# 9/11/2024 Paper Review as Meeting Prep

# Anomaly Detector through Data Visualization

**Anomaly identification through data visualization: regression analysis revisited**


**Assion Lawson-Body,** *University of North Dakota, assion.lawsonbody@und.edu*
**Patrick Glenn,** *University of North Dakota, patrick.glenn@und.edu*
**Laurence Lawson-Body,** *University of North Dakota, laurence.lawsonbody@und.edu*
**Abdou Illia,** *Eastern Illinois University, aillia@eiu.edu*
**Emmanuel Ufiteyezu,** *University of Rwanda, ufitemmy54@gmail.com*

# Why are data visualizations important?

- Often only way to make sense of work
- Brings new insights into descriptive statistics domain
- Makes measurement errors of regression analysis insignificant and irrelevant
- Aids in proving a point

# Data Visualization Techniques

- Need clear purpose and convey essential message
- Visibility
  - "Picture is worth a thousand word"
  - Fixes information load issue
- Intuitive & Informative
- Mass notification
  - Focus on exceptions, outliers, and data highlights
  - Alert system
  - Advise possible actions
- Emergency management
  - Geographic distribution
  - Detection of temporal patterns in data
- Accessibility
  - Effort required to access visualization

# Anomalies

- **Point anomaly** – single anomalous instance in a larger dataset
- **Collective anomaly** – multiple occurrences in line with one another but offset from the dataset
- **Contextual anomaly** – falls within normal/typical distribution but proves to be anomaly based on when it occurs
- Could be data collection errors
- May indicate an organizational change is needed
- Could simply be seasonality – important to acknowledge when forecasting/predicting

# How this could relate to our work...

- Final deliverable to convey information about data
- Throughout process working with NRAO to decide between approaches, get more background information, etc.

# Visual Anomaly Detection for Images: A Survey

Visual Anomaly Detection for Images: A Survey

Jie Yang, Ruijie Xu, Zhiquan Qi, Yong Shi
University of Chinese Academy of Sciences, Beijing 101408, China,
Email: yangjie181@mails.ucas.ac.cn

# Anomaly Detection Logistics

- Typically unsupervised
  - Abnormal image samples are rare/difficult to collect
  - Abnormal samples do not have stable statistical laws
- **Image level detection:** whole image normal or abnormal
- **Pixel level detection:** locate abnormal regions of image

# Image Level Anomaly Detection

- **Density Estimation**
  - Classify as anomaly if: testing image does not meet probability distribution model that is estimated with normal image samples
  - Requires large number of training samples
  - Poor scalability
- **One-class Classification**
  - Attempts to reconstruct decision boundary of target class (for "normal" images)
  - Do not require large number of training samples
  - Poor scalability
  - Deep CNN – transfer learning; fine-tune the pre-trained convolution network to extract discriminative image features and take nearest neighbor classification method to construct the one-class classifier

# Image Level Anomaly Detection

- **Image Reconstruction**
  - Map image to low-dimensional vector representation (latent space) + try to find inverse mapping or reconstruction for original image
  - Assumes reconstruction errors of anomalies are large while those of normal image are small
  - Autoencoder - may compress redundant information in input data while retaining and distinguishing non-redundant information
  - General Adversarial Network (GAN) - trained with normal image beforehand and then anomaly can be detected by calculating difference between test image and normal image closest to test image
  - To improve detection, constrain/regularize feature distribution of latent code in joint model of autoencoder and adversarial network

# Image Level Anomaly Detection

- **Self-Supervised Classification**
  - Attempt to mine available supervision information from large-scale unsupervised data
  - Learn visual representation (typically via CNN) and representations can be transferred to anomaly detection
  - Learns features from normal samples – those that are abnormal don't have these characteristics
  - Rotation and translation are difficult for anomaly detection
  - *Contrastive learning* is a special case of this classification that works well (source 44 in this paper)

# Pixel Level Anomaly Detection

- **Image reconstruction**
  - Compress / reconstruct input image with deep convolution autoencoder
  - Learn to reconstruct normal images
  - Identify anomalies by evaluating pixel differences between input image and reconstructed images
  - Detecting anomalies by comparing differences between test image and nearest neighbor is Inefficient
  - **Leads to large number of false alarms**

# Pixel Level Anomaly Detection

- **Feature Modeling**
  - Detect anomaly in feature space
  - Label as abnormal if regional feature corresponding to the local region of the test image deviates from the modeled feature distribution
  - Assumes trained on only normal images
  - May not predict abnormal images
  - "Anomaly score"
  - Successful implementation – hierarchical convolution encoder

# How this could relate to our work

- Are there labeled data with typical anomalies? If so, how many?
- How unique, varied are expected anomalies?
- Is the intent for the anomaly detection to be image or pixel based?
- How much data do we have to train with to test for anomalies? How much of this data is labeled normal and/or abnormal?
- Is it worse to have a "false alarm" anomaly or to miss detection of an anomaly?

# Transparency and Explainability of AI Systems: From Ethical Guidelines to Requirements

## Transparency and explainability of AI systems: From ethical guidelines to requirements

Nagadivya Balasubramaniam [a], Marjo Kauppinen [a], Antti Rannisto [a,b], Kari Hiekkanen [a,c], Sari Kujala [a]

# Why are there ethical considerations?

- Black-box nature of AI systems
- Transparency + trust depend on each other
- Understandability contributes to transparency
- Traceability accentuates importance of tracing decisions in AI systems

# Things to keep in mind...

- **Communicate models' input, capabilities, intended purpose, and limitations**
- Develop AI that is transparent across processes and functions
- Clarity of purpose of AI system affects definition of explainability requirements
- With multidisciplinary teams, explainability requirements are steeper

# Results Ranked AI Requirements of Study with Software Development and Recuitment

1. Transparency
2. Explainability + Privacy
3. Fairness
4. Security
5. Safety
6. Accountability
7. Reliability

# Steps

1. Validate work through simulation, etc.
2. Discuss with key stakeholders
3. Pilot evaluation (ensure solution fits well with current practice)

# How could relate to our work?

- Obtain model specifications
- Clearly describe what data our models' use to make predictions and classifications
- Discuss limitations of various model techniques (certainty vs accuracy, speed vs accuracy, precision vs. accuracy, etc.) and determine what is most important in their context
- Make sure tools we deliver are user-friendly