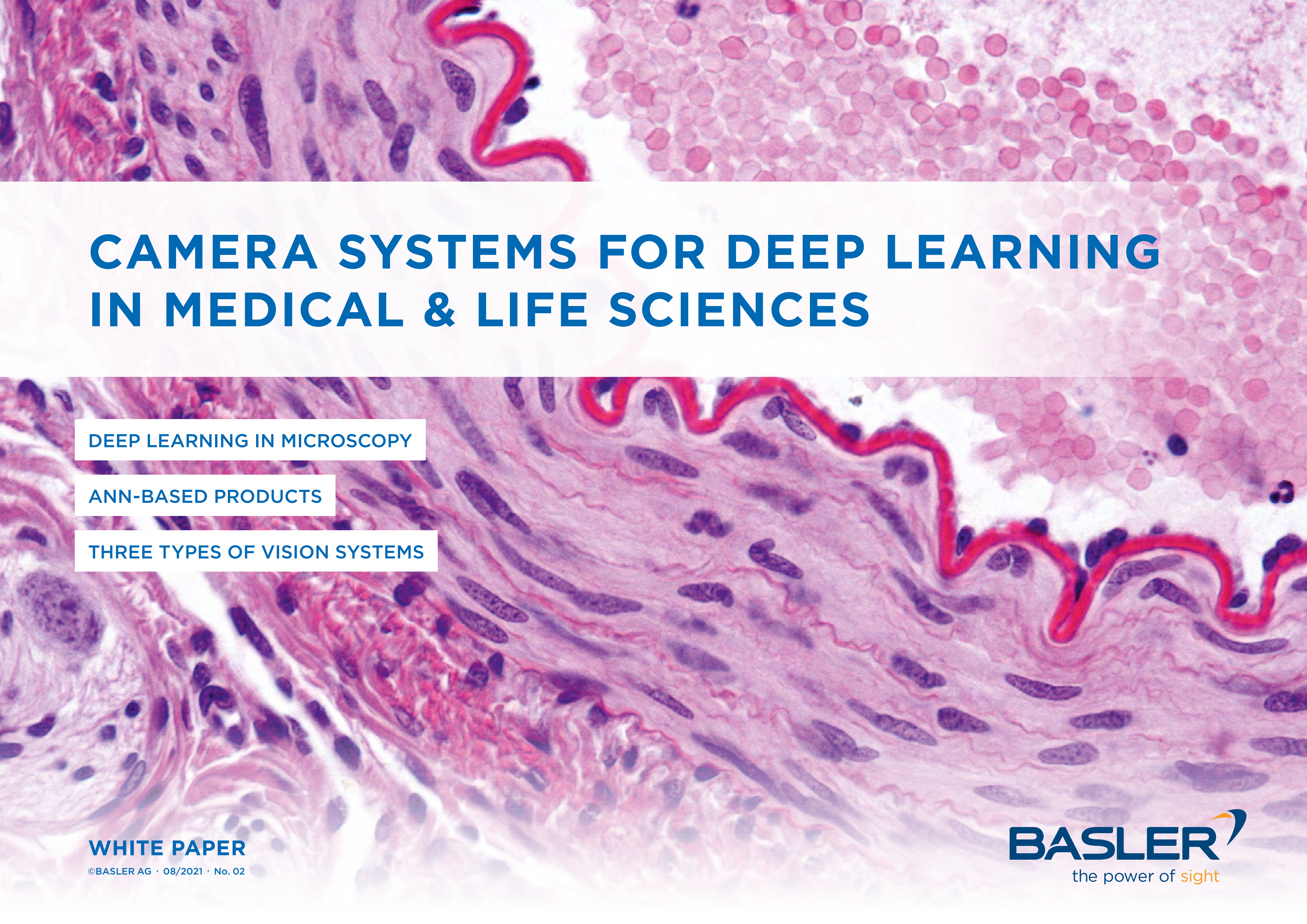


CAMERA SYSTEMS FOR DEEP LEARNING IN MEDICAL & LIFE SCIENCES

DEEP LEARNING IN MICROSCOPY

ANN-BASED PRODUCTS

THREE TYPES OF VISION SYSTEMS

WHITE PAPER

BASLER
the power of sight

The development and dissemination of algorithms for deep learning, especially in the field of Artificial Neural Networks (ANNs), is increasing rapidly. Even if not all applications benefit from this form of artificial intelligence, ANNs can in many cases significantly increase the accuracy and robustness of an application – and not only in the field of factory automation.

In the field of life sciences, deep learning algorithms enable completely new applications that were previously considered unfeasible for machines and are now being solved with impressive accuracy and reliability. This White Paper deals with the possible applications of camera-based vision systems for deep learning models in medical and life sciences.

1. INTRODUCTION

Deep learning algorithms already support medical imaging in many application areas, such as the classification of mammography examinations, prostate segmentation, segmentation of lesions in the brain or respiratory tract segmentation. The imaging methods include all methods – from X-ray and ultrasound examinations to computer and magnetic resonance tomography and optical coherence tomography.

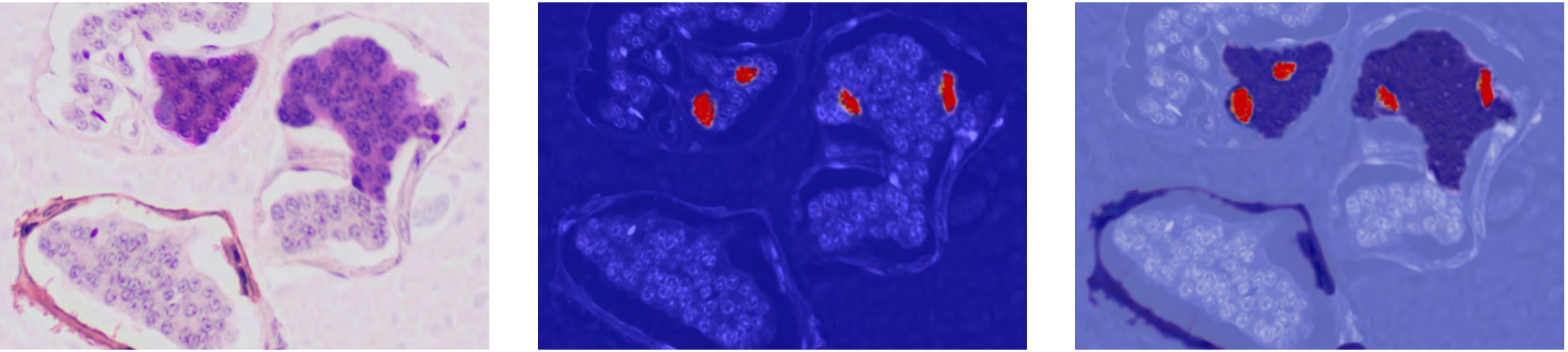
2. DEEP LEARNING IN MICROSCOPY

In the field of medical microscopy, two areas in particular offer great potential for deep learning applications: digital pathology and virtual staining of tissue sections.

2.1 Example: Digital Pathology

Digital pathology deals with the analysis of histological and cytological preparations in order to identify and analyze pathological changes in tissues or cells. These histo- or cytodiagnostic examinations are used, among other things, for the early detection of inflammation or developing tumors as part of preventive medical check-ups. Digital pathology is a core area of microscopy and covers a wide range of applications. Naturally, the potential of ANNs is very high in this area: especially the cooperation between artificial intelligence and human expertise seems promising. Automated diagnostic procedures have several advantages: for pathologists, the amount of scanned tissue sections is reduced and the need for manual inspection directly at the microscope is reduced, the result and the diagnostic statement and the corresponding therapeutic measures for patients are improved and, ultimately, costs are reduced.

***Figure 1:** Classification of suspicious regions in whole histological slices with deep learning similar to Wang et al.[1]*



Slide image *Probability map of cancer regions* *Overlay*

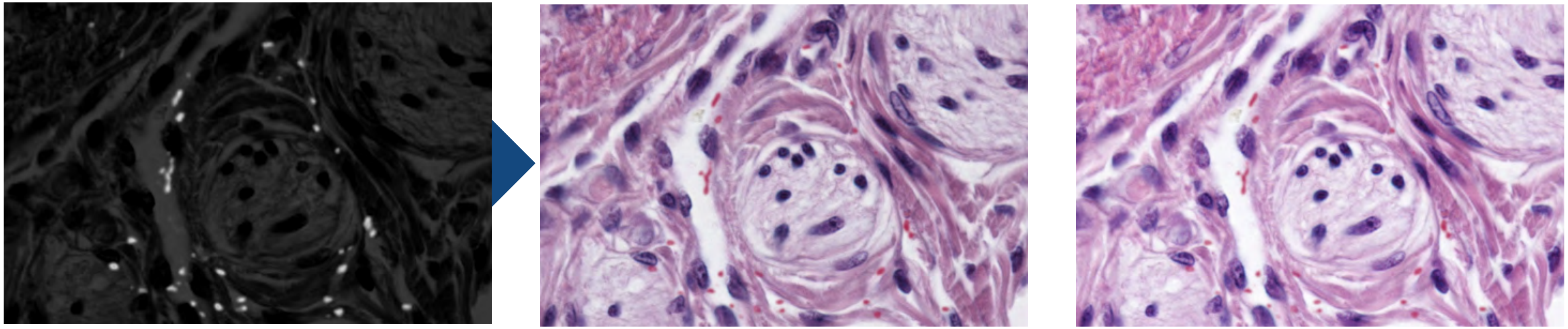
[1] *Wang et al.: Deep Learning for Identifying Metastatic Breast Cancer, 2016*

2.2 Example: Virtual Staining in Microscopy

In order to detect and assess pathological changes in human tissue, very fine, so-called histological sections of the tissue are chemically stained and examined under a light microscope. Depending on the dye used, different structures appear in the tissue. This process of dyeing is very time-consuming and cost-intensive. For example, if a tissue sample has to be stained during a surgical procedure in order to quickly obtain detailed information about a pathological change, this time expenditure is cumbersome and sometimes critical.

In this case, an artificial neural network, which "virtually colors" microscope images, can provide valuable support. This network is trained to use unstained images from a fluorescence microscope as input and convert them into stained light microscopic images. The advantage of this method: an automated solution that saves time and reduces costs.

Figure 2: *Illustration of the principle of virtual coloring.*



Unstained auto-fluorescence image (input) *Virtually stained image (output)* *Histologically stained image (reference)*

3. HOW CAN ANN-BASED PRODUCTS BE MADE COMMERCIALLY VIABLE?

Put simply, the main task of an artificial neural network is to imitate the visual cortex of the human brain – i.e. the visual perception of the human being – and to do so much faster and more precisely. For an ANN to work reliably, it must be trained. This training is usually based on and supported by thousands of marked images – the so-called "supervised learning". For the network architecture open-source software libraries such as TensorFlow, Caffee, Theano or Thorch are used in most cases.

The use of an ANN therefore goes through two phases: phase 1 comprises the described training, and phase 2 the execution. In this White Paper, we assume an already trained ANN which is now to be implemented in a series product.

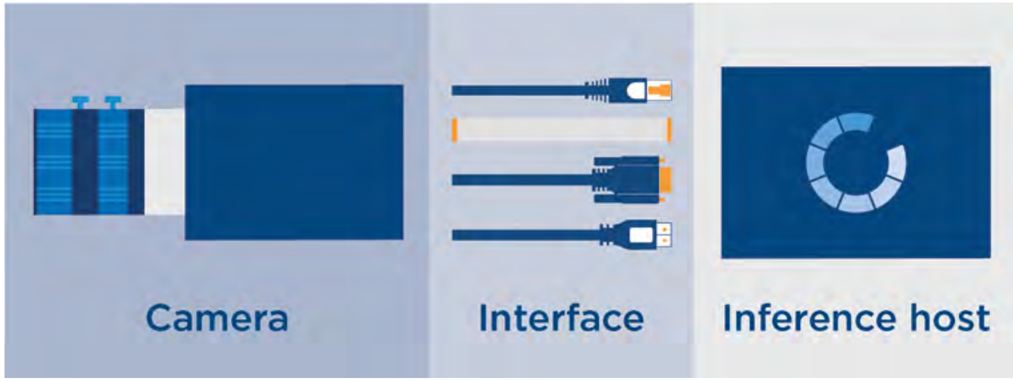


Figure 3: *Schematic representation of a vision system for deep learning: a camera for image acquisition, an interface for data transmission and an inference host for the ANN inference.*

The basic prerequisite for bringing an AI-based system to market is a vision system on which the typical image processing pipeline can run. This comprises the following four steps:

1. Image acquisition (input)
The camera is set up and programmed with the appropriate parameters such as exposure time or Region of Interest (ROI). The sensor takes an image.

2. Image pre-processing
The sensor image is pre-processed using specific algorithms such as debayering, color correction, sharpness optimization, etc. The image is being prepared for the execution – the inference – of the ANN.

3. ANN Inference
The artificial neural network now uses the pre-processed image as input and classifies it according to the specifications for which it was trained.

4. Presentation of results (output)
The output of the ANN inference must now be presented to the operator in a clear and visually prepared form.

While the training of an ANN with the help of the above mentioned software libraries is already very easy, the big challenge is to provide a hardware architecture for this trained ANN which on the one hand meets the requirements of the market, but on the other hand is also convincing from a cost point of view. The hardware must meet the requirements for the described scanning and processing steps. This requires three components in a vision system:

- a camera,
- an interface and
- a processing unit on which the inference is processed (“inference host”).

Figure 4: *Image processing pipeline for deep learning vision systems.*

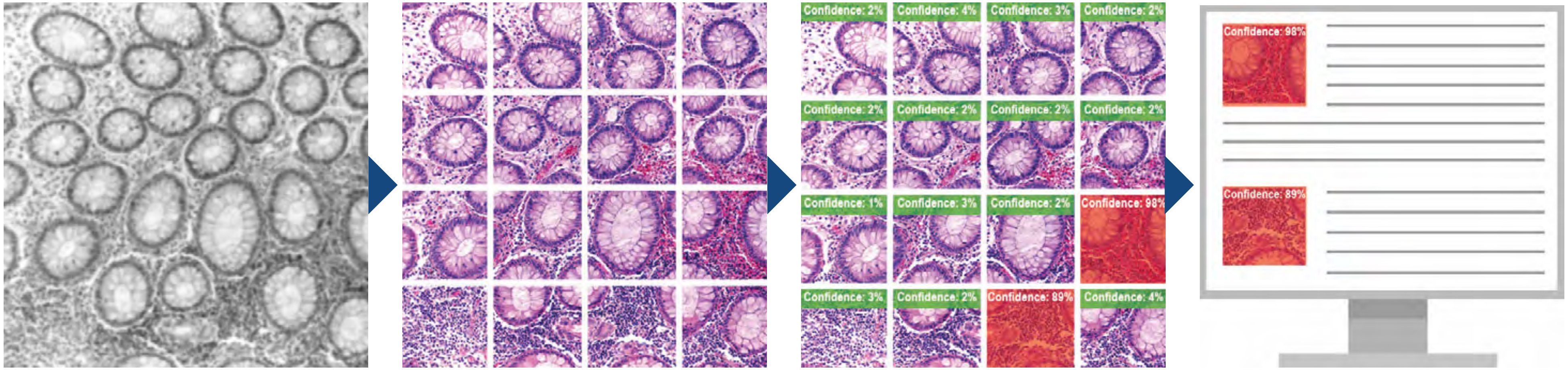
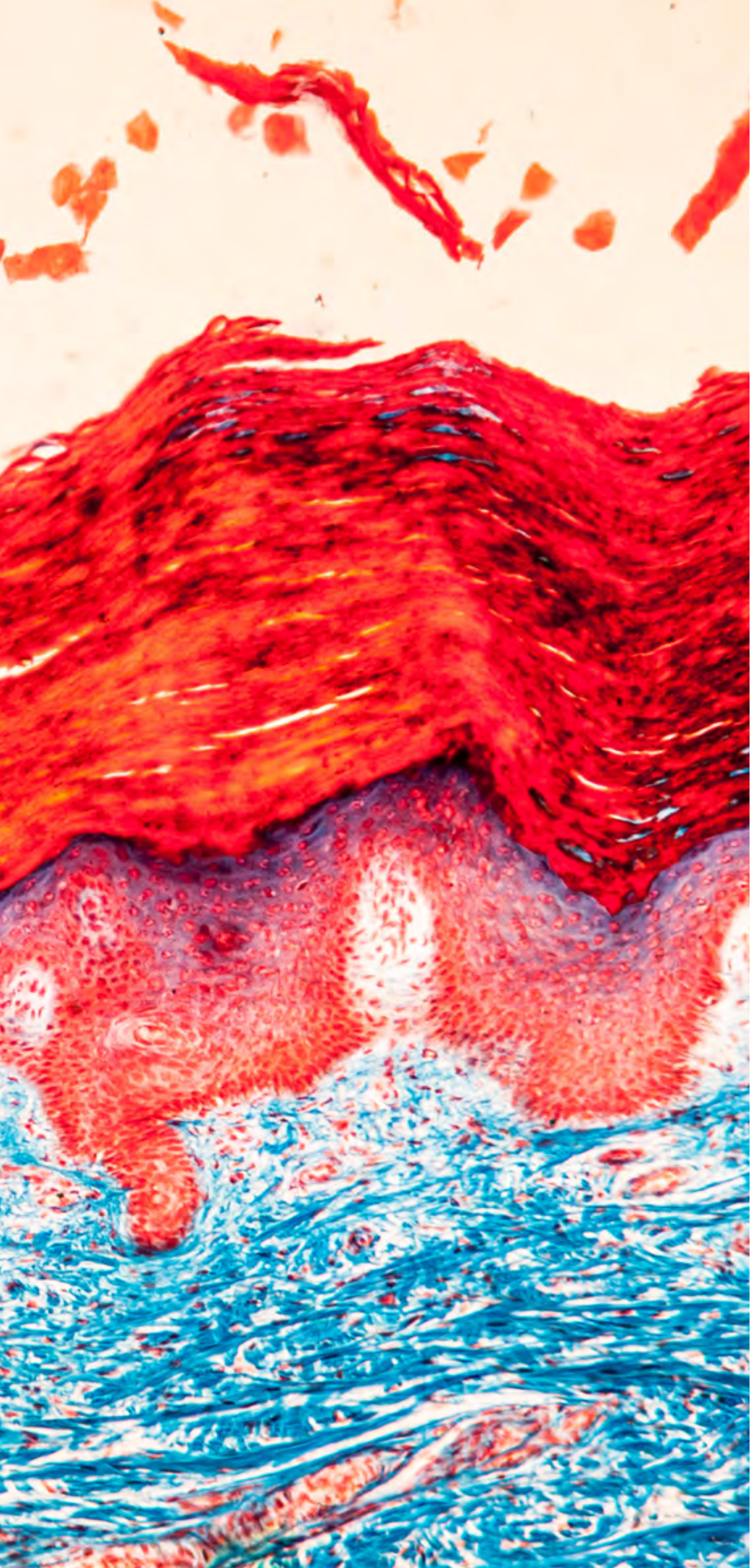


Image acquisition *Image pre-processing* *ANN inference* *Presentation of results*



4. THE THREE TYPES OF VISION SYSTEMS FOR DEEP LEARNING

Depending on the application and requirements, three different vision system architectures can be outlined for ANNs today, which differ in terms of performance, engineering effort and Total Cost of Ownership (TCO), in some cases significantly:

- embedded systems
- PC-based systems
- FPGA frame grabber-based systems

4.1 Embedded Vision System

The fundamental characteristic of an embedded system is its comparatively low TCO, which is made up of the following components:

Camera	Board level model: inexpensive and compact
Sensor	CMOS sensor (e.g. Sony or ON Semiconductor)
Data interface	MIPI CSI-2 (with advantages compared to USB3 and GigE)
Processor unit	ARM-based as Single Board Computer (SBC), System-on-Module (SoM) or System-on-Chip (SoC)

Important for deep learning applications is a hardware on which an ANN can be used quickly and efficiently. Due to their flexibility, long-term availability and good documentation, for example, these are currently the I.MX8 boards from NXP.

The **conclusion on embedded vision systems** for deep learning:

+	-
Low unit costs	Complex and resource-intensive system design and integration Lower performance, since computing capacity is often limited to relatively simple image processing tasks or simple network architectures

4.2 PC-based System

The basic feature of a PC-based system is its low integration and engineering effort thanks to commercially available hardware components.

These are illustrated by the example of a medical application:

Camera	ISO 13485:2016 compliant camera (e.g. Basler MED ace)
Sensor	Powerful CMOS sensor (e.g. Sony IMX)
Data interface	USB 3.0
Processor unit	Standard x86 with graphics processor (GPU) or CPU

The image data generated by the sensor already undergoes pre-processing within the camera – on the FPGA (Field-programmable Gate Array) – using special features such as 5×5 debayering, color or sharpness correction. This pre-processed image is transmitted to the inference host via the data interface. Since a large part of the computing power for image pre-processing is already done within the camera, more computing capacity is available for the ANN.

Common software libraries for deep learning reference such as OpenCV or Tensorflow and a good software infrastructure for camera control allow smooth and uncomplicated plug-and-play between the individual components.

The **conclusion on PC-based systems** for Deep Learning:

+	-
Easy integration and engineering	Comparatively high TCO
Short time-to-market	Data rate depending on interface (e.g. USB 3.0 max. 350 MB/s)

4.3 FPGA Frame Grabber-based System

The fundamental characteristic of an FPGA frame grabber-based system is its performance in terms of robustness, availability and performance.

Camera	Area or line scan camera (e.g. Basler boost)
Sensor	Powerful CMOS sensor (e.g. Sony IMX)
Data interface	CoaXPress (CXP), CameraLink
Processor unit	FPGA-based frame grabber

Thanks to the new CoaXPress 2.0 standard, the CXP-12 interface can transfer and process image data with a bandwidth of up to 12.5 Gbps – almost three times faster than USB 3.0.

Unlike USB 3.0 or GigE, however, CXP is not a conventional interface that can be connected to a PC. A frame grabber is required to capture and pre-process an image. This contains a large FPGA on which a pre-trained ANN can be placed. Visual Applets facilitate the use of frame grabber-based deep learning considerably by making it very easy to implement trained ANNs in the software on the FPGA.

The **conclusion about frame grabber-based systems** for deep learning:

+	-
High performance up to 12.5 Gbps thanks to CXP-12 and image processing (including inference) in FPGA	Higher TCO than with embedded systems
Long-term availability with regard to required certification for use in the Medical & Life Sciences market	
Lower TCO than with high-end GPUs	

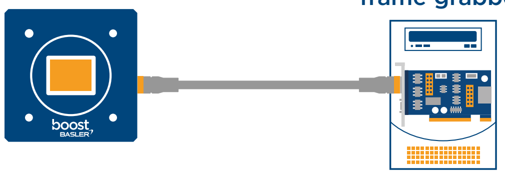
SETUP	EASE OF INTEGRATION	PERFORMANCE	TCO	VISION ARCHITECTURE
board-level-camera · MIPI CSI-2 · processor board 	+	++	$	Embedded
boxed camera · USB3 · PC 	+++	++	$$$	PC-based
boxed camera · CXP-12 · PC + frame grabber 	+++	+++	$$	Frame grabber-based

Figure 5: *Three different types of architectures when designing a vision system for deep learning: embedded, PC-based and FPGA frame grabber-based. While embedded architecture offers the most attractive TCO, the frame grabber-based system offers the best performance.*

BASLER

the power of sight

5. SUMMARY

If one wants to efficiently use the many new application areas of deep learning in Medical & Life Sciences, the following criteria play an important role in the design of a vision system:

- Ease of integration
- Robustness
- Price
- Performance
- Availability

If these factors are taken into account right from the start when selecting a deep learning vision system, this makes it much easier to translate a proof of concept into a competitive series product with ANNs.

AUTHOR

Peter Behringer

Head of Global Partner Management & Head of Medical Business

peter.behringer@baslerweb.com

Basler AG
Germany, Headquarters
Tel. +49 4102 463 500
Fax +49 4102 463 599
sales.europe@baslerweb.com
baslerweb.com

Basler, Inc.
USA
Tel. +1 610 280 0171
Fax +1 610 280 7608
sales.usa@baslerweb.com

Basler Asia Pte Ltd.
Singapore
Tel. +65 6367 1355
Fax +65 6367 1255
sales.asia@baslerweb.com