What are local anomalies?
Local anomalies or outliers are points that are anomalous with respect to their immediate neighbourhood instead of the data set as a whole. Examples:
- Credit card fraud
- Network intrusion
- Other abnormal behaviour in a local region (e.g. stock market time period)
- Cancer cells
- Image data with occlusion and shadows
- Free text anomalies
Examples: blue X in graph below are locally distinct but not globally (some red points are bigger global outliers)

Standard (global) outlier example
From real world data set of object images
- 24,000 pictures of 1000 objects, 24 pictures per object, varying the direction of lighting
- Images are 192x144 pixels = 27648 ambient dimensionality, but low intrinsic dimensionality
- Top Global outliers correspond to distinct and bright shapes that are most different from every other object in the data set

Local outlier example
- The 24 pictures of each object form a cluster in the data set
- When a picture is changed substantially via lighting changes or the appearance of shadows, the picture may move a short distance away from the cluster, forming a local outlier
- The top local outliers of the whole data set are among the most occluded by shadows and low illumination – difficult cases for image recognition

The Algorithm (PINN)
Allows us to find local outliers in very large and high-dimensional real data

We use dimensionality reduction to project data to a lower dimension
- Allows efficient indexing
- Near-linear algorithm, improvement from O(n²)
- Distances between points are preserved through projection
- First known method for high-dimensional local outliers:

Results
Accuracy and performance on real world data sets from a wide range of different type (Image, Text, Medical, Mixed)
Red and black lines refer to the new method (with parameter changes). Pink is the standard approach without optimisation
Blue is the baseline standard approach Random Projection. Green is baseline Principal Components Analysis.