GOAL CONDITIONED TRAJECTORY PREDICTION FOR SOCIAL NAVIGATION

Jay Patrikar
jaypat@cs.cmu.edu

Arti Anantharaman
artia@andrew.cmu.edu

Carlos Gonzalez
cggonzal@andrew.cmu.edu

May 21, 2021

1 Research Questions

Today, manned and unmanned vehicles are separated, limiting the utility and flexibility of operations and reducing efficiency. While established rules to separate manned and unmanned aerial traffic exist, one domain that is not well defined is in terminal airspaces around airports. To enable unmanned aircrafts to use the same infrastructure as GA while at the same time ensuring safe interaction with manned aircraft, it is necessary to achieve safe manned-unmanned vehicle teaming. Not only will this improve the system performance and have each agent (robot/human) learn from each other in various aircraft operations, it also has the potential to reduce the manning needs of manned aircraft especially in emergency scenarios.

Operations in the vicinity of an airport, heliport, or forward operating base traffic pattern (See Fig. 1) are a basic capability that every beginning pilot needs to master and where conflicts between aircraft are common and need to be resolved. Mastering Visual Flight Rules (VFR) operations for autonomous aircraft has significant operational advantages at unimproved sites, as well as in achievable traffic density compared to Instrument Flight Rules (IFR) or completely separated operations between manned and unmanned systems.

One of the challenges that a pilot planning to land at an untowered airport has to face is the ability to predict the motion of other agents. This prediction is conditioned on multiple sources of information like radio voice communications, current position and past trajectories of the agents, knowledge of the local traffic pattern, weather, and past experience. In this project, we propose to develop a motion prediction system to predict the trajectories of multiple agents conditioned on the position and past history of all the agents, weather and traffic pattern information. To simulate the knowledge from the voice communication, we will assume a known final position. We plan to incorporate three main contexts 1) Weather Context 2) Social Context 3) Intermediate and Final Goal Context to develop a learning-based socially-compliant trajectory prediction system.

In mathematical terms, given the context ϕ and past trajectories $Y_{1:t}^{1:A}$ for all agents $\{1, ..., A\}$ for time $[1, t]$, we are interested in calculating the conditional probability of observing the future trajectories $X_{t+1:t+h}^{1:A}$ for the time horizon h.

$$P(X_{t+1:t+h}^{1:A} | Y_{1:t}^{1:A}, \phi) \tag{1}$$

The proposed research questions are then:

- Social Context : How the past trajectories of other agents affect the trajectory prediction?
- Environmental Context : How do environmental factors like wind affect the trajectory prediction?
- Goal Context : How do perceived goals of agents affect the trajectory prediction?
- Multi-future Context : How do we account for multiple futures in the trajectory prediction?

2 Related Work

Trajectory prediction in domains such as pedestrian movements (Amirian *et al.* [2019]; Mohamed *et al.* [2020]) and vehicular traffic (Lee *et al.* [2017]; Rhinehart *et al.* [2018, 2019]) is a well studied problem. Recently, multiple deep

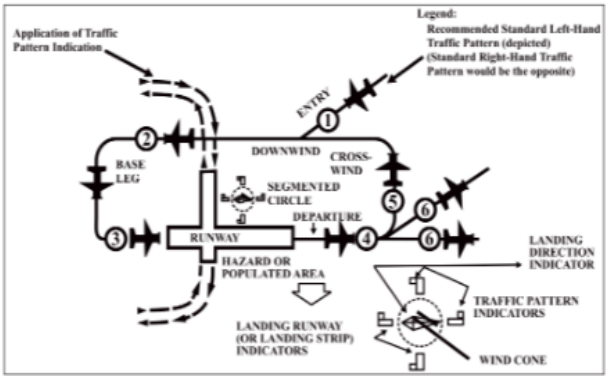

Figure 1: Standard Airport Traffic Pattern

learning based algorithms have produced promising results in both the domains but long horizon prediction still remains a challenge. To the best of authors knowledge, no such work exists for aircraft trajectory prediction. The domain-level challenges include prediction in 3D, prediction in absolute coordinate and prediction with a novel weather context. We pick two baselines from the trajectory prediction domain to demonstrate the advantages of our proposed algorithm:

Trajectory prediction has been studied by Mangalam *et al.* [2020] and Monti *et al.* [2020]. Mangalam *et al.* [2020] uses a network they call "Y-Net" that is composed of 3 separate networks, as shown in figure 2, in order to predict a trajectory and incorporate uncertainty into the prediction. The first network U_e attempts to extract information from the past trajectory and a segmentation map. The second network U_g generates a heat map for the way points and the goal. The third network U_t uses the goal and waypoint heatmap output from U_g in order to predict a future trajectory.

Monti *et al.* [2020] uses an approach based on recurrent neural networks and Variational Autoencoders (VAE) that relies on 2 graph neural networks (GNN) as seen in figure 3. The first GNN defines the future objectives while the second models the hidden states and agent interactions. The core assumption made in this approach is that Recurrent Neural Networks are the best model for representing and predicting time series data.

However, Bai *et al.* [2018] argues that convolutions are a better way of extracting time series data and provides convincing results that robust and state of the art sequence modeling can be performed using convolution based networks instead of recurrent based networks.

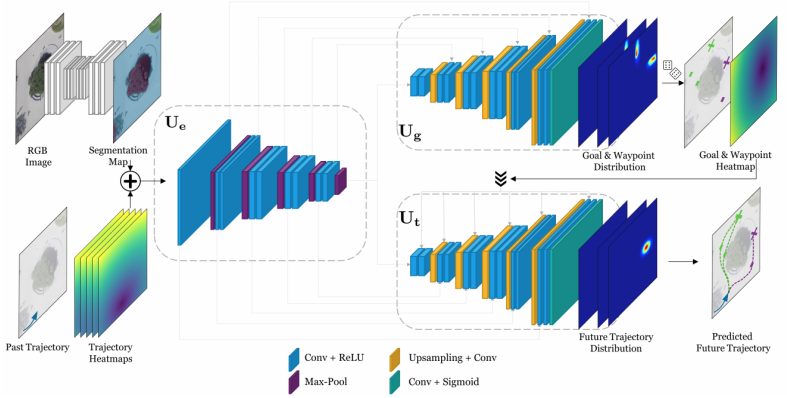

Figure 2: Y-Net Architecture from Mangalam *et al.* [2020]

Thus the challenge in trajectory prediction has 3 main parts. First, how will the data be represented such that the maximum amount of information can be extracted from the time axis? Second, what part of the model's architecture will extract the time series data? And third, how do you represent the goal and way points such that a model has enough room to learn but is not penalized greatly when a trajectory deviates by a small amount? One final thing to note is that a fourth challenge is added in the case where an agent must reason about other agents in the environment thus emphasizing the need for the model to take into account the social context of the problem.

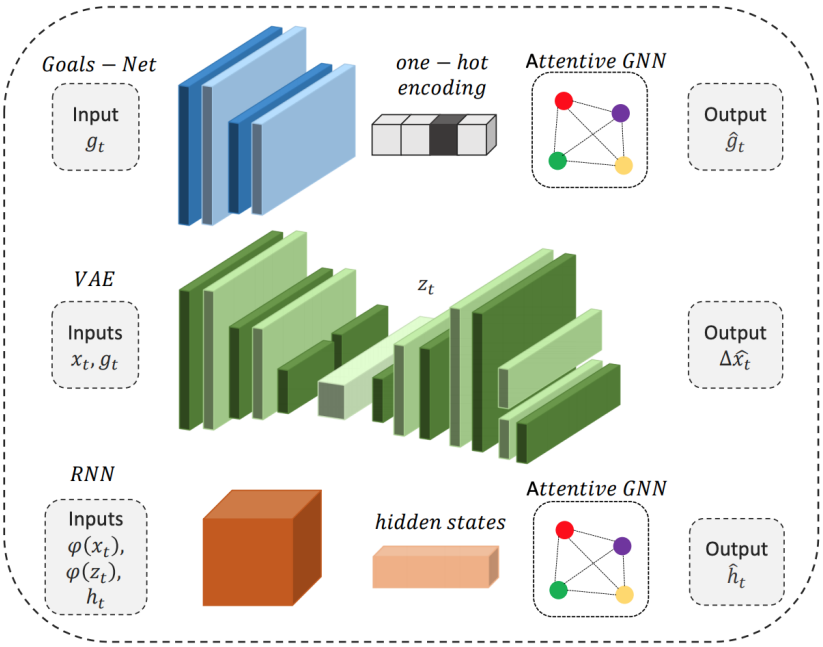

Figure 3: DAG-Net Architecture from Monti *et al.* [2020]

3 Proposed Approach

The proposed architecture is shown in Fig 4. The dataset consists of "scenes" where each scene represents a window of data in time. To simulate knowledge of radio communications, we plan to leak the future to get the final goal location for each agent. Trajectory data will be obtained from the ADS-B transponders while the weather data will be obtained from the METAR streams at the airport. The data from both the sources will be synced using UTC timestamps. For encoding the timeseries trajectory data for each agent, we plan to use a Temporal Convolutional Network (TCN) as detailed in Bai *et al.* [2018]. Bai *et al.* [2018] have shown TCNs to be atleast either superior or equivalent to corresponding RNN architectures. Due to the convolutional architecture they are scalable due to parallelization, and do not have the same gradient issues as RNNs. A simple TCN encoder-decoder setup should be able to generate meaningful trajectories.

In order to incorporate the social context, we plan to use a Graph Attention Network (GAN) as defined in Veličković *et al.* [2017] that uses the TCN encoded output as its node features. A multi-headed attention network can reason about the effect each node (agent) has on the other agents by looking at the data. Earlier graph-based approaches hard-code this effect based on relative distances of agents to each other but this assumption might not always hold true. The GAN thus encodes the relative effects of each agent thereby encoding the social effect on the trajectory.

In order to incorporate the weather and final goal context, the values can be appended to the TCN encoded output before they are incorporated into the GAN. The values thus become part of the features of each node (agent). This may ensure that the future trajectories of all the agents are not only influenced by their past trajectories but also on the final goals of all the agents and weather. Intuitively, this knowledge of final goals and weather should improve the TCN baseline for all the agents.

To capture the multi-modality of the data, we propose using a Conditional Variational Autoencoder (CVAE) at the end. The CVAE will be trained in a style similar to other conditional trajectory prediction approaches Lee *et al.* [2017]. The proposed CVAE architecture is shown for one agent in Figure 5.

In regards to the wind information, we have written a script that fetches the METAR string corresponding to a given UTC timestamp and extracts the wind speed and wind direction from it. This is a useful script because we can retrieve the wind parameters that prevailed at any point in the dataset. The wind speed and wind direction are, by default, in the True North frame. Since the length of the runway serves as the x-axis in all our experiments, we rotated the wind direction from the True North frame to the runway frame. The wind was resolved into two components:

1. Headwind: Wind speed along the the positive x-axis of the runway

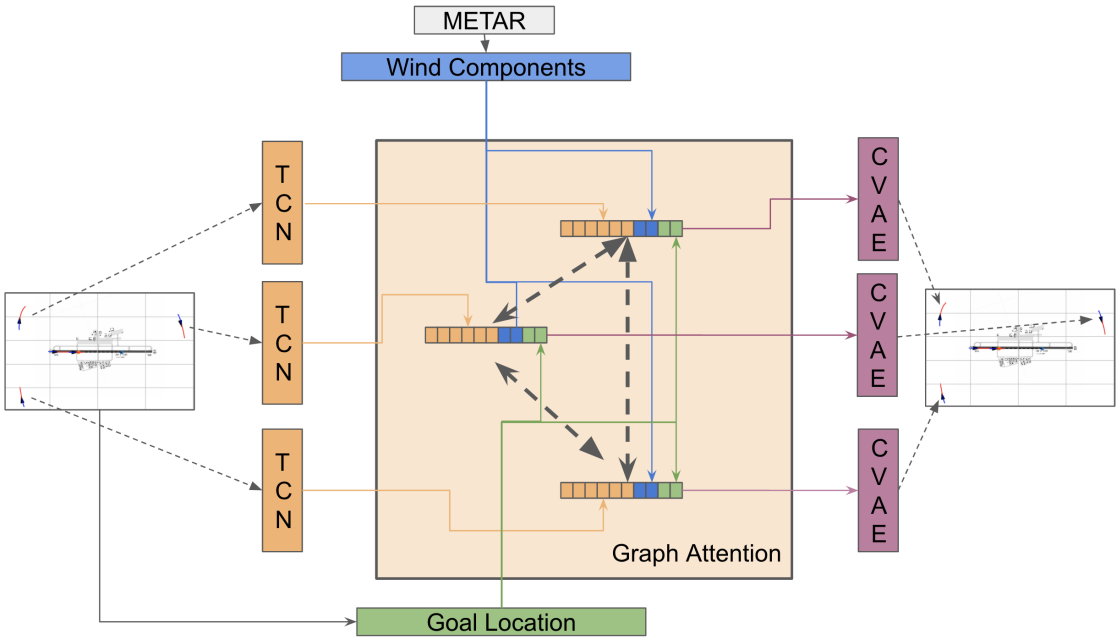

Figure 4: Proposed Complete Architecture

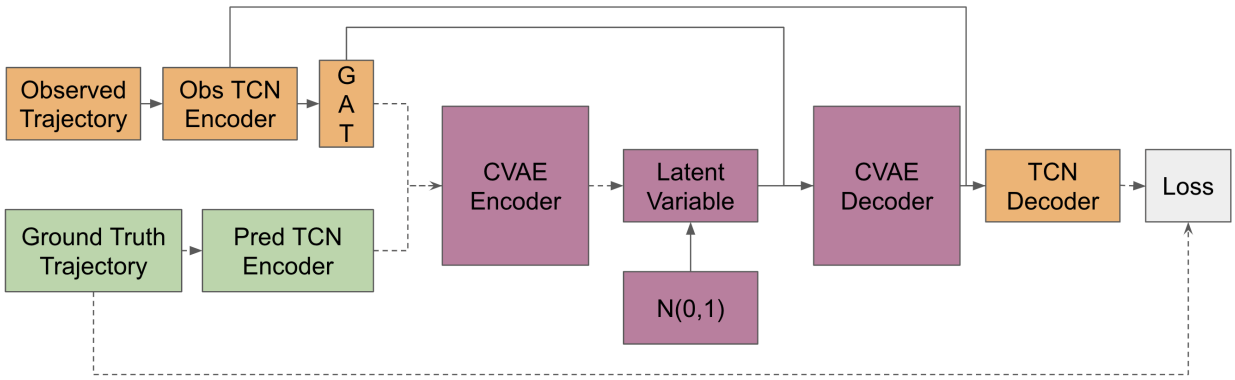

Figure 5: Proposed CVAE Architecture

2. Crosswind: Wind speed along the y-axis, i.e., perpendicular to the runway

This data has been appended to the dataset. The model now has to be trained by incorporating the obtained wind information.

3.1 Evaluation Plan

Evaluation of the proposed approach is carried out on the collected dataset. We use 7 days of data and split it 70-30 for training and testing. The code is written using Pytorch and Python.

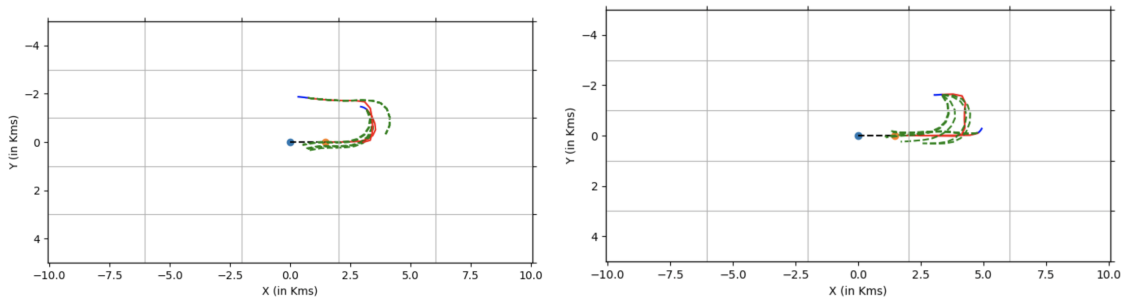

Figure 6: Some results using 8 sec as input (blue) and 120 sec as prediction output (green) using the MotivAir network are plotted against the ground truth (red).

We plan to use two metrics to evaluate the proposed architecture. Average Displacement Error (ADE) and Final Displacement Error (FDE) are two commonly used metrics for estimating how well the generated trajectories distribution represents the underlying distribution. To take into account the multi-future nature of trajectory generation the algorithm is queried K-times and the best ADE and FDE are reported for each agent. A value of K=5 is often used.

$$ADE = \frac{\sum_{a \in A} \sum_{t \in \{t+1:t+h\}} \sqrt{(\hat{X}_t^a - X_t^a)^2}}{|A| \times h} \tag{2}$$

$$FDE = \frac{\sum_{a \in A} \sqrt{(\hat{X}_{t+h}^a - X_{t+h}^a)^2}}{|A|} \tag{3}$$

Equations 2 and 3 show the ADE and FDE calculations respectively where $\hat{X}$ is the predicted output and X is the ground truth.

4 Results

4.1 Quantitative Results

Table 1 shows our results. There is significant improvement with addition of social context. Addition of the environmental context did not show significant improvement but that can be attributed to the fact that we only used 7 days of data that may not have captured the distribution. Comparative results are show wit DAG-Net as a baseline.

Method	ADE(Km)	FDE(Km)
Motivair 2D	0.46	0.80
Motivair 2D w/o Wind	0.45	0.81
MotivAir 2D w/o Social w/o Wind	0.61	0.44
DAGNet 2D	1.04	1.06

Table 1: Results from experiments

4.2 Qualitative Results

The qualitative results are shown in Figure 6. The results show 5 trajectories that capture the distribution of the underlying data. The figure on the left shows how the social context is captured by the model. The top aircraft is predicted with a longer downwind to improve separation between this and the aircraft on final. The figure on the right shows the multi-modality of the data which captures the stochasticity of the underlying data. The aircraft turning base is shown with multiple possible trajectories.

5 Conclusion

In this paper we proposed a new architecture for predicting the trajectories of multiple agents that incorporates social interaction, environmental context, time varying context, and multi-future context. We evaluated the results of the

architecture on our data set and compared it to the state of the art DAGNet architecture Monti *et al.* [2020]. We showed that we were able to outperform DAGNet with all the variants of our data set that included a social component using GAT Veličković *et al.* [2017].

In the future we would like to include goal conditioning into our architecture similar to that proposed by Rhinehart *et al.* [2019]. We hope that the addition of goal conditioning will lead to even better results on our data set. Additionally, we would like to investigate why the results shown in table 1 did not show a significant increase in accuracy when the environmental context was included.

6 Team

6.1 Jay

Jay incorporated a CVAE structure Lee *et al.* [2017] into the pipeline and generated some preliminary results. Jay is a PhD student in the Robotics Institute at CMU. His research mainly focuses on finding methods and techniques to enable mobile robots to operate in safety critical real-world environments without sacrificing performance. Jay also has two post-graduate degrees in Robotics and Aerospace Engineering with a graduate degree in Aerospace Engineering. For this project, he will be focusing on the intermediate goal conditioning to improve long horizon prediction.

6.2 Carlos

Carlos incorporated the social aspect into the pipeline. The pipeline currently has a graph neural network based on Veličković *et al.* [2017] that uses a GAT based approach to add social context into the model. The pipeline still requires some tweaks relating to internal dimensionality checks, but once that is resolved we will be able to train the model and compare the results to the model that did not include the GNN for social interactions. We hope that once we incorporate the GNN and the weather context, we will see significantly improved results when compared to the baseline models.

Carlos did his undergrad at CMU in Electrical and Computer Engineering (ECE) and is currently doing his Masters in ECE. He worked with the AirLab on the AlphaPilot and the DARPA Subterranean challenge projects. In the AlphaPilot project he helped to set up the Computer Vision pipeline and so has a background in Machine Learning and Computer Vision. He will be focusing on the social aspect of the project which reasons about how the different agents, in this case aircraft, will act when placed in an environment with other agents.

6.3 Arti

Arti is a Master's student in the Robotic Systems Development (MRSD) program at CMU. She has a background in motion planning, controls, haptics, sensor-robot registration, and software development. In regards to this project, she will be focusing on incorporating wind parameters in the trajectory prediction and evaluating resultant model performance. Over the past few weeks, she has extracted wind speed and direction from METAR strings, rotated them to the runway frame, and appended the wind components to the dataset. Her goal for the next milestone is to modify the TCN-GAT architecture to include weather data and train the model to determine if context improves trajectory prediction.

References

Amirian, Javad, Hayet, Jean-Bernard, & Pettré, Julien. 2019. Social ways: Learning multi-modal distributions of pedestrian trajectories with gans. *Pages 0–0 of: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops*.

Bai, Shaojie, Kolter, J. Zico, & Koltun, Vladlen. 2018. *An Empirical Evaluation of Generic Convolutional and Recurrent Networks for Sequence Modeling*.

Lee, Namhoon, Choi, Wongun, Vernaza, Paul, Choy, Christopher B, Torr, Philip HS, & Chandraker, Manmohan. 2017. Desire: Distant future prediction in dynamic scenes with interacting agents. *Pages 336–345 of: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*.

Mangalam, Karttikeya, An, Yang, Girase, Harshayu, & Malik, Jitendra. 2020. *From Goals, Waypoints & Paths To Long Term Human Trajectory Forecasting*.

Mohamed, Abduallah, Qian, Kun, Elhoseiny, Mohamed, & Claudel, Christian. 2020. Social-stgcnn: A social spatio-temporal graph convolutional neural network for human trajectory prediction. *Pages 14424–14432 of: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*.

Monti, Alessio, Bertugli, Alessia, Calderara, Simone, & Cucchiara, Rita. 2020. *DAG-Net: Double Attentive Graph Neural Network for Trajectory Forecasting*.

Rhinehart, Nicholas, Kitani, Kris M, & Vernaza, Paul. 2018. R2p2: A reparameterized pushforward policy for diverse, precise generative path forecasting. *Pages 772–788 of: Proceedings of the European Conference on Computer Vision (ECCV)*.

Rhinehart, Nicholas, McAllister, Rowan, Kitani, Kris, & Levine, Sergey. 2019. Precog: Prediction conditioned on goals in visual multi-agent settings. *Pages 2821–2830 of: Proceedings of the IEEE International Conference on Computer Vision*.

Veličković, Petar, Cucurull, Guillem, Casanova, Arantxa, Romero, Adriana, Lio, Pietro, & Bengio, Yoshua. 2017. Graph attention networks. *arXiv preprint arXiv:1710.10903*.