Exploration of Anomaly detection in CCTV footage: Computer Vision

Wilson Ler (lws803) , Sean Decker (skdecker)
CS229, Stanford University

introduction

Although closed-circuit television (CCTV) cameras are ubiquitous in the modern world, the footage from these cameras often goes unexamined. Machine learning can help to overcome this waste of possibly valuable footage. Specifically, we propose that the use of anomaly detection to filter CCTV footage in real time can be used to detect possibly important events. Along with classification, we propose that the model also use a saliency window test in order to explain its classifications. This would allow human moderators to not need to treat the model as a black box.

data & features

UC San Diego's Statistical Visual Computing Lab's UCSD Anomaly Detection Dataset

- Grey-scale CCTV frames captured from above a pedestrian walkway
- Normal: groups of pedestrians walking in various crowd densities
- Anomalies: corresponds to either the appearance of non pedestrians on the walkways or anomalous pedestrian motion patterns

Fig 1) Example training set class example: a normal crowd with no anomalies.

- Training set comes with only class-examples, we use randomly added Gaussian blur to therefore simulate the presence of anomalies in the training dataset

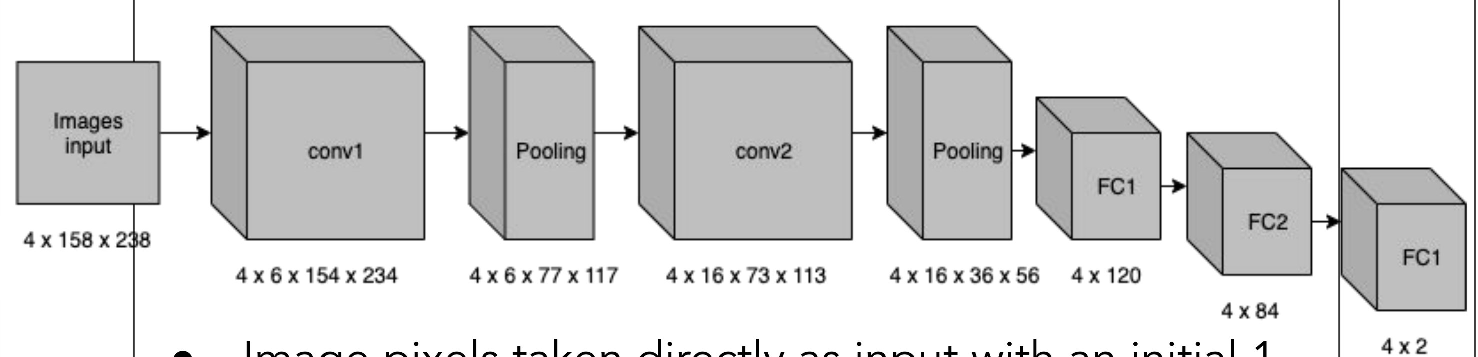
Fig 2) Example of training class with Gaussian blur and Gaussian noise added in order to simulate an anomaly. Noise is added with specific size window

methods - model

naive solution: logistic regression

- This is a classification problem, logistic regression is the simplest binary classifier
- Problem: training data to learn only 1 class, not 2
- We want a singular classifier, not a binary classifier

our solution: neural network



- Image pixels taken directly as input with an initial 1 to 1 convolution layer to preserve as many features as possible
- Relu is used for activation before being passed into a 2 x 2 window max pool
- Next convolve with 3 x 3 polynomials. Relu is used for activation and then again passed into maxpool.
- 3 hidden layers before output

methods - explanation tools

saliency window test

- In order to have our model also be able to explain why it makes the predictions that it does, we built a tool that allowed the model to test what parts of a given image are most important to the predicted score
- The saliency window test works as follows:
  - Store output from target anomalous image as Y
  - Within a sliding window loop, take a non-anomalous image and crop the exact window dimensions + location to replace in the target image
  - Feed the resultant image into the model and obtain the score Y*
  - Calculate the difference between Y* - Y
  - If Y* - Y is negative, multiply the value by a red mask on output image to generate a heat patch. Larger Y* - Y difference implies a more distinct heat patch

preliminary results

naive solution: logistic regression

- Trained with raw training data, gradient descent until convergence
- Because the training data is entirely class examples, naive logistic regression only learns to label all inputs as non-anomalies

our solution: neural network

- Training our model:
  - Stochastic Gradient Descent, adding Gaussian anomalies to ½ of training images
  - Train for 20 epochs, learning rate of 0.1
- Model does well in identifying images that have golf carts in them

Fig 3) Example Testing frame with anomaly present, i.e. the golf cart.

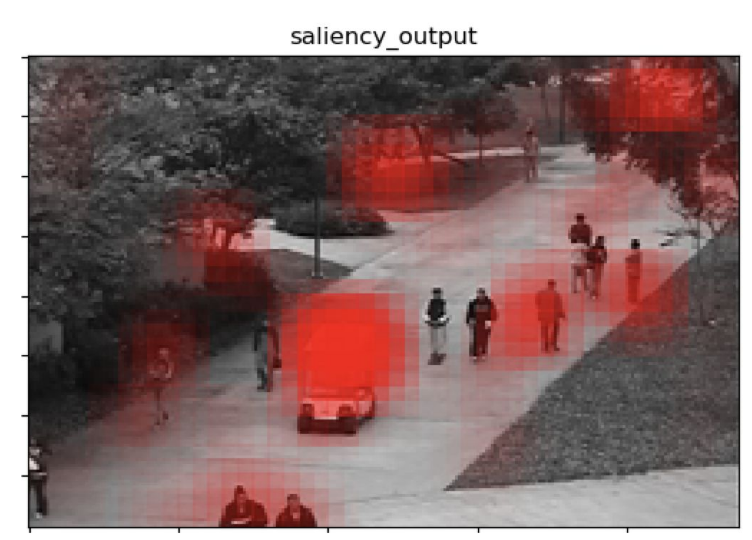


Fig 4) Saliency Test output for the figure above. More red in an area means that the pixels in that area are more important to the output.

- Our network however does very poorly identifying anomalies that deal with bikes or scooters

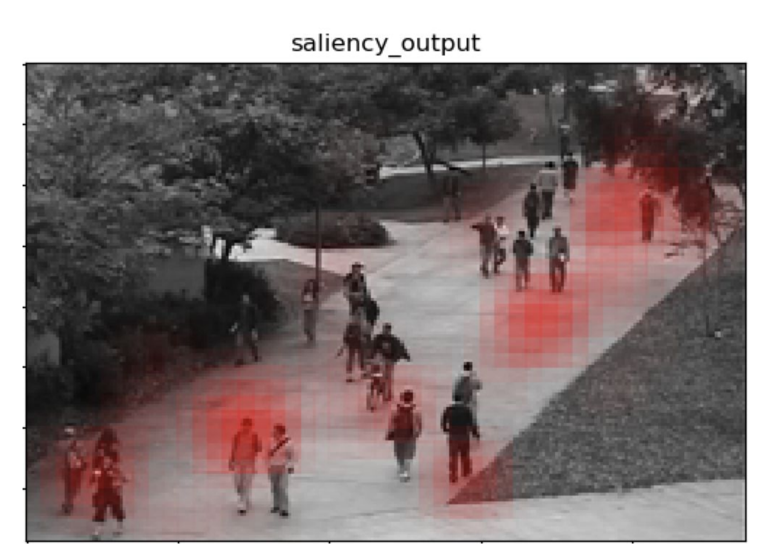


Fig 5) Saliency Test output for the figure with a biker. Note that the biker is not identified as anomalous.

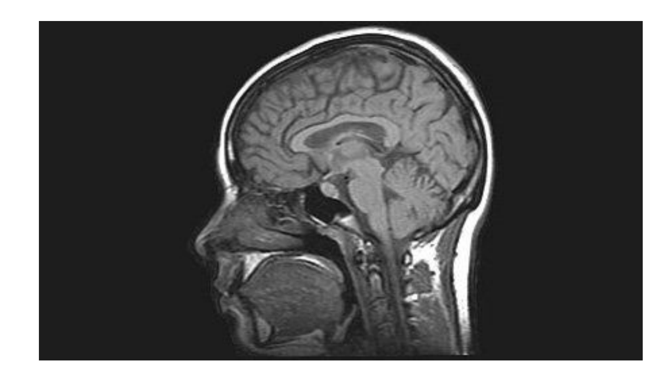
- Possibly due to fixed size of noise added to images

discussion

- Our model works well identifying larger anomalies like golf carts and wheelchairs, but does poorly on smaller anomalies
- We believe this is due to how our current model trains with very few features as input. Training images are low resolution so few pixels differentiate a bike from a pedestrian
- Fixes:
  - Adding a pretrained network to identify features
  - Having rgba values for images
  - Training on sequences of frames instead of single frames

future

- Our model would expand to detecting anomalies in medical images very easily we predict
- Sonogram and CT scan images are also gray-scale, low resolution images
- our model would only need inputs or healthy examples of these scans
- the fact that our model is built to explain the anomalies, making for a useful assistant.



references

- Perera, Pramuditha, and Vishal M. Patel. "Learning Deep Features for One-Class Classification." IEEE Transactions on Image Processing 28.11 (2019): 5450–5463. Crossref. Web.
- "Anomaly Detection and Localization in Crowded Scenes." Anomaly Detection and Localization in Crowded Scenes, \textit{Statistical Visual Computing Lab: UC San Diego},
- Lundberg, Scott M and Lee, Su-In. "A Unified Approach to Interpreting Model Predictions." Advances in Neural Information Processing Systems 30 (2017): 4765-4774. Curran Associates, Inc. Web.
- http://papers.nips.cc/paper/7062-a-unified-approach-to-interpreting-model-predictions.pdf.