Master Thesis proposal

Topic:
Sim to real: unsupervised learning for industrial quality detection with CAD data

Background:
A challenge to applying deep learning-based computer vision technologies for production quality inspection lies in a large number of diverse products and the restricted annotated training data. A possible method to overcome this challenge is utilizing synthetic annotated images generated from CAD models.

Aim:
This thesis aims to develop a 'Sim to Real' object detection model, which can only train with synthetic annotated images generated from CAD models, e.g., Fig. 2, but achieve promising performance on real image captures from camera, e.g., Fig. 1. The model will be evaluated on a provided industrial quality inspection dataset.

Challenges and possible approaches:
In industrial quality inspection, the classification accuracy of the detection model is very important. The model should apply reliable classification on the camera data by filling the domain gap between the cad data and camera data.

The student can first test state-of-the-art object detection models on the provided industry datasets, e.g., Swin Transformer [1], YOLOR [2], FPN [3], Mask RCNN [4], etc., to set a baseline performance.

Then the student can experience different methods to improve the classification accuracy of the baseline models. The possible methods can be inspired by:

- GAN-based: RetinaGAN [5], etc.
- Attention/transformer-based methods: SENet [6], CoAM [7], etc.
- Methods focus on local feature: CoAM [7], strong-weak distribution alignment[8], image processing methods, etc.
- Others.

Notice that the camera images are not available for training. The student is free to test different kinds of methods, but it is important to understand: what methods can possibly help fill the domain gap and improve classification accruacy? and how?

Fig. 1. Real images captured from camera from pedal car front wheel assembly line.

Fig. 2. Synthetic images generated from CAD models of pedal car front wheel assembly.

[1] Z. Liu *et al.*, “Swin Transformer: Hierarchical Vision Transformer using Shifted Windows,” *arXiv:2103.14030 [cs]*, Aug. 2021, Accessed: Nov. 02, 2021. [Online]. Available: http://arxiv.org/abs/2103.14030

[2] C.-Y. Wang, I.-H. Yeh, and H.-Y. M. Liao, “You Only Learn One Representation: Unified Network for Multiple Tasks,” *arXiv:2105.04206 [cs]*, May 2021, Accessed: Nov. 02, 2021. [Online]. Available: http://arxiv.org/abs/2105.04206

[3] T.-Y. Lin, P. Dollár, R. Girshick, K. He, B. Hariharan, and S. Belongie, “Feature Pyramid Networks for Object Detection,” *arXiv:1612.03144 [cs]*, Apr. 2017, Accessed: Apr. 21, 2021. [Online]. Available: http://arxiv.org/abs/1612.03144

[4] K. He, G. Gkioxari, P. Dollár, and R. Girshick, “Mask R-CNN,” *arXiv:1703.06870 [cs]*, Jan. 2018, Accessed: Nov. 02, 2021. [Online]. Available: http://arxiv.org/abs/1703.06870

[5] D. Ho, K. Rao, Z. Xu, E. Jang, M. Khansari, and Y. Bai, “RetinaGAN: An Object-aware Approach to Sim-to-Real Transfer,” *arXiv:2011.03148 [cs]*, Jul. 2021, Accessed: Nov. 02, 2021. [Online]. Available: http://arxiv.org/abs/2011.03148

[6] X. Wang, Z. Cai, D. Gao, and N. Vasconcelos, “Towards Universal Object Detection by Domain Attention,” *arXiv:1904.04402 [cs]*, Jul. 2019, Accessed: Nov. 02, 2021. [Online]. Available: http://arxiv.org/abs/1904.04402

[7] O. Wiles, S. Ehrhardt, and A. Zisserman, “Co-Attention for Conditioned Image Matching,” p. 10.

[8] K. Saito, Y. Ushiku, T. Harada, and K. Saenko, “Strong-Weak Distribution Alignment for Adaptive Object Detection,” *arXiv:1812.04798 [cs]*, Apr. 2019, Accessed: Apr. 21, 2021. [Online]. Available: http://arxiv.org/abs/1812.04798