# Automated Anomaly Detection Using Auto Encoders Applied ML 2025

Carl-Johannes Johnsen

Duplicati - carl@duplicati.com

Department of Computer Science - carl-johannes@di.ku.dk University of Copenhagen



May 26, 2025 Slide 1/18 What are we trying to solve?



# What are we trying to solve?





### What are we trying to solve?



Can we build a program that can do *automatic anomaly detection* of *X-Ray images* of food while keeping up with the *high-throughput demands* of food production?



#### Automated feature extraction







Good results (>90 %)



Good results (>90 %) ... but could be better.



Good results (>90 %) ... but could be better.

The approach is supervised.



Good results (>90 %) ... but could be better.

The approach is supervised.

Anomalies are far in between.



Good results (>90 %) ... but could be better.

The approach is supervised.

Anomalies are far in between.

There's a high probability that we cannot cover the entire problem domain.



Good results (>90 %) ... but could be better.

The approach is supervised.

Anomalies are far in between.

There's a high probability that we cannot cover the entire problem domain.

Which gave me an idea ...



#### Can we construct a model purely from "good" data samples?



Can we construct a model purely from "good" data samples?

Alina Sodes Master's thesis investigated whether this is possible.



Can we construct a model purely from "good" data samples?

Alina Sodes Master's thesis investigated whether this is possible.

She found that Auto Encoders could be a viable approach.



# Auto Encoder (AE)







Preprocess image to counter X-ray effects.



- Preprocess image to counter X-ray effects.
- Apply the AE to the image.



- Preprocess image to counter X-ray effects.
- Apply the AE to the image.
- 3 Compare the input to the output.



- Preprocess image to counter X-ray effects.
- Apply the AE to the image.
- 3 Compare the input to the output.
- If they match, the image was a "good" sample.

















|--|--|--|













Accuracy of 86 %



# How do we improve the performance?



#### How do we improve the performance?

After Alina Sode's master thesis, I further investigated using AEs.



### How do we improve the performance?

After Alina Sode's master thesis, I further investigated using AEs.

Alina Sode trained on the entire image, which means that the model has to learn the shape.



# The data





# The data





Virtually peel the potatoes



# Virtually peel the potatoes











# AE - bad sample







We can clearly see the difference, so it should be possible for the machine as well.



We can clearly see the difference, so it should be possible for the machine as well.

Alina Sode looked at SE, IMED, SSIM, ZNCC, GMSD, with SSIM working best.



We can clearly see the difference, so it should be possible for the machine as well.

Alina Sode looked at SE, IMED, SSIM, ZNCC, GMSD, with SSIM working best... in some cases.



We can clearly see the difference, so it should be possible for the machine as well.

Alina Sode looked at SE, IMED, SSIM, ZNCC, GMSD, with SSIM working best... in some cases.

I further looked at RMSE, PSNR, FSIM, ISSM, SRE, SAM, UIQ and SH.



We can clearly see the difference, so it should be possible for the machine as well.

Alina Sode looked at SE, IMED, SSIM, ZNCC, GMSD, with SSIM working best... in some cases.

I further looked at RMSE, PSNR, FSIM, ISSM, SRE, SAM, UIQ and SH.

From these, FSIM worked the best on these images.



We can clearly see the difference, so it should be possible for the machine as well.

Alina Sode looked at SE, IMED, SSIM, ZNCC, GMSD, with SSIM working best... in some cases.

I further looked at RMSE, PSNR, FSIM, ISSM, SRE, SAM, UIQ and SH.

From these, FSIM worked the best on these images.

FSIM works better than SSIM, due to the errors we see.



# AE + FSIM - distribution





# AE + FSIM - ROC





# Conclusion

- CNNs work great but are supervised.
- AEs can work great and are semi-supervised.
- Preprocessing is important.
- Choosing the right similarity measure is important.

