

03-IMAP-D3BV	Deep Learning und 3D Bildverarbeitung
	<i>Deep Learning and 3D Computer Vision</i>

Lehrform (*teaching format*) / **SWS** (*hours per week*): 2VL + 2UE

Kreditpunkte (*credit points*): 6

Turnus (*frequency*): i.d.R. jedes SoSe

Inhaltliche Voraussetzungen (*content-related prior knowledge/skills*): KEINE

Sprache (*language*): Deutsch

Lehrende (*teaching staff*): AG Multi-Sensor Interactive Systems (Prof. Dr. Udo Frese)

Studiengang (<i>degree program</i>)	Module	Semester
Informatik (Master)	IMAP, IMK-VMC, IMAP-DMI, IMVP-AI	ab 1.Sem.
AI and Intelligent Systems (Master)	AI-M-MLCS	ab 2.Sem.
Digital Media (Master)	Free Choice	ab 1.Sem.
Systems Engineering I/II (Master)	M07-IM-AuR-INF, M07-IM-PT-INF, M07-IM-Me-INF	ab 1./2.Sem.
ProMat (Master)	Informatikwerkzeuge	ab 1.Sem.
Informatik (Bachelor)	(nur <i>Freie Wahl</i>)	
Zertifikatsstudium DiMePäd	DM in Lernumgebungen	ab 1.Sem.

Lernergebnisse:

Die wichtigsten Methoden moderner Bildverarbeitung verstehen

- Bildverarbeitung mit Deep Learning (Convolutional Neural Networks)
- 3D Bildverarbeitung („von Pixeln zu Metern“)

Anwendungsprobleme mit diesen Methoden lösen können

- Passende Verarbeitungsketten für Anwendungsprobleme entwerfen
- Anwendungsprobleme als Deep Learning Aufgabe formulieren
- Trainingsdaten beschaffen, annotieren und aufbereiten
- Dreidimensionale geometrische Zusammenhänge in Bildern modellieren
- Deep Learning und 3D Bildverarbeitungssysteme systematisch entwickeln
- Implementierung mit TensorFlow / Keras

Learning Outcome:

Understand the most important methods of modern computer vision

- computer vision with deep learning (convolutional neural networks)
- 3D computer vision (from pixel to meter)

Solve application problems with these methods

- design computer vision pipelines for specific applications

- formulate application problems in a deep learning framework
 - record, annotate and prepare training data
 - model 3d geometric issues in images
 - systematically develop deep learning and 3d computer vision systems
 - implement with tensorflow / keras
-

Inhalte:

Bildverarbeitung mit Deep Learning.

All die folgenden Inhalte werden in ihrer formalen Definition, aber auch in einem intuitiven Verständnis für die Idee dahinter, die Bedeutung im Gesamtthema und die Interaktion mit anderen Inhalten vermittelt.

- Paradigmen „analytisch entwickelt“ vs. „maschinell gelernt“
- Die Architektur von künstlichen neuronalen Netzen die Bilder verarbeiten („Big Picture“ - Überblick noch ohne die Details, die später folgen)
- Formen der Ausgabe: Klassifizierung, Semantic Segmentation, Heatmap, Boundingboxes, Objekt-Id pro Pixel, applikationsabhängige Werte
- Schichten: Convolution, Aktivierung, Pooling, Unpooling, Fully Connected, Batch Normalization
- Losses: absoluter, quadratischer, relativer Fehler, Maximum-Likelihood, Crossentropy, Gewichtungen, Kombination mehrerer Losses
- Optimierung durch Gradientenabstieg, Sicht eines Netzes mit Loss als Graph von Tensoroperationen, Tensorformate, Backpropagation auf solch einem Graphen
- Rezeptives Feld als Architekturkenngroße
- Typische CNN-Backbonearchitekturen und ihre Nutzung im „pretrained“-Ansatz
- Decoder-Encoder Architektur für Bilder als Ausgabe, Bedeutung der Querverbindungen
- Objekterkennung: Ausgabeform für Boundingboxes, one-shot vs. two-shot Ansatz
- Vorgehen bei der Datenbeschaffung, Annotation und Aufbereitung, Probleme von Datensätzen
- Vorgehen bei der Entwicklung und Evaluation von Deep Learning Bildverarbeitungssystemen
- Mediale Anwendungen von Deep Learning, besonders zur Bildgenerierung
- Generative Adversarial Networks (die Grundidee)
- Verarbeitung von 3D-Voxel-Bildern und Tiefenbildern

3D – Bildverarbeitung

- Paradigma von „Pixeln zu Metern durch Gleichungslösen“
- Punktfeatures
- Kameragleichung
- Geometrische Rekonstruktion (welche 3D Eigenschaften lassen sich aus wie vielen 2D Punkten rekonstruieren)
- Quadratische Ausgleichsrechnung als generischer algorithmischer Ansatz dafür

Contents:

Computer vision with deep learning.

All following topics include the formal definition, the idea behind, the role in the big picture and relations to other topics.

- paradigms “analytically derived” vs. “maschine learning”
- the architecture of artificial neural networks which process images (“big picture” overview, where details follow later)
- different outputs: classification, semantic segmentation, heatmap, bounding boxes,

instance segmentation, application specific values

- layers: convolution, activation, pooling, unpooling, fully connected, batch normalization
- losses: absolute, quadratic, relative error, maximum-likelihood, crossentropy, weights, combined losses
- tensors, gradient descent optimization, viewing a NN as a graph of tensor operations, backpropagation
- receptive field
- typical CNN backbone architectures and their use in pretraining
- encoder-decoder architecture for image output, role of shortcuts
- object detection: how to output bounding boxes, one-shot vs. two-shot approach
- how to record, annotate and preprocess data, typical dataset problems
- development and evaluation of deep learning based computer vision systems
- applications in media, in particular image generation
- Generative Adversarial Networks (basic idea)
- 3D voxel images and depth images in CNNs

3D computer vision

- paradigm “from pixels to meters by solving equations”
 - point features
 - camera model
 - geometric reconstruction (which 3D properties can be derived from how many 2D points)
 - least squares as a generic approach to geometric reconstruction
-

Hinweise (*remarks*): In der Tabelle sind nur die primären/spezifischsten Module aufgelistet, denen diese Veranstaltung zugeordnet ist.