MI3: Machine-Initiated Intelligent Interaction for Interactive Classification and Data Reconstruction

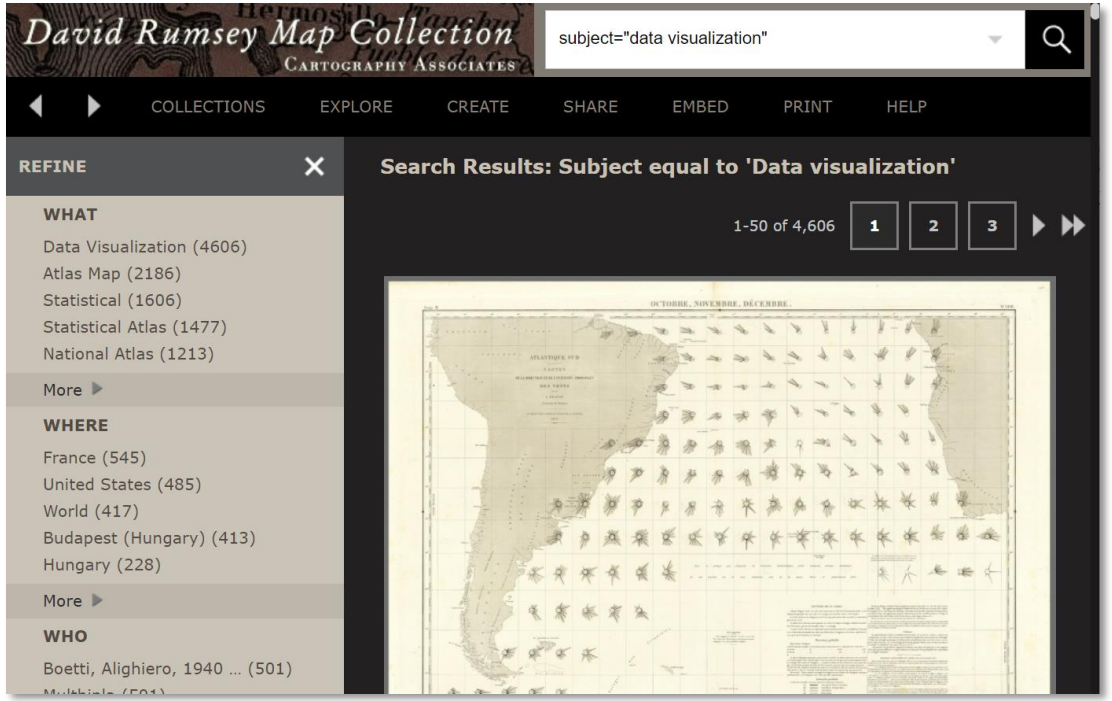


Yu Zhang Bob Coecke Min Chen

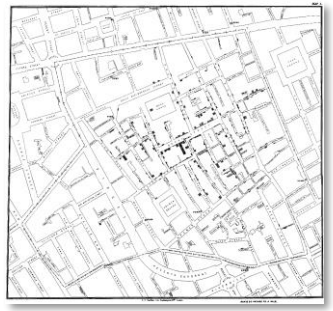
University of Oxford

Historical Visualizations

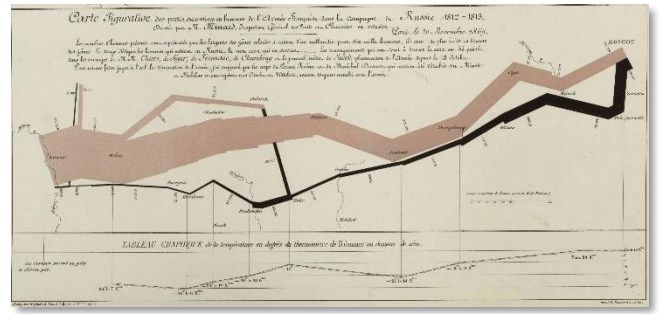
- Historical visualization databases



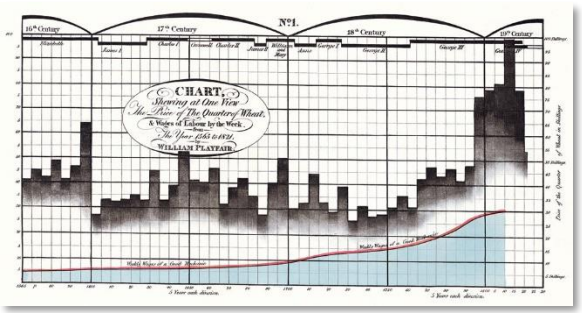
David Rumsey Map Collection



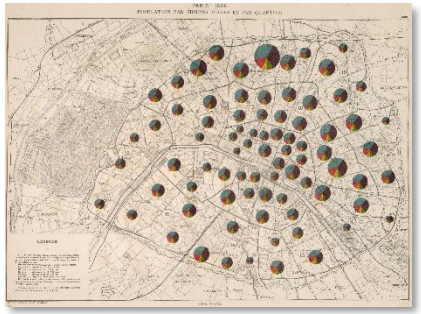
John Snow's cholera map, 1855



Charles Minard's map of Napoleon's Russian campaign, 1869



William Playfair's chart of wheat price and wage, 1784



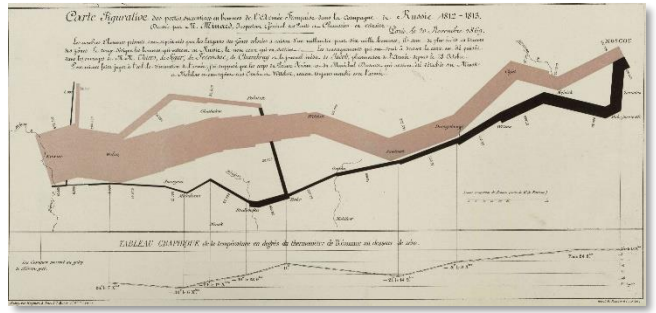
Jacques Bertillon's thematic map with pie-chart glyphs, 1886

Historical Visualizations

- Historical visualization databases

David Rumsey Map Collection

John Snow’s cholera map, 1855



Charles Minard’s map of Napoleon's Russian campaign, 1869

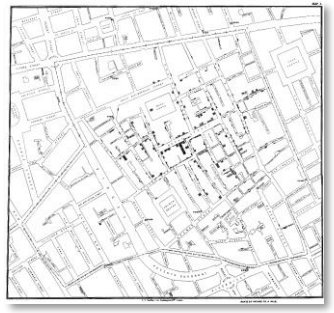
William Playfair’s chart of wheat price and wage, 1784

Jacques Bertillon’s thematic map with pie-chart glyphs, 1886

Historical Visualizations

- Historical visualization databases

David Rumsey Map Collection



John Snow’s cholera map, 1855

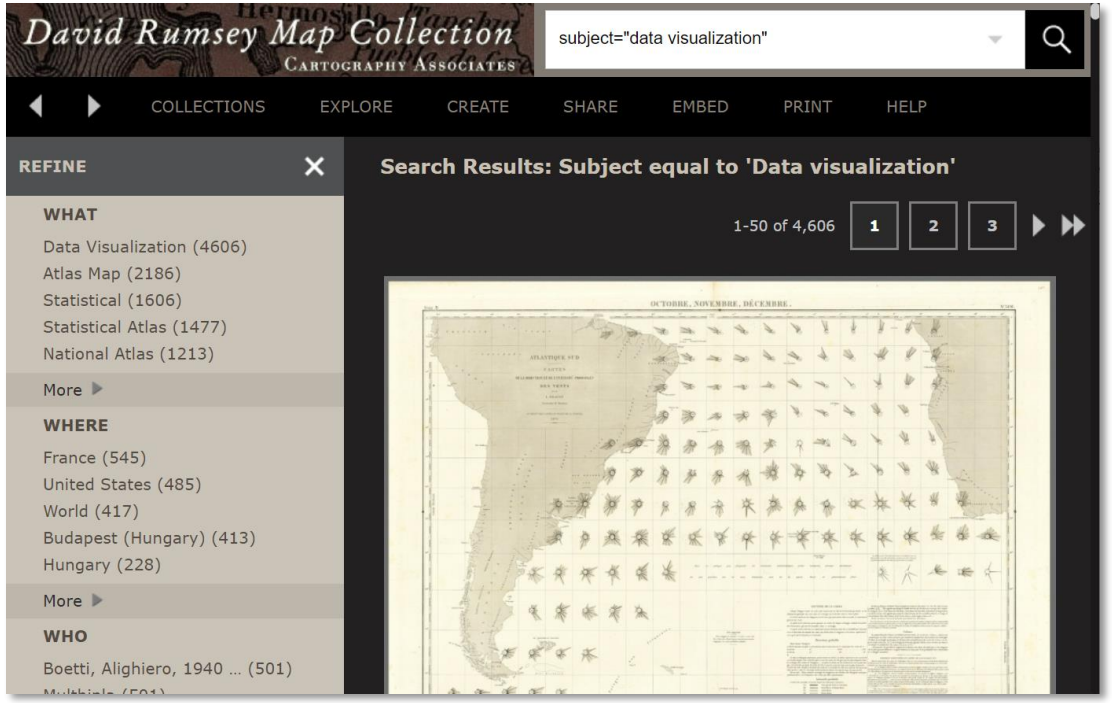
Charles Minard’s map of Napoleon's Russian campaign, 1869

William Playfair’s chart of wheat price and wage, 1784

Jacques Bertillon’s thematic map with pie-chart glyphs, 1886

Historical Visualizations

- Historical visualization databases



David Rumsey Map Collection

John Snow's cholera map, 1855

Charles Minard's map of Napoleon's Russian campaign, 1869

William Playfair's chart of wheat price and wage, 1784

Jacques Bertillon's thematic map with pie-chart glyphs, 1886

Data Reconstruction

- **Input**: a bitmap image of a visualization
- **Output**: the underlying dataset that generated the visualization

John Snow's cholera map, 1855

#victims	latitude	longitude
1	51.514005	-0.139363
8	51.512863	-0.139129
…	…	…

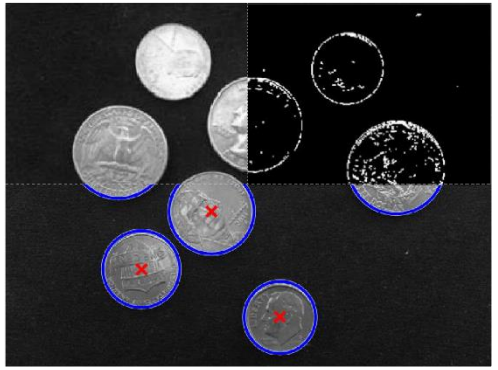
Existing Data Reconstruction Approaches

- **Manual reconstruction**
  - Overly rely on human labor

Existing Data Reconstruction Approaches

- **Manual reconstruction**
  - Overly rely on human labor

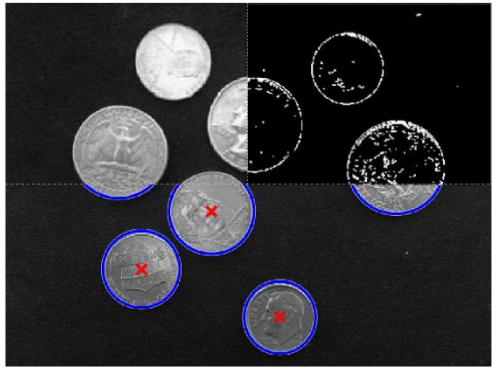
- **Image processing algorithms**
  - Overly rely on algorithm



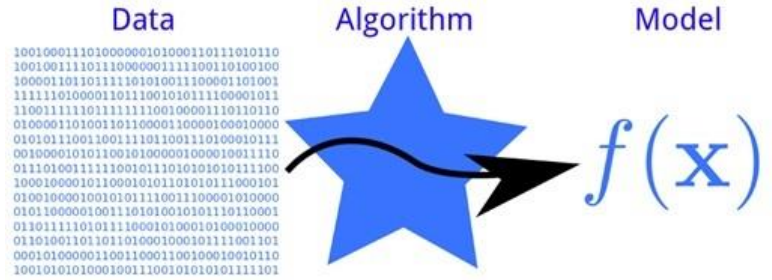
Existing Data Reconstruction Approaches

- **Manual reconstruction**
  - Overly rely on human labor

- **Image processing algorithms**
  - Overly rely on algorithm



- **Machine-learned models**
  - Overly rely on algorithm



Existing Data Reconstruction Approaches

- **Manual reconstruction**
  - Overly rely on human labor

- **Image processing algorithms**
  - Overly rely on algorithm

How to join human and machine strengths?

- **Machine-learned models**
  - Overly rely on algorithm

Outline

- **An Interactive Machine Learning Approach**
  - Case study of John Snow's cholera map
- **Generalization**
  - Modules for a universal data reconstruction software
- **Evaluation**
  - Interaction simulation

Outline

- **An Interactive Machine Learning Approach**
  - Case study of John Snow's cholera map
- **Generalization**
  - Modules for a universal data reconstruction software
- **Evaluation**
  - Interaction simulation

An Interactive Machine Learning Approach

- **Data reconstruction as Q & A ← Machine-Initiated**
  - Machine raises questions and user answers

An Interactive Machine Learning Approach

- **Data reconstruction as Q & A ← Machine-Initiated**
  - Machine raises questions and user answers

- **Design Principles ← Intelligent Interaction**

An Interactive Machine Learning Approach

- **Data reconstruction as Q & A ← Machine-Initiated**
  - Machine raises questions and user answers

- **Design Principles ← Intelligent Interaction**
  - Questions should be simple for the user but hard for the machine

An Interactive Machine Learning Approach

- **Data reconstruction as Q & A ← Machine-Initiated**
  - Machine raises questions and user answers

- **Design Principles ← Intelligent Interaction**
  - Questions should be simple for the user but hard for the machine
  - Machine should provide suggestions and learn from the user

An Interactive Machine Learning Approach

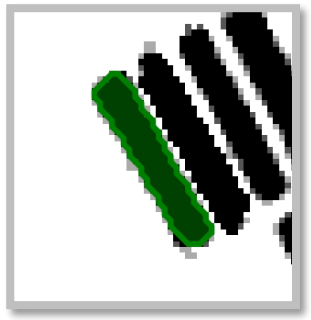
- **Data reconstruction as Q & A ← Machine-Initiated**
  - Machine raises questions and user answers

- **Design Principles ← Intelligent Interaction**
  - Questions should be simple for the user but hard for the machine
  - Machine should provide suggestions and learn from the user

An Active Learning Process

An Interactive Machine Learning Approach

- **Convert data reconstruction into classification**



Example classification task:
“Does the visual object encode data?”
Yes/No

- How to convert

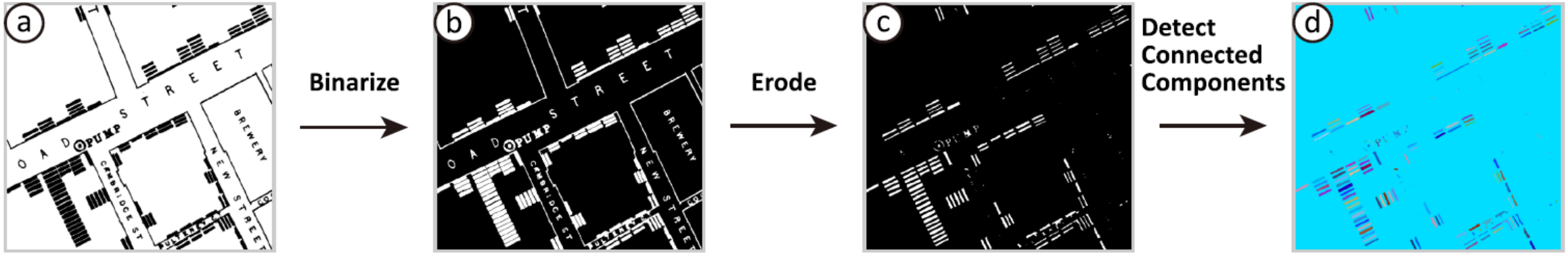
Example preprocessing to facilitate the conversion

An Interactive Machine Learning Approach

- Convert data reconstruction into classification

Example classification task:
"Does the visual object encode data?"
Yes/No

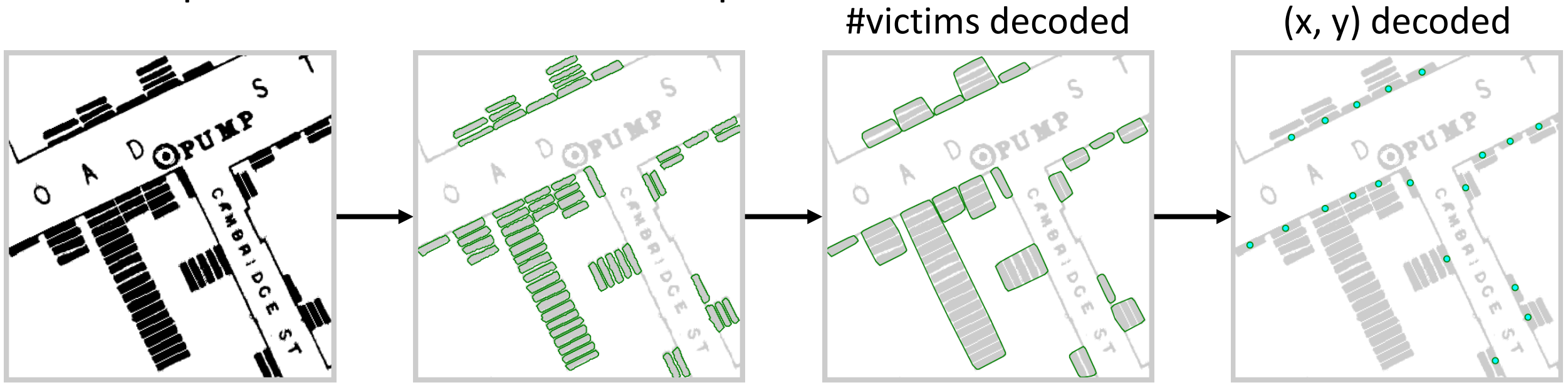
- How to convert



Example preprocessing to facilitate the conversion

Convert Data Reconstruction into Classification

- Example: John Snow's cholera map



"Does the visual object encode data?"

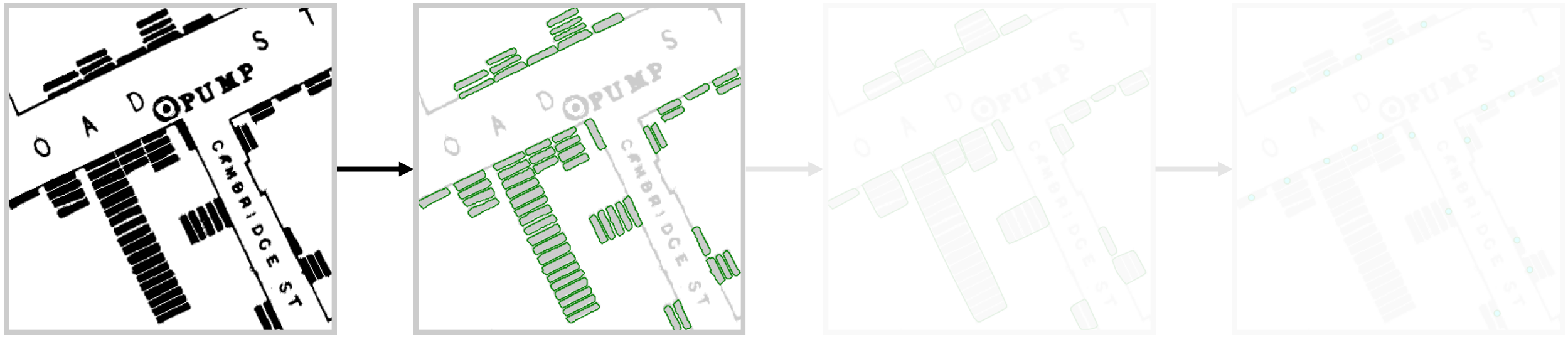
"Should the two visual objects be grouped?"

"Is this an anchor point?"

Convert Data Reconstruction into Classification

#victims decoded

(x, y) decoded



"Does the visual object encode data?"

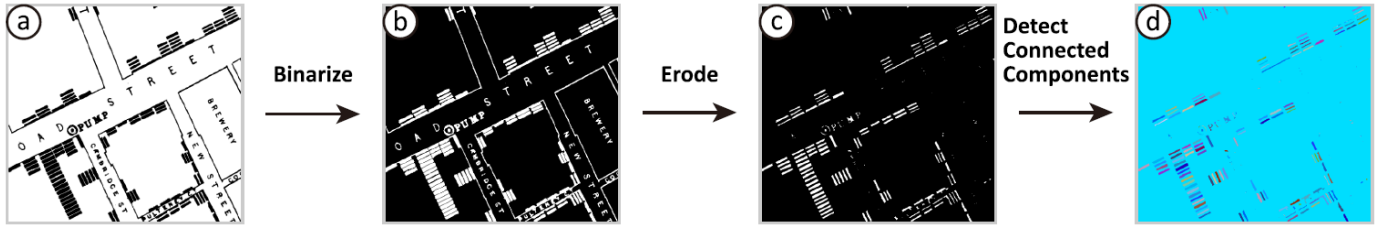
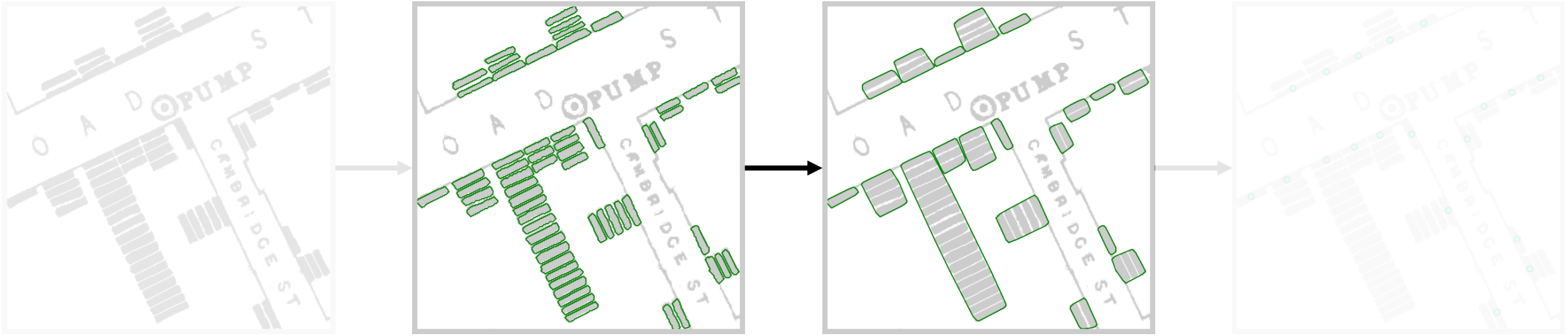


Image processing procedure for candidate visual object detection

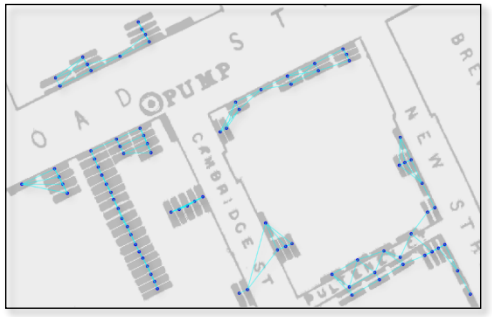
Convert Data Reconstruction into Classification

#victims decoded

(x, y) decoded



“Should the two visual objects be grouped?”

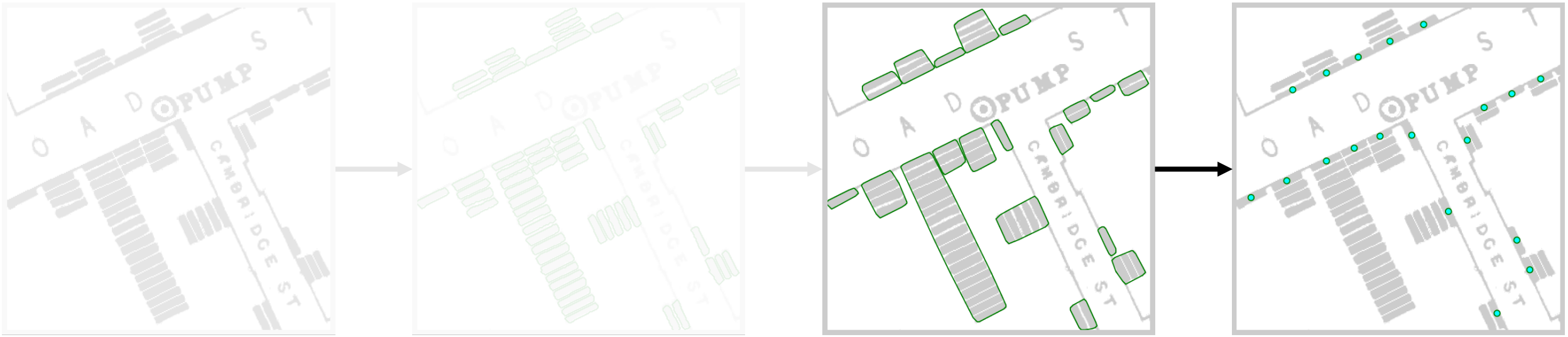


Build a graph with visual objects and classify edges

Convert Data Reconstruction into Classification

#victims decoded

(x, y) decoded

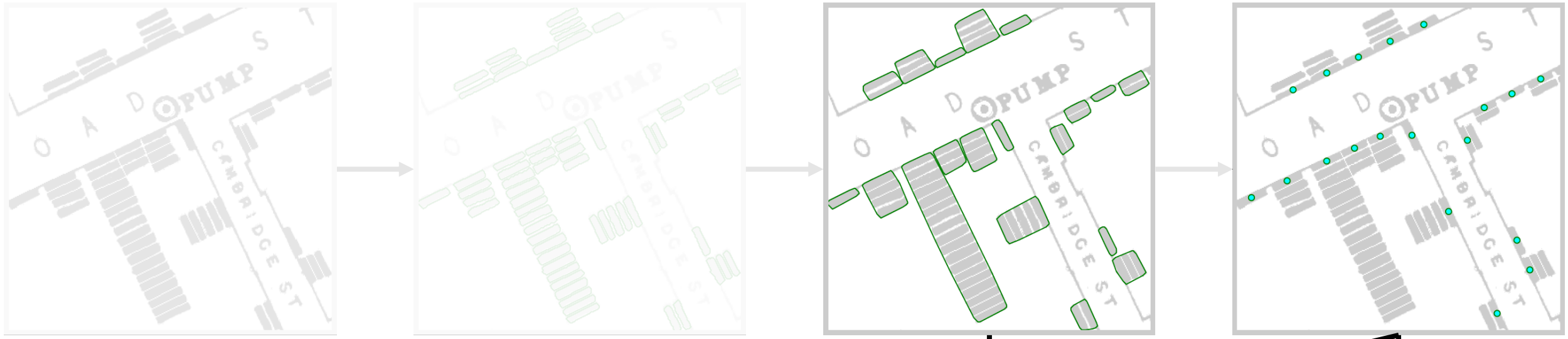


“Is this an anchor point?”

Data Reconstruction Outcome

#victims decoded

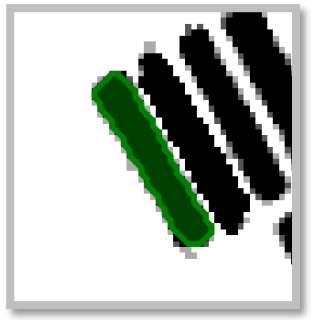
(x, y) decoded



#victims	latitude	longitude
1	51.514005	-0.139363
8	51.512863	-0.139129
…	…	…

An Interactive Machine Learning Approach

- **Convert data reconstruction into classification**

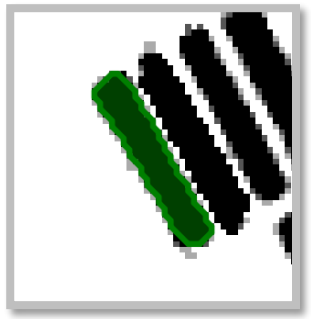


Example classification task:
“Does the visual object encode data?”
Yes/No

- **Save user interaction efforts in classification**

An Interactive Machine Learning Approach

- **Convert data reconstruction into classification**

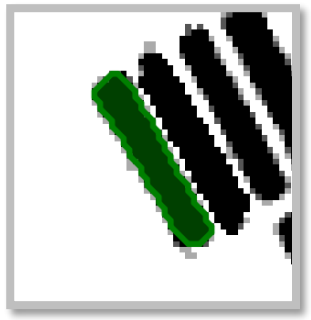


Example classification task:
"Does the visual object encode data?"
Yes/No

- **Save user interaction efforts in classification**
  - Active sampling *"Please answer these questions first that I'm most uncertain about."*
    - Example implementation: uncertainty-based sampling in active learning

An Interactive Machine Learning Approach

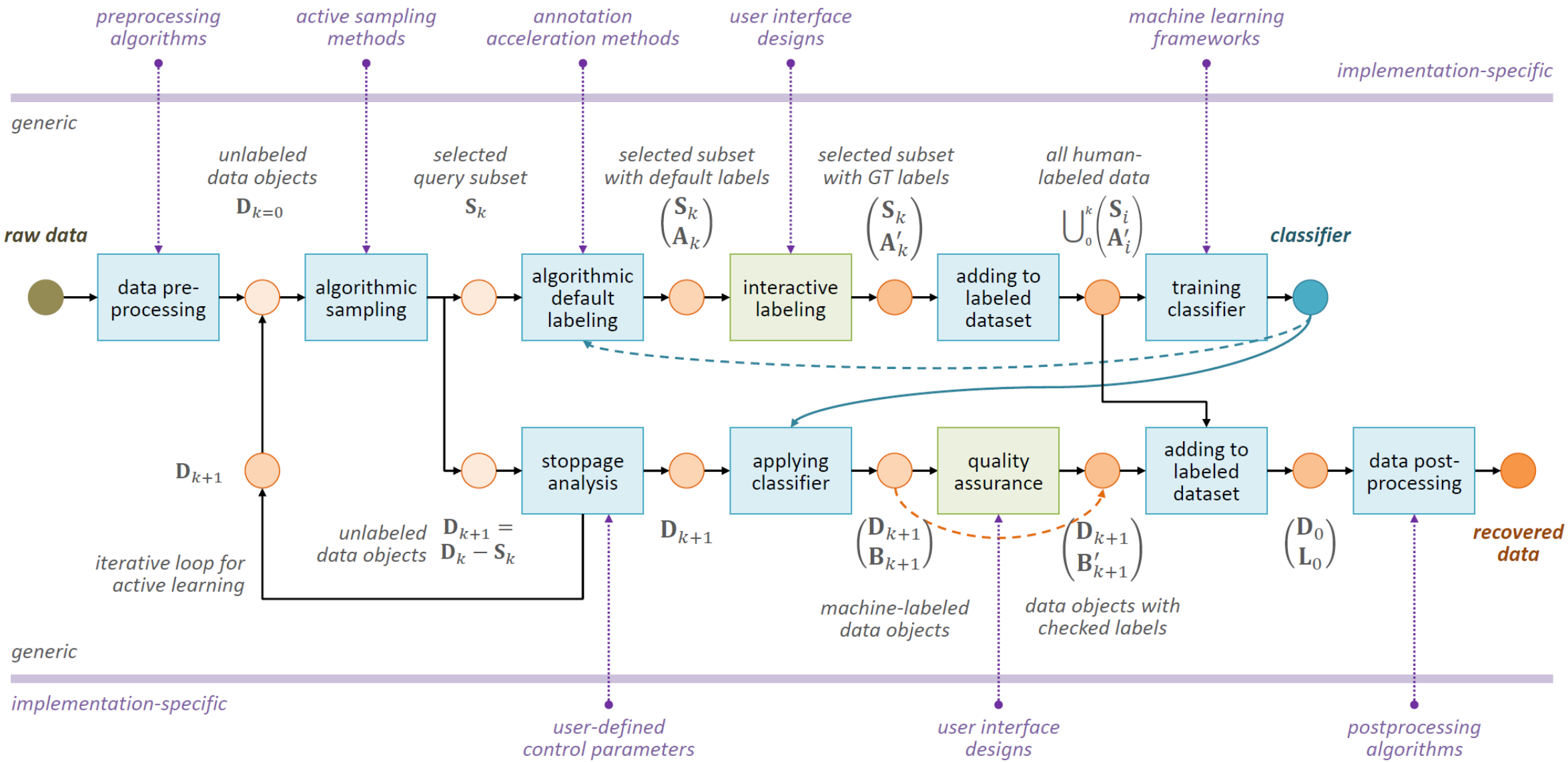
- **Convert data reconstruction into classification**



Example classification task:
"Does the visual object encode data?"
Yes/No

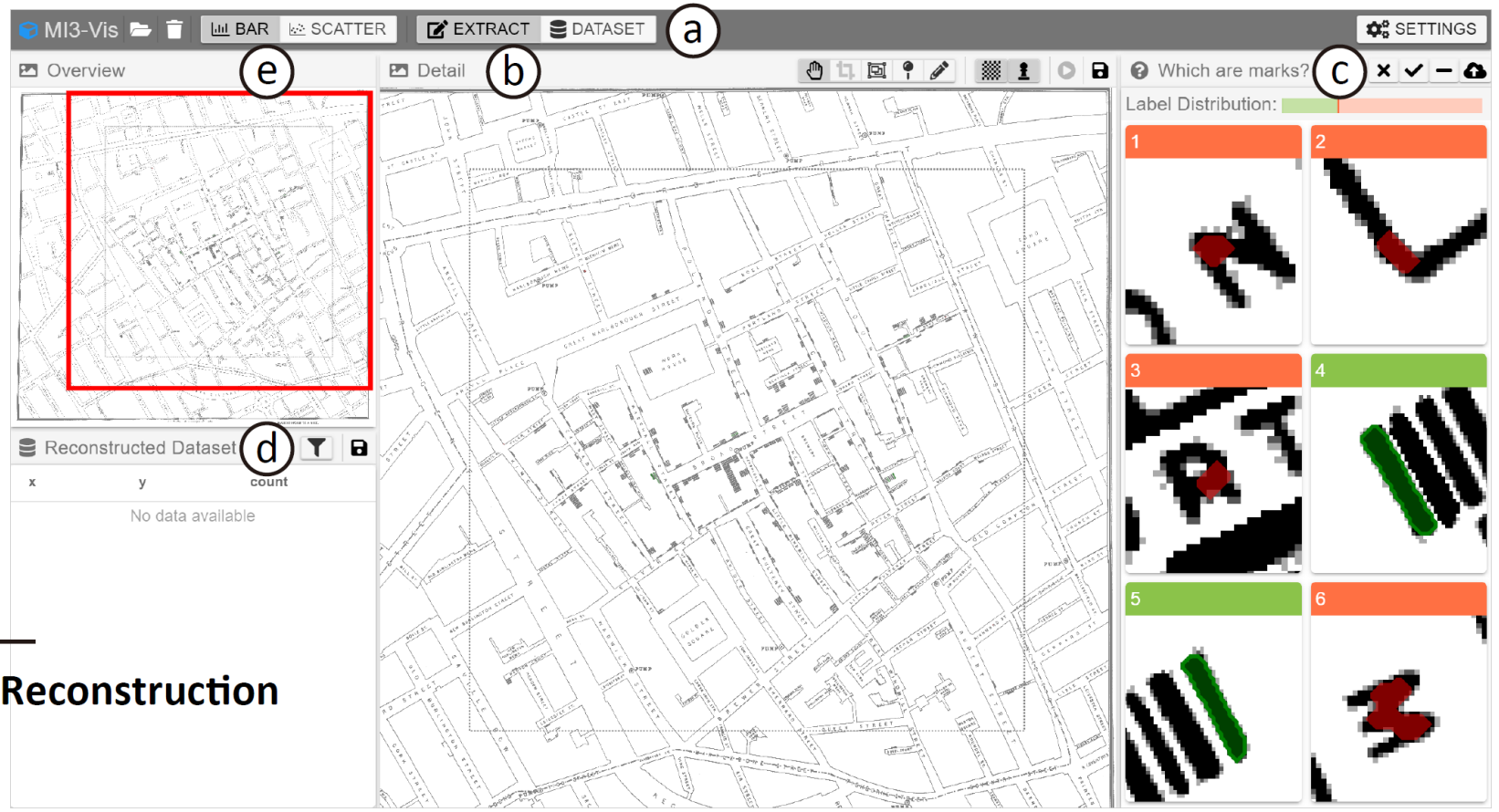
- **Save user interaction efforts in classification**
  - Active sampling *"Please answer these questions first that I'm most uncertain about."*
  - Default labeling *"I have presented tentative answers. Are they Correct?"*
    - Example implementation: graph-based label propagation

A Data Reconstruction Workflow



A Data Reconstruction System

Data reconstruction system



Reconstructed Dataset

latitude	longitude	#victims
51.514005	-0.139363	1
51.513002	-0.139293	1
51.513423	-0.139286	1
51.512884	-0.139282	1
51.513595	-0.139268	1
51.512799	-0.139257	2
51.513517	-0.139183	1
51.512943	-0.139157	1
51.512863	-0.139129	8
51.513225	-0.139054	1
51.512934	-0.138981	2
51.512973	-0.138904	1
51.513614	-0.138853	1
51.513335	-0.138822	1
51.513661	-0.138770	1
51.513527	-0.138765	1
51.513699	-0.138764	2
51.513378	-0.138724	2
51.513588	-0.138694	1
51.513108	-0.138640	1
51.513197	-0.138599	1
51.512805	-0.138588	1
51.513409	-0.138504	1
51.512250	-0.138444	1
51.513170	-0.138393	3

After Reconstruction

Reconstructed dataset <latitude, longitude, #victims>[]

Outline

- **An Interactive Machine Learning Approach**
  - Case study of John Snow's cholera map
- **Generalization**
  - Modules for a universal data reconstruction software
- **Evaluation**
  - Interaction simulation

Composable Data Reconstruction Modules

- **Component-based visual object detection**

“Does the visual object encode data?”

- **Visual object grouping**

“Should the two blocks be grouped?”

- **Position measurement**

“Is this an anchor point?”

Composable Data Reconstruction Modules

- Visual object **detection** modules
  - Component-based visual object detection
  - Tracking-based visual object detection
  - …
- Visual object **organization** modules
  - Visual object grouping
  - Visual object segmentation
  - …
- Visual variable **measurement** modules
  - Position Measurement
  - Area measurement
  - …

- **> 150 historical visualization images**
- **18 composable modules**

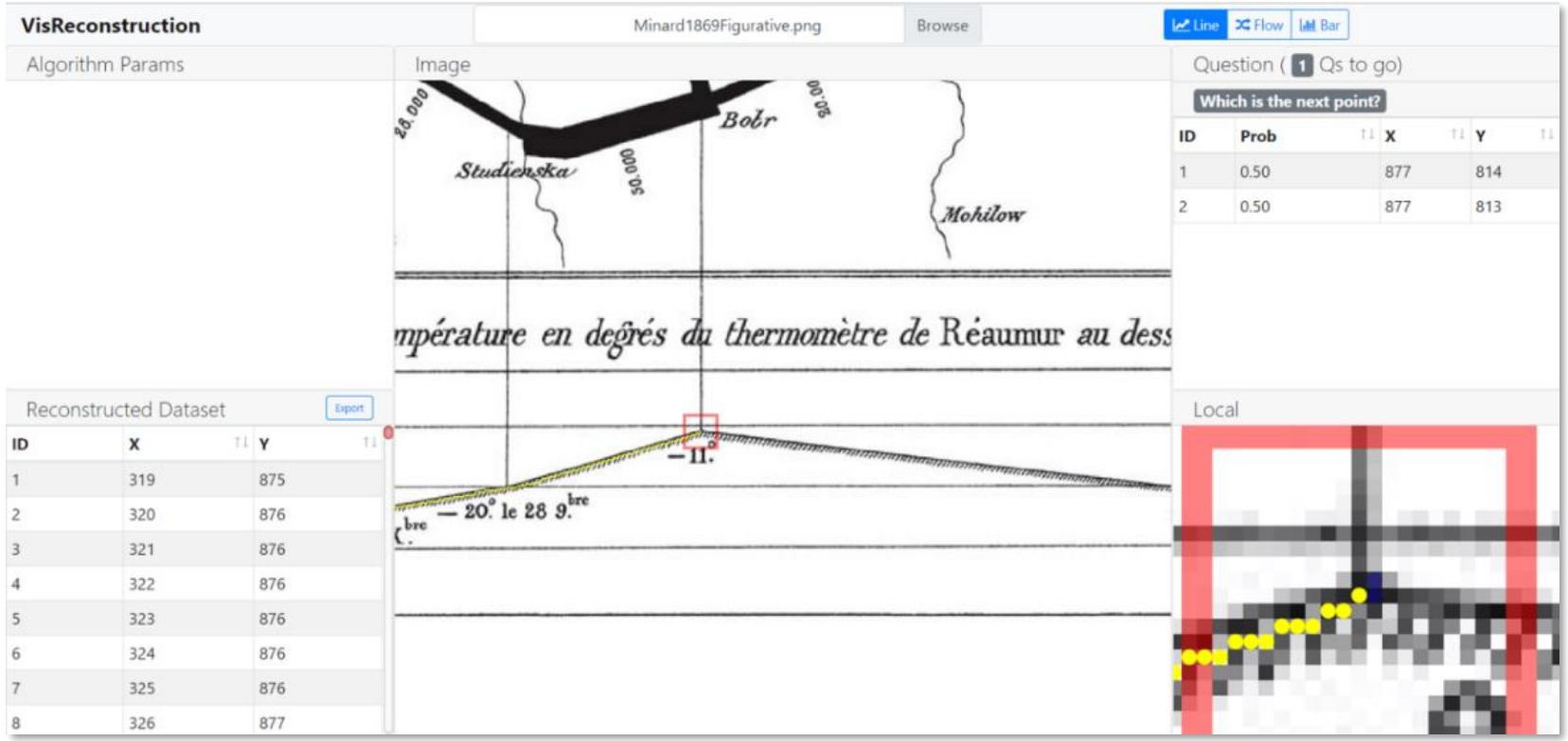
Composable Data Reconstruction Modules

- Visual object **detection** modules
  - Component-based visual object detection
  - Tracking-based visual object detection
  - …
- Visual object **organization** modules
  - Visual object grouping
  - Visual object segmentation
  - …
- Visual variable **measurement** modules
  - Position Measurement
  - Area measurement
  - …

- **> 150 historical visualization images**
- **18 composable workflows**

Composable Data Reconstruction Modules

- Line chart data reconstruction
  - Tracking-based visual object detection
  - Position measurement



Reconstructed dataset <x, y>[]

Outline

- **An Interactive Machine Learning Approach**
  - Case study of John Snow's cholera map
- **Generalization**
  - Modules for a universal data reconstruction software
- **Evaluation**
  - Interaction simulation

Evaluation Setup

- **Compared approaches**
  - Manual reconstruction
  - Pre-trained machine learning models

Evaluation Setup

- **Compared approaches**
  - Manual reconstruction
  - Pre-trained machine learning models
  - ✗ active sampling + ✗ default labeling
  - ✓ active sampling + ✗ default labeling
  - ✗ active sampling + ✓ default labeling
  - ✓ active sampling + ✓ default labeling

Evaluation Setup

- **Compared approaches**
  - Manual reconstruction
  - Pre-trained machine learning models
  - ✗ active sampling + ✗ default labeling
  - ✓ active sampling + ✗ default labeling
  - ✗ active sampling + ✓ default labeling
  - ✓ active sampling + ✓ default labeling
- **Method**: simulation

Evaluation Setup

- **Compared approaches**
  - Manual reconstruction
  - Pre-trained machine learning models
  - ✗ active sampling + ✗ default labeling
  - ✓ active sampling + ✗ default labeling
  - ✗ active sampling + ✓ default labeling
  - ✓ active sampling + ✓ default labeling
- **Method**: simulation
- **Metric**: #interactions (correction, confirmation, creation)
- **Dataset**: John Snow's cholera map

Evaluation Setup

- **Compared approaches**
  - Manual reconstruction
  - Pre-trained machine learning models
  - ✗ active sampling + ✗ default labeling
  - ✓ active sampling + ✗ default labeling
  - ✗ active sampling + ✓ default labeling
  - ✓ active sampling + ✓ default labeling
- **Method**: simulation
- **Metric**: #interactions (correction, confirmation, creation)
- **Dataset**: John Snow's cholera map

Evaluation Results

- **Compared approaches**
  - Manual reconstruction
  - Pre-trained machine learning models
  - ✗ active sampling + ✗ default labeling
  - ✔ active sampling + ✗ default labeling 🏆
  - ✗ active sampling + ✔ default labeling 🏆
  - ✔ active sampling + ✔ default labeling 🏆🏆

Evaluation Results

- **Compared approaches**
  - Manual reconstruction
  - Pre-trained machine learning models
  - ✗ active sampling + ✗ default labeling
  - ✔ active sampling + ✗ default labeling 🏆
  - ✗ active sampling + ✔ default labeling 🏆
  - ✔ active sampling + ✔ default labeling 🏆🏆

Next Steps

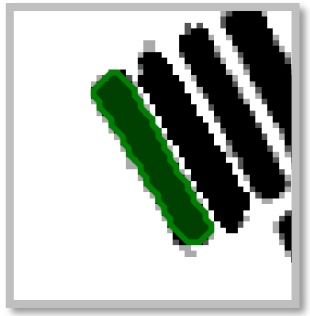
- Increase coverage on data reconstruction modules

Next Steps

- Increase coverage on data reconstruction modules
- Evaluate with relaxed assumption

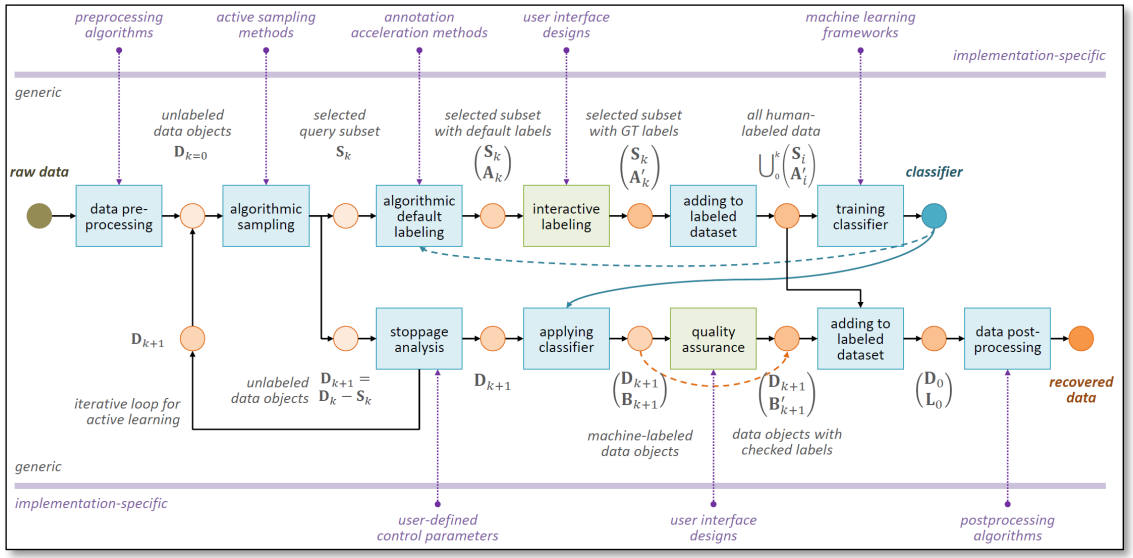
Summary

- Data reconstruction as classification
  - Composable reconstruction modules



Example classification task:
"Does the visual object encode data?"
Yes/No

- Machine-assistance in classification
  - Active sampling
  - Default labeling



Thank you!


Yu Zhang


Bob Coecke


Min Chen

University of Oxford