Data-Driven City Traffic Planning Simulation

Tam V. Nguyen*
Dept. of Computer Science
University of Dayton

Thanh Ngoc-Dat Tran†
University of Science
Vietnam National University
Ho Chi Minh City

Viet-Tham Huynh‡
University of Science
Vietnam National University
Ho Chi Minh City

Bao Truong§
Dept. of Computer Science
University of Dayton

Minh-Quan Le¶
University of Science
Vietnam National University
Ho Chi Minh City

Mohit Kumavat‖
Dept. of Computer Science
University of Dayton

Vatsa S. Patel**
Dept. of Computer Science
University of Dayton

Mai-Khiem Tran††
University of Science
John von Neumann Institute
Vietnam National University
Ho Chi Minh City

Minh-Triet Tran‡‡
University of Science
John von Neumann Institute
Vietnam National University
Ho Chi Minh City

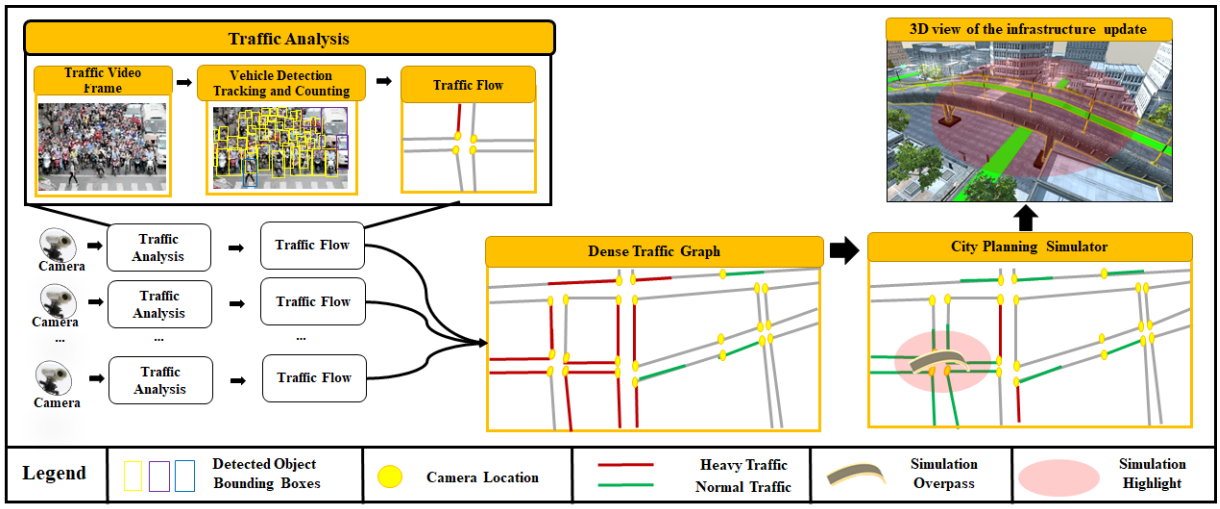


Figure 1: The overview of our proposed system. There are two main components in the proposed system, namely, traffic analysis, and city planning simulation.

Abstract

Big cities are well-known for their traffic congestion and high density of vehicles such as cars, buses, trucks, and even a swarm of motorbikes that overwhelm city streets. Large-scale development projects have exacerbated urban conditions, making traffic congestion more severe. In this paper, we proposed a data-driven city traffic planning simulator. In particular, we make use of the city camera system for traffic analysis. It seeks to recognize the traffic vehicles and traffic flows, with reduced intervention from monitoring staff. Then, we develop a city traffic planning simulator upon the analyzed traffic data. The simulator is used to support metropolitan transportation planning. Our experimental findings address traffic planning challenges and the innovative technical solutions needed to solve them in big cities.

Keywords: Data-driven, computer vision, simulation, user experience, evaluation

Index Terms: Human-centered computing—Visualization—Visualization techniques; Human-centered computing—Visualization—Visualization design and evaluation methods

*e-mail: tamnguyen@udayton.edu
†e-mail: tndthanh@selab.hcmus.edu.vn
‡e-mail: hvtham@selab.hcmus.edu.vn
§e-mail: truongb1@udayton.edu
¶e-mail: lmquan@selab.hcmus.edu.vn
‖e-mail: kumavatm1@udayton.edu
**e-mail: patelv20@udayton.edu
††e-mail: tmkhiem@selab.hcmus.edu.vn
‡‡e-mail: tmtriet@fit.hcmus.edu.vn

1 Introduction

The growing population and its centralization in metropolitan areas are among some factors creating a serious issue of traffic congestion in big and growing cities. Large-scale urban development projects have exacerbated conditions, making traffic congestion more severe. Additionally, traffic congestion is one of the leading contributors to noise and dust pollution in the city [5, 12]. Altogether, traffic congestion poses major barriers to urban quality of life, but the

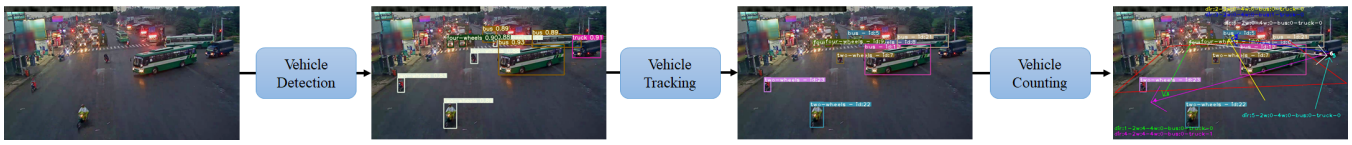


Figure 2: Overview of our traffic data analysis component. First, the object detector component is responsible for vehicle detection resulting in bounding boxes. Then, the tracking component tracks the detected vehicles in bounding boxes. Finally, the vehicles travelled in different motions are counted.

solutions are complex. There are two main existing problems with traffic in the big cities. First, the big cities need resources to solve infrastructure problems. Second, monitoring staff watch traffic activities on multiple screens, whose data are collected from numerous monitoring cameras installed on streets in the big cities.

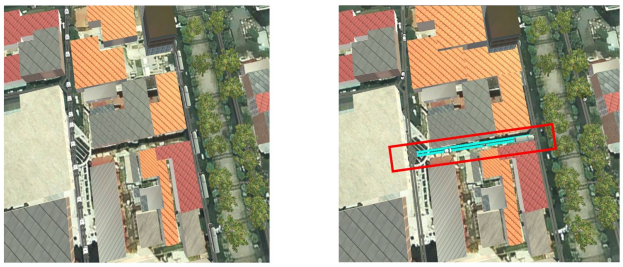
Therefore, the overarching goal of this work is to use a data-driven method for for the city traffic planning simulator. In particular, we make use of the city camera system which was originally used to detect traffic congestion and guide vehicles to alternate routes. Our findings do not only address specific urban challenges and the innovative technical solutions, but also provide models use in other contexts, including worldwide cities where traffic and congestion can benefit from AI. Figure 1 shows the overview of our proposed data-driven traffic simulation system. There are two main components in the proposed system, namely, traffic data analysis, and city traffic planning simulation.

It is worth noting that our novelty is the integration of analyzed data via computer vision into a simulator for city traffic planning. First, this work applies state-of-the-art computer vision algorithms from object detection, to trajectory-based tracking in order to improve the performance of traffic flow estimation. At the moment, most simulators used the mock-up data or random data. Regarding the actual traffic data collection, the consulting firms still deploy humans using manual clicker to count vehicles in a period of several days or weeks. This is very inefficient and tedious. Second, the project analyzes the camera data in the context of a graph; the traffic of one node can affect to another node in the traffic graph. Significantly, the visual data taken from one city camera grid can be used for other cities where the traffic situation remains the same. Third, we develop a city traffic planning simulator based on the actual traffic data. The simulator is used to support metropolitan transportation planning. The addition or removal of any traffic infrastructure can be observed from the simulator. The simulator results can guide engineers and city authorities in urban development planning. Fourth, the spatial-temporal navigation in the simulator makes use of the historically recorded analyzed traffic data. The impact of any additional/removal of infrastructure can be observed from different time periods.

The remainder of this paper is organized as follows. Section 2 summarizes the related works. In Section 3, we introduce the proposed framework. Section 4 presents the experiments. Finally, Section 5 concludes the paper and paves way to the future work.

2 Related Works

Computer vision is a field of artificial intelligence (AI) that trains computers to "see" and understand the visual world. In the task of analyzing traffic data, we consider computer vision algorithms such as object detection, action recognition, and semantic segmentation. First, object detection is an important task in computer vision which recognizes objects along with their location (in the form of bounding box) in the image. There are a number of object detection methods such as YOLO, Faster RCNN, and Focal-loss object detector [18, 26, 27] based on Convolutional Neural Networks (CNN) [15]. In addition, there are many variants to improve the runtime performance or accuracy [24, 40]. Action recognition in video is another important task in computer vision. It aims to recognize the humans and their activities in the videos. To date, there exist many approaches based on handcrafted features [16, 19, 37] and deep features [29, 34]. Meanwhile, image semantic segmentation is another computer vision fundamental task which a semantic label for each image pixel. This task is very challenging since it implicitly integrates the tasks of object detection, segmentation, and multi-label recognition into one single process. There are some works based on data-driven parsing [33] or deep learning [28, 41].



(a) Before adding road *(b) After adding road*

Figure 3: The illustration of adding a new infrastructure item into the simulator. The red rectangle highlights the newly added road.

Simulation is a powerful tool for training/education/planning in many different domains, i.e., public speaking simulation [22], non-destructive evaluation training [20], and sheet music interface for live performance [14]. There also exist many city traffic planning simulators [4, 8, 38] in literature. Daniel et al. [8] proposed a simulator which support to plan zones, roads, public transport like trains, trams, and buses. However, this simulator only focused on predicting traffic emissions based on the complete road network. Cai et al. [4] introduced a simulator for urban driving. However, the traffic data are randomly generated. Weyl and Glake [38] presented traffic simulation for city planning. However, there is no graphics available in the simulator. Here, the main drawback of these aforementioned simulators is the lack of the real traffic data analysis. In addition, the graphics are not customized for the city map and terrain. Also, these systems do not support virtual reality for the immersive experience.

3 Proposed Framework

3.1 Data Traffic Analysis

Figure 2 shows the overview of our traffic data analysis component. Video frames are first fed into the detection module, i.e., YOLO (You Only Look Once) detector [26] with the YOLOv5 implementation [40]. Then, we obtain a list containing bounding boxes and vehicle labels belong to these boxes for each frame. DeepSORT [39] works as our main tracking module. From the outcomes of tracking, for each bounding box, we retrieve its particular tracking ID. The counting module considers bounding boxes having the same tracking ID i as a trajectory. From the tracked vehicles through video frames, we can estimate the velocity v_i for each tracked vehicle i. The traffic

Figure 4: The modeling stage in our simulator. From left to right: (left) the 2D layout of the city map, (middle) the overlaid terrain onto the 2D city map, (right) our built 3D models such as buildings, roads, and landmark points for the city planning simulator.

density k is computed via:

$$k = \frac{N_t}{l} \tag{1}$$

, where N_t is number of vehicles at the time measurement t and l is the length of roadway. From the tracked points of the detected vehicles, we estimate the traffic flow. First, we compute the space mean speed $\bar{v_s}$ which is defined as the harmonic mean of speeds passing a point during a period of time. It also equals the average speeds over a length of roadway:

$$\bar{v_s} = \frac{N_t}{\sum_{i=1}^{N_t} \frac{1}{v_i}} \tag{2}$$

, where N_t again is number of vehicles at the time measurement. The output traffic flow q from one traffic camera is estimated: $q = k\bar{v_s}$, where k is already computed from the previous stage.

Note that each tracked vehicle i has n_i tracked points. We then compute the movement vector $\vec{m_i}$ for vehicle i based on its first tracked point p_i^1 and the final tracked point $p_i^P n_i$. Following the computation of movement vectors of vehicles, we leverage the K-Means algorithm to cluster those vectors into K dominant movement directions $[d_1, d_2, \ldots, d_K]$. We also apply the Elbow method [32] to find the optimal value of K in K-Means for each video:

$$[d_1, d_2, \ldots, d_K] = \text{Elbow\_KMeans}([m_1, m_2, \ldots, m_{N_t}]). \tag{3}$$

Then, we assign a direction dv_i to a vehicle i by calculating the similarity between the particular movement vector to every dominant directions and selects the direction k that gets the highest cosine similarity score:

$$dv_i = \arg\max_k \frac{\vec{m}_j \cdot \vec{d}_k}{|\vec{m}_j||\vec{d}_k|}. \tag{4}$$

We consider each single camera as a vertex in a graph G. After the geospatial calibration, each camera is responsible for one traffic vertex in the city map. The graph vertices are linked via the real map obtained the OpenStreetMap data [21]. Note that the graph links (edges) contain the weights, i.e., length of roadway l, and the number of vehicles N_t at the time measurement t. Larger roads tend to have larger ls. Meanwhile, the N_ts values are dynamically varied. The time-series data including graph G_t at different time ts are stored for the later use in the simulator that is described below.

3.2 City Traffic Planning Simulator

In this work, we proposed the data-driven city planning simulator. First, the "data-driven" term means the traffic data are not mock-up or dummy data. Instead, the data are analyzed from the actual traffic videos as mentioned in the previous subsection. We spawn the N_t vehicles into the simulator at time t. We assign the aforementioned speed v_i and direction dv_i for each vehicle i.

Second, the virtual city scenes are modeled from actual city landscapes. Therefore, city authorities are easily able to observe the impact of any update, e.g., the addition/removal of a bridge, and the change of road direction (e.g., one-way to two-way or vice versa). Due to the graph-based setting in the first component, the proposed simulator is able to visualize the update impact within the whole city. Our user interface is easy to use. The simulator supports users with common infrastructure items. For instance, the users are able to add a new infrastructure item into the traffic graph. Items to be added include road, overpass, bridge, tunnel, and traffic light, among others. The users are also able to change the traffic direction, i.e., from one way to two way, or vice versa. The addition/removal of any infrastructure is updated on the graph G by adding/removing the edges or change the graph weights. In addition, the neighboring nodes are affected as well. Following the graph update, the estimated traffic is recomputed to fit well the new graph. As depicted in Figure 3, the added road shares the traffic load with the main road.

3.3 Implementation

For the implementation, we develop the simulator with Unity3D engine [35]. We choose Ho Chi Minh City in Asia to develop our prototype. The reasons are two-fold. First, Ho Chi Minh City has busy traffic with 7.3 million motorbikes for more than 8.4 million residents that overwhelm city streets [10]. Second, the city provides a public camera grid showing the traffic at various locations [23]. This facilities our data collection process.

For the traffic analysis, we collected a dataset of 26,821 video frames and manually annotated 244,106 bounding boxes of popular vehicle classes such as car, bus, truck, and two-wheeled vehicle. The collected dataset is used to train the vehicle detection and tracking models as mentioned in Section 3.1. Regarding 3D modeling, we build the city infrastructures such as roads and buildings in the virtual environment in Blender [3]. Then, we import real world height maps [25] into Blender. The imported height map along with the satellite images from Google Map are used as the reference for us to touch up and refine the city landmark points and buildings. These 3D models are imported to the city simulator in Unity3D. Figure 4 visu-

Figure 5: The graphical user interface of the proposed city planning simulator in 3 different navigation modes. From left to right: Mode 1 - the users teleport to the predefined landmark points in the virtual environment, Mode 2 - the users travel in the virtual world in the third-person view, Mode 3 - the users experience the immersive environment via the first person view.

alizes the key steps in the city infrastructure modeling. We further model vehicles such as motorbikes, buses, trucks and cars. Regarding the traffic simulation, we use the analyzed traffic data to estimate the traffic flow. Then, we spawn the 3D vehicles with corresponding travel paths. We also attach the corresponding camera videos in the simulator for the references. Regarding VR headset, we deploy our simulator on Oculus Quest which is lightweight and portable. Thus it is very suitable for us to conduct experiments. Note that traversing or navigating within a virtual city is very challenging [6]. Therefore, we develop several modes, namely, Mode 1 (teleporting between predefined landmark points), Mode 2 (third-person view navigation), and Mode 3 (first-person view navigation), which will be investigated in the next section.

4 Evaluation

Our study received approval from the Institutional Review Board (IRB) at the university. We follow [13] and [31] for the design methodology and sample size, respectively. 25 people participated in this study, 12 of these participants identified themselves as female. The participants are university students and staff, whose ages range from 19 to 44 ($\mu = 26.6$). We provided participants the instructions for the experiment after they completed the consent form. The participants evaluate 3 aforementioned modes or variants, namely, Mode 1 (teleporting between predefined landmark points), Mode 2 (navigating in the third-person view), and Mode 3 (navigating in the first-person view). Figure 5 shows the user interface of the three evaluating modes. Each participant took part in a 30-minute session, namely, a 10-minute trial for each mode. We then showed the questionnaire and asked for feedback regarding the following perspectives:

- Ease of use: How easy is the method?
- Convenience: How convenient is the method?
- Realism: How does the mode look real to you?
- Functionality: Are you satisfied with the functions available in the system?
- Preference: How much do you prefer a certain mode over other modes?

The '*ease of use*' is the most popular criterion in literature [1, 9, 36]. Meanwhile, '*convenience*' and '*realism*' were included in [30]. In addition, the '*preference*' criterion was studied in [7,20]. Also, the criterion '*functionality*' was mentioned in [2,9,11,20]. Therefore, we included all the aforementioned criteria in our study. The participant rated each interaction mode on a 5-point Likert scale [17] from the best (5) to the worst (1) for each criterion.

Figure 6 shows the average scores of different modes for the aforementioned criteria. Mode 1 (teleporting mode) and Mode 2 (third-person view navigation) are highly rated for *ease of use*. The participants simply use AWSD keys to navigate and observe the virtual environment. Meanwhile, regarding Mode 3 (first person view navigation), the participants need to learn how to navigate in the virtual reality by using VR headset and controllers. Especially, they need time to get used to the first person view. Note that the consecutive wrong activities that are not recognized by the VR controllers frustrate the users.

Regarding the *convenience*, Mode 2 and Mode 3 achieve the highest rate. These modes are convenient and provides the great experience to users. Meanwhile, Mode 1 (teleporting between predefined landmark points) is not as convenient as Mode 3 since the users can only visit the limited places in the virtual world. In terms of *realism*, Mode 3 obtains the top rates. The main reason that Mode 3 outperforms others can be explained via the navigation support. Actually, the use of controllers is very interesting to participants. Therefore, it is realistic for the users to navigate within the city planning simulator. Meanwhile, Mode 1 achieves the lowest rate due to the limited navigation.

Regarding functionality and preferences, Mode 2 and Mode 3 are preferred by the participants. The participants highly rated the ability to observe and interact with the superimposed objects in the virtual environment. Furthermore, our participants appreciated the ability to navigate inside the virtual world.

The participants appreciate the efforts to integrate the real analyzed traffic data into the simulator. In addition, they also appreciate the embedded videos showing the real traffic scenes associated to the corresponding locations in the virtual world. In addition, we received many valuable comments to further improve our simulator. First, there is a need to increase the high fidelity of the VR headset. Using VR headset for a long time might cause a cybersickness. Second, the name of the road should be shown as a hologram to help users identify the current location. There should be additional feature to travel to certain location in the simulator by inputting the address. There is also a need of a minimap feature with compass for the ease of navigation. The participants also suggested that the Mode 2 can be used for city manipulation whereas Mode 3 can be used only for immersive navigation. This research has opened opportunities for future research to explore how the VR system can be applied in other big cities. Our participants recommended applying

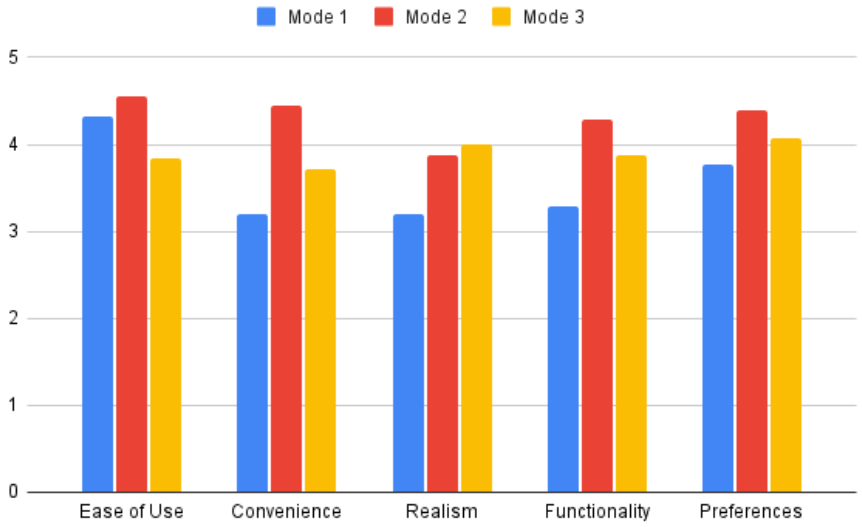


Figure 6: The average rating scores from the user study evaluation for the three modes.

our VR system in a more complicated environment (for example, desserts, countryside, suburban). Mode 1 is easy to use and can be deployed on the web browser. Meanwhile, Mode 2 and Mode 3 can be further developed on the VR headsets. These arguments suggest that our system has a lot of potentials for future usage.

5 Conclusion

In this paper, we introduce a data driven city traffic planning simulator. In particular, we analyze the real traffic data and feed the analyzed data into a city planning simulator. By using our simulator, the users are able to make certain modifications on the city infrastructures and observe the changes in traffic. We assess different variants of the proposed system in terms of the ease of use, convenience, realism, functionality, and preference. The user study indicates that participants favor the simulator in the third person-view for the ease of use and convenience, and the full VR first-person view support for realistic navigation.

In the future, we will investigate methods to further improve the current system. In particular, we aim to extend this work to different big cities in the world. Additionally, future studies should integrate different factors into the simulator such as the living costs, the gas price, the birth rates, and the land price. Finally, we believe this work could attract more future research looking to data-driven simulation in the rise of virtual reality era.

Acknowledgments

This work was partially funded by National Science Foundation (NSF) under Grant 2025234. This research was also funded by Vingroup and supported by Vingroup Innovation Foundation (VINIF) under project code VINIF.2019.DA19

References

[1] J. Ahn, S. Choi, M. Lee, and K. Kim. Investigating key user experience factors for virtual reality interactions. *Journal of the Ergonomics Society of Korea*, 36(4):267–280, 2017.

[2] M. K. Bekele and E. Champion. A comparison of immersive realities and interaction methods: Cultural learning in virtual heritage. *Frontiers in Robotics and AI*, 6:91, 2019.

[3] Blender software. `https://www.blender.org/`, Retrieved 13 June 2022.

[4] P. Cai, Y. Lee, Y. Luo, and D. Hsu. SUMMIT: A simulator for urban driving in massive mixed traffic. In *IEEE International Conference on Robotics and Automation*, pp. 4023–4029. IEEE, 2020.

[5] California's Pandemic Shutdowns Reveal How Traffic Pollutes Hispanic, Asian Communities. `https://e360.yale.edu/digest/pandemic-shutdowns-reveal-severity-of-car-pollution\-in-californias-hispanic-asian-communities`, Retrieved 13 June 2022.

[6] S. Chen, F. Miranda, N. Ferreira, M. Lage, H. Doraiswamy, C. Brenner, C. Defanti, M. Koutsoubis, L. Wilson, K. Perlin, and C. T. Silva. Urbanrama: Navigating cities in virtual reality. *IEEE Transactions on Visualization and Computer Graphics*, pp. 1–1, 2021.

[7] I. Cicek, A. Bernik, and I. Tomicic. Student thoughts on virtual reality in higher education—a survey questionnaire. *Information*, 12(4):151, 2021.

[8] M. Daniel, R. Dostál, S. Kozhevnikov, A. Matysková, K. Moudrá, A. M. Pereira, and O. Přibyl. City simulation software: Perspective of mobility modelling. In *2021 Smart City Symposium Prague (SCSP)*, pp. 1–7. IEEE, 2021.

[9] V. Gopalan, A. N. Zulkifli, and J. Abubakar. A study of students motivation based on ease of use, engaging, enjoyment and fun using the augmented reality science textbook. *Journal of the Faculty of Engineering*, 31:27–35, 2016.

[10] Ho Chi Minh City will not ban motorbikes, promises leader . `https://e.vnexpress.net/news/news/hcmc-will-not-ban-motorbikes-promises-leader-3891039.html`, Retrieved 13 June 2022.

[11] W. Huang, R. D. Roscoe, M. C. Johnson-Glenberg, and S. D. Craig. Motivation, engagement, and performance across multiple virtual reality sessions and levels of immersion. *Journal of Computer Assisted Learning*, 37(3):745–758, 2021.

[12] In South Asia, Vehicle Exhaust, Agricultural Burning and In-Home Cooking Produce Some of the Most Toxic Air in the World. `https://insideclimatenews.org/news/19042022/south-asia-air-pollution/`, Retrieved 13 June 2022.

[13] K. Kaur. Designing virtual environments for usability. In *International Conference on Human-Computer Interaction*, pp. 636–639, 1997.

[14] S. Kohen, C. Elvezio, and S. Feiner. Mixr: A hybrid AR sheet music interface for live performance. In *2020 IEEE International Symposium on Mixed and Augmented Reality Adjunct*, pp. 76–77. IEEE, 2020.

[15] A. Krizhevsky, I. Sutskever, and G. E. Hinton. Imagenet classification with deep convolutional neural networks. In P. L. Bartlett, F. C. N. Pereira, C. J. C. Burges, L. Bottou, and K. Q. Weinberger, eds., *Ad-*

vances in Neural Information Processing Systems, pp. 1106–1114, 2012.

[16] I. Laptev. On space-time interest points. *International Journal of Computer Vision*, 64(2-3):107–123, 2005.

[17] R. Likert. A technique for the measurement of attitudes. *Archives of psychology*, 1932.

[18] T. Lin, P. Goyal, R. B. Girshick, K. He, and P. Dollár. Focal loss for dense object detection. In *IEEE International Conference on Computer Vision*, pp. 2999–3007, 2017.

[19] T. V. Nguyen, J. Feng, and K. Nguyen. Denser trajectories of anchor points for action recognition. In *Proceedings of the 12th International Conference on Ubiquitous Information Management and Communication*, pp. 1:1–1:8, 2018.

[20] T. V. Nguyen, S. Kamma, V. Adari, T. Lesthaeghe, T. Boehnlein, and V. Kramb. Mixed reality system for nondestructive evaluation training. *Journal of Virtual Reality*, 25(3):709–718, 2021.

[21] OpenStreetMap. `https://www.openstreetmap.org/`, Retrieved 13 June 2022.

[22] F. Palmas, J. Cichor, D. A. Plecher, and G. Klinker. Acceptance and effectiveness of a virtual reality public speaking training. In *IEEE International Symposium on Mixed and Augmented Reality (ISMAR)*, pp. 363–371, 2019.

[23] Public traffic cameras in Ho Chi Minh City (in Vietnamese). `http://giaothong.hochiminhcity.gov.vn/`, Retrieved 13 June 2022.

[24] S. Qiao, L. Chen, and A. L. Yuille. Detectors: Detecting objects with recursive feature pyramid and switchable atrous convolution. In *IEEE Conference on Computer Vision and Pattern Recognition*, pp. 10213–10224, 2021.

[25] Real World Height Maps. `http://terrain.party/`, Retrieved 13 June 2022.

[26] J. Redmon, S. Divvala, R. Girshick, and A. Farhadi. You only look once: Unified, real-time object detection. In *Conference on Computer Vision and Pattern Recognition*, pp. 779–788, 2016.

[27] S. Ren, K. He, R. B. Girshick, and J. Sun. Faster R-CNN: towards real-time object detection with region proposal networks. *IEEE Trans. Pattern Anal. Mach. Intell.*, 39(6):1137–1149, 2017.

[28] E. Shelhamer, J. Long, and T. Darrell. Fully convolutional networks for semantic segmentation. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 39(4):640–651, 2017.

[29] K. Simonyan and A. Zisserman. Two-stream convolutional networks for action recognition in videos. In *Advances in Neural Information Processing Systems*, pp. 568–576, 2014.

[30] M. Slater, C. Gonzalez-Liencres, P. Haggard, C. Vinkers, R. Gregory-Clarke, S. Jelley, Z. Watson, G. Breen, R. Schwarz, W. Steptoe, D. Szostak, S. Halan, D. Fox, and J. Silver. The ethics of realism in virtual and augmented reality. *Frontiers in Virtual Reality*, 1:1, 2020.

[31] F. Suárez-Warden, M. Rodriguez, N. Hendrichs, S. García-Lumbreras, and E. G. Mendívil. Small sample size for test of training time by augmented reality: An aeronautical case. *Procedia Computer Science*, 75:17 – 27, 2015.

[32] M. Syakur, B. Khotimah, E. Rochman, and B. D. Satoto. Integration k-means clustering method and elbow method for identification of the best customer profile cluster. In *IOP conference series: materials science and engineering*, vol. 336, p. 012017, 2018.

[33] J. Tighe, M. Niethammer, and S. Lazebnik. Scene parsing with object instance inference using regions and per-exemplar detectors. *International Journal of Computer Vision*, 112(2):150–171, 2015.

[34] D. Tran, L. D. Bourdev, R. Fergus, L. Torresani, and M. Paluri. Learning spatiotemporal features with 3d convolutional networks. In *IEEE International Conference on Computer Vision (ICCV)*, pp. 4489–4497, 2015.

[35] Unity3D Game Engine. `https://unity.com/`, Retrieved 13 June 2022.

[36] V. Venkatesh. Determinants of perceived ease of use: Integrating control, intrinsic motivation, and emotion into the technology acceptance model. *Information systems research*, 11(4):342–365, 2000.

[37] H. Wang, D. Oneata, J. Verbeek, and C. Schmid. A robust and efficient video representation for action recognition. *International Journal of Computer Vision*, 119(3):219–238, 2016.

[38] J. Weyl, U. A. Lenfers, T. Clemen, D. Glake, F. Panse, and N. Ritter. Large-scale traffic simulation for smart city planning with mars. In *Proceedings of the 2019 Summer Simulation Conference*. Society for Computer Simulation International, 2019.

[39] N. Wojke, A. Bewley, and D. Paulus. Simple online and realtime tracking with a deep association metric. In *2017 IEEE International Conference on Image Processing (ICIP)*, pp. 3645–3649, 2017.

[40] YOLOv5. `https://github.com/ultralytics/yolov5`, Retrieved 13 June 2022.

[41] H. Zhao, J. Shi, X. Qi, X. Wang, and J. Jia. Pyramid scene parsing network. In *IEEE Conference on Computer Vision and Pattern Recognition*, pp. 6230–6239, 2017.