Object Recognition

Chapter 1: Introduction

Prof. Dr. Johannes Maucher

HdM CSM

Version 1.5
April, 20th 2020

Document History

Version Nr.	Date	Changes
1.0	29.01.2013	Initial Version
1.1	25.03.2014	Adaptations for SS 14
1.2	20.03.2017	Adaptations for SS 17
1.3	19.03.2018	Adaptations for SS 18
1.4	19.03.2019	Adaptations for SS 19
1.5	20.04.2020	Adaptations for SS 20

Chapter 1: Introduction

Organization

Date	Wednesday, 08.15h-11.30h
Form	Online Zoom Session
Assignments	Submit solutions. Minimum number of credits must be achieved
Language	German
Exam	written exam
Credits	4 SWS / 5 ECTS
Room	137
Homepage	http://maucher.pages.mi.hdm-stuttgart.de/ai/page/or/
Gitlab Repo	: https://gitlab.mi.hdm-stuttgart.de/maucher/or
Checker Quests	Link to Checker Quests

Contents of this lecture

Introduction	Course Structure, Motivation, Definitions, Applications
Image Processing Basics	Filtering, Noise Surpression, Pyramids and Scale, Template Matching, Edge Detection
Global Features	Pixel Intensities, Color Histograms, Multidimensional Receptive Field Histograms, Probabilistic Recognition
Subspace Features	PCA, LDA, Face Recognition with Eigenfaces and Fisherfaces
Local Features	SIFT, SURF, HOG
Specific Object Recognition	with local features, Efficient Similarity Search, Indexing Features with Visual Vocabularies, Geometric Verification
Generic Object Recognition I	Conventional: Bag of Words, Image Classificiation
Object Detection:	Conventional: Vioal-Jones Face Detection; Pedestrian Detection
Generic Object Recognition II:	Deep Neural Networks (CNNs)
Object Localisation:	Deep Neural Networks: Overfeat, R-CNN, Yolo, ...
Semantic Segmentation	with Deep Neural Networks
Pose Detection	with Deep Neural Networks
Tracking	in Video

Computer Vision: Definition and Goal

Goal of Computer Vision

Computer Vision seeks to enable machines to see and understand data from images and videos

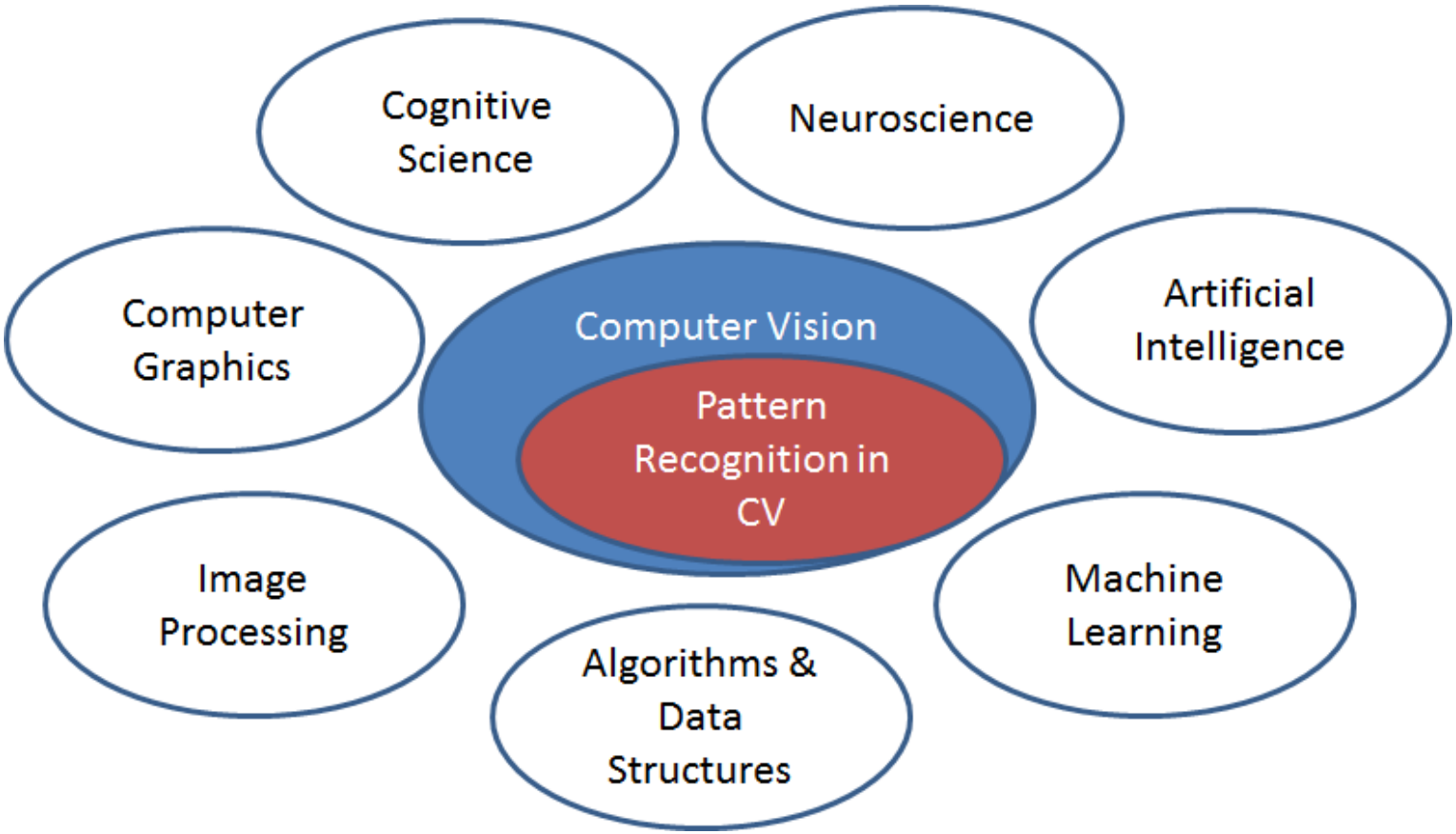
Definition

A branch of artificial intelligence and image processing concerned with computer processing of images from the real world. Computer vision typically requires a combination of low level image processing to enhance the image quality (e.g. remove noise, increase contrast) and higher level pattern recognition and image understanding to recognise features present in the image.

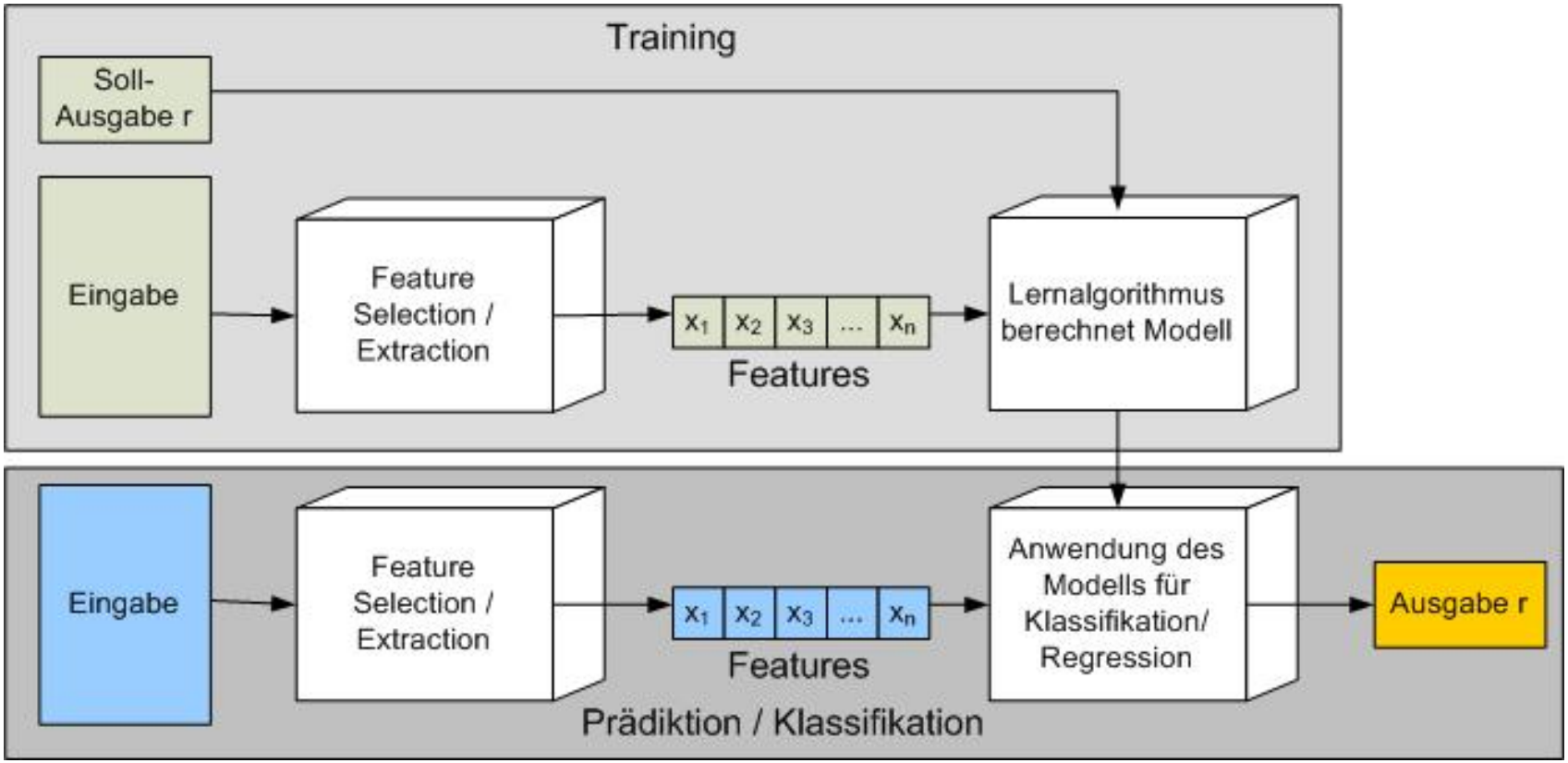
Source: The Free Online Dictionary of Computing (2003-OCT-10)

- Object recognition is the central task in computer vision
- Other computer vision tasks are e.g. image restoration, 3D reconstruction, motion analysis, image rendering etc.

Related Science



Object Recognition General Scheme



High-level Categorization

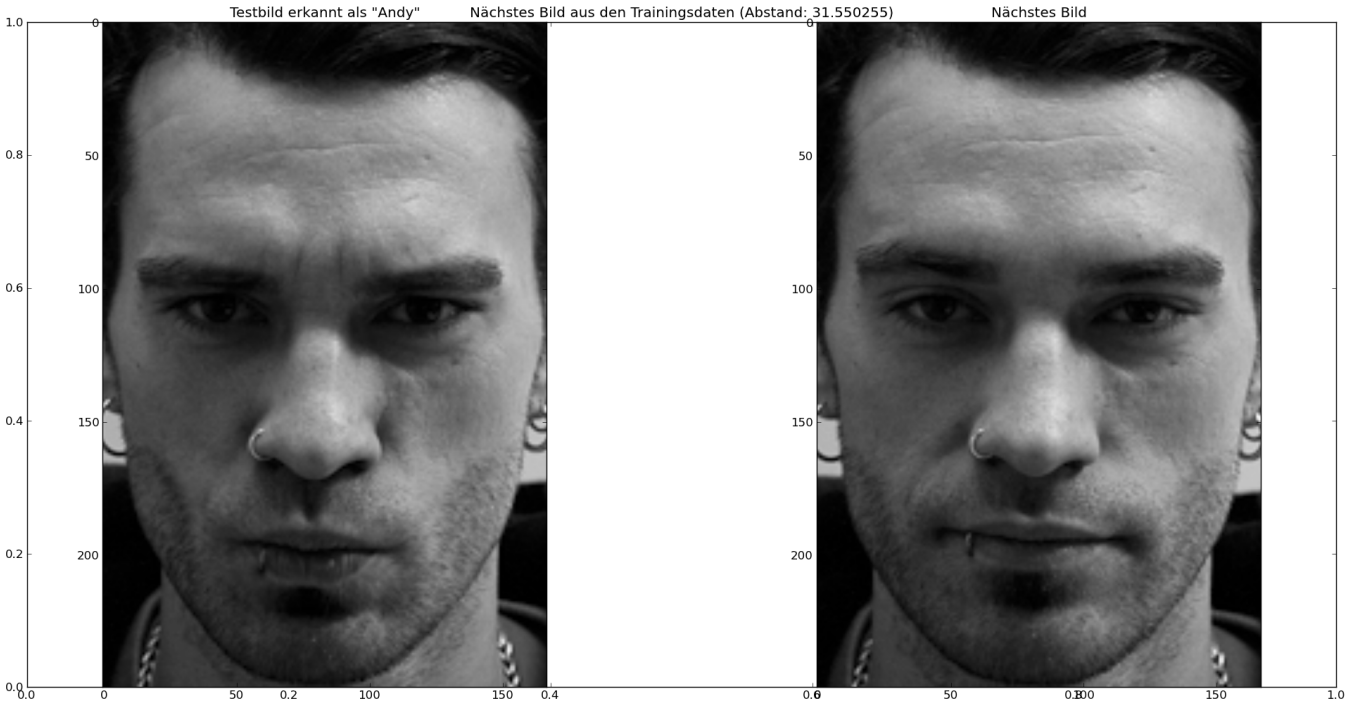
	Recognition of specific objects = Identification	**Recognition of generic object categories**	
		Single Category = Detection	Multiple Categories
Global Features			
Subspace Features			
Local Features			

Categories of Object Recognition

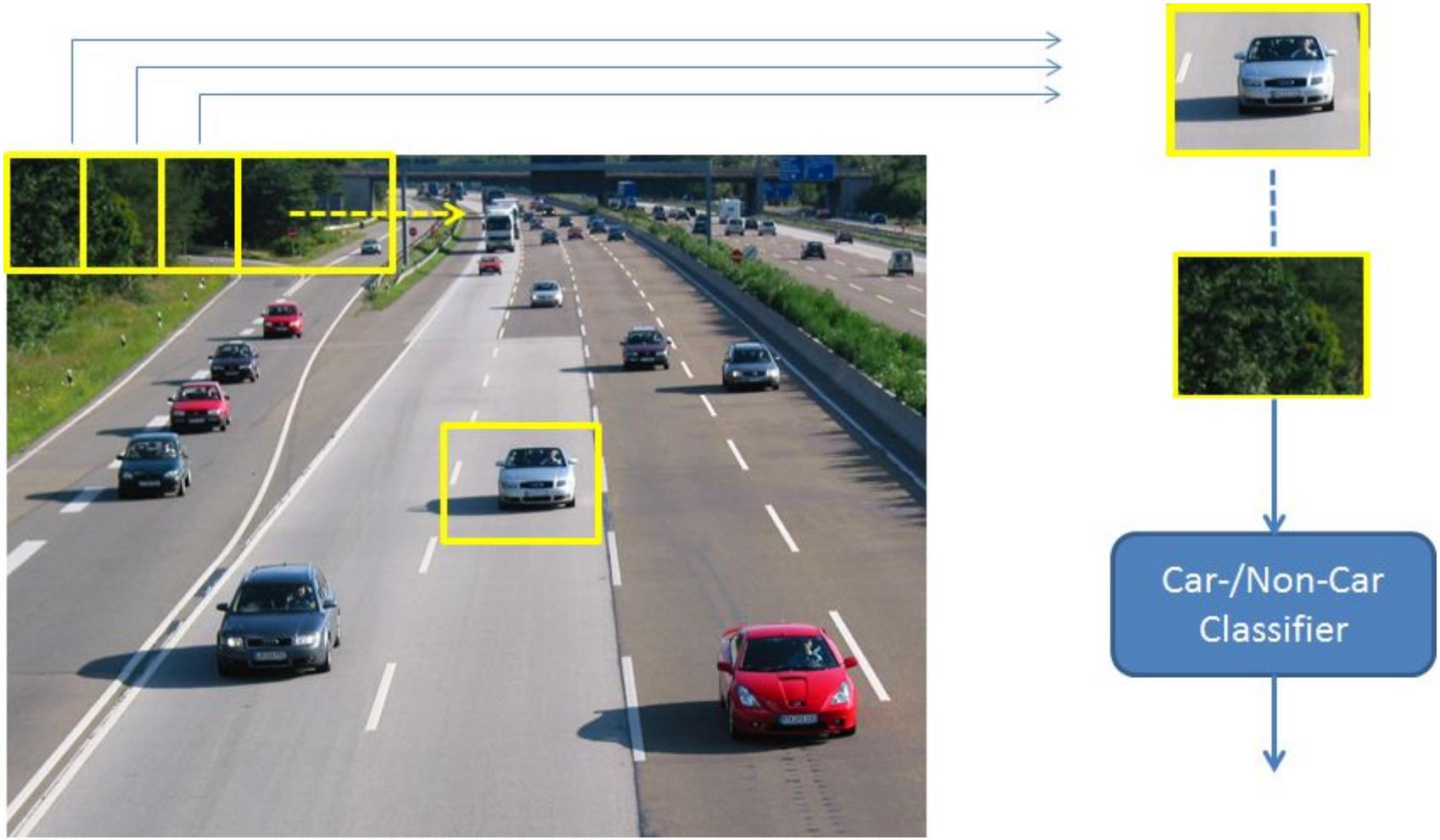
- Recognition of specific objects also called Identification
  - Identify a particular instance, e.g. the face of a particular person or a particular building.
- Recognition of a generic object category also called Object Categorization:
  - Objects of what categories are in the image, e.g. a face, a bottle, an orange, a car.
  - For each of the categories that shall be recognized, the system must have a model that is aware of the in-class variations within each category.
  - Object Detection: If only a single object category is of interest and the task is to find instances of this class in an image.
  - Localisaton: Which objects are in the image and where are they (bounding boxes)?
  - Semantic Segmentation: For each pixel in the image: Determine to which object it belongs.

Specific Object Recognition: Face Recognition - Training Data

Specific Object Recognition: Face Recognition - Test



Recognition of single category: Object Detection



Recognition of single category: Face Detection

Generic Object Recognition Training: Learn model of each category

What you see (input)



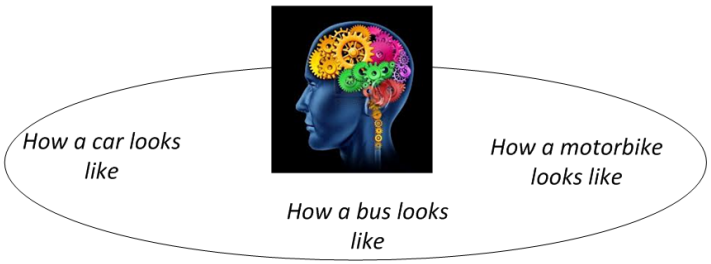




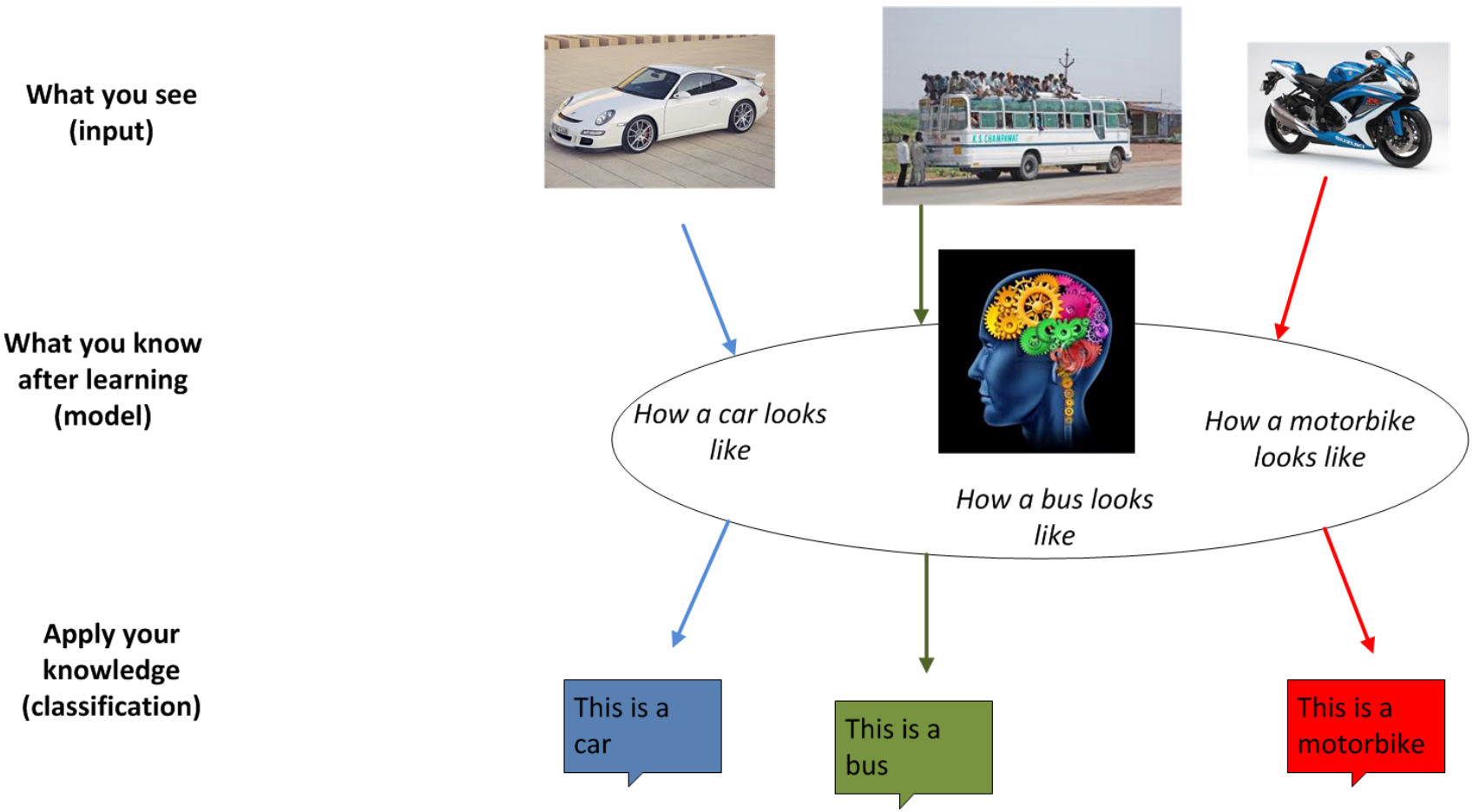
What your supervisor teaches you (label)

This is a bus | This is a bus | This is a motorbike | This is a car | This is a bus | This is a car | This is a motorbike

What you know after learning (model)

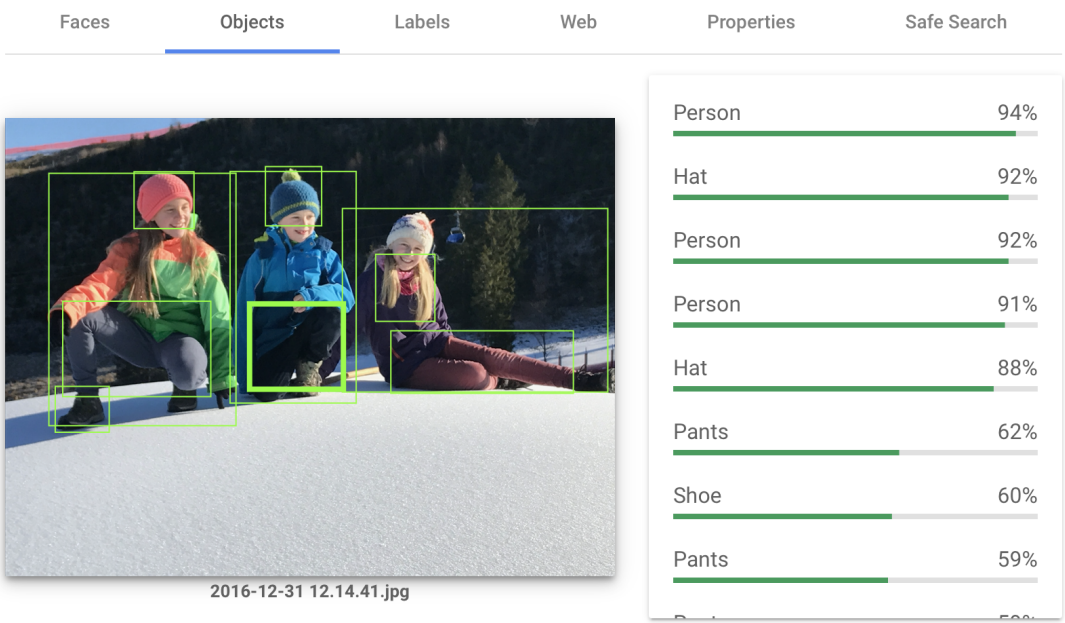


Generic Object Recognition Inference: Determine category of given object



Object Classification and Identification

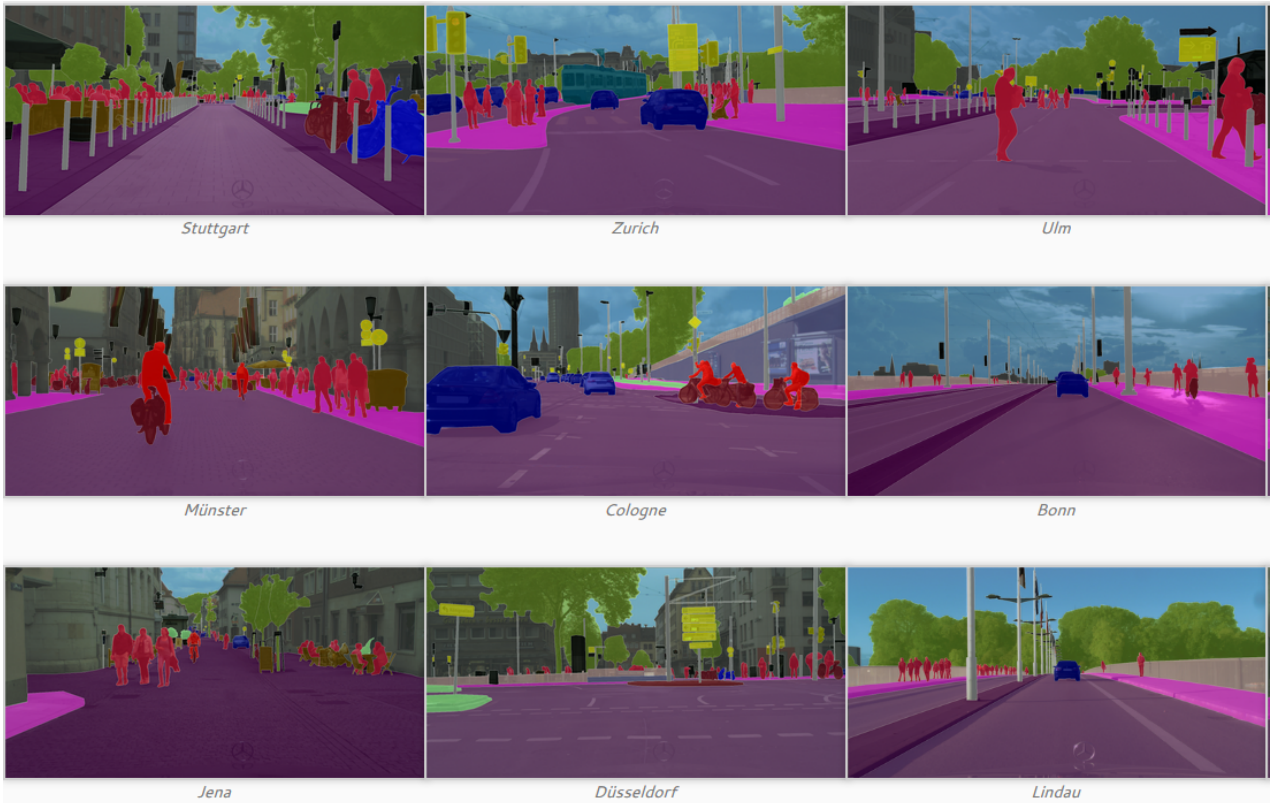
Google Vision API



https://cloud.google.com/vision

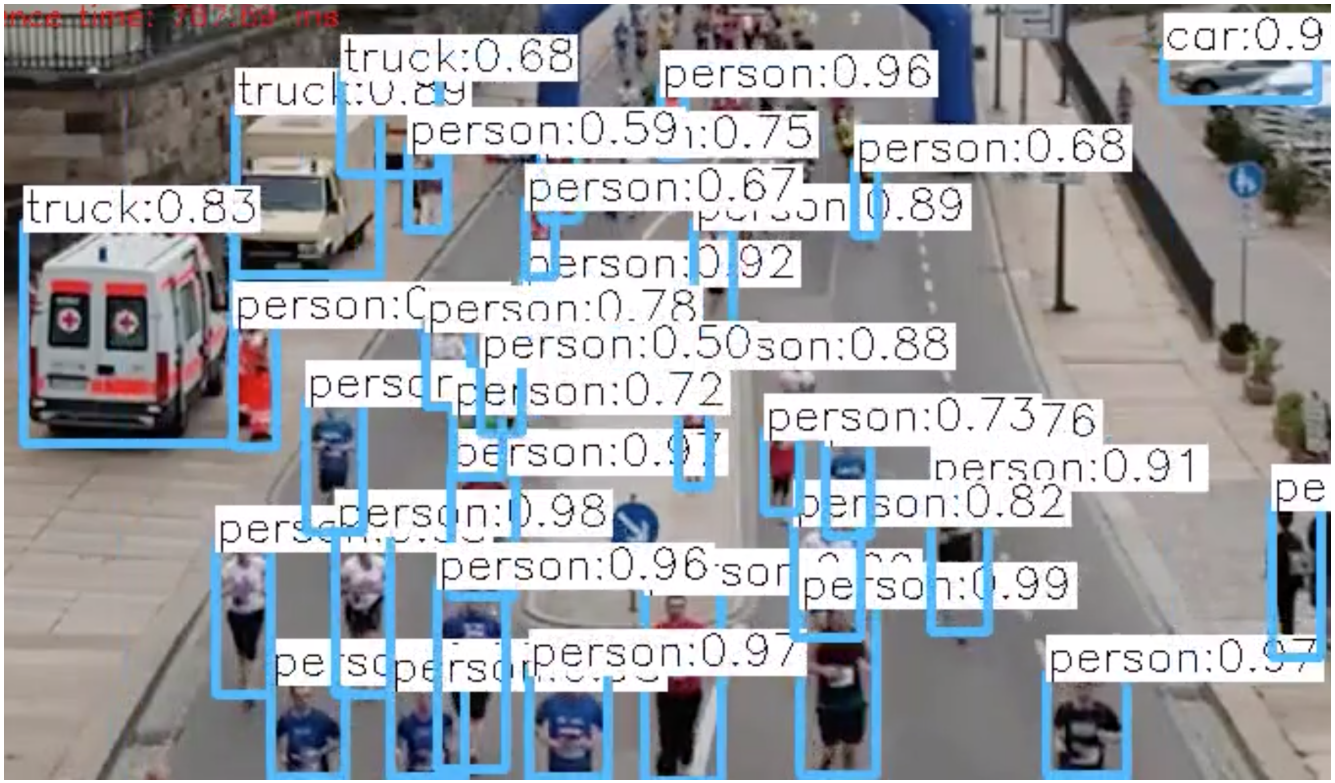
Generic Object Recognition

Semantic Segmentation



- Segmentation partitions an image into subregions
- Semantic Segmentation identifies the object-type of the subregion

Object Localisation and Tracking



Multiperson Pose Estimation

Feature Extraction

- Object Recognition performance depends on
  - the choice of a suitable recognition algorithm, e.g. Neural Network, Support Vector Machines, Adaboost, Naive Bayes, etc.
  - a suitable method to model the features of an image
- In computer vision recognition performance depends much more on the feature extraction than on the algorithm choice $\Rightarrow$ *That's why we will deal more with image processing and vision than with machine learning.*
- In some efficient recognition methods feature extraction and the recognition algorithm are closely interwoven, e.g. Viola-Jones Face Detector.

Categorization of Feature Descriptors

- Global: A single descriptor is calculated for the entire image, e.g.
  - a vector containing all pixel values of all channels in the image
  - a color histogram which measures the frequency of colors in the entire image
- Local: Local descriptors are calculated at all keypoints found in an image. Keypoints are e.g. corners. A local descriptor encodes information in a small region around a keypoint.

The challenge of feature modelling is to determine which type of information, explicitly or implicitly encoded in an image, is most informative for reliable and robust object recognition?

Example for global feature descriptor

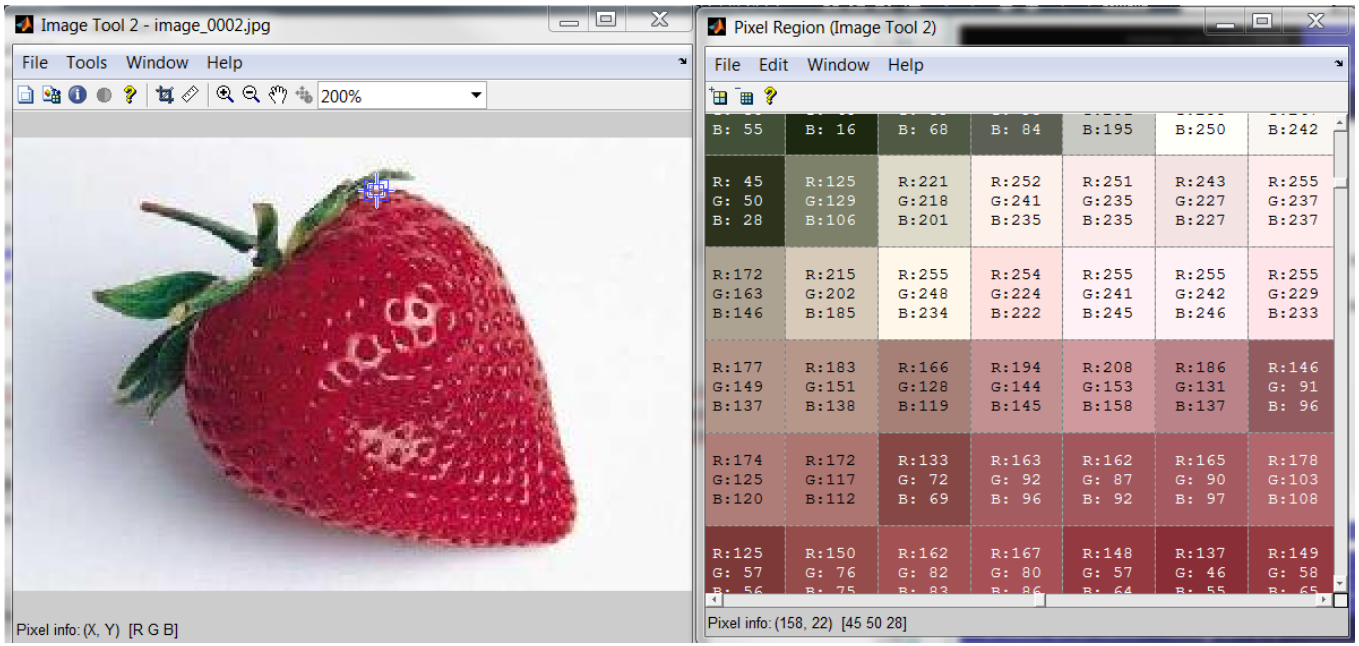


Abbildung: Global Descriptor: Vector of all pixel values in an image

Example for local feature descriptors

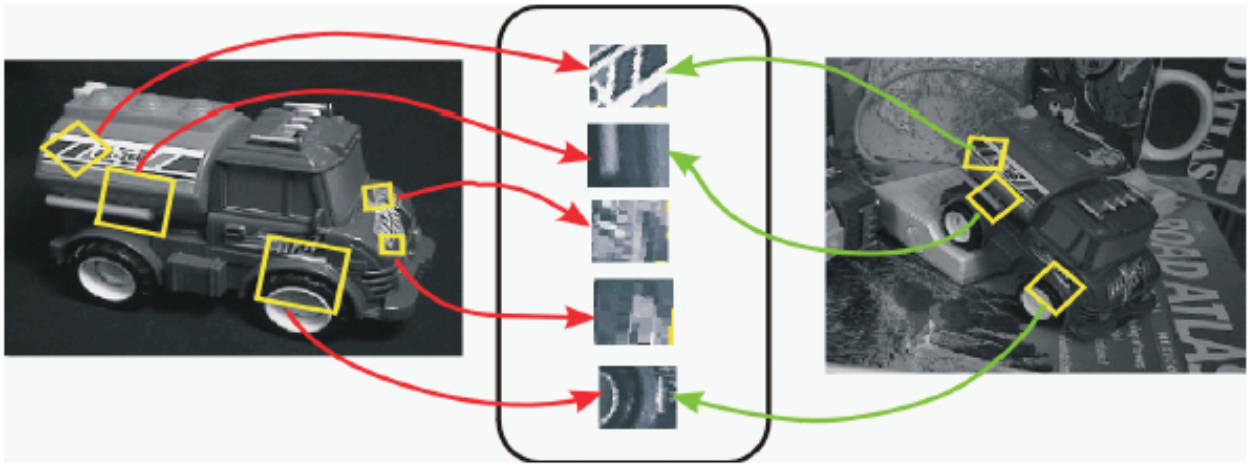
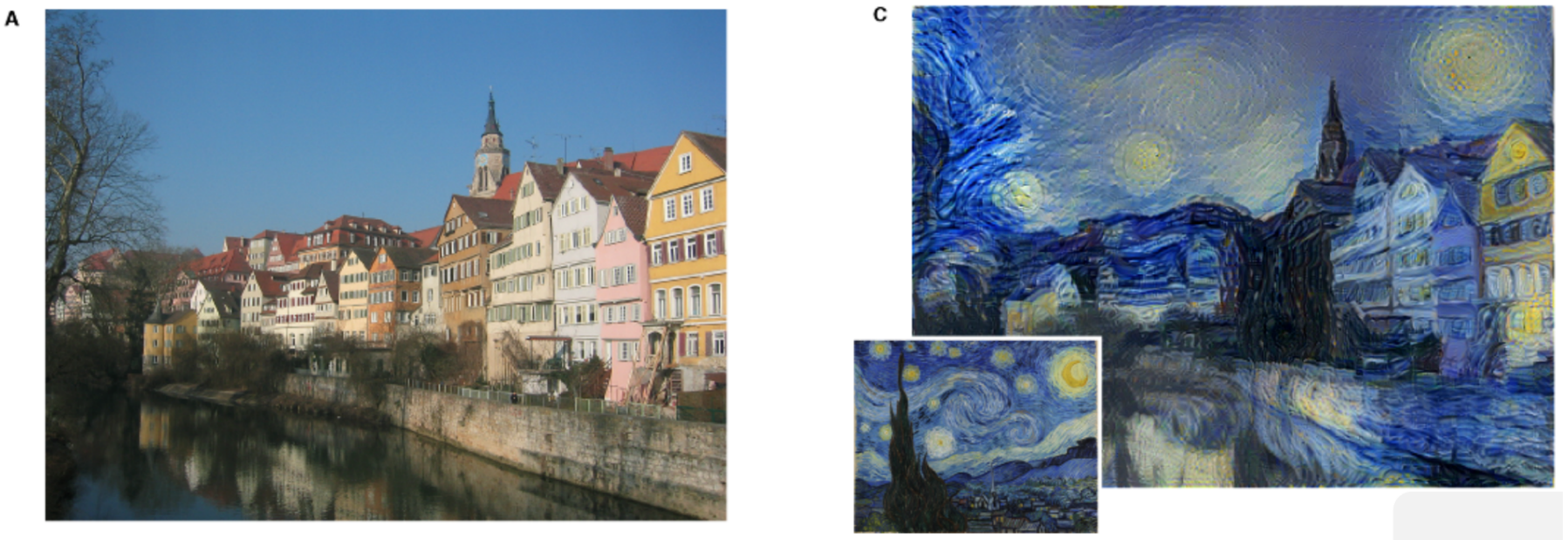


Abbildung: Local Features at keypoints and matching of local features

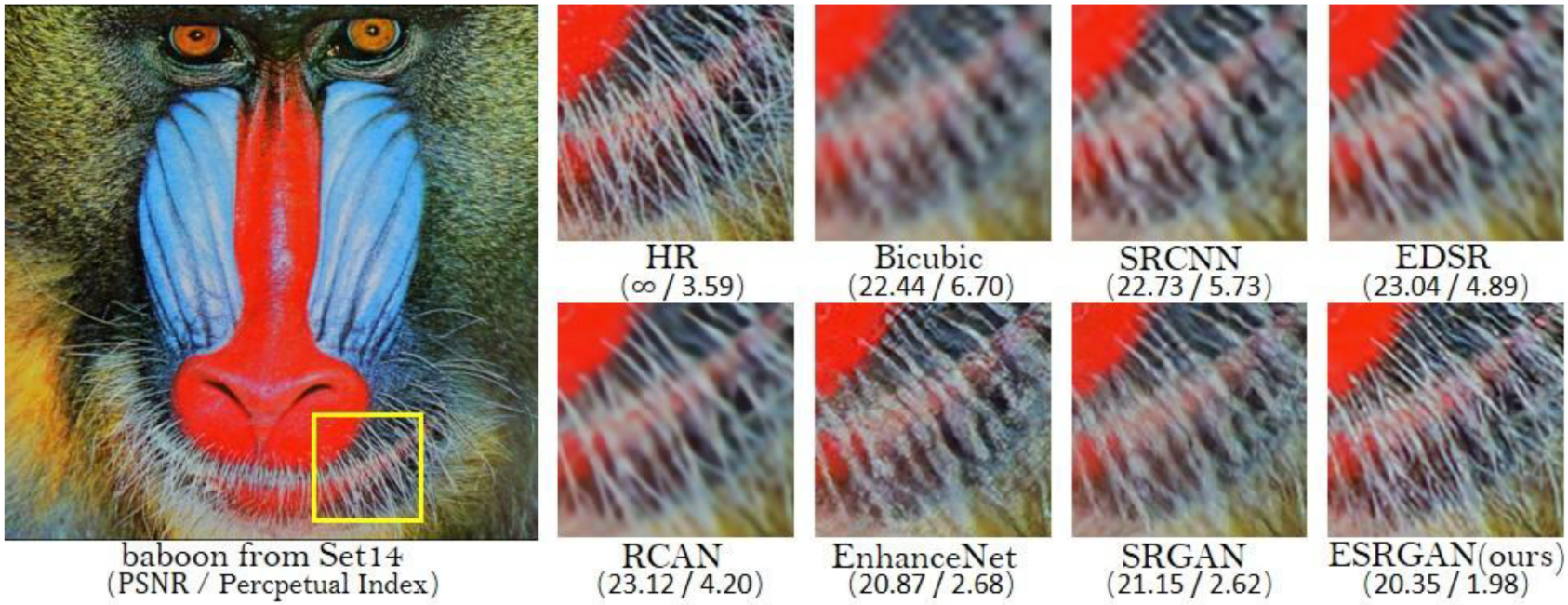
Deep Neural Networks for Object Recognition

- Deep Learning is currently the most researched and discussed topic in Computer Science
- Deep Learning has initially been researched and applied for Object Recognition tasks e.g. in the context of
  - Social Networks
  - Autonomous driving / Driver assistance
- In Deep Neural Networks a feature extractor and a classifier a learned in a single network end-to-end.
- Progress in Deep Learning architectures is mainly driven by the ILSVRC Challenge on object recognition.

Further Topics: Style Transfer

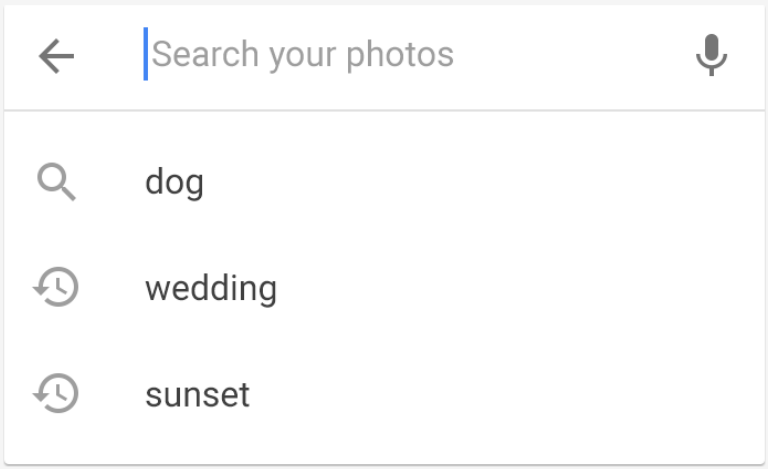


Further Topics: Super Resolution



Quelle: Quelle: Xintao Wang1 , Ke Yu1 , Shixiang Wu2 , Jinjin Gu3 , Yihao Liu4 , Chao Dong2 , Chen Change Loy5 , Yu Qiao2 , Xiaoou Tang1: ESRGAN: Enhanced Super-Resolution Generative Adversarial Networks(2018) Github: https://github.com/xinntao/ESRGAN 2

Image Retrieval: Google Photos and Google image search



Google Image Search: https://images.google.com

Adobe Sensei

Adobe Sensei

- Semantic Segmentation / Background Subtraction with *Select Subject* in PS cc
- Image Postprocessing in Ps CC, e.g. DeepFill
- Automatic Tagging for more efficient search in Adobe Stock
- https://www.youtube.com/watch?v=xT9NxX$_n w_k$

Sony Smile Shutter

Smile Shutter

How it works

Switch on Smile Shutter and let your Cyber-shot take the photo for you! When Smile Shutter is activated your Cyber-shot will automatically take pictures when people smile. It's great for capturing spontaneous and natural smiles and ensures you snap people looking their best. What's more, it's incredibly easy and fun to use!

Using Smile Shutter is as simple as 1, 2, 3:

1. Set

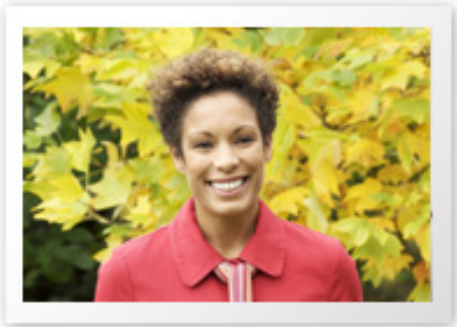
Set your camera to Smile Shutter mode.

2. Point

Point the camera at your subject and press the shutter button.

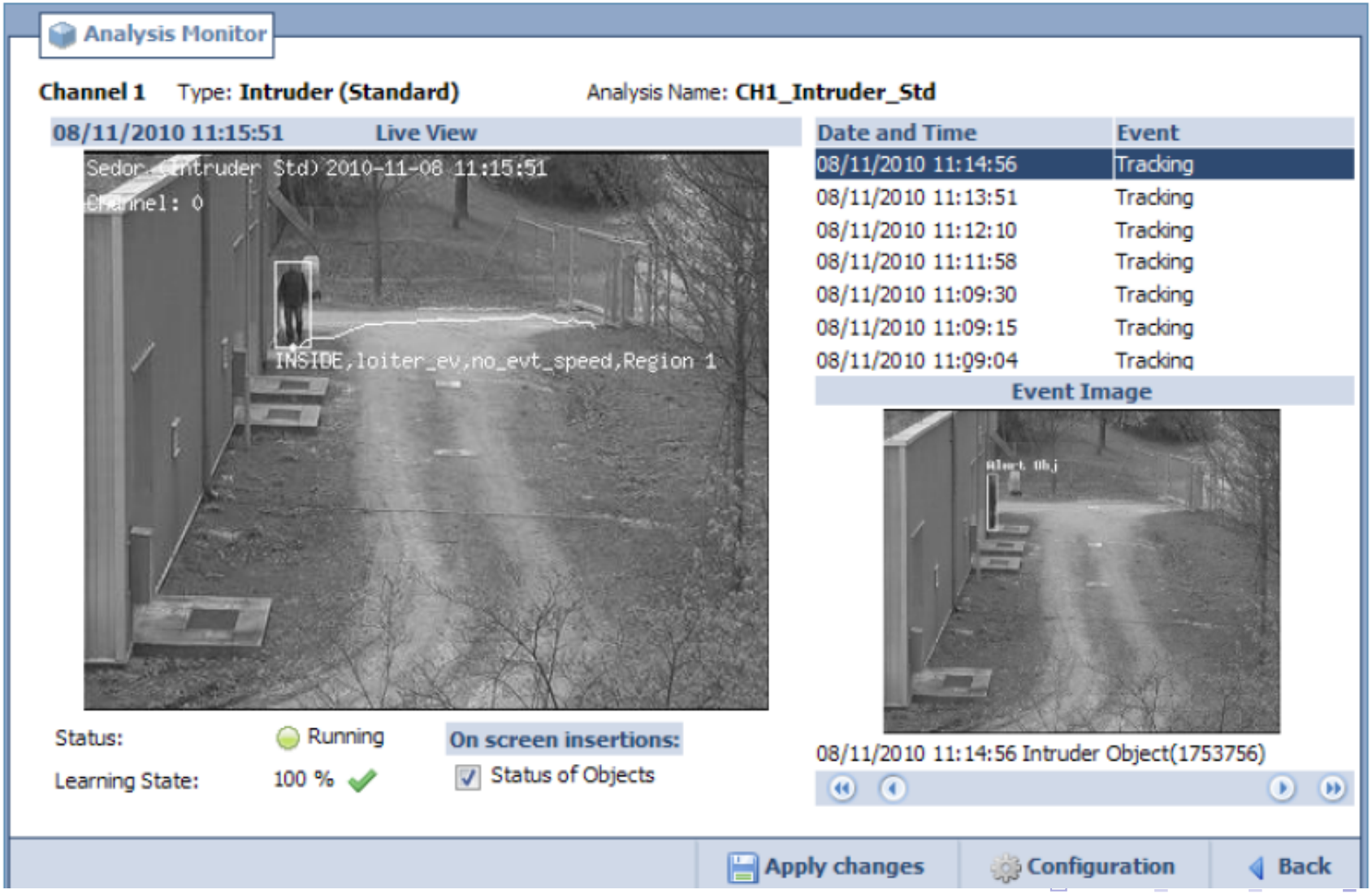
3. Say 'cheese'

As soon as they smile your camera automatically takes the shot for you!



Autostitch IPhone and IPad App

Intruder Detection (Dallmeier Sedor)



Fahrerassistenzsysteme (Mercedes S-Klasse)

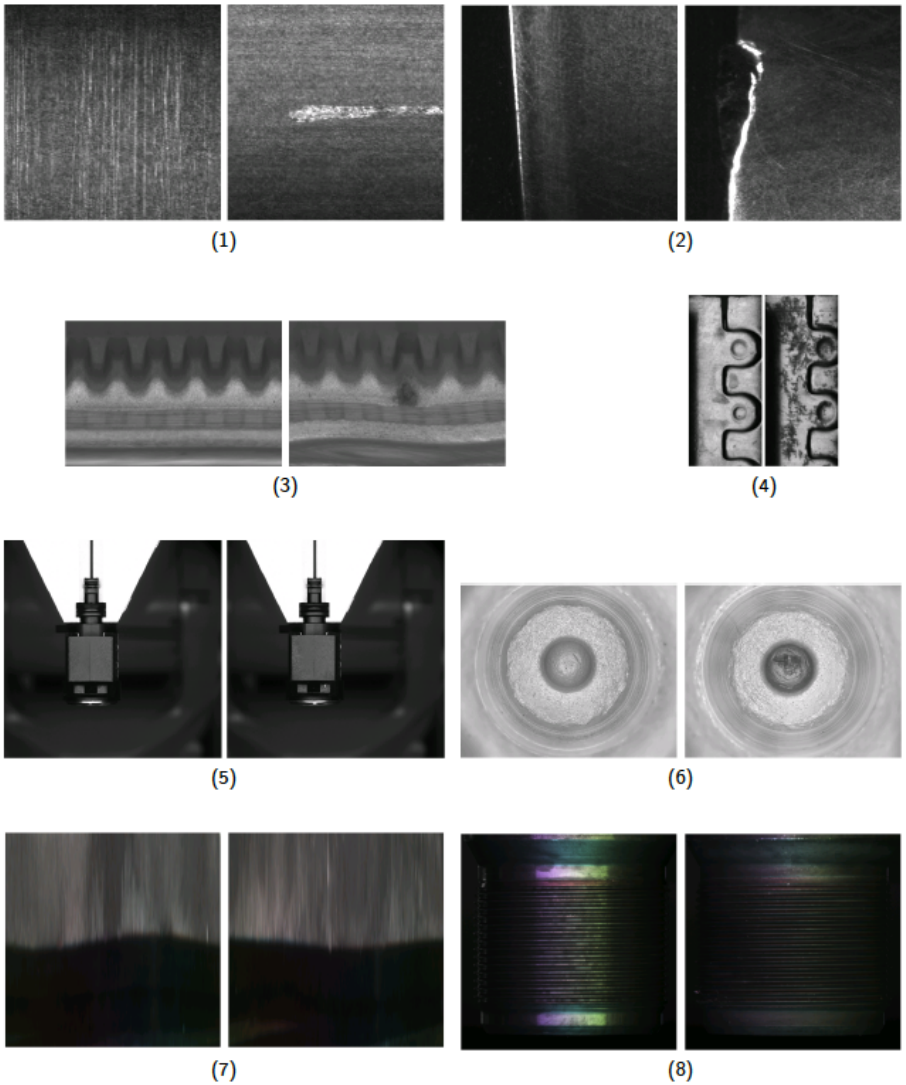
◢ Die Funktionen BAS PLUS und PRE-SAFE® Bremse wurden um eine Fußgängererkennung erweitert. Außerdem wurde die autonome Bremsung auf Vorausfahrzeuge deutlich verbessert. Durch Fusion von Daten aus Stereokamera und Radarsensoren können Fußgänger im Bereich vor dem Fahrzeug erkannt werden. Bei erkannter Gefahr warnt das System den Fahrer optisch und akustisch. Reagiert er und bremst, verstärkt BAS PLUS bei Bedarf die Bremswirkung situationsgerecht bis hin zur Vollbremsung. Reagiert er nicht, bremst die PRE-SAFE® Bremse das Fahrzeug autonom ab. Die PRE-SAFE® Bremse mit Fußgängererkennung ist bis 72 km/h aktiv und kann Fußgängerunfälle aus einer Geschwindigkeit von mehr als 50 km/h vermeiden. Durch autonomes Bremsen kann das System ebenfalls einen Auffahrunfall auf stehende Fahrzeuge aus über 50 km/h vermeiden. Damit erfüllt es bereits jetzt die ab 2014 gültigen strengen EuroNCAP-Regelungen zur autonomen Notbremsung in der Stadt mit maximaler Punktzahl.

Kinect

Further Object Recognition Applications

- Categorization, Semantic Search, Automatic Annotation and Captioning in Image Databases (Apple Fotos, Google Fotos, ...)
- Automatic Inspection
- Optical Character Recognition
- Automatic Survey Evaluation
- Person Identification (Iris, Face, ...)
- Object Detection in Medical Images

Optical Inspection



Challenges


Scale

Pose

Intraclass Appearance

Occlusion

Clutter

Illumination

Viewpoint

Challenges

Complexity:

- *„The fact that about half of the cerebral cortex in primates is devoted to processing visual information gives some indication of the computational load one can expect to invest for this complex task.“* [1]
- Paritcularly in the case of
  - Image Retrieval with large image databases,
  - Image Recognition, were an image must be categorized in one of thousands of categories

  highly efficient algorithms are required.
- In the case of supervised learning labeled training data is required. In general labeling is cost-intensive.

References I

[1] D.J. Felleman and D.C. van Essen. Distributed hierarchical processing in the primate cerebral cortex. *Cereb Cortex*, pages 1–47, 1991.