it is more difficult to understand the intent of the use. But subject-deployed images reflect a more natural purpose.

As with the first question on production, the purpose could arise from a hybrid between researcher and subject. For example, a subject might be given a set of sample images and asked to pick which one would best accompany a specific news story.

We can also think about cases where there is third-party production. To extend a previous example, we might want to know how subjects respond to propaganda images from an opposition party. We could study this on social media by examining reactions to propaganda posts (e.g. likes, shares). Here a third-party (the opposition party) is deploying the images, not the research subjects (who are doing the liking) or the researcher.

3.3 Who Interprets?

The third question also lends itself to a general "subject or researcher" continuum (with room for hybrid or third-party cases). Given an image, who in the research pipeline is tasked with interpreting the meaning? In most quantitative studies, the researchers themselves extracts relevant information from or about an image, giving it meaning. A researcher may outsource this task to a research team or crowdsourced employees. The task of image labeling, or annotating, provides generalizable meaning (e.g. which pictures show street protest from a diverse set of images). For example, we might ask labelers to look at images and count the number of people present, or we might ask them to guess the emotion being felt by a person in the picture. Here the research team, including crowdsourced workers, are interpreting the images.

Alternatively, we could ask research subjects themselves to tell us what an image means. We can think of the research subject as the person taking the picture; as the person in the picture; or as the person being exposed to the picture for some purpose. We might ask, for example, what they were trying to convey when composing an image, or whether they thought a photograph was humorous when they took it. We might ask someone pictured to tell us their gender, instead of guessing it ourselves. In the example of subjects looking at propaganda, we could ask them to tell us what information they gleaned from the image before decided whether or not to hit the "like" button.

There is a tricky line to ponder over whether crowdsourced employees are researchers or research subjects. Typically, they apply an ontology developed by the researchers, and thus can be considered an extension of the research team. Yet they also make their own interpretative judgements as they label, and the differences in their labeling might be the focus of the research. For example, recent research demonstrates that Republican and Democratic image labelers report very different emotional reactions to left-leaning social movement images Williams et al. (n.d.). The issue of interpretive differences also applies to text analysis. Sap et al. (2019) demonstrates, for example, that tagging texts for hate speech varies depending on how the tagger interprets African American English. Researchers benefit from being clear about who the subjects are. If they are studying the behavior of the labelers, then the labelers are the subjects. If they are using the information provided by the labelers to address another question, then the labelers are part of the research team.

3.4 Active or Passive Subjects?

The fourth question asks whether research subjects are passive or active in the research. In a typical observational study of behavior on social media, for example, the subjects are generally passive in the research. The researcher collects data that subjects have posted online, following a digital trace. The subjects likely do not know they are being studied. A study might have slightly more active subject participation if the subject is asked to download an app that tracks their image sharing activities. Experimental studies asks the subject to be active in the research process, in that they must view an image and then respond in some fashion.

Yet we can also imagine studies where the subjects are even more active than in the experiment example. For example, we might work with subjects to develop the research design. Or we might ask subjects to take pictures of the political signs they see in their neighborhood. We might then ask them to participate in an interview where they describe their photos and explain why they took them. Once the researcher writes up a draft of the results, they could ask participants to read the draft and give feedback. There are also studies that straddle the line between active or passive. A participant-observation study of an activist art studio, for example, might involve the researcher also being an active subject.

4 A Research Example in Light of the Guiding Questions

As described at the outset of this chapter, many projects with an "images as data" flavor are positioned firmly in the large-N, quantitative realm. They rarely engage explicitly with the four questions laid out here. Yet for the most part, the answers to the questions are fairly obvious from reading the paper. For example, Casas et al. (2018) collect images that are shared online to see which images received more shares. The subjects are the social media users who view the images "in the wild" and decide to hit share. The images were produced by other social media users and were purposefully posted by those users, meaning that the answer to first two questions is "third-party" production and distribution. The unit of analysis is the post itself, with the number of shares the post received as the outcome of interest. Therefore, and answering the fourth question, the subjects (those doing the sharing) were passive in the study process. Although they were actively sharing posts, they were unaware they were in a study and were not asked by the researchers to do anything specific.

It is the third question that is the trickiest for papers like Casas et al. (2018). The researchers labeled images for the presence of certain objects or scenes, including whether or not the images contained a street protest. They also labeled images for how strongly the images evoked a set of emotions. The coding ontology was developed by the researchers and labeling was completed by undergraduate students and crowdsourced (Mechanical Turk) workers. Measures of inter-rater reliability checked that the annotators generally had similar rates of seeing protest, though there was greater variability around the emotional reactions Casas et al. (2018, Appendix B).

What is important to think about is whether the interpretations of the image meaning that the researchers had, filtered through the annotators and labeling forms, matched how subjects viewing the original posts interpreted them. Did they see protest in the images at the same rate as the annotators? Do their emotional reactions match those of the annotators? Did the original viewers making the decision to retweet notice protest in the picture, or did they focus on some other aspect of the image, perhaps the presence of a sign with provocative language? It is extremely difficult to say given the research procedures, especially with the entirely passive nature of subject participation. Considering the guiding questions prior to embarking on images-as-data research could strengthen the validity of such work. In this example, focus groups or a survey of social media users about the posts they share could add a helpful element of active subject participation.

5 Computational Social Science Methods and Images

With large amounts of data, scholars use what are increasingly referred to as "computational social science" tools. These often include machine learning and artificial intelligence but at a minimum include automated processing of materials. While the Casas et al. (2018) paper described above takes a quantitative approach, it uses no machine learning or artificial intelligence tools. Yet the questions raised by the above discussion about interpretation carry through and can be magnified when these technologies are employed. Before explaining the challenges of interpretation in the computational sphere, I first discuss some automated tools that do not involve machine learning. I then briefly describe two main classes of machine learning tools that large-N images-as-data researchers encounter: supervised and unsupervised learning.

5.1 Automated processing

Many tools exist to automatically process images that do not rely on machine learning. Popular image processing packages in Python and other programming languages can perform a range of tasks, from converting raw image files into arrays to resizing, cropping, and rotating. Digitized images are represented as numeric values, where each value indicates the color intensity of a single pixel in the image (either red, green or blue, or a gray-scale value if the picture is in black and white instead of full color). These processing tools allow you to, for example, deduplicate images by comparing how similar the numerical representations of images in a corpus are to one another. The cropping and resizing tools are often used to prepare (or "preprocess") images to be used in a machine learning pipeline, as the algorithms described below are often built to only work with images of a certain size (e.g. 224 by 224 pixels). Yet these might also be tasks that researchers want to accomplish outside of a machine learning pipeline. Perhaps, for example, you only want to work with the top half of each image. You could automatically crop the images with a program instead of cropping them all by hand. Other non-machine-learning tools can help automatically extract information about the images. For example, the Athec package from Yilang Peng returns a range of information about the aesthetics of an image, including measures for colorfulness and visual complexity (Peng, 2022).

5.2 Supervised learning

Imagine that you have a dataset of 2 million geolocated images scraped from the photo-sharing website Flikr. And suppose that your research question involves using those photos to estimate the prevalence of public art projects around the world. A supervised machine-learning algorithm could help you to quickly classify the images into groups based on the presence of public art in the photos. The algorithm would answer the question, "Is there public art in this photo?" The current forefront of these algorithms learn how to classify images into categories (e.g. public art or no public art) based on a set of manually labeled images. That is, you provide the algorithm with images that you already know contain public art (*true positive images*) and images that you already know do not contain public art (*true negative images*). In order to identify these true positives and true negatives, annotators will look at a sample of images from the full corpus. Once trained using the manual labels, the algorithm can then classify the remaining, unlabeled images. For more on this process see Webb Williams et al. (2020) or Torres et al. (2021).

The prior example involves binary classification, or putting images into one of two classes (e.g. the presence of public art, or not). Supervised learning can also take multiple classes, such as a classification schema that could separate out different types of public art (e.g.

sculpture, murals, knit bombs). Again, to train an algorithm for these tasks, a subset of labeled images for each class is required.

There are many challenges to supervised machine learning, from the computing power it takes to train the algorithms to the difficulties of building training sets that accurately reflect the domain of interest (that is, making sure the true positive and true negative images are representative of the remainder of the corpus to be labeled). But beyond these more technical issues is the fundamental issue of interpretation. How do we know what public art looks like? Does graffiti count, for example? A project like this requires the researchers to have a clear sense of their target classes, and to make sure that annotators interpret images in the same way. And here is where qualitative tools underpin quantitative research with images. In essence, we are engaging in content analysis with this type of research – labeling images for the presence or absence of certain markers of meaning. How we develop, adapt, and teach our coding schema to others will affect our labels, and therefore also our machine learning algorithms.

Commercial image taggers, including services provided by Amazon, Google, and Microsoft (among others) offer their own supervised-learning, trained algorithms to return labels/classifications after "looking" at provided images. These can be quick and relatively inexpensive tools to analyze image content. Yet the issues with image labels persists – if you use these tools, you know very little about how the original image labels were generated. How were annotators trained? What types of biases did they bring to the task, and how is that bias carried forward by the trained algorithm? A full accounting of the risks and rewards of commercial image tagging services is beyond the scope of this chapter. Chapter 16 from Webb Williams et al., 2020 contains a general overview of commercial taggers. In addition, Schwemmer et al., 2020 does an excellent job addressing the risks of commercial taggers, particularly relating to biases in labels, while Buolamwini et al., 2018 finds that commercial gender taggers perform much worse for individuals with darker skin tones. Researchers tempted to use these tools should not forget to think through the four guiding questions, particularly on the question of interpretation.

5.3 Unsupervised learning

What if we wanted to explore our hypothetical 2 million Flikr images without a pre-specified classification schema? How could we group together images that are similar to quickly see what broad types of images are in the corpus? For this task, we could use unsupervised classification tools. For these methods, we do not provide the algorithm with any pre-labeled data. Instead, we ask the algorithm to cluster the images according to a set of rules. Any set of numbers can be used to build clusters, where the numbers are leveraged to calculate the distance between observations (for example using the popular k-means algorithm). Digitized images are easily transformed into a string of numbers for clustering. The questions of interpretation arise in (a) determining what the optimal number of clusters is and (b) interpreting what the clusters represent. Point (a) derives from the fact that we do not know what the true number of clusters should be (if we already knew what the different types of images were in the corpus, we would not need unsupervised classification for corpus exploration!). If we specify 10 clusters for an algorithm, it will return 10, even if some of the clusters contain only one image or do not seem to have any commonalities to the human eye. Point (b) suggests that the interpretation question follows us throughout our quantitative explorations. Once we have our 10 clusters, how do we know what each cluster represents? What may seem like an obvious title for a cluster to one person (e.g. "these are all pictures of the sky") might not be what someone else would use to describe the same images (e.g. "these are all pictures of clouds").

When using these quantitative tools, researchers must keep the questions of interpretation in mind. Whose interpretations we use, and how we develop what meaning we care about, will impact our research outcomes. In making this point about the inevitability of interpretation questions in quantitative studies, I join others in noting that producing excellent large-N, images-as-data research relies on (or should rely on) qualitative fundamentals (see, for example, Tanweer et al., 2021).

5.4 Best Practices

No matter the number of images in an image study, a few best practices apply. First is researcher reflexively, or a consideration of the biases, goals, and perspectives the researchers themselves carry as they approach the work. Second is transparency – other scholars must be able to understand the research procedures, the source of images and image labels, and so on. The four guiding questions in this piece can help researchers be clear about their research procedures. Third is validation – a point often made by text-as-data researchers (see, for example, Grimmer et al., 2013). A validation step requires checking how accurate the machine-driven process is based on a human ground-truth. It also means checking how different labelers might produce different labels and checking for systematic biases in labels based on image types. For example, if three out of three independent lablers respond that a photograph contains a protest, we can be relatively sure that there is a agreement about what a protest looks like and whether it is present in the given picture. With large numbers of images, having multiple, diverse labelers is key so that disputes over annotations can be resolved (e.g. if two out of three annotators saw protest, we could use a majority rule to say that protest is present). In crowdsourcing situations, the researcher must spot-check the results and put into place other measures to ensure good quality labels (e.g. speed checks so that labels from suspiciously fast annotators are dropped). As a recent example, see Boussalis et al. (2021), who analyze still frames from videos for which emotions are shown on faces. The authors engage in careful validation to demonstrate that the perceived emotions are fairly stable across labelers; they also test whether there are systematic biases in perceiving emotion based on the gender of individuals in pictures.

6 Trends in Images as Data Work

Within the large-N, images-as-data world, social scientists have analyzed a range of topics. Understandably, scholars in the realm of communications and psychology often have a leg up on political scientists or sociologists because communications and psychology theories already treat images seriously (to name just a few sources: Barry, 2002; Graber, 1996; Whitehouse et al., 2006). Communications scholars were also poised early on to think about questions related to social media and the Internet. Political scientists who do not have political communication as a primary area of study often have catching up to do in terms of both theory and methods if their work addresses social media data. Yet many political scientists have begun working in the images-as-data sphere, combining cutting-edge tools (quantitative and qualitative) with a desire to understand political phenomena.

Much of the current and ongoing research activities using images-as-data tools in political science ask questions about political *activities* and/or political *actors* and/or *political communication*. Activities of study include protests (Casas et al., 2018; Won et al., 2017; Zhang et al., 2019), voting (Cantú, 2019; Torres et al., 2021; Wu et al., n.d.), and violence ("Bodies as Battleground: Gender Images and International Security", n.d.). These studies might use images to estimate the size or ethnic composition of protests, or as a measure of voter suppression. "Actors" often means politicians or political candidates, and how they either portray themselves (Franklin Fowler et al., 2016; Joo et al., 2019) or are portrayed by others (Peng, 2018). Within the broadest category of political communication fall studies of more general phenomena, such as propaganda (Lu et al., n.d.).

Many of these images-as-data projects rely on social media sources or news media (print or television). For example, the Wesleyan Media Project[8] – following in the footsteps of the Wisconsin Advertising Project – has created a valuable repository of political advertising in

[8] https://mediaproject.wesleyan.edu/

the United States, including televised and social media ads (Franklin Fowler et al., 2016). Video ads can be broken up into single frames and analyzed as static images (or more complicated methods can track movement or other features within video – see Dietrich (2020) and Neumann et al. (n.d.)), but each video could have thousands of frames. Thus big data processing tools are appealing. Other sources of big data include images collected from Twitter (Casas et al., 2018); Russian-sponsored Facebook ads in the run-up to the 2020 election in the USA ("Social Media Advertisements", n.d.); the Wayback Machine from the Internet Archive ("Internet Archive Wayback Machine", n.d.); and content scraped from websites. Other studies leverage even more creative outlets of imagery, including collecting militant recruitment posters (Loken, 2021), tracking graffiti (Lerner, 2019) or asking children to draw political scenes (Bos et al., 2021).

6.1 Combining images with other channels

Increasingly, scholars are combining images with other forms of data, including text (Wu et al., n.d.; Zhang et al., 2019). These studies acknowledge that images and texts often work in combination to convey information or otherwise create a reaction in the viewer. Other work analyzes the combination of a string of images (a video) with audio data, analysing music or the tone of voices accompanying the images (Hwang et al., 2019).

These endeavors are undoubtedly useful for social science. They also magnify the challenges of interpretation described above. Who decides whether a particular melody is happy or sad, for example? What if the movie-maker intended to convey irony, but the annotator did not pick up on this subtlety? When combining tools, the best practices of reflexively, transparency, and validation still apply, as do the four guiding questions laid out above.

7 Conclusion

This chapter has argued that a wide range of social science work falls under the general moniker of "images as data," from experimental work to ethnography. The four proposed questions to guide research are applicable to many different methodological perspectives. Yet the bulk of the current political science field that claims the "images as data" title involve quantitative, large-N analyses. While these studies are important and have allowed for significant advances in many subfields, they often fall short in answering the guiding question of "who interprets?" I argue that attempting to maintain a pluralistic perspective by incorporating both quantitative and qualitative perspectives will strengthen the validity of these large-N studies. As computational social science tools become more readily available and more mainstreamed in our disciplines, we must stay thoughtful and critical about their relative strengths and weaknesses.

References

Anastasopoulos, L. J., Badani, D., Lee, C., Ginosar, S., & Williams, J. (2016). Photographic Home Styles in Congress: A Computer Vision Approach, 1–52.

Barry, A. M. (2002). Perception and Visual Communication Theory. *Journal of Visual Literacy*, *22*(1), 91–106.

Bodies as Battleground: Gender Images and International Security. (n.d.).

Bos, A., Greenlee, J., Holman, M. R., Oxley, Z., & Lay, J. C. (2021). This One's for the Boys: How Gendered Political Socialization Limits Girls' Political Ambition and Interest. *American Political Science Review*, 1–18.

Boussalis, C., Coan, T. G., Holman, M. R., & Muller, S. (2021). Gender, Candidate Emotional Expression, and Voter Reactions During Televised Debates. *American Political Science Review*, *First View*, 1–16.

Buolamwini, J., & Gebru, T. (2018). Gender Shades: Intersectional Accuracy Disparities in Commercial Gender Classification. *Proceedings of Machine Learning Research*, *81*, 1–15.

Callen, M., & Long, J. D. (2015). Institutional Corruption and Election Fraud: Evidence from a Field Experiment in Afghanistan. *American Economic Review*, *105*(1), 354–381.

Cantú, F. (2019). The Fingerprints of Fraud: Evidence from Mexico's 1988 Presidential Election. *American Political Science Review*, *113*(3), 710–726.

Casas, A., & Webb Williams, N. (2018). Images that Matter: Online Protests and the Mobilizing Role of Pictures. *Political Research Quarterly*, *72*(2), 360–375.

Dietrich, B. J. (2020). Using Motion Detection to Measure Social Polarization in the U.S. House of Representatives. *Political Analysis*, *29*(2), 250–259.

Dzieza, J. (2019). Why CAPTCHAS Have Gotten So Difficult: Demonstrating you're not a robot is getting harder and harder. *The Verge*.

Emmel, N., & Clark, A. (2011). Learning to Use Visual Methodologies in Our Research: A Dialogue Between Two Researchers. *Forum Qualitative Sozialforschung / Forum: Qualitative Social Research*, *12*(1), Art. 36.

Franklin Fowler, E., Franz, M. M., & Ridout, T. N. (2016). *Political Advertising in the United States 1st Edition*. Routledge.

Graber, D. A. (1996). Say it with Pictures. *The Annals of the American Academy of Political and Social Science*, *546*(The Media and Politics), 85–96.

Grimmer, J., & Stewart, B. M. (2013). Text as Data: The Promise and Pitfalls of Automatic Content Analysis Methods for Political Texts. *Political Analysis*, *21*(3), 267–297.

Homan, M. D., Schumacher, G., & Bakker, B. N. (n.d.). *Do We Mimic Politicians? Examining Voter Responses to the Emotional Displays of Politicians*.

Hwang, J., Imai, K., & Tarr, A. (2019). *Automated Coding of Political Campaign Advertisement Videos: An Empirical Validation Study.*

Internet Archive Wayback Machine. (n.d.).

Jean, N., Burke, M., Xie, M., Davis, W. M., Lobell, D. B., & Ermon, S. (2016). Combining satellite imagery and machine learning to predict poverty. *Science*, *353*(6301), 790–4.

Joo, J., Bucy, E. P., & Seidel, C. (2019). Automated Coding of Televised Leader Displays: Detecting Nonverbal Political Behavior with Computer Vision and Deep Learning. *International Journal of Communication*, *13*, 4044–4066.

Joo, J., & Steinert-Threlkeld, Z. (2022). Image as data. *Computational Communication Research*, *4*(1), 11–67.

Knoblauch, H., Baer, A., Laurier, E., Petschke, S., & Schnettler, B. (2008). Visual Analysis. New Developments in the Interpretative Analysis of Video and Photography. *Forum Qualitative Sozialforschung / Forum: Qualitative Social Research*, *9*(3), Art. 14.

Lazer, D., & Radford, J. (2017). Data ex machina: Introduction to big data.

Lerner, A. M. (2019). The Co-optation of Dissent in Hybrid States: Post-Soviet Graffiti in Moscow. *Comparative Political Studies*, *November.*

Livny, A. (2021). Can Religiosity be Sensed with Satellite Data? An Assessment of Luminosity during Ramadan in Turkey. *Public Opinion Quarterly.*

Loken, M. (2021). Using images as data in political violence research. *Journal of Human Rights*, *20*(3), 373–379.

Lu, Y., & Pan, J. (n.d.). The Pervasive Presence of Chinese Government Content on Douyin Trending Videos. *Computational Communication Research.*

Ludes, P., Nöth, W., & Fahlenbrach, K. (2014). Critical Visual Theory - Introduction. *TripleC*, *12*(1), 202–213.

Min, B., Mensan Gabab, K., Fall Sarrc, O., & Agalassoud, A. (2013). Detection of Rural Electrification in Africa using DMSP-OLS Night Lights Imagery. *International Journal of Remote Sensing*, *34*(22), 8118–8141.

Neumann, M., Franklin Fowler, E., & Ridout, T. N. (n.d.). Body Language and Gender Stereotypes in Campaign Video. *Computational Communication Research*.

Peng, Y. (2018). Same Candidates, Different Faces: Uncovering Media Bias in Visual Portrayals of Presidential Candidates with Computer Vision. *Journal of Communication*, *68*(5), 920–941.

Peng, Y. (2022). Athec: A python library for computational aesthetic analysis of visual media in social science research. *Computational Communication Research*, *4*(1), 323–349.

Philipp, H., Müller-Crepon, C., & Cederman, L.-E. (n.d.). *Roads to Rule, Roads to Rebel: Relational State Capacity and Conflict in Africa*.

Sap, M., Card, D., Gabriel, S., Choi, Y., & Smith, N. A. (2019). The Risk of Racial Bias in Hate Speech Detection. *Proceedings ofthe 57th Annual Meeting ofthe Association for Computational Linguistics*, 1668–1678.

Schwemmer, C., Knight, C., Bello-Pardo, E. D., Oklobdzija, S., Schoonvelde, M., & Lockhart, J. W. (2020). Diagnosing Gender Bias in Image Recognition Systems. *Socius*, *6*.

Social Media Advertisements. (n.d.).

Tanweer, A., Gade, E. K., Krafft, P. M., & Dreier, S. K. (2021). Why the Data Revolution Needs Qualitative Thinking. *Harvard Data Science Review*, 1–39.

Torres, M. (n.d.). *The Bag of Visual Words: a framework for the unsupervised and semi-supervised analysis of images*.

Torres, M., & Cantú, F. (2021). Learning to See: Convolutional Neural Networks for the Analysis of Social Science Data. *Political Analysis*, *First View*, 1–19.

Varshney, K. R., Chen, G. H., Abelson, B., Nowocin, K., Sakhrani, V., Xu, L., & Spatocco, B. L. (2015). Targeting Villages for Rural Development Using Satellite Image Analysis. *Big Data*, *3*(1), 41–53.

Webb, J., & Bedi, S. (2017). A Brief Introduction to Visual Research Methods in Library and Information Studies. *Visual research methods: An introduction for library and information studies* (pp. 3–28). Facet Publishing.

Webb Williams, N., Casas, A., & Wilkerson, J. D. (2020). *Images as Data for Social Science Research: An Introduction to Convolutional Neural Nets for Image Classification*. Cambridge University Press.

Whitehouse, A. J., Maybery, M. T., & Durkin, K. (2006). The development of the picture-superiority effect. *British Journal of Developmental Psychology*, *24*(4), 767–773.

Williams, N. W., Casas, A., Aslett, K., & Wilkerson, J. D. (n.d.). *When Republicans See Red but Democrats Feel Blue: Why Labeler-Characteristic Bias Matters for Image Analysis*.

Won, D., Steinert-Threlkeld, Z. C., & Joo, J. (2017). Protest Activity Detection and Perceived Violence Estimation from Social Media Images. *Proceedings of the 25th ACM International Conference on Multimedia*.

Wong, C., Bowers, J., Rubenson, D., Fredrickson, M., & Rundlett, A. (2020). Maps in people's heads: Assessing a new measure of context. *Political Science Research and Methods*, *8*(1), 160–168.

Wu, P., & Mebane, W. R. J. (n.d.). MARMOT: A Deep Learning Framework for Constructing Multimodal Representations for Vision-and-Language Tasks. *Computational Communication Research*.

Zhang, H., & Pan, J. (2019). CASM: A Deep Learning Approach for Identifying Collective Action Events with Text and Image Data from Social Media. *Sociological Methodology*, *49*(1), 1–57.