ADDI: Data Analysis Paper

Taylor Arnold and Lauren Tilton

Introduction

This document shows a variety of different approaches for summarizing and visualizing automatically produced visual annotations by the ADDI Project. Specifically, annotations were produced by applying computer vision algorithms to the digitized images from five collections of U.S. documentary photography held by the Library of Congress. Here, we jump directly into a series of the data analyses of these annotations. Please see the corresponding Methods Paper in the GitHub repository for details about the larger project, the specific collections, our approach, the computer vision algorithms, and how the data were processed and organized.

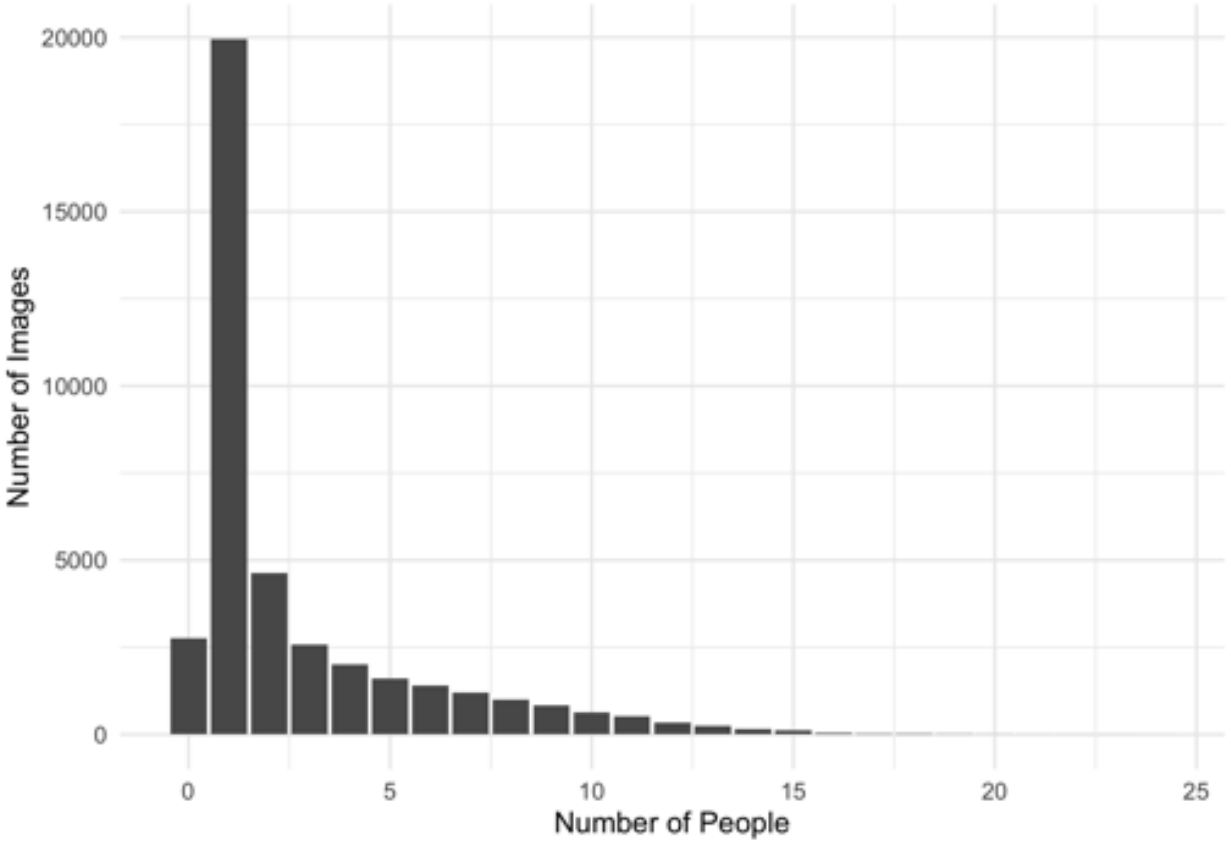
We have written the data analysis paper as a Rmarkdown file. This allows us to mix code, output, and text all within a single document. The goal of this paper is to illustrate examples of how the annotations produced by the computer vision algorithms can be used for access and analysis along with the ADDI Prototype. The easiest way to enable future work to extend off of these idea is, in our opinion, to provide full working code that others can run, adapt, and modify. A full output of the analysis produced by the document is provided in two formats. One version is given as an HTML file with embedded figures; this version contains a visible version of the R code for each analysis. A second version is given as a PDF file and contains only the text and output.

The remainder of this document is organized into three discrete analyses. The first looks at the detection and classification of portraits in the Bain Collection. We investigate how to detect the orientation of portrait photography. Next, we look at how image embedding-based recommendations function to connect across the five collections. Finally, we conclude by showing how the classification of photography as indoors and outdoors provides a useful classification for studying the composition of documentary photography. These three analyses are far from an exhaustive study of all the possibilities afforded by these datasets. We have simply made an attempt to select a variety of tasks that show the potential for looking at several annotation types while also looking at different parts of the corpus.

Detecting Portraits in the Bain Collection

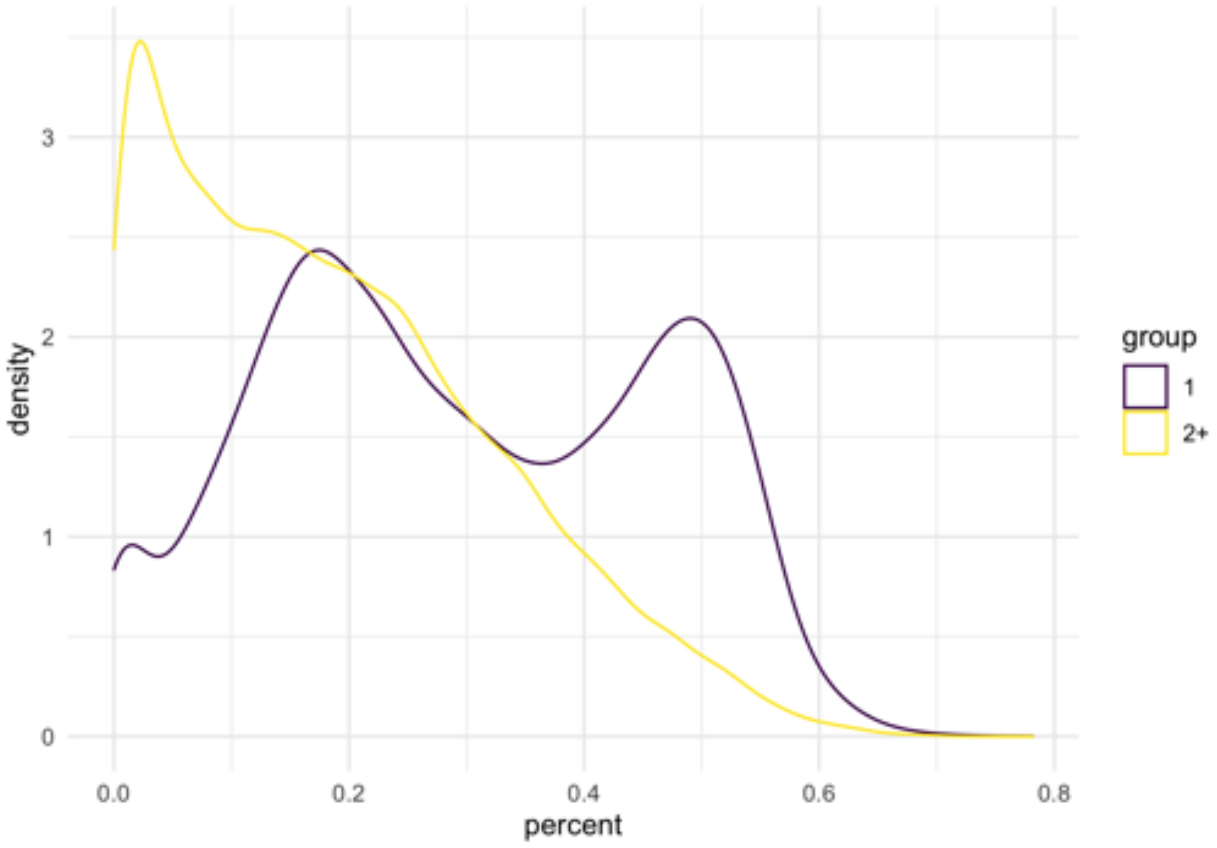
The George Grantham Bain Collection consists of nearly 40,000 black-and-white photographs taken by one of the earliest news picture agencies in the United States. Most images come from the first two decades of the 20th century. They document a wide range of activities, including quotidian scenes of shops, gas stations, and lunch counters, major political rallies, University football games, weddings, and funerals. One type of image that we found to be particularly prominent when browsing through the collection are formal portrait photographs. There are clearly many of these in the collection, but they are not directly identified by metadata fields or a consistent description in the photograph titles. Our goal in this section is to determine how we can use the computer vision annotations to identify and describe the the portrait photographs found within the Bain collection.

We start by looking at the results of the region segmentation algorithm. As described in the Methods Paper, this algorithm attempts to associate each pixel in an image with a object type or background region (such as the ground or sky). Though technically not an "object" per say, one of the object types detected by the algorithm are people. Below, we will look at the number of images in the Bain collection based on the number of people detected by the region segmentation algorithm.



Two interesting thing stand out with the above graphic image. First of all, notice that there are very few images that contain no people. This leads us to assume that there are not many images that contain only the built or natural environment. It also indicates potential patterns about how Bain visually defined news, as at least including and potentially centering people. Secondly, we see that images with a single person are a clear outlier in this plot. There are far more images with one person compared to any other number of people. These, likely, are where most of the portrait photographs can be found.

To investigate further, we need to understand more about the people detected in the images. We can do this by looking at the proportion of the image frame that is taken up by people. Specifically, we will look at a density plot of the proportions of the images taken up by people based on whether there is only one person or multiple people. Images without people are excluded. Our primary interest is the shape of the images with one person; the other images will help as a point of comparison. The density plot is shown below.



Interestingly, we see that for images with only one person, the proportion of frame image taken up by the person concentrates around two different values. One is around about 20% of the image and the other is around 50% of the image. As a comparison, notice that the density curve for images with two or more people has a sharp peak around 3%, with a steady decrease for larger proportions.

To understand more, we will look closer at the images with a single person from each of the modes of the density plot. Note that the process of moving back and forth between aggregating the data and looking at individual images is a common and fruitful mode of analysis throughout our work. Here, we start with 20 random photographs that have a single detected individual that takes up between 15% and 20% of the image frame.

Now, for comparison, we select 20 random photographs that have a single detected individual that takes up between 45% and 50% of the image frame.

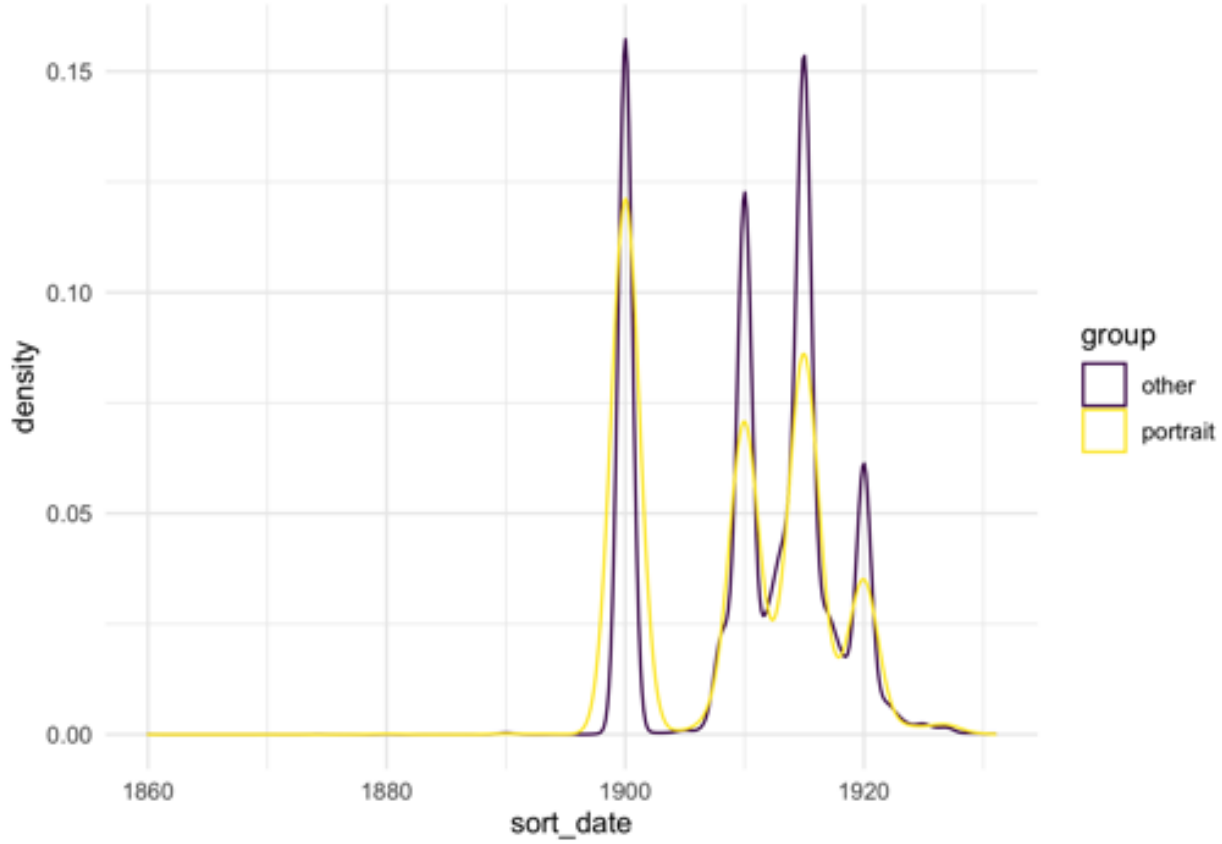
Looking at the images, we can now understand the difference between these two types of photographs containing a single detected person. In the first mode, most of the images feature the entire body of a single person shown in an interesting place. For example, a baseball player on the field, a man in a radio station, or a woman in front of a sewing machine. In contrast, the second mode primarily contains portraits of people shown from the chest upwards. The majority of these images appear to be shot in a studio setting and a neutral background. The person in the frame is often dressed in a formal wear or an official uniform. They seem to be looking directly at, or just slightly off, the camera.

It seems, then, that we can use the number of people (1) and the percentage of the frame taken up by a person (around 50%) to determine if an image in this collection is a studio portrait. However, note that these general patterns of the two modes are generally accurate but not perfect. One image in the second mode is a interestingly framed image of a man playing the piano; two of the images in the first mode are full-length shots taken in a studio setting. So, in using these derived annotations, we should keep in mind that there will be some errors. This does not stop us from using the results in aggregate or as a general method for search and discovery. However, if we are to add this information directly into the archival metadata tag directly, we

will want to be clear that the categorization is algorithmically generated.

The analysis also opens up interesting questions about news and visual culture. Are their social roles that are documented in certain ways compared to others? Our initial data suggests that activities such as dance and sports (which we could generalize to a category called performers) as well as women more generally are often photographed with visual information to clearly communicate to the audience the role of the person. In other words, the scene is their skill and helps the viewer understand why they are being featured. On the other hand, the studio portrait with a neutral or decorative background draws the eye to the face and clothes. There is little extra information to indicate exactly who the person is. Like the portraits that line government buildings with a name engraved in brass, the style of the portrait is designed to convey the person's prominence. It appears that certain roles in society such as military officials and politicians are being granted the visual power and cultural prominence of the close up. There is significantly more analysis to bolster this initial observation, but the initial differences are opening up questions and potential (historical) patterns regarding the relationship of framing, social position, and power.

Now that we have identified the set of portrait photographs, what can we do with them? From an access perspective, we could identify these photographs and create an exhibit or digital public project focused specifically on these images. As a form of analysis, we might try to identify how other archival metadata compares to portrait photographs. As one example, we can look at the distribution of photographs from the Bain collection by time based on whether an image is a portrait or not. Below is a density plot of these results.



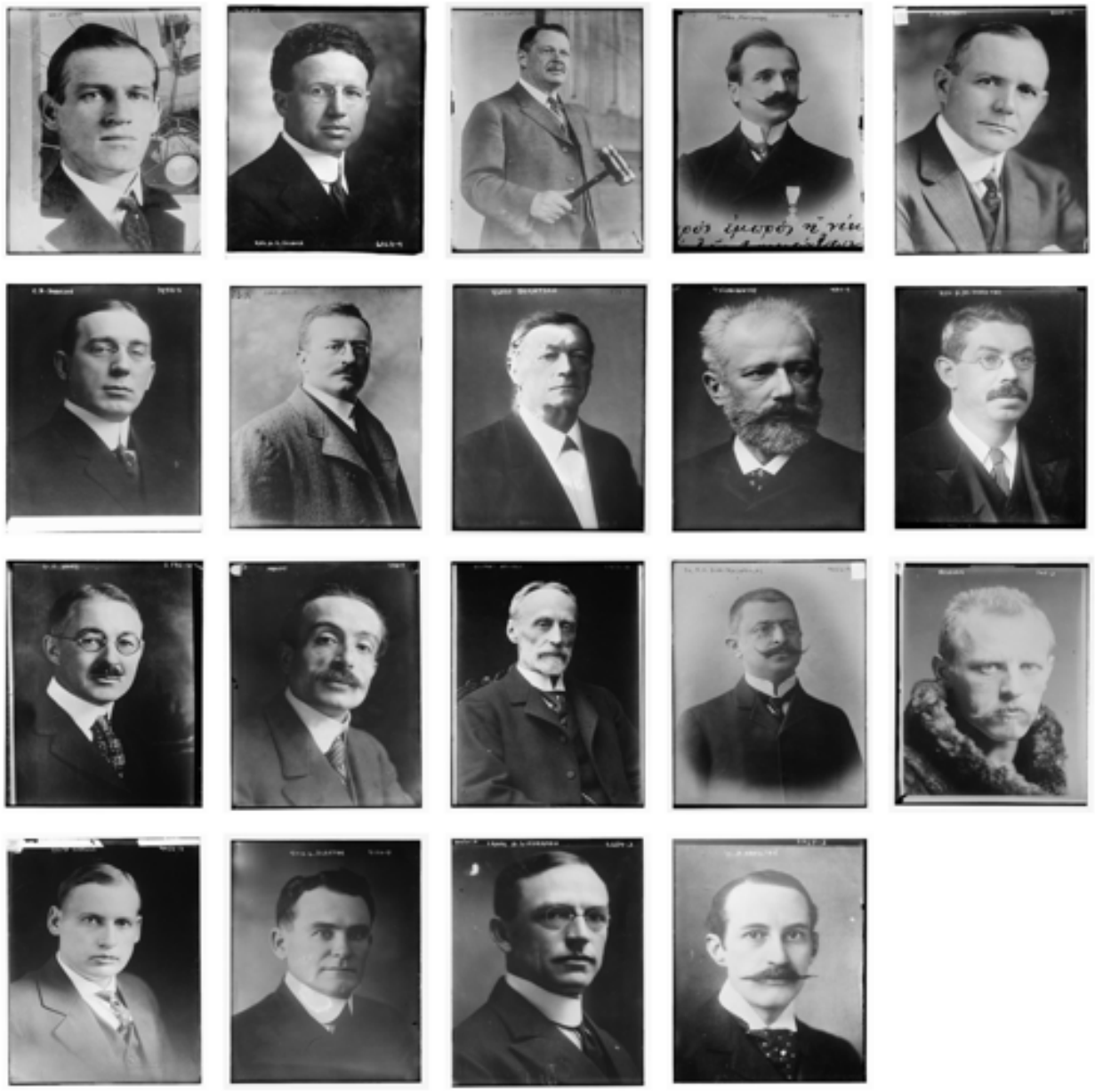
Looking at the density plot shows that the sort dates of photographs seem to cluster around "round" dates, such as 1900, 1910, 1915, and 1920. This is likely an artifact of the data collection rather than an interesting feature of the data itself. Unfortunately, a significant portion of the collection was lost in a fire. Specifically on the topic of the portrait photography, what is most interesting is that what we have tagged as studio portrait photographs seem to be equally distributed across the same time periods of the rest of the collection. So, the set of portrait photographs are an important element of the Bain collection through the early 20th

century rather than being a feature of only a few years.

Finally, another way of looking at the portrait photography in the Bain Collection is by using the pose detection algorithm to estimate the orientation of people looking at the camera. This will help us understand if people are rotated to the left, right, or squarely looking into the camera. The pose detection algorithm (further described in the Methods Paper) can help with this by allowing us to compare the position of different body parts relative to one another. As a starting point, we compared the distance between one's nose with their right and left ears. Based on which ear is closer in two-dimensional space to the nose is a way of detecting how the face is framed relative to the camera. Applying this algorithm to the portraits in the Bain Collection shows that there seems to be no particular preference for poses to the left or right.

```
##
##  pose left pose right
##        147         132
```

In order to understand these results, let's look at some of the images based on their pose. Below are 20 randomly selected images that appear to be posed to the right.



For comparison, below are 20 randomly selected images that appear to be posed to the left.

Looking at these images we can see that they do seem to correctly identify the orientation of people's faces. However, our algorithm only uses to the location of the ears and therefore is unable to to detect which way people's actual eyes are being directed. One trope we see in the above images is that may poses have one's face directed to one side of the frame, but their eyes cutting across the frame in the other direction. Further exploration of this compositional pattern is necessary for it also has the potential to connect back to our earlier observations. We can see again that the portraits in this style are primarily men and many appear to be White, although we want to proceed with caution about assuming race by doing additional research. The analysis opens up more questions about the role of portraiture, gender, and race in early 20th century visual culture.

As a final step, we can repeat the same process using the locations of the eyes themselves relative to the location of the nose key points. Similarly, this algorithm does not display any strong preference for poses to the left or right.

##

```
##  look left look right
##         324         335
```

Looking at examples can once again be helpful. Below are 20 randomly chosen examples of poses based on the eyes to the right.

And in the final set, below are 20 randomly chosen examples of poses based on the eyes to the left.

Looking at these results, we see that the eye-based calculation does find poses which are more strongly oriented to one side of the image or another. In all of the example cases, we see that the entire head of oriented in the chosen directly. Still, several examples show people who are looking off across the camera with their pupils. This highlights that orienting a person one way while gazing across the frame of the image is a common element of these studio portrait photographs. Further close and computational analysis, as well as a refinement of the classification of portrait photography, in order to better understand this phenomenon is a planned topic for future work.

Comparing Recommendations Across Collections

In the second analysis, we turn to looking at the entire set of photographic collections. Our goal here is to understand how the image embeddings create connections across the different collections of images. This will both help us describe the visual similarity between collections as well as highlight the behavior of a recommendation system designed around the embeddings. For more details about image embeddings and

how they are used to measure the similarity between any two images, please see the Methods Paper.

As a starting point, it will be helpful to see a few examples of the nearest neighbors (most similar photos) from the image embedding algorithm. Below are eight randomly selected starting images and their five closest neighbors based on the image embedding algorithm.

Overall, the recommendations seem to find images that have a close resemblance to the starting image. In some cases, such as the second-to-last row, the algorithm seems to find photographs from the same time and locations from slightly different angles. In other cases, it seems to pick up on the objects in the images, such as the cars in the first and sixth rows. For others, it is the framing of the image as well, as the images of men in the woods (row 4) and the full-frame images of men in uniform (row 8). Finally, we see that in rows 3 and 5, the recommendations are at least partially influenced by the quality of the images. In both of these rows, the recommendations all have similar defects (possibly water damage) that contribute to the similarity between the records. The method opens up a variety of analysis including identifying themes, a set of images from the same shooting assignment, and material histories of photography and the archive. For more examples of the recommendation system at work, we recommend looking at the interactive visualization that was also part of the ADDI project. It can be found here:

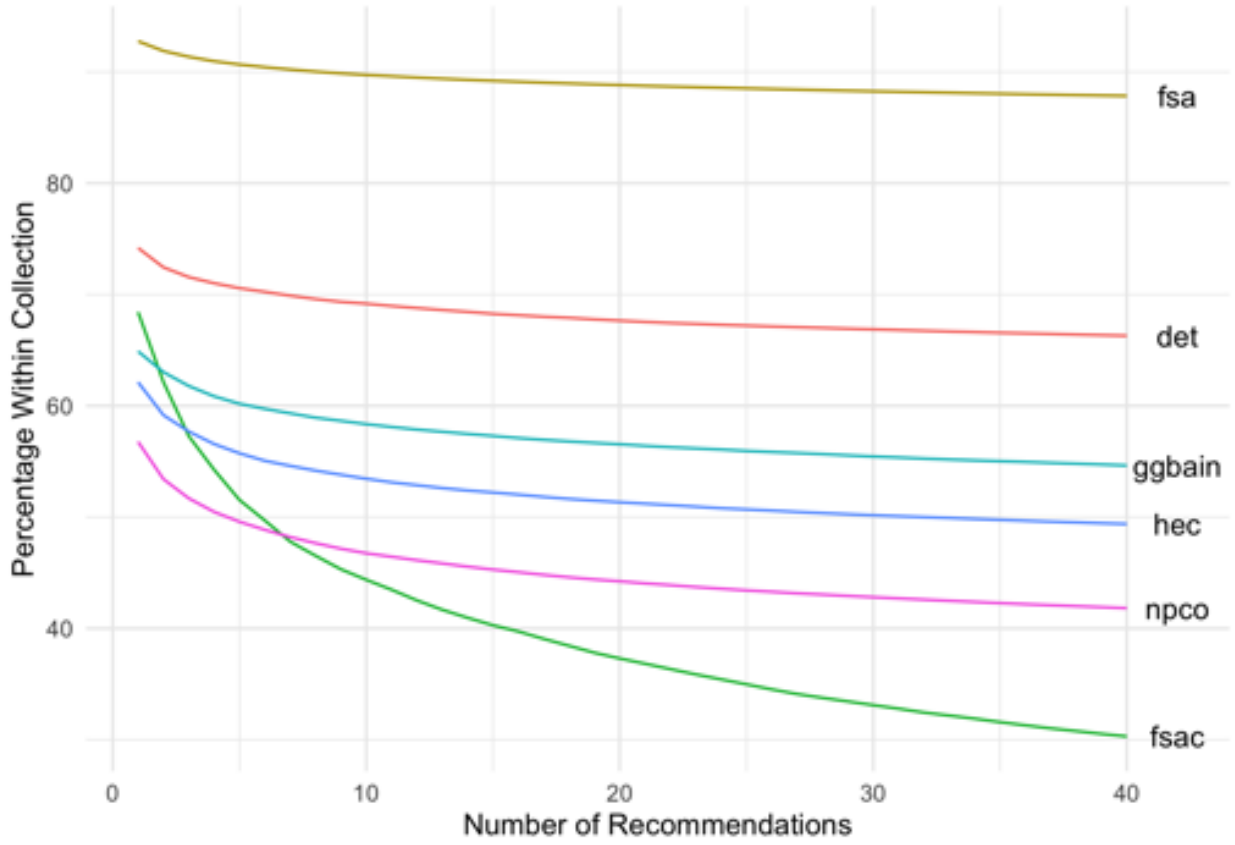
- https://distant-viewing.github.io/addi/06_interactive_viz/build/?id=2014712029

Now that we have some understanding of the recommendation algorithm, we will investigate the types of connections it makes across different collections. We will start by taking 10 recommendations from each image in our dataset and showing the percentage of the recommendations from one collection that are made to another collection. For the purpose of this table, and the other analyses below, we follow the LoC's organizational logic by grouping the FSA color photographs as a different collection from the black-and-white photographs.

```
## # A tibble: 6 x 7
## # Groups:   collection [6]
##   collection   det   fsa  fsac ggbain   hec  npco
##   <chr>      <dbl> <dbl> <dbl>  <dbl> <dbl> <dbl>
## 1 det           69    10     0      7     5     8
## 2 fsa            1    90     0      1     4     3
## 3 fsac           4    42    44      2     6     2
## 4 ggbain         6     6     0     58    14    15
## 5 hec            4    17     0     10    53    15
## 6 npco           7    14     0     14    18    47
```

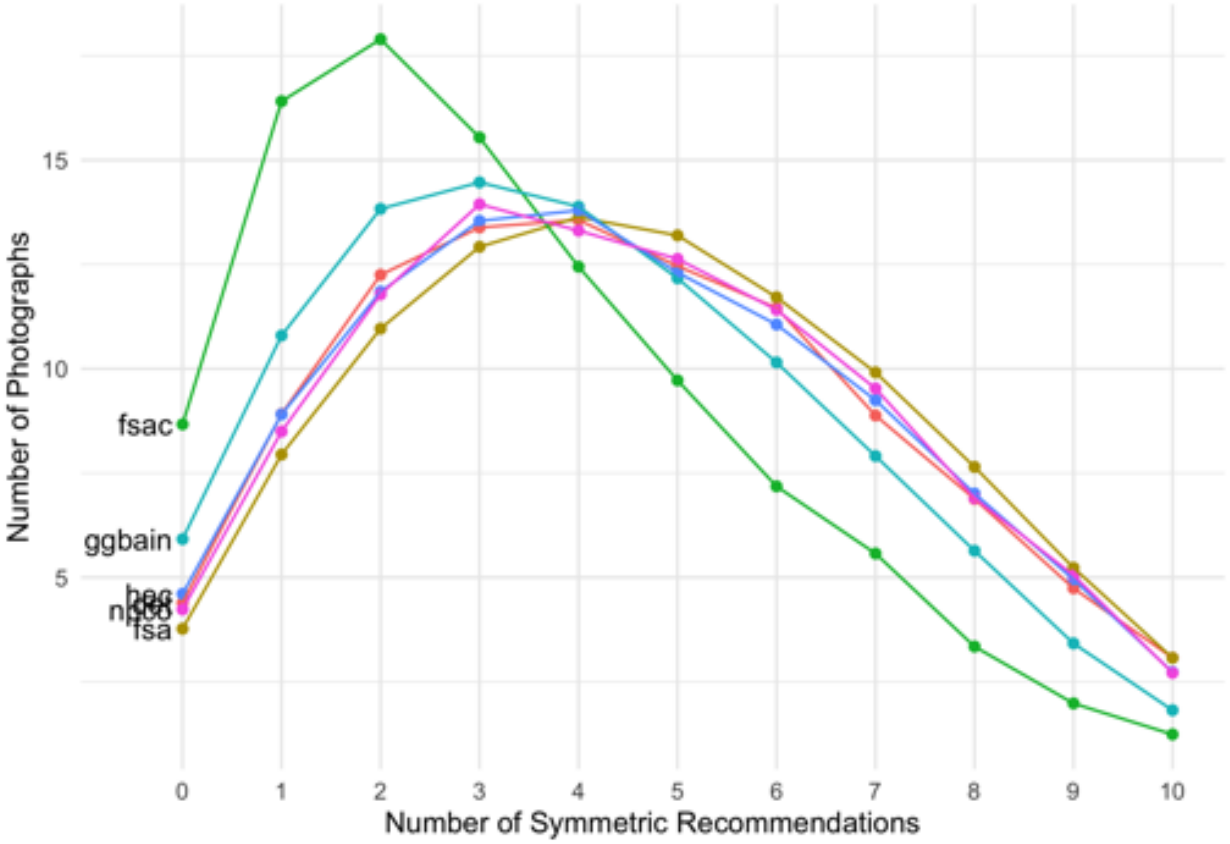
The results show that in each case an image is more likely to be recommend to another one in the same collection. However, there are also a large number of cross-collection recommendations. Notice that the color FSA images are almost as likely to be recommended to FSA black-and-white images as they are to other color ones. This is a great indication that the algorithm is often able to focus on thematic and compositional features (which should be similar across these two groups) rather than just a focus on simple textures and colors (which would favor making all the color images near one another). A next step would be to understand what makes these collections so connected. Is it primarily the topic of the image (such as farming and agriculture)? Are there compositional elements that indicate a a certain style and therefore speak to how the images work in visual culture? Other than the two FSA groups, there are no particularly strong patterns of recommendations across the collections.

The table above is based on using 10 neighbors for each image. We can plot the percentage of recommendations made within a collection as a function of the number of neighbors chosen. We will look at this for each of the collections and see how the numbers change.



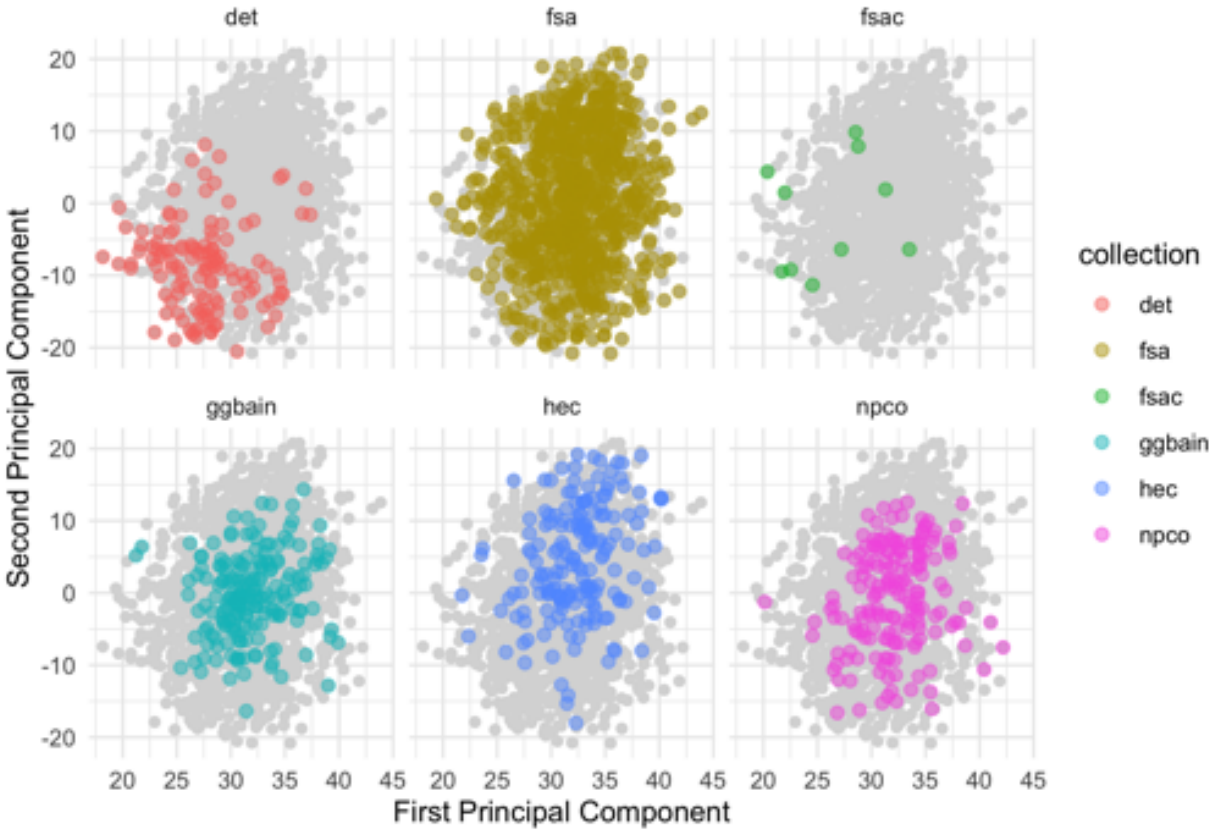
We see that the pattern relative to changing the number of neighbors for most of the collections follows a predictable pattern: the number of links within the collection slowly decreases as the number of neighbors increases. However, the decrease is slow and the ordering of the five larger collections is stable regardless of the number of neighbors chosen. The color FSA images have a slightly different pattern. The number of connections within the collection decreases rapidly (from the other table, we know this is towards recommendations to the FSA black-and-white images) as the number of neighbors increases. One hypothesis for the reason for this is that the images were often taken in sets, so there are a few images that are all very similar (the first few neighbors) but after this many of the connections are thematic and therefore as likely to be a black-and-white FSA image as a color one. Also, there are far fewer color FSA photographs (1,615) than any other collection (the next smallest is the Detroit Publish Company's 25,169 photos).

Another way to understand the structure of the recommendations is to look at the symmetric nearest neighbors. These recommendations are based on taking the K top nearest neighbors for each image and then making recommendations only on those recommendations where two images recommend each other. This is not particularly useful to be implemented directly as a recommendation system because (i) we usually want a fixed number of recommendations for each item and (ii) the goal is to *not* get stuck in loops. However, it is a useful metric for understanding the types of recommendations being made as well as a very useful tool for clustering analysis. Below is a figure that shows the distribution of the number of links that are symmetric within each collection if we choose 10 starting recommendations.



This data again shows how uniform the structure of the black-and-white collections are and the slight differences with regards to the small datasets of color images. Generally, though, most images have at least a few symmetric recommendations and a few asymmetric recommendations. This indicates that the recommendations are finding meaningful relationships (otherwise, the recommendations would look like noise and have very few symmetric recommendations) without being too stuck in small loops of very similar images (otherwise, most links would be symmetric). The results indicate that building a recommender system using image embeddings is an exciting way to move across the collections in meaningful ways.

As a final analysis of the embedding, we can plot the first two principal components of the image embedding and show where each collection is in the space. Below is a plot showing the first two dimensions using a random sample of the data in order to not crowd out the larger story.



We see, in line with the result above, that there are no clear clumps of the images and most are contained in a tight, dense point cloud. There are some edges associated with certain collections, most clearly the Detroit Publishing Company. This approach affirms above and the meaningful use of a recommender system to move across collections. A further understanding of these embeddings is another line of future work and will involve developing a better understanding of each of the clusters and regions within this plot.

Indoor and Outdoor

Our final analysis in this report investigates the use of image region segmentation to determine whether an image was taken indoors or outdoors. We focus on these two categories because they are a common way to categorize photography, and therefore a way of looking for images that is popular. It is also a helpful step for further refining additional categories that we might want to subset and recommend such as events in a town (i.e parade or march), the domestic sphere (i.e. dining room or living room), conflict (i.e. battles or damage) and the environment (i.e. mountains or oceans). We will start by looking at a single collection (Bain) and then proceed to see the distribution of indoor and outdoor images across all of the collections.

The image region segmentation algorithm we applied to each of the images attempts to associate each pixel in an image with either an object, person, or background region of "stuff". The latter includes regions of uncountable elements such as the sky, ground, or water. One helpful element of the region segmentation algorithm is that stuff categories are arranged hierarchically. All of the stuff categories are grouped into "indoor stuff" or "outdoor stuff"; these are further divided into several subcategories. Below are printed the granular categories along with their aggregated groups and classes.

```
## # A tibble: 54 x 3
##    class        group       super
##    <chr>        <chr>       <chr>
##  1 bridge       building    indoor
```

```
##  2 building     building    indoor
##  3 house        building    indoor
##  4 roof         building    indoor
##  5 tent         building    indoor
##  6 ceiling      ceiling     indoor
##  7 floor        floor       indoor
##  8 floor-wood   floor       indoor
##  9 food         food        indoor
## 10 fruit        food        indoor
## 11 cabinet      furniture   indoor
## 12 counter      furniture   indoor
## 13 door-stuff   furniture   indoor
## 14 light        furniture   indoor
## 15 mirror-stuff furniture   indoor
## 16 shelf        furniture   indoor
## 17 stairs       furniture   indoor
## 18 table        furniture   indoor
## 19 cardboard    rawmaterial indoor
## 20 paper        rawmaterial indoor
## 21 banner       textile     indoor
## 22 blanket      textile     indoor
## 23 curtain      textile     indoor
## 24 pillow       textile     indoor
## 25 rug          textile     indoor
## 26 towel        textile     indoor
## 27 wall         wall        indoor
## 28 wall-brick   wall        indoor
## 29 wall-stone   wall        indoor
## 30 wall-tile    wall        indoor
## 31 wall-wood    wall        indoor
## 32 window       window      indoor
## 33 window-blind window      indoor
## 34 dirt         ground      outdoor
## 35 gravel       ground      outdoor
## 36 pavement     ground      outdoor
## 37 platform     ground      outdoor
## 38 playingfield ground      outdoor
## 39 railroad     ground      outdoor
## 40 road         ground      outdoor
## 41 sand         ground      outdoor
## 42 snow         ground      outdoor
## 43 flower       plant       outdoor
## 44 grass        plant       outdoor
## 45 tree         plant       outdoor
## 46 sky          sky         outdoor
## 47 mountain     solid       outdoor
## 48 rock         solid       outdoor
## 49 fence        structural  outdoor
## 50 net          structural  outdoor
## 51 river        water       outdoor
## 52 sea          water       outdoor
## 53 water        water       outdoor
## 54 things       things      things
```

One thing we notice about these categories is that outdoor things tend to only occur outside. They consist of natural elements such as "snow", "flowers", and "grass" as well as elements of the built environment such as "pavement" and "railroad". These elements, in almost all cases, will always be outside. The indoor categories are more difficult to place. They include elements such as "fruit", "house", and "towel" that *could* be inside, but will also be commonly photographed outside ("bridge" is a strange outlier in the indoor category as it will usually be found outside). Our approach to this unbalanced nature of the categories is to use the following algorithm to classify images: if at least 10% of the image contains stuff from the outdoor category tag it as being outdoors; otherwise tag it as "other". We will not definitively call the "other" category indoors because, through experience (as we show below) we know it will make many false negatives.

Let's apply this logic to the photographs in the Bain collection. Looking at the output of this method, we see that roughly half of the images are classified as "outdoor" and half are classified as "other".

```
## # A tibble: 2 x 2
##   group       n
##   <chr>   <int>
## 1 other   22034
## 2 outdoor 18012
```

In order to understand how well this algorithm performs, we will look at some sample images tagged by the algorithm. First, we will show some sample images that have been labelled as outdoors.

All of these images do appear to be images that are taken outside. Further, they show a variety of different types of outdoor scenes. These range from sports games, to parades, scenes of destroyed buildings, and boat on the water. Similarly, we can look at a set of images tagged as “other”.

The majority of images in the other category are correctly indoors. However, there are some mistakes. In all of these cases there are tightly framed people against the backdrop of a building. We see neither the sky, nor the ground, nor any other regions of natural materials. Therefore, according to the limited categories of the segmentation algorithm, there is no way to be able to identity them differently from a tightly framed photograph taken indoors. This uncertainty is why we choose to label them as "other", though for the purpose of aggregation, the other category is associated with indoor photographs sufficiently well to be able to see general trends.

As an example of what we might be able to learn about collections of documentary photography based on the proportion of outdoor photographs, we will apply this algorithm to all of the collections in our dataset and report the percentage of images tagged as "outdoors" within each.

```
## # A tibble: 6 x 2
##   collection prop_outdoor
##   <chr>             <dbl>
```

```
## 1 hec                42.2
## 2 ggbain             45.0
## 3 npco               62.2
## 4 fsa                66.1
## 5 fsac               79.0
## 6 det                89.1
```

The results show that this is a strong indicator of the differences between the collections. The Harris and Ewing Collection, for example, has only 42% of its images tagged as being outdoors. In contrast, over 89% of the Detroit Publishing Company's images are classified as being outside. Understanding why this difference exists is one way to look closer at what was considered newsworthy in D.C. and Detroit.

We plan to extend the analysis of indoor and outdoor photography in our future work. Specifically, we plan two types of extensions. First, we would like to produce a custom annotation algorithm for better classifying indoor and outdoor images that goes beyond (or builds off of) the region segmentation. As a fundamental way of describing the context of an image, knowing whether an image was taken inside or outside is an important task for the analysis of cultural heritage collections but has not been of much interest in the field of computer vision. It therefore seems like a good candidate for such as custom model. In addition, as a second extension, we hope to apply an even more accurate algorithm to further detect clusters within and across each collection of images.

Closing

The data analyses presented here are just the beginning of the use of computational methods to further study and increase access to these five important collections. They delve into further details about how the specific algorithms in the Methods Paper can contribute to the study of early 20th century photography from questions about content, themes, and tropes to composition, framing, and style. While computer vision offer a way of looking at scale, we aimed to demonstrate how toggling between looking closely at individual and small sets of photos as well as at scale alongside critical assessment of the algorithm is key to the analysis; an approach that we call distant viewing. In addition to this paper offering direct extensions of these three specific applications of the computer vision annotations to the five collections, we also are excited to see how these approaches can be applied to other collections guided by the cautions and care that is outlined in the Methods Paper. If you are doing this kind of work, please reach out. We are excited to see what is possible together.