Source Code!

Locally Varying Distance Transform for Unsupervised Visual Anomaly Detection

Wen-Yan Lin*[1], Zhonghang Liu*[1] and Siying Liu[2]

[1] *Singapore Management University* [2] *Institute for Infocomm Research, Singapore* **denotes joint first author*

Instability of Visual Anomaly Detection

Anomaly detectors behave erratically on image data. We attempt to explain and correct this problem. High-level overview:

- **Anomaly detection is often based on low dimensional statistics:** If anomalies are a small fraction of a dataset, anomalous regions of a sample space will be less densely populated than normal regions.
- **Image data is high dimensional:** The size of the sample space increases exponentially with the number of dimensions. This means both normal and anomalous regions of the sample space will be sparsely populated.
- **Visual anomaly detectors resort to implicit or explicit dimensionality reduction:** This may be causing the instabilities.

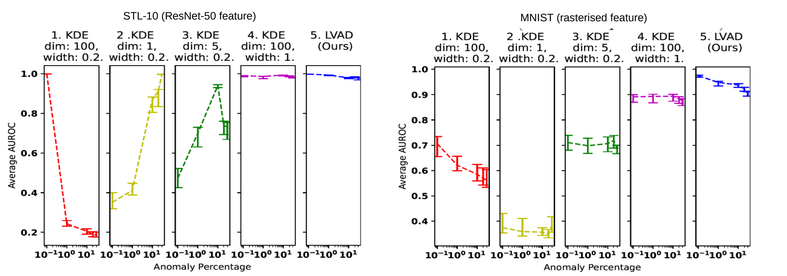
Types of Instability

- **Parameter Sensitivity:** Papers frequently change parameters / neural networks for every dataset.
- **Dataset Sensitivity**: Anomaly detectors are often effective on only a few datasets or at specific anomaly percentages.
- **Performance inversion:** This is an extreme form of dataset sensitivity where the detector is more effective at (difficult) high anomaly percentages than at (easy) low anomaly percentages.

Ilustrating Instability

We vary the parameters of a simple, density based anomaly detector, which employs PCA for dimensionality reduction.

- **Retaining many dimensions makes desnity estimation brittle.** Anomaly detection is very unstable at high anomaly percentages..
- **Removing many dimensions causes performance inversion.** Likely because dimensionality reduction tends to only preserve anomalous variations when anomaly percentages are high.
- **Large (high-dimensional) density estimation kernels provide stability but reduce accuracy.**



Re-thinking Anomaly Detection

How can dimensionality reduction be formalized?

- Image classes have notably large within-class variations.
- Perhaps the common assumption that image classes are derived from a single generative process is incorrect.
- Instances of a class seem to be derived from an ensemble of generative process.
- Formulate anomaly detection as scoring an instance's potential membership to a number of generative processes, which vary in importance?
- **Instead of a single embedding, normality is represented by multiple embeddings, each accommodating a specific type of normality.**
- Amenable to statistical interpretation, reducing the reliance on heuristic dimensionality reduction.

Our Approach

Assume local data clusters are outcomes of individual, high dimensional generative processes.

- Shell Theory [1] suggests instances of each generative process are uniquely close to their mean. i.e. the likelihood an instance belongs to a cluster can be determined solely from its distance to the cluster mean.
- A new embedding scheme based on a set of 1-D distance-from-mean projections for deciding memberships in respective clusters.
- Scores are agglomerated with a weighted average.

$$a(\mathbf{x}) = p\left(Y=1\,|\,\mathbf{x}\right) = \sum_{j=1}^{m} a_j \times p\left(Y_j=1\,|\,\mathbf{x}\right)$$

where:

- a_j is the weight given to cluster-j. It is set to the percentage of instances belonging to cluster-j.
- $p\left(Y_j{=}1|x\right)$ is the proability that instance **x** belongs to cluster-j. This is estimated from data.

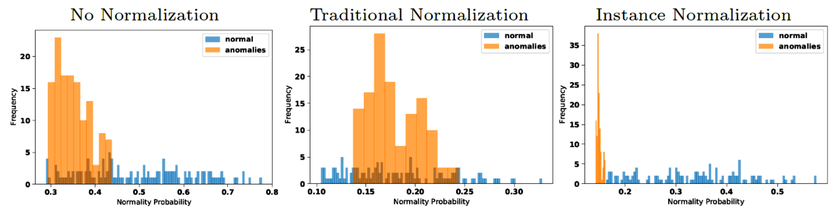
Need to address two challenges:

- **Noise cancellation:** Shell based distance interpretation requires noise cancellation [2]. We use instance normalization as a noise cancellation procedure.
- **Statistical inference:** Traditional probability density based inference is ill-conditioned on shells. We propose a cumulative density alternative.

Noise Cancellation / Instance Normalization

If feature dimensions are ergordic in the mean, instance normalization removes noise's impact on feature distance. This allows the theoretically predicted shells to emerge. The instance normalization equation is:

$$\mathbf{x} = n(\widetilde{\mathbf{x}}, \mathbf{v}) = \frac{\widetilde{\mathbf{x}} - \mathbf{v}}{\|\widetilde{\mathbf{x}} - \mathbf{v}\|} \quad \text{where} \quad \mathbf{v}_i = \left[\, m_i \; m_i \; \ldots \; m_i \,\right]^T, \quad m_i = \frac{1}{d}\sum_{k=1}^{d} \widetilde{\mathbf{x}}_i[k]$$



Cumulative Density Inference

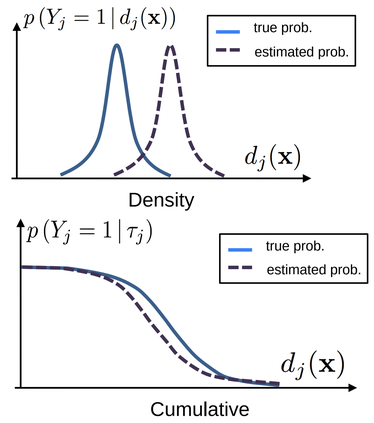
If distance to cluster-mean is a unique identifier of cluster membership,

$$p\left(Y_j=1\,|\,\mathbf{x}\right) = p\left(Y_j=1\,|\,d_j(\mathbf{x})\right)$$

where $d_j(\boldsymbol{x})$ is the distance of instance x from the mean of cluster-j.

In theory, $p\left(Y_j{=}1|d_j(\boldsymbol{x})\right)$ can be estimated from data density. In practice, the hollow nature of shells makes naive density based inference ill-conditioned.

A slight misestimation of probability density, will cause many true cluster instances to be misclassified as not belonging to the cluster.



Rather than desnity, we employ cummilative density, $p\left(Y_j{=}1|\tau_j\right)$, where τ_j is the upper bound of an instance's distance to the mean of cluster-j.

A cummlative density based inference is less senstive to estimation errors. It can be estimated from data using Baye's rule:

$$p\left(Y_j=1\,|\,\tau_j\right) = \frac{p\left(\tau_j\,|\,Y_j=1\right)+\epsilon}{\left(p\left(\tau_j|Y_j=1\right)+\epsilon\right)+p\left(\tau_j|Y_j=0\right)\times\frac{p(Y_j=0)}{p(Y_j=1)}}$$

LVAD is stable across many different datasets and anomaly percentages.

Dataset	Algorithm	Anomaly Percentage (%)					Ave.	Diff.
		0.1	1	10	20	30		
STL-10 (ResNet-50)	LVAD (ours)	0.998	0.993	0.979	0.983	0.977	**0.986**	**0.021**
	Shell-Renorm	0.803	0.829	0.997	0.999	0.999	0.925	0.196
	OC-SVM	0.996	0.995	0.967	0.877	0.777	0.922	0.219
	DAGMM	0.574	0.477	0.826	0.911	0.883	0.734	0.434
	Deep Unsup.	0.384	0.956	0.906	0.869	0.886	0.796	0.572
	RSRAE	0.995	0.992	0.972	0.971	0.944	0.975	0.051
MNIST (Rasterized Pixels)	LVAD (ours)	0.974	0.948	0.938	0.923	0.904	**0.937**	**0.070**
	OC-SVM	0.937	0.901	0.885	0.856	0.824	0.881	0.113
	DAGMM	0.624	0.708	0.629	0.616	0.613	0.638	0.095
	Deep Unsup.	0.525	0.891	0.847	0.779	0.835	0.775	0.366
	RSRAE	0.966	0.948	0.851	0.794	0.763	0.864	0.203

* A small Diff. indicates stability

ACKNOWLEDGMENT

We gratefully acknowledge the funding support by Lee Kong Chian Fellowship.

REFERENCE

[1] Lin, W. Y., Liu, S., Ren, C., Cheung, N. M., Li, H., & Matsushita, Y. Shell Theory: A Statistical Model of Reality. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2021.
[2] Lin, W. Y., Liu, S., Dai, B., & Li, H. Distance Based Image Classification: A solution to generative classification's conundrum? Int. Journal on Computer Vision (to appear)