НАУЧНО-ТЕХНИЧЕСКИЙ ВЕСТНИК ИНФОРМАЦИОННЫХ ТЕХНОЛОГИЙ, МЕХАНИКИ И ОПТИКИ
ноябрь–декабрь 2024 Том 24 № 6 http://ntv.ifmo.ru/
SCIENTIFIC AND TECHNICAL JOURNAL OF INFORMATION TECHNOLOGIES, MECHANICS AND OPTICS
November–December 2024 Vol. 24 No 6 http://ntv.ifmo.ru/en/
ISSN 2226-1494 (print) ISSN 2500-0373 (online)

НАУЧНО-ТЕХНИЧЕСКИЙ ВЕСТНИК
ИНФОРМАЦИОННЫХ ТЕХНОЛОГИЙ, МЕХАНИКИ И ОПТИКИ

doi: 10.17586/2226-1494-2024-24-6-972-981

Surface defect detection with limited data based on SSD detector and Siamese networks

Nikita P. Novgorodcev[1], Kseniia A. Baturina[2]✉, Valeria A. Efimova[3], Anatoly A. Shalyto[4]

[1,2,3,4] ITMO University, Saint Petersburg, 197101, Russian Federation

[1] novgorodcevnp@gmail.com, https://orcid.org/0009-0000-5352-1852
[2] kseniya.baturina@mail.ru✉, https://orcid.org/0009-0003-7555-282X
[3] valeryefimova@gmail.com, https://orcid.org/0000-0002-5309-2207
[4] shalyto@mail.ifmo.ru, https://orcid.org/0000-0002-2723-2077

Abstract
This study presents an algorithm for the problem of detecting defects on hard surfaces when trained with zero or a small number of examples, addressing the challenge of limited data availability. The existing defect detection methodology using machine vision is enhanced. A hybrid approach is proposed, combining the strengths of the SSD detector and Siamese Neural Networks (SNN). The SSD detector extracts feature vector representations from images, while the SNNs are used to construct the feature space. The new approach demonstrates high accuracy in detecting both known and previously unseen defects in the training dataset. Based on testing across seven different datasets, the model showed good performance in scenarios with a limited number of training examples. A comparative analysis with existing models highlights the high performance of the proposed approach and its potential as an innovative and effective solution for the universal detection of defects on hard surfaces.

Keywords
computer vision, defects detection, zero-shot learning, object detection, Siamese networks

For citation: Novgorodcev N.P., Baturina K.A., Efimova V.A., Shalyto A.A. Surface defect detection with limited data based on SSD detector and Siamese networks. *Scientific and Technical Journal of Information Technologies, Mechanics and Optics*, 2024, vol. 24, no. 6, pp. 972–981. doi: 10.17586/2226-1494-2024-24-6-972-981

УДК 004.855.5

Обнаружение дефектов твердых поверхностей при ограниченном объеме данных на основе SSD-детектора и сиамских сетей

Никита Павлович Новгородцев[1], Ксения Александровна Батурина[2]✉, Валерия Александровна Ефимова[3], Анатолий Абрамович Шалыто[4]

[1,2,3,4] Университет ИТМО, Санкт-Петербург, 197101, Российская Федерация

[1] novgorodcevnp@gmail.com, https://orcid.org/0009-0000-5352-1852
[2] kseniya.baturina@mail.ru✉, https://orcid.org/0009-0003-7555-282X
[3] valeryefimova@gmail.com, https://orcid.org/0000-0002-5309-2207
[4] shalyto@mail.ifmo.ru, https://orcid.org/0000-0002-2723-2077

Аннотация
Введение. Представлен алгоритм решения задачи обнаружения дефектов твердых поверхностей при обучении на нулевом или малом числе примеров, который решает проблему ограниченного объема данных. Усовершенствуется существующая методология обнаружения дефектов методом с использованием машинного зрения. **Метод.** Предложен гибридный подход, сочетающий преимущества SSD-детектора и сиамских нейронных сетей. SSD-детектор позволяет извлекать векторные представления признаков из изображений, а сиамские нейронные сети применяются для построения пространства извлеченных признаков. **Основные результаты.** Показано, что новый подход обладает высокой точностью как на известных, так и на не встречавшихся ранее в обучающей выборке дефектах. По результатам тестирования на 7 различных наборах данных представленный

алгоритм продемонстрировал хорошие возможности в сценариях с ограниченным числом примеров для обучения. **Обсуждение.** Сравнительный анализ с существующими моделями показал высокую производительность предлагаемого алгоритма и его потенциал как инновационного и эффективного решения задач универсального обнаружения дефектов твердых поверхностей.

Ключевые слова
компьютерное зрение, обнаружение дефектов, обучение с нулевым количеством примеров, обнаружение объектов, сиамские сети

Ссылка для цитирования: Новгородцев Н.П., Батурина К.А., Ефимова В.А., Шалыто А.А. Обнаружение дефектов твердых поверхностей при ограниченном объеме данных на основе SSD-детектора и сиамских сетей // Научно-технический вестник информационных технологий, механики и оптики. 2024. Т. 24, № 6. С. 972–981 (на англ. яз.). doi: 10.17586/2226-1494-2024-24-6-972-981

Introduction

The detection and precise localisation of objects in images or videos are fundamental tasks with diverse applications across industries, encompassing medicine [1], transportation [2], and retail [3]. Notable instances of object detection applications include automated face detection, traffic violation identification, and quality control in manufacturing illustrating the versatility of object detection technologies and propelling advancements in computer vision. Within the expansive domain of computer vision, defect detection emerges as a critical undertaking with profound implications for various sectors, spanning industrial production [4, 5] and medical diagnostics [6].

In practical industrial settings, pixel-level detection of defects translates into more understandable and usable metrics for engineers, such as defect type, its physical size, and depth. This research focuses only on defect detection on 2D images captured through the following pipeline. Typically, factories install cameras over conveyor belts with fixed height, focal length, and resolution to capture images. These images are then used to create datasets. Our solution processes these images, detects defects, and generates bounding boxes. The algorithm automatically converts these pixel coordinates into millimetres or other professional terms to meet manufacturing precision requirements. For instance, camera calibration, its height and resolution setup allow for conversion from pixel coordinates to real-world measurements, enabling engineers to accurately pinpoint defect locations. This pipeline ensures that our detection algorithms provide valuable data that can be seamlessly integrated into factory workflows, enhancing the overall efficiency and precision of quality control.

The ongoing challenge for researchers and practitioners lies in designing algorithms capable of adapting to the intricacies of specific domains. This study addresses the multifaceted challenges inherent in defect detection, particularly on solid surfaces, focusing on those defects that are visually observable without considering their depth where nuanced methodologies are essential for effective operation with limited data and adaptation to diverse defect patterns. Our primary objective is to introduce an innovative algorithm that excels in zero-shot and few-shot defect detection, surmounting the limitations of existing models.

To achieve this goal, we propose a novel defect detection methodology that seamlessly integrates the strengths of the Single Shot MultiBox Detector (SSD300) [7] and Siamese Neural Networks (SNNs) [8]. This approach aims to extract rich feature representations from images using SSD300 while the Siamese networks construct a feature space conducive to defect detection. Our model employs a distance-based defect detection method demonstrating commendable mean Average Precision (mAP) performance on familiar and previously unseen defect datasets.

In pursuit of our objectives, the study evaluates the proposed algorithm on seven distinct datasets, showcasing its predictive capabilities and robustness across diverse scenarios. We aim to contribute to developing a universal defect detection solution that accommodates limited data and offers real-time zero-shot detection.

The subsequent sections of this paper provide a comprehensive exploration of the landscape of defect detection methods, detail our proposed methodology, including the neural network architectures and the employed loss function, and present implementation details, results, and comparative analyses with existing solutions. Through this endeavour, we seek to illuminate the transformative potential of our approach in revolutionising defect detection on solid surfaces.

Related works

Defect detection on solid surfaces is a critical task in various industrial settings. Accurate and efficient detection methods can significantly enhance manufacturing precision and operational efficiency. This paper evaluates the SSD300+Siamese approach for universal defect detection building upon the foundation of existing research and methodologies.

We conducted experiments using the North Eastern University Steel Surface Defects Database (NEUSSDD) [9], the Wood Defects dataset [10], and the MVTec Anomaly Detection Dataset [11]. These datasets present distinct challenges and encompass a range of defect classes making them suitable for comprehensive model assessment. The NEU dataset provides a benchmark for steel surface defects, the Wood Defects dataset includes a variety of wood surface flaws and the MVTec dataset is designed for anomaly detection in industrial inspection scenarios. The images in Fig. 1, 2, and 3 illustrate examples of the types of defects we aim to detect using our model. These examples highlight the diversity and complexity of defects in different materials, demonstrating the necessity for robust and adaptable detection methods.

The NEU dataset [9] consists of 1800 grayscale images of steel surfaces, with various manufacturing defects

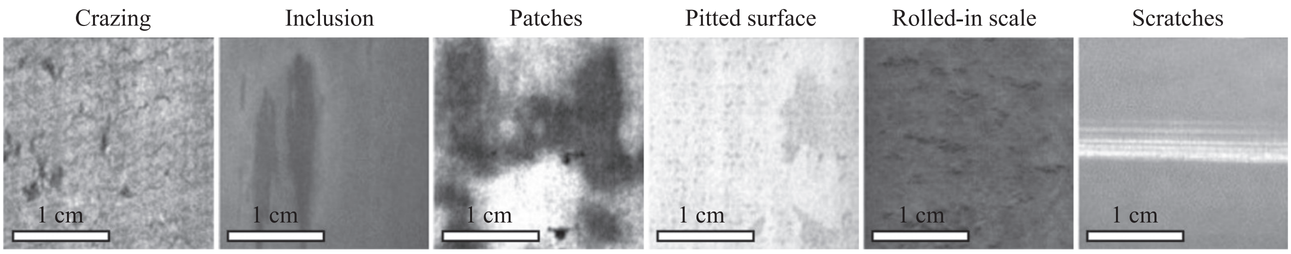

Fig. 1. Examples of NEU dataset defects

Fig. 2. Example of Wood Defects dataset defects

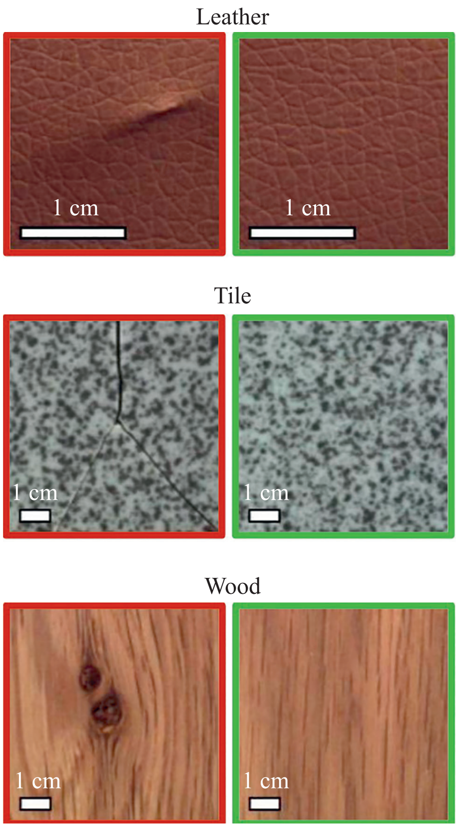

Fig. 3. Examples of MVTec dataset surfaces, with and without defects. Images containing defects are outlined in red, while others are outlined in green

occurring during production. Six defect categories are present: surface scratch, pitted surface, rolled-in surface, patches, inclusion, and crazing. This dataset is a benchmark for evaluating the model performance in detecting defects on steel surfaces.

The Wood Defects dataset [10] comprises over 43,000 labelled images of wood surface defects, covering ten common defect types such as live knots, dead knots, cracks, resins, marrows, etc. The dataset includes semantic maps and bounding box labels for each image, facilitating semantic segmentation and localisation tasks. This dataset challenges the model with diverse defect patterns encountered in wood production.

The MVTec dataset [11] is designed to benchmark anomaly detection methods, focusing on industrial inspection scenarios. It comprises over 5,000 high-resolution images across fifteen object and texture categories, each containing defect-free training images and test images with various anomalies. Pixel-precise annotations of anomalies provide a detailed evaluation of the model ability to detect and localise defects.

Building on the foundation of established methods, our research draws from several key works in the field. These studies form the basis for our approach and highlight the progress and challenges in defect detection on solid surfaces.

SSD300. The work by Liu et al. [7] introduced the SSD300, a pioneering method for object detection in images using a single deep neural network. SSD300 discretizes the output space of bounding boxes into a set of default boxes over different aspect ratios and scales per feature map location. This approach eliminates the need for object proposal generation and subsequent pixel or feature resampling stages, making SSD300 easy to train and integrate into systems requiring a detection component. The method combines predictions from multiple feature maps with different resolutions to efficiently handle objects of various sizes. The SSD300 algorithm achieves comparable accuracy to methods with an additional object proposal step while being significantly faster. However, SSD300 is designed for generic object detection and may need to be optimised for the specific challenges of universal defect detection on solid surfaces.

The SSD300 architecture utilises a feed-forward convolutional network to generate bounding boxes and scores representing object class instances. Multi-scale detection is facilitated by appending convolutional feature layers to the truncated backbone progressively decreasing in size. For each feature layer, a set of convolutional filters produces detection predictions for category scores or shape offsets relative to default box coordinates. The SSD300 model excels in object detection tasks but may not be tailored to the nuances of defect detection on solid surfaces.

While SSD300 is effective for generic object detection, its performance in defect detection scenarios with limited data and specific challenges of solid surfaces may be suboptimal. The model reliance on default box representations may not capture the intricacies of defects, necessitating a more specialised approach.

Signature Verification using a "Siamese" Time Delay Neural Network. The algorithm by Bromley et al. [8] proposes a signature verification method utilizing a SNN architecture. This structure comprises two identical sub-networks connected at their outputs, with the joining neuron measuring the distance between feature vectors extracted from two signatures. Signatures that fall within a distance threshold of the stored representation are accepted, while those beyond the threshold are rejected as forgeries. Although the Siamese network demonstrates proficiency in signature verification, its direct application to defect detection on solid surfaces may be limited due to differences in data characteristics.

The Siamese network developed for signature verification utilizes two identical sub-networks that process input signatures. The network measures the distance between feature vectors extracted from two signatures, allowing for effective verification. Training involves using a modified back propagation algorithm, and various architectures are tested for their performance in resisting forgeries and remaining insensitive to changes in signature characteristics.

While the Siamese network excels in signature verification, its direct application to defect detection on solid surfaces may face challenges. Signatures and defects have distinct characteristics, and the Siamese network may not capture the diverse visual patterns associated with defects. Additionally, its sensitivity to temporal features may not align with the static nature of defect images.

Detecting Object Defects with FCSNNs. Nagy and Czuni (2021) [12] present a convolutional SNN-based approach for defect detection, emphasising the network ability to generalise knowledge across different object classes without re-training. The Fusioning Convolutional Siamese Neural Network (FCSNN) introduced in this work leverages SNNs to recognise defects in new object types.

The FCSNN architecture involves feature extraction using pre-trained VGG16 models, fully connected layers, a fusion of features through concatenation, and the computation of a similarity score. The evaluation is performed on traffic signs and castings datasets, demonstrating competitive accuracy compared to other deep learning models.

While FCSNN exhibits strengths in defect detection, it is limited to binary classification and may face challenges when dealing with multi-class scenarios. Additionally, the reliance on pre-trained models and fusion strategies may limit its adaptability to diverse defect datasets.

Convolutional Ensembling based Few-Shot Defect Detection Technique. Karmakar et al. (2022) [13] propose a few-shot defect detection technique based on convolutional ensembling aiming for real-time implementation. The approach employs a knowledge base of pre-trained convolutional models and introduces a novel ensembling strategy to enhance accuracy while minimising parameter count.

The method combines ResNet 50, EfficientNet B5, and DenseNet 201 for feature embeddings. Features are reshaped, stacked, and processed through a convolutional block and a Multi-Layer Perceptron for classification. The ensembling strategy involves stacking and combining features from different feature extractors.

While the approach excels in few-shot scenarios, it may lack effectiveness in zero-shot defect detection. The reliance on pre-trained models and ensembling techniques may introduce computational overhead limiting its real-time applicability in specific settings.

Few-shot learning combine attention mechanism-based defect detection in bar surface. Lv and Song (2019) [14] proposed a deep learning-based approach to address the challenge of defect detection in the production process of bar steel. In this context, the conventional reliance on human eye observation prompted the need for an automated solution. The authors emphasised the advantages of deep learning methods, mainly unsupervised or weakly supervised techniques, which are more suitable for scenarios with limited samples, a common constraint in industrial settings.

The method introduced in the paper involves dividing images of the bar surface into fragments, computing features using a Convolutional Neural Network, and employing an attention mechanism to highlight crucial features while reducing the impact of noise. Few-shot learning is utilised to train the model on a small set of images with and without defects, enabling it to generalise to new images.

Despite its merits, the approach has limitations. It focuses on defect detection in the specific context of bar steel production and may not be readily adaptable to universal defect detection on solid surfaces. The reliance on few-shot learning might not be optimal for scenarios where the availability of labelled data is slightly more abundant, as in our study.

Siamese Basis Function Networks for Data-efficient Defect Classification in Technical Domains [15]. Siamese networks and radial basis function networks are used for data-efficient defect classification in technical domains. They focused on developing models using three specialised datasets and demonstrating general applicability to classical datasets.

The method involves training the model on a small dataset to generate basis functions which are then used to represent new data. The nearest neighbour method is employed for defect classification, determining the class of an unknown sample based on the closest sample from the training set.

While the proposed approach is practical for data-efficient classification, it does not directly address the defect detection task. The method relies on the nearest-neighbour approach which may not be suitable for scenarios where defects need to be detected and located precisely. Additionally, the emphasis on data efficiency might limit its performance when dealing with larger datasets.

Classification and Fast Few-Shot Learning of Steel Surface Defects with Randomized Network. Nagy and Czuni (2022) [16] present a classification and fast few-shot learning method for steel surface defect detection. Their work addresses challenges related to a low number of available shots for new defect classes, catastrophic forgetting of known information, and the time-consuming nature of retraining or fine-tuning existing models.

The authors introduce a novel architecture that combines EfficientNet deep neural networks with randomised classifiers. Random matrices serve as weighting factors for the neural networks, resulting in randomised neural networks. These networks are trained on a small dataset comprising several samples of each defect class. The authors propose the "Fast Few-Shot Learning" method to enable fast learning on new data, facilitating rapid adaptation to new classes with only a few samples.

Despite the success in achieving high classification accuracy, Nagy and Czuni's approach is not designed for real-time defect detection. Additionally, the method focuses on classification rather than defect detection, and its reliance on random matrices may limit interpretability and generalisation.

Segment Any Anomaly without Training via Hybrid Prompt Regularization. Cao et al. (2023) [17] present a novel framework called Segment Any Anomaly + (SAA+) for zero-shot anomaly segmentation with hybrid prompt regularisation. The framework aims to enhance the adaptability of modern foundation models to anomaly segmentation tasks.

The proposed SAA+ framework leverages non-parameter foundation models like Segment Anything to incorporate diverse multi-modal prior knowledge for anomaly localisation. Hybrid prompts, derived from domain expert knowledge and target image context, act as regularisations for adapting non-parameter foundation models to anomaly segmentation. The model performs state-of-the-art anomaly segmentation benchmarks in a zero-shot setting.

While successful in zero-shot anomaly segmentation, the SAA+ framework is not designed for real-time applications. It focuses on segmentation rather than defect detection, and its dependency on prompt-based and regularisation-based approaches may limit its effectiveness in scenarios with diverse defect classes and limited data.

Summary

While each method contributes significantly to defect detection on solid surfaces, our proposed SSD300+Siamese approach is a comprehensive solution. By seamlessly integrating real-time processing, zero-shot learning, and multi-class defect detection, our method addresses key aspects crucial for universal defect detection. A detailed description of our solution, experimental results, and comparative analysis will be presented in the subsequent sections, providing a comprehensive understanding of the strengths and advantages of our proposed approach.

Method

Our proposed defect detection methodology is a novel adaptation of the SSD300 architecture, aiming to achieve universal defect detection on solid surfaces. Our modified approach retains the SSD300 backbone but makes crucial adjustments to enhance defect detection capabilities. Specifically, we remove class prediction layers and introduce new layers before the bounding boxes localisation. These additional layers are pivotal in creating a feature space conducive to efficient defect pattern recognition.

A key aspect of our methodology is the integration of SNNs which are trained using the Triplet loss function. This training method involves three images: an anchor image representing the target object, a positive image depicting a similar object, and a negative image showcasing an object dissimilar to the anchor. Training the network on such triplets enables it to learn features that distinguish objects from one another, facilitating the identification of desired defect patterns. In our context, these patterns correspond to defects highlighted by bounding boxes.

The use of Siamese networks is particularly advantageous for our defect detection task due to their inherent ability to effectively handle and compare pairs of images. Unlike conventional fully connected layers, which process embeddings independently, Siamese networks are designed to focus on the relative similarities and differences between pairs of images. This is crucial for defect detection, where the key challenge lies in identifying subtle variations between defective and non-defective regions. By learning a discriminative feature space, Siamese networks enhance the model capability to differentiate between similar and dissimilar patterns, making them better suited for tasks that require high sensitivity to minor variations, such as defect detection on solid surfaces.

As depicted in Fig. 4, the resulting architecture consists of intermediate layers trained on image triples and bounding box search layers that utilise the feature space for defect pattern identification.

The architecture comprises the following key stages:

— SSD300 Backbone: The initial stage involves the transfer of the input image to the model, generating

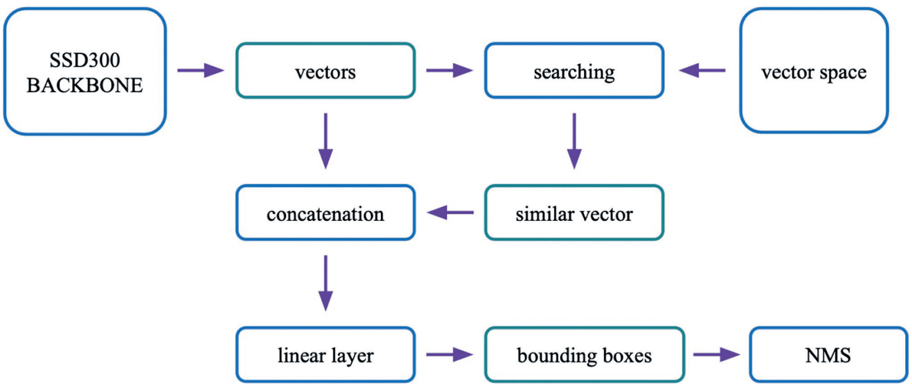

Fig. 4. SSD300+Siamese Networks for Defect Detection

vectors for each bounding box without offsets. These vectors provide a preliminary representation of potential defect locations.
- Searching: In this stage, the obtained vectors are compared with examples of defects pre-given to the model. The closest vector in the feature space becomes the target vector representing the most similar example. This process enables efficient search and identification of defect patterns.
- Concatenation: The obtained vector is concatenated with the found template vector enriching the feature representation. This step ensures the model captures and integrates relevant information from the input image and the learned defect patterns.
- Linear Layer: The concatenated vector undergoes a linear layer operation to obtain offsets for the corresponding bounding box, refining localisation. This layer introduces fine-tuning to align the model predictions with the precise location of defects in the image.
- Non-Maximum Suppression (NMS): Overlapping boxes and those with uncertainty are filtered out in this stage, improving precision. NMS is a critical post-processing step that ensures the final set of predicted bounding boxes is accurate, non-redundant, and well-localised. It helps eliminate duplicate detections and selects the most confident predictions by retaining the one with the highest confidence score while suppressing others with significant overlap.

Training Process

Labelled examples of each defect class must be loaded to ensure the effective operation of the model. This step is critical as it allows the model to extract essential features for defect detection on unseen data. The loaded examples are converted into feature vectors, serving as templates for defect patterns.

The model generates 8,732 vectors for each image, each representing features of a different part of the image. These vectors are compared with template vectors, and similar ones are saved as templates with the vector class indicated. Vectors far from the templates in the feature space are excluded from further processing. The distance threshold at which a vector is considered distant is a parameter.

Performance Evaluation

In our evaluation, we utilise the mAP metric at the Intersection Over the Union (IoU) threshold of 50 %, referred to as mAP50. This metric comprehensively measures the model precision and recall, considering the overlap between predicted bounding boxes and ground truth annotations. A mAP score of 50 % indicates a moderate level of overlap between predicted and ground truth bounding boxes. The choice of mAP50 as the primary evaluation metric was driven by the specific requirements of our task, where detecting defects is more important than precisely defining their boundaries. mAP50 is particularly suitable in this context as it prioritises successful detection over exact localisation. In contrast, mAP95, which demands higher boundary precision, is unnecessarily strict for our purposes and may penalise minor localisation errors that don't impact the overall outcome. Thus, mAP50 provides a more relevant and effective measure for evaluating our model performance.

With a focus on experiments and results, the subsequent section will delve into a detailed exploration of the performance of our proposed defect detection methodology across various datasets. The experimental outcomes will provide insights into the model effectiveness in addressing challenges such as limited data, diverse defect patterns, and real-time detection requirements. The presented results aim to substantiate our approach transformative potential in universal defect detection on solid surfaces.

Experiments and results

Datasets

To evaluate the proposed SSD300+Siamese approach for universal defect detection on solid surfaces, we conducted experiments on two training datasets (NEUSSDD [9] and Wood Defects dataset [10]). Additionally, we tested the model on the MVTec Anomaly Detection Dataset [11]. Each dataset presents distinct challenges and encompasses a range of defect classes, making them suitable for comprehensive model assessment.

Additionally, the model was tested on four unseen datasets, namely Wheat leaf dataset[1], Meat dataset[2], Car defect dataset[3], and MSWeldDefect dataset[4], to assess its generalisation capabilities further. The examples of the types of defects presented in these datasets are presented in Fig. 5. The subsequent section will present the detailed outcomes obtained from these experiments.

Training

The training process involved two primary datasets, NEU Steel Surface Defects and Wood Defects, and utilised a pre-trained SSD300 model on ImageNet. The model underwent training for 100 epochs, employing a weighted sum of IoU loss and Triplet loss. The Triplet loss facilitated the creation of a feature space conducive to efficient defect pattern recognition.

During training, null vectors were created for each class in the model, later replaced by vectors obtained from the SSD300. This enabled the model to develop triplets independently from different sources supporting multi-dataset training. An iterative example of saving vectors during training is illustrated in Fig. 6.

Model Metrics on Trained Datasets

The model performance was evaluated on the NEU, Wood, and MVTec datasets using varying numbers of templates. The mAP at 50 % IoU (mAP50) was employed as the evaluation metric, providing insights into the precision and recall of the defect detection.

As presented in Table 1, results highlight the model varying performance depending on the dataset and

---

[1] Available at: https://www.kaggle.com/datasets/olyadgetch/wheat-leaf-dataset (accessed: 30.10.2024).

[2] Available at: https://www.kaggle.com/crowww/meat-quality-assessment-based-on-deep-learning (accessed: 30.10.2024).

[3] Available at: https://www.kaggle.com/knightnikhil/cardefect (accessed: 30.10.2024).

[4] Available at: https://www.kaggle.com/datasets/ruthka/maskrcnn-mswelddefect (accessed: 30.10.2024).

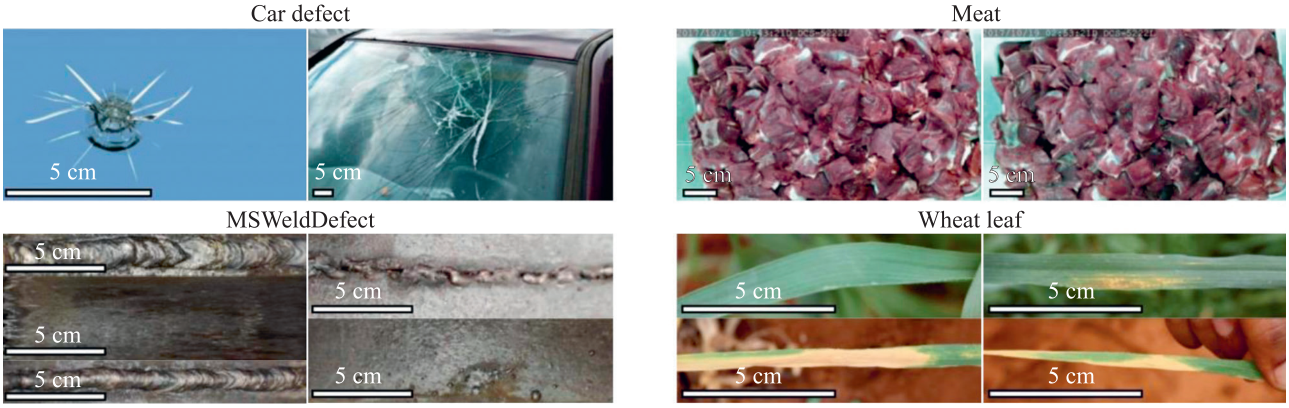

Fig. 5. Car defect, Meat, MSWeldDefect and Wheat leaf datasets examples

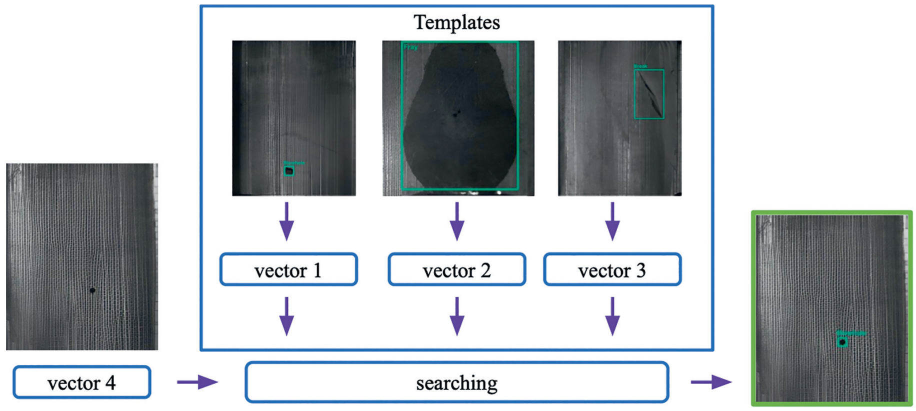

Fig. 6. An example of searching for a similar defect among the templates

Table 1. Metrics for different datasets, s is the number of samples given to the model

Dataset	5 s, mAP50	10 s, mAP50	15 s, mAP50	20 s, mAP50
MVTec	0.51	0.67	0.72	0.81
NEU	0.62	0.75	0.84	0.92
Wood	0.64	0.77	0.89	0.91

the number of templates provided. Notably, the model demonstrated exceptional accuracy in detecting defects on the NEU dataset, showcasing its effectiveness in handling diverse steel surface anomalies. These results underscore the model adaptability to specific datasets and the importance of providing adequate templates for more intricate defect classes. Fig. 7 demonstrates the model ability to accurately detect defects in Wood and MVTec datasets, highlighting its robustness and adaptability across various materials.

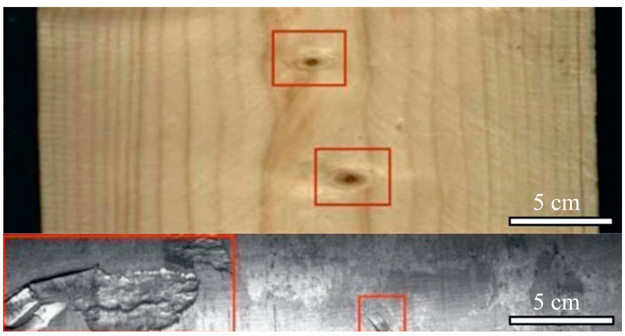

Fig. 7. Example results from the Wood and MVTec datasets. The identified defects are highlighted with red bounding boxes

Cross-Dataset Testing

Cross-dataset testing was conducted on datasets the model had not been explicitly trained on to assess the model generalisation capabilities. As shown in Table 2, the results indicate the model accuracy in detecting defects on different datasets, showcasing its potential for real-world applications.

The results of Table 2 show that the model exhibits a high level of accuracy in detecting defects in specific datasets. At the same time, its performance is comparatively lower in more complex datasets such as those containing leaves. These results indicate that the model performs well in processing simple data that does not include many details.

Real-time Processing Speed

To evaluate the real-time capabilities of the model, a speed analysis was performed on the NEU dataset, varying

Table 2. The model performance on the data it was not trained on using 10 samples to construct embeddings

Test dataset	Result, accuracy
Wheat leaf	0.68
Meat	0.99
Car defect	1.0
MSWeldDefect	1.0

the number of templates provided. As depicted in Fig. 8, the model demonstrated impressive real-time processing capabilities even with a substantial number of templates. Testing was done on an RTX 3090.

The results in Fig. 8 highlight the model efficiency. It maintains real-time processing capabilities across various template quantities, making it suitable for practical, dynamic applications.

Comparative Analysis

A comparative analysis with existing defect detection models is presented in Table 3, emphasising key attributes, such as real-time processing, zero-shot learning, and multi-class defect detection. The models presented in Table 3 were introduced in the studies discussed in the Related works section.

Due to the differences in the datasets used and experimental conditions, reproducing the results of other works is challenging. Our solution implements aspects not found in any of the previous works, such as the combination of real-time processing, zero-shot learning, and multi-class defect detection in a single framework. Therefore, it is impossible to conduct a quantitative analysis directly comparing our approach with existing methods. Nonetheless, the comparative analysis provided in Table 3 reinforces the innovative and versatile nature of our proposed methodology, highlighting its unique strengths in addressing the challenges of defect detection on solid surfaces.

Table 3 highlights the unique strengths of our proposed SSD300+Siamese approach, excelling in real-time processing, zero-shot learning, and multi-class defect detection. As demonstrated in Table 3, none of the presented models, except for our proposed approach, can achieve real-time multi-class detection, particularly in zero-shot learning. This analysis underscores the significance of our contribution to the field and the potential of our methodology as a comprehensive solution for universal defect detection.

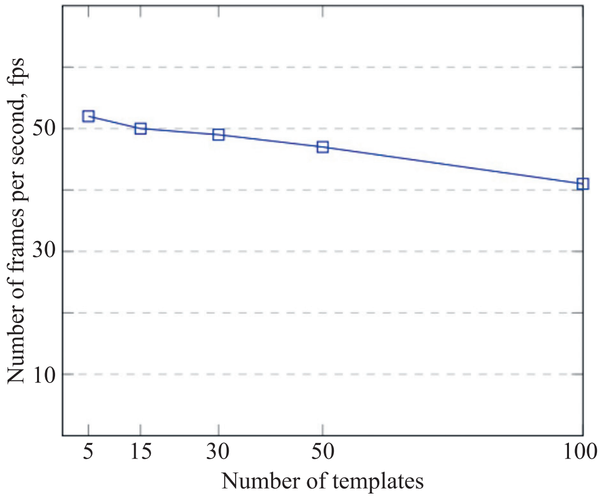

Fig. 8. The dependence of the algorithm speed on the number of templates

Conclusion and Discussion

In this study, we introduced an innovative SSD300+Siamese approach for universal defect detection on solid surfaces, leveraging the strengths of the SSD300 architecture for object detection and Siamese networks for feature learning. The model showcased exceptional accuracy across diverse datasets, emphasising its robust performance on explicitly trained datasets, including the NEU Dataset for steel surfaces and a Wood Surface Defects Dataset. Cross-dataset testing further underscored the model adaptability to new datasets, affirming its potential for real-world applications. Our proposed approach demonstrates a novel solution for universal defect detection, with distinguishing features, such as real-time processing capabilities, proficiency in zero-shot learning, and the ability to perform multi-class defect detection. These aspects set our model apart from existing solutions, positioning it as a cutting-edge contribution to universal defect detection. However, the evaluation of data and concept drift was not extensively covered in this study, and additional experiments are needed to assess these aspects, which could be a focus of future research. We also considered the application of transformer models; however, their effective implementation typically requires significant amounts of training data. Given the current limited dataset, we chose to focus on methods better suited to these constraints.

Table 3. Comparison with existing solutions

Model	Real-time	Zero-shot	Multi-class	Detection
FCSNN [12]	+	—	—	—
ConvEnsembling [13]	+	—	+	—
CombAttention [14]	—	—	+	+
SBFN [15]	+	—	+	—
RandNet [16]	+	—	+	—
HPR [17]	—	+	—	+
SSD300+Siamese	+	+	+	+

Future research may explore the use of transformers as data availability increases, potentially yielding further improvements in defect detection performance. Further research may also investigate more precise region-of-interest extraction, refinement of the Siamese network for enhanced computational efficiency, and transfer learning techniques to facilitate the model adaptation to new datasets.

References

1. Kamiya N., Zhou X., Chen H., Muramatsu C., Hara T., Fujita H. Model-based approach to recognize the rectus abdominis muscle in CT images. *IEICE Transactions on Information and Systems*, 2013, vol. E96D, no. 4, pp. 869–871. https://doi.org/10.1587/transinf.e96.d.869
2. Bai T., Gao J., Yang J., Yao D. A study on railway surface defects detection based on machine vision. *Entropy*, 2021, vol. 23, no. 11, pp. 1437. https://doi.org/10.3390/e23111437
3. Saberironaghi A., Ren J., El-Gindy M. Defect detection methods for industrial products using deep learning techniques: a review. *Algorithms*, 2023, vol. 16, no. 2, pp. 95. https://doi.org/10.3390/a16020095
4. Srividhya R., Shanmugapriya K., SindhuPriya K. Automatic detection of surface defects in industrial materials based on image processing. *International Journal of Engineering & Technology*, 2018, vol. 7, no. 3.34, pp. 61–64. https://doi.org/10.14419/ijet.v7i3.34.18717
5. Guijo D., Onofre V., Del Bimbo G., Mugel S., Estepa D., De Carlos X., Adell A., Lojo A., Bilbao J., Orus R. Quantum artificial vision for defect detection in manufacturing. *arXiv*, 2022. arXiv:2208.04988. https://doi.org/10.48550/arXiv.2208.04988
6. Prajwala N.B. Defect detection in pharma pills using image processing. *International Journal of Engineering & Technology*, 2018, vol. 7, no. 3.3, pp. 102–106. https://doi.org/10.14419/ijet.v7i3.3.14497
7. Liu W., Anguelov D., Erhan D., Szegedy C., Reed S., Fu C.-Y., Berg A.C. SSD: Single shot multibox detector. *Lecture Notes in Computer Science*, 2016, vol. 9905, pp. 21–37. https://doi.org/10.1007/978-3-319-46448-0_2
8. Bromley J., Guyon I., LeCun Y., Säckinger E., Shah R. Signature verification using a «Siamese» time delay neural network. *Advances in Neural Information Processing Systems*, 1993, vol. 6.
9. Song K.-C., Hu S., Yan Y. Automatic recognition of surface defects on hot-rolled steel strip using scattering convolution network. *Journal of Computational Information Systems*, 2014, vol. 10, no. 7, pp. 3049–3055.
10. Kodytek P., Bodzas A., Bilik P. Supporting data for Deep Learning and Machine Vision based approaches for automated wood defect detection and quality control. *Zenodo*, 2015. Available at: https://zenodo.org/records/4694695 (accessed: 30.10.2024)
11. Bergmann P., Batzner K., Fauser M., Sattlegger D., Steger C. The MVTec anomaly detection dataset: a comprehensive real-world dataset for unsupervised anomaly detection. *International Journal of Computer Vision*, 2021, vol. 129, no. 4, pp. 1038–1059. https://doi.org/10.1007/s11263-020-01400-4
12. Nagy A.M., Czúni L. Detecting object defects with fusioning convolutional siamese neural networks. *Proc. of the 16th International Joint Conference on Computer Vision, Imaging and Computer Graphics Theory and Applications (VISAPP). V. 5*, 2021, pp. 157–163. https://doi.org/10.5220/0010263301570163
13. Karmakar S., Banerjee A., Gidde P., Saurav S., Singh S. Convolutional ensembling based few-shot defect detection technique. *Proc. of the Thirteenth Indian Conference on Computer Vision, Graphics and Image Processing (ICVGIP '22)*, 2022, pp. 1–7. https://doi.org/10.1145/3571600.3571607
14. Lv Q., Song Y. Few-shot learning combine attention mechanism-based defect detection in bar surface. *ISIJ International*, 2019, vol. 59, no. 6, pp. 1089–1097. https://doi.org/10.2355/isijinternational.isijint-2018-722
15. Schlagenhauf T., Yildirim F., Brückner B. Siamese basis function networks for data-efficient defect classification in technical domains. *Lecture Notes in Computer Science*, 2023, vol. 13765, pp. 71–92. https://doi.org/10.1007/978-3-031-26236-4_7
16. Nagy A.M., Czúni L. Classification and fast few-shot learning of steel surface defects with randomized network. *Applied Sciences*, 2022, vol. 12, no. 8, pp. 3967. https://doi.org/10.3390/app12083967

Литература

1. Kamiya N., Zhou X., Chen H., Muramatsu C., Hara T., Fujita H. Model-based approach to recognize the rectus abdominis muscle in CT images // IEICE Transactions on Information and Systems. 2013. V. E96D. N 4. P. 869–871. https://doi.org/10.1587/transinf.e96.d.869
2. Bai T., Gao J., Yang J., Yao D. A study on railway surface defects detection based on machine vision // Entropy. 2021. V. 23. N 11. P. 1437. https://doi.org/10.3390/e23111437
3. Saberironaghi A., Ren J., El-Gindy M. Defect detection methods for industrial products using deep learning techniques: a review // Algorithms. 2023. V. 16. N 2. P. 95. https://doi.org/10.3390/a16020095
4. Srividhya R., Shanmugapriya K., SindhuPriya K. Automatic detection of surface defects in industrial materials based on image processing // International Journal of Engineering & Technology. 2018. V. 7. N 3.34. P. 61–64. https://doi.org/10.14419/ijet.v7i3.34.18717
5. Guijo D., Onofre V., Del Bimbo G., Mugel S., Estepa D., De Carlos X., Adell A., Lojo A., Bilbao J., Orus R. Quantum artificial vision for defect detection in manufacturing // arXiv. 2022. arXiv:2208.04988. https://doi.org/10.48550/arXiv.2208.04988
6. Prajwala N.B. Defect detection in pharma pills using image processing // International Journal of Engineering & Technology. 2018. V. 7. N 3.3. P. 102–106. https://doi.org/10.14419/ijet.v7i3.3.14497
7. Liu W., Anguelov D., Erhan D., Szegedy C., Reed S., Fu C.-Y., Berg A.C. SSD: Single shot multibox detector // Lecture Notes in Computer Science. 2016. V. 9905. P. 21–37. https://doi.org/10.1007/978-3-319-46448-0_2
8. Bromley J., Guyon I., LeCun Y., Säckinger E., Shah R. Signature verification using a "Siamese" time delay neural network // Advances in Neural Information Processing Systems. 1993. V. 6.
9. Song K.-C., Hu S., Yan Y. Automatic recognition of surface defects on hot-rolled steel strip using scattering convolution network // Journal of Computational Information Systems. 2014. V. 10. N 7. P. 3049–3055.
10. Kodytek P., Bodzas A., Bilik P. Supporting data for Deep Learning and Machine Vision based approaches for automated wood defect detection and quality control // Zenodo. 2015 [Электронный ресурс]. URL: https://zenodo.org/records/4694695 (дата обращения: 30.10.2024)
11. Bergmann P., Batzner K., Fauser M., Sattlegger D., Steger C. The MVTec anomaly detection dataset: a comprehensive real-world dataset for unsupervised anomaly detection // International Journal of Computer Vision. 2021. V. 129. N 4. P. 1038–1059. https://doi.org/10.1007/s11263-020-01400-4
12. Nagy A.M., Czúni L. Detecting object defects with fusioning convolutional siamese neural networks // Proc. of the 16th International Joint Conference on Computer Vision, Imaging and Computer Graphics Theory and Applications (VISAPP). V. 5. 2021. P. 157–163. https://doi.org/10.5220/0010263301570163
13. Karmakar S., Banerjee A., Gidde P., Saurav S., Singh S. Convolutional ensembling based few-shot defect detection technique // Proc. of the Thirteenth Indian Conference on Computer Vision, Graphics and Image Processing (ICVGIP '22). 2022. P. 1–7. https://doi.org/10.1145/3571600.3571607
14. Lv Q., Song Y. Few-shot learning combine attention mechanism-based defect detection in bar surface // ISIJ International. 2019. V. 59. N 6. P. 1089–1097. https://doi.org/10.2355/isijinternational.isijint-2018-722
15. Schlagenhauf T., Yildirim F., Brückner B. Siamese basis function networks for data-efficient defect classification in technical domains // Lecture Notes in Computer Science. 2023. V. 13765. P. 71–92. https://doi.org/10.1007/978-3-031-26236-4_7
16. Nagy A.M., Czúni L. Classification and fast few-shot learning of steel surface defects with randomized network // Applied Sciences. 2022. V. 12. N 8. P. 3967. https://doi.org/10.3390/app12083967

17. Cao Y., Xu X., Sun C., Cheng Y., Du Z., Gao L., Shen W. Segment any anomaly without training via hybrid prompt regularization. *arXiv*, 2023. arXiv:2305.10724. https://doi.org/10.48550/arXiv.2305.10724

17. Cao Y., Xu X., Sun C., Cheng Y., Du Z., Gao L., Shen W. Segment any anomaly without training via hybrid prompt regularization // arXiv. 2023. arXiv:2305.10724. https://doi.org/10.48550/arXiv.2305.10724

Authors

Nikita P. Novgorodcev — Student, ITMO University, Saint Petersburg, 197101, Russian Federation, https://orcid.org/0009-0000-5352-1852, novgorodcevnp@gmail.com

Kseniia A. Baturina — Student, Laboratory Assistant, ITMO University, Saint Petersburg, 197101, Russian Federation, https://orcid.org/0009-0003-7555-282X, kseniya.baturina@mail.ru

Valeria A. Efimova — PhD, Assistant, ITMO University, Saint Petersburg, 197101, Russian Federation, sc 57207459404, https://orcid.org/0000-0002-5309-2207, valeryefimova@gmail.com

Anatoly A. Shalyto — D.Sc., Full Professor, Chief Researcher, ITMO University, Saint Petersburg, 197101, Russian Federation, sc 56131789500, https://orcid.org/0000-0002-2723-2077, shalyto@mail.ifmo.ru

Авторы

Новгородцев Никита Павлович — студент, Университет ИТМО, Санкт-Петербург, 197101, Российская Федерация, https://orcid.org/0009-0000-5352-1852, novgorodcevnp@gmail.com

Батурина Ксения Александровна — студент, лаборант, Университет ИТМО, Санкт-Петербург, 197101, Российская Федерация, https://orcid.org/0009-0003-7555-282X, kseniya.baturina@mail.ru

Ефимова Валерия Александровна — кандидат технических наук, ассистент, Университет ИТМО, Санкт-Петербург, 197101, Российская Федерация, sc 57207459404, https://orcid.org/0000-0002-5309-2207, valeryefimova@gmail.com

Шалыто Анатолий Абрамович — доктор технических наук, профессор, главный научный сотрудник, профессор, Университет ИТМО, Санкт-Петербург, 197101, Российская Федерация, sc 56131789500, https://orcid.org/0000-0002-2723-2077, shalyto@mail.ifmo.ru

Received 27.04.2024
Approved after reviewing 02.10.2024
Accepted 15.11.2024

Статья поступила в редакцию 27.04.2024
Одобрена после рецензирования 02.10.2024
Принята к печати 15.11.2024