Top Level Newsletter: Connected Vehicle

(Published by IEEE Orange County Section)

August 2021
Volume 20.0

Editor-in-Chief: Kay Das, IEEE Life Member,
Systems Research Development and Innovation

Vol 20 comprises:

(1) Tesla AI Chief Explains Why Self-Driving Cars Don't Need Lidar (complete article)

(2) 6G Wireless Systems: Vision, Requirements, Challenges, Insights, Opportunities

(3) A Novel Simulation Framework for the Design and Testing of Advanced Driver Assistance Systems

(4) Learning to Automatically Catch Potholes in Worldwide Road Scene Images

This series of newsletters is intended to provide the IEEE member with a top level briefing of the many different subjects relevant to the research, development and innovation of the connected vehicle. This newsletter additionally takes an early peek at 6G and ways in which its development will differ from 5G.

A note on the Connected Vehicle Newsletter development: Volume 20 departs from the norm and includes a complete article. Volume 19 returned to the usual format. Volume 18 was a top-level synopsis of the 74 most recent entries from the twelve previous volumes (vol 17 to vol 6) since March 2020.

The objective is to provide a platform for fast learning and quick overview so that the reader may be guided to the next levels of detail and gain insight into correlations between the entries to enable growth of the technology. Intended audiences are those that desire a quick introduction to the subject and who may wish to take it further and deepen their knowledge. This includes those in industry, academia or government and the public at large. Descriptions will include a range of flavors from technical detail to broad industry and administrative issues. A (soft) limit of 300 to 600 words is usually set for each entry, but not rigorously exercised, as in this newsletter where the full article is presented.

As descriptions are not exhaustive, hyperlinks are occasionally provided to give the reader a first means of delving into the next level of detail. The reader is encouraged to develop a first level understanding of the topic in view. The emphasis is on brief, clear and contained text. There will be no diagrams in order to keep the publication concise and podcast-friendly. Related topics in the case of Connected Vehicle technology, such as 5G cellular and the Internet of Things will be included. The terms Connected Vehicle and Automated Driving will be used inter-changeably. Articles from other published sources than IEEE that add to the information value will occasionally be included.

This newsletter forms part of the regional Advanced Technology Initiative (ATI) of which connected vehicles form a constituent part. Technical articles solely from IEEE journals/magazines are referred to by their Digital Object Identifier (DOI) or corresponding https link. The link for each article is provided. Those readers who wish to delve further to the complete paper and have access to IEEE Explore (www.ieeexplore.ieee.org) may download complete articles of interest. Those who subscribe to the relevant IEEE society and receive the journal may already have physical or electronic copies. In case of difficulty please contact the editor at kaydas@mac.com. The objective is to provide *top level guidance* on the subject of interest. As this is a collection of summaries of already published articles and serves to further widen audiences for the benefit of each publication, no copyright issues are foreseen.

Readers are encouraged to develop their own onward sources of information, discover and draw inferences, join the dots, and further develop the technology. Entries in the newsletter are normally either editorials or summaries or abstracts of articles. Where a deepening of knowledge is desired, reading the full article is recommended.

1. Tesla AI Chief Explains Why Self-Driving Cars Don't Need Lidar, Ben Dickinson
The Machine, Making sense of AI, 3 July 2021, (FULL ARTICLE)
https://venturebeat.com/2021/07/03

What is the technology stack you need to create fully autonomous vehicles? Companies and researchers are divided on the answer to that question. Approaches to autonomous driving range from just cameras and computer vision to a combination of computer vision and advanced sensors. Tesla has been a vocal champion for the pure vision-based approach to autonomous driving, and in this year's Conference on Computer Vision and Pattern Recognition (CVPR), its chief AI scientist Andrej Karpathy explained why.

Speaking at CVPR 2021 Workshop on Autonomous Driving, Karpathy, who has been leading Tesla's self-driving efforts in the past years, detailed how the company is developing deep learning systems that only need video input to make sense of the car's surroundings. He also explained why Tesla is in the best position to make vision-based self-driving cars a reality.

A general computer vision system: Deep neural networks are one of the main components of the self-driving technology stack. Neural networks analyze on-car camera feeds for roads, signs, cars, obstacles, and people. But deep learning can also make mistakes in detecting objects in images. This is why most self-driving car companies, including Alphabet subsidiary Waymo, use lidars, a device that creates 3D maps of the car's surrounding by emitting laser beams in all directions. Lidars provided added information that can fill the gaps of the neural networks.

However, adding lidars to the self-driving stack comes with its own complications. “You have to pre-map the environment with the lidar, and then you have to create a high-definition map, and you have to insert all the lanes and how they connect and all the traffic lights,” Karpathy said. “And at test time, you are simply localizing to that map to drive around.” It is extremely difficult to create a precise mapping of every location the self-driving car will be traveling. “It’s unscalable to collect, build, and maintain these high-definition lidar maps,” Karpathy said. “It would be extremely difficult to keep this infrastructure up to date.” Tesla does not use lidars and high-definition maps in its self-driving stack. “Everything that happens, happens for the first time, in the car, based on the videos from the eight cameras that surround the car,” Karpathy said.

The self-driving technology must figure out where the lanes are, where the traffic lights are, what is their status, and which ones are relevant to the vehicle. And it must do all of this without having any predefined information about the roads it is navigating. Karpathy acknowledged that vision-based autonomous driving is technically more difficult because it requires neural networks that function incredibly well based on the video feeds only. “But once you actually get it to work, it’s a general vision system, and can principally be deployed anywhere on earth,” he said. With the general vision system, you will no longer need any complementary gear on your car. And Tesla is already moving in this direction, Karpathy says. Previously, the company’s cars used a combination of radar and cameras for self-driving. But it has recently started shipping cars without radars. “We deleted the radar and are driving on vision alone in these cars,” Karpathy said, adding that the reason is that Tesla’s deep learning system has reached the point where it is a hundred times better than the radar, and now the radar is starting to hold things back and is “starting to contribute noise.”

Supervised learning: The main argument against the pure computer vision approach is that there is uncertainty on whether neural networks can do range-finding and depth estimation without help from lidar depth maps. “Obviously humans drive around with vision, so our neural net is able to process visual input to understand the depth and velocity of objects around us,” Karpathy said. “But the big question is can the synthetic neural networks do the same. And I think the answer to us internally, in the last few months that we’ve worked on this, is an unequivocal affirmative.”

Tesla’s engineers wanted to create a deep learning system that could perform object detection along with depth, velocity, and acceleration. They decided to treat the challenge as a supervised learning problem, in which a neural network learns to detect objects and their associated properties after training on annotated data. To train their deep learning architecture, the Tesla team needed a massive dataset of millions of videos, carefully annotated with the objects they contain and their properties. Creating datasets for self-driving cars is especially tricky, and the engineers must make sure to include a diverse set of road settings and edge cases that don't happen very often. "When you have large, clean, diverse datasets, and you train a large neural network on it, what I've seen in practice is… success is guaranteed," Karpathy said.

Auto-labeled dataset: With millions of camera-equipped cars sold across the world, Tesla is in a great position to collect the data required to train the car vision deep learning model. The Tesla self-driving team accumulated 1.5 petabytes of data consisting of one million 10-second videos and 6 billion objects annotated with bounding boxes, depth, and velocity.

But labeling such a dataset is a great challenge. One approach is to have it annotated manually through data-labeling companies or online platforms such as Amazon Turk. But this would require a massive manual effort, could cost a fortune, and become a very slow process.
Instead, the Tesla team used an auto-labeling technique that involves a combination of neural networks, radar data, and human reviews. Since the dataset is being annotated offline, the neural networks can run the videos back in forth, compare their predictions with the ground truth, and adjust their parameters. This contrasts with test-time inference, where everything happens in real-time and the deep learning models can't make recourse.

Offline labeling also enabled the engineers to apply very powerful and compute-intensive object detection networks that can't be deployed on cars and used in real-time, low-latency applications. And they used radar sensor data to further verify the neural network's inferences. All of this improved the precision of the labeling network. "If you're offline, you have the benefit of hindsight, so you can do a much better job of calmly fusing [different sensor data]," Karpathy said. "And in addition, you can involve humans, and they can do cleaning, verification, editing, and so on." According to videos Karpathy showed at CVPR, the object detection network remains consistent through debris, dust, and snow clouds. Karpathy did not say how much human effort was required to make the final corrections to the auto-labeling system. But human cognition played a key role in steering the auto-labeling system in the right direction.

While developing the dataset, the Tesla team found more than 200 triggers that indicated the object detection needed adjustments. These included problems such as inconsistency between detection results in different cameras or between the camera and the radar. They also identified scenarios that might need special care such as tunnel entry and exit and cars with objects on top. It took four months to develop and master all these triggers. As the labeling network became better, it was deployed in "shadow mode," which means it is installed in consumer vehicles and run silently without issuing commands to the car. The network's output is compared to that of the legacy network, the radar, and the driver's behavior. The Tesla team went through seven iterations of data engineering. They started with an initial dataset on which they trained their neural network. They then deployed the deep learning in shadow mode on real cars and used the triggers to detect inconsistencies, errors, and special scenarios. The errors were then revised, corrected, and if necessary, new data was added to the dataset.

“We spin this loop over and over again until the network becomes incredibly good,” Karpathy said. So, the architecture can better be described as a semi-auto labeling system with an ingenious division of labor, in which the neural networks do the repetitive work and humans take care of the high-level cognitive issues and corner cases. Interestingly, when one of the attendees asked Karpathy whether the generation of the triggers could be automated, he said, “[Automating the trigger] is a very tricky scenario, because you can have general triggers, but they will not correctly represent the error modes. It would be very hard to, for example, automatically have a trigger that triggers for entering and exiting tunnels. *That’s something semantic that you as a person have to intuit* that this is a challenge… It’s not clear how that would work.”

Tesla’s self-driving team needed a very efficient and well-designed neural network to make the most out of the high-quality dataset they had gathered. The company created a hierarchical deep learning architecture composed of different neural networks that process information and feed their output to the next set of networks. The deep learning model uses convolutional neural networks to extract features from the videos of eight cameras installed around the car and fuses them together using transformer networks. It then fuses them across time, which is important for tasks such as trajectory-prediction and to smooth out inference inconsistencies.

The spatial and temporal features are then fed into a branching structure of neural networks that Karpathy described as heads, trunks, and terminals. “The reason you want this branching structure is because there’s a huge amount of outputs that you’re interested in, and you can’t afford to have a single neural network for every one of the outputs,” Karpathy said. The hierarchical structure makes it possible to reuse components for different tasks and enable feature-sharing between the different inference pathways.

Another benefit of the modular architecture of the network is the possibility of distributed development. Tesla is currently employing a large team of machine learning engineers working on the self-driving neural network. Each of them works on a small component of the network and they plug in their results into the larger network.

Vertical integration: In his presentation at CVPR, Karpathy shared some details about the supercomputer Tesla is using to train and finetune its deep learning models. The compute cluster is composed of 80 nodes, each containing eight Nvidia A100 GPUs with 80 gigabytes of video memory, amounting to 5,760 GPUs and more than 450 terabytes of VRAM. The supercomputer also has 10 petabytes of NVME superfast storage and 640 tbps networking capacity to connect all the nodes and allow efficient distributed training of the neural networks. Tesla also owns and builds the AI chips installed inside its cars. “These chips are specifically designed for the neural networks we want to run for [full self-driving] applications,” Karpathy said.

Tesla’s big advantage is its vertical integration. Tesla owns the entire self-driving car stack. It manufactures the car and the hardware for self-driving capabilities. It is in a unique position to collect a wide variety of telemetry and video data from the millions of cars it has sold. It also creates and trains its neural networks on its proprietary datasets, its special in-house compute clusters, and validates and finetunes the networks through shadow testing on its cars. And, of course, it has a very talented team of machine learning engineers, researchers, and hardware designers to put all the pieces together. “You get to co-design and engineer at all the layers of that stack,” Karpathy said. “There’s no third party that is holding you back. You’re fully in charge of your own destiny, which I think is incredible.”

This vertical integration and repeating cycle of creating data, tuning machine learning models, and deploying them on many cars puts Tesla in a unique position to implement vision-only self-driving car capabilities. In his presentation, Karpathy showed several examples where the new neural network alone outmatched the legacy ML model that worked in combination with radar information. And if the system continues to improve, as Karpathy says, Tesla might be on the track of making lidars obsolete. And I don’t see any other company being able to reproduce Tesla’s approach.

Open issues: But the question remains as to whether deep learning in its current state will be enough to overcome all the challenges of self-driving. Surely, object detection and velocity and range estimation play a big part in driving. But human vision also performs many other complex functions, which scientists call the “dark matter” of vision. Those are all important components in the conscious and subconscious analysis of visual input and navigation of different environments.
Deep learning models also struggle with making causal inference, which can be a huge barrier when the models face new situations they haven’t seen before. So, while Tesla has managed to create a very huge and diverse dataset, open roads are also very complex environments where new and unpredicted things can happen all the time.

The AI community is divided over whether you need to explicitly integrate causality and reasoning into deep neural networks or if you can overcome the causality barrier through “direct fit,” where a large and well-distributed dataset will be enough to reach general-purpose deep learning. Tesla’s vision-based self-driving team seems to favor the latter (though given their full control over the stack, they could always try new neural network architectures in the future). It will be interesting to observe how the technology fares against the test of time.

2. 6G Wireless Systems: Vision, Requirements, Challenges, Insights, and Opportunities

By Harsh Tatari et al

Proceedings of the IEEE Year: July 2021, Volume: 109, Issue: 7

Abstract:
Mobile communications have been undergoing a generational change every ten years or so. However, the time difference between the so-called “G’s” is also decreasing. While fifth-generation (5G) systems are becoming a commercial reality, there is already significant interest in systems beyond 5G, which we refer to as the sixth generation (6G) of wireless systems. In contrast to the already published papers on the topic, we take a top-down approach to 6G. More precisely, we present a holistic discussion of 6G systems beginning with lifestyle and societal changes driving the need for next-generation networks. This is followed by a discussion into the technical requirements needed to enable 6G applications, based on which we dissect key challenges and possibilities for practically realizable system solutions across all layers of the Open Systems Interconnection stack (i.e., from applications to the physical layer). Since many of the 6G applications will need access to an order-of-magnitude more spectrum, utilization of frequencies between 100 GHz and 1 THz becomes of paramount importance.

As such, the 6G ecosystem will feature a diverse range of frequency bands, ranging from below 6 GHz up to 1 THz. We comprehensively characterize the limitations that must be overcome to realize working systems in these bands and provide a unique perspective on the physical and higher layer challenges relating to the design of next-generation core networks, new modulation and coding methods, novel multiple-access techniques, antenna arrays, wave propagation, radio frequency transceiver design, and real-time signal processing. We rigorously discuss the fundamental changes required in the core networks of the future, such as the redesign or significant reduction of the transport architecture that serves as a major source of latency for time-sensitive applications. This is in sharp contrast to the present hierarchical network architectures that are not suitable to realize many of the anticipated 6G services. While evaluating the strengths and weaknesses of key candidate 6G technologies, we differentiate what may be practically achievable over the next decade, relative to what is possible in theory. Keeping this in mind, we present concrete research challenges for each of the discussed system aspects, providing inspiration for what follows.

Excerpt: …Starting from the technical capabilities needed to support the 6G applications, we discuss the new spectrum bands that present an opportunity for 6G systems. While a lot of bandwidth is available in these new bands, how to utilize it effectively remains a key challenge. For instance, frequency bands at 100 GHz and above present formidable challenges in the development of hardware and surrounding system components, limiting the application areas where all of the spectra can be utilized. We discuss the deployment scenarios where 6G systems will most likely be used, as well as the technical challenges that

must be overcome to realize the development of such systems. *This includes new modulation methods, waveforms and coding techniques, multiple-access techniques, antenna arrays, RF transceivers, real-time signal processing, and wave propagation aspects.* We note that these are all substantial challenges in the way of systems that can be realized and deployed. Nevertheless, addressing these challenges at the PHY layer is only a part of resolving the potential issues. Improvements in the network architecture are equally important. The present *core network* design is influenced—and encumbered—by historical legacies. For example, the submillisecond latency required by many of the new services cannot be handled by the present transport network architecture. To this end, flattening or significant reduction of the architecture is necessary to comply with 6G use case requirements. The basic fabric of mobile Internet—the Transmission Control Protocol/Internet Protocol *(TCP/IP)—is not able to guarantee quality-of-service (QoS) needed for many 6G applications, as it is in effect based on best effort services*. These and many other aspects require a complete rethink of the network design, where the present transport networks will begin to disappear and be virtualized over existing fiber, as well as be isolated using modern software-defined networking (SDN), and virtualization methodologies. At the same time, the core network functions will be packaged into a microservice architecture and enabled on the fly. (673 words)
DOI: https://doi.org/10.1109/JPROC.2021.3061701

3. A Novel Simulation Framework for the Design and Testing of Advanced Driver Assistance Systems , by Florian Schiegg et al

Excerpt of article published in 2019 IEEE Vehicular Technology Conference (VTC2019-Fall)

Abstract:
The number and complexity of newly developed automated driving systems has been constantly rising over the past decade. Especially the introduction of vehicle-to-everything (V2X) communication is expected to further potentiate this development. In order to be deployed, the functional safety of the developed systems has to be assured previously. However, the testing in a representative number of field tests is costly and time-consuming. For this reason, virtual test drives have risen as an important option for design and testing of automated driving technologies, leaving only the final validation to test with real vehicles and thus significantly reducing the overall expenditure. The authors of the work at hand introduce a simulation framework based on the vehicle simulator CarMaker, complemented with the middleware platform Robot Operating System (ROS) and fed with real traffic data, which allows to automatically test advanced driver assistance systems for a large number of real world scenarios by varying topology, vehicle and communication parameters, among others. The simulation framework is then used to demonstrate the benefit of collective perception (i.e. sharing of on-board sensor data among nearby vehicles by V2X communication) for a vehicle merging into a freeway, with metrics such as the vehicle awareness on spot and the time it has to plan and execute its maneuver. (210 words)
DOI: https://doi.org/10.1109/VTCFall.2019.8891221

4. **Learning to Automatically Catch Potholes in Worldwide Road Scene Images,**

by Javier Yebes et al.

Excerpt of article published in 2021 IEEE Intelligent Transportation Systems Magazine

(Volume: 13, Issue: 3, Fall 2021)

Abstract:

Among several road hazards that are present in any paved way in the world, potholes are one of the most annoying and involving higher maintenance costs. There is an increasing interest on *the automated detection* of these hazards enabled by technological and research progress. Our work tackled the challenge of pothole detection from images of real world road scenes. The main novelty resides on the application of latest progress in Artificial Intelligence to learn the visual appearance of potholes. We built a large dataset of images with pothole annotations. They contained road scenes from different cities in the world, taken with different cameras, vehicles and viewpoints under varied environmental conditions. Then, we fine-tuned four different object detection models based on Deep Neural Networks. We achieved mean average precision above 75% and we used the pothole detector on the Nvidia DrivePX2 platform running at 5-6 frames per second. Moreover, it was deployed on a real vehicle driving at speeds below 60 km/h to notify the detected potholes to a given Internet of Things platform as part of AUTOPILOT H2020 project.

Pothole detectors researched were ultrasonic-based, accelerometer-based, image-based, and depth and vision-based. In another approach a 2D LiDAR and a camera were combined. (201 words)

DOI: 10.1109/MITS.2019.2926370