https://doi.org/10.1016/j.eswa.2021.116000

CenterNet-based Defect Detection for Additive Manufacturing

Wang Ruoxin and Cheung, Chi Fai*
State Key Laboratory of Ultra-precision Machining Technology, Department of Industrial and Systems Engineering, The Hong Kong Polytechnic University, Hung Hom, Kowloon, Hong Kong, China.

ABSTRACT:

Additive manufacturing (AM) has been widely used in the fabrication of optical components. However, surface defects generated during the AM process have an adverse effect on surface quality. Although some studies have explored the defect features based on the processing of information including images, acoustic signals, thermal history, etc., they focus mainly on defect classification or one type of defect detection. Over recent years, convolution neural networks have displayed promising performance in object detection in images in various fields. Therefore, in this paper, to detect and characterize surface defects more comprehensively and accurately, a novel defect detection model based on CenterNet is presented to extract the defect features, including type, location and count simultaneously, in which there are four output heads to predict heatmaps, object size, local offset, and density map, respectively. Moreover, count loss is added in the original objective function to boost the detection performance. To perform the model validation, surface defect dataset is captured through scanning electron microscope on the surfaces of the workpiece made of 316L fabricated by AM. A series of experiments was conducted and the proposed model achieved better detection accuracy on defect dataset compared with other state-of-the-art models.

Key words: Additive manufacturing, Defect detection, Selective laser melting, Convolutional neural network (CNN), Density map estimation, Surface defects, Machine learning, Precision measurement

* Corresponding author. Tel.: +852 2766 7905; fax: +852 2362 5267. E-mail address: benny.cheung@polyu.edu.hk (Chi Fai Cheung).

1 Introduction

With the demand for products possessing highly complex structures and high value addition, additive manufacturing (AM) has become more prominent as compared to traditional manufacturing techniques due to its high efficiency and low cost. AM not only makes use of less material and time in design and manufacture but also avoids assembly time when fabricating multiple parts because of the capability of converting 3D CAD models into products directly. As a result, AM has been widely applied in many areas, such as aerospace, medicine, tissue engineering, energy and optics. Thanks to the flexibility of AM in components design and fabrication, it exhibits more suitability for optical manufacturing, in which the scales of the geometrical features vary in one setting. Applications of AM in optical components thus have undergone rapid developments, such as embedded optical elements, optical crystal, full three-dimensional printed confocal imagers and eagle eye. There are many AM techniques applied in optical manufacturing, including fused deposition modelling (FDM),

stereolithography (SLA), multi-photon stereolithography (MPS), inkjet printing, direct inkjet writing and selective laser melting (SLM). SLM (Yap et al., 2015) is one kind of powder bed fusion (PBF) technology, which scans a thin layer of power spread on the build stage using a laser beam to melt a specific region thereby forming a layer of a target component and then deposits another layer of powder to complete a new layer of fabrication. The component is built completely by repeating this process. Although SLM has many advantages, such as better surface finish, better resolution and higher material utilization, which enable superiority to produce complex and small components, many defects are generated because of complex physical processes, including laser beam spraying, material evaporation, powder melting and solidification (Nagarajan et al., 2019; Malekipour et al., 2018). Defect features have a determinant effect on the surface integrity and mechanical properties of additive manufactured components. As a result, defect detection has attracted much attention.

There are some computer vision methods for AM defect detection. Scime and Beuth (2018) adopted a filter bank which includes Gaussian, Uniform Averaging Disk, Difference of Gaussian, Oriented Edge Detectors, Oriented Line Detectors, Streak Detectors and Gabor, to obtain defect features and detected six common powder bed anomalies. Khanzadeh et al. (2018) used functional principal component analysis method to capture morphological features and five machine learning methods to identify abnormal porosity. Anomalous computed tomography voxels were extracted by Gobert et al. (2018) through 3D Gaussian filter and k-means method was utilized for clustering them. However, these traditional methods only capture low-level features and fail to extract the deep sematic information from images, which is important to understand the process of AM and achieve further improvement. In recent years, convolutional neural networks (CNNs) with powerful capacity of feature extraction have been widely used in defect detection and performed well. Wang et al. (2020) proposed a smart surface inspection system based on faster RCNN to detect the defects of turbo blades with a precision of 0.81. Yang et al. (2020) developed an optimized Visual Geometry Group (VGG) model to classify defects of laser welding with a testing accuracy of 99.87%. Song et al. (2019) detected micro-scratches on metal components using a deep CNN with 0.40mm average error. Liong et al. (2019) automatically segmented the defects of leather using mask RCNN and found 91.5% segmentation accuracy. Therefore, in this paper, the development of a novel CenterNet-based model with count loss (CenterNet-CL) for the purpose of detecting and characterize surface defects is presented and the model was built based on CenterNet so as to count, classify and localize defects, in which four heads are responsible for density map, heatmap, object size and local offset prediction, respectively. They share the weights of a backbone network. The loss function is modified to boost the detection performance through adding count loss. To evaluate the performance of the CenterNet-CL model, a series of experiments was conducted to construct the defect dataset, in which workpiece images were acquired from the components produced by SLM and types of ground-truth defects and their positions were labelled and tagged. The CenterNet-CL model was found to yield 48% AP^{50} and 18%AP in regard to the defect dataset which is significantly higher than some traditional backbone networks such as dilated convolution layers (DCL), etc. and other state-of-the-art methods.

2 Related works

Object detection is a most popular and challenging problem in computer vision, which aims to locate and classify objects in images. Currently, there are two detection frameworks: one-stage detectors, such as YOLO and two-stage detectors, such as R-CNN. Two-stage detectors normally adopt techniques to obtain a set of proposals and then train region classifiers for each proposal. However, one-stage detectors predict the class of each location of image directly. Due to the time-consuming region proposal generation, two-stage detectors commonly cost more time but yield better detection performance on most public benchmarks. However, one-stage detectors have more promising future in real-time detection scenarios. Since the use of CNNs is a powerful technique to learn the representations from data directly, they are applied in many fields for object detection, such as unmanned driving (Cebollada et al.,2020), robot vision (Soyguder, 2011), face recognition (Tripathi & Jalal, 2021), video surveillance (Mabrouk & Zagrouba, 2018) and security (Roldán et al., 2020). There also are some CNN-based methods applied in AM defect detection.

2.1 One-stage detection

There are many studies which focus on exploring one-stage framework detection. OverFeat (Sermanet et al., 2013) replaced the fully connected layer with a fully convolutional layer based on AlexNet (Krizhevsky et al., 2017) to allow random size input and offset max-pooling was developed to reduce the subsampling ratio and achieve better resolution. Due to the sharing computation of overlapping regions, it was time efficient. However, the classifiers and regressors were not optimized simultaneously. After that, a real-time detector, YOLO (Redmon et al., 2016), was proposed, which divides the image into a fixed number of grid cells while the class probability, coordinates of bounding box and confidence scores in each grid are determined. The detection speed of YOLO could reach 45 FPS (Frames Per Second) on VOC2007, a dataset proposed by Everingham et al. (2007), with 63.4% mAP (mean Average Precision) in real time, since it was an end-to-end trainable model with lightweight architecture and could jointly classify and localize. However, YOLO tended to predict false positives in the background and made more localization errors.

In order to improve the detection accuracy and maintain real-time speed, a single shot multibox detector (SSD) was developed by Liu et al. (2016), which set several default boxes with different aspect ratios in different scale feature maps at each grid cell. It was based on VGG-16, which was proposed by Simonyan and Zisserman's (2014) model with localization loss and confidence loss. It achieved 74.3% mAP on VOC2007 test at 59 FPS. Inspired by the SSD, Redmon and Farhadi (2017) made many improvements based on YOLO, including adding batch normalization, increasing the resolution of input images, and getting better priors of anchor box dimensions through k-means. In order to detect more object categories, they adopted the WordTree hierarchy to combine ImageNet and MS COCO which are two public image datasets created by Deng et al. (2009) and Lin et al. (2014) and train model on classification and detection jointly. The previous studies all rely on anchor boxes to detect objects.

Recently, a novel anchor-free framework called CornerNet was proposed by Law and Deng (2018), which detected the top-left corner and the bottom-right corner of a target box to obtain the prediction box using two separate modules. The basic feature map extracted by two stacked

Hourglass networks was shared by these two prediction modules, which included corner pooling and corresponding heatmaps, pair embeddings and offsets together. Heatmaps were responsible for predicting the score of corners. The embedding vector made the distance between the two corners (top-left corner and bottom-right corner) of the same objects the shortest, and offsets were used to generate a tighter bounding box. CenterNet was proposed by Zhou et al. (2019) to improve the performance of CornerNet, which regards object detection as a standard key point estimation problem. Images are fed into an end-to-end convolutional network to generate a heatmap, local offset and object size. This anchor-free method is a promising direction since it avoids many drawbacks of anchor boxes, such as long training time, hard to determine the hypermeters of anchor boxes, causing an imbalance of positive and negative samples.

2.2 Two-stage detection

Since high detection accuracy has been possible, many two-stage detectors based on deep learning have been developed, such as R-CNN, SPPNet, fast R-CNN, faster R-CNN, and RFCN. R-CNN (Girshick et al., 2014) firstly extracts 2,000 proposals using selective search and crops them to obtain the feature representations through a CNN. Hence, class-specific linear SVMs classify each region and bounding box regressors modify the original proposals to obtain an accurate location. Based on the R-CNN, SPPNet (He et al., 2015) introduced spatial pyramid pooling to allow arbitrary size input images, which could obtain a fixed-length representation for any image scale and avoid rescale pre-processing, thereby improving the recognition accuracy. Similar to SPPNet, fast R-CNN (Girshick, 2015) also generates fixed-length region features through a novel technique, ROI pooling. The main idea is replacing warping at the image level with warping at the feature level. Fast R-CNN speeds up the detection time by about 3 times in training and 10 times in testing. Instead of exploring the pooling layer, faster R-CNN Ren et al. (2015) proposed a Region Proposal Network (RPN) to generate the region proposal rather than using selective search, which shares convolutional computation of a backbone network to improve the detection efficiency. RPN first generates a few anchor boxes with different aspect ratios and scales. In each anchor, the feature vector is fed into two sibling full connected layers, one for classification and the other for box regression. Faster R-CNN is a purely CNN-based network through generating region proposals with a CNN. In order to further improve the efficiency, RFCN was proposed by Dai et al. (2016), in which a set of position-sensitive score maps are extracted to construct a shared ROI sub-network.

2.3 Defect detection in AM

Some researchers have paid attention to detecting defects of AM. Zhang et al. (2018) adopted a CNN to extract the features of defects and classified them into three categories, namely melt pool, plume and spatter through a support vector machine method. Caggiano et al. (2019) used two parallel CNNs to combine laser scanning and a powder recoating image thereby boosting classification performance. Wong et al. (2021) used U-Net (a U-shape CNN) to segment AM defects automatically based on X-ray computed tomography images. Although much effort has been made to explore the defect features, only information of defect type or

defect distribution, or the information about one type of defect is inadequate to characterize the surface and evaluate the surface quality of a workpiece. In order to fill this research gap, a novel CNN is presented to detect and characterize defect types, localization and count, simultaneously, which provides comprehensive information to quantify the surface quality of workpieces fabricated by AM. The proposed CenterNet-CL approach is developed based on CenterNet with an hourglass (Newell et al., 2016) backbone network, which is a state-of-the-art detection model and yields high detection accuracy and speed in the MS COCO dataset. The density map estimation and detection task were integrated through adding Bayesian Loss (Ma et al., 2019) to evaluate the density map quality and boost the detection performance.

3 Defect detection model based on CenterNet-CL

3.1 Structure of the model

The architecture of the proposed defect characterization model is illustrated in Fig.1, in which an output head architecture is designed to predict the density map and it serves as an intermediate supervision to modulate the weights of the first Hourglass network based on CenterNet (Zhou et al., 2019). Surface defect images of width W_o and height H_o, as input images $I \in R^{W_o * H_o * 3}$, are fed into a stacked hourglass backbone neural network. The first Hourglass network is followed by four output heads, including heatmaps, object size, local offset, and density map. Here, peaks in the heatmap $\hat{P} \in R^{W*H*C}$ correspond to object centres, where W and H represent the output width and height, respectively and C is the number of defect types. The local offset $\hat{O} \epsilon R^{W*H*2}$ compensates for the discretization error caused by the output stride k. The object size $\hat{S} \epsilon R^{W*H*2}$ gives the width and height of objects. Except density map $\hat{D} \epsilon R^{W*H*1}$, which represents the distribution of defects, the remaining output features are processed by a subsequent Hourglass network to further learn and evaluate higher-order spatial relationships. The architecture of the density map output head is shown in the dotted box, which is composed of four convolutional layers. '256-bn-relu' represents the convolutional layer with 256 filters, followed by a batch normalization layer and RELU activation function. '128-relu' means that there is no batch normalization layer. The final outputs of the model predict the defect class $\hat{c}$, bounding box, $\widehat{bb}$ count $\widehat{ct}$ and confidence score $\hat{p}$.

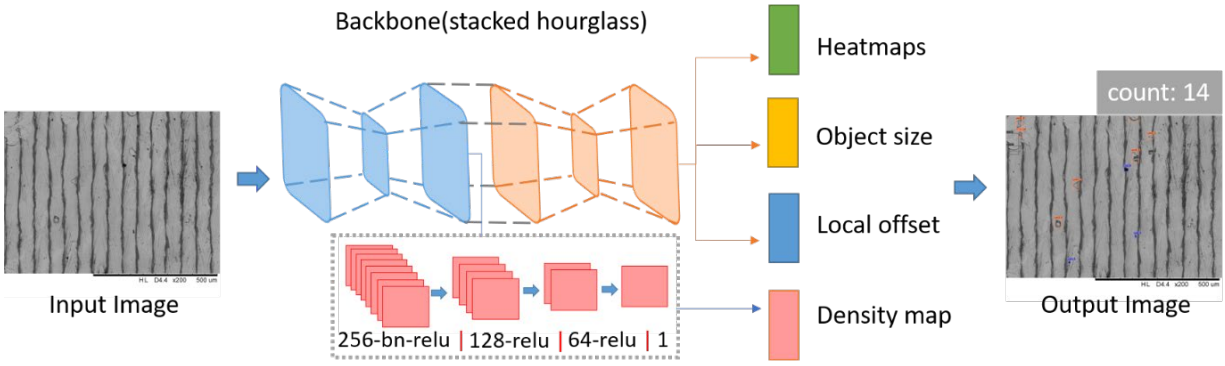


Fig. 1. The architecture of the defect characterization model

3.2 Loss function

Focal loss. Assume that $\hat{p}_{x,y,c}$ is the prediction score at location (x, y) for class c, where

$x \in (1, W), y \in (1, H), c \in (1, C)$, $W = \frac{W_o}{k}, H = \frac{H_o}{k}$, k is the output downsample ratio. The "ground-truth" key point (x_o, y_o) in the input image is mapped into the "ground-truth" heatmap location $(\tilde{x}, \tilde{y})$, $\tilde{x} = \left\lfloor \frac{x_o}{k} \right\rfloor, \tilde{y} = \left\lfloor \frac{x_o}{k} \right\rfloor$. Here $\lfloor x \rfloor$ means $floor(x)$. The "ground-truth" heatmap $p_{x,y,c}$ is determined by using unnormalized 2D gaussian kernel as Eq. (1), where σ equals radius/3 and the radius is varied according to object size, following the same setup in Law and Deng (2018) and Zhou et al. (2019).

$$p_{x,y,c} = \exp\left(-\frac{(x-\tilde{x})^2 + (y-\tilde{y})^2}{2\sigma^2}\right) \tag{1}$$

The objective is to minimize the focal loss L_{foc}(Lin et al., 2017), defined by Eq. (2), in which N is the number of objects in an image.

$$L_{foc} = -\frac{1}{N}\sum_{c=1}^{C}\sum_{x=1}^{W}\sum_{y=1}^{H}\begin{cases} (1-\hat{p}_{x,y,c})^{\alpha}\log(\hat{p}_{x,y,c}) & \text{if } p_{x,y,c}=1 \\ (1-p_{x,y,c})^{\beta}\hat{p}_{x,y,c}{}^{\alpha}\log(1-\hat{p}_{x,y,c}) & \text{otherwise} \end{cases} \tag{2}$$

Offset loss. Offset loss L_{off} is utilized to evaluate the error caused by the output stride. It adopts the traditional $L1$ loss, and offset prediction $\hat{O}_n$ is shared by all classes c. The loss function is illustrated in Eq. (3).

$$L_{off} = \frac{1}{N}\sum_{n}^{N}\left|\hat{O}_n - O_n\right| \quad s.t.$$
$$O_n = \left(\frac{x_n}{k} - \left\lfloor\frac{x_n}{k}\right\rfloor, \frac{y_n}{k} - \left\lfloor\frac{y_n}{k}\right\rfloor\right) \tag{3}$$

Size loss. Size loss also makes use of $L1$ loss function to evaluate the bounding box accuracy. Let $bb_c^n = (x_t^n, y_t^n, x_b^n, y_b^n)$ be the "ground truth" bounding box of object n with class c. (x_t^n, y_t^n) and (x_b^n, y_b^n) are the top-left and bottom-right coordinate of the bounding box. As a result, the object size is $S_n = (x_b^n - x_t^n, y_b^n - y_t^n)$. The model is trained to minimize the size loss function, as shown in Eq. (4) so as to obtain the accurate $\hat{S}_n$.

$$L_{size} = \frac{1}{N}\sum_{n}^{N}\left|\hat{S}_n - S_n\right| \tag{4}$$

Count loss. Since defects sometimes have blurry boundaries with background and the number of defects on surface is an important criterion for evaluating surface integrity, count loss is added to correct the detection prediction. The count loss makes use of the Bayesian loss function (Ma et al., 2019) to evaluate the density map estimation quality.

$$L_{cnt} = \sum_{n=1}^{N}\|1 - E[c_n]\|_1 \tag{5}$$

$$E[c_n] = \sum_{m=1}^{M} p(y_n|l_m)\widehat{D}(l_m) \tag{6}$$

where l_m denotes m_{th} 2D pixel location and M is the pixel number of the density map; $\widehat{D}(l_m)$ denotes the estimated density map; $p(y_n|l_m)$ is the posterior probability of l_m determined by Bayes' theorem; c_n denotes the total count associated with y_n.

As a result, the overall loss function of the defect characterization model is defined as Eq. (7)

$$L = L_{foc} + \lambda_{off}L_{off} + \lambda_{size}L_{size} + \lambda_{cnt}L_{cnt} \tag{7}$$

3.3 Evaluation Metrics

Two important and common criteria in detection problems are always used to evaluate the detection accuracy, namely mean average precision (mAP) and average recall (AR). For a detailed definition refer to Everingham et al. (2007). Here precision and recall are defined by Eq. (8) and Eq. (9) respectively.

$$Precision = \frac{TP}{TP + FP} \tag{8}$$

$$Recall = \frac{TP}{TP + FN} \tag{9}$$

where TP is true positive; FP is false positive; FN is false negative. AP is the area value under the P-R curve of one class, which is drawn by Precision and Recall. mAP is a further evaluation metric for multi-class detection, which averages the AP value among all classes. Similar to mAP, AR is the average value among all classes.

In this paper, some evaluation metrics of the MS COCO dataset, a public dataset containing common object images proposed by Lin et al. (2014) are adopted including AP, AP^{50}, AP^{75}, AR^1, AR^{10}, AR^{100}. The descriptions of these metrics are summarized in Table 1, in which IoU (Intersection over Union), an evaluation metric in Everingham et al. (2007), is defined by Eq. (10). It represents the overlap ratio between predicted bounding box $\widehat{bb}$ and "ground-truth" bounding box bb.

$$IoU\left(\widehat{bb}, bb\right) = \frac{area\left(\widehat{bb} \cap bb\right)}{area\left(\widehat{bb} \cup bb\right)} \tag{10}$$

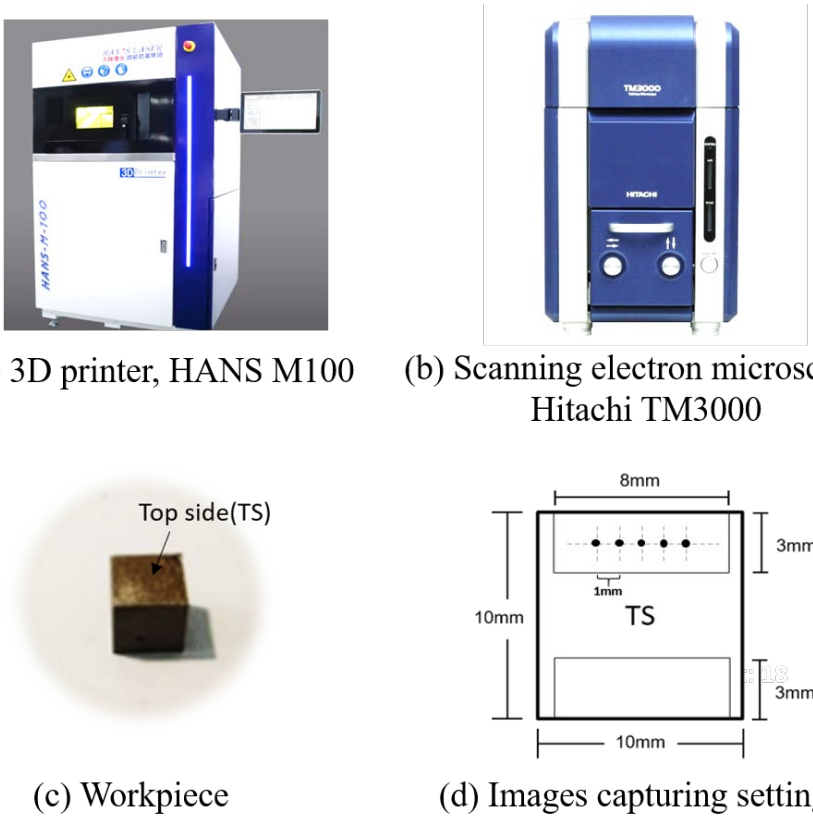
Table 1
The description of evaluation metrics

Metric	Description
AP	mAP averaged over ten IOUs: {0.5 : 0.05 : 0.95}
AP^{50}	mAP at IOU=0.5
AP^{75}	mAP at IOU=0.75
AR^1	AR given 1 detection per image
AR^{10}	AR given 10 detections per image
AR^{100}	AR given 100 detections per image

4 Experimental work and results

4.1 Dataset

A series of experiments was conducted to build a dataset for the selective laser melting technique of additive manufacturing. Eighteen components with the size 10 mm x 10 mm x10 mm were fabricated by additive manufacturing equipment, HANS M100 as shown in Fig. 2(a). The additive manufacturing conditions are listed in Table 2. The surface images were captured on the top side of components using a scanning electron microscope, i.e. Hitachi tabletop microscope TM3000 as shown in Fig. 2(b). The images on the top side of workpiece were acquired as shown in Fig. 2(c). Each image was obtained at the position shown in Fig. 2(d). In each workpiece, there were two image capturing zones with the size 3 mm x 8 mm, in which five points were examined and each point was separated by 1 mm. The ground-truth was tagged by Labelme (Wada et al., 2016), in which four types, namely crack, un-melted powder, porosity and collapse as well as the locations of surface defects were tagged. The profile of surface defects is shown in Fig. 3.



(a) 3D printer, HANS M100 (b) Scanning electron microscope, Hitachi TM3000

(c) Workpiece (d) Images capturing setting

Fig. 2. The experimental equipment and setup

Table 2
Additive manufacturing experimental conditions

Method	Selective laser melting
Material	316L stainless steel
Particle size	0-25 µm
Laser spot diameter	25 µm
Laser power	70 W
Scanning speed	700 mm/s

Hatch spacing	0.06 mm
Scan path	Raster Path
Layer thickness	20 μm

4.2 Implementation details

Input image resolution of 512×512 was fixed and the output down-sampling ratio k was set as 4. As a result, an output resolution of 128×128 was obtained. In order to improve the robustness of the defect characterization model, random scaling, random flip and cropping were used to implement data augmentation. The hyperparameters α, β in L_{foc} were set as 2 and 4, respectively, following Law and Deng (2018). Loss function was optimized by Adam (Kingma et al., 2014) and the initial learning rate was 1.25e-4 for 140 epochs, with the learning rate dropping 10 times at 90 and 120 epochs. $\lambda_{off}, \lambda_{size}$ to 1,0.1 were set, respectively, following Zhou et al. (2019) and λ_{cnt} was set as 0.01. The counting loss is the loss in the first stage of the Hourglass network. On the contrary, the rest of the losses were accumulated in two stages of the Hourglass network. When in the inference time, no augmentation was applied and the original resolution was kept for the images.

4.3 Results and discussion

Some output examples which give the surface defect type, position and count are illustrated in Fig. 3. Different types of defects are indicated by different colours and position is indicated by the bounding box, as shown in Fig. 4. In order to verify the performance of the proposed defect characterization model, a series of experiments was conducted on different backbone networks, such as Hourglass network (Hourglass-104), Deep Layer Aggregation (DLA-34) (Xiao et al., 2018), and ResNet (Res-101) (Yu et al., 2018). All backbone networks were tested under two conditions, namely with count loss (w/ CL) and without count loss (w/o CL). The results are listed in Table 4. On the whole, Hourglass-104 performed better on mAP and AR than DLA-34 and ResNet-101, and the best accuracy was obtained on AP, AP^{50}, AR^{1}, and AR^{10}, reaching 0.18, 0.48, 0.22, and 0.31, respectively, which implies that deeper semantic information benefits the defect detection task. For different backbone networks, count loss boosts the performance in most of the evaluation metrics. The Hourglass-104 backbone network with count loss (H-104-wCL) has a similar accuracy on AP while yielding better performance on AP, AP^{50} and AR-related evaluation metrics as compared with the results of the Hourglass-104 backbone network without count loss (H-104-woCL). For the DLA-34 backbone network, count loss yielded better detection accuracy but worse recall. The reason for the low recall is that the network is shallow and density map estimation has more influence on detection performance, thus losing specific location and defect fine detailed information. The results of ResNet-101 with a count loss is better in most of the evaluation metrics, including AP, AP^{75}, AR^{10}, and AR^{100}. As a result, count loss obtained from the additional density map output head had a positive effect on defect detection.

Table 3

Representation of defect detection

Items	Notation
Defect types	Dt

Un-melted powder	um
Porosity	pr
Collapse	cs
Crack	ck
Detection results	Dt^{score}

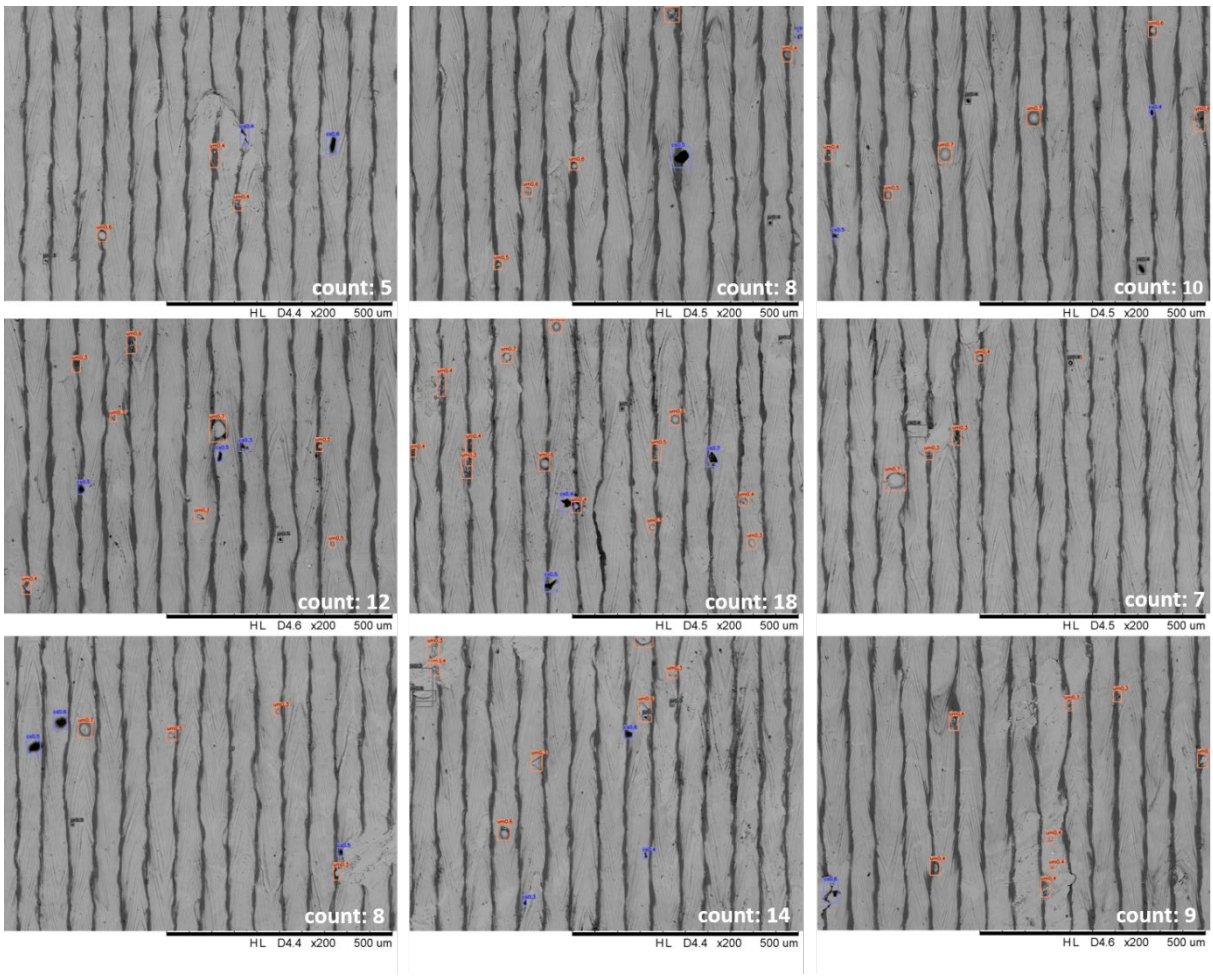


Fig. 3. Results of defect detection and characterization

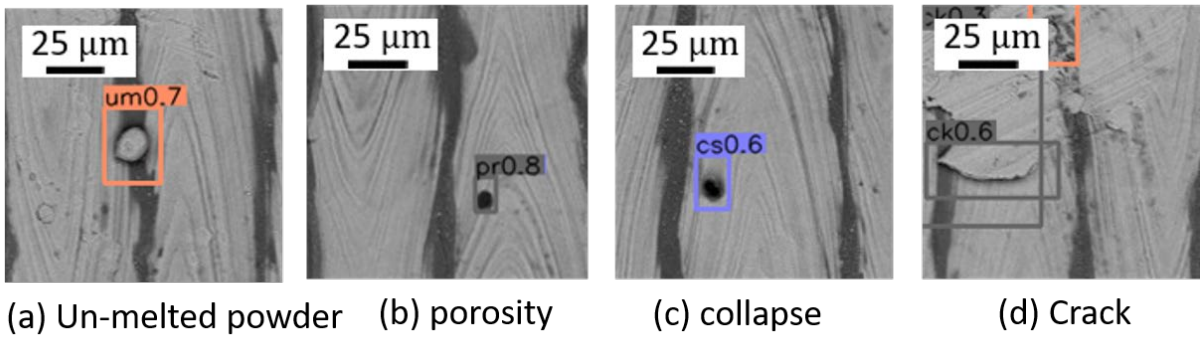


Fig. 4. The types of defects detected

Table 4

Comparison of two conditions on different backbone networks

	AP	AP^{50}	AP^{75}	AR^{1}	AR^{10}	AR^{100}
Hourglass-104 (w/ CL)	0.18	0.48	0.07	0.22	0.31	0.31
Hourglass-104 (w/o CL)	0.18	0.44	0.14	0.19	0.28	0.28
DLA-34 (w/ CL)	0.17	0.39	0.13	0.12	0.21	0.22
DLA-34 (w/o CL)	0.13	0.34	0.10	0.13	0.31	0.32
Res-101 (w/ CL)	0.07	0.20	0.03	0.05	0.11	0.11
Res-101 (w/o CL)	0.06	0.23	0.02	0.06	0.10	0.10

4.4 Ablation studies

Hyperparameters

To explore the hyperparameter setting of the proposed defect characterization model, some ablation studies were conducted. The performance in terms of different σ, was compared which was used in the generation of density map ground-truth. Table 5 clearly shows that AP, AP^{50} is the best when $\sigma = 4$. However, AP^{75} is the best when $\sigma = 8$. As a result, if the low requirement is used or overlap ratio, σ is then set to 4. When $\sigma = 2$, AR, AR^{10}, and AR^{100} have the best performance, which means setting σ to 2 if the aim is to find defects as fully as possible in images. In addition, we explored the influence of dilated convolution layers (DCL) in the density map output head. Table 6 illustrates that DCL has an adverse effect on the defect detection task. Furthermore, the experiments were conducted on H-104-wCL with different crop sizes. The results are listed in Table 7, which show that higher resolution of images boost the defect detection performance on all AP and AR evaluation metrics. Besides, we analysed the sensitivity of our model to the loss weight λ_{cnt}. In Table 8, 0.01 exhibits the best detection performance. For smaller λ_{cnt}, the AP and AR both decrease dramatically. However, the value drops only slightly for larger λ_{cnt}.

Table 5

Influence of σ on the Hourglass network

σ	AP	AP^{50}	AP^{75}	AR^{1}	AR^{10}	AR^{100}
2	0.179	0.483	0.069	0.217	0.305	0.305
4	0.187	0.500	0.099	0.174	0.270	0.271
8	0.149	0.350	0.117	0.102	0.287	0.291

Table 6

Influence of dilated convolutional layer on the density map head of the Hourglass network

	AP	AP^{50}	AP^{75}	AR^{1}	AR^{10}	AR^{100}
w/o DCL	0.187	0.500	0.099	0.174	0.270	0.271

w/ DCL	0.094	0.274	0.005	0.083	0.274	0.276

Table 7
Influence of input image crop size on the Hourglass network

Crop size	AP	AP^{50}	AP^{75}	AR^{1}	AR^{10}	AR^{100}
512*512	0.179	0.483	0.069	0.217	0.305	0.305
384*384	0.070	0.197	0.035	0.081	0.121	0.125

Table 8
Influence of λ_{cnt} on the Hourglass network

λ_{cnt}	AP	AP^{50}	AP^{75}	AR^{1}	AR^{10}	AR^{100}
0.1	0.119	0.297	0.070	0.080	0.216	0.216
0.01	0.187	0.500	0.099	0.174	0.270	0.271
0.005	0.061	0.203	0.015	0.044	0.149	0.150

Dataset

Except studying the influence of hyperparameters of models, we also pay attention to the detection performance of models on other kinds of dataset. Therefore, we built a new dataset, containing the surface images of CoCr workpieces. The geometry of defects on CoCr workpieces are different from those on 316L workpieces, as shown in Fig. 5. Furthermore, we could see that this model could detect almost all defects on the surface of CoCr components, which means that our model has good detection performance on similar defect datasets.

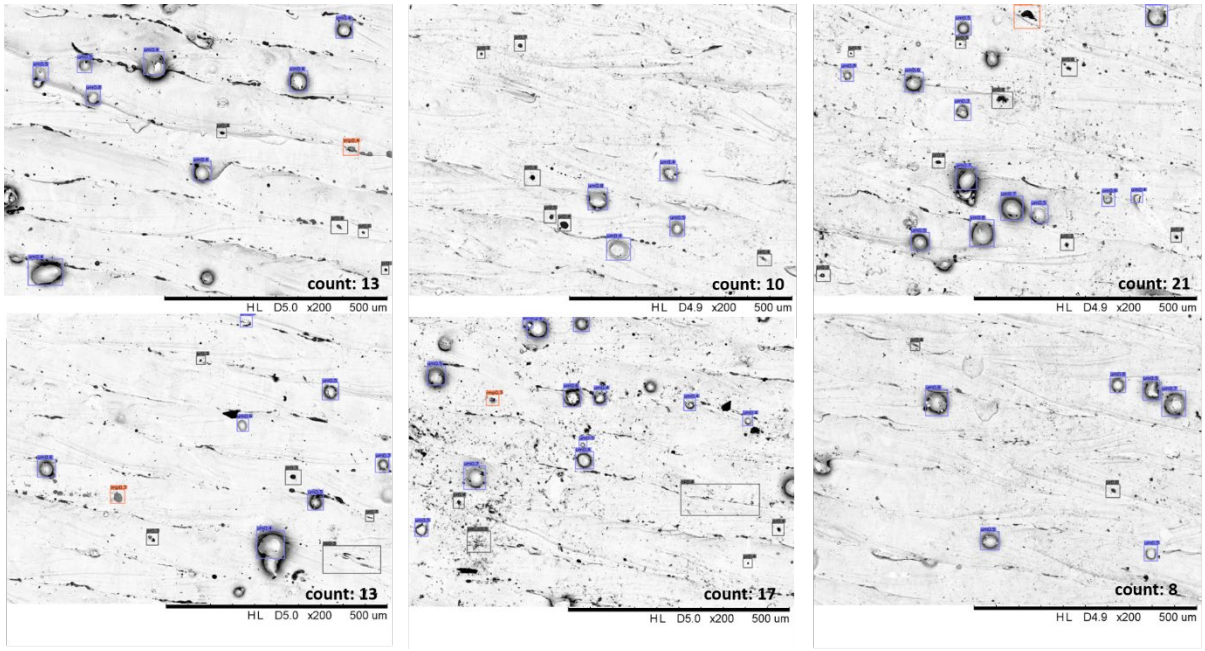


Fig. 5. The results of workpieces made of CoCr

Models

To verify the superiority of our model, CenterNet-CL, we compared the detection

performance with other state-of-the-art models. All experiments were run on the MMDetection toolbox (Chen et al., 2019) and the experimental setup is default. We compared our model with one-stage detection models, such as SSD512 (Liu et al., 2016), YOLOv3 (Redmon & Farhadi, 2018), two-stage models, such as Faster R-CNN (Ren et al., 2015), Cascade_R-CNN (Cai & Vasconcelos, 2018), Cascade RPN (Vu et al., 2019) and anchor-free models, such as CornerNet (Law & Deng, 2018), FCOS (Tian et al., 2019), RepPoints (Yang et al., 2019), and CenterNet. The results in Table 9 show that our model achieves better detection accuracy.

Table 9
Detection performance comparison with state-of-the-art models

Model	Backbone	AP^{50}
Faster R-CNN (Ren et al., 2015)	ResNeXt-101	0.10
SSD512 (Liu et al., 2016)	VGG16	0.06
YOLOv3 (Redmon & Farhadi, 2018)	DarkNet-53	0.10
Cascade_R-CNN (Cai & Vasconcelos, 2018)	ResNet-50	0.42
CornerNet (Law & Deng, 2018)	Hourglass-104	0.23
Cascade RPN (Vu et al., 2019)	ResNet-50	0.29
FCOS (Tian et al., 2019)	ResNet-50	0.27
RepPoints (Yang et al., 2019)	ResNeXt-101-DCNv2	0.31
CenterNet (Zhou et al., 2019)	Hourglass-104	0.44
PAA (Kim & Lee, 2020)	ResNet-50	0.35
SABL (Wang et al., 2020)	ResNet-101	0.21
GFL (Li et al., 2020)	ResNeXt-101-DCNv2	0.17
CenterNet-CL (ours)	Hourglass-104	0.48

5 Conclusion

In this paper, surface defects on addictive manufactured 316L workpieces were firstly classified, localized and counted, simultaneously. A novel convolutional neural network, CenterNet-CL, was built based on CenterNet and an additional prediction output head responsible for density map estimation and count loss was added to improve the detection performance. A dataset was built, containing surface defect images captured by scanning electron microscope and the ground-truth labelled by hand. A series of experiments was conducted to validate the model performance in terms of different aspects, including backbone network, σ, dilated convolutional layer, crop size and different datasets. Compared with existing state-of-the-art models, including one-stage, two-stage and anchor-free models, the proposed model, CenterNet-CL, has better detection performance. Although this method performed well, it is just for special domains. In the future, we will expand it into more general situations, such as defects of freeform surface, different additive manufacturing parameters or various materials. In addition, we will build a larger dataset to enrich the diversity of defects and improve detection accuracy.

Acknowledgements

The work described in this paper was mainly supported by a grant from the Research Grants

Council (Project No. 15202717) and Innovation and Technology Commission (ITC) (Project No.: ITS/076/18FP) of the Government of the Hong Kong Special Administrative Region, China. The authors would also like to express their sincere thanks to the Research Committee of The Hong Kong Polytechnic University for their financial support of the project through a PhD studentship (project account code: RK36).

References

Caggiano, A., Zhang, J., Alfieri, V., Caiazzo, F., Gao, R., & Teti, R. (2019). Machine learning-based image processing for on-line defect recognition in additive manufacturing. *CIRP Annals, 68*(1), 451-454.

Cai, Z., & Vasconcelos, N. (2018). *Cascade r-cnn: Delving into high quality object detection.* In Proceedings of the IEEE conference on computer vision and pattern recognition.

Cebollada, S., Payá, L., Flores, M., Peidró, A., & Reinoso, O. (2020). A state-of-the-art review on mobile robotics tasks using artificial intelligence and visual data. *Expert Systems with Applications*, 114195.

Chen, K., Wang, J., Pang, J., Cao, Y., Xiong, Y., Li, X., . . . Xu, J. (2019). MMDetection: Open mmlab detection toolbox and benchmark. *arXiv preprint arXiv:1906.07155*.

Dai, J., Li, Y., He, K., & Sun, J. (2016). R-fcn: Object detection via region-based fully convolutional networks. *arXiv preprint arXiv:1605.06409*.

Deng, J., Dong, W., Socher, R., Li, L.-J., Li, K., & Fei-Fei, L. (2009). *Imagenet: A large-scale hierarchical image database.* In Proceedings of the 2009 IEEE conference on computer vision and pattern recognition.

Everingham, M., Van Gool, L., Williams, C. K., Winn, J., & Zisserman, A. (2007). The PASCAL visual object classes challenge 2007 (VOC2007) results.

Girshick, R. (2015). *Fast r-cnn.* In Proceedings of the IEEE international conference on computer vision.

Girshick, R., Donahue, J., Darrell, T., & Malik, J. (2014). *Rich feature hierarchies for accurate object detection and semantic segmentation.* In Proceedings of the IEEE conference on computer vision and pattern recognition.

Gobert, C., Reutzel, E. W., Petrich, J., Nassar, A. R., & Phoha, S. (2018). Application of supervised machine learning for defect detection during metallic powder bed fusion additive manufacturing using high resolution imaging. *Additive Manufacturing*, *21*, 517-528.

He, K., Zhang, X., Ren, S., & Sun, J. (2015). Spatial pyramid pooling in deep convolutional networks for visual recognition. *IEEE transactions on pattern analysis and machine intelligence, 37*(9), 1904-1916.

Kim, K., & Lee, H. S. (2020). Probabilistic anchor assignment with iou prediction for object detection. *arXiv preprint arXiv:2007.08103*.

Khanzadeh, M., Chowdhury, S., Marufuzzaman, M., Tschopp, M. A., & Bian, L. (2018). Porosity prediction: Supervised-learning of thermal history for direct laser deposition. *Journal of manufacturing systems*, *47*, 69-82.

Kingma, D. P., & Ba, J. (2014). Adam: A method for stochastic optimization. *arXiv preprint arXiv:1412.6980*.

Krizhevsky, A., Sutskever, I., & Hinton, G. E. (2012). Imagenet classification with deep convolutional neural networks. *Advances in neural information processing systems, 25*, 1097-1105.

Law, H., & Deng, J. (2018). *Cornernet: Detecting objects as paired keypoints.* In Proceedings of the European conference on computer vision (ECCV).

Li, X., Wang, W., Wu, L., Chen, S., Hu, X., Li, J., . . . Yang, J. (2020). Generalized focal loss: Learning qualified and distributed bounding boxes for dense object detection. *arXiv preprint arXiv:2006.04388*.

Lin, T.-Y., Goyal, P., Girshick, R., He, K., & Dollár, P. (2017). *Focal loss for dense object detection.* In Proceedings

of the IEEE international conference on computer vision.

Lin, T.-Y., Maire, M., Belongie, S., Hays, J., Perona, P., Ramanan, D., . . . Zitnick, C. L. (2014). *Microsoft coco: Common objects in context.* In Proceedings of the European conference on computer vision.

Liong, S. T., Gan, Y. S., Huang, Y. C., Yuan, C. A., & Chang, H. C. (2019). Automatic defect segmentation on leather with deep learning. *arXiv preprint arXiv:1903.12139*.

Liu, L., Ouyang, W., Wang, X., Fieguth, P., Chen, J., Liu, X., & Pietikäinen, M. (2020). Deep learning for generic object detection: A survey. *International journal of computer vision, 128*(2), 261-318.

Liu, W., Anguelov, D., Erhan, D., Szegedy, C., Reed, S., Fu, C.-Y., & Berg, A. C. (2016). *Ssd: Single shot multibox detector.* In Proceedings of the European conference on computer vision.

Ma, Z., Wei, X., Hong, X., & Gong, Y. (2019). *Bayesian loss for crowd count estimation with point supervision.* In Proceedings of the IEEE/CVF International Conference on Computer Vision.

Mabrouk, A. B., & Zagrouba, E. (2018). Abnormal behavior recognition for intelligent video surveillance systems: A review. *Expert Systems with Applications*, *91*, 480-491.

Malekipour, E., & El-Mounayri, H. (2018). Common defects and contributing parameters in powder bed fusion AM process and their classification for online monitoring and control: a review. *The International Journal of Advanced Manufacturing Technology, 95*(1), 527-550.

Nagarajan, B., Hu, Z., Song, X., Zhai, W., & Wei, J. (2019). Development of micro selective laser melting: The state of the art and future perspectives. *Engineering, 5*(4), 702-720.

Newell, A., Yang, K., & Deng, J. (2016). *Stacked hourglass networks for human pose estimation.* In Proceedings of the European conference on computer vision.

Redmon, J., Divvala, S., Girshick, R., & Farhadi, A. (2016). *You only look once: Unified, real-time object detection.* In Proceedings of the IEEE conference on computer vision and pattern recognition.

Redmon, J., & Farhadi, A. (2017). *YOLO9000: better, faster, stronger.* In Proceedings of the IEEE conference on computer vision and pattern recognition.

Redmon, J., & Farhadi, A. (2018). Yolov3: An incremental improvement. *arXiv preprint arXiv:1804.02767*.

Ren, S., He, K., Girshick, R., & Sun, J. (2015). Faster r-cnn: Towards real-time object detection with region proposal networks. *arXiv preprint arXiv:1506.01497*.

Roldán, J., Boubeta-Puig, J., Martínez, J. L., & Ortiz, G. (2020). Integrating complex event processing and machine learning: An intelligent architecture for detecting IoT security attacks. *Expert Systems with Applications, 149*, 113251.

Scime, L., & Beuth, J. (2018). Anomaly detection and classification in a laser powder bed additive manufacturing process using a trained computer vision algorithm. *Additive Manufacturing, 19*, 114-126.

Sermanet, P., Eigen, D., Zhang, X., Mathieu, M., Fergus, R., & LeCun, Y. (2013). Overfeat: Integrated recognition, localization and detection using convolutional networks. *arXiv preprint arXiv:1312.6229*.

Simonyan, K., & Zisserman, A. (2014). Very deep convolutional networks for large-scale image recognition. *arXiv preprint arXiv:1409.1556*.

Song, L., Lin, W., Yang, Y. G., Zhu, X., Guo, Q., & Xi, J. (2019). Weak micro-scratch detection based on deep convolutional neural network. *IEEE Access, 7,* 27547-27554.

Soyguder, S. (2011). Intelligent control based on wavelet decomposition and neural network for predicting of human trajectories with a novel vision-based robotic. *Expert Systems with Applications, 38*(11), 13994-14000.

Tian, Z., Shen, C., Chen, H., & He, T. (2019). *Fcos: Fully convolutional one-stage object detection.* In Proceedings of the IEEE/CVF International Conference on Computer Vision.

Tripathi, R. K., & Jalal, A. S. (2021). Novel local feature extraction for age invariant face recognition. *Expert*

Systems with Applications, 175, 114786.

Vu, T., Jang, H., Pham, T. X., & Yoo, C. D. (2019). Cascade rpn: Delving into high-quality region proposal network with adaptive convolution. *arXiv preprint arXiv:1909.06720*.

Wada, K. (2016). Labelme: Image polygonal annotation with python: GitHub repository

Wang, Y., Liu, M., Zheng, P., Yang, H., & Zou, J. (2020). A smart surface inspection system using faster R-CNN in cloud-edge computing environment. *Advanced Engineering Informatics, 43,* 101037.

Wang, J., Zhang, W., Cao, Y., Chen, K., Pang, J., Gong, T., . . . Lin, D. (2020). *Side-aware boundary localization for more precise object detection.* In Proceedings of the the European Conference on Computer Vision.

Wong, V. W. H., Ferguson, M., Law, K. H., Lee, Y. T. T., & Witherell, P. (2021). Automatic volumetric segmentation of additive manufacturing defects with 3D U-Net. *arXiv preprint arXiv:2101.08993.*

Wu, X., Sahoo, D., & Hoi, S. C. (2020). Recent advances in deep learning for object detection. *Neurocomputing, 396*, 39-64.

Xiao, B., Wu, H., & Wei, Y. (2018). *Simple baselines for human pose estimation and tracking.* In Proceedings of the European conference on computer vision (ECCV).

Yang, Z., Liu, S., Hu, H., Wang, L., & Lin, S. (2019). *Reppoints: Point set representation for object detection.* In Proceedings of the IEEE/CVF International Conference on Computer Vision.

Yang, Y., Pan, L., Ma, J., Yang, R., Zhu, Y., Yang, Y., & Zhang, L. (2020). A high-performance deep learning algorithm for the automated optical inspection of laser welding. *Applied Sciences, 10*(3), 933.

Yap, C. Y., Chua, C. K., Dong, Z. L., Liu, Z. H., Zhang, D. Q., Loh, L. E., & Sing, S. L. (2015). Review of selective laser melting: Materials and applications. *Applied physics reviews, 2*(4), 041101.

Yu, F., Wang, D., Shelhamer, E., & Darrell, T. (2018). *Deep layer aggregation.* In Proceedings of the IEEE conference on computer vision and pattern recognition.

Zhang, Y., Hong, G. S., Ye, D., Zhu, K., & Fuh, J. Y. (2018). Extraction and evaluation of melt pool, plume and spatter information for powder-bed fusion AM process monitoring. *Materials & Design, 156*, 458-469.

Zhou, X., Wang, D., & Krähenbühl, P. (2019). Objects as points. *arXiv preprint arXiv:1904.07850*.