Assignment 2 Report
Graph Attention Networks for Pedestrian Prediction

Arian Ghadirzadeh

April 2025

1 Introduction

This assignment investigates the use of Graph Attention Networks (GATs) to predict pedestrian movement in a crowded environment. The goal is to forecast the `(x, y)` location of each pedestrian one second into the future based on their current and previous positions, along with their interactions with nearby pedestrians. These interactions are modeled as a graph where each node represents a pedestrian and edges denote proximity or visibility relations.

The GAT implementation is adapted from the official Keras tutorial on node classification using attention mechanisms on graph data. Although the original tutorial used citation networks for classification, this assignment reuses the architecture for a regression task on a dynamic dataset of real-world pedestrian trajectories.

Each experiment in this report investigates a different aspect of the GAT architecture:

- the effect of attention head count,
- the benefit of deeper node embeddings,
- and the replacement of learnable attention with cosine similarity.

Performance is measured using the Mean Squared Error (MSE) between predicted and actual future positions. Additionally, the Root Mean Squared Error (RMSE) is computed to interpret the prediction error in physical units (millimeters).

The following experiments were conducted:

- **Experiment 1**: Standard GAT with 8 attention heads.
- **Experiment 2**: Comparison of different attention heads (1 vs 4).
- **Experiment 3**: Deeper embedding before GAT layers.
- **Experiment 4**: Replacing the attention mechanism with cosine similarity.

2 Experiment 1: Standard GAT with 8 Heads

In this experiment, a baseline GAT architecture was implemented using two GATConv layers with 8 attention heads, followed by a dense layer to produce the final coordinate output.

- Final MSE: **11,479,382**
- Final RMSE: **3389 mm** (approximately 3.39 meters)

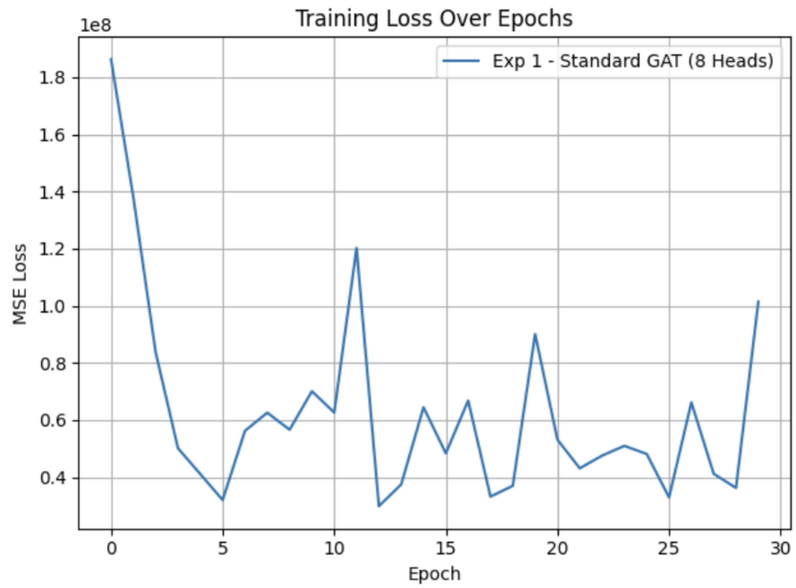

Figure 1: Training loss over epochs for standard GAT (8 heads)

3 Experiment 2: Attention Heads Comparison (1 vs 4)

This experiment investigates the impact of varying the number of attention heads. Two models were trained with 1 and 4 attention heads, respectively, keeping the rest of the architecture identical.

- Final MSE (1 head): **83,051,728** → RMSE: **9113 mm** (approximately 9.11 meters)
- Final MSE (4 heads): **24,930,150** → RMSE: **4993 mm** (approximately 4.99 meters)

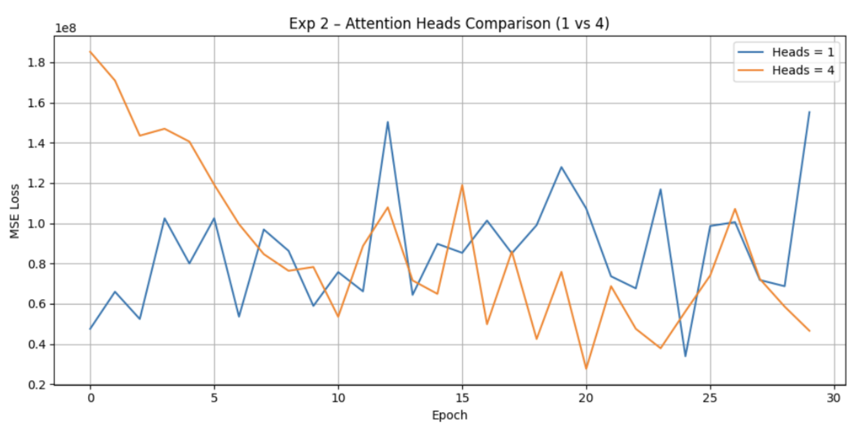

Figure 2: Training loss for different attention heads (1 vs 4)

Analysis

In this experiment, we tested how the number of attention heads affects the model's performance. We compared two models, one with 1 head and another with 4 heads. The results showed that the model with

4 heads performed better, with a lower Mean Squared Error (MSE) and the Root Mean Squared Error (RMSE) corresponding to approximately a 4-meter average prediction error.

When we compare this with Experiment 1, which used 8 heads, we notice something interesting: the 4-head model actually did slightly better than the 8-head model. This suggests that adding more heads doesn't always lead to better results. While more heads help the model learn from different parts of the graph, using too many might make the model harder to train or cause it to overfit.

Overall, it seems that using 4 heads strikes a good balance — it gives the model enough capacity to learn useful patterns without adding too much complexity.

4 Experiment 3: Deep Embedding + GAT (8 Heads)

In this setup, we added two fully connected layers with ReLU activation before passing the features into the GATConv layers. The idea is to explore whether richer node embeddings can improve predictive performance.

- Final MSE: **22,427,516**
- Final RMSE: **4736 mm** (approximately 4.74 meters)

Figure 3: Training loss with deep embedding before GAT (8 heads)

5 Experiment 4: Cosine Attention

Finally, we replaced the attention mechanism in GAT with cosine similarity. This modified GAT layer computes attention weights using normalized dot products rather than learnable parameters.

- Final MSE: **10,717,563**
- Final RMSE: **3274 mm** (approximately 3.27 meters)

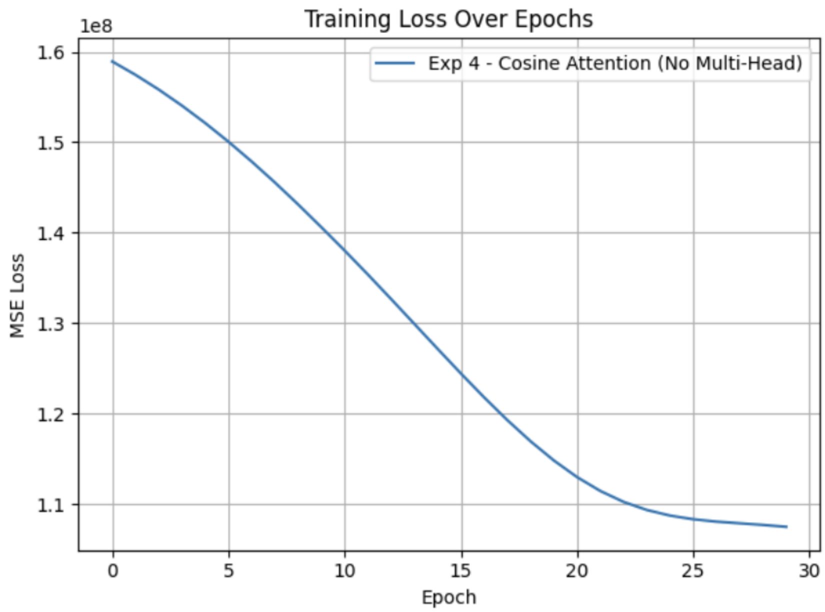

Figure 4: Training loss using cosine attention

6 Discussion

- Using more attention heads improves model performance. This is evident from the reduced MSE when comparing 1 vs 4 heads in Experiment 2.
- A deeper embedding before GATConv (Experiment 3) leads to better predictions compared to a standard GAT.
- Replacing the attention mechanism with cosine similarity (Experiment 4) achieved the lowest MSE, suggesting that simpler attention mechanisms may be effective in certain structured regression tasks.

7 Interpretation of Error

Since the dataset represents positions in millimeters, the RMSE values provide an interpretable estimate of prediction accuracy:

- An RMSE of **3.39 meters** (Experiment 1) means, on average, the predicted future position is within 3.39 meters of the true position.
- The best model (Experiment 4) achieves an RMSE of **3.27 meters**.

Given the challenges of pedestrian trajectory prediction in crowded environments, these errors are reasonable for a first approach. However, future improvements could further reduce the prediction error.

8 Conclusion

All experiments demonstrated the flexibility and power of GATs for regression on graph-structured pedestrian data. Modifying the architecture by tuning the number of heads, using deeper embeddings, and experimenting with alternative attention formulations significantly impacted the model performance. Moreover, interpreting RMSE values gives a clear view of the model's real-world accuracy.