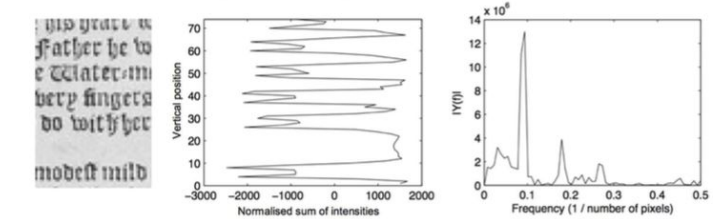
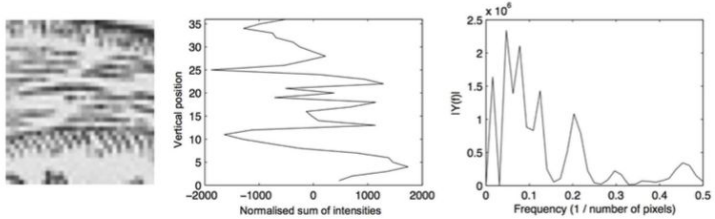
Extracting Illustrations

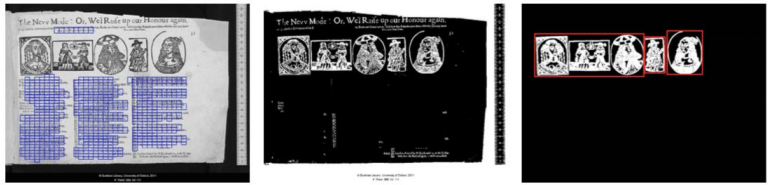


(a) An area of text – a distinctive peak is observed



(b) An area of image – no distinctive peak is observed

Fig. 3. Fourier transform of horizontal sum



(a) Regions of text (b) Text removed from the binary image (c) Bounding boxes detected

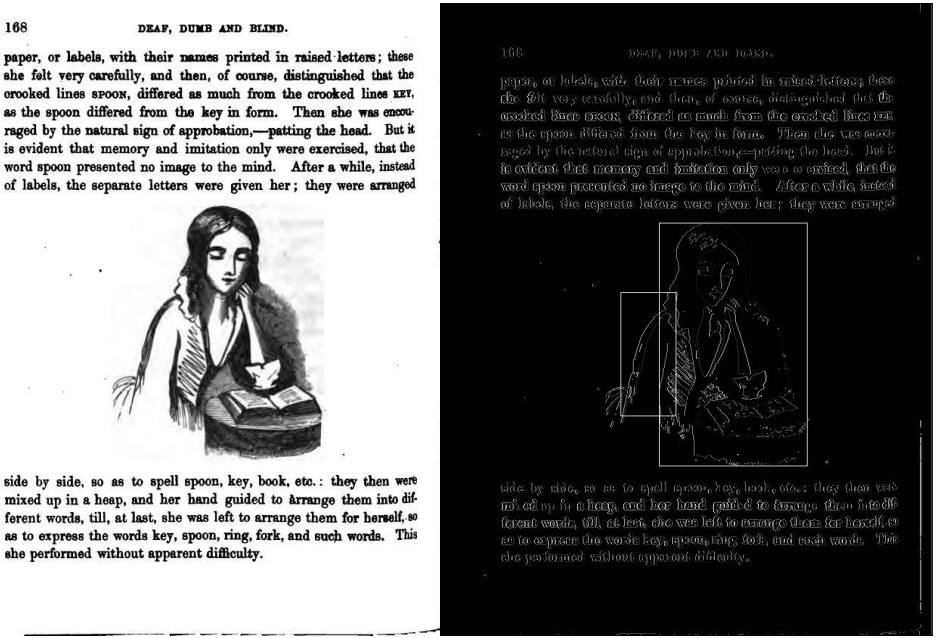
Fig. 4. Text detection and removal, and candidate bounding boxes

Extracting Illustrations

paper, or labels, with their names printed in raised letters; these she felt very carefully, and then, of course, distinguished that the crooked lines SPOON, differed as much from the crooked lines KEY, as the spoon differed from the key in form. Then she was encouraged by the natural sign of approbation,—patting the head. But it is evident that memory and imitation only were exercised, that the word spoon presented no image to the mind. After a while, instead of labels, the separate letters were given her; they were arranged

side by side, so as to spell spoon, key, book, etc.: they then were mixed up in a heap, and her hand guided to arrange them into different words, till, at last, she was left to arrange them for herself, so as to express the words key, spoon, ring, fork, and such words. This she performed without apparent difficulty.

Extracting Illustrations



168 DEAF, DUMB AND BLIND.

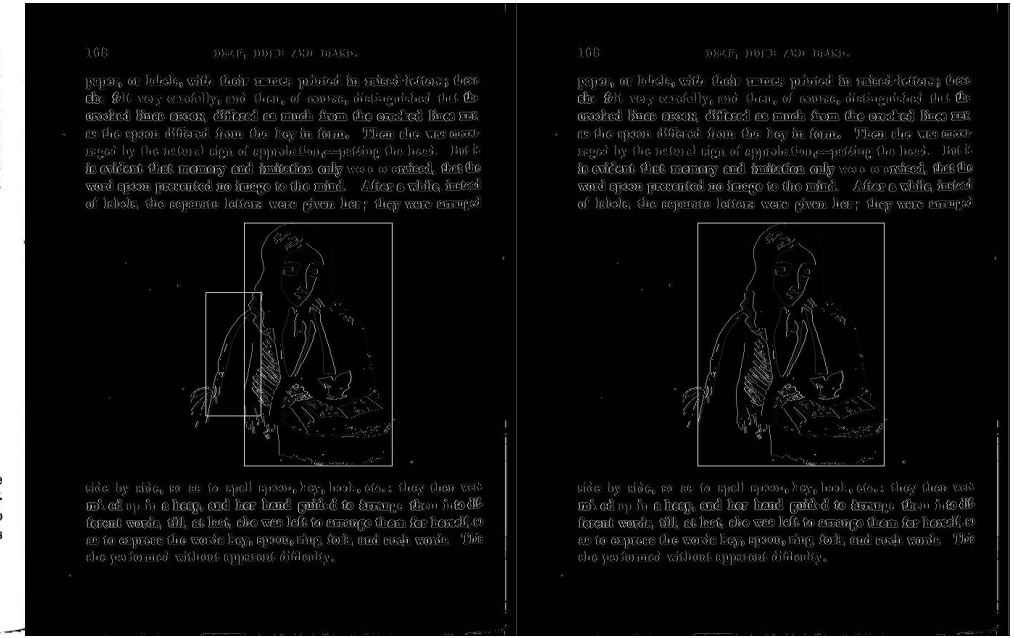
paper, or labels, with their names printed in raised letters; these she felt very carefully, and then, of course, distinguished that the crooked lines SPOON, differed as much from the crooked lines KEY, as the spoon differed from the key in form. Then she was encouraged by the natural sign of approbation,—patting the head. But it is evident that memory and imitation only were exercised, that the word spoon presented no image to the mind. After a while, instead of labels, the separate letters were given her; they were arranged

side by side, so as to spell spoon, key, book, etc.: they then were mixed up in a heap, and her hand guided to arrange them into different words, till, at last, she was left to arrange them for herself, so as to express the words key, spoon, ring, fork, and such words. This she performed without apparent difficulty.

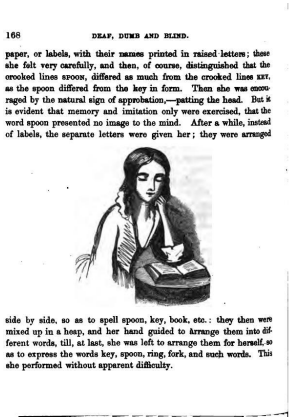
Extracting Illustrations

paper, or labels, with their names printed in raised letters; these she felt very carefully, and then, of course, distinguished that the crooked lines SPOON, differed as much from the crooked lines KEY, as the spoon differed from the key in form. Then she was encouraged by the natural sign of approbation,—patting the head. But it is evident that memory and imitation only were exercised, that the word spoon presented no image to the mind. After a while, instead of labels, the separate letters were given her; they were arranged

side by side, so as to spell spoon, key, book, etc.: they then were mixed up in a heap, and her hand guided to arrange them into different words, till, at last, she was left to arrange them for herself, so as to express the words key, spoon, ring, fork, and such words. This she performed without apparent difficulty.

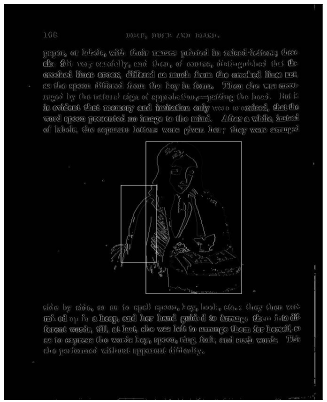


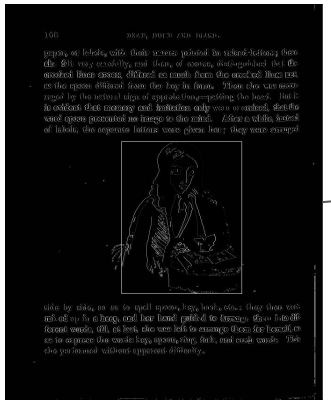
Extracting Illustrations


168 DEAF, DUMB AND BLIND.

paper, or labels, with their names printed in raised letters; these she felt very carefully, and then, of course, distinguished that the crooked lines SPOON, differed as much from the crooked lines KEY, as the spoon differed from the key in form. Then she was encouraged by the natural sign of approbation,—patting the head. But it is evident that memory and imitation only were exercised, that the word spoon presented no image to the mind. After a while, instead of labels, the separate letters were given her; they were arranged

side by side, so as to spell spoon, key, book, etc.: they then were mixed up in a heap, and her hand guided to arrange them into different words, till, at last, she was left to arrange them for herself, so as to express the words key, spoon, ring, fork, and such words. This she performed without apparent difficulty.





Instance Clustering

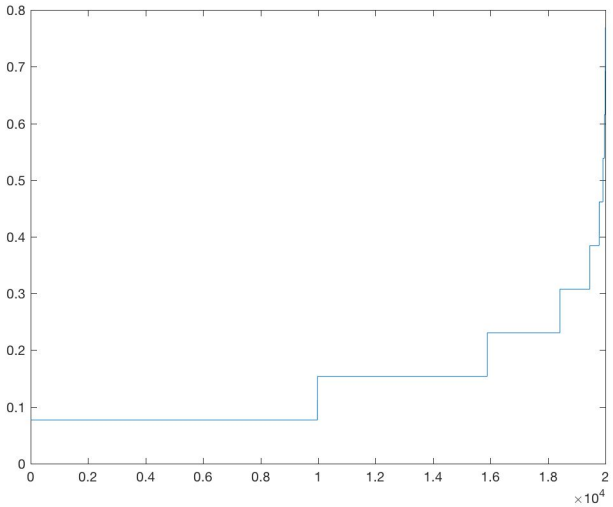
Goal: Recognize pages that are the same across different editions

First Attempt: Histogram Comparison with different similarity metrics

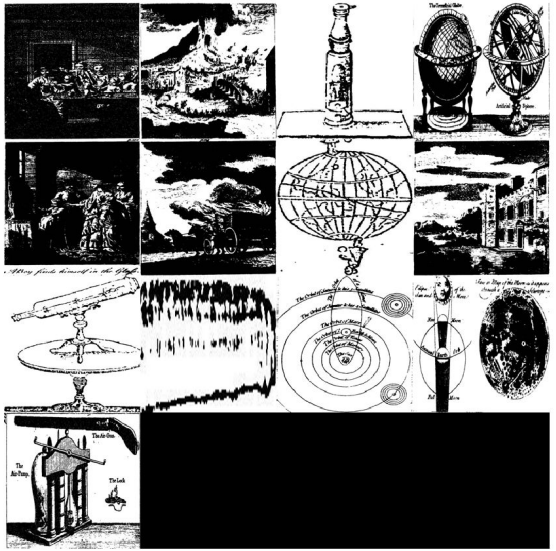
Instance Clustering

Goal: Recognize pages that are the same across different editions

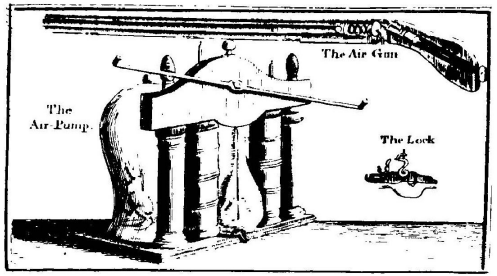
Second Attempt: BOF SIFT with 1-NN Clustering

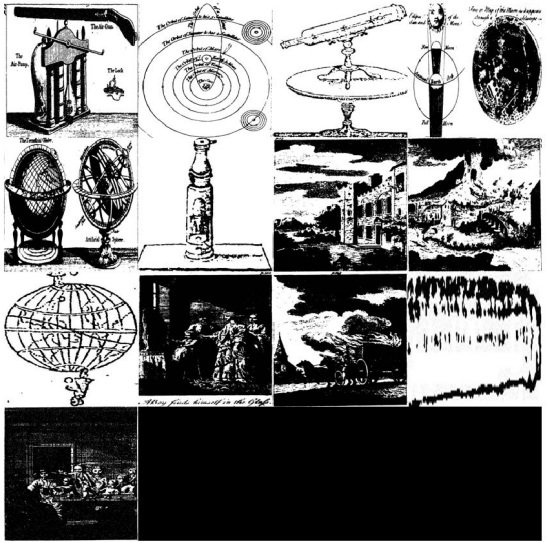


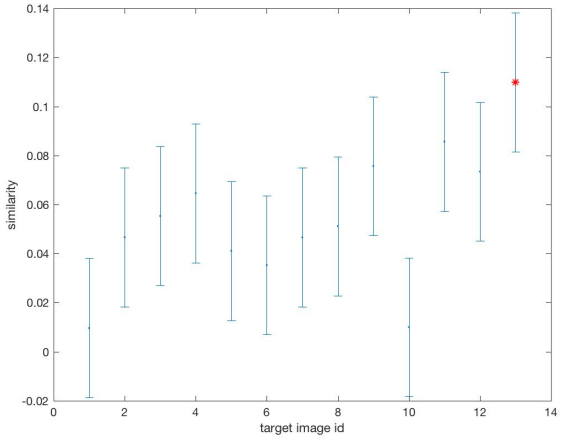
Visual word v. percentage of images present in



Instance Clustering



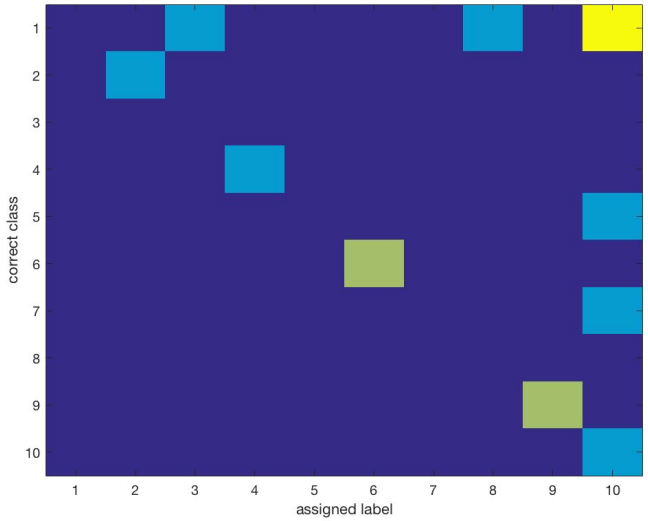




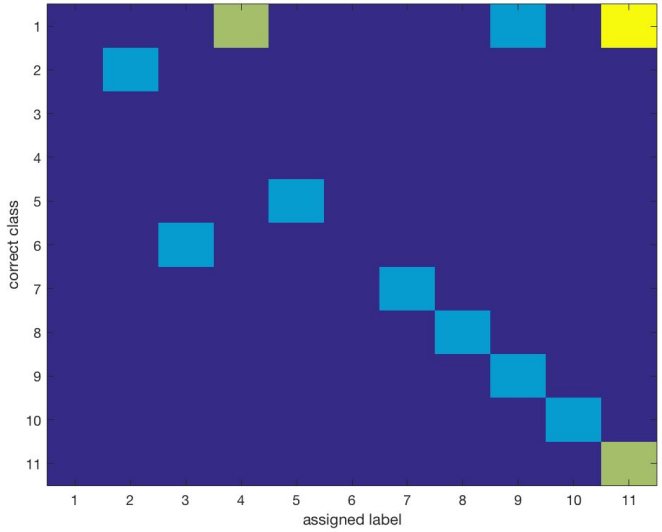
Instance Clustering

Different size training sets

1 image/page training set



2 image/page training set

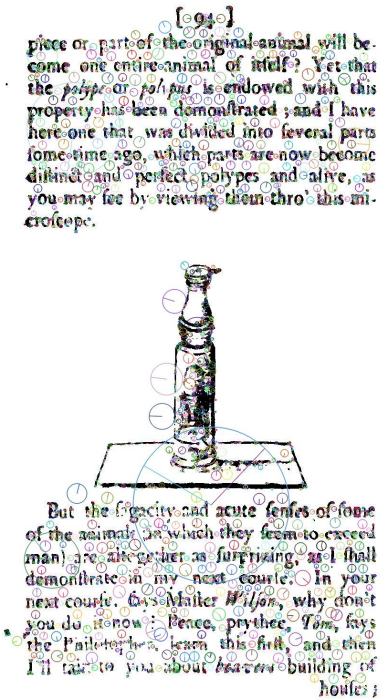


Semantic Clustering

Goal: Recognize similar features across multiple images and group together images with similar features

Approach: Use SIFT to find keypoints + create a BOW dictionary and then cluster based on probabilities of words in an image

Semantic Clustering

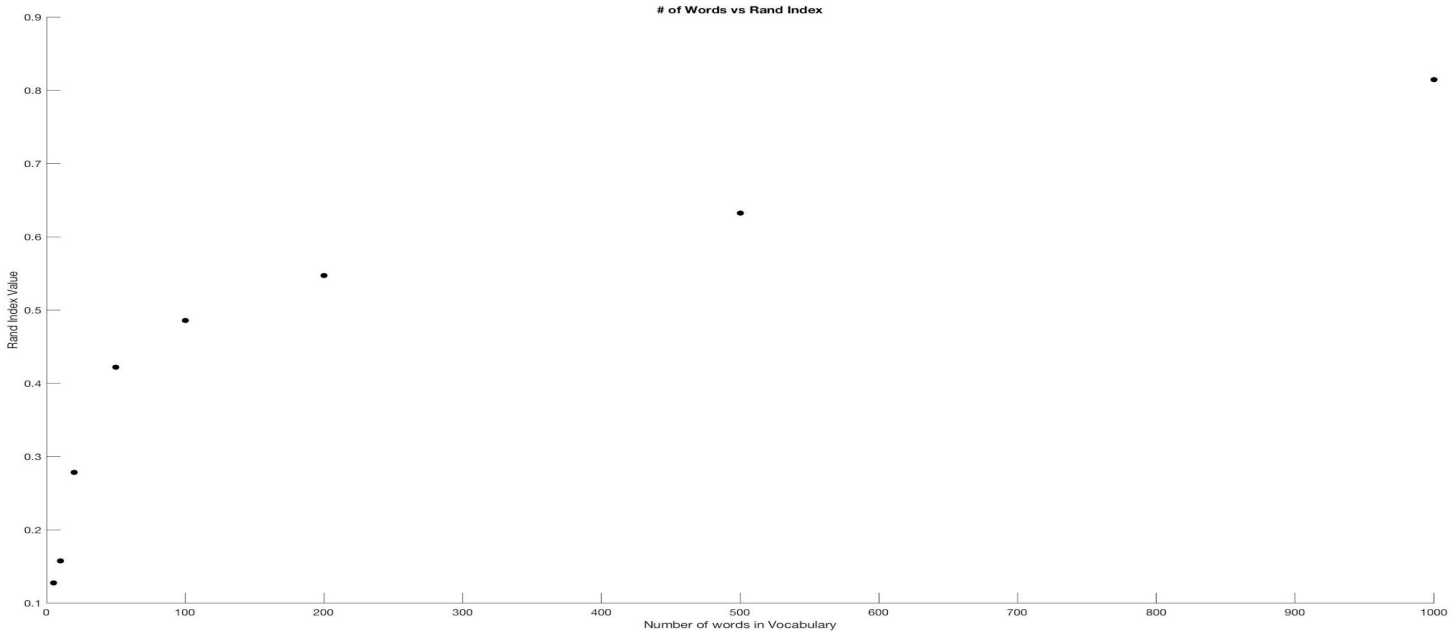


[94]

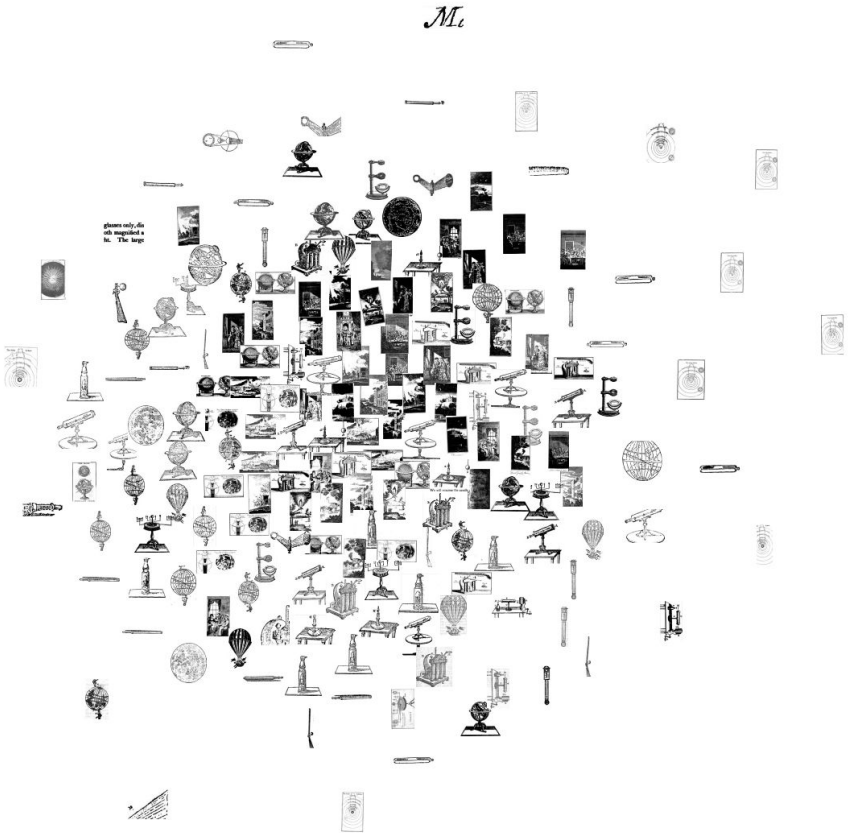
piece or part of the original animal will become one entire animal of itſelf? Yet that the *polype* or *polypus* is endowed with this property has been demonſtrated; and I have here one that was divided into ſeveral parts ſome time ago, which parts are now become diſtinct and perfect polypes and alive, as you may ſee by viewing them thro' this microſcope.

But the ſagacity and acute ſenſes of ſome of the animals (to which they ſeem to exceed man) are altogether as ſurprizing, as I ſhall demonſtrate in my next courſe. In your next courſe, ſays Maſter *Wilſon*, why don't you do it now? Hence, prythee, *Tom*, ſays the Philoſopher, learn this firſt, and then I'll talk to you about the curious building of houſes;

Semantic Clustering



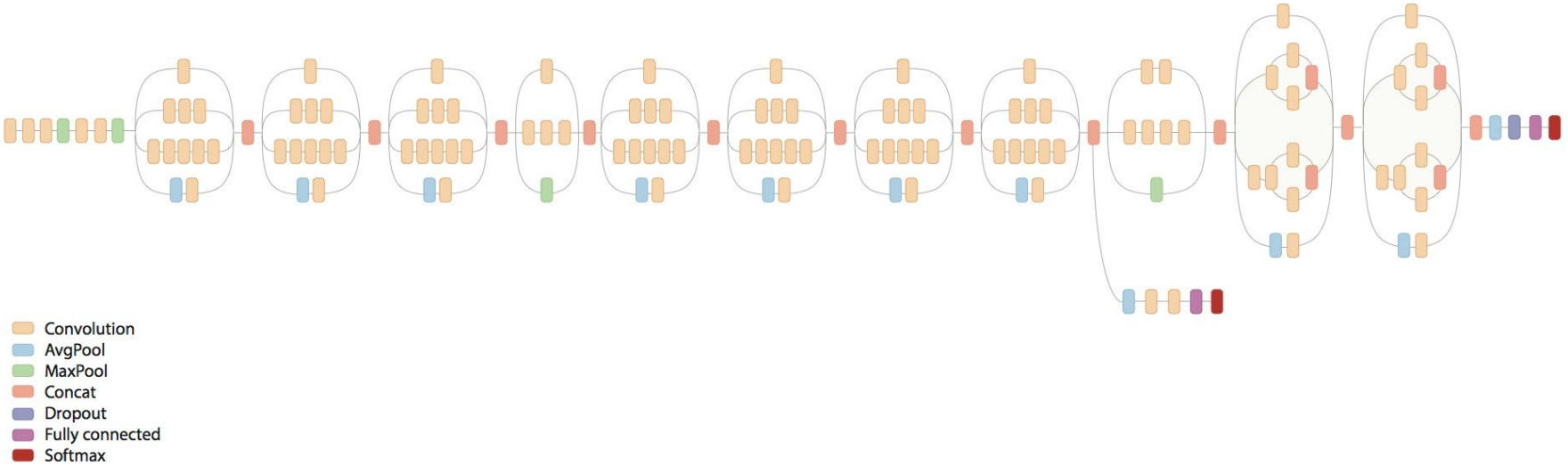
Semantic Clustering



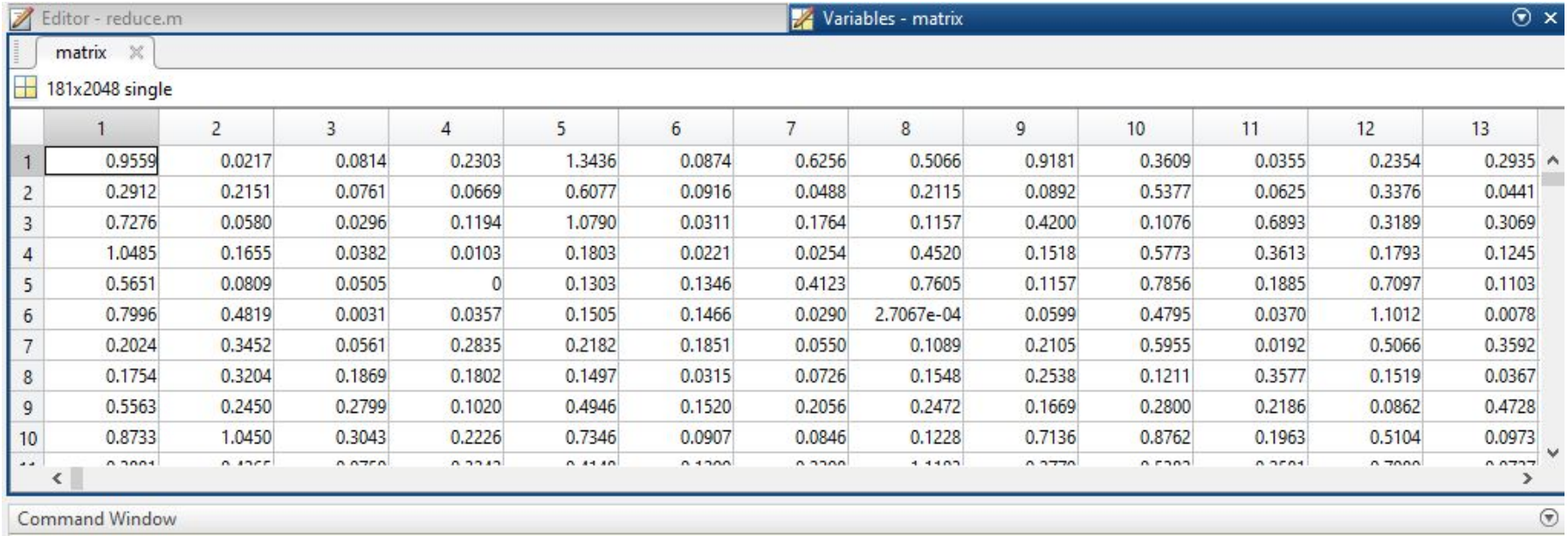
Clustering with TensorFlow's Inception

- Large-scale open-source CNN
- Trained on ImageNet
  - Academic database; mostly Flickr images; sponsors CV competitions
  - Order of 1,000,000 images with bounding box annotations and SIFT features
- ImageNet based on WordNet
  - ImageNet provides 1000 examples per "synset" (synonym set)
  - WordNet synsets are hierarchical clusters of semantic relations
  - Most are nouns with simple part-whole or similarity relations
- IDEA: use existing labeled data to embed input illustrations in high-dimensional feature space (2048 synset nodes)
- HOPE: "transfer learning" as tagged photo database gives insight into 19C illustration features

Inception v3 architecture (classify_image_mk.py)



Output is N x 2048 likelihood matrix [0,~5.38 … ?]


Editor - reduce.m

Variables - matrix

matrix

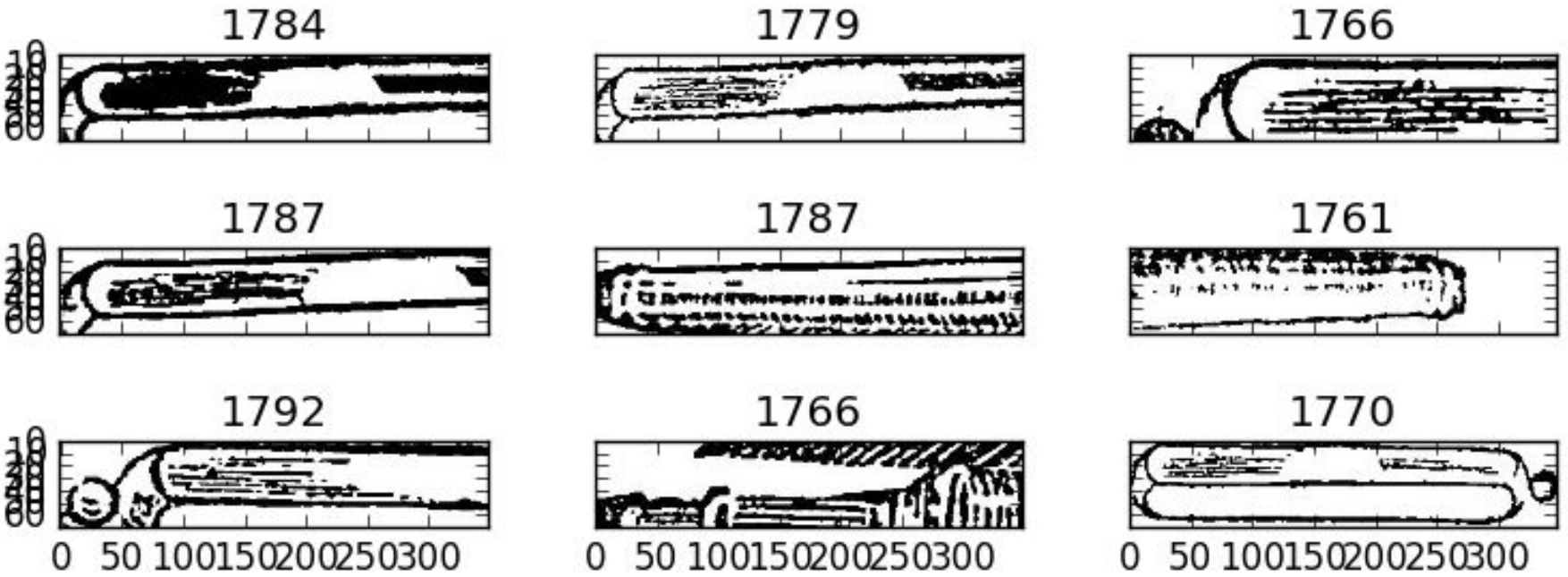
181x2048 single

	1	2	3	4	5	6	7	8	9	10	11	12	13
1	0.9559	0.0217	0.0814	0.2303	1.3436	0.0874	0.6256	0.5066	0.9181	0.3609	0.0355	0.2354	0.2935
2	0.2912	0.2151	0.0761	0.0669	0.6077	0.0916	0.0488	0.2115	0.0892	0.5377	0.0625	0.3376	0.0441
3	0.7276	0.0580	0.0296	0.1194	1.0790	0.0311	0.1764	0.1157	0.4200	0.1076	0.6893	0.3189	0.3069
4	1.0485	0.1655	0.0382	0.0103	0.1803	0.0221	0.0254	0.4520	0.1518	0.5773	0.3613	0.1793	0.1245
5	0.5651	0.0809	0.0505	0	0.1303	0.1346	0.4123	0.7605	0.1157	0.7856	0.1885	0.7097	0.1103
6	0.7996	0.4819	0.0031	0.0357	0.1505	0.1466	0.0290	2.7067e-04	0.0599	0.4795	0.0370	1.1012	0.0078
7	0.2024	0.3452	0.0561	0.2835	0.2182	0.1851	0.0550	0.1089	0.2105	0.5955	0.0192	0.5066	0.3592
8	0.1754	0.3204	0.1869	0.1802	0.1497	0.0315	0.0726	0.1548	0.2538	0.1211	0.3577	0.1519	0.0367
9	0.5563	0.2450	0.2799	0.1020	0.4946	0.1520	0.2056	0.2472	0.1669	0.2800	0.2186	0.0862	0.4728
10	0.8733	1.0450	0.3043	0.2226	0.7346	0.0907	0.0846	0.1228	0.7136	0.8762	0.1963	0.5104	0.0973

Command Window

Idea 1: K-NN in R^d (d << 2048)

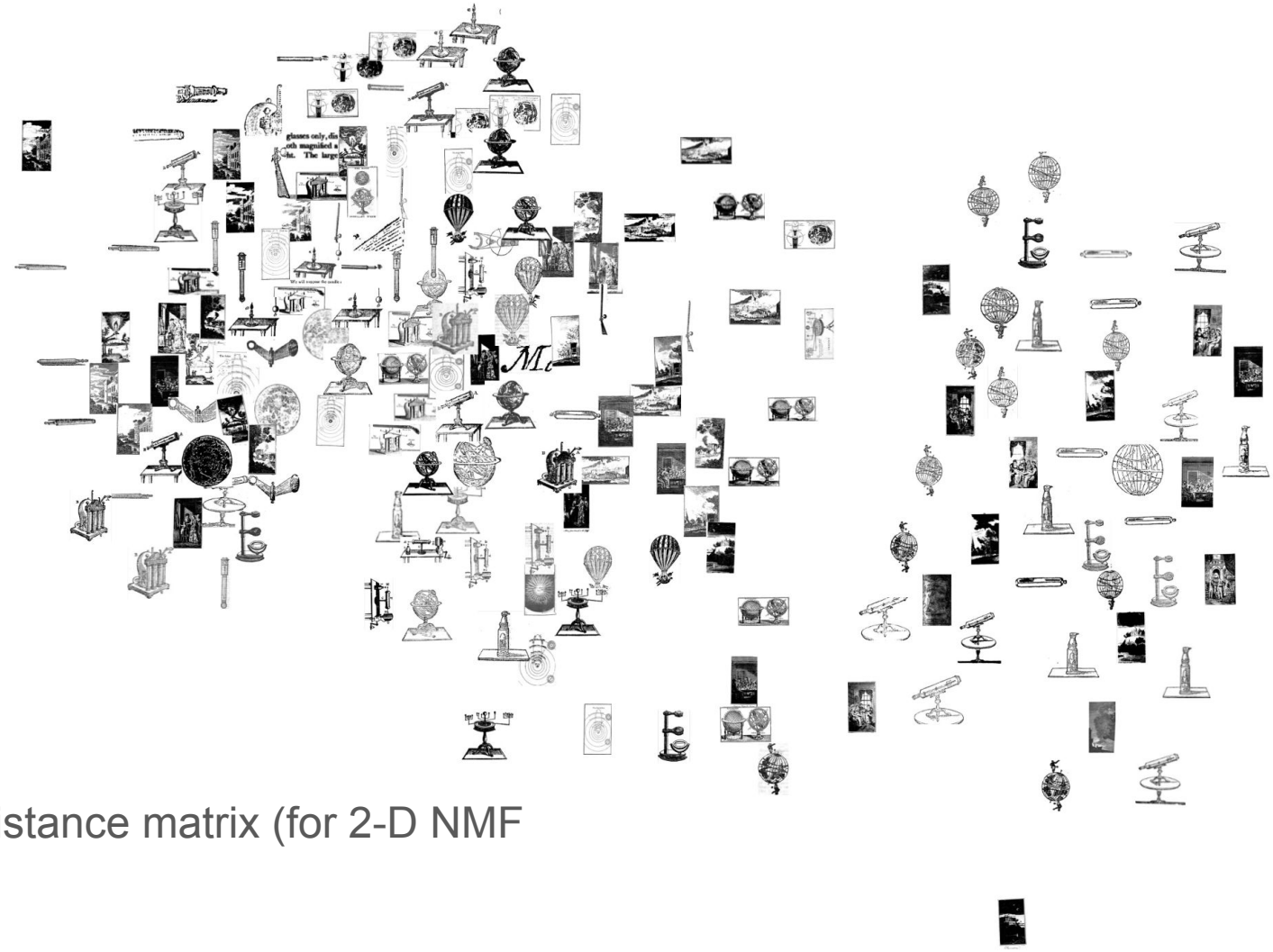
- Approximate Nearest Neighbors
  - From annoy import AnnoyIndex
  - Created by Spotify; generate forest of trees using randomly projected hyperplanes (approximate box decomposition trees)
  - Can use static files (.ann) as indexes
  - Data becomes d-dimensional vectors (after factorizing matrix)
- Gives a sort of “instance” or “type” clustering
- VISUALIZE: pick random image from dataset and show its neighborhood
- EXPECT: image from close publication year will be close by
- FINDING: this is generally the case



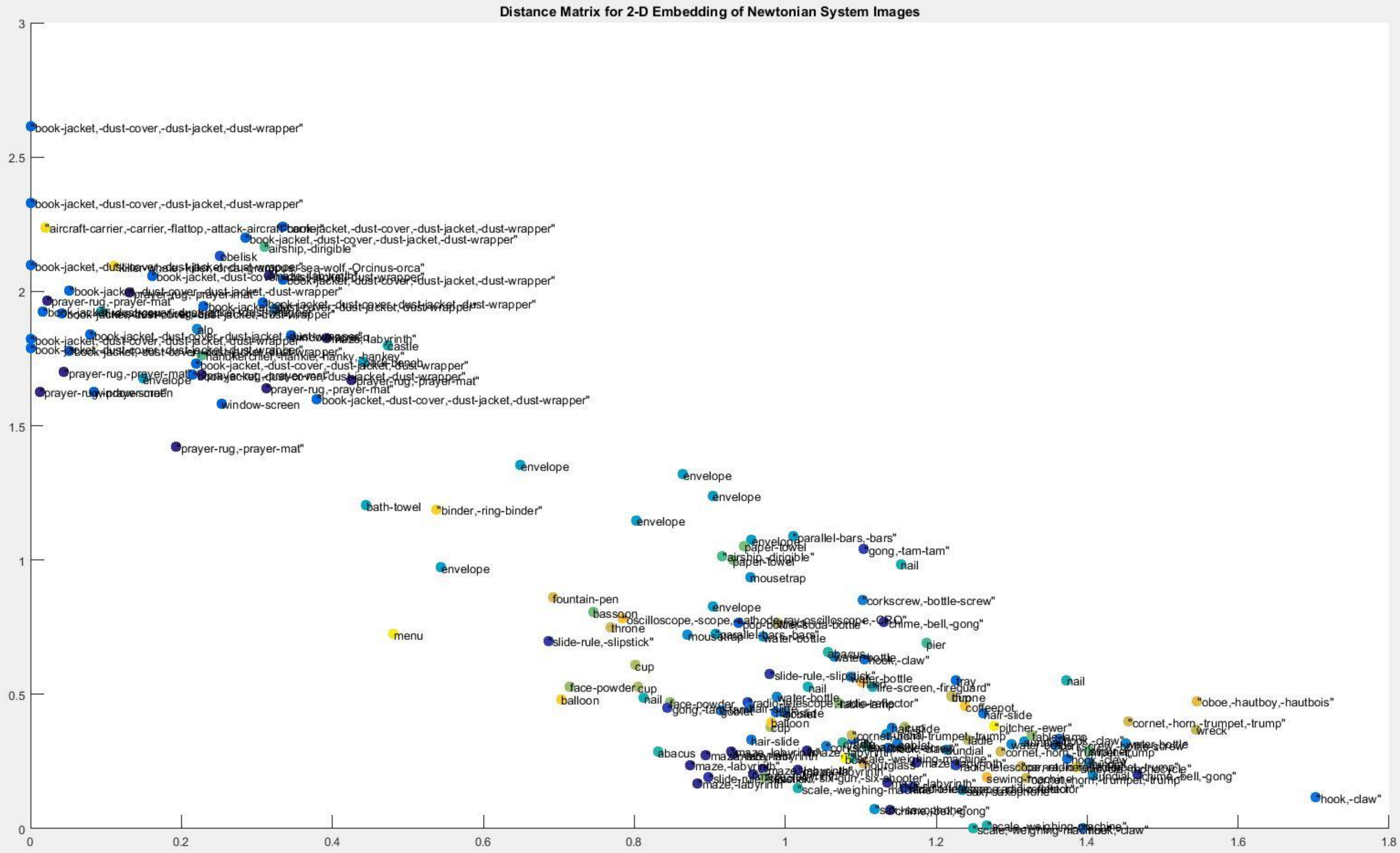
Result 1: 8-NN labeled by publication year (input top left)

Idea 2: Factorize into 2-D embedding

- Non-negative matrix factorization
  - from sklearn.decomposition import NMF
  - Two components, default random state
  - Center the matrix and project
  - Makes sense since we have likelihoods (non-negative)
  - Useful in CV and document clustering tasks
- Pros: tractable; scatterplot has two general clusters
- Cons: don’t know what information we are losing



Result 2: FDG of distance matrix (for 2-D NMF embedding)



Distance Matrix for 2-D Embedding of Newtonian System Images
"book-jacket,-dust-cover,-dust-jacket,-dust-wrapper"
"book-jacket,-dust-cover,-dust-jacket,-dust-wrapper"
"aircraft-carrier,-carrier,-flattop,-attack-aircraft-carrier"
"book-jacket,-dust-cover,-dust-jacket,-dust-wrapper"
"airship,-dirigible"
obelisk
"book-jacket,-dust-cover,-dust-jacket,-dust-wrapper"
"prayer-rug,-prayer-mat"
alp
castle
"handkerchief,-hankie,-hanky,-hankey"
"book-jacket,-dust-cover,-dust-jacket,-dust-wrapper"
envelope
"prayer-rug,-prayer-mat"
window-screen
"book-jacket,-dust-cover,-dust-jacket,-dust-wrapper"
"prayer-rug,-prayer-mat"
envelope
envelope
envelope
bath-towel
"binder,-ring-binder"
envelope
"parallel-bars,-bars"
paper-towel
"gong,-tam-tam"
"airship,-dirigible"
nail
mousetrap
envelope
fountain-pen
"corkscrew,-bottle-screw"
bassoon
envelope
throne
"chime,-bell,-gong"
menu
"slide-rule,-slipstick"
pier
cup
"hook,-claw"
"slide-rule,-slipstick"
face-powder
cup
nail
"fire-screen,-fireguard"
balloon
nail
"oboe,-hautboy,-hautbois"
coffeepot
hair-slide
"cornet,-horn,-trumpet,-trump"
"pitcher,-ewer"
wreck
hair-slide
abacus
"maze,-labyrinth"
"cornet,-horn,-trumpet,-trump"
"scale,-weighing-machine"
"sewing-machine"
"chime,-bell,-gong"
"maze,-labyrinth"
"hook,-claw"
"scale,-weighing-machine"
"scale,-weighing-machine"
0
0.2
0.4
0.6
0.8
1
1.2
1.4
1.6
1.8
0
0.5
1
1.5
2
2.5
3

Result 2: Top Semantic Clusters

synset	count	percentage
book-jacket,-dust-cover,-dust-jacket,-dust-wrapper	19	11.5854
maze,-labyrinth	13	7.9268
prayer-rug,-prayer-mat	8	4.8780
envelope	8	4.8780
hook,-claw	6	3.6585
water-bottle	6	3.6585
hair-slide	5	3.0488
sundial	5	3.0488

WordNet Search - 3.1

Word to search for: book jacket Search WordNet

Display Options: (Select option to change) Change

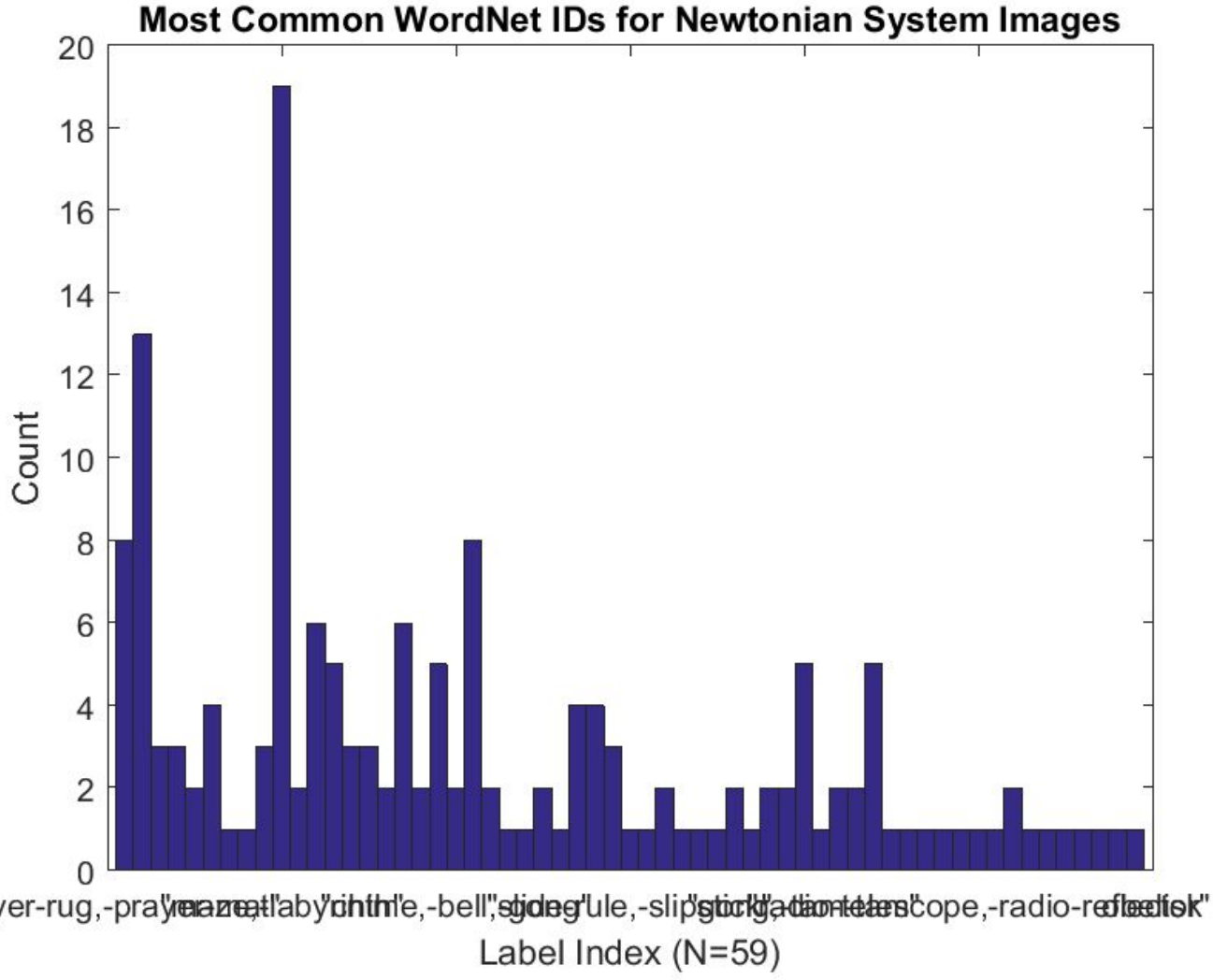
Key: "S:" = Show Synset (semantic) relations, "W:" = Show Word (lexical) relations

Display options for sense: (frequency) {offset} <lexical filename > [lexical file number] (gloss) "an example sentence"

Display options for word: word#sense number (sense key)

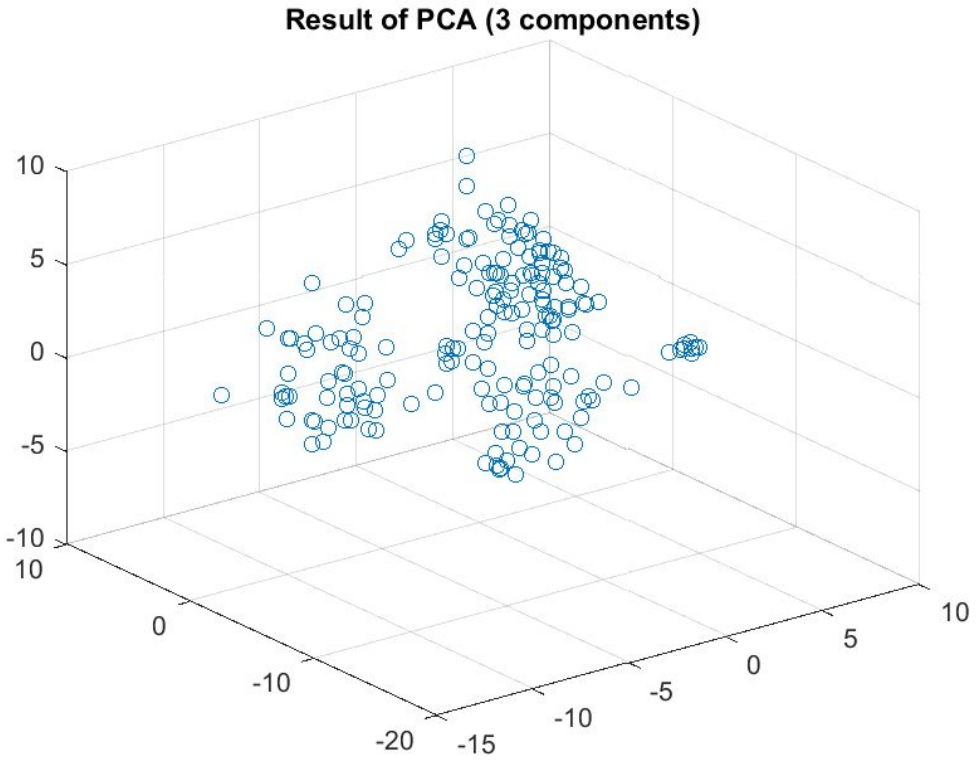
Noun

- (1){07262988} <noun.communication>[10] S: (n) **book jacket#1 (book_jacket%1:10:00::)**, dust cover#1 (dust_cover%1:10:00::), dust jacket#1 (dust_jacket%1:10:00::), dust wrapper#1 (dust_wrapper%1:10:00::) (a paper jacket for a book; a jacket on which promotional information is usually printed)


Most Common WordNet IDs for Newtonian System Images
Count
0
2
4
6
8
10
12
14
16
18
20
"prayer-rug,-prayer-mat"
"cornet,-horn,-trumpet,-trump"
"abacus"
"bell"
"sliding-rule,-slip-stick"
"stick"
"radio-telescope,-radio-reflector"
"fork"
Label Index (N=59)

Impasse

- Clustering is “OK” but no way to validate; no pure clusters and no clear idea of how many clusters there should be
- No sense if images lying on manifold or not
- IDEA: estimate the underlying dimensionality of the similarity vectors
- EXTENSION: project on first few principal components and hierarchically cluster; examine dendrogram level for our default/coarse hypothesis (i.e. there are as many clusters as the average number of illustrations per edition)
- TOOLS: Laurens van der Maaten’s Matlab toolkit, OJ Woodford’s montage extension

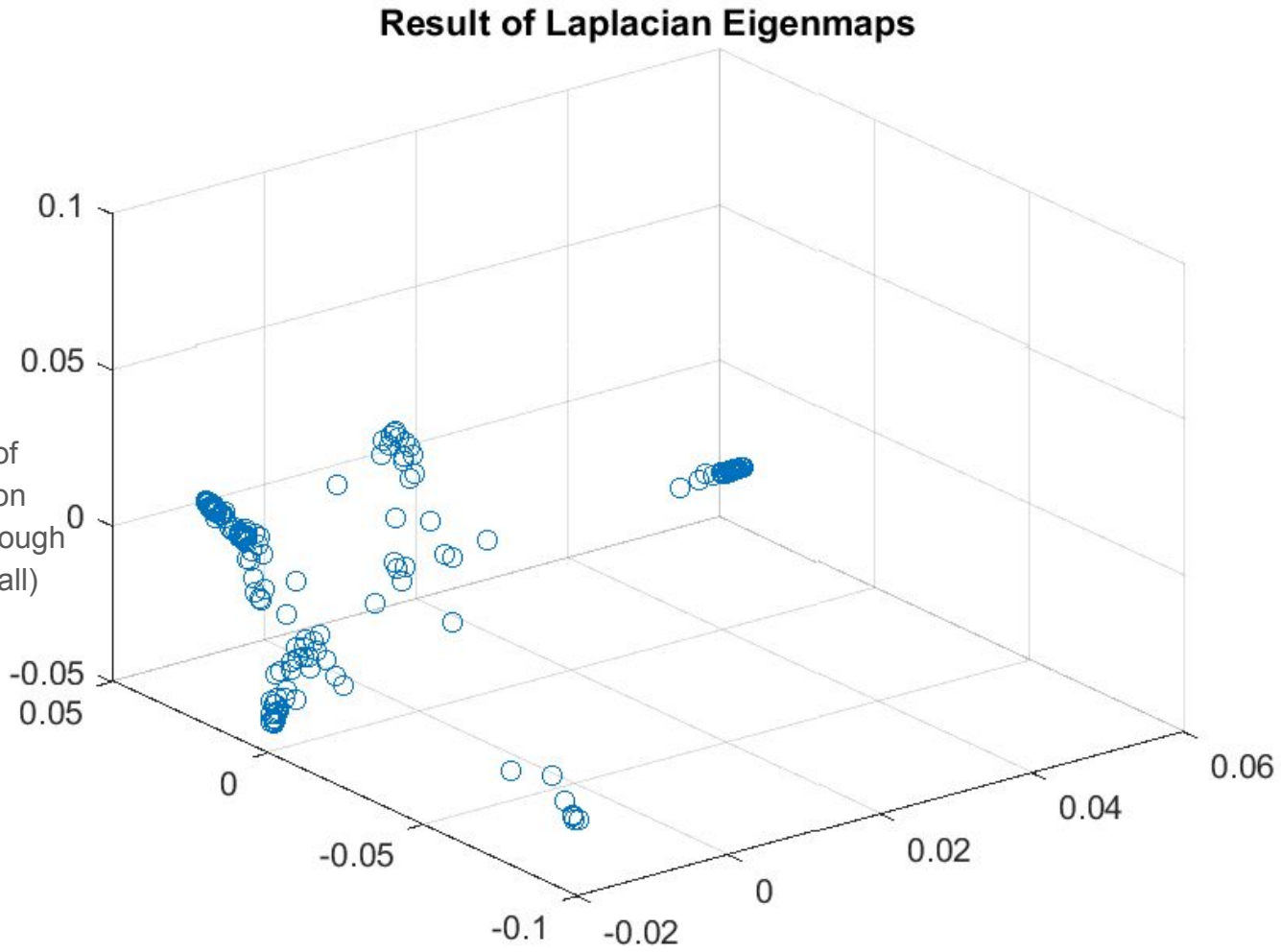


13 top eigenpairs explain >90% of attribute variance

But . . .

The data are highly non-linear.

Solution? Build graph Laplacian of k-NN points. Just like with diffusion maps. Much better spacing! (although diffusion maps did not do well at all)



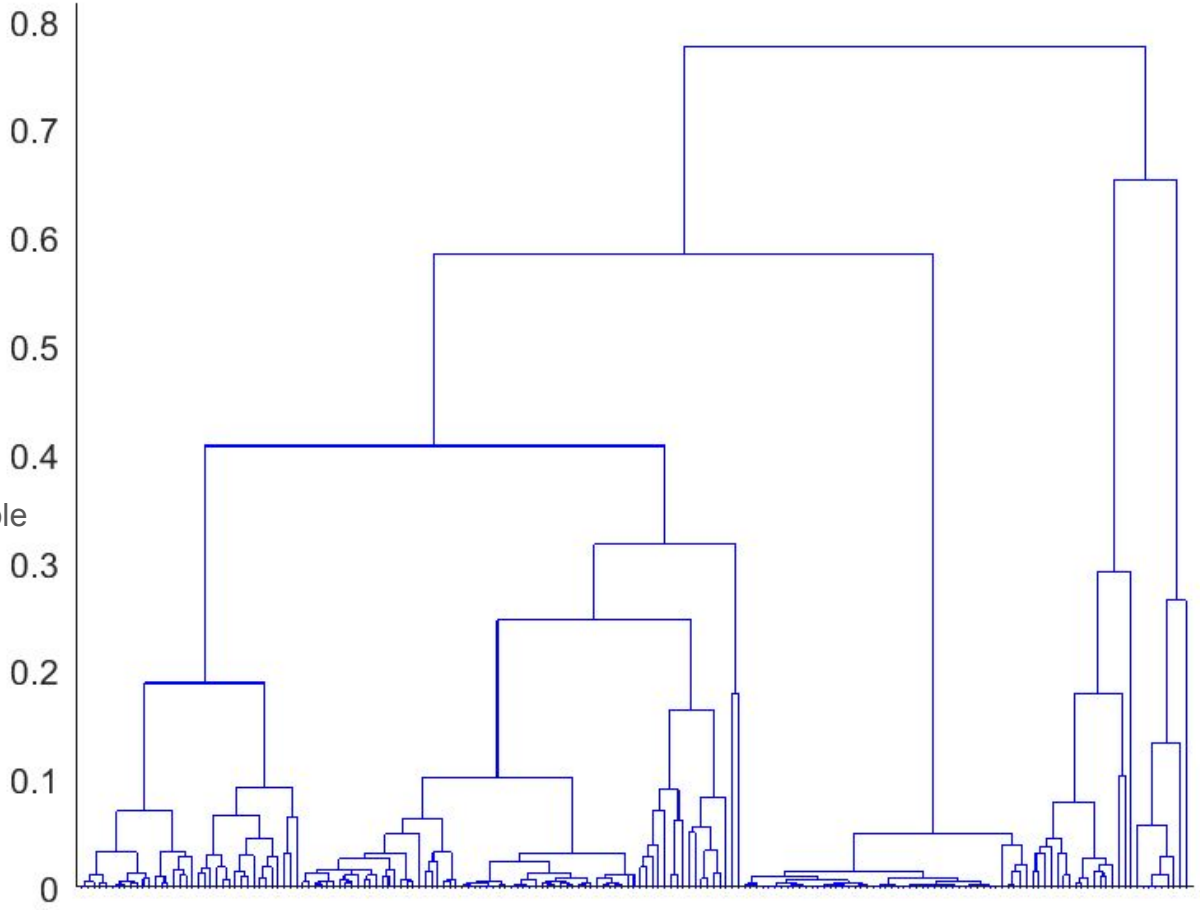
Average-link

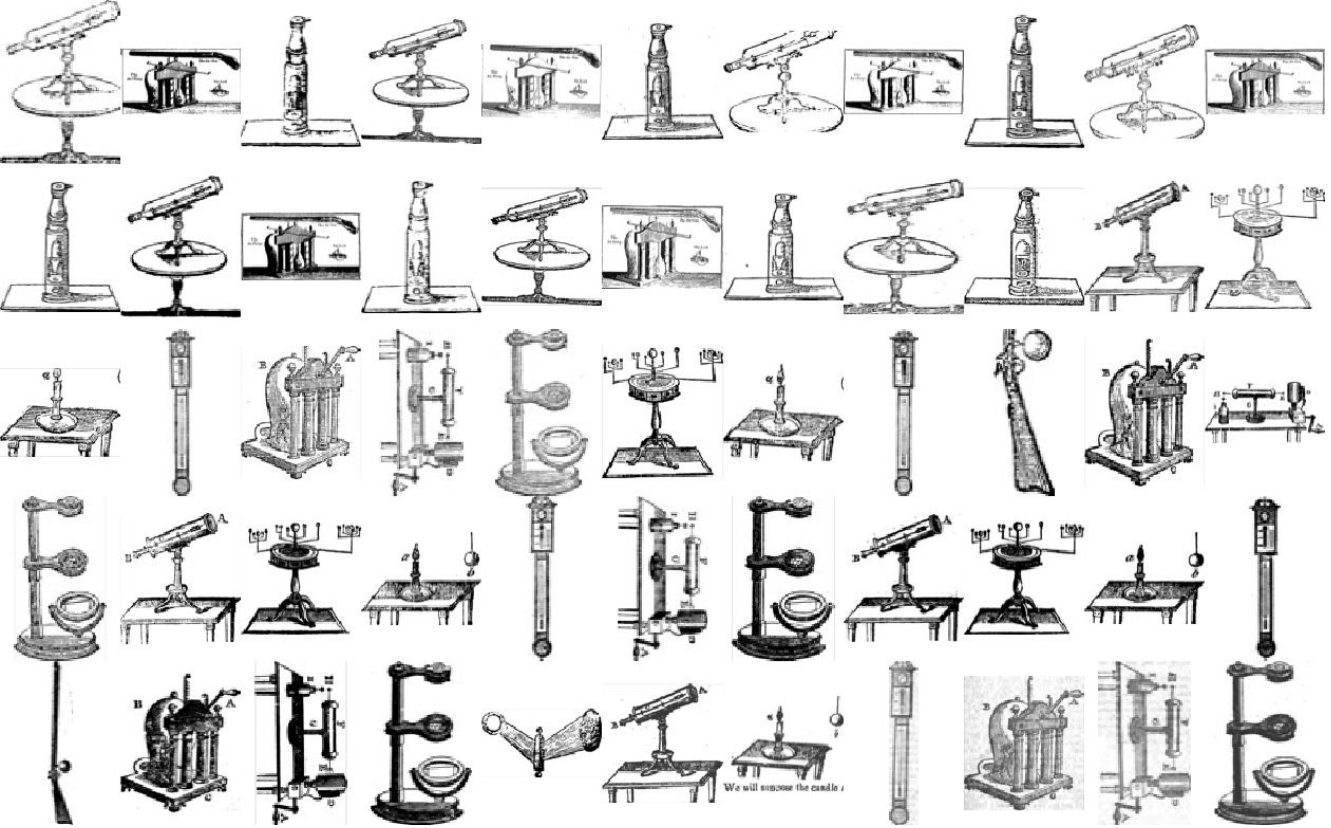
Since we can now visualize clear clusters in R^3 attribute space, we hierarchically cluster the embedded data (no points lost in the mapping).

Looks like 0.19 or so in the pairwise distance matrix gives us a reasonable number of clusters.

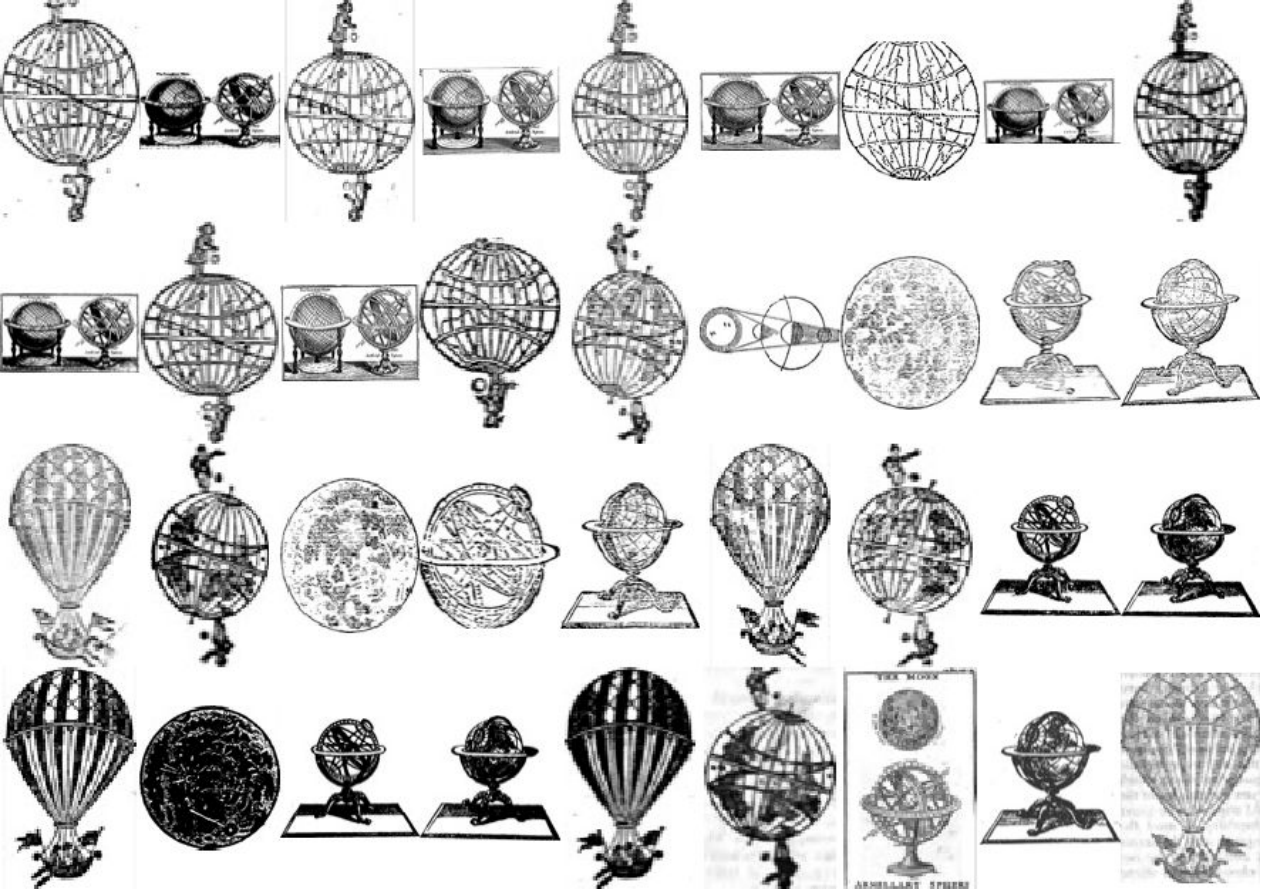
PLAN: color the points by cluster ID and (subjectively) evaluate the grouped image thumbnails.

N.B. single-link NOT good. Why?

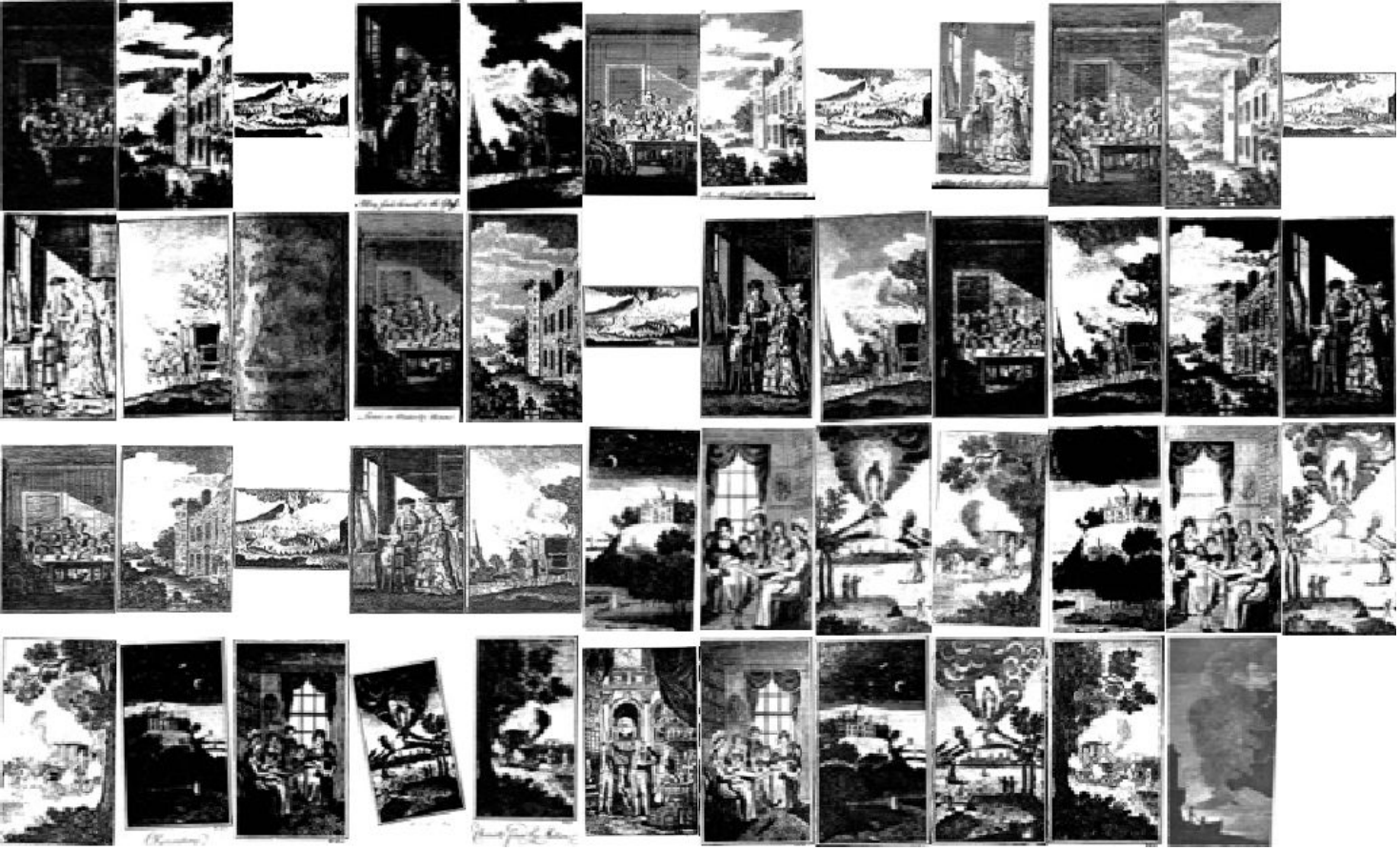




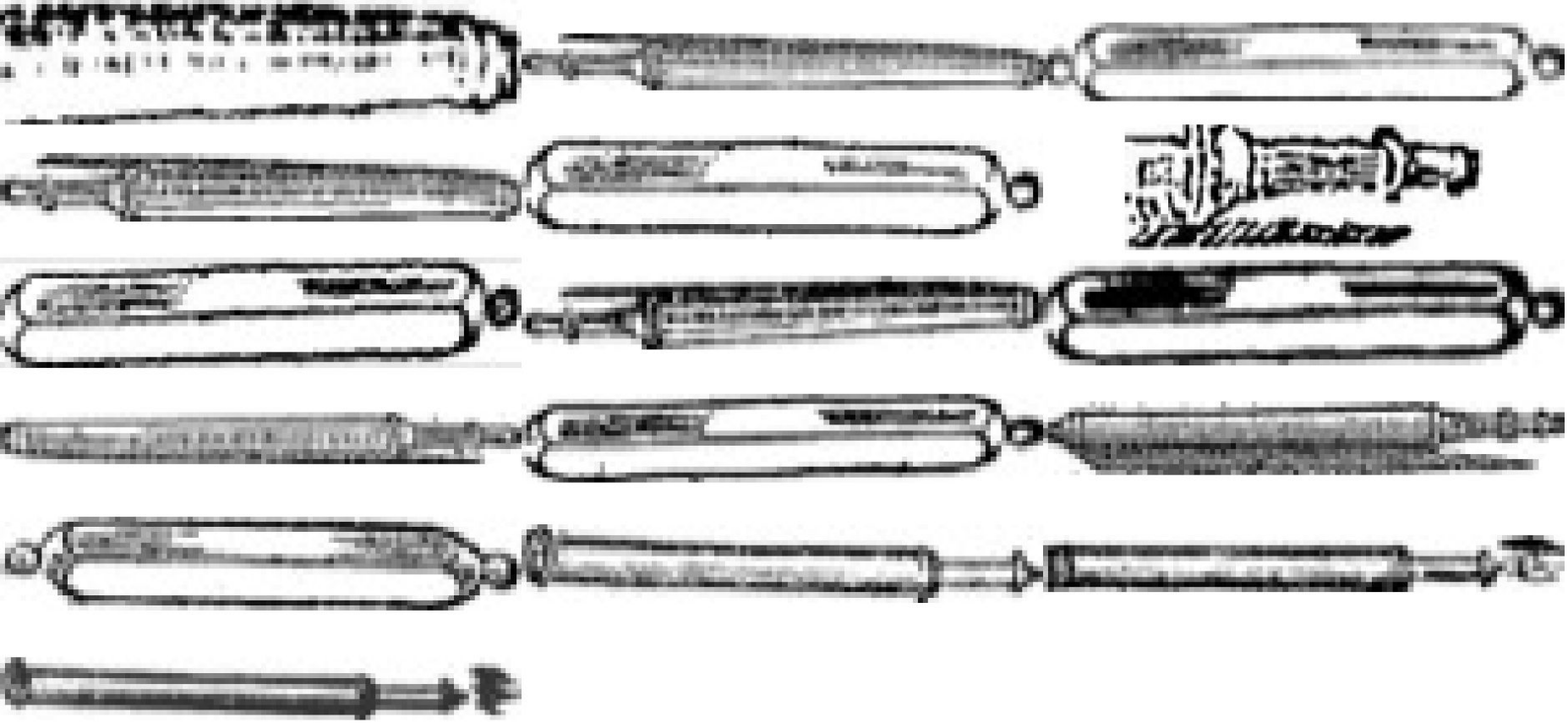
Cluster 1: “Instruments”



Cluster 5: “globes”



Cluster 6: “Full-page”

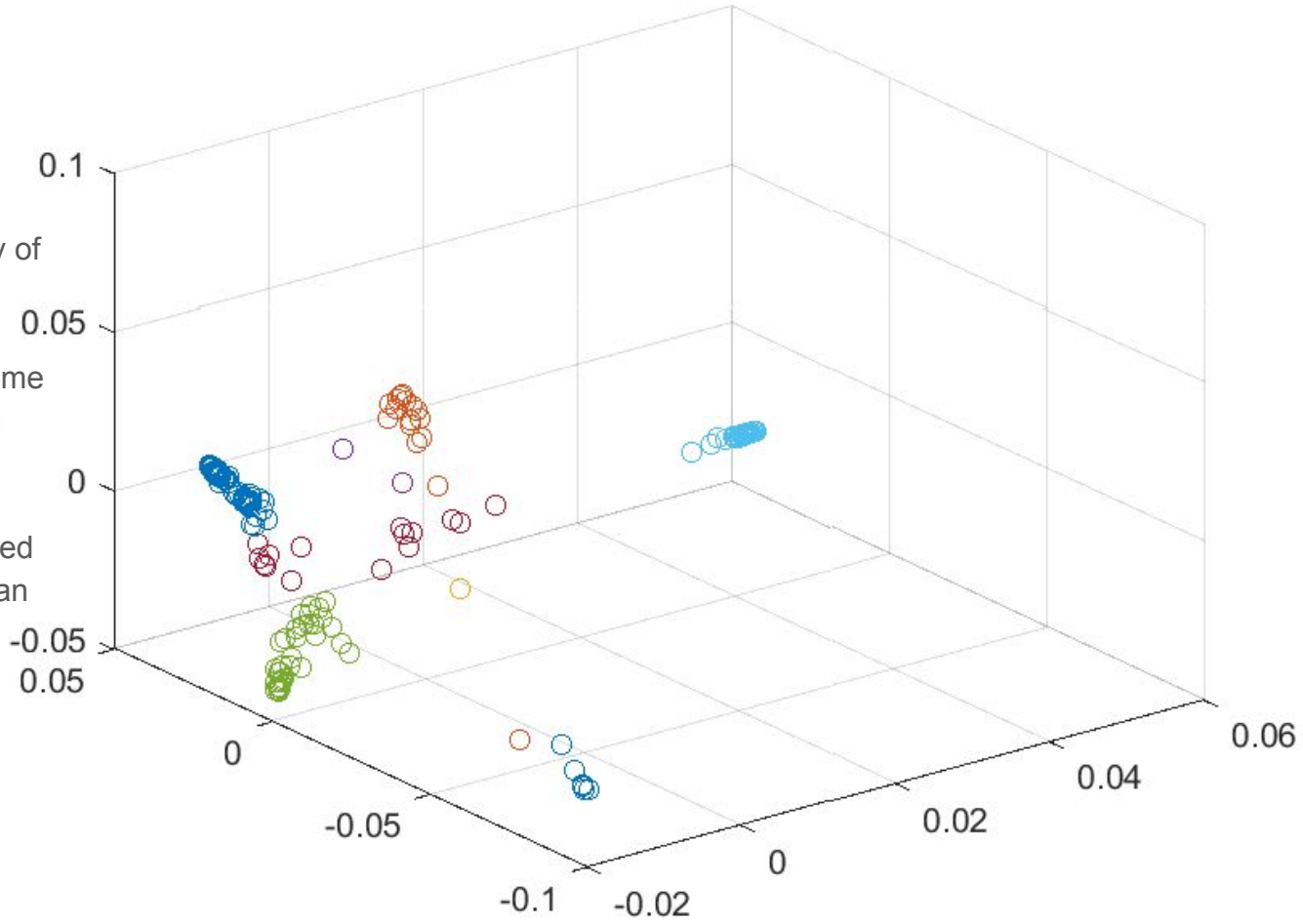


Cluster 9: “Telescopes”

Validation

Subjective judgments of similarity of images subject to bias and not possible in unsupervised setting. Confirm that the clusters have some separation in the R^3 embedding using Laplacian eigenmaps.

TODO: if the clustered and cleaned data are linearly separable, we can train a classifier.



Directions

Map “suggestions” from our model onto the UI discussed in the proposal.

Incorporate human judgments of similarities between illustrations.

Formalize relations between illustrations (exact match, thematic, tracing) using graph database.