EPISODE 1117

[INTRODUCTION]

[00:00:00] JM: Image annotation is necessary for building supervised learning models for computer vision. An image annotation platform streamlines the annotation of these images. Well-known annotation platforms include scale AI, Amazon Mechanical Turk, and CrowdFlower. There are also large consulting-like companies that will annotate images in bulk for you. If you have an application that requires lots of annotations such as self-driving cars, then you might be compelled to outsource this annotation to such a company.

SuperAnnotate is an image annotation platform that can be used to build these image annotation outsourcing firms. This episode explores SuperAnnotate and the growing niche of image annotation. Vahan and Tigran Petrosyan are the founders of SuperAnnotate and they join the show for today's interview. And if you are interested in helping out on Software Engineering Daily, helping with writing and research and preparation and articles, send me an email, jeff@softwareengineeringdaily.com. Send me some of your writing samples. Also, I am an investor. I'm looking for investments. If you're building a software company and it's especially around infrastructure or developer tooling, I'd love to hear from you. Send me an email, jeff@softwareengineeringdaily.com.

[SPONSOR MESSAGE]

[00:01:20] JM: If you are a retailer, a big sales day like Cyber Monday can make or break your business. If you sell accounting software, the tax deadline day is like your Super Bowl. And if you're a sports broadcaster, then the Super Bowl is your Super Bowl. Every company has days or seasons that are more critical than the rest. If your systems are ready for the moments that matter most to you, then you're going to be doing much better. And that's the theme of this year's Chaos Conf, the world's largest chaos engineering conference. Chaos Conf is a free online conference taking place on October 6th through 8th, and you can register at chaosconf.io.

Ever since the first Chaos Conf in 2018, the objective has been to create a community around resilience and SRE best practices. Attendees range from having a decade of experience to those who are totally new to chaos engineer. And this year's keynote speakers are Gene Kim, and has been a guest on the show several times; and Adrian Cockroft, who's the VP of cloud architecture strategy at AWS. There will be 20 sessions for all experienced levels focusing on the practice of reliability, completing the DevOps loop and how to build a data-driven culture of reliability. You can register for free at chaosconf.io and the first 1,000 registrants will receive a limited edition swag pack. Claim yours at chaosconf.io and be prepared for the moments that matter.

[INTERVIEW]

[00:02:51] JM: Guys, welcome to Software Engineering Daily.

[00:02:53] TP: Yeah, pleasure to be here, Jeff.

[00:02:55] JM: So we all know that there is a lot of applications of computer vision. There's self-driving cars, facial recognition, medical imaging. But in order to make use of that data, we need to annotate it. Can you explain what annotation, computer vision annotation consists of?

[00:03:15] TP: Right. So you're right for all these type of applications. If you want to solve a certain object detection problem, let's say for autonomous vehicles, you want to detect cars, or trees, or pedestrians, you need to label images. Annotation is basically selecting objects and naming them in the images. More of the annotations currently is done manually. So humans in many parts of the world, currently, there are hundreds of thousands of people doing this. Going to the images, selecting objects and images, like in Photoshop, you would select some object. And then you would name it with a corresponding object name.

[00:03:55] JM: And how does that annotation get done at scale? Because there are lots and lots of images that need to be annotated.

[00:04:03] A: Right. Usually these companies, what they do is they hire outsourcing teams. For example, there are companies in India would have thousands of annotators. Sometimes they can reach to tens of thousands of users who are assigned to those images and they start doing this. When you do that on scale, then you have a large scale of data.

Of course nowadays there are more and more solutions that try to automate this part more and make the human interference less and less. Basically, annotation of images start at the point where the computer already predicted some level of annotations. And then humans have to approve. Or if the detection accuracy is, let's say, less than a certain point, then human have to interfere.

[00:04:50] VP: And just to add maybe what Tigran was saying. So the software is becoming very essential to annotate those large scale annotation projects. Otherwise, if everybody is sitting on their homes and then doing the annotation on a local computer, then this project management and analytics is becoming such a big issue that it's going to be impossible to do a high-quality annotation even with a lot of AI and ML.

[00:05:19] JM: In order to annotate images, the image needs to be divided into bounding boxes or it needs to be segmented. Can you explain the role of object segmentation in having computer vision be annotated?

[00:05:41] TP: There are different types of annotations, like one is the boxes. Putting boxes around the object. Another one is putting some polygons, points around the objects to be as closely resembling as possible. Another one is segmentation. We call is semantic segmentation or panoptic segmentation. And this is the case where you basically segment images in different parts to make sure that you can annotate faster or to make sure that you segment those in a way that it grasps the object completely within those segments.

This way, this is what we do in our company as one part of the annotation to make sure the annotation is as fast as possible. And this was Vahan's PhD research when he was doing it in Stockholm to make sure that this segmentation is as fast, as accurate as possible to make sure that we can accelerate this pixel-accurate annotation tasks. Vahan, maybe you can elaborate more on this.

[00:06:44] VP: When you link boxes, basically, you're approximately saying that there is an object in this area. If you're doing more precise annotation, then in those cases they can have like different use cases. One of the fascinating use cases for me was when you have a lot of satellite images and then you have a lot of power lines, for example. So you want to predict how trees are coming close to the power lines, because that can just shut down the entire electricity of the whole city or village.

What you want to do is you need to have like infrastructure inspection where you can predict pixel-accuracy. How those trees are coming nearby the power lines? So this is when the pixel-accurate annotation is coming to help basically, because you cannot truly detect if it found your boxes. Yeah. Those are like different use cases of annotations.

[00:07:46] JM: The people who are annotating images, these have historically been Mechanical Turk, Amazon Mechanical Turk people. Is that still the case today or is there some different mechanism for finding people to label images?

[00:08:03] TP: Right now, there are many outsourcing companies already emerged in the last few years. Yeah. So starting Amazon Mechanical Turks was a way to go, but eventually more organized, more managed annotation teams emerged because the annotation tasks become more and more complicated, and there was a learning process needed for each annotation task. And they have to be managed in order to get to really high-level of accuracies.

And this is why many of the companies who have from hundreds to thousands of annotators, they help these companies to get to the high-level of accuracies. And these companies of course need tools. And in order to get those tools or the companies have some tools that they have built internally or they need some other tools that they can help them. This is where the tools like us come into play.

[00:08:59] JM: I see. So you are a set of tools that could be used by these pools of labeling people.

[00:09:08] TP: Right. Basically, not only them, but also the computer vision companies together. Because we let the annotation to be streamlined between the data scientists, computer vision engineers and this large annotation service groups. It's very important to organize and manage this process, because you need to make sure that the quality of annotation and speed is in a right level. And on the other hand, we also allow some machine learning capabilities such as training the models within our tools. Every time someone does, let's say, 100 or 1000 annotators, they can immediately train within our system. And then they can predict those objects in the next thousand images, for example. This way we want to accelerate this process as much as we can.

In addition, we help the customer, the end scientist, also to make sure they're annotating the right data. There is something called active learning where it's very much important to understand which data needs to be annotated in order to get to the high-level of detection accuracies. For example, you can annotate every single frame in the video, but you might not get a high-level of accuracy rather than, let's say, annotating certain frames from that videos. This is where we also help these companies to decide which data to annotate.

[00:10:35] JM: One point of comparison just from a company perspective that's worth making is to scale AI, or course. And I think the scale company as being a fully vertically integrated system for both the labeling software and the people who are annotating the images, the Mechanical Turks, as it were. How would you differentiate yourself from the fully integrated platform of scale?

[00:11:05] TP: First of all, I think scale's main business model is providing services. So they have large groups of people that will do the services. What we do differently is we also provide the platform for the companies who already have their group of annotators. If they need services, we also help them to find the right service groups that will do annotations for them. This is one part.

The second is probably our segmentation technology that was Vahan's PhD research, and this was the tech that helped us to accelerate this pixel-accurate annotation process by at least 5 to 10X for certain tasks. This was our unique value proposition when we started to entering this market. And eventually this also became one of the core technologies we have.

Overall, yes. Our platform is quite open. There is a free version that people can register and use. And then once they're ready to go to a starter, or pro, or enterprise package, we can basically let them do that. It's more open to the community and we also want to help the researchers. Since we came out of universities, we are partnering with universities and letting them annotate data for free within our platform. This is another part that we want to basically give back to the community of academia.

[00:12:33] VP: Just to add something what Tigran is saying, our platform is like open for computer vision researchers on one end and then those annotation companies on the other end. So it's acting like a marketplace where it allows you to find the service you need because if you have very, very special expertise and you know some group of people that are highly skilled in medical annotation, let's say, then you're not going to be able to use scale as platform because we cannot truly find those highly specialized people to do the annotation for you. In this case, we're agnostic to either data scientists, or conservation engineer, or the annotation service.

[00:13:20] JM: Tell me about how SuperAnnotate fits into an overall application workflow for labeling data? You could pick any application. Maybe satellite imagery, or self-driving cars. Whatever you like.

[00:13:36] VP: Yeah. Maybe I can give like an example with the garbage collecting companies, for example, when they're sorting the garbage. That's like a very, very kind of application that it's becoming more and more automated by conservation, where when you have a garbage and then you need to sort out different garbage in different bins. But then that part when you're trying to automate, usually a camera is taking the video that you need to take specific frames from that video in a smart way to send it to our platform automatically.

Once that is sent to our platform, our annotators will get alerted and then they will see the images that are assigned to them so they will go and annotate. And then once they're done with them, then they will send back that annotation to our client immediately. And then imagine if you have a beta that is coming in a streamline when every single minute there is new data, then you want to have all these close-globe system where you need to annotate more and more data at the same time, you need basically to improve your model. The companies keep getting more data and improving their model.

In all that life cycle of computer vision, when they're getting the data from the video towards improving the computer vision model, the annotation is a very, very important part of that human in the loop process. And we're kind of connecting towards that end-to-end platform to allow this conservation application to build very quickly and seamlessly.

[00:15:32] JM: Tell me more about labels. If I have a range of labels that I might apply to a certain image, how am I selecting which label to apply to that image?

[00:15:44] TP: Each use case sometimes is quite unique. Even in autonomous vehicles, usually every company has it's unique way of annotating. Maybe one just wants to annotate cars pixel accurately. Another one is putting boxes. Another one might even put the color of the car or like – I don't know, some other details there. It's very unique. Every user decides which way they need to be annotated in order to solve their unique object-detection problem.

The way they decided is basically very much depend on what's their business use case, what they want to achieve in a short term and long term. They also have to consider, for example, what level of detection accuracy they need. And then based on that, they need to understand how much data they need to annotate for that. Our experience shows that the way they label data may vary so much even within the same industry. This is very, very interesting and fascinating for us to see across.

[00:16:50] JM: Let's take a step back. Give me your respective perspectives for how computer vision has evolved over the years and where we are today.

[00:16:59] VP: Yeah. Basically, the modern computer vision pretty much started from the [inaudible 00:17:05] dataset, which was collected, if I'm not mistaken, in 2009. And then from 2012 when the deep learning came across, people just trying to get one label from the data, telling that this image is a cat image or a dog image. Once that deep learning algorithm start to involve, then they started predicting the location of different objects. This is where the bounding proxies were used.

At some point then they started doing pixel-accurate annotations, which was a – Before it was very expensive together. So there was not that much data to learn. Pixel-accurate annotations became like very popular few years ago. And now that stage is evolving even further when people are not truly doing pixel-accurate annotation, but then they're trying to tell their action. For example, if you're annotating the human, you can put a human on it or you can predict all different joins on the human. This is like human process estimation where you can basically predict what the person is doing. This is kind of becoming very, very important problem on them.

A new emerging problem that is very, very relevant at the moment is trying to build relationship between different objects, because you can say that this is a hand and then this is a knife. But what's the relationship between a hand and knife? And then since the industry is evolving and then there is more details that can be predicted from the image, then that's kind of how I see the history is involved and can then, as a result, annotations, as Tigran was saying, is become more and more complicated.

[SPONSOR MESSAGE]

[INTERVIEW CONTINUED]

[00:20:44] JM: You are pursuing a PhD before founding SuperAnnotate. How did your research topics lead to what you've done with SuperAnnotate?

[00:20:57] VP: Yes. It's an interesting story to be honest. My professor really was a world-class researcher. Is a world-class researcher in reinforcement learning. Unfortunately, I didn't really go with his path. And then he was extremely nice enough to just let me to pursue my career and research in conservation. As a result, when he gave me this freedom to do my own research, I had found a very big passion in image segmentation, which resulted like a lot of different application. At the beginning, I was wondering maybe I will apply my take in image editing. But then that didn't really end up being a good solution. I found that annotation is even a bigger pain than just editing a few images, if you want. That's how I found the image annotation space after attending several conferences in conservation field. And as a result, after seeing the pain on my side and also from all the others, researcher side, we decided that maybe it's something that we shouldn't really pursue a PhD. And I asked Tigran to kind of drop-out from his PhD, and then together we found the company.

[00:22:22] TP: Maybe add a little bit on my side. I was doing my PhD in biomedical imaging. I was in Switzerland at the time. Just a few months left before I graduated. I was very excited about machine learning automation. I even gave a TEDx Talks in Switzerland at the time to show how automation is actually helping in the long run the humans to be faster, more effective, smarter. Basically helping improve our lives. So I was kind of a big advocate of that.

Eventually when I saw this opportunity with Vahan's technology, with opportunity huge market, opportunity how we can help all these different industries to make the annotation faster, more efficient. This was super exciting for me to join Vahan. And then help with my little experience from like biomedical imaging, and general image processing, machine learning, and also some business related expertise that I've gained before to basically help drive the company. And this was the time that we both dropped out of our PhDs and started the company.

[00:23:32] JM: And what was your process for finding people to use SuperAnnotate? What was the go-to-market process?

[00:23:42] TP: Initially, when we were building the product, it was very internal. So we had the product that was working well, but only internally. We didn't publish it until early this year. Before that, we were providing annotation services. We had the group of annotators in East Europe that we have collected, gathered and trained. And then we were providing services. And now our developers were sitting next to the annotators. While they were annotating, they were helping developers to build the tools that they need in order to annotate better and more collaboratively.

Eventually, when this process came to the point, we realized that actually we can release this tool for all the community to use it by themselves without our intervention. And this was the point that we decided to release the platform. And in addition to the services, we could also provide a platform for their internal use.

[00:24:45] JM: Now that you've had some time working with a variety of different companies, what are the common problems that you see teams encountering when they want to manage annotation tasks?

[00:24:57] VP: One of the main problems is how to have an efficient quality management system. When the first round of annotation is gone, how do you make the second round of annotation a lot more efficient and then a lot more accurate? Quality assurance can take almost the same time as the annotation, the first round of annotation itself at the beginning. And this is a word we see a lot of opportunity that other tooling providers don't really have. And that's kind of a place where I see we can do a lot of work to improve the software and then eventually to get the highest quality of annotations.

[00:25:47] TP: Maybe another point to add here is since the COVID hit a few months ago, most of these service companies, they were doing annotations in a same room. So it was easier to manage. There were people always talking together, understanding the tasks, helping each other to annotate. Since these people started to work from home, and managing large groups of people will annotate and basically QA and administer these jobs became very complicated. This is where we also saw a huge opportunity, because our platform fits very well to organize the work of the large groups of people without really having a presence. Because the platform allows a very flexible user management system where you have different roles, from annotated QA, admin, customer level, QA level, etc. You have data management system. We have automated task distribution system and some machine learning capabilities in order to speed up the process. I would maybe add also some chat communication system, which is quite important to make sure that the quality is high in this process.

[00:27:02] JM: So you're describing essentially one of the problems being project management or workflow management around the annotations?

[00:27:10] TP: Right. This is one of the big problems, especially now when people have to work from homes.

[00:27:17] JM: Are there tools that allow you to better manage the annotation workflows? Is that something that you've built into SuperAnnotate?

[00:27:25] TP: Right. This is one of the main features we have built in the platform where if you have thousands of people working in a same project, and within them could be like let's say 1,000 people are annotators the first level. And then let's say 300 people are doing QAs. And there are few admins in the customer side or a few admins in the annotation management side. So we have a system that organizes that all. And it's all automatic.

For example, annotator does the first round and then it sends to the QA. QA does the corrections. They can leave comments or do corrections and then send it back to the annotator. Or if they agree, they can do thumbs up, and then it goes to the admin. And then ultimately admin can confirm or send back to the QA, etc. And then there is a very robust analytics tool built where you see how each annotator is performing. Whether they're active? What's their speed of annotation? How many hours they worked? All this is tracked within our dashboards, where if you're a manager, basically you don't really have to see people working at that moment. You just look at the screen of the dashboards and you see everything what's going on in the project. This is very, very robust that people really love in our platform.

[00:28:50] JM: It's cool. So you have this way where you can partner with annotation companies. Can you tell me more about the workflow for these annotation companies? I mean, they have like tons and tons of people who they hire to do the annotation. So what's the partnership and the onboarding look like for these different companies?

[00:29:16] TV: When it comes to these service companies, we partnered with them in different ways. So the first is we partner with them as a service provider. For example, whenever we have a customer who needs a large group of annotators, we basically match them with a service provider. We do some organized pilot projects with them. And then let the customer check the quality, speed, etc. And then decide whether they want to go with them or not. And then also we let the customer to do testing with different companies that had experience in their space. And we help both sides to find the right fit, like customer and the service provider. This is one way. This was the part of the marketplace of annotation that Vahan was talking about earlier.

Another way of partnering is since these service companies, they have their own clients. But the clients don't have their tools. So the service provider need to provide tools for their clients. And this is where we come in. We provide our tool to these large groups so they can manage all their work to provide service to their client. That's the main two ways that basically we help the service providers. When it comes to the final customer, the computer vision-based company, they basically decide we help them to find the right annotation group that will do the best job for them. And then when it's done, both sides are happy basically. The annotation service provider found a client. And then the company got the right fit in order to finish the work in a highest quality and speed manner that they want.

[00:31:03] JM: I'd like to talk more about the actual usage of your platform and how images are annotated and what happens after the image annotation. Let's say we've got an image of a street and it's taken from a self-driving car. And that image lands on the computer of somebody who is responsible for annotating that image. What is that person seeing and what's the workflow for annotating the image and where does it go after being annotated?

[00:31:36] VP: First of all, the annotator need to often read like 40, 50 pages of instructions how to do the annotation. And once they see the image, then they should have clearly followed the instructions of how to annotate. And once they are sending to their QA and QA approves, then it can directly go to the customers, either S3 bucket or whatever the data is sitting in.

So, this way when we do the annotation, then we can allow them to just send the annotations directly to the client. And then in this way, customer doesn't really lose any time to get the annotation. Then once the annotator is finishing the job, then basically it can put the training right away. So if you have a training already put in your cloud and then you can basically wait until the annotations get more and more. So if you leave your training on, so to speak, then you'll get higher and higher accuracy just because you get more and more data that comes to your bucket or whatever cloud place you have.

[00:32:55] TP: It's basically an iterative process where our annotator does the annotation. It goes back to the customer. Once they get enough data, they can set training. They will create new types of annotations or new data to streamline to our system. And then it annotated again. Improves their models back. They do another training. And then this process iteratively continues until the customer is happy with their detection accuracies.

[00:33:23] JM: Tell me more about the engineering behind SuperAnnotate. Maybe you could start with the software architecture from a high level.

[00:33:31] VP: Overall, the software architecture, everything is built on the cloud. There is no need to do any installation or anything. We have on a highest level, we have something called teams. Each team consists of group of people. And then within each team, you can have project. As many projects as you want. So that's like the second layer of architecture that goes down for each team. And then each team consists of members and each project consists of different members of the team.

And then basically when you're getting an image for a certain project and you're assigning the images within the team members and within the project members, and as a result, there is like two big levels, like teams and projects and then they're connected to each other. Each team have several members and then each team has several projects that are more thinking one environment.

Was this answering your question?

[00:34:41] JM: To some extent. I'm looking a little bit deeper for the architecture, like choices of frontend framework, and database, and which cloud provider you're using and stuff like that. I'm hoping it will open up to a broader conversation.

[00:34:58] VP: Yeah. We're using AWS as a cloud provider. Our frontend is done in AngularJS, and the backend is mostly Node.js. We have some small part that is done in Python in backend where we're communicating to our Python SDK. And since the SDK is written in Python, but then the main backend is written in Node.js. Then they're kind of talking to each other. But the frontend is mainly done in Angular.

[00:35:31] JM: Gotcha. Which cloud provider are you using?

[00:35:34] VP: AWS.

[00:35:35] JM: And what have you chosen for the backend architecture? Are you using just VMs, or containers, or Kubernetes? Tell me a little bit more about that?

[00:35:47] VP: It's not VMs or Kubernetes. So it's sitting in – The data is sitting in our S3 bucket. And then we're using AWS cognito services. And, yeah, it's like a monolith that can be employed in different countries in AWS. And then if somebody has, let's say, GDPR compliances, then we can spin something in Europe. If somebody has other requirements, then we can spin something in different countries as well.

[00:36:20] JM: And the deployment model? So it's just deployed to VMs? Like Elastic Beanstalk or something?

[00:36:28] VP: Yeah. We're using Elastic Beanstalk, but we're also using heavily AWS SageMaker, because we're putting online trainings with just a few clicks. Imagine if you have annotated 1000 images. Then you can already learn from those 1000 images and then do predications on the next set of 1000 images. For those services, we're using AWS SageMaker.

[00:36:54] JM: Right. That brings up an interesting question, because it varies from label to label how many images you would need to learn what constitutes that label. If I label 1000 trees, maybe that's enough to generalize to how a tree gets labeled. But if I label 1000 people crossing a street, that may not be enough to generalize. How do you know how many images are necessary to get a consistent label for one of these types of things that need to be labeled?

[00:37:31] VP: Yeah. It really depends from the application. As you were saying, the difference between trees and pedestrians. If you're in autonomous riding space, then you probably need like millions and millions of images of pedestrians to be annotated. But it's really – We give some type of MAP score course or some machine learning accuracy scores to the client when we're putting a training. So the client is the one that is deciding if this is a good enough accuracy or not.

But even if the accuracy is not good enough, our predictions is a suggestion to the annotator. So they can annotate things faster. We're not really providing the full model at the end of the day. We are a platform that helps to get the most accurate data in a fast amount of time. So even though our models can be, let's say, 2% less accurate than 100%, then it will automate 98% of the job and then leave some 2% of the job to be correct.

[00:38:44] TP: Just to add on this point that, for example, in use cases of detecting trees, maybe 90% detection accuracy is enough to do your analysis. But for your autonomous vehicle company, you probably need 99.99 – I don't know what percentage in order to detects humans, because every time you miss a detection, it can be fatal. That's why autonomous vehicles are constantly getting new annotated data in more and more edge cases, more better conditions, geographies. So that's how autonomous driving basically drove this industry quite much. And then there were, of course, many more applications came up. But the detection accuracies that needs to be very, very high in autonomous vehicles, this is why it's driving a lot of data to be annotated constantly.

[00:39:42] JM: Do you expect just more and more growth in this annotation industry? How big is the annotation industry?

[00:39:52] TP: There are different estimates from like various marketing research companies. It's usually within around 2.5 billion already next year when it comes to tooling and data preparation and annotation. And it's expected to grow by around 30% year-by-year in the next few years. And the major driving force of course is more automation coming from autonomous vehicles, from retail automation, medical imaging. For, let's say, disease detection automatically, from robotics, etc. For all these applications, you need to constantly get more and more data to be annotated. And this is the basically main driving force, while why infrastructure of having label data, training, iterative training, modeling, becoming a really huge deal nowadays.

[00:40:49] JM: Do you think there'll be domain-specific annotation platforms? Like I remember when I interviewed Scale, they kind of had this where they had – In fact, they had a tremendous amount of success in just focusing on the self-driving market. But I can imagine satellite imagery being its own platform. Medical imagery being its own platform, and verticalized companies developing into that space. I don't know if the market is big enough or if it's one of those things where SuperAnnotate is the company that is able to capitalize on those domain-specific solutions. But what do you think about that?

[00:41:26] TP: Yeah, absolutely. Very good question. I see how the tools can be eventually more and more verticalized, because let's say medical imagine have their own specialties that you need to address. Maybe there's some security, more advanced security parts or ability to give the solution on-premise. For satellite imagery, you have to, let's say, get the image or part of the image directly from some large image, scaled, annotate and send back. There can be some little tricks or differences between different verticals. I clearly see how this can get more and more verticalized in a long run. And we already – Right now, we're putting all these features for different applications in one big system. But eventually we see how the solutions can be separated for different verticals.

[SPONSOR MESSAGE]

[00:42:31] JM: Today's show is sponsored by StrongDM. Managing your remote team as they work from home can be difficult. You might be managing a gazillion SSH keys and database passwords and Kubernetes certs. So meet StrongDM. Manage and audit access to servers, databases and Kubernetes clusters no matter where your employees are. With StrongDM, you can easily extend your identity provider to manage infrastructure access. Automate onboarding, off-boarding and moving people within roles. These are annoying problems. You can grant temporary access that automatically expires to your on-call teams. Admins get full auditability into anything anyone does. When they connect, what queries they run, what commands are typed? It's full visibility into everything. For SSH and RDP and Kubernetes, that means video replays. For databases, it's a single unified query log across all database management systems. StrongDM is used by companies like Hurst, Peloton, Betterment, Greenhouse and SoFi to manage access. It's more control and less hassle. StrongDM allows you to manage and audit remote access to infrastructure. Start your free 14-day trial today at strongdm.com/sedaily. That's strongdm.com/sedaily to start your free 14-day trial.

[INTERVIEW CONTINUED]

[00:43:59] JM: As far as the – Let's just take a simple aspect of your application, which is that if I create – If I'm ingesting a bunch of images of satellite imagery and all of those different images get bounding boxes displayed on them, or get the polygons in those images highlighted, is that kind of highlighting that kind of image segmentation? Is that stuff pretty much commoditized at this point? Are there very simple open source platforms for image segmentation, for example?

[00:44:39] VP: Not really for image segmentation. I mean, putting bounding boxes, there are platforms, open source platforms that you can do the annotation. But here, the important aspect is to gather that in a good user and project management system. So it's not really only the canvas that you're doing the annotation. Even if you have the canvas for image segmentation, the important problem would be if you have tens of thousands of images, how are you going to make sure that people in different parts of the world communicate to each other to get the high-quality data? In that sense, I think there are canvas for bounding boxes. But then, overall, it's not really a commoditized market. The entire AI infrastructure is far from being mature itself.

[00:45:35] TP: Yeah. Basically, if you need scalable project to build open source solutions that feed for usually one person to work in a specific task, it will not work well. You need to get to more such comprehensive platforms to deal with.

[00:45:55] JM: Now, the actual open source tooling for rendering polygons, or identifying, breaking up the different polygons in an image, that is available though, right? There are publicly available models for processing those images and finding the different bounding areas, right?

[00:46:13] TP: Mm-hmm.

[00:46:13] JM: Okay. Yeah. Just wondering how far along the open source ecosystem is.

[00:46:19] VP: Yeah. For one person, if you're doing, let's say, 1000 images that one person can sit for a month and then gather that data, I think that part can be organized pretty well with open source tools. But then the problem is coming when you're doing a real conservation project or you need clearly high accuracies. And the quality is number one priority.

[00:46:45] JM: Do you have the tools for a quality maintenance? So like when I talk to scale, one of the problems is like let's say I am a person who's managing the images and I'm going to sit down to work for an 8-hour a day. And at the beginning of that 8-hour a day, I'm doing a really good job. I'm doing very high accuracy annotation. But overtime, I make it tired. I may do a worse and worse job. And one way to hedge against this is you could send the images to three different people and do a best two out of three sort of thing. Another thing you could do is develop a rating system to somehow see which kinds of people don't get tired and which kinds of people do get tired throughout the day. Tell me about how you do quality maintenance.

[00:47:33] TP: In these situations, we have these different levels of quality assurance system. For example, when annotators do the first round, it goes to the QA. We also track how many images were sent back to the annotator or how many images were corrected. In this way, we kind of track whether this annotator's quality of work was diminishing or it was getting better. Maybe their speed was getting better or getting slower. This kind of system we have built basically helps to track all these characteristics within our platform. Also, the customers or QAs could give thumbs up or thumbs down. An image is counting those. Also gives some feeling about whether the quality is getting better or worse.

Vahan, do you have something to add?

[00:48:24] VP: Basically, that's the manual way of doing the quality. The second way, what we have in our platform, is automatically detecting suspicious annotations and then marking those annotations in red color. When you're sending those annotations back to the next level, then the quality assurance person immediately sees the – Where it can be wrong annotations. They're like spending less time. They're not really going through the entire image through every single annotated object, but rather they have been highlighted what is a potential threat. And then they just go and correct those.

[00:49:07] JM: So you have like different kinds of roles of people who can maintain the different layers of quality.

[00:49:14] VP: Correct. Yeah. Overall, there is like 7 different roles into the platform. But then when you're doing annotation, then there is like 3 or 4 of the annotation.

[00:49:26] JM: That's a great insight. And it makes me wonder if there're any other insights about workflow management that you've had. Like what are the ways in which managing annotation teams have surprised you? Or what are the ways in which managing annotation teams have become more sophisticated over the years?

[00:49:47] VP: I think it's mainly because the instructions are becoming more and more complicated. Then managing is also becoming a little bit more complicated because you have to make sure that they fully understood the instructions. One way that we're actually doing it differently, what our competitors are doing, is that we can pin one image that is ideally annotated. And then that will be distributed through all the annotators. As a result, when you're reading, let's say, 40 pages of instructions, it's a lot better than to look a few examples yourselves. And then if there is some common mistakes, then you can pin those images as well so they can see also common mistakes so they will learn visually a lot better than if you would just give them like 40 pages of instructions.

[00:50:43] JM: You've been spending time on video annotation. How does video annotation differ from image annotation?

[00:50:50] VP: One of the most important things in video annotation is like ability to track different objects in different shapes, right? The tracking is the number one component. You cannot really break it down to different images and then annotate it one by one. You can either jump 10 or 20 frames and then do some linear interpolation between those frames. Or the next stage is like use some tracking algorithms that we are building right now in order to accelerate that process as well. Let's say you're going to put one [inaudible 00:51:32] on a pedestrian and then you'll click play and it will just follow that pedestrian over the next 100 frames. So you don't really need to touch the annotation for the next 100 frames. This is becoming very crucial for those tasks where tracking is essential and you cannot really spend that much time to annotate every single frame.

[00:51:59] JM: Tell me more. Is there any tooling that you've needed to build around video annotation that is significantly different from image annotation?

[00:52:09] VP: Yeah. We had some internal tools that we were using our own tools. But then now we're releasing external tools as a SaaS model next month. So things are a lot different there. So what we're doing differently is allowing you to track multiple objects instead of one object, because most of the companies when they're trying to track, they only allow you to track one object.

[00:52:40] JM: I see. And what kinds of applications do you expect to be built around video annotation that could not be built around image annotation?

[00:52:50] VP: Yeah. A lot of security applications can be done with video annotation. In cashier-less check-outs or in retail surveillance when you're trying to detect various different actions. So you really need like a few frames. Let's say, 10, 15 frames in order to know if the human, let's say, is picking and putting something in the basket or not. You cannot really do those things within the annotation.

[00:53:23] TP: Maybe of course the most obvious one is the autonomous driving, where you have basically videos on the streets. Terabytes of videos that you need to know which one to annotate and how much to annotate. This is I think one of the big problems currently in the autonomous vehicle space.

Seems like eventually more and more, instead of taking images, people are just getting videos. Even in medical imaging, you record someone sometimes in video in real-time tracking. So the more advanced it gets, the system, it starts to already track images in real-time and then the video annotation is getting more and more important.

I've seen another example automating warehouses. There they fly drones to read what objects are there. What is not? And there they have gathered a lot of videos from drones. And this is another application they need annotating videos. Also, in satellite imagery actually. Yeah. When you fly drones or satellite images in real-time, you deal with videos mostly.

[00:54:35] JM: And the workflow where you could label some number of images, like what we talked about earlier with the trees. If you label a hundred images of a tree, then maybe you can have a SageMaker model that can just detect the images of a tree. My sense is that that kind of process in terms of video is more immature. Like if I'm trying to understand like gate recognition.

How a human is walking, for example. That's a little bit further behind than just recognizing static images. Or correct me if I'm wrong. How far are we with that kind of training for movement?

[00:55:19] VP: Yeah. Certainly, the training for the movement, in general, the video training is taking a lot longer time. But then there are quite good algorithms right now that can do – Kind of video tracking is – I think it's quite advanced at the moment. When it comes to the annotation tools, since if you want to do everything on the cloud, it's becoming very challenging to work with the video with the cloud applications especially when you're trying to put an AI in it. Then it's becoming even more challenging.

In my opinion, we are still one of two years behind to have very complete video annotation tools. Because that space is booming at the moment, I think. Most of the image annotation projects will be video-based eventually in the future.

[00:56:15] JM: Okay. Well, we're beginning to wrap up. Tell me a little bit more about what's in the near future. What are you hoping to build in the near future for SuperAnnotate?

[00:56:25] VP: Tigran, would we say it?

[00:56:28] TP: Yeah. Go ahead.

[00:56:29] VP: We have one great idea that we're just in a stealth mode for the moment. But maybe Tigran can –

[00:56:39] TP: Yeah. Go ahead, Vahan.

[00:56:41] JM: You guys are both afraid to say what the stealth mode awesome feature is.

[00:56:47] VP: Yeah, exactly.

[00:56:49] TP: You could cut around it.

[00:56:50] VP: Just giving you some lens. We're trying to build a community of researchers that will collaborate with the same dataset and then improving the same dataset and kind of we're building a community of researchers that will gather in one place and then work on a common project and keeping including the common project when it comes to the computer vision space.

[00:57:15] TP: Yeah. This is just the community part of the platform. Eventually we realized that it's not just about giving platform for these large groups to work on annotation project. But rather have these groups. Especially, universities is a great place, academics to collaborate within our system to keep their data, to improve them with time, to share with the research community. We were quite excited about this already a while because we came out of academia and we want to give back to academia. So this is something we're excited about just from this perspective.

Of course, going forward, we want to also to extend to all the other types of annotation tools starting like in immediate, which is of course video and then going to Lidar. Maybe eventually covering all types of annotation space, from speech, from text, etc. Having one space where all types of annotations are done and connected with the right people with skilled workforce to be – Annotation to be done in one space. That's kind of a vision. And then eventually having not just annotation space, but also addressing the full pipeline of annotation, training, iteration, etc. Addressing the full AI pipeline eventually.

[00:58:46] JM: Okay. That's very exciting. Guys, thank you so much for coming on the show. It's been a real pleasure gaining an understanding of the annotation space.

[00:58:53] VP: Yeah, pleasure talking to you, Jeff.

[END OF INTERVIEW]

[00:59:05] JM: If you are selling enterprise software, you want to be able to deliver that software to every kind of customer. Some enterprises are hosted on-prem. Some enterprises are on AWS. There might be a different cloud provider they use entirely, and you want to be able to deliver to all of these kinds of enterprises.

Gravity is a product for delivering software to any of these kinds of potential environments or data centers that your customers might want to run applications in. You can think of Gravity as something that you use to copy+paste entire production environments across clouds and data centers. It puts a bubble of consistency around your applications so that you can write it once and deploy it anywhere. Gravity is open source so you can look into the code and understand how it works.

Gravity is trusted by leading companies, including MuleSoft, Splunk and Anaconda. You can go to gravitational.com/sedaily to try Gravity Enterprise free for 60 days. That's gravitational.com/sedaily to find out how applications can run the way that your customers expect in their preferred data center. That's gravitational.com/sedaily.

Thanks to the team behind Gravity, the company Gravitational, for being a sponsor of Software Engineering Daily.

[END]