

Imagery2Flow: Predicting Human Mobility Flows in Cities Using Deep Learning on Satellite Imagery

Corresponding Author: Professor Song Gao

This file contains all reviewer reports in order by version, followed by all author rebuttals in order by version.

Version 0:

Reviewer comments:

Reviewer #1

(Remarks to the Author)

This paper proposes a deep learning method that takes the satellite imagery as inputs and generates the human mobility OD flows. The idea is novel since no previous studies have tried OD flow generation by using satellite imagery. However, I have the following concerns about this manuscript.

1. About the rationale and the motivation. The authors' motivation to do such OD flow generation is that "origin-destination (OD) mobility flow information between different sub-regions of cities that relies on traditional surveys is costly and time-consuming to get updated". However, the used ground-truth OD flow is the 2020 Origin-Destination Employment Statistics (LODES)[50] dataset in each MSA. In other words, we already have the OD flow data, the proposed method is only to reproduce the existing dataset, which is somehow not that meaningful to me.
2. About the problem definition. The authors define the OD flow as a graph, not dynamic graph with respect to time. To me, OD flow between regions should vary from day to day, time to time, thus the problem definition is not so sound to me.
3. Following 2, there are many related works about OD flow generation/prediction in the research area of computer science/AI that modes OD flows as a dynamic graph problem (i.e., $T \times N \times N$, N is the number of regions and T are timestamps). More comprehensive literature review on OD flow generation/prediction should be conducted.
4. About the position and the core contribution. To me, this is a deep learning approach paper, so the main focus should be on the effectiveness and the accuracy of the proposed method. Calling back to 3, as an approach paper, more baselines should be included besides the basic Radiation model and Gravity model.
5. Following 4, the impact factors on city OD flow analyzed in Section "2.3 Correlation to Urban Morphology" and "2.4 Spatial Heterogeneity" is based on the predicted OD flow, which is not necessary contents. If the authors would like to claim their proposed approach is accurate, they can just show the performances on many metrics. If the authors would like to reveal the insights on the impact factors on OD flows, the analysis can be done based on the ground-truth OD flow data.
6. The above two points make me feel the contribution is somehow blurry and weak.
7. Basic definitions about the spatial unit should be written in a clearer way. "Each region of interest (ROI) is partitioned into N spatial areas according to their administrative boundaries (e.g., census tracts)", so what is each area among N ?
8. Some typos exist including but not limited to:
Section 2.1, 1st paragraph, Given->given.
Section 2.3, While higher population->while
Section 4.3, the indent after Eq.4 and Eq.5 before where should be deleted.

(Remarks on code availability)

Reviewer #2

(Remarks to the Author)

In this manuscript, the authors proposed Imagery2Flow, a methodology that leverages satellite imagery to predict commute flows. Visual representations derived from satellite imagery are used to extract geospatial features, which are subsequently integrated into a Graph Attention Network to effectively model and predict commute flows. The proposed approach is validated using commute flow data from the top 10 largest Metropolitan Statistical Areas (MSAs) in the United States. Notably, the transferability tests demonstrate the model's cross-city generalization, and the study explores how urban factors (e.g., centrality and compactness) influence human mobility, providing insights into the interplay between urban morphology and movement patterns. While the study presents an innovative model for commute flow prediction, improvements are needed in the clarity of methods and scope of experiments.

Major questions:

1. How can the proposed model account for the temporal characteristics of human mobility such as prediction of OD flows over different time periods (e.g., morning commute or weekend activity)? Considering temporal variations can demonstrate the practical applicability and broader generalizability of the proposed model.
2. The authors utilized ResNet-50 for visual feature extraction. To demonstrate the generalization of the proposed framework, it would be beneficial to include experimental results using various backbone CNN models.
3. Can the proposed methodology be applicable at different levels of spatial granularity (e.g., census block groups, ZIP Code Tabulation Areas)? Depending on the spatial scale of the region-of-interest (ROI), multiple satellite images may be included. In such cases, how is the representation of the ROI handled? Clarifying this would help assess the versatility and scalability of the approach.
4. Clarify the nature of the node embeddings utilized in the GAT module. Specifically, are the embeddings derived from satellite imagery used as the graph's node embeddings? A more detailed explanation would greatly help in understanding this aspect of the methodology. If the satellite imagery embeddings are indeed employed as node embeddings, it would suggest that the study enhances commute flow prediction performance by leveraging visual representations as auxiliary information.
5. Some recent works have enhanced commute flow prediction accuracy using GNNs by incorporating node features derived from external data sources [1-4], such as OpenStreetMap (OSM) data or Primary Land Use Tax Lot Output (PLUTO) datasets. Does the use of visual representations from satellite imagery provide unique advantages over these conventional data sources? It would be beneficial to include a performance comparison between the approach using satellite imagery and those utilizing OSM or PLUTO datasets.
6. An ablation study is necessary to better understand the contribution of each component in the proposed model. Additionally, what drives the performance differences between the proposed model and the baseline deep gravity model? Providing a detailed analysis of these differences would help clarify the strengths and unique advantages of the proposed approach.

Minor comments:

1. In Section 2.3, clarify the description of BCSI values. The current statement "A smaller BCSI value indicates a more compact city, with a value of 1 representing a perfect circle." is incorrect. A value of 1 generally represents irregular shape.
2. In Appendix C, the visibility of the circles used to illustrate the heterogeneous effect needs to be improved.

References:

- [1] Zhao, Y., Cheng, S., Gao, S., Wang, P., & Lu, F. (2024). Predicting origin-destination flows by considering heterogeneous mobility patterns. *Sustainable Cities and Society*, 106015.
- [2] Liu, Z., Miranda, F., Xiong, W., Yang, J., Wang, Q., & Silva, C. (2020, April). Learning geo-contextual embeddings for commuting flow prediction. In *Proceedings of the AAAI conference on artificial intelligence* (Vol. 34, No. 01, pp. 808-816).
- [3] Shi, Q., Zhuo, L., Tao, H., & Yang, J. (2024). A fusion model of temporal graph attention network and machine learning for inferring commuting flow from human activity intensity dynamics. *International Journal of Applied Earth Observation and Geoinformation*, 126, 103610.
- [4] Nejadshamsi, S., Bentahar, J., Eicker, U., Wang, C., & Jamshidi, F. (2025). A geographic-semantic context-aware urban commuting flow prediction model using graph neural network. *Expert Systems with Applications*, 261, 125534.

(Remarks on code availability)

The submitted code file includes all relevant materials, including the training process for the image encoder proposed in the paper, the training of the OD flow predictor, and the visualization of results. The README file provides guidance on package requirements, code structure, and instructions to reproduce the results presented in the paper.

Version 1:

Reviewer comments:

Reviewer #1

(Remarks to the Author)

I have checked the rebuttal letter and the revised manuscript, however, the authors' rebuttal cannot address my concerns. I can't recommend this manuscript to be accepted due to the following reasons:

1. The significance of this research is unclear, the OD flow data is already available in 10 metropolitan areas. It is not so difficult to obtain, so why we need this approach.
2. Defining the OD flow as a static one is also a huge limitation.
3. The novelty of the proposed methodology is kind of limited, which is just the simple combination of CNN and GNN.
4. The evaluation of the proposed method is somehow not adequate. Recent baselines and potential better solutions are not adequately included.

(Remarks on code availability)

Reviewer #2

(Remarks to the Author)

Manuscript#: NCOMMS-24-68099

Title: Imagery2Flow: Predicting Human Mobility Flows in Cities Using Deep Learning on Satellite Imagery

Comments for Author

The authors have addressed several concerns effectively, including conducting ablation studies, demonstrating model generalizability using various backbones, and clarifying methodological details. In addition, the authors' focus on exogenous geographic contexts in modeling static OD flow patterns is both valid and insightful, particularly for uncovering the structural determinants of mobility. While the study is well-executed and the results are clearly presented, I believe the technical novelty and distinction from prior work could be more articulated to fully highlight the contribution of the authors' work.

1. While the application is valuable, the methodology appears to build upon well-established frameworks. The proposed model pipeline closely resembles prior studies that leverage satellite imagery or mobility data for urban region embedding [1, 2, 3], as well as GAT-based approaches for static commute flow prediction [4, 5]. In particular, the use of satellite image embeddings as node features in graph-based commute flow prediction models, in itself, may not represent a substantial methodological innovation.

2. Additional details on the reproduction of baseline models are necessary. Since prior studies typically focused on OD flow prediction within a limited number of cities, the hyperparameters used in those studies may not be well suited for the broader or more diverse set of cities considered in this work. For a fair comparison, it is important that such baselines be re-tuned under the current experimental setting.

3. It would be beneficial to include more recent and competitive baseline models.

References:

- [1] Jean, N., Burke, M., Xie, M., Alampay Davis, W. M., Lobell, D. B., & Ermon, S. (2016). Combining satellite imagery and machine learning to predict poverty. *Science*, 353(6301), 790-794.
- [2] Yeh, C., Perez, A., Driscoll, A., Azzari, G., Tang, Z., Lobell, D., ... & Burke, M. (2020). Using publicly available satellite imagery and deep learning to understand economic well-being in Africa. *Nature communications*, 11(1), 2583.
- [3] Zhang, M., Li, T., Li, Y., & Hui, P. (2021, January). Multi-view joint graph representation learning for urban region embedding. In *Proceedings of the twenty-ninth international conference on international joint conferences on artificial intelligence* (pp. 4431-4437).
- [4] Zhao, Y., Cheng, S., Gao, S., Wang, P., & Lu, F. (2024). Predicting origin-destination flows by considering heterogeneous mobility patterns. *Sustainable Cities and Society*, 106015.
- [5] Liu, Z., Miranda, F., Xiong, W., Yang, J., Wang, Q., & Silva, C. (2020, April). Learning geo-contextual embeddings for commuting flow prediction. In *Proceedings of the AAAI conference on artificial intelligence* (Vol. 34, No. 01, pp. 808-816).

(Remarks on code availability)

The submitted code package includes all relevant components, including the module for extracting embeddings from satellite imagery, the OD flow predictor, and the learning objective. As a researcher with experience in deep learning, I had no difficulty running the code. However, it would be helpful if the README file provided more detailed guidance on package requirements, code structure, and instructions for reproducing the results presented in the paper.

Version 2:

Reviewer comments:

Reviewer #2

(Remarks to the Author)

Manuscript#: NCOMMS-24-68099

Comments for Author

The authors have effectively addressed several concerns, particularly by tuning the baseline and incorporating recent models as baselines, which significantly strengthens the validity of the study. However, one concern still remains regarding technical novelty.

The authors utilized features extracted from satellite imagery as region embeddings for the nodes. While this is a reasonable design choice, the methodological novelty of this component seems limited. Specifically, the region embeddings are obtained via a vanilla contrastive learning approach, and it is unclear how compelling or innovative this representation strategy is within the proposed framework.

Nevertheless, the study provides valuable insights by demonstrating how graph-based modeling with satellite imagery can be effectively applied to uncover the structural determinants of human mobility. Furthermore, it shows the feasibility of building a transferable end-to-end model for static OD flow prediction using publicly accessible remote sensing data.

(Remarks on code availability)

The provided README file was clear and well-structured, which made it straightforward to reproduce and utilize the authors' implementation and results.

Open Access This Peer Review File is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made.

In cases where reviewers are anonymous, credit should be given to 'Anonymous Referee' and the source.

The images or other third party material in this Peer Review File are included in the article's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder.

To view a copy of this license, visit <https://creativecommons.org/licenses/by/4.0/>

Response to Reviewer 1

General Comment: This paper proposes a deep learning method that takes the satellite imagery as inputs and generates the human mobility OD flows. The idea is novel since no previous studies have tried OD flow generation by using satellite imagery. However, I have the following concerns about this manuscript.

Response: Thanks very much for your thoughtful review and constructive feedback that helps improve our manuscript. Your comments focused mainly on the motivation, experimental design, and result analysis, and how these points supported our contributions. We fully agree with your opinion that clear statements, additional validation analysis are necessary to highlight key findings.

We have carefully addressed each of your comments and revised the manuscript accordingly. We have expanded our experiments to strengthen the advancement and generalization of our proposed *Imagery2Flow* framework. Additionally, we explored the design idea of separating the spatial interaction learner using GAT (encoder) and the OD flow predictor (decoder) using a gradient boosting regressor as suggested, which improved the model performance significantly and illustrated the flexibility of our framework. The discussion was further enhanced to highlight the key contributions, practical implications, and the limitation of our current framework. We hope that the revision can address your concerns.

Comment 1: About the rationale and the motivation. The authors’ motivation to do such OD flow generation is that “origin-destination (OD) mobility flow information between different sub-regions of cities that relies on traditional surveys is costly and time-consuming to get update”. However, the used ground-truth OD flow is the 2020 Origin-Destination Employment Statistics (LODES)[50] dataset in each MSA. In other words, we already have the OD flow data, the proposed method is only to reproduce the existing dataset, which is somehow not that meaningful to me.

Response: Thanks for your comment. Our research focuses on the origin-destination (OD) mobility flow prediction task, where the objective is to develop a transferable deep learning model to predict the unknown OD flow volumes, based on the characteristics of the region of interest (ROI) captured from the satellite imagery [1]. Using the survey-based LODES dataset is primarily for validating the effectiveness of our proposed deep learning framework.

Although our motivation arises from the fact that traditional data acquisition is labor-intensive and time-consuming, we would like to clarify that our method is not just to reproduce the existing dataset. More precisely, we leverage the existing dataset to train and validate our OD flow prediction model. Through the establishment of a mapping between urban characteristics and OD flows, we aim to uncover the impact of spatial heterogeneity of urban characteristics on human mobility. It contributes to transportation, urban planning and public safety domains by achieving accurate prediction of crowd OD flow patterns and urban travel demand assessment from satellite imagery using advanced AI approaches. In practical applications, our trained model can predict the missing OD data within a city as long as the satellite imagery is available or transferring the model to other cities to predict their OD flows without historic mobility data.

Building on the good performance of state-of-the-art OD flow prediction models, a good spatiotemporal transferability further enhances its practical significance. In the revised manuscript, Section 2.2 provides the performance details of the model on the test dataset within the same city, while Section 2.5 presents the model’s transferability when the entire dataset from unseen time periods or cities is used as testing during cross-time or cross-city transfer. Regarding temporal transferability, the models trained based on existing survey data can be applied to periods where survey data is not available, helping to reveal the patterns of urban population flow in diverse time periods. At the same time, the spatial transferability of our model ensures that it can be easily adapted to new, previously unseen urban environments, solving the bottleneck issue in data acquisition, making it a valuable tool for real-world applications. (see **Section 3 Paragraph 4, [Page 14-15]**)

We have also enhanced the discussion to address your concerns.

Comment 2: About the problem definition. The authors define the OD flow as a graph, not dynamic graph with respect to time. To me, OD flow between regions should vary from day to day, time to time, thus the problem definition is not so sound to me.

Response: Thank you for highlighting the dynamic nature of OD flow. We agree with you that population mobility in real-life scenarios is affected by temporal characteristics such as holidays and weather, and the study of dynamic OD flow can offer critical insights on traffic analysis. This is not the focus of our work. While daily flow volumes within regions fluctuate dynamically, there exists a consistent spatial pattern in the OD flows. Specifically, the variation in the number of people traveling between regions generally follows a similar trend over a long period (e.g., yearly). In designing our study, we focused primarily on how the spatial heterogeneity of geographical contexts affects human mobility spatial patterns [1], which can be more effectively captured using static OD flows.

As noted in the discussion section, “Different from forecasting research on time series, the *Imagery2Flow* focuses on the exogenous mechanism of geographic contexts on OD flow patterns, compared to mechanistic mobility models (e.g., the gravity model and the radiation model) and deep-learning models” (Section 3 Paragraph 1, [Page 14]), we concentrate our research exclusively on static OD flows, represented by the commuting flows. Additional encoding of daily fluctuations may obscure or complicate the identification of the key drivers for travel demand estimation in urban environments.

Although the OD flow modeling approach in this work is static (similar to the Deep Gravity Model [7]), the insights into human mobility studies are still critical. The proposed framework can serve as a foundation for future advanced models involving temporal or sequential elements. Satellite images themselves have captured time information and can reflect quarterly, annually, or long-term land changes on the earth, such as the proliferation and densification of buildings and roads, as well as the expansion of urban settlements. The long-term patterns are captured in the temporal transferability of our *Imagery2Flow* framework, which can be applied to remote sensing imagery from 10 or 20 years ago to infer the human mobility flows with a good accuracy. As a result, our proposed model has learned the underlying relationship between complex geographic contexts in satellite images and human mobility flows (Section 2.5 Paragraph 5, [Page 13]).

We acknowledge that short-term variability and cyclical effects of human mobility are noteworthy, though they were not the focus of this research. We have included a discussion of our limitation and potential extensions in future work (Section 3 the last Paragraph, [Page 15]).

Comment 3: Following comment 2, there are many related works about OD flow generation/prediction in the research area of computer science/AI that modes OD flows as a dynamic graph problem (i.e., $T \times N \times N$, N is the number of regions and T are timestamps). More comprehensive literature review on OD flow generation/prediction should be conducted.

Response: Thanks very much for your suggestion. We agree that a more comprehensive review of OD flow generation/prediction is indeed necessary. In our manuscript, we categorized the models into traditional mechanism models and machine learning (and deep learning models), and then further reviewed the data applied to OD generation/prediction tasks.

Existing research shows that graph neural networks (GNNs) have a natural applicability for spatial interaction modeling [3, 4] such as OD flows. For the graph construction of static OD flows, the nodes usually represent the geographic units, and the edges can be constructed based on geographical adjacency [5], transport connectivity [4, 6], historical OD flows [6, 8], distance or other types of spatial relations [10, 11]. For the dynamic OD flows, Spatial-temporal graph neural networks (STGNNs) [3, 12] have been applied to consider spatial and temporal dependencies to update node features when the topological structure of the network remains unchanged. When the nodes and edges of the network change over time, dynamic graph neural networks [14–16] captures the network’s topological evolution and simulates information interactions. Our study focuses on static OD flow prediction (i.e., travel demand estimation) from urban spatial contexts.

We have further added literature reviews about machine learning and deep learning for modeling OD flows, including time series problems and dynamic graph solutions, to explain our approach. Please see Section 1 Paragraph 2 [Page 2] in our revised manuscript.

Comment 4: About the position and the core contribution. To me, this is a deep learning approach paper, so the main focus should be on the effectiveness and the accuracy of the proposed method. Calling back to 3, as an approach paper, more baselines should be included besides the basic Radiation model and Gravity model.

Response: Thanks for your advice. We have expanded our baselines by including a wider range of models, including traditional physical/mechanistic models, machine learning models and deep

learning models. The Radiation model and Gravity model in the original manuscript are widely used traditional physical/mechanistic models. Since OD flow prediction is fundamentally a regression task, we have also incorporated two ensemble tree models [17], Extreme Gradient Boosting (XGBoost) [18] and Light Gradient Boosting Machine (LGBM) [19], for performance comparison. Regarding deep learning models, SIGCN [8] builds the graph based on known flows and uses GCN to extract spatial features. GMEL [5] uses geo-adjacency graph and multi-task learning strategy to extract geo-contextual embedding, then a gradient boosting regressor is used to predict the flow volume. The Deep Gravity model in the original manuscript concatenates the features of origin and destination and distance, then feed them into fully connected layers to model OD flow intensity. Please see **Section 4.4 [Page 18]** for the detailed information of the baselines.

To compare the effectiveness of our proposed model, the input features of machine learning and deep learning baselines are consistent with *Imagery2Flow*, that is, the visual features extracted from satellite images. The performance was notably enhanced when the original bilinear decoder was replaced by the LGBM regressor. Therefore, *Imagery2Flow-LGBM* is also included in our analysis as a variant of the originally proposed *Imagery2Flow* based on the bilinear decoder. The input features of XGBoost and LGBM are the same as the LGBM decoder in *Imagery2Flow-LGBM*, which is the concatenation of visual features and geographic distances between origin and destination. Please see Table 1 (**Section 1 [Page 7]**) for a comprehensive comparison of model performance. Our proposed *Imagery2Flow-LGBM* model achieves the best performance, by leveraging the graph structure to learn spatial interaction patterns and utilizing a gradient boosting regressor to precisely model the OD distribution based on input visual features from satellite imagery. We also added discussion of the differences between baselines to better understand the performance gap (**Section 2.2 Paragraph 3 [Page 5]**).

Additionally, to assess the effectiveness of each module in our *Imagery2Flow* framework, we added an ablation experiment. The results are displayed in the Table 1 and Table 2 below and also in the revised manuscript **Section 2.2 Paragraph 2 [Page 5]** and **Appendix D Table D3-D4 [Page S-16]**. We first replaced the visual features with numerical socioeconomic features that are widely used in current human mobility research, demonstrating the effectiveness of satellite imagery features and their ability to reflect human activities. Second, the BMC Loss is replaced by the conventional MSE Loss. The experiment results demonstrate the effectiveness of our model design. Without using a regressor as decoder, our original *Imagery2Flow* with bilinear decoder performs the best; while using a regressor as decoder, the variant *Imagery2Flow-LGBM* performs the best in terms of having smallest RMSE/MAE and largest CPC metric values.

Table 1: The socioeconomic features collected from OpenStreetMap.

Categories	Features	Contents
Land use areas	5	The total area (in km ²) for each land-use class, including residential, commercial, industrial, retail, and natural areas.
Road network lengths	3	The total length (in km) for each type of roads including residential, main and other.
Facility counts	10	The total count of POIs and buildings for each possible category of transport, food, health, education, retail facilities.

Table 2: The Ablation experiment on M1 using 10-m satellite imagery.

	RMSE	MAE	CPC
Imagery2Flow	<i>16.47</i>	<i>8.97</i>	<i>0.79</i>
POI2Flow	25.62	11.51	0.74
Imagery2Flow-MSELoss	18.56	10.58	0.76
Imagery2Flow-LGBM	13.44	7.36	0.83
POI2Flow-LGBM	13.51	7.42	0.83
Imagery2Flow-MSELoss-LGBM	13.60	7.42	0.83

Bold indicates the best performance of LGBM decoder, while *italics* denotes the best performance of bilinear decoder.

Comment 5: Following 4, the impact factors on city OD flow analyzed in Section “2.3 Correlation to Urban Morphology” and “2.4 Spatial Heterogeneity” is based on the predicted OD flow, which is not necessary contents. If the authors would like to claim their proposed approach is accurate, they can just show the performances on many metrics. If the authors would like to reveal the insights on the impact factors on OD flows, the analysis can be done based on the ground-truth OD flow data.

Response: Thanks for your comment. We apologize for any confusion caused. Section 2.3 Correlation to Urban Morphology explores the relationship between population mobility and urban morphology based on the commuting flows. The parameter λ in the exponential law distribution of flows with distance decay in each MSA is calculated based on the ground-truth OD flow data. Since the Kullback-Leibler divergence and Jensen’s Shannon divergence of the actual flows and the predictions are very small (see **Appendix B Table B1 [Page S-6]**), thus the distributions of the predicted ones and the ground-truth are very similar (Fig. 3). Therefore, we hope to demonstrate that the OD flow learned by satellite imagery can also capture the information of urban form and spatial structure that is comparable to the actual flow. Our method can help urban planners connect satellite image to urban morphology from the perspective of human mobility, which wasn’t considered before in the literature as most of existing works focused on land cover/land use/population mapping. We refined our statements in **Section 2.3 Paragraph 2 [Page 8]** and the last Paragraph **[Page 10]**.

Upon analyzing the results, we observed that the prediction accuracy of flows varies across different locations, yet these differences are not adequately represented by the metrics in Section 2.2. After further attribution, we found that these differences were related to socioeconomics of residents in different neighborhoods, land use and vegetation, so Section 2.4 Spatial Heterogeneity was added to analyze the spatial distribution differences of model errors based on the predicted OD flows. Moreover, Section 2.4 reveals the limitations of remote sensing images in mobility modeling and identifies potential factors, while also providing insights for future improvements. We added the statements in **Section 2.4 Paragraph 1 [Page 10]**.

Comment 6: The above two points make me feel the contribution is somehow blurry and weak.

Response: We hope our new experiments and clarifications in addressing above two points make our contribution more clear now. From the comparison to the performance of added baselines (Table 1 in **Section 1 [Page 7]**), the effectiveness of the architecture of *Imagery2Flow* and its variant *Imagery2Flow-LGBM* has been verified. The novel idea of learning human mobility OD flows purely from satellite imagery and the effectiveness of each module has been validated with our added ablation experiment (**Appendix D Table D4 [Page S-16]**). Compared with the socioeconomic survey data that are hard to get across different regions, the medium-resolution remote sensing images have a global coverage and openly accessible, which has a great potential for a global study in future, especially for regions without annual survey or statistical data. The updated remote sensing images can be employed to estimate the latest population mobility information using our pre-trained *Imagery2Flow* model.

Through the analysis of urban morphology and spatial variation in prediction, we gained deeper insights into how various urban characteristics influence human movements, which helps refine urban planning strategies for future development (as discussed in **Section 3 Paragraph 3 [Page 14]**).

Comment 7: Basic definitions about the spatial unit should be written in a clearer way. “Each region of interest (ROI) is partitioned into N spatial areas according to their administrative boundaries (e.g., census tracts)”, so what is each area among N?

Response: We apologize for the confusing statement. We have revised the definition as “Each region of interest (ROI) (i.e., study area) is partitioned into N smaller geographic units, whose boundaries can be defined by administrative divisions (e.g., census tracts) or square grids. In our study, a ROI is a metropolitan statistical area and the census tracts are the smaller geographic units for predicting OD flows. That is, we have N census tracts in each ROI.” (**Section 2.1 Paragraph 1 [Page 3]**) Alternatively, one may choose square grids as the spatial units to evenly partition each ROI.

Comment 8: Some typos exist including but not limited to:

Section 2.1, 1st paragraph, Given $\rightarrow$ given.

Section 2.3, While higher population $\rightarrow$ while

Section 4.3, the indent after Eq.4 and Eq.5 before where should be deleted.

Response: Thanks for pointing out the typos. We have carefully reviewed the entire manuscript and corrected all the identified typos.

Thank you very much for all your constructive comments that help us improve the quality and clarity of our manuscript.

References

- [1] Can Rong, Jingtao Ding, and Yong Li. An Interdisciplinary Survey on Origin-destination Flows Modeling: Theory and Techniques. *ACM Comput. Surv.*, 57(1):4:1–4:49, October 2024.
- [2] Filippo Simini, Gianni Barlacchi, Massimiliano Luca, and Luca Pappalardo. A Deep Gravity model for mobility flows generation. *Nature Communications*, 12(1):6576, November 2021.
- [3] Di Zhu, Yu Liu, Xin Yao, and Manfred M. Fischer. Spatial regression graph convolutional neural networks: A deep learning paradigm for spatial multivariate distributions. *GeoInformatica*, 26(4):645–676, October 2022.
- [4] Yuebing Liang, Fangyi Ding, Guan Huang, and Zhan Zhao. Deep trip generation with graph neural networks for bike sharing system expansion. *Transportation Research Part C: Emerging Technologies*, 154:104241, 2023.
- [5] Zhicheng Liu, Fabio Miranda, Weiting Xiong, Junyan Yang, Qiao Wang, and Claudio Silva. Learning Geo-Contextual Embeddings for Commuting Flow Prediction. *Proceedings of the AAAI Conference on Artificial Intelligence*, 34(01):808–816, April 2020.
- [6] Chen Fu, Zhou Huang, Bronte Scheuer, Jiayuan Lin, and Yi Zhang. Integration of dockless bike-sharing and metro: Prediction and explanation at origin-destination level. *Sustainable Cities and Society*, 99:104906, 2023.
- [7] Shayan Nejadshamsi, Jamal Bentahar, Ursula Eicker, Chun Wang, and Faezeh Jamshidi. A geographic-semantic context-aware urban commuting flow prediction model using graph neural network. *Expert Systems with Applications*, 261:125534, 2025.
- [8] Xin Yao, Yong Gao, Di Zhu, Ed Manley, Jiaoe Wang, and Yu Liu. Spatial Origin-Destination Flow Imputation Using Graph Convolutional Networks. *IEEE Transactions on Intelligent Transportation Systems*, 22(12):7474–7484, December 2021.
- [9] Yibo Zhao, Shifen Cheng, Song Gao, Peixiao Wang, and Feng Lu. Predicting origin-destination flows by considering heterogeneous mobility patterns. *Sustainable Cities and Society*, 118:106015, 2025.
- [10] Yuxia Wang, Xin Yao, Yu Liu, and Xia Li. Generating population migration flow data from inter-regional relations using graph convolutional network. *International Journal of Applied Earth Observation and Geoinformation*, 118:103238, 2023.
- [11] Junlei Yuan, Yuxin Zhao, Disheng Yi, Sijia Jin, Huijun Zhou, and Jing Zhang. Exploring multi-relational spatial interaction imputation with distance-decay effects. *International Journal of Digital Earth*, 17(1):2300316, December 2024.
- [12] Jun Yang, Xiao Han, Tan Ye, Yinghao Tang, Weidong Feng, Aili Wang, Huijun Zuo, and Qiang Zhang. Spatiotemporal virtual graph convolution network for key origin-destination flow prediction in metro system. *Mathematical Problems in Engineering*, 2022(1):5622913, 2022.
- [13] Qingli Shi, Li Zhuo, Haiyan Tao, and Junying Yang. A fusion model of temporal graph attention network and machine learning for inferring commuting flow from human activity intensity dynamics. *International Journal of Applied Earth Observation and Geoinformation*, 126:103610, 2024.

- [14] Yixuan Yang, Shiyao Zhang, Chenhan Zhang, and James J.Q. Yu. Origin-destination matrix prediction via hexagon-based generated graph. In *2021 IEEE International Intelligent Transportation Systems Conference (ITSC)*, page 1399–1404, 2021.
- [15] Ziheng Huang, Weihao Zhang, Dujuan Wang, and Yunqiang Yin. A gan framework-based dynamic multi-graph convolutional network for origin-destination-based ride-hailing demand prediction. *Information Sciences*, 601:129–146, 2022.
- [16] Dapeng Zhang and Feng Xiao. Dynamic Auto-Structuring Graph Neural Network: A Joint Learning Framework for Origin-Destination Demand Prediction. *IEEE Transactions on Knowledge and Data Engineering*, 35(4):3699–3711, April 2023. Conference Name: IEEE Transactions on Knowledge and Data Engineering.
- [17] Leo Grinsztajn, Edouard Oyallon, and Gael Varoquaux. Why do tree-based models still outperform deep learning on typical tabular data? In *Advances in Neural Information Processing Systems*, volume 35, pages 507–520, 2022.
- [18] Tianqi Chen and Carlos Guestrin. Xgboost: A scalable tree boosting system. In *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, KDD ’16, page 785–794, New York, NY, USA, 2016. Association for Computing Machinery.
- [19] Guolin Ke, Qi Meng, Thomas Finley, Taifeng Wang, Wei Chen, Weidong Ma, Qiwei Ye, and Tie-Yan Liu. Lightgbm: A highly efficient gradient boosting decision tree. In *Advances in Neural Information Processing Systems*, volume 30 of *NIPS’17*, page 3149–3157, 2017.

Response to Reviewer 2

General Comment: In this manuscript, the authors proposed Imagery2Flow, a methodology that leverages satellite imagery to predict commute flows. Visual representations derived from satellite imagery are used to extract geospatial features, which are subsequently integrated into a Graph Attention Network to effectively model and predict commute flows. The proposed approach is validated using commute flow data from the top 10 largest Metropolitan Statistical Areas (MSAs) in the United States. Notably, the transferability tests demonstrate the model’s cross-city generalization, and the study explores how urban factors (e.g., centrality and compactness) influence human mobility, providing insights into the interplay between urban morphology and movement patterns. While the study presents an innovative model for commute flow prediction, improvements are needed in the clarity of methods and scope of experiments.

Response: Thank you very much for acknowledging the novelty of our model and sharing your valuable feedback on our manuscript. Your primary concerns are the adequacy of our experiments and the clarity of analyses and discussions, particularly how satellite imagery is superior to other data sources in human mobility modeling. We therefore conducted additional baseline models comparison and ablation experiments to further demonstrate the effectiveness and generalizability of our proposed *Imagery2Flow* framework. In addition, inspired by your comments, we replaced the bilinear decoder with a regressor and proposed a variant *Image2Flow-LGBM* that follows the design idea of separating the spatial interaction learner and the OD flow predictor, which further improved the model performance. The experiment results demonstrate the effectiveness of our model design. For models without using a regressor as decoder, our original *Imagery2Flow* with bilinear decoder performs the best; while using a regressor as decoder, the variant *Imagery2Flow-LGBM* performs the best in terms of having smallest RMSE/MAE and largest CPC metric values.

We have carefully addressed each of your comments and revised the manuscript accordingly. We much appreciate your constructive comments that help us further improve the quality and clarity of our manuscript.

Major comments:

Major Comment 1: How can the proposed model account for the temporal characteristics of human mobility such as prediction of OD flows over different time periods (e.g., morning commute or weekend activity)? Considering temporal variations can demonstrate the practical applicability and broader generalizability of the proposed model.

Response: Thank you for highlighting the dynamic nature of OD flow data. We agree with you that population mobility in real-life scenarios is affected by temporal characteristics such as time of a day or weekday versus weekends, and the study of dynamic OD flow can offer critical insights on traffic analysis. However, similar to the Deep Gravity Model [7], the temporal aspect is not the focus of our work. While hourly/daily flow volumes within regions fluctuate dynamically, there exists a consistent spatial pattern in the OD flows regarding the overall travel demands. Specifically, the variation in the number of people traveling between regions generally follows a similar trend over a long period (e.g., yearly). In designing our study, we focused primarily on how the spatial heterogeneity of geographical contexts affects human mobility spatial patterns [1], which can be more effectively captured using static OD flows.

As noted in the discussion section, “Different from forecasting research on time series, the *Imagery2Flow* focuses on the exogenous mechanism of geographic contexts on OD flow patterns, compared to mechanistic mobility models (e.g., the gravity model and the radiation model) and deep-learning models” (Section 3 Paragraph 1, [Page 14]), we concentrate our research exclusively on static OD flows, represented by the commuting flows. Additional encoding of daily fluctuations may obscure or complicate the identification of the key drivers for travel demand estimation in urban environments.

Although the OD flow modeling approach in this work is static (similar to the Deep Gravity Model [7]), the insights into human mobility studies are still critical. The proposed framework can serve as a foundation for future advanced models involving temporal or sequential elements. Satellite images themselves have captured time information and can reflect quarterly, annually, or long-term land changes on the earth, such as the proliferation and densification of buildings and roads, as

well as the expansion of urban settlements. The long-term patterns are captured in the temporal transferability of our *Imagery2Flow* framework, which can be applied to remote sensing imagery from 10 or 20 years ago to infer the human mobility flows with a good accuracy. As a result, our proposed model has learned the underlying relationship between complex geographic contexts in satellite images and human mobility flows (**Section 2.5 Paragraph 5, [Page 13]**).

We acknowledge that short-term variables and self-cyclical effects of human mobility are noteworthy, though they were not the focus of this work. Different data sources provide varying information, and the insights derived from them need to be tailored to meet specific expectations. For example, social sensors such as smartphones and georeferenced social media posts contain fine-grained time, location, and human activity information, making them naturally more well-suited for temporal tracking and dynamic mobility modeling, compared to the satellite imagery. We have included a discussion of our limitation and potential extensions in future work (**Section 3 the last Paragraph, [Page 15]**).

Major Comment 2: The authors utilized ResNet-50 for visual feature extraction. To demonstrate the generalization of the proposed framework, it would be beneficial to include experimental results using various backbone CNN models.

Response: Thanks for your valuable advice. We have expanded our experiments to include not only CNN-based models but also Vision Transformers (ViTs) as visual feature extractor for a more comprehensive comparison. The backbone comparison results are shown below and in the revised manuscript **Appendix D Table D5**, suggesting that ViT-B/32 and ViT-L/16 transformers perform slightly better than the CNN backbones. Therefore, we reconducted all the experiments, utilized the ViT-L/16 for subsequent transferability experiments, and updated all the experimental results accordingly. The updated contents were added to the revised manuscript in **Section 2.2 Paragraph 2, [Page 5] and Appendix D [Page S-16]**.

Additionally, your subsequent comments inspired us to design models with different decoders for performance improvements. We have added all the relevant descriptions in **Section 2.1 Paragraph 4, [Page 4]**, and further presented the updated *Imagery2flow* with ViT-L/16 backbone and its variant (*Imagery2flow-LGBM*) with a regressor as the decoder in the Results and Discussion sections (see Table 1 [**Page 7**], Fig. 2 [**Page 8**]). Therefore, the table below (same as **Appendix D Table D5**) shows the effectiveness of ViT backbones under different decoders (Bilinear vs. LGBM).

We really appreciate your insightful comments that improve our modeling framework. It also demonstrates the flexibility of our proposed *Imagery2Flow* as other scholars may replace one or more of the three modules (i.e., visual feature extraction/spatial context encoder, spatial interaction learner, and OD flow predictor/regressor) with more advanced ones in future.

Table 3: Performance comparison of using various backbones for the satellite image feature extraction in M1.

Backbone	Bilinear			LGBM		
	RMSE	MAE	CPC	RMSE	MAE	CPC
10-m						
ResNet-18	17.02	9.57	0.79	13.77	7.49	0.83
ResNet-50	16.98	9.42	0.79	13.79	7.53	0.83
ResNeXt-50	16.85	9.27	0.79	13.68	7.48	0.83
EfficientNet-v2-m	17.32	9.62	0.78	14.27	7.72	0.82
ViT-B/32	16.66	9.12	0.79	13.64	7.44	0.83
ViT-L/16	16.47	8.97	0.79	13.44	7.36	0.83
30-m						
ResNet-18	17.19	9.61	0.78	13.98	7.61	0.83
ResNet-50	17.34	9.66	0.78	14.18	7.78	0.82
ResNeXt-50	17.16	9.45	0.78	14.01	7.65	0.83
EfficientNet-v2-m	17.70	9.95	0.77	14.35	7.80	0.82
ViT-B/32	16.65	9.16	0.79	13.78	7.53	0.83
ViT-L/16	16.66	9.28	0.79	13.72	7.49	0.83

Major Comment 3: Can the proposed methodology be applicable at different levels of spatial granularity (e.g., census block groups, ZIP Code Tabulation Areas)? Depending on the spatial scale

of the region-of-interest (ROI), multiple satellite images may be included. In such cases, how is the representation of the ROI handled? Clarifying this would help assess the versatility and scalability of the approach.

Response: Thanks for this comment. In our research, given the ground-truth data availability, we conducted experiments at the Census Tract level of spatial granularity. However, our *Imagery2Flow* framework can be applied at different spatial scales. After downloading the satellite imagery for each region of interest using Google Earth Engine, the whole imagery for the study area was clipped by the census tract boundaries and resized into 224*224 in the data augmentation step with random crop transformation. When the spatial unit increases to the scale of ZIP Code Tabulation Areas (ZCTA) or even counties, directly resizing the clipped satellite image may result in loss of information. In this case, each ZCTA may be initially divided into smaller square grids and the remote sensing images are then cropped according to the grids to obtain smaller satellite image patches. The visual features extracted from those image patches within grids can be fused to form the visual features at the ZCTA level. Common methods for feature fusion include summarizing or averaging the visual features of satellite image patches or passing them through multiple multilayer perceptron layers or more advanced feature fusion neural networks [1].

We have added the updated contents in **Section 4.2 Paragraph 3 [Page 16]**.

Major Comment 4: Clarify the nature of the node embeddings utilized in the GAT module. Specifically, are the embeddings derived from satellite imagery used as the graph’s node embeddings? A more detailed explanation would greatly help in understanding this aspect of the methodology. If the satellite imagery embeddings are indeed employed as node embeddings, it would suggest that the study enhances commute flow prediction performance by leveraging visual representations as auxiliary information.

Response: Thanks very much for your thoughtful review and clear understanding of our proposed method. In the initial graph structure, nodes are spatial units (i.e., census tracts), and edges are the spatial relations between nodes (e.g., geographic adjacency and distance). High-dimensional visual feature embeddings extracted from remote sensing images are used as initial features of nodes, which offer important information about the spatial contexts of each geographic unit. Following the GAT encoding, the node representations are updated by incorporating structural knowledge and neighborhood information through graph representation learning [2], facilitated by message passing and aggregation mechanisms. The updated node embeddings are then used as the feature vector of each spatial unit and fed into the decoder for OD flow prediction [3–6].

In addition, inspired by your comments for enhancing the flow prediction performance, we replaced the bilinear decoder with a regressor and proposed a variant *Image2Flow-LGBM* that follows the design idea of separating the spatial interaction learner and the OD flow predictor, which further improved the model performance. The experiment results demonstrate the effectiveness of our model design. For models without using a regressor as decoder, our original *Imagery2Flow* with bilinear decoder performs the best; while using a regressor as decoder, the variant *Imagery2Flow-LGBM* performs the best in terms of having smallest RMSE/MAE and largest CPC metric values.

We have updated the relevant statements in the revised manuscript **Section 2.1 Paragraph 3 [Page 3]** and added detailed analysis of the differences between models and presented these findings in conjunction with a performance comparison in **Section 2.2 Paragraph 3 [Page 5]**.

Major Comment 5: Some recent works have enhanced commute flow prediction accuracy using GNNs by incorporating node features derived from external data sources [1-4], such as OpenStreetMap (OSM) data or Primary Land Use Tax Lot Output (PLUTO) datasets. Does the use of visual representations from satellite imagery provide unique advantages over these conventional data sources? It would be beneficial to include a performance comparison between the approach using satellite imagery and those utilizing OSM or PLUTO datasets.

Response: Thanks for your advice and suggestion of all the relevant literature. We have not only added them into our literature review but also conducted a new ablation experiment on different input data to understand the effectiveness of our modeling framework. Since PLUTO only covers the New York City, we referenced the Deep Gravity work [7] to add the OpenStreetMap (OSM) data across ten cities into our new experiments:

Table 4: The socioeconomic features collected from OpenStreetMap.

Categories	Features	Contents
Land use areas	5	The total area (in km ²) for each land-use class, including residential, commercial, industrial, retail, and natural areas.
Road network lengths	3	The total length (in km) for each type of roads including residential, main and other.
Facility counts	10	The total count of POIs and buildings for each possible category of transport, food, health, education, retail facilities.

Each spatial unit (i.e., census tract) has a socioeconomic numerical vector with a dimension of 18, serving as node feature instead of the visual features extracted from satellite images. The results are as follows (in Response to **Comment 6**). The results show that the visual features of the Sentinel-2 satellite imagery have better performance than the POI data, no matter whether the decoder is bilinear layer or Light GBM. Additionally, compared to the OSM data, remote sensing images have a better data quality, a global coverage, and clear timestamp information, which can ensure that the model has a potential to infer human mobility OD flows globally when statistical/survey data is not available. But the performance would require ground-truth data for further validation in future work, which is always a big challenge to get such validation data at large scale. We have updated these contents into **Section 3 Paragraph 2 [Page 14]**.

Major Comment 6: An ablation study is necessary to better understand the contribution of each component in the proposed model. Additionally, what drives the performance differences between the proposed model and the baseline deep gravity model? Providing a detailed analysis of these differences would help clarify the strengths and unique advantages of the proposed approach.

Response: Thanks for your valuable comment. To assess the effectiveness of each module, we have added an ablation experiment as shown in the table below. The results are further discussed in the revised **Section 2.2 Paragraph 2 [Page 5]** and **Appendix D**. We first replaced the visual features with numerical socioeconomic features extracted from OSM POIs, denoted as POI2Flow, which demonstrated the effectiveness of satellite imagery visual features and their ability to infer human activities. Moreover, the BMC Loss utilized in the Imagery2Flow and Imagery2Flow-LGBM is replaced by the conventional MSE Loss, denoted as Imagery2Flow-MSELoss and Imagery2Flow-MSELoss-LGBM, respectively.

Table 5: The ablation experiment on M1 using 10-m satellite imagery.

	RMSE	MAE	CPC
Imagery2Flow	<i>16.47</i>	<i>8.97</i>	<i>0.79</i>
POI2Flow	25.62	11.51	0.74
Imagery2Flow-MSELoss	18.56	10.58	0.76
Imagery2Flow-LGBM	13.44	7.36	0.83
POI2Flow-LGBM	13.51	7.42	0.83
Imagery2Flow-MSELoss-LGBM	13.60	7.42	0.83

Bold indicates the best performance of LGBM decoder, while *italics* denotes the best performance of bilinear decoder.

Additionally, we expanded our baselines by including traditional physical/mechanistic approaches and machine learning/deep learning models. The Radiation model and Gravity model in the original manuscript are widely used traditional physical/mechanistic models. Since OD flow prediction is fundamentally a regression task, we have also incorporated two ensemble tree models [17], Extreme Gradient Boosting (XGBoost) [18] and Light Gradient Boosting Machine (LGBM) [19], for performance comparison. Regarding deep learning models, SIGCN [8] builds the graph based on known flows and uses GCN to extract spatial features. GMEL [5] uses geo-adjacency graph and multi-task learning strategy to extract geo-contextual embedding, then a gradient boosting regressor is used to predict the flow volume. The Deep Gravity model in the original manuscript concatenates the features of origin and destination and distance, then feed them into fully connected layers to model OD flow intensity. Please see the revised **Section 4.4 [Page 18]** for the detailed information of the baselines.

To compare the effectiveness of our proposed model, the input features of machine learning and deep learning baselines are consistent with *Imagery2Flow*, that is, the visual features extracted from satellite images. The performance was notably enhanced when the original bilinear decoder was replaced by the LGBM regressor. Therefore, *Imagery2Flow-LGBM* is also included in our analysis as a variant of the originally proposed *Imagery2Flow* based on the bilinear decoder. The input features of XGBoost and LGBM are the same as the LGBM decoder in *Imagery2Flow-LGBM*, which is the concatenation of visual features and geographic distances between origin and destination. Please see Table 1 (**Section 1 [Page 7]**) for a comprehensive comparison of model performance. The *Imagery2Flow-LGBM* model achieves the best performance, by leveraging the graph structure to learn spatial interaction patterns and utilizing a gradient boosting regressor to precisely model the OD distribution based on input visual features from satellite imagery. We also added discussion of the differences between baselines to better understand the performance gap (**Section 2.2 Paragraph 3 [Page 5]**).

Minor comments:

Minor Comment 1: In Section 2.3, clarify the description of BCSI values. The current statement “A smaller BCSI value indicates a more compact city, with a value of 1 representing a perfect circle.” is incorrect. A value of 1 generally represents irregular shape.

Response: We are sorry for this incorrect description and we have corrected the description in the revised manuscript. The Boyce-Clark Shape Index (BCSI) is used to quantify and analyze the shape of geographic features. According to the formula $BCSI = \sum |\frac{r_i}{\sum_{i=1}^n r_i} \times 100 - \frac{100}{n}|$, BCSI measures the irregularity of a shape based on the radial distances from its centroid. The n rays are projected from the shape’s centroid at equal angular intervals to its boundary. The length of each ray r_i is measured from the centroid to the intersection point on the shape’s edge. A circle will have all r_i values equal to its radius, resulting in a BCSI of 0, indicating a highly regular shape.

We revised and added descriptions to BCSI in **Section 2.3 Paragraph 3 [Page 8-9]** as below:
 “The Boyce-Clark Shape Index (BCSI) [58] is adopted to quantify the shape irregularity of cities, by projecting n rays from the shape’s centroid at equal angular intervals to its boundary and measuring the variation in radial distances. The formula for BCSI is computed by $BCSI = \sum |\frac{r_i}{\sum_{i=1}^n r_i} \times 100 - \frac{100}{n}|$. As the radial distances r_i become more consistent, the BCSI value decreases, indicating a more compact shape. A BCSI value of 0 represents a perfect circle, which has equal radial lengths from the centroid to the boundary, while higher BCSI values represent more irregular shape.”

Minor Comment 2: In Appendix C, the visibility of the circles used to illustrate the heterogeneous effect needs to be improved.

Response: Thanks for pointing out the visibility issue in Appendix C. The original figures utilized different colors to symbolize different factors that are prone to the spatial variation of the prediction errors. In the revised manuscript, we use the letters ‘A’, ‘B’ and ‘C’ to label the circles. Meanwhile, we have adjusted the color of circles into black to enhance the distinction between background map elements. The revised figures have improved the visibility.

Thank you very much for all your constructive comments that help us improve the quality and clarity of our manuscript.

References

- [1] Weiwei Song, Shutao Li, Leyuan Fang, and Ting Lu. Hyperspectral image classification with deep feature fusion network. *IEEE Transactions on Geoscience and Remote Sensing*, 56(6):3173–3184, 2018.
- [2] William L. Hamilton. Graph representation learning. *Synthesis Lectures on Artificial Intelligence and Machine Learning*, 14(3):1–159.
- [3] Qingli Shi, Li Zhuo, Haiyan Tao, and Junying Yang. A fusion model of temporal graph attention network and machine learning for inferring commuting flow from human activity intensity

- dynamics. *International Journal of Applied Earth Observation and Geoinformation*, 126:103610, 2024.
- [4] Shayan Nejadshamsi, Jamal Bentahar, Ursula Eicker, Chun Wang, and Faezeh Jamshidi. A geographic-semantic context-aware urban commuting flow prediction model using graph neural network. *Expert Systems with Applications*, 261:125534, 2025.
 - [5] Zhicheng Liu, Fabio Miranda, Weiting Xiong, Junyan Yang, Qiao Wang, and Claudio Silva. Learning Geo-Contextual Embeddings for Commuting Flow Prediction. *Proceedings of the AAAI Conference on Artificial Intelligence*, 34(01):808–816, April 2020.
 - [6] Yibo Zhao, Shifen Cheng, Song Gao, Peixiao Wang, and Feng Lu. Predicting origin-destination flows by considering heterogeneous mobility patterns. *Sustainable Cities and Society*, 118:106015, 2025.
 - [7] Filippo Simini, Gianni Barlacchi, Massimiliano Luca, and Luca Pappalardo. A Deep Gravity model for mobility flows generation. *Nature Communications*, 12(1):6576, November 2021.

Revision of “Imagery2Flow: Predicting Human Mobility Flows in Cities Using Deep Learning on Satellite Imagery”

Response to Reviewer 1

General Comment: I have checked the rebuttal letter and the revised manuscript, however, the authors’ rebuttal cannot address my concerns:

Comment 1: The significance of this research is unclear, the OD flow data is already available in 10 metropolitan areas. It is not so difficult to obtain, so why we need this approach.

Response: Thanks for this comment again. As a common practice in machine learning studies [1–4], we need to use some ground-truth data to validate our proposed model. Once it is trained, it can be applied to regions or timestamps with missing OD flow data or when such data are difficult to obtain. Our research addresses such a great challenge in many countries. If choosing other metropolitan areas or regions without existing OD flow data as ground truth, We couldn’t even train or validate any proposed models.

Comment 2: Defining the OD flow as a static one is also a huge limitation.

Response: As noted in the discussion section in the revised manuscript, “Different from forecasting research on time series, the *Imagery2Flow* focuses on the exogenous mechanism of geographic contexts on OD flow patterns learned from satellite imagery, compared to mechanistic mobility models (e.g., the gravity model and the radiation model) and deep-learning models” (Section 3 Paragraph 1, [Page 14]), we concentrate our research exclusively on static OD flows, represented by the commuting flows. Although the OD flow modeling approach in this work is static (similar to the Deep Gravity Model [2]), the insights into human mobility studies are still critical. The proposed framework can serve as a foundation for future models involving temporal or sequential elements with other additional data input (e.g., mobile phone location data but hard to access). Satellite images themselves have its own temporal resolution and may reflect monthly, quarterly, annually, or long-term land changes on the earth, such as the proliferation and densification of buildings and roads, as well as the expansion of urban settlements.

In addition, our research results have demonstrated that the long-term patterns are indeed captured in the temporal transferability of our *Imagery2Flow* framework, which has been applied to remote sensing imagery from 10 or 20 years ago to infer the human mobility flows with a good accuracy (Table 2). As a result, our proposed model has learned the underlying relationship between complex geographic contexts in satellite images and human mobility flows (Section 2.5 Paragraph 5, [Page 13]). We acknowledge that short-term variability and cyclical effects of human mobility are noteworthy, though they were not the focus of this research. Different data sources provide varying information, and the insights derived from them need to be tailored to meet specific expectations. We have included a discussion of our limitation and potential extensions in future work (Section 3 the last Paragraph, [Page 15]).

Comment 3: The novelty of the proposed methodology is kind of limited, which is just the simple combination of CNN and GNN.

Response: The contribution of our work lies in the novel idea of learning human mobility OD flows purely from satellite imagery and the effectiveness of each module has been validated with our added ablation experiment (Appendix D Table D4 [Page S-16]). Compared with the socioeconomic survey data (used in the other existing models) that are hard to get across different regions and countries, the medium-resolution remote sensing images have a global coverage and openly accessible, which has a great potential for a global study in future, especially for regions without annual survey or statistical data. The updated remote sensing images can be employed to estimate the latest population mobility information using our pre-trained *Imagery2Flow* model.

Comment 4: The evaluation of the proposed method is somehow not adequate. Recent baselines and potential better solutions are not adequately included.

Response: Thanks for your comment. We have made every effort to include existing and latest open-source models for comparison (Table 1). As mentioned in Section 1 paragraph 2, while there are more recent studies on OD flow prediction, they all rely on auxiliary data for model construction such as transport connectivity, flight and metro activities. In contrast, our model only utilizes the satellite imagery and geographical adjacency network.

References

1. Rong, C., Ding, J. & Li, Y. An Interdisciplinary Survey on Origin-destination Flows Modeling: Theory and Techniques. *ACM Comput. Surv.* **57**, 4:1–4:49. doi:[10.1145/3682058](https://doi.org/10.1145/3682058) (Oct. 2024).
2. Simini, F. *et al.* A Deep Gravity Model for Mobility Flows Generation. *Nature Communications* **12**, 6576. doi:[10.1038/s41467-021-26752-4](https://doi.org/10.1038/s41467-021-26752-4) (Nov. 2021).
3. Zhao, Y. *et al.* Predicting origin-destination flows by considering heterogeneous mobility patterns. *Sustainable Cities and Society* **118**, 106015. doi:<https://doi.org/10.1016/j.scs.2024.106015> (2025).
4. Yao, X. *et al.* Spatial Origin-Destination Flow Imputation Using Graph Convolutional Networks. *IEEE Transactions on Intelligent Transportation Systems* **22**, 7474–7484. doi:[10.1109/TITS.2020.3003310](https://doi.org/10.1109/TITS.2020.3003310) (Dec. 2021).

Response to Reviewer 2

General Comment: The authors have addressed several concerns effectively, including conducting ablation studies, demonstrating model generalizability using various backbones, and clarifying methodological details. In addition, the authors’ focus on exogenous geographic contexts in modeling static OD flow patterns is both valid and insightful, particularly for uncovering the structural determinants of mobility. While the study is well-executed and the results are clearly presented, I believe the technical novelty and distinction from prior work could be more articulated to fully highlight the contribution of the authors’ work.

Response: We much appreciate your valuable and constructive comments, which have significantly strengthened our paper. We have further provided a more detailed description of the technical details and contributions of our research, and explain the key distinctions from prior works more clearly.

Comment 1: While the application is valuable, the methodology appears to build upon well-established frameworks. The proposed model pipeline closely resembles prior studies that leverage satellite imagery or mobility data for urban region embedding [1–3], as well as GAT-based approaches for static commute flow prediction [4, 5]. In particular, the use of satellite image embeddings as node features in graph-based commute flow prediction models, in itself, may not represent a substantial methodological innovation.

Response: Thanks for your further comments. The contribution of [1, 2] lies in the use of satellite imagery to predict the spatial distributions of socioeconomic data (specifically related to poverty and wealth) but does not primarily lies in the advancement of the methodology itself to address the survey data scarcity issues that resonate with the core motivation of our work. In addition, while most previous studies have focused on predicting and mapping socioeconomic indicators such as population and poverty, our work demonstrates that satellite imagery can also predict human mobility flows across the geographic regions and reveal underlying patterns of spatial interaction, a novel perspective that has not been explored in prior research. This opens a new direction for leveraging remote sensing data in understanding human behavioral dynamics and geographic impacts.

Compared with other OD flow prediction studies using GAT approaches, we focus on building a transferable end-to-end model using publicly available remote sensing data to predict OD flows across a wide range of major cities. Meanwhile, the timely updated remote sensing images can be employed to estimate the latest or historic human mobility information using our pre-trained *Imagery2Flow* model. In contrast, other prior studies use known OD flows and other auxiliary data such as POIs and transportation infrastructure [3–5], to construct the model, which increases the difficulty of model transferability and limits the application scenarios. We have further elaborated the difference of these models and highlights the significance of our research in the in the Introduction section and Discussion section.

Comment 2: Additional details on the reproduction of baseline models are necessary. Since prior studies typically focused on OD flow prediction within a limited number of cities, the hyperparameters used in those studies may not be well suited for the broader or more diverse set of cities considered in this work. For a fair comparison, it is important that such baselines be re-tuned under the current experimental setting.

Response: Thank you for the suggestion. We agree that tuning the baselines under our current experimental setup is important for a fair comparison. Therefore, we have further conducted a grid search to identify the optimal parameters. We focused our hyperparameter tuning efforts on deep learning methods with visual features (i.e. Deep Gravity-V, SIGCN, GMEL), as they represent the most competitive baselines, enabling a more precise comparison with our proposed *Imagery2Flow* model. The hyperparameters, including the learning rate, hidden dimension, and number of GNN layers (if applicable), were determined via grid search. The search space was defined based on the values reported in the original papers and those used in our model. Training was performed until a convergence. The optimal hyperparameters are as follows: the learning rate is set to 1e-4 for DeepGravity-V and GMEL, and 1e-3 for SIGCN; the hidden dimensions are 512 for all three models; the number of GNN layers is 3 for GMEL and 2 for SIGCN (Section 4.4).

To ensure consistency and reduce overfitting to specific cities, we selected M1 with 10-m imagery as a representative subset of dataset for validation-based tuning. The parameter configuration that yielded the optimal RMSE on the validation set was then applied across all the geographic regions and spatial resolutions. This strategy was adopted for both our proposed Imagery2Flow and all the baselines.

The baseline results have been updated and presented in the updated Table 1. While parameter tuning led to improved performance of some baselines across cities, the overall conclusions drawn in Section 2.2 remain consistent and the Imagery2Flow-LGBM remains the best. It should be noted that although the performance improvement of our model over these three baselines may not appear substantial, the baselines are constructed and learned from known OD flow graphs, incorporating the total inflow/outflow volumes as constraints or using existed flows as prior graph structure (which is consistent with [4]). The design of baselines is hard to be transferred to flow data scarcity cities, which limits their applicability across different regions and countries.

Comment 3: It would be beneficial to include more recent and competitive baseline models.

Response: Thank you for the suggestion. We have made every effort to include latest relevant and open-source OD flow prediction models for comparison. The model in [5] has already been incorporated in our baseline comparison, while the model in [4] is not publicly available. Besides, as mentioned in Section 1 paragraph 2, while there are more recent studies on OD prediction, they all rely on auxiliary data for model construction such as transport connectivity, flight and metro activities. In contrast, our model only utilizes satellite imagery and the geographical adjacency network, which is different from these studies.

References

1. Jean, N. *et al.* Combining satellite imagery and machine learning to predict poverty. *Science* **353**, 790–794. doi:[10.1126/science.aaf7894](https://doi.org/10.1126/science.aaf7894) (Aug. 2016).
2. Yeh, C. *et al.* Using publicly available satellite imagery and deep learning to understand economic well-being in Africa. en. *Nature Communications* **11**. Publisher: Nature Publishing Group, 2583. doi:[10.1038/s41467-020-16185-w](https://doi.org/10.1038/s41467-020-16185-w) (May 2020).
3. Zhang, M. *et al.* Multi-view joint graph representation learning for urban region embedding in *Proceedings of the Twenty-Ninth International Joint Conference on Artificial Intelligence* (Yokohama, Yokohama, Japan, Jan. 2021), 4431–4437.
4. Zhao, Y. *et al.* Predicting origin-destination flows by considering heterogeneous mobility patterns. *Sustainable Cities and Society* **118**, 106015. doi:<https://doi.org/10.1016/j.scs.2024.106015> (2025).
5. Liu, Z. *et al.* Learning Geo-Contextual Embeddings for Commuting Flow Prediction. *Proceedings of the AAAI Conference on Artificial Intelligence* **34**, 808–816. doi:[10.1609/aaai.v34i01.5425](https://doi.org/10.1609/aaai.v34i01.5425) (Apr. 2020).

Final Revision of “Imagery2Flow: Predicting Human Mobility Flows in Cities Using Deep Learning on Satellite Imagery”

Response to Reviewer 2

Comments for Author: The authors have effectively addressed several concerns, particularly by tuning the baseline and incorporating recent models as baselines, which significantly strengthens the validity of the study. However, one concern still remains regarding technical novelty. The authors utilized features extracted from satellite imagery as region embeddings for the nodes. While this is a reasonable design choice, the methodological novelty of this component seems limited. Specifically, the region embeddings are obtained via a vanilla contrastive learning approach, and it is unclear how compelling or innovative this representation strategy is within the proposed framework.

Nevertheless, the study provides valuable insights by demonstrating how graph-based modeling with satellite imagery can be effectively applied to uncover the structural determinants of human mobility. Furthermore, it shows the feasibility of building a transferable end-to-end model for static OD flow prediction using publicly accessible remote sensing data.

Response: Thanks for your comment. While the proposed approach builds on state-of-the-art techniques including contrastive learning, the key innovation of our work lies in a transferable end-to-end deep learning model for timely and cost-effective prediction of human mobility patterns by leveraging publicly available satellite imagery. This approach enables us to understand the impacts of urban environments on human mobility flows from up-to-date earth observation satellite. With regard to our region embedding approach, due to its good performance and wide applicability, we utilize the SimCLR contrastive learning method to learn image embeddings which support various augmented views of the satellite imagery from the same geographic area. To test this representation strategy, we replaced the visual features extracted from satellite imagery with numerical socioeconomic features extracted from OpenStreetMap points of interest (POI) data. The results in Supplementary Table S4 showed that the visual features extracted from Sentinel-2 satellite imagery via SimCLR have a better performance than the POI data, no matter whether the decoder is bilinear layer or Light GBM. It demonstrated that a vanilla contrastive learning approach was sufficient.

Comments for Author: The provided README file was clear and well-structured, which made it straightforward to reproduce and utilize the authors’ implementation and results.

Response: We thank the reviewer for reviewing our code repository again and providing positive feedback.