SUPER DATASCIENCE
MAKING THE COMPLEX SIMPLE

SDS PODCAST EPISODE 906: HOW PROF. JASON CORSO SOLVED COMPUTER VISION'S DATA PROBLEM

Jon Krohn: 00:00 This is episode number 906 with Dr. Jason Corso, professor at the University of Michigan and co-founder of Voxel51.

00:19 Welcome back to the Super Data Science Podcast. We've got an exceptional guest for you today. It's Dr. Jason Corso. He's professor of robotics, electrical engineering, and computer science at the prestigious University of Michigan with over 20 years of research experience spanning video understanding, robotics, and AI. He's published over 150 academic papers in that time that together have been cited over 20,000 times. In addition to his academic work, he's also co-founder and chief science officer at Voxel51, a leading platform for visual AI development. His work bridges academic innovation and obviously real world impact earning him major honors that I don't have time to list, but there's a ton of them.

00:59 Today's episode skews a bit more towards hands-on practitioners like data scientists and ML engineers, particularly anyone tackling computer vision problems. That said, Jason is a charismatic and exceptional communicator, so perhaps any listener to this podcast will enjoy today's episode. In it, Jason details how his research spin out, Voxel51, is solving the biggest bottleneck in computer vision, the surprising way autonomous vehicles learn to handle accidents they've never seen, and why the secret to better AI models isn't better algorithms, it's something else that's hiding in plain sight. All right, let's jump right into our conversation. Jason, welcome to the Super Data Science Podcast. Where are you calling in from today?

Jason Corso: 01:38 Great to be here, Jon. I'm calling in from Western New York in Buffalo.

Jon Krohn: 01:41 Nice. It is a town that, it's the airport that I frequent most in the world. I love that airport. It is tiny, efficient. It's just the right size. It has everything you need, but you can get in and out of there very quickly, and so I frequently am going between New York and Toronto and Buffalo is my airport of choice. Lots of interesting, there's a famous art museum in Buffalo, isn't there?

Jason Corso: 02:09 There is. It recently had a renovation, it's called the AKG now.

Jon Krohn: 02:13 The AKG. That's right.

Jason Corso: 02:15 But it used to be called the Albright-Knox Gallery, and it was actually designed by a local architect by the name of E.B. Green. In fact, when I used to live here some 15, 20 years ago, I owned one of his houses as well and it was such a cool little house on the west side.

Jon Krohn: 02:32 Wow.

Jason Corso: 02:32 Buffalo is a great old city with old architecture and stuff like that.

Jon Krohn: 02:35 For sure. It's something that probably people don't expect if they've never been there, but it's stunning. And from the highway, you can look, there's a lot of great vantage points of the architecture in the city as you drive around it and it's beautiful.

Jason Corso: 02:50 Absolutely. Good people, good views, and we got the Bills as well, so there you go.

Jon Krohn: 02:55 Yeah, the Buffalo Bills, for sure. That was my football team growing up as well, getting that in Toronto. We'd get American broadcasts over the lake, over Lake Ontario, and so my local football team growing up was the Bills as well. So I lived through the, was it four or five years that the Bills made the Superbowl in the '90s and never won? Four.

Jason Corso: 03:18 It was four years, although it predates my Buffalonian status, but it still is painful for the region, I think. Although these days, the city rallies behind or the whole region rallies behind the Bills. But anyway.

Jon Krohn: 03:32 For sure. They're a strong club. So despite living in Buffalo, you are actually a University of Michigan professor. You're a professor of robotics over there as well as a professor of electrical engineering and computer science. Tell us a bit, Jason or Professor Corso about the work that you do over there at Michigan.

Jason Corso: 03:51 Right on. Please call me Jason. I live this dual life. I also have a family. My family's here, but I'm a 20 some year veteran of generally computer vision. The angle I take in computer vision is what I call physically grounded cognitive systems, so I'm interested in problems since my dissertation. My dissertation was called Techniques for Vision-Based Human Computer Interaction, and so we had cameras watching humans and the humans would do things and then that would create interaction scenarios. In fact, we built this thing called the 4D Touchpad. I think it was at a workshop at CVPR and maybe like 2003 or something like that, and it used gesture tracking and in some sense created multi-touch prior to there being iPads and so on. But that's the type of work that I've done over the years.

04:45 My research group has focused on areas like video captioning. We have one of the first, if not the first paper at CVPR 2013 on video captioning, we apply that nowadays to guidance of humans doing activities. I'm cooking a dish in my kitchen or whatever, I'm about to reach for the salt, but the recipe is for sugar and my AI can tell me, "Don't use salt, use sugar," or more socially relevant perhaps is an exciting project we have right now, which is applying the same type of AI agentic guidance type scenarios in rural healthcare scenarios. A major problem in the US and really worldwide is just a shortage of trained physicians. So our project is trying to enable the upskilling of RNs or physician's assistants or nurse practitioners who can go out into rural America in say a mobile clinic or whatnot and do anything from a cardiac ultrasound to deep vein thrombosis in the lower limbs with AI guiding them through every step of the process.

05:57 So it's an exciting area. I really enjoy computer vision. I enjoy the boundaries of computer vision and the fact that humans are looking to be, or we are trying to build systems that work alongside humans to upskill them or to basically create a better world in some sense. Maybe that's pie in the sky, but that's a real driver. We do things by humans for humans and of humans, so that's been a 20-year driver.

Jon Krohn: 06:27 And it sounds like cutting edge research. It sounds like it would be very impactful. It doesn't sound pie in the sky either. It sounds like a real tangible way to be making the world a better place with AI, which is perhaps the thing that we love most on this show above all.

Jason Corso: 06:43 Right on. Absolutely.

Jon Krohn: 06:44 You mentioned something there CVPR. For our listeners who don't know what that means, it's the Conference on Computer Vision and Pattern Recognition, and you can correct me if I'm wrong on this, Jason, but I think it's hands down, the biggest most important academic conference on computer vision in the world.

Jason Corso: 06:59 It's definitely one of the two. Yeah, so there have been two historically that are among the top, so CVPR is one of them. The other one is called ICCV, International Conference on Computer Vision. Generally, there are two conferences per year, and so CVPR happens every year. ICCV happens every other year, and then the third one is

ECCV, European Conference on Computer Vision, and that alternates with ICCV. But generally, those are the two key conferences in computer vision. There are many others that are amazing, especially now as the field has been exploding. In some sense, I wish there were even more conferences that were a little smaller, because I mean, when I was a grad student, CVPR had something like 500 papers max, maybe 1,000 attendees, actually probably even had less than 500 papers 20 years ago. Nowadays, I think there's like 2,500 papers on average every year, 10,000 plus attendees. It's great to see the growth, but it's also hard to know where to go, how to focus, what you're going to learn at the conference and so on.

Jon Krohn: 07:59 For sure. Interesting to hear about that growth over time. And so, with all of this experience with this 20 years of experience that you have in computer vision, you identified about a decade ago some key problems that could be solved in this space with a technological solution, and you founded, you co-founded a company called Voxel51. You were CEO of that company for it looks like about seven years, and then for the past few years you've been chief science officer. Tell us about Voxel51, how it got started, how it came out of your university research.

Jason Corso: 08:32 Absolutely. So as an individual, I identify as a creator, so I've always had one hand on the keyboard while I'm lecturing or whatever or meeting with someone because I just love to build things. And over time in the research lab, we began to notice that data was playing a key role alongside algorithmic or model work. Usually when you're an academic and you're writing a paper, generally that paper is going to be about a new model or a new algorithm, and although there have been some early works in data sets like Caltech 256 or in an ImageNet and so on, the number of data papers was significantly dwarfed by the number of model papers, and it remains true today. It's just the general, the mindset.

09:24 However, we began to notice that as model capabilities began to improve for a given problem, say object detection, even more concretely like pedestrian avoidance for autonomous vehicles, just as an example, we began to notice that you can pull a model architecture off the shelf, one of maybe half a dozen or so, and the performance you got out of the system you ended up training was more of a function of the data set you used to train that model than it was which of the six model architectures you chose.

09:54 And we basically began to build this conviction around data is at least as important if not more important than the model architecture you choose. And ultimately, Voxel51 grew out of that observation or that vision. This notion that wait, people need data. And it's not like we were the first to think this or the only to think this. But one of our initial mantras was better data, better models, and we truly believe in that.

10:22 And importantly, as a creator or a builder, there just was not enough tooling around how one works with data, how one analyzes data. As a grad student, I had this data set where I actually went into the cafeteria early one morning and took some photos of the layout. I was trying to do basically semantic mapping of the environment, and I think my data set had 100 images, maybe even a little less than that. So I could look and study the impact of any algorithmic modification on every single sample of that data set when I was doing this, whatever it was, 20-some years ago.

11:00 Fast forward to even just 10 years after that, 10 years ago when ImageNet came around, a million data samples, or actually the full ImageNet is 20-some million samples. Nowadays, you have 5 billion samples per data set, even in some open source data sets like the LAION-5B you may have heard of or I know the Florence-2 data set or Florence-2 model was trained on a 5 billion sample data set. It's just impossible. It was becoming impossible to basically put your eyeballs on enough of the data samples to build an intuition over when you work over with your model, how it's going to be impacted by the data and vice versa and so on.

11:37 So ultimately, Voxel51 is a company that tries to speed up the work you do with your data and the work you do with your models by providing the right dev tool in some sense for visual AI. We are an open source tool. We release the open source tool in August of 20... It's been a while ago, now that I'm forgetting. It's either 2019 or 2020. Don't quote me on which one of those it is. We have about 3 million installs of that or more, and we've always tried to have the IC, the ideal user, ideal customer of that tool is really a heavy technical data scientist or computer vision scientist. And so, it's super flexible and you can write plugins for it or extensions for it for the front end and the back end and anyway, great. So that's a quick overview of where we were and why we got started. I'm not sure if you have any questions about that.

Jon Krohn: 12:36 No, it's a great story on the origin. You identify a pain point through your expertise and then you're able to create a product to solve that pain point, and you've already had a huge amount of success with 3 million downloads and the GitHub repo, which so we'll talk about, I realize the open source is a little bit different, but it also gives some perspective on how important solutions like this are when the GitHub repo has 10,000 stars like your stuff.

Jason Corso: 13:03 Absolutely. Yep.

Jon Krohn: 13:05 And have you ever heard of, we actually recently in an episode, in episode 901, which came out a few weeks ago, we had someone on the show named Lilith Bat-Leah who, she runs workshops at ICML and ICLR, two other big machine learning conferences, which obviously you know, Jason, but just for our audience. And so, she runs working groups at those on data centric machine learning. So it's a DMLR, data centric machine learning research. It's like the acronym that's used there. Have you come across that acronym before? It sounds similar to what you're describing.

Jason Corso: 13:48 Absolutely. I mean, data centric ML can mean a lot of things, but even at Voxel, we used to use that in our outbound community driven marketing material as well. I say it can mean a lot of things, not to put it down, but it's critical to recognize that when we think of building technological software systems, we think of writing code, and then we hope that we can debug a software system like the tool we're using right now to record this. But when you think of machine learning, there's code and then there's the data that goes into the code. In some sense, it gets transformed into data weights or coefficients or something like that. But these two things are inseparable. And as the evolution from what some folks have called software 1.0 just code the software 2.0, which is essentially just a different type of code, it's just humans can't really write it. We write other code to train it from data.

14:53 So data has been used to train up machine learning systems for decades now, but I think there's just when we collectively as a research community or a user community, think of data-centric machine learning, I think what I was saying earlier where we tend to emphasize that await, it's not just that there's data in this code and the code is more important, it's that you really cannot separate or divorce the two things. The data and the structure of the model structure, even like the ops underneath it, these are wed together in a way that is critical to understand all facets. And if you really want to build a successful ML system or AI system, you really need the right tooling around analyzing the data, analyzing the models, analyzing the ops, and they do need to work in concert so that you have a good sense of what's going on.

Jon Krohn: 15:42 It seems really obvious when we say it out loud like this, but it is amazing how much attention in terms of new releases go onto some exciting new model. But in something like this, in computer vision where you think about it is probably easy for users to imagine something like the machine vision problem of an autonomous vehicle where you have sensors on a vehicle that are driving through streets and you wanted to make safe driving decisions obviously as close to 100% of the time as possible. In that kind of scenario, it makes it easy to imagine how the model, if it has some percentage improvement over some other model, that's great, but the model isn't going to be valuable at all if you don't have data covering the whole gamut of situations that that autonomous vehicle is going to run into.

16:36 If you only train the vehicle on situations where there's no car accidents, there's no possible world where that AI system could know how to handle seeing an accident happen right in front of it. And so, hopefully that's a very simplistic example that I just gave, but hopefully that it allows us to visualize pretty easily the critical importance of data in having any AI model work effectively in the real world, which computer vision systems I think basically always are operating in the real world.

Jason Corso: 17:09 So I mean, I think you hit it on the head really, and I think it's a great example especially because I have an 18-year-old, recently taught her how to drive and so on.

I'm wondering how many miles does she have to drive before she's going to see even a near miss or even just a situation where a kid runs out into the street. So I think the number is something like it's in the tens of millions of miles driven for every accident that's recorded by the US government or NHTSA. And so, actually finding the hardest part about this world of building highly successful, 99.999 whatever percent accurate systems is getting the data, then getting the data labeled and then train the model and figuring out what are the failure modes, what are the success cases, what are my failure modes and where do I need to add more data and begin this process?

18:08 So actually something I'm really excited about at Voxel51 right now is this new direction we've taken our product, I think annotation companies, the problem of taking raw data and labeling boxes or labels or classes or whatever on your data samples to train the machine learning algorithm on it, those were probably the first wave of companies in computer vision, at least in modern computer vision, machine learning based computer vision. But Voxel51 never identified as an annotation company. We're always, in some sense, we explicitly decided strategically, we are not an annotation company. We actually don't even support, you can do a lot of things in 51 including load varieties of different labels into the software tool and visualize them and so on, but until this coming summer, you will not have been able to edit them in our tool. We were so against it almost like we had almost like a one button mouse challenge that Apple had over the years.

19:05 And nowadays, that annotation problem has been so central, but I think it's transforming based on these decades of progress. And so, with performant foundation models, now we have this tool, this new product line called Verified Auto Labeling, which can take this raw media, automatically generate labels on it via foundation models, and obviously there's going to be a spectrum of performance for certain classes that the foundation models have seen a lot like pedestrians or bicycles or other vehicles. It's going to work pretty well.

19:46 For other scenarios like teddy bears or certain types of hats or coffee mugs, whatever, maybe it'll work less, but the critical aspect and what's I think really exciting from our perspective is the V in the verification part of that. Well, our workflow is you take your raw media, you apply foundation models and we have a battery of them you can apply against it, and then we have our custom ML that will rank the outputs from those foundation models so that you can in batch have high confidence that you're automatically going to accept something like 70% of them, and already that's a huge amount of money that you're saving in time and so on. And then, for the remaining 30%, we can still rank them again so you can have your human labels or human QA people only spending time on the challenging scenarios, the scenarios like the ones you're pointing out like the corner cases, the hard cases that we really need humans to look at and humans to verify and so on.

Jon Krohn: 20:46 Very cool. So this Verified Auto Labeling, this new initiative that you said summer, so we're talking Northern Hemisphere summer for our international listeners, so it's like around the time that this episode is coming out is when this Verified Auto Labeling starts to be something that people can use in Voxel51?

Jason Corso: 21:02 Yeah, actually it recently went into alpha with some customers, so it's already in alpha and we are improving it and it will be in every release over the coming two months probably. It will get more functionality and more users to adopt it.

Jon Krohn: 21:18 Fantastic. And it sounds like this is then solving what is the biggest bottleneck in computer vision, and you're doing that using intelligent techniques so that it makes it way more time efficient and cost-effective orders of magnitude relative to having humans be annotating the data.

Jason Corso: 21:37 Absolutely, yeah. I mean, the tagline that I, not approved by marketing but that I like these days is curation is the new annotation. I mean, annotation 1.0 if we want to use that analogy was basically, I don't really know how to filter my data, so I'm just going to send it all to humans to label and I don't have to pay for all of that and it's time-consuming as well, and then I'm going to get it back and give it to my machine learning engineers and maybe 1%, maybe 10% of that's useful. It's hard to say. You don't really know. It's like walking in a dark hallway without any light switch on. Which door are you going to try?

22:16 I think nowadays, we're probably in the annotation 1.5 era, where it's obvious to apply a foundation model even for something like pre-filtering just so you can rank your data, so you're going to send it to humans to label. And I think Verified Auto Labeling is farther along the line in saying, "Wait a second, sure, pre-filter your data, only apply certain things," but we're also saying, "Wait, you still don't have to have humans annotate everything." You can rank them, filter them, and so on. So you can just accept the automatic labels out of the box.

22:51 And where do I think we're actually going? What is actually annotation 2.0 and this is not what we're releasing this summer, but it's likely coming in the future if I have my way, is this notion that instead of the humans asking the foundation models what they should label or what the labels are or what have you, it's more agentic where there's a problem statement given behemoth amount of unlabeled data, and then the models are able to actually ask the humans questions just when it's necessary and it's more driven by the AI agent, if you will, so even fewer less human involvements needed.

Jon Krohn: 23:30 That is awesome, Jason. So it sounds like Voxel51 has figured out how to leverage the latest technology in terms of what we can do with automation to allow people to get the highest quality data for building high performance computer vision models at a fraction of the effort and the cost. That's cool.

Jason Corso: 23:51 Absolutely. Nail on the head right there. Yep.

Jon Krohn: 23:54 Nice. Fantastic. I mean, this is a Friday episode so it's not amongst our longer episodes. And so, that was a short but very rich episode with you, Professor Corso, Jason. We could certainly, I mean we could easily have had an episode that was like a Joe Rogan style three or four hours with you, I'm sure on computer vision if we had to because you have such a rich understanding of the space. Before I let my guests go, I always ask them for a book recommendation, what do you have for us?

Jason Corso: 24:32 Cool. So I am an avid reader and I was just saying in my head, if you want to do a three or four-hour episode one time, maybe when we're both stuck in the Buffalo Airport and it's snowstorm, we can record then. But avid reader here, I guess one of the best books I've read in the last few months is a book called Quit by Annie Duke, and it really puts the, I'm a hard worker, I'm a grinder, so I'm always one to really just want to see a project through. It really puts that type of grit, which every founder needs in some sense, and every professor really needs these days as well up against this notion that make sure you're being smart about how you spend your time and how you're planning, pre-planning when you might want to deep switch or quit an angle and go on a different angle. I think for really any adult, I think the lessons learned that she writes in this book are fantastic.

Jon Krohn: 25:36 I like that a lot. It's tricky for those of us, probably a lot of our listeners, if the way that you choose to spend your free time is listening to a technical podcast about data science and AI, you're probably somebody who has a lot of grit and is really pushing their career. But this kind of thing, it becomes important, especially as opportunities accumulate as you focus and have more grit, more and more opportunities come up and you can't keep doing everything. It's a tricky thing. Even just things like I used to be pre-pandemic, I used to be able to be inbox zero and respond to anything that should be responded to. And that's a simple, silly example. I mean, it sounds like this quit thing is more about big strategic decisions, but just figuring out what you have to let go so that you can make space for even bigger things.

Jason Corso: 26:27 Absolutely. I mean, there's no, in terms of the strategy and the tactics of decision making, there's no thing too small to think about, frankly. So the notion of email inbox or inbox zero, whatever, is highly relevant. I think the way I would put it is one of our investors has used the term indigestion. If you're so successful, you're going to get indigestion over just trying to do too much, and it's definitely something to watch out for, I think. And the author, Annie Duke really does a good job of explaining that.

Jon Krohn: 27:02 Nice. Sounds like a great book. And then, so for people who want more literary recommendations from your avid reading brain or more insights on what you're up to with Voxel51, computer vision research, where should they be following your work?

Jason Corso: 27:17 Absolutely. So I guess number one would be to follow me on LinkedIn. I do try to post two to three times per week, various opinions and so on. Also, you can find the 51 open source repository at github.com/voxel, V-O-X-E-L-5-1, /fiftyone, the word fifty-one, and you can also find me on Bluesky at Jason Corso as well.

Jon Krohn: 27:44 I knew you'd be on Bluesky. I called that before we even hit the record button when I said that this question would come off at the end. I was like, you have the right profile.

Jason Corso: 27:53 Absolutely.

Jon Krohn: 27:53 To also have a Bluesky account. I'm going to have to get in there at some point. I'm not academic enough anymore that I have to have one, but I'd like to still pretend that I'm academic enough. Nice. Thank you so much, Jason. This has been an awesome episode. Thanks for joining us, and I look forward to that three to four-hour episode that we record in the snowstorm.

Jason Corso: 28:15 Right on. It was great to chat, Jon. Thanks for having me.

Jon Krohn: 28:18 Thanks to Jason Corso for coming on the show and providing such an informative and entertaining episode. In it, Professor Corso covered how he discovered that model performance depends more on the quality of training data than on the choice of algorithm architecture leading to Voxel51's founding principle, better data, better models. He also talked about how Voxel51's Verified Auto Labeling uses foundation models to automatically label data, then ranks output so teams can accept about 70% automatically and focus human reviewers only on challenging edge cases. Overall, saving massive time and cost.

28:51 He also talked about how we've moved from annotation 1.0, sending everything to humans through annotation 1.5, where we pre-filter with AI and are now approaching annotation 2.0 where AI agents actively ask humans questions only when necessary. All right, I hope you enjoyed today's episode. Be sure not to miss any of our exciting upcoming episodes. Subscribe to this podcast if you haven't already, but most importantly, I just hope you'll keep on listening. Until next time, keep on rocking it out there and I'm looking forward to enjoying another round of the Super Data Science Podcast with you very soon.