

# Computer Vision and Convolutional Neural Networks

Subjects: [Agricultural Engineering](#) | [Computer Science](#), [Artificial Intelligence](#)

Contributor: Ya-Hong Wang , Wen-Hao Su

Computer vision (CV) combined with a deep convolutional neural network (CNN) has emerged as a reliable analytical method to effectively characterize and quantify high-throughput phenotyping of different grain crops, including rice, wheat, corn, and soybean. In addition to the ability to rapidly obtain information on plant organs and abiotic stresses, and the ability to segment crops from weeds, such techniques have been used to detect pests and plant diseases and to identify grain varieties. The development of corresponding imaging systems to assess the phenotypic parameters, yield, and quality of crop plants will increase the confidence of stakeholders in grain crop cultivation, thereby bringing technical and economic benefits to advanced agriculture.

grain crops

convolutional neural network

computer vision

## 1. Introduction

Global food security remains an important issue for human development <sup>[1]</sup>. By 2050, the global population is likely to exceed 9 billion, which means that agricultural production will need to increase by at least 70% from its current level to meet the growing demand for food <sup>[2]</sup>. Grains are the main component of the human diet, and rice, wheat, corn, and soybean account for more than 80% of global grain production <sup>[3]</sup>. Intelligent perception of crop phenotypic information helps to achieve precise field management, such as the selection of new varieties of high-yield and high-quality crops, and the minimization of agricultural inputs without affecting crop output. Plant phenotypes are the recognizable morphological, physiological, and biochemical characteristics and traits resulting from gene-environment interactions, including plant structure, composition, growth, and development <sup>[4]</sup>. This means that phenotypic assessment not only involves the traits expressed by crop genes, but also reflects complex traits such as physiology, biochemistry, quality, stress resistance, or ones that are influenced by the external environment.

Computer vision (CV), when combined with pattern recognition algorithms and automatic classification tools, exhibits outstanding performance. Traditional plant phenotype detection relies on manual observation and measurement to obtain a description of the external morphology of the plant, and then assess the relationship between genes or external environment and phenotype. However, this approach can only detect individual traits from a small sample of crops, thus the acquisition process is inefficient and the amount of data available is very limited. With the increasing demand for high-volume plant phenotypic information, researchers urgently need high-precision, high-throughput, and low-cost techniques to replace traditional manual methods of obtaining relevant data. A variety of imaging techniques are available to collect complex traits related to growth, yield, and adaptation

to biotic or abiotic stresses (e.g., diseases, insects, water stress, and nutrient deficiencies), including color imaging (e.g., machine vision), imaging spectroscopy (e.g., multi-spectral and hyperspectral remote sensing), thermal infrared imaging, fluorescence imaging, 3D imaging, and laminar imaging [5].

Over the past few decades, computer vision has been widely applied to analyze the phenotypic characteristics of grain crops and thus ease the food supply problem. Although a review of the phenotypic assessment of grain crops based on computer vision was published in 2018, the research mainly summarized the application of traditional machine-learning algorithms such as the support vector machine (SVM) and the back-propagation neural network (BPNN) [6]. In addition, some researchers have reviewed the research on pest and disease analysis of crops [7], crop and weed identification [8], and physical and chemical phenotypic characteristics of crops [9], but they only mentioned a particular phenotyping task. Importantly, some new network architectures and strategies applied to the field of the convolutional neural network (CNN) and computer vision are rarely covered in the extensive reviews covering crop phenotype detection since 2019. Several papers have been published in the last three years that provide comprehensive reviews of deep learning techniques for such computer vision tasks as image classification [10], object detection [11], and semantic and instance segmentation [12]. These reviews effectively summarize the basic principles, development history, and future trends for the latest CNNs in computer vision, but none of them provide information related to agriculture, which highlights a gap between these technological theories and phenotyping applications.

Focusing on the state-of-the-art CNN algorithms rather than traditional machine learning (the specific differences are shown in **Figure 1**), this research is an important early step in the search for phenotyping of grain crops. Given the importance of the four most productive grain crops (rice, wheat, maize, and soybean) in the world, the related work on computer vision-based CNN models for the detections of crop organs, crops in weeds, plant diseases, insect infestations, abiotic stresses, and grain varieties since 2019 has been reviewed. The goal is to provide a comprehensive overview of novel CNN models combined with CV for phenotype detection in grain crops and to provide researchers and breeders with clear guidance for related decisions. This will greatly boost the productivity of grain crops.

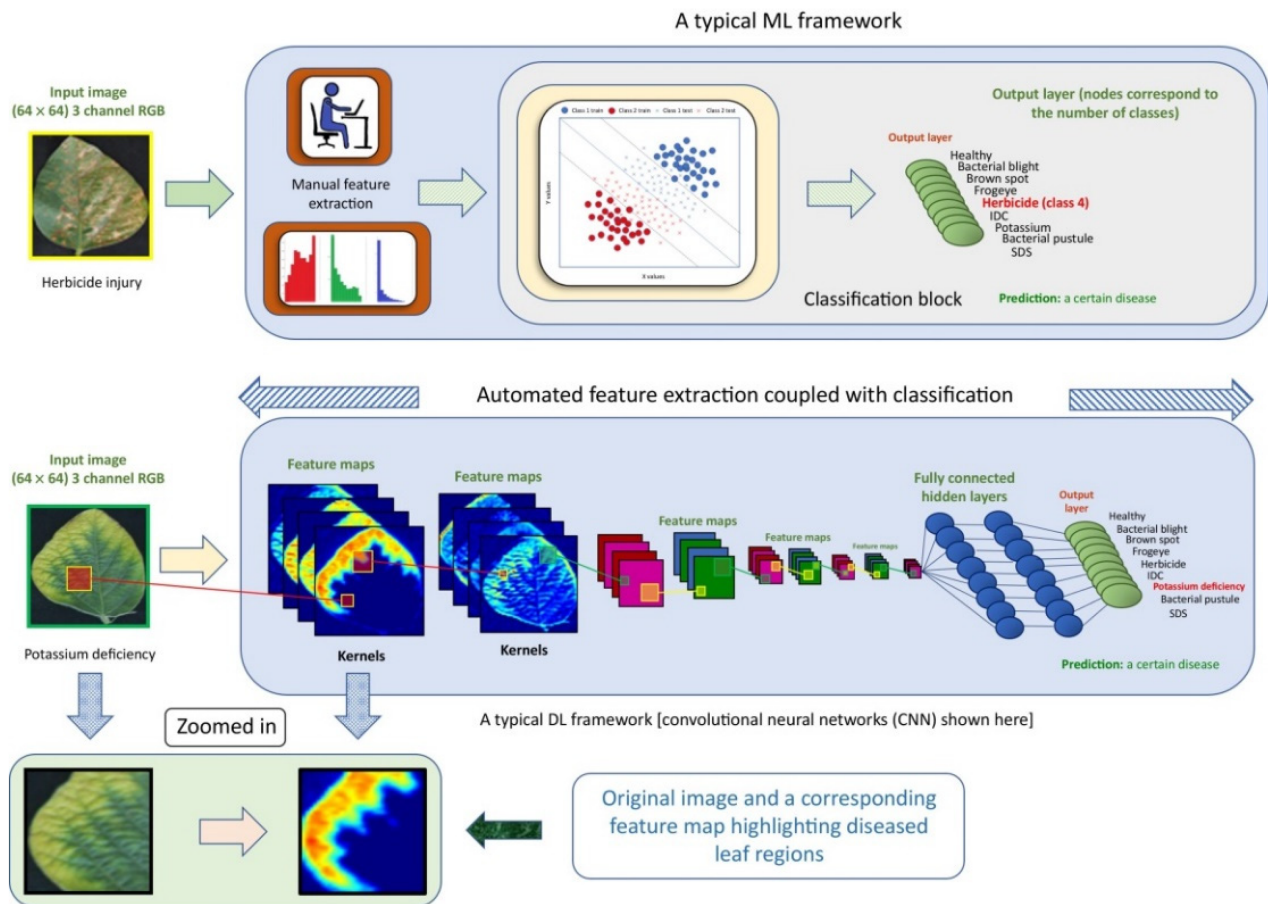


Figure 1. Key differences between machine learning (ML) and deep learning (DL) paradigms [13].

## 2. Computer Vision (CV) and Convolutional Neural Networks (CNNs)

### 2.1. CV

In recent years, both the hardware and software of CV systems have been significantly developed. The hardware, including cameras, lights, and communication devices, is the foundation of CV, while the software, such as image processing algorithms, is the core of the system. A typical image acquisition system is indispensable to illumination devices. The illumination devices can be divided into point light sources, strip light sources, ring light sources, backlight light sources, structure light sources, and combined light sources. These light sources can be further classified as light-emitting diode (LED) light sources, halogen light sources, and high-frequency fluorescent light sources. In addition, the camera can be characterized as a global shutter or a roll-up shutter camera.

### 2.2. CNN

Since 2012, CNNs have dominated solutions to CV tasks, showing superior performance over traditional machine-learning methods [14]. CNNs are deep learning architectures with spontaneous feature learning for image processing and image recognition. After the parameter optimization of training and learning, the CNN performs

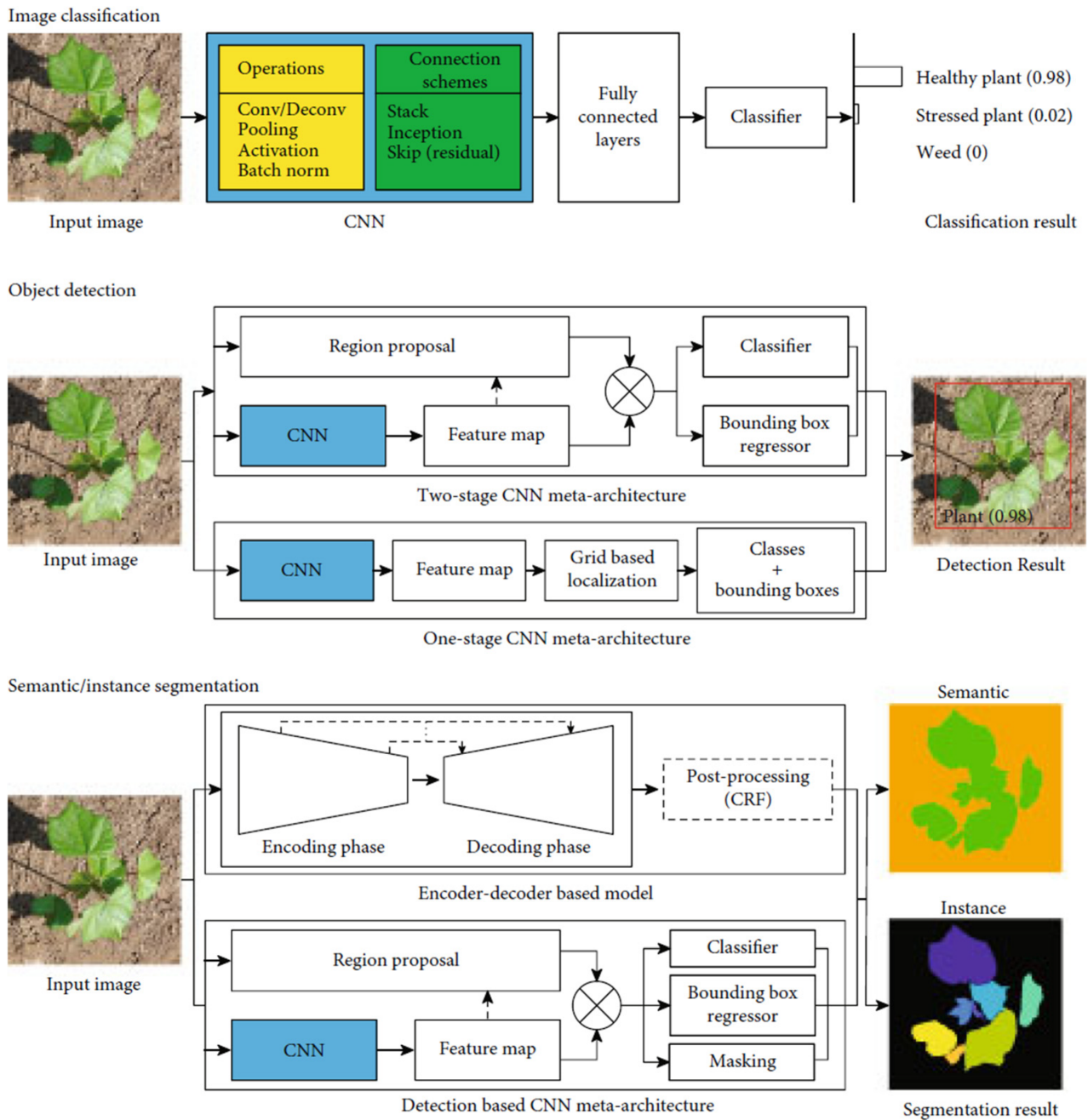
multiple layers of nonlinear transformations on the input data, continuously coupling the low-level features, and finally obtains a high-level semantic representation. Compared with traditional machine learning, a CNN can use a deeper neural network model to train the input data to simplify the data processing process.

A typical CNN consists of a convolutional layer, a pooling layer, and a fully connected layer [15]. The neurons in the convolutional layer are arranged in a matrix to form a multi-channel feature map. A neuron in each channel is connected to only a part of the feature map before that layer [16]. The final input of the neuron is obtained by convolving it with a convolution kernel and then using an activation function. CNNs emphasize weight sharing as a key component. Neurons located on the same channel feature map of the same convolutional layer are obtained by applying the same convolutional kernel to the previous feature map of the layer. Guided by local features in higher feature maps, the convolutional layer searches for links between them, while pooling layers combine data with the same semantics. Because the graphical information formed by adjacent positions may be slightly jittered, the pooling operation extracts the main information from the upper feature map. Maximum pooling and average pooling are common pooling operations. The model is able to keep translation and rotation invariant while preserving features [15]. After alternating between convolution and pooling, a fully connected layer often appears. Each neuron in the fully connected layer is connected to every neuron in the upper layer. All the information is combined to turn the multi-dimensional features into one-dimensional features, which are handed over to the final regressor and classifier to produce the final result.

## 2.3. CNNs Combined with CV Tasks

### 2.3.1. Image Classification

Image classification aims to assign predefined class labels to images. The CNN is currently the most popular neural network that combines a set of mathematical operations (e.g., convolution, pooling, and activation), using various connection schemes, such as plain stacking, start, and residual connections, to learn operational parameters from annotated images in order to classify image datasets (**Figure 2**). The current development of modern CNNs for image classification can be divided into three phases: (1) the appearance of modern CNNs (2012–2014); (2) the development and refinement of CNN architecture intensification (2014–2017); and (3) the introduction of reinforcement learning and artificial intelligence for CNN architecture design (start of 2017).



**Figure 2.** Diagrams of CNN architecture mechanisms for image classification, object detection, and semantic and instance segmentation [17].

In 2012, the first modern CNN architecture named AlexNet was proposed. The algorithm demonstrated strong performance in image classification in the ImageNet Large Scale Visual Recognition Challenge (ILSVRC 2012) competition in that year [18]. The report of this model introduces a new era of image classification and other CV tasks using CNNs. From 2014 to 2017, researchers developed several representative CNNs such as the residual neural network (ResNet) [19], the visual geometry group network (VGG) [20], and the dense convolutional network (DenseNet) [21] for image classification. These CNNs significantly improved the learning ability and recognition complexity by using efficient computational algorithms and modified connectivity schemes. From 2017, more studies focused on the use of reinforcement learning to search for the best CNN architecture that could yield higher

performance [22]. This process introduces a reinforcement learning framework to find the optimal convolutional image elements on small datasets, followed by stacking and transferring the resulting image elements in a different way to a large unknown dataset.

Researchers have investigated the mechanisms of CNNs for image classification. A recent study improved AlexNet to create a new variant (ZFNet) using a visualization tool. This tool is a framework integrated with CNNs that can map neuronal activity back to the input pixel space. Thus, pixel-level activations can be visualized after each convolutional layer, which is particularly useful for understanding the CNN mechanism for further upgrades. CNNs can learn general representations of images rather than features solely for classification. Subsequent research developed various gradient-based methods, including guided backpropagation, gradient-weighted class activation mapping (Grad-CAM), and layer-by-layer relevance propagation (LRP). Meanwhile, some general frameworks (e.g., LIME and occlusion maps) can also be used to display important image regions for classification results [23][24].

### 2.3.2. Object Detection

Object detection is defined as determining the location of objects in a given image and the class to which each object belongs. As shown in **Figure 2**, object detection using CNNs can be divided into two categories: single-level and two-level CNN architectures. In the early framework development, OverFeat is the most representative model [25], and won the localization task of the 2013 ILSVRC competition. Then, a series of region-based region-convolution neural network (R-CNN) frameworks was introduced, including the original R-CNN [26], Fast R-CNN [27], and Faster R-CNN [28]. There are three key techniques in the RCNN architectures, including the region proposal network (RPN), region of interest (ROI) pooling operation, and multi-task loss function. The R-CNN family has been widely adopted as object detectors for various domain datasets.

### 2.3.3. Semantic and Instance Segmentations

Semantic segmentation aims to assign a class to each pixel in an image, but objects in the same classes are not distinguished. Instance segmentation outputs the mask and class of the target. Typically, CNN architectures for semantic and instance segmentations can be divided into two categories, including encoder-decoder-based frameworks and detection-based frameworks, as shown in **Figure 2**. The encoder-decoder-based model is the most primitive intelligent image segmentation network for improving segmentation accuracy. In the encoder stage, the CNN extracts semantic features from input samples. In the decoder stage, deconvolution is used to assign the extracted features to the label of each pixel. Representative models based on encoder-decoder include full convolutional networks (FCNs) [29], DeepLab [30], and U-Net [31]. Frameworks including R-CNN, Faster R-CNN, and Mask R-CNN have been widely used for instance segmentation [32][33].

---

## References

1. Tilman, D.; Balzer, C.; Hill, J.; Befort, B.L. Global food demand and the sustainable intensification of agriculture. *Proc. Natl. Acad. Sci. USA* 2011, 108, 20260–20264.
2. Steensland, A.; Thompson, T.L. 2020 Global Agricultural Productivity Report: Productivity in a Time of Pandemics. *Global Agricultural Productivity Report: Productivity in a Time of Pandemics*; College of Agriculture and Life Sciences: Blacksburg, VA, USA, 2020.
3. Yu, Q.Y.; Xiang, M.T.; Wu, W.B.; Tang, H.J. Changes in global cropland area and cereal production: An inter-country comparison. *Agric. Ecosyst. Environ.* 2019, 269, 140–147.
4. Pan, Y.H. Analysis of concepts and categories of plant phenome and phenomics. *Acta Agron. Sin.* 2015, 41, 175–186.
5. Vithu, P.; Moses, J.A. Machine vision system for food grain quality evaluation: A review. *Trends Food Sci. Technol.* 2016, 56, 13–20.
6. Patrício, D.I.; Rieder, R. Computer vision and artificial intelligence in precision agriculture for grain crops: A systematic review. *Comput. Electron. Agric.* 2018, 153, 69–81.
7. Ngugi, L.C.; Abelwahab, M.; Abo-Zahhad, M. Recent advances in image processing techniques for automated leaf pest and disease recognition—A review. *Inf. Process. Agric.* 2021, 8, 27–51.
8. Wang, A.C.; Zhang, W.; Wei, X.H. A review on weed detection using ground-based machine vision and image processing techniques. *Comput. Electron. Agric.* 2019, 158, 226–240.
9. Sun, D.; Robbins, K.; Morales, N.; Shu, Q.; Cen, H. Advances in optical phenotyping of cereal crops. *Trends Plant Sci.* 2022, 27, 191–208.
10. Khan, A.; Sohail, A.; Zahoor, U.; Qureshi, A.S. A survey of the recent architectures of deep convolutional neural networks. *Artif. Intell. Rev.* 2020, 53, 5455–5516.
11. Dhillon, A.; Verma, G.K. Convolutional neural network: A review of models, methodologies and applications to object detection. *Prog. Artif. Intell.* 2020, 9, 85–112.
12. Mo, Y.; Wu, Y.; Yang, X.; Liu, F.; Liao, Y. Review the state-of-the-art technologies of semantic segmentation based on deep learning. *Neurocomputing* 2022, 493, 626–646.
13. Singh, A.K.; Ganapathysubramanian, B.; Sarkar, S.; Singh, A. Deep Learning for Plant Stress Phenotyping: Trends and Future Perspectives. *Trends Plant Sci.* 2018, 23, 883–898.
14. Diba, A.; Sharma, V.; Pazandeh, A.; Pirsiavash, H.; van Gool, L. Weakly supervised cascaded convolutional networks. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition 2017, Honolulu, HI, USA, 21–26 July 2017*; pp. 914–922.
15. Alzubaidi, L.; Zhang, J.; Humaidi, A.J.; Al-Dujaili, A.; Duan, Y.; Al-Shamma, O.; Santamaria, J.; Fadhel, M.A.; Al-Amidie, M.; Farhan, L. Review of deep learning: Concepts, CNN architectures, challenges, applications, future directions. *J. Big Data* 2021, 8, 53.

16. Liu, J.; Wang, X. Plant diseases and pests detection based on deep learning: A review. *Plant Methods* 2021, 17, 22.
17. Jiang, Y.; Li, C. Convolutional Neural Networks for Image-Based High-Throughput Plant Phenotyping: A Review. *Plant Phenomics* 2020, 2020, 4152816.
18. Krizhevsky, A.; Sutskever, I.; Hinton, G.E. Imagenet classification with deep convolutional neural networks. *NIPS* 2012, 25, 84–90.
19. Szegedy, C.; Vanhoucke, V.; Ioffe, S.; Shlens, J.; Wojna, Z. Rethinking the inception architecture for computer vision. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition* 2016, Las Vegas, NV, USA, 27–30 June 2016; pp. 2818–2826.
20. Simonyan, K.; Zisserman, A. Very deep convolutional networks for large-scale image recognition. *arXiv* 2014, arXiv:1409.1556.
21. Huang, G.; Liu, Z.; van der Maaten, L.; Weinberger, K.Q. Densely connected convolutional networks. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition* 2017, Honolulu, HI, USA, 21–26 July 2017; pp. 4700–4708.
22. Zoph, B.; Le, Q.V. Neural architecture search with reinforcement learning. *arXiv* 2016, arXiv:1611.01578.
23. Montavon, G.; Samek, W.; Muller, K.R. Methods for interpreting and understanding deep neural networks. *Digit. Signal Process.* 2018, 73, 1–15.
24. Arrieta, A.B.; Diaz-Rodriguez, N.; del Ser, J.; Bennetot, A.; Tabik, S.; Barbado, A.; Garcia, S.; Gil-Lopez, S.; Molina, D.; Benjamins, R.; et al. Explainable Artificial Intelligence (XAI): Concepts, taxonomies, opportunities and challenges toward responsible AI. *Inf. Fusion* 2020, 58, 82–115.
25. Sermanet, P.; Eigen, D.; Zhang, X.; Mathieu, M.; Fergus, R.; LeCun, Y. Overfeat: Integrated recognition, localization and detection using convolutional networks. *arXiv* 2013, arXiv:1312.6229.
26. Girshick, R.; Donahue, J.; Darrell, T.; Malik, J. Rich feature hierarchies for accurate object detection and semantic segmentation. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition* 2014, Columbus, OH, USA, 23–28 June 2014; pp. 580–587.
27. Girshick, R. Fast r-cnn. In *Proceedings of the IEEE International Conference on Computer Vision* 2015, Santiago, Chile, 7–13 December 2015; pp. 1440–1448.
28. Ren, S.Q.; He, K.M.; Girshick, R.; Sun, J. Faster R-CNN: Towards Real-Time Object Detection with Region Proposal Networks. In *Proceedings of the 28th International Conference on Neural Information Processing Systems*, Montreal, QC, Canada, 7–12 December 2015.
29. Long, J.; Shelhamer, E.; Darrell, T. Fully convolutional networks for semantic segmentation. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition* 2015, Boston, MA, USA, 7–12 June 2015; pp. 3431–3440.

30. 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 Trans. Pattern Anal. Mach. Intell.* 2018, 40, 834–848.
31. Ronneberger, O.; Fischer, P.; Brox, T. U-net: Convolutional networks for biomedical image segmentation. In *Proceedings of the International Conference on Medical Image Computing and Computer-Assisted Intervention 2015*, Munich, Germany, 5–9 October 2015; pp. 234–241.
32. Hariharan, B.; Arbeláez, P.; Girshick, R.; Malik, J. Simultaneous detection and segmentation. In *Proceedings of the European Conference on Computer Vision 2014*, Zurich, Switzerland, 6–12 September 2014; pp. 297–312.
33. He, K.; Gkioxari, G.; Dollár, P.; Girshick, R. Mask R-CNN. In *Proceedings of the IEEE International Conference on Computer Vision 2017*, Venice, Italy, 22–29 October 2017; pp. 2961–2969.

---

Retrieved from <https://encyclopedia.pub/entry/history/show/78858>