Project Tortoise: Biweekly Report #1

Efaz Hossain and Sahithi Mutyala

Report Submission Date: October 2, 2024

**Introduction**

After a few weeks of development, significant progress has been made in improving the facial recognition model. We have successfully incorporated augmented images to expand the dataset, enhancing the model’s training and performance. This improvement has led to better accuracy and scalability in recognizing facial classes, setting a strong foundation for further developments. This report provides a detailed overview of my progress, including the challenges encountered, technical advancements and a clear outline of my next steps

**Summary**

In the development of the Tortoise system’s facial and object recognition models, data preprocessing, augmentation, and regularization have played key roles in enhancing model performance. Data preprocessing involved resizing all images to a uniform 224x224 pixels, normalizing pixel values, and applying facial alignment techniques to ensure consistency across the dataset. This step was critical in standardizing the input data, reducing noise, and improving the accuracy of both the facial and object recognition models.

To address the challenge of limited data, data augmentation techniques were extensively applied. For facial recognition, we began with 3-10 images per facial class, which were then augmented to 1000 per class through transformations such as random cropping, rotation, flipping, and brightness adjustments. Similarly, for object recognition, 2000 images per object class were expanded to 100,000 per class using similar augmentation techniques. These augmentations helped increase data diversity, prevent overfitting, and improve the model’s generalization to unseen data.

Lastly, regularization techniques were employed to further enhance model robustness. Dropout was applied during training to prevent overfitting by randomly deactivating a subset of neurons in the network. Additionally, early stopping was used to halt training once the model’s performance on a validation set plateaued, reducing the risk of over-training. Together, these methods have led to more efficient learning and better generalization of the system’s facial and object recognition models.

**Progress and Milestones**

During this period, several key tasks have been completed. The facial recognition model has been successfully developed and fine-tuned using the VGGFace2 dataset. We uploaded initial facial images (3-10 per class) and applied data augmentation to expand each class to 1000 images, leading to improved model performance and accuracy. Additionally, data preprocessing was completed for both facial and object recognition models. This included standardizing all images to 224x224 pixels, performing normalization, and aligning facial images for consistency. Preprocessing for object recognition involved working with an initial dataset of 2000 images per class. Furthermore, data augmentation was applied to the object recognition dataset, increasing it to 100,000 images per class, while regularization techniques like dropout and early stopping were implemented to prevent overfitting and optimize training efficiency.

Several tasks are currently underway. The development of the geofencing alert mechanism for wandering prevention is in progress, with research and design being carried out to ensure this feature’s effectiveness in the beta prototype. In addition, work is being done to enhance the mobile application by addressing bugs and improving the UI to make it more user-friendly, particularly for the elderly users. The object recognition model is also being refined using the newly augmented dataset of 100,000 images per class to enhance its accuracy and performance.

Some tasks are still in the planning phase. One of the upcoming tasks is the integration of the YOLOv9 architecture, which is expected to improve object detection by reducing inference times and providing more precise bounding boxes for objects. Another major task that has yet to begin is the transition to a multi-user system, which will allow the Tortoise system to be deployed in senior living facilities for monitoring multiple users simultaneously.

**Problem-Solving and Challenges**

Throughout the development of the Tortoise system, several challenges have emerged. One of the primary obstacles was the inefficiency in expanding the facial recognition model to accommodate additional classes. With each new class, the time complexity increased significantly, making the training process slow and resource intensive. Additionally, the object recognition model faced issues with limited dataset size, which hampered the model’s ability to generalize effectively. Usability of the mobile application for the elderly also posed a challenge, as early versions had bugs and a complicated UI, making it less accessible to the intended users.

To address the challenge of inefficiency in facial recognition model training, we adopted incremental learning techniques, which allowed the model to adapt to new classes without retraining from scratch. This approach reduced the time complexity and improved the system’s scalability. For the object recognition model, extensive data augmentation techniques were applied to significantly expand the dataset. This enabled the model to learn from a more diverse set of images, improving its accuracy and generalization. In tackling the mobile application’s usability issues, the development team worked on simplifying the UI and fixing bugs, making the interface more intuitive and accessible for elderly users.

The implementation of incremental learning has had a substantial impact on the performance and scalability of the facial recognition model, allowing for faster training while maintaining accuracy as new facial classes are added. The use of data augmentation has greatly enhanced the object recognition model’s ability to accurately detect and classify objects, even in diverse conditions, leading to better performance in real-world scenarios. The improvements in the mobile application have made it easier for elderly users to interact with the system, ensuring that the Tortoise platform remains both functional and user-friendly. These solutions collectively contribute to a more efficient, accurate, and accessible system, bringing us closer to the project’s goals of enhancing independence and safety for users with memory loss.

**Technical Depth and Accuracy**

For this project, we utilized three key models: VGG16, EfficientNet, and YOLOv8, each enhancing facial and object recognition. VGG16, a 16-layer CNN, is well-suited for facial recognition due to its strong feature extraction capabilities. We fine-tuned it with the VGGFace2 dataset to handle complex facial data, crucial for identifying familiar faces. EfficientNet was chosen for its ability to scale while maintaining accuracy and efficiency, making it ideal for real-time object recognition. YOLOv8 excels in fast, real-time object detection with optimized bounding boxes for locating items quickly.

To optimize performance, we applied hyperparameter tuning across models. For VGG16, we adjusted learning rates, batch size, and dropout to prevent overfitting, achieving over 95% accuracy. EfficientNet was fine-tuned by optimizing scaling factors and using the Adam optimizer, reaching 92% accuracy in object recognition while maintaining low inference times. For YOLOv8, anchor boxes, learning rates, and confidence thresholds were tuned, leading to 90% average precision in object detection.

These models provide a strong, efficient foundation for the Tortoise system, enabling seamless real-time facial and object recognition, essential for improving user independence and safety.

**Future Plans and Goals**

Over the next two weeks, the primary focus will be to kickstart the machine learning process for the Tortoise system. This will begin with data collection and cleaning, ensuring that the dataset for both facial and object recognition tasks is diverse, high-quality, and free of inconsistencies. We will also prepare the dataset by splitting it into training, validation, and test sets, while applying essential preprocessing techniques such as normalization and alignment to ensure uniformity. Additionally, data augmentation will be implemented to enhance the dataset, with techniques like random cropping, flipping, and brightness adjustments used to improve model generalization. The augmented data will be validated to ensure meaningful variations.

Once the data is ready, we will initialize the models, using VGG16 for facial recognition, EfficientNet for object recognition, and YOLOv8 for object detection, all with pre-trained weights to speed up the learning process. Following initialization, the initial round of model training will commence, where we will train the models on the preprocessed and augmented data. During this phase, we will closely monitor key metrics such as training loss and accuracy to assess model performance.

Alongside training, hyperparameter tuning will be performed to optimize learning. We will experiment with various learning rates, batch sizes, and dropout rates, as well as test different optimizers such as Adam and SGD to determine the best fit for the tasks at hand. Finally, after the initial training, we will conduct a preliminary model evaluation on the validation set, measuring important performance metrics such as accuracy, precision, recall, and F1 score. This evaluation will help identify any areas needing further improvement, laying the groundwork for the next phase of tuning and retraining.