



# Deep Learning in Image Processing

## Neuronal Networks (CNN) on FPGAs in Industrial Applications

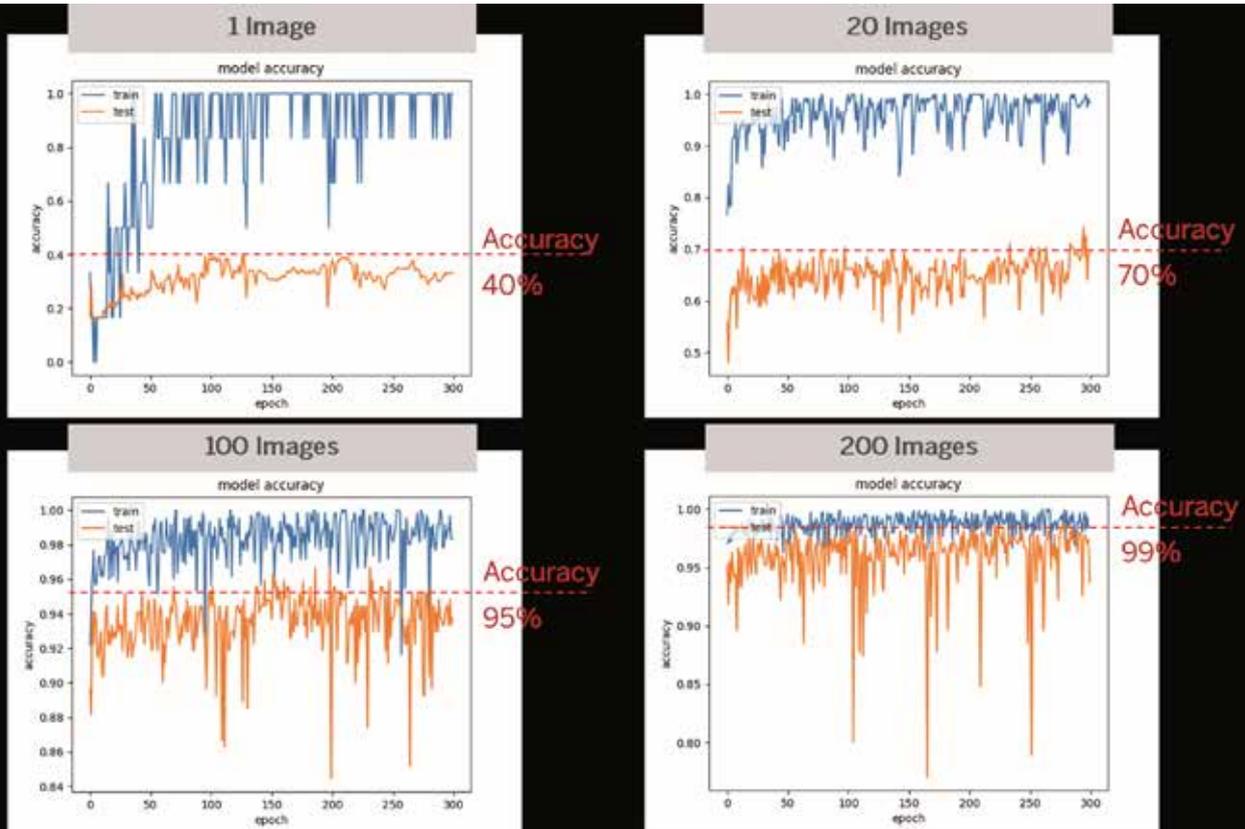


Image 1: Increase of prediction accuracy by increasing the training data/  
Source: Silicon Software GmbH

Deep learning in image processing is characterized by very high prediction accuracy in recognition, will improve the quality of today's image processing systems and open up new applications. The solution approach with deep neural networks like CNN (Convolutional Neural Networks) therefore takes over more and more tasks of classical image processing based on algorithmic description. Deep learning is becoming increasingly important in image classification (defect, object and characteristic classification), while image preprocessing, post-processing and signal processing continue to be carried out using existing methods. If tasks can be solved exclusively, more easily or with better results with Deep Learning, this displaces classical image processing - especially with complicated

variables such as reflective surfaces, poorly illuminated environments, varying lighting or moving objects. The translation invariance is one of its strengths, which means a very high expenditure with classical programming. Classic algorithms are advantageous when localization of objects or errors in an image, dimensional checking, code reading or post-processing are required.

Deep learning differentiates between the neural network's training and learning, implementation of the network — for example, on an FPGA — and inference, the execution of the network's CNN algorithmic upon images with output of a classification result. The more data that is used for training, the higher the predictive accuracy for classification

Traditional Machine Vision	Deep Learning
<b>Typical applications:</b> <ul style="list-style-type: none"> <li><input type="checkbox"/> Dimensional measurements</li> <li><input type="checkbox"/> Code reading</li> <li><input type="checkbox"/> Presence or absence checking</li> <li><input type="checkbox"/> Robot guidance</li> <li><input type="checkbox"/> Print inspection</li> <li><input type="checkbox"/> Automotive</li> <li><input type="checkbox"/> Electronics</li> </ul>	<b>Typical applications:</b> <ul style="list-style-type: none"> <li><input type="checkbox"/> Surface inspection (cracks, scratches)</li> <li><input type="checkbox"/> Food, plant, wood inspection</li> <li><input type="checkbox"/> Plastics, injection molding</li> <li><input type="checkbox"/> Textile inspection</li> <li><input type="checkbox"/> Medical imaging</li> <li><input type="checkbox"/> Agriculture</li> </ul>
<b>Typical characteristics:</b> <ul style="list-style-type: none"> <li><input type="checkbox"/> Rigid objects</li> <li><input type="checkbox"/> Fixed orientation</li> <li><input type="checkbox"/> Customer provides formal specification with tolerances</li> <li><input type="checkbox"/> Reliability 100%</li> </ul>	<b>Typical characteristics:</b> <ul style="list-style-type: none"> <li><input type="checkbox"/> Deformable objects</li> <li><input type="checkbox"/> Variable orientation</li> <li><input type="checkbox"/> Customer provides vague specification with examples of Good and Bad parts</li> <li><input type="checkbox"/> Reliability 99%</li> </ul>

*Image 2: Suitability of deep learning compared to classical image processing/ Source: Silicon Software GmbH*

will be. Due to the very high amount of data, GPUs are the right choice for training neural networks.

### Speed versus accuracy

With machine vision's special conditions surrounding rapid execution (inference) of CNNs with very low latencies, are the various processor technologies equally well suited for it? High bandwidth, low heat output, and long-term availability are in demand, along with speed and real-time demands for which conventional CPUs or GPUs alone are hardly suitable. These represent more appropriate solutions for image processing tasks in the non-industrial sphere, where lower throughput performance in favor of a more complex recognition requirement sometimes suffices. If we merely compare technical aspects, these technology platforms show different performance values which exclude their use in applications with high demands. Thus, inference time of a GPU is considerably shorter than the one on a CPU or special chips such as a TPU (TensorFlow Processing Units) or the Intel Movidius processor, but as a performance indicator averages lower than 50 MB/s.

When selecting a suitable network, small or mid-sized networks often suffice for typical image processing applications when only a few characteristics are to be classified. AlexNet, SqueezeNet or MobileNet are typical representatives for this. These are in good relationship between prediction accuracy, implementation size and computing speed or bandwidth in the machine vision. Here, it is clear to see that an acceptable loss of identifica-

tion with a concurrent gain in data throughput is achievable by choosing appropriate network architectures for the application demands in question – a possibility to optimize resources and to increase the classification quality.

### FPGAs and SoCs for inference

FPGAs as standalone processors or as SoCs together with ARM processors are best equipped for the demands upon inference of many image processing tasks, particularly of machine vision. FPGAs prove themselves with high parallelism of data calculation, guaranteed robust image acquisition and – in comparison to CPUs and GPUs – high processing power, image rate, and bandwidth. In so doing, CNNs on FPGAs classify at high throughput rates, something that fulfills particularly the time requirements of inline inspection.

The FPGA enables processing of image data – from acquisition to output and for device control – directly on a frame grabber or embedded Vision device without burdening the CPU, a quality which is particularly well suited for process-intensive applications like CNNs. Thus, smaller PCs without GPUs can be used, reducing overall system costs. The energy efficiency of FPGAs in the industrial temperature range is ten times higher in comparison to that of GPUs, making it ideal for embedded devices. This markedly expands the deep learning field of use with regard to Industry 4.0 as well as drones and autonomous driving.

The higher computational accuracy of GPUs and

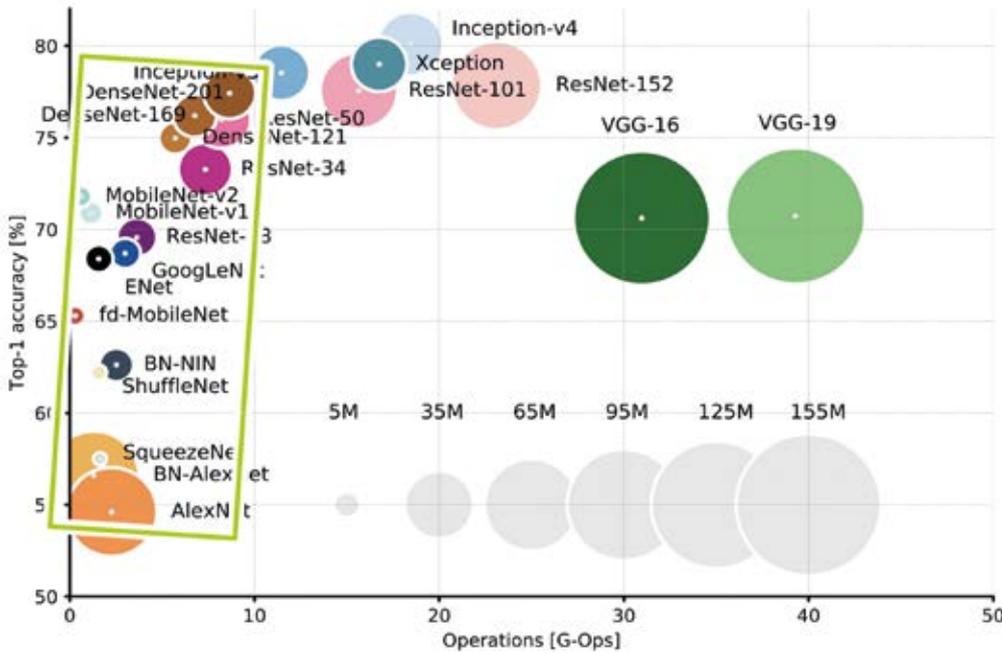


Image 3: "An analysis of Deep Neural Network Models for Practical Applications"/ Source: Alfredo Canziani/ Adam Paszke/ Eugenio Culurciello (2017)

thus higher prediction accuracy is bought by significantly shorter availability, higher power consumption, but also by lower data throughput. An exemplary comparison of data processing performance is 7.3 times higher for an FPGA-based solution than for a comparable system solution with GPUs.

### Optimize FPGA resources

For deep learning, there are various methods for saving resources without reducing the quality of classification. One important method is image scaling that reduces the internal data throughput.

Or the depth of calculation: Experience has shown that the depth of calculation only marginally affects the later prediction accuracy. The reduction from 32 bit to 8 bit and from floating point to fixed point/integer enables the FPGA to use the resources for larger network architectures or a higher data throughput. Thus it is possible to increase production speed in weld seam inspection or robotics, for example. Moreover, effective image preprocessing that reduces data enables use of smaller networks or FPGAs. These often suffice for simple classification tasks for error detection with few characteristics.

A 32 bit floating point GPU's higher computational accuracy is of little importance for deep learning inference, while 8 bit fixed point FPGAs achieve sufficiently precise prediction accuracy for most deep learning applications with negligible error tolerance. As regards requirements for especially precise computational accuracy, 16 bit fixed point can be implemented on a larger FPGA as a resource compromise.

For processing speeds required in production, high-output frame grabbers and embedded Vision devices such as cameras and sensors with larger FPGAs are already available. Using more comprehensive FPGA resources, more complex

	Frame grabber	GPU
Model	marathon VCL (Kintex7)	NVIDIA GTX1080
Availability	15 years +	~ 3 years
Max. TDP	15W	180W
Resolution	8bit Fixed-Point	32bit Floating-Point
Accuracy	98.7%	99.0%
Time per Image	0.128ms	0.94ms
Inferences per Sec.	7812	1063

Image 4: The FPGA performance is 7.3 times greater than that of the GPU/ Source: Silicon Software GmbH

architectures, and thus applications, can be processed. The higher data bandwidth enables processing of an entire image or additional image pre- and post-processing on the FPGA. It is high enough to analyze the entire data output of a GigE Vision camera using deep learning, to name one example.

## Result

Compared to classical image processing, a relatively high training effort in deep learning is more than offset by reliability and speed. FPGA technology on frame grabbers and (embedded) Vision devices enable use of neural networks for applications with industrial demands on real-time ability and low latencies (important for inline inspection), data throughput, bandwidth, and low heat output (important for embedded Vision), also for high resolution images. The long-term availability of FPGAs and frame grabbers guarantees a high level of investment security. Users profit from long-term savings due to rapid adaptability and lower overall system costs.

**microEnable 5 marathon deepVCL is a high-speed image processing solution for Deep Learning applications by VisualApplets.**

Due to its simple architecture, this is a robust and well established standard frame grabber for all industrial requirements and applications. With the large selection of cameras, there is always an adequate solution – especially for linescan cameras. An economical solution for higher bandwidth requirements, too.

The cable length has been considerably enhanced using advanced electronics.

- ◆ Compatible with all CL standardized cameras
- ◆ Compatible with “CNN Ready”
- ◆ Enables real-time image processing for Deep Learning
- ◆ Simple device control via the front trigger
- ◆ Secure control by the optional opto-decoupled trigger
- ◆ Onboard image preprocessing functions
- ◆ Industrial multi-device, multi-camera support
- ◆ Custom FPGA programming with VisualApplets supporting Xilinx Kintex FPGAs

