Prof. Dr.-Ing. Jörg Franke

Institute for Factory Automation and Production Systems

Friedrich-Alexander University Erlangen-Nuremberg

Friedrich-Alexander-Universität
Technische Fakultät

Comparative Analysis of Deep Learning Methods for Automatic Visual Inspection of Electric Motor Components

Final Presentation on the Applied AI Project
Muhammad Abdullah, Microsoft Teams, 24.01.2024

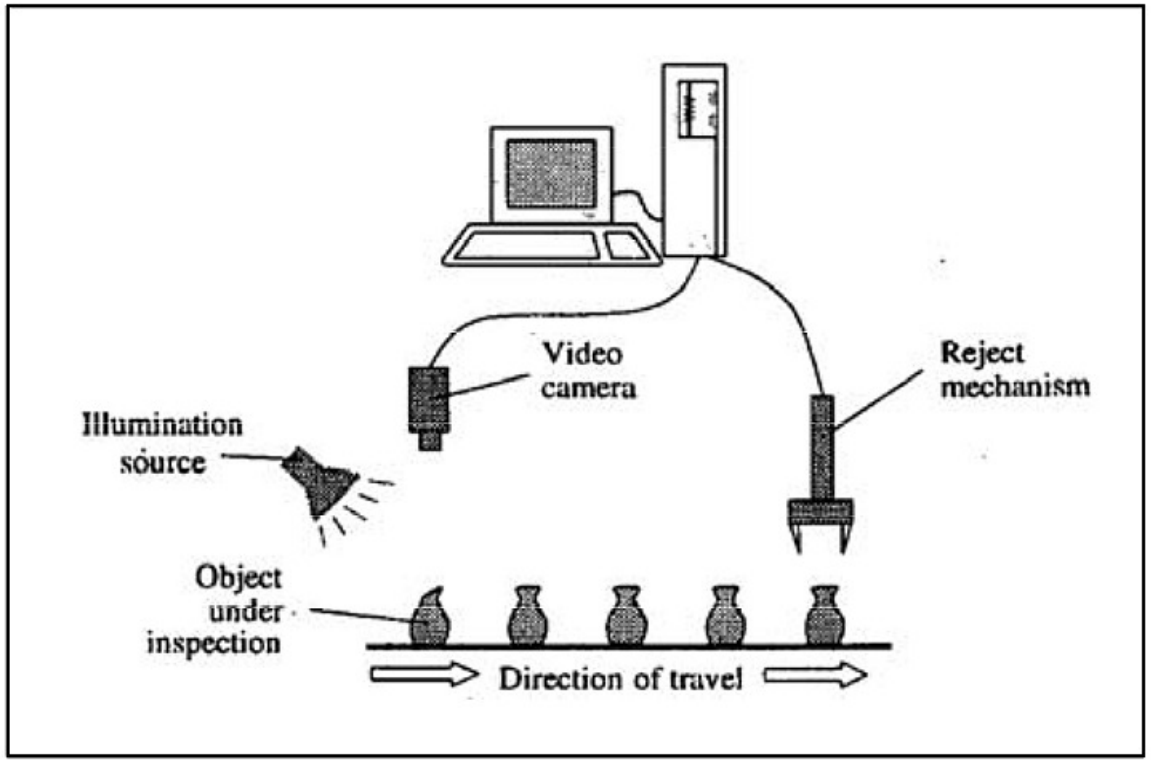
Visual inspection and automated visual inspection are introduced while explaining why automated visual inspection is better.

Visual Inspection

- Determine if a product deviates from given set of specification [2]
  - Geometric dimensions
  - Assembly integrity
  - Surface finish
- A quality control task
- Problems with human inspectors
  - Lower consistency
  - High labor costs

Automated Visual Inspection (AVI)

- Inspection using machine vision [2]
  - Better consistency
  - Suitable for environments unsafe for people
  - Lower labor costs



An illustration depicting the process of automated visual inspection [1]

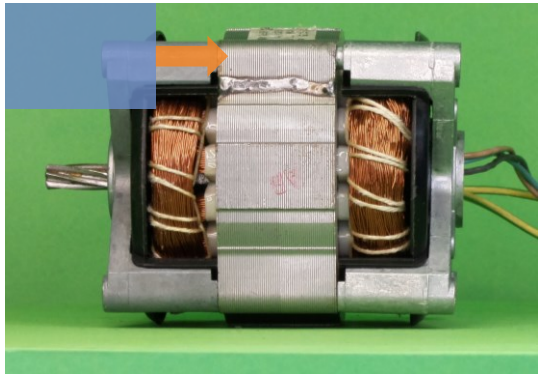
Deep learning approaches have significantly improved the performance of machine vision tasks used for automated visual inspection.

Deep Learning Approaches

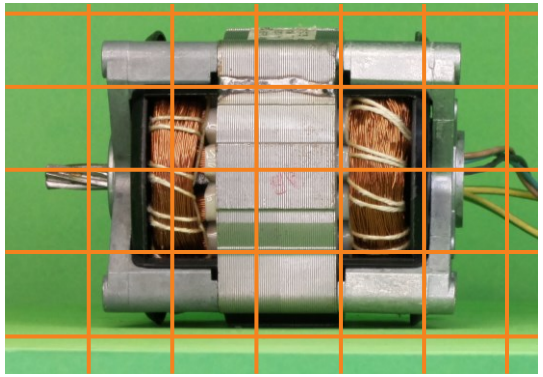
- Convolutional Neural Networks (CNN)
  - Learn inductive biases such as locality and translation equivariance
  - Good generalization
- Vision Transformer (ViT)
  - Image is worth 16x16 words [3]
  - No inductive bias
  - Large scale training trumps inductive bias
- Representation Learning
  - Probabilistic modeling
  - Out-of-distribution detection

Problem Statement

- Compare these approaches for inspection of electric motor components
- First of its kind study with a limited amount of data



CNN performs convolution along the sliding window

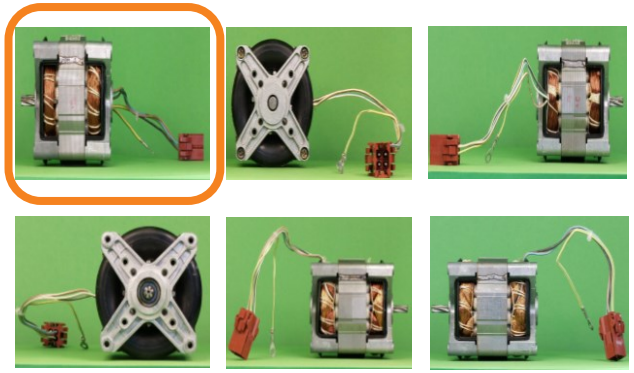


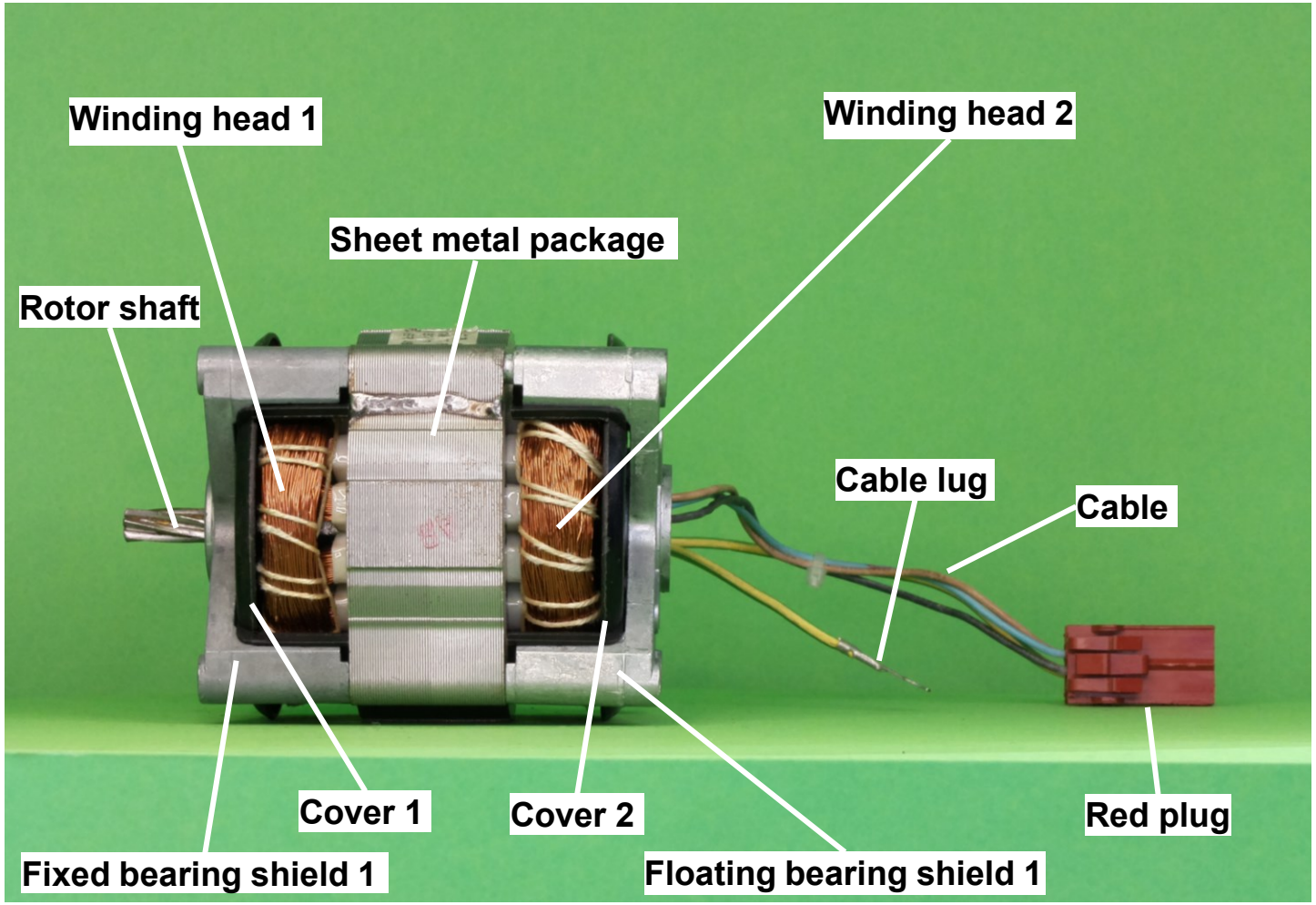
ViT divides image into patches and pass them through transformer encoder

The FAPS proprietary multi-view electric motor dataset has been used for comparing different deep learning approaches.

Multi-view Electric Motor Dataset [4]

- 481 motor combinations
- Six different views
- 11 different object classes
  - View-dependent
- Labels
  - Object detection
  - Object classification
- 5 surface defects



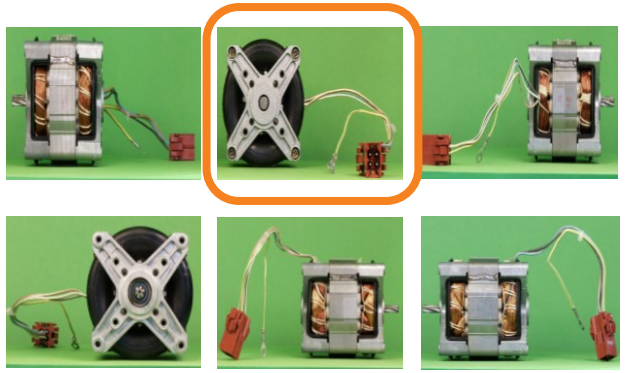


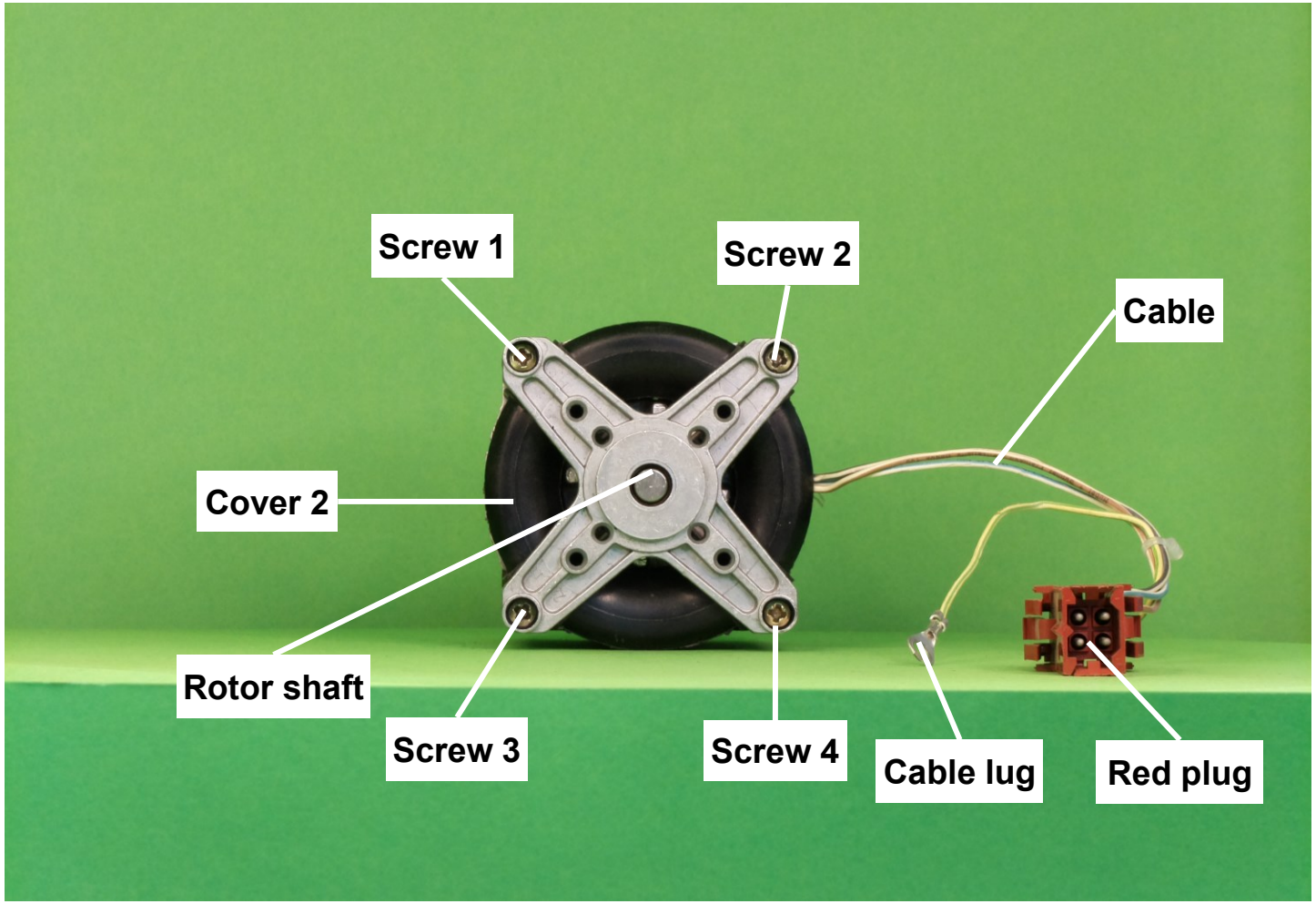
An exemplary 0° view image

The FAPS proprietary multi-view electric motor dataset has been used for comparing different deep learning approaches.

Multi-view Electric Motor Dataset [4]

- 481 motor combinations
- Six different views
- 11 different object classes
  - View-dependent
- Labels
  - Object detection
  - Object classification
- 5 surface defects





An exemplary 90° view image

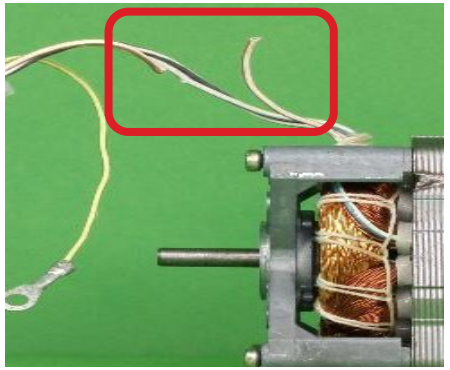
Binary classification has been performed on five components classifying them as defective or non-defective.

Binary Classification of five Surface Defects

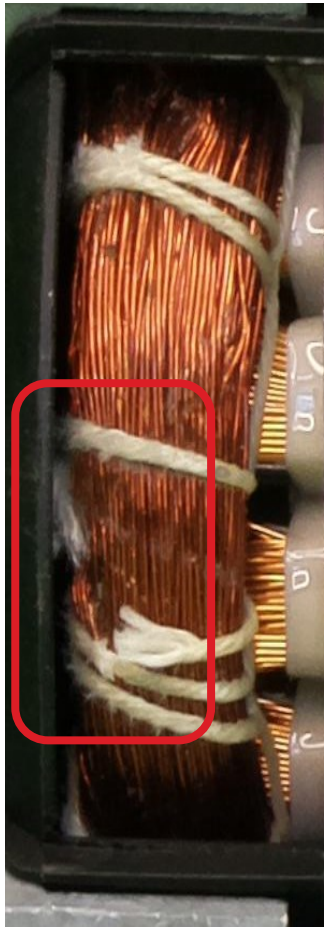
- Broken cable
- Broken insulation cover
- Broken bandaging threads
- Unusual gaps among metal sheets
- Screws not properly screwed

Broken insulation cover

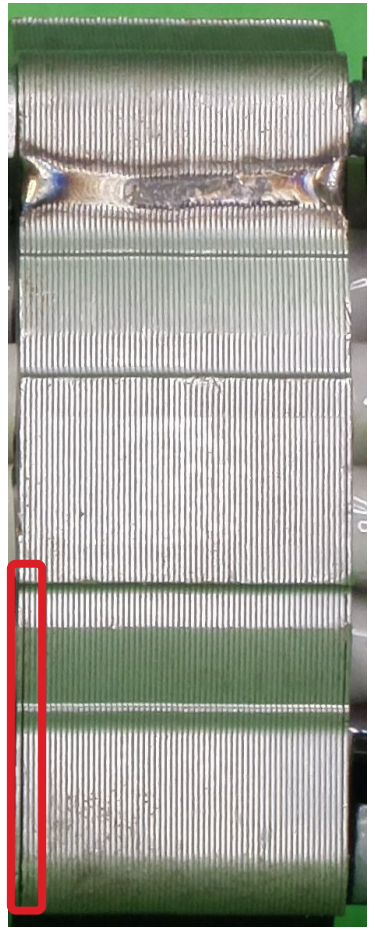
Screw not properly screwed



Broken cable



Broken bandaging threads



Gaps among sheets

The dataset is highly imbalanced in train, validation, and test split and have a few unique defective samples as well.

Distribution of samples in train, validation, and test splits

Type	Train	Validation	Test
Cable	1999	441	444
Cover	2353	575	579
Screw	1132	250	269
Winding Head	2672	592	592
Sheet Metal Package	1336	296	296

Distribution of defective samples in train, validation, and test splits

Type	Train	Validation	Test
Cable	251 (12.5%)	53 (12%)	27 (6%)
Cover	174 (7.4%)	43 (7.5%)	42 (7.3%)
Screw	384 (34%)	83 (33.2%)	84 (31.2%)
Winding Head	66 (2.5%)	12 (2.0 %)	25 (4.2%)
Sheet Metal Package	181 (13.5%)	23 (7.8%)	36 (12.2%)

The dataset is augmented, and defective sample are oversampled during training to counter imbalance.

Strategies to Counter Imbalance Data

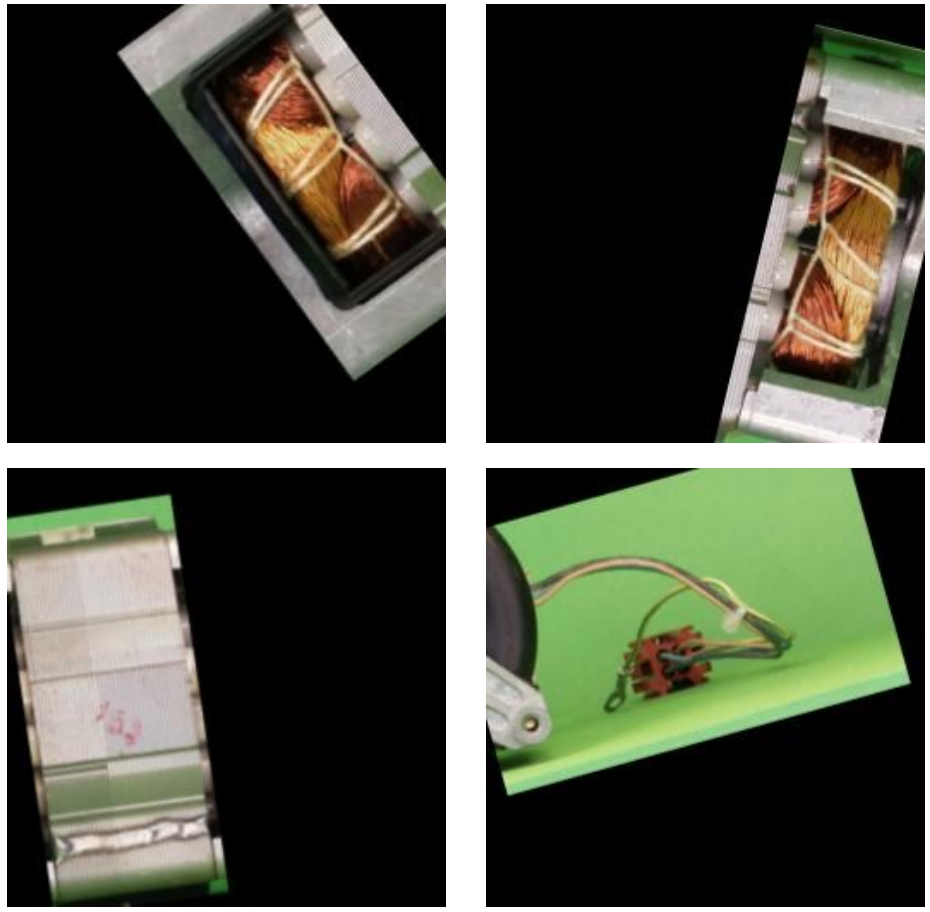
- Weighted cross entropy
- Focal loss
- Oversampling of defective samples

Image Augmentations

- Image mirroring (x-axis and y-axis)
- Affine transformation
  - Translate, scale, rotate

Image Input

- Experiments with both grayscale and RGB

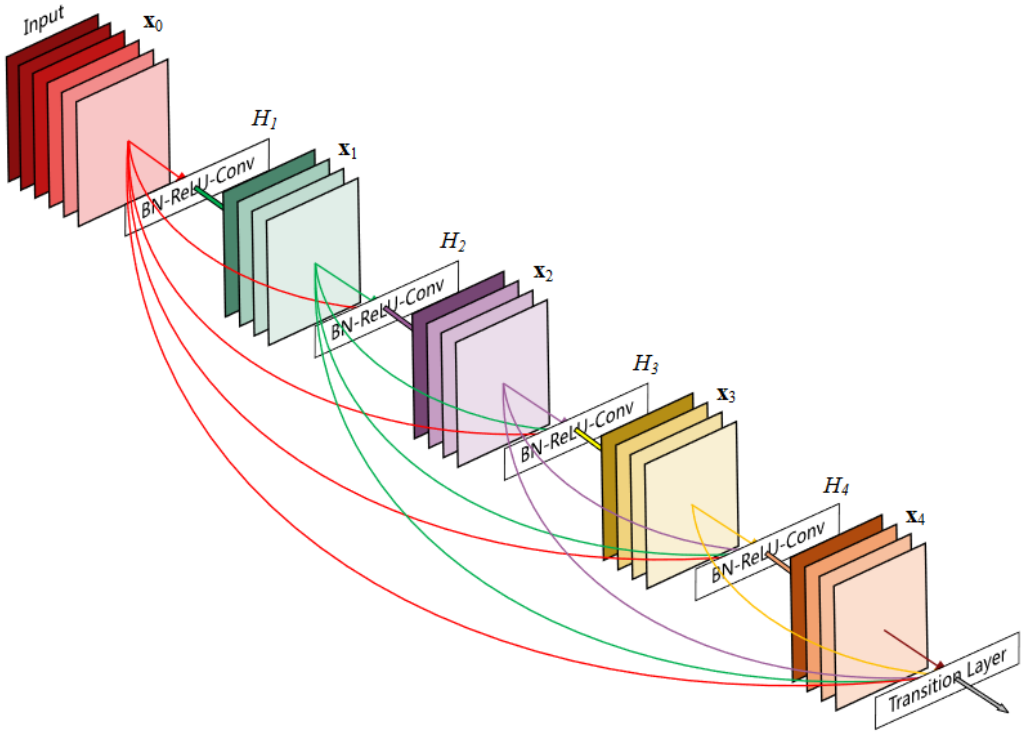


A few examples of augmented images

DenseNets are dense convolutional networks that enable maximum information flow by taking input from all preceding layers in a dense block.

Densely Connected Convolutional Networks (DenseNet) [5]

- Obtains additional inputs from all preceding layers
- Maximum information flow
- Concatenation instead of summation
- Easy to train
- Dense connections have a regularizing effect
- Reduces overfitting on tasks with smaller training set

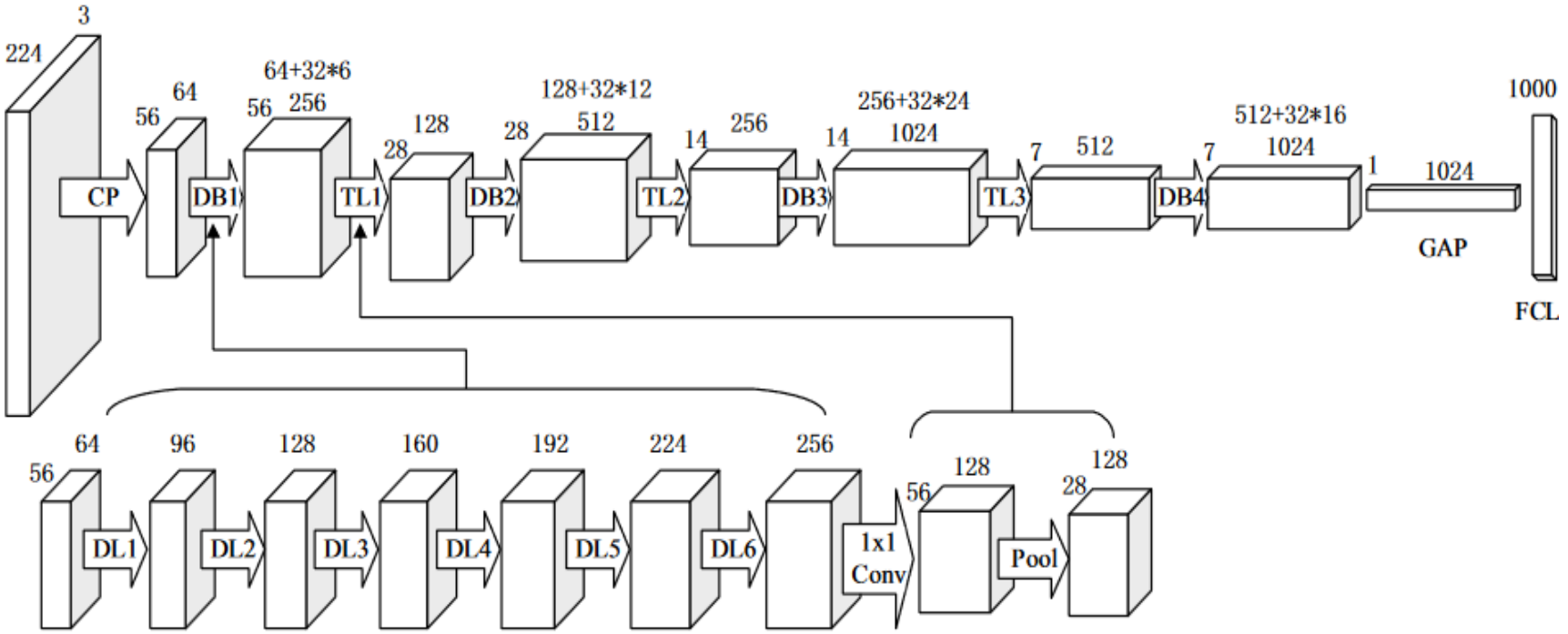


A dense block containing 5 layers. Spatial dimensions are not changed inside dense block (No pooling) [5]

DenseNet-121 is basic version of DenseNet that has 121 dense layers with number of parameters less than ResNet-18.

DenseNet-121

- 121 dense layers
- 4 dense blocks
- 7.2 M param
- <ResNet-18 (11M)



Architecture of DenseNet-121 [6]

DenseNet-121 binary classification results indicate strong performance for screw and cover but sheet metal package, cable, and winding head struggle.

Binary classification results of defective samples on test set

Type	Precision	Recall	F1
Screw	1.0	**1.0**	1.0
Cover	0.86	**1.0**	0.92
Sheet Metal	0.61	**0.66**	0.63
Cable	0.80	**0.60**	0.69
Winding Head	**0.19**	**0.51**	0.28

Screw confusion matrix

	Predicted defective	Predicted non-defective
Actual defective	252	0
Actual non-defective	0	498

Cover confusion matrix

126	0
21	1563

Sheet metal confusion matrix

71	37
45	735

Cable confusion matrix

49	32
12	1239

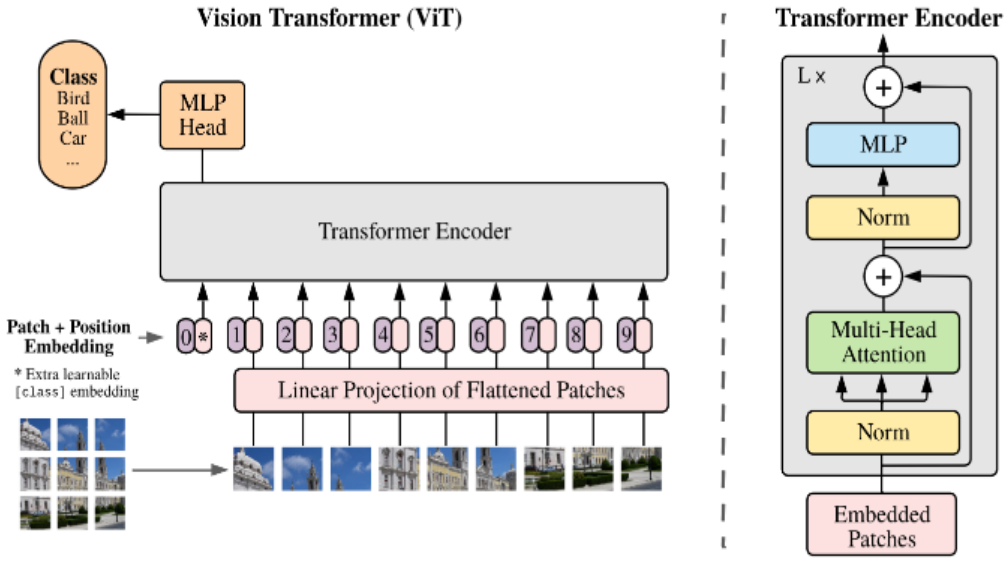
Winding head confusion matrix

38	37
163	1538

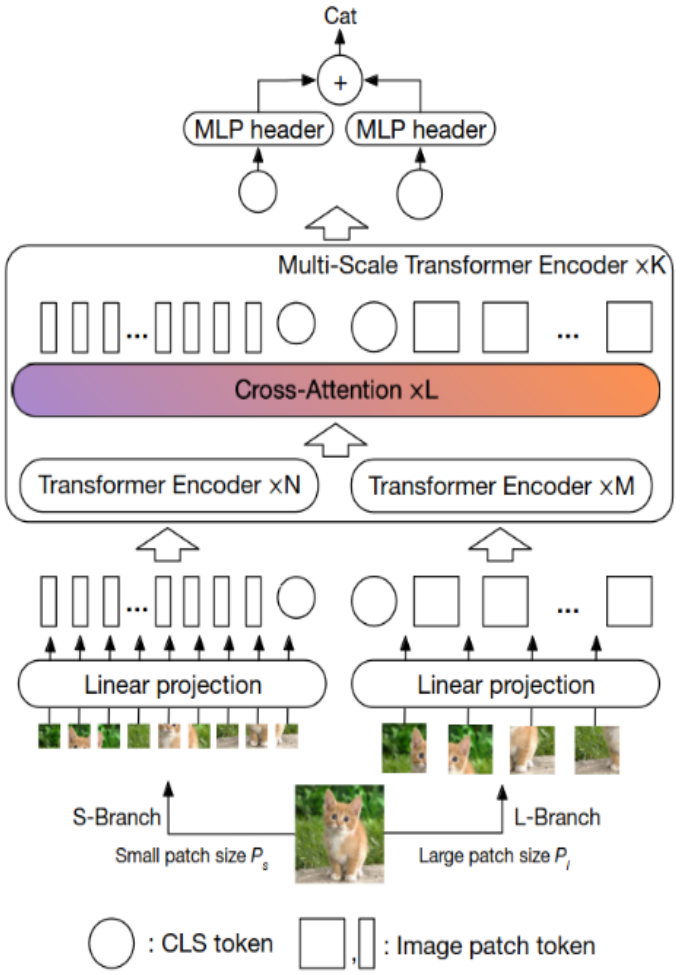
CrossViT is a dual branch vision transformer that take multi-scale features into account while performing classification.

Cross-Attention Multi-Scale Vision Transformer (CrossViT)

- Multi-scale feature representations
- CrossViT-S (~26 M param) – fewer params and flops than ViT-B (86M)



Vision Transformer [3]



Cross-Attention Multi-Scale Vision Transformer [7]

CrossViT-S binary classification results for five surface defects indicate poor performance than DenseNet-121.

Binary classification results of defective samples on test set

Type	Precision	Recall	F1
Screw	1.0	**1.0**	1.0
Cover	0.38	**0.58**	0.46
Sheet Metal	0.32	**0.64**	0.43
Cable	**0.07**	**0.46**	0.12
Winding Head	**0.17**	**0.53**	0.26

Screw confusion matrix

	Predicted defective	Predicted non-defective
Actual defective	252	0
Actual non-defective	0	498

Cover confusion matrix

72	52
116	1456

Sheet metal confusion matrix

68	39
142	631

Cable confusion matrix

37	44
509	738

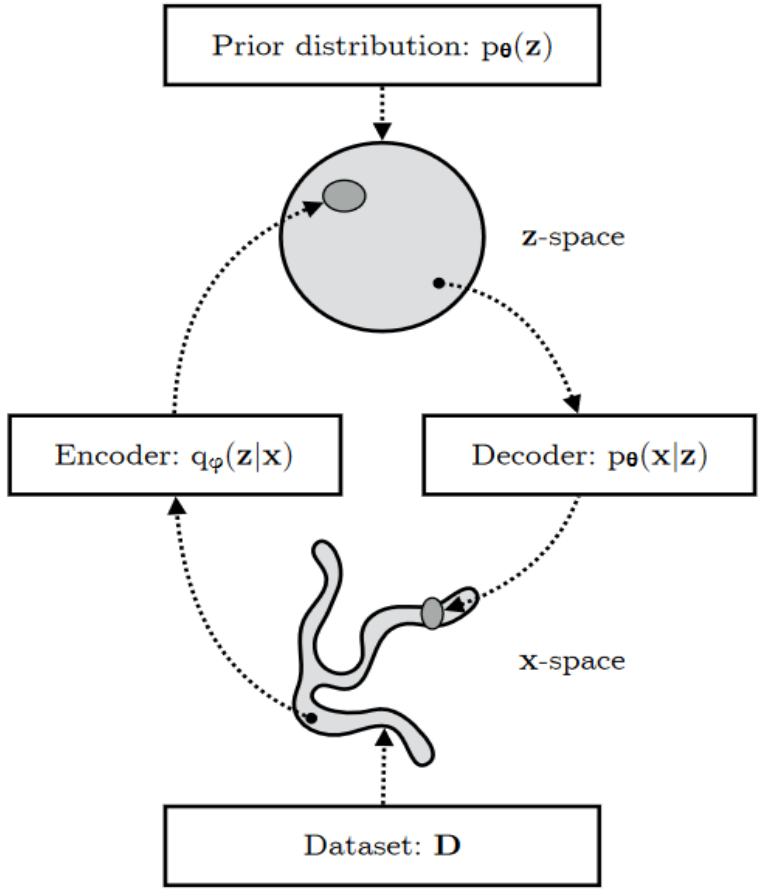
Winding head confusion matrix

40	35
197	1504

Variational Autoencoder learns the representation of non-defective samples during training. During inference, defective samples are detected using reconstruction error.

Variational Autoencoder

- Generative model
- Only non-defective samples are required
- During training time
  - Model learns the representation of non-defective
- During inference time
  - Model try to reconstruct samples
  - Defective samples tend to have higher reconstruction error



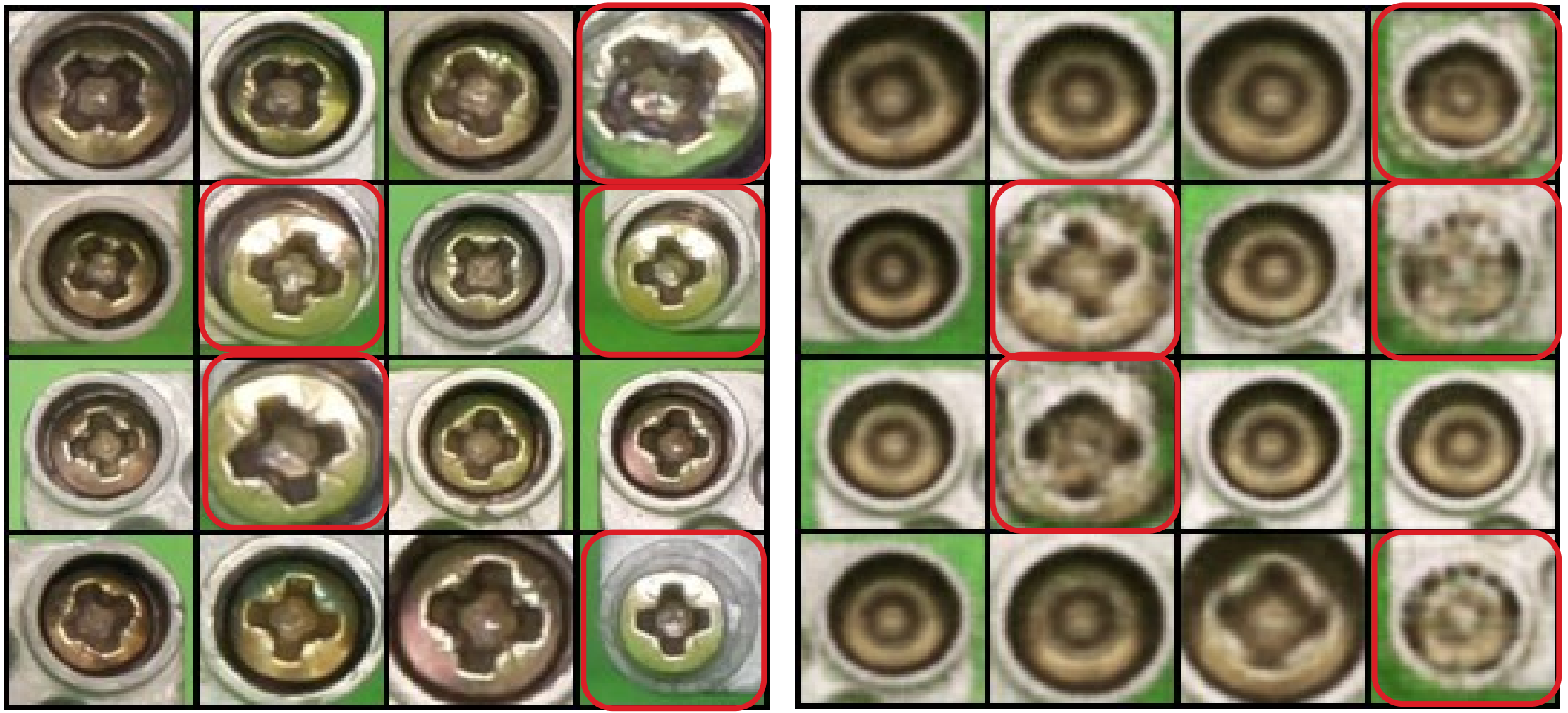
An illustration depicting working of VAE [8]

Defective screws have larger reconstruction error than non-defective ones and can be detected by selecting suitable threshold.

Defective and non-defective screw samples from validation set

Corresponding reconstructed samples

Defective

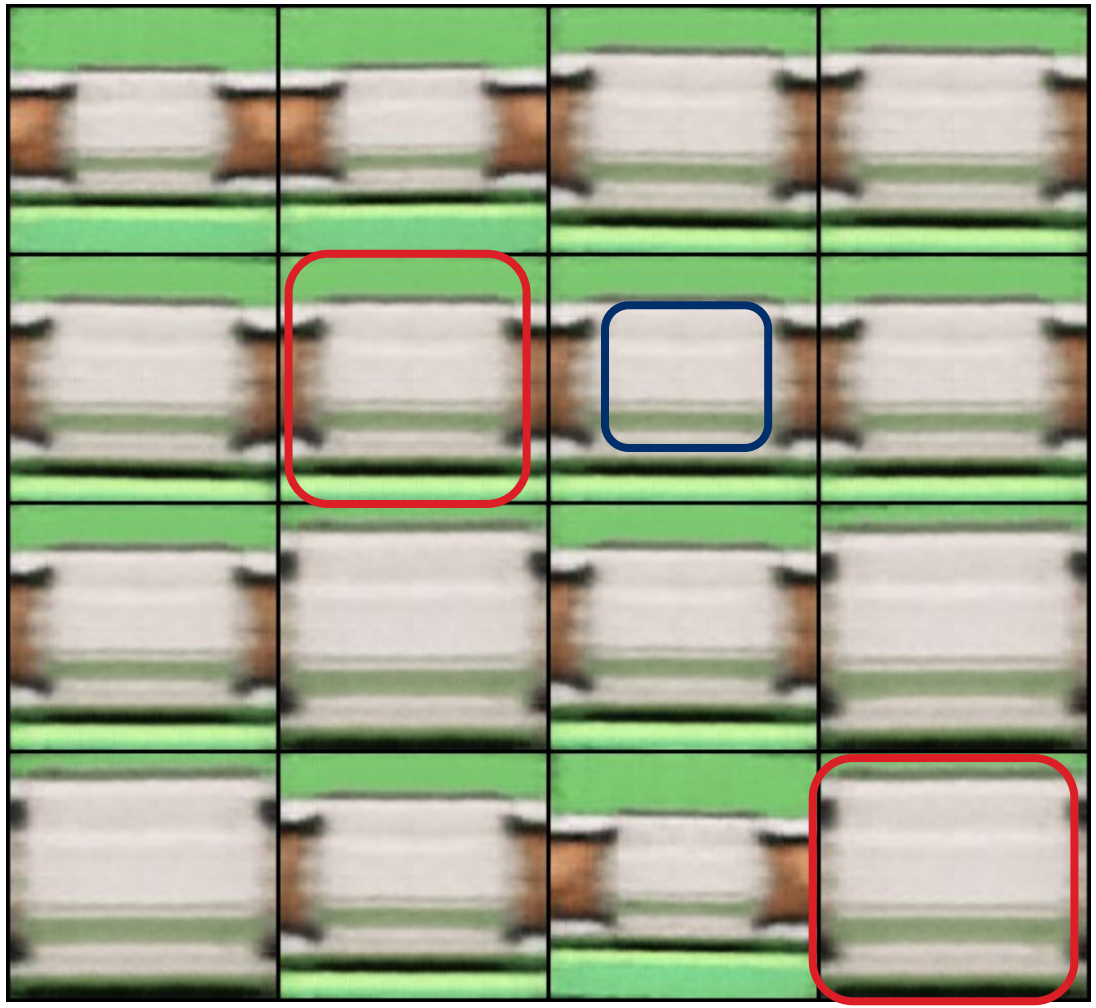


Defective sheet metal packages samples and non-defective ones have similar reconstruction error while stickers on surface also cause problem in reconstruction.

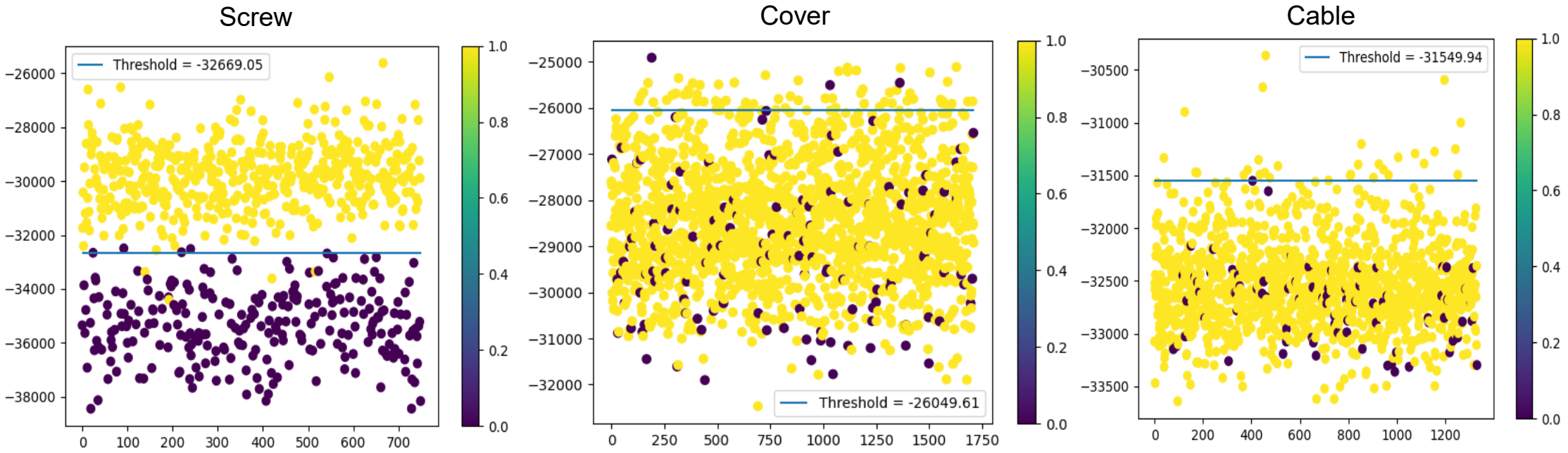
Defective and non-defective sheet metal samples from val. set

Corresponding reconstructed samples

Defective



Only defective screws can be detected based on reconstruction error as rest of the surface defects exhibit similar reconstruction error as non-defective ones.



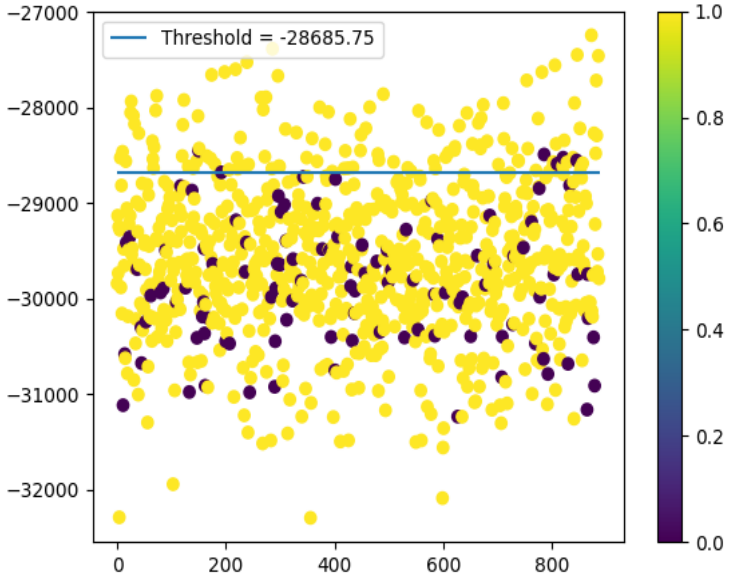
- *Defective samples*
- *Non-defective samples*

x-axis: Samples from validation set

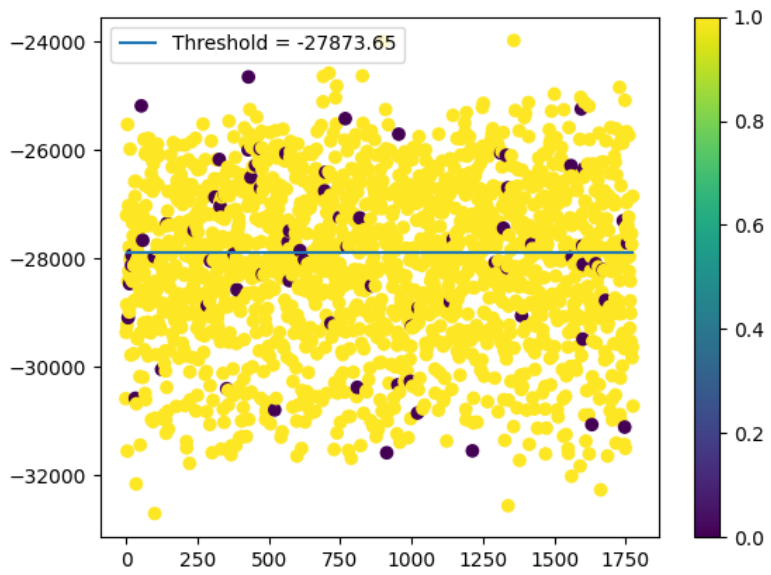
y-axis: BCE reconstruction error (ignore the negative sign)

Only defective screws can be detected based on reconstruction error as rest of the surface defects exhibit similar reconstruction error as non-defective ones.

Sheet Metal Package



Winding Head



- *Defective samples*
- *Non-defective samples*

x-axis: Samples from validation set

y-axis: BCE reconstruction error (ignore the negative sign)

CNNs show better performance than ViT on limited data while anomaly detection based on VAE struggles with complex shapes but works for simple shapes.

Summary

- Automated visual inspection tasks tend to have limited data
- CNNs still perform better than ViT on limited data
- VAE based anomaly detection can work but finding suitable reconstruction error threshold is difficult
- ViT also pose deployment challenges because of their large sizes
- Amount of good quality data take precedence over architecture exploration

Future Outlook

- Self-supervised pretraining can enable better performance
- Knowledge distillation can be used to train smaller models that are easier to deploy
- Generative models can help in synthetic data creation

Prof. Dr.-Ing. Jörg Franke

Institute for Factory Automation and Production Systems

Friedrich-Alexander University Erlangen-Nuremberg

Friedrich-Alexander-Universität
Technische Fakultät

THANK YOU

Bibliography

1. Prabuwono, A. S., Away, Y., & Hasniaty. (2004). Visual Inspection Application in Industry by Integrated Computing System. International Conference on Information Integration and Web-Based Applications & Services. https://api.semanticscholar.org/CorpusID:12501892
2. T. S. Newman and A. K. Jain, “A survey of automated visual inspection,” Computer Vision and Image Understanding, vol. 61, no. 2, pp. 231–262, 1995
3. A. Dosovitskiy, L. Beyer, A. Kolesnikov, D. Weissenborn, X. Zhai, T. Unterthiner, M. Dehghani, M. Minderer, G. Heigold, S. Gelly, J. Uszkoreit, and N. Houlsby, “An image is worth 16x16 words: Transformers for image recognition at scale,” 2021
4. A. Christoph, “Hybrid artificial intelligence in industrial production using the example of visual inspection of spatial assembly units,” 2023
5. G. Huang, Z. Liu, L. van der Maaten, and K. Q. Weinberger, “Densely connected convolutional networks,” 2018
6. Y.-D. Zhang, S. C. Satapathy, X. Zhang, and S.-H. Wang, “COVID-19 diagnosis via DenseNet and optimization of transfer learning setting,” Cognitive Computation, Jan. 2021
7. C.-F. Chen, Q. Fan, and R. Panda, “Crossvit: Cross-attention multi-scale vision transformer for image classification,” 2021
8. D. P. Kingma and M. Welling, “An introduction to variational autoencoders,” Foundations and Trends in Machine Learning, vol. 12, no. 4, p. 307–392, 2019

Alexander Christoph EfficientNetV2-S (24 M parameters) [3] binary classification results for five surface defects are listed here.

Mine Better or eq.

Mine Worse

Alexander binary classification results of defective samples on test set

Type	Precision	Recall	F1
Screw	1.0	1.0	1.0
Cover	0.94	0.97	0.95
Sheet Metal	0.47	0.74	0.57
Cable	0.69	0.35	0.46
Winding Head	0.30	0.27	0.28

Screw confusion matrix

	Predicted defective	Predicted non-defective
Actual defective	267	0
Actual non-defective	0	540

Cover confusion matrix

129	3
8	1645

Sheet metal confusion matrix

80	28
90	750

Cable confusion matrix

59	106
26	1231

Winding head confusion matrix

24	63
56	1753

Defective winding head samples and non-defective ones have similar reconstruction error.

Defective and non-defective winding head samples from val. set

Corresponding reconstructed samples