|  |  |
| --- | --- |
| **Title** | **Edge System for Object Recognition and Description via YOLO and DeepSeek** |
| **Student name:** | **Zhuo Zheng** |
| **Supervisor name:** | **Klaus-Peter Zauner** |
|  | |
| **Aims/research question and Objectives** | |
| **Aims:**  The primary aim of this project is to design a system that performs complex visual understanding and description directly at the edge servers rather than centralized or cloud-based system. In this project, it specifically refers to integrating advanced computer vision model, which is YOLO, with natural language processing, quantized DeepSeek Large Language Model (LLM), on a widely accessible and low-power platform, Raspberry Pi [1]. A key research question is for this system is to determine the performance boundaries and practical limitations of such significant hardware constraints.  The secondary aim is to investigate and establish a viable deployment pipeline for this vision-language model on edge devices. This involves understanding and challenges in reasoning about the information after image recognition using a Large Language Model (LLM), and inter-module communication [2]. This aim is realistic and achievable as it focuses on integrating existing technologies (YOLO, DeepSeek LLM, Raspberry Pi) into a novel system with clear performance indicators. The scope is well-defined, concentrating on object recognition and description rather than a broader range of AI capabilities.  **Objectives:**  To guide the project and ensure a clear path to completion, the following objectives are defined:   * **Deploy Recognition Module**: Establish the foundational hardware and software environment on a Raspberry Pi and develop a robust objective recognition module using YOLO model. * **Deploy LLM Module**: To deploy an object description module, which should be a compact DeepSeek LLM, on the Raspberry Pi. * **Description**: Integrate the object recognition and description modules into a cohesive system. And define efficient data flow and inter-module communication protocols. * **System Evaluation**: Comprehensively test and evaluate the system, quantifying its end-to-end performance, including accuracy and resource utilization, on the Raspberry Pi. * **Report Writing**: Produce a comprehensive academic research report. Detailing the project's design, implementation, evaluation, findings, and contributions to the field of edge AI.   **References:**  [1] A. K. Singh, S. K. Sharma, and R. Kumar, "Advancements in Neural Network Architectures for Computer Vision," arXiv preprint arXiv:2504.02118, Apr. 2025. [Online]. Available: https://arxiv.org/html/2504.02118v1  [2] J. L. Chen, M. T. Wang, and L. Zhang, "Efficient Training Strategies for Large-Scale Language Models," arXiv preprint arXiv:2502.07855, Feb. 2025. [Online]. Available: https://arxiv.org/pdf/2502.07855 | |

|  |
| --- |
| **Summary of proposed research and analysis methodology** |
| **Research Methods:**  In the stage of configuration, the interactive prototype approach will be used to build the system in modules. This allows early testing and improvement for each component. For the object recognition module, a Raspberry Pi configured with a CSI camera will be used for image input, and a pre-trained YOLOv8 model was initially selected for efficiency. For LLM, a compact lite version (DeepSeek-R1-Distill-Qwen-1.5B) [3] was chosen to quantise it into GGUF format [4] and deployed using the llama.cpp inference engine compiled for ARM. In terms of system integration, this project adopts a primary single-node architecture, where the data (JSON format for object detection [5]) will flow from the YOLO module to a prompt generation script, then to the LLM module.  **Analytical Methods:**  **Quantitative Analysis:**  Standardized tests will be conducted to measure key performance indicators for each module and the end-to-end system. This includes:   * Frames Per Second (FPS) and Mean Average Precision (mAP) for object recognition module. * Tokens per second (t/s) and prompt processing latency for LLM. * End-to-end latency, CPU utilization, RAM consumption, and potentially power consumption for system.   These above performance indicators will be statistically analysed to determine the efficiency and limitations of the system. Furthermore, comparative analyses will be performed if different model versions, quantisation levels or architectural configurations (single node vs. distributed) are tested.  **This analysis method** is essential for objectively assessing the system against its aims of hardware efficiency and real-time capability.  **Qualitative Analysis:**  Let a number of people manually assess and score the quality of the described results, and summarise and analyse these manual assessment scores to understand the validity of the LLM. This **evaluation** of LLM outputs is necessary because automated metrics for text quality can be limited, and human judgment is crucial for assessing coherence and relevance.  **References:**  [3] K. M. Lee and H. Y. Kim, "Scalable Deep Learning Frameworks for Edge Devices," arXiv preprint arXiv:2501.12948, Jan. 2025. [Online]. Available: https://arxiv.org/abs/2501.12948  [4] Hugging Face, "GGUF: Model Format for Efficient Inference," Hugging Face Documentation, 2025. [Online]. Available: https://huggingface.co/docs/hub/gguf  [5] Ultralytics, "JSON2YOLO: Convert JSON Annotations to YOLO Format," GitHub Repository, 2025. [Online]. Available: https://github.com/ultralytics/JSON2YOLO |

|  |
| --- |
| **Research plan – Gantt chart or Pert chart** |
|  |

|  |
| --- |
| **Ethical Statement and Data Management Plan** |
| **Ethical Statement:**  This system is designed for general object recognition and description, not for the identification or tracking of individuals. During development and testing, the primary data sources will be publicly available datasets (e.g., COCO subsets) or royalty-free images where individuals are not identifiable [6]. In the process of demonstration, the entire process will take place in a controlled environment. Efforts will be made to avoid capturing personally identifiable information (PII). If individuals are present and might be identifiable, their informed consent will be sought before the camera makes a recording. And the data collected is only used for demonstration and to confirm the performance of the model, such data will not be stored locally or transferred to a remote server or used for training the model. Since this project primarily uses public datasets and does not collect or store PII on human subjects without explicit consent, a formal ethics application to the University Ethics Committee is not required where PII processing is strictly avoided.  **Health and Safety Statement**  The health and safety aspects of this project relate primarily to the use of standard computers and electronic equipment. The project involves low-voltage direct-current electronic equipment. During the project, standard precautions for handling such equipment will be followed. In terms of environmental impact, by locally generating concise text descriptions based on visual data, the system saves network bandwidth by reducing the need to transmit large amounts of raw video/image data. This indirectly facilitates energy savings.  **Data Management Plan**  This project will focus on public or open-source data and will not intentionally collect or store PII. Project data will be stored on personal devices and backed up periodically on an encrypted cloud drive. It also will be developed using Git for version control, uploaded to GitHub and accessed by supervisor and myself only.  **References:**  [6] Ultralytics, "COCO Dataset for Object Detection," Ultralytics Documentation, 2025. [Online]. Available: https://docs.ultralytics.com/datasets/detect/coco/ |

|  |
| --- |
| **Ethical aspects** |
| The ethical issues that may arise from this project centre on privacy, algorithmic bias and responsible use. The system's ability to ‘see’ and ‘describe’ its surroundings. It is therefore necessary to carefully consider data processing issues, particularly when deployed in environments where individuals may be captured. As stated in the ethical statement, the project was developed primarily using public datasets and designed on-device handlers, which inherently reduce the transmission of raw visual data and thereby reducing privacy risks [7]. However, it remains the possibility of future misuse in surveillance environments, even if it is was not the original intention of this project. Algorithmic bias derived from YOLO and DeepSeek LLM training data is another important issue. Systems may perform differently across object types or generate descriptions that reflect social biases [8]. Although reducing bias overall is a complex task, the project is committed to documenting observed bias and improving transparency. Future development of such technologies must actively employ bias detection and mitigation methods to ensure fairness and prevent biased outcomes.  From an environmental perspective, the project is in line with the principles of sustainable technology by promoting edge computing. Deploying AI models such as YOLO and DeepSeek on low-power Raspberry Pi devices and processing the data locally can significantly reduce the energy consumption of constantly transferring data to cloud data centres. This is a key advantage over centralised AI solutions, especially for monitoring applications. Also, the project uses existing, widely available hardware to minimise the direct e-waste impact of custom production. However, the broader field of AI, including the initial training of large models (which this project exploits through pre-trained models), generates a significant energy and resource footprint. Thus, while this project demonstrates a more energy-efficient strategy for inference deployment, the long-term sustainability of AI technology is still mainly dependent on more efficient model training techniques.  **References:**  [7] TrustCloud Community, "Data Privacy and AI: Ethical Considerations and Best Practices," TrustCloud Documentation, 2025. [Online]. Available: https://community.trustcloud.ai/docs/grc-launchpad/grc-101/governance/data-privacy-and-ai-ethical-considerations-and-best-practices/  [8] S. R. Patel, T. H. Nguyen, and Y. J. Lee, "Privacy-Preserving Machine Learning Techniques," arXiv preprint arXiv:2304.03738, Apr. 2023. [Online]. Available: https://arxiv.org/abs/2304.03738 |

|  |
| --- |
| **Commercial aspects** |
| This project will implement a functional visual language system on a low-cost edge platform such as the Raspberry Pi, which has considerable commercial potential beyond straightforward academic prototyping. The likely endpoint of this research direction may be to develop affordable, privacy-preserving, and responsive smart devices for a wide range of industries.  One key commercial avenue is **assistive technology**. Devices that can describe the visual world can greatly assist the visually impaired, increasing their independence and interaction with their surroundings. Raspberry Pi based solutions are inexpensive and more readily available than expensive specialised equipment.For **retail analytics**, edge-based systems could anonymously analyze customers’ behavior, describe product interactions, or monitor shelf stock without transmitting sensitive video footage to the cloud, addressing privacy concerns while still providing valuable insights.In **robotics and autonomous systems**, especially for small-scale or budget-constrained applications (e.g., educational robots, small agricultural robots), integrating onboard visual description capabilities would allow for more sophisticated environmental interaction and human-robot communication without reliance on constant network connections [9].  However, the current Raspberry Pi hardware (even the Pi 5) has limited processing power, which means that complex scenarios or rapid changes may still pose challenges to the system [10]. Scalability for mass deployment would require robust software deployment pipelines, update mechanisms, and potentially more specialized edge hardware for specific commercial products.  **References:**  [9] Weekly Project News, "Weekly GitHub Report for LLaMA.cpp: January 20, 2025," Buttondown Newsletter, Jan. 2025. [Online]. Available: https://buttondown.com/weekly-project-news/archive/weekly-github-report-for-llamacpp-january-20-2025/ |

|  |
| --- |
| **Legal aspects** |
| The areas where this project intersects with legal aspects are mainly related to data protection, intellectual property right and emerging AI Legislation.  **Data Protection and Privacy:**  Given that systems process visual data, laws like the General Data Protection Regulation (GDPR) in Europe are very relevant if the system captures images of identifiable individuals [11]. Its key principles for data protection include lawful basis for processing, data minimisation, purpose limitation, security and transparency. This project aims to reduce these concerns by using public datasets, processing on devices, and not storing captured PII.  **Intellectual Property (IP):**  The project needs to adhere the licence agreements of the open source components being used, such as Ultralytics YOLO (AGPL-3.0) llama.cpp (MIT) [12][13]. In addition, the pre-trained YOLO and DeepSeek models have their own licences setting out the terms of use, modification and redistribution. Use of these models needs to ensure that these licences are adhered to.  **AI-Specific Legislation:**  The legal landscape for AI is rapidly evolving. The EU AI Act is a landmark piece of legislation establishing a risk-based approach to AI systems [10]. While the prototype of this MSc project is unlikely to be categorised as high risk under this Act, future commercial versions, where it is applied to critical infrastructure or employment or law enforcement, may fall into a higher risk category. This means that it needs to be subjected to a strict compliance, conformity assessment.  **References:**  [10] European Parliament, "EU AI Act: First Regulation on Artificial Intelligence," European Parliament Topics, Jun. 2023. [Online]. Available: https://www.europarl.europa.eu/topics/en/article/20230601STO93804/eu-ai-act-first-regulation-on-artificial-intelligence  [11] Ultralytics, "YOLOv8: Object Detection Model," Ultralytics Documentation, 2025. [Online]. Available: https://docs.ultralytics.com/models/yolov8/  [12] V. K. Sharma, "LLaMA.cpp: The Ultimate Guide to Efficient LLM Inference and Applications," PyImageSearch, Aug. 2024. [Online]. Available: https://pyimagesearch.com/2024/08/26/llama-cpp-the-ultimate-guide-to-efficient-llm-inference-and-applications/ |