Image Recognition Computer Vision Platform that can Recognize, Identify, Caption and Classify the Details of a Digital Image in Seconds

Brad Folkens
Chief Executive Officer

CloudSight Inc.
www.cloudsight.ai

Contact:
Brad Folkens
310.582.5901
contact@cloudsight.ai

Interview conducted by:
Lynn Fosse, Senior Editor
CEOCFO Magazine

> "We won't stop until we're at a point where a computer can understand and interpret an image and really reason through that."- Brad Folkens

CEOCFO: *Mr. Folkens, what is the idea behind CloudSight?*
Mr. Folkens: There has been a lot of activity in recent years behind this idea of computer vision. We are focused on the idea of visual cognition and understanding. If you think about computer vision, it is kind of like how audio recognition was going on in the nineties and early 2000s. There is a lot of effort focused on recognizing individual bits of speech but not really understanding the audio as a whole. If you remember using those telephone systems where they would ask you to say something and never quite got it right, even until SIRI came along where everybody started finally recognizing that voice recognition was not about recognizing bits of speech but rather understanding and cognitive ability as a whole. It was more about the context than it was bits of speech. We ask ourselves the question of why everybody is focused on this idea of image recognition when they should really be focused on this idea of visual cognition or understanding so that is what we focus on in the company.

CEOCFO: *What have you developed to do this?*
Mr. Folkens: We developed an API (Application Program Interface). Basically an API allows other companies to use our technology to understand the process of digital media. Companies can send us images and we will send back the meaning of those images. If you send us an image of a certain picture, we can describe what is in the picture from the level of understanding instead of just recognizing little bits of it and instead of saying there is a 20% probability that there is a cat or a dog in the picture, we will tell you exactly what kind of breed of cat. Perhaps if you would like to know what the entire scene is about in the picture, we can give you a good understanding of the scene. Companies send us their images and we send them back the meaning and the understanding. To date, we have had over 500 million images that have been submitted through our service that we recognize.

CEOCFO: *How are you able to get down to the breed of cat?*
Mr. Folkens: We started this company in 2012 and we spent the last several years building out the biggest database that has ever been curated for the purposes of visual cognition and understanding. Having recognized over 500 million images now, we have built up a tremendous database, larger than any other database specifically curated from this purpose. There is a database Image Net which is used for training computer vision in the academic space and that is only 14 million images. With us having recognized over 500 million images, it gives you an idea of the difference in scale. This image database that we have built up trains deep learning neuro networks which is a mathematical approximation of the human brain that has become popular in the past few years but it trains a mathematical or software based brain that understands now what it is looking at based on all of this training data that we set. This is kind of synonymous to when kids are young and you start showing them the world around them and they start understanding eventually; they point to a

cat and say that is a cat and when they get older they can tell you that is a Shebalino. They start learning more and more about the details. Several years ago we started going with database and we could point to a dog and say that is a dog and when we got more and more data that we set in the system, the system started learning breeds of dogs and even if it does not know exactly the breed, it describes it so you can find it through visual search on the internet. We came out with this app called CamFind several years ago and that provided us a great window to the world to start learning about the world through visual search. It is kind of like a Google for the physical world. We also made another app called TapTapSee which is for the blind and visually impaired. What that app does is if you are blind or visually impaired you can take a picture of your surroundings and it describes them to you. This has helped the blind community and in fact in 2014 we won an award from the American Foundation for the Blind. The award is called the Access Award which is awarded to companies that make tremendous breakthroughs for the blind and visually impaired community. It was through these two apps and then later on adding our API customers, we have been able to grow this database and curate this database for the purposes of training people on neuro networks. That is what has allowed us to have the kind of accuracy and service that we have today.

CEOCFO: *Who is making use of your services?*
Mr. Folkens: Many companies use our services and we have them listed on our website, www.cloudsight.ai. Today companies like Kohls, Mars, Proctor & Gamble, Grey Group, Belk and Walmart all use us to bring a lot of different functionalities to the user; one of them is visual search and the other functionality that they use is recognizing their digital media assets. Any time a company wants to understand their visual media they come to us for that. Those are just some of the customers that we have that use the service.

CEOCFO: *What is the competitive landscape?*
Mr. Folkens: There are many companies in the space of computer vision these days but computer vision is much like the medical field because there are a lot of different specialties that companies decide to go into. We are focused on this idea of visual cognition and understanding. There are a lot of other companies that are focused on classification and there are other companies that are focused on self-driving cars, the automotive industry. There many different niches out there and we are a little different from some of the other ones in that we focus on this more general idea of visual cognition and understanding.

CEOCFO: *What is your business model?*
Mr. Folkens: The business model is known as SaaS or Platform as a Service. What we do is provide a web API and customers send us images and we charge them per-image-request basis. With our enterprise customers we have some custom contracts. Most of the customers pay us per image request or for batches of images. It makes it simple to calculate in terms of business opportunity.

CEOCFO: *What has changed as people have been using your product and your offering has grown?*
Mr. Folkens: Many clients have come to us on a variety of different challenges that they want to solve in their businesses, so we have had some exciting ways to work with customers on proprietary solutions that have grown. The most exciting thing that tends to happen is a thing called transfer-learning, so if we show the system a picture of a broken coffee mug and then I show pictures of a Starbuck's coffee mug, and if I take a Starbuck's coffee mug and I break it, then the system will actually figure out that it is a broken Starbuck's coffee mug. This interesting idea of transfer-learning is you start to train the computer with more and more images, it starts to learn these different concepts and it is able to start piecing the concepts together like a font which is an interesting thing to see. We have experienced a lot of synergy as we started attracting more and more clients.

CEOCFO: *How can CloudSight make the leap in understanding?*
Mr. Folkens: There is a critical mass that it takes for anybody to achieve understanding. You walk around and show a child a bunch of things and eventually they start piecing together the world, achieving understanding. Right now we are at a primitive stage with computers. What we are essentially showing computers are these still images which is absent of any other senses. It is absent of three dimensions, it is absent of time. We are showing computer still images and there is a lot experimentation being done with video too but that is only a piece of the puzzle. If you think about your experience as a human you have so many different senses that you use, including motions to put together an understanding of the world. We are still in a primitive phase with computers but I think a lot of that is going to advance rapidly in the next few years.

CEOCFO: *Would you tell us about the new patents?*
Mr. Folkens: The new patents are very exciting. We started patenting a lot of our technique back in 2012-2013. The patent process is very long and arduous. We are finally starting to have some of these relieved. We were one of the first

to pioneer this idea of hybrid intelligence in the commercial space. Essentially we have humans that are correcting the computer when the computer did it wrong. This is very powerful. When you walk around with a child and the child calls a cat a dog you will say oh no that is a cat. You correct them and they start to learn about the world around them. This is an idea in the computer science world called supervised learning. Supervised learning is where we have seen the most effectiveness in training these deep learning neural networks how to understand the world around them. Instead of pioneering in the commercial space this idea of hybrid intelligence years ago, we were able to get a lot of that IP secured. We have quite a few more patents that are in the pipeline. We have over thirty patents pending worldwide and some of them are finally starting to issue.

CEOCFO: *What surprised you as the company has grown and evolved?*
Mr. Folkens: What surprised us the most is initially when we started out on this mission, computer vision was in such a primitive state. We thought computers would be able to recognize a lot more than they did back in 2012 when we first started this. Our idea at the time was to go after visual search so this idea of a Google that can see. What happened is we were serving a lot of computers and technologies at the time in 2012 and we found that none of them were any good other than creating anything more than a novelty. There was this product at the time called Google Goggles which was a visual search for Google, from Google. You would take a picture of something and you were supposed to find it on the internet. The problem was it never worked and only worked about 5% of the time if you were lucky. We looked at that and thought it was horrible and that it was 2012 and computers should be able to see. When we were working on this idea for CamFind, which was our visual search tool that we were putting together in 2012, there were not any very good image recognition technology providers so that is what we decided to start building, our own image recognition platform. That eventually grew into what it is today. We never thought that it would be our core business eventually. The thing with visual search is that there is a big behavioral change that has to happen with users and we are finally starting to see that more now with a lot of companies getting interested in visual search but the idea of the platform itself grew into a much larger opportunity for us as a company and that is our core business now. That was something that we did not expect when we first started off.

CEOCFO: *How are you reaching out?*
Mr. Folkens: We have many different marketing things we do internally but I think our greatest success has been through word of mouth. We have had a tremendous amount of growth that we have seen just in the past couple of years. Initially people found us through the app Can Find. As the business started growing and more and more people started recognizing us for the platform itself CloudSight. Now that CloudSight has grown it has generated a name for itself and we have gotten a lot of word-of-mouth customers coming to us.

CEOCFO: *What is next for CloudSight?*
Mr. Folkens: We are excited about this idea of visual cognition and understanding. We won't stop until AI can see and that's our big mission, to make artificial intelligence see. What does that mean? Well right now computers can recognize a tremendous amount of detail. What we do is we call it fine grain object detection and we have two different channels in our API. We won't stop until we're at a point where a computer can understand and interpret an image and really reason through that. Right now if you take a look at a picture and there is some sort of comedic thing going on in the picture, the picture is funny - that is an interpretation. It seems that we do not think of that too much but it just comes naturally to us. A computer cannot get those subtle meanings yet so that is what we are excited about coming next.