Quality Engineering

ISSN: (Print) (Online) Journal homepage: https://www.tandfonline.com/loi/lqen20

A fresh look at computer vision for industrial quality control

Bart De Ketelaere, Niels Wouters, Ioannis Kalfas, Remi Van Belleghem & Wouter Saeys

To cite this article: Bart De Ketelaere, Niels Wouters, Ioannis Kalfas, Remi Van Belleghem & Wouter Saeys (2022) A fresh look at computer vision for industrial quality control, Quality Engineering, 34:1, 152-158, DOI: 10.1080/08982112.2021.2001828

To link to this article: https://doi.org/10.1080/08982112.2021.2001828

Published online: 13 Jan 2022.

QUALITY ENGINEERING
2022, VOL. 34, NO. 1, 152–158
https://doi.org/10.1080/08982112.2021.2001828

QUALITY QUANDARIES

OPEN ACCESS Check for updates

A fresh look at computer vision for industrial quality control

Bart De Ketelaere, Niels Wouters, Ioannis Kalfas, Remi Van Belleghem, and Wouter Saeys

Department of Biosystems – MeBioS, KU Leuven, Leuven, Belgium

KEY POINTS
This article draws attention to key developments in the field of computer vision for industrial quality control and provides some illustrative examples.

Introduction

Computer vision (CV) is a discipline that aims at developing techniques that enable computers to see and interpret the content of digital images. Although this task might appear to be trivial because we, as humans, are extremely capable of doing so, it remains a very active field of research more than 50 years after its origins. This paper provides recent directions in CV both from a hardware as from a software point of view, and gives the reader key considerations, advantages and limitations.

Industrial use for quality control

CV is used in many industrial applications for quality control. Quality traits that are considered mainly relate to color, size, shape and the absence of defects. Most straightforward applications are those where identical parts are to be manufactured, and where the CV system is used to detect deviations in any of these quality aspects. One can think of automotive or semiconductor parts that have a tight tolerance on their characteristics. More complex are those cases where the products show a large intra-class variability. This variability can be caused by the product itself (agro-food products are an excellent example—no two apples are identical), but can also be caused by a different position or orientation of the samples with respect to the camera itself. Because capturing images is easy and fast, CV is mostly used for online quality control at relatively high speeds, a situation that is often too challenging for humans.

Components of a computer vision system

A CV system generally consists of four components, being illumination, camera system, computer hardware and software. In its classical form, a vision system tries to mimic human vision in terms of color sensitivity. To achieve this a light source is used to mimic natural light. Nowadays, most often this light source is a white LED lamp because it provides a rather flat intensity all over the visible part of the electromagnetic spectrum (400 to 700 nm). A camera is then used to capture the scene of interest. Most cameras used in industry are silicon-based Complementary Metal-Oxide Semiconductors (CMOS), and can be panchromatic or color cameras. The panchromatic cameras integrate all light in their photosensitive region (400 nm to 1000 nm), which is broader than the visible region alone. Color cameras, in contrast, typically use a color filter array for arranging Red, Green and Blue (hence the name RGB) color filters on their digital image sensors. The digital image is then fed to a computer where it can be stored and (pre-)processed. The preprocessing stage mainly handles issues such as noise, non-uniform illumination, geometric distortion, improper focus, amongst others, and highlights regions and features that can be used for the actual processing task (Brosnan and Sun 2004).

Until recently, the main approach to process digital images was to first calculate important, handcrafted features, which should be distinctive but invariant to photometric and geometric transformations, and then using these features as an input for classification or regression algorithms. Extensive research has been (and still is) devoted to the calculation of such features from images because their quality mainly

CONTACT Bart De Ketelaere bart.deketelaere@kuleuven.be Department of Biosystems – MeBioS, KU Leuven, Leuven, Belgium
This article has been corrected with minor changes. These changes do not impact the academic content of the article.

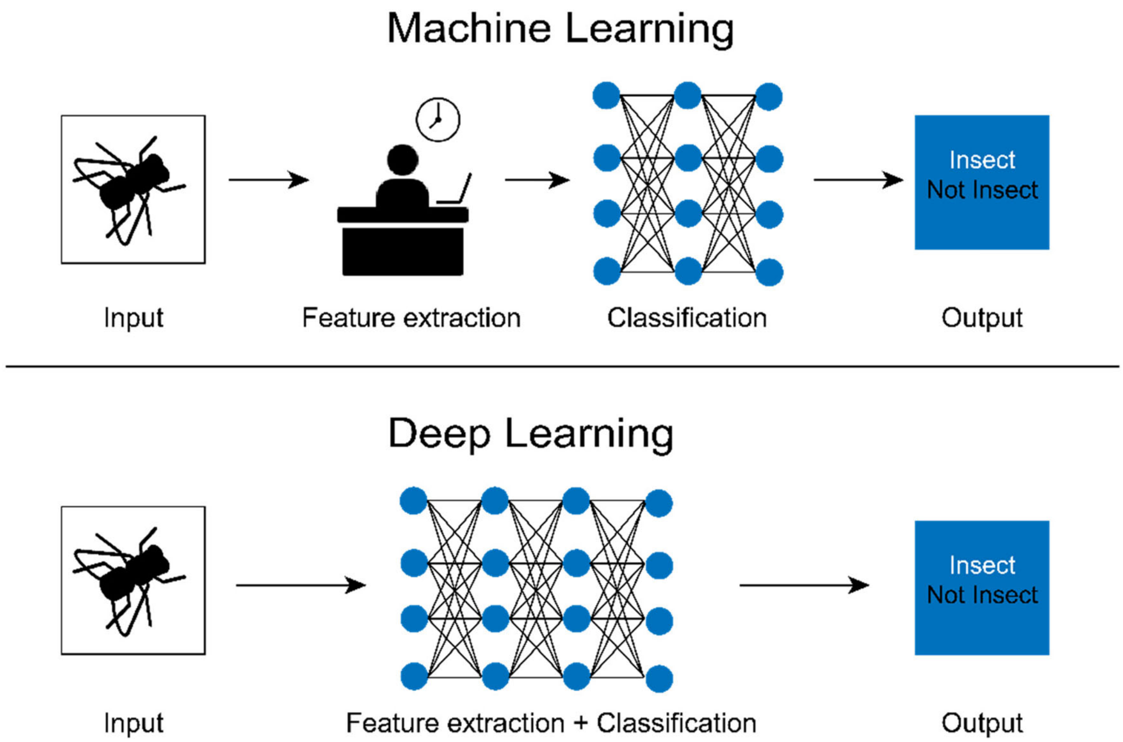

Figure 1. (A) The classical computer vision approach includes a first step where features are extracted from an image through a manual/human operation, and these features are then fed to a Machine Learning algorithm; (B) Computer vision based on Deep Learning provides a one-pass procedure during which the algorithm itself extracts the relevant features and processes them into a result.

determines the final result. An excellent review on the topic is provided in Li and Allinson (2008). Once the features have been extracted from the image, data analysis tools are used to classify the image, or to make predictions based on it. Traditionally, statistical and machine learning algorithms are used such as discriminant analysis, support vector machines, fuzzy logic, regression, but also neural networks.

Using these traditional red-green-blue (RGB) camera systems and performing feature-based image analysis has been the standard for many years, and still provides valuable results in many application fields. Improvements in camera sensitivity, resolution and computational power allowed for more challenging applications in terms of speed and general task complexity, but inherent limitations to the hard- and software hamper the full potential of CV. It is from this perspective that two rather recent developments are revolutionizing the field.

The *first development* is the rapid emergence of the field of Artificial Intelligence (AI) with specialized network architectures tailored to the analysis of images. These AI algorithms, and especially Deep Learning algorithms that prove to be powerful for image analysis, have a massive number of parameters to be tuned. This task was nearly infeasible more than ten years ago, but is enabled nowadays due to the availability of large databases of labeled images that can be used to train these networks. Also, the power of Graphical Processing Units (GPUs) that are capable of calculating a large number of (rather easy) operations in parallel – the typical setting encountered when training AI algorithms – has been instrumental.

The *second development* is related to the imaging system itself. As mentioned before, traditional imagers are based on silicium technology, inherently limiting their sensitivity to the 400–1000 nm range. Furthermore, this range is at most split into three different color regions (RGB) providing only a rough characterization of the scene that is imaged. This is especially true when realizing that a substantial number of organic molecules absorb light at specific wavelength that do not necessary align with the RGB bands. As such, the rough and rigid division of the light spectrum into three broad R, G and B bands may not provide sufficient information for all vision tasks. Instead, spectral cameras can split captured light into more, narrow color bands. Cameras capturing 4 to 10 distinct bands are typically termed 'multispectral', whereas 'hyperspectral' cameras capture a spectrum of up to several hundreds of bands, thus providing a detailed spectral fingerprint of the scene – for every pixel. Specialized cameras based on alternative detectors even allow to broaden the wavelength range in the (near) infrared beyond 1000 nm, providing more

Figure 2. (A) Four pictures of a target pest insect trapped on a commercial sticky plate. (B) Four examples of insects of another class trapped on a sticky plate.

detailed chemical information of the sample, or allowing to visualize aspects that remain invisible to the human eye.

In the next two sections, we discuss both major developments more in detail and give some considerations for the practitioner.

Deep learning revolutionizes image analysis

As mentioned before, calculating features from images and feeding them to a Machine Learning algorithm has been the standard in CV for many years (Figure 1A). It is especially the calculation of relevant features from the images that requires most attention, as these should be representative for the application and calculated in a precise and robust manner for every object in the image. Because appropriate calculation of important features takes away complexity from the ultimate classification or prediction model, fairly 'simple' algorithms, having a limited number of parameters to tune, are desirable. In the early years of CV, feedforward artificial neural networks (ANN) were already proposed for this purpose. These ANNs usually had only one or two so-called hidden layers composed of processing nodes that are called neurons and that connect the input (in this case the extracted features from the image) to the output (a class or number). They were trained in a "supervised" way meaning they were presented with input data and matching output labels. After seeing multiple instances of the input data, they were able to predict the output labels of novel incoming data.

As an illustrative example, we can look at the automatic classification of flying insects based on images. This is of high relevance in agriculture, where pest insects have to be detected, and public health, where viral vector insects such as mosquitos have to be monitored. In both cases, the common practice is to trap insects on so-called sticky plates that are collected and manually inspected by trained personnel. It provides an excellent example of what CV is all about – using a computer to see and interpret digital images. Figure 2 shows in total eight different images belonging to two insect species (classes). Insects belonging to the left class are harmful to a target crop (in this case Belgian Endive), whereas insects from the right class are regular flies that are not harmful to that crop. It can readily be seen that the orientation of the insects is completely random, posing challenges to the feature extraction step in the classical computer vision approach.

In contrast to the classical approach, Deep Learning (DL) involves using an architecture that encompasses many intermediate layers (hence the name *deep*) instead of just a few (Figure 1(B)), yet still training it in a supervised manner. Most DL computer vision algorithms use convolutional neural networks (CNNs) as a basis. CNNs have a multi-layered architecture that gradually reduces the input data (e.g., the image) to the most relevant representation. This representation is then compared to known data to classify the complex input. As the name suggests, convolutional layers are central in CNNs. In essence, they consist of the application of a mathematical filter to an input to produce an activation. When this filter is repeatedly applied to different parts of an input, it results in a map of activations. These activations denote the locations and strengths of a detected feature in the input. Besides the convolutional layers, CNNs often also encompass other types of layers, such as pooling layers and fully connected layers. A pooling layer pools data points from a previous layer into a single, new data point, while a fully connected layer connects every neuron (unit) in a given layer to every neuron in another layer. It is the combination of these types of layers and filters that determines the architecture of a given network. One of the first networks that received wide attention was AlexNet, which has five convolutional layers and three fully connected layers (Krizhevsky, Sutskever, and Hinton 2012). This marked the start of many different architectures being proposed, such as GoogleNet and VGGNet (Simonyan and Zisserman 2015), ResNet (He et al. 2016), DenseNets (Huang et al. 2017) and Xception (Chollet 2017).

The advantage of complex DL architectures is that they can handle both the feature extraction and the final classification or prediction phase (Figure 1(B)). This simplifies the task for the user who no longer has to define and extract relevant features. However, this flexibility comes at the price of a more complex model that is more difficult to understand. Therefore, a significant amount of data is required to finetune the often large number of parameters that need to be estimated in the DL approach.

The data-greedy property of DL methods has been partially alleviated by the use of so-called pre-trained networks. These networks are trained on large image datasets that are publicly available. ImageNet is the most often used (Deng et al. 2009). It contains about 14 million of annotated images covering as much as 20,000 categories. One can then use these pretrained networks to analyze different yet similar tasks – a procedure that is called "transfer learning". Two main approaches are seen here. The first uses the parameters of a pre-trained network for initializing the new network, while the second freezes the parameters of the pre-trained network and only re-trains the final layer (or few layers) to represent the classes of interest in the new task (Weiss, Khoshgoftaar, and Wang 2016). As DL methods are highly nonlinear and complex, the interpretation is not straightforward. To overcome this important limitation, much attention is nowadays given to development of so-called "explainable Artificial Intelligence" (XAI).

What is essential to the practitioner is the way networks are trained for future use under practical conditions. Evidently, the set of images that is used to train a network should be representative for what is expected in the field. In order to cover more variation than what is sometimes available in the training set, data augmentation is often used to provide a richer dataset. The augmentation step can include geometric transformations, color space augmentations, kernel filtering, mixing of images and randomly erasing parts of the image (Shorten and Khoshgoftaar 2019). Next to the training set, a validation set is necessary to tune the hyperparameters of the network and select an appropriate final model which generalizes well beyond the training set. Once the final model has been built, a representative test set is given to the final model to verify that the model is generalizing well on new data. As this is a crucial step before deploying the model in production, the selection of an appropriate test set should be performed carefully. Given the high complexity and substantial number of parameters that need to be estimated, DL algorithms are prone to overfitting, finding patterns that are not per se linked to the output alone, but also to the dataset itself. It thus has to be avoided at all times that there is any dependence between the test set and the calibration/validation set. Typically, it is advised to use for the test set a completely different production run, taken at a different time. This way, the test performance of the model will be closest to what can be expected in practice. An example with considerations on the topic is given in Kalfas, De Ketelaere, and Saeys (2021).

Hyperspectral imaging systems allow to see the invisible

A second major trend in computer vision is the availability of advanced imaging systems that go beyond traditional panchromatic or Red, Green and Blue (RGB) – a color scheme mimicking the human eye, where the different cone cells are either sensitive to red, green or blue light, providing us with color perception. By mimicking the human eye, patterns that are invisible to us will be invisible for the computer vision systems well. As many CV applications aim at replacing human visual inspection, this is acceptable for many applications. However, for other applications this limitation is not desirable. An illustrative example is given in Figure 3. On the left side, a typical color printed food package filled with slices of cheese is shown, as seen by a classical RGB camera. At the bottom, part of a slice of cheese is trapped in the sealing area, causing the package to lose its much desired air tightness. Because the packaging material is printed, it is virtually impossible to see the presence of unwanted material in the sealing area. Hence, computer vision with traditional camera systems is inappropriate for the application.

(A) (B)

Figure 3. (A) Image of a typical sealed food package taken with an RGB camera. (B) The same package now seen through a hyperspectral camera, enabling to detect an anomaly in the sealing region of the package (Source: http://www.engilico.com. All Rights Reserved.).

In order to broaden the scope of computer vision, the interest in systems that allow to "see" more than the human eye has spurred over the last two decennia (Lu et al. 2020). One of the most promising directions is the development of multi- and hyperspectral imaging systems. Their main idea is to build imaging systems that expand the working principle of RGB cameras in two directions.

First, these systems split up the covered wavelength range of the camera into a larger number of more narrow color bands. This is depicted in Figure 4. At the left side, the working principle of an RGB camera is schematically presented. For each pixel in the image, three different values are generated, representative for the amount of Red, Green and Blue light. At the right side, the hyperspectral counterpart is visualized. Instead of having only R, G and B values for each pixel, a large number of intensities is obtained, resulting in a spectral fingerprint. The resulting hyperspectral image thus consists of two spatial dimensions (x and y), and one spectral dimension that together build up the so-called hypercube.

The reason for the rapid increase in hyperspectral technology is the fact that the hardware has improved drastically and has become more affordable. Hyperspectral cameras in the silicon range, i.e., up to 1000 nm, are nowadays available through multiple providers lowering the entry barrier for industrial adoption. Moreover, during the last years also the wavelength range up to 1700 nm and even 2500 nm has become available for routine industrial use. This availability has increased the interest of companies wanting to perform quality measurements on a wide variety in products for which computer vision was not appropriate before.

Second, the acquisition, analysis and interpretation of the data generated by hyperspectral cameras underwent a true revolution, with storage space becoming virtually unlimited, computer power that keeps on increasing, and novel algorithms allowing for accurate analysis of the hyperspectral images.

Above Figure 3(B) shows the same sealed food package, now seen through a very specific wavelength of a hyperspectral camera that is sensitive outside the visible part of the electromagnetic spectrum. The color printed packaging material is much more transparent in this wavelength range, which allows to "see" through it at the given wavelength. It is clear that based on such an image, computer vision algorithms will be able to detect these types of defects.

These new trends in vision systems did also alter the characteristics of the images acquired, and pose new challenges to the data analysis tools to optimally exploit the information. Roughly, the algorithms that

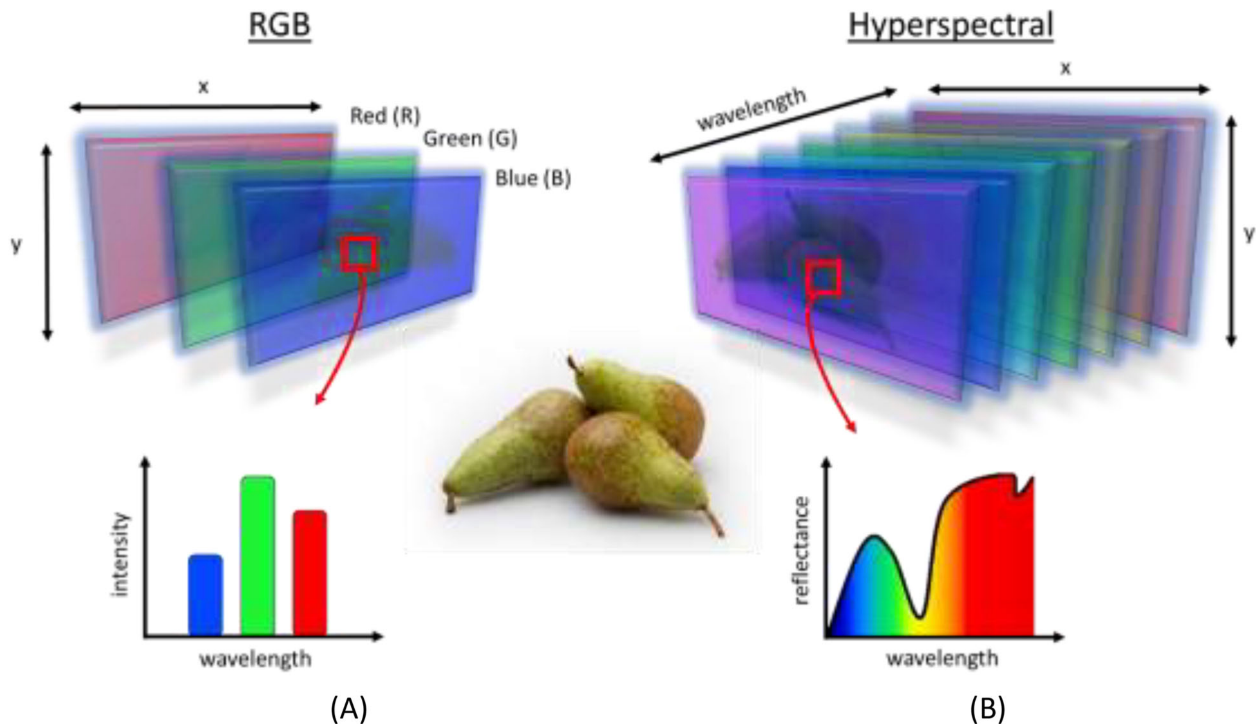

Figure 4. (A) In RGB imaging, each pixel contains three intensities, one for red, one for green and one for blue. (B) In hyperspectral imaging, a full spectral fingerprint for each pixel in the image is generated. This complex data structure is called a *hypercube*.

are used can be divided into two categories. These categories are linked to the dual spectral-spatial structure of hyperspectral data. The first category focuses on the spectral dimension of the data, and starts by processing each spectrum at a given pixel. In the field of chemometrics, which encompasses methods to analyze spectral data, advanced methods have been developed to preprocess (e.g., through scatter correction and various normalization methods (Saeys et al. 2019) and process the spectral data (e.g., using Partial Least Squares and derived methods). Afterwards, post-processing can be used to compensate for the fact that neighboring pixels convey similar information. The second category starts from the spatial dimensions and applies image processing techniques on each individual wavelength image. Next, it combines all this information to come to the final analysis. Recent research initiatives are devoted to the development of algorithms that inherently combine the spectral and spatial characteristics.

Hyperspectral deep learning

Given the two discussed major trends in computer vision, it is expected that their combination, which is still in its infancy, will lead to a new step-up in capabilities of modern computer vision. Despite being a logical path forward, it poses significant data analysis challenges. Deep Learning has revolutionized the computer vision field, but was only possible through the availability of vast databases of labeled images. This typically is not the case for hyperspectral data. Furthermore, the higher complexity of hyperspectral data compared to classical RGB images could require even more complex DL architectures, which, in turn, require even more data. Recent literature (Paoletti et al. 2019) proposes several avenues for solving this bottleneck. One of the most straightforward approaches is to condense the spectral information that is gained through hyperspectral imaging into a very limited number of key descriptors, e.g., by projecting the spectra on a lower dimensional space through Principal Component Analysis (PCA). Also other (nonlinear) algorithms are being developed for compressing the data (e.g., autoencoders). If the spectral data can be summarized by three descriptors, one could feed the compressed hypercubes into the available pre-trained networks and perform transfer learning based on a limited number of samples, just like what is classically done for normal RGB images.

Conclusions

Despite being more than 50 years old, computer vision is a very active field of research, and the recent breakthrough developments both in terms of hardware (hyperspectral cameras) and software (Deep Learning) have increased its performance significantly, leading to its increased adoption for industrial use for very complex cases. It is expected that the combination of these developments will even further strengthen the position of computer vision for industrial quality control.

About the authors

Bart De Ketelaere is a Research Manager at the Katholieke Universiteit Leuven (Belgium). He combines a master and a PhD in bio-engineering with a master in statistics and has developed a keen interest in industrial quality control. He is elected Honorary Member of the European Network of Business and Industrial Statistics (ENBIS) to recognize his significant contribution in the area of industrial statistics. He is the (co-)author of 146 ISI publications and is inventor in more than 10 patents. He is strongly involved in the valorization of research and in the creation of spin-off companies.

Niels Wouters is a postdoctoral research fellow in the MeBioS research group at KU Leuven. He specializes in sensor systems and data processing for agro-food applications, including hyperspectral vision technology.

Ioannis Kalfas is a PhD researcher in the MeBioS research group at KU Leuven. He specializes in data analysis and modelling using machine learning and deep learning methods.

Remi Van Belleghem is a PhD researcher in the MeBioS research group at KU Leuven. He explores how to improve quality control in the agro-food industry by utilizing deep learning methods for hyperspectral image analysis.

Wouter Saeys is a professor at the KU Leuven Department of Biosystems, where he leads the Biophotonics group with a focus on applications in the AgroFood chain. His main research interests include light transport modelling and optical characterisation of biological materials, chemometrics, agricultural automation and robotics. In 2013, he received the 'Young Statistician Award' from the European Network of Business and Industrial Statistics (ENBIS) for his work on multivariate calibration of spectral sensors in the agrofood industry.

References

Brosnan, T., and D. W. Sun. 2004. Improving quality inspection of food products by computer vision—a review. *Journal of Food Engineering* 61 (1):3–16. doi:10.1016/S0260-8774(03)00183-3.

Chollet, F. 2017. Xception: Deep learning with depthwise separable convolutions. *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, Honolulu, HI, USA, 1800–7.

Deng, J., W. Dong, R. Socher, L. Li, K. Li, and L. Fei-Fei. 2009. ImageNet: A large-scale hierarchical image database. IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Miami, FL, USA; 248–255, doi: 10.1109/CVPR.2009.5206848.

He, K., X. Zhang, S. Ren, and J. Sun. 2016. Deep residual learning for image recognition. IEEE Conference on Computer Vision and Pattern Recognition (CVPR): 770–8, doi: 10.1109/CVPR.2016.90, Las Vegas, NV, USA.

Huang, G., Z. Liu, L. Van Der Maaten, and K. Q. Weinberger. 2017. Densely Connected Convolutional Networks. Conference on Computer Vision and Pattern Recognition (CVPR), Honolulu, HI, USA; 2261–2269.

Kalfas, I., B. De Ketelaere, and W. Saeys. 2021. Towards in-field insect monitoring based on wingbeat signals: The importance of practice oriented validation strategies. *Computers and Electronics in Agriculture* 180:105849. doi: 10.1016/j.compag.2020.105849.

Krizhevsky, A., I. Sutskever, and G. E. Hinton. 2012. ImageNet classification with deep convolutional neural networks. NIPS 2012: Neural Information Processing Systems, Lake Tahoe, Nevada.

Li, J., and N. M. Allinson. 2008. A comprehensive review of current local features for computer vision. *Neurocomputing* 71 (10–12):1771–87. doi:10.1016/j.neucom.2007.11.032.

Lu, Y., W. Saeys, M. Kim, Y. Peng, and R. Lu. 2020. Hyperspectral imaging technology for quality and safety evaluation of horticultural products: A review and celebration of the past 20-year progress. *Postharvest Biology and Technology* 170:111318. doi:10.1016/j.postharvbio.2020.111318.

Paoletti, M. E., J. M. Haut, J. Plaza, and A. J. Plaza. 2019. Deep learning classifiers for hyperspectral imaging: A review. *ISPRS Journal of Photogrammetry and Remote Sensing* 158:279–317. doi:10.1016/j.isprsjprs.2019.09.006.

Saeys, W., D. Nguyen, R. Van Beers, and B. Nicolai. 2019. Multivariate calibration of spectroscopic sensors for postharvest quality evaluation: A review. *Postharvest Biology and Technology* 158:110981. doi:10.1016/j.postharvbio.2019.110981.

Shorten, C., and T. M. Khoshgoftaar. 2019. A survey on image data augmentation for deep learning. *Journal of Big Data* 6 (1):1–48. doi:10.1186/s40537-019-0197-0.

Simonyan, K., and A. Zisserman. 2015. *Very deep convolutional networks for large-scale image recognition*. 2015 3rd International Conference on Learning Representations (ICLR), San Diego, CA, USA.

Weiss, K., T. M. Khoshgoftaar, and D. Wang. 2016. *A survey of transfer learning*. *Journal of Big Data* 3 (1):40. doi: 10.1186/s40537-016-0043-6.