Review

A Survey of Image Segmentation for Industrial Applications with a Focus on Quality Control

Ramona Kühlechner *

Posted Date: 8 April 2026

doi: 10.20944/preprints202604.0507.v1

Keywords: image segmentation; quality control; manufacturing; industry

 doi:10.20944/preprints202604.0507.v1

Review

A Survey of Image Segmentation for Industrial Applications with a Focus on Quality Control

Ramona Kühlechner

Independent Researcher, Austria; ramona.kuehlechner2@gmail.com

Abstract

Precise segmentation of defects is a key component of industrial quality control. This paper presents a comprehensive overview of contemporary methods utilising convolutional neural networks that have demonstrated practical efficacy. Depending on the application, semantic, instance-based, panoptic and hybrid segmentation methods are used to reliably detect material defects. Finally, prospects for industrial use are discussed, including the optimisation of hybrid methods, real-time capability and integration into existing production processes to ensure efficient, robust and practical defect detection.

Keywords: image segmentation; quality control; manufacturing; industry

1. Introduction

In modern industrial manufacturing, ensuring consistently high product quality is a key success factor. Quality control (QC) [1] and quality assurance (QA) [1] are necessary to detect production errors at an early stage, avoid rejects and meet customer requirements and legal specifications. Traditional manual inspection methods, such as visual inspections or measurements, reach their limits, especially in highly automated production environments. These methods are considered time-consuming, subjective and difficult to scale.

Automated processes are therefore becoming increasingly important. With the advance of digitalisation, computer vision is establishing itself as a key technology in industrial quality assurance. Camera systems and image processing algorithms enable the automatic inspection of products during or after production and reduce the need for manual inspection.

In particular, image segmentation methods, which analyse and classify image areas with pixel precision, offer great potential for the automated detection of manufacturing defects [2]. These include surface defects, dimensional deviations, cracks or breaks. Contemporary deep learning methodologies, especially convolutional neural networks (CNNs), have proven to be formidable instruments in this domain, since they adeptly extract picture information, exhibit resilience to interference, and can be readily tailored to various industrial contexts. Different types of segmentation enable precise detection of defects and reduce the amount of manual annotation required, in some cases through the use of synthetic training data.

Despite the increasing prevalence of deep learning-based segmentation methods, challenges remain in the areas of real-time capability, adaptability to different product variants, and minimisation of false alarms, especially in safety-critical scenarios. However, modern methods show that highly accurate and flexible segmentation is increasingly feasible even under demanding production conditions.

The aim of this thesis is to summarise current state-of-the-art (SOTA) methods of image segmentation for industrial quality assurance, evaluate their advantages and disadvantages, and compare the most relevant approaches based on defined criteria and typical application scenarios. Particular attention is paid to segmentation types, training strategies, and practical applications in industrial production processes.

 doi:10.20944/preprints202604.0507.v1

2. Fundamentals of Image Segmentation

Image segmentation is a fundamental technique in computer vision, facilitating comprehensive analysis of visual data. Unlike pure object recognition, which solely identifies the existence and location of objects, segmentation offers pixel-level accuracy in defining structures within an image. This makes it an indispensable tool for industrial quality control systems, where even the smallest defects or material deviations must be reliably detected and localised. Depending on the objective, a distinction is made between semantic segmentation [3], instance segmentation [4] and panoptic segmentation [5]. The effectiveness of modern segmentation methods today relies heavily on deep learning architectures supported by large data sets and powerful frameworks. The following section therefore begins by presenting the most important software ecosystems that enable the efficient development and application of segmentation methods, before going on to describe key data sets that serve as the basis for training and evaluation.

2.1. Frameworks

Many validated open-source frameworks and specialised tools facilitate the deployment and effective use of image segmentation methods. They provide significant support in the creation, training and inference of segmentation models, thereby contributing to a considerable simplification of the development process. Well-maintained and versatile software ecosystems are particularly important in industry, where robustness, efficiency and adaptability to specific requirements are crucial.

PyTorch [6] and TensorFlow [7] are among the most important deep learning frameworks. Both offer extensive libraries for the development and training of neural networks that support both classic CNN architectures and modern transformer-based models [8]. Thanks to its dynamic computation graphs, PyTorch is particularly popular for research and experimental developments, as it offers very intuitive and flexible modelling options. TensorFlow, on the other hand, has advantages for industrial real-time applications, as it offers a stable production environment and enables the efficient execution of models on various hardware platforms (GPU, TPU, edge devices).

The Detectron2 framework [9], created by Facebook AI Research, is regarded as a premier tool in instance and semantic segmentation. It provides a modular and powerful platform on which different segmentation methods can be implemented. It impresses with its high flexibility, extensive documentation and numerous pre-trained models that can be used for domain-specific adaptations. It is used in particular for tasks that place high demands on segmentation accuracy or involve complex image content.

Another established tool is the Segment-Anything-Model (SAM) architecture from Meta AI [10]. This framework aims to provide a universal segmentation model that has been trained on a large amount of diverse image data. It enables both precise segmentation of individual objects and rapid adaptation to new application areas using prompting mechanisms.

In addition, there are specialised frameworks and tools for specific application areas of image segmentation, such as DeepMask and SharpMask [11,12]. These were developed by Facebook Research as Torch implementations for object detection and segmentation. MultiPath Network is also a Torch implementation of a network for object detection and segmentation based on the work A MultiPath Network for Object Detection [13].

OpenCV [14] is an important preprocessing tool for image data and an interface between camera systems and deep learning models. It offers basic and advanced image processing operations such as scaling and filtering. In addition, OpenCV can be easily integrated into existing industrial image processing systems, enabling efficient processing of image data for downstream segmentation models.

2.2. Datasets

In an industrial context, where specific requirements and error types must be taken into account, the accessibility and quality of datasets are crucial for the development and evaluation of image segmentation models. Industrial image datasets often contain a variety of defect types, different

 doi:10.20944/preprints202604.0507.v1

materials and varying image conditions. This results in special requirements for the robustness and generalisability of models.

In addition to such industry-specific datasets, generally available segmentation datasets are also used, especially in the pre-training phases of transfer learning [15]. Here, robust feature extractors are first trained on large, broadly annotated datasets before fine-tuning with domain-specific data. This significantly reduces the training effort and improves recognition performance, even when only a small amount of industrial data is available.

One of the most widely used general datasets is COCO (Common Objects in Context) [16], which, in addition to object recognition and keypoint detection, also provides extensive instance segmentation with more than 330,000 images and 80 categories. Another classic in computer vision is Pascal VOC [17], which is used for classification, object recognition and, in particular, semantic segmentation and comprises around 20,000 images with 20 classes. For scenes in urban areas, the Cityscapes dataset [18] plays a central role, as it contains both semantic and instance-based segmentations of street scenes, with 5,000 finely annotated images and 19 classes. LVIS (Large Vocabulary Instance Segmentation) [19] is an extension of COCO that contains significantly more classes and is particularly suitable for fine-grained instance segmentation.

2.3. Quality Control and Assurance in Industry

Industrial quality assurance [1] encompasses all organised and methodical measures necessary to ensure that products and manufacturing processes meet specified quality standards. These include both testing and preventive measures that can take place during production (inline testing) [20,21] or after completion of the manufacturing process (end-of-line testing) [22]. The aim is to detect defective products at an early stage, reduce waste and minimise the costs of rework or complaints. These measures ensure product quality, improve processes and increase customer satisfaction.

Automated inspection systems increasingly use technologies such as AI-based image processing or anomaly detection. Such systems can detect surface defects, dimensional deviations, missing components and other quality deviations with high precision and speed. However, conventional rule-based image processing systems often reach their limits when dealing with complex or variable visual patterns.

To overcome these limitations, companies are increasingly turning to segmentation-based models based on deep learning. These models offer greater flexibility and accuracy, especially when detecting subtle defects or deviations that are difficult to define using fixed rules. The following sections provide an overview of such models, their types and specific objectives in the context of industrial inspections.

The aforementioned principles, encompassing frameworks and pertinent data sets, constitute the foundation for comprehending the practical implementation of image segmentation models. Various segmentation-based deep learning models have been created to efficiently exploit these ideas in industrial applications. These differ in architecture, processing speed and detection accuracy and are crucial for overcoming specific challenges in automated quality assurance. The following section presents and compares the model types used in modern industrial applications.

2.4. Traditional Methods

Conventional image segmentation methods rely on traditional image processing techniques and statistical analysis. They do not usually require large amounts of data or extensive training processes like modern deep learning approaches, but instead use direct properties of image data such as intensity, colour or spatial relationships. One of the most important classic segmentation methods is presented below.

 doi:10.20944/preprints202604.0507.v1

Thresholding

Thresholding [23] is a fundamental and widely employed technique for image segmentation. It entails establishing one or more threshold values T to categorise picture pixels $I(x,y)$ into distinct classes, such as foreground and background. Simple global thresholding can be described mathematically as:

$$S(x,y) = \begin{cases} 1 & \text{if } I(x,y) \geq T, \\ 0 & \text{if } I(x,y) < T. \end{cases} \tag{1}$$

Local or adaptive thresholding methods take regional image differences into account and define the threshold value $T(x,y)$ depending on the pixel's environment, e.g. by means of averaging or Gaussian weighting within a window:

$$S(x,y) = \begin{cases} 1 & \text{if } I(x,y) \geq T(x,y), \\ 0 & \text{if } I(x,y) < T(x,y). \end{cases} \tag{2}$$

Region Growing

Region Growing [24] is an iterative segmentation-based approach. Starting from a seed point s_i, neighbouring pixels p are added to region R_i if a similarity criterion $f(p,R_i)$ is satisfied, e.g. intensity difference:

$$f(p,R_i) = |I(p) - \mu_{R_i}| < \epsilon \tag{3}$$

where μ_{R_i} is the mean intensity of region R_i and ϵ is a threshold value. The region grows until no further pixels satisfy the condition.

Clustering

Clustering methods segment the image by grouping pixels into clusters C_k based on a feature vector $\mathbf{x}_j$ (e.g., intensity, colour, or texture). In K-means [25], the sum of the squared distances is minimised:

$$\min_{C_1,\ldots,C_K} \sum_{k=1}^{K} \sum_{\mathbf{x}_j \in C_k} \|\mathbf{x}_j - \boldsymbol{\mu}_k\|^2 \tag{4}$$

where $\boldsymbol{\mu}_k$ is the mean value of the features in cluster C_k.

Fuzzy C-Means [26] generalises this by assigning each pixel a membership function $u_{jk} \in [0,1]$, and the cost function is:

$$J_m = \sum_{j=1}^{N} \sum_{k=1}^{K} u_{jk}^m \|\mathbf{x}_j - \boldsymbol{\mu}_k\|^2, \tag{5}$$

with the fuzzification parameter $m > 1$.

Watershed Algorithm

The watershed algorithm [27] interprets the image intensities as topographical relief $I(x,y)$. Catchment areas W_i are defined by letting 'water' flow into the landscape from the local minimum. The pixel assignment can be formally described by minimising the distance to the nearest minima:

$$W_i = \{(x,y) \mid \text{flowed from } (x,y) \text{ to minimum } m_i\}. \tag{6}$$

The segmentation boundaries arise at the pixels where catchment areas meet. Marker-based variants additionally use manually or automatically defined starting points M_i to control the region expansion.

 doi:10.20944/preprints202604.0507.v1

3. Types of Image Segmentation

Image segmentation is a crucial procedure in the domains of computer vision and machine learning. The objective of segmentation is to partition an image $I : \Omega \subset \mathbb{R}^2 \to \mathbb{R}^c$ into significant parts $\{R_k\}_{k=1}^K$ such that each region conveys pertinent information and facilitates a streamlined representation of the picture.

Formally, this can be described by partitioning the image domain Ω:

$$\Omega = \bigcup_{k=1}^{K} R_k, \quad R_i \cap R_j = \emptyset \text{ for } i \neq j. \tag{7}$$

3.1. Semantic Segmentation

Semantic segmentation [3] assigns a semantic class $c \in \{1, \ldots, C\}$ to each pixel $p \in \Omega$. This can be formally described as a classification function $f : \Omega \to \{1, \ldots, C\}$:

$$f(p) = c, \quad \forall p \in \Omega. \tag{8}$$

For an image with multiple objects of the same class, all pixels p_i are assigned to the same class c, regardless of which instance they belong to:

$$f(p_i) = f(p_j) = c \quad \text{if } p_i, p_j \text{ belong to objects of the same class.} \tag{9}$$

In neural networks, the function f is approximated by a parametrised model f_θ that predicts the class membership of each pixel based on its features x_p:

$$\hat{c}_p = f_\theta(x_p), \quad x_p \in \mathbb{R}^d \tag{10}$$

where $\hat{c}_p$ is the predicted class for pixel p. Training is typically performed by minimising a loss function L, e.g. cross-entropy:

$$\mathcal{L}(\theta) = -\sum_{p \in \Omega} \sum_{c=1}^{C} y_{p,c} \log \hat{y}_{p,c}, \tag{11}$$

where $y_{p,c}$ is the ground truth assignment and $\hat{y}_{p,c} = \mathbb{P}(\hat{c}_p = c)$ is the predicted probability.

A summary of notable semantic segmentation methods is presented in 1.

Fully Convolutional Networks (FCNs)

Fully Convolutional Networks (FCNs) [28] were the inaugural deep learning architectures composed exclusively of convolutional layers and tailored for semantic segmentation. Rather of employing fully connected layers at the conclusion of a traditional CNN, FCNs utilise transposed convolutions [28] to restore spatial resolutions. The model integrates semantic and geographical information by amalgamating characteristics from various depth levels.

U-Net

The U-Net [29] was initially designed for biomedical image segmentation and is especially effective in situations with constrained training data. The architecture comprises a contractive pathway for context acquisition and a symmetric expanding pathway that facilitates accurate localisation. A defining aspect is the skip connections between contractive and expanding layers, which retain precise information.

Efficient Neural Network (ENet)

In 2016, Paszke et al. presented ENet [30], a notably resource-efficient neural network designed for semantic segmentation. Unlike many previously developed architectures, ENet focused on real-time capability and efficiency, so that the model can also be used on devices with limited computing

 doi:10.20944/preprints202604.0507.v1

power, such as mobile platforms or embedded systems. The architecture employs an encoder-decoder framework while utilising considerably fewer parameters than similar approaches. This is achieved, among other things, through the use of asymmetric and factorised convolutions, early downsampling steps and bottleneck modules. Despite its low complexity, ENet delivers robust results in semantic segmentation, making it particularly attractive for time-critical applications such as autonomous driving.

V-Net

In 2016, Milletari et al. [31] presented V-Net, an architectural framework founded on three-dimensional convolutional neural networks, designed explicitly for medical segmentation.In contrast to two-dimensional approaches, V-Net processes volumetric data directly in 3D, which allows spatial context information to be captured more effectively. The architecture is structured as an encoder-decoder network, with a symmetrical design with skip links that mitigate information loss during depth reduction. Another significant innovation of V-Net was the introduction of the Dice Loss function, which targets the segmentation result directly and is particularly suitable for datasets with highly unbalanced class distributions. This concept contributed significantly to improving the performance of neural networks in medical segmentation and served as the basis for many subsequent architectures.

Efficient Residual Factorized Network (ERFNet)

In 2017, Romera et al. [32] presented ERFNet (Efficient Residual Factorized Network), an architecture for semantic segmentation that is specifically designed for efficient computation and real-time applications.ERFNet employs residual and factorised convolutional blocks to diminish model complexity while preserving the capacity to extract profound features. This combination enables the network to attain an optimal equilibrium between precision and processing resources, rendering it especially appealing for applications like autonomous driving or embedded devices. The architecture shows that targeted simplification and optimisation of convolution blocks can achieve high segmentation performance even with limited hardware resources.

SegNet

SegNet [33] is another encoder-decoder network designed specifically for semantic segmentation tasks. Its distinctive feature is the use of max pooling indices from the encoder to efficiently reconstruct the feature maps in the decoder. This not only saves memory but also improves segmentation accuracy, as important structural information is retained during upsampling.

Pyramid Scene Parsing Network (PSPNet)

In 2017, Zhao et al. [34] presented the Pyramid Scene Parsing Network (PSPNet), an architecture that adeptly leverages global scene context for semantic segmentation. The Pyramid Pooling module is a fundamental element of the model, aggregating characteristics at several scales to consider both local details and global contextual information. This multi-level pooling structure enables PSPNet to better understand complex scenes and reliably segment objects of different sizes and distances. The architecture is based on a deep ResNet backbone for feature extraction, complemented by the Pyramid Pooling strategy, and delivers state-of-the-art results on many semantic segmentation benchmark datasets.

DeepLab

In 2017, Chen et al. [35] introduced DeepLab v1, an architecture for semantic image segmentation that attracted attention primarily through the use of dilated convolutions. This technique allows for the expansion of the filters' receptive field without compromising the spatial precision of the feature maps, facilitating precise segmentation in high-resolution images. DeepLab employs an encoder-decoder methodology, wherein the encoder derives profound semantic characteristics and the decoder reconstructs the segmentation to the original image resolution. Employing this strategy, DeepLab attained notable enhancements in the precision of segmentation outcomes, particularly for objects

 doi:10.20944/preprints202604.0507.v1

of varying dimensions. The release of v1 laid the foundation for further versions, which introduced additional improvements such as Atrous Spatial Pyramid Pooling (ASPP) [35] and more effective decoder modules.

Image Cascade Network (ICNet)

In 2018, Zhao et al. [36] introduced the Image Cascade Network (ICNet), an architecture for semantic segmentation optimised for real-time, high-resolution applications. The model follows a multi-stage cascade approach in which the input images are processed at different resolutions. Through the astute integration of these features, ICNet attains elevated segmentation precision while preserving a minimal computing burden. The architecture is especially appropriate for applications like autonomous driving or video processing, where real-time performance is essential for handling big image sizes.

Bilateral Segmentation Network (BiSeNet)

In 2018, Yu et al. [37] introduced BiSeNet (Bilateral Segmentation Network), an architecture for semantic segmentation that enables high accuracy in real-time applications. The model combines two separate paths. One is the spatial path, which maintains spatial details, and the other is a context path, which extracts deep semantic information. This bilateral structure allows both fine edges and global context to be used efficiently. BiSeNet is characterised by low latency and efficient computation, making it particularly suitable for applications such as autonomous driving or mobile systems that require fast and accurate segmentation.

Attention U-Net

In 2022, Zhu et al. [38] introduced Attention U-Net, a further development of the classic U-Net that is specially optimised for medical image segmentation. The central innovation consists of attention modules that are integrated into the skip connections of the encoder-decoder network. Those components allow the network to concentrate on relevant features and diminish extraneous information, thus enhancing segmentation precision, particularly for intricate or noisy images. By combining U-Net's proven symmetric architecture with attention-based mechanisms, Attention U-Net can achieve more accurate segmentations without significantly increasing network complexity, making it particularly suitable for demanding medical applications.

SEgmentation TRansformer (SETR)

SETR [39] is a pioneer among pure Transformer-based segmentation models. It substitutes the conventional CNN encoders with a purely Vision Transformer architecture that treats image patches as tokens, akin to natural language processing. The decoder architecture subsequently reconstructs the segmentation masks at high resolution. This architecture allows SETR to attain excellent performance, particularly for large-scale structures.

Segmenter

In 2021, Strudel et al. [40] introduced Segmenter, a model that applies the concepts of Vision Transformers (ViT) [8] to semantic segmentation. Instead of classic convolutional neural networks, Segmenter uses patch-based Transformer encoders that effectively model the global dependencies in the image. A special decoder reconstructs the complete segmentation map from the patch embeddings. Thanks to its Transformer-based architecture, Segmenter can utilise long-range contextual information, which leads to more accurate segmentations, especially in complex scenes. The model shows that combining transformer mechanisms with segmentation tasks can achieve SOTA results, especially in scenarios where global image information is crucial.

SegFormer

SegFormer [41] is a modern model for semantic image segmentation that combines the advantages of transformer architectures with efficient feature extraction mechanisms. SegFormer eliminates traditional positional encodings and employs a hierarchical transformer encoder architecture that

 doi:10.20944/preprints202604.0507.v1

captures information across various scales. This encompasses both global contextual information and intricate characteristics of the objects, resulting in enhanced segmentation accuracy, particularly in complicated scenarios. The model is distinguished by its low computing complexity and considerable adaptability, rendering it appropriate for diverse applications, including autonomous driving and medical picture processing.

Table 1. Overview: Semantic Segmentation Models.

Model	Year	Description
FCN (32s,16s,8s) [28]	2015	First fully convolutional network for pixel-wise segmentation
U-Net [29]	2015	Encoder-decoder with skip connections, biomedical focus
ENet [30]	2016	Lightweight real-time network
V-Net [31]	2016	3D extension of U-Net for volumetric data
ERFNet [32]	2017	Efficient residual factorized convs
SegNet [33]	2017	Encoder-decoder with pooling indices for efficient upsampling
PSPNet [34]	2017	Pyramid scene parsing, global context pooling
DeepLab v1 [35]	2017	Atrous convolutions, multi-scale context
ICNet [36]	2018	Cascade for real-time semantic segmentation
BiSeNet [37]	2018	Bilateral path for speed + accuracy balance
DeepLab v2–v4	2018–2020	Improved ASPP + encoder-decoder refinement
SETR [39]	2021	First pure Vision Transformer for segmentation
Segmenter [40]	2021	Transformer encoder + lightweight decoder
SegFormer [41]	2021	CNN-Transformer hybrid, efficient
Attention U-Net [38]	2022	U-Net with attention gates for better localization

3.2. Instance Segmentation

Unlike semantic segmentation, instance segmentation differentiates not only between semantic categories but also recognises specific object instances within each category [4]. Formally, let $I : \Omega \subset \mathbb{R}^2 \to \mathbb{R}^c$ be an image, and let $\mathcal{C} = \{1, \ldots, C\}$ be the set of classes. In instance segmentation, each pixel $p \in \Omega$ is assigned a pair (c, i):

$$f(p) = (c_p, i_p), \quad c_p \in \mathcal{C}, \quad i_p \in \mathbb{N}, \tag{12}$$

where c_p represents the semantic class and i_p represents the instance number within that class. This means that objects of the same class but different instances are assigned different masks:

$$\{R_{c,i}\}_{i=1}^{N_c}, \quad R_{c,i} = \{p \in \Omega \mid f(p) = (c, i)\}, \tag{13}$$

where N_c is the number of instances of class c.

Modern neural network architectures, such as *Mask R-CNN*, approximate the function f_θ parameterised by θ and combine object localisation (bounding boxes) with precise pixel mapping:

$$\hat{f}_\theta(p) = (\hat{c}_p, \hat{i}_p), \quad \forall p \in \Omega. \tag{14}$$

The training loss function takes into account both the classification of the objects (L_{cls}) and the segmentation accuracy of the mask (L_{mask}):

$$\mathcal{L}(\theta) = L_{\text{cls}} + L_{\text{box}} + L_{\text{mask}}. \tag{15}$$

Instance segmentation facilitates the accurate analysis and monitoring of distinct items, even when they are part of the same category. A summary of notable instance segmentation methodologies is presented in 2.

R-CNN

 doi:10.20944/preprints202604.0507.v1

R-CNN was developed by Girshick et al. [42] and represents one of the first approaches that proposes regions in the image and then classifies them with a CNN. The method first extracts region proposals using an external algorithm such as Selective Search and then processes each region individually through a CNN to assign classes and localise objects. R-CNN laid the foundation for modern object detection and segmentation models, but was very computationally intensive due to the sequential processing of regions.

Fast R-CNN

Fast R-CNN [43] improved the original R-CNN architecture by applying feature extraction to the entire image and then using regions-of-interest (RoI) pooling to extract features for each region. This drastically reduces the computational effort while enabling end-to-end training. Fast R-CNN thus significantly accelerated detection without sacrificing accuracy.

Sequential Grouping Networks (SGN)

Liu et al. [44] introduced Sequential Grouping Networks (SGN), an architecture that detects objects by sequentially grouping segmentation proposals. SGN systematically processes the proposed segments and consolidates pertinent regions to generate the final instance segmentation. This method enhances accuracy for significantly overlapped or fragmented entities.

Mask - R-CNN

Mask R-CNN extends Faster R-CNN with an additional mask prediction branch [45], which creates a precise segmentation mask for each region. This architecture thus enables instance segmentation, i.e. the separation of individual objects within the same class. Mask R-CNN continues to use RoIAlign to increase the accuracy of the masks and combines detection and segmentation in an end-to-end trainable framework.

PANet

The Path Aggregation Network (PANet) [46] was originally developed as an extension to Mask R-CNN to increase the accuracy of instance segmentation. PANet adds a bottom-up path to the Feature Pyramid Network (FPN) structure, which efficiently aggregates finer details from deeper network layers. In addition, an adaptive feature pooling method is introduced, which enables better mask prediction across different object sizes.

MaskLab

Chen et al. [47] developed MaskLab, a framework for instance segmentation that combines classification, semantics and mask formation. MaskLab uses both region-based features and contextual information to precisely align masks with object boundaries. MaskLab demonstrates high accuracy, especially in complex scenes with overlaps.

Cascade Mask R-CNN

Cascade Mask R-CNN [48] is an advancement of Mask R-CNN that addresses the problem of accuracy degradation when processing difficult objects. It connects several stages of detectors and mask predictions in series, with each stage building on the outputs of the previous stage. This cascading structure improves precision in both object detection and instance segmentation, as errors from earlier stages can be corrected in later stages.

Hybrid Task Cascade (HTC)

Chen et al. [49] introduced Hybrid Task Cascade (HTC) in 2019, which extends the advantages of Cascade Mask R-CNN by more closely integrating detection and segmentation tasks. HTC combines multiple cascaded stages and uses cross-task feature fusion, allowing semantic information from mask segmentation to flow back into the detection pipeline. This further increases both object detection and segmentation accuracy. HTC is regarded as one of the most robust systems for instance segmentation on intricate datasets.

 doi:10.20944/preprints202604.0507.v1

You Only Look At CoefficienTs (YOLACT)

YOLOACT (You Only Look At CoefficienTs) [50] is a real-time instance segmentation model that offers significantly higher speed than previous models such as Mask R-CNN, while maintaining comparable accuracy. The model separates the generation of proto masks and the prediction of mask coefficients, enabling parallel processing. The final masks are then generated by a weighted combination of both components.

Tensormask

In 2019, Chen et al. [51] introduced TensorMask, a model for instance segmentation that treats segmentation as a dense prediction of masks over a grid of tensors. Unlike classical Mask R-CNN-based approaches, which generate masks per region, TensorMask generates masks for the entire image in a single pass, directly mapping the spatial structure of the objects into the tensors. This enables more accurate and consistent mask generation, especially for closely spaced or overlapping objects. TensorMask thus combines the advantages of dense predictions with modern deep learning architectures for instance segmentation.

Segmenting Objects by Locations (SOLO)

SOLO (Segmenting Objects by Locations) [52] is an instance segmentation approach based on an anchor-free concept. The model divides an image into a regular grid, with each grid point tasked with generating a mask for an object in its respective region. Unlike classical methods, SOLO does not require separate region proposals or RoI operations, making the architecture simpler and more efficient while achieving precise segmentation of individual objects.

Segmenting Objects by Locations (SOLOv2)

SOLOv2 (Segmenting Objects by Locations) is a further development of SOLO and pursues an anchor-free approach to instance segmentation [53]. The model divides the image into a regular grid and assigns each grid point the task of generating a mask for an object within its region. SOLOv2 improves both accuracy and efficiency over its predecessor, including through more dynamic feature assignments and improved mask prediction.

Conditional Instance Segmentation (CondInst)

In 2020, Tian et al. [54] introduced CondInst (Conditional Instance Segmentation), a framework for instance segmentation that generates masks directly through conditional convolution filters that are generated individually for each object. This eliminates the need for classic RoI-based processing, enabling more efficient and flexible mask predictions. CondInst can generate masks for any number of objects simultaneously and demonstrates high accuracy while reducing computation time.

DEtection TRansformer (DETR)

Carion et al. [55] developed DETR (DEtection TRansformer), which formulates object detection as an end-to-end transformer problem. DETR replaces classical region proposal methods with attention-based global image representations, enabling precise detection of objects in complex scenes. This model paved the way for transformer-based segmentation approaches.

Deformable DETR

Zhu et al. [56] introduced Deformable DETR, an evolution of DETR that addresses the long convergence time and computational costs of the original model. Deformable DETR replaces the standard self-attention mechanisms with deformable attention, which focuses on a limited set of relevant key positions. This allows the model to train faster while delivering accurate object detection and segmentation in high-resolution images.

Conditional DETR

Chen et al. [57] and Meng et al. [58] extended DETR with Conditional DETR, which improves convergence speed and enables more stable training processes. By conditioning on object prior

 doi:10.20944/preprints202604.0507.v1

information, the model can learn faster and more accurately, especially in complex scenes and with a large number of objects. The v2 version incorporates more optimisations and attains superior performance in several object detection and segmentation benchmarks.

Table 2. Overview: Instance Segmentation Models.

Model	Year	Description
R-CNN [42]	2014	Region proposals + CNN classification
Fast R-CNN [43]	2015	Faster training with ROI pooling
Faster R-CNN [59]	2015	Introduced RPN for detection backbone
SGN [44]	2017	Sequential grouping of pixels into instances
Mask R-CNN [45]	2017	Adds mask head for pixel-wise instance masks
PANet [46]	2018	Improves Mask R-CNN with bottom-up path
MaskLab [47]	2018	Combines semantic + detection features
Cascade Mask R-CNN [48]	2018	Multi-stage refinement for robust masks
HTC [49]	2019	Joint box and mask optimization cascade
YOLACT [50]	2019	Real-time instance segmentation with prototypes
TensorMask [51]	2019	Dense sliding-window instance masks
SOLO [52]	2019	Anchor-free instance segmentation
SOLOv2 [53]	2020	Improved SOLO with dynamic assignment
CondInst [54]	2020	Dynamic filters for instance-specific masks
DETR [55]	2020	Transformer for detection, extended to masks
Deformable DETR [56]	2020	Deformable attention for faster convergence and high-res images
Conditional DETR [57,58]	2021	Improved DETR convergence and mask quality

3.3. Panoptic Segmentation

Panoptic segmentation [5] integrates the principles of semantic and instance segmentation. Formally, let $I : \Omega \subset \mathbb{R}^2 \to \mathbb{R}^c$ be an image and $\mathcal{C} = \{1, \ldots, C\}$ the set of classes. For each pixel $p \in \Omega$, a pair (c_p, i_p) is assigned:

$$f(p) = (c_p, i_p), \tag{16}$$

where

$$i_p = \begin{cases} 0, & \text{if } c_p \in \text{stuff classes}, \\ 1, 2, \ldots, N_{c_p}, & \text{if } c_p \in \text{things classes (instances)}. \end{cases} \tag{17}$$

This simultaneously segments and classifies things (objects with instances) and stuff (flat regions without instances). The regions for instances and semantic classes can be written as:

$$R_{c,i} = \{p \in \Omega \mid f(p) = (c, i)\}, \quad c \in \mathcal{C}, \quad i \in \mathbb{N}_0. \tag{18}$$

Neural networks for panoptic segmentation, such as EfficientPS, approximate the function f_θ parameterised by θ:

$$\hat{f}_\theta(p) = (\hat{c}_p, \hat{i}_p), \quad \forall p \in \Omega. \tag{19}$$

The training loss function integrates both semantic and instantiation aspects:

$$\mathcal{L}(\theta) = L_{\text{sem}} + L_{\text{inst}} + L_{\text{consistency}}, \tag{20}$$

where L_{sem} evaluates the semantic classification, L_{inst} evaluates the instance masks, and $L_{\text{consistency}}$ evaluates the consistency between the two components.

 doi:10.20944/preprints202604.0507.v1

Panoptic segmentation thus enables a holistic and consistent interpretation of image content in complex scenes. An overview of representative panoptic approaches is provided in 3.

Unified Panoptic Segmentation Network (UPSNet)

Xiong et al. [60] presented UPSNet (Unified Panoptic Segmentation Network), a methodology for panoptic segmentation that integrates semantic and instance information into a cohesive framework. UPSNet incorporates distinct branches for semantic and instance segmentation, with a unified loss function to consolidate the outcomes. This facilitates uniform panoptic segmentation while preserving elevated efficiency.

Adaptive Instance Selection Network (AdaptIS)

Sofiiuk et al. [61] created AdaptIS, a framework for instance segmentation employing adaptive instance-aware convolutions. AdaptIS generates masks for individual objects without relying on RoI pooling, making the model more flexible and efficient with objects of different sizes.

Efficient Panoptic Segmentation Network (EPSNet)

Chang et al. [62] presented EPSNet (Efficient Panoptic Segmentation Network), which executes panoptic segmentation in a resource-efficient and effective manner. EPSNet uses a unified backbone architecture for semantic and instantiation branches and performs optimised feature fusion. This allows EPSNet to achieve high accuracy with lower computational effort, making it particularly interesting for mobile and embedded systems.

Fast Panoptic Segmentation Network (FPSNet)

De et al. [63] developed FPSNet (Fast Panoptic Segmentation Network), a framework for panoptic segmentation that is optimised for efficient processing and fast inference. FPSNet integrates semantic and instance data into a streamlined network, yielding robust panoptic segmentation outcomes while substantially decreasing calculation times, thereby rendering it appropriate for real-time applications.

Panoptic-Deeplab

Cheng et al. [64] introduced Panoptic-DeepLab, which combines the advantages of DeepLab-based semantic segmentation with instance segmentation. The model employs atrous Spatial Pyramid Pooling (ASPP) [64] and decoder-based feature fusion to effectively handle both semantic and instance data. Panoptic-DeepLab produces strong outcomes in panoptic segmentation tasks.

Efficient Panoptic Segmentation (Efficientps)

Mohan et al. [65] developed EfficientPS, an efficient framework for panoptic segmentation. It combines a lean backbone with separate branches for semantics, instances and panoptic fusion, achieving high accuracy with low computational load, which makes it particularly attractive for real-time applications.

Mask2Former

Cheng et al. [66] presented Mask2Former, a cohesive transformer-based model that facilitates semantic, instance, and panoptic segmentation. MaskFormer conceptualises segmentation as a mask classification challenge, wherein each pixel is designated to a particular mask class. This architecture combines transformer-based global context modelling with flexible mask prediction and delivers SOTA performance on multiple segmentation benchmarks.

Mask DINO

Li et al. [67] introduced Mask DINO, an advanced framework for instance and panoptic segmentation based on transformer architectures. Mask DINO extends the DINO detection pipeline with mask prediction, where each pixel is precisely assigned to an instance mask. By combining self-attentive mechanisms and dynamic mask classification, Mask DINO achieves high accuracy and robustness, especially in complex scenes with many overlapping objects. The model is considered a state-of-the-art solution for modern segmentation tasks.

 doi:10.20944/preprints202604.0507.v1

Video Panoptic Segmentation Network (VPSNet)

Wen et al. [68] introduced VPSNet (Video Panoptic Segmentation Network), a model explicitly engineered for panoptic segmentation in video content. VPSNet enhances traditional panoptic designs by incorporating temporal consistency and object tracking across frames, facilitating stable and coherent panoptic segmentation in video sequences. VPSNet is very advantageous for autonomous systems and video analysis.

Table 3. Overview: Panoptic Segmentation Models.

Model	Year	Description
Panoptic FPN [5]	2019	FPN with semantic + instance heads
UPSNet [60]	2019	Unified panoptic segmentation network
AdaptIS [61]	2019	Pixel-wise instance parameter regression
FPSNet [63]	2020	Lightweight fast panoptic segmentation
EPSNet [62]	2020	Efficient panoptic segmentation with unified backbone
Panoptic-DeepLab [64]	2020	Bottom-up approach for things + stuff
EfficientPS [65]	2021	Efficient panoptic segmentation CNN
Mask2Former [66]	2022	Transformer with masked attention
Mask DINO [67]	2023	Extends Mask2Former + DETR
VPSNet [68]	2025	Video panoptic segmentation network

3.4. Hybrid and Combined Models

These models cannot be strictly assigned to a single category, as they combine different approaches. The following section therefore presents methods that are neither exclusively semantic, instance-based nor panoptic. Particular attention is paid to models that combine different paradigms, for example by linking object recognition and segmentation, CNNs and transformers, prompt-based methods or generative approaches. An overview of representative hybrid approaches is provided in 4.

Diffusion Network (Difnet)

DifNet is a diffusion-based model for image segmentation that aims to refine the boundaries of objects through iterative information propagation [69]. Initial segmentation estimates are diffused and adjusted over several steps, which improves segmentation accuracy, especially at object edges. DifNet is characterised by its ability to efficiently integrate both global context information and local details.

SEG-YOLO

SEG-YOLO [70] is an extension of the well-known YOLO architecture [71], which combines real-time object detection with segmentation capability. The classic bounding box-based YOLO model is supplemented by additional segmentation heads, allowing masks to be generated directly for detected objects. As with object detection, YOLO-Seg aims for high speed with acceptable accuracy, making it particularly interesting for real-time applications.

DeepLabCut

[72] Nath et al. [72] developed DeepLabCut, a tool for markerless animal pose estimation. It is based on deep learning segmentation and keypoint tracking techniques that enable the precise identification of body points in animals in videos. DeepLabCut is particularly useful in behavioural research and neuroscience, as it allows highly accurate analysis of movements without invasive markers.

SegDiff

SegDiff [73] is a diffusion-based model for image segmentation that applies the principles of generative diffusion processes to the task of segmentation. Instead of directly predicting masks, SegDiff

 doi:10.20944/preprints202604.0507.v1

generates segmentation iteratively by reconstructing and refining noisy masks over several steps. This approach allows for particularly flexible and precise modelling of complex structures and fine details.

Swin-Unet

Swin-Unet [74] integrates the U-Net architecture [29] with Swin transformers functioning as encoders. Swin transformers [75] utilise a hierarchical representation and sliding windows, facilitating the effective processing of both local and global image information. The recursive structure of U-Net is retained, allowing the architecture to benefit from both the advantages of skip connections and the powerful global context modelling provided by transformers.

Segment Anything Model (SAM)

The Segment Anything Model (SAM) [10] was introduced by Meta AI and aims to create a universal model for segmenting arbitrary objects in arbitrary images, regardless of the object domain or class. It is based on a Transformer-based encoder-decoder architecture and uses prompts such as points, boxes, or masks as input. SAM is characterized by its ability to generalize to new, unseen data, making it particularly attractive for zero-shot and few-shot applications.

FastSAM

FastSAM [76] is an optimized variant of the Segment Anything Model (SAM) that aims to achieve significantly faster inference times while maintaining segmentation accuracy. Through more efficient network architectures, reduction of redundant computations, and accelerated feature extraction, FastSAM can segment large images or video streams in real time. The model retains SAM's flexibility to segment multiple objects using points, bounding boxes, or text instructions, making it particularly suitable for applications with high performance requirements such as robotics or interactive image processing.

Diffumask

DiffuMask is another diffusion-based segmentation model based on the idea of iteratively reconstructing image masks from noisy estimates [77]. By gradually refining the masks, DiffuMask can reliably capture complex object shapes and fine structures. The model combines the advantages of probabilistic diffusion processes with modern deep learning architecture to achieve accurate and consistent segmentation results.

Grounded-SAM

Grounded-SAM [78] extends the Segment Anything Model (SAM) with the ability to link segmented objects to semantic descriptions. While SAM is primarily specialized in generating object masks based on input points or bounding boxes, Grounded-SAM additionally allows segmentation to be specifically controlled by text or context information. This allows users to select specific objects in complex scenes without the need for separate training data for each class. This model combines the flexibility of SAM with targeted controllability through semantic inputs.

PS-YOLO-seg

PS-YOLO-Seg [79] is an extension of the well-known YOLO architecture [71] for instance and object detection, which additionally integrates a segmentation component. The model combines YOLO's fast and efficient object localization with precise mask predictions, so that objects are not only detected but also segmented with pixel-level accuracy. PS-YOLO-Seg is particularly suitable for applications that require both real-time performance and accurate segmentation, such as robotics or autonomous driving systems.

GS-YOLO-Seg

[80] GS-YOLO-Seg builds on the basic principles of YOLO [71] and extends them with a segmentation-capable architecture similar to PS-YOLO-Seg [79], but with an additional focus on

 doi:10.20944/preprints202604.0507.v1

improving accuracy for highly overlapping objects. Through optimized feature aggregation and special mask modules, GS-YOLO-Seg can accurately segment instances even in complex scenes without compromising high processing speed. This model is especially appropriate for applications requiring the reliable differentiation of objects in dense settings.

Table 4. Overview: Hybrid / Combination Segmentation Models.

Model	Year	Description
DifNet [69]	2018	Diffusion-based refinement for object boundaries
SEG-YOLO [70]	2019	YOLO detection with added segmentation heads
DeepLabCut [72]	2019	Hybrid: segmentation + keypoint-based pose estimation
SegDiff [73]	2021	Diffusion-based generative segmentation model
Swin-Unet [74]	2022	U-Net with hierarchical Swin Transformer encoder
DiffuMask [77]	2023	Iterative diffusion-based mask generation
SAM [10]	2023	Prompt-based universal segmentation model
FastSAM [76]	2023	Optimized, real-time variant of SAM
Grounded-SAM [78]	2024	SAM extended with text/context-driven segmentation
PS-YOLO-Seg [79]	2025	YOLO with instance segmentation heads
GS-YOLO-Seg [80]	2025	Enhanced YOLO segmentation for overlapping objects

4. Method Comparison: Advantages and Limitations

Image segmentation can be categorised into four primary groups: semantic segmentation, instance segmentation, panoptic segmentation, and hybrid approaches. Each of these methods has specific advantages and disadvantages that depend on the respective use case, the available computing resources, and the requirements for speed and accuracy.

5. Application Examples in the Industry

Image segmentation and deep learning-based quality assurance are used in a wide range of industrial applications. In automotive production, surface inspections are performed on car body parts to automatically detect paint defects, scratches, or dents. In electronics manufacturing, segmentation-based algorithms enable the detection of solder joint defects, missing components, or short circuits on printed circuit boards. Computer vision is also increasingly being used in the food industry to check product shape, size, or surface quality, for example in fruits, baked goods, or packaged foods. Other examples include industrial textile production, where weaving errors such as broken threads or irregularities are automatically detected, and metal processing, where cracks, scratches, or dimensional deviations are reliably identified. These applications show that automated segmentation systems not only make quality assurance more efficient and consistent, but can also complement and, in many cases, replace human inspection tasks.

5.1. Common Requirements and Issues in Industrial Environments

The use of image segmentation systems in industrial quality assurance processes places special demands on the technologies used. In production lines, image processing must keep pace with the production speed or cycle time in order to ensure that the manufacturing process runs without delays. This often requires real-time processing, with the decisive criterion being compliance with the time specifications of the production system.

The system must also work reliably under difficult conditions, such as changing lighting conditions, contamination, reflections, or varying positions and orientations of the objects to be inspected. Image segmentation solutions should be flexibly adaptable to different inspection characteristics, product variants, and production conditions. Seamless integration into the existing production infras-

 doi:10.20944/preprints202604.0507.v1

tructure is also necessary in order to communicate with the respective control systems, databases, and machines.

In addition, factors such as cost, maintenance effort, and the processing of large amounts of data play an important role. The selection of suitable image segmentation methods in an industrial environment therefore requires careful consideration of technical performance and practical feasibility.

6. Meta Analysis of Image Segmentation Methods

In order to evaluate the practical application of current image segmentation methods for industrial quality assurance tasks, a comparison based on relevant criteria is necessary. Depending on the application, different factors such as segmentation accuracy, processing speed, and hardware requirements are of crucial importance.

6.1. Search Strategy and Selection Criteria

A systematic literature search was performed in the scientific databases IEEE Xplore, ScienceDirect, SpringerLink, and the preprint platform arXiv to locate pertinent publications. This research aimed to locate contemporary studies on image segmentation techniques within the industrial domains of quality control and quality assurance, and to conduct a comparative analysis of their performance metrics.

Relevant search terms and keywords were specifically defined for the literature search in order to precisely narrow down the thematic focus. Among others, the keywords “image segmentation,” “quality control,” “industrial inspection,” “image segmentation quality control,” and combinations such as “semantic segmentation quality control industry,” “instance segmentation quality control manufacturing,” and “panoptic segmentation quality control production.” The search phrases were amalgamated utilising Boolean operators to attain an exact and thorough search outcome.

Explicit inclusion and exclusion criteria were delineated to guarantee the pertinence and integrity of the studies incorporated. Only peer-reviewed articles and high-quality preprints (e.g., from arXiv) from the publication period of 2018 to 2025 were considered in the evaluation. Peer-reviewed works were identified based on their publication in established, indexed journals or conferences. High-quality preprints were selected based on the reputation of the authors, citation frequency (where available), and recognition by the professional community.

Only papers with a clear reference to industrial quality control and assurance were considered. Studies without experimental performance data and papers dealing exclusively with non-industrial applications such as medical imaging or autonomous vehicles were excluded from the analysis. Review papers and publications without their own empirical investigations were also not considered. Table 5 provides an overview.

Table 5. Overview of selected papers on segmentation in industrial quality control.

Paper	Method	Applications	Advantage	Segmentation type
Yao et al. [81]	Weakly supervised segmentation, centroid loss	Industrial quality control	Pixel annotation not necessary, robust with little data	Semantic
Chen et al. [82]	Combination of anomaly detection + segmentation	Defect detection in cast and manufactured parts	Higher detection accuracy	Semantic
Shi et al. [83]	Semi-supervised learning	Surface inspection of industrial products	Efficient and accurate with little annotated data	Semantic
Tabernik et al. [84]	Surface inspection	Industrial quality control	Pixel-accurate defect detection	Semantic
Schack et al. [85]	Semantic segmentation	Fresh concrete quality control	Detection of air pockets and material distribution	Semantic
Valente et al. [86]	DeepLab-v3+ segmentation	Print defect mapping	High accuracy through synthetic training data	Semantic
Knott et al. [87]	Weakly supervised panoptic segmentation	Automated defect classification of fruit	Combination of semantic + instance, low annotation required	Panoptic
Nivaggioli et al. [88]	Synthetic training data + Panoptic segmentation	Industrial scenes	Reduced manual annotation effort, realistic training data	Panoptic
Ji et al. [89]	Instance segmentation on microscopic images	Food crystal quality control	Automated precise control	Instance
Marchi et al. [90]	3D + Color image segmentation	Overlapping screws in manufacturing	Robust detection despite overlaps	Instance
Jin et al. [91]	Real-time defect detection in moving objects	Production processes	Real-time capability in production	Instance
Kriegler et al. [92]	Instance Segmentation CNN	Laser cutting quality in batteries	High-precision defect detection	Instance
Chiu et al. [91]	Mask R-CNN + data augmentation	Wafer defect classification	High classification accuracy (97.7%)	Hybrid / Combination
Ferguson et al. [93]	CNN + Transfer Learning	Manufacturing defects	Better accuracy with small datasets	Hybrid / Combination

 doi:10.20944/preprints202604.0507.v1

6.2. Qualitative Evaluation

The overview of selected works shows a wide variety of approaches to image segmentation in industrial quality control. Semantic segmentation dominates many applications, especially in the detection of surface defects or material deviations, as in [81–86]. The advantages of these methods lie primarily in the reduction of manual annotations, robustness with limited data sets, and the ability to obtain precise, pixel-accurate information about defects. Panoptic segmentation is used when both the class membership and the identity of individual objects are relevant, as in [87,88]. It enables the combination of semantic and instance information and reduces the effort required for manual annotations by using synthetic training data.

Instance-based segmentation is primarily used in high-precision applications such as microscopic analysis [89], 3D object recognition [90], and laser cutting control [92]. The advantage here is the precise localization of individual objects, even in cases of overlap or moving objects, which enables automated quality checks.

Hybrid approaches that combine multiple methods, such as Mask R-CNN with data augmentation [91] or CNN-based transfer learning [93], offer a good balance between accuracy and efficiency, especially in scenarios with small or heterogeneous datasets.

In general, qualitative analysis shows that the choice of segmentation method depends heavily on the use case. Semantic segmentation is suitable for general surface inspections, while instance and panoptic methods are suitable for applications that require precise object boundaries and individual objects, and hybrid approaches offer a flexible solution for complex industrial scenarios.

7. Discussion

The methods developed to date for industrial defect detection and quality assurance are largely based on convolutional neural networks (CNNs). These models have proven themselves in numerous applications because they extract image features locally, can be implemented efficiently, and offer high performance in the segmentation of surface defects, cracks, or corrosion on different materials. In particular, they enable real-time detection and classification of defects, which is of great importance for industrial production lines.

However, CNN-based methods reach their limits, especially when it comes to capturing global relationships in the image and processing complex, heterogeneous data. The rigid local recording of CNN filters makes it difficult to detect defects that are characterized by subtle changes or varying contextual information. Certain types of defects are also difficult to detect based on local features alone, which limits the accuracy and robustness of the models.

In this context, transformer-based architectures represent a promising advance. The self-attention mechanism makes it possible to effectively model global dependencies in image data and capture more complex relationships. This opens up new possibilities for more precise and versatile segmentation of errors. Studies show that transformer models are more robust to noise and variations in the data. Advances in the real-time capability of transformer-based methods also show that these powerful architectures are increasingly becoming practical for resource-constrained and time-critical applications.

An application-specific approach is recommended for practical industrial use. Mask R-CNN or DeepLabv3+ are suitable for high-precision offline or safety-critical inspections, while lightweight U-Net variants or YOLACT models are more suitable for real-time and inline inspections. Flexible or changing product scenarios benefit from the combination of object detection and segmenting models or, experimentally, from the use of universal models such as Segment Anything. Hybrid image processing systems that combine classic image analysis and deep learning to further increase robustness and adaptability are also a useful addition.

 doi:10.20944/preprints202604.0507.v1

8. Conclusion and Outlook

The review of contemporary techniques for image segmentation in industrial quality control indicates that convolutional neural networks (CNNs) remain crucial. They provide exceptional precision in identifying and segmenting surface imperfections, fissures, and material discrepancies, and have been integrated into numerous applications. Weakly supervised and semi-supervised approaches reduce the annotation effort and enable the efficient use of small data sets. Semantic, instance, or panoptic/hybrid segmentation is utilised based on the application. Semantic segmentation is effective for flat faults, instance segmentation enables the recognition of distinct objects and overlapping components, and hybrid methodologies integrate both benefits to tackle more intricate situations. The use of synthetic data, transfer learning, and data augmentation further contributes to increased accuracy and robustness.

The rapid development in the field of computer vision and object segmentation opens up numerous possibilities for industrial quality assurance in the future. Resource-efficient models allow direct use on production lines without a cloud connection, while Explainable AI (XAI) [94] increasingly delivers comprehensible and interpretable segmentation results that are relevant for industrial quality standards. In addition, the expansion to 3D data and multi-sensor setups will improve the capture of complex structures, and no-code platforms [95] simplify the integration and adaptation of segmentation models even without in-depth AI knowledge. Zero- and few-shot learning approaches [96] reduce the training data requirements and enable the flexible use of universal models such as Segment Anything.

In summary, it can be said that the combination of these developments will further increase automation and process reliability in industrial quality assurance. Hybrid methods, optimized real-time applications, and targeted integration into industrial workflows will make industrial image segmentation more efficient, accurate, and practical in the future, with the focus remaining on balancing high performance and practical application.

References

1. Machado, N.C.; Illes, B.; Glistau, E. Logistik und Qualitätsmanagement.
2. Owen, D.G. Manufacturing defects. *SCL Rev.* **2001**, *53*, 851.
3.
4.
5.
6. Paszke, A.; Gross, S.; Massa, F.; Lerer, A.; Bradbury, J.; Chanan, G.; Killeen, T.; Lin, Z.; Gimelshein, N.; Antiga, L.; et al. PyTorch: An Imperative Style, High-Performance Deep Learning Library. In Proceedings of the Advances in Neural Information Processing Systems, 2019, Vol. 32, pp. 8024–8035.
7. Abadi, M.; Barham, P.; Chen, J.; Chen, Z.; Davis, A.; Dean, J.; Devin, M.; Ghemawat, S.; Irving, G.; Isard, M.; et al. TensorFlow: A system for large-scale machine learning. In Proceedings of the 12th USENIX Symposium on Operating Systems Design and Implementation (OSDI 16), 2016, pp. 265–283 title=TensorFlow: A system for large–scale machine learning, author=Abadi, Martin and Barham, Paul and Chen, Jianmin and Chen, Zhifeng and Davis, Andy and Dean, Jeffrey and Devin, Matthieu and Ghemawat, Sanjay and Irving, Geoffrey and Isard, Michael and others, booktitle=12th USENIX Symposium on Operating Systems Design and Implementation (OSDI 16), pages=265–283, year=2016.
8. Khan, S.; Naseer, M.; Hayat, M.; Zamir, S.W.; Khan, F.S.; Shah, M. Transformers in vision: A survey. *ACM computing surveys (CSUR)* **2022**, *54*, 1–41.
9. Wu, Y.; Kirillov, A.; Massa, F.; Lo, W.Y.; Girshick, R. Detectron2. https://github.com/facebookresearch/detectron2, 2019.
10. Kirillov, A.; Mintun, E.; Ravi, N.; Mao, H.; Rolland, C.; Gustafson, L.; Xiao, T.; Whitehead, S.; Berg, A.C.; Lo, W.Y.; et al. Segment anything. In Proceedings of the Proceedings of the IEEE/CVF international conference on computer vision, 2023, pp. 4015–4026.
11. Pinheiro, P.O.; Collobert, R.; Dollár, P. Learning to Segment Object Candidates. In Proceedings of the NIPS, 2015.

 doi:10.20944/preprints202604.0507.v1

12. Pinheiro, P.O.; Lin, T.Y.; Collobert, R.; Dollár, P. Learning to Refine Object Segments. In Proceedings of the ECCV, 2016.
13. Zagoruyko, S.; Lerer, A.; Lin, T.Y.; Pinheiro, P.O.; Gross, S.; Chintala, S.; Dollár, P. A MultiPath Network for Object Detection. In Proceedings of the BMVC title=Fastai: a layered API for deep learning, author=Howard, Jeremy and Gugger, Sylvain, journal=Information, volume=11, number=2, pages=108, year=2020, publisher=MDPI, 2016.
14. Bradski, G. The OpenCV Library. *Dr. Dobb's Journal of Software Tools* **2000**.
15. Torrey, L.; Shavlik, J. Transfer learning. In *Handbook of research on machine learning applications and trends: algorithms, methods, and techniques*; IGI Global Scientific Publishing, 2010; pp. 242–264.
16.
17. Everingham, M.; Gool, L.V.; Williams, C.K.I.; Winn, J.; Zisserman, A. The Pascal Visual Object Classes (VOC) Challenge. In Proceedings of the International Journal of Computer Vision, 2010, Vol. 88, pp. 303–338. https://doi.org/10.1007/s11263-009-0275-4.
18. Cordts, M.; Omran, M.; Ramos, S.; Rehfeld, T.; Enzweiler, M.; Benenson, R.; Franke, U.; Roth, S.; Schiele, B. The Cityscapes Dataset for Semantic Urban Scene Understanding. *CoRR* **2016**, *abs/1604.01685*, [1604.01685].
19. Gupta, A.; Dollár, P.; Girshick, R.B. LVIS: A Dataset for Large Vocabulary Instance Segmentation. *CoRR* **2019**, *abs/1908.03195*, 88–97, [1908.03195]. https://doi.org/10.1016/j.patrec.2008.04.005.
20. Azamfirei, V.; Psarommatis, F.; Lagrosen, Y. Application of automation for in-line quality inspection, a zero-defect manufacturing approach. *Journal of Manufacturing Systems* **2023**, *67*, 1–22. https://doi.org/10.1016/j.jmsy.2022.12.010.
21. Wu, Z.G.; Lin, C.Y.; Chang, H.W.; Lin, P.T. Inline Inspection with an Industrial Robot (IIIR) for Mass-Customization Production Line. *Sensors* **2020**, *20*, 3008. https://doi.org/10.3390/s20113008.
22. Kim, H.; Frommknecht, A.; Bieberstein, B.; Stahl, J.; Huber, M.F. Automated end-of-line quality assurance with visual inspection and convolutional neural networks. *tm - Technisches Messen* **2023**, *90*, 196–204. https://doi.org/10.1515/teme-2022-0092.
23. Sahoo, P.K.; Soltani, S.; Wong, A.K. A survey of thresholding techniques. *Computer vision, graphics, and image processing* **1988**, *41*, 233–260.
24. Hojjatoleslami, S.; Kittler, J. Region growing: a new approach. *IEEE Transactions on Image processing* **1998**, *7*, 1079–1084.
25.
26.
27. Kornilov, A.S.; Safonov, I.V. An Overview of Watershed Algorithm Implementations in Open Source Libraries. *Journal of Imaging* **2018**, *4*. https://doi.org/10.3390/jimaging4100123.
28. Long, J.; Shelhamer, E.; Darrell, T. Fully convolutional networks for semantic segmentation. In Proceedings of the Proceedings of the IEEE conference on computer vision and pattern recognition, 2015, pp. 3431–3440.
29. Ronneberger, O.; Fischer, P.; Brox, T. U-Net: Convolutional Networks for Biomedical Image Segmentation, 2015, [arXiv:cs.CV/1505.04597].
30. Paszke, A.; Chaurasia, A.; Kim, S.; Culurciello, E. Enet: A deep neural network architecture for real-time semantic segmentation. *arXiv preprint arXiv:1606.02147 title=Bisenet: Bilateral segmentation network for real-time semantic segmentation, author=Yu, Changqian and Wang, Jingbo and Peng, Chao and Gao, Changxin and Yu, Gang and Sang, Nong, booktitle=Proceedings of the European conference on computer vision (ECCV), pages=325–341, year=2018* **2016**.
31. Milletari, F.; Navab, N.; Ahmadi, S.A. V-net: Fully convolutional neural networks for volumetric medical image segmentation. In Proceedings of the 2016 fourth international conference on 3D vision (3DV). Ieee, 2016, pp. 565–571.
32. Romera, E.; Alvarez, J.M.; Bergasa, L.M.; Arroyo, R. Erfnet: Efficient residual factorized convnet for real-time semantic segmentation. *IEEE Transactions on Intelligent Transportation Systems* **2017**, *19*, 263–272.
33. Badrinarayanan, V.; Kendall, A.; Cipolla, R. Segnet: A deep convolutional encoder-decoder architecture for image segmentation. *IEEE transactions on pattern analysis and machine intelligence* **2017**, *39*, 2481–2495.
34. Zhao, H.; Shi, J.; Qi, X.; Wang, X.; Jia, J. Pyramid scene parsing network. In Proceedings of the Proceedings of the IEEE conference on computer vision and pattern recognition, 2017, pp. 2881–2890.
35. Chen, L.C.; Papandreou, G.; Kokkinos, I.; Murphy, K.; Yuille, A.L. Deeplab: Semantic image segmentation with deep convolutional nets, atrous convolution, and fully connected crfs. *IEEE transactions on pattern analysis and machine intelligence* **2017**, *40*, 834–848 title=V–net: Fully convolutional neural networks for volumetric medical image segmentation, author=Milletari, Fausto and Navab, Nassir and Ahmadi, Seyed–

 doi:10.20944/preprints202604.0507.v1

Ahmad, booktitle=2016 fourth international conference on 3D vision (3DV), pages=565–571, year=2016, organization=Ieee.

36. Zhao, H.; Qi, X.; Shen, X.; Shi, J.; Jia, J. Icnet for real-time semantic segmentation on high-resolution images. In Proceedings of the Proceedings of the European conference on computer vision (ECCV), 2018, pp. 405–420.
37. Yu, C.; Wang, J.; Peng, C.; Gao, C.; Yu, G.; Sang, N. Bisenet: Bilateral segmentation network for real-time semantic segmentation. In Proceedings of the Proceedings of the European conference on computer vision (ECCV), 2018, pp. 325–341.
38. Zhu, Z.; Yan, Y.; Xu, R.; Zi, Y.; Wang, J. Attention-Unet: A deep learning approach for fast and accurate segmentation in medical imaging. *Journal of Computer Science and Software Applications* **2022**, *2*, 24–31.
39. Zheng, S.; Lu, J.; Zhao, H.; Zhu, X.; Luo, Z.; Wang, Y.; Fu, Y.; Feng, J.; Xiang, T.; Torr, P.H.; et al. Rethinking semantic segmentation from a sequence-to-sequence perspective with transformers. In Proceedings of the Proceedings of the IEEE/CVF conference on computer vision and pattern recognition, 2021, pp. 6881–6890 title=Rethinking semantic segmentation from a sequence–to–sequence perspective with transformers, author=Zheng, Sixiao and Lu, Jiachen and Zhao, Hengshuang and Zhu, Xiatian and Luo, Zekun and Wang, Yabiao and Fu, Yanwei and Feng, Jianfeng and Xiang, Tao and Torr, Philip HS and others, booktitle=Proceedings of the IEEE/CVF conference on computer vision and pattern recognition, pages=6881–6890, year=2021.
40. Strudel, R.; Garcia, R.; Laptev, I.; Schmid, C. Segmenter: Transformer for semantic segmentation. In Proceedings of the Proceedings of the IEEE/CVF international conference on computer vision, 2021, pp. 7262–7272 title=Pyramid scene parsing network, author=Zhao, Hengshuang and Shi, Jianping and Qi, Xiaojuan and Wang, Xiaogang and Jia, Jiaya, booktitle=Proceedings of the IEEE conference on computer vision and pattern recognition, pages=2881–2890, year=2017.
41. Xie, E.; Wang, W.; Yu, Z.; Anandkumar, A.; Alvarez, J.M.; Luo, P. SegFormer: Simple and efficient design for semantic segmentation with transformers. *Advances in neural information processing systems* **2021**, *34*, 12077–12090 title=SegFormer: Simple and efficient design for semantic segmentation with transformers, author=Xie, Enze and Wang, Wenhai and Yu, Zhiding and Anandkumar, Anima and Alvarez, Jose M and Luo, Ping, journal=Advances in neural information processing systems, volume=34, pages=12077–12090, year=2021.
42. Girshick, R.; Donahue, J.; Darrell, T.; Malik, J. Rich feature hierarchies for accurate object detection and semantic segmentation. In Proceedings of the Proceedings of the IEEE conference on computer vision and pattern recognition, 2014, pp. 580–587.
43. Girshick, R. Fast r-cnn. In Proceedings of the Proceedings of the IEEE international conference on computer vision, 2015, pp. 1440–1448.
44. Liu, S.; Jia, J.; Fidler, S.; Urtasun, R. Sgn: Sequential grouping networks for instance segmentation. In Proceedings of the Proceedings of the IEEE international conference on computer vision, 2017, pp. 3496–3504 title=Conditional convolutions for instance segmentation, author=Tian, Zhi and Shen, Chunhua and Chen, Hao, booktitle=European conference on computer vision, pages=282–298, year=2020, organization=Springer.
45. He, K.; Gkioxari, G.; Dollár, P.; Girshick, R. Mask r-cnn. In Proceedings of the Proceedings of the IEEE international conference on computer vision, 2017, pp. 2961–2969.
46. Liu, S.; Qi, L.; Qin, H.; Shi, J.; Jia, J. Path aggregation network for instance segmentation. In Proceedings of the Proceedings of the IEEE conference on computer vision and pattern recognition, 2018, pp. 8759–876 title=Path aggregation network for instance segmentation, author=Liu, Shu and Qi, Lu and Qin, Haifang and Shi, Jianping and Jia, Jiaya, booktitle=Proceedings of the IEEE conference on computer vision and pattern recognition, pages=8759–8768, year=2018.
47. Chen, L.C.; Hermans, A.; Papandreou, G.; Schroff, F.; Wang, P.; Adam, H. Masklab: Instance segmentation by refining object detection with semantic and direction features. In Proceedings of the Proceedings of the IEEE conference on computer vision and pattern recognition, 2018, pp. 4013–4022 title=End–to–end object detection with transformers, author=Carion, Nicolas and Massa, Francisco and Synnaeve, Gabriel and Usunier, Nicolas and Kirillov, Alexander and Zagoruyko, Sergey, booktitle=European conference on computer vision, pages=213–229, year=2020, organization=Springer.
48. Cai, Z.; Vasconcelos, N. Cascade r-cnn: Delving into high quality object detection. In Proceedings of the Proceedings of the IEEE conference on computer vision and pattern recognition, 2018, pp. 6154–6162.
49. Chen, K.; Pang, J.; Wang, J.; Xiong, Y.; Li, X.; Sun, S.; Feng, W.; Liu, Z.; Shi, J.; Ouyang, W.; et al. Hybrid task cascade for instance segmentation. In Proceedings of the Proceedings of the IEEE/CVF conference on computer vision and pattern recognition, 2019, pp. 4974–4983.
50.

 doi:10.20944/preprints202604.0507.v1

51. Chen, X.; Girshick, R.; He, K.; Dollár, P. Tensormask: A foundation for dense object segmentation. In Proceedings of the Proceedings of the IEEE/CVF international conference on computer vision, 2019, pp. 2061–206 title=Upsnet: A unified panoptic segmentation network, author=Xiong, Yuwen and Liao, Renjie and Zhao, Hengshuang and Hu, Rui and Bai, Min and Yumer, Ersin and Urtasun, Raquel, booktitle=Proceedings of the IEEE/CVF conference on computer vision and pattern recognition, pages=8818–8826, year=2019.
52. Wang, X.; Kong, T.; Shen, C.; Jiang, Y.; Li, L. Solo: Segmenting objects by locations. In Proceedings of the European conference on computer vision. Springer, 2020, pp. 649–665.
53.
54. Tian, Z.; Shen, C.; Chen, H. Conditional convolutions for instance segmentation. In Proceedings of the European conference on computer vision. Springer, 2020, pp. 282–298.
55. Carion, N.; Massa, F.; Synnaeve, G.; Usunier, N.; Kirillov, A.; Zagoruyko, S. End-to-end object detection with transformers. In Proceedings of the European conference on computer vision. Springer, 2020, pp. 213–229.
56. Zhu, X.; Su, W.; Lu, L.; Li, B.; Wang, X.; Dai, J. Deformable detr: Deformable transformers for end-to-end object detection. *arXiv preprint arXiv:2010.04159 title=Deformable detr: Deformable transformers for end-to-end object detection, author=Zhu, Xizhou and Su, Weijie and Lu, Lewei and Li, Bin and Wang, Xiaogang and Dai, Jifeng, journal=arXiv preprint arXiv:2010.04159, year=2020* **2020**.
57. Chen, X.; Wei, F.; Zeng, G.; Wang, J. Conditional detr v2: Efficient detection transformer with box queries. *arXiv preprint arXiv:2207.08914* **2022**.
58. Meng, D.; Chen, X.; Fan, Z.; Zeng, G.; Li, H.; Yuan, Y.; Sun, L.; Wang, J. Conditional detr for fast training convergence. In Proceedings of the Proceedings of the IEEE/CVF international conference on computer vision, 2021, pp. 3651–3660.
59. Ren, S.; He, K.; Girshick, R.; Sun, J. Faster R-CNN: Towards real-time object detection with region proposal networks. *IEEE transactions on pattern analysis and machine intelligence* **2016**, *39*, 1137–1149.
60. Xiong, Y.; Liao, R.; Zhao, H.; Hu, R.; Bai, M.; Yumer, E.; Urtasun, R. Upsnet: A unified panoptic segmentation network. In Proceedings of the Proceedings of the IEEE/CVF conference on computer vision and pattern recognition, 2019, pp. 8818–8826.
61. Sofiiuk, K.; Barinova, O.; Konushin, A. Adaptis: Adaptive instance selection network. In Proceedings of the Proceedings of the IEEE/CVF international conference on computer vision, 2019, pp. 7355–7363.
62. Chang, C.Y.; Chang, S.E.; Hsiao, P.Y.; Fu, L.C. EPSNet: efficient panoptic segmentation network with cross-layer attention fusion. In Proceedings of the Proceedings of the Asian conference on computer vision title=EPSNet: efficient panoptic segmentation network with cross-layer attention fusion, author=Chang, Chia-Yuan and Chang, Shuo-En and Hsiao, Pei-Yung and Fu, Li-Chen, booktitle=Proceedings of the Asian conference on computer vision, year=2020, 2020.
63. De Geus, D.; Meletis, P.; Dubbelman, G. Fast panoptic segmentation network. *IEEE Robotics and Automation Letters* **2020**, *5*, 1742–1749 title=The mapillary vistas dataset for semantic understanding of street scenes, author=Neuhold, Gerhard and Ollmann, Tobias and Rota Bulo, Samuel and Kontschieder, Peter, booktitle=Proceedings of the IEEE international conference on computer vision, pages=4990–4999, year=2017.
64. Cheng, B.; Collins, M.D.; Zhu, Y.; Liu, T.; Huang, T.S.; Adam, H.; Chen, L.C. Panoptic-deeplab: A simple, strong, and fast baseline for bottom-up panoptic segmentation. In Proceedings of the Proceedings of the IEEE/CVF conference on computer vision and pattern recognition, 2020, pp. 12475–12485.
65. Mohan, R.; Valada, A. Efficientps: Efficient panoptic segmentation. *International Journal of Computer Vision* **2021**, *129*, 1551–1579.
66. Cheng, B.; Misra, I.; Schwing, A.G.; Kirillov, A.; Girdhar, R. Masked-attention mask transformer for universal image segmentation. In Proceedings of the Proceedings of the IEEE/CVF conference on computer vision and pattern recognition, 2022, pp. 1290–1299.
67. Li, F.; Zhang, H.; Xu, H.; Liu, S.; Zhang, L.; Ni, L.M.; Shum, H.Y. Mask dino: Towards a unified transformer-based framework for object detection and segmentation. In Proceedings of the Proceedings of the IEEE/CVF conference on computer vision and pattern recognition, 2023, pp. 3041–3050.
68. Wen, J.; Zhang, Q.; Zhang, G. VPSNet: 3D object detection with voxel purification and fully sparse convolutional networks. *The Journal of Supercomputing* **2025**, *81*, 466.
69. Jiang, P.; Gu, F.; Wang, Y.; Tu, C.; Chen, B. Difnet: Semantic segmentation by diffusion networks. *Advances in Neural Information Processing Systems* **2018**, *31*.
70.

 doi:10.20944/preprints202604.0507.v1

71. Redmon, J.; Divvala, S.; Girshick, R.; Farhadi, A. You only look once: Unified, real-time object detection. In Proceedings of the Proceedings of the IEEE conference on computer vision and pattern recognition, 2016, pp. 779–788.
72. Nath, T.; Mathis, A.; Chen, A.C.; Patel, A.; Bethge, M.; Mathis, M.W. Using DeepLabCut for 3D markerless pose estimation across species and behaviors. *Nature protocols* **2019**, *14*, 2152–2176.
73. Amit, T.; Shaharbany, T.; Nachmani, E.; Wolf, L. Segdiff: Image segmentation with diffusion probabilistic models. *arXiv preprint arXiv:2112.00390* **2021**.
74. Cao, H.; Wang, Y.; Chen, J.; Jiang, D.; Zhang, X.; Tian, Q.; Wang, M. Swin-unet: Unet-like pure transformer for medical image segmentation. In Proceedings of the European conference on computer vision. Springer, 2022, pp. 205–218.
75. Liu, Z.; Lin, Y.; Cao, Y.; Hu, H.; Wei, Y.; Zhang, Z.; Lin, S.; Guo, B. Swin transformer: Hierarchical vision transformer using shifted windows. In Proceedings of the Proceedings of the IEEE/CVF international conference on computer vision, 2021, pp. 10012–10022.
76. Zhao, X.; Ding, W.; An, Y.; Du, Y.; Yu, T.; Li, M.; Tang, M.; Wang, J. Fast segment anything. *arXiv preprint arXiv:2306.12156* **2023**.
77. Wu, W.; Zhao, Y.; Shou, M.Z.; Zhou, H.; Shen, C. Diffumask: Synthesizing images with pixel-level annotations for semantic segmentation using diffusion models. In Proceedings of the Proceedings of the IEEE/CVF International Conference on Computer Vision, 2023, pp. 1206–1217.
78. Ren, T.; Liu, S.; Zeng, A.; Lin, J.; Li, K.; Cao, H.; Chen, J.; Huang, X.; Chen, Y.; Yan, F.; et al. Grounded sam: Assembling open-world models for diverse visual tasks. *arXiv preprint arXiv:2401.14159* **2024**.
79. Qiu, Z.; Huang, X.; Deng, Z.; Xu, X.; Qiu, Z. PS-YOLO-seg: A Lightweight Instance Segmentation Method for Lithium Mineral Microscopic Images Based on Improved YOLOv12-seg. *Journal of Imaging* **2025**, *11*, 230.
80. Qiu, Z.; Huang, X.; Sun, Z.; Li, S.; Wang, J. GS-YOLO-Seg: A Lightweight Instance Segmentation Method for Low-Grade Graphite Ore Sorting Based on Improved YOLO11-Seg. *Sustainability* **2025**, *17*, 5663.
81. Yao, K.; Ortiz, A.; Bonnin-Pascual, F. A weakly-supervised semantic segmentation approach based on the centroid loss: Application to quality control and inspection. *IEEE Access* **2021**, *9*, 69010–69026.
82. Chen, M.C.; Yen, S.Y.; Lin, Y.F.; Tsai, M.Y.; Chuang, T.H. Intelligent Casting Quality Inspection Method Integrating Anomaly Detection and Semantic Segmentation. *Machines* **2025**, *13*. https://doi.org/10.3390/machines13040317.
83. Shi, C.; Wang, K.; Zhang, G.; Li, Z.; Zhu, C. Efficient and accurate semi-supervised semantic segmentation for industrial surface defects. *Scientific Reports* **2024**, *14*, 21874.
84. Tabernik, D.; Šela, S.; Skvarč, J.; Skočaj, D. Segmentation-based deep-learning approach for surface-defect detection. *Journal of Intelligent Manufacturing* **2020**, *31*, 759–776 title=Efficient and accurate semi–supervised semantic segmentation for industrial surface defects, author=Shi, Chenbo and Wang, Kang and Zhang, Guodong and Li, Zelong and Zhu, Changsheng, journal=Scientific Reports, volume=14, number=1, pages=21874, year=2024, publisher=Nature Publishing Group UK London.
85. Schack, T.; Coenen, M.; Haist, M. Image-based quality control of fresh concrete based on semantic segmentation algorithms. *Civil Engineering Design* **2024**, *6*, 96–105.
86. Valente, A.; Wada, C.; Neves, D.; Neves, D.; Perez, F.; Megeto, G.; Cascone, M.; Gomes, O.; Lin, Q. Print defect mapping with semantic segmentation. In Proceedings of the Proceedings of the IEEE/CVF winter conference on applications of computer vision, 2020, pp. 3551–3559 title=Print defect mapping with semantic segmentation, author=Valente, Augusto and Wada, Cristina and Neves, Deangela and Neves, Deangeli and Perez, Fabio and Megeto, Guilherme and Cascone, Marcos and Gomes, Otavio and Lin, Qian, booktitle=Proceedings of the IEEE/CVF winter conference on applications of computer vision, pages=3551–3559, year=2020.
87. Knott, M.; Odion, D.; Sontakke, S.; Karwa, A.; Defraeye, T. Weakly Supervised Panoptic Segmentation for Defect-Based Grading of Fresh Produce. In Proceedings of the Proceedings of the Computer Vision and Pattern Recognition Conference, 2025, pp. 5462–5471 title=Weakly Supervised Panoptic Segmentation for Defect–Based Grading of Fresh Produce, author=Knott, Manuel and Odion, Divinefavour and Sontakke, Sameer and Karwa, Anup and Defraeye, Thijs, booktitle=Proceedings of the Computer Vision and Pattern Recognition Conference, pages=5462–5471, year=2025.
88. Nivaggioli, A.; Hullo, J.; Thibault, G. Using 3D models to generate labels for panoptic segmentation of industrial scenes. *ISPRS Annals of the Photogrammetry, Remote Sensing and Spatial Information Sciences* **2019**, *4*, 61–68.

 doi:10.20944/preprints202604.0507.v1

89. Ji, X.; Allebach, J.P.; Shakouri, A.; Zhu, F. Efficient Microscopic Image Instance Segmentation for Food Crystal Quality Control. In Proceedings of the 2024 IEEE 26th International Workshop on Multimedia Signal Processing (MMSP), 2024, pp. 1–6. https://doi.org/10.1109/MMSP61759.2024.10743276.
90. Marchi, E.; Fornasier, D.; Miorin, A.; Foresti, G.L. Segmentation networks for detecting overlapping screws in 3D and color images for industrial quality control. *Integrated Computer-Aided Engineering* **2025**, *32*, 244–257. https://doi.org/10.1177/10692509251328780.
91. Chiu, M.C.; Chen, T.M. Applying Data Augmentation and Mask R-CNN-Based Instance Segmentation Method for Mixed-Type Wafer Maps Defect Patterns Classification. *IEEE Transactions on Semiconductor Manufacturing* **2021**, *34*, 455–463. https://doi.org/10.1109/TSM.2021.3118922.
92. Kriegler, J.; Liu, T.; Hartl, R.; Hille, L.; Zaeh, M.F. Automated Quality Evaluation for Laser Cutting in Lithium Metal Battery Production Using an Instance Segmentation Convolutional Neural Network. *Journal of Laser Applications* **2023**, *35*, 042072. https://doi.org/10.2351/7.0001213.
93. Ferguson, M.; Ak, R.; Lee, Y.T.T.; Law, K.H. Detection and segmentation of manufacturing defects with convolutional neural networks and transfer learning. *Smart and sustainable manufacturing systems* **2018**, *2*, 137–164.
94. Ribeiro, M.T.; Singh, S.; Guestrin, C. " Why should i trust you?" Explaining the predictions of any classifier. In Proceedings of the Proceedings of the 22nd ACM SIGKDD international conference on knowledge discovery and data mining, 2016, pp. 1135–1144.
95. Upadhyaya, N. Low-Code/No-Code platforms and their impact on traditional software development: A literature review. *No-Code Platforms and Their Impact on Traditional Software Development: A Literature Review (March 21, 2023)* **2023**.
96. Chen, J.; Geng, Y.; Chen, Z.; Pan, J.Z.; He, Y.; Zhang, W.; Horrocks, I.; Chen, H. Zero-shot and few-shot learning with knowledge graphs: A comprehensive survey. *Proceedings of the IEEE* **2023**, *111*, 653–685.