**COMP[1682.1] Project Proposal**

**["Sorcerer Fight: A human-computer interaction entertainment application using computer vision"]**

**[Truong Duong Hong Phuoc]**

**BSc H COMPUTING (FINAL YEAR ENTRY) (FPT, CAN THO)**

**001391502**

1. **Overview**

Write 100 - 300 words describing as succinctly as possible the outline and rationale of your intended project using third-person grammar (no ‘I’, no ‘my’).

Including:

* Context and context

In modern times, computer vision is utilised in various fields, such as security, payment systems, entertainment, etc. This research focuses on applying computer vision in the field of hand detection and illustrating its application in game development products. The problem enhances algorithms to detect hands with significant accuracy and, at the same time, maintain the process in real time while not affecting performance. Develop a robust machine that adapts to real-time processing while preserving performance and high accuracy. Further, the game illustrates that it will also slant the wrong sides of war abstractly by providing people's conditions during wartime. The game will not be exposed clearly, but it will be abstract enough to make people feel about it.

In simple terms, the project uses computer vision to track hands in games without causing any delays. And the game itself will touch on the hardships of war without showing it directly.

* Problems in the above context

challenging that might arise in achieving accurate hand detection in real-time while barely sacrificing the performance of the game. Moreover, covering the complexity and worst side of war within a game setting requires a lot of work in design, story, and context to assassinate the emotional impact of the war and crimes that games relate to in real life.

* Solution

A possible solution for computer vision involves algorithms that optimise hand detection accuracy while maintaining real-time processing capability. Furthermore, the game development will focus on the subtle narrative elements and gameplay that cover the thematic exploration of wartime experience with abstraction.

* Benefits of this solution

by implementing computer vision techniques combined in the game aspect, it provides a significant and unique experience, providing interactive gameplay that makes the user feel fit and newer in experience rather than traditional games, further incorporating thought-provoking themes and fostering deeper player engagement in the emotional field.

War is might not present in many countries but not completely disappear in the modern world for example according to (middleeasteye, n.d.) War is now Committed in Gaza, Israel. This war affects people's lives, and commits crimes, Violent. Media is the greatest way to tell people about war, what it does, and how cursed it is. To propagate and remind people how bad war was, a game is great, and culture propagates for the job. The purpose of the project is not only to solve academic problems in implementing technology in the context of various categories, such as game development but also to condemn war commitment for any non-reasonable reason. The project will show most of the worst aspects of war. The project remains not lively but abstract, not to violate ethical principles.

Keywords: [“Computer vision”, “Game development”, “Design pattern”]

1. **Aim**

This project aims for a better gaming experience and creativity by integrating deep learning combined with Computer vision, giving a fresh view and proving the potential and relation between the Gaming field and trendy tech.

1. **Scope**

The project involves computer vision aspect and game development. The computer vision will be used to detect hands and recognize hand gestures for only 1 hand, which will be adopted as an alternative to the controller. For the Game aspect, the gameplay will remain at least 30 minutes of fully developed gameplay, including UX/UI, Game story, etc.

1. **Related works / Literature review**

**<introduce to computer vision>**

In the project, computer vision plays a crucial role in game mechanics; the project will use it as input, partly replacing the controller. Giving fresh and advanced experience for the player. The game's concept maintains the user experience while giving an accurate image of the bad effects of war. Due to the importance of computer vision, it is crucial to use and implement. According to (Chen et al., 2024) computer vision covers many features related to images, such as gathering images by using cameras, sensors or similar technology for example, in the project, it is required to open the camera to gather information to detect and use it as input for control, furthermore, CV including manipulate the images to a suite for future process, for example, it is easier to remove colour to identify light of the objects, it also covers feature extraction which means identity and isolating characteristic and patterns within images. It also covers image recognition and understanding images, which utilise extracted features and patterns to identify or classify the wanted object. Interpreting and comprehending the content or context conveyed by images through analysis and inference.

**<Role of Deep Learning in Computer Vision>**

to use computer vision as a control for the game, it requires a machine to match the recognition to a kind of input type, which is suited for deep learning due to the need for a machine that remembers objects and images to give out the suitable input for the control. According to (Robert, 2024), the reason deep learning matches computer vision is End to end learning, feature learning and representation, handling variability and complexity, scaleability and big data, transfer learning and fine-tuning, continuous learning and adaptability. End-to-end learning, which means that deep learning is able to learn from raw pixels, is capable of learning through abstract representation layers, which is convenient due to the tradition requiring hand-crafted features. Deep learning models are better at learning distinguishing points and capturing patterns of visual data. The original deep learning architectures allow it to learn abstractly while maintaining the meaning, which enables the models to adapt from low-detail to high-detail pictures, which means it is able to work with complex visuals. Furthermore, deep learning models are able to generate new data, which makes them strong due to their capability to learn large amounts of data in different kinds for example, they are able to learn with light conditions, object appearance, viewpoints, occlusion, and more. Another advantage of deep learning is that it is capable of processing big data. The availability of large-scale labelled datasets, Transfer learning, helps deep learning models improve by using pre-trained models on big datasets to learn new tasks with fewer data. This makes learning faster and more accurate, especially when getting lots of labelled data is complex or expensive. Deep learning models can keep learning and adapting to new data, staying updated and performing well over time. This adaptability is essential for changing environments and tasks. In summarize, due to abstract learning and ability to adapt large amounts of learning data deep learning models are suitable to implement with computer vision (Robert, 2024)

**<why not ML, Compare ML and DL in implementing the Computer vision>**

**Although** (Robert, 2024)said that deep learning is great to implement with computer vision, the performance between other methods of deep learning comparison is still the mists; according to (Nafea et al., 2024) research comparison with machine learning and deep learning, they both provide advantages and disadvantages. On the first hand, machine learning provides popular algorithms such as support vector machines, random forests, and naïve Bayes, and each of them gains great performance in specific tasks. In general, ML provides good performance in the ability to manage complex decisions and dimension features. Well in established algorithms with reasonable results, great effective in labelled training data for train model. Otherwise, the limitation it gives is that it requires handcraft, which means, in some cases, needing a professional in the field. Furthermore, it is sensitive to outliers and noise in data sets. Labelling the data takes much effort to archive. Lastly, ML has difficulty adapting to large scale data. In summary, machine learning requires a handy process and is time-consuming while providing a great output in a limited amount of data. On the other hand, deep learning usually combines with computer vision. It provides popular algorithms such as CNNs, which adapt great performance in classification and detect objections due to the features' hierarchies. RNNs provide significant work with sequential data, but it is also challenging to process long sequences. Lastly, variants of CNNs and RNNs like ResNet, inception,… might involve increasing the required amount of training data.

**<Rought in implementing deep learning and computer vision>**

According to (sciencedirect, n.d.) Implementing deep learning with computer vision faces many problems, including a lack of clear understanding of performance between models, training with a limited dataset, hard archives, real-time processing, and the need for more powerful models. This research focuses on implementing DL and CV in a Real-time Processing Application demonstrated as a Game.

The integration of deep learning (DL) Combined with Computer vision (cv) Provides a Promising value for any application, including the Field of Game development, which requires Real-time processing, this integration has many challenges, according to (sciencedirect, n.d.) important problem Might face during develop the Integration are there is no clear Performance Compare between architecture, the Limited of the dataset, Hard to archive real-time processing application and maintain high accurate. In response, this research focuses on implementing DL and CV with a thin real-time processing context Demonstrated as a Game. With this approach, Hoping for a better gaming experience.

By integration of DL and CV, there is present, of Numerous challenges study reference (sciencedirect, n.d.) problem lies in the absence of clear performance of architectures and models; without comprehensive insights, there is a struggle into select a suitable approach for a gaming application. Furthermore, the robust training for DL and CV models requires a large amount of data, so acquiring a dataset tailored to the gaming context can be challenging. So far, the requirement for real-time processing and maintaining accuracy can be strictly due to sophisticated optimisation techniques and strong processing hardware.

To solve the problem of the research, the possible solution is, firstly, to address the lack of clarity regarding the performance of architectures and models; extensive comparative studies and benchmarking exercises should be conducted. This involves evaluating the effectiveness of different DL and CV approaches to identify the most suitable techniques. Secondly, addressing the need for a Robust training dataset, consider generating a Dataset based on the concept of the game and the special requirements of the game context. Additionally, considering using a model that works well with a small amount of data might give some advantages for a limited dataset. Thirdly, achieving real-time processing and maintaining accuracy involves implementing efficient algorithms, parallel Processing Techniques, and hardware acceleration solutions to minimise processing latency.

[Deep learning in computer vision: A critical review of emerging techniques and application scenarios - ScienceDirect](https://www.sciencedirect.com/science/article/pii/S2666827021000670): Problem [1]

**<Input Hand for the project hand gestrure>**

it is required in the project to detect hand gestures to use as input party to replace the control. According to (Yadav et al., 2024), it is commonly hand detected mostly review in an environment that is under control. In fact, the environment is mostly unable to control in real applications, which is challenging to deal with the complexity of variations, for example, hand detection in different light conditions, hand rotations, and scales,… those factors that weren’t able to be controlled directly affect the performance of the computer vision including motion blur and dominance background. To solve the problem that the authors have stated, they introduce deep learning model hand localisation modes. They have used a customised object detector, which is combined with tracking modules for efficiency. Furthermore, the scaled-based database evaluates models on different scales. By using many layers to separate the human background, human body, and hand to detect. Furthermore, by giving points for each box on the screen, for example, 26 columns and 26 rows, each cell will have 0 points for the background, 1 for the body and 2 for the bare hand. The point is given by their custom hand detector.

**<Hand gesture recognisation>**

According to (Qi et al., 2024, page number), the main process of the hand gesture recognition process is data acquisition, which is used by a device such as a camera to gather information. Gesture detection and segmentation detect the hand and its position, then segment the hand region. Gesture recognition, which means extracting the image's features to recognise the gesture type. In the details of the process of hand gesture recognition, data acquisition is first, which means processing is used by cameras or anything able to gather the information that is needed for future processes and analyses. Gesture detection and segmentation: This is the phase in which the computer uses the gathered information, such as images, to determine the position of the hand and how the hand shapes. Feature extraction and classification: features are extracted from the data, which could be images. The data is extracted from the hand region, and gestures are classified based on the feature needed. Further, the author also provides additional steps, but in general, this is required for the gesture recognition process.

**< Real-time processing>**

According to (Qi et al., 2024), To archive real-time processing, there is a framework which integrates many components for real-time visual control. In this context, the author(s) use the Yolo V8 model, which is well-known for object detection, identifying and classifying objects. The author uses the model to detect objects in real-time by a process in each frame of the gathered data, which is video. Further, the author also uses the YoloV8 to track objects in real time. Due to the performance of the model, it is possible to process the data in real time after training the amount of the dataset.

Otherwise, according to (Lu et al., 2024), authors achieve real-time processing by implementing end-to-end architecture combined with 3d DenseNet and LSTM for real-time detection and recognition of hand gestures. Basically, the process of the system uses end-to-end deep neuron network architecture to process in real-time while detecting and recognise the object and then using the 3DdenseNet to extract feature of the current frame features and LSTM for extracting global features, which means extraction, in general, will consider the previous frame as well.

**<game development process>**

The project is required for the professional development process to follow and conduct the final product. According to (Guðnason et al., n.d.), the process of developing a game includes Market analysis, pre-production, designing the game, developing the game, beta testing, feedback base integration design, marketing, and publishing the game. The project will adapt and follow but in fewer steps, which include pre-production, designing the game, developing the game, testing, publish the game.

**<game pattern>**

Patterns are common in game development; as a matter of fact, patterns are used in many systems and units of the game. For example, there might be only one character available to implement. It is suitable for singleton due to the properties of the pattern. According to …. Applying design patterns increases the performance of the game and helps in the structure of the game. The patterns that the author uses for the game are state patterns and Finite State Machines. The final result gives out quite a huge gap between performance with the game that has a pattern and the game that does not use it.

**<Game affects human behaviour>**

According to …. Author said that games are able to affect human behaviour due to identifying with the visual character, increasing memorising, start to learn something easily, which leads to effect thinking then will affect behaviour. It could be good or bad, but as a matter of fact, it really affects human behaviour on a larger scale. It could be a negative effect or a positive effect on society. Through this, the project adapts to the human perception of war and crime. Trying to have a positive impact on people by raising awareness of the right things.

1. **Objectives**

4.1 Research Report Document

4.1.1 A Document that Shows all of the related research, including Various papers that Relate to the current Research

4.1.2 Find 15 Relate paper to the relate-work of current research [0.5]

4.1.3 Reading papers that found in 4.1.2 [2.0]

4.1.4 Writing Proposal [4.0]

4.1.4 Writing Research Report Document [1.5]

4.2 Game Design Document

4.2.1 A document that Tells all about the Product, including Game genes, Mechanics, Technologies, Engine,… [3.5]

4.2.2 Design and Conceptualization [0.5]

4.2.3 Artistic and Technical Specifications [1.5]

4.2.4 final document of game design including all detail in game [3.0]

4.3 Implementation

4.3.1 the product will be complete after this phase and Ready for Processing

4.3.2 implement the Back-end including developing Core game mechanics and basic gameplay [1.5]

4.3.3 Implement the GUI, Animation, cut screen, Light, etc. [2.0]

4.4 Testing game product

4.4.1 Identify test object [1]

4.4.2 Performing the test plan and recorg in test Log [3.0]

4.4.4 Fixing the bug that found in 4.4.2 [5.0]

4.4.5 Repeat the testing game if needed []

4.5 Evaluation and Reflection Report

4.5.1 evaluate the final product [4.0]

4.5.2 Evaluate the process of developing the product[2]

4.5.3 Evaluate Feedback [1.5]

4.5.4 Evaluate report [2.0]

4.5.4 Writing the Evluation report [5]

* 1. Project Framework or Any Methodology Used

The framework adopted in the project is agile; by applying the agile framework and method, agile is well-known for adaptability and flexibility. Which suits the requirements of the project? An agile framework can provide flexible plans in case the project has trouble. Agile also maintains great quality for the final product by fully implementing the functions before moving. By aiming the core function, it gives great performance within a limit of time. Personally, I use Unity to implement the game for coding. Unity has great advantages in simplicity, ease of conduct, and long-term development. Unity is a great platform for implementing the product.

1. **Legal, Social, Ethical and Professional**

**Introduction:**

The main aim of this research is to solve the problem of applying computer vision in the program to demonstrate it as a game-developed product, proving the potential of game development compatible with trendy technology. Furthermore, the project also condemns acts of war in an abstract way. It is critical to consider many aspects of ethics and society.

* 1. **Legal**

intellectual property rights: the project might involve property rights; there are plenty of free models, textures, and systems out there. The project will commit that it will not take any outsourced features without a license and agreement. The project also maintains a large amount of in-house sources.

Data privacy:

This research will not involve collecting personal user data information or any kind of sensitive data. For example, when the user uses the camera, the product recognises any video, audio, speech,.. or kind of data that the user provides. The project might require turning on the camera, but It wouldn’t be able to record.

* 1. **Social**

User accessibility:

* The project might be difficult to gather input from users who have hand(s) disabilities. Due to a lack of training data, it would be difficult to ensure that the system works smoothly with the users. Furthermore, the project only adapts to the window operation system.

Global Impact

* the final product will give the users a point of view of War, showing the worst case in order to protest against war. The project gives end-users a look at the war by not mentioning it directly but only showing its impact.

Digital divide:

The game is tailored for PC, leveraging its superior performance for an optimal gaming experience, unlike traditional console games, which may not suit hand movements and cameras.

* 1. **Ethical**

Ethical standard:

The project will adhere strictly to ethical standards throughout its development and implementation. Given the sensitive nature of the thematic elements surrounding war, the project will prioritise ethical considerations in several standards, such as sensitive content and culture appropriate.

User privacy and consent:

The user's privacy and consent will be protected when it is involved in our game. The game requires a camera to turn on to detect the hand. It raises concerns about the privacy of end-users. For privacy, the project will have a policy to protect and ensure user privacy. Users will gain full control of the camera. They are able to turn it on or off at their will at any time.

Bias and Discrimination:

In this game, the project will not show any favouritism or discrimination against any individual or organisation. Every behaviour, race. It is not based on fact. Stories are only created to fit the world and circumstances.

* 1. **Professional**

**Quality of the final product:**

To ensure the quality of the product, the project will adapt to the quality assurance of the (1). The QA is adapted in many aspects of the product. Basically, it is divided into four big relevant parts: unit, integration, functional, and performance. This will be adapted in the project, which will help to ensure the final quality.

**Quality for the resource:**

To ensure the quality of the resources, the project will follow the rules which improve the resource quality provided in (2) and (3), which is to apply patterns and make the resource clear and easy to read.

**professionalism**:

Ensuring the developer of the project will follow the standard of IGDA standard, which is standard for developing a game.

[Game Development Testing and QA Best Practices | Unity](https://unity.com/how-to/testing-and-quality-assurance-tips-unity-projects#how-qa-and-test-game) (1)

[Clean up your code: How to create your own C# code style | Unity Blog](https://blog.unity.com/engine-platform/clean-up-your-code-how-to-create-your-own-c-code-style)(2)

[Level up your code with game programming patterns | Unity Blog](https://blog.unity.com/games/level-up-your-code-with-game-programming-patterns)(3)

[Game Industry Standards – IGDA](https://igda.org/resourcelibrary/game-industry-standards/)(4)

1. **Planning (see Appendix A)**

*You will describe two significant contents:*

*- What model will you use to make a schedule (for ex, Scrum, Scaled, Crystal, Kanban.... you should look at the development process in Enterprise Web Software Development or Software Lifecycle).*

*- What do you plan to do for each phase of your project?*

*And show your schedule created on project management software (for ex, MS Project).*

*You should put the Gantt chart (based on your schedule) in Appendix A.*

1. **Initial References**