Machine Vision for Aesthetic Quality Control of Reflective Surfaces

Hansen, Anne Juhler; Philipsen, Mark Philip; Knoche, Hendrik; Moeslund, Thomas B.

Published in:
Proceedings of the International Conference on Artificial Intelligence and Computer Vision (AICV2021)

DOI (link to publication from Publisher):
10.1007/978-3-030-76346-6_36

Creative Commons License
Unspecified

Publication date:
2021

Document Version
Accepted author manuscript, peer reviewed version

Machine Vision for Aesthetic Quality Control of Reflective Surfaces

Anne Juhler Hansen, Mark P. Philipsen, Hendrik Knoche, and Thomas B. Moeslund

Aalborg University, Rendsburggade 14, 9000 Aalborg
{ajha, mpph, hk, tbm}@create.aau.dk

Abstract. Systems for automatic inspection of product quality are in high demand. However, their prevalence is limited by complex development and great expenses. Since inspection systems must be engineered to specific products and environments, such systems are generally only viable with high volume product series. Inspired by human visual inspection of highly reflective brushed aluminium objects we capture images across multiple viewpoints. We employ a spatio-temporal weighting of defects, where defects that occur consistently across viewpoints are considered more severe, and compare to the confidence scores produced by an off-the-shelf object detector (YOLOv5). Our results show the challenges with training object detectors on a realistic low-volume dataset of reflective brushed surfaces. Despite the poor detection performance and difficulty in distinguishing between design and defects, our method proves to classify our small test set with an area under the precision-recall curve of 66.5%.

Keywords: Aesthetics Quality Control, Image Acquisition, Defect Detection

1 Introduction

Quality control of the visual appearance of high quality products is resource intensive but necessary since customers find surface defects on products aesthetically displeasing [1]. Currently, aesthetic quality control in industry is performed by human assessors, resulting in additional labour costs and subjectivity inherent to human visual assessment [2]. Rapid development in computer vision and machine learning may benefit industrial applications of defect detection by embracing technology such as deep learning [3], [4]. We investigate the feasibility of using deep learning (YOLOv5) [5] on low-volume products to create an automated system for detecting visual defects that can support assessors in their aesthetic quality evaluations. Some materials, such as brushed aluminium, are highly reflective and anisotropic, meaning the material drastically change appearance depended on illumination and viewing angle (see Figure 1+2). This variation in the appearance tremendously complicates the defect detection problem. When design features (such as polishing strokes) share visual similarities with defects the problem of making a distinction between defects and surface texture becomes even more challenging [6]. Since defects (scratches, dents, holes, etc.) have clear visibility from some viewpoints while also being invisible from other angles, human quality assessors compensate for this by rotating the objects during visual inspection [2]. With inspiration from the human visual inspection process we propose a data acquisition

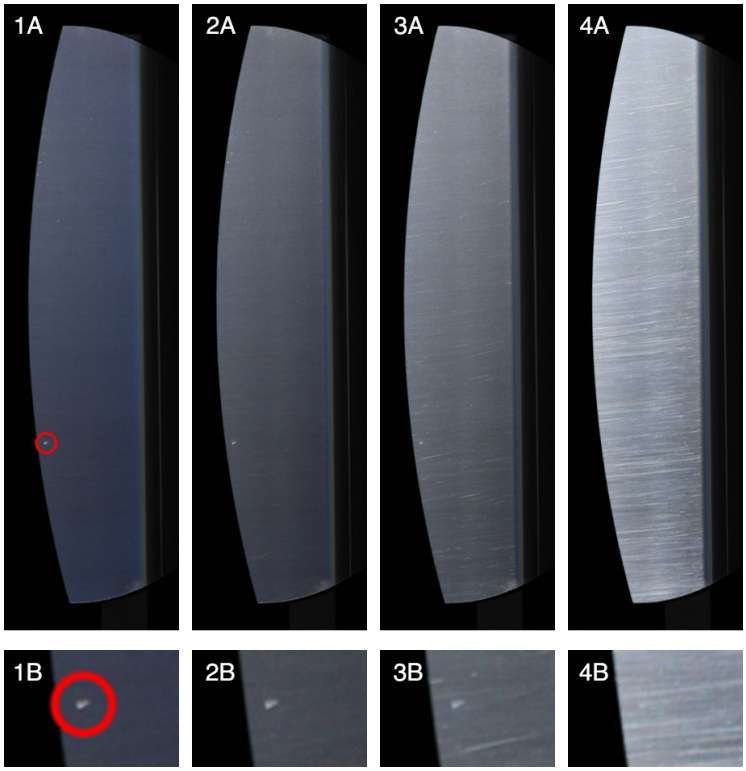


Fig. 1: Defects are visible from certain angles (1A+1B) but as the surface is rotated (2A+2B) the defect becomes less distinct (3A+3B). In 4A+4B the polishing is evident while the defect is hard to detect.

setup using an robotic arm to rotate an object and capture a spatio-temporal image sequence. We use the information from the spatio-temporal image sequence to calculate the *angle of opportunity* (AoO), i.e. the span of view angles from where a potential defect is visible [6]. Combined with the confidence score produced by the YOLOv5 object detector this results in a combined severity score where defects that occur consistently across viewpoints are weighted as more severe compared to defects that only appear from very specific and limited points of view.

1.1 Contribution

We investigate the feasibility of using a deep learning based spatio-temporal detection system for item-level aesthetic quality inspection inspired by the current human quality assessment process. The contributions can be summarized as follows:

– Defect detection with a low-volume products of samples does not work well with a state-of-the-art off-the-self detector.

- Data acquisition setup using a robotic arm to model the visual inspection process of human quality assessors.
- Defect detection pipeline for aesthetic quality control of highly reflective brushed aluminium surfaces.
- Spatio-temporal scoring system for defects based on a combination of the angle of opportunity combined with the confidence score from an off-the-shelf object detector improves performance.

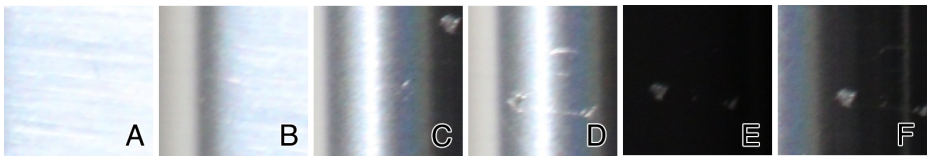


Fig. 2: Brushed aluminium surfaces are reflective and drastically change appearance depended on illumination and viewing angle. (A) OK surface with visible polishing, (B) OK corner, (C+D) defects on the corner, (E+F) defects in low illumination.

2 Related Work

Aesthetic quality is an important parameter within retail where customers demand quality, hence, companies spend plenty of resources searching for visual defects and trying to optimize the defect detection process [1]. Visual appearance can be quantified by the optical properties of materials such as color, gloss, translucency and texture [7], but due to the wide variety in visual appearance of defects it is expensive for industrial applications to establish general and comprehensive databases of defects. Existing solutions are commonly engineered to solve one distinct problem at a time i.e. limiting quality control based on material properties [3] or constrained to individual product types in a controlled setup with fixed object placement and illumination [8]. This results in automatic quality inspection often only being economical with long running and high volume product series, thereupon lacking solutions for low volume high quality manufacturing [9]. Many different defect types exists [10] and hence considerable amount of data is required in order to adequately solve classification tasks within this domain. Existing datasets for defect detection include large amounts of data. One example is the Severstal: Steel Defect Detection data set containing a total of 12568 images and among these 6666 of them include at least one defect [11]. This data set has experienced large interest from researchers and has resulted in a Dice coefficient above 0.9. Likewise the DAGM dataset for optical inspection on textured surfaces consists of 10 different defect classes with a minimum of 150 images of individual defects [12]. Another example is the MVTec AD dataset, which include a total of 5354 images and 70 different types of defects (such as scratches, dents, contamination, and various structural changes) [13]. This dataset is intended for unsupervised anomaly detection and thus includes defect-free images. However, for detecting unique defects in the context

of brushed aluminium products the defect types of the MVTec AD dataset (i.e. textured surfaces of various materials as well as natural images) might not generalize well.

2.1 Defect Detection

Real-time defect detection have previously been performed on different reflective surfaces including highly reflective curved plastic surfaces in the automotive industry [14], diagnosing the penetration state of laser welding of tailor-rolled blanks [15] or finding defects on highly reflective ring components [16]. Tiny casting defects can be detected with a convolutional neural network (CNN) trained from image-level labels with a novel training strategy using an object-level attention mechanism [17], or the surface quality of welds can be predicted using support vector machines for classifying features and auto-encoders for reducing the dimensionality of images [18]. Fusing the results of different object detection principles can improve detection results and provide information fusion [19]. This has been considered for automatic inspection of thermal fuses where incorporating machine vision with artificial neural networks is used for detection of four common defects [20]. In summary, it is common to narrow the scope of defect types to a selected few (due to the access to and amount of training data) and use different inspection algorithms for different types of defect (due to the distinctive features of the defects). Considering the computational expense involved in the construction of deep neural networks, any simplification that can be achieved in the machine vision pipelines is desirable. This has been considered for quality inspection of bottles using image processing methods, region of interest detection in combination with a deep neural network [21]. This shows how integrating conventional image processing methods and pre-processing can benefit computationally expensive methods.

2.2 Generic Object Detectors

Object detectors fall into two categories, single- and multi-stage. These methods differ by the fact that single stage detectors predicts the object location and class simultaneously, while multi-stage detectors rely on a dynamic object proposal step which are then classified. Over the years, several CNN-based object detector algorithms and designs have been proposed [22]. The multi-stage R-CNN method [23], and its subsequent improvements [24, 25, 26] have lead to major improvements in the object detection field. However, the dynamic object proposal step leads to longer processing times. Comparatively, the YOLO single-stage object detectors [5, 27, 28, 29, 30] are capable of real-time object detection, at near comparable detection and classification performance rates of the R-CNN based methods.

3 Data Acquisition

In collaboration with a premium manufacturing company images have been captured of 50 brushed aluminium items. Each item consist of two flat surfaces (front; 173mm$\times$57mm, side; 133mm$\times$70mm) and a convex 90°corner. The ground truth of aesthetic defects in the 50 items has been validated by the manufacturer's expert assessors. Data acquisition

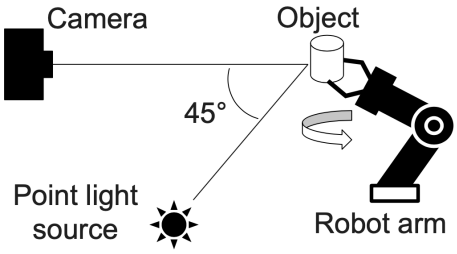
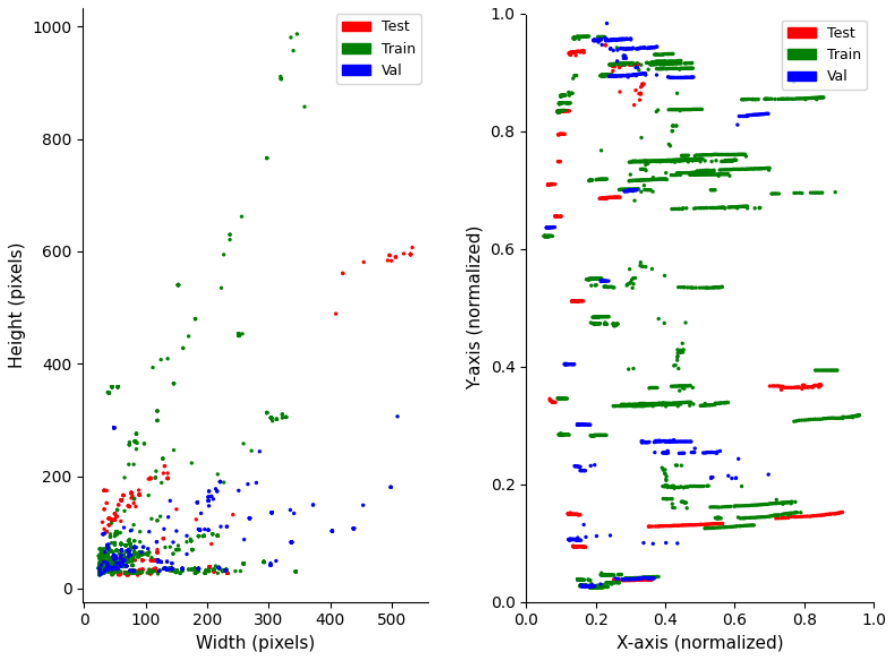


Fig. 3: A robotic data acquisition setup is used to capture image sequences covering a range of viewpoints for each item, an object detector locates individual defects, and item-level defect classification is made based on detection confidence scores and spatio-temporal consistency scores.

consist of a camera, a light source, and a robotic arm for rotating brushed aluminium items in front of the camera and light. RGB images are captured using a Canon 5D (JPEG format with an image resolution of 6050×3300px). The light is a Elinchrom 100W Modeling Lamp with a color temperature of 3200K and a luminous flux of 2700 lumen. It is placed at a 45°angle from the line between the item and camera (see Figure 3). Polishing patterns and defects are most visible when the light rays originate from a single point (versus e.g. diffuse light) hence we select a narrow spot light to best resemble a point light source. The items are fixed upright to a robotic arm (Universal Robots UR10e) that is rotating the items (see Figure 3). Empirical findings exposed that human visual assessment is performed by rotating the items along the direction of the polishing, thus our data collection process includes rotation along the yaw-axis (see Figure 1). The robotic arm is rotated 1 degree between images for front and side and in steps of 5 degrees along the corner. The step size is larger around the corner since these images tend to be overexposed due to the curvature and the highly reflective surface. Since the reflected light is dependent on the rotation of the surface, some images have low illumination and others are overexposed due to the reflective surface (see Figure 2). This complicates the data capturing process but also signifies the importance of capturing a spatio-temporal image series.

3.1 Dataset

The defects are manually labeled using bounding boxes, which differ substantially in size (see Figure 4(a) where the distribution of ground truth bounding box measures are plotted). The majority of defects are relatively small, however our dataset does contain defects up to 1000 pixel high. Larger defects that can span the entire height of the item are rare but exists. In this work these defects are left out since we later sub-sample the images as a pre-processing step. Images from a total of 50 items have been collected. These items along with their respective frame sequences are divided into a training, validation, and test set. The test set consist of 10 items, while the training and validation sets consist of 33 and 7 items, respectively. The number of images and images with



(a) (b)

Fig. 4: (a) Distribution of defect sizes in terms of their bounding box measures. (b) Defect positions across a normalized image coordinate system.

defects in each part is listed in Table 1. It should be noted that the amount of images with defects is substantially larger than the amount of individual defects as the defects are captured in several images through the image sequence. In Figure 4(b) the location of defects is plotted in a normalized image coordinate system. This show how defects, according to the ground truth bounding boxes, can be observed moving across the image as the product is rotated. The long tracks indicate a wide angle of opportunity. Notice also that many defects does not seem to produce long tracks due to their visibility being greatly depended on illumination and view angle.

Table 1: Dataset specifications: Some images have been removed from the training set because they contain defects that are larger than 1000 pixel wide or tall. The size of the complete dataset including these images are (enclosed).

Dataset	**Training**	**Validation**	**Test**
# Items	33	7	10
# Images	2381 (2425)	415	942
# Images with defects	1879 (1948)	563	486

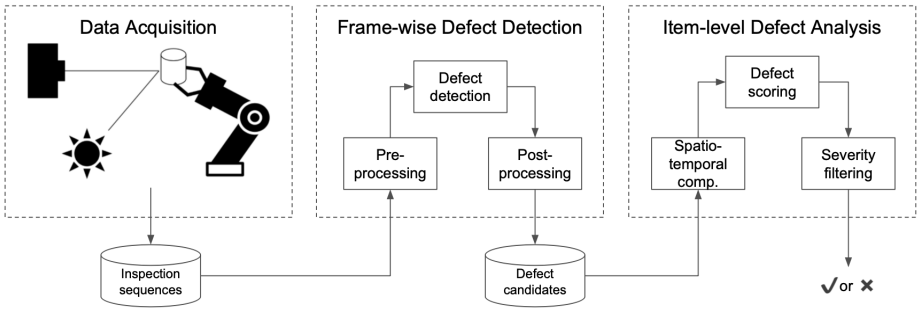


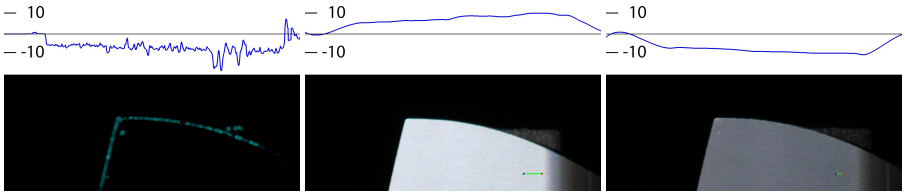
Fig. 5: (Left) Systematic data acquisition resulting in image sequences covering different viewpoints. (Middle) Object detection for frame-wise defect detection. (Right) Item-level defect analysis based on detections and a spatio-temporal consistency score.

4 Method

Here we describe how an off-the-shelf object detector, specifically YOLOv5, can be applied to a severely imbalanced defect detection problem. The proposed method mimics human aesthetic quality assessors by making use of a spatio-temporal (i.e. AoO) data collection and analysis scheme. Figure 5 shows an overview of the components that are involved in our proposed aesthetic quality control system. First data is acquired by a robot that gradually rotates the products in front of a camera and light source. The result is a sequence of frames that cover multiple different viewpoints. For pre-processing prior to defect detection, the high resolution frames are tiled in order to produce image sizes appropriate for a standard object detector. The detector identifies defects across tiles before they are combined and mapped back to the original high resolution frame. In order to produce an item-level decision, the detected defects across frame sequences are collected in *tracks*. This is done by estimating the translation of the product surface due to the rotation using optical flow. As mentioned earlier the AoO is correlated with the severity of a defect here represented by track length. This results in a combined confidence score based on detector confidence and track length.

4.1 Pre- and Post-processing

The dynamic nature of our dataset (including large image resolution, class imbalance, and creating reliable annotations) is challenging since CNNs are not typically well suited for high resolution images or extremely imbalanced datasets [31]. Tiling high resolution images to produce smaller more manageable image patches is standard practice when applying object detectors to high resolution images. The reasons for doing this are keeping the size of the detection networks manageable and avoiding extreme imbalance between background and foreground. Tiling can be done by extracting patches in non-overlapping or overlapping patterns [32], where the overlapping patterns attempt to preserve objects on the boundaries between tiles. We employ a non-overlapping tiling



(a) (b) (c)

Fig. 6: The graphs in the top row show the magnitude of the optical flow along the x-axis. (a) The image intensity indicate the flow direction, with blue and red indicating leftward or rightward motion, respectively. (b)+(c) The noisy flow estimates are filtered along the x-axis for all items. The smoothed flow function is used to propagate the flow vector across the x-axis and estimating the translation due to the rotation of items. All images are cropped on the y-axis for visualisation.

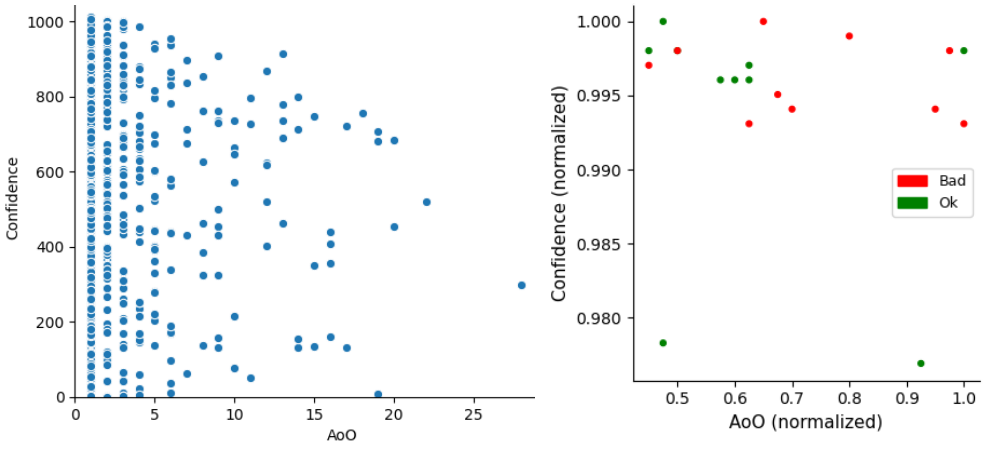
scheme with tile sizes of 1024×1024 pixels and to ensure that the entire image is being covered, the right and bottom edges are padded with black pixels. Post-processing consists of two operations; first any conflicts between overlapping bounding box predictions are resolved by applying non-maximum suppression. Non-maximum suppression ensures that defects are associated with only a single bounding box based on the bounding box overlap quantified using the Intersection over Union (IoU). The confidence score of the detector is used to choose which box to discard. Secondly, the predictions are mapped from the tile coordinate system to the original image coordinate system.

4.2 Spatio-Temporal Compensation

In order to assign detections to tracks it is necessary to compensate for the translation that occurs to the defects on the item surface as the item is rotated. Alternatively, consecutive detections could be matched using similarity or multi-order filters, but since the rotation of items is well defined and consistent it is an option to use flow vectors from optical flow to predict the expected future position of defects and create tracks. Matches are made based on the vicinity of new detections to these predictions. Optical flow is primarily measurable along the edges and very noisy (see Figure 6(a)). To counteract this a smoothed flow map is created using the median values across all images from the observed angle for all available items in the training set. Thereafter, a polynomial function is fitted to the flow values across image columns i.e. the x-axis. The smoothed flow offset along the x-axis can be seen on Figure 6(b) and Figure 6(c) together with the trajectory of selected pixels after the items have been rotated.

5 Results

We present our defect scoring based on three different scores for performing both frame-level and item-level defect detection. (1) The mean detector confidence score across all



(a) (b)

Fig. 7: (a) The main features associated with tracks, specifically AoO and confidence. (b) A summary of the most extreme features from the tracks belonging to each item in the test set. Red and green points represent defective and OK items, respectively.

detections in the highest scoring track. (2) Angle of Opportunity (AoO) based on the length of the highest scoring track. (3) The combined score based on both detector confidence and AoO. The captured tracks can be summarized as seen in Figure 7(a), where all of the tracks found during analysis of *item07* are plotted according to the length and mean confidence associated with each track. As mentioned before the length of the tracks is considered equal to AoO. Considering that the number of true defects on a typical item is around 1-5 it is clear from Figure 7(a) that the number of false detections, and thus tracks, are high. For an item-level assessment we represent an item by the highest scoring track, corresponding to a point in the upper right corner of Figure 7(a). Figure 7(b) shows each of the items in the test set represented by their highest scoring track. Defective items are colored red and items without serious defects are green. Table 2 presents the defect detection results from the frame-level detection subsystem. With the small defect size it proves difficult to satisfy the standard IoU criteria, for this reason we also show precision and recall with a modest IoU of 0.1. As evident from the extremely low precision, the detector also proves to have difficulties in distinguishing between desired design features (e.g. the polishing texture) and defects. With item-level defect detections from the detector and our proposed method for temporal consistency scoring (AoO), defective items can at best be detected such that we get an area under the precision-recall curve of 66.5% as seen in Table 3.

6 Discussion

Unlike the previous datasets discussed in Section 2 who worked on large datasets we investigated a use-case with low-volume products and reflective surfaces with varying

Table 2: Frame level defect detection performance on test and (validation) set using the YOLOv5 detector at different IoU.

Threshold	Precision [%]	Recall [%]
IoU = 0.5	0.29 (0.180)	10.91 (16.70)
IoU = 0.1	1.49 (0.407)	56.79 (37.83)

Table 3: Defect detection performance on test and (validation) set using detector confidence, AoO, and a combination.

Method	AUC [%]
Confidence	53.9 (83.1)
AoO	**66.5** (87.5)
Combined	60.7 (84.4)

visual appearance. Therefore, it is expected that the data intensive deep learning method cannot achieve similar performance to previous defect detection tasks e.g. within bottle defect classification (accuracy of 99.60%) or thermal fuse bur defect detection (accuracy of 98.43%). Our detector produces a large number of false positive with high confidence (see Figure 7). This is likely caused by the ambiguity in the visual appearance of defects and the limited size of the training set. The number of long tracks on the other hand lies in a more reasonable range. The main weakness in our pipeline is the poor performance of the detector in terms of accurately localizing defects, coping with very large objects and distinguishing between the intended surface texture and unwanted defects. However, we find that the AoO is a powerful indicator of the presence of a defect. This suggests that if a detector can be designed to work within the requirements of a low-volume data regime, and thereby lower the amount of false positive, the performance of the system will improve further. An indication of this is shown in Table 3, where the AUC performance of the combined score (detector confidence and AoO) is 6.8 percentage points larger than only using the detector confidence scores, while still being 5.8 percentage points behind only using AoO. The localization of defects and the problems with handling large objects may be addressed by transitioning from a bounding box detector to a CNN for pixel-level segmentation. Limitations of the data capturing system include overexposure in images due to the highly reflective surfaces being captured. Future work could include the use of multiple exposure values when capturing images in order to make sure all areas of the images are visible. Also, future work should include improving the performance of the system and making it applicable in industrial contexts. Performance could be improved using data augmentation and loss weighting schemes to better cope with the imbalanced data issues and the problems of the current detector producing a large number of false positives.

7 CONCLUSIONS

We investigate the feasibility of a deep learning-based defect detection framework, for low-volume highly reflective brushed aluminum products. Inspired by the human visual assessment workflow a data acquisition framework is developed using a robotic arm to capture images across multiple views. A spatio-temporal detection system is proposed using a state-of-the-art object detector, YOLOv5, combined with an optical flow-based tracking to determine the angle of opportunity per defect.

In our evaluation we find that the object detector produces a large amount of false positive detections due to the difficult task of detecting defects in our low sample and imbalanced setting. This results in an area under the precision-recall curve of 53.9% for the detector, and 60.7% for the full spatio-temporal detection framework. However, we find that by using just the AoO an AUC of 66.5% is obtained. Furthermore, our results indicate that if an object detector can be designed to produce fewer false positives the performance of our full spatio-temporal defect detection framework for low-volume products will improve.

Acknowledgements. This work was funded by Manufacturing Academy of Denmark (MADE) and the Innovation Fund Denmark.

References

[1] Y.-l. Deng, S.-p. Xu, and W.-w. Lai. A novel imaging-enhancement-based inspection method for transparent aesthetic defects in a polymeric polarizer. *Polymer Testing*, 61: 333–340, 2017.

[2] A. J. Hansen, H. Knoche, and T. B. Moeslund. Getting crevices, cracks, and grooves in line: Anomaly categorization for aqc judgment models. In *2018 Tenth International Conference on Quality of Multimedia Experience (QoMEX)*, pages 1–3. IEEE, 2018.

[3] T. Czimmermann, G. Ciuti, M. Milazzo, M. Chiurazzi, S. Roccella, C. M. Oddo, and P. Dario. Visual-based defect detection and classification approaches for industrial applications—a survey. *Sensors*, 20(5):1459, 2020.

[4] X. Xie. A review of recent advances in surface defect detection using texture analysis techniques. *ELCVIA: electronic letters on computer vision and image analysis*, pages 1–22, 2008.

[5] Ultralytics. Yolov5, 2020. URL https://github.com/ultralytics/yolov5. last accessed: August 20, 2020.

[6] A. J. Hansen, H. Knoche, and T. B. Moeslund. Defect or design? leveraging the angle of opportunity for classifying scratches on brushed aluminium surfaces. *Submitted*, 2021.

[7] M. Pointer. CIE TC1-65 - a framework for the measurement of visual appearance. *CIE Publication*, pages 175–2006, 2006.

[8] H. S. El-Mesery, H. Mao, and A. E.-F. Abomohra. Applications of non-destructive technologies for agricultural and food products quality inspection. *Sensors*, 19(4):846, 2019.

[9] E. Verna, G. Genta, M. Galetto, and F. Franceschini. Planning offline inspection strategies in low-volume manufacturing processes. *Quality Engineering*, pages 1–16, 2020.

[10] EN ISO 8785. Geometrical Product Specification (GPS) – Surface imperfections – Terms, definitions and parameters. *European Committee for Standardization*, 1999.

[11] Severstal: Steel Defect Detection, Kaggle competition. https://www.kaggle.com/c/severstal-steel-defect-detection/overview. Accessed: 2020-12-30.

[12] DAGM optical inspection, Kaggle competition. https://www.kaggle.com/mhskjelvareid/dagm-2007-competition-dataset-optical-inspection. Accessed: 2020-12-30.

[13] P. Bergmann, M. Fauser, D. Sattlegger, and C. Steger. Mvtec ad–a comprehensive real-world dataset for unsupervised anomaly detection. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 9592–9600, 2019.

[14] G. Rosati, G. Boschetti, A. Biondi, and A. Rossi. Real-time defect detection on highly reflective curved surfaces. *Optics and Lasers in Engineering*, 47(3-4):379–384, 2009.

[15] Z. Zhang, B. Li, W. Zhang, R. Lu, S. Wada, and Y. Zhang. Real-time penetration state monitoring using convolutional neural network for laser welding of tailor rolled blanks. *Journal of Manufacturing Systems*, 54:348–360, 2020.

[16] R. A. Boby, P. S. Sonakar, M. Singaperumal, and B. Ramamoorthy. Identification of defects on highly reflective ring components and analysis using machine vision. *The International Journal of Advanced Manufacturing Technology*, 52(1-4):217–233, 2011.

[17] C. Hu and Y. Wang. An efficient cnn model based on object-level attention mechanism for casting defects detection on radiography images. *IEEE Transactions on Industrial Electronics*, 2020.

[18] J. Günther, P. M. Pilarski, G. Helfrich, H. Shen, and K. Diepold. Intelligent laser welding through representation, prediction, and control learning: An architecture with deep neural networks and reinforcement learning. *Mechatronics*, 34:1–11, 2016.

[19] C. Hofmann, F. Particke, M. Hiller, and J. Thielecke. Object detection, classification and localization by infrastructural stereo cameras. In *VISIGRAPP (5: VISAPP)*, pages 808–815, 2019.

[20] T.-H. Sun, F.-C. Tien, F.-C. Tien, and R.-J. Kuo. Automated thermal fuse inspection using machine vision and artificial neural networks. *Journal of Intelligent Manufacturing*, 27(3): 639–651, 2016.

[21] J. Wang, P. Fu, and R. X. Gao. Machine vision intelligence for product defect inspection based on deep learning and hough transform. *Journal of Manufacturing Systems*, 51:52–60, 2019.

[22] L. Liu, W. Ouyang, X. Wang, P. Fieguth, J. Chen, X. Liu, and M. Pietikäinen. Deep learning for generic object detection: A survey. *International Journal of Computer Vision*, 128(2): 261–318, October 2019. doi: 10.1007/s11263-019-01247-4. URL https://doi.org/10.1007/s11263-019-01247-4.

[23] R. Girshick, J. Donahue, T. Darrell, and J. Malik. Rich feature hierarchies for accurate object detection and semantic segmentation. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 580–587, 2014.

[24] R. Girshick. Fast r-cnn. In *2015 IEEE International Conference on Computer Vision (ICCV)*, pages 1440–1448, 2015. doi: 10.1109/ICCV.2015.169.

[25] K. He, G. Gkioxari, P. Dollár, and R. Girshick. Mask r-cnn. In *Proceedings of the IEEE international conference on computer vision*, pages 2961–2969, 2017.

[26] S. Ren, K. He, R. Girshick, and J. Sun. Faster r-cnn: Towards real-time object detection with region proposal networks. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 39(6):1137–1149, 2017. doi: 10.1109/TPAMI.2016.2577031.

[27] A. Bochkovskiy, C.-Y. Wang, and H.-Y. M. Liao. Yolov4: Optimal speed and accuracy of object detection, 2020.

[28] J. Redmon and A. Farhadi. Yolo9000: Better, faster, stronger. In *2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 6517–6525, 2017. doi: 10.1109/CVPR.2017.690.

[29] J. Redmon and A. Farhadi. Yolov3: An incremental improvement. *arXiv preprint arXiv:1804.02767*, 2018.

[30] J. Redmon, S. Divvala, R. Girshick, and A. Farhadi. You only look once: Unified, real-time object detection. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 779–788, 2016.

[31] J. M. Johnson and T. M. Khoshgoftaar. Survey on deep learning with class imbalance. *Journal of Big Data*, 6(1):27, 2019.

[32] F. Ozge Unel, B. O. Ozkalayci, and C. Cigla. The power of tiling for small object detection. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops*, pages 0–0, 2019.