Enriching Historic Photography with Structured Data using Image Region Segmentation

Taylor Arnold[1] and Lauren Tilton[2]

25-26 May 2020

[1] Assistant Professor of Statistics and Linguistics
University of Richmond
statsmaths.github.io

[2] Assistant Professor of Digital Humanities
University of Richmond
laurentilton.com

Introduction

- Guiding question: How can we increase discovery and access of digitied photographic collections through structured data and the semantic web.

- Guiding question: How can we increase discovery and access of digitied photographic collections through structured data and the semantic web.
- Challenge: Lack of structured linguistic descriptions $\longrightarrow$ difficult to link/search between and across collections; limits computational analysis

- **Guiding question**: How can we increase discovery and access of digitied photographic collections through structured data and the semantic web.
- **Challenge**: Lack of structured linguistic descriptions $\longrightarrow$ difficult to link/search between and across collections; limits computational analysis
- **Manual approach**: manual data construction $\longrightarrow$ extensive resources and becomes more difficult as digitized datasets increase in size.

- Guiding question: How can we increase discovery and access of digitied photographic collections through structured data and the semantic web.
- Challenge: Lack of structured linguistic descriptions $\longrightarrow$ difficult to link/search between and across collections; limits computational analysis
- Manual approach: manual data construction $\longrightarrow$ extensive resources and becomes more difficult as digitized datasets increase in size.
- Our approach: Use computer vision techniques!

Challenges of using computer vision for automated data creation:

- built using modern datasets

Challenges of using computer vision for automated data creation:

- built using modern datasets
- may produce annotations that are inaccurate or inappropriate for historic data

Challenges of using computer vision for automated data creation:

- built using modern datasets
- may produce annotations that are inaccurate or inappropriate for historic data
- incorrectly extracted data records are particularly concerning when making data available to the public

Challenges of using computer vision for automated data creation:

- built using modern datasets
- may produce annotations that are inaccurate or inappropriate for historic data
- incorrectly extracted data records are particularly concerning when making data available to the public
- even when labeled as automated studies have shown that people have trouble accurately interpreting probabilistic data and are overly confident in predictions (Khaw, Luminita, Woodford, 2019)

Challenges of using computer vision for automated data creation:

- built using modern datasets
- may produce annotations that are inaccurate or inappropriate for historic data
- incorrectly extracted data records are particularly concerning when making data available to the public
- even when labeled as automated studies have shown that people have trouble accurately interpreting probabilistic data and are overly confident in predictions (Khaw, Luminita, Woodford, 2019)
- risk reinforcing racial, gender, and socioeconomic biases inherent in the training data behind machine learning techniques

Data

FSA-OWI

Apply to 1610 color photographs from the Farm Security Administration-Office of War Information Collection (FSA-OWI):

- it is a part of one of the most famous and researched photography archives from the United States

FSA-OWI

Apply to 1610 color photographs from the Farm Security Administration-Office of War Information Collection (FSA-OWI):

- it is a part of one of the most famous and researched photography archives from the United States
- held by a library that is invested in open access and encourages experimentation with their digital collections

FSA-OWI

Apply to 1610 color photographs from the Farm Security Administration-Office of War Information Collection (FSA-OWI):

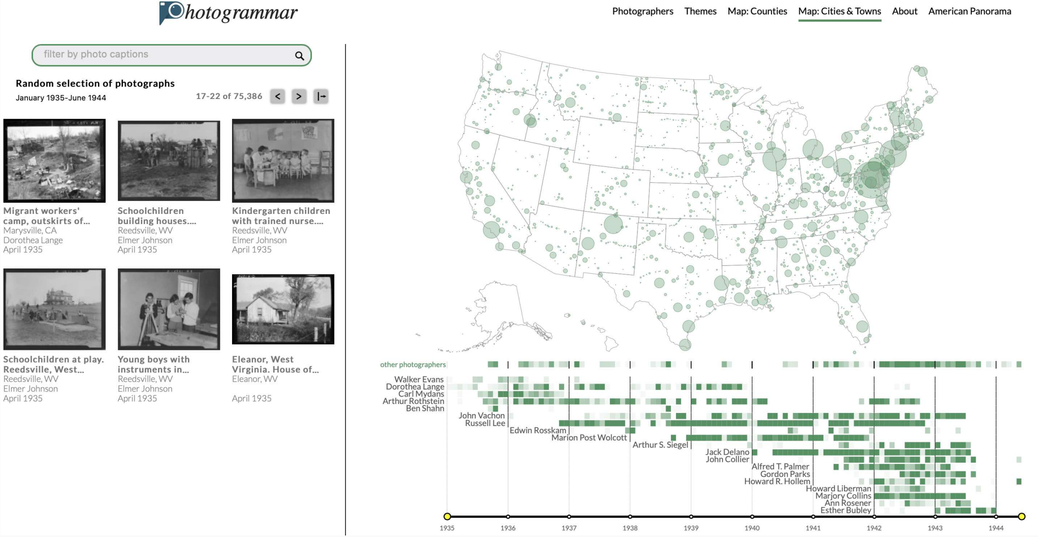
- it is a part of one of the most famous and researched photography archives from the United States
- held by a library that is invested in open access and encourages experimentation with their digital collections
- indicative of many documentary photography collections held in GLAM institutions

FSA-OWI

Apply to 1610 color photographs from the Farm Security Administration-Office of War Information Collection (FSA-OWI):

- it is a part of one of the most famous and researched photography archives from the United States
- held by a library that is invested in open access and encourages experimentation with their digital collections
- indicative of many documentary photography collections held in GLAM institutions
- a collection we have worked with through out Photogrammar project since 2010

Photogrammar



Previous Approaches

Computer vision methods for creating structured data from still photography:

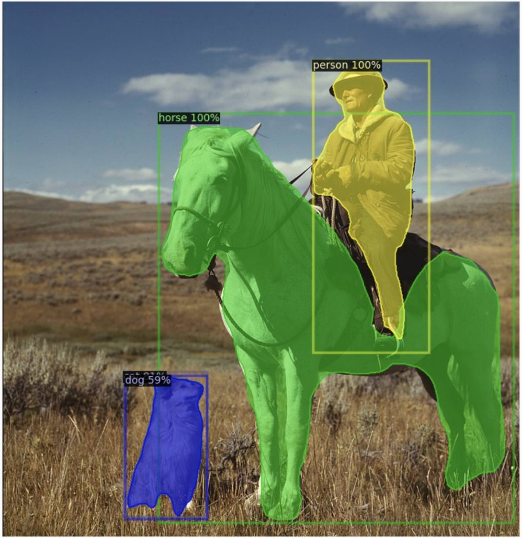
- object detection
- automatic free-form captions
- image embeddings

Object Detection

Mask R-CNN instance object classification algorithm (X101-FPN).

Wu, Y., Kirillov, A., Massa, F., Lo, W.-Y., and Girshick, R. (2019). *Detectron2.*

Object Detection



Object Detection

Object Detection

Automatic Captions

Show, attend and tell

Xu, Kelvin, et al. “Show, attend and tell: Neural image caption generation with visual attention.” *International conference on machine learning.* 2015.

Automatic Captions

a man is wearing a hat and a hat

a woman is sitting on a stage with a microphone

a man is sitting on top of a pile of oranges

a couple of giraffes are in a room

Image Embedding

VGG-19

K. Simonyan, A. Zisserman "Very Deep Convolutional Networks for Large-Scale Image Recognition." *arXiv:1409.1556*

Image Embeddings

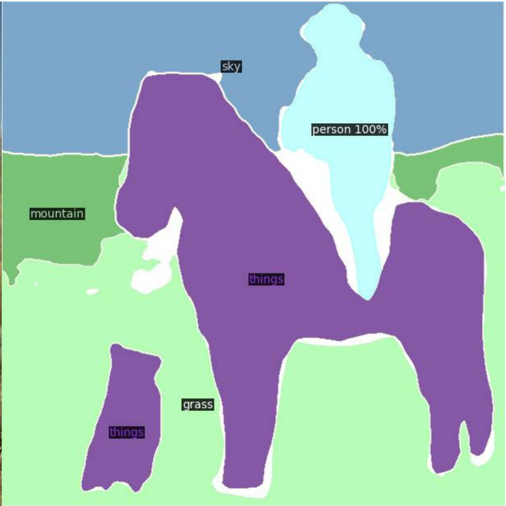
Image Segmentation

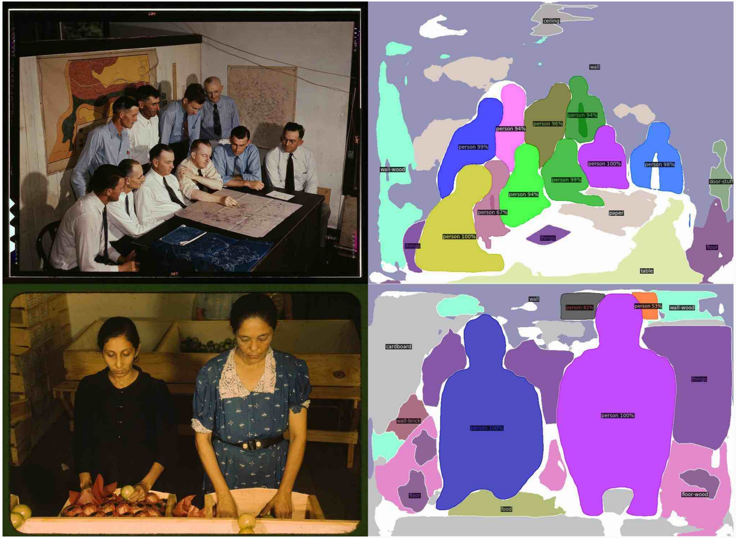
Image Embedding

COCO-stuff

Caesar, H., Uijlings, J., & Ferrari, V. (2018). "COCO-stuff: Thing and stuff classes in context." In: *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition* (pp. 1209-1218).

Group	Meta	Categories
indoor	ceiling	ceiling-tile
indoor	floor	floor-wood; floor-stone; floor-tile; floor-marble; carpet
indoor	food	fruit; vegetable; salad
indoor	furniture	cabinet; cupboard; counter; desk; door; light; mirror; shelf; stairs; table
indoor	rawmaterial	cardboard; metal; paper; plastic
indoor	textile	banner; blanket; curtain; cloth; clothes; napkin; mat; pillow; rug; towel
indoor	wall	wall-brick; wall-stone; wall-tile; wall-wood; wall-panel; wall-concrete
indoor	window	window-blind
outdoor	building	bridge; house; roof; skyscraper; tent
outdoor	ground	dirt; gravel; pavement; platform; playingfield; railroad; road; sand; snow; mud
outdoor	plant	flower; grass; tree; bush; leaves; branch; moss; straw
outdoor	sky	clouds
outdoor	solid	mountain; rock; hill; stone; wood
outdoor	structural	fence; net; railing; cage
outdoor	water	river; sea; waterdrops; fog


sky
person 100%
mountain
things
grass
things


ceiling
wall
person 94%
person 96%
person 94%
person 99%
wall-wood
person 100%
person 98%
door-stuff
person 99%
person 94%
person 67%
paper
things
person 100%
things
floor
table
wall
person 82%
person 53%
wall-wood
cardboard
things
person 100%
wall-brick
person 100%
floor
floor-wood
food

Creating Structured Data

Creating structured data

Our method:

- apply the image segementation algorithm
- record any detected ‘stuff’ region along with the percentage of the image covered
- record total percentage of ‘thing’ regions
- record number of ‘people’ regions

Web Annotation Data Model: Example I

```
<http://photogrammar.org/anno1> a oa:Annotation ;
  dcterms:creator <http://photogrammar.org/tarnold2> ;
  dcterms:created "2020-02-19T12:00:00Z" ;
  oa:hasBody [
    a                        pgram:ImageSegmentationRegion ;
    pgram:regionName         <http://example.org/stuff/things> ;
    pgram:regionPercent      32 ;
  ] ;
  oa:hasTarget <http://photogrammar.org/resource/1a35022v> ;
  oa:motivatedBy oa:tagging .
```

Web Annotation Data Model: Example II

```
<http://photogrammar.org/anno2> a oa:Annotation ;
  dcterms:creator <http://photogrammar.org/tarnold2> ;
  dcterms:created "2020-02-19T12:01:00Z" ;
  oa:hasBody [
    a                        pgram:ImageSegmentationRegion ;
    pgram:regionName         <http://example.org/stuff/people> ;
    pgram:regionPercent      6 ;
    pgram:regionCount        1 ;
  ] ;
  oa:hasTarget <http://photogrammar.org/resource/1a35022v> ;
  oa:motivatedBy oa:tagging .
```


snow
bridge
wall-brick
person


sky
food
mountain
railroad

Evaluation

Quantitative evaluation

Hand labelled accuracy of the annotations generated by each method.

	Accuracy	**Images Labelled**	**Unique Results**
Captions	31.5%	100%	1040
Objects	70.7%	37.3%	178
Stuff & People	97.5%	98.9%	1140

Future Directions

Future Directions

- further encode information about the detected regions ⟶ dominant colors of each region type; describe regions by quadrant
- develop a structured language for creating image captions from the structured data
- hierarchical version of a tagged object detection algorithm that simulates the stuff-based regions