Matrox Imaging White Paper

Deep Learning: Its Proper Role and Use in Machine Vision

www.matrox.com/imaging

MATROX IMAGING

Abstract

Artificial intelligence, specifically machine learning by way of deep learning, is having a tremendous beneficial impact on the world at large. Deep learning technology is used in a myriad of applications, from giving virtual assistants the ability to process natural language, to enhancing the e-commerce experience through recommendation engines, to assisting medical practitioners with computer-aided diagnosis, to performing predictive maintenance in the aerospace industry. Deep learning technology is also a key enabler of Industry 4.0—the fourth industrial revolution that has occurred in manufacturing, specifically with the use of smart and autonomous systems fueled by data and machine learning—where machine vision technology is an important contributor. Crucial to note is that deep learning alone is not capable of tackling all manners of machine vision tasks, and requires careful preparation and upkeep to be truly effective. This white paper will detail how machine vision—the automated computerized process of acquiring and analyzing digital images primarily for ensuring quality, tracking, and guiding production—benefits from deep learning as the latter is making the former more accessible and capable.

Machine vision and deep learning: The challenges

Machine vision deals with identification, inspection, guidance, and measurement tasks commonly encountered in the manufacturing and processing of consumer and industrial goods. Conventional machine vision software addresses these tasks with specific algorithm and heuristic-based methods; these methods often require specialized knowledge, skill, and experience in order to be implemented properly. Moreover, these methods or tools sometimes fall short in terms of their ability to handle and adapt to complex and varying conditions. Deep learning is of great help but requires a painstaking training process based on previously collected sample data in order to produce the level of results generally required in industry (i.e., 3σ or at least 99.7% process accuracy). Moreover, more training is occasionally needed to account for unforeseen situations that can adversely affect production output. It is important to appreciate that deep learning is primarily employed to classify data and not all machine vision tasks lend themselves to this approach.

It is important to appreciate that deep learning is primarily employed to classify data and not all machine vision tasks lend themselves to this approach.

Where deep learning does and does not excel

As noted, deep learning is the process through which data—such as images or their constituent pixels—are sorted into two or more categories. Deep learning is particularly well suited for recognizing objects or object traits, for example, identifying that widget A is different from widget B, and so on. The technology is also especially good at detecting defects, whether the presence of a blemish or foreign substance, or the absence of a critical component in or on a widget that is being assembled. It also comes in handy for recognizing text characters and symbols such as expiry dates and lot codes.

The technology is also especially good at detecting defects, whether the presence of a blemish or foreign substance, or the absence of a critical component in or on a widget that is being assembled.

While deep learning excels in complex and variable situations such as finding irregularities in non-uniform or textured image backgrounds or within an image of a widget whose presentation changes in a normal and acceptable manner (see Figure 1), deep learning alone cannot locate patterns with an extreme degree of positional accuracy and precision. Analysis using deep learning is a probability-based process and is therefore not practical or even suitable for jobs that require exactitude. High-accuracy high-precision measurement is still very much the domain of traditional machine vision software. The decoding of barcodes and two-dimensional symbologies, which is inherently based on specific algorithms, is also not an area appropriate for deep learning technology (see Figure 2).

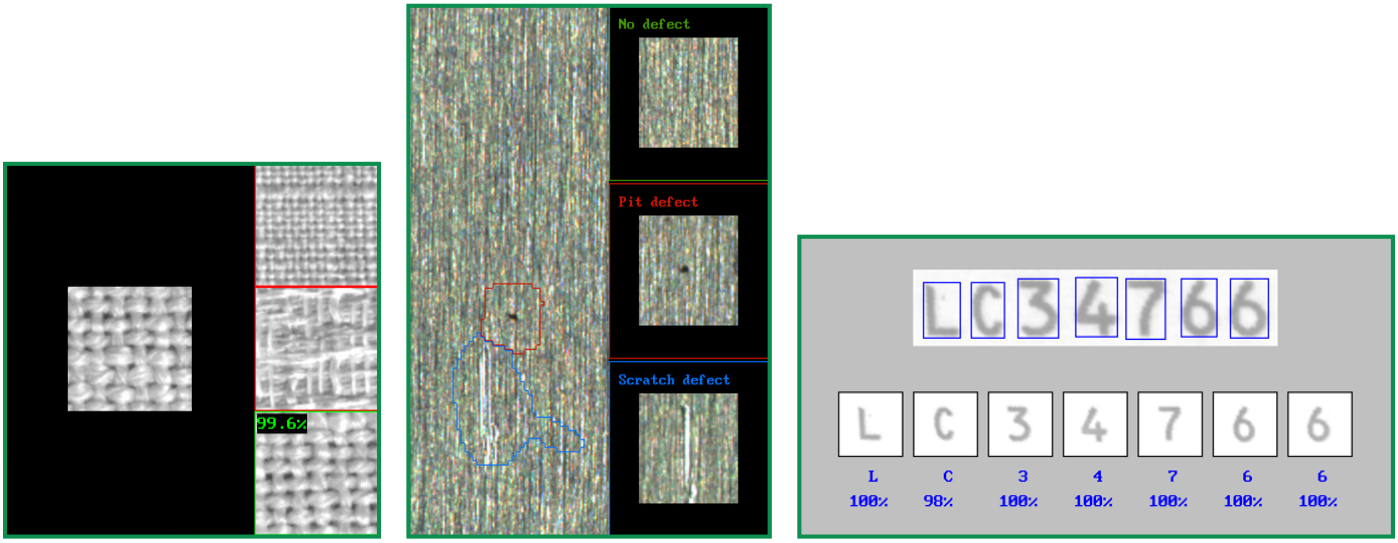

Figure 1. Where deep learning excels: Identification (left), defect detection (middle), and OCR (right).

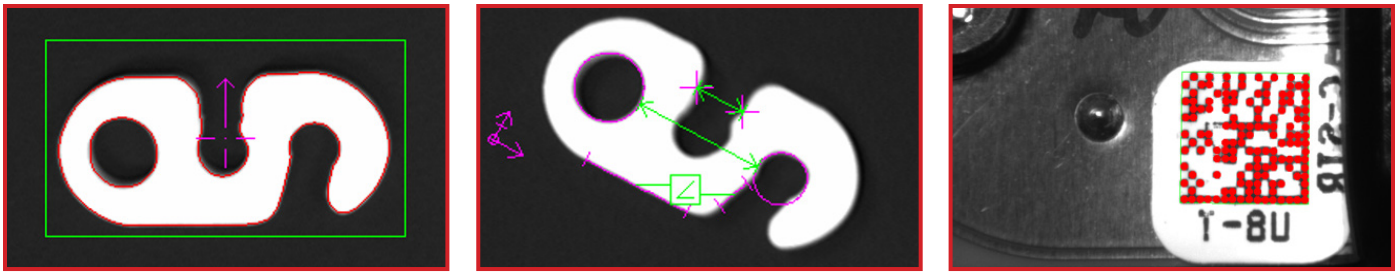

Figure 2. Where deep learning does not excel: High-accuracy high-precision pattern matching (left), metrology (middle), and code reading (right).

Deep neural networks

Deep learning is the latest manifestation of machine learning, which is itself subdivided into three distinct types, namely, supervised learning, unsupervised learning, and reinforcement learning. Supervised deep learning is the most common one used in business applications today.

Deep learning technology makes use of deep neural networks to perform its classification function. These neural networks take inspiration from the way the human brain processes sensory input in order to interpret data.

Deep learning technology makes use of deep neural networks to perform its classification function. These neural networks take inspiration from the way the human brain processes sensory input in order to interpret data. Specifically, deep learning technology leverages the convolutional neural network (CNN)—or close relatives thereof—to analyze images. The CNN may also be referred to as the *model*.

A deep neural network consists of a number of layers (see Figure 3). The *input layer* defines the image attributes the neural network must handle. The *hidden layers*—that can count two or more—extract features (i.e., edges, corners, etc.) of progressively higher complexity and establishes a feature space. Finally, the *output layer* establishes the classification based on the features retained and delivers what is called the *inference* or *prediction* result.

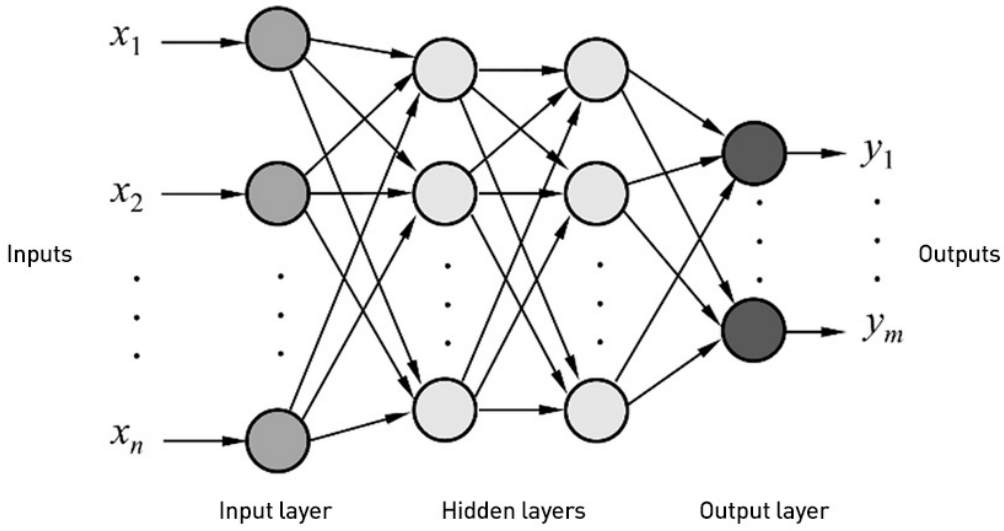

Figure 3. A deep neural network.

Preparing for training

To work, a CNN must first be trained using reference images in order to accomplish its function. Before undertaking the training, users must take the time to carefully collect and prepare a reference image set. Collection begins with a set-up comprising a camera, lens, and appropriate lighting that is identical, or nearly identical, to the one to be used by the deployed system. Using a substantially different set-up, like a mobile phone, will not lead to an applicable model.

The reference image set or dataset must be sufficiently large and representative of the expected application conditions; failure to do so will result in poor prediction results. The exact size of the dataset depends on the application complexity—such as the degree of subtlety in the differences, or how difficult it is to discern differences, from one desired class to another—and variability of the application conditions. The desired type of training to perform also governs the size of the dataset.

Recommendations vary, however a typical dataset should be composed of roughly five hundred images per desired class. In certain instances, where applications are more straightforward (e.g., images with a single consistent foreground object and a uniform and steady background) or where different conditions can be readily synthesized, it might be possible to train using an initial dataset with fewer than the recommended number of images.

Acquiring images to produce a balanced dataset is evidently challenging; some of the possible process variations, like the appearance of irregularities, occur very infrequently. It is possible to overcome this inadequacy by synthesizing images using classical image processing, a process known as *data augmentation* (see Figure 4).

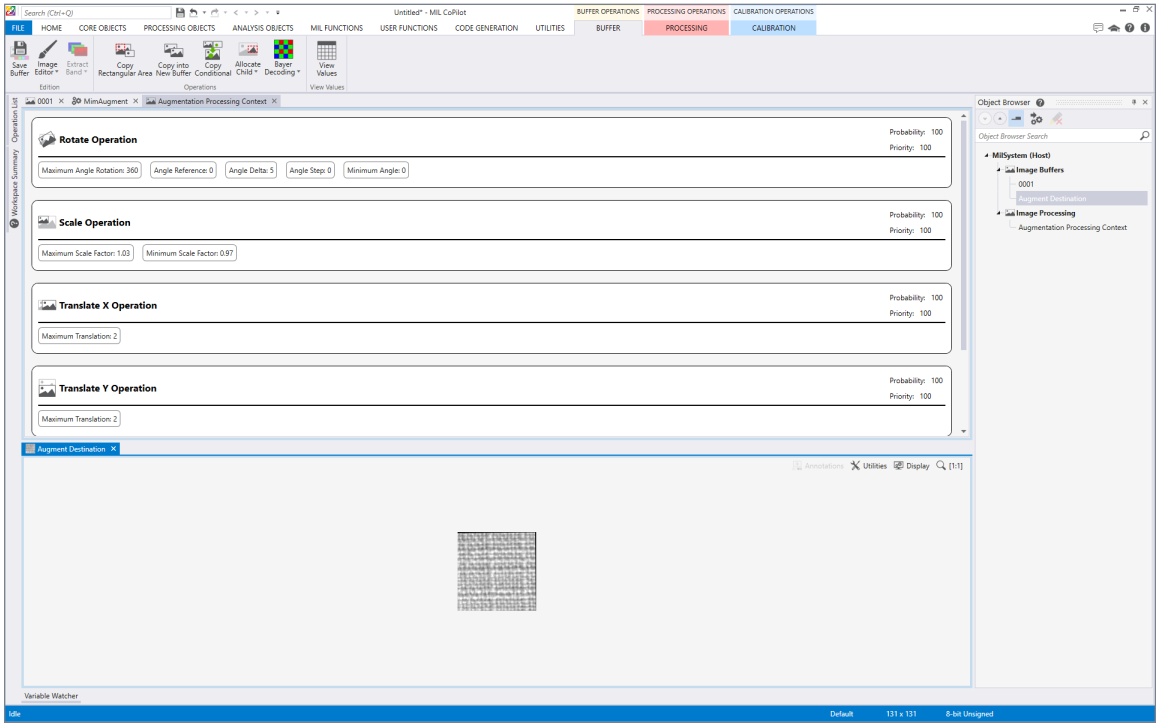

Figure 4. Augmenting the training dataset using classical image processing.

A dataset is ideally composed of so many images because it needs to be divided into at least two, preferably three, subsets prior to training. The first subset—the *trainset*—is the one used to actually train the model, which involves adjustments to mathematical weights. The *devset* is the second subset and is used to monitor the training process by tracking the difference between the delivered and expected classification outcomes. The third and optional subset is called the *testset*, and is used at the very end to independently assess the performance of the trained model; if the testset is not used, the devset would also take on this performance-assessment role.

In all cases, the dataset should anticipate all possible variations to be met, including potential changes in the appearance and presentation of the subject matter itself, as well as its environment. Care must be taken not to introduce ambiguities that could lead to an unwanted bias for a specific class, which would result in poor predictions.

In all cases, the dataset should anticipate all possible variations to be met, including potential changes in the appearance and presentation of the subject matter itself, as well as its environment (i.e., differences in illumination). Care must be taken not to introduce ambiguities that could lead to an unwanted bias for a specific class, which would result in poor predictions. Consider the following example: A series of images are taken of widgets; the images of widgets that meet the quality control standard are taken with a certain illumination, whereas the images of widgets that fail the same standard are taken with a different illumination. In this example, once the training is complete and the deep learning system deployed, it is entirely likely that the system will incorrectly classify images of good widgets should they appear with the latter illumination. This example illustrates one of the fundamental shortcomings of poor CNN training and the impact that can have on the predictions made, and thus the results obtained.

Once acquired, reference images must be carefully sorted into groups, one for each desired class; each reference image must also be labeled according to the specific desired class it belongs to. Effective labelling is required to establish the ground truth, which is critical for successful training (see Figure 5). Labelling of the dataset is so intrinsic to the effective training of the model that it must be performed by a subject matter expert, such as a quality control technician.

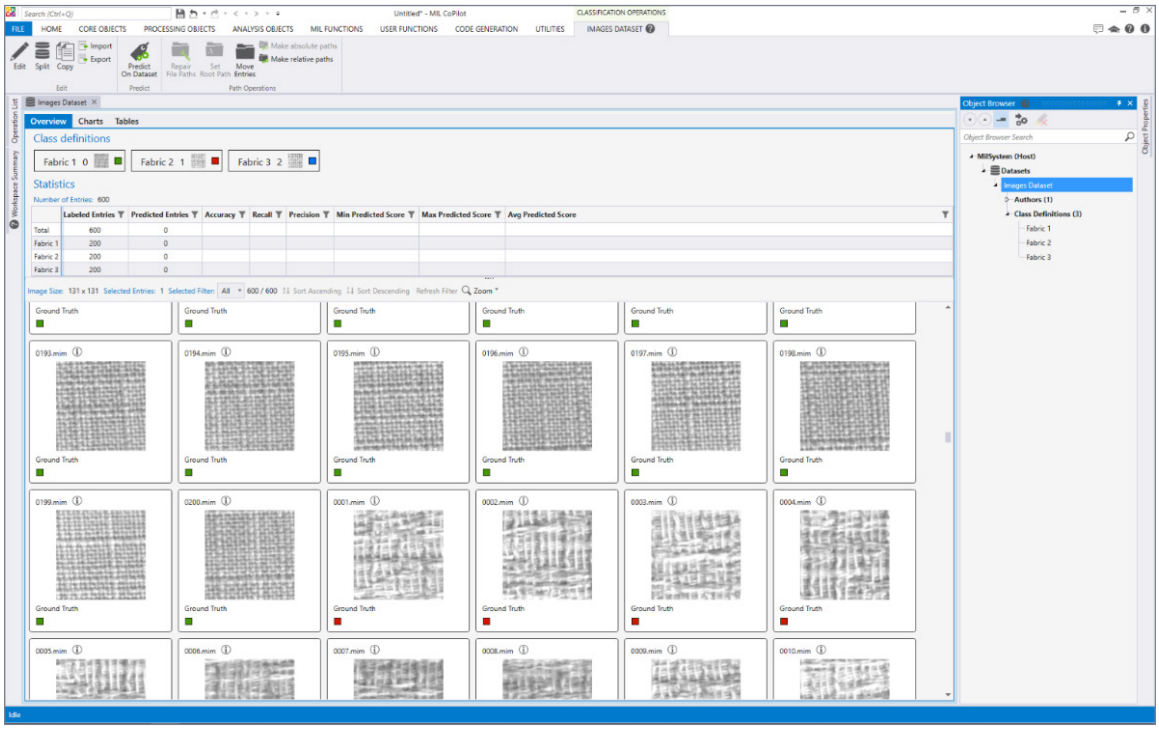

Figure 5. Organizing and labeling the training dataset.

Training and evaluating a model

The process of training a model occurs in one of three manners: from scratch, repurposed by way of *transfer learning*, or improved for through *fine-tuning*. The approach used depends on not only the objective but also the quantity of reference images.

Best used when the size of the training dataset is limited, transfer learning leverages what was learned in the training of a model for another use, specifically the feature extraction part.

Best used when the size of the training dataset is limited, transfer learning leverages what was learned in the training of a model for another use, specifically the feature extraction part. Fine-tuning is used when additional reference images of unforeseen conditions or borderline cases are acquired and labelled by the subject matter expert.

Training a model from scratch involves setting and adjusting numerous *hyperparameters*. Some of these are set before training commences. Fortunately, the default settings provided by commercial software are generally adequate. The adjustment of the remaining hyperparameters is automated through an iterative process and consists of altering the mathematical weights in the model to minimize the classification error (see Figure 6). Though the process can be lengthy, it is greatly aided by the massive parallel processing power provided by the graphics processing unit (GPU) in a computer.

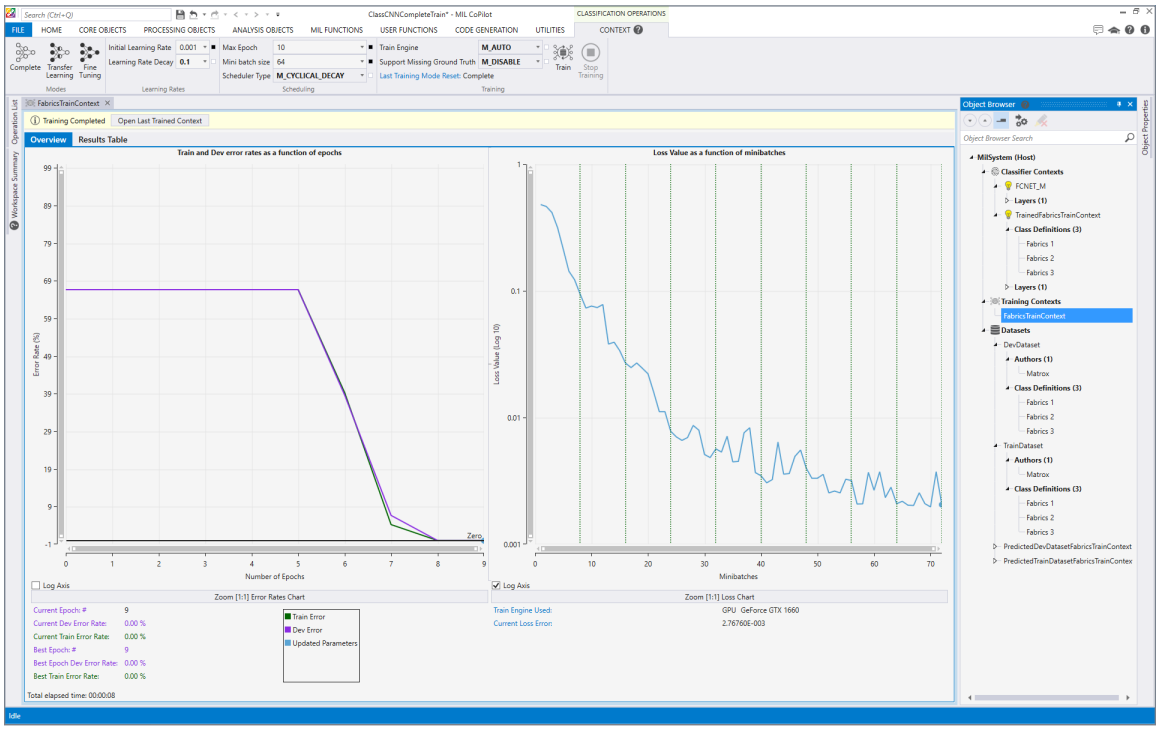

Figure 6. Graphs used to track the training process.

Once an optimal model is produced, its performance in terms of speed, accuracy, and robustness should be evaluated prior to deployment. The accuracy of a model is evaluated through the use of a *confusion matrix*, a table which summarizes the delivered and expected classification results per class. Robustness is determined by the separation in the distribution of the classification scores for each class (see Figure 7).

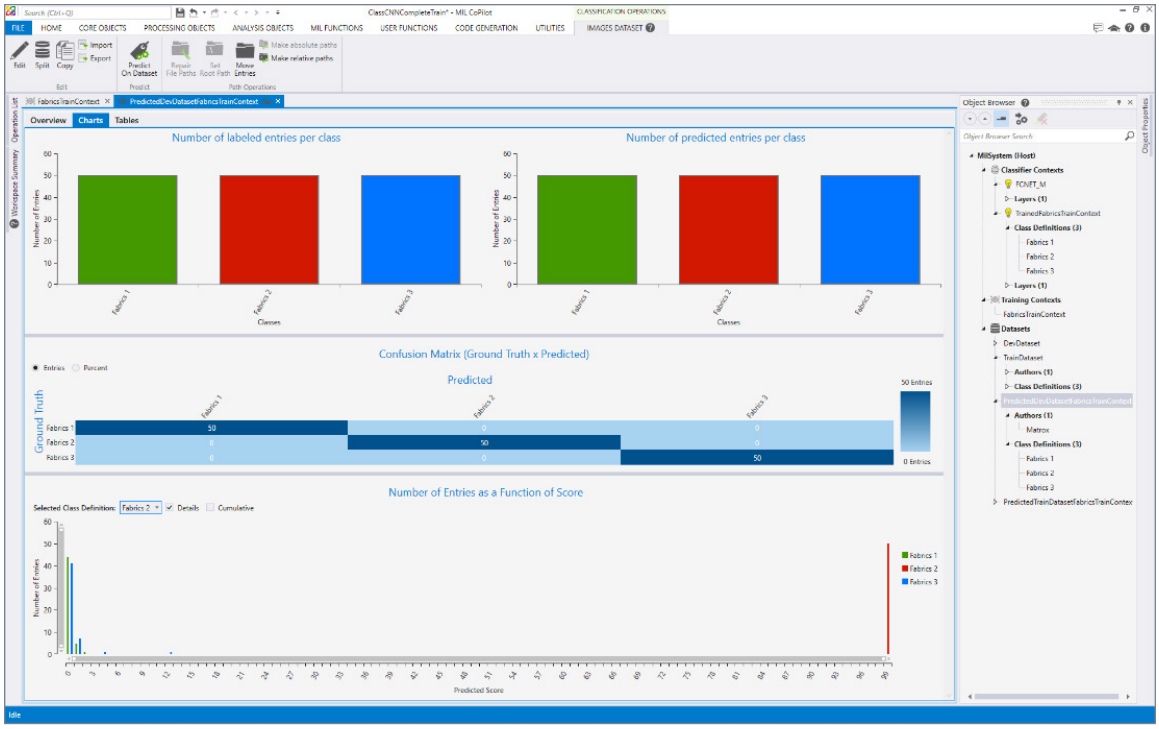

Figure 7. Confusion matrix (middle) and score distribution (bottom) to evaluate model quality.

Optimal unions

Now armed with the understanding of the suitability of deep learning for machine vision, it is clear that deep learning is a compliment to—and not a replacement for—the classic software used in automated visual identification, inspection, guidance, and measurement systems deployed in production plants and factories. Software that includes conventional tools as well as tools based on deep learning is thus the informed choice for developing machine-vision systems.

Software that includes conventional tools as well as tools based on deep learning is thus the informed choice for developing machine-vision systems.

While deep learning is a sufficiently mature technology so as not to absolutely require a machine-learning expert to put to use, the use of deep learning does require attentive preparatory work and deep application knowledge to be effective. Machine vision software with a user-friendly graphical interface is key for ensuring productivity with deep learning, whether preparing the training dataset, monitoring the training process, or analyzing the training results. Commercial software adds the element of dependable technical assistance for users, as they gain access to the knowledge and skill accrued from dealing with numerous applications across a multitude of industries over time.

Matrox Imaging software

Matrox Imaging offers two established software development packages that include classic machine vision tools as well as image classification tools based on deep learning. Matrox Imaging Library (MIL) X is a software development kit for creating applications by writing program code. Matrox Design Assistant X is an integrated development environment where applications are created by constructing and configuring flowcharts (see Figure 8). Both software packages include image classification models that are trained using the MIL CoPilot interactive environment (see Figure 6), which also has the ability to generate program code. Users of either software development package get full access to the Matrox Vision Academy online portal, offering a collection of video tutorials on using the software, including image classification, that are viewable on-demand. Users can also opt for Matrox Professional Services to access application engineers as well as machine vision and machine learning experts for application-specific assistance.

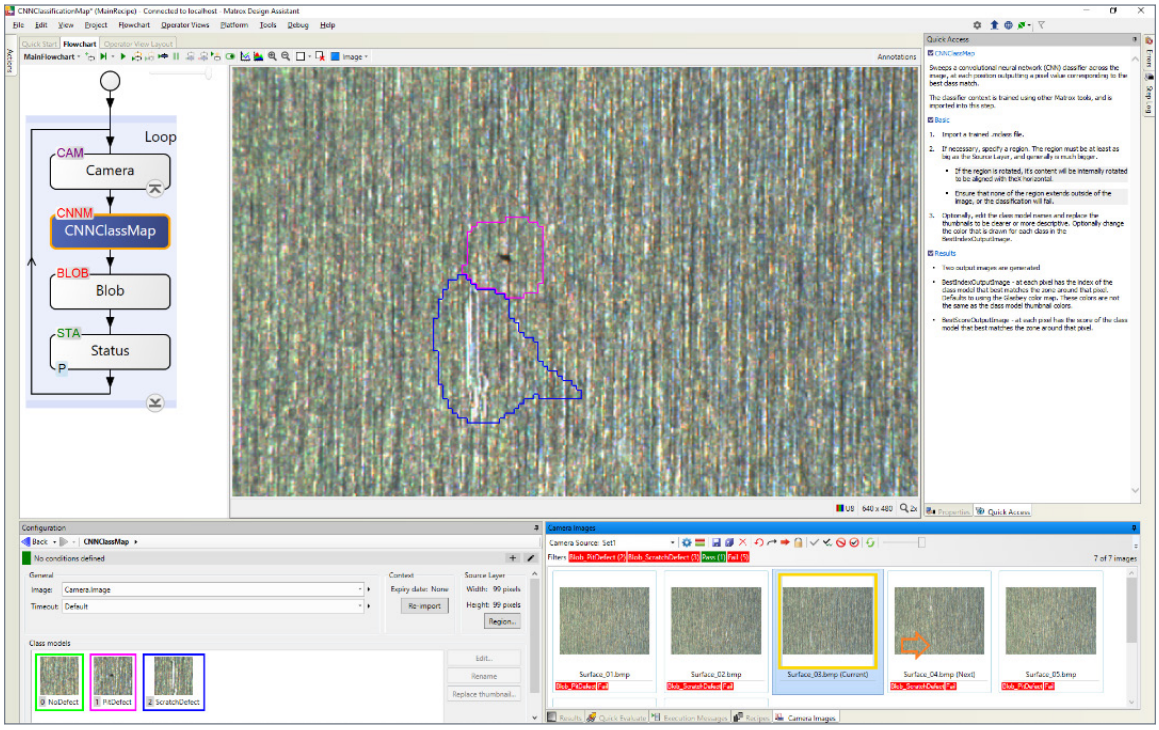

Figure 8. Image classification using deep learning with Matrox Design Assistant X.

About Matrox Imaging

Founded in 1976, Matrox® is a privately held company based in Montreal, Canada. Imaging, Graphics, and Video divisions provide leading component-level solutions, leveraging the others' expertise and industry relations to provide innovative, timely products.

Matrox Imaging is an established and trusted supplier to top OEMs and integrators involved in machine vision, image analysis, and medical imaging industries. The components consist of smart cameras, vision controllers, I/O cards, and frame grabbers, all designed to provide optimum price-performance within a common software environment.

Contact Matrox

imaging.info@matrox.com

North America Corporate Headquarters: 1 800-804-6243 or 514-822-6020
Serving: Canada, United States, Latin America, Europe, Asia, Asia-Pacific, and Oceania

www.matrox.com/imaging

MATROX IMAGING