Gümüşhane University Journal of Science

GUFBD / *GUJS* (2024) 14(2): 538-549 doi: 10.17714/gumusfenbil.1391054 Research Article

Small object detection method with k-detector for metal parts surface defect detection

Metal parçaların yüzey kusurlarını tespit için k-dedektörü ile küçük nesne tespit yöntemi

Yavuz Selim BALCIOĞLU*

Gebze Teknik Üniversitesi, İşletme Fakültesi, Yönetim Bilişim Sistemleri Bölümü, 41400, Kocaeli

• Received: 15.11.2023 • Accepted: 07.02.2024

Abstract

In the process of its development, intelligent manufacturing often focuses on production flexibility, client customization, and quality control, all of which are required for the manufacture of powder-based metallurgy. In particular, the identification and categorization of defects are crucial steps in the production processes involved in powder-based metallurgy. Intelligent strategies to detect faults in metal parts continue to be a challenge in automated industrial production lines. These techniques have been a particular concern for microscopic metal component producers for a long time. Due to its precision and speed, the YOLOv4 approach has been widely used for object detection. On the other hand, the identification of tiny targets, particularly imperfections on the surface of metal parts, continues to present a number of obstacles and difficulties. To increase the overall performance of detection, this research provided a technique for the detection of tiny objects based on YOLOv4 for such objects. To increase the effectiveness of the detection process, this involves expanding the size of the k detector while simultaneously eliminating unnecessary branches of the YOLO head network. Experiments have shown that the KD-YOLO model performs better than its predecessors, YOLOv4, YOLOv5, and PP-YOLO, in terms of the total number of parameters, classification accuracy and detection precision.

Keywords: Object detection, Surface detection, Quality control

Öz

Gelişim sürecinde, akıllı imalat genellikle üretim esnekliğine, müşteriye özel üretime ve kalite kontrolüne odaklanmaktadır, bunların hepsi toz bazlı metalurji üretimi için gereklidir. Özellikle, toz bazlı metalurjide üretim süreçlerinde hataların tespiti ve kategorize edilmesi kritik adımlardır. Metal parçalardaki hataları tespit etmek için akıllı stratejiler, otomatik endüstriyel üretim hatlarında hala bir meydan okumadır. Bu teknikler, özellikle mikroskopik metal bileşen üreticileri için uzun zamandır özel bir endişe kaynağı olmuştur. Hassasiyeti ve hızı nedeniyle, YOLOv4 yaklaşımı nesne tespiti amacıyla yaygın olarak kullanılmıştır. Öte yandan, özellikle metal parçaların yüzeyindeki kusurlar gibi küçük hedeflerin tanımlanması, birçok engel ve zorluk sunmaya devam etmektedir. Bu araştırma, tespit performansını genel olarak artırmak için, YOLOv4'e dayalı küçük nesnelerin tespiti için bir teknik sunmaktadır. Tespit sürecinin etkinliğini artırmak için, bu, YOLO baş ağının gereksiz dallarının kaldırılması ile k detektörünün boyutunun genişletilmesini içermektedir. Deneyler, KD-YOLO modelinin toplam parametre sayısı, sınıflandırma doğruluğu ve tespit hassasiyeti açısından önceki modelleri YOLOv4, YOLOv5 ve PP-YOLO'dan daha iyi performans gösterdiğini göstermiştir.

Anahtar kelimeler*: Obje tanıma, Yüzey tespiti, Kalite kontrol*

*Yavuz Selim BALCIOĞLU; ysbalcioglu@gtu.edu.tr

ISSN: 2146-538X e-ISSN: 2146-538X https://dergipark.org.tr/tr/pub/gumusfenbil

1. Introduction

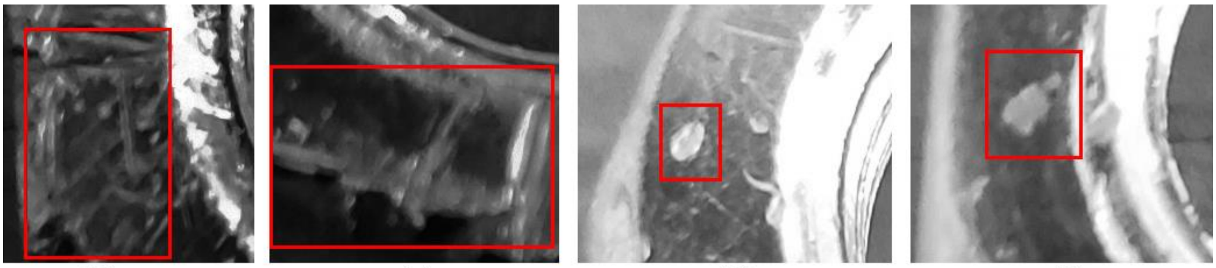
Powder metallurgy (Panda et al., 2018) is highly dependent on the quality of the deliverables, customer customization, and strict quality control, all of which have received a lot of attention in the development of intelligent manufacturing. Detection and categorization of defects are crucial for production operations in the powder metallurgy sector. Detection of defects is essential for effective product quality management (Silvius & Schipper, 2014). The traditional detection technique is time-consuming (Moreira Monteiro et al., 2021), and poor detection efficiency and a large proportion of missed inspections could be the end consequences of using manual detection for a prolonged period of time. Moreover, most broken metal parts are either recycled or fixed, and those that cannot be fixed are usually thrown away.

Defective metal components must be sorted according to the kind of defect before they can be processed efficiently. It is a gold mine of data to spot manufacturing flaws. In spite of this, the vast majority of technologies for detection still rely on human intervention (Smıetańska & Podzıewskı, 2019). Because of this, powder-based metallurgical processing, manufacturers have always had a strong incentive to come up with better ways to find flaws in metal products as they move along an automated industrial production line.

When metal components using powder metallurgy, it is possible for a number of different types of flaws to manifest in the final product, including interfacial tension and faulty solder joints. See Fig. 1 for a breakdown of the types of edge defects (big and tiny) that are the subject of this investigation.

Due to its mix of speed, accuracy, and minimal resource usage, You Only Look Once v4 (also known as YOLOv4) has been a popular alternative for object identification applications (Parico & Ahamed, 2021) ever since it was first made available. Identification of tiny targets, such as surface flaws on metal components, remains challenging and complex. To enhance the detection performance, due to the findings of this study, a method for the identification of small objects called KD-YOLO that is based on YOLOv4 was proposed. See Fig. 2 for an illustration of the KD-YOLO tiny object identification result.

The study's contribution to the field is that the YOLOv4 network model served as the basis for the KD-YOLO architecture that was suggested. However, four adjustments were made to accommodate especially the detection of tiny objects.



(a) (b) (c) (d)

Figure 1. The large region missing (a) - (b) and the little area missing (c) – (d) are both common surface flaws in metal nuts.

To begin, we present a novel feature Graph Convolutional Neural Network (GCNN) with a k approach to satisfy tiny item recognition and further improve the network's prediction accuracy without slowing it down. We altered the features of the b2gray module and performed network pruning to raise the bar for tiny object recognition precision. The KD-YOLO module, which makes use of graph convolution neural networks, replaces the traditional feature color with a grayscale representation (GCNN). Deep convolutional neural networks, also known as DCNNs (Atwood & Towsley, 2016), have the ability to learn hierarchical features in several layers, allowing them to collect input from objects of varied sizes. Specifically, the spatially rich characteristics that are found in shallow layers have higher resolution, making them more helpful for finding small things. Unfortunately, many of the fine features of tiny objects may be lost since GCNN and Feature Pyramid Network (FPN) use the deconvolution layer for the top-most feature maps. Furthermore, systems based on the characteristics of the dimension of the color perform connections for each prediction layer, which indicates that the greater the number of layers, the greater the computational cost.

Figure 2. KD-YOLO flaw detection findings on a sample metal nut picture.

When the number of colors used to represent an item is increased, more details of that item are kept in memory. As a result, reliable detection of tiny objects using simply the highest-level layers is challenging. The suggested feature k approach is more suited for detecting small objects.

Next, an adaptively optimized number and size of anchor boxes are achieved by training using the k-means clustering algorithm. Defect detection performance for tiny objects may be enhanced by using a priori boxes scaled to the size of the sample. Next, the previous anchor boxes are put to use in the detection process, which may make the prediction scale more flexible.

Lastly, in order to reduce the number of model parameters and computations required for the detection of very small objects, the duplicated YOLO head network connections have been removed, with the large-scale feature layer being the only exception. Typical detection techniques are not suitable for detecting tiny objects since the average defect area is just 0.5 mm2, while metal components are typically 10 mm2 in size. Thus, the whole network is pruned, except for the top-level characteristics.

The main contributions of our study can be summarized as follows:

Novel Model Architecture: The study presents the KD-YOLO model which modifies the standard YOLOv4 architecture. The key distinction lies in the Path Aggregation Network (PANet) architecture where shallow feature information is preserved to improve the detection of very small objects.

- K-Method Feature: We introduce the "k-method" to enhance the detection of small objects. By integrating various types of feature information, both deep semantically rich features and shallow geographically rich features are utilized for superior detection performance.
- K-Means++ for Preliminary Anchor Estimation: Our study uses the k-means++ clustering method to estimate prior anchor boxes, improving the model's ability to determine bounding boxes.
- Enhanced Data Augmentation: We utilize the mosaic data augmentation method, based on the cut-mix theory, to increase the diversity and volume of their training data, enhancing model performance.
- Evaluation and Comparative Analysis: We evaluate our KD-YOLO model on a dataset of images from a domestic manufacturing facility and compare its performance with the original YOLOv4, YOLOv5, and PP-YOLO models. The KD-YOLO model shows superior performance in our experiments.
- Application to Real-world Problems: The KD-YOLO model is applied for the detection of flaws in metal components, demonstrating the practical utility of their research.

2. Related work

Recently, machine learning techniques have seen widespread use in the field of surface defect identification and quality assurance. For instance, by using an algorithm from the Region-Based Convolutional Neural Network (RCNN), family (R-CNN (Chu et al., 2021), Fast R-CNN (MathWorks.com, 2019), Faster R-CNN (Lin et al., 2017), and Mask R-CNN (He et al., 2019), the issue of defect detection may become a two-stage object detection problem. This can be accomplished by employing the R-CNN method. This is accomplished by using a detector based on the method. Compared to several other algorithms, it has superior detection accuracy. However, compared to a single-stage detector like the single-shot multibox detector (SSD) or YOLO, this method requires more computer processing time. The YOLOv4 method is especially useful for identifying objects because it works quickly and makes good use of parallel computations.

Recently, more error-free detectors, such as Fast R-CNN (Girshick, 2015) and Faster R-CNN. In particular, the Spatial Pyramid Pooling Net (SPPNet) (He et al., 2015) is the source of inspiration for the Faster R-CNN (Cao et al., 2019), which is responsible for introducing the Region Proposal Network (RPN) (Cao et al., 2019). A fully convolutional network with the ability to predict object limitations in addition to object scores at each location simultaneously is referred to as an RPN. However, because their receptive fields are fixed, the top-most feature maps are incompatible with objects in pictures of varying sizes. Because there is not much information left on the topmost characteristics, particularly for smaller items, it is difficult to utilize this information in our investigation regarding the identification of surface flaws in metal components. Liao et al. (2021) propose an end-to-end detector based on cutting-edge deep learning, YOLOv4, to automatically detect surface defects in printed circuit boards. Through the evaluation of the detector based on a custom dataset, the research demonstrates that the proposed detector has a higher performance in terms of mean average precision (mAP) and detection speed compared to similar research work. They improve on the original YOLOv4 by including some new approaches that improve how surface patches are accumulated and how defects are searched. Chen et al. (2020) propose DeepNDEC, an YOLO-V4 backbone network enhanced with focal loss that promotes a balance in accuracy between positive and negative samples, overcoming issues of imbalance. They designed and implemented a target detection algorithm for electrical power equipment that can successfully classify images of one of 16 objects, including oil leakage, electric equipment, vehicles, rocks, and others. Liu et al. (2021) propose a surface defect detection algorithm based on YOLOv4, which has both improved recognition accuracy and computational efficiency. They pushed the mean average precision for steel strip surface defect detection to 85.41%, from 84.64% for the previous state-of-the-art, in the context of comparing against edge point clouds. With a mean average precision of 84.64%, YOLOv4 outperforms the original YOLOv3.

Deep learning and machine learning are two examples of automated inspection approaches that are beginning to replace human inspectors in the area of industrial product inspection. Defect detection in manufactured goods is increasingly striving towards real-time detection. Although it works better at identifying industrial defects, deep learning still has to deal with problems such as incomplete defect data sets, short samples, and small goals.

3. Proof of theorem for k-methods

The following assertion demonstrates the existence of a valid dual quadratic under the condition that the components in the support are suitably spaced apart, and as a direct result of this, the theorem of the K method follows as a direct consequence.

This notion will be demonstrated throughout the rest of this section. Our approach involves first interpolating v on T using a low-frequency kernel and then correcting the interpolation to guarantee that the derivative of the dual quadratic is zero on T. This is how we get our results. The essential component that we use is;

$$K(t) = \left[\frac{\sin\left(\left(\frac{fc}{2}+1\right)\pi t\right)}{\left(\frac{fc}{2}+1\right)\sin(\pi t)}\right], 0 < k < 1 \tag{1}$$

And k(0) = 1. If fc is an even number, then K(t) is the square of the Fejer kernel, which is a trigonometric polynomial with frequencies that satisfy the condition that |k| fc/2. As a direct result of this, K takes the form.

If the reader is paying close attention, they could notice that the selection of the interpolation kernel seems to be quite random. In the point of fact, one may also utilize the Fej'er kernel or any other power of the kernel by using very comparable approaches for the proof. We have discovered that the second equation gives a decent constant, since it strikes a reasonable balance between the transfer between localization in temporal and in frequency. This makes it a useful constant.

In order to build the dual quadratic, we first interpolate v with K and then with K', which is K's derivative. When the following q functions are investigated (2, 3, 4), the boxing expressions of the image mappings are eliminated at the location where the k value is discovered.

$$q(t) = \sum_{tj\epsilon T} \sigma jK(t-tj) + \beta jK(t-tj), \tag{2}$$

$$q(tk) = \sum_{tj\epsilon T} \sigma jK(tk-tj) + \beta jK(tk-tj) = vk, \ \ \forall tk \in T, \tag{3}$$

$$q(tk) = \sum_{tk\epsilon T} \sigma jK(tk-tj) + \beta jK(tk-tj) = 0, \ \ \forall tk \in T. \tag{4}$$

Because of this, as we shall see in the next section, it follows that the magnitude of q achieves a local maximum at these sites, which, in turn, is used to demonstrate that the hypothesis is valid.

The demonstration of this proposition relies on these equations, each of which will be the topic of discussion in the subsequent section. The first one demonstrates that it is feasible to interpolate any sign pattern perfectly, provided that the support is spread out and distributed in a certain way.

The proofs of the formulas discussed above make extensive use of the fact that the interpolation kernel and its derivatives decay at an increasingly fast rate as one moves further away from the origin. The intermediate result, which may be seen below (5, 6), demonstrated that. The distribution range of the spatial states of the picture files may be stated by looking at equation 7, which can be found here.

$$|k(\tau)(t)| \leq Be(t) = \left\{Bl(t) = \frac{\pi^l He(t)}{(fc+2)t^4}\right\} \frac{1}{2}\tau c \leq k \leq \sqrt{\frac{2}{\pi}}, \sqrt{\frac{2}{\pi} \leq k \leq \frac{1}{2}}, \tag{5}$$

Where $K_0 = 1$, $K_1 = 3$, $K_2 = 17$, $K_3 = 76$,

$$K_0(t) = \sigma^4(t), K_1(t) = \sigma^4(t)\big(2+2b(t)\big), K_2(t) = \sigma^4(t)\big(4+7b(t)+6b^2(t)\big), K_3(t) = \sigma^4(t)\big(8+24b(t)+30b^2(t)+16b^3(t)\big), \tag{6}$$

And

$$a(t) = \frac{2}{\pi(1-\frac{\pi^2 t2}{6})}, b(t) = \frac{1}{fc}\frac{a(t)}{t}. \tag{7}$$

First, we take into account the sum of all positive ti that are less than K, and designate by t+ the positive element in K that is closest to 0. The following equation (8) is an illustration of the expression of the input values over the sampled model of the K value.

$$\sum_{ti\in K:0<ti\leq 1/2}\left|K^{(l)}(t-ti)\right| = \left|K^{(l)}(t-ti)\right| + \sum_{ti\in K\setminus\{ti\}:0<ti\leq 1/2}\left|K^{(t)}(t-ti)\right|. \tag{8}$$

Let us suppose that K+<3Δ minf_0 〚(if K+>3Δmin〛 is less than two dimensions, (if K+ is more than two dimensions, the logic is essentially similar). If fc≥128 to24Δ minf_0 〚<0.33< √(2/π)〛. When we talk about the minimal separation requirement equation 9, we indicate that the second term on the right-hand side is the maximum allowed.

$$\sum_{j=3}^{24} Bl(j\Delta min - k) + \frac{\pi l}{(fc+3)^{24-l}} \sum_{j=24}^{\infty} \frac{Kl}{(j\Delta min+t)}, \tag{9}$$

which might have an upper limit established due to the fact that calculations were made to determine the maximum and lowest ranges of the K value (10, 11) over the picture being targeted.

$$\sum_{j=24}^{\infty} \frac{K_l^{\infty}}{(j\Delta min+t)} \leq \sum_{j=23}^{\infty} \frac{K_l^{\infty}}{(j\Delta \min \quad)} = \frac{K_l^{\infty}}{\Delta min}\left(\sum_{j=1}^{\infty}\frac{1}{j} - \sum_{j=1}^{23}\frac{1}{j}\right) = \frac{K_l^{\infty}}{\Delta max}\left(\frac{\pi}{99} - \sum_{j=1}^{22}\frac{1}{j}\right) \tag{10}$$

$$\left|K^{(l)}(t-ti)\right| \leq \{(min\Delta \leq k \leq 3\Delta \min.)(max\Delta \leq k \leq 3\Delta \max.)\}, k \leq 3\Delta min, k \leq 3\Delta max, \tag{11}$$

4. Methods

The backbone, neck, and head networks constitute the standard architecture for object detection. The backbone network is utilized for the extraction from an image. Common examples are the VGG Net, ResNet (McNeely-White et al., 2020) and Inception Net (Punn & Agarwal, 2020). The neck network, which may be either a Feature Pyramid Network (FPN) (Wang & Zhong, 2021), a Path Aggregation Network (PANet) (Liu et al., 2018), or a Bidirectional Feature Pyramid Network (BiFPN) (Cao et al., 2021), has seen extensive application for combining features from several layers through either a bottom-up or a top-down route. The head network may be utilized for prediction once data are gathered and processed from the neck and backbone networks. One-stage object detectors like YOLO and SSD, and four-stage detectors like the R-CNN series are common ways to classify the head network.

The KD-YOLO network architecture (shown in Fig. 3) consists of four modules: the input model; the backbone, a modified PANet; and a detector network. The PANet architecture is the primary distinction between YOLOv4 and KD-YOLO. The KD-PANnet YOLO's were modified so that shallow feature information could be saved. This was done because the lack of shallow feature information could cause the identification of very small objects to be inaccurate. There are two phases to this adjustment (see Fig. 3). YOLO's network is responsible for feature extraction from incoming pictures first. Second, a convolutional neural network is used to classify the items inside each of the image's regressed bounding boxes. It is suggested to combine PANet with KD-YOLO in order to improve the accuracy and speed of detecting metal component faults. Finally, flaws in metal components may be found with the help of the detection network. The FPN findings are used as the foundation for the construction of a bottom network, which is then used to supplement the shallow location information contained in FPN discoveries. Because of this, PANet is now a k-method backbone network that operates both top-down and bottom-up.

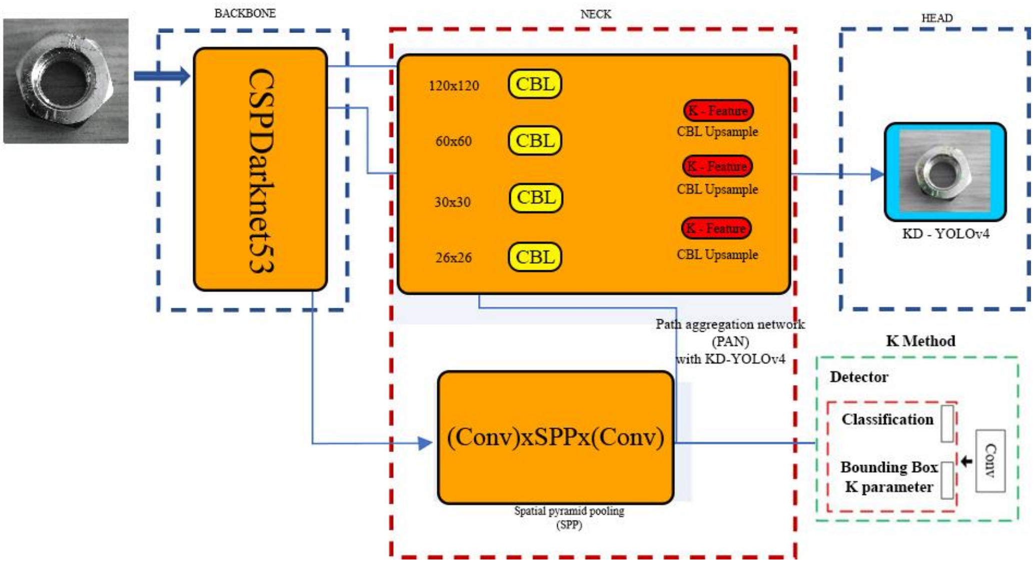


Figure 3. The suggested architecture of KD-YOLO

The k method's feature for tiny items: handling feature size concerns is critical for detecting tiny objects. In particular, semantically rich features have the potential to enhance shallow spatially rich features. It is possible to obtain both deep semantically rich features and shallow geographically rich features via the integration of many types of feature information, which is a trait that the k-method has. As a result, the form of k is crucial for this investigation. The sample study of detecting defective parts in the analysis phase is shown in Fig. 4. The highest color value determined in the 144 square area divided into 12 * 12 grid areas was 98. The error was detected in two different square areas belonging to 90% and above the color value. As a result of the analysis, the defective parts are seen compared to the original pictures of the product.

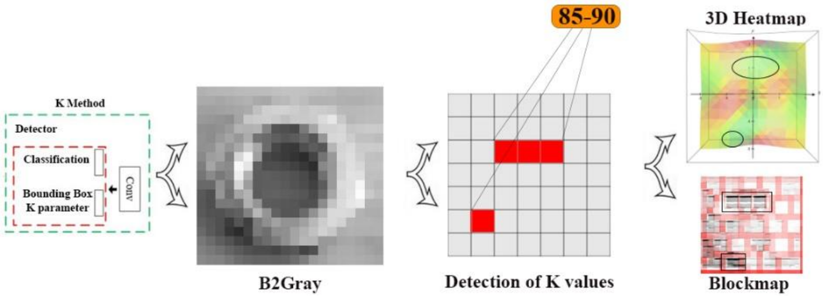


Figure 4. Structure of the K method (Detector)

The YOLOv4 algorithm for tiny object recognition begins by upsampling the 30x30 feature map acquired from the SPP network to 60x60, and then adding a k layer in between the feature maps as seen in Fig. 4. The original 60x60 feature map is then simply upsampled to 120x120. Finally, it makes contact with the 60x60 and 120x120 feature layers.

The YOLOv4 PANet integrates both low-and high-level data. However, it might have drawbacks when used to detect tiny objects. Upsampling from 60x60 to 120x120 may cause predictive data loss. When going from 256 channels to 128, there may be some information loss. Furthermore, on the 120x120 scale, the amount of model computation for the head network may increase. In Fig. 5, we can see the proposed KD-Feature YOLO's Fusion network in action.

Figure 5 shows the layout of the improved PANet. The feature maps of 60x60 with 256 channels, 26x26 with 128 channels, and 120x120 with 128 channels all have a feature of kSPP, rather than being upsampled directly from 60x60 to 120x120. The suggested KD-YOLO feature of k-network is shown in Fig. 4 to have an extra k layer in comparison to the original YOLOv4 architecture.

The network of pruners, enhanced when using the box as an anchor, precedes it with a priori. A rough estimate of the number of previous anchor boxes, as well as their aspect ratio, can be obtained by applying the K-means++ clustering method. The ability to determine bounding boxes may then be improved with the help of the previous anchor boxes. K was limited to a range of one to twelve. Figure 5 shows that these data suggest that the knee point of K is five, leading to an average IoU and a complexity of the model that are both manageable. The results show that there are five cluster centers with larger box-bound sizes (24, 24), (20, 36), (38, 12), (24, 28), and (30, 36).

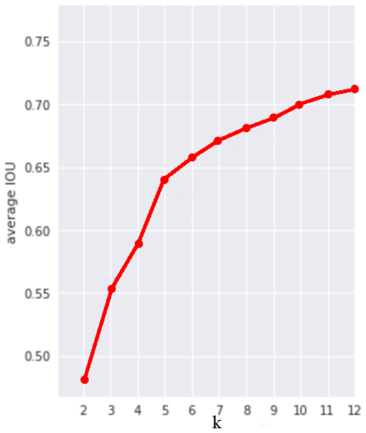


Figure 5. Shows the average IOU for a variety of anchor box counts.

4.1. Head kd-yolo

In the original YOLOv4 method, there are 5 prediction bounding boxes for each of the 120x120, 60x60, and 30x30 pixel feature maps. Larger receptive fields are ideal for small object detection in KD-YOLO, and deep-layer features like 60x60 and 30x30 with semantic-rich information are ideal. Research has shown that the detection of tiny objects might be challenging for sensors with a narrow receptive field. This study stopped using smaller feature maps in favor of a 120x120 feature map, which also reduced the complexity of the network.

5. Results

5.1. The Selection and Preparation Process of the Training and Testing Data

In this study, the training and testing data for the KD-YOLO model were meticulously selected and prepared from a domestic manufacturing facility, ensuring relevance and applicability to real-world industrial settings. A total of 2,400 high-resolution images (3264x2448 pixels) were captured using a single industrial camera, focusing on a range of metal parts with surface defects. The defects represented in these images comprise no more than 4% of the total area, reflecting realistic scenarios where defects are typically small and discrete. To prepare this dataset for effective model training and testing, we employed an 8:2 split, ensuring a comprehensive learning process while retaining a substantial portion for unbiased evaluation. The augmentation of the dataset was a critical step in our methodology. This system is equipped with a single NVIDIA GeForce GTX1650ti graphics card, 32GB of RAM, and an AMD Ryzen 7 4800 7 core processor. We used a learning rate of 0.001, an attenuation coefficient of 0.0005, and an iteration rate of 40000. Python 3.9 and the Darknet framework were used for this. Techniques like geometric distortion and mosaic data augmentation, based on the cut-mix theory, were employed. This approach helped in enhancing the dataset's diversity, adding contextual richness, and simulating various potential defect scenarios. The mosaic technique involved combining images in a 12x12 grid format, which proved particularly beneficial in increasing the complexity and variability of the training data, challenging the model to identify defects under diverse conditions. This rigorous process of data selection, augmentation, and preparation was crucial in training the KD-YOLO model to recognize and classify defects with high precision. By using real images from a manufacturing environment, we ensured the model's training was grounded in practical, industry-relevant scenarios. The diverse nature of the data, enhanced by our augmentation techniques, was instrumental in developing a robust model capable of handling the intricacies of real-world defect detection in metal components.

Data augmentation techniques, such as geometric distortion and mosaic, were used to add to the original data set. The cut-mix theory is the basis for mosaic data augmentation. As opposed to the four pictures utilized by the mosaic, there are only two in the cut-mix. Mosaic's use is beneficial since it adds more contextual information to picture databases. The data set was first mirrored and rotated to increase its size by a factor of ten. The photos were then combined using the mosaic technique, as seen in Fig. 6.

When evaluating the performance of an object detector, two important metrics to consider are its average precision (AP) and its mean average precision (MAP) in relation to the ground truth and the predicted bounding box (IOU). In this research, we conducted two tests to test the accuracy of the optimized feature of the K network. Data from the experiments are shown in Table 1.

Table 1 summarizes the average findings of many separate tests. As the graph shows, increasing the feature map's dimension leads to more accurate predictions. The performance of the KD-YOLO-4/512 is superior to that of its predecessor, the Model KD-YOLO-2/256. Both models employ the same resblock body and feature map dimension as YOLOv4. Table 1 shows that MAP has increased from 73.61 percent to 85.42 percent. KD-YOLO has a lower IOU than KD-YOLO-2/256 since it does so with a less-than-ideal choice for the previous anchor box. Table 2 shows that compared to YOLOv4, the suggested technique has a higher performance in terms of both accuracy and processing time. Potential causes include a better harmony between semantic and spatial data, as well as a larger total number of factors. As seen in Fig. 7, the KD-YOLO method converges after 40000 iterations.

Figure 6. Example of mosaic data augmentation

Table 1. Comparison of mAP and IOU

METHOD	PRECISION	F1 SCORE	RECALL	MAP	AVERAGE IOU
KD-YOLO 2/128	82.83	80.67	81.25	80.56	51.25
KD-YOLO 2/256	85.25	84.17	84.08	84.03	54.63
KD-YOLO 4/512	86.58	85.25	85.58	85.42	53.29
YOLOV4	76.08	73.58	74.41	73.61	49.04

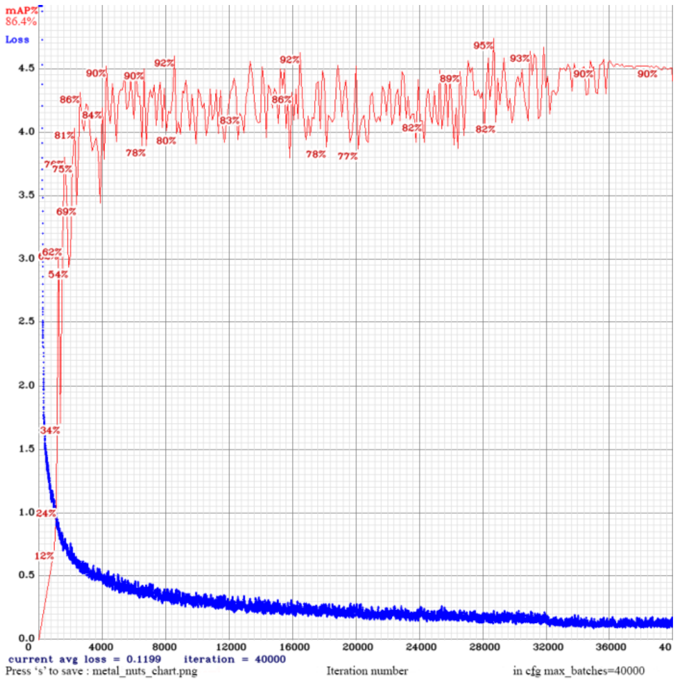


Figure 7. Training loss and mAP of KD-YOLO-4/512

Experiments were carried out to compare the performance of the KD-YOLO model with that of YOLOv4, YOLOv5, and PP-YOLO; the results are summarized in Table 2. All of these models were trained using the same data in this experiment. Based on the findings, the KD-YOLO model outperformed YOLOv4, YOLOv5, and PP-YOLO. For starters, KD-YOLO had the best mAP. Second, although the F1 values for KD-YOLO and PP-YOLO are quite close to each other, YOLOv4 is less accurate than KD-YOLO. Third, YOLOv4 has the worst performance of the four examined here. As a result, our results demonstrate that both deep and superficial information is necessary for tiny object recognition. In conclusion, our approach not only simplifies the model but also enhances the precision with which small objects are detected.

Table 2. Comparison of detection accuracy between KD-YOLO and the other models

MODEL	F1	MAP	BFLOP
YOLOV4	0.82	0.8	47.6
YOLOV5	0.85	0.84	100.4
PP-YOLO	0.86	0.83	118.5
KD-YOLO	0.89	0.91	70.1

6. Conclusions

In this investigation, KD-YOLO is offered as a method to improve chip surface categorization and defect detection. In this research, a novel technique called the k-feature method was used to improve the whole PANet, which was previously built on top of the CspDarknet53 meta network. Thus, by strategically combining the right layers, the receptive field can be improved using this technique. In addition, k-means++ was used as a preliminary anchor estimation technique in this investigation. Finally, the data set was preprocessed using the Mosaic data augmentation approach. Compared to the original YOLOv4 model, KD-YOLO performs better in terms of classification and detection accuracy, as well as the number of parameters. The suggested features of the k approach and pruning model may help with the identification of small objects. There is room for improvement in the accuracy of the detection and in the ability to make the model easier to understand. We expect to see both of these things in future research.

6.1. Practical Implications and Benefits

The KD-YOLO model, with its enhanced real-time processing capabilities, holds significant potential for practical applications in industrial settings, particularly in the field of production efficiency and quality control. In our study, the model's ability to swiftly process and analyze high-resolution images for defect detection in a real-world manufacturing environment showcases its practical utility. This rapid processing capability is essential in industrial production lines, where even small delays can accumulate, leading to significant time and resource expenditure. By integrating the KD-YOLO model into the quality control process, manufacturers can expect a more efficient workflow, with real-time detection allowing for immediate identification and rectification of defects. This immediacy not only helps in reducing waste but also ensures that the quality of the final product meets the high standards required in today’s competitive market. Furthermore, the model’s accuracy in defect detection minimizes the risk of false negatives, which are critical in maintaining product reliability and customer trust. Overall, the implementation of the KD-YOLO model in industrial environments exemplifies a move towards more automated, efficient, and reliable production processes, marking a significant step forward in the integration of advanced AI technologies in manufacturing.

6.2. Limitations of the KD-YOLO Model

In our study, we have identified several key limitations of the KD-YOLO model which are crucial for its future development and application in real-world scenarios. Firstly, the model was tested in a controlled manufacturing environment, and its performance in varied lighting conditions or complex backgrounds typical of real-world industrial settings is yet to be evaluated. This raises concerns about the model’s adaptability and robustness in less controlled environments, highlighting the need for training on more diverse datasets. Secondly, while the model excels in detecting small-scale defects, its efficacy across a diverse range of object sizes, especially larger defects, is not extensively validated. This suggests a potential need for further tuning to maintain high accuracy across all defect sizes. Thirdly, the significant computational resources required by the KD-YOLO model pose a challenge for real-time applications and in environments with limited processing capabilities. Optimizing the model for computational efficiency, possibly through model pruning or advanced hardware accelerators, is essential for its broader applicability. Lastly, the model's current performance evaluation is based on a dataset with limited variability, which might affect its generalization to different types of datasets with varying qualities of metal parts. Moreover, the model's ability to detect extremely small or faint defects, distinguish defects closely resembling background material, and accurately identify overlapping defects are areas that need attention. Addressing these failure scenarios is critical for enhancing the model's reliability and accuracy in high-stakes industrial quality control applications, ensuring it can effectively handle a wider range of real-world defect detection scenarios.

Recognizing the identified limitations of our KD-YOLO model, we envisage several promising avenues for future research to further refine and enhance its applicability. Primarily, algorithmic enhancements are crucial for better generalization capabilities across diverse datasets. This includes adapting the model to effectively handle variations in lighting, background complexity, and a broader range of defect types in metal parts, which are common in real-world industrial settings. Moreover, computational optimization is another key area for development, especially to facilitate real-time applications in manufacturing environments where processing power may be limited. This would involve streamlining the model architecture and exploring more efficient methods of feature extraction and object detection, without compromising the detection accuracy, particularly for small-sized defects. Additionally, future studies could focus on integrating the KD-YOLO model with other emerging technologies, such as augmented reality for enhanced visual inspection, or Internet of Things (IoT) for creating a more interconnected and automated defect detection system. By pursuing these directions, we aim to not only overcome the current limitations but also expand the practical utility and robustness of the KD-YOLO model in diverse industrial applications.

Author contribution

All stages of this study were carried out by YSB.

Declaration of Ethical Code

In this study, we undertake that all the rules required to be followed within the scope of the Higher Education Institutions Scientific Research and Publication Ethics Directive have been complied with, and that none of the actions specified under the title of Actions Contrary to Scientific Research and Publication Ethics of the said directive have been carried out.

The author of this article declare that the materials and methods used in this study do not require ethics committee approval and/or legal-specific permission.

Conflicts of interest

The author declare that they have no conflict of interest

References

Atwood, J., & Towsley, D. (2016). Diffusion-convolutional neural networks. *Advances in neural information processing systems*, *29*.

Cao, D., Dang, J., & Zhong, Y. (2021). Towards accurate scene text detection with bidirectional feature pyramid network. *Symmetry*, *13*(3), 486.

Cao, Z., Yang, H., Zhao, J., Pan, X., Zhang, L., & Liu, Z. (2019). A new region proposal network for far-infrared pedestrian detection. *IEEE Access*, *7*, 135023-135030.

Chi, W., Ma, L., Wu, J., Chen, M., Lu, W., & Gu, X. (2020). Deep learning-based medical image segmentation with limited labels. *Physics in Medicine & Biology*, *65*(23), 235001.

Chu, P., Li, Z., Lammers, K., Lu, R., & Liu, X. (2021). Deep learning-based apple detection using a suppression mask R-CNN. *Pattern Recognition Letters*, *147*, 206-211.

Girshick, R. (2015). Fast r-cnn. In *Proceedings of the IEEE international conference on computer vision* (pp. 1440-1448).

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.

He, K., Girshick, R., & Dollár, P. (2019). Rethinking imagenet pre-training. In *Proceedings of the IEEE/CVF International Conference on Computer Vision* (pp. 4918-4927).

Liao, X., Lv, S., Li, D., Luo, Y., Zhu, Z., & Jiang, C. (2021). YOLOv4-MN3 for PCB surface defect detection. *Applied Sciences*, *11*(24), 11701.

Lin, T. Y., Dollár, P., Girshick, R., He, K., Hariharan, B., & Belongie, S. (2017). Feature pyramid networks for object detection. In *Proceedings of the IEEE conference on computer vision and pattern recognition* (pp. 2117-2125).

Liu, S., Qi, L., Qin, H., Shi, J., & Jia, J. (2018). Path aggregation network for instance segmentation. In *Proceedings of the IEEE conference on computer vision and pattern recognition* (pp. 8759-8768).

Liu, Y., Wang, Q., Zhao, K., & Liu, Y. (2021). Real-time defect detection of hot rolling steel bar based on convolution neural network. *Chin. J. Sci. Instrum, 42*, 211-219.

MathWorks.com (2019) '*R-cnn, fast r-cnn, and faster r-cnn basics*', ©1994-2019 The MathWorks, Inc.

Monteiro, A. M., Vale, J. M., Cepêda, C. M., & de Almeida Leite, E. M. (2021). Internal control system quality and decision-making success: The role of the financial information quality. *Universal Journal of Accounting and Finance*, *8*(10), 3310-3322.

Panda, A., Dobránsky, J., Jančík, M., Pandová, I., & Kačalová, M. (2018). Advantages and effectiveness of the powder metallurgy in manufacturing technologies. *Metalurgija*, *57*(4), 353-356.

Parico, A. I. B., & Ahamed, T. (2021). Real time pear fruit detection and counting using YOLOv4 models and deep SORT. *Sensors*, *21*(14), 4803.

Punn, N. S., & Agarwal, S. (2020). Inception u-net architecture for semantic segmentation to identify nuclei in microscopy cell images. *ACM Transactions on Multimedia Computing, Communications, and Applications (TOMM)*, *16*(1), 1-15.

Silvius, A. J., & Schipper, R. P. (2014). Sustainability in project management: A literature review and impact analysis. *Social business*, *4*(1), 63-96.

Śmietańska, K., & Podziewski, P. (2019). A human quality control system in furniture manufacturing–a pilot study. *Annals of Warsaw University of Life Sciences SGGW Forestry and Wood Technology*, *108*, 93-96.

Wang, C., & Zhong, C. (2021). Adaptive feature pyramid networks for object detection. *IEEE Access*, *9*, 107024-107032.