BA & MA Topics

Leibniz Universität Hannover

Fachgebiet Visual Analytics

Prof. Dr. Ralph Ewerth

Analysis of Student Drawings

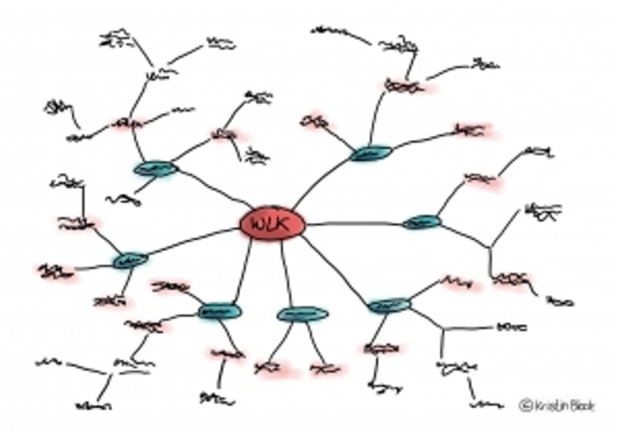
Leibniz Universität Hannover

Science teaching often uses multiple modalities

- Text describing processes and relationships
- Images showing structure or processes

Objective of the student project

- based on a dataset of drawings generated by students in class (physics or chemistry)
- develop methods for automatic
  - clustering of drawings representing similar level of conceptual understanding
  - automatic grading



Prior knowledge: Image analysis, computer vision

Skills: python, OpenCV, (pytorch, tensorflow)

Type of work: Master thesis, (Lab)

Historical Visual Question Answering

Leibniz Universität Hannover

Problem

- Answering questions related to historical photos is difficult as not only visual (objects), but also contextual information (time, place, persons, purpose) are required

Possible Task

- Create a dataset for the VQA task on historical images
- Crawl images from different sources (Flickr, Wikimedia, …)
- Generate question-answer (QA) pairs from image captions
- Evaluate different question generation (QG) methods

Prior knowledge: Deep Learning, Computer Vision

Skills: Python, Pytorch or Tensorflow

Type of work: Master Thesis, (Lab)

"Martin Luther King's 'I have a dream' speech in Memorial Park, Washington (1963)."

Identification of Narrative Patterns in News Videos

Leibniz Universität Hannover

Given a **corpus of news videos** from state and alternative media:

- **Extract a <u>selection</u> of features** from
  - Video (e.g., age, gender, emotion, text and action recognition)
  - Text from video via OCR (e.g., named entities)
  - Speech (e.g., named entity recognition, sentiment)
  - Audio (e.g., voice emotion, music style/genre)
- **Combine them to identify narrative patterns [Wu et al. 2018]**
- **Compare narrative patterns** in state and alternative media

Prior knowledge: Computer science, deep learning
Skills: python, (pytorch), (tensorflow)
Type of work: Master thesis, (Bachelor thesis), (Lab)

Features:
Video: war, anchor, medium shot, …
Text (OCR): "Krieg", "Ukraine", "Russland", …
Audio: tagesschau tune, emotion: neutral, …
Speech: greeting, anchor name, "Krieg", …

Narrative Patterns:
- Identification of event, time, and location
- Authoritative voice (news report in studio)
- …

[Wu et al. 2018] Thinking Like a Director: Film Editing Patterns for Virtual Cinematographic Storytelling, ACM Trans. Multim. Comput. Commun. Appl, 2018

Named Entity Linking in Scientific Videos

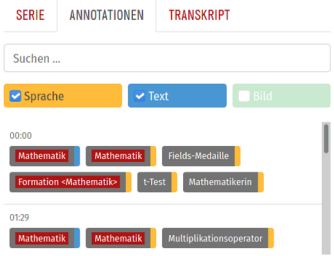
Leibniz Universität Hannover

- TIB AV-Portal contains thousands of scientific videos
- **Problem:** Topics within the videos are unknown

- Given: Transcripts from speech and text overlay
- Extract concepts of six core domains (e.g., mathematics)
- Link concepts to knowledge bases (e.g., Wikidata, GND)
- Fine-tune models for named entity recognition and linking

Prior knowledge: Natural Language Processing, Knowledge Graphs

Skills: python, (pytorch, tensorflow)
Type of work: Master thesis, (Lab)



Multimodal Hate Speech Detection for Videos

Leibniz Universität Hannover

Possible tasks

- Detect offensive languages or content in videos
- Analyse social media for hate speech
- Incorporate text, images, videos

Is this right for you?

- Master Thesis
- Skills: Python, Natural Language Processing, Computer Vision, Neural Networks

Image source: https://youthforeurope.eu/wp-content/uploads/2019/01/1080x580-hate-speech-2-garn-press-1080x580-1024x550.jpg

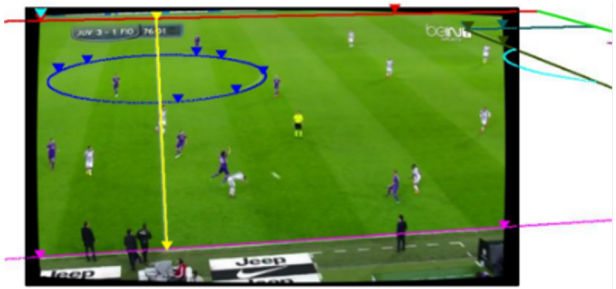
3D Sports Field Registration (Camera Calibration) - 1

Leibniz Universität Hannover

- given broadcast videos or individual images for team sports like soccer, handball, or basketball
- estimate camera position, orientation, and focal length (+radial lens distortion coefficients)

Objective of the student project

- Extend a given framework
  - for other team sports (currently soccer)
  - for temporal consistent predictions



Prior knowledge: Passed Deep Learning course; ideally basics in camera calibration

Skills: PyTorch

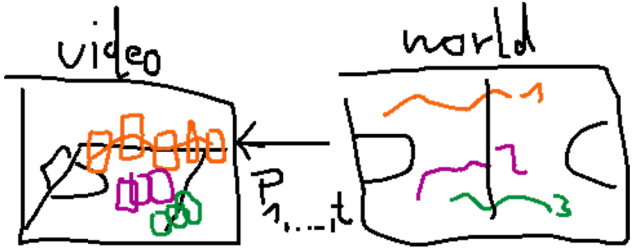
Type of work: Master Thesis, Lab

Further Reading: https://www.soccer-net.org/tasks/calibration

Fundamentals: R. Hartley / A. Zisserman. Multiple View Geometry in Computer Vision. Cambridge University Press, ISBN 0-521-62304- 9, 2000a

3D Sports Field Registration (Camera Calibration) - 2

- **Data**:
  - 2D/3D player trajectories in world space are known
  - Tracked players (bounding boxes or pose) in video
- **Idea / Task:** Jointly learn
  - (1) the assignment between tracked players and player trajectories
  - (2) camera parameters



Prior knowledge: Passed Deep Learning course; ideally basics in camera calibration

Skills: PyTorch

Type of work: Master Thesis

Fundamentals: R. Hartley / A. Zisserman. Multiple View Geometry in Computer Vision. Cambridge University Press, ISBN 0-521-62304- 9, 2000a

Event Detection in News Images

Leibniz Universität Hannover

Possible tasks

- Given an image as input estimate the event represented in the image
- Incorporate news body text to estimate the event represented in the image

Prior knowledge: Deep Learning, Computer Vision, Natural Language Processing

Skills: Python, PyTorch

Type of work: Master Thesis, Lab

London Olympics 2012

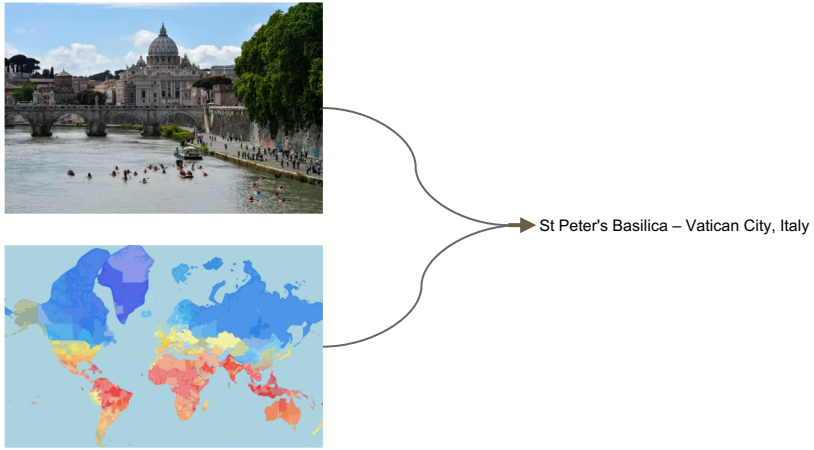
Geolocation Estimation using Auxiliary Information

Leibniz Universität Hannover

Possible tasks

- Extend existing models for geolocation estimation of images using auxiliary information, such as: Temperature, Precipitation, Population, and GDP to name but a few
- Experiment the explainability of the proposed model for the geolocation estimation

Prior knowledge: Deep Learning, Computer Vision
Skills: Python, PyTorch
Type of work: Master Thesis, (Lab)



Draft classification in artworks

Leibniz Universität Hannover

Possible task

- Classification and retrieval of drafts in art collections
- Application of style transfer for data generation
- Contrastive learning to match draft and image

Prior knowledge: Deep Learning, Computer Vision

Skills: Python, Tensorflow or PyTorch

Type of work: Master Thesis, Bachelor Thesis, Lab

Image-text relationship in art literature

Leibniz Universität Hannover

Possible task

- Pre-processing of scanned books and essays
- Training of a multi modal embedding for image and text from art history

Prior knowledge: Deep Learning, Computer Vision

Skills: Python, Tensorflow or PyTorch

Type of work: Master Thesis, Bachelor Thesis, Lab

SCHIFF

(Das Schiff der Kirche)

Vgl. ↗Ecclesia, ↗Narr, ↗Navicella

Abk.: K. = Kirche; S. = Schiff

I. Quellen. A. Patrist.: Die Gleichsetzung v. S. u. K. erscheint zuerst bei Tertull, De bapt. 12 (CSEL 20, 212), E. 2. Jh., u. findet allg. Verbreitung seit dem 4. Jh. bes. im Abendl.: Hippol, De Antichr. 59 (GCS Hippol I 2, 39), u. Aug, Serm. 75, 3 (PL 38, 475). Das im Meer umhergeworfene S. wird zum Bild der Anfangssituation der K. (Petrus Chrysol, Serm. 20 [PL 52, 254]), das nicht untergeht, da Christus als Steuermann das S. der K. lenkt (Ps.-Ambros, Serm. 46, 4, 10 [PL 17, 697]). In den sog. S.katalogen (Hippol, De Antichr. 59, u. Epist. Clementis ad Jacob. 14, 15 [PG 2, 49]) wird der Aufbau der K. m. den Funktionen u. der Besatzung eines S. verglichen. Der Mastbaum als Zeichen des guten S. wird dem Kreuz gleichgesetzt (Ambros, De virginitate 18, 118 [PL 16, 297]), das das S. der K. auf dem Meer lenkt. Das Tropaion als Symb. des Sieges ist notwendig für die sichere Fahrt über das Meer u. Ausdruck der Heilsgewißheit der K. (Just, Apol. I 55, 3 [Ed. de Otto I 1, 150]). Die Allegorie v. Kreuz als dem v. Mastbaum u. Antenne gebildeten Tropaion des Sieges steht im Mittelpunkt der

s steuert das Schiff der Kirche, Lombardische m 1480, Morg. Libr. Ms. 799 fol. 234v.

Knowledge graph for art documents

Leibniz Universität Hannover

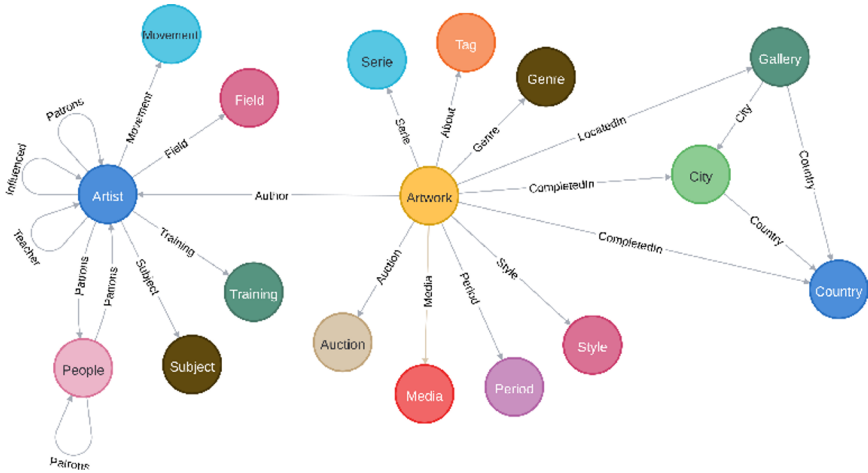
Possible task

- Knowledge discovery based on feature representation of images
- Model a knowledge graph from meta and text information
- Evaluation of graph based retrieval methods

Prior knowledge: Deep Learning, Computer Vision

Skills: Python, Tensorflow or PyTorch

Type of work: Master Thesis, Bachelor Thesis, Lab



Pace Analysis of Educational Videos

Leibniz Universität Hannover

The speed at which the information is presented to the student in educational videos can affect their learning by overloading their processing capacity

Possible tasks

- Analyze instructor's pace (speech rate, syllable duration, pauses)
- Analyze slides's pace (was the presentation of a slide too slow or too fast?)

Is this right for you?

- Master Thesis
- Skills: Python, Natural Language Processing, Computer Vision

Image source: Image source: https://www.shutterstock.com/video/clip-1048614721-young-student-watching-lesson-online-studying-home

Analysis of Informal Language in Educational Videos

Leibniz Universität Hannover

People learn better when the instructor use informal language instead of third-person constructions

Possible tasks

▪Extract automatically features that represent the attempt of the instructor to connect with the learner:, e.g. humor, praise, self-disclosure, asking questions, enthusiasm, inclusive language

Is this right for you?

▪Master Thesis
▪Skills: Python, Natural Language Processing,
Computer Vision

Created by Lia Rahdiah
from Noun Project

Patents Similarity Ranking using Multimodal Features

Leibniz Universität Hannover

Problem: Investigating multimodal features like visual (patent images) and text (caption, keywords) in similarity ranking.

Possible tasks:

1. **Learn to rank:** Giving relevant and non relevant document features to different machine learning models and highlight the best contribution of particular feature combination.
2. **Detection ambiguous record:** Using similarity to detect records which are visually or textually similar but belong to different category or class, this will help to eradicate or avoid in learning.

Prior knowledge: Computer Vision , Machine Learning

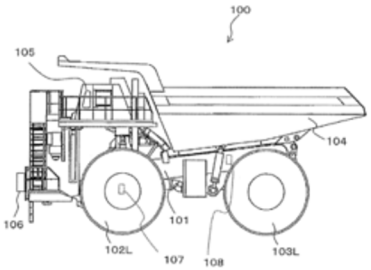
Skills: Python, PyTorch

Type of work: Master , Bachelor, Lab

Explore more: Deep learning for patent analysis

Multimodal approach for patent analysis

"FIG. 1 is a side view of an electrically-driven mining vehicle according to a first embodiment."



Example of an Image and caption from a patent document

Clustering Patent Image Types

Problem: Having large dataset of patent images without image type labels then use clustering algorithms to segregate images in to different classes.

Possible tasks:

1. **Clustering based on multimodal features:** In this task visual embeddings and textual embeddings form pretrained models will be used to cluster patent images on to different classes.

   Data set can be enhanced using variational autoencoder to improve clustering results.

   (last part of this task is advance and recommended for Master student only)

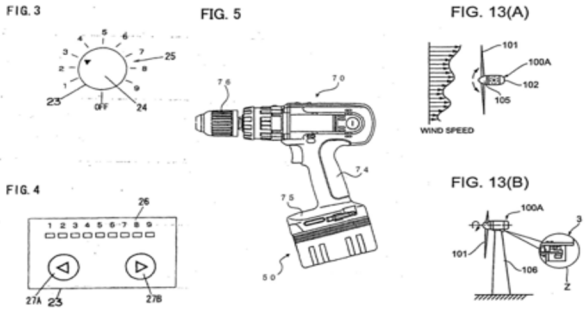
Prior knowledge: Computer Vision , Machine Learning

Skills: Python, PyTorch

Type of work: Master , Bachelor, Lab

Explore more: [Deep learning for patent analysis]

[Image clustering]



Example: Images from **same class (technical)**

Temporal Segmentation of Instructional Videos based on Multimodal similarities

Leibniz Universität Hannover

Problem: Divide instructional or educational videos in to different temporal segments and give it a possible title from keywords that will indicate which sub topic is under discussion.

Possible tasks:

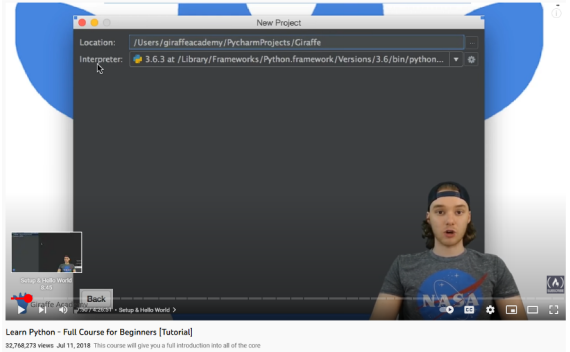
1. **Computing similarity in time span:** An instructional or educational video has visual frames, subtitles and audio which can give us multimodal features and similarities between these features can lead us to club data in a time span which can represent a temporal segment and keywords which can be collected from frames using ocr (optical character recognition), subtitles aur audio using asr (automatic speech recognition)

Prior knowledge: Computer Vision , Machine Learning

Skills: Python, PyTorch

Type of work: Master , Bachelor, Lab

Explore more: [Lecture Video Segmentation from Extracted Speech Content]



Example: Temporal divided tutorial (Learn python)

How Videos are changing Human Mood and Personality?

Leibniz Universität Hannover

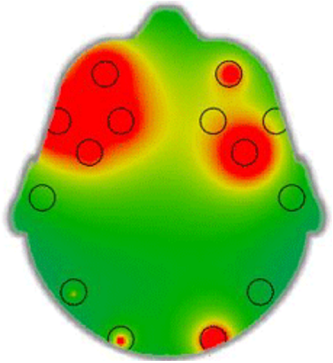
Problem

- Classification of visuals spans that have positive or negative effects on personality.
- Short term mood swing due to video segments via all modalities.

Possible tasks

1. Mood classification for short video extractions from movies (What kind of content videos contain i.e Sadness, Surprise, Fear, Anger), considering AMIGOS dataset [1]
2. Identify the shot span from a given video for target mood or personality classes

Prior knowledge: Computer Vision , Deep Learning
Skills: Python, PyTorch
Type of work: Master , Bachelor, Lab



[1] AMIGOS dataset source

Categorizing and Linking Equations in Learning Material

Leibniz Universität Hannover

Context

Automated analysis of learning material contents can be used for enriching their metadata which in turn could be used to improve reliant search systems. Equations are a key piece of learning material especially in physics and chemistry. Categorizing equations and linking symbols in equations to surrounding entities would be a useful feature for these use cases.

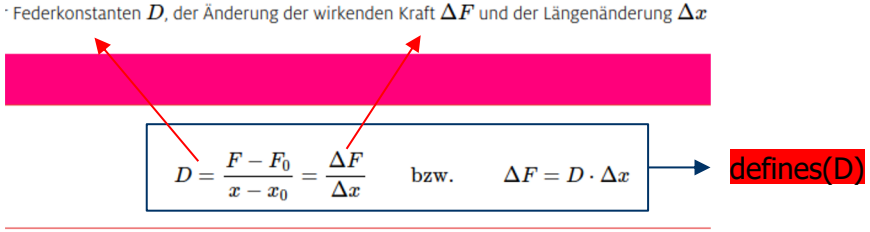
Problem

- Classify MathJAX equation descriptions from popular learning repositories: Leifi{Physik/Chemie} as definitions or illustrations.
- Link symbols to entities in the surrounding text.

Prior Knowledge: Seq-Seq models, Python

Skills: Python, PyTorch, SkLearn, NLTK

Type of work: Master, Bachelor



News recommendation using image-text Relations and user characteristics

Leibniz Universität Hannover

Problem: News representation is key to accurate news recommendation. In this project, the goal is incorporate heterogeneous information and different modalities in news for better news recommendation.

Possible Tasks:

1. **For labor:** Implementing and analyzing an existing framework on a dataset for news recommendation.
2. **Bachelor/Master thesis:** Develop a new model possibly incorporating image-text relations and user characteristics to achieve state-of-the-art performance.

Skills: python, deep learning frameworks like pytorch, Machine learning, computer vision, natural language Processing

Further Reading:
https://arxiv.org/pdf/2104.07407.pdf,
https://arxiv.org/abs/1906.08595

Generating task-agnostic data for understanding Vision-Language models

Leibniz Universität Hannover

Problem: In recent years, there has been a keen interest in developing large vision-language models by training in self-supervised fashion on millions (billions) of image-text pairs. But, many questions still remain unanswered: Do they understand context? What happens with small manipulations to either modalities?

Possible Tasks:

1. **For labor:** Fine-tuning and evaluating the models on a different downstream task which requires additional context.
2. **Bachelor/Master thesis:** Develop manipulation (on language or images) methods to create a dataset that causes failures or exposes weakness of the pre-trained models. An idea can be to use generative models to create images or text that change the prediction.

Skills: python, deep learning frameworks like pytorch, Machine learning, computer vision, natural language Processing

Further Reading: https://arxiv.org/abs/2112.07566

Incorporating context or auxiliary information in Vision-language models

Leibniz Universität Hannover

Problem: In recent years, there has been a keen interest in developing large vision-language models by training in self-supervised fashion on millions (billions) of image-text pairs. But, is it possible that all possible information is encoded through un-constrained training? What kind of auxiliary inputs or information can we add to learn better multimodal representation?

Possible Tasks:

1. **For labor:** Investigate ways (better evaluations/tasks?) to find weaknesses in large vision-language pre-trained models.
2. **Bachelor/Master thesis:** Develop a method to inject a new input(s) or auxiliary information during training or fine-tuning a vision-language model

Skills: python, deep learning frameworks like pytorch, Machine learning, computer vision, natural language Processing

Further Reading: https://arxiv.org/abs/2205.04363

Multimodal fake news detection in social media

Leibniz Universität Hannover

Problem

- Given an multimodal social media post, the goal is to detect whether this post is fake or real

Possible tasks

- Develop a method for incorporating multimodal input and related evidence for detecting the misinformation
- Extend existing datasets with retrieved related evidence which helps for validation task.

Prior knowledge: computer science, deep learning
Skills: Python, (PyTorch, Tensorflow)
Type of work: master, Lab

real

fake