

Multi-Task Representation Learning

Learning diverse and useful representations
from multiple tasks

Position paper

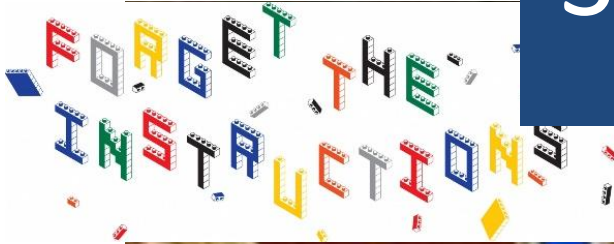
**Mohamed-Rafik Bouguelia, Sepideh Pashami
and Slawomir Nowaczyk**

~~Diagnostics?~~ ~~Maintenance?~~

FORGET THE
INSTRUCTION



Self-Monitoring



Overall Goals



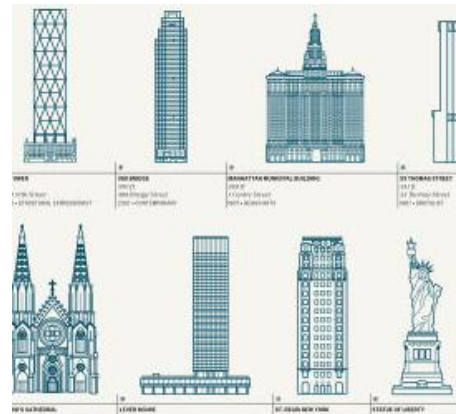
General methods



Learn from data and experts

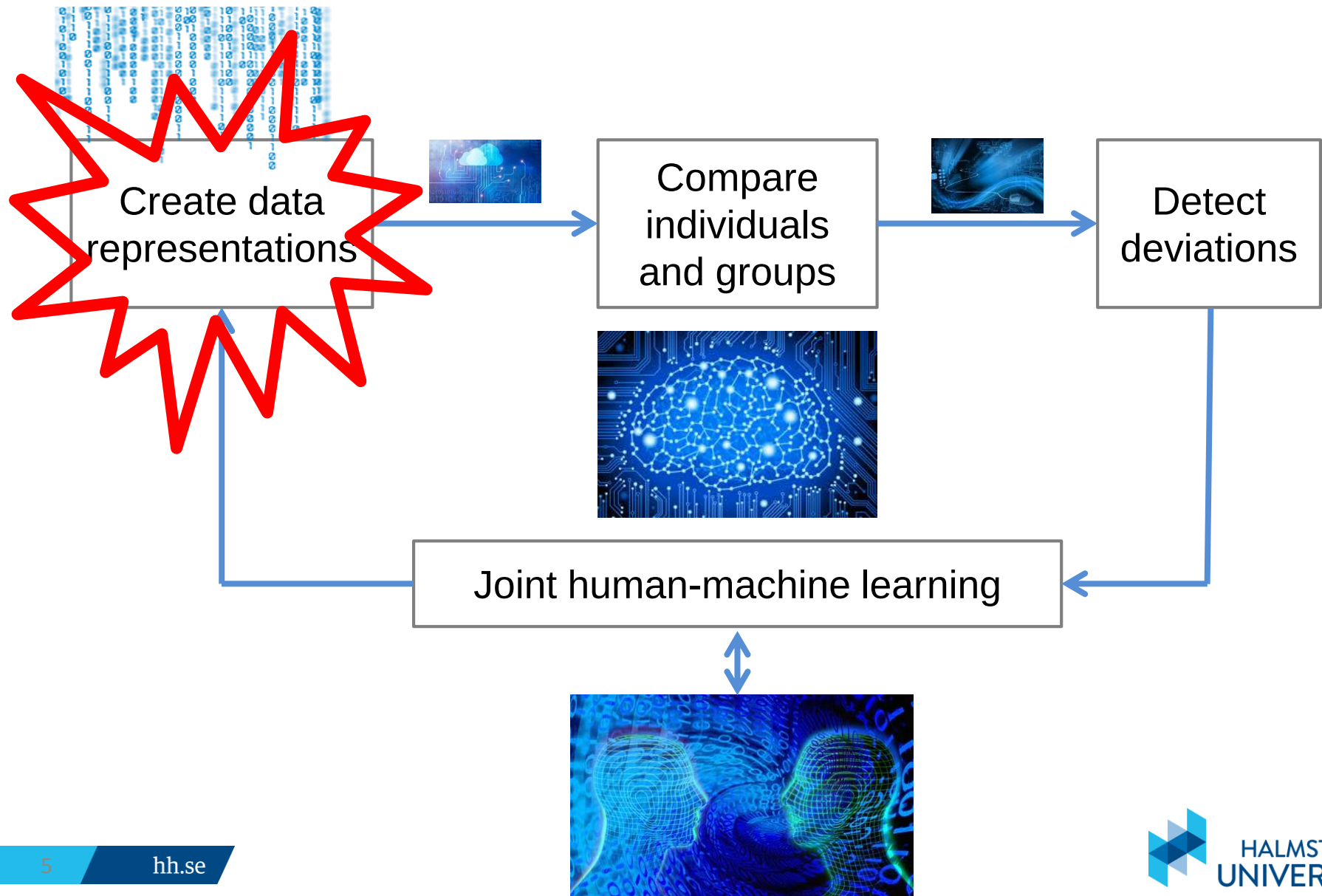


Reusing knowledge



Multiple domains

SeMI Approach

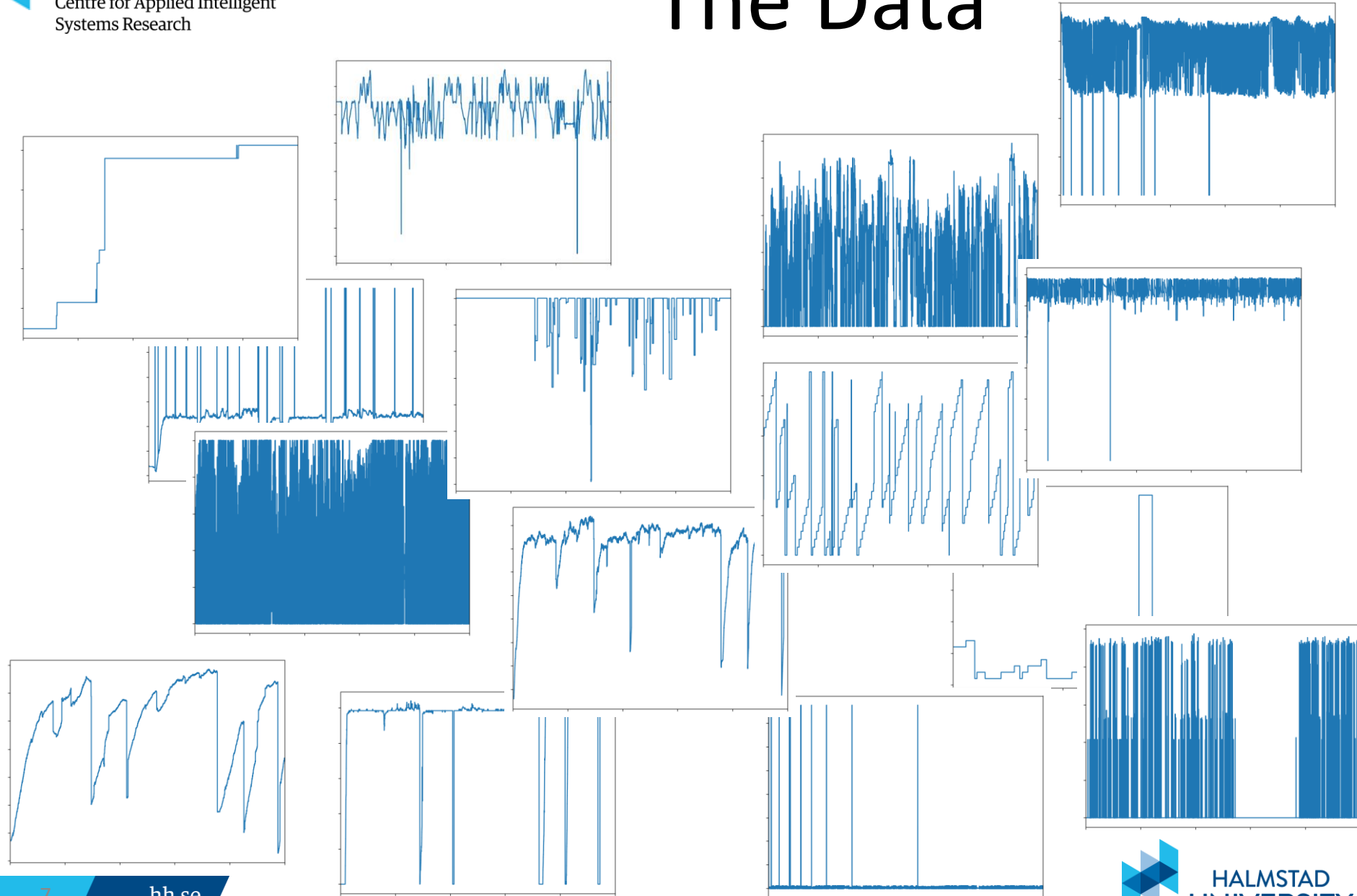


Main Objective

Build an intelligent
and aware system

Handle events that are
unknown at the time of
system design

The Data



Representation Learning

- Not a new field, but somewhat revived by DNNs
 - some very impressive results for images and video
 - earlier work used, e.g., network data (cf. Tina's invited talk)
- Can those ideas be generalised to other types of data?
 - heterogeneous sensor network data
 - describing operation of physical systems
 - with self-monitoring as the main purpose

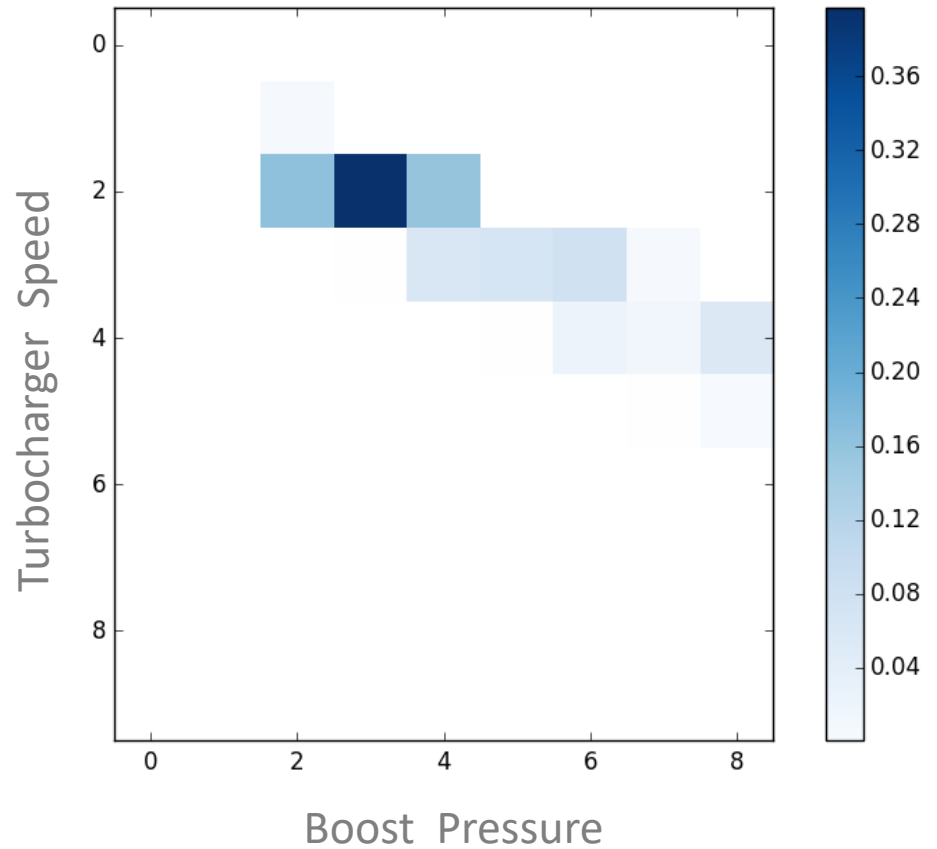
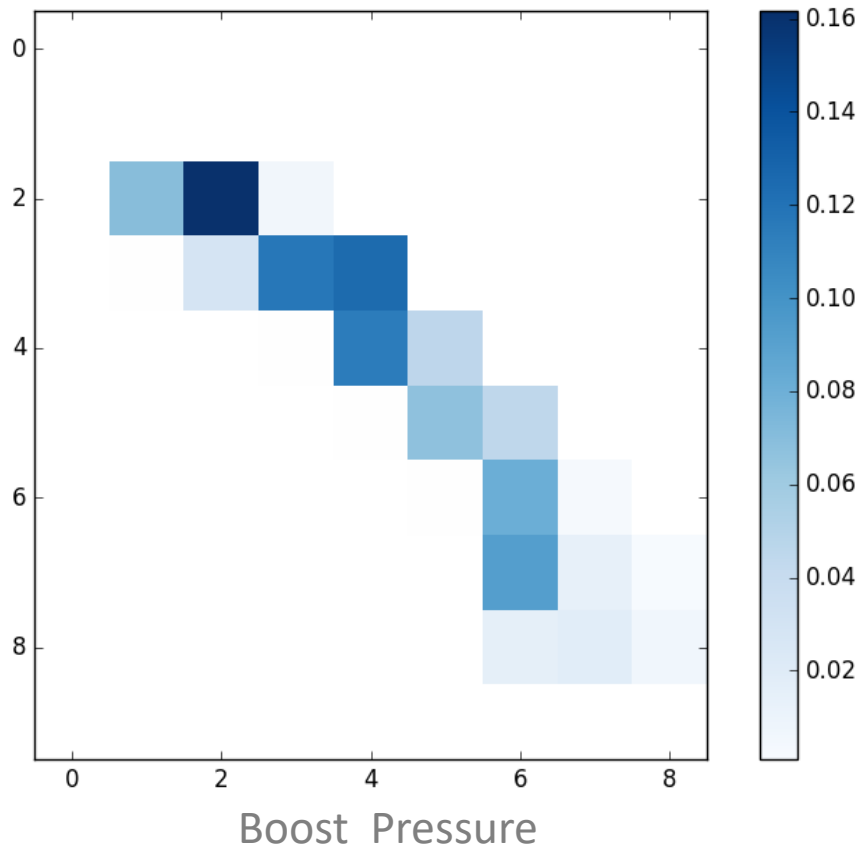
Motivation

- Traditional feature engineering is not sufficient
 - tedious, time consuming, expensive, error-prone and subjective
 - does not necessarily capture all the different aspects of the data
 - is not likely to generalize to other tasks sufficiently well
- Big Data & AI
 - a lot of data is being collected all around us
 - often without a clear idea of how and why
 - later re-used for solving different problems

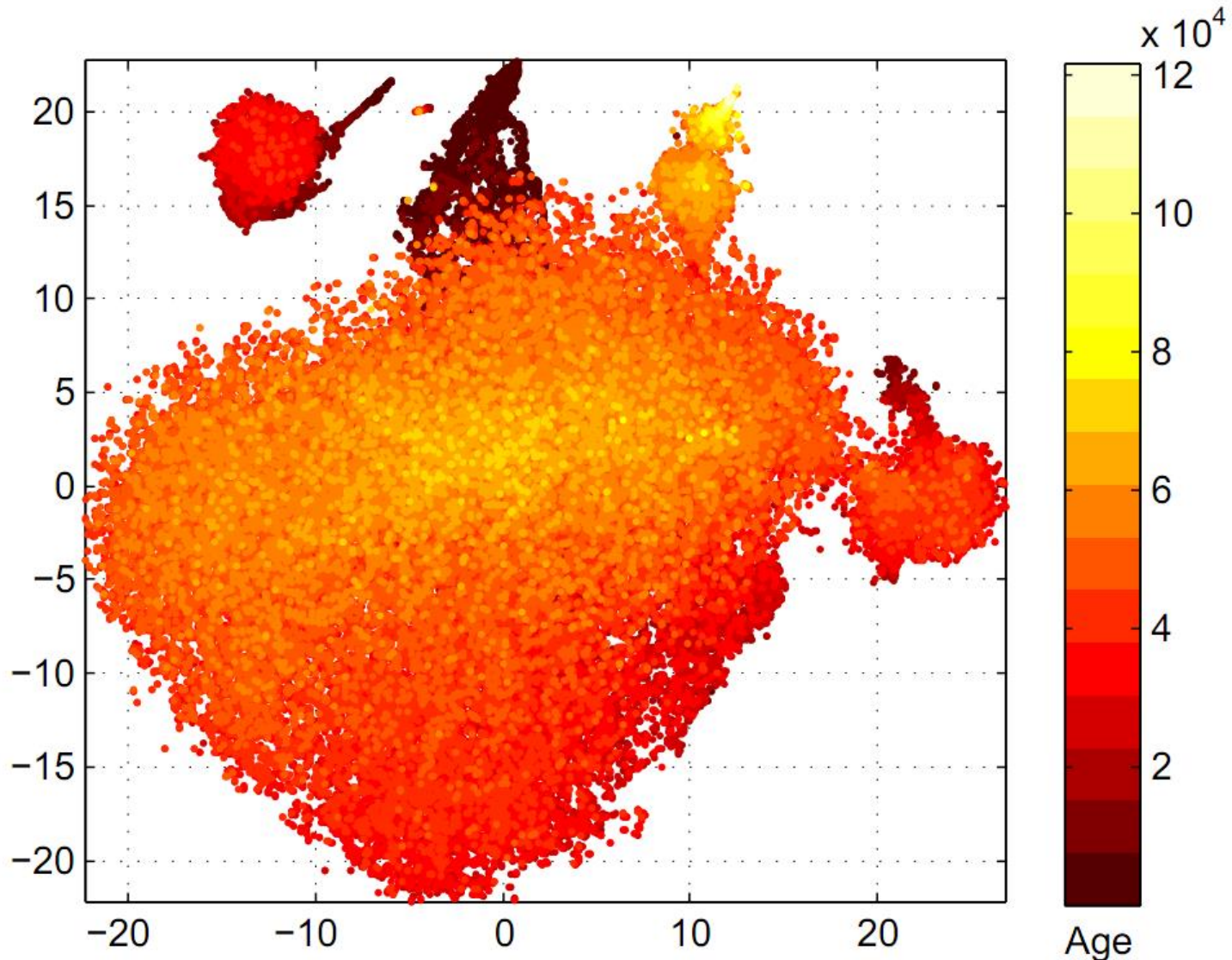
Representation Learning

- Transforming raw data into useful representations
 - what is “useful” very clearly depends on the goal
 - good representations need to work for many tasks
- Unsupervised Representation Learning
 - yields very general features, capturing all the information
 - results are not specifically designed as important for given task(s)
- Supervised Representation Learning
 - learn representations that are tailored for a specific task
 - no direct provision for applying them to another task(s)

Volvo Truck Example



100.000 Volvo Trucks



Visualised using t-SNE

Overall Goal

- Volvo predictive maintenance solution
 - forecast future failures for an individual truck
- Limited amount of training data
 - labelled samples are rare or expensive to get
- Automatically learn high-level features
 - allows re-use of knowledge for other tasks
 - more informative and less prone to overfitting
 - lead to models that are easier to understand
 - data compression & computational complexity

Open Questions

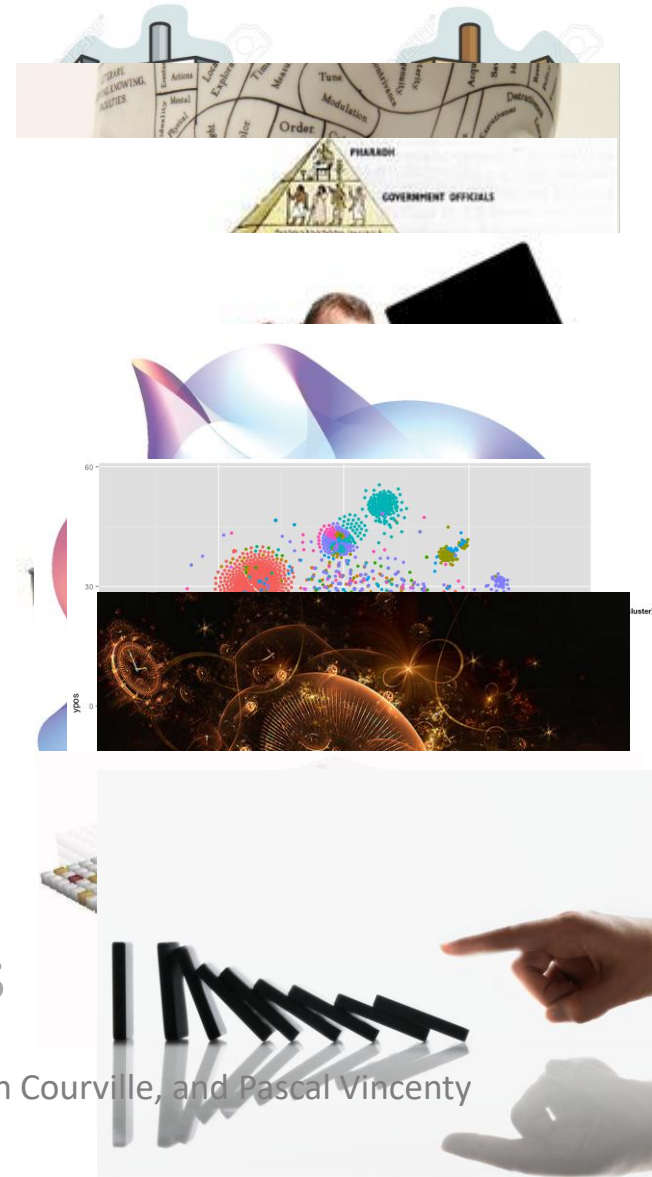
- In images we have “lines” and “corners”
 - what are the equivalents in streaming data?
- How do we evaluate new representations?
 - in a general way, not focusing on a single task
 - internal evaluation of the resulting structure
 - external evaluation based on extra information
- Can we measure generality of a feature?
 - across multiple (unknown *a priori*) tasks
 - balance *generality*, *precision* and *compactness*?

Objectives

- Map raw sensor data into a low-dimensional space
 - compression
- Allow for accurate and interpretable predictions
 - discard irrelevant information
- Extract general features and re-use knowledge
 - useful across many related, but not identical tasks

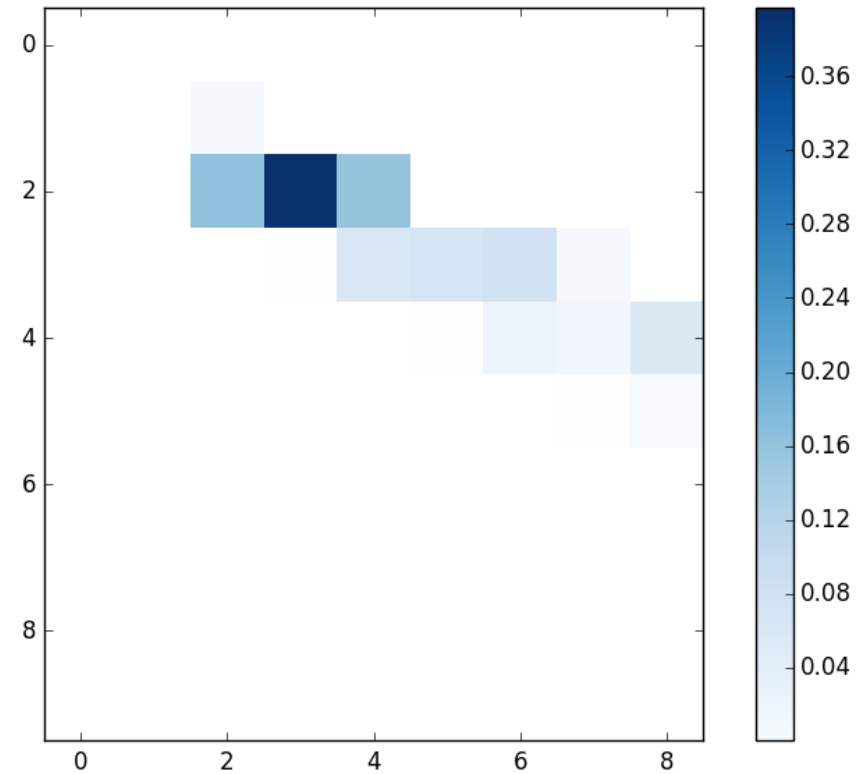
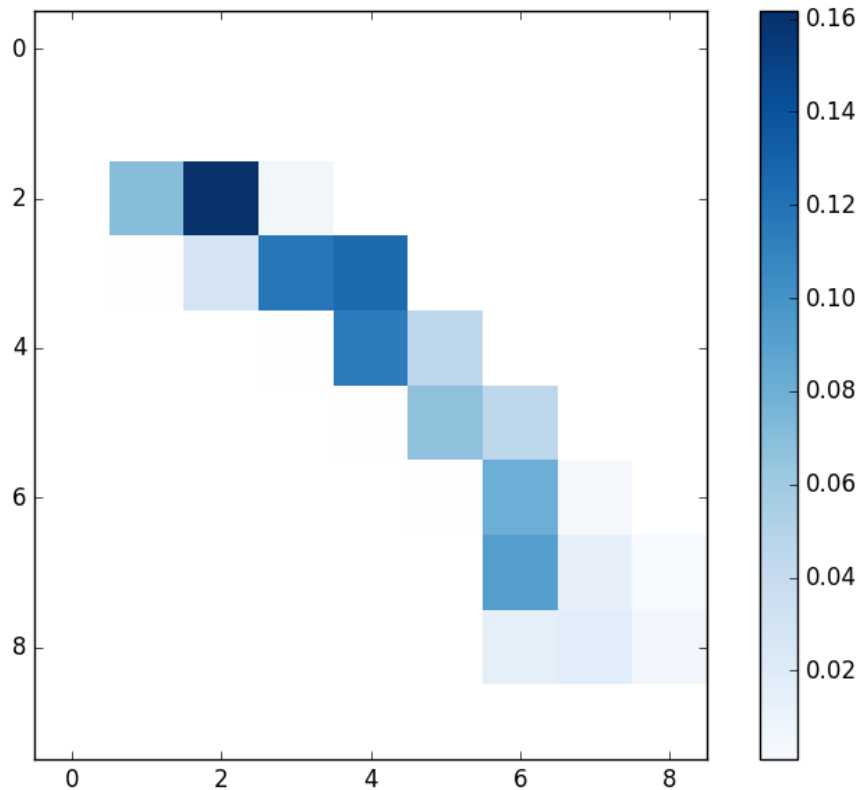
Representation Learning

- Smoothness
- Multiple explanatory factors
- Hierarchical organization
- Shared factors across tasks
- Manifolds
- Natural clustering
- Temporal and spatial coherence
- Sparsity
- Simplicity of factor dependencies



Representation Learning: A Review and New Perspectives, Yoshua Bengio, Aaron Courville, and Pascal Vincenty

Volvo Data: 90-dimensional histograms



Evaluation Methods

- Internal evaluation measures
 - Accuracy, ARI, NMI, ...
- Evaluation using different classification tasks
 - Country: Canada, USA & Other
 - Truck type: Rigid & Tractor
 - Gearbox: AT2612D, ATO2612D & Other
 - Marketing type
 - Emission level standard
 - Fuel capacity

Gearbox: AT2612D, ATO2612D & Other

NMI:

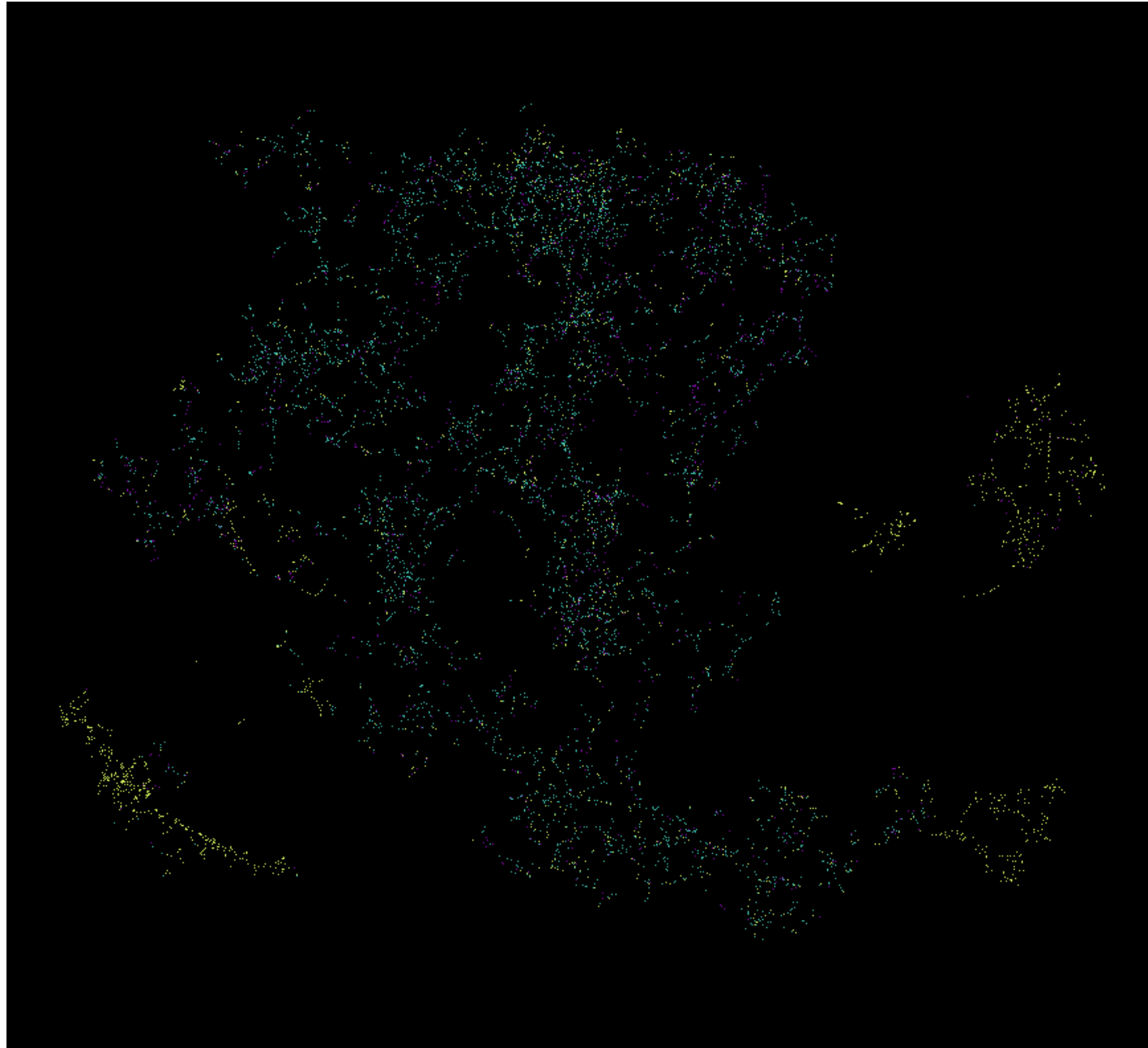
0.0636

0.1274

0.0614

0.0006

t-SNE



NMI:

0.0636

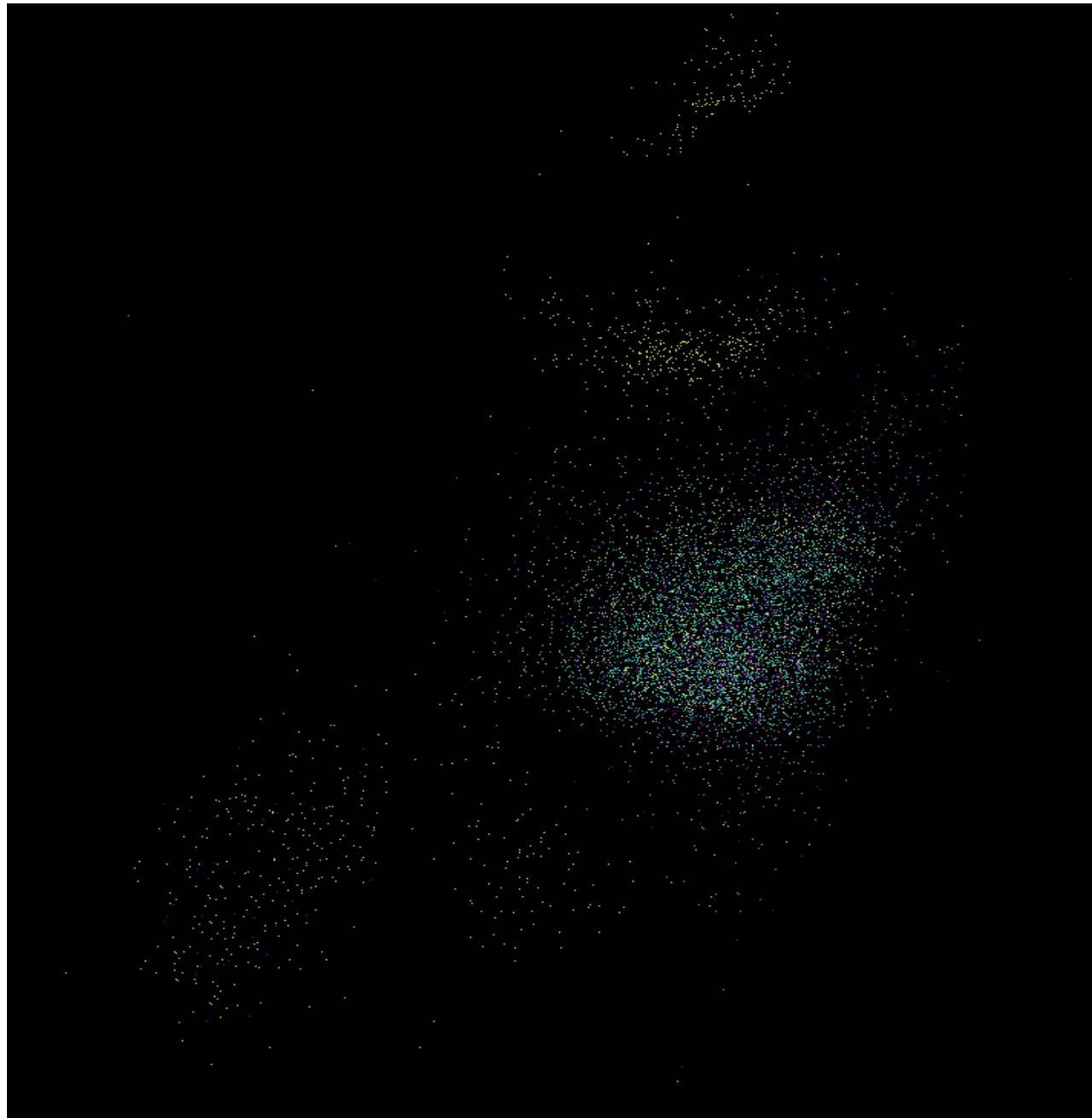
0.1274

0.0614

0.0006

CAE

1 layer 32 filters



NMI:

0.0636

0.1274

0.0614

0.0006

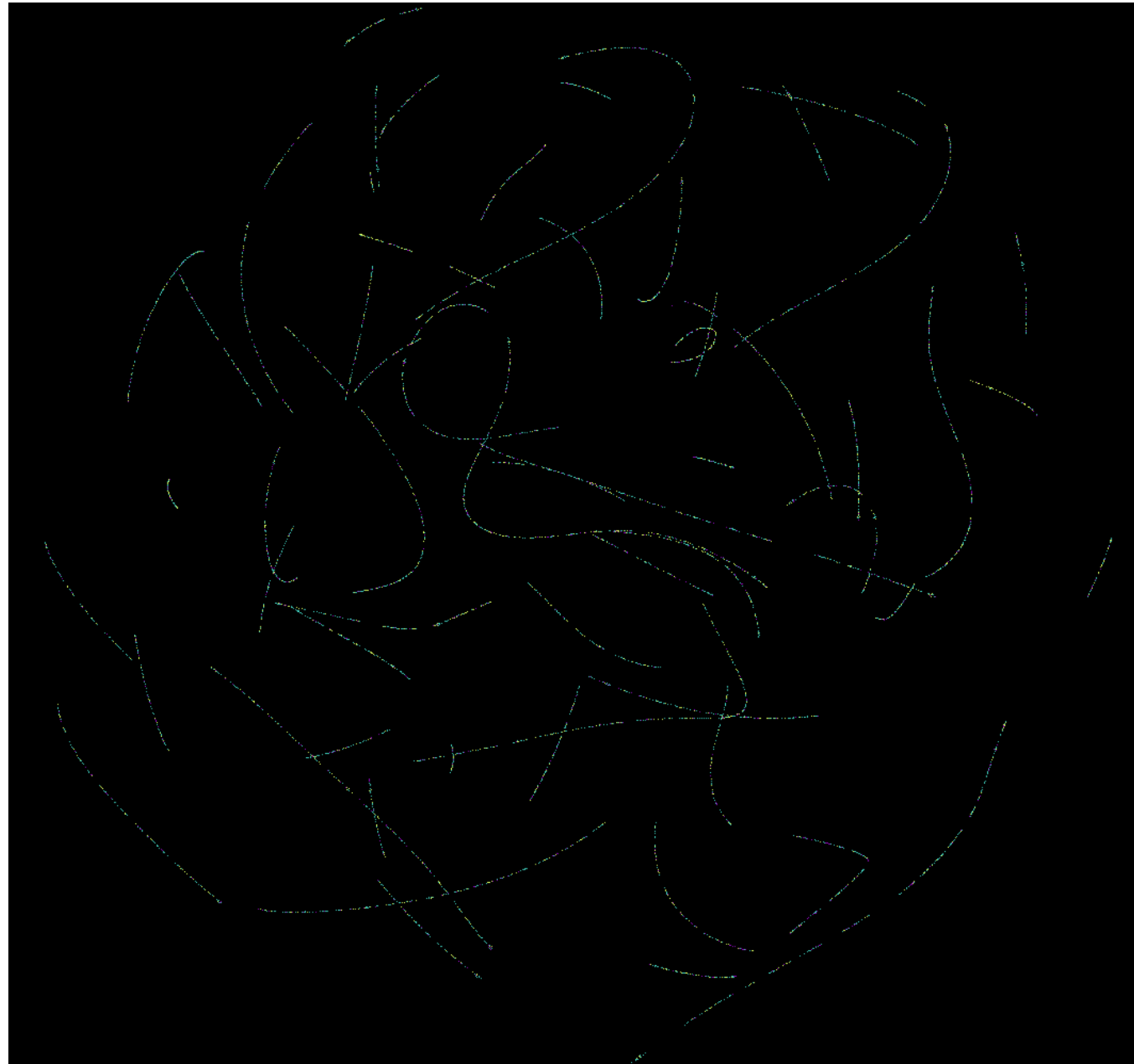
CAE

1 layer 32 filters
(using RPCA)



NMI:
0.0636
0.1274
0.0614
0.0006

Hybrid
CAE + t-SNE



Problem formulation

- Supervised learning
 - Training examples sampled from unknown D
 - Labelled by an unknown function T
 - Goal: hypothesis H minimizing $P_{x \sim D}(T(x) \neq H(x))$
- Multi-task representation learning
 - Unknown distribution of data examples D_E
 - Unknown distribution over tasks D_T
 - A set of labelling for the training data: $T_i \sim D_T$
 - Goal: group of representations with the lowest error on unseen data across all the expected tasks.

Challenges:

Representation Generation

- Immense space of possible solutions to consider
- For a given problem, is it sufficient to consider a single family of representations, or are multiple needed?
- What kind of problem properties can/should be influencing the choice of one family over another?
- How to balance specificity and diversity in a way that leads to an optimal/improved accuracy?

Challenges:

Representation Evaluation

- Define clear and measurable objective for the complete learning process
- How to go beyond evaluating outcome representation in terms of its usefulness for a single task?
- To what degree can/should unsupervised evaluation measures, such as reconstruction error, be used?
- How to improve the generality of representations?

Research Questions

- When to generate multiple diverse representations, rather than a single multi-dimensional representation?
- How many should be created?
- How to explicitly promote the diversity within the set of representations being created?
- How to measure and optimize the trade-off between accuracy on each task and representation diversity?

Research Questions

- How to generate a family of representations that ensures a good coverage of the space of mapping functions?
- How to measure the coverage?
- Is the family expressive and non-redundant enough?
- How to relate the expressiveness of the family with the complexity of the expected tasks (e.g. VC dimension)?

Research Questions

- Can one identify and group related tasks, and learn good representations for each group separately?
- How the similarity among tasks should be measured?
- How to balance different quality aspects?
 - Data compression (or compactness of representations)
 - Accuracy relative to each task,
 - Generality of representations across tasks?
- How to quantify the expected benefit of MTRL?

Research Questions

- How to define complexity of tasks, together with other problem instance properties, in a principled way?
- What problem instance properties influence
 - The choice of algorithms
 - The representation families
 - The quality measures
- How to measure the complexity or difficulty of tasks?

Approaches & Ideas

- Complexity of future tasks based on training tasks
 - generalize the VC dimension to a family of tasks
- Modeling the similarity between different tasks
 - compare parameters of models learned from individual tasks
 - estimate how well a one task model performs across other tasks
- Develop well-founded evaluation measures
 - extended PAC learning theory, where assumptions related to learning from both data and tasks can be established.

Approaches & Ideas

- Generating representations from adequate families, based on the problem properties.
- Investigate the Functional Data Analysis, now gaining popularity in mathematical statistics
- Ensure a trade-off between the “exploration of all possible representations” and the “exploitation of the best generated representations”

Approaches & Ideas

- Encourage diversity among representations and generality of the representations across tasks
 - On the algorithm level
 - Create representations using sufficiently different algorithms
 - Explicitly control the bias and variance for a family of algorithms
 - On the mapping level
 - Consider representations as functions from one feature space to another
 - Compare the functions based on their mathematical properties
 - On the example level
 - Capture the relative positions for the data points in the training sets
 - Measure how those positions change across different representations

Emission level standard:
EU+JPN, USA14, USA34, Other

