Informatics PhD projects at King's College London, AY 24-25 — Computer vision

The PhD project proposals listed below will be considered for 2024/25 studentships available in the Department of Informatics to start 1 October 2024 or later during the 2024/25 academic year. Please note that this list is not inclusive and potential applicants can alternatively identify and contact appropriate supervisors outlining their background and research interests or proposing their own project ideas.
The PhD projects are listed in two groups. In the first group are the projects with allocated studentships: each project in this group has one allocated studentship. The remaining studentships will be considered for the projects listed in the second group. The number of those remaining studentships is smaller than the number of the projects in the second group. The allocation of studentships will be based on the merits of individual applications. Applications for PhD studies in the Department of Informatics, for all listed projects as well as for other projects agreed with supervisors, are also welcome from students applying for other funding (within other studentship schemes) and from self-funded students. See also this list of funding opportunities available at King's for post-graduate research in Computer Science.

Scholarship Allocated

(Back to Top)

- Synthetic video generation: counterfactual explanations
- Text promptable surgical video generation

Synthetic video generation: counterfactual explanations

Supervisor: Luis C. Garcia Peraza Herrera

Areas: Artificial Intelligence (AI), Machine Learning (ML), Computer vision

(Back to Scholarship Allocated)

Project Description

Counterfactual explanations provide valuable insights into machine learning models. They reveal the minimum changes required in the input to yield a different output, as illustrated in Fig. 1 below. In the case of deep learning models using images as input [1, 2], the counterfactual explanation is also presented as an image:

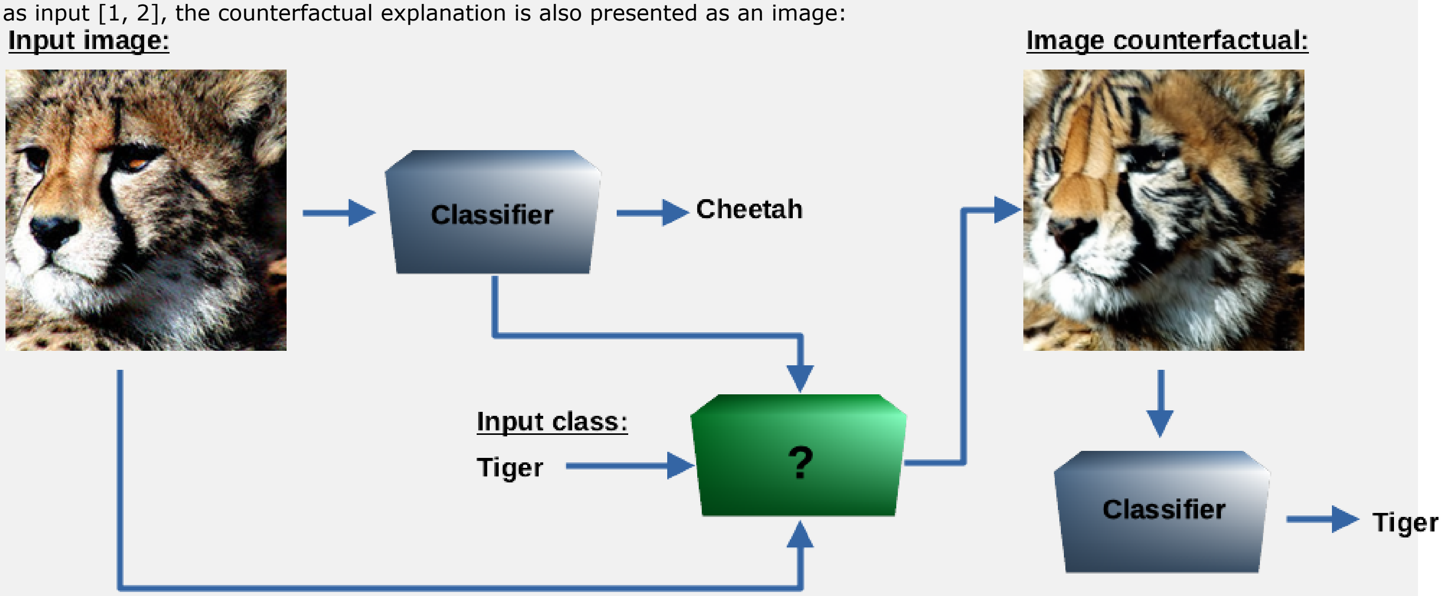

Figure 1. Image counterfactual explanation. The objective of this project is to extend this concept to video data. Instead of dealing with static images, we aim to devise machine learning methods (represented by [?] in Fig. 1 above) for generating video counterfactual explanations. A video counterfactual explanation 1) minimally alters a given input video, and 2) causes the video classifier to predict a different and specific class compared to the original input video. Although our project will focus on developing methods to create video counterfactual explanations specifically tailored to video classifiers, these methods can potentially be applied to other domains as well (e.g. understanding why autonomous robotic systems predict certain actions based on video input). This technology has several applications in the medical domain. Particularly in the realm of computer vision for surgery, the ability to generate synthetic videos has a multitude of potential applications. It holds the capacity to create synthetic datasets for training deep learning models and develop simulators that replicate surgical scenarios, offering clinicians a platform for sharpening their surgical skills.

References

[1] Boreiko et al. Sparse Visual Counterfactual Explanations in Image Space, DAGM GCPR, 2022.
[2] Augustin et al. Diffusion visual counterfactual explanations, NeurIPS, 2022.

Text promptable surgical video generation

Supervisor: Luis C. Garcia Peraza Herrera

Areas: Artificial Intelligence (AI), Machine Learning (ML), Computer vision

(Back to Scholarship Allocated)

Project Description

The goal of this PhD project is to develop an innovative framework for generating synthetic surgical videos through command prompts. This research aims to advance the field of surgical simulation by creating realistic and diverse datasets for training and evaluating computer vision models in surgery.

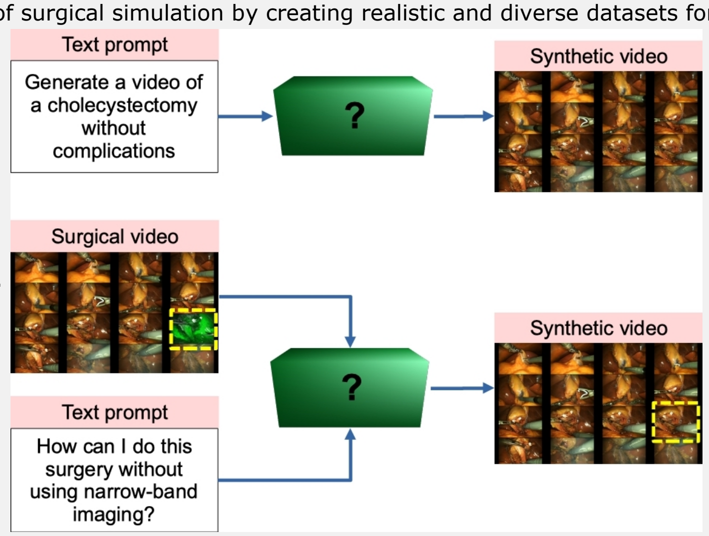

Figure 1. Text promptable surgical video generation. Our ultimate aspiration is to establish a system akin to DALL-E, whereby we can seamlessly request the generation of synthetic surgical videos on-demand. We aim to explore methods that use command prompts as a guiding mechanism, investigating the integration of procedural commands to control the content, complexity, and variability of the simulated surgeries. This involves improving the visual quality, mimicking real-world variations in surgical procedures, and incorporating dynamic factors such as tissue deformation, blood, and instrument interactions. The generated videos should cover a wide spectrum of medical procedures, surgical tools, and operating conditions to improve the robustness and generalization of the simulation.

References

[1] Garcia-Peraza-Herrera et al. Image Compositing for Segmentation of Surgical Tools without Manual Annotations, IEEE TMI, 2021.
[2] Singer et al. Make-A-Video: Text-to-Video Generation without Text-Video Data, ICLR, 2023.

Scholarship Not Allocated

(Back to Top)

- Co-Improving Generative AI Systems
- SLAM in Wearable Robots
- Improving active learning strategies for limited annotation budgets

Co-Improving Generative AI Systems

Supervisor: Dr Hector Menendez Benito (1st) and Dr Karine Even-Mendoza (2nd)

Areas: Artificial Intelligence (AI), Machine Learning (ML), Computer vision, Data science

(Back to Scholarship Not Allocated)

Project Description

Generative AI is non-deterministic; therefore, creating an appropriate test suite to evaluate whether a generative AI is reaching its expected outcome becomes a difficult task. Thus, it can hardly be measurable with metrics currently used to evaluate the quality of software. In the context of generative AI, we need to consider that for every prompt the user introduces, a different outcome can be created. Consequently, multiple repetitions of the prompt are required to understand whether the learning process has been fulfilled. It also applies to the different parameters of the models. In many cases, this requires manual evaluation, which is difficult and costly to scale. In the context of image generation, this problem can be alleviated under specific constraints by using a combination of different AIs. In our previous work, we created a tool called StableYolo that was able to select proper parameters for the generative AI process (under the Stable Diffusion model) by using automatic feedback from a visualization model (YOLO). This automatic feedback was focused mainly on photorealistic images, and in combination with search, it was able to identify proper parameters for the system and engineer both the positive and negative prompts to the best possible combination of words. This PhD proposal aims to focus on the generalization problem of this strategy. The main goal is to investigate how different artificial intelligence models can be combined to improve their quality. The student will start by extending the idea of generative AI in images, focusing not only on a photorealistic environment but also on other possible environments. This will also attempt to generate multiple objectives for the optimization process that aim to improve not only the quality of the problems but also to identify new words and combinations of AIs to support the description process. The main idea is to create a general framework to support how AIs should be combined to reinforce each other. The project will be divided into three parts as follows. First part: model identification and matching. During the first part, the student will focus on studying the state of the art regarding different models for generative AI. Within this model, the student will try to match which ones should support each other. In a similar fashion, the student will identify the parameters of the models and study how these parameters affect the output's quality. With this information, the student will be able to create a search algorithm that co-evolves and involves both models. Second part: formal auditing of the generative model. This part focuses on creating or identifying different metrics to measure the effect of the optimization process, define boundaries during the optimization, and create a new set of strategies that will support identifying other kinds of problems within the systems, for instance, bias or fairness issues. Third part: improving and explaining the models. The last part focuses on how the models can be directly improved and not only turned based on the outputs of the other models. The end goal of this strategy is to create better AI systems with a focus on adversarial machine learning combined with explainability.

References

- H. Berger, A. Dakhama, Z. Ding, K. Even-Mendoza, D. Kelly, H. D. Menendez, R. Moussa, and F. Sarro, StableYolo: Optimizing Image Generation for Large Language Models. In Symposium on Search-Based Software Engineering (SSBSE) 2023 Springer.
- X. Hou, Y. Zhao, Y. Liu, Z. Yang, K. Wang, L. Li, X. Luo, D. Lo, J. Grundy, H. Wang. Large language models for software engineering: A systematic literature review. arXiv preprint arXiv:2308.10620. 2023 Aug 21
- J. M. Zhang, M. Harman, L. Ma and Y. Liu, "Machine Learning Testing: Survey, Landscapes and Horizons", in IEEE Transactions on Software Engineering, vol. 48, no. 1, pp. 1-36, 1 Jan. 2022, doi: 10.1109/TSE.2019.2962027.
- Jo, A. (2023). The promise and peril of generative AI. Nature, 614(1), 214-216.
- Saharia, C., Chan, W., Saxena, S., Li, L., Whang, J., Denton, E. L., ... & Norouzi, M. (2022). Photorealistic text-to-image diffusion models with deep language understanding. Advances in Neural Information Processing Systems, 35, 36479-36494.

SLAM in Wearable Robots

Supervisor: Letizia Gionfrida

Areas: Computer vision, Robotics

(Back to Scholarship Not Allocated)

Project Description

The utilization of wearable robotic technology presents a significant potential for revolutionizing the support provided to individuals with mobility impairments (1). Departing from conventional systems reliant on manual control and mode switching, contemporary approaches incorporate biological signals within human-in-the-loop assistive frameworks. However, the state-of-the-art human-in-the-loop assistance, typically reliant on sEMG signals known for their noise and limited data output, necessitates advancement. Ultrasound has proven effective for human-in-the-loop assistance in wearable robots (2,3). However, the ability to evaluate the wearable effectiveness relies on brightness (b-mode) data collected in costly lab settings. This research aims to propel the development of a vision-centric ankle exosuit tailored for navigating unstructured environments (4). To achieve this, the intention is to replace costly Motion Capture (Mocap) systems with an economical alternative utilizing AprilTags alongside advanced visual localization algorithms. This approach not only significantly reduces expenses, rendering sophisticated localization technology accessible to a broader spectrum of medical facilities, but also presents a novel amalgamation of technologies. The strategy involves outfitting an ultrasound probe with a compact, coin-sized computer board and a standard RGB camera (5). This configuration enables a real-time algorithm capable of self-localization through the recognition of strategically placed AprilTags, offering unparalleled operational flexibility and ease. Evaluation of this setup aims to attain millimetre-level localization accuracy, a substantial advancement in precision for various medical procedures beyond conventional biomechanics laboratories. The validation of this claim will be supported by comparative data showcasing the system's accuracy in contrast to existing technologies. Furthermore, recognizing the significance of addressing potential challenges like environmental interferences or calibration requirements, the proposed methodology prioritizes the design and implementation of auto-calibration algorithms (1). These algorithms aim to mitigate the impact of changing environments, ensuring the robustness of the localization algorithms at a product level. The subsequent phase aims to innovate ultrasound imaging by developing a methodology for converting 2D ultrasound images into comprehensive 3D models. This integration involves utilizing specialized localization technology with ultrasound probes to precisely track their position and orientation. Consequently, this spatial context facilitates the alignment of 2D ultrasound slices into a cohesive 3D space, enhancing clarity, reducing noise, and generating anatomically accurate models. The crux of this proposal centres on the advancement of an intricate reconstruction algorithm. This algorithm is designed to seamlessly integrate aligned 2D B-mode ultrasound slices into a continuous 3D volume, managing variations in thickness and spacing for anatomical accuracy. Subsequent refinements involve enhancing details, eliminating artefacts, and employing visualization techniques to make the models readily interpretable for medical professionals. Ultimately, the goal is to seamlessly integrate these 3D models into existing clinical workflows, ensuring compatibility with contemporary medical imaging platforms. This integration not only enhances ultrasound imaging's diagnostic capacities but also introduces new avenues for modelling muscle behaviours for wearable robotic applications. By bridging the gap between 2D and 3D imaging, this proposal aspires to establish a new standard, offering a more detailed and accurate perspective of B-mode ultrasound scans to model tailored assistance.

References

Jia, Rui, et al. "A computer-aided tracking and motion analysis with ultrasound (CAT & MAUS) system for the description of hip joint kinematics." International journal of computer assisted radiology and surgery 11 (2016): 1965-1977.
Nuckols, Richard W., et al. "Individualization of exosuit assistance based on measured muscle dynamics during versatile walking." Science robotics 6.60 (2021): eabj1362.
Gionfrida, Letizia, et al. "Age-Related Reliability of B-Mode Analysis for Tailored Exosuit Assistance." Sensors 23.3 (2023): 1670.
Hallock, Laura A., Akira Kato, and Ruzena Bajcsy. "Empirical quantification and modeling of muscle deformation: Toward ultrasound-driven assistive device control." 2018 IEEE International Conference on Robotics and Automation (ICRA). IEEE, 2018.
Koolwal, Aditya B., et al. "A fast SLAM approach to freehand 3-D ultrasound reconstruction for catheter ablation guidance in the left atrium." Ultrasound in medicine & biology 37.12 (2011): 2037-2054.

Improving active learning strategies for limited annotation budgets

Supervisor: Luis Carlos Garcia-Peraza

Areas: Artificial Intelligence (AI), Machine Learning (ML), Computer vision

(Back to Scholarship Not Allocated)

Project Description

In machine learning, determining the subset of data points (e.g. images, videos) for annotation emerges as a critical decision-making process. The selected data points carry the responsibility of providing a representative snapshot of the diverse scenarios anticipated during real-world testing. Despite the multitude of proposed strategies for data point selection, an enduring observation persists, suggesting that random selection, especially in low-budget scenarios, often proves to be an optimal approach. Active learning problem Figure 1. The active learning problem. The overarching objective of this project is to propel active learning strategies tailored specifically for situations characterized by highly limited annotation budgets. This pursuit is particularly relevant in fields with stringent budget constraints, such as medicine.

References

[1] Mahmood et al. Low-Budget Active Learning via Wasserstein Distance: An Integer Programming Approach, ICLR, 2022.
[2] Chen et al. Making Your First Choice: To Address Cold Start Problem in Medical Active Learning, MIDL, 2023.