

Which of these do not belong?

Anomaly Detection In Images Using A Deep Neural Network & Isolation Forest



Mark Wilber

Galvanize Hiring day - October 1st, 2015

Motivation

Much of anomaly detection is done using supervised learning.

- Existing data sets, labeled anomalies (fraud, network intrusions, etc.)
- Train classifiers on anomalies we know about: “known unknowns”

What about things we have not yet observed?

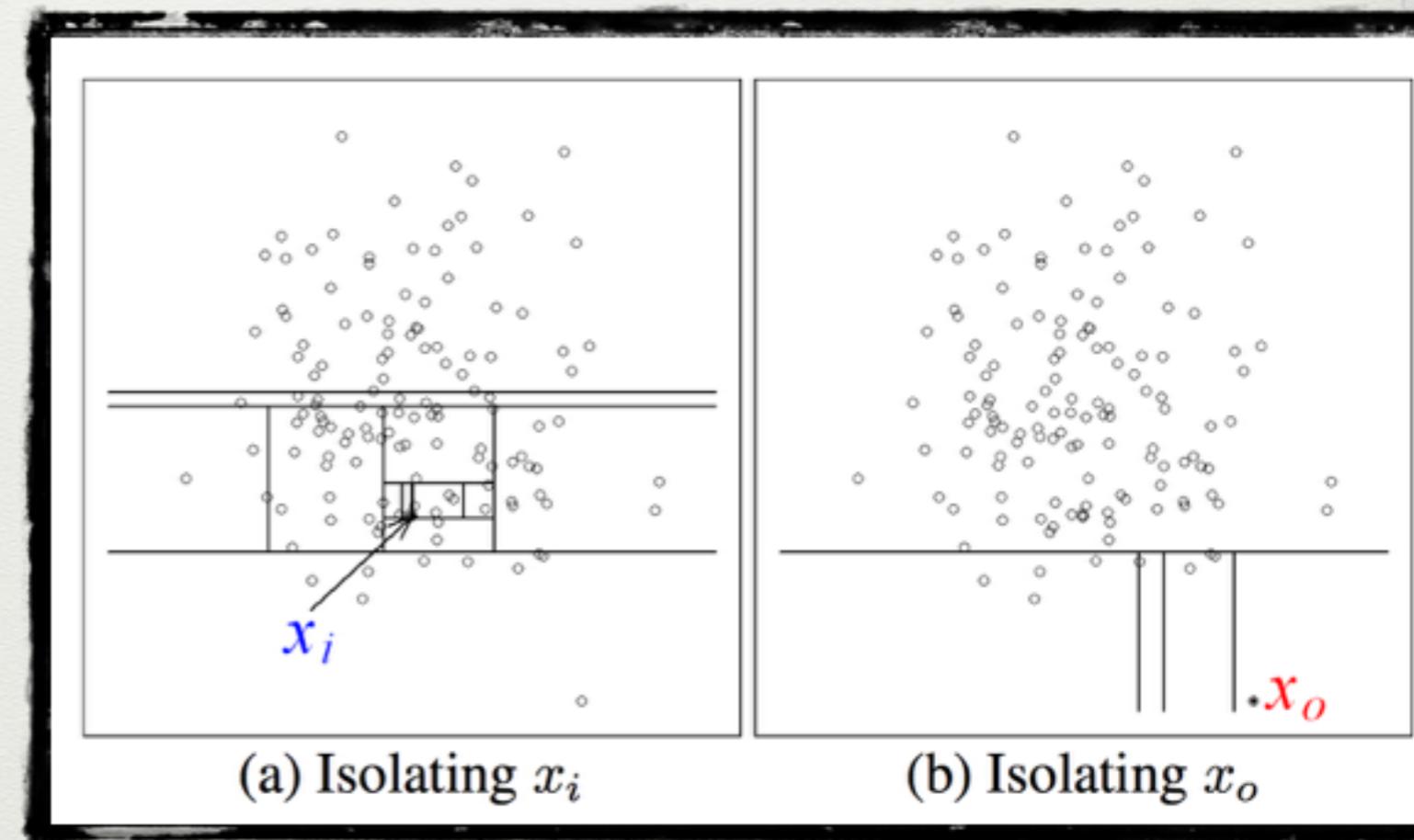
Solution

Isolation Forest¹

Un-supervised method of statistically isolating points that have no close neighbors in feature space

Ensembles binary trees:

- branch on randomly-selected features
- randomly-selected feature values
- track depth at which a point is isolated.



¹Liu, F.T. et al, ICDM '08b, 2008

Solution

Isolation Forest

Anomaly scores derived from average tree length

- values < 0.5 uninteresting
- values > 0.6 usually are

Advantages:

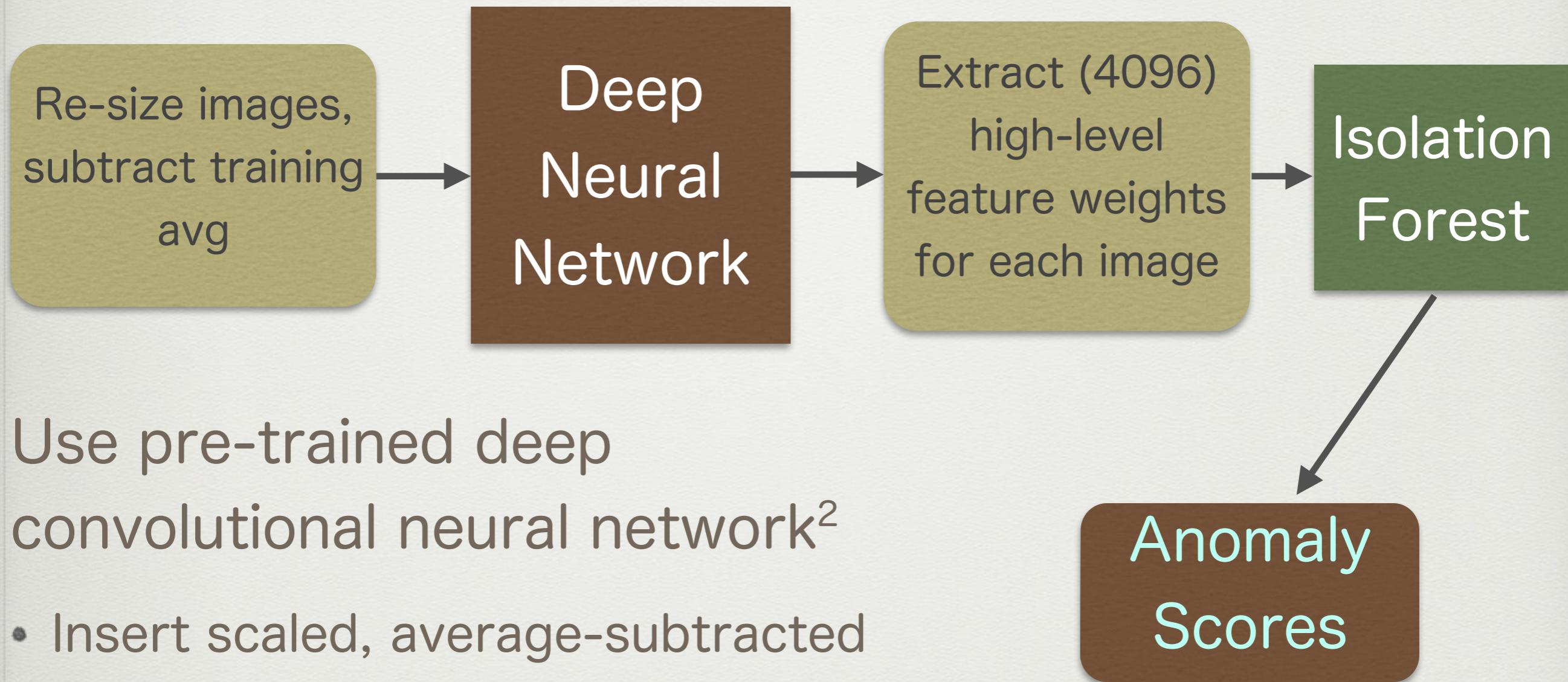
- no need to compute many distances to determine proximity
- don't need prior knowledge of the number of clusters

Image Features

For image datasets, could work with raw pixels, but ...

- Identical images translated by a few pixels can be located far apart in p-dimensional space
- Should take advantage of modern image feature representation — deep convolution neural nets

Featurizing Images



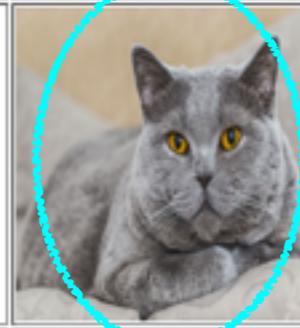
²Krizhevsky, A, et al, Imagenet classification with deep convolutional neural networks, 2012

Results

61 images: tigers + 3 leopards + 3 house cats + 1 house



Results

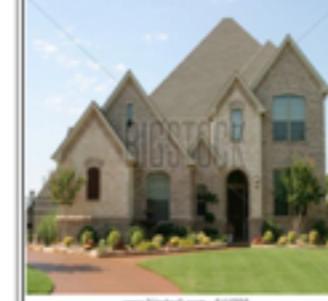
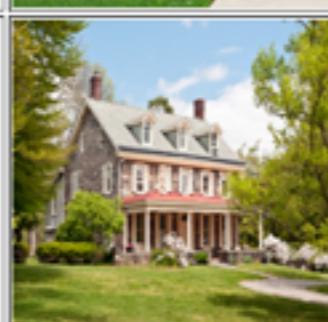
fc7 rank	image	fc6 score	fc7 score	fc8 score	fc7 rank	image	fc6 score	fc7 score	fc8 score
0		0.55	0.62	0.65	1		0.52	0.57	0.64
2		0.52	0.52	0.61	3		0.51	0.52	0.60
4		0.50	0.52	0.59	5		0.50	0.51	0.59
6		0.49	0.51	0.59	7		0.49	0.50	0.59
8		0.48	0.50	0.58	9		0.47	0.50	0.57

Results

61 images: houses + 1 hospital + 1 house boat + 1 “boat”



Results

fc7 rank	image	fc6 score	fc7 score	fc8 score	fc7 rank	image	fc6 score	fc7 score	fc8 score
0		0.55	0.65	0.71	1		0.51	0.64	0.62
2		0.50	0.61	0.61	3		0.49	0.59	0.61
4		0.49	0.55	0.59	5		0.49	0.53	0.56
6		0.48	0.53	0.55	7		0.48	0.52	0.54
8		0.48	0.51	0.54	9		0.48	0.51	0.54

Questions?

