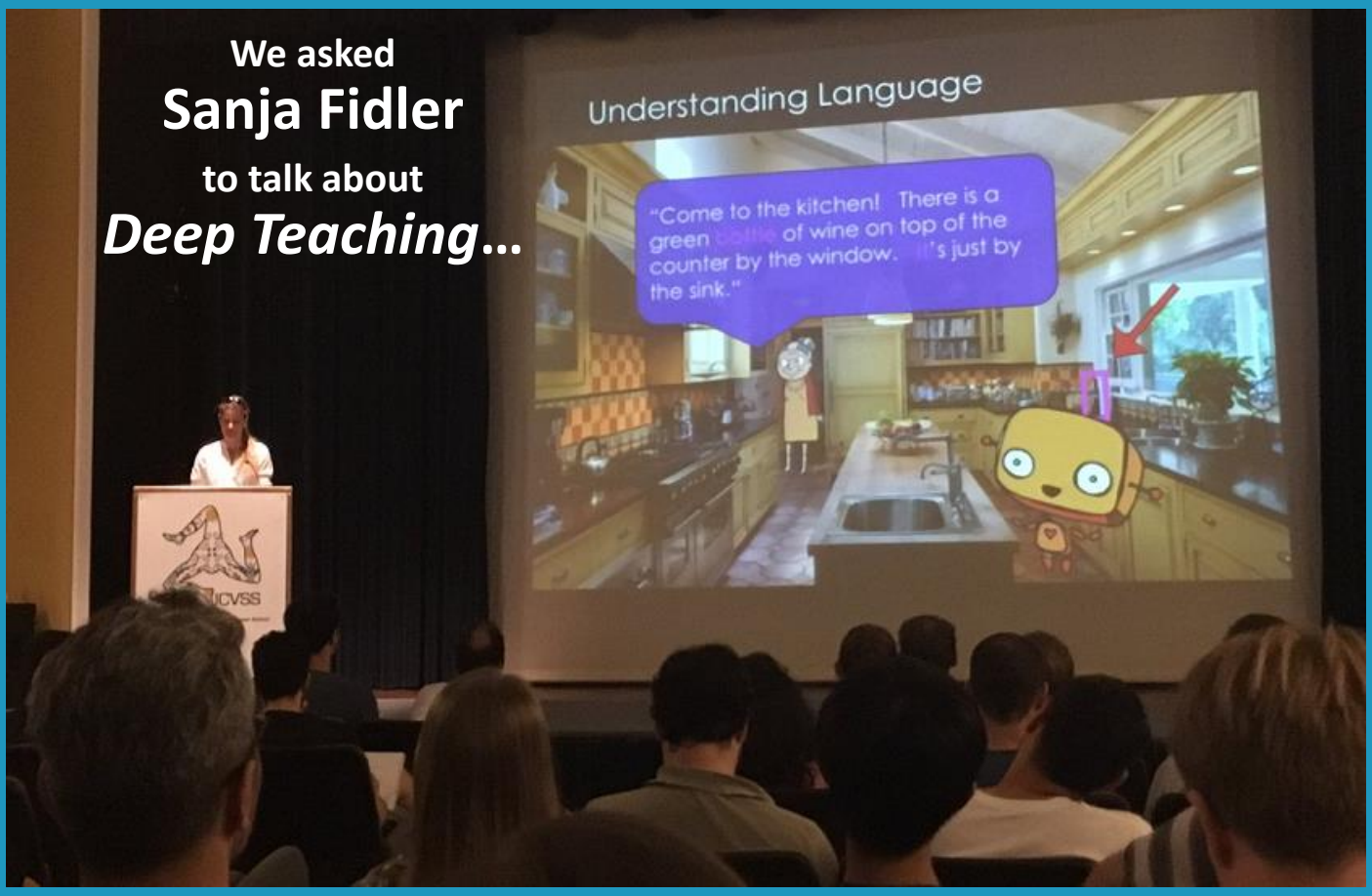
CVPR DAILY
July 21-26 HONOLULU 2017
Computer Vision & Pattern Recognition

Tuesday 25



Exclusive Interviews with:
Vittorio Ferrari
Harry Shum

Women in Computer Vision:
Nour Karessli

Presenting work by:
Phillip Isola
Sergi Caelles
Laura Leal-Taixé
Silvia Zuffi
Vamsi Ithapu
Linjie Li

Roxanes's Picks for today

Read an important community message in the last page!

In cooperation with
Computer Vision News
The Magazine of The Algorithm community

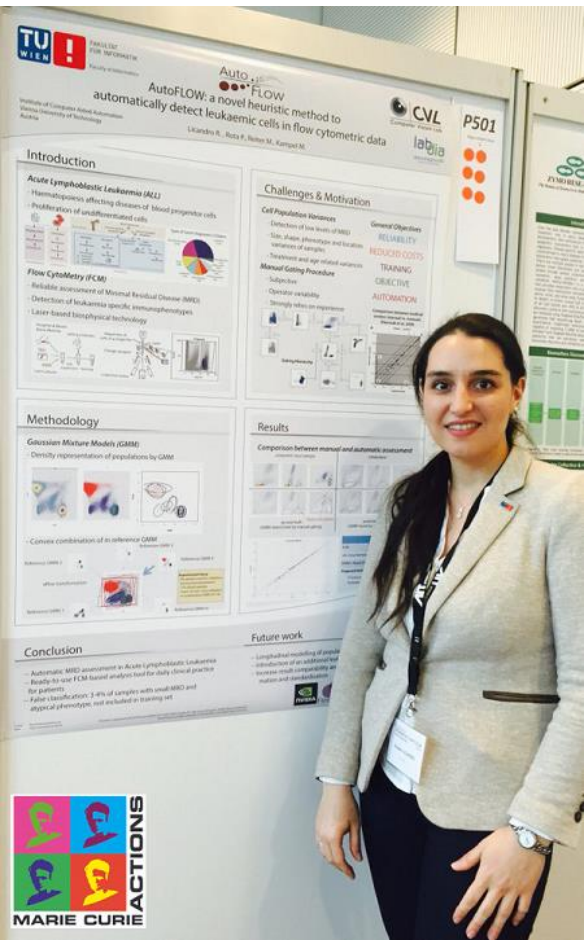
A publication by
RSIP VISION

Roxane's Picks

CVPR DAILY
July 21-26 2017 HONOLULU Tuesday

For today, Tuesday 25

Roxane Licandro



Roxane Licandro is currently a Marie Skłodowska-Curie researcher and affiliated with the Computer Vision Lab at TU Wien and with the Computational Imaging Research Lab at the Medical University of Vienna in Austria. Her research focus lies on medical computer vision and machine learning for cancer research and neuroscience.

"As member of the Women in Computer Vision Workshop organizing committee I cordially invite everybody to attend the CVPR workshop on 26th July afternoon with renown speakers and a fantastic panel.

TIP: If you need a break from the beach, you can visit the Bishop Museum. It is the largest museum in Hawaii and has the world's broadest collection of Polynesian artefacts.

Roxane's picks of the day:

- ## Morning

O4-1A.9 09:03 Page 33 of the Pocket Guide:
Geometric Deep Learning on Graphs and Manifolds Using Mixture Model CNNs
P4-1.28 10:00 Page 34 of the Pocket Guide:
Convex Global 3D Registration With Lagrangian Duality
P4-1.54 10:00 Page 35 of the Pocket Guide:
Optical Flow Requires Multiple Strategies (but Only One Network)
P4-1.62 10:00 Page 35 of the Pocket Guide:
Deep Photo Style Transfer

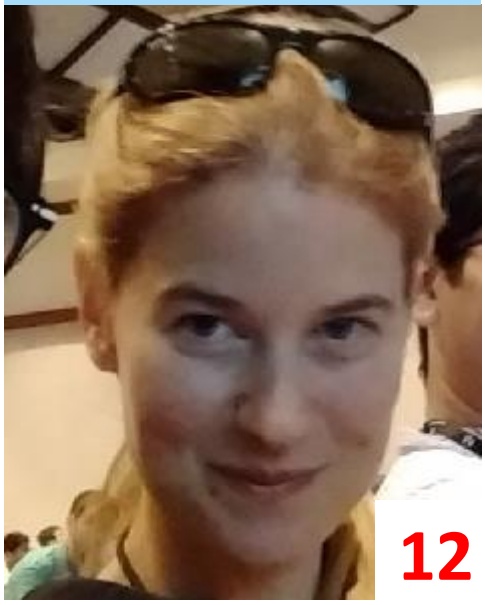
- ## Afternoon

P4-2.46 14:30 Page 40 of the Pocket Guide:
Multi-Way Multi-Level Kernel Modeling for Neuroimaging Classification
P4-2.47 14:30 Page 40 of the Pocket Guide:
WSISA: Making Survival Prediction From Whole Slide Histopathological Images

Roxane's Picks

02

Sanja Fidler

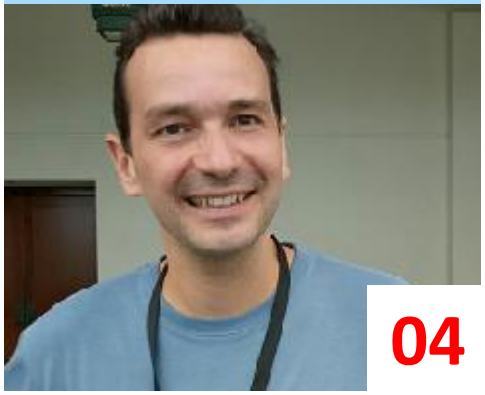


12

Harry Shum

22

Vittorio Ferrari

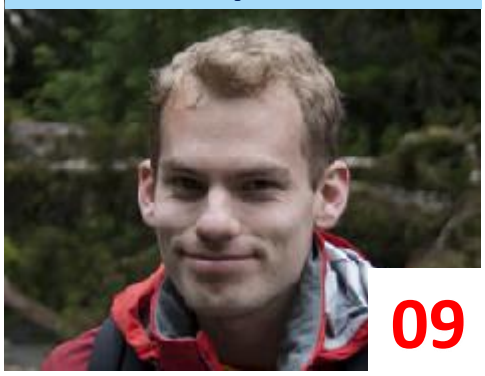


04

Linjie Li

27

Phillip Isola



09

Women in Comp. Vision
Nour Karessli

18

Sergi - Laura

28

Aloha, CVPR!

This last CVPR Daily for this year is so rich that there is no space for my editorial. So I turned it into a lovely goodbye picture in the last page. Please do not miss it for any reason. Yaay!

Ralph Anzarouth
Editor , **Computer Vision News**
Marketing Manager, **RSIP Vision**

CVPR Daily
Publisher: **RSIP Vision**

Vittorio Ferrari

CVPR DAILY
July 21-26 HONOLULU 2017 Tuesday

Vittorio Ferrari is a professor at the **University of Edinburgh** and a research scientist at **Google Zurich**, in both of which he runs a research group.

"I salute the students that are braving the new world in these days."

Vittorio, do we have in common an Italian background?

Almost right - I am Swiss, from the Italian part of Switzerland.

So we are neighbours! I am from Milan, and there we say that we are nearly Swiss.

Actually, in Lugano, I have been told by the Swiss Germans that I am nearly Italian. [*We both have to laugh*]

So you grew up in the Italian part of Switzerland, then decided to become a scientist and your career brought you to different places.

Yes. Interestingly enough, when I was a kid, there was no university in Lugano, my hometown. So even just to study, before being a scientist, you had to move. That started my journey about half a life ago.

Where has this journey led you, what are you working on now?

I work on various problems in computer vision, but my general life mission since 5-6 years is to try to learn computer vision models that are able to localise objects in images with high quality and least human intervention possible at the same time. This sometimes is described as weakly supervised learning. Recently I have been working a lot on human-in-the-loop learning, where there is a bit of human supervision and intervention during training. For example, as a

typical outcome you would get a model that is capable of labelling every pixel in the image that is containing an object, but at training time you perhaps only need image-level labels. And yet, you can get a localisation model almost out of thin air.

Why are you so passionate about these kind of problems?

There are two really good reasons. One is an intrinsic, scientific reason and the

"Oh, I am more passionate now!"

other is the impact it can have. The impact it can have is that we will one day be able to train from tens of millions - no, billions - of training examples that cover tens of thousands of object detectors. This is in fact necessary to reach human-level ability, you need lots of samples and lots of classes. And one day we will be able to do that at a cost that is within the ability… maybe not of everyone, but at least within the ability of a millionaire [he laughs], in the million dollar range. Now this would be absolutely impossible, if you want a complete annotation of every pixel in an image. You cannot do a million objects, basically. The problem will only be solvable once we have all the annotated data, and we will not annotate it by hand the way we are doing it in the fully supervised world. That's why reducing the annotation time is not just a sport, it's an enabler of solving computer vision. Now if you go to the scientific reason, which I am even more passionate about, it is a very interesting information-theory type of trap. When you have a weakly supervised learning problem, let's say, this image where we stand now: this is a couch, this is Ralph, this is Vitto, this is a plant in Hawaii. You have these labels, and there is actually combinatorially many assignments of the pixels in the image to these labels. And all of them are consistent with the labelling of the image, but some of them make more sense in terms of regularity. It's very interesting that theoretically, there are many solutions that are valid, so that strictly and information-theoretically speaking, it is impossible to reconstruct pixel-level labelling of an image from image-level labels. And yet, there exist some assignments that are more likely to make sense in the visual world. For instance, all the pixels on your face probably all take the same label, they're all face. For me it is very exciting that although we know that there is no perfect closed-form solution that will work, there is certain families that make more sense in the visual world and that lead to good results at test time. So somehow I like the fact that you start by saying that the problem is impossible, and yet you try to solve it.

"Like a kid in a candy store…"

You sound still as passionate as when you started to study…

Oh, I am more passionate now! When I started my PhD, I felt like a kid in a candy store. You jump at everything that looks cool, and you grab something, lick it a bit, then you take something else… so there is no continuity of mission. Now I am equally motivated, but because I focused the energy of my team over multiple years on a family of problems, I also see a lot more progress. And I appreciate the fine details of these families of problems. So in fact I actually feel more passionate now compared to when I started.

Do you have tips on how to keep the passion over a long period of time?

I started my PhD 17 years ago, and one way to keep the passion over 17 years is to change the family of problems every once in a while, to freshen up. Stay in machine learning and computer vision problems, but change at the beginning every 3 years and later every 6 years, and eat from the diversity of the computer vision fruit - it's a really big fruit.

Is that something that others do as well?

If I may dare to mention the really great, like Zisserman and Malik, who manage to keep the passion for a long time, they all have one distinguishing mark: they worked on a lot of problems, and typically they make one landmark contribution in each era.

Can it happen that somebody enters into a field and realises that they shouldn't have?

Oh, absolutely. It happens when you're younger, and it happens when you're older. You have to be able to feel whether putting energy into an area is going to lead to things you want. Which are always the same two: happiness for yourself, so you have fun, and the second is your publishing and that people are interested in what you write. These two criteria are often in contradiction. So you need to feel it as fast as you can. I would say if you're not happy after six months after entering a field, you should change.

How do you rebound in the case it doesn't work?

When you are the first implementer - a PhD or a postdoc, before you are a professor, rebound is somewhat easier. You have to have the self-discipline of going to your advisor and saying, look, this thing doesn't work. And then normally it's about having a picture of the new thing. So just saying "I hate what I'm doing, and I quit because it doesn't work", then the alternative is the void, and the void scares everybody. So you need to set aside some time. Normally when I felt I wasn't doing so well in an area, especially when I was younger, I would just say: this week, I only read. I don't program anything. And just read as diverse stuff as possible, and then decide what to work on. When you are a older and you are a group leader, then it's harder to rebound, because you are very much in love with your own vision [he laughs] and you don't quite see why it doesn't work.

And you have a responsibility for the people who are with you.

Yes, telling your students: you know what, because of various reasons, perhaps technical impossibility reasons or because somebody else already implemented your idea, you have to change direction. This is tricky, but it's important to do it. As you said, you have a responsibility for the student. And sometimes the best interest of the student is to radically change topic. As a group leader you must make these choices.

"humble down sometimes"

Might it also be the case where the student opens the eyes of the professor and says that something is not going to work?

Absolutely, sometimes it's exactly the PhD student or postdoc that has to go to his boss and say: you know, Vitto, this thing you like so much - it ain't gonna fly [*he laughs*]. The professor

"The noise is noise with respect to the center point, but it's signal with respect to scale"

has to be able to humble down sometimes. Sometimes the first implementor, the person that is doing the actual work, is actually not seeing that it could be working, and the group leader sees that it could be working. And then the group leader should stay on track. But sometimes the group leader is just illusioned, is lost in his or her own ego, is in love with the own idea, and then you have to say: you know what, you're right. And this dynamic between bottom-up and top-down leadership, it is a big feature of a healthy group.

Do you think there are any changes between your generation of students and the generation of students that you see now?

Ok, ehm… Temporally skip. I give you just a quick comment, we can continue after. Just off the record comment… this is such a awesome series of questions! [*we both laugh*] It's just so fun. Ok.

[*Laughing*] **Why don't you want to put this into the interview?**

[*Still laughing*] Ok, put it in. This is so cool. So before we re-start… what was the question again?

[*I repeat the question*]

Ok, this is an awesome question! See, it's very interesting. Let me first say how the environment changed. When I was a student, everything was much slower. You had an idea, and you thought: it's awesome! Then you had approximately a year, or a year and a half, time from an idea to a publication because the density of people working in your area was low and, you know, in the end you wouldn't be scooped. These days, if you work on something hot, especially on neural networks understanding something - the very middle of the field - between an idea and somebody scooping you, you have maybe six months. So it's becoming more stressing. But at the same time, because it's so much denser, if you do something good you get a lot more citations. And citations are a big currency, a big mark of success, that you trade for positions and professorships. So in a way it's harder and easier, at the same time. But it's certainly more stressful, and I salute

the students that are braving the new world in these days.

Now, in terms of skills. I believe that this generation is able to get somewhere faster. Nowadays, they have more tools, and they can recombine software pieces. So in a way it's exciting, the pace at which they can go. Perhaps if I can dare to make a recommendation, something that I felt that back in the days were perhaps a little bit better. The students today have a tendency to be very rushed to say: I am working on whatever is hot now, and what happened three years ago is forgotten. And this is very short-term sometimes. So perhaps back in the days, the students were trying to think a little bit more like: how can I change things globally? And they looked a little bit more beyond their field. So perhaps this has changed, but this is also a reaction to the environment. Today things just have to go quicker.

"How can they be so silly, or what a genius…"

What is the biggest surprise that you ever experienced from a student?

Oh, voilà! I would need a lot of time to answer, because there were so many times my students surprised me. So many times! Sometimes positive, and sometimes negative. But they often surprise you. And it's very important - back to information theory - anytime a student surprises you, positively or negatively, and you think - how can they be so silly, or what a genius - both times, take a step back as an advisor and update your own neural network in your head [*laughs*]. Update the student model, because that's where you learn, the surprise. So, I will just answer with something that comes to my mind which is fun, it's part of our papers we have at CVPR.

"That's awesome! How did you think of this?"

It's a technical contribution, but I thought it was really fun. My student was working on this project, where we try to learn object class detectors by annotating objects using the center point, instead of drawing a box around it. And my student was saying: you know, we should ask two people to click in the middle. And I said, forget it! It's useless! It's just a little bit of noise cancellation. The student said: well, you know what, Vitto. If they are both asked to click in the middle, they are going to make an error. And I said, so what? So the student said: but the errors they make is related to how big the object is. Because if the object is big, the two annotators are going to click further apart from each other, and on the smaller object they will click closer. And therefore we can estimate how big the object is based on the errors the annotators make. And I was like: that's awesome! How did you think of this? We are exploiting errors to get information about the object scale out. And in weakly supervised learning, object scale is one of the big holy grails. If you have it, it makes it a lot easier to learn. And so, you know, we dumped it into this paper and it became one of the coolest bits of the paper. I thought - how did you think of exploiting the errors in humans? When I think about it, I want to cancel out the errors, not turn them into information. I was really impressed by the student when he said that. It was very clever: the noise is noise with respect to the center point, but it's signal with respect to scale.

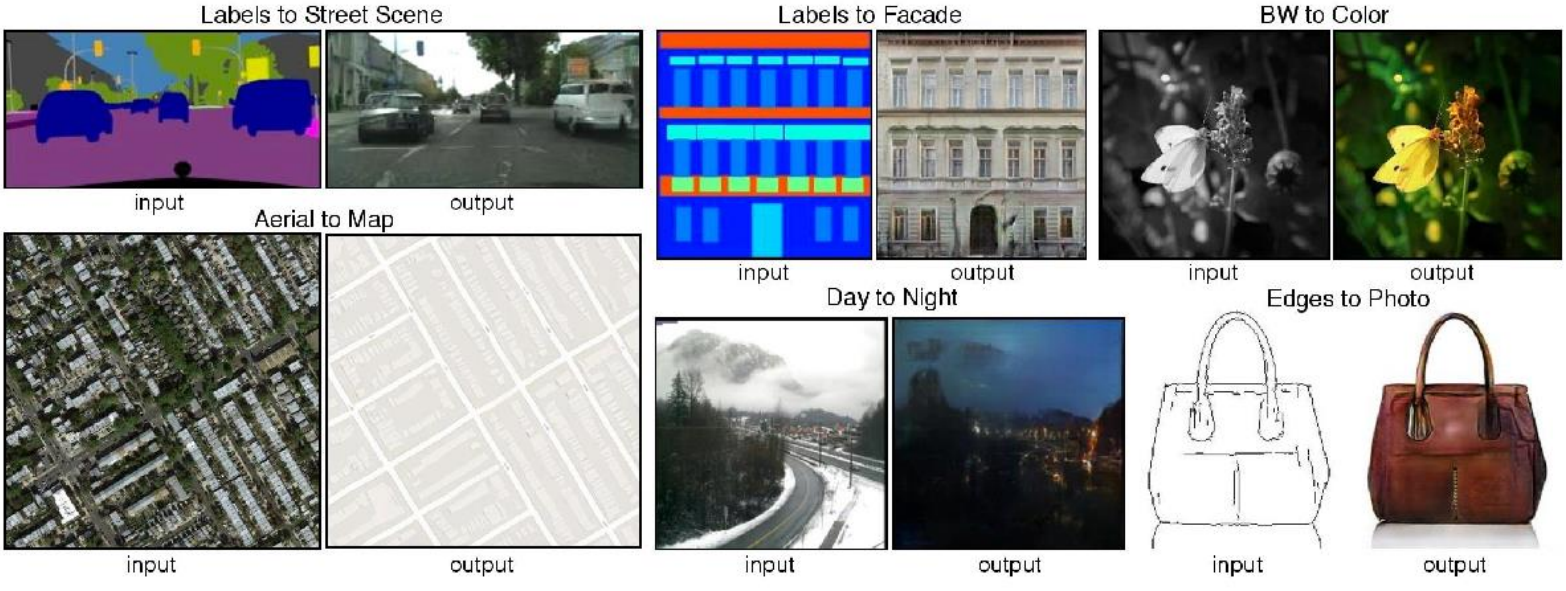
Image-to-Image Translation with Conditional Adversarial Networks

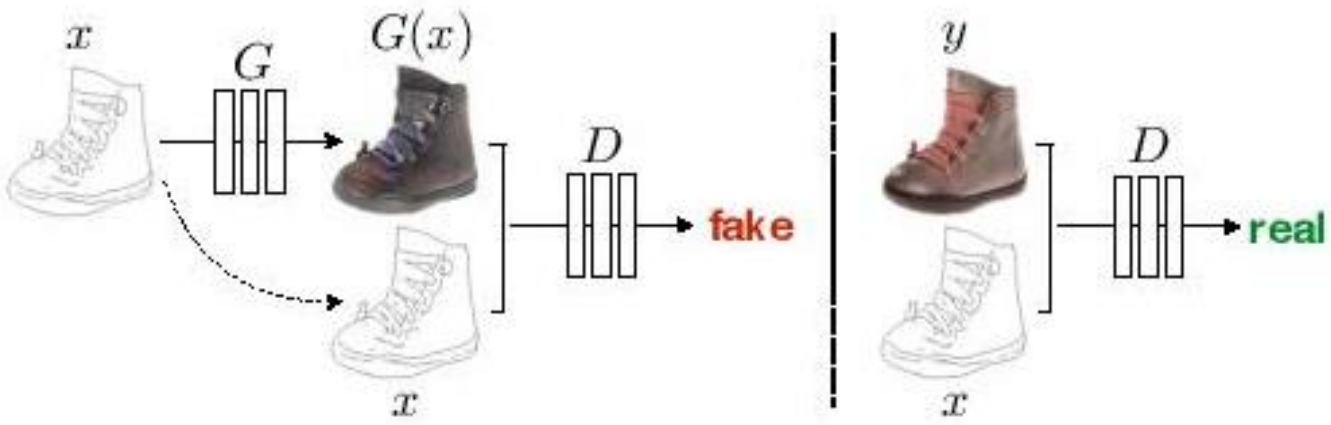
Phillip Isola is a postdoc with **Alyosha Efros** at **UC Berkeley**.

"There's a lot more problems that are conditional than unconditional, especially practical problems in computer vision and graphics"

Phillip Isola is presenting his paper "**Image-to-Image Translation with Conditional Adversarial Networks**", which is joint work together with Jun-Yan Zhu, Tinghui Zhou, and Alexei A. Efros. Their idea is to use **generative adversarial networks (GANs)** to solve image-to-image mapping problems, and in their paper they demonstrate that these are a general-purpose tool that can be applied to a lot of problems.

GANs, which were introduced by Ian Goodfellow et al. in 2014, and are a popular idea at the moment, and a large part of our community has gotten quite excited about them - "*rightfully so*", Phillip says. He told us that previously a lot of people have done work on unconditional GANs, which were used to generate random images. But Phillip and his co-authors thought that it might be more compelling to look at the conditional case, where you use a GAN for regression problems to learn a mapping from inputs X to outputs Y.





"There's a lot more problems that are conditional than unconditional, especially practical problems in computer vision and graphics", Phillip told us. For example semantic segmentation or edge detection are both conditional image-to-image mapping problems, or things like image colourisation (taking a black-and-white photo and producing a coloured version of it). In all of these problems you want to learn a mapping from pixel-to-pixels, i.e., images-to-images.

Phillip explained to us that what happened in the last couple of years is that CNNs have turned out to be a very generic way of processing images and are used for a lot of problems. But usually a CNN is only modelling structure in the input space. CNNs with the standard regression loss are treating every output pixel (of the semantic segmentation map or edge map) as conditionally independent given the input, so they don't model semantic structure in the output space. As a reaction, the community has already done a lot of structured regression problems modelling structure in the output space, for example using conditional random fields. But what Phillip's current work is doing is using adversarial discriminators as a way of learning a structured loss function to model structure in the output space. You thus have a neural network that models structure in the input space, and a neural network that models structure in the output space, to do generic things that can process images. "A year ago this was all very new and unexpected. The field has developed these ideas all together and we are one of them." In their paper, Phillip and his co-authors show that this kind of approach is suitable for many image-to-image mappings, and they demonstrate that this works well on a lot of problems without any change in the architecture or method.

"We then realised that we could remove almost all the bells and whistles"

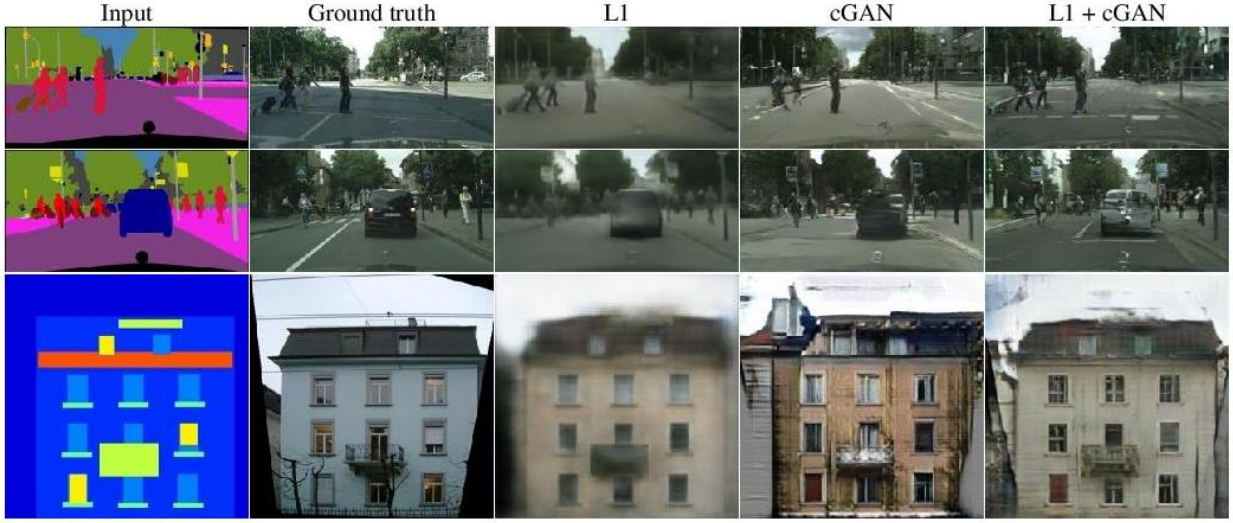
Phillip also told us about some insights they got from working on this problem: "The process was that we added a bunch of bells and whistles and got something working, and then realised we could remove almost all the bells and whistles".

The final model therefore only has a couple of tricks which turned out to be necessary. One thing they found is that while GANs are hard to optimise and can be unstable in the unconditional case, the conditional case is a lot more constrained: the conditional color distribution given a black-and-white image has much lower entropy than just the distribution over all possible random images. Because you have paired inputs-outputs, you are now also in a supervised learning setting: you can mix your GAN objective with the more traditional supervised objective. That's what they did in this work - they added L1 regression as an extra term in the objective to stabilise things. This leads to faster convergence and learning is more stable. "The nice thing is that if you average in this L1 regression with a small weight, then it doesn't really change the final results", Phillip told us, "and you still get nice clean GAN-quality results".

For future work, Phillip thinks there is still a lot of exciting things to do in the conditional GAN setting for image-to-image problems, and they already have a follow-up paper, called CycleGAN. Here they start with the observation that in the conditional GAN setting they needed paired supervised data. Given coloured images for example, they can train a mapping from black-and-white to coloured images in a supervised fashion, because there a million images to use for this. But if you want to learn a mapping between two domains, like paintings and photos, then you don't know the pairing. You might for example want to learn the mapping from a photo to a Monet-style image - but since these don't exist, this can't be trained in a supervised fashion. So without the paired data, you can't apply things quite the same way. "But it turns out that some small changes allow you to also learn the mapping in the case where you don't have paired data, but you just have two stylistically different domains", Phillip told us.

If you want to learn more about Phillip's work, make sure to visit his poster (number 65) "Image-To-Image Translation With Conditional Adversarial Networks" today at 10:00. TIP: ask him also about a fun tool made by Christopher Hesse with their code, for translating sketches of cats into photos of cats.

Current paper *CycleGAN*



CVPR DAILY July 21-26 HONOLULU 2017 Tuesday

"The teacher should adapt"

Sanja Fidler is assistant professor at the **University of Toronto** and a cofounder of the recently opened **Vector Institute**.

Sanja, what is the Vector Institute?

It's a research institute that we opened in March, and it's focused on fundamental research in the area of machine learning, specifically deep learning. It's government funded, and there are a lot of companies that contribute as well. We do pure research. Everyone can do whatever they want in different areas like computer vision, machine learning, NLP, and so on. We have many different topics. We hire around 20 faculties like research scientists that can create their own groups of students that they can hire. The idea is to foster what Toronto already has and bring it to the next level.. It is known for deep learning.

"Keep the talent in Canada"

What is the purpose of the group?

Originally, the idea was to establish research and keep the talent in Canada. Canada has less industry and less academic possibilities than the US, which is larger.

There is less focus on Canada.

Right - The idea was to keep all of these amazing, talented students in Canada.

That raises two questions. First, how did they get your talent, since you are originally not from Canada? The second question is how did it fall on your shoulders to found this institute?

Originally, I am from Slovenia. The first time I came to Canada was for a Postdoc position in 2011. My PhD work was very related to deep learning and theoretical representations of objects. I thought that would be a really nice place for me to do research. I didn't know much about Toronto, but I thought it would be good. I arrived in January, and it was so cold!…. But you dress appropriately, and you get used to it.

What was your drive to go on an adventure on the other side of the world?

Between my family and friends, no one ever left. Then on the last year of my PhD, I was invited by Professor Trevor Darrell to visit his lab. I went there for 7 months, and that was just amazing.

What convinced you to stay?

It was really the group. I really connected

"In Toronto, everyone is coming. There is always someone you can talk to who is interested in your research. It's awesome!"

with the group. It gave me the opportunity to talk to other researchers. The scale was much different, and people were much more engaged. These universities are structured in a way that allows people to be great. There are a lot of faculty and visitors. You are exposed to cutting edge research all of the time.

A research community like that can attract people to stay.

At the University of Ljubljana, the group was maybe a couple of students. We never really got visitors. You're kind of there on your own. Maybe you go to conferences, but that's it. In Toronto, everyone is coming. There is always someone you can talk to who is interested in your research. It's awesome! I said that's it. I'm going to finish my PhD, and then I am going abroad.

What was the most exciting part about moving?

I really love research. The opportunities there are just incredible. It's my passion. I can never switch off my brain. Even when on vacation, I am always thinking about my work.

Did you sacrifice anything by moving?

I really miss my family, but I find ways to see them often. I go home basically after every deadline. Before a deadline, I work really hard, and then the next week, I visit home and relax. I get to see my sister and her kids, my parents, and my friends.

This brings me to my next observation, if this impressive research community could attract you from across the world then maybe the Vector Institute can also bring in talent from all over.

Yes - That's we hope!

At what moment in your career, did you start to feel less like a student and more like a teacher?

Ooh - That's a tough one. I still feel like a student, just a different kind of student. When you are a student, you are learning the field. Now as a professor, I am learning how to teach. I always feel like I am learning.

"…deep teaching…"

What is more difficult, deep learning or "deep teaching"?

[*laughs*] Probably deep teaching… Deep learning is really interesting so it's easy to pick up.

What is more satisfying, a successful paper of yours or from a student?

[*replies with certainty*] A student's - The best thing is to see the paper being accepted and seeing the student become super excited. That's the

Apparently, being cool is a family trait: Sanja with niece Ajda and nephew Marsel

moment that makes it all worth it.

Will you have the same satisfaction in 20 or 30 years after teaching hundreds of students? How can you keep the spark of excitement in years to come?

I don't know about the future, but so far I feel the same excitement that I felt in the beginning of my PhD. When's there a new idea, I tremble with excitement. I enjoy collaborating with and teaching students. It doesn't seem like it's going to go away.

You seem very upbeat. What advice can you give to people whose papers were rejected or to those who are still waiting to have their papers accepted?

I feel that good research is always going to find a way to get published to become visible to people. If you really believe that you are doing something great, who cares about what a bunch of reviewers say? Maybe there is noise in the process. Maybe your paper actually isn't ready. Sometimes you need to agree with the reviews and make your paper better for next time based on their feedback. Next time it is going to be ever better! The key is to stay upbeat. You should not taking it personally. You need to believe in yourself. Don't get depressed from reading the reviews. Sometimes you get reviews which can get pretty nasty. This happens, but you need to remember why you still believe in your work. You can also learn something from the process.

Some people get stressed.

I guess for the first paper. For me, it's not stress, but it's more about feeling super curiosity about what will happen and feeling excited. You should not be stressed.

You have had a high percentage of papers accepted by the conferences. How does it feel to see you work succeed?

I always wait to submit papers until they are ready. Then you have a higher chance of it being accepted. Although then it can put pressure on yourself for the future. You can't always compete with your past achievements. If you did really well this year, you might want to do even better the next year. It can cause stress.

"I just really love what I do. My passion comes from a place of curiosity"

Are you more competitive with yourself or with others?

I am definitely more competitive with myself. I wouldn't call myself competitive. I just really love what I do. My passion comes from a place of curiosity.

After the CVPR oral (given by Lluis Castrejon on the left), which got the best paper honorable mention. On the right are Kaustav Kundu and Raquel Urtasun

It seems like you have many goals.

Yes - I set high standards and work to be as good as possible. I guess it might impose some stress on the students.

Did you ever see a talented student quit?

Yes, actually I have. That is the most frustrating part of the job, I'd say. There are two cases that I can think of now. One quit because of personal reasons, not because they were unsuccessful. His wife couldn't find a job, and they had visa issues in Canada

thestar.com

Life · Fashion & Style

U of T scientists create software to analyze outfits

New program, which they hope to turn into an app, determines whether an outfit is stylish and offers suggestions.

University of Toronto researchers Raquel Urtasun, left, and Sanja Fidler are creating an app that assesses clothing and recommends how to be more fashionable. (STEVE RUSSELL / TORONTO STAR) | ORDER THIS PHOTO

at that time. There weren't a lot of job opportunities in Canada back then so they wanted to move to the US. Then he found an industrial job. He was a very good student so I'm still trying to get him back.

The other student was really, really talented. He said he wanted a taste of the industry before deciding to do the PhD. He might still come back. He went to work for a product team. It was a surprising choice because he has a lot of talent.

Students go through processes that can be quite frustrating. I think they want a taste of something different. In comparison, industry offers a more stable life than research.

What drives students to stick with academia?

Through these years, you learn a lot about yourself. You realize your limitations and boundaries before discovering what you really want to do.

What did you discover about yourself?

[*laughs*] Ahh, you're putting me on the spot! The learning process is never-ending. I didn't always know what to do. I didn't know if I wanted to work in industry or become a professor.

What convinced you to stay in academia?

In the beginning, I wasn't always sure. When I started, I was very afraid. I didn't know if I would be good at teaching and guiding students. I knew I wanted to try it. After the first year, I really enjoyed it, and now I cannot see myself doing anything else.

You have had extraordinary teachers. Which quality do you admire the most in your own teachers?

I had a lot of teachers that I learned from even from when I was a little kid. As a supervising student, I learned the most from Raquel Urtasun, who is also a Co-Founder of the Vector Institute. She guides students in a really natural way. She teaches them how to learn and how to approach problems. I am a little bit more chaotic. I tend to throw many ideas out there. Perhaps it confuses students. I learned that you shouldn't rush into things. You should go slowly and help them realize things by themselves rather than just by telling them. I think that is probably the best thing I learned from my teachers.

What is the most precious thing that you learned from your students?

[*laughs*] Wow, I've had so many! Every student is very different, and every student needs a different type of approach. Some like things to be very structured. I like to brainstorm so maybe I wasn't as structured. It changed the way that I interact. If I have ideas that I want to convey then I do it slowly.

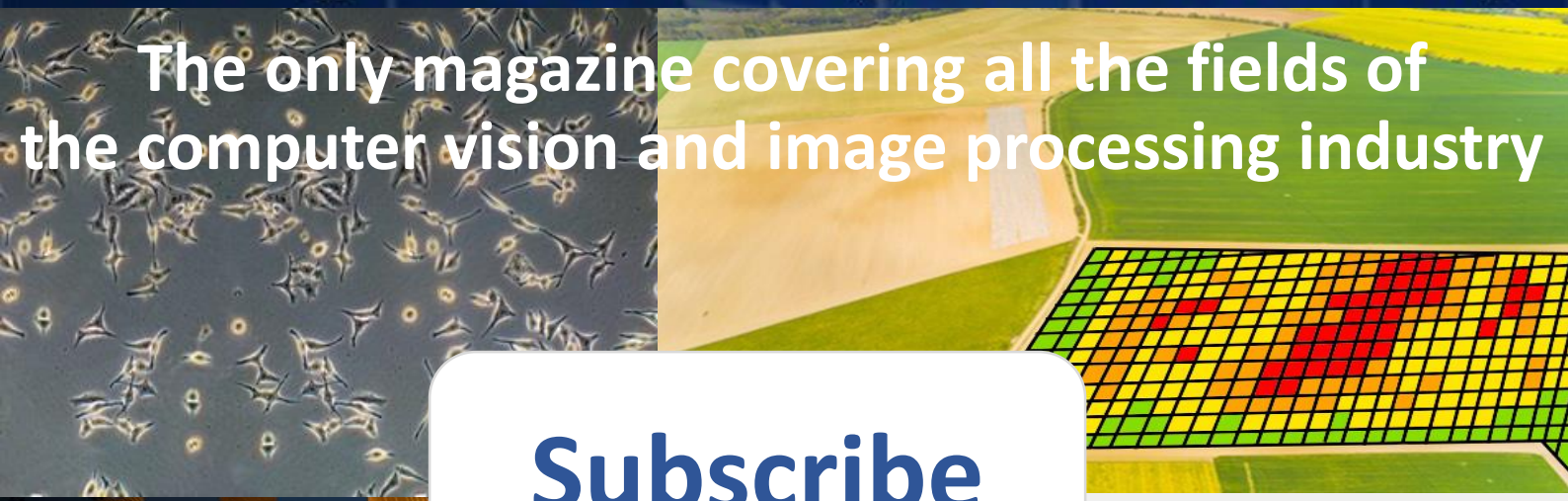
Have you seen benefits of this?

Yes - I've seen progress, and a lot of projects are going well. Again, some people like it one way, and some like it another way. The teacher should adapt.

Improve your vision with

Computer Vision News

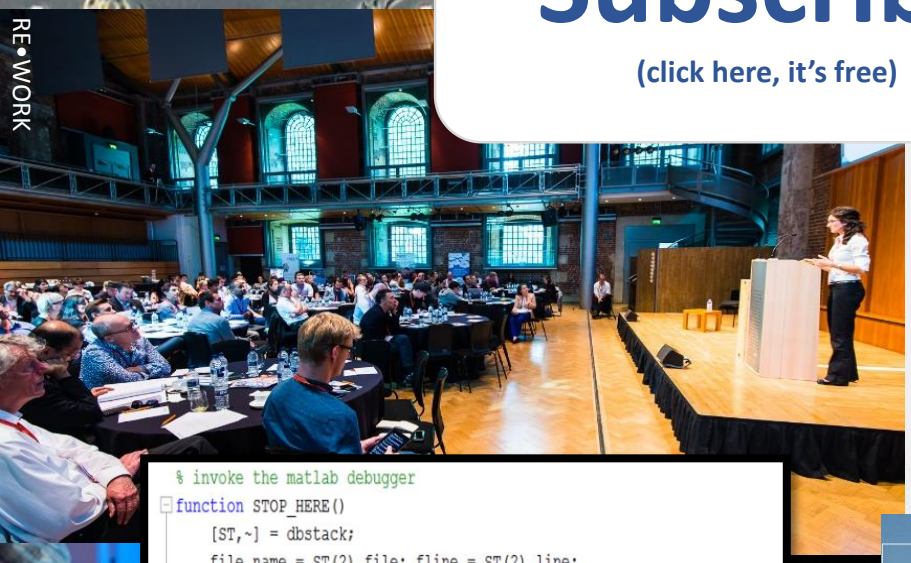
The Magazine Of The Algorithm Community

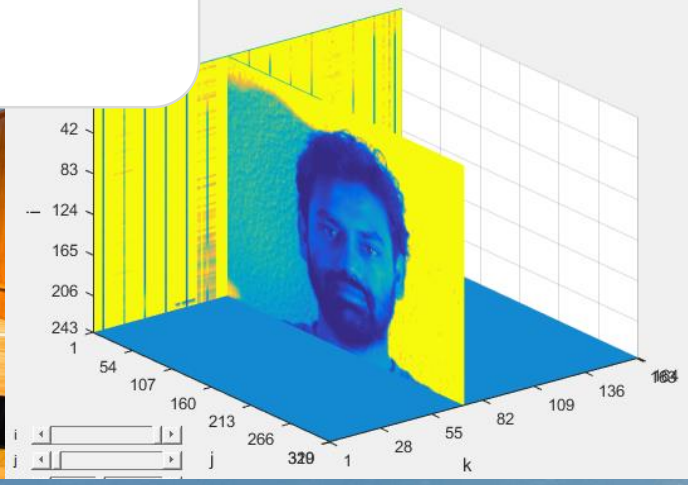


The only magazine covering all the fields of the computer vision and image processing industry

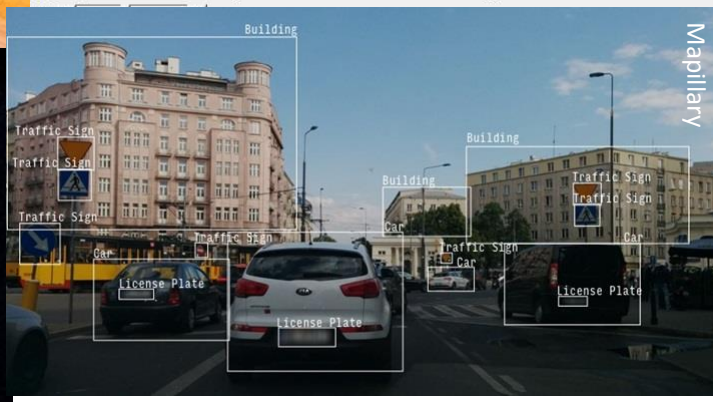
Subscribe

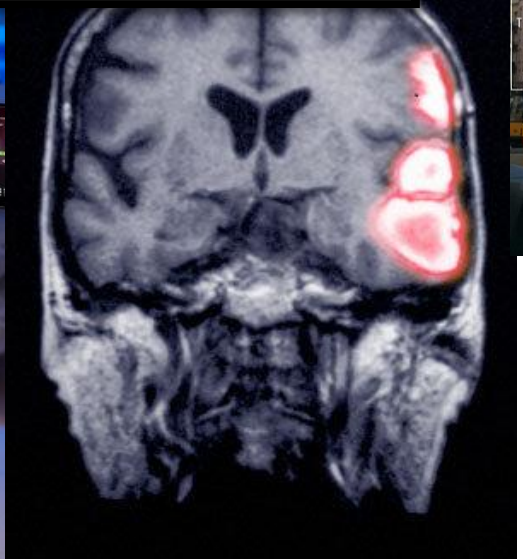
(click here, it's free)






```
% invoke the matlab debugger
function STOP_HERE()
    [ST,~] = dbstack;
    file_name = ST(2).file; fline = ST(2).line;
    stop_str = ['dbstop in ' file_name ' at ' num2str(fline+1)];
    eval(stop_str)
```





A publication by

RSIP VISION

Women in Computer Vision

Nour Karessli is a computer vision engineer at **EyeEm**, who are located in Berlin. She published "**Gaze Embeddings for Zero-Shot Image Classification**" here at CVPR, together with Zeynep Akata, **Bernt Schiele** and **Andreas Bulling**.

Nour, where are you working at the moment?

I work at EyeEm, which is a photography company, and we work on cutting-edge technology for computer vision. We connect a community of talented photographers with iconic brands and sell photos. I finished my master's degree in July last year and started at EyeEm in August, so it's been a year.

What was the focus of your master?

My master thesis was about gaze embeddings for zero-shot learning for classification. I did it at the Max Planck Institute in Saarland.

I understand you did not start your studies there?

I started my master's there, and before that I was doing a bachelor in Syria, at the Damascus University.

"We make use of the human ability to distinguish between different classes unconsciously"

You are doing a presentation today. What is the work that you are presenting?

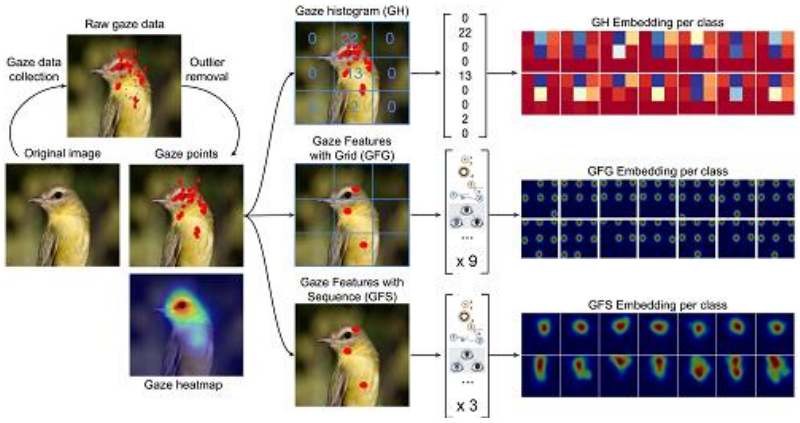
The work I am presenting is a paper about gaze embedding for zero-shot image classification. In this paper we use human gaze information to guide the classification task in a zero-shot setting. We make use of the human ability to distinguish between different classes unconsciously.

What is the novelty of this work?

Previous approaches used object discriminative properties collected by experts, and then the annotators had to go through the objects and annotate these attributes. This is very costly, especially for fine-grained classification. It's also difficult because the categories are visually very similar, and thus our suggestion is to use the gaze information. It's cheaper and faster, because it's implicit. You just ask the annotator to look at the image and distinguish between the objects, and then the human - without thinking about it - will focus on the important features.

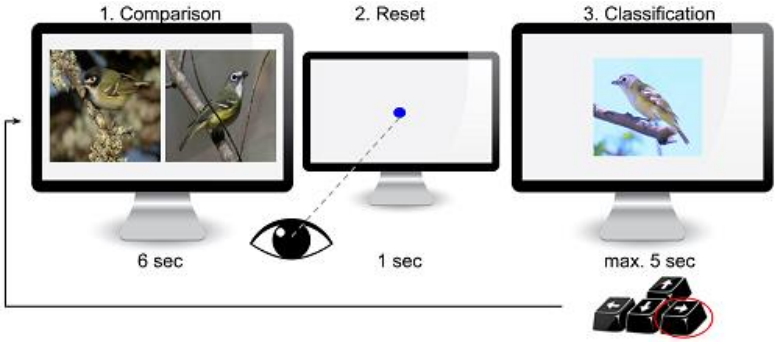
What particular example did you use in this work?

The main objective was comparing two types of birds, where we give the annotator six seconds to explore the objects and find the differences. Then we show one instance of the two previous classes, and give the annotator five seconds to make a decision to which class they think this image belongs. The five seconds was kind of short, and usually in gaze studies they use longer fixations. In our case it was a short time because the annotator had to give a fast reply, so we had to process the gaze data in a different way. We had to take out the outliers and have a shorter time for the fixations.



How did you solve the problem?

After we collected the gaze data, we processed the raw data and obtained the gaze points out of that. Then we wanted to come up with a class representation, which is needed in zero-shot learning to aid the classification task. To do this, we used the gaze data for the individual images of the class to get an image representation, and then averaged all these images to one class representation. So we had three types of representations, the details are all in the paper. We extracted many features from the gaze points - the location on the images, the duration, the pupil diameter of the annotator, and the sequence information between the points, which is the angle between subsequent points. Using this information we noticed that the location, duration and sequence was more helpful than using the pupil diameter. Studies say that pupil diameter helps indicating the concentration level of the annotator, but apparently as the annotator became familiar with the categories, their concentration dropped. So it wasn't very helpful to use this information, and we had better performance using only the other features.



What are the next steps for your work?

In future work we want to explore how to combine the gaze information from different images to represent one class in a better way. We would also like to do more experiments on more datasets. Our work compared species level of birds and pets, but we could explore more or larger fine grained datasets, for example asking how we can compare on the subspecies level.

You seem very passionate about this subject. What do you particularly like?

I always had this interest in computer vision and how vision works in humans, and how we understand that one object is different than another, just by one glimpse. So it was interesting for me to study the human behavior. Zero-shot learning is particularly interesting because as humans this is easy. For

example if I describe a giraffe to you in words, you would know how a giraffe looks without ever seeing one before, just by me telling you that it has a long neck and brown/yellow color. And then when you see it you are able to recognize it. But this is not the case with computers, and it's very interesting to be able to somehow transfer this knowledge to computers.

Does the fact that you are working with this subject change the way you look at the eyes of people?

It kind of affected my way of looking at people, because I was always wondering what the trigger is that we use when gazing at objects, and what makes us recognize them fast - the visual system is very fast. So this work changed my view on humans and how they focus on different regions. When I had to go through the data which we collected from 5 participants, I saw that they have different focus regions. For example one would only focus on the head always, or the body, and so you see that there is bias of the participant data.

"Eyes are the most important feature in the human face. The way that you look at people is a very strong way of communication"

Did you also try it yourself?

Yes, I did it myself - it took a long time, because I had to do all of them. I learned a lot about different birds species and dogs and cats, it was interesting. I tried it myself to make sure it's comfortable for others to do the experiment.

Is this something that interests you also before you started you scientific studies, the way that people watch?

Actually for me personally I always found that the eyes are the most important feature in the human face. The way that you look at people is a very strong way of communication.

How do I look?

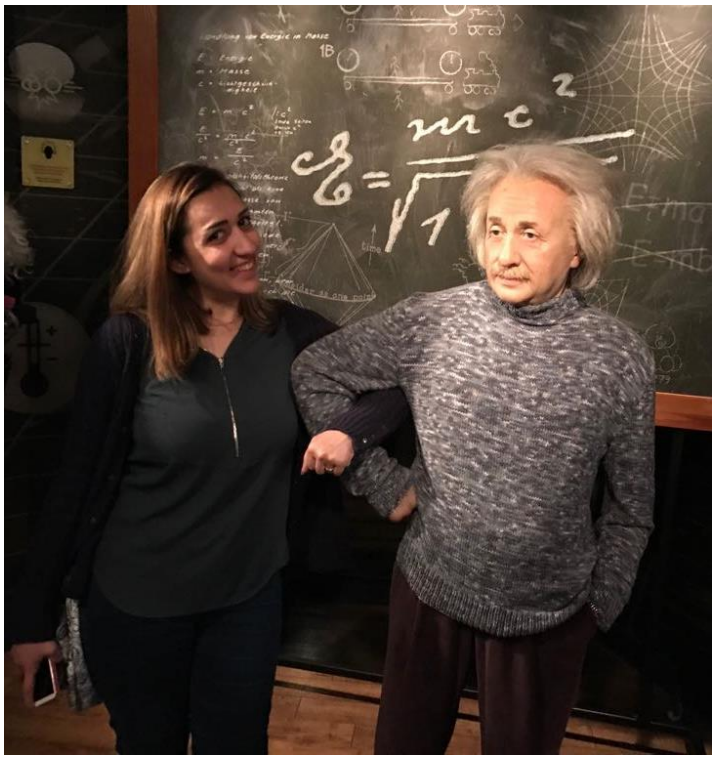
You look perfectly! [*We both laugh*]

But you know what the person is thinking of you, or if they like you, by the way they look at you, right?

You can obtain a lot of information by looking at the way they look at you.

How come that babies naturally look into your eyes - and not into your ears, for example?

It could be because they are always moving and thus attracting the attention. And they are also just nice to look at and sparkling [*she laughs*].



You have a special story: of the 5,000 CVPR participants, you are maybe the only one who lived in a war zone only three years ago. Can we tell our readers where you come from?

I originally come from Damascus, the capital of Syria, where I was born and raised. I came to Germany in 2014 to do a master's degree, and I already had plans to continue studying abroad. Germany especially has been very advanced in computer science recently. But of course, the war situation was the main motivation to leave the country.

"In other places that are not a war zone, you can build dreams without much worrying about the basic things"

You underwent many difficult things and had very strong experiences. It must be very difficult being a young woman in a war zone, dreaming of going away?

You're always feeling uncertain about everything. Whatever plan you come up with, you are always uncertain whether it will happen or not. In other places that are not a war zone, you can build dreams without much worrying about the basic things. But in the war zone, for example the electricity is unstable or the water station is unstable. So you are more focused on the basic life needs instead of focusing on your dreams.

Were you also worried about food?

In Damascus, especially in the city center, it was a bit better than in other places. But in the rural areas, it was sometimes under siege, and it was always hard for the people to get food.

Did these experiences make you stronger or weaker?

I think it made me stronger. Because I now know that as a human we are surprisingly adaptive to any situation. At the beginning, you will have a lot of fear and concern about even going out of your house. But then you grow a resistance and you are stronger to face these fears and just be able to continue your life. And I think I am now much stronger, because I know how to face my fears.

What is the latest challenge you had to overcome?

When I applied here, I wasn't sure if I would get the Visa or not, because of my Syrian passport. And even though I got the Visa, I was worried all the way here until I passed the borders, because I was unsure whether they would let me through.

I think that CVPR 2017 would have lost a lot if you wouldn't have had this Visa.

"I am now much stronger, because I know how to face my fears"

CVPR DAILY July 21-26 HONOLULU 2017 Tuesday

"A significant part of what makes us, as humar beings, so unique in the universe, is our curiosity"

Harry Shum is Executive Vice President, **Artificial Intelligence and Research Group, Microsoft**.

During your keynote, you discussed the relation between Research, Product and Business - and how these 3 elements should perfectly interact in order to have success in the marketplace. In your opinion, in our industry, which of these are working well together and which are not?

That is a great question! The product has a key role in the connection between those things. We mostly are trained as technologists. We actually are good at developing technologies. Most of us don't have that opportunity to get our hands dirty and deploy things into the wild or to learn what products are really doing.

To me, the most challenging thing is actually thinking about a product form that people actually care about. You asked me about these three elements. Everything is difficult. If you ask a business person, they will say technology is hard. For technologists, business is hard. The product is where you connect those things. The researchers need to push further to get out of their comfort zone. If you care about your product, business people can't just talk about the business model they need or what kind of products they should have built to get to the market. To me, this is where it's most interesting and exciting talking about the entire cycle.

You were talking about the idea of curiosity. It was clear that one of the things that drives your passion in this field is your curiosity for the potential of what can be achieved. How do you nurture curiosity in people, and how do you nurture curiosity in corporations?

A significant part of what makes us, as human beings, so unique in the universe, is our curiosity. Think about why we invented so many things. Many of those things come out of curiosity. How did people create the wheel? Why did people create the printing press? Now we know why people do things in AI or computer vision. I think a lot of that is because of curiosity.

I do agree with you that curiosity needs to be nurtured. This should start from a very young age, within your family, then when you go to school, and onto university. When you work in a company, in a place like Microsoft Research, I think we are very fortunate. We have a lot of people that

are not only smart, but who are curious, who want to do new things. To me, these are very important characteristics of curiosity. You always ask questions. If you are not curious then there's nothing new here.

In research, we always say that the most important question that is asked is: What's new? For example, what can Microsoft Pix do that others cannot? For me, those things need to be nurtured.

We have here at CVPR a lot of young startups trying to become the next Microsoft…

Of course… and some of them will!

Would you advise them to put curiosity at the highest level when they recruit new people?

Absolutely! I think it's especially important for startups because you need to find a way to, first of all, survive then thrive. By following those big companies ahead of them, they have to start with something new: something that others, the big companies, have not paid attention to. For me, I think curiosity is the most important thing for the startups. They need to have new ideas. Otherwise, why did you do a startup?

Do you have any friends or colleagues that had an extraordinary amount of curiosity when they started, but then it faded away as they got older?

I think as we get older, we do know more because of the experiences we have had. More often than not, by applying previous knowledge, you get a lot of things right. Curiosity is very important in your mental state. Did you want to have breakthroughs? Did you want to do those new things? Do you still care about how to do something extraordinary that hasn't been done before?

I read a book, **A Mathematician's Apology** by **Hardy**. Hardy said that when you get old as a mathematician, your intellectual power goes down after a certain age. I actually don't know if we apply this at computer vision. I feel that I'm at the prime of my career! [*we laugh together*]

Can you give some advice to young students on how NOT to become a 40 year old mathematician? [*we both laugh again*] **How do you keep this mindset alive?**

It's very, very simple. There's two kinds of advice I always give, even to my children. Be curious, and work hard. Nothing can replace those things.

What advice would you give to **students? Let's say someone who feels like a techie that finished high school and needs to choose a path in university. Knowing that half of the jobs that will exist in 5 years do not yet exist, how can he choose the best path?**

I don't think it's that complicated. If you look at human history, there are always new technologies and new professions coming up. I think it's all

about the basic skills that we really need to master. Things like language will always be there. That is the foundation of what our kids should have to learn, in high school and middle school. Beyond that there is a very clear trend that is already there. Many of the great minds are already talking about what they call "computational thinking". I think computational thinking is what we need to learn. You can argue that computational thinking is also a language skill. Instead of English or Chinese, we are talking about programming languages. It looks at the way you think, the numbers coming together, systems working together. That is something that the next generation will have, much like mathematics 50 years or 100 years ago. I'm not talking about necessarily anything beyond calculus, but even writing Python programs is not something that every kid can do. That is something that we, as a society, need to make more explicit. I'm glad to see more American children in high school learning programming. I think that is the right thing to do.

I moderated a panel yesterday about the shortage in STEM workforce. I know that you in the big corporations are fighting to get the best talents. As a result, Microsoft, Google, Facebook, Apple and Amazon recruit many of them. The problem that I see is that there are huge pools of talent that have no access to education or these opportunities. I think about how many women or people living in some regions of the world might have the same opportunities. How can you as a corporation help reach out to those who do not have the opportunities to showcase their talents?

We are just as passionate as you are. We think about these important issues. One thing I mentioned in my speech and in the **Microsoft Mission Statement** is that we want to empower every person and organization on the planet, and not just here in Hawaii. We say that, and we mean that. That actually reflects in what kind of products we make and the kind of features that we build. That can include taking care of people with learning disabilities or dyslexia. We are talking about people in different continents. Our CEO travels frequently and regularly around the world. A good example of this is what we recently announced about making Wi-Fi available in the United States. We have been engaging with many others in many other countries too. We have to democratize technology. Microsoft has been on that path forever, ever since **Bill Gates** said to put Microsoft Software on every desktop. This is democratizing technology. I think we have this responsibility to help everyone on earth to do that, I completely agree with you.

Steve Cruz, Harry Shum and Terrance Boult yesterday at CVPR2017 (photo courtesy of Steve Cruz, Vision and Security Technology Lab)

3D Menagerie: Modeling the 3D Shape and Pose of Animals

Silvia Zuffi is a postdoctoral researcher at the **Institute of Applied Mathematics and Information Technologies** in Milan, Italy. Her current work focuses on 3D models of animals, and her CVPR paper is a joint work with **Angjoo Kanazawa**, **David W. Jacobs** and **Michael J Black**.

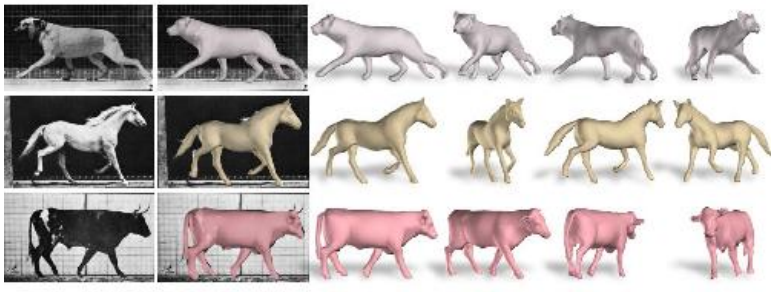
We talked with Silvia about her current work, and she told us that "*there are many models of the human body, but so far, there is none for animals*". However, there are many applications in different fields where animal shape and motion capture can be useful, like studying animal motions in biology, or applications in entertainment.

The main difficulty that has prevented the development of 3D animal models so far is the data acquisition. You cannot simply follow the same pipeline as you would do for creating 3D scans of humans. There, data acquisition can be done by inviting people to your research lab, asking them to stand still in a specific pose, and then making a 3D scan of their body - but you cannot ask the same thing of a tiger.

To efficiently overcome this problem, Michael Black came up with the idea of scanning toys instead of real animals. The advantage is that they do not move and you can easily scan them, but this also means that they might not be in the pose which you want to model, because you cannot scan thousands of toys in all possible shapes. For training the model, you need to put all the 3D scans in vertex-to-vertex correspondence, which is called "registration". This is easy to do for humans, if you have many scans of the same pose. Toys are all in different poses, however. To make the 3D models useful for visual and graphical applications, Silvia and her co-authors propose a method which allows them to align the scans of animals with different shapes and poses.

One topic of future research that Silvia sees is to tackle the problem of learning how the animal body deforms with changes of pose. This cannot be learned well with toys since they are static, so it remains an open question of how to get scans of moving animals, to get pose-dependent deformation - without having to bring wild animals into a lab. She concluded our interview with the words: "*It was great to build the model. Now we want to extend it to a lot more type of animals*".

If you want to learn more about Silvia's work, go to her spotlight presentation today at 08:54 in the Kalākaua Ballroom.



Vamsi Ithapu

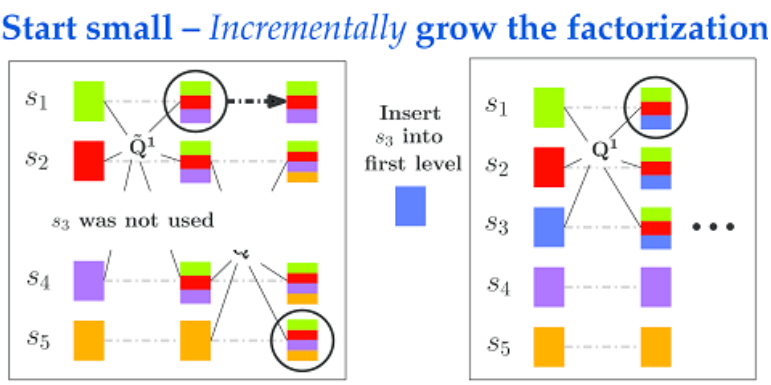
CVPR DAILY
July 21-26 HONOLULU 2017 Tuesday

The Incremental Multiresolution Matrix Factorization Algorithm

Vamsi Ithapu is a PhD student at the **University of Wisconsin Madison**, and he is working on explainability of deep neural networks. His current work is a joint work with **Risi Kondor**, **Sterling C Johnson** and **Vikas Singh**.

Vamsi told us that the motivation and the context of his current work started with the success of deep learning. He describes deep learning as “a beast which needs to be tamed”. That is, we need machine learning and computer vision tools which can decode the learned deep neural networks. “There is a lot of software which you can use to train a good networks, but we also need to understand what the deep representations really mean, and their relationship to human semantics”, Vamsi said. His work contributes to taking first steps towards that goal. It is asking: given a trained network, what interesting deep representations did this network learn? Do the deep networks see what humans see with respect to different categories? Vamsi and his co-author’s approach to answering these questions is to take symmetric matrices from deep network representations, and factorising the complex hierarchical block structure in this data. The existing tools for this, like PCA, use low-rank and global methods and are not adequate, Vamsi told us. Therefore he proposes a novel factorisation, which he calls incremental multiresolution matrix factorisation. This is the first Mallat-style wavelet on symmetric matrices. Constructing wavelets on matrices themselves, instead of non-euclidean data like grass and trees, is a relatively new development, Vamsi told us. They visualise the factorisation that this method produces in a nice way, which they called MMF graph.

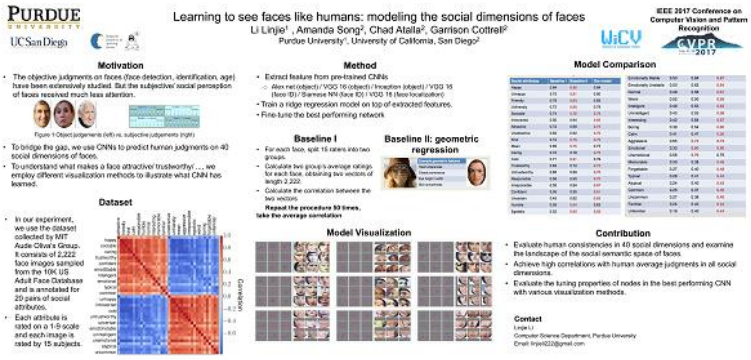
If you want to learn more about Vamsi’s work and see examples of these visualisations, visit his talk titled "**Decoding Deep Networks**" at the Explainable Computer Vision workshop on Wednesday, July 26. He will talk about how to use the factorisations they propose in their paper to study deep neural networks and the evolution of semantics in a neural network, and how these compare to human semantics.

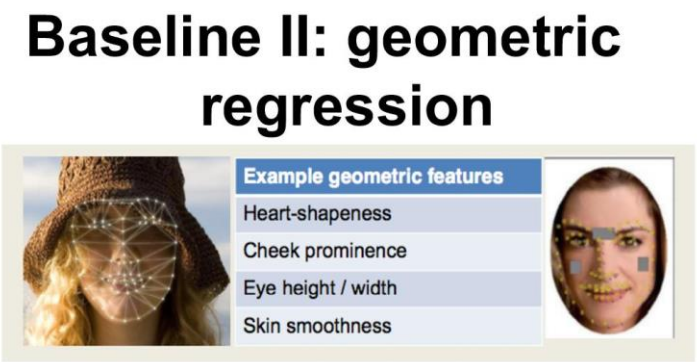


Learning to see faces like humans: modeling the social dimensions of faces

Linjie Li is currently pursuing her PhD in Computer Science at **Purdue University**. Prior to her Ph.D., she obtained a master's degree in Electrical Engineering from **UCSD**. She was working as a research assistant at **GURU lab** in UCSD, focusing on machine learning, computer vision and neural networks.

In the era of the digital age, we are constantly forming first impressions on others by browsing each other's photos online. Although first impressions seem to be subjective, psychological studies have shown that there is often a consensus among human in how they perceive attractiveness, trustworthiness, and dominance in faces. Are deep learning models, which have successfully conquered various visual tasks, also capable of predicting subjective social impressions of faces? To answer this question, we systematically examine 40 social features on faces and use deep learning models to predict human first impression on faces. Employing the internal representations from pretrained neural networks (for object classification, face identification, face landmark detection), we build a ridge regression model on top of the extracted features and our model can successfully predict human social perception whenever human have consensus. We further visualise the key features defining different social attributes to facilitate an understanding of what makes a faces salient in a certain social dimension.





This work, prepared with Amanda Song, **Garrison Cottrell** and **Chad Atalla**, will be presented tomorrow (Wednesday) at the **Women in Computer Vision** (WicV) Workshop.

Laura Leal-Taixé
Sergi Caelles

CVPR DAILY
July 21-26 2017 HONOLULU Tuesday

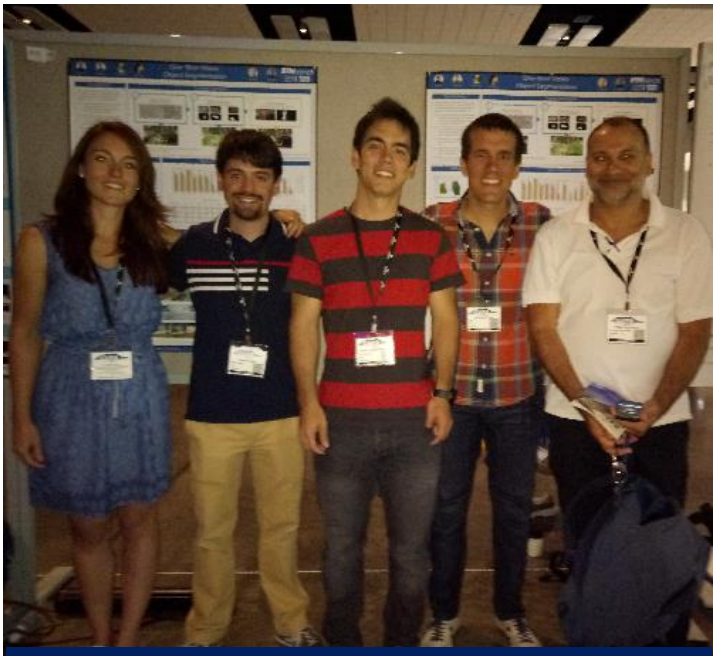
One-Shot Video Object Segmentation

Sergi Caelles and **Laura Leal-Taixé** are both authors of this CVPR paper, which was joint work with **Kevis-Kokitsi Maninis**, **Jordi Pont-Tuset**, **Daniel Cremers** and **Luc Van Gool**. Sergi is a PhD student at **ETH Zürich** and Laura is a postdoctoral researcher at the **Technical University of Munich**.

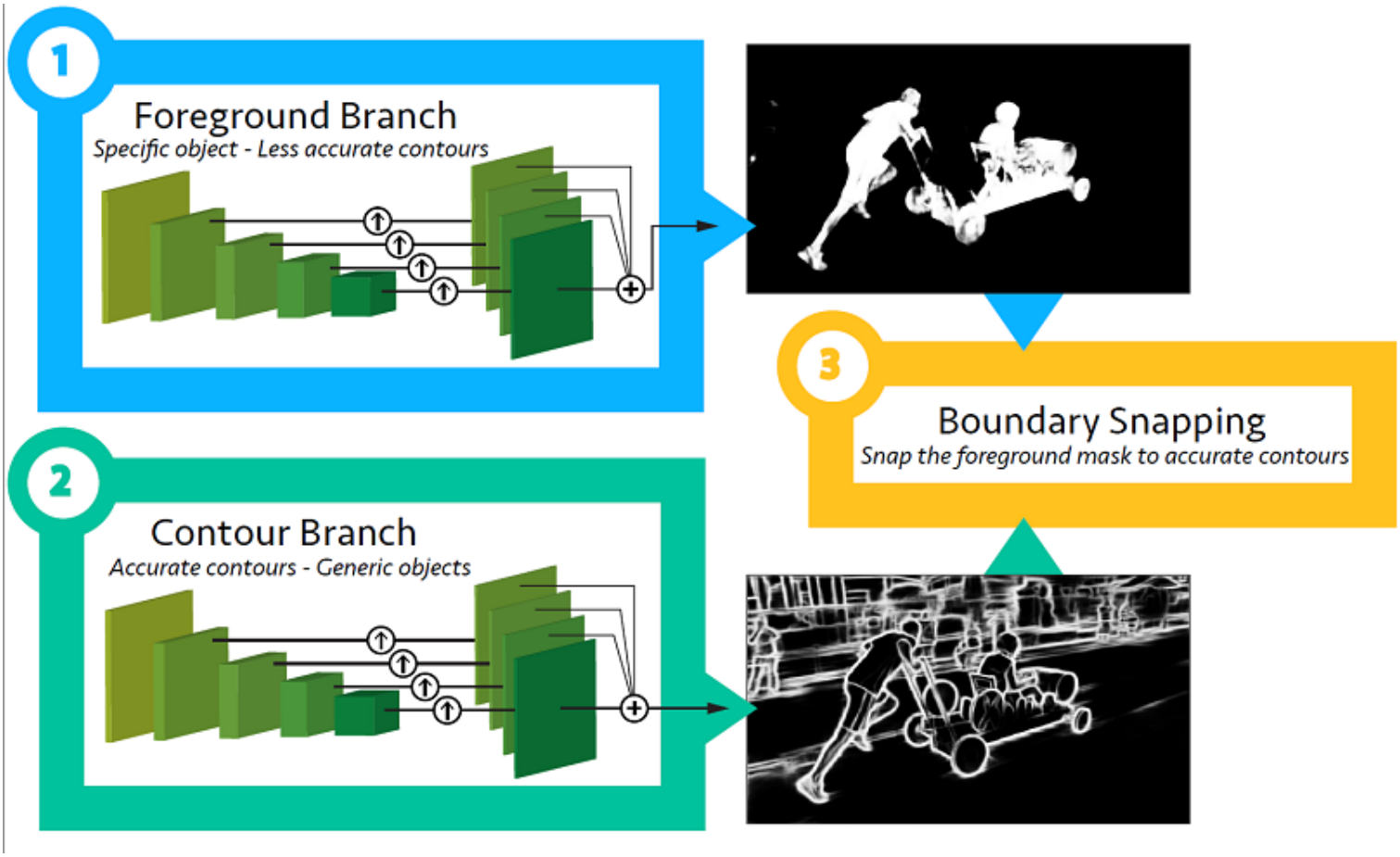
Their work focuses on semi-supervised video object segmentation: given a video, the goal is to segment a specific object for the whole video, given the first frame. They are the first to approach this task using deep learning. When we asked why nobody has taken this kind of approach before, Laura says: "*Ideas are not so easy to come by*". However they method itself is not very complex, she explained, but it's a fast method and it clearly outperforms the state of the art. "*Sometimes, simplicity works really well*", Sergi says. He told us that one of the most challenging parts of this work was to use all the parts of the model in the right way, although the model architecture was not very complex.

Their approach consists of a separation into first training the parent network with the DAVIS training set, using a fully convolutional neural network, to separate objects from the background. Then they fine-tune on the first segmentation mask of the video. "*This really is the key point: that you learn the appearance of the object during this fine-tuning*", Laura told us.

In the current work, they are considering the appearance model of only a specific object. In a follow-up paper, Sergi tells us, they are introducing the concept of instance and segmentation into the mix. They then not only learn the appearance of the object, but also the category of the object and that it is a particular instance of the object.



The authors Laura Leal-Taixé, Sergi Caelles, Kevis-Kokitsi Maninis and Jordi Pont-Tuset catching up with our editor at their poster, yesterday at CVPR.



Food for thought. At the moment of closing this last CVPR Daily of this year, we receive the following thoughts from **Albert Ali Salah**: As we rely more and more on algorithms, one topic that is gaining importance is algorithmic accountability, yet it is not so much known in the CVPR community yet. The questions we ask are: Are the algorithms we use biased in any way, for instance, do they treat white people better than the colored, or prefer males over females? Jurafsky touched upon some of these issues in human behavior, but computers also have biases... I think it is a question worth asking in the magazine...

I will discuss accountability and why we need it in my presentation on the 26th. I will present the paper that won the ChaLearn competition this year, on deciding (from short videos) whether someone was invited for a job interview or not, and then explaining the decision.

Ralph and Sea Turtle

CVPR DAILY
July 21-26 HONOLULU 2017 Tuesday

The editor of CVPR Daily and his friend Sea Turtle thank you for reading our magazine at CVPR 2017 in Honolulu. In true Aloha spirit, we hope you enjoyed our work! See you in Salt Lake City at CVPR 2018. Mahalo!

Image courtesy of Erika Roberts at waikikidiving.com
Dive with them, they are awesome!