Towards Sim-to-Real Industrial Parts Classification with Synthetic Dataset

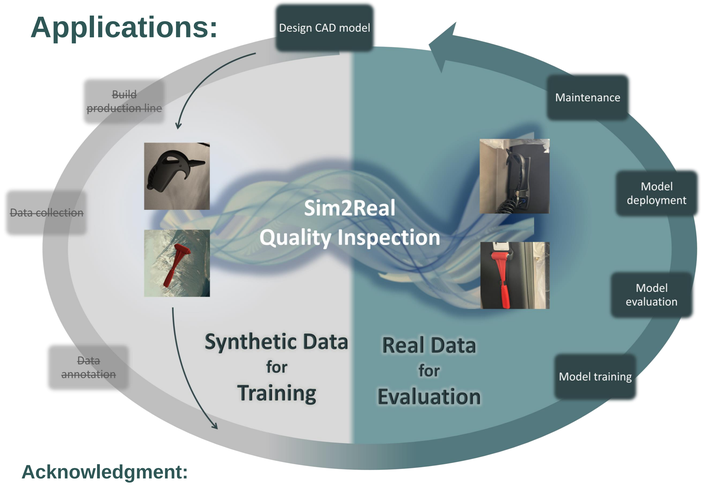
Xiaomeng Zhu[1,3], Talha Bilal[1], Pär Mårtensson[1], Lars Hanson[2], Mårten Björkman[3], Atsuto Maki[3]
[1]Scania CV AB, [2]University of Skövde, [3]KTH Royal Institute of Technology

CVPR VISION'23 VANCOUVER, CANADA

Introduction of the Synthetic Industrial Parts dataset (SIP-17)

- Designed for Sim-to-Real challenge in Industrial Parts Classification.
- Comprises 17 objects that represent six industry use cases. Including isolated parts and assembled parts.
- Syn_O: synthetic data generated from CAD models without random backgrounds, lights, and post-processing. 1,500 images per category.
- Syn_R: synthetic data generated from CAD models with random backgrounds, lights, and post-processing. 1,500 images per category.
- Real: 566 real images for testing.

Applications:



Acknowledgment:

This work is partially supported by the Wallenberg AI, Autonomous Systems and Software Program (WASP) funded by the Knut and Alice Wallenberg Foundation.

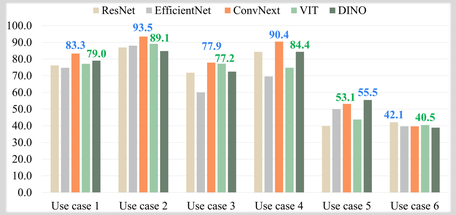
Benchmark:

Five models:

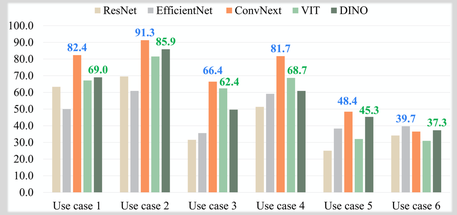
- ResNet 152 [1]
- EfficientNet B7 [2]
- ConvNext base [3]
- Vision Transformer (ViT) base p16 [4]
- DINO self-supervised learning with ViT base p16 [5]

Results:

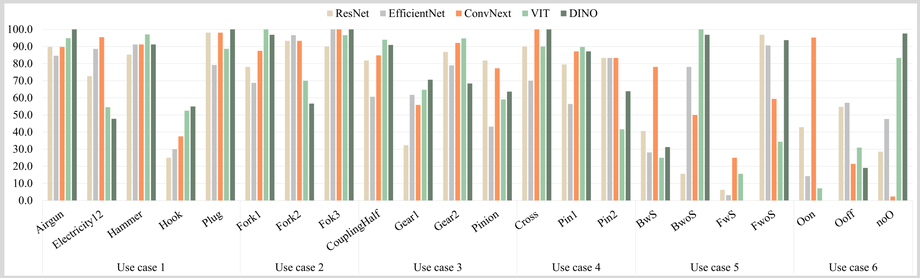
Syn_R accuracy (%)



Syn_O accuracy (%)



Syn_R class-wise accuracy (%)



Reference:

[1] K. He, X. Zhang, S. Ren, and J. Sun. Deep residual learning for image recognition. In CVPR, pages 770–778, 2016.
[2] M. Tan and Q. Le. Efficientnet: Rethinking model scaling for convolutional neural networks. In ICML, pages 6105–6114, 2019.
[3] Z. Liu, H. Mao, C.-Y. Wu, C. Feichtenhofer, T. Darrell, and S. Xie. A convnet for the 2020s. In CVPR, pages 11976–11986, 2022.
[4] A. Dosovitskiy, L. Beyer, A. Kolesnikov, D. Weissenborn, X. Zhai, T. Unterthiner, M. Dehghani, M. Minderer, G. Heigold, S. Gelly, et al. An image is worth 16x16 words: Transformers for image recognition at scale. arXiv preprint arXiv:2010.11929, 2020.
[5] M. Caron, H. Touvron, I. Misra, H. Jégou, J. Mairal, P. Bojanowski, and A. Joulin. Emerging properties in self-supervised vision transformers. In ICCV, pages 9650–9660, 2021.

SIP-17: