**Claude.AI**

I'd be happy to help you prepare for your thesis defense presentation. Based on the slides you've shared, I'll create a detailed presentation guide for each slide.

**Thesis Defense Presentation Guide**

**General Presentation Tips:**

* Stand confidently and make eye contact with your committee members and audience
* Speak clearly and at a moderate pace
* Use the slides as a guide, but avoid reading directly from them
* Be prepared to elaborate on any slide if questions arise
* Aim for about 1-2 minutes per slide (total presentation ~20-25 minutes)
* Practice your presentation multiple times to ensure smooth transitions

**Detailed Guide for Each Slide:**

**Slide 1: Title Slide**

* Begin with a warm greeting: "Good afternoon, professors and committee members."
* Introduce yourself: "My name is Ankita Sarkar, and I'm presenting my thesis on Missing Spatial-Temporal Multi-Modal Traffic Data Flow Imputation and Prediction."
* Briefly mention your appreciation for the opportunity to present your research

**Slide 2: Introduction**

* Start with the problem significance: "Traffic flow data plays a crucial role in urban planning and intelligent transportation systems."
* Highlight the challenge: "However, sensor failures and transmission errors lead to missing data, which significantly impacts the accuracy of traffic analysis."
* State your thesis goal clearly: "My research addresses this challenge by developing a method for missing data imputation in spatial-temporal multimodal traffic datasets."
* Introduce your solution briefly: "I've developed a Graph Recurrent Imputation Network, or GRIN, which effectively handles these missing data points."

**Slide 3: Background and Motivation**

* Connect with real-world applications: "Let me explain why this research matters. Accurate traffic data is essential for urban mobility planning, traffic forecasting, and intelligent transportation systems."
* Use the visual to illustrate the problem: "As you can see in this graph, missing data points create significant gaps in our understanding of traffic patterns."
* Explain the causes of missing data: "These gaps occur due to sensor failures, communication issues, and power outages."
* Emphasize consequences: "Without addressing these gaps, traffic prediction and management systems operate with reduced accuracy, affecting millions of urban commuters."

**Slide 4: Literature Review**

* Start with classical approaches: "Previous work in this field includes classical imputation models that focus on either temporal or spatial dependencies."
* Transition to newer methods: "More recently, deep learning approaches using RNNs, LSTMs, and CNNs have shown promising results."
* Highlight graph neural networks: "Graph Neural Networks have emerged as powerful tools for capturing spatial dependencies in traffic data."
* Identify research gaps: "However, existing models have limitations in handling multimodal data and capturing complex spatial-temporal dependencies, which my research addresses."

**Slide 5: Proposed Solution: GRIN**

* Introduce your model confidently: "To address these limitations, I propose GRIN - a Graph Recurrent Imputation Network."
* Explain the core concept: "GRIN integrates Graph Neural Networks with Recurrent Neural Networks to capture both spatial and temporal dependencies in traffic data."
* Highlight your innovations: "The key innovations of my approach are the incorporation of multimodal features including building density, type diversity, and geographic location, along with bidirectional imputation to enhance accuracy."

**Slide 6: GRIN Architecture**

* Walk through the visual diagram: "This diagram illustrates the GRIN architecture. Let me walk you through how information flows through the system."
* Explain the encoder: "The encoder uses a Message Passing Graph Recurrent Network, where nodes exchange information with their neighbors."
* Describe the learning process: "Through multiple hidden layers, the network learns from these embeddings to accurately impute missing values."
* Connect back to the problem: "This architecture is specifically designed to handle the spatial-temporal nature of traffic data."

**Slide 7: Methodology**

* Introduce your dataset: "For this research, I used data from the Traffic4cast 2022 challenge, focusing on London with loop counter data and building footprints."
* Explain preprocessing steps: "My methodology involved several preprocessing steps, including dataset splitting, mask generation for simulating missing data, and construction of a sequential graph-based time series dataset."
* Highlight the adjacency matrix: "A critical component was calculating the adjacency matrix to represent spatial relationships between traffic nodes."
* Mention feature engineering: "I engineered features related to building density, type, and location to enhance the model's understanding of traffic patterns."

**Slide 8: Feature Engineering**

* Elaborate on spatial features: "For spatial features, I calculated building density and diversity metrics for each traffic node area. These features are relevant because urban density directly influences traffic patterns."
* Explain temporal features: "Temporal features include time of day and day of week, capturing the cyclical nature of traffic."
* Describe multimodal integration: "A key contribution of my work is the integration of these diverse features. By combining spatial and temporal information, GRIN develops a more comprehensive understanding of traffic patterns."

**Slide 9: Experiments and Results**

* Set up the experimental framework: "To evaluate GRIN, I used standard metrics including RMSE and MAE, comparing against several baseline methods."
* Present results with confidence: "As you can see from these results, GRIN consistently outperformed baseline methods, with a 15% improvement in imputation accuracy."
* Highlight visualizations: "These visualizations compare the imputed traffic flow with actual data, showing how closely GRIN can approximate real traffic patterns even with significant missing data."

**Slide 10: Data Exploration and Analysis**

* Share observed patterns: "During data exploration, I observed several interesting traffic patterns in London, particularly around congestion hotspots like London Bridge, Elephant and Castle, and Potters Bar."
* Discuss feature importance: "Analysis showed that building density and time of day were particularly strong predictors of traffic flow."
* Address challenges: "I encountered several challenges with the dataset, including irregular sampling rates and varying sensor reliability, which informed my approach to preprocessing."

**Slide 11: Conclusions**

* Summarize findings confidently: "In conclusion, GRIN effectively imputes missing traffic data by leveraging both spatial and temporal dependencies."
* Emphasize the multimodal aspect: "The incorporation of multimodal features significantly improved imputation accuracy compared to baseline methods."
* State your contributions clearly: "This research contributes a novel approach to traffic data imputation and provides valuable insights for intelligent traffic management systems."

**Slide 12: Future Work**

* Suggest model improvements: "Looking forward, there are several exciting directions for future work, including exploring different GNN architectures and incorporating real-time data streams."
* Mention broader applications: "This approach could be extended to other cities and traffic datasets, and integrated with traffic forecasting models for comprehensive traffic management solutions."
* Show enthusiasm for continuation: "I'm particularly interested in exploring how this work could be applied to emerging smart city initiatives."

**Slide 13: Q&A**

* Thank the audience: "Thank you for your attention. I'd be happy to answer any questions about my research."
* Stand confidently and be ready to address questions thoughtfully
* Have backup slides ready for anticipated technical questions

**Final Preparation Tips:**

* Prepare for common questions about methodology choices, limitations, and alternative approaches
* Know your data and results thoroughly
* Be honest if you don't know an answer, but suggest how you might investigate it
* Consider preparing a one-page handout with key findings for committee members

I hope this guide helps you prepare for your defense! Would you like me to elaborate on any specific part of the presentation?

**++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++**

**Perplexity.ai**

Okay! Congratulations on reaching your thesis defense! I will help you create a set of presentation slides based on your thesis report, "Missing spatial-temporal multi-modal traffic data flow imputation and prediction," and provide detailed notes on how to present them effectively.

**Thesis Defense Presentation Slides**

Here's a suggested structure for your presentation slides, incorporating key aspects of your thesis:

**Slide 1: Title Slide**

* **Title:** Missing Spatial-Temporal Multi-Modal Traffic Data Flow Imputation and Prediction
* **Your Name:** Ankita Sarkar
* **Date:** March 6, 2025
* **University Logo** (if applicable)

**Slide 2: Introduction**

* **Problem Statement:**
  + Traffic flow data is critical for urban planning and intelligent transportation.
  + Missing data from sensor failures and transmission errors is a significant challenge.
* **Thesis Goal:**
  + To address missing data imputation in spatial-temporal multimodal traffic datasets.
* **Solution Overview:**
  + Using a Graph Recurrent Imputation Network (GRIN).

**Slide 3: Background and Motivation**

* **Why is Traffic Data Important?** (Visualize with a city traffic image)
  + Urban mobility planning
  + Traffic forecasting
  + Intelligent Transportation Systems
* **The Problem of Missing Data:** (Show a graph with missing data points)
  + Sensor failures, communication issues, power outages
  + Impacts accuracy of traffic prediction and management

**Slide 4: Literature Review**

* **Classical Imputation Models:**
  + Temporal Dependency Models (HA, ARIMA, PPCA)
  + Spatial Dependency Models (Regression, Kriging)
  + Spatial-Temporal Dependency Models (Kernel PPCA, KNN)
* **Deep Learning-Based Models:**
  + RNNs, LSTMs, CNNs, GANs
  + Mention GRU-D, BRITS, CNN-based context encoder
* **Graph Neural Networks**
  + Explain how GNNs capture spatial dependencies in traffic data
* **Limitations of Existing Models**
  + Highlight the gaps your thesis addresses

**Slide 5: Proposed Solution: GRIN**

* **Graph Recurrent Imputation Network (GRIN)**
  + Integrates Graph Neural Networks (GNNs) and Recurrent Neural Networks (RNNs).
  + Captures spatial and temporal dependencies.
* **Key Innovations:**
  + Multimodal features: building density, building type diversity, geographic location.
  + Imputation in both forward and backward directions

**Slide 6: GRIN Architecture (Model Overview)**

* **(Visual Diagram of GRIN Architecture)**
  + Show the flow of data through the GNN and RNN components.
  + **Encoder Architecture:** Message Passing Graph Recurrent Network
* **Explain Message Passing:**
  + Nodes exchange information with neighbors.
  + Hidden layers learn from embeddings.

**Slide 7: Methodology**

* **Dataset:**
  + Traffic4cast 2022 challenge (London).
  + Loop counter data, building footprints.
* **Data Preprocessing:**
  + Dataset split
  + Mask Generation
  + Sequential Graph-Based Time Series Dataset Construction
  + Adjacency Matrix Calculation
* **Feature Engineering:**
  + Building density, type, location.

**Slide 8: Feature Engineering (Details)**

* **Spatial Features:**
  + Explain how building density and diversity are calculated.
  + Why are these features relevant to traffic flow?
* **Temporal Features:**
  + Time of day, day of week.
* **Multimodal Integration:**
  + How spatial and temporal features are combined in the model.

**Slide 9: Experiments and Results**

* **Experimental Setup:**
  + Evaluation metrics (e.g., RMSE, MAE).
  + Baselines compared against.
  + Explaination of the experimental settings
* **Key Results:**
  + Present quantitative results (tables, graphs).
  + Highlight improvements over baselines.
* **Visualizations:**
  + Show imputed traffic flow compared to actual data.

**Slide 10: Data Exploration and Analysis**

* **Traffic Patterns:**
  + Discuss observed traffic patterns in the dataset.
  + Congestion examples (London Bridge, Elephant and Castle, Potters Bar)
* **Feature Importance:**
  + Which features contributed most to imputation accuracy?
* **Data Challenges:**
  + Discuss any difficulties encountered with the dataset.

**Slide 11: Conclusions**

* **Summary of Findings:**
  + GRIN effectively imputes missing traffic data.
  + Multimodal features improve accuracy.
* **Contributions:**
  + Novel approach to traffic data imputation.
  + Insights for intelligent traffic management.

**Slide 12: Future Work**

* **Model Improvements:**
  + Explore different GNN architectures.
  + Incorporate real-time data.
* **Broader Applications:**
  + Extend to other cities and traffic datasets.
  + Integrate with traffic forecasting models.

**Slide 13: Q&A**

* **Thank You**
* **Questions?**

**Detailed Presentation Notes**

Here's a guide on how to present each slide effectively:

**Slide 1: Title Slide**

* "Good morning/afternoon, everyone. My name is Ankita Sarkar, and I'm excited to present my master's thesis on missing spatial-temporal multi-modal traffic data flow imputation and prediction."
* Briefly introduce yourself and the topic.

**Slide 2: Introduction**

* "Traffic data is essential for managing our cities, but it's often incomplete. My thesis focuses on solving this problem by accurately filling in the gaps in traffic data."
* Clearly state the problem and your proposed solution.

**Slide 3: Background and Motivation**

* "Imagine trying to plan a route when half the road signs are missing. That's the challenge we face with incomplete traffic data. This slide highlights why accurate traffic data is crucial and the real-world impact of missing information."
* Use a relatable analogy to explain the importance of your work.
* Emphasize the consequences of missing data.

**Slide 4: Literature Review**

* "Before diving into my solution, let's look at what others have done. This slide summarizes existing approaches, from classical statistical methods to deep learning techniques."
* Briefly describe each category of models.
* Highlight the limitations of previous approaches and how your thesis builds upon them.

**Slide 5: Proposed Solution: GRIN**

* "My solution is the Graph Recurrent Imputation Network, or GRIN. It's a hybrid model that combines the strengths of graph neural networks and recurrent neural networks to capture both spatial and temporal patterns in traffic."
* Clearly explain the core idea behind GRIN.
* Emphasize the novelty of your approach, especially the use of multimodal features.

**Slide 6: GRIN Architecture (Model Overview)**

* "This slide shows the architecture of GRIN. The GNN component captures spatial relationships between traffic sensors, while the RNN component models the temporal flow of traffic."
* Walk the audience through the diagram, explaining the key components and their interactions.
* Explain the concept of message passing in GNNs.

**Slide 7: Methodology**

* "To evaluate GRIN, I used data from the Traffic4cast 2022 challenge, focusing on London. This slide outlines the steps I took to prepare the data and train the model."
* Briefly explain each step in the methodology, highlighting any unique aspects of your approach.

**Slide 8: Feature Engineering (Details)**

* "I engineered novel spatial features, such as building density and type diversity, to provide GRIN with additional context about the traffic environment. These features capture the relationship between the urban landscape and traffic flow."
* Explain how you calculated the spatial features and why they are relevant.

**Slide 9: Experiments and Results**

* "This slide presents the results of my experiments. GRIN outperformed existing methods in imputing missing traffic data, demonstrating the effectiveness of the proposed approach."
* Use visuals (tables, graphs) to present your results clearly.
* Highlight the key improvements achieved by GRIN.

**Slide 10: Data Exploration and Analysis**

* "Before modelling, I explored the dataset to understand inherent traffic patterns and anomalies".
* "These visualisations show some of the real-world events, such as congestion on roads at certain times on certain days".

**Slide 11: Conclusions**

* "In conclusion, my thesis demonstrates that GRIN is a promising approach for imputing missing traffic data. By integrating spatial and temporal information with multimodal features, GRIN achieves high accuracy and provides valuable insights for traffic management."
* Summarize your key findings and contributions.
* Emphasize the potential impact of your work.

**Slide 12: Future Work**

* "This research opens up several avenues for future work. We can explore different GNN architectures, incorporate real-time data, and extend the model to other cities and traffic datasets."
* Discuss potential improvements and broader applications of your work.

**Slide 13: Q&A**

* "Thank you for your attention. I'm now happy to answer any questions you may have."
* Be prepared to answer questions about your methodology, results, and future work.

**General Presentation Tips:**

* **Practice, Practice, Practice:** Rehearse your presentation multiple times to ensure a smooth delivery.
* **Know Your Audience:** Tailor your language and level of detail to your audience's background.
* **Engage with the Audience:** Make eye contact, use a conversational tone, and encourage questions.
* **Use Visual Aids Effectively:** Keep slides clean and uncluttered, and use visuals to illustrate key points.
* **Be Confident:** Believe in your work and present it with enthusiasm.
* **Time Management:** Stick to your allotted time and prioritize the most important information.

Good luck with your thesis defense! Let me know if you have any other questions.

**++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++**

**Deepseek**

**Below is a suggested structure for your thesis presentation slides, followed by detailed notes on how to present each slide effectively. The slides are based on the content of your thesis report.**

**---**

**### \*\*Slide 1: Title Slide\*\***

**- \*\*Title\*\*: Missing Spatial-Temporal Multi-Modal Traffic Data Flow Imputation and Prediction**

**- \*\*Your Name\*\*: Ankita Sarkar**

**- \*\*Date\*\*: [Insert Date of Defense]**

**- \*\*Reviewers\*\*: Jun.-Prof. Dr. rer. nat. Thomas Liebig, MSc. Siba Mohsen**

**- \*\*Institution\*\*: Technische Universität Dortmund, Data Science**

**\*\*Presentation Notes\*\*:**

**- Start with a confident introduction: "Good [morning/afternoon], my name is Ankita Sarkar, and today I will present my master's thesis on 'Missing Spatial-Temporal Multi-Modal Traffic Data Flow Imputation and Prediction.'"**

**- Briefly mention your reviewers and institution.**

**- Keep this slide simple and professional.**

**---**

**### \*\*Slide 2: Agenda\*\***

**- \*\*1. Introduction\*\***

**- \*\*2. Problem Statement\*\***

**- \*\*3. Literature Review\*\***

**- \*\*4. Methodology\*\***

**- \*\*5. Dataset and Feature Engineering\*\***

**- \*\*6. Experiment Design\*\***

**- \*\*7. Results and Discussion\*\***

**- \*\*8. Conclusion and Future Work\*\***

**\*\*Presentation Notes\*\*:**

**- Briefly outline the structure of your presentation.**

**- Say something like: "I will start with an introduction to the problem, followed by a review of existing literature, the methodology I used, the dataset, experiment design, results, and finally, conclusions and future work."**

**---**

**### \*\*Slide 3: Introduction\*\***

**- \*\*Importance of Traffic Data\*\*: Crucial for urban mobility planning, traffic forecasting, and intelligent transportation systems.**

**- \*\*Challenges\*\*: Missing data due to sensor failures, transmission errors, or sparse data collection.**

**- \*\*Objective\*\*: Address missing data imputation using a Graph Recurrent Imputation Network (GRIN) with multimodal features.**

**\*\*Presentation Notes\*\*:**

**- Explain why traffic data is important and the challenges posed by missing data.**

**- Highlight the objective of your study: "The goal of this thesis is to address the missing data problem using a GRIN model, which integrates graph neural networks and recurrent neural networks, along with novel multimodal features."**

**---**

**### \*\*Slide 4: Problem Statement\*\***

**- \*\*Formal Definition\*\*:**

**- Graph \( G = (V, E) \), where \( V \) = nodes, \( E \) = edges.**

**- Goal: Predict future congestion levels using past loop counter data.**

**- \*\*Congestion Classes\*\*: Green (uncongested), Yellow (moderate), Red (heavy).**

**- \*\*Influence of Infrastructure\*\*: Buildings, water bodies, POIs.**

**\*\*Presentation Notes\*\*:**

**- Define the problem formally: "We represent the traffic network as a graph, where nodes are sensors and edges are road segments. The goal is to predict future congestion levels using past data."**

**- Mention the congestion classes and how surrounding infrastructure influences traffic patterns.**

**---**

**### \*\*Slide 5: Literature Review\*\***

**- \*\*Classical Imputation Models\*\*: Temporal, spatial, spatial-temporal, and exogenous dependency-based models.**

**- \*\*Deep Learning Models\*\*: LSTM, RNN, CNN, GAN.**

**- \*\*Graph Neural Networks (GNNs)\*\*: GCN, STGCN, DSTGCN.**

**\*\*Presentation Notes\*\*:**

**- Summarize the key approaches in the literature: "Classical models focus on temporal and spatial dependencies, while deep learning models like LSTM and GANs have gained popularity. GNNs, such as GCN and STGCN, are particularly effective for traffic data."**

**- Highlight the gap your research addresses: "Most existing models do not fully integrate spatial and temporal dependencies, which is where our GRIN model comes in."**

**---**

**### \*\*Slide 6: Methodology\*\***

**- \*\*GRIN Model\*\*: Combines GNNs and RNNs for spatial-temporal imputation.**

**- \*\*Encoder Architecture\*\*: Message Passing Graph Recurrent Network (MPNN + GRU).**

**- \*\*Decoder Architecture\*\*: Spatio-Temporal Graph Convolution.**

**- \*\*Feature Engineering\*\*: Building density, diversity, and type encoding.**

**\*\*Presentation Notes\*\*:**

**- Explain the GRIN model: "Our model combines graph neural networks and recurrent neural networks to capture both spatial and temporal dependencies."**

**- Briefly describe the encoder and decoder architectures.**

**- Highlight the novel features: "We introduced building density, diversity, and type encoding to improve imputation accuracy."**

**---**

**### \*\*Slide 7: Dataset and Feature Engineering\*\***

**- \*\*Data Source\*\*: Traffic4cast 2022 NeurIPS competition (London).**

**- \*\*Features\*\*:**

**- Building density: Number of buildings per unit area.**

**- Building type diversity: Number of unique building types.**

**- Location type: Residential, commercial, educational, etc.**

**- \*\*Exploratory Data Analysis\*\*: Congestion patterns, speed classes.**

**\*\*Presentation Notes\*\*:**

**- Describe the dataset: "We used data from the Traffic4cast 2022 competition, focusing on London."**

**- Explain the feature engineering process: "We added features like building density and type diversity to capture the influence of surrounding infrastructure."**

**- Mention the exploratory analysis: "We analyzed congestion patterns and speed classes to understand the data better."**

**---**

**### \*\*Slide 8: Experiment Design\*\***

**- \*\*Dataset Split\*\*: Training, validation, and test sets.**

**- \*\*Mask Generation\*\*: Training and evaluation masks for missing data.**

**- \*\*Adjacency Matrix\*\*: Euclidean distance and Gaussian RBF kernel.**

**- \*\*Model Parameters\*\*: 8a38982 trainable parameters, 300 epochs, batch size 32.**

**\*\*Presentation Notes\*\*:**

**- Explain how the dataset was split: "We divided the data into training, validation, and test sets to ensure robust evaluation."**

**- Describe the mask generation process: "We created masks to simulate missing data and evaluate the model's imputation performance."**

**- Mention the adjacency matrix calculation: "We used Euclidean distance and a Gaussian RBF kernel to construct the adjacency matrix."**

**- Highlight the model parameters: "The model has over 8 million trainable parameters and was trained for 300 epochs with a batch size of 32."**

**---**

**### \*\*Slide 9: Results and Discussion\*\***

**- \*\*Performance Metrics\*\*: RMSE, MAPE, MAE.**

**- \*\*Key Findings\*\*:**

**- GRIN outperforms traditional models in imputation accuracy.**

**- Multimodal features improve imputation performance.**

**- \*\*Visualizations\*\*: Congestion patterns, speed classes.**

**\*\*Presentation Notes\*\*:**

**- Present the results: "Our model achieved lower RMSE, MAPE, and MAE compared to traditional models, indicating better imputation accuracy."**

**- Highlight the impact of multimodal features: "The inclusion of building density and type diversity significantly improved the model's performance."**

**- Show visualizations: "Here are some visualizations of congestion patterns and speed classes, which demonstrate the model's effectiveness."**

**---**

**### \*\*Slide 10: Conclusion and Future Work\*\***

**- \*\*Conclusion\*\*:**

**- GRIN effectively imputes missing traffic data.**

**- Multimodal features enhance imputation accuracy.**

**- \*\*Future Work\*\*:**

**- Integrate additional modalities (weather, accidents).**

**- Improve spatial modeling with dynamic graphs.**

**- Generalize to other cities.**

**\*\*Presentation Notes\*\*:**

**- Summarize the key findings: "Our GRIN model successfully imputes missing traffic data, and the addition of multimodal features improves accuracy."**

**- Discuss future work: "In the future, we plan to integrate more data sources, improve spatial modeling, and test the model in other cities."**

**---**

**### \*\*Slide 11: Questions\*\***

**- \*\*Title\*\*: Thank You!**

**- \*\*Subtitle\*\*: Questions?**

**\*\*Presentation Notes\*\*:**

**- End with a confident conclusion: "Thank you for your attention. I am now happy to take any questions."**

**- Be prepared to answer questions about your methodology, results, and future work.**

**---**

**### \*\*Detailed Presentation Notes\*\*:**

**1. \*\*Practice\*\*: Rehearse your presentation multiple times to ensure smooth delivery. Time yourself to stay within the allotted time (usually 15-20 minutes for a thesis defense).**

**2. \*\*Engage the Audience\*\*: Make eye contact, speak clearly, and avoid reading directly from the slides.**

**3. \*\*Explain Visuals\*\*: When showing graphs or visualizations, explain what they represent and how they support your findings.**

**4. \*\*Anticipate Questions\*\*: Think about potential questions the reviewers might ask and prepare answers in advance.**

**5. \*\*Stay Calm\*\*: If you don’t know the answer to a question, it’s okay to say, "That’s a great question, and I’ll need to look into that further."**

**Good luck with your defense!**

**++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++**

**Chatgpt**

**Thesis Defense Presentation: Missing Spatial-Temporal Multi-Modal Traffic Data Flow Imputation and Prediction**

**Slide 1: Title Slide**

* **Title:** Missing Spatial-Temporal Multi-Modal Traffic Data Flow Imputation and Prediction
* **Your Name:** Ankita Sarkar
* **Institution:** Technische Universität Dortmund
* **Supervisors:** Jun.-Prof. Dr. rer. nat. Thomas Liebig, MSc. Siba Mohsen
* **Date:** Thesis Defense Date

**Slide 2: Introduction**

* **Why is traffic data important?**
  + Urban mobility planning
  + Traffic forecasting
  + Intelligent transportation systems
* **Problem Statement:**
  + Missing data due to sensor failures, transmission errors, sparse collection
  + Challenge in accurate traffic forecasting

**Slide 3: Research Objective**

* **Main Goal:** Develop a model to accurately impute missing spatial-temporal multimodal traffic data
* **Key Contributions:**
  + Implementing Graph Recurrent Imputation Network (GRIN)
  + Integrating spatial (geographic) and temporal (time-series) dependencies
  + Using multimodal features like building density and geographic influence

**Slide 4: Literature Review**

* **Classical Approaches:**
  + ARIMA, kNN, Bayesian Networks, Kriging models
* **Deep Learning Approaches:**
  + LSTMs, RNNs, GANs, CNN-based context encoders
* **Graph Neural Network (GNN) Approaches:**
  + Graph Convolutional Networks (GCN)
  + Spatio-Temporal Graph Convolutional Networks (STGCN)
  + Graph Attention Networks (GAT)

**Slide 5: Problem Definition**

* **Graph Representation:**
  + Nodes = Traffic Sensors
  + Edges = Road Segment Connectivity
* **Traffic Data Components:**
  + Temporal Features (Speed, Flow, Congestion Levels)
  + Spatial Features (Node Location, Building Density, POI Influence)
* **Prediction Task:**
  + Impute missing values and predict congestion class (Green, Yellow, Red)

**Slide 6: Methodology**

* **Dataset:** Traffic4cast 2022 Challenge (London City Data)
* **Model Used:** Graph Recurrent Imputation Network (GRIN)
* **Steps:**
  1. **Data Preprocessing**
  2. **Graph Construction** (Adjacency Matrix, Euclidean Distance, Similarity Score)
  3. **Feature Engineering** (Building Density, POI, Road Graph)
  4. **Model Training and Evaluation**

**Slide 7: Graph Recurrent Imputation Network (GRIN)**

* **Why GRIN?**
  + Captures both spatial and temporal dependencies
  + Uses GCNs for spatial relations, GRUs for time-series imputation
  + Performs bidirectional imputation for better accuracy
* **Architecture:**
  + Spatial-Temporal Encoder (MPNN + GRU)
  + Spatial Decoder (Graph Convolutional Network)
  + Imputation Executor (Bidirectional Prediction)

**Slide 8: Dataset Description**

* **Source:** Traffic4cast 2022 (London)
* **Data Points:**
  + Nodes: 59,110
  + Edges: 132,414
  + Loop Counters: 3,751
* **Time Period:** 2019-07-01 to 2020-01-31
* **Sampling Rate:** 15-minute intervals

**Slide 9: Data Preprocessing**

* **Handling Missing Data:**
  + Mask Generation (Training Mask, Evaluation Mask)
* **Graph Construction:**
  + Euclidean Distance Computation
  + Adjacency Matrix Calculation
  + Similarity Score (Gaussian Radial Basis Function - RBF)
* **Feature Engineering:**
  + Building Density Calculation
  + POI Influence Mapping
  + Location Type Encoding (Residential, Commercial, Industrial, etc.)

**Slide 10: Experiment Design**

* **Train-Validation-Test Split**
  + Time series forecasting setup (Window-Horizon Approach)
  + 70% Training, 15% Validation, 15% Testing
* **Evaluation Metrics:**
  + RMSE (Root Mean Squared Error)
  + MAE (Mean Absolute Error)
  + MAPE (Mean Absolute Percentage Error)

**Slide 11: Results**

* **Imputation Accuracy:**
  + GRIN performed better than traditional methods (kNN, ARIMA, simple RNNs)
* **Prediction Accuracy:**
  + Improved congestion class prediction compared to baseline models
  + Effective use of spatial features in multimodal learning

**Slide 12: Conclusion**

* **Key Findings:**
  + GRIN effectively imputes missing traffic data using spatial-temporal dependencies
  + Multimodal features like building density enhance performance
* **Impact:**
  + Supports intelligent traffic management
  + Can be extended for real-world deployment

**Slide 13: Future Work**

* **Potential Enhancements:**
  + Integrating real-time traffic data streams
  + Experimenting with more advanced GNN architectures (Graph Transformers)
  + Expanding the dataset to multiple cities

**Slide 14: Q&A**

* **Thank You!**
* **Any Questions?**

**Thesis Defense Presentation Notes**

**Slide 1: Title Slide**

* Greet the audience: "Good [morning/afternoon/evening], everyone. Thank you for being here today."
* Introduce yourself: "I am Ankita Sarkar, and I will be presenting my master's thesis titled 'Missing Spatial-Temporal Multi-Modal Traffic Data Flow Imputation and Prediction.'"
* Mention your institution and supervisors.
* Outline the structure: "I will take you through the problem, methodology, key findings, and future directions."

**Slide 2: Introduction**

* Explain the importance of traffic data: "Accurate traffic data is essential for urban mobility planning, traffic forecasting, and smart transportation."
* Highlight the problem: "However, missing data due to sensor failures and sparse collection poses a significant challenge."
* Briefly introduce your solution: "My study addresses this using Graph Recurrent Imputation Network (GRIN), which integrates Graph Neural Networks (GNNs) and Recurrent Neural Networks (RNNs)."

**Slide 3: Research Objective**

* State the primary goal: "The goal of this research is to develop a robust imputation model that accurately reconstructs missing traffic data."
* List the key contributions:
  1. Implementing GRIN for spatial-temporal dependencies.
  2. Introducing novel multimodal features such as building density and POI influence.
  3. Evaluating model performance on real-world traffic datasets.

**Slide 4: Literature Review**

* Explain traditional methods:
  + "Classical approaches include ARIMA, kNN, and Bayesian Networks."
  + "Deep learning models like LSTMs and CNNs have also been explored."
* Introduce GNNs: "Graph-based models like GCNs and STGNNs have proven effective but lack multimodal integration."

**Slide 5: Problem Definition**

* Define the problem in graph terms: "Traffic data can be represented as a graph with sensors as nodes and roads as edges."
* Explain the missing data issue: "Many nodes have missing values, making accurate prediction difficult."
* Describe the prediction goal: "We aim to reconstruct missing values and predict congestion levels."

**Slide 6: Methodology**

* Explain the dataset: "We use the Traffic4cast 2022 dataset, which includes loop counter data from London."
* Summarize the steps:
  1. Preprocessing
  2. Graph Construction
  3. Feature Engineering
  4. Model Training and Evaluation

**Slide 7: GRIN Model**

* Explain why GRIN is suitable: "It captures spatial and temporal dependencies using GCNs and GRUs."
* Describe the architecture: "The model consists of an encoder, decoder, and imputation executor."

**Slide 8: Dataset Description**

* Provide dataset statistics: "The dataset contains 59,110 nodes, 132,414 edges, and 3,751 loop counters."
* Mention data coverage: "Covers July 2019 - January 2020 at 15-minute intervals."

**Slide 9: Data Preprocessing**

* Explain the missing data handling: "We generate masks to differentiate missing and available data."
* Describe graph construction: "We use Euclidean distance and similarity scores to build an adjacency matrix."

**Slide 10: Experiment Design**

* Explain dataset split: "We use a time-series forecasting setup with 70% training, 15% validation, and 15% testing."
* List evaluation metrics: "We measure performance using RMSE, MAE, and MAPE."

**Slide 11: Results**

* Present key findings:
  + "GRIN significantly outperforms classical imputation models."
  + "Spatial features enhance accuracy in congestion prediction."
* Show model effectiveness: "GRIN provides better reconstructions for missing traffic data."

**Slide 12: Conclusion**

* Summarize key contributions:
  + "GRIN successfully imputes missing data and predicts congestion levels."
  + "Multimodal features improve accuracy."
* Highlight real-world impact: "This approach supports intelligent traffic systems."

**Slide 13: Future Work**

* Suggest enhancements:
  + "Integrate real-time traffic data."
  + "Experiment with graph transformers."
  + "Expand the dataset to multiple cities."

**Slide 14: Q&A**

* Thank the audience: "Thank you for your time. I welcome any questions!"
* Engage with questions confidently and concisely.