DPLA photographs: exploratory workset

For this week's workset, I want to start on the collective endeavor of understanding what kind of photographs are in the DPLA's massive collection.

There are three routes into this; I want you to each do an individual part of finding one interesting subcollection, and then to write three *collaborative* posts.

Personal post:

Describe an interesting subset of images in the DPLA that might be suitable for some kind of large-scale analysis or enhancement in an exhibit.

Dig through and try to find a *collection* of images that might be interesting to look into more: speculate about them. It could be small (a dozen pictures) or huge (there seem to be tens of thousands of images of plants.) The important thing is that it be a fruitful set to:

1. Try to connect these images in some way, some how, to a historical question. What can these images in aggregate show?

2. Try and understand something about how they came to be digitized. Are there similar photos that weren't scanned, either within this institution or elsewhere? What are the holes and gaps in the archive?

Group posts

Let's also try to get up three collective posts. We'll create groups in class; I'd write these as a group in a Google Doc or something, and send them along.

Think about these as a resource for others approaching the problem in the future. Try to write them reasonably well; you include most images that you talk about.

How do neural networks see digital pictures?

Google has released code that "embeds" images into a multidimensional space. As shown in class, it's able to assign a variety of labels with some level of accuracy.

The task for this group is: look through a lot of these labels and reflect both philosophically and at a fine grained level at what the algorithms are good at seeing, and what they're bad at. What kinds of things should we be thinking about pulling out?

If you're looking at the labels, someone will have to do a little to look at the kind of labels that Imagenet uses; etc.

What is in the DPLA? How can you describe its contents?

There are something 5,000,000 images in the DPLA. What are they of? How many people's faces are there? What institutions contributed the most? What kind of things that should be included are not? Are there other big collections of photographs we should consider working with?

What sorts of questions should we ask of large collections of digital pictures, or how else should they be used?

This is the hardest question, but a big one; how can we think about the kinds of questions to ask about large collections of digital photographs? Are there places that change over time is a useful metric? What could be done with facial recognition? Etc. This is just sort of a brainstorming post.

The large here is critical. There have been lots of projects working with small collections of digital pictures, or pictures of a particular place. The question is–and this is a good one for historians–what do we need in terms of aggregate description, search, extraction, etc.?

What might be useful ways to read *horizontally* across collections? That use metadata that may not be applied consistently?

Try to drop in some citations to prior art of things people have done.