



ARTICLES

## AI as a Historical Lens: An Experiment in Periodization of Russia's State Photography Archive with Neural Networks

Seth Bernstein 

Department of History, University of Florida, Gainesville, USA  
Email: [sethbernstein@ufl.edu](mailto:sethbernstein@ufl.edu)

### Abstract

Chronology is an important framing mechanism in history and changes significantly based on who defines historical eras. The area studies field has recently grappled with the need to decenter perspectives and reconsider the sources that scholars use. This article uses deep learning artificial intelligence methods to process 169,634 images from the Russian State Documentary Film and Photo Archive (RGAKFD), a major archive of photography in the region, as containing a statist chronological logic, one defined by political change in the center. By peering under the hood of the algorithm's predictions, by thinking with the machine, it is possible to see patterns in the images that may not seem crucial to the human eye. Looking at RGAKFD as a potential source of data for AI raises parallels between algorithmic bias and the Moscow-centric bias of sources, while also providing opportunities to use such methods as a tool for exploratory research.

**Keyword:** Artificial intelligence; Digital History; Soviet Union; Russia; Photography

The Russian State Documentary Film and Photo Archive (Rossiiskii gosudarstvennyi arkhiv kinofotofonodokumentov, RGAKFD) in Krasnogorsk is one of the largest repositories of photographs and cinema in the world and a flagship repository for visuals in the Russian, East European, and Eurasian Studies (REEES) field. Its collection includes official photographs from state proceedings and those used in the press from the late nineteenth century to the present. Even scholars who have not worked on location have undoubtedly seen images from the archive or have themselves ordered digital copies for publications. In January 2003, RGAKFD first published a website, complete with twenty-three images, fifteen from its collection and eight of the archive itself.<sup>1</sup> Today, the archive's website has published digital copies of over 200,000 images with catalog information.

As the major government archive of visual culture, it is a repository that reflects the perspective of the Russian/Soviet state. After Russia's full-scale invasion of Ukraine in 2022, scholars have increasingly taken up the idea that the REEES field needs to be decolonized or decentered.<sup>2</sup> I prefer “decenter” over “decolonize,” as the former encompasses a broad attempt to think beyond official narratives that might apply to the Russian periphery as

<sup>1</sup> “Fotogalereia,” Rossiiskii gosudarstvennyi arkhiv kinofotodokumentov, January 4, 2003, <https://web.archive.org/web/20030215101535/http://rgakfd.ru/fotogal.htm> (accessed February 13, 2025).

<sup>2</sup> The number of works that have made this call are very large at the time of writing. See, for instance, the forum: “Approaches to Decolonization” in *Canadian Slavonic Papers* 65, no. 2 (2023): 141–244.

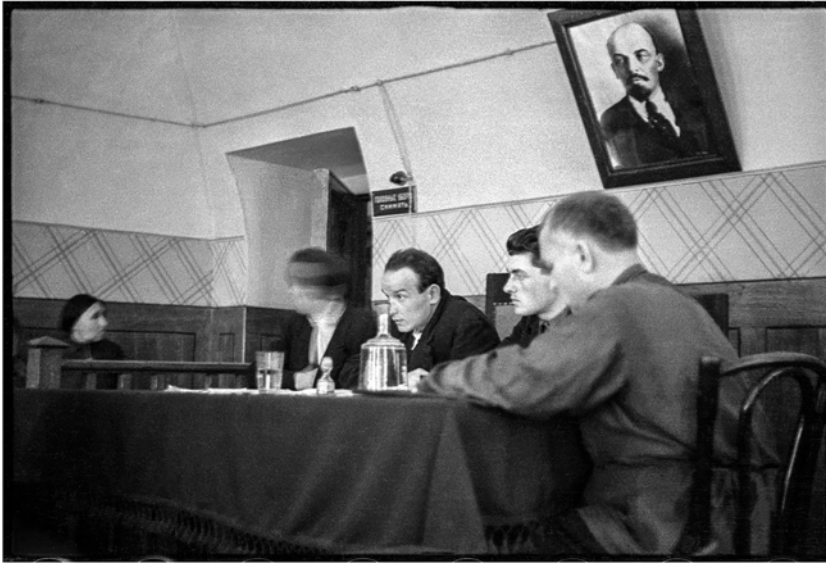
well as to non-Russian areas. Some elements of decentering are clear, such as reconsidering Moscow-centric labels and encouraging research topics that focus on actors beyond the capitals. There are important stories that are difficult or impossible to tell without state archives, though. Even researchers who interpret documents against the grain may encounter the perspective of the state in subtle and diffuse ways. Moreover, the very structure of these archives can channel researchers to sources that embed a logic of the state that is difficult to perceive. In the case of RGAKFD, is it possible to see this vantage point manifested across hundreds of thousands of images?

Periodization is an important part of any narrative about the past. A basic element of any statist account is that major political changes in the capital mattered for the country. Arguably the core of history as a discipline is disputes about such narratives, which usually amount to questions of continuity and rupture and the setting of boundaries between eras. Scholars in the Soviet/Eastern Europe/Eurasia area studies field argue about the (dis)continuities between the imperial Russian state and its Soviet successors, between Iosif Stalin and Nikita Khrushchev, between Vladimir Putin and his predecessors from centuries past. As historian Susan Smith-Peter suggests regarding the decentering of the field, a periodization of Russian intellectual history may change if figures from Siberia or other regions are considered on an equal footing with those in the capitals.<sup>3</sup> Generally, scholars engage with periodization as Smith-Peter proposes, by analyzing a selection of documents about significant historical figures, policies, or discourses to establish change or stability over time. This article presents another way to consider continuity and change by analyzing a large amount of data in aggregate with the help of computer algorithms. The introduction considers the problem of AI, artificial intelligence, in the humanities and the REES field broadly. The first section takes an overview of the data and the technology. The second analyzes how deep learning models respond to different sets of periodization. The third examines how the models make their predictions and how that might change ideas about the visual cues that tie together periods. I have also included a technical appendix for those interested in experimenting with these methods.

For this article, I accessed all the images from the twentieth century on RGAKFD's site and used them to create computer models that predict the era of photographs. Using a deep learning algorithm, a form of artificial intelligence that finds patterns in data without human intervention, I created several models that classify data by category. In this case, I had the models analyze photographs grouped by period. Each model used largely the same photographs but processed the images as grouped into different arrangements of years. Three models separated the images into regular periods, like decades. Two used periodization that adhered to textbook versions of Russian/Soviet political history. Deep learning algorithms are excellent at finding patterns in data but, similar to human observers, are better at predicting when the patterns are stronger. If the RGAKFD archive is defined by political change at the center, a model that reflects that idea should be the best at predicting the era of photographs. At the risk of spoiling a finding, this hypothesis was correct and political periodization allowed the algorithm to find stronger patterns than models generated with images grouped by regular intervals. Why was that the case? By peering under the hood of the algorithm's predictions, it is possible to see patterns in the images that may not seem crucial to the human eye.

This article will go into details of the technology later, and the appendix gets quite technical. For the time being, it might help to demystify things by showing an example. [Figure 1](#) is an image taken by an American visitor to the USSR in 1934, not part of the RGAKFD

<sup>3</sup> Susan Smith-Peter, "Periodization as Decolonization," H-Net, January 4, 2023, <https://networks.h-net.org/node/10000/blog/decolonizing-russian-studies/12148542/periodization-decolonization> (accessed February 13, 2025).



**Figure 1.** Official Proceeding, USSR (possibly Moscow) by Richard J. Scheuer, 1934. My thanks to Joan Neuberger for sharing the image and to the Scheuer family for providing permission for its use.

collection. I analyzed the image with a model that I am calling “Big Political Eras” that classifies photographs into five periods of ten to thirty years, divided by important political shifts (from 1917 revolutions to the period of the Soviet government before World War II) (Figure 1). The model correctly predicts that the image belongs to the period between 1917 and 1940 (inclusive). It is also possible to visualize the parts of the image that the model associated with that period. The visualization is called a “class activation map,” which the algorithm can then show as an overlay on the image. (See Figure 1a and Figure 1b; here and elsewhere, the class activation map and overlay image derived from the original are available as supplementary materials at <https://doi.org/10.1017/slr.2025.10148>). What the model sees as predictive is both expected and not. The edge of the Vladimir Lenin portrait is highlighted, but so is the wallpaper, the wooden paneling and the folds of a man’s revolutionary tunic.

By examining machine-predicted periodization in RGAKFD’s photographs, this article stakes out a position in the area of artificial intelligence in the humanities. It has become commonplace for scholars to use large-scale textual data to see change over time in the structure and content of writing.<sup>4</sup> The development of methods and tools for

<sup>4</sup> Computational linguistics is a longstanding field of study whose tools have become more common in humanities disciplines with the advent of accessible platforms for their use. Franco Moretti coined the term “distant reading,” and his collection of articles in the book of the same name, *Distant Reading* (London, 2013), provides excellent examples of the technique. The number of scholars working along similar lines is hard to count. One notable example is Frank Fischer, et al., “Programmable Corpora: Introducing DraCor, an Infrastructure for the Research on European Drama,” last modified July 10, 2019, in *Proceedings of DH2019: “Complexities,”* Utrecht University, doi:10.5281/zenodo.4284002 (accessed February 13, 2025). A set of corpora for computational analysis of the structure and text of theatrical works with especially strong datasets in east European languages like Bashkir, Russian, Tatar, and Ukrainian.

macro-analysis of visual culture has been slower to develop, although new avenues for “distant viewing” have emerged in recent years.<sup>5</sup> There are ways that one could approach the task of distant viewing of a visual archive with simple but powerful statistical analysis of the apparent qualities in images. For instance, it is possible to measure the average contrast in brightness in thousands of images to gain a sense of stylistic change over time. A more sophisticated—and to humanities scholars, frustratingly opaque—method is to let the algorithm itself classify the data using deep learning. One such algorithm is a neural network, inspired by the unconscious workings of biological nervous systems, which predicts patterns in data without human input. Unlike a procedural algorithm, where the programmer tells the computer all the steps to complete a task, for a neural network algorithm, the programmer sets the basic parameters—how many criteria (neurons) to use, how many times to revise predictions, what data to consider—and the algorithm determines the best way to classify the data.

Over the past several years, such artificial intelligence models have increasingly occupied the thoughts of humanities scholars. The breakthrough of the company OpenAI’s ChatGPT web application at the end of 2022 brought artificial intelligence into every humanities department.<sup>6</sup> Virtually all scholars are considering how artificial intelligence will change their job, especially as teachers, even if their understanding of the technology is limited. Seemingly every journal in the humanities is publishing special issues on the topic, including *American Historical Review*’s forum in 2023 on artificial intelligence in history.<sup>7</sup>

The REEES field has not yet succumbed to AI mania it seems. This is not to say that there are no works in the field that employ or analyze sophisticated digital methods. However, this research has not been a major presence in the leading journals in area studies scholarship.<sup>8</sup> While other humanities disciplines have worried about AI, REEES scholars have been understandably distracted in the aftermath of Russia’s invasion of Ukraine. Make no mistake, though, computational methods will influence this field, not only despite the war but perhaps because of it. Scholars’ inability to access sources in the region has already encouraged the use of digitized materials. Meanwhile, the closure of some repositories and the potential that governments will limit access to existing digitized sources may further narrow the choice of databases. The search for novel methods to analyze well-tread sources and the availability of algorithms to apply to digital scans and transcriptions will make computational analysis an increasingly attractive option for researchers.

<sup>5</sup> Taylor Arnold and Lauren Tilton have used this term to describe a toolkit they are developing for large-scale analysis of visual corpora. See Taylor Arnold and Lauren Tilton, *Distant Viewing: Computational Exploration of Digital Images* (Cambridge, Mass., 2023).

<sup>6</sup> Just one example of thousands: Beth McMurtrie and Beckie Supiano, “ChatGPT Has Changed Teaching. Our Readers Tell Us How,” *The Chronicle of Higher Education*, December 11, 2023, <https://www.chronicle.com/article/chatgpt-has-changed-teaching-our-readers-told-us-how> (accessed February 13, 2025).

<sup>7</sup> R. Darrell Meadows and Joshua Sternfeld, “Artificial Intelligence and the Practice of History: A Forum,” *The American Historical Review* 128, no. 3 (September 2023): 1345–49. See also the associated articles from the forum.

<sup>8</sup> An exception to the lack of digital scholarship in flagship REEES publications is Hilah Kohen, Katherine M. H. Reischl, Andrew Janco, Susan Grunewald, and Antonina Puchkovskaia, “Reading Race in Slavic Studies Scholarship through a Digital Lens,” *Slavic Review* 80, no. 2 (Summer 2021): 234–44; and Tatyana Gershkovich, Madeline Kehl, and Simon DeDeo, “Public Patterns in Private Writing: Computational Insights into Russophone Diaries,” *Russian Review* (forthcoming), <https://doi.org/10.1111/russ.70026>. For a broad overview of the use of artificial intelligence in the field, see Daria Gritsenko, Mikhail Kopotev, and Mariëlle Wijermars, “Digital Russian Studies: An Introduction” in *The Palgrave Handbook of Digital Russian Studies* edited by Daria Gritsenko, Mariëlle Wijermars, and Mikhail Kopotev (Basingstoke, 2021). See also individual works in part II of this volume. The journal *Studies in Russian, Eurasian and Central European New Media*, previously known as *Digital Icons*, has published much innovative research that analyzes online media in the last twenty years, although this research tends to use close reading as its method rather than computational approaches.

Discussions about the impact of AI on research in faculty meetings and the academic press have primarily focused on teaching. When research-focused discussions have occurred, they envision such algorithms as speedy but inscrutable research assistants. Historian Lara Putnam's concerns about digitization continue to be relevant as AI techniques enable even easier classification of relevant sources. Scholars working in a paper archive might page through dozens of files to locate a single necessary document. In contrast, a text-searchable database can find the document momentarily. When an archive is digitized, there seems to be no need to visit its distant physical location where a researcher would look through tangentially related files. Yet remote access to digitized sources means that scholars lose the hard-won contextual knowledge that came with on-site, analog research.<sup>9</sup> AI algorithms have intensified this dynamic, sorting information so that a researcher can find relevant documents without using precise keywords; a search in a digitized archive for "Gulag" could uncover sources that do not include that specific term but instead have related words like "camp" or "incarceration." Unlike a keyword search, it is very difficult to know how the algorithm made such a prediction, further decontextualizing the source.

Despite these drawbacks, the efficiency of such algorithms and their capacity to see connections in data mean that they are here to stay. For projects that can tolerate errors, they are a boon that can enable work that would have otherwise been prohibitively costly or extensive. It might take years for a single scholar to count the number of times images of Lenin or Stalin appear in Soviet films. A reliable model running on a decent computer could process this task in a few days and would do so with consistency; it would not be entirely accurate, but its mistakes would be the same across the data.

The increasing proficiency of algorithms to process humanities data at scale is creating new expectations. The answer to the eternal question "why is this case important?" can increasingly be answered with evidence drawn from databases and information provided by AI models. Historian Benjamin Schmidt posits that the capacity of AI models to deal with massive data will at once create competition for humanities scholars from technically savvy fields that are more capable of leveraging computational methods and increase pressure to situate otherwise deeply qualitative work in a context realized through such techniques.<sup>10</sup>

Alongside the scrutiny of AI applications' potential impact on research and learning, there are growing concerns about their production. Communications scholar Kate Crawford and artist Trevor Paglen demonstrate that these models are not value-neutral but embed the bias of the people who choose and process the data that train AI.<sup>11</sup> A language model might default to describing doctors as "he" because that is the most common pronoun in the millions of texts used as training data. Often such bias is not a deliberate choice but a product of programmers' concern for quantity over quality, the preference for integrating sources that are readily available because they have already been digitized without a critical eye toward potential shortcomings. This problem of data bias is exacerbated by AI's opacity. Companies like OpenAI and Google are reluctant to release data sets that helped them develop their models. Even if researchers had access to such data, the exact sources that contribute to the results AI tools provide are difficult to understand. Predictably, this issue has generated the most attention in relation to issues of credit and remuneration:

<sup>9</sup> Lara Putnam, "The Transnational and the Text-Searchable: Digitized Sources and the Shadows They Cast," *The American Historical Review* 121, no. 2 (April 2016): 377–402.

<sup>10</sup> Benjamin Schmidt, "Representation Learning," *The American Historical Review* 128, no. 3 (June 2023): 1350–53.

<sup>11</sup> Kate Crawford and Trevor Paglen, "Excavating AI: The Politics of Images in Machine Learning Training Sets," <https://excavating.ai/> (accessed February 18, 2025). See also Andrew Prescott, "Bias in Big Data, Machine Learning and AI: What Lessons for the Digital Humanities?," *Digital Humanities Quarterly* 17, no. 2 (2023), <http://www.digitalhumanities.org/dhq/vol/17/2/000689/000689.html> (accessed February 18, 2025).

How should the makers of AI models pay the millions of individuals who involuntarily contributed training data in the form of shared images or texts when it is impossible to evaluate the impact of any single source on the models' predictions? These questions of bias and authorship in AI are similar to issues surrounding the decentering of the REES field, where Moscow-centric perspectives are not always overt in texts or images. As AI algorithms vacuum up data to create regionally and historically specific models, images from the center will be the ones that are the most readily accessible. The prevalence of data generated by the Russian/Soviet state has the potential to skew such models in ways that will be hard to trace.

The risk of this technology lies in problems of data selection, but it is also possible to use these same algorithms to analyze patterns in data. Joshua Sternfeld proposes considering deep learning models as a kind of historian that can “enable methods for historicizing those biases” that exist in the data that create them.<sup>12</sup> The term “bias” perhaps suggests a degree of consciousness that may not apply to archival collections. Yet every archive contains assumptions, underlying or explicit, about what data belongs in the collection and how it should be classified—by chronology or geography or theme or otherwise. Such an assumption might be termed a pattern, and deep learning algorithms are excellent at finding patterns, often those that are difficult for human eyes to perceive. These might be poses, objects, landscapes, or technical aspects of an image, such as graininess. Even in cases where the model predicts incorrectly, its justification for the prediction can be revealing of visual elements that it perceives as defining an era.

### The Photo Archive as Big Data

RGAKFD is a large archive, with more than a million images in its physical holdings, but in important ways its size belies the specificity of the collection. It includes photographs from important events (such as the Yalta Conference of 1945) and official portraits of figures like actors and military officers. It also holds photographs that depict non-elites but were intended for use in major Soviet media outlets. A huge number of the images where geographical data is available were taken in Moscow, but even the photographs taken outside the capital reflect an official vantage point. RGAKFD's catalog gives dates for most of its images, and this factor made this article possible. There is a significant chronological bias in the archive. More than a third of the images come from the 1940s, and the 1930s make up an additional sixth of the images (Table 1). In other words, the archive focuses disproportionately on the Stalin period and especially on WWII. At the same time, the photographs surely capture not only the Kremlin's intended visual themes and concerns but also tangential elements, such as technology or socially-rooted aesthetics.

The RGAKFD images are a useful dataset for reasons beyond their volume and scope. It is relatively simple to find the images. I was able to access 210,632 photographs by using the programming language Python to ask the site for each image by its electronic identification number. These digital copies represent approximately one-sixth of all the photographs housed in the physical archive.<sup>13</sup> The images provided freely online are of too low quality to use in publications generally. For generating computer vision models, identifiable but low-quality images can work no worse than high quality images. Indeed, lower quality images are arguably more useful, because they demand less computing power to process.

An additional advantage of the RGAKFD dataset is that it has robust metadata. To train an AI model, one needs to label data by category to provide examples for the algorithm.

<sup>12</sup> Joshua Sternfeld, “AI-as-Historian,” *The American Historical Review* 128, no. 3 (September 2023): 1376.

<sup>13</sup> “Obshchaia informatsiia,” Rossiiskii gosudarstvennyi arkhiv kinofotodokumentov, <http://rgakfd.ru/obshchaya-informaciya> (accessed February 18, 2025).



**Table I.** RGAKFD Image Database by Identifiable Decade

Decade	# Photographs
1900s	1,440
1910s	11,387
1920s	6,902
1930s	27,867
1940s	62,170
1950s	16,089
1960s	11,503
1970s	11,966
1980s	12,994
1990s	7,316
Total	169,634

Training AI models with historical data is harder than working with images from the recent past. Imagine starting an AI training project with an archive of unlabeled photographs. If the images were recent and from a familiar context, it would be easy to hire research assistants to categorize them. In contrast, labelling historical photographs requires specialist knowledge. Archival metadata can serve as ersatz labels, reducing or eliminating the need for manual annotation. In effect, the archivists who created the collection are the annotators.

Metadata in the RGAKFD images differ from item to item. Most images provide a date, and some provide a locale (city and/or event), a list of named persons, a description, and a thematic categorization (“Meetings of Government Bodies”). Such annotations can provide shortcuts in labeling for computer vision models. It would be possible to make a quick and dirty model by isolating, for instance, the photographs where Stalin is a named person.

I gathered the images that I was confident were produced in the twentieth century. I processed the metadata with a Python library called Dateparser, which converts various textual representations of dates into a universal format, (so that January 1, 1900 and 1.1.1900 become the same to a computer), and isolated the year that RGAKFD provides in the image metadata.<sup>14</sup> Of 210,632 images I accessed, 191,925 had chronological data. The images without any timestamp tend to be difficult to place chronologically, such as portraits or reproductions of postcards. Dateparser identified 169,634 images as having dates in the twentieth century, and I trained the models on these images. Several thousand of the remaining images are from the 1800s (410) or 2000s (2,137). For a significant number (19,744), Dateparser gave no date at all or produced an obviously incorrect year (for example, 2061). Most of these mistakes occurred because the metadata gives approximate dates of production (“1930s” or “1923–1925”). Normalizing these dates would have both cost time and created new problems; would a “1910s” image be before or after the revolution? Images with fuzzy dates tend to be photographs of city scenes and studio portraits, which are hard to pinpoint chronologically. A smaller number of photographs depict specific events, like parades or meetings, where the date cannot be determined from the photograph alone. The lack of dates in some images and their approximation in others probably reflect the random absence or loss of identifying data in the archive’s catalog. The kinds of images that do not have dates are still well-represented among the remaining images, and there is little reason to think that their exclusion altered the results significantly.

<sup>14</sup> Dateparser—Python Parser for Human Readable Dates, <https://dateparser.readthedocs.io/en/latest/> (accessed February 18, 2025).

By sorting the images into folders by era, I created “labelled data.” The deep learning algorithm is a pattern-finding machine. It takes a sample of labelled data as examples (“training data”) and attempts to figure out what makes each label cohesive. When a computer processes images, it sees a numerical matrix, where each pixel corresponds to a value; in a grayscale image, for instance, the value goes from 0 (black) to 255 (white). The algorithm looks for repetition between the arrangement of pixels and predicts what category fits the image. Not all repetitions are meaningful, though. The real work of the algorithm is to provide a weight for each pattern or combination of patterns, to check the resulting prediction, and to adjust its weighting before another round of prediction. In my case, I used a shortcut called “transfer learning,” where I grafted my classification task onto a robust image classification model, Google’s Inceptionv3. Because Inceptionv3 is already trained to recognize visual patterns that tend to be predictive in photographs, retraining it to classify periodization in the RGAFKD photographs provided much better results than if I had attempted to train models from scratch.<sup>15</sup>

The algorithm also uses “validation data” to allow the human programmer to assess the quality of the model. A risk in machine learning is that a model will become too specialized in the material it used for training and will be unable to classify material that it has not seen. This is called overfitting. To see whether the model has become too specialized, the algorithm tests the model’s performance on data it has not seen. A rule of thumb is that 15 to 20 percent of a dataset should be set aside as validation data. The machine learning algorithm does not adjust itself based on the accuracy of predictions in the validation data. Instead, the validation process allows the human to make sure the model does not only work for the closed world of the training data. For more information about the steps I took to generate the models, see the technical appendix.

This explanation may seem divorced from reality, so here is an example: I hired a student to gather digital scans of Lenin and Stalin posters from the internet and put them in folders labeled as such. Google has a simple deep learning tool called Teachable Machine.<sup>16</sup> I uploaded the Lenin and Stalin folders, with 438 and 496 images respectively, as two different categories “Lenin” and “Stalin.” The application trained a model with 85 percent of the images and set aside the rest for validation. It searched for repetitions in how the images ordered pixels that might explain why one image is a “Lenin” picture and another is a “Stalin.” It might have identified Lenin, for instance, in the patterns of pixels that create a goatee and bald head. The resulting model was able to predict “Lenin” and “Stalin” pictures on the training data with 98 percent accuracy and around 80 percent accuracy on the validation data. There was a real pattern, and the algorithm discovered it. I then created another model where the images were mixed at random into two categories and labeled them as “Lenin/Stalin 1” and “Lenin/Stalin 2.” The resulting model was also 98 percent accurate in predicting the training data but just 50 percent accurate when it was given validation data. The machine learning algorithm found a pattern in the data that was entirely arbitrary to human understanding. When faced with images it had not seen, the model guessed at random between the two labels.

As this example demonstrates, neural network models are great at classifying data with clear criteria: if we tell an algorithm the value of pieces of art and X, Y, and Z characteristics about each, it can make predictions about the value of other pieces of art based on the same criteria; the novelists Stephen King and Toni Morrison use certain words and sentence structures, and an algorithm that has enough of their works should be able to make a good

---

<sup>15</sup> Christian Szegedy, Vincent Vanhoucke, Sergey Ioffe, Jon Shlens, Zbigniew Wojna, “Rethinking the Inception Architecture for Computer Vision,” *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition* (Las Vegas, 2016): 2818–26. See the appendix for more information on transfer learning.

<sup>16</sup> Teachable Machine, <https://teachablemachine.withgoogle.com/> (accessed February 18, 2025).



guess about the authorship of other, unlabeled works by those authors. In contrast, it is hard to find patterns where they do not exist. If a person would struggle to identify the era of a photograph, so would a computer.

For this study, I created five period classification models by organizing the images from RGAKFD into two recognizable sets of political periods and three sets of regular intervals (decades). For the “Political Eras Model,” I sorted images into seven eras: the pre-revolutionary period, the first years of the Soviet government, the years of Stalin’s rule, the Thaw period, the years from Khrushchev’s ouster through the early 1980s, the years of Perestroika, and the period after the fall of the USSR. For the “Big Political Eras Model,” I sorted images into five eras: the pre-revolutionary period, the pre-1941 Soviet Union, from the year of Germany’s invasion until the year of Stalin’s death, from the post-Stalin leadership until Gorbachev’s ascension, and from Gorbachev to the end of the century. The periodization of these two models could mirror the organization of a course in Soviet history. My assumption was that if the photographs follow a logic of periodization that presents itself in visual patterns, a machine learning model should be able to uncover it and provide more accurate predictions with data sorted in that way.

Such models are useful in ways that are both obvious and perhaps less apparent. Imagine that RGAKFD discovered a cache of undated photographs or a library received a donation of images from the Soviet Union. A program that was able to periodize photographs accurately would save time in creating search metadata; the results would still include errors but could save many hours even if the catalog required corrections. A more abstract application of neural network models is as analytical tools. It is possible to use the models to study at scale which set of periods works best with the data and what elements hold any single period together. By peering into the elements of a photograph that produced an (in)accurate prediction, researchers can gain ideas about the patterns that algorithms find in data sets, connections built on so much information that it would be impossible for a single person or even a research team to process alone.

This technology could be useful with a number of categories of analysis. I use the example of political era periodization because chronological metadata was consistently available for the RGAKFD photographs and because of my disciplinary training. An art historian might create neural network models that predict the authorship of a photograph or the style of a painting; those models’ predictions could be used to highlight affinities between artistic movements. The adaptability of neural networks to different kinds of analysis should excite scholars, because they could allow subject experts to create their own labeled datasets and train models to deploy their knowledge on a massive scale.

Although I did not label the RGAKFD data myself, I had to account for problems with the unbalanced weighting of categories. The RGAKFD data set has significantly more photographs from the 1930s and 1940s, especially from the period of WWII, than other periods. In my first experiments with this data set, I used all of the photographs to make models—more is better, after all. However, I discovered that the disproportionate number of images from the Stalin period skewed the models’ predictions. The models often guessed that photographs were from the 1930s and 1940s and, because so much of the archive came from those periods, the accuracy ratings were quite good. But when the models encountered images from other periods, they also put the photographs in the Stalin period at very high rates. The models were playing “rock, paper, scissors” with an opponent who throws rock most of the time; it could win most of the games by throwing paper at the expense of figuring out other patterns. The solution was to balance the data through sampling, randomly selecting images so that the training data for each period was roughly the same and no category included more than 20 percent more images than any other. I give the number of images per category in each periodization model (Political Eras, Big Political Eras, Decades, Twenty Years, Fifty Years) in [Table 2](#).

**Table 2.** Number of Photographs by Model and Period (Balanced)

Political Eras	#	Big Political Eras	#	Decades	#	Twenty Years	#	Fifty Years	#
1900–16	4,462	1900–16	8,765	1900s	1,157	1900–19	10,270	1900–1949	57,495
1917–28	4,462	1917–40	10,518	1910s	1,388	1920–39	12,324	1950–1999	47,913
1929–53	4,462	1941–53	10,518	1920s	1,388	1940–59	12,324	Total	105,408
1954–63	4,462	1954–84	10,518	1930s	1,388	1960–79	12,324		
1964–84	4,462	1985–99	10,518	1940s	1,388	1980–99	12,324		
1985–91	4,462	Total	50,837	1950s	1,388	Total	59,566		
1992–99	3,719			1960s	1,388				
Total	30,491			1970s	1,388				
				1980s	1,388				
				1990s	1,388				
				Total	13,649				

**Table 3.** Accuracy and Loss by Model

	# Categories	# Training Images	Training Data		Validation Data	
			Accuracy	Loss	Accuracy	Loss
Political Eras 1	7	30,491	51.69%	1.30	57.42%	1.30
Political Eras 2	7	30,491	52.08%	1.29	52.82%	1.30
Big Political Eras 1	5	50,837	57.80%	1.06	55.16%	1.14
Big Political Eras 2	5	50,837	57.84%	1.05	54.67%	1.10
Decades 1	10	13,649	45.07%	1.57	37.34%	1.83
Decades 2	10	13,649	44.72%	1.57	38.25%	1.79
Twenty Years 1	5	59,566	56.18%	1.09	54.85%	1.18
Twenty Years 2	5	59,566	56.38%	1.10	54.03%	1.18
Fifty Years 1	2	105,408	77.13%	0.48	84.14%	0.38
Fifty Years 2	2	105,408	77.12%	0.48	81.96%	0.43

Balancing the data in this way has a disadvantage. An ideal experiment would use the exact same images as training data for each of the models. The inability to do so introduces a variable, where it is possible that one periodization provided worse training for the algorithm than another not because of the periodization itself, but because the randomly sampled training images were exceptionally (in)consistent compared to another model's training data. To hedge against this possibility, I created two sets of images to train each model and checked that the resulting predictions for each were approximately the same. As a pure experiment to test the thesis that sorting by political eras produces the best results, it also might have been worth balancing the total number of images each model saw, since the “Fifty Years” model used nearly eight times more images than the “Decades” model. Ultimately, I decided that balancing the data in that way was not worth the cost of producing less robust models.

### Periodization with Neural Networks

The research question that began this paper was whether a deep learning algorithm would be able to find the period of RGAFKD photographs better when they were sorted by a historically-informed periodization or by regular partitions like decades. The results suggest that the standard political chronology works slightly better (Table 3).

**Table 4.** Big Political Eras Confusion Table

		Predicted Era									
		Pre-Revolutionary		Pre-World War II		World War II and Postwar		Mature Socialism		Perestroika and Post-Soviet	
		V1	V2	V1	V2	V1	V2	V1	V2	V1	V2
Photograph Era	Pre-Revolutionary	73%	80%	7%	11%	6%	3%	10%	5%	4%	1%
	Pre-WWII	5%	8%	39%	62%	25%	15%	22%	10%	9%	5%
	WWII and Postwar	3%	6%	16%	38%	53%	41%	20%	11%	7%	4%
	Mature Socialism	3%	5%	9%	22%	10%	11%	59%	46%	20%	16%
	Perestroika and Post-Soviet	2%	3%	6%	15%	5%	5%	28%	22%	59%	55%

**Table 5.** Twenty Years Confusion Table

		Predicted Era									
		1900		1920		1940		1960		1980	
		V1	V2	V1	V2	V1	V2	V1	V2	V1	V2
Photograph Era	1900	71%	68%	9%	14%	6%	5%	12%	10%	2%	3%
	1920	6%	5%	44%	58%	21%	13%	23%	17%	5%	8%
	1940	4%	4%	19%	32%	49%	39%	24%	18%	4%	7%
	1960	3%	4%	8%	14%	9%	8%	69%	56%	10%	18%
	1980	3%	3%	7%	12%	5%	4%	41%	25%	43%	55%

When a model processes an image, it delivers a set of probabilities that the image belongs to each category. For instance, the Fifty Years model might give an 80 percent probability that an image is from the category of “1900–1949” and a 20 percent probability that it is from “1950–1999.” The accuracy rating is the percentage of images where the model assigned the highest probability to the correct era. The loss rating is a measure of how far the probabilities were from the correct ones, where a lower number means that its predictions were more correct. The best results will have high accuracy and low loss. A model might have a perfect accuracy rating because it gives the highest probability to the correct label all the time but might still have a high loss rate because it assigns a high probability to incorrect labels. Similarly, a model’s having a low accuracy rating but a relatively low loss rating would indicate that it was often assigning a close second probability to the correct label. While the deep learning algorithm trains the model, it keeps a record of the accuracy and loss of its predictions on the training data and validation data. [Table 3](#) presents the best accuracy and loss that each model produced after training on the data for 28 cycles or “epochs.” This is not necessarily the metrics from the final training epoch, as the final iterations of models might make adjustments that produce worse results than earlier iterations. The programmer can set the compiler so that it can save the model with the best results, rather than the latest iteration of the model.

Lots of numbers. How can we tell which periodization worked best? Naturally, it is easier to choose from fewer options than more. For that reason, it is unsurprising that models with fewer categories are more accurate. Comparing the Big Political Eras model with Twenty

Table 6. Political Eras Confusion Table

		Prediction													
		Pre-Revolutionary		Revolutionary		Stalinism		Thaw		Brezhnevite		Perestroika		Post-Soviet	
Photograph Era		V1	V2	V1	V2	V1	V2	V1	V2	V1	V2	V1	V2	V1	V2
	Pre-Revolutionary	78%	65%	6%	9%	5%	7%	6%	9%	3%	8%	1%	2%	1%	1%
	Revolutionary	14%	6%	41%	45%	18%	19%	16%	13%	7%	10%	3%	4%	1%	3%
	Stalinism	7%	2%	10%	10%	44%	46%	27%	22%	8%	12%	3%	5%	1%	2%
	Thaw	5%	1%	4%	4%	11%	12%	54%	51%	18%	21%	4%	7%	4%	4%
	Brezhnevite	6%	2%	4%	5%	8%	8%	31%	26%	40%	42%	8%	13%	4%	5%
	Perestroika	5%	1%	3%	3%	5%	5%	20%	16%	25%	22%	33%	44%	9%	9%
	Post-Soviet	4%	1%	4%	4%	4%	4%	18%	17%	18%	15%	8%	16%	44%	42%

Years is telling, because both have five chronological categories. The former is slightly better at predicting, even though the vagaries of balancing the distribution meant that it used 8,000 fewer photographs as training data. In other words, even though it had less evidence, the machine learning algorithm was able to find better patterns with textbook periodization. All the models find patterns in the images, but standard political eras work better because the shift in political regimes appears to be correlated with a shift in the kind of photographs held by RGAKFD.

In addition to looking at this aggregate data, it is worth examining how the models performed on various eras. The data can be visualized with “confusion tables,” a matrix of the proportion of the actual label and the predicted label. The confusion tables for the Big Political Eras models and the Twenty Year models reveal that both were best able to classify the earlier years of the twentieth century, but the former were able to do so at a higher rate, especially in the second version. This result is intuitive. The photographs of the various eras after the Bolshevik Revolution overlapped a great deal in their iconography and technology, whereas photographs from before 1917 would have been distinct. The Twenty Years model is probably less facile with the earlier period because it includes the years of revolution, 1917–19. Also telling is that both models rarely assign photographs from other periods to the earliest category, suggesting that the models find patterns in the early years that do not apply to other categories and vice versa. This is also true of the Political Eras models, where the classification of images from the Pre-Revolutionary period was as accurate or more accurate than the other models.

The other understandable outcome is that the models tend to make mistakes by guessing adjacent periods. There are some hard to explain results, though. What makes the Big Political Eras and Twenty Years models associate the last four decades of the twentieth century with the non-adjacent period before WWII? In this sense, the Political Eras model, which was less accurate than the other models, has the most understandable distribution, since the mistakes are largely in adjacent eras.

### Seeing Like an Algorithm

It is possible not only to calculate an algorithm’s accuracy, but also to see what elements of the photographs led to those predictions. “Class activation maps” can show the areas of the image that made the model associate it with a certain class, in this case a period. The maps can be layered onto the image to display elements that were determinative in the model’s prediction. Sometimes the class activation maps highlight seemingly random elements, suggesting that the model had no idea and made a wild guess. In many cases, however, the visualization demonstrates a logic, even when the prediction is wrong. This section dissects a handful of these visualizations of the model’s predictions.

The Political Eras model correctly assigns a photograph from the Fifth Congress of the Soviets in 1918 to the category representing the first years of Soviet rule. The factors behind the prediction are similar to human rationale. The model points to both horse heads in the photograph as factors in its prediction. A human, too, might have understood that the presence of horses in an urban scene would place the image in one of the earlier periods. Also notable is that the photograph has a handwritten label that the model seems to highlight. Even if a viewer was not familiar with the Congress of the Soviets, the handwriting would indicate an image made earlier in the century, when such labels were common (Figure 2; see Figure 2a and 2b at <https://doi.org/10.1017/slr.2025.10148>).

A telling contrast is with later group photos (Figure 3; see Figure 3a and 3b at <https://doi.org/10.1017/slr.2025.10148>). With high confidence, all the models place this 1969 group photo of Kazakh delegates to the Third All-Union Congress of Collective Farm Workers in whatever category includes the first half of the 1960s, and similar photographs produce



Figure 2. Fifth Congress of the Soviets. RGAKFD D-415 ch/b. <http://photo.rgakfd.ru/photo/54602>.



Figure 3. Members of the Kazakh Delegation to the Third All-Union Congress of Collective Farm Workers. RGAKFD D-415 ch/b. <http://photo.rgakfd.ru/photo/555746>.

the same result. The overlay from the Political Eras model suggests that it associates this type of photo with an ornate backdrop framing a group of seated comrades. In this case, it appears that this type of image is disproportionately found in RGAKFD's photographs from the early 1960s.





**Figure 4.** Construction of an Ice Cellar at an Agricultural Commune. RGAKFD I-7323 ch/b. <http://photo.rgakfd.ru/photo/422444>.

In an example from 1927 of the construction of an ice cellar, every model places this photograph in whatever category includes the years of WWII. The class activation map from a Big Political Eras model shows the area of the fortification-like woodworks as an important element in making this prediction. The ax is also significant, perhaps suggesting similarities with soldiers holding a rifle or a shovel for digging a trench. The upper left corner includes what appears to be the foliage of a tree but the Big Political Eras model finds it to be notable, possibly because it resembles smoke. A human observer would probably not classify this photograph in the WWII period. The people in the photograph are identifiable as civilians, and their clothes suggest that they are from the early twentieth century. The visual logic of the models is understandable, but limited. The trench might raise the possibility that it is a war photo for a human viewer, but would not create absolute certainty as in this case (Figure 4; see Figure 4a and 4b at <https://doi.org/10.1017/slr.2025.10148>).

The models fail in this case to predict the correct category but reveal patterns. They look at objects and poses, rather than individuals. The predictions are derived from elements that make a genre of photograph that is prominent in an era or objects that only appear in a certain period. One might expect that political leaders, above all Stalin, would be a defining factor of some eras, but the models seem to focus on everything but facial features. The Big Political Eras models incorrectly classify a picture of young Stalin from 1905 in the period from 1917 to 1940. It is tempting to see this misclassification as an association of Stalin's visage with the period of his political ascendancy, but the model does not highlight Stalin except, in lesser focus, for the edge of his hair, an ear, and his forehead (Figure 5; see Figure 5a and 5b at <https://doi.org/10.1017/slr.2025.10148>). The models classify a 1910 image of Sergei Kirov, the future Leningrad Communist Party leader and an ally of Stalin, as being from the same period (Figure 6; see Figure 6a and 6b at <https://doi.org/10.1017/slr.2025.10148>). One might explain this error as caused by Kirov's appearance in photographs from the later period, yet the models do not highlight his face. It is his belt and sleeve, visual elements characteristic of the Stalin-era leadership, and the pattern on



**Figure 5.** Young Stalin. RGAKFD V-1402 ch/b. <http://photo.rgakfd.ru/photo/51424>.

the door that the model sees as relevant. The Decades models classify an image from 1933 of German Communist leader Clara Zetkin and Russian Communist Leader and Lenin's widow, Nadezhda Krupskaya, as being from the 1920s. A person encountering that image might make a similar guess, given the prominence of both women in politics in the 1920s. The model highlights Krupskaya's hair, but may be more interested in her pose, since the other elements the model finds significant involve fabric or the image backdrop (Figure 7; see Figure 7a and 7b at <https://doi.org/10.1017/slr.2025.10148>).

These results suggest both problems for classifying historical photographs with deep learning algorithms and opportunities for leveraging the technology. When historians consider important chronological divisions, the default is to think about the changes or continuities across political regimes. Analyzing a state photography archive, a person might see the most visible changes in the appearance and disappearance of political figures. The machine learning models do respond to periodization when dealing with the RGAKFD photographs, and a standard political division of the twentieth century works better than generic divisions. The models do not seem to care about people, though, despite making good predictions and surprising connections with an underlying logic. Instead, the models appear to register differences in style that transpired over time and may have accelerated with political changes. Is the AI model seeing subtle patterns across thousands of images that a human observer would have trouble seeing? These might include aspects of material culture or fashion that are pervasive but hard to perceive, and employing this technology



**Figure 6.** Sergei Kirov. RGAKFD D-220 ch/b. <http://photo.rgakfd.ru/photo/60156>.

might help scholars connect the backdrop of history to the headlining events that typically define scholarship. Historians of Soviet culture have investigated how political transitions led to cultural change in other instances, especially in literature.<sup>17</sup> Similarly, it is possible to see some misclassifications as recognitions of visual genres that persisted across periods and lineages in the presentation of state control.

This paper explored how machine learning algorithms can allow researchers to test and visualize assumptions embedded in data. The experiment of building prediction models for periodizing historical photographs produced results that are both expected and surprising. I found that dividing images from Russia's state photography archive into a standard political periodization of Soviet history tended to work better than regular chronologies, although all the models were more accurate than random guesses. How the models were able to make these guesses is sometimes counterintuitive. Where many people would look

<sup>17</sup> Among the many works on aesthetic shifts, Katerina Clark, *The Soviet Novel: History as Ritual* (Bloomington, 2000), shows the consolidation of the socialist realist canon in the novel under Stalin. The liberalization of culture after Stalin's death saw an emphasis on sincerity in prose and private writing. See Anatoly Pinsky, "The Diaristic Form and Subjectivity under Khrushchev," *Slavic Review* 73, no. 4 (Winter 2014): 805–27.



**Figure 7.** Clara Zetkin and Nadezhda Krupskaya. 1933. RGAKFD 2-113819 ch/b. <http://photo.rgakfd.ru/photo/898446>.

at political iconography and personages, the models based their predictions less on these factors than on cues that a human might ignore.

It is possible to see these signals as patterns that only have meaning to the algorithm, but they might also inspire a second look at images. Are the specific Stalinist sleeve-wrinkle or the ornate backgrounds of posed photographs that somehow have a home in the Khrushchev period characteristic visual elements that would otherwise go unnoticed? When I was developing the models for this paper, I showed students from a digital history class, many of them veterans of my courses in Soviet history, the mistaken predictions the computer made. The class spent ten minutes browsing the misclassifications and, as in this article, trying to understand what the model had seen in various images, speculating about the logic of the model or critiquing it where the predictions seemed entirely off base. It forced the class to think about aesthetic echoes and the visual clues that define one period versus another.

The models this paper used suggest that RGAKFD's photographs reflect a standard political periodization, but AI models might provide insights to scholars who hoped to revise chronologies as well. Data from a non-Moscow photography archive might produce more accurate models if they were divided with a different periodization. Moreover, the historically-informed groupings this article used were hardly exhaustive. It is possible that another arrangement of the RGAKFD photographs would produce even better predictions and inspire debate among scholars about how they should divide the twentieth century. Beyond contiguous periodization, scholars might use AI models to see the links between other forms of categorization. A possibility I considered for this paper was to arrange the photographs into two categories: the years of the two world wars and all other years. Peering into the areas of the image that were significant to the model might lead to new ideas about what differentiates images of war and peacetime.



The goal of this paper was not to produce the most predictive model for identifying official photographs of the Soviet Union by period, although there is potential value in using computer vision for this purpose. Even the models I have produced for this paper may have a use for determining the era of unlabeled photographs from Russia. Libraries and scholars have used such models to categorize images from other contexts.<sup>18</sup> This is an area where humanities scholars have much to offer the makers of artificial intelligence tools. My method in this paper is not so different from that of many technology companies making AI tools currently; I scraped an entire archive for my data set and randomly divided the images into training and validation data. Rather than simply throwing all the images into the algorithm, though, I could have curated a large sample that I found to be representative of each period. That approach would raise questions about my decisions to include or omit some photographs but would have advantages. By leveraging my domain knowledge as a historian, the resulting model could have produced results closer to my own thinking. The elements it highlighted in the class activation maps might also have been more understandable to me. For the purposes of this article, however, I decided that presenting a machine learning algorithm with the entire archive has value on its own, allowing a large-scale analysis of the repository.

The experiment is a reminder that neural network models do not simply exist but are the product of the choices that their human creators make about data. This paper began by comparing the problem of determining the sources and logic of AI predictions to the issue of decentering in the REEES field. When AI models are produced with state-centric sources, these issues are not only similar but connected. The specific patterns that the RGAKFD periodization models identified in images were hard to see as Moscow-centric, yet in aggregate the models built on a standard periodization of the Soviet/Russian state were more effective in classifying images. For any scholar hoping to decenter the REEES field, the RGAKFD collection would be a poor place to look for new perspectives. Yet as researchers try to employ algorithms like neural networks, centrally-produced archives will be attractive sources of training data because they tend to be large and already digitized. Moreover, the resulting models may still do a good job of classifying photographs when applied to materials from other contexts. A model trained on RGAKFD might correctly identify the period of creation of many photographs from state and private archives in Ukraine or Uzbekistan or the Russian regions. These are the very archives where such an approach might be seen as a cost-saving shortcut to a digital catalog, since robust metadata is more likely to be absent. This way of generating machine learning models has the potential to make systematic mistakes; a researcher using an AI-produced catalog to find pictures from one era might miss an entire genre of photographs that the model misinterpreted based on the patterns it derived from RGAKFD's images. It is possible to pull back the curtain on AI decision making as an analytical tool, though. Looking at the visual patterns that an RGAKFD-trained model emphasized could reveal Moscow's gaze on other regions. By thinking with the algorithm, a researcher can understand not only why a model is appropriate or inappropriate in different contexts but the assumptions embedded in visual corpora.

**Supplementary material.** The supplementary material for this article can be found at <https://doi.org/10.1017/slr.2025.10148>.

<sup>18</sup> Harish Maringanti, Dhanushka Samarakoon, Bohan Zhu, "Machine Learning Meets Library Archives: Image Analysis to Generate Descriptive Metadata," <https://research.lyrasis.org/server/api/core/bitstreams/e11773df-b65f-4f85-84f2-258860a60264/content> (accessed February 18, 2025). For a similar historical image classification project, see Jhe-An Chen, Jen-Chien Hou, Richard Tzong-Han Tsai, Hsiung-Ming Liao, Shih-Pei Chen, Ming-Ching Chang, "Image Classification for Historical Documents: A Study on Chinese Local Gazetteers," *Digital Scholarship in the Humanities* 39, no. 1 (April 2024): 61–73.



**Figure 8.** Eastern Workers in 1945 Encounter Red Army Soldiers. RGAKFD 0-255952 ch/b. <http://photo.rgakfd.ru/photo/166644>.

## Appendix: Method

The machine learning algorithm this article uses is called a convolutional neural network (CNN). The name “neural network” is a metaphor based on the workings of organic nervous systems: a person touches a hot stove (input) and millions of neurons produce signals that result in a feeling of pain (output). Similarly, a researcher can give a neural network model an image (input) and through a series of tests (neurons), it will produce signals that result in a classification (output).

To produce a useful neural network model, a researcher has to train it with data that has already been labeled (training data). In this case, I gave the algorithm the label of an era for each image in the RGAKFD dataset. I told the algorithm how many tests (neurons) to use in the model. For scholars accustomed to having control over their research methodology, it may be a frustration that the programmer has little control or knowledge about what neurons the algorithm uses or which ones it finds meaningful. The compiling algorithm’s job is to assign a value (a weight) to each neuron or combination of neurons and to see whether that value produces an accurate prediction. Because the algorithm has access to the correct labels, after trying the photographs and seeing how correct its guesses were, it can adjust how it values the neurons and try again.

An example: one neuron might assess the brightness of pixels in the corner of photographs; which photographs have similar brightness in that area? The neural network model could initially give this neuron definitive significance, so when the level of brightness in the corner of two photographs is close, the model predicts that these photographs belong to the same category. If it turns out that the model made many incorrect guesses, the compiler algorithm would adjust how the model weighs the neuron that tests corner brightness. The algorithm will continue to change the influence of this and other neurons in the network until it runs through the data a set number of times (epochs) or reaches a level of predictive accuracy that the researcher has specified. This is a very simple example, of course. Neural networks typically include multiple layers of neurons. A first layer might include multiple tests for the brightness in different corners while a test in the second layer might take the aggregate similarity of those multiple corners, a test in the third layer might take the aggregate similarity plus one of the corner measurements from the first layer, and so on.

CNNs are a subset of neural networks that use filters (convolutions) to accentuate data points and make it easier for the algorithm to isolate patterns. When dealing with images, a common type of convolution alters the color or darkness of pixels based on their neighbors. Looking at the images below, one can see how the convolutions might allow the neural network to isolate aspects of the image when the rest is black, perhaps the curve of a hat that was common when the image was taken in 1945 (Figures 8 and 8a).

My models use transfer learning, meaning that I grafted my dataset onto a robust, existing model to produce a new model. This step is critical, as a model built from scratch would make very poor predictions. The principle is similar to training a person to play a new sport. A person who has played standard five-on-five basketball will





**Figure 8a.** Convolution of Eastern Workers in 1945 Encounter Red Army Soldiers.

adapt well to three-on-three. A person who has played no basketball but has played other team sports will have basic athletic skills that could apply to basketball. An adult will generally fair better than a child, and a child better than a baby. Similarly, I used as a starting point the Inceptionv3 model, a sophisticated, Google-developed computer vision model trained on 1.3 million images. The unaltered Inceptionv3 model has 48 layers of neurons and can identify 1000 objects. Retraining Inceptionv3 means that the model forgets the specific things it could predict without forgetting general patterns that tend to be predictive when dealing with images. To extend the previous metaphor, if my models had been built from scratch, it would be like giving infants a basketball and seeing who scores the most points (answer: they all score zero). Using Inceptionv3 as a base for my models is like giving a basketball to football players (your choice which kind); those players may need to be trained in dribbling and shooting, but they will have a high level of fitness and understand rule-based sports.

My experiment uses another technique that is worth discussing, the random manipulation of images to make it harder for a model to specialize in a certain dataset (overfitting). An example can illustrate why this principle works: Imagine a situation where a researcher feeds an algorithm a thousand unmodified images of Lenin as “Lenin photographs,” each with Lenin positioned in a similar way, in approximately the same part of the frame, and occupying approximately the same amount of space in the frame, and a thousand images without Lenin as “Non-Lenin photographs.” If Lenin is situated just so in other photographs, the resulting model will do an excellent job identifying them as “Lenin photographs.” If Lenin is bigger or smaller, in another part of the frame or leaning, the model would be less capable of identifying it as a “Lenin photograph.” Image manipulation simulates the variation in images that might exist outside of the training data.

The process I used simulates difference in images by using the following manipulations at random:

- rotates the images up to 40 percent;
- shifts the width and height of the images by up to 20 percent;
- shears the images by up to 20 percent;
- zooms into the images by up to 20 percent;
- creates a mirrored version of some of the images;

A machine learning model can over fit and get to near perfect accuracy in its predictions on even a massive dataset by isolating patterns in the data that a human would find arbitrary. For instance, the algorithm may find that isolating the brightness of a single pixel in the middle of each photograph always allows the algorithm to achieve total accuracy in categorizing training images of Lenin, and it therefore overvalues that factor. Despite predicting at astounding levels on data it has seen, this over fitted model makes less accurate predictions on images outside of the training data because it searches for that single pixel rather than assessing patterns that are truly determinative. The benefit of using image manipulation seems obvious, but the tradeoff is in the computational

power needed to compile the models. Performing image manipulation is computationally intensive and time consuming. If I was producing many models, it would have been worth considering whether image manipulation contributed significantly to accuracy and, if it did not, omitting that stage of the process.

The platform I used to make these models is TensorFlow, an open-source deep learning library developed by Google and available as a module in the programming language Python (<http://www.tensorflow.org>). Like much of my understanding of neural networks overall, the parameters I used come from DeepLearning.AI's excellent series of courses on TensorFlow, which are available to audit for free on the Coursera platform (<https://www.coursera.org/learn/introduction-tensorflow/>). I assigned the algorithm to use 1024 neurons (tests) on top of the Inceptionv3 architecture. A larger number might produce better results on the training data because it would find very granular results, but it would have a harder time generalizing to a broader dataset. In contrast, a smaller number of neurons could make the model more universal, but it also might miss important specificities of the data. For the loss optimizer, the part of the algorithm that calculates if a change in how the model has weighted the neurons has produced a better or worse result, I used Root Mean Squared Propagation or RMSProp. I trained the models for 28 epochs, meaning that the algorithm spent 28 rounds looking at the training data, and after each round stopped to assess how it had done by testing itself on validation images that it did not use for training. I arrived at the number 28 arbitrarily and, in every case, the algorithm stopped producing significant predictive gains before epoch 28. Within each epoch, the algorithm made small adjustments after seeing a group of 64 images (a batch). I have made my code, the resulting models, and their predictions on the images available here: <https://zenodo.org/records/14176714>.

My goal in making these models as an experiment meant that I was not so invested in the exact parameters used to construct them. Optimizing the parameters would be more important for producing a model for use in real-life classification of images. However, I wanted to know the relative strength of models against one another and to get a sense of how they were making their predictions. It is possible that experimenting with various parameters would have given better or worse results, but doing so would have produced a similar effect for each set of periods.

Another factor worth considering is the computational resources used to make these models. There are thousands of feasible ways to carve up the twentieth century into chronologies. Why not run all to see which works the best? It is possible but computationally expensive. The production of each model and its predictions in this study took roughly 8.5 hours on average using an allotment of two central processing unit (CPU) cores and one graphics processing unit (GPU) on HiPerGator, the University of Florida's research computing cluster. The compilation of the model itself with manipulated images took between four and eight hours, while generating the predictions based on that model took an additional hour or two. These processes would have taken significantly longer without a GPU. It is possible to optimize the process by reducing the number of training epochs, which were probably excessive, and by investigating whether image manipulation is necessary to avoid overfitting in this case. Those wishing to replicate or use these models but lacking access to a research computing service can adapt the code to Google's Colab platform, which provides free but limited cloud access to computers with GPUs and the option of paying for more time on a GPU.

I am thankful to Paula Chan, Nic Delorme, Andy Janco, Dan Maxwell, and Joan Neuberger for their careful readings of this article. In addition to reading the article as a draft, Joan Neuberger helped arrange one of the figures that appears in it. Dan Maxwell also organized my access to the University of Florida HiPerGator research computing cluster. Elizaveta Stovba expertly coordinated the acquisition of high-resolution images for this publication. Sonny Russano provided research assistance with funding from the Center for European Studies at the University of Florida. I am amazed and grateful that *Slavic Review* was able to find three blind reviewers who waded through an unusually technical text by an unknown colleague to contribute insightful suggestions that have improved the work.

**Seth Bernstein** is Associate Professor of History at the University of Florida. He is the author of *Return to the Motherland: Displaced Soviets in WWII and the Cold War* (Cornell, 2023) and *Raised under Stalin: Young Communists and the Defense of Socialism* (Cornell, 2017). He is currently working on a digital toolkit for analyzing Soviet secret police documents and a book about the arrests of Soviet Jews in 1953.