
PROF. DR. MARTIN STEINEBACH

OPEN THESIS TOPICS

Our research group is always interested in motivated students who want to do their Bachelor's or Master's thesis. In the following we provide open thesis topics we can currently offer.

In case you are interested in these thesis subjects – or have your own thesis proposals – please don't hesitate to contact us:

E-Mail: Martin.Steinebach@sit.fraunhofer.de

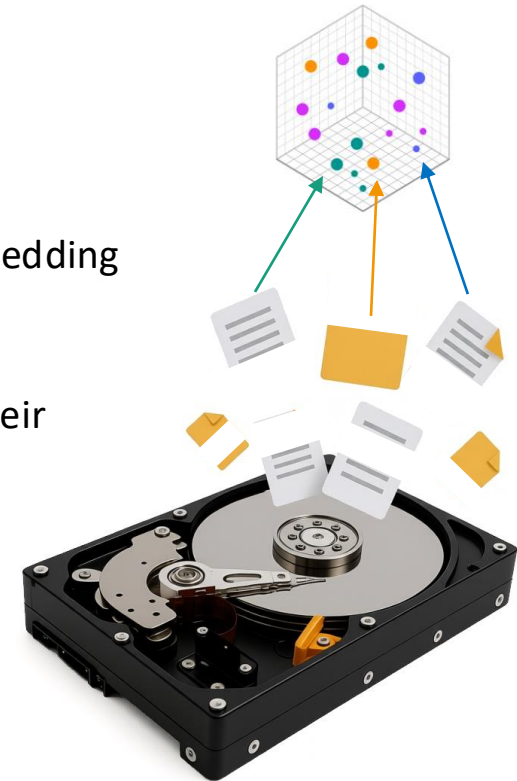
WWW: <https://www.sit.fraunhofer.de/de/mediasecurity/>
<https://www.sit.fraunhofer.de/de/itforensics/>

Please send a CV with info about programming skills and experiences as well as a transcript of records with your application.

Version 18.07.2025

LEARNED FILE FRAGMENT EMBEDDINGS FOR CONTENT-AWARE FILE CARVING

- Digital forensic investigators face storage mediums where data is damaged or intentionally removed. Recovering files (i.e. „File Carving“) may be crucial for forensic analyses, cybercrime investigations or in general legal cases.
- Existing approaches mainly aim to identify file types of file fragments. Reassembling fragments back to original entire files is challenging and still mostly done using heuristics or brute force procedures.
- The goals of this thesis are:
 - Analyze if modern Machine-Learning approaches can learn to extract representations (in form of embedding vectors) from file fragments, which unify information not only about their file type but also about coherence/matching potential to other fragments
 - Develop said embedding approach, allowing to quickly find connected file fragments by comparing their respective embedding vectors (maybe also incorporating vector database for fast lookups)
 - Evaluate accuracy and performance against existing methods (e.g. PhotoRec, Foremost, X-Ways Forensics, ...) on synthetic or real data drive images
 - Related Work (specifically for JPEG files):
<https://testojs.aca-p.com/index.php/acj/article/view/24>



AI-GENERATED TEXT DETECTION

- As the amount of AI-generated content is rapidly increases, there is a growing necessity to develop detectors that can distinguish between AI-generated and genuine text.
- The goals of this thesis are:
 - To analyze state-of-the-art detection methods and to identify potential avenues for improvement.
 - To provide an implementation of a classification system that can improve the results on real-world datasets (e.g. DetectRL)
- Earlier work regarding AI-generated text detection:
 - <https://github.com/junchaoIU/DetectRL>
 - <https://aclanthology.org/2023.ranlp-stud.1.pdf>



AUDIO DEEPAKE GENERATION

- The quality of audio deepfakes is rapidly improving in conjunction with the advances in artificial intelligence. Audio deepfake generation methods can be classified into two main categories: voice conversion (VC) and text-to-speech synthesis (TTS).
- The goals of this thesis are:
 - To analyze state-of-the-art generation methods (VC and/or TTS) and to identify potential avenues for improvement.
 - To provide an implementation of a generation system that can enhance the quality of the generated audio recording and increase its similarity to the target speaker
 - OR evaluate the characteristics of various generation methods
- Earlier work regarding audio deepfake generation:
 - <https://www.mdpi.com/2076-3417/13/5/3100>
 - <https://ieeexplore.ieee.org/stamp/stamp.jsp?arnumber=10096399>



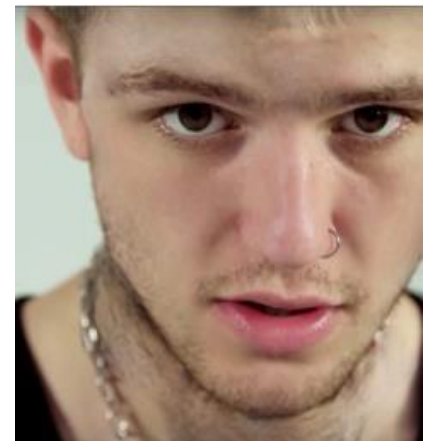
AUDIO DEEPFAKE DETECTION

- As the amount of AI-generated audio content on social media rapidly increases, there is a growing necessity to develop detectors that can distinguish between AI-generated and genuine audio.
- The goals of this thesis are:
 - To analyze state-of-the-art detection methods and to identify potential avenues for improvement.
 - To provide an implementation of a classification system that can improve the results on e.g. the in-the-wild dataset
- Earlier work regarding audio deepfake detection:
 - <https://www.mdpi.com/1999-4893/15/5/155>
 - <https://dl.acm.org/doi/pdf/10.1145/3658664.3659662>



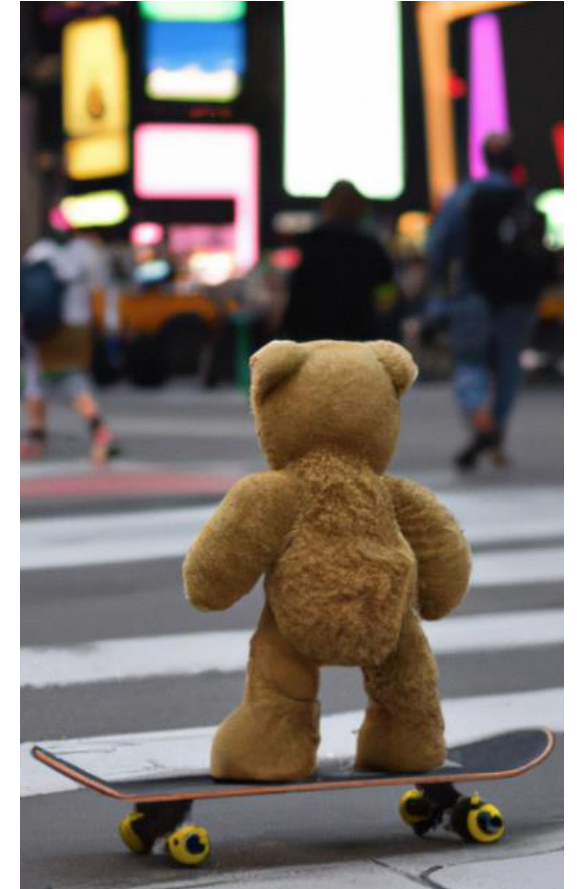
DETECTING SYNTHETIC IMAGE CONTENT CREATED BY „INPAINTING“

- Detail in digital images can be enhanced or created using ML-based image synthesis methods in terms of „inpainting“. Such enhancements can be applied for malicious purposes such as forging digital evidence or distributing fake news
- Current methods for identifying inpainting feature varying detection performance with respect to different image inpainting algorithms
- The goals of this thesis are:
 - To analyze the shortcomings of current detection algorithms for “inpainting”
 - To improve the forgery detection performance by exploiting common characteristics of a selection of synthesis algorithms
- Earlier work regarding the detection of splicing boundaries:
 - https://openaccess.thecvf.com/content_CVPR_2020/papers/Li_Face_X-Ray_for_More_General_Face_Forgery_Detection_CVPR_2020_paper.pdf



DETECTING SYNTHETIC IMAGES CREATED BY „FULL IMAGE SYNTHESIS“

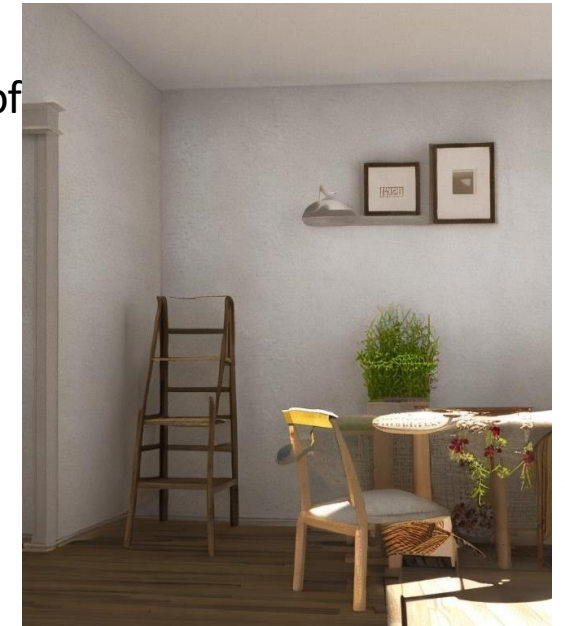
- Digital images can be created using ML-based image synthesis. In contrast to „inpainting“ *full* images can be synthesized from scratch in a photorealistic appearance. Such enhancements can be applied for malicious purposes such as making up digital evidence or fake news
- Current methods for “full image synthesis” identification feature varying detection performance with respect to different image synthesis algorithms
- The goals of this thesis are:
 - To analyze the shortcomings of current detection algorithms for “full image synthesis”
 - To implement a detection method exploiting characteristics of a selection of synthesis algorithms to improve the forgery detection performance
- Earlier works regarding the annotation of synthetically generated images:
 - <https://arxiv.org/pdf/2211.00680v1.pdf>



openai.com

RECOGNIZING ROOMS/LOCATIONS INSIDE BUILDINGS BASED ON REFERENCE DATA

- There exist several solutions for recognizing objects in images and videos as well as techniques to match visual data. For some use-cases training an ML-based solution is challenging as training data is scarce.
- Training an ML-based classifier on 3D synthesized data to estimate the facial landmarks of human faces has shown to outperform state-of-the-art methods
- The goals of this thesis are:
 - To analyze whether synthesizing training data in 3D space can improve the performance of methods trying to match the interior of a room
 - To evaluate the transferability on real data
- Requirements: Knowledge in 3D modelling and/or game engines
- Earlier works regarding the transfer of synthesized image data:
 - https://openaccess.thecvf.com/content/ICCV2021/papers/Wood_Fake_It_Till_You_Make_It_Face_Analysis_in_the_ICCV_2021_paper.pdf



Stable Diffusion

IMAGE MANIPULATION DETECTION

- Digital images are subject to manipulation. Powerful image editing tools make it possible to manipulate images without specialized skills. Therefore, image forensics is needed to verify the integrity and authenticity of digital images.
- Different types of manipulations leave different traces, which can not be easily detected by a single method. Moreover, such traces can be weakened or even eliminated by post-processing.
- The goal of this thesis is to
 - develop or improve a deep learning based or classical forensic algorithm for specific image manipulations
 - implement the forensic algorithm
 - evaluate the implemented algorithm using different datasets
- Related Work:
 - <https://www.sciencedirect.com/science/article/pii/S1051200417301938>

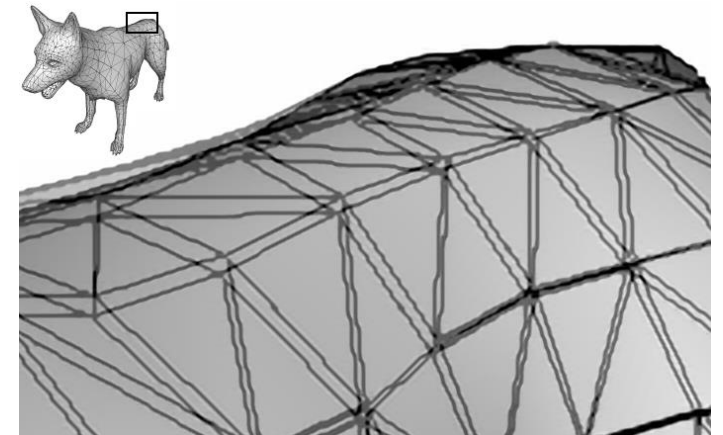


BENCHMARKS FOR LABELING AI-GENERATED CONTENT

- AI-generated content is increasingly prevalent across the internet and various media platforms. It is also becoming highly realistic and increasingly difficult to distinguish from authentic content. This poses significant challenges to content trust and introduces new risks of misinformation, manipulation, fraud and impersonation.
- Therefore, labeling AI-generated content has become essential. The EU AI Act stipulates that AI-generated content must be clearly and recognizably marked using appropriate technical measures to ensure traceability of its origin, for example through digital watermarking techniques.
- The objectives of this thesis is to
 - identify emerging challenges and requirements for digital watermarking in the context of labeling AI-generated content
 - design and implement a benchmarking tool for testing the resilience of existing watermarking algorithms for images and audios
 - test and evaluate the performance of the implemented benchmarking tool

3D-MODEL WATERMARK

- Decentralized production becomes more important with 3D printers
- 3D models are sent to the printers and printed/manufactured on-site
- Sharing the print data, or scanning a 3D model and duplicating it, is possible and that is where this process fails to progress
- The goal of this thesis
 - Is to design a digital watermarking technique for 3D models with the following properties
 - Written for G Code
 - Watermark extractable after 3D printing and scanning
 - Watermark is imperceptible to a person
 - Implement the designed watermark and evaluate it
 - A good starting point is e.g. <https://arxiv.org/abs/2109.07202>



SOCIAL MEDIA DATA ANALYSIS (1)

- We offer a variety of topics related to analyzing social media data. Possible topics include:

- Extracting Semantic Information from Images

You will be asked to select tools for image captioning and/or text identification in images. First, you will evaluate these tools using existing datasets. Then, you will apply them to a selected set of social media data (which you can either crawl yourself or use existing datasets). The goal is to compare the scalability and variations in outcomes across these extraction tools, and to assess their applicability to large-scale social media data analysis.

- Extracting Information from Audio

You will extract audio content (ex. from videos) and analyze it using tools such as speech-to-text, speaker recognition, and emotion detection. First, you will evaluate the tools on labeled data, then apply them to unlabeled social media audio (a dataset will be provided, or you may crawl your own) to assess scalability and variation in results. For example, you can compare transcription performance across different audio qualities, speaker attributes, and noise levels to evaluate tool suitability for large-scale social media analysis.

SOCIAL MEDIA DATA ANALYSIS (2)

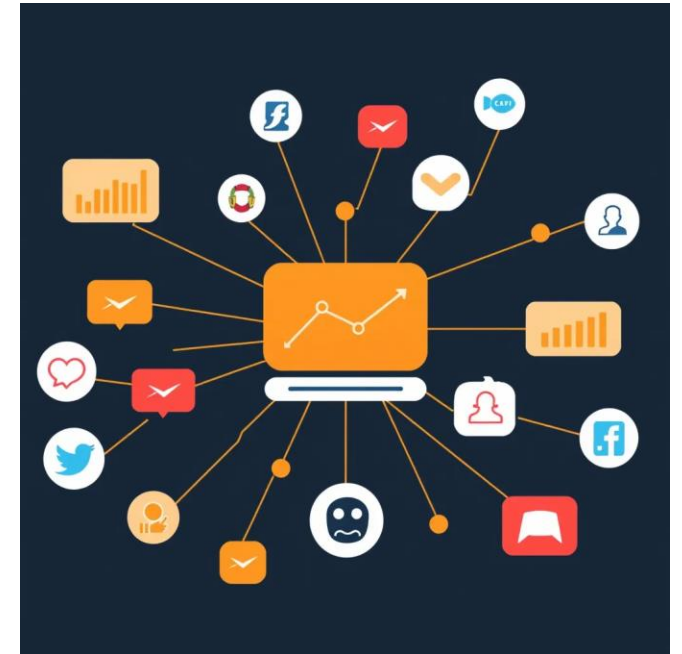
- Extracting Topical Dynamics Within a Telegram Channel/Group Over Time

You will analyze how topics and themes evolve over time within a Telegram channel or group. Various topic extraction methods should be evaluated, particularly with respect to scalability and performance. Additionally, these methods should be applied to multiple channels or groups to gain insights into information flow and thematic development.

(paper: <https://arxiv.org/pdf/2406.09556>)

- Network Analysis of Social Media Data

You will first construct different types of networks using social media data. The networks can be built from metadata (such as posts being forwarded) or content data (like posts discussing similar topics or expressing similar opinions). The structure of the network will also depend on the selected dataset. You will then analyze these networks to gain insights into information flow, influential actors, information sources, and the spread of information.



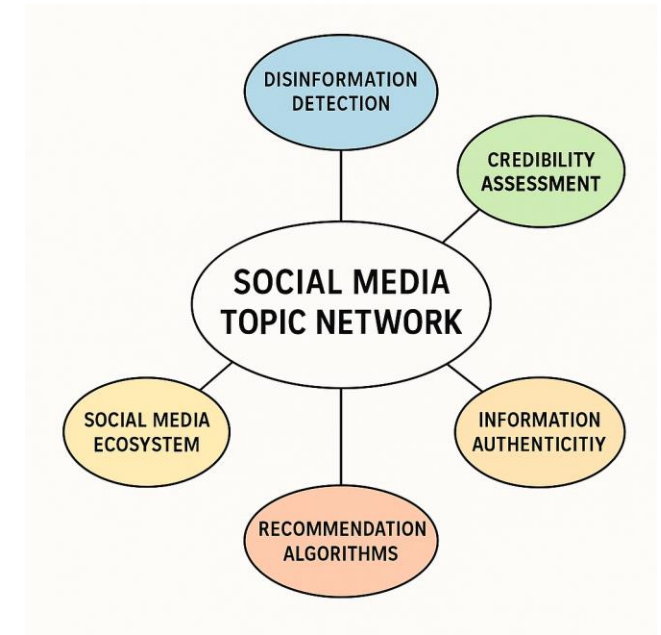
SOCIAL MEDIA DATA ANALYSIS (3)

■ Other relevant topics

Topics such as disinformation detection, fake news detection, credibility assessment, information authenticity, recommendation algorithms, and studies on the general social media ecosystem are also welcome.

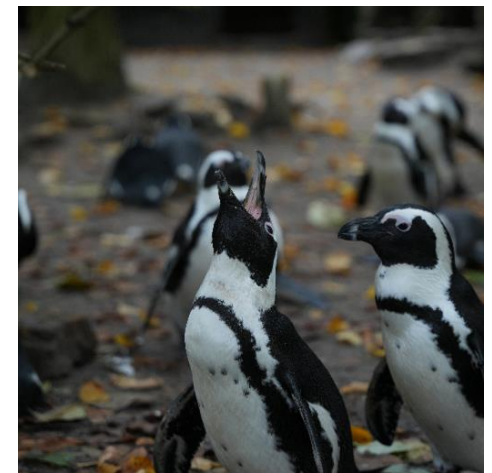
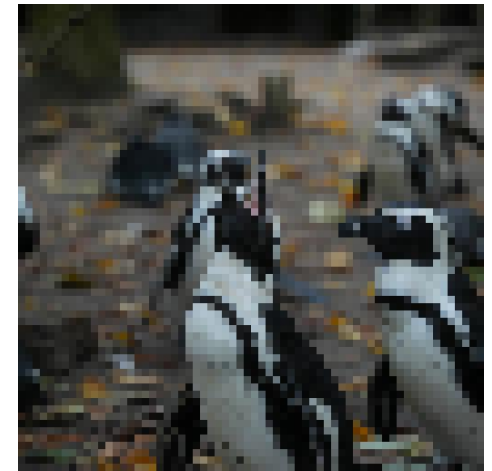
However, these should be addressed using technical approaches rather than purely qualitative methods.

Students from other majors outside of information science or information technology are also welcome, though the topic may need to be further discussed and adapted to align with their major.



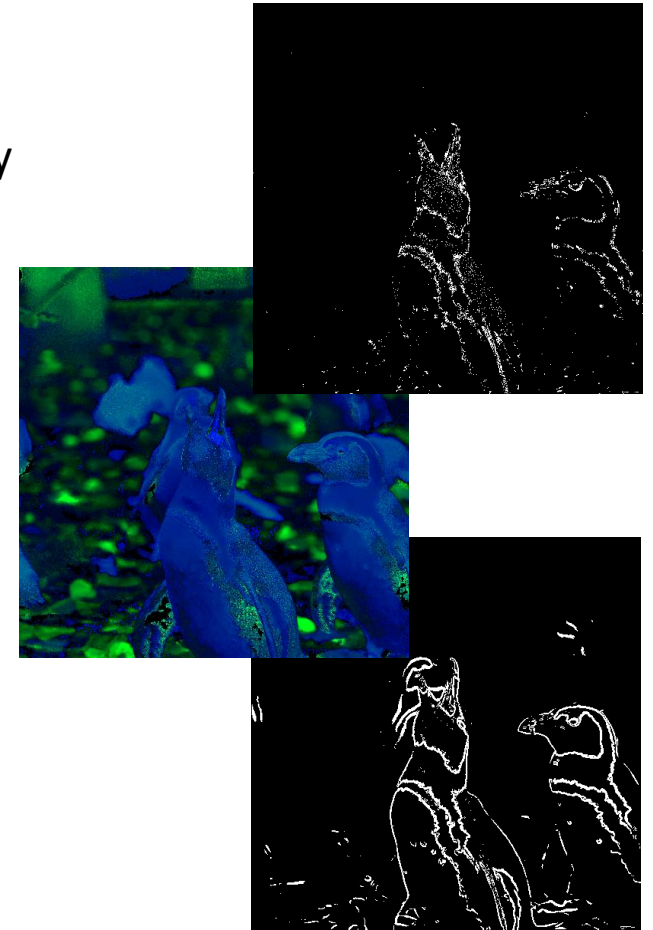
EVALUATION OF AI-BASED IMAGE SUPER-RESOLUTION

- Super-Resolution has important applications for beneficial purposes such as geographic maps or medical imaging (MR/CT/US). At the same time, it can be potentially misused to make AI-generated content appear more authentic, since these are often created in lower resolution
- To address this, a dataset is created from real images and their corresponding super-resolution images
- The goal is to evaluate different metrics, study the images in the frequency domain and search for hallucinated details and structures that differ from the original image
- Additionally, adversarial testing is conducted to hide or perturb the detected features
- Interesting papers
 - Overview of methods <https://arxiv.org/abs/2501.07855>
 - A recent method <https://arxiv.org/abs/2503.18446>



WIP: MULTI MODAL AI-IMAGE (FUSION) DETECTOR

- Current AI-image detection approaches typically process all image information through a single, uniform pipeline, even though color, noise, and edge features operate on different spatial scales. To more effectively exploit these complementary attributes and extract the maximal amount of relevant information, we propose a detector that treats color, noise, and edge information as separate input modalities
- The key challenge is to design a representation for each modality that captures its distinctive characteristics and to find an architecture capable of extracting these features efficiently
- Finally, the modality representations are fused into a unified embedding, which is then evaluated against state-of-the-art methods to assess its performance in identifying AI-generated images
- Related work using different modality inputs:
 - <https://www.nature.com/articles/s41598-025-93616-y>



CYBERGROOMING DETECTION IN ONLINE COMMUNICATION

- Cybergrooming is a serious and growing societal problem, where offenders exploit online communication platforms to target vulnerable individuals, particularly minors. Developing automated methods to detect such deceptive behavior can significantly contribute to safer online environments.
- The objective of this thesis is to investigate and develop methods for detecting cybergrooming behavior in online communication using SOTA NLP techniques.
- Possible research directions include cybergrooming content detection, early risk detection, real-time detection system, detection under low-resource conditions (e.g. transfer learning).
- A relevant benchmark task is the [PAN 2012 Sexual Predator Identification](#).
- While the primary focus is cybergrooming, related online harm domains may also be explored. However, the thesis should remain within the scope of content detection, early detection, real-time detection, or low-resource learning.
- Available dataset include, but not limited to, PAN 12.