

**Department of Computing**

**COMP5622 Research Seminar III**

**Seminar Report**

**Part I: [to be completed by student]**

|  |  |  |
| --- | --- | --- |
| **Student name:** | SHEN Leming | |
| **Student ID:** | 22040929R | |
| **Chief Supervisor:** | Dr Yuanqing Zheng | |
| **Submission date:** | 2023-12-13 | |
| **Word count:** | 2075 | **words (excluding references)** |

**List of research seminars attended:**

|  |  |  |  |
| --- | --- | --- | --- |
| **No.** | **Date** | **Title** | **Speaker** |
| 1 | 2023-12-11 | Cryptography in Blockchain | Prof. Joseph LIU |
| 2 | 2023-10.24 | Online gaming security, an example for network security research at KAUST | Prof. Marc DACIER |
| 3 | 2023-10.20 | Data-driven Scientific Discovery and Ubiquitous Computing: Systems, Algorithms, Applications | Prof. Qin (Christine) LV |
| 4 | 2023-10-04 | Recent Developments in Multi-modal Human-Machine Communication | Prof. Gerhard Rigoll |
| 5 | 2023-09-05 | My view on Research Methodology | Prof. Chengqi ZHANG |
| 6 | 2023-09-26 | Advanced Robotics for Effective Stroke Rehabilitation Treatment in Home Environment | Prof. Shane Xie |
| 7 | 2023-10-18 | Why is it easy to understand each other during a conversation? A neuro-cognitive model of language processing integrating facilitation mechanisms | Prof. Philippe BLACHE |
| 8 | 2023-11-14 | Self-improvement and Self-evolving of Large Language Models | Prof. Liu Qun |
| 9 | 2023-11-27 | AI for production automation | Prof. Dr.-Ing. Jorg Kruger |
| 10 | 2023-12-13 | Active Mobility Research as Climate Action | Prof. Maria Attard |

**Part II: [to be completed by Chief Supervisor]**

|  |  |
| --- | --- |
| **Grade of report:** | **Pass / Fail** |
| **Supervisor Signature:** |  |

**Part III: Seminar report [one of the attended seminars listed on Part I]**

|  |  |
| --- | --- |
| **Title:** | Recent Developments in Multi-modal Human-Machine Communication |
| **Speaker:** | Prof. Gerhard Rigoll |
| **Date of seminar:** | 2023-10-04 |

**(A): [Active participation in seminar]**

|  |
| --- |
| **A question you asked in the seminar:** |
| Thanks for the fabulous sharing of your recent research works which inspire me a lot. I have one question about the multi-view image recognition task. I suddenly came up with a similar scenario when we perform human activity recognition tasks, where IMU sensors are attached at different body parts capturing IMU data with different patterns. I’m wondering if it’s possible to adopt a similar model or training strategy to the multi-view image classification you mentioned in your recent works? |
| **Response from the seminar speaker:** |
| Absolutely, you can draw parallels between multi-view image classification and human activity recognition using IMU sensor data. While the modalities (image pixels vs. IMU sensor readings) are different, the underlying concept of leveraging information from multiple perspectives remains similar. You can explore architectures that fuse information from different IMU sensors, similar to how multi-view models combine features from different views. Recurrent Neural Networks (RNNs) or 3D Convolutional Neural Networks (CNNs) can be employed to capture temporal dependencies in IMU data. Just as data augmentation is used in image classification to improve model generalization, consider augmenting IMU data to enhance the model's ability to recognize activities under different conditions. Remember that the success of the model will depend on the characteristics of your dataset, the complexity of the activities, and the quality of the IMU sensor data. Experimentation and iterative model refinement are key aspects of developing effective models for human activity recognition using IMU sensors. |

**(B): [Written report – no less than 1500 words (excluding references, please attach the seminar poster of this report]**

**Abstract**

In the realm of multi-modal human-machine communication, users engage with machines through diverse human communication channels, including voice, vision, and haptics. This interdisciplinary domain intersects with pivotal research fields like speech recognition and computer vision. The foundational principles for these areas are intricately tied to advanced methods in pattern recognition and machine learning. The impressive strides in advanced machine learning methods over the past decade have significantly empowered human-machine communication, particularly within the well-established domain of deep learning. This presentation provides a concise overview of the evolution of classical algorithms in speech recognition and computer vision — a journey from traditional statistical methods to the prevalent use of pure neural recognition algorithms today. Additionally, we delve into recent research findings from our institution, exploring topics such as face recognition from partial or occluded face data, recognition for low-resolution faces, and action recognition, including gait identification using graph neural networks.

**1 Introduction**

In the dynamic landscape of human-machine interaction, the paradigm has shifted towards the incorporation of multiple communication channels, marking the advent of multi-modal interfaces. Users now seamlessly engage with machines through a spectrum of modalities, spanning voice, vision, and haptics. This intersectional domain amalgamates key research fields, prominently including speech recognition and computer vision, forging a path into the future of intelligent and responsive machines.

**Evolution of Human-Machine Interaction**. At the crux of these advancements lie the foundational principles rooted in pattern recognition and machine learning. Over the past decade, the relentless evolution of machine learning methodologies, especially within the realms of deep learning, has catalyzed a transformative era in human-machine communication. Notable works like "Speech Recognition with Deep Recurrent Neural Networks" [1] showcase the shift towards neural architectures in speech recognition, laying the groundwork for subsequent innovations.

**Revolutionizing Recognition**. Delving into the evolution of recognition algorithms, we witness the paradigm shift from rule-based and statistical models to the data-driven and adaptive nature of neural networks. Pioneering studies such as "ImageNet Classification with Deep Convolutional Neural Networks" [2] exemplify the transformative impact of deep learning on computer vision, demonstrating the superiority of convolutional neural networks in image classification tasks.

**Recent Research Findings**. Beyond the retrospective, this discourse ventures into the forefront of recent research emanating from our institution. Building on the foundations laid by "DeepFace: Closing the Gap to Human-Level Performance in Face Verification" [3], we explore the nuances of face recognition when faced with partial or occluded data. Additionally, works like "Low-Resolution Face Recognition in the Wild" [5] inspire our exploration into recognizing faces at lower resolutions, addressing challenges encountered in real-world scenarios.

**Innovative Approaches**. Notably, our exploration extends to the innovative domain of gait identification through the lens of graph neural networks. While seminal works like "Graph Neural Networks: A Review of Methods and Applications" [5] provide a comprehensive overview, our research endeavors push the boundaries by applying these techniques to the distinctive realm of gait recognition. The integration of graph neural networks offers novel insights into understanding and recognizing intricate patterns within the dynamics of human gait.

As we navigate through the chapters, a narrative unfolds, encapsulating the synergy of classical wisdom and avant-garde innovation that defines the landscape of multi-modal human-machine interaction. The chapters that follow provide a detailed exploration of our research endeavors, shedding light on both the historical context and the cutting-edge applications that shape the future of this dynamic field.

**2 Latest Research**

**2.1 ATTENTION-BASED PARTIAL FACE RECOGNITION [6]**

This paper presents a comprehensive investigation into the use of partial faces for face recognition in unconstrained environments. It addresses the limitations of traditional face recognition approaches, which are designed for controlled imaging conditions and struggle to handle partially visible faces in real-world scenarios. The proposed approach utilizes a modified ResNet-50 architecture to extract feature and attention maps, which are then re-calibrated and aggregated to obtain a single feature vector. The model is trained with adapted losses to handle partial faces and is evaluated under various benchmark protocols, including naturally and synthetically occluded partial faces. The study demonstrates that the proposed approach outperforms baseline methods, showcasing its effectiveness in recognizing partially occluded faces.

A diagram of a call center

Description automatically generated

Figure 1: Illustration of Proposed Partial FR Approach

The introduction highlights the challenges faced in recognizing partially visible faces in real-world scenarios and emphasizes the need for a model capable of handling partial faces. Traditional partial face recognition approaches are discussed, and the limitations of existing methods are highlighted, leading to the proposal of a new approach to partial face recognition using fixed-size images. The approach focuses on predicting attention maps capable of focusing on specific regions of interest in the face, independent of their positions.

The methodology section details the network architecture, including the Extract, Attend, and Aggregate modules (Figure 1). The Extract module uses a truncated ResNet-50 to obtain feature and attention maps. The Attend module re-calibrates the attention maps and performs attentional pooling, while the Aggregate module fuses the feature information into a single global feature vector. The study also discusses the loss functions used to train the model.

The experiments section describes the training and benchmark protocols used to evaluate the proposed approach. The results show that the proposed approach outperforms baseline methods, especially in the case of naturally occurring occlusions. An ablation study is conducted to analyze the impact of different parameters on the model's accuracy, demonstrating the effectiveness of the proposed approach.

In conclusion, the study presents a comprehensive approach to partial face recognition, demonstrating the model's ability to outperform baselines and provide satisfactory results for challenging partial face recognition protocols. The proposed method successfully addresses the challenges of recognizing partially occluded faces, making it a promising approach for unconstrained environments. The paper provides valuable insights into the development of a novel approach to partial face recognition, offering a comprehensive analysis of the model's performance and its potential applications in real-world scenarios.

**2.2 GAITGRAPH: GRAPH CONVOLUTIONAL NETWORK FOR**

**SKELETON-BASED GAIT RECOGNITION [7]**

The paper "GaitGraph: Graph Convolutional Network for Skeleton-Based Gait Recognition" presents a novel approach to gait recognition using a graph-based representation of human skeleton poses. The traditional methods for gait recognition often rely on silhouette images, which may lose spatial information and capture other visual clues beyond gait features. The proposed GaitGraph leverages human pose estimation to extract robust skeleton poses from RGB images, providing a cleaner representation of gait. By incorporating Graph Convolutional Network (GCN), GaitGraph enables powerful spatio-temporal modeling for gait recognition (Figure 2).

A diagram of a diagram of a number of lines

Description automatically generated with medium confidence

Figure 2: Overview of the Pipeline

The paper discusses the challenges in gait recognition, such as sensitivity to various factors like surface type, clothing, carried items, and clutter in the environment. It highlights the limitations of existing methods, which mostly rely on silhouette images extracted using background subtraction, posing challenges in complex real-world scenarios. The emergence of robust human pose estimators and their potential for new model-based methods for gait recognition is also emphasized.

The proposed GaitGraph approach is compared with existing methods, demonstrating significant improvements in gait recognition across various views and walking conditions. The paper also discusses the experimental validation of GaitGraph, showcasing its state-of-the-art performance in model-based gait recognition. Additionally, the availability of the code and models for GaitGraph is emphasized, enhancing its accessibility for future research and applications in gait recognition.

In summary, the paper introduces GaitGraph as a modern interpretation of model-based gait recognition, leveraging robust human pose estimation and powerful spatio-temporal modeling using GCN. The approach demonstrates superior performance in gait recognition compared to existing methods, particularly in modeling temporal gait features. The availability of the code and models further enhances its accessibility for future research and applications in gait recognition.

The paper also delves into the technical details of the GaitGraph approach, including the network architecture, graph convolutions, human pose extraction, and network implementation details. It provides an in-depth analysis of the experimental results, comparing GaitGraph with other state-of-the-art methods on the CASIA-B gait dataset. The paper also includes a comprehensive list of references, showcasing the extensive research and related work in the field of gait recognition.

Overall, the paper provides a comprehensive overview of the challenges in gait recognition, the development of GaitGraph, its experimental validation, and its comparison with existing methods, highlighting its potential for advancing the field of gait recognition.

**2.3 SUSCEPTIBILITY TO IMAGE RESOLUTION IN FACE**

**RECOGNITION AND TRAINING STRATEGIES TO ENHANCE**

**ROBUSTNESS [8]**

The paper "Susceptibility to Image Resolution in Face Recognition and Training Strategies to Enhance Robustness" by Martin Knoche, Stefan Hörman, and Gerhard Rigoll, published in Leibnitz Transactions on Embedded Systems in 2022, addresses the impact of image resolution on face recognition performance. The authors analyze the susceptibility of face verification to different image resolutions and propose two intuitive approaches to enhance performance, particularly for very low image resolutions (Figure 3). They also conduct multi-resolution learning and introduce three evaluation protocols focusing on low, mid, and high resolutions to measure performance in cross-resolution verification scenarios.

The paper discusses the experimental setup, including the baseline network structure, training process, and popular datasets used for testing face verification performance. The baseline network comprises a modified ResNet-50 pretrained on ImageNet and an ArcFace layer for classification. The authors also delve into the downsampling method and performance measurement methods for face verification accuracy.

Furthermore, the paper provides a comprehensive review of related work in generic face recognition, image resolutions, and cross-resolution face recognition. It discusses the existing methods and approaches in the field, including the use of loss functions to improve performance, synthetic downsample methods, and the challenges of cross-resolution face recognition.

A diagram of a person's face

Description automatically generated

Figure 3: Illustration of Methods in Face Recognition with CR Images

Overall, the paper offers valuable insights into the impact of image resolution on face recognition and presents practical strategies to enhance the robustness of face recognition models across different resolutions. It contributes to the advancement of face recognition technology and provides a benchmark for cross-resolution face verification.

**3 Discussion**

The seminar showcased a compelling array of research papers, each contributing distinct perspectives to the multifaceted realm of human-machine interaction. The papers collectively emphasized the evolving landscape of recognition technologies, shedding light on challenges and innovative solutions. In this discussion chapter, we delve into the insights gleaned from the presented papers, offering critical reflections and potential avenues for future exploration.

**Face Recognition and Image Resolution**. The exploration of image resolution's impact on face recognition, as presented by Knoche et al., underscores the nuanced challenges faced in real-world scenarios. The susceptibility of face verification to varying resolutions introduces a critical consideration for system designers. The proposed training strategies exhibit an intuitive response to the challenges posed by very low image resolutions. However, a deeper investigation into the scalability of these strategies across diverse datasets and real-world conditions would enhance the generalizability of the findings.

**Gait Recognition through Graph Convolutional Networks**. The GaitGraph paper introduces a refreshing departure from conventional gait recognition methodologies by leveraging graph convolutional networks. The move from silhouette-based approaches to human skeleton pose representations reflects a commitment to richer, more informative data sources. The adoption of Graph Convolutional Networks empowers the model with spatio-temporal modeling capabilities, providing a holistic understanding of gait dynamics. However, the generalizability of GaitGraph across diverse populations and environmental conditions warrants further exploration, considering potential biases and challenges in real-world deployment.

**Partial Face Recognition in Unconstrained Environments.** The investigation into recognizing partially occluded faces in unconstrained environments marks a pivotal shift in addressing the limitations of traditional face recognition systems. The proposed ResNet-50 architecture, designed to handle partial faces, showcases promising results under various benchmark protocols. The study's emphasis on adapting losses for training and evaluating under realistic occlusion scenarios positions it as a significant advancement. However, the robustness of the model to handle varying degrees and types of partial occlusion, especially in dynamic, uncontrolled settings, remains a focal point for future investigations.

**Critical Reflections and Future Directions**. While each paper contributes substantially to its respective domain, there are overarching considerations that merit attention. The generalization of proposed methodologies to diverse demographic groups, potential biases in training data, and the ethical implications of deployment in real-world scenarios are critical aspects deserving of deeper scrutiny. Additionally, future research could explore the integration of these recognition technologies into holistic multi-modal systems, offering a more comprehensive understanding of user-machine interactions.

**4 Conclusion**

In conclusion, the seminar's diverse set of papers underscores the dynamism within recognition technologies and sets the stage for continued exploration and innovation. The critical reflections presented here aim to catalyze discussions, inspire future research, and contribute to the ongoing evolution of human-machine interaction.

**Reference**

1. Graves, A., Mohamed, A. R., & Hinton, G. (2013, May). Speech recognition with deep recurrent neural networks. In 2013 IEEE international conference on acoustics, speech and signal processing (pp. 6645-6649). Ieee.
2. Krizhevsky, A., Sutskever, I., & Hinton, G. E. (2012). Imagenet classification with deep convolutional neural networks. Advances in neural information processing systems, 25.
3. Taigman, Y., Yang, M., Ranzato, M. A., & Wolf, L. (2014). Deepface: Closing the gap to human-level performance in face verification. In Proceedings of the IEEE conference on computer vision and pattern recognition (pp. 1701-1708).
4. Li, P., Prieto, L., Mery, D., & Flynn, P. J. (2019). On low-resolution face recognition in the wild: Comparisons and new techniques. IEEE Transactions on Information Forensics and Security, 14(8), 2000-2012.
5. Zhou, Jie, Ganqu Cui, Shengding Hu, Zhengyan Zhang, Cheng Yang, Zhiyuan Liu, Lifeng Wang, Changcheng Li, and Maosong Sun. "Graph neural networks: A review of methods and applications." AI open 1 (2020): 57-81.
6. Hörmann, S., Zhang, Z., Knoche, M., Teepe, T., & Rigoll, G. (2021, September). Attention-based partial face recognition. In 2021 IEEE International Conference on Image Processing (ICIP) (pp. 2978-2982). IEEE.
7. Teepe, T., Khan, A., Gilg, J., Herzog, F., Hörmann, S., & Rigoll, G. (2021, September). Gaitgraph: Graph convolutional network for skeleton-based gait recognition. In 2021 IEEE International Conference on Image Processing (ICIP) (pp. 2314-2318). IEEE.
8. Knoche, M., Hörmann, S., & Rigoll, G. (2022). Susceptibility to image resolution in face recognition and training strategies to enhance robustness. Leibniz Transactions on Embedded Systems, 8(1), 01-1.

