VISTA 2.0: An Open, Data-driven Simulator for Multimodal Sensing and Policy Learning for Autonomous Vehicles, 10/28

Ao Qu

Part I. Literature (for further reading about simulators for autonomous vehicles)

A. Amini et al., "Learning Robust Control Policies for End-to-End Autonomous Driving From Data-Driven Simulation," in IEEE Robotics and Automation Letters, vol. 5, no. 2, pp. 1143-1150, April 2020, doi: 10.1109/LRA.2020.2966414.

Alexander Amini, Tsun-Hsuan Wang, Igor Gilitschenski, Wilko Schwarting, Zhijian Liu, Song Han, Sertac Karaman, and Daniela Rus. 2022. VISTA 2.0: An Open, Data-driven Simulator for Multimodal Sensing and Policy Learning for Autonomous Vehicles. In 2022 International Conference on Robotics and Automation (ICRA). IEEE Press, 2419–2426. https://doi.org/10.1109/ICRA46639.2022.9812276

Wang, T. H., Amini, A., Schwarting, W., Gilitschenski, I., Karaman, S., & Rus, D. (2022, May). Learning interactive driving policies via data-driven simulation. In 2022 International Conference on Robotics and Automation (ICRA) (pp. 7745-7752). IEEE.

Kaur, P., Taghavi, S., Tian, Z., & Shi, W. (2021). A survey on simulators for testing self-driving cars. 2021 Fourth International Conference on Connected and Autonomous Driving (MetroCAD). https://doi.org/10.1109/metrocad51599.2021.00018

Dosovitskiy, A., Ros, G., Codevilla, F., Lopez, A., & Koltun, V. (2017, October). CARLA: An open urban driving simulator. In Conference on robot learning (pp. 1-16). PMLR.

Rong, G., Shin, B. H., Tabatabaee, H., Lu, Q., Lemke, S., Možeiko, M., ... & Kim, S. (2020, September). Lgsvl simulator: A high fidelity simulator for autonomous driving. In 2020 IEEE 23rd International conference on intelligent transportation systems (ITSC) (pp. 1-6). IEEE.

Part II. Recent News

Is simulation the way forward to fully autonomous driving? Engineering.com. Retrieved November 14, 2022, from https://www.engineering.com/story/is-simulation-the-way-forward-to-fully-autonomous-driving

Yahoo!. The driving training simulator market is expected to grow from US$ 1,950.00 million in 2021 to US$ 2,646.73 million by 2028. Yahoo! Finance. Retrieved November 14, 2022, from https://finance.yahoo.com/news/driving-training-simulator-market-expected-130300645.html

Gordon, R. (2022, August 29). MIT scientists release open-source photorealistic simulator for autonomous driving. SciTechDaily. Retrieved November 14, 2022, from https://scitechdaily.com/mit-scientists-release-open-source-photorealistic-simulator-for-autonomous-driving/

Part III.

Jinhua questions

Q: You mentioned the three levels of data driven engine, from the physical, then to the environmental, then to the behavior. Do they require different type of methodology to tackle each of them, or do they actually share a fundamental similar principle that you capture?

A(Amini): Iit's quite interesting because actually a lot of the backbones for data driven simulation has not really been developed yet because these techniques are so new. So a lot of what we're trying to do is actually establish that principle backbone of how we can unify a lot of these themes to avoid kind of very scattered research across these different domains. The way we've targeted this so far has been unifying it in the representation space of the data. So taking your data, whether that be camera data, but it could also be other types of multimodal data modalities and trying to unify that into a common space. So all of these sensor modalities can be shared so that when you want to simulate, let's say, a new sensor modality, you can leverage from this common space, that gives you a very unified way of dealing with the physical simulation and also with the environmental stimulation, because both of those live in that physical space. The behavioral simulation, though, is quite different because it's not only about the environment, but also about the agents that interact in the environment. Now, oftentimes what we found is that those agents rely not only on other agents, the motion of other agents, but the environment itself. So it is important to also rely on that shared physical space. And we're still finding ways to incorporate the behavioral models with other agents. But that's still a very open question and a very exciting line of research.

A(Rus): In fact, the picture that Alexander has on the slide here is very relevant because we really have a hierarchy of simulation capabilities. And you can think of it as a pyramid or a cylinder. However you want to imagine the hierarchy, but the higher up you go, In the hierarchy, the higher the level of abstraction in some sense. But the abstractions build on the representations that Alexander has talked about. And so we have common representations that enable us to put different sensors underneath them. From these representations, we can also make connections with the environment, and then we can put higher level behaviors and obstructions on top of that.

Q: The second question is about the edge case construction. Here, everybody talking about this is a key challenge. They are critical for safety, but it's highly inefficient to learn if you just use real world data. The way you are doing this is you absorb this real world data this year, the structure, and then you inject variations to create these new edge cases. So tell us a little bit more about the strategy of this construction of new edge cases, and how do you make sure that the things that you constructed are relevant to reality? Or there would still be a surprise that you wouldn't know how to construct very extreme edge cases, for example?

A(Amini): I can share maybe two ways that we've tackled this problem so far. The first way is a very traditional reinforcement learning mentality where basically you say that the construction of these edge cases is not for humans to decide. It's actually something that should be decided through interacting with the environment. And if you don't observe these types of edge cases, then perhaps they're not of enough utility for the system. So the mentality in this case is let's unleash our virtual agents into these types of data driven environments and let's allow them to interact and find all these edge cases. If they don't observe them and through a much larger scale, a more sustainable scale of data collection, we can actually do this in virtual simulation compared to real world settings, then that gives us already a great foundation of dealing with edge cases because in this interaction with the environment we can already be exposed to a lot of these cases. Now the other regime is something that we've started more recently, which is taking a much more principled approach towards targeted edge case curation. And in this case, we're actually trying to optimize, almost solving an optimization process between two objectives. One objective is trying to increase the uncertainty or the ambiguity or the stress of the mobility system that we're trying to curate. And we try to create edge cases that will propagate the maximum amount of uncertainty for those systems. Now, on the other end, it's a dual optimization process because not only do we want to maximize the uncertainty of those edge cases, we also want to minimize the deviation and capture the realism still of the environments. Because of course, we can create many types of very challenging edge cases, but they're almost impractical for real world use and it's almost not fair to the system. So a lot of what we're doing now, especially towards the end of the talk, I was hinting at this with this uncertainty propagation of the systems is to not only propagate the uncertainty through the simulation platforms, but also propagating through the decision making to achieve uncertainty with decisions so that we can actually tackle these types of questions and solve these types of very challenging edge cases that are still realistic.

A(Rus): In fact, this is where we are with reasoning about edge cases. But at the beginning when we started this work, we were very interested in accidents and near accident situations. So we were interested in simpler things, but things that occur a lot. For instance, I was rear ended on the mass spike yesterday and this was while I was changing lanes and this is not an uncommon event. We actually did start with some common types of situations that we wanted to avoid. But then we had people asking us, well, suppose a plane is to do an emergency landing on the highway. How would you respond to it? Can your system adapt to this? And like, if you think about this, the idea of a plane landing on the highway is something that happens very rarely. But when it happens, I guess we as humans understand the semantics of the emergency landing, understand what a plane is and we kind of understand what the response to a situation like this might be. And so we started thinking about how we would connect real time performance with unexpected events that the system wouldn't have seen before through reasoning and decision making. And so there is a lot to be done here, primarily because our robot cars don't have the same range of knowledge and insight that humans have. But at the same time we're not deploying these systems yet or we're deploying them in very limited scenarios where we don't really expect the plane to be landing on the highway.

Q: At the beginning, Daniela mentioned there's a strong asymmetry between the large corporations' access to data versus academics and also even small companies, for example. Right. So I have a feeling that your work can potentially really change that situation, at least contribute to rebalancing between the two. And then from the large corporation point of view, their advantage of having large data set will relatively diminish in that way. Right. So for that I'd like to see your view about how the whole play between academia and large corporates evolve in this.

A(Amini): Actually, this is one of the things that really inspired us and made us pursue this decision because we do think that this is really a fundamental technology and really a pivotal reframing of how simulation is done. Even in these industry companies, very large scale industry self driving leaders are taking a radically different approach to simulation. And you can see especially bigger companies like Google and Waymo, decades of research into self driving, but invested into the model based simulation. They're continually trying to build up the realism of these model based simulation engines. And within a few years, actually, by leveraging and properly restructuring the problem formulation, we can leverage the real data and build much more realistic engines from this data. So we're really inspired by this reframing of the problem and how that will reshape the way that these communities think about it's possible, how it's possible to capture the very harsh, uncertainties and edge cases of these realworld systems, not only in mobility, but I think there's huge potential for data driven simulation far beyond these types of systems as well. And by open sourcing it, we actually wanted to, like you said, even the playing ground between these two fields. We've already started to see a shift in industries now moving away from model based simulation engines, now closer towards data driven simulators because they actually do have these large data sets. They have the tools and the mechanisms to build very high fidelity engines like Vista within their own organizations. But I think even taking a step back from, let's say, industry versus academia, I think this presents another very exciting possibility, which is for the governance of autonomy as well. Because I think right now what we're missing is a unified framework for evaluating policies that are deployed onto real world cities and streets. What I'm excited about with this work is actually it can start to serve as the framework, a unified framework that's very high fidelity, very photorealistic, both in terms of what we see, the physical environments as well as the behaviors that exist on the roads to establish this baseline for governance and how we can have mobility systems, autonomous mobility systems that can be seamlessly integrated into society with our existing transportation networks and so on.

A(Rus): I observe that for many businesses, the businesses with the best data are the businesses that have the most success with machine learning. And machine learning is very powerful today. Now, with these open source tools, what we're doing is essentially we're leveling the playing field because we're saying we have this way that makes it really easy for people to have data in the space, although Alexander said that the work is extensible to other domains. But we're leveling the playing field. We're making it really easy for data to be generated. And this is good quality data. We also have a suite of algorithms that allow us to test the quality of the data to figure out, given the space, where do we have data, where do we have too much data, where are we missing data? And so with tools like this, we can sort of bring everyone onto the same level from the point of view of data. And that means that we open the possibility of innovation and of making progress in many more directions. We are open to other university researchers who may not have autonomous vehicles to collect data or special sensors to collect data, but they could build on top of the Vista engine and they could add new capabilities. We are also open to small businesses who may not have machine learning expertise in mind, although I would say that Vista is open source and freely available for research purposes, not for commercial purposes.

Audience

Q: Is your simulation limited by the quality of data? Do you have any methodology to validate this against the real world data once the system is done processing it?

A(Amini): The fundamental approach that we're taking to this is a very technological approach. We definitely agree with this assessment that not all data is equal. And especially when simulating and generating new data, it's especially important to keep that in our minds so that we're not actually polluting whatever we're training or evaluating based on incorrect or faulty data. And the approach that we're taking for this is actually one that's especially suited for data driven simulators. And that's the ability to have data driven uncertainty estimation techniques that can be built into the AI systems that are actually generating the data and processing that data so that we can actually propagate that uncertainty all the way. From the data space through to the predictive models that are actually generating the data and even all the way to the other end in the rendering and the actual simulation of that output data feed. And we can actually sense the uncertainty, propagating that uncertainty throughout all stages of this. And this gives us a very quantitative understanding, both in terms of quality of the data, how noisy that data is, for example, not only the data from the sensors, but also the labels as well. Both sides of that are critically important, as well as, the imbalance of your data. So we've built a suite of computational and quantitative methodologies for data driven neural networks that can not only understand and propagate the uncertainties, but also propagate these imbalances in the data distributions as well.

A(Rus): I would just like to emphasize that the reason Vista works so well is because we do start with high quality data. But high quality means really high resolution, high fidelity. Once we have a car trace that has high fidelity and high resolution, then it's easy to do all the processing that Alexander described. If the data from the sensors is fuzzy and low resolution, then it's much harder to reconstruct the 3D scene, as you have seen.

Q: Are you going to make some sort of an interface for public data sets? People can use Vista plug in their own data sets into it, some sort of crowdsourcing for the input data that would really increase the number of data points. Have you all considered other sort of actors such as bikes, pedestrians, scooters, motorcycles?

A(Amini): For the first question, we actually already have preliminary results where basically the input to Vista can be not only our data that's collected, but you can think of it almost as a conversion function that can take an arbitrary data set that's coming from. We have already achieved this conversion functionality with many open-sourced datasets such as Google’s self-driving dataset. For the second question, we haven't done this yet, but extremely interesting to do. For example, a car will behave very differently next to another car versus next to a bicyclist. And how can you capture that into your behavioral models as well? Not only the physical simulation of those agents?

Q: Hardware requirements for processing all the data?

A(Amini): This was something we wanted to prioritize when we were building the simulator. We did not like this trend of if you want to train and simulate or test and simulation, you need these large supercomputers. Our simulator runs local simulations. That means that you could have a very large data set, tens of thousands of kilometers. But your simulation of the agent is only happening just locally around you, just based on what you can see around you. We could actually simulate and train our vehicles on tens of thousands of kilometers of road within a matter of a few hours on a single desktop machine with a commercial grade NVIDIA GPU.

Q: Do you envision this open source system being something like Tesla has? Because Tesla has the maximum VMT that they gather from the cameras.

A(Rus): Not all the data is equal and most of the data collected directly from cars is pretty boring. Tesla is very good at following a lane, and we don't know what they have in house for other kinds of situations. But our solution is aimed at accelerating the ability to get good data, because not all data is equal. There is the data that brings you new information, that brings new knowledge into the system. Our system aims to get this kind of new data faster.

Q: You are going to be aware of IEEE2846 standards. They've defined like eight scenarios such as car following and pedestrians walking onto the street. Are you following some sort of a pattern where these vehicles in these scenarios will behave in a certain way or is it more of just inputting the data, let the system learn on its own and then deploy it to the best possible manner on the street?

A: We are adopting a hybrid approach. We want to not only rely on the simulations that have been handcrafted by us as humans, we want to actually explore the full range of this environment to see what's possible. There is a huge advantage towards having a simulation platform that can have these tunable knobs that you can control and modify to your heart's desire and you can actually craft the types of scenarios and environments that you care about the most. And I think that's especially important for evaluation purposes, not only for training, but especially in the case of evaluation where you want to make sure that whatever you're deploying into reality meets some standard baselines or behaves in the ways that you would expect to be congruent with the rules and regulations of the streets. What we've created in Vista has been very much tuned to both of those visions. We've wanted to create something that can both explore the vast variety of the data, to see these things that we haven't thought of yet, but also gives us that tunable knob so that we could do that if we wanted to.

Part IV. Summary of Memos.

Themes from Other Memos

1. A data-driven simulation means that it would collect data from multiple sources. How does this simulator account for varying quality of data (e.g., camera resolution, frame rate, etc.) when reconstructing the environment?
2. The simulator is capable of conducting "counterfactual" experiments such as perturbing the real trajectory. Does the simulation fidelity remain consistent when the perturbation gets larger? Taking one step further, would it be possible to consider other "what-if"s such as drivers looking at their phone or aggressive driving behavior?
3. Currently, the tool is mostly designed for developing and testing autonomous vehicle algorithms. How could this tool be used by non-automation researchers and practitioners in helping to spearhead autonomous vehicle deployment? For example, can city planners use such tools to optimize curbside space and parking locations? The goal here is really to get more people on board with the development and eventual rollout of autonomous vehicles.
4. If we can successfully build a real-world approximation in the data-driven simulator, can we not solve for broader society level issues and do we not enter the next stage of the

metaverse that – till date – only highly immersive games have managed to approach with their world design? What kinds of breakthroughs are we waiting for in this data-driven model?

My Reflection

First, identifying and training on edge cases is crucial to the successful deployment of autonomous vehicles. So far, we have solved probably more than 95% of traffic scenarios but the less than 5% that hasn't been encountered and properly solved is what truly hinders autonomous vehicles from hitting the road. It's almost impossible to tackle all the edge cases because they are just infinite. However, is there a way to evaluate if AV has the same level of capability of handling edge cases as human drivers? Personally, I don't think it would be possible to have collision-free guarantee on AVs but it would be interesting to investigate what portion of edge cases would be enough to demonstrate its super-human level driving skills.

Second, this digital twin approach can not only facilitate the development and testing of autonomous vehicles, but also has the promise of helping us build better transportation systems and better cities. By collecting real-world data and feeding them into the simulator, we are curating a digital asset of the cities we live in and many interventions such as traffic signal control and road design can be fully tested before we launch them in reality. I think this is some gap that the civil engineering community and computer science community should work together to bridge.

Part V. Other Information

Other questions: If industry and academics commit to more collaborations and adopt a standardized simulator where they share data and trained policy, we can see that there is clearly an advantage because that would facilitate the development of technology, but is there a good way to cope with the liability issue? Can we trace back to whom should be held accountable if an accident happens as a result of the policy trained on such a simulator? Should they actually be responsible for the accident?