Neural Networks: Machine Vision for the Visual Archive

Peter Leonard, PhD

Yale UNIVERSITY LIBRARY
Digital Humanities Laboratory

DHLab

Yale

Machine Vision
Use Case #1:

Given a picture of interest,
how can I see more like it?

Yale

Selected Projects in Image Similarity for Humanities / Art History

John Resig
TinEye MatchEngine (Commercial; Black Box)

- ukiyo-e.org: Japanese Woodprints
- PHAROS consortium of museum photo archives

Carl Stahmer
Arch Vision (Speeded up robust features; Bag of Visual Words)

- Early English Broadside / Ballad Impression Archive

Resig, John. "Aggregating and Analyzing Digitized Japanese Woodblock Prints." *Japanese Association of Digital Humanities conference*, 2013
Stahmer, Carl. "Arch-V: A platform for image-based search and retrieval of digital archives." *Digital Humanities 2014: Conference Abstracts*, 2014

Yale

Meserve-Kunhardt Collection

The Meserve-Kunhardt Collection originated as one of the **largest private collections** of 19th-century photography.

Now part of the Beinecke Rare Book and Manuscript Library, the archive of over 73,000 images is a world-renowned record of America from the **Civil War** to the **Gilded Age**.

Yale

Convolutional Neural Networks for Image Captioning

2012 — today

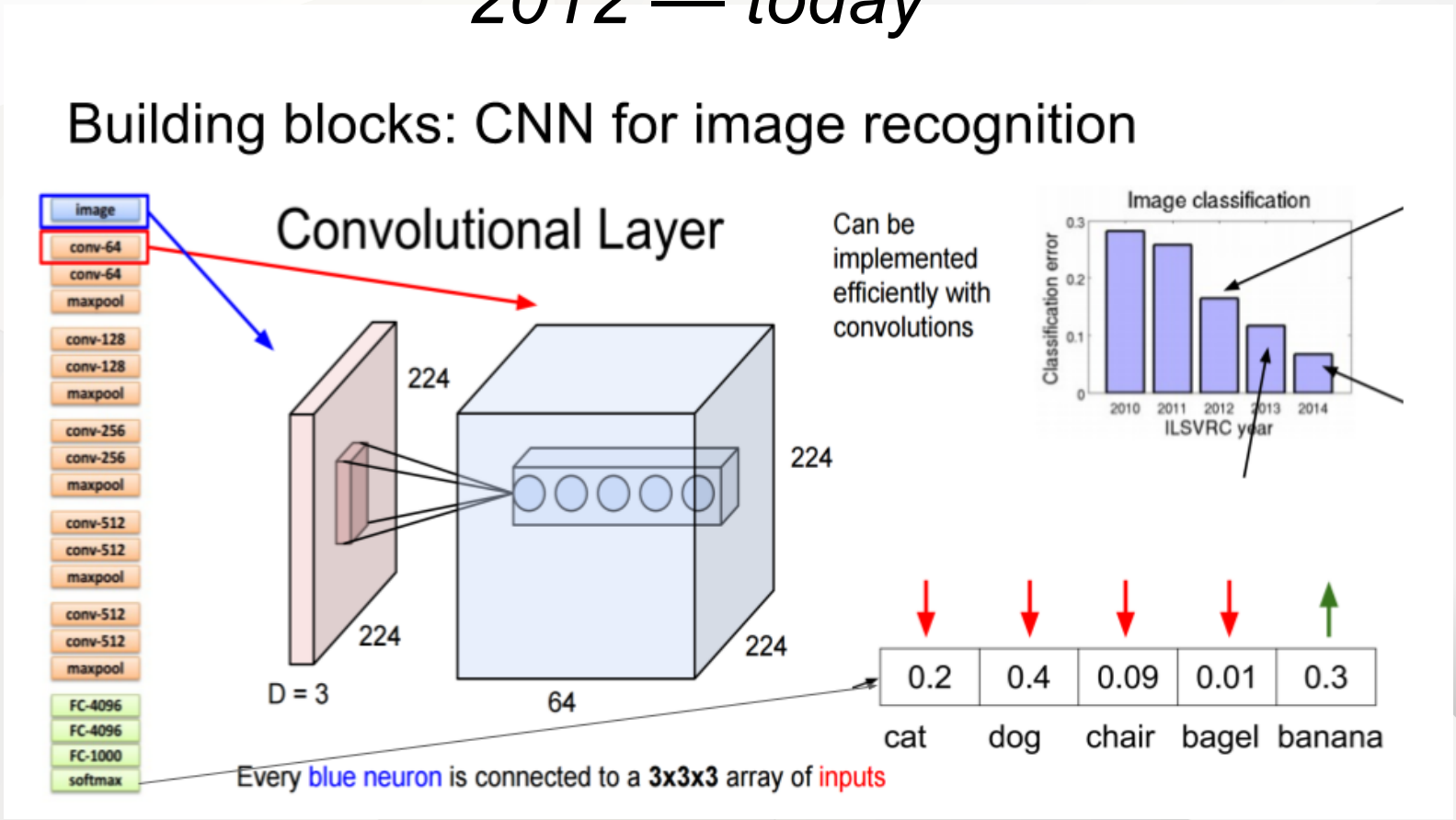
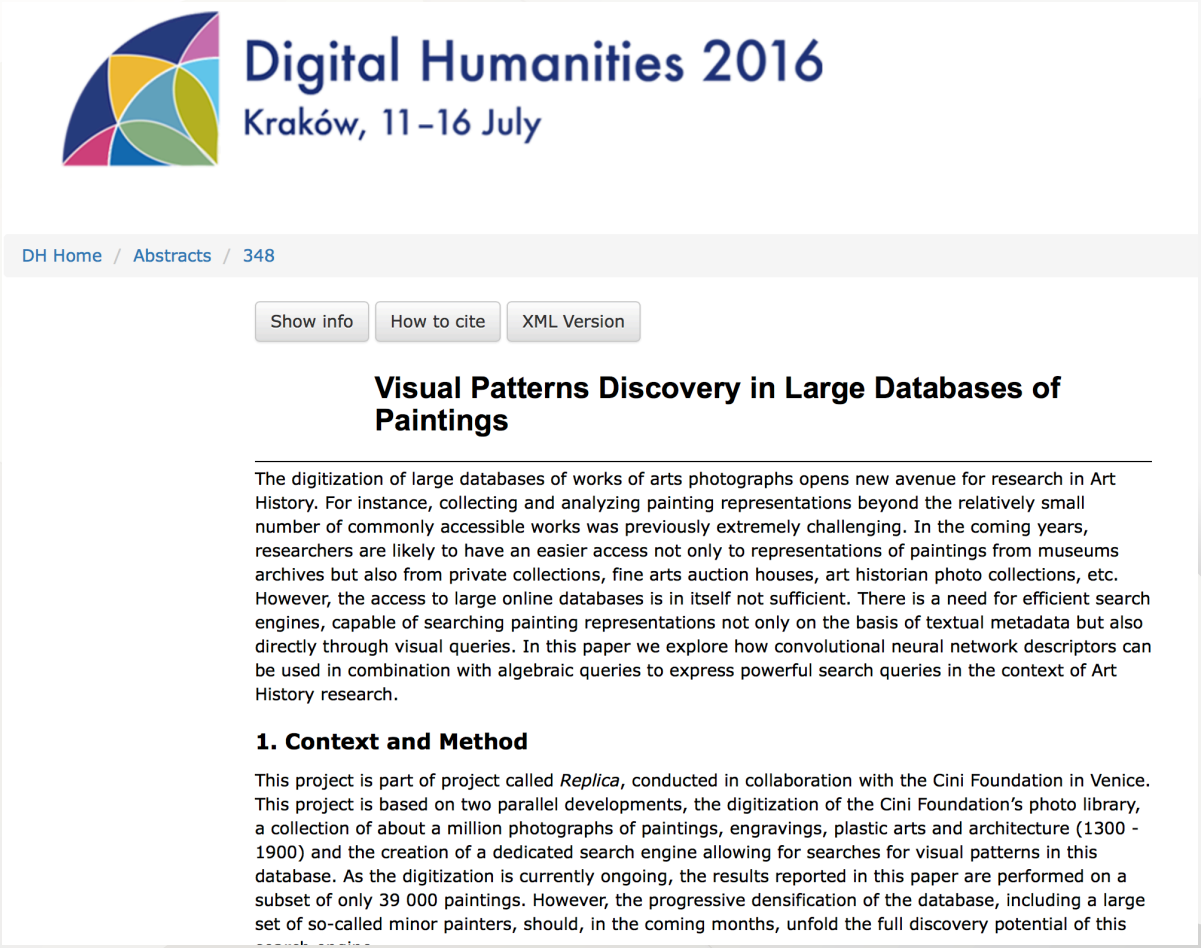


Diagram: Pankaj Kumar

Yale


Howe
ELLIOTT & FRY
Copyright
55, BAKER ST.
PORTMAN SQUARE
Sarony
680 BROADWAY, N. Y.
Nathaniel Hawthorne

Penultimate Layer in Captioning Network: More abstract = more useful?



Digital Humanities 2016
Kraków, 11–16 July

DH Home / Abstracts / 348

Show info | How to cite | XML Version

Visual Patterns Discovery in Large Databases of Paintings

The digitization of large databases of works of arts photographs opens new avenue for research in Art History. For instance, collecting and analyzing painting representations beyond the relatively small number of commonly accessible works was previously extremely challenging. In the coming years, researchers are likely to have an easier access not only to representations of paintings from museums archives but also from private collections, fine arts auction houses, art historian photo collections, etc. However, the access to large online databases is in itself not sufficient. There is a need for efficient search engines, capable of searching painting representations not only on the basis of textual metadata but also directly through visual queries. In this paper we explore how convolutional neural network descriptors can be used in combination with algebraic queries to express powerful search queries in the context of Art History research.

1. Context and Method

This project is part of project called *Replica*, conducted in collaboration with the Cini Foundation in Venice. This project is based on two parallel developments, the digitization of the Cini Foundation's photo library, a collection of about a million photographs of paintings, engravings, plastic arts and architecture (1300 - 1900) and the creation of a dedicated search engine allowing for searches for visual patterns in this database. As the digitization is currently ongoing, the results reported in this paper are performed on a subset of only 39 000 paintings. However, the progressive densification of the database, including a large set of so-called minor painters, should, in the coming months, unfold the full discovery potential of this

di Lenardo, I., Seguin, B., Kaplan, F. (2016). Visual Patterns Discovery in Large Databases of Paintings. In *Digital Humanities 2016: Conference Abstracts*.

Yale

Neural Neighbors:
Pictorial Tropes in the Meserve-Kunhardt Collection

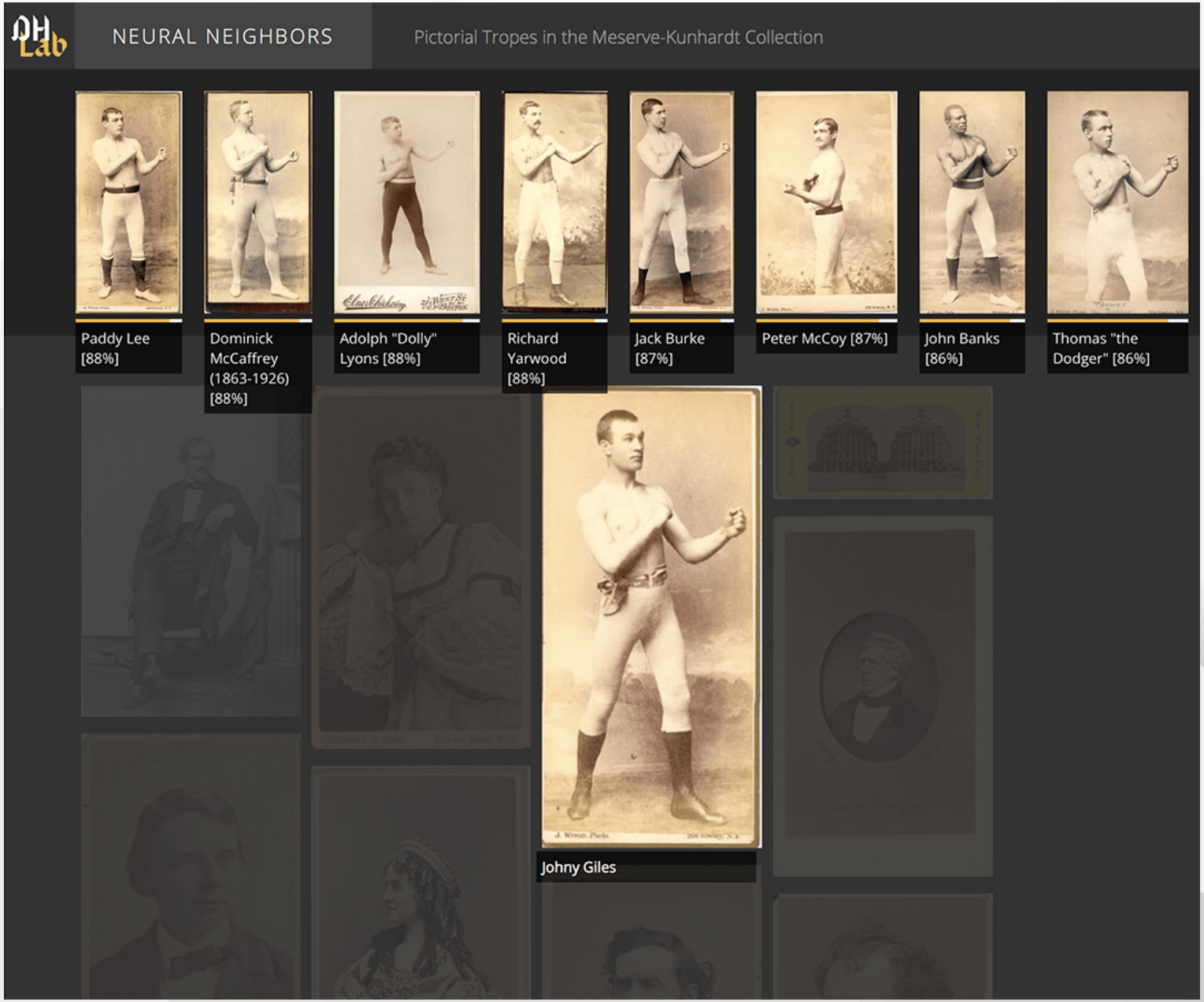
by Doug Duhaime, Monica Ong Reed & Peter Leonard

1. Take a **Convolutional Neural Network** designed for image captioning [Google's Inception]…
2. …**trained** on 1.2 million images [ImageNet 2012]…
3. …but extract the **penultimate layer** in the network [pool_3:0]…
4. …which contains a **more abstract** set of visual features [2,048 dimensions]…
5. …and find the **approximate nearest neighbors** of each image in this high-dimensional space [github.com/spotify/annoy]…
6. … there's no step 6.

yaledhlab.github.io/neural-neighbors/

An experiment with the Beinecke Rare Book & Manuscript Library

Yale


DH Lab
NEURAL NEIGHBORS
Pictorial Tropes in the Meserve-Kunhardt Collection
Paddy Lee [88%]
Dominick McCaffrey (1863-1926) [88%]
Adolph "Dolly" Lyons [88%]
Richard Yarwood [88%]
Jack Burke [87%]
Peter McCoy [87%]
John Banks [86%]
Thomas "the Dodger" [86%]
Johny Giles

Machine Vision

Use Case #2:

How can a large-scale visual collection organize itself?

Yale

PixPlot

by Doug Duhaime

- Process images with the *Inception* Convolutional Neural Network
  - CNN trained on ImageNet 2012
  - Penultimate layer yields 2,048 abstract feature dimensions
- Project images into two-dimensional manifold with UMAP
  - *Uniform Manifold Approximation and Projection*
  - Faster than t-SNE; better at preserving global structure
- Visualize 2-D projection using WebGL
  - Performant to > 100,000 images
- Similar images will appear nearby
- Use *k*-means clustering to find ‘constellations’ in the ‘galaxy’

github.com/YaleDHLab/pix-plot

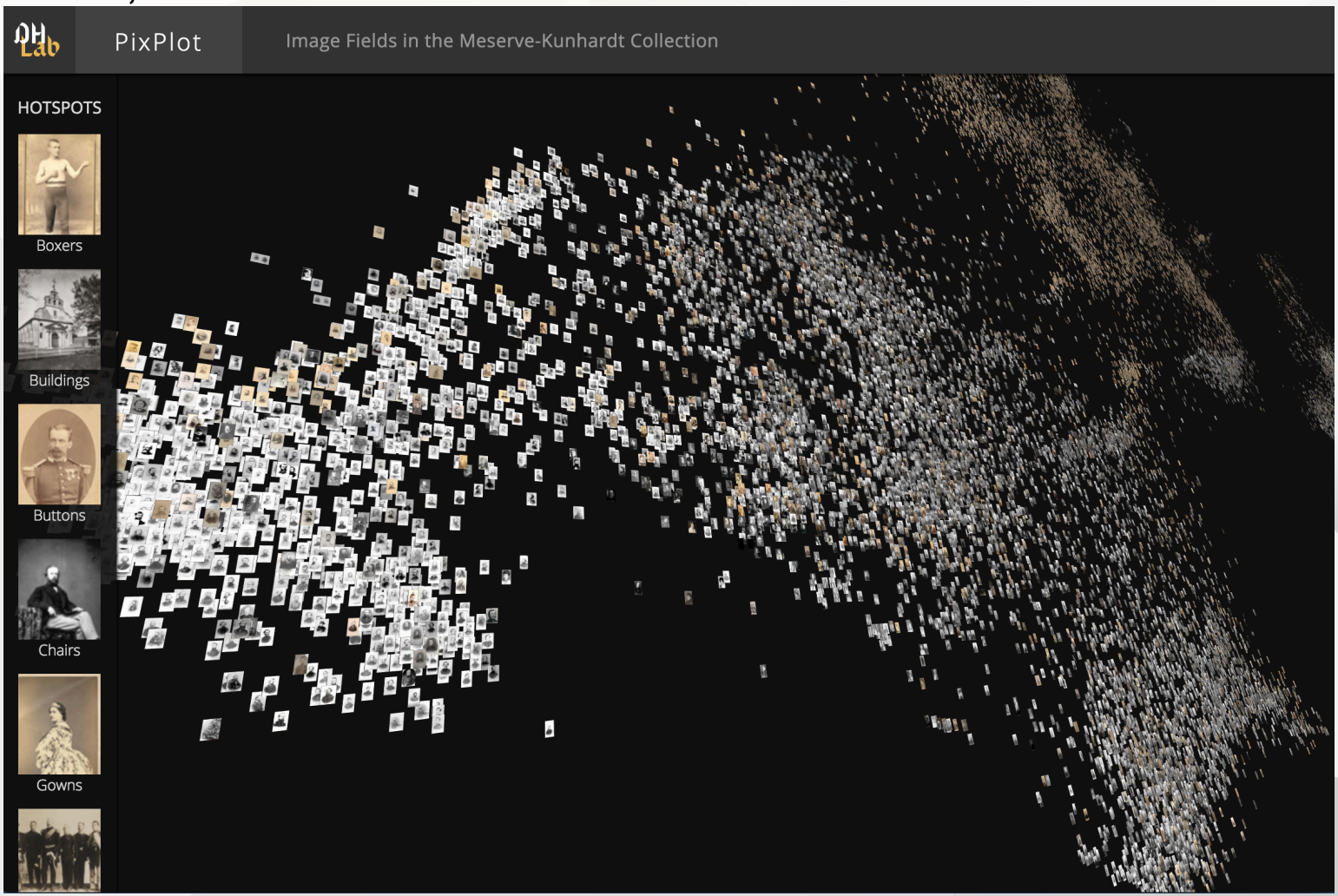
Special thanks to Cyril Diagne, Benoît Seguin and Damon Crockett

Yale

PixPlot: *Tens of thousands of images in your browser*

Corpus: Meserve-Kunhardt Collection

Size: 27,000

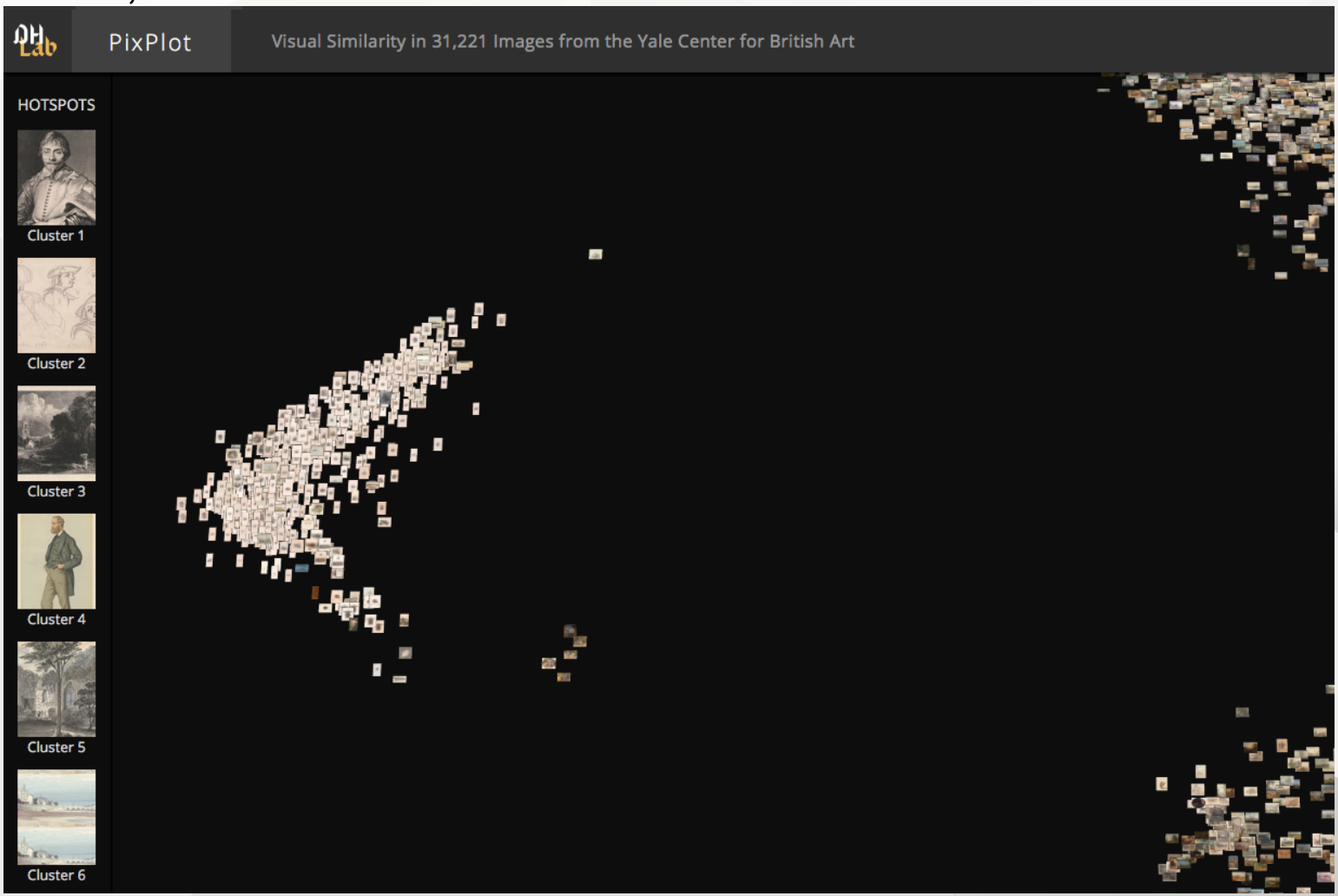


Yale

PixPlot: *Tens of thousands of images in your browser*

Corpus: Yale Center for British Art

Size: 31,221



Future Development Directions

- Deep links back to system of record
- Higher-resolution images at large zoom levels
  - IIIF (International Image Interoperability Framework)
- Animation between different states
  - t-SNE vs. UMAP, different layers in network…
- Curatorial Tools
  - Represent metadata facets with border colors
  - “Lasso tool” to allow sub-selections for curation
- Transfer Learning for domain-specific captioning
  - Retraining final layer with curated training data

github.com/YaleDHLab/pix-plot

Yale

Machine Vision

Use Case #3:

Can neural networks dream ?

Yale

Recurrent Neural Networks

```
     Title: CHINESE MEAT OF 2 OR. SALAD
Categories: Chinese, Candies, Fat
     Yield: 1 Servings

          Unsweetened chocolate canned
     1    Medium potatoes
          Chopped onions
          -or wine vinegar
     4 c  Sesame seeds
     6    Sprigs freshly ground black pepper
     2 tb Lemon juice
     2 tb Tarragon
     1 ts Soy sauce
     2 c  Butter or margarine
     1 c  Parsley, divided
     1 c  Diced cilantro

Prepare sour cream into egg coloring; drain first pans in a large bowl. Add the
the flour and seasonings and season with 1 teaspoon of
flour mixture.  Add spices, sugar, cinnamon, yeast, celery, tart cheddar cheese,
chili
meat and celery syrup, garlic and beat in eggs. Cover and simmer for 40
minutes or until cheese is creamy forms. Turn out on cookie sheet.
In a large bowl, combine with sour cream,
and salt and pepper. Remove from heat;
drain on lemon. When ice cream can be softened dumplings.
```

gist.github.com/nylki
github.com/karpathy/char-rnn

Generative Adversarial Networks

Consider two neural networks:

- The **Forger** looks at tens of thousands of 19th-century photographs and tries to **dream up new ones**.
- The **Detective** looks at all of these photographs (real counterfeit) and tries to **determine which are fake**.

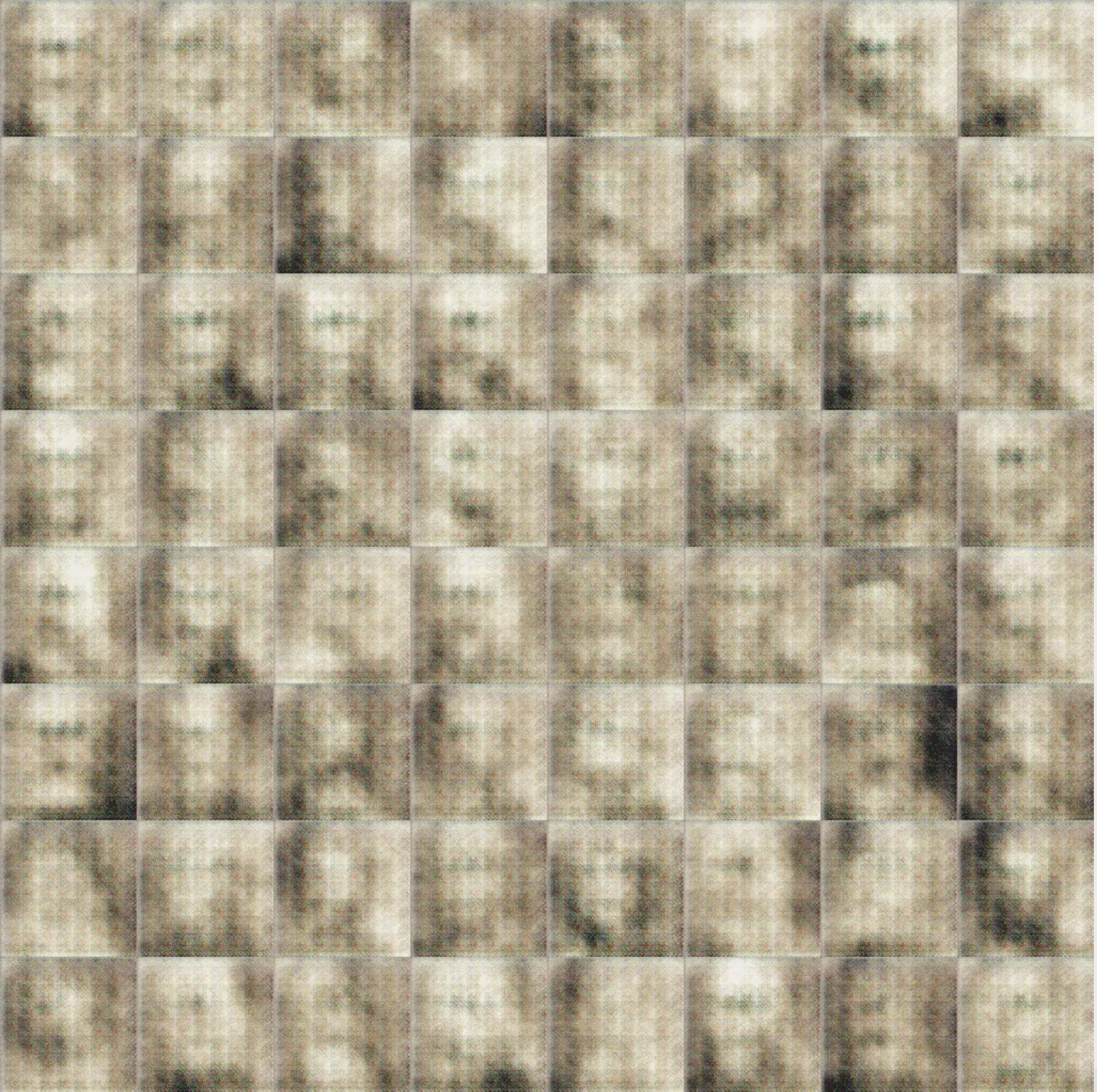
Each learns from the other in order to do its job better:

- The **Forger** learns to **dream more accurately**.
- The **Detective** learns to **be more discerning**.

All of this happens thousands of times a second.
For days and days.

carpedm20/BEGAN-tensorflow

Yale



Generative Adversarial Networks

These people never existed.

Yale

Thank You
& Questions

Yale UNIVERSITY LIBRARY
Digital Humanities Laboratory