LANDING AI

From Big Data to Good Data: Making AI Accessible to Machine Vision

CONTENTS

Introduction

Artificial intelligence (AI) is undisputedly the technology of the future. Big Silicon Valley tech companies have paved the way by developing complex algorithms that can learn from the massive amounts of data they collect every day — whether it's our search behavior in Google or the movies and series we watch on Netflix. In doing so, these companies have not only advanced the field of machine learning but have also helped prove that AI works.

AI technologies such as deep learning offer tremendous potential for accurate, effective, and flexible automated inspection processes. Still, AI adoption in machine vision systems has not picked up as fast as expected. To truly reap the benefits of deep learning, manufacturers must shift the paradigm away from code-driven development to a data-driven approach. By filling the gaps where rules-based machine vision falls short, deep learning software can push automated inspection systems into a new era of efficiency and capabilities.

Machine Vision Synergy

Rules-based machine vision systems have helped automate disparate inspection tasks over the past several decades, thus improving quality and increasing efficiency without slowing down the manufacturing process.

One major driver for this success is that mass production of standardized goods in highly automated factories offers ideal conditions for machine vision implementation. Manufactured products are standardized, and manufacturing processes are implemented in a controlled environment, making automated machine vision inspection a viable option. Specifically, the speed of the production line, the shape of the objects, the type of defects to be identified, and the lighting conditions remain consistent throughout each production run.

However, rules-based machine vision reaches its limits when the product or the process includes more variability. Examples include the processing of organic products, such as fruit, vegetables, and wood, which by their very nature are not standardized. In some production processes, product defects can look different in images acquired at different times of day due to changes in ambient lighting. In other cases, cosmetic defects, such as scratches, can be hard to differentiate from smudges, dust, or other surface contaminants. Deep learning can help improve each of these applications, making machine vision solutions more flexible and adaptable to variations.

What can cause rules-based machine vision to break down?

- Organic products
- Changes in ambient lighting
- Cosmetic defects vs. smudges, dust, or surface contaminants

An Uphill Climb

Unlike traditional machine vision system designs, which use rules-based algorithms to define defects numerically, **AI solutions — also called deep learning (DL) machine vision solutions — create models of what good and bad parts look like in images, based on statistical analyses of images that have been expertly tagged by human operators**. In essence, DL software learns what is good and bad from its human teachers, following their lead but taking into account many more visual cues than human or traditional machine vision solutions bring to bear. As a result, DL machine vision solutions can enable more applications than traditional machine vision solutions, and at higher speeds.

Therefore the quantity of images of specific defects available to define a problem — for example, a missing cap on a bottle — can be limited.

Lack of data can result in deep learning machine vision solutions that are not effective or systems that require a lot of time to develop because additional defect data must be acquired to improve the DL model, either through quality assurance systems on existing production lines or through simulation of defective images. This lack of acceptable defect datasets, plus a dearth of experienced AI designers, can slow the development and validation of inspection systems from prototyping to actual series production.

Deep learning software learns what is good and bad from its human teachers, following their lead but taking into account many more visual cues than human or traditional machine vision solutions bring to bear.

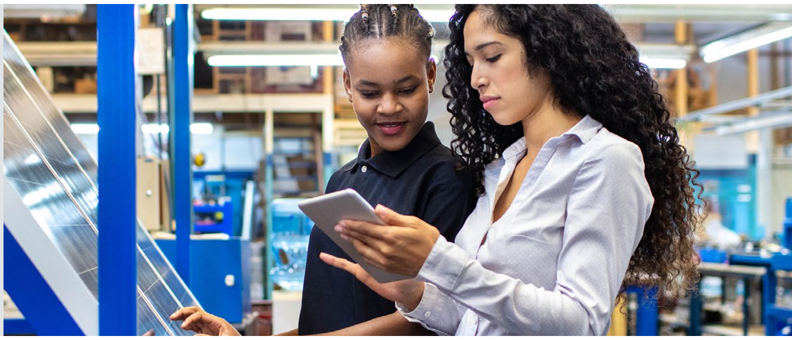
While AI offers much promise in manufacturing, machine vision professionals face many hurdles when it comes to implementing deep learning in practice. Scale is one challenge. AI pioneers like Google have billions of users, all of them doing similar things — for example, searching for keywords or watching videos on YouTube. Their AI is fed by huge amounts of data, and the type of data is homogeneous. In machine vision, this is not the case, as applications range from printed circuit board inspection to fruit sorting to bottle fill-level control. The data is heterogeneous, and this diversity means that each application segment is smaller. Additionally, modern manufacturing produces a small percentage of defective products; the majority are built to correct specifications.

For all these reasons, adoption of AI in industrial machine vision applications has not been as fast as expected, and many manufacturers are hesitant to make the move — not because they doubt the potential benefits of deep learning for their manufacturing enterprises but because the barriers to success seem too high.

Better Data, Better Results

AI can adapt to the specificities of factory-floor applications. According to Dr. Andrew Ng, CEO of Landing AI, the AI industry has been mainly focused on collecting massive amounts of data. However, he adds that **shifting from a "big data" approach to a "quality data" approach can be the paradigm change that lowers the barrier to more widespread AI deployment in manufacturing**.

According to AI theory, if just 10% of data is mislabeled, manufacturers need 1.88 times as much new data to achieve a certain level of accuracy. If 30% of the data is mislabeled, manufacturers need 8.4 times as much data compared to a situation with clean data.



In practice, a data-centric AI design approach means that the designer focuses on the quality of the data used to train the AI model rather than trying to tweak the resulting AI/DL model by changing specific values or changing the statistical methods used to sample the images and create the model.

Ng suggests that manufacturers focus on correctly classifying, grading, and labeling defect images rather than just focus on the number of images. According to AI theory, if just 10% of data is mislabeled, manufacturers need 1.88 times as much new data to achieve a certain level of accuracy. If 30% of the data is mislabeled, manufacturers need 8.4 times as much data compared to a situation with clean data. So it quickly becomes apparent that a data-centric AI/DL design platform that is machine learning operations (MLOps) compliant can save considerable time and energy, while placing the responsibility for system optimization back on the customer, who needs to provide quality data.

AI Design Environments Built with Humans in Mind

New MLOps design environments, such as Landing AI's LandingLens deep learning software, make deep learning system design available to non-programmers while giving them the tools to optimize highly effective data-centric machine vision design. **The software's tools are designed for manufacturing experts, not AI programmers, with a goal of leveraging the knowledge of experts on the factory line to maximize the accuracy and quality of the learning data**.

Different experts may have different opinions — and these may vary even within the same facility, let alone across a global enterprise. Effective DL design environments offer a way to automatically identify outliers in the labeled data, helping build consensus among quality inspectors and improving final system performance. This is often called creating a "defect book," with the best examples of what constitutes a defective product, guaranteeing that all participants agree on goals and what success will look like.

Consistent high quality product is the goal of every manufacturer, and their AI/ML teams realize that focusing on better data will drive better models which leads to faster and more reliable results.

After the acquisition of sample images, the first step is to label the images by identifying the presence and location of defects; both criteria are important. The criteria for these choices, however, may be ambiguous. For example, if a dent on a metal surface is a nonconformity, what about a small scratch?

Leverage In-house Knowledge

With optimally labeled training data, it becomes possible to train deep learning–based machine vision systems reliably, even if the quantity of data is limited. This data-driven approach has other benefits:

- **It builds on the in-house knowledge of manufacturing experts**, not outside data scientists, to tweak machine learning algorithms.
- Because the data is acquired, graded, and labeled by subject matter experts, **the system is tailored to the specific needs of the application**.
- This data-centric approach **ensures that the resulting model will be effective at automatically inspecting products**, reducing the time needed to optimize the system as it moves from development laboratory to production line.
- Finally, a data-centric process doesn't only train the commissioned machine vision solution to solve today's manufacturing problems; it also **simplifies the introduction of new products into existing production lines and helps system integrators adapt production to changing conditions**, such as new or aging equipment that impacts quality of the final product.

Machine vision is implemented in nearly all industries. Because of this utility, each application and each vision system is unique, facing its own special challenges, from data acquisition to data quality to support. In each case, a focus on data quality rather than quantity can help manufacturing industries reap the benefits of deep learning for their inspection systems now and into the future.

Quality managers, subject matter experts, and developers can work together during the development process to:

- **Reach a consensus on defects and labels**
- **Build a model**
- **Analyze results**
- **Make further optimizations.**

About LandingLens

LandingLens is an industry-first data-centric artificial intelligence (AI) visual inspection platform. It helps improve inspection accuracy and reduce false positives. The end-to-end platform standardizes deep learning solutions that reduce development time and scale projects easily to multiple facilities across the globe. Our focus remains on our customers and continual product innovation to solve the real-world problems of the manufacturing audience. To learn more, visit: www.landing.ai and follow Landing AI on Twitter and LinkedIn.

About Landing AI

Landing AI™ is pioneering the next era of AI in which companies with even limited data sets can realize the business and operational value of AI and move AI projects from proof-of-concept to full scale production. Guided by a data-centric AI approach, Landing AI's flagship product is LandingLens™, an enterprise MLOps platform that offers to build, iterate, and operationalize AI powered visual inspection solutions for manufacturers. With data quality being key to the success of production AI systems, LandingLens™ enables users to achieve optimal data accuracy and consistency. Founded by Dr. Andrew Ng, co-founder of Coursera, former chief scientist of Baidu, and founding lead of Google Brain, Landing AI is uniquely positioned to lead the development of AI from a technology that benefits a few to a technology that benefits all.

LANDING AI

sales@landing.ai

Part Number: 20004-01-EN