Deep learning saves researchers years in race to save teens from bad driving
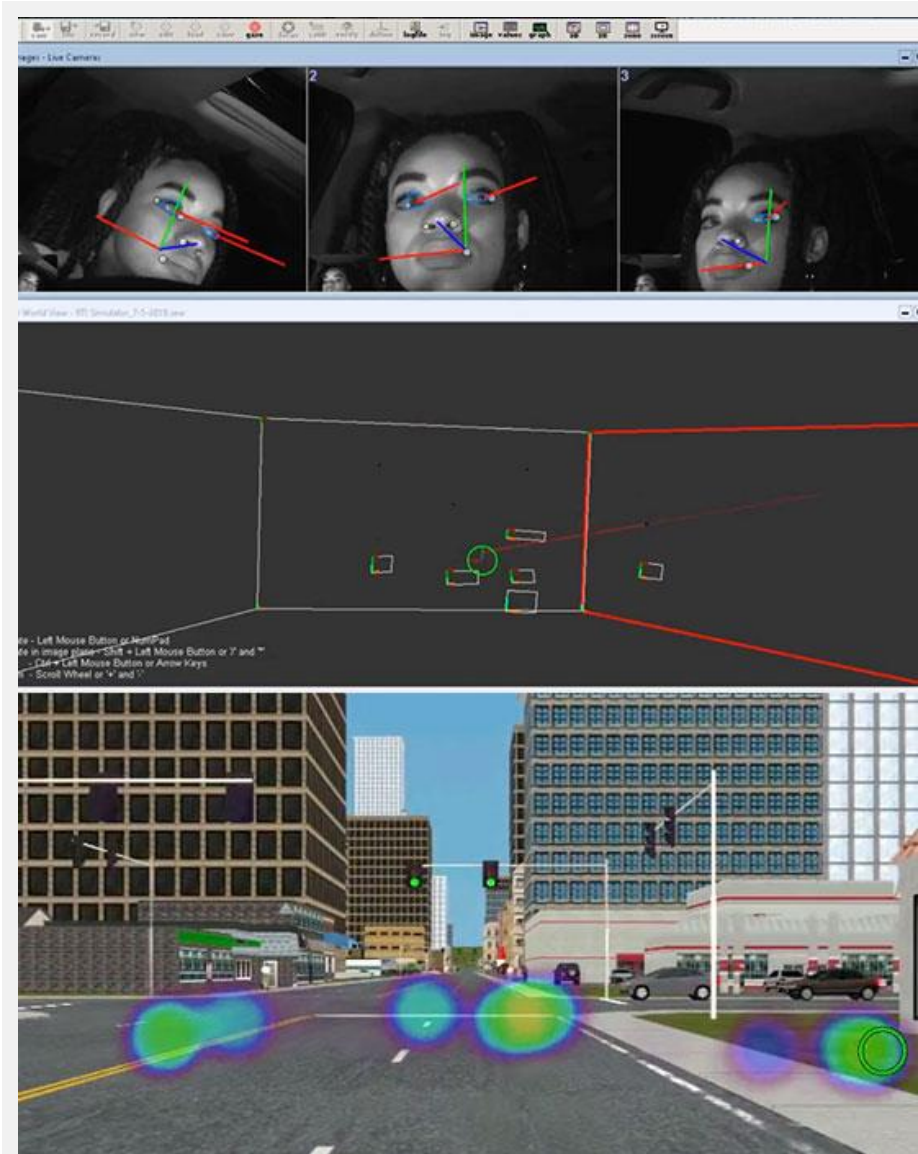
Matt Windsor

- May 17, 2021

A major NIH-funded study of teen drivers at UAB records participants's eyes as they progress through several simulated drives. This data is crucial to understanding how new drivers respond on the road, but the volume of information is massive. A new solution uses artificial intelligence to dramatically reduce the time needed to analyze all this data. *Images courtesy TRIP Lab.* Advances in machine learning and computer vision

have made it possible for cars to drive themselves. Some of the same techniques could help new drivers improve more quickly, according to experts in driving behavior and machine learning at UAB. **In a new paper**, these researchers describe how they are accelerating critical studies of teen driving behavior.

They aren't just shaving off a few minutes: using a machine learning technique called deep learning, they demonstrated how they could reduce the total time needed to analyze data from an important ongoing study by more than two and a half years. What takes a human 90 seconds on average, their system can finish in six seconds on a single GPU, the researchers report. And their vision doesn't end with faster analysis: it could eventually lead to an entirely new type of driver instruction.

Danger behind the wheel

New solutions to reduce teen motor vehicle crashes are desperately needed. The first six months that teens spend behind the wheel are the riskiest of their entire lives. Per mile driven, teens ages 16-19 **are three times more likely** to be in a fatal crash than drivers ages 20 and older. The younger they are, the higher the risk, according to research. For 16-year-olds, the crash rate **per mile driven** is 1.5 times higher than it is for 18-19-year-olds. In **this chart** compiled by the Insurance Institute for Highway Safety, it is clear that the crash rate per million miles traveled dips sharply from its peak at age 16.

Many states, including Alabama, have moved to graduated models of licensing in recent years, adding restrictions on when and how a newly licensed driver can drive in the first year or so after they pass their driving test. A **2015 study** that assessed the effect of age and experience on crashes among young drivers found that the youngest drivers had the highest crash rates, but the risk decreased after the first few months of driving. After six months of licensure, there did not seem to be any continued drop in crash risk. "Future studies should investigate whether this increase is accounted for by a change in driving exposure, driving behaviors and/or other factors," the researchers wrote in the journal Annals of Accident Prevention.

The full-body SUV driving simulator in UAB's TRIP Lab is the first of its kind in the world, giving study participants an immersive driving experience.

"Is it age or is it inexperience? We really don't know. Some of the early evidence suggests that experience outweighs age, but it is hard to parse that out based on existing studies."

Is it age or experience?

In other words, what's more important: a person's age at licensure, or their months of experience behind the wheel? Those are the questions that UAB researchers are trying to answer with a study called Roadway Environments and Attentional Change in Teens, or REACT. "Is it age or is it inexperience? We really don't know," said Despina Stavrinos, Ph.D., associate professor in the Department of Psychology and director of the UAB Translational Research for Injury Prevention Laboratory (TRIP Lab). "Some of the early evidence suggests that experience outweighs age, but it is hard to parse that out based on existing studies."

REACT has enrolled 190 teens in four groups: 16-year-olds with and without driving experience and 18-year-olds with and without driving experience. Every three months during a period of 18 months, each teen completes three virtual drives in TRIP Lab's state-of-the-art driving simulator: an actual Honda Pilot SUV surrounded by massive screens that project a highly realistic driving environment. The Pilot has no wheels; it is mounted on a motion base that rocks the simulator forward and backward while accelerating and braking. During their drives, participants are confronted with a series of potential hazards, including pedestrians unexpectedly stepping into an intersection, cars

abruptly entering the roadway from side streets and deer leaping out of the woods onto the freeway.
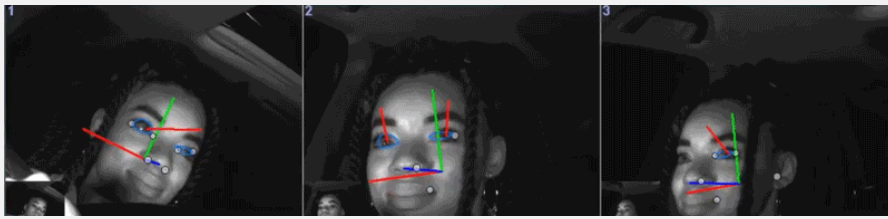
Software tracks the position of participants' eyes while they drive through a simulated world; distractions that cause participants to take their eyes off the road are part of the study. *Image courtesy TRIP Lab.*

Looking for trouble

Despina Stavrinos, Ph.D.

Stavrinos and her team don't just want to know if their teen drivers can avoid these obstacles; they want to watch what they are watching. Eye-tracking cameras mounted on the simulator's dash record participants' gaze throughout, and software overlays this information on top of the recording of the drive itself.

"Experienced drivers are constantly scanning the environment for danger," Stavrinos said. "Inexperienced drivers tend to focus on one point, usually looking straight ahead. They don't really have the background and experience. That's what we're trying to quantify."

Every time the participants drive, some of the potential hazards activate while others remain stationary but in view. These potential hazards are randomized so that participants do not memorize the pattern from one drive to the next. But during the 18 months the participants all see the same core set of hazards, "so we can compare driver maneuvers: where they look, when they look, how many times they look and how

long they look," Stavrinos said. "We know when a hazard comes into view from the annotation, and we can see when those vectors of gaze intersect with the region that's been annotated. That gives us a visual reaction time."

Eye tracking shows that experienced drivers (top) pay attention to the entire roadway and surrounding streets, while inexperienced drivers (bottom) tend to keep their gaze straight forward. *Images courtesy TRIP Lab.*

200 million frames

Benjamin McManus, Ph.D.

The study is funded by a $2.2 million grant from the National Institutes of Health. All along, Stavrinos knew that analyzing the data would involve a significant amount of time

from a team of coders — trained graduate and undergraduate students — who would watch the driving footage and note exactly when these hazards came into participants' paths. In practice, this means advancing through the video frame by frame at key points and using special software to carefully draw a polygon around the obstacles as they appear in the driver's view — then continuing to draw the polygon over that obstacle in each frame until it disappears.

There are just two problems. One is that TRIP Lab only has a single license for the expensive software (roughly $10,000 per license) needed to code and annotate each video frame, so only one coder can work on a single computer at a time. It takes about 90 seconds, on average, for a coder to finish each frame. Now for the other problem: nearly 200 participants times seven appointments times three simulated drives during each appointment generates lots and lots of individual frames.

"We want to know how driving changes over time, but a single participant's driving session takes up to 20 gigabytes," explained Benjamin McManus, Ph.D., assistant director for the TRIP Lab. Eventually, the data from the entire project could surpass hundreds of terabytes of storage on Cheaha, UAB's supercomputer and high-performance data storage resource. "We will easily exceed 200 million frames on this project," McManus said.

"It comes down to human factors issues, really," Stavrinos said. "A trained research assistant can't just stare at a screen and code frame by frame for hours on end. We encouraged our research assistants to take breaks frequently so their attention would stay on point. But it quickly became clear we had to do something more. That's when I started to explore what other options were out there."

Per mile driven, teens ages 16-19 are three times more likely to be in a fatal crash than drivers age 20 and older. The younger they are, the higher the risk, according to research.

An unusual collaboration

Thomas Anthony

Stavrinos had some exposure to machine learning as part of her work **on autonomous driving**, which uses object detection techniques heavily. Stavrinos and McManus went to a transportation research conference and noted that apart from autonomous-driving hotbeds such as Stanford University and MIT, machine learning had not made inroads in simulated driving studies — the only ethical method of placing drivers in potentially hazardous driving situations. The most realistic and highest fidelity driving simulators across the world "still use undergraduate assistants as manual coders," Stavrinos noted. "This remains the gold standard in our field. But we wanted to innovate the field and give machine learning a try, so we asked ourselves, 'Who can do this at UAB?'" To find out, Stavrinos and McManus did what everyone does in this kind of situation. "We Googled 'machine learning at UAB,' and when you do that, Thomas Anthony comes up," Stavrinos said.

Anthony is the founder-director of the Big Data and Analytics Lab at UAB and co-founder of the startup Analytical AI. Stavrinos contacted Anthony, who suggested they start with a pre-trained deep learning system and refine it for this particular use case. Anthony had students and researchers with the expertise to help, and they worked part-time with the TRIP Lab to begin the task. That led to a full-time partnership with one of Anthony's computer engineers, Piyush Pawar, who eventually became a TRIP Lab employee. Pawar is the lead author on their paper, **Hazard Detection in Driving Simulation using Deep Learning**, which formalizes a presentation they initially made at SoutheastCon 2021, organized by IEEE, the world's largest technical professional organization. "You really don't see this kind of collaborative effort between psychology and computer engineering in public institutions often," Stavrinos said. "This is truly special within the transportation safety field."

"A trained research assistant can't just stare at a screen and code frame by frame for hours on end. We encouraged our research assistants to take breaks frequently so their

attention would stay on point. But it quickly became clear we had to do something more. That's when I started to explore what other options were out there."

Layers and conspiracy theories

Deep learning, as the name implies, is a type of machine learning that relies on multiple, stacked layers. The layers are known as neurons for their superficial resemblance to the way a biological neuron functions: Give it a strong enough signal and it will pass that signal along, otherwise the signal will die there. A deep learning model trains multiple layers of these neurons, each of which learns to recognize some element of the feature of interest. For Anthony and Stavrinos, those elements are the digital pedestrians and vehicles that are programmed to interact with study participants while they drive.

Anthony's team proposed taking an existing model trained to recognize objects and then perform the training necessary to have it work on simulated images. "That is the standard practice in the field, because it is very tough to start from scratch," Anthony said. "You always start with a pre-trained network so you don't have to teach it basic concepts. The model knows what 'green' is, so you just need to teach it about a 'green car.'"

Training the model — technically, a mask region-based convolutional neural network, or mask RCNN — took a couple of weeks, Anthony said. As with all machine learning projects, the major time investment was in fine tuning the hyperparameters (settings) used. "That is all about finding the things you need to find and not overfitting to any particular case," Anthony said. "Neural networks are good at conspiracy theories. If they find a green car responsible for a crash, they can easily decide that all green cars are hazards. Neural network tuning is still considered as much an art as a science."

"You really don't see this kind of collaborative effort between psychology and computer engineering in public institutions often. This is truly special within the transportation safety field."

25,200 hours vs. 1,488 hours

As the researchers explain in their IEEE paper, the system can identify hazards and draw polygons in an average of six seconds per frame, compared with 90 seconds for human coders. In the version of the model explained in their paper, the researchers achieved 83.16% accuracy in correctly identifying frames that needed to be annotated in an early test sample of drives from 40 participants. The researchers have further refined the model to exceed 90% accuracy. Extrapolating those figures to the entire study, the researchers estimate that their model could annotate all frames in 1,488.2 hours, compared with 25,200 hours if done by a human coder. That corresponds to about two months for the machine, compared with just under three years for the human.

"Currently we are still refining the algorithms to get as close to the human annotator success rate as possible," Anthony said. "Human annotators are currently taken as the gold standard, but they have their own biases as well and can be slightly inconsistent or variable, even if they follow the annotation protocol to the T."

"We could teach poor drivers what good drivers are looking at in any given situation, so they can train themselves to act more like good drivers. That is potentially very powerful."

Implications for research — and training new drivers

That time savings could represent a significant financial savings, Stavrinos said. This could give the researchers freedom to do additional analyses and/or accelerate the translation of their data into published research reports. "We're definitely adding this to other projects as well," Stavrinos said. "It's already been applied to our studies of drivers with disabilities, and we are also planning to use it on our autonomous driving studies as well."

The model could be used to help other researchers. "This problem is not unique to us," Anthony said. "Eventually, once we build out the system, we could become a data-annotation clearinghouse. You can send us their files, and we can get them annotated for you. That would help researchers all across the country."

Ultimately, the deep learning model is not about annotating video frames but about reducing an overwhelming sea of data to some workable subset. "This is a data-reduction solution," Stavrinos said. "When you think scientifically where this could go, it could be applied to predicting good versus bad drivers based on a sample of their driving. It could be used to improve our understanding of distracted driving, especially under different levels of cognitive load."

Perhaps most intriguing is Stavrinos' vision of adapting the system to create what she calls "attention-training programs." "We could teach poor drivers what good drivers are looking at in any given situation, so they can train themselves to act more like good drivers," she said. "That is potentially very powerful."