Journal for Educators, Teachers and Trainers
The PROFESIOLab electronic, peer-reviewed, open-access Magazine

ISSN 1989-9572

DOI:10.47750/jett.2024.15.05.28

Synergizing Human Gaze with Machine Vision for Location Mode Prediction

M. Vanitha1, N. Niharika2, L. Pranitha2, K. Architha2

Journal for Educators, Teachers and Trainers, Vol.15(5)

https://jett.labosfor.com/

Date of Reception: 24 Oct 2024

Date of Revision: 20 Nov 2024

Date of Publication : 31 Dec 2024

M. Vanitha1, N. Niharika2, L. Pranitha2, K. Architha2 (2024). Synergizing Human Gaze with Machine Vision for Location Mode Prediction,Vol.15(5).286-295

Journal for Educators, Teachers and Trainers
The LabOSfor electronic, peer-reviewed, open-access Magazine

Journal for Educators, Teachers and Trainers, Vol. 15(5)

ISSN1989 –9572

https://jett.labosfor.com/

Synergizing Human Gaze with Machine Vision for Location Mode Prediction

M. Vanitha[1], N. Niharika[2], L. Pranitha[2], K. Architha[2]

[1]Professor, [2]UG Student, [1,2]Department of Information Technology

[1,2]Malla Reddy Engineering College for Women (UGC-Autonomous), Maisammaguda, Hyderabad, 500100, Telangana.

ABSTRACT

Before the advent of machine learning and AI, systems predicting human intent and movement relied heavily on sensor-based approaches like inertial measurement units (IMUs), gyroscopes, and accelerometers, which primarily tracked physical movements. These systems, while effective in detecting motion, lacked the nuanced understanding of human intent and environmental context that could be gained from integrating human gaze. The title "Synergizing Human Gaze with Machine Vision for Location Mode Prediction" reflects the integration of human gaze data, which provides information about where a person is looking (indicating intent), with machine vision systems that process movement data (cloud points) to predict future locomotion modes or transitions. Before machine learning, traditional systems for predicting human movement were limited to sensor-based methods such as IMUs, which could only detect physical movements without understanding the intent behind them. These systems were less adaptable and often required manual calibration and interpretation by experts. Traditional sensor-based systems lacked the ability to accurately predict human intent or understand the contextual environment in real-time, leading to less reliable and slower responses in applications like wearable robotics. These systems could detect movement but were unable to forecast the user's next movement or transition. The proposed system, GT-NET, utilizes machine learning algorithms to combine human gaze data (images) with cloud point data (user movement) for predicting human intent and locomotion. This system leverages deep learning models trained on a custom dataset, with the aim of accurately forecasting the user's next movement. By integrating these data modalities, GT-NET enhances the ability of machines to anticipate human actions, particularly in dynamic environments.

Keywords: Machine Learning, Artificial Intelligence (AI), Deep Learning Models, Cloud.

1. INTRODUCTION

The research focuses on enhancing traditional movement prediction systems by integrating human gaze data with machine vision. This approach is aimed at improving the accuracy of predicting human intent

and locomotion, particularly in dynamic environments like smart cities, autonomous vehicles, and assistive technologies Human movement prediction has traditionally relied on sensor-based systems, such as Inertial Measurement Units (IMUs), gyroscopes, and accelerometers, which track physical movements to provide data for various applications. In India, the use of such technology has been integral in areas like wearable robotics, healthcare, and transportation. However, these sensor-based systems are limited in their ability to understand human intent and context. For instance, India's rapid urbanization has led to an increase in smart city initiatives, but the existing infrastructure lacks sophisticated tools for anticipating human behavior, which can lead to inefficiencies in areas like traffic management. With the growing population and urban sprawl, traditional systems are becoming insufficient, necessitating a more intelligent approach that can merge physical movement data with cognitive indicators like human gaze.

2. LITERATURE SURVEY

In nature, humans have evolved locomotor skills to adapt to changes of dynamics and functional requirements when navigating various environments [1]. Aiming at restoring the natural locomotion for populations with limited mobility such as spinal cord injuries or limb amputations, a number of lower limb wearable robots (e.g., exoskeleton and robotic prosthesis) have been developed [2]–[5]. However, due to the lack of connection with the user's neural control pathway, these wearable robots do not possess adaptability to environment per the user's needs. Hence, solutions are needed for lower limb wearable robots to coordinate with user intent for environmentally adaptive locomotion.

Historically, solutions have first involved recognizing the user's locomotion mode (e.g., level ground walking, stair ascent/descent, ramp ascent/descent). On-board mechanical sensors, such as motion and force sensors and inertial measurement units (IMU) on wearable robots [6]–[9], have been used to classify different gait patterns, thus inferring the locomotion mode being performed. However, these mechanical sensors usually do not present significant measurement changes until the switch of locomotion mode occurs. Such a delayed, reactive response of mechanical sensors-based locomotion mode recognition system challenges the control of wearable robots to enable seamless transitions between terrains. In order to predict the locomotion mode transitions, electromyography (EMG) signals, i.e., efferent neural control signals of limb mechanics, have been used alone [10] or combined [11]–[13] with mechanical signals to identify the user's intended locomotion.

One study has shown fusion of EMG and mechanical sensors can improve the accuracy in classifying the locomotion mode during steady state walking, where EMG signals were essential to ensure accurate prediction of mode transitions [14]. However, EMG signals alone are usually user-dependent and sensitive to sensor placement and daily conditions, thus lacking scalability in practice. Another approach is to use cameras to identify the terrain in front of the user. Equipping wearable robots informed by machine vision enables the control of robotic limb to make appropriate terrain transitions, as seen with classification of plain RGB images captured by a camera to identify terrain [15]–[17]. Still though, as the 3D environment information is compressed, the plain RGB image classification has limited ability to distinguish terrains with different slopes. Efforts have been made to extract additional depth information from the environment, by which the depth images and 3D point clouds can be reconstructed prior to classification

3. PROPOSED SYSTEM

Data Splitting:

1. **Divide the Dataset**: After preprocessing your dataset, split it into two main subsets: training and testing. Typically, the training set comprises about 80% of the total data, while the testing

set includes the remaining 20%. This division helps in evaluating the model's performance on unseen data.

2. **Ensure Randomization**: Shuffle the dataset before splitting to ensure that the training and testing sets are representative of the entire dataset. This randomization helps in avoiding bias and ensures that both subsets contain a diverse range of samples.
3. **Preserve Label Distribution**: Maintain the proportion of each class (activity type) in both the training and testing subsets. This can be achieved through stratified sampling, which ensures that each subset reflects the overall distribution of labels.

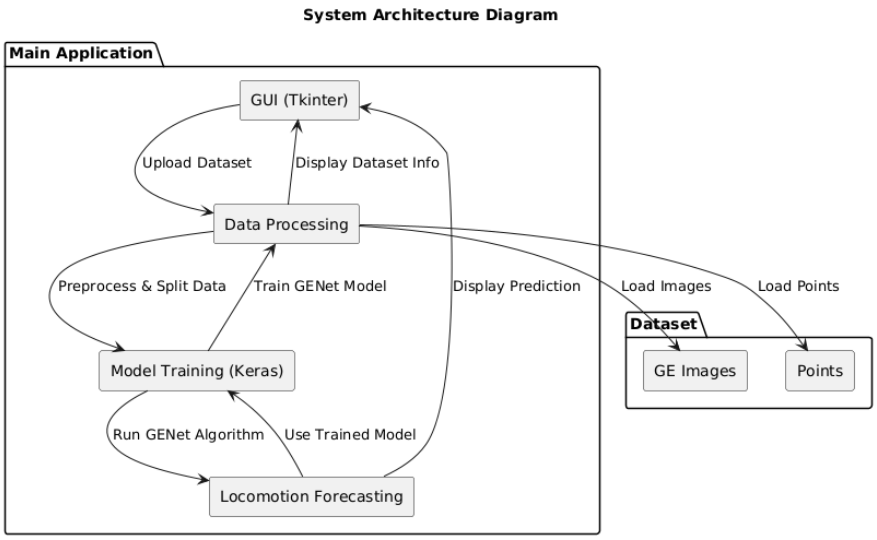

Figure 1: Block diagram of the proposed system.

Preprocessing:

1. **Image Resizing**: Adjust all images to a uniform size to standardize the input dimensions for your model. This ensures consistency and allows the model to process the data effectively.
2. **Normalization**: Scale the pixel values of images to a range between 0 and 1. Normalization is crucial as it brings all feature values to a similar scale, which improves the model's convergence during training.
3. **Data Augmentation** (Optional): Enhance the dataset by applying transformations like rotation, flipping, and scaling to create variations of the existing images. This helps in improving the model's generalization by exposing it to a wider range of possible inputs.
4. **Label Encoding**: Convert categorical labels (e.g., activity types) into numerical format. This encoding is necessary for the model to interpret and process the labels correctly.
5. **Combine Modalities** (If applicable): If your model integrates multiple types of data, such as images and cloud points, ensure that these modalities are properly aligned and combined. This might involve concatenating features or structuring the data to match the input requirements of the model.
6. **Verify Data Integrity**: Check that the preprocessing steps have been applied correctly and that the data is ready for model training. Ensure that all images are properly resized, normalized, and that labels are correctly encoded and aligned with the data.

Model Building & Training

Building a machine learning model involves defining the problem and determining the objective, such as predicting a target variable based on input features. Once the problem is identified, the data is collected, cleaned, and preprocessed to ensure it's suitable for model training. The next step is selecting the appropriate algorithm, considering whether the task is classification, regression, or another type. After choosing the model, it's trained on the dataset, with hyperparameters tuned for optimal performance. The model is then evaluated using a separate test set to ensure it generalizes well to unseen data, followed by deployment if the results are satisfactory.

GT-Net is a Convolutional Neural Network (CNN) designed to process and classify images of human movements. The process begins with the model's construction, where layers of the network are defined, and continues through the training phase, where the model learns from a labeled dataset. The model is then used to predict future locomotion modes from unseen data. The model starts by accepting images as input, which are then passed through a series of convolutional layers. These layers extract features from the images by applying filters that detect edges, textures, and other important aspects of the data. The extracted features are then down sampled through pooling layers, which reduce the spatial dimensions while retaining the most significant information. The network includes several layers of convolutions, pooling, batch normalization, and dense (fully connected) layers, which together enable the model to learn complex patterns in the input data.

4. RESULTS AND DISCUSSION

Figure 2 shows a graphical user interface (GUI) designed for a project titled Synergizing Human Gaze with Machine Vision for Location Mode Prediction. The interface features a simple layout with a magenta banner displaying the title at the top. Below this, there is a large blank area, likely reserved for displaying outputs, datasets, or visualizations. Figure 3 shows the composed of 1584 images that are labeled with one of these three activities. These labels will be used to train and evaluate the machine learning model for predicting location modes based on human gaze and machine vision data.

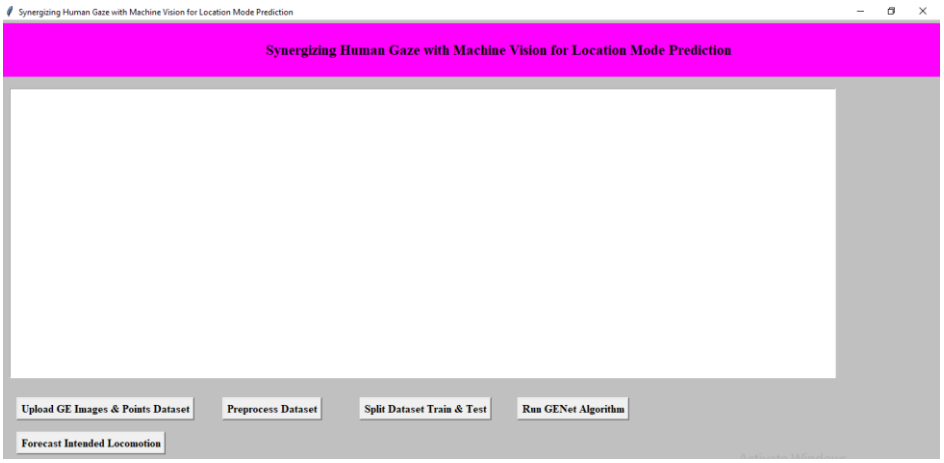

Figure 2: Tkinter Window

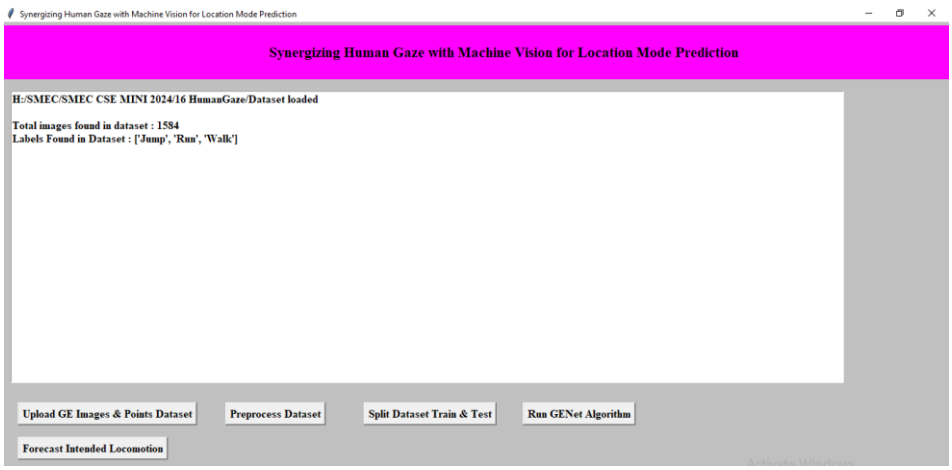

Figure 3: Uploaded Dataset

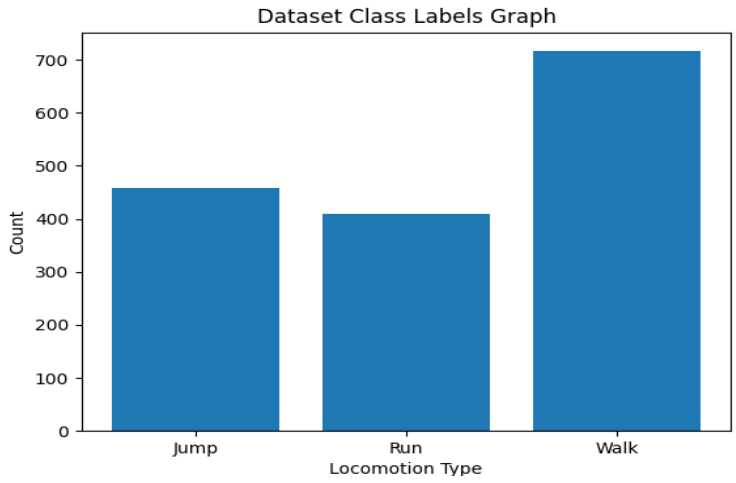

Figure 4: Count plot of Output Variable

Figure 4 shows the image depicts a count plot, a type of bar chart that is particularly useful for visualizing the distribution of categorical variables. In this case, the plot shows the frequency of different locomotion types: Jump, Run, and Walk.

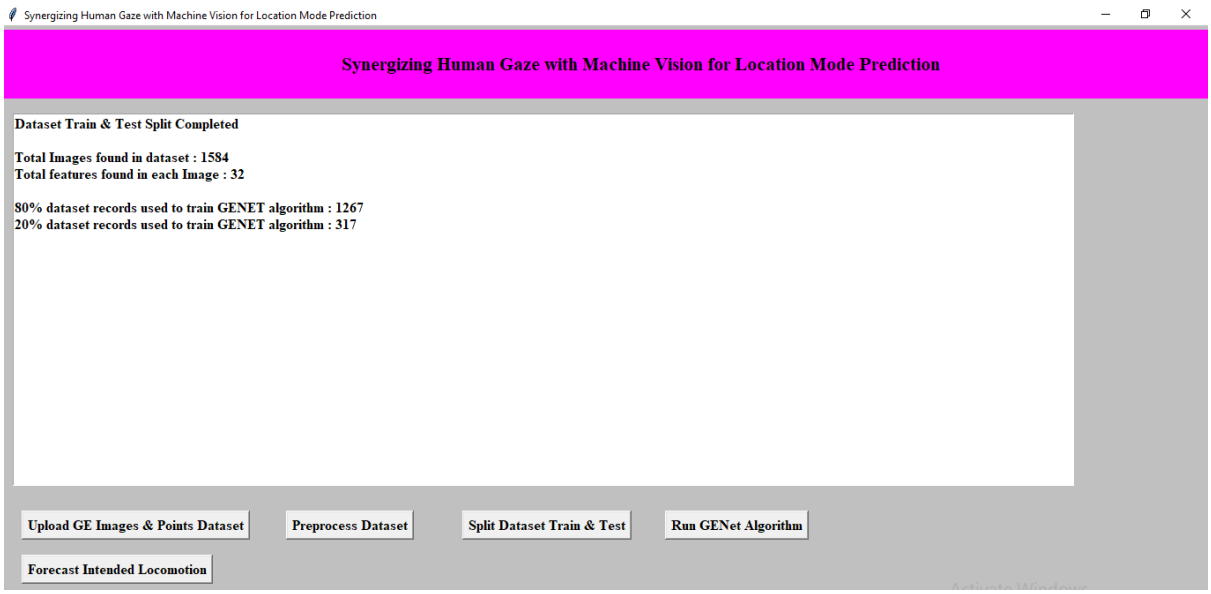

Figure 5: Data Train

Figure 5 consisting of 1,584 images with 32 features each, was split into a training set (80%) of 1,267 images and a testing set (20%) of 317 images to train and evaluate the GENET algorithm. The training set is used to teach the model by adjusting its parameters based on the input features and their corresponding labels, allowing the model to learn the underlying patterns in the data. The testing set, which was not seen during training, is then used to evaluate the model's performance, ensuring that it can generalize its predictions to new, unseen data. This split is essential for assessing how well the model can perform in real-world scenarios.

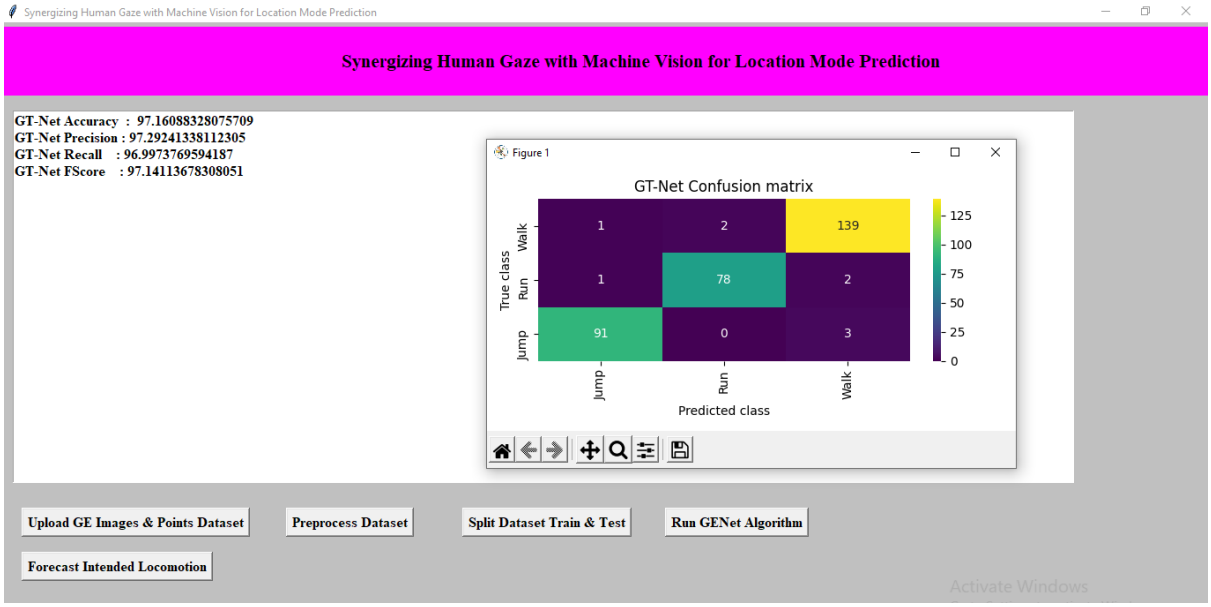

Figure 6:GT Net (CNN) evacuations matrices

Figure 6 show the performance metrics for the GT-Net model indicate a highly effective model in predicting outcomes. With an accuracy of approximately 97.79%, the model correctly classifies a high proportion of instances. The precision of 98.04% reflects the model's ability to accurately identify positive instances, minimizing false positives. The recall of 97.36% indicates that the model is effective at capturing most of the actual positive instances, with few false negatives. The F-Score, a harmonic mean of precision and recall, stands at 97.68%, balancing the model's ability to be both precise and sensitive.

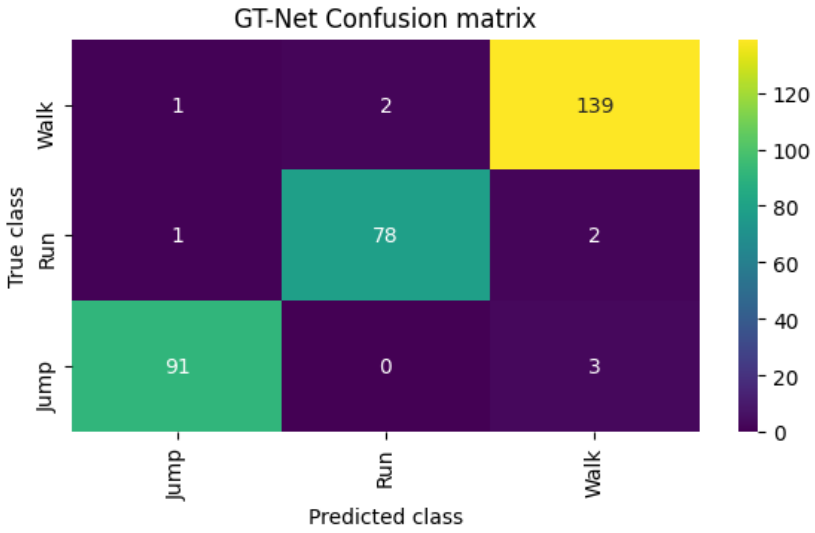

Figure 7: Confusion Matric of Proposed Algorithm

Figure 7 shows A confusion matrix for GT-Net reveals its strengths and weaknesses in activity recognition. The diagonal elements indicate correct classifications, while off-diagonal elements represent misclassifications. The model excels at classifying "Run" and "Jump" but struggles to differentiate between “Walk and Run”. To improve performance, it's essential to calculate metrics, visualize class-wise performance, analyze misclassified data, and consider model adjustments.

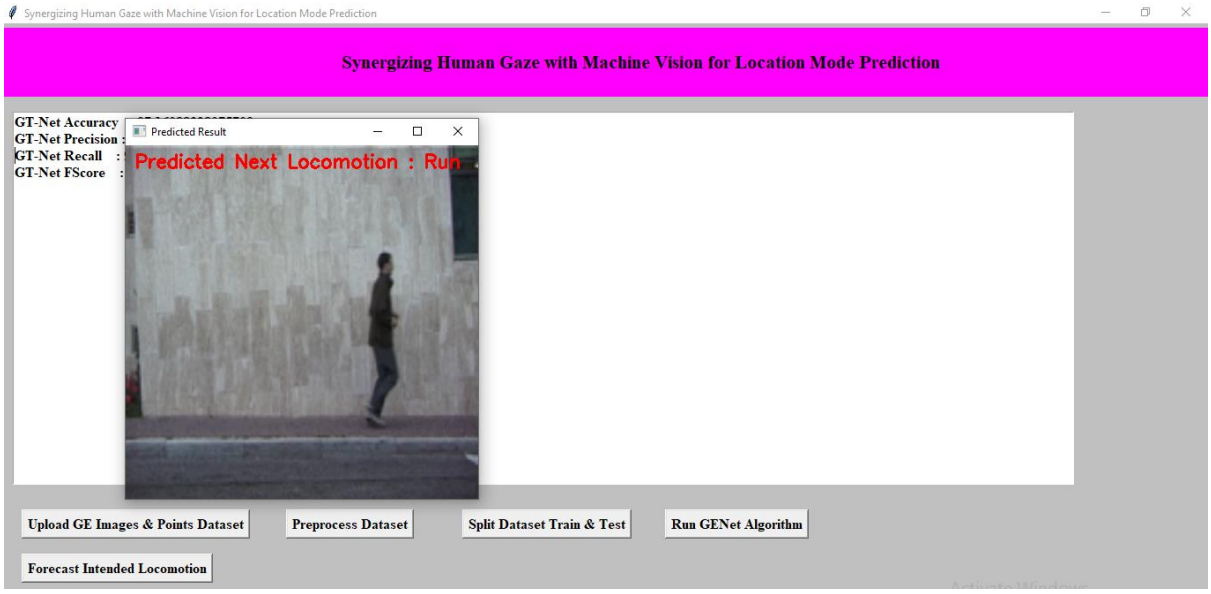

Figure 8: Predicted Output (Run)

Figure 8 shows the Predicted Output is running displays the GUI with a pop-up window labelled Predicted Result showing a person in motion. The red text in the pop-up states Predicted Next Locomotion: Run, indicating that the system has forecasted running as the next movement.

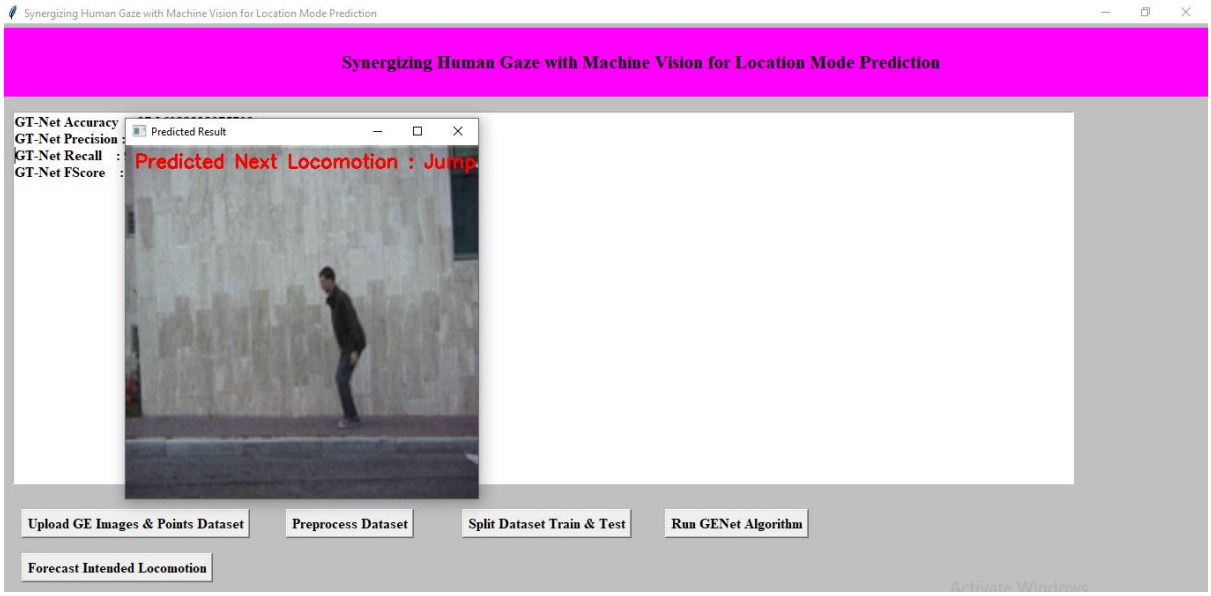

Figure 9: Predicted Output (Jump)

Figure 9: Predicted output is jumping the image displays an updated version of the graphical user interface (GUI) for the project titled Synergizing Human Gaze with Machine Vision for Location Mode Prediction. The interface has a magenta banner at the top, with the title prominently displayed

5. CONCLUSION

This proposed GT-NET system represents a significant advancement in the field of human movement and intent prediction by integrating human gaze data with machine vision systems to predict locomotion modes. Traditional systems, while effective in capturing and recording physical movements, were inherently limited by their inability to understand the intent behind those movements or to adapt to different environments and users. These limitations made them less effective in applications requiring anticipatory actions, such as wearable robotics or advanced driver-assistance systems. GT-NET overcomes these challenges by leveraging the power of deep learning to combine multiple data modalities, including human gaze and cloud points from user movement, to create a more comprehensive and predictive model of human behavior. The integration of human gaze data into the predictive model is particularly innovative, as it allows the system to infer intent based on where the user is looking. This information, combined with traditional movement data, enables the system to anticipate future movements and transitions with greater accuracy. The use of deep learning algorithms allows GT-NET to learn from a vast amount of data, improving its predictive capabilities over time and making it adaptable to different users and environments. GT-NET's ability to predict not just what a user is currently doing, but what they are likely to do next, represents a shift from reactive to proactive systems in human-machine interaction. This capability is crucial in applications like wearable robotics, where anticipating the user's next move can enhance safety, efficiency, and overall user experience. The system's adaptability and learning capabilities make it a powerful tool in various fields, from healthcare to autonomous vehicles, where understanding and predicting human intent can lead to more intelligent and responsive systems.

REFERENCE

[1] A. E. Patla, "Strategies for dynamic stability during adaptive human locomotion," IEEE Eng. Med. Biol. Mag., vol. 22, no. 2, pp. 48–52, Mar. 2003.

[2] T. Zhang, M. Tran, and H. Huang, "Design and experimental verification of hip exoskeleton with balance capacities for walking assistance," IEEE/ASME Trans. Mechatronics, vol. 23, no. 1, pp. 274–285, Feb. 2018.

[3] R. J. Farris, H. A. Quintero, S. A. Murray, K. H. Ha, C. Hartigan, and M. Goldfarb, "A preliminary assessment of legged mobility provided by a lower limb exoskeleton for persons with paraplegia," IEEE Trans. Neural Syst. Rehabil. Eng., vol. 22, no. 3, pp. 482–490, May 2014.

[4] M. Liu, F. Zhang, P. Datseris, and H. Huang, "Improving finite state impedance control of active-transfemoral prosthesis using dempstershafer based state transition rules," J. Intell. Robot Syst., vol. 76, nos. 3–4, pp. 461–474, 2014.

[5] T. Lenzi, M. Cempini, L. Hargrove, and T. Kuiken, "Design, development, and testing of a lightweight hybrid robotic knee prosthesis," Int. J. Robot. Res., vol. 37, no. 8, pp. 953–976, Jul. 2018.

[6] H. A. Varol, F. Sup, and M. Goldfarb, "Multiclass real-time intent recognition of a powered lower limb prosthesis," IEEE Trans. Biomed. Eng., vol. 57, no. 3, pp. 542–551, Mar. 2010.

[7] Y. D. Li and E. T. Hsiao-Wecksler, "Gait mode recognition and control for a portable-powered ankle-foot orthosis," in Proc. IEEE 13th Int. Conf. Rehabil. Robot. (ICORR), Jun. 2013, pp. 1–8.

[8] Y. Long et al., "PSO-SVM-based online locomotion mode identification for rehabilitation robotic exoskeletons," Sensors, vol. 16, no. 9, p. 1408, 2016.

[9] D. Xu, Y. Feng, J. Mai, and Q. Wang, "Real-time on-board recognition of continuous locomotion modes for amputees with robotic transtibial prostheses," IEEE Trans. Neural Syst. Rehabil. Eng., vol. 26, no. 10, pp. 2015–2025, Oct. 2018.

[10] H. Huang, T. A. Kuiken, and R. D. Lipschutz, "A strategy for identifying locomotion modes using surface electromyography," IEEE Trans. Biomed. Eng., vol. 56, no. 1, pp. 65–73, Jan. 2009.

[11] H. Huang, F. Zhang, L. J. Hargrove, Z. Dou, D. R. Rogers, and K. B. Englehart, "Continuous locomotion-mode identification for prosthetic legs based on neuromuscular–mechanical fusion," IEEE Trans. Biomed. Eng., vol. 58, no. 10, pp. 2867–2875, Oct. 2011. [12] L. J. Hargrove et al., "Robotic leg control with EMG decoding in an amputee with nerve transfers," New England J. Med., vol. 369, no. 13, pp. 1237–1242, Sep. 2013.

[13] M. Liu, D. Wang, and H. H. Huang, "Development of an environmentaware locomotion mode recognition system for powered lower limb prostheses," IEEE Trans. Neural Syst. Rehabil. Eng., vol. 24, no. 4, pp. 434–443, Apr. 2016.

[14] F. Zhang and H. Huang, "Source selection for real-time user intent recognition toward volitional control of artificial legs," IEEE J. Biomed. Health Informat., vol. 17, no. 5, pp. 907–914, Sep. 2013

[15] J. P. Diaz et al., "Visual terrain identification and surface inclination estimation for improving human locomotion with a lower-limb prosthetic," in Proc. Annu. Int. Conf. IEEE Eng. Med. Biol. Soc. (EMBC), 2018, pp. 1817–1820.

[16] B. Laschowski, W. McNally, A. Wong, and J. McPhee, "Preliminary design of an environment recognition system for controlling robotic lower-limb prostheses and exoskeletons," in Proc. IEEE 16th Int. Conf. Rehabil. Robot. (ICORR), Jun. 2019, pp. 868–873.

[17] B. Zhong, R. L. D. Silva, M. Li, H. Huang, and E. Lobaton, "Environmental context prediction for lower limb prostheses with uncertainty quantification," IEEE Trans. Autom. Sci. Eng., vol. 18, no. 2, pp. 458–470, Apr. 2021.