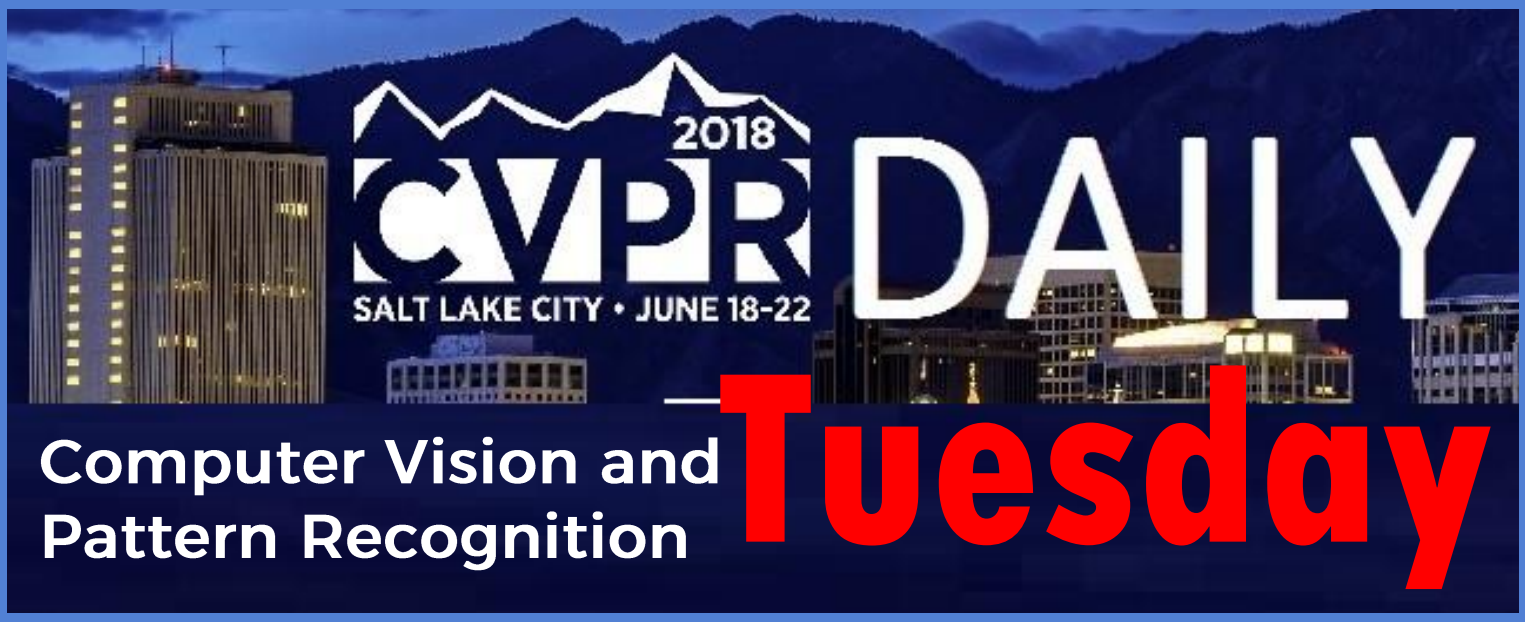
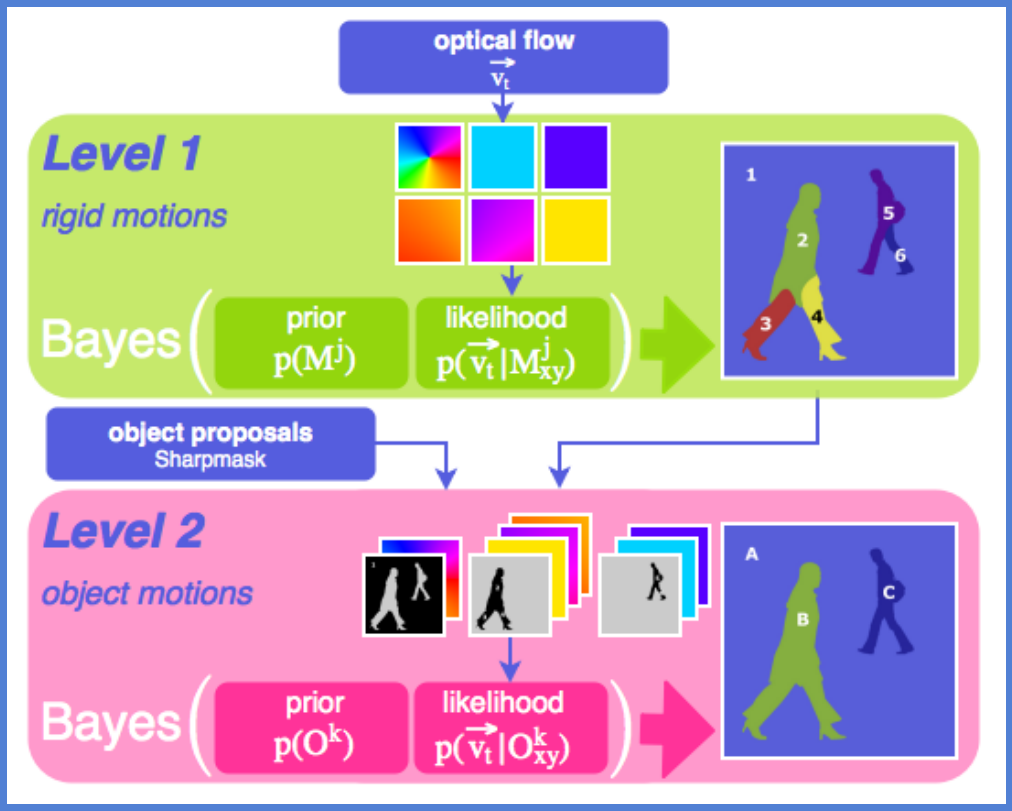


Computer Vision and Pattern Recognition Tuesday

Today's Picks by:
Carlo Dal Mutto



Presenting Work by:
Andrea Zunino
Fabien Baradel
Federico Pernici
Jingya Wang
Pia Bideau
Sarah Adel Bargal

Program co-Chair:
Michael Brown

Women in Science:
Nasrin Mostafazadeh

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

A publication by
RSIP VISION

For today, Tuesday 19

Carlo Dal Mutto is CTO of Aquifi Inc, working on advanced solutions that combine 3D and AI for removing errors in manufacturing and logistics. Carlo olds a PhD from the University of Padova, Italy. He is author of two books, several papers and is the inventor of over 20 patents regarding depth sensing, 3D reconstruction and deep learning applied to 3D data.

"KCNN is a deep-learning based approach for extremely efficient keypoint detection. KCNN is a framework able to effectively emulate different keypoint detectors (e.g., KAZE, SIFT) and the proposed CNN architecture is suited for efficient implementation on embedded systems such as FPGAs, resulting in less than 1 millisecond computation time.
KCNN received yesterday the best paper award at the Embedded Vision Workshop. We are very proud of this award and we really believe that efficient embedded implementation of deep learning techniques constitutes a fundamental pillar for many real-world applications."

Carlo's picks of the day:

Orals:

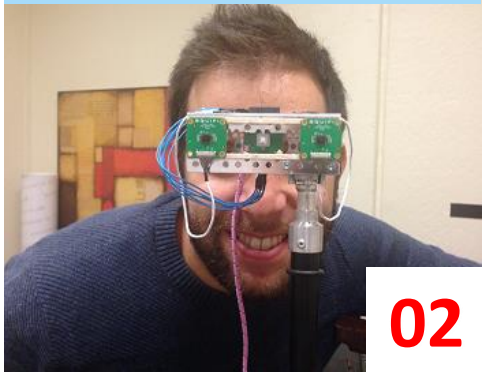
08:50 (O1-1B) [A7] "*Finding Tiny Faces in the Wild With Generative Adversarial Network*"
08:50 (O1-1C) [C13] "*A Certifiably Globally Optimal Solution to the NonMinimal Relative Pose Problem*"
09:40 (O1-1A) [A4] "*Learning by Asking Questions*"
14:50 (O1-2A) [A6] "*Deep Layer Aggregation*"
14:50 (O1-2C) [E10] "*Learning to Find Good Correspondences*"
14:50 (O1-2B) [C12] "*SPLATNet: Sparse Lattice Networks for Point Cloud Processing*"
14:50 (O1-2C) [E13] "*OATM: Occlusion Aware Template Matching by Consensus Set Maximization*"

Spotlights:

15:48 (S1-2C) [G2] "*Deep Adversarial Metric Learning*"
15:48 (S1-2C) [F9] "*Self-Supervised Feature Learning by Learning to Spot Artifacts*"

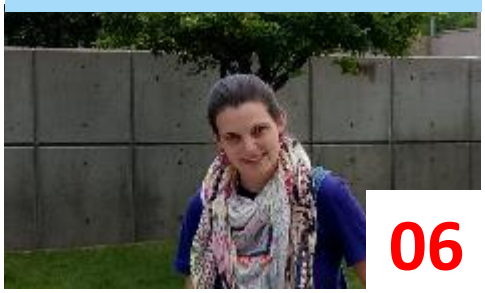
Demo:

Carlo invites us with Paolo Di Febbo to Hall C demo: 10:10-12:30 "*KCNN: Extremely-Efficient Hardware Keypoint Detection With a Compact Convolutional Neural Network*"

Carlo's Picks

02

Welcome to CVPR18

04

Pia Bideau

06

Federico Pernici

08

Sarah Adel Bargal
Andrea Zunino

10

Jingya Wang

12

Fabien Baradel

14

Woman in Science
Nasrin Mostafazadeh

16

Good morning CVPR!

For the third year, **CVPR** and **Computer Vision News** (published by **RSIP Vision**) join forces to bring you a **CVPR Daily** magazine during the main program of **CVPR2018**.

In this online magazine you will find some of the highlights of yesterday's program, ideally mixed with several previews of today's program. Be sure to have a look at our advices for today's presentation. After thanking all the readers and all those who accepted to talk to us in order to prepare this mag, my warmest gratitude goes to **Nicole Finn** and her team at **CtoC**, both for the perfect organization and for sending you once again the CVPR Daily news. Enjoy the reading and the conference!

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

CVPR Daily
Publisher: **RSIP Vision**
Copyright: **RSIP Vision**

All rights reserved
Unauthorized reproduction
is strictly forbidden.

Our editorial choices are fully independent from IEEE, CVPR and the conference organizers.

General Chair - Michael Brown

Michael, we are here once again at CVPR. Have you got a message for our participants?

Firstly, I would like to welcome everybody to CVPR this year. I hope they enjoy a great event. We have a huge set of offerings, from 70 workshops and tutorials, to a massive expo, to an academic papers programme with over 900 papers. This is the biggest CVPR ever: 6,500 people! I just hope everyone enjoys the cool air, the mountains, and the academic programme.

Do you have any advice about how best to take advantage of the overwhelming scientific offering that this conference has to give?

One thing I would advise everyone to do is study the programme guides. It is important to look at the programme guide because, I will be honest, I think CVPR has grown to the stage that within the 5 days that we have it is just not possible to do everything.

You have to make a choice.

[*laughs*] Yes, you have to make a choice! We have two programme guides to help you – a pocket guide for the workshops and tutorials, and a pocket guide for the main academic papers programme. We will also be giving out maps to help you figure out which exhibitors you want to meet and the poster locations. Also, I would highly encourage everyone to go to a tutorial or workshop. Make the most of your time. Do spend a little bit of time in Salt Lake City too, because there are a lot of interesting things to do. Of course, we can't compete with Hawaii in terms of having the ocean.

Not yet!

[*laughs*] Not yet, but there is a lake! A lot of people don't realise that it is a very eclectic city. There are plenty of interesting restaurants and breweries, so enjoy yourself. Have a good time!

Do you have any advice for first-time participants on how to take advantage of the Disneyland of possibilities that is open in front of them?

[*laughs*] First-time participants may be overwhelmed! There really is a lot going on and it depends on what you are coming for. CVPR has grown and in particular, the expo has grown so much. We have a lot of people who are coming for the expo now. We also have new people coming for the first time to learn what computer vision is about.

For those people, I highly encourage them to consider the tutorials. Those tutorials cut across everything in computer vision. Look at the interesting papers in the papers programme too. In the workshops, look for the keynote speakers and the various papers that they can go and see.

"Make the most of your time!"

The expo has been growing and growing over the years. Does this mean that industry is becoming a more important part of CVPR?

Yes, industry is becoming an important part of CVPR. Partly because this is the industry that wants to hire our students and our professors. It shows that our talents and our skill sets are in demand. Most trade shows that you go to, companies show off what they do so that you buy their products. I think CVPR is unique as a lot of these companies are hiring. They are actually here to find talent. It's very exciting, especially for students, grad students, and people looking to change their career paths. This is a really great place to be right now.

It feels like the balance between academia and industry in CVPR is becoming more balanced over the years.

It is becoming more balanced. I think the other thing we're seeing is the research coming out of the labs is very high quality. A lot of great research is coming from the industry, in addition to academics.

Finally, is there anything you'd like to say about the organisation of the event?

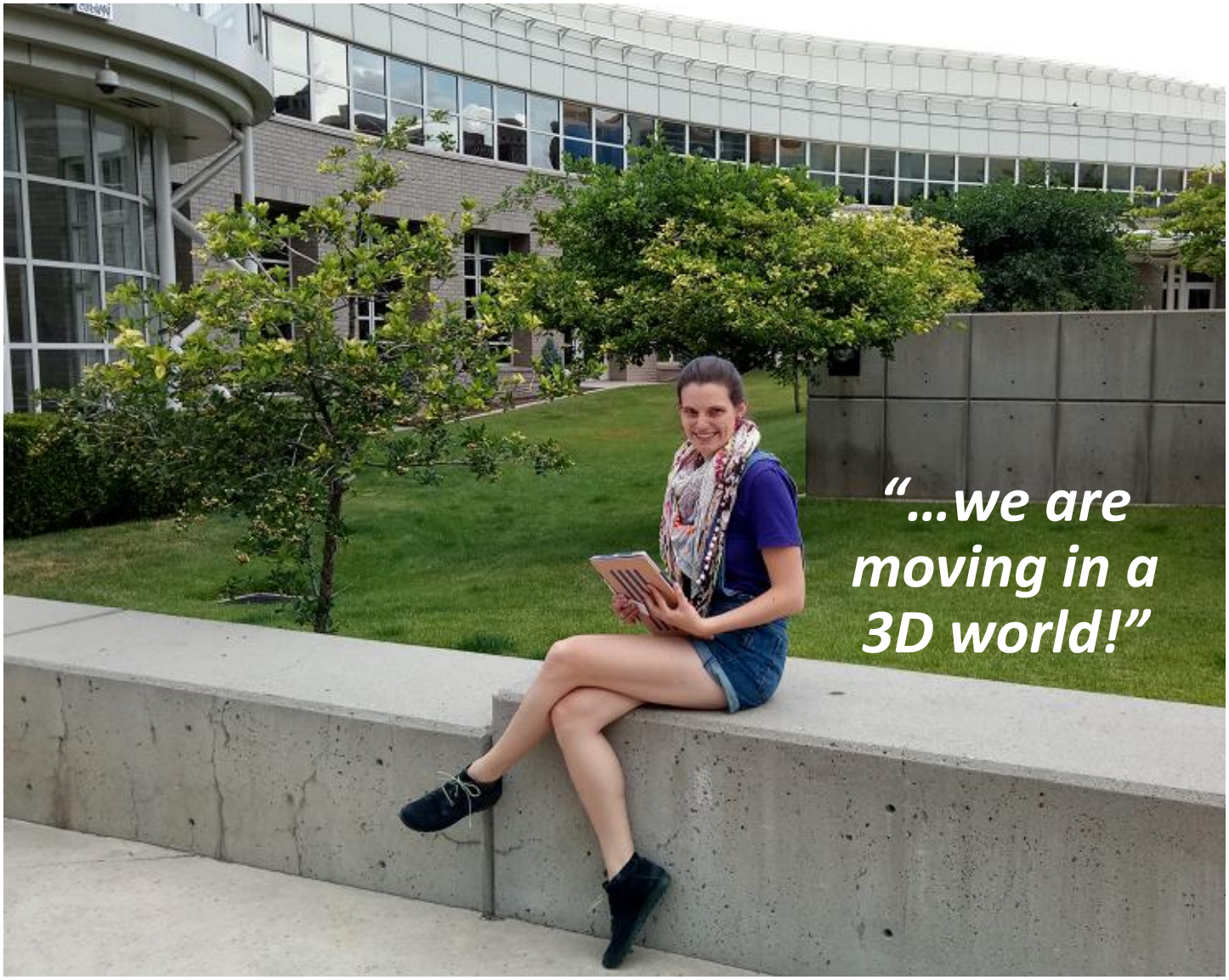
Yes, it's been absolutely fantastic. My co-chairs Shmuel Peleg and Bryan Morse, and of course, Nicole and Liz from CtoC Events, they've been fantastic in helping us to run this event and staying with us to help grow it. Something that a lot of people don't know is that Nicole's been with us for many years and many of the innovations that you see are really because of Nicole. We are also using Hall-Erickson to help us expand the expo and to manage all of the companies that are coming in.

"A lot of these companies are hiring!"

Our editor with Michael Brown

Pia Bideau

Tuesday
CVPR 2018 SALT LAKE CITY • JUNE 18-22 DAILY



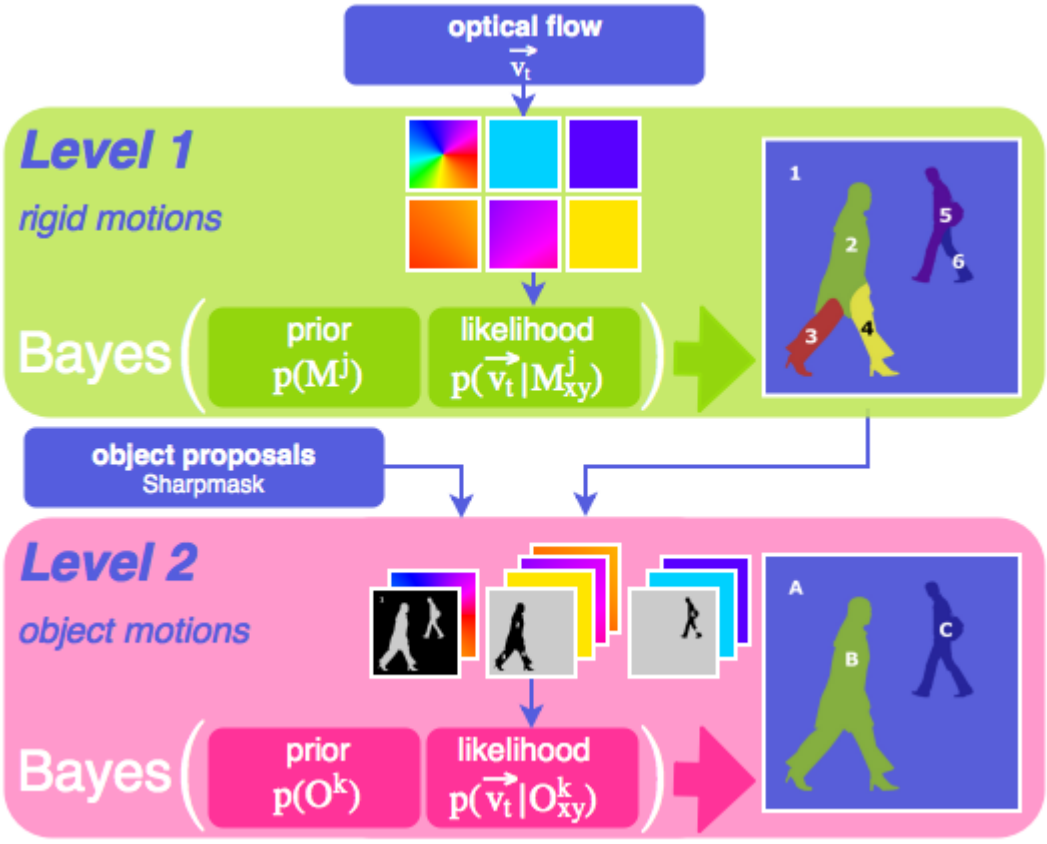
Pia Bideau is a PhD student at the University of Massachusetts at Amherst. She speaks to us ahead of her poster presentation today.

Pia's work is about **motion segmentation**, which is the **segmentation of independently moving objects in a video**. Both the camera and objects can be moving – for example, cars, humans, animals – and we want to detect all those objects that are moving differently to the camera.

Pia explains that motion is very important for human vision, for scene understanding, and to get to know the world we are moving in. When a person is walking through a forest or just walking along the street, it is easy to detect objects as soon as they move. A moving car will draw your attention. A squirrel will run up a tree. You will detect them immediately just because they are moving.

The challenge of this work comes because **we are moving in a 3D world**. We have objects that are very close to us and objects that are very far away. Objects that are close might move more on the image plane than objects

"If you solve all those problems together, they can support each other!"



that are very far away, but actually they do not necessarily move. A tree which is close to us is displacing a lot, and a tree which is far away is moving just a little, but none of them actually move. Only the person or the animal is moving. There are a lot of connected topics like object segmentation, depth estimation, and motion segmentation which we have to consider all at the same time.

Pia explains: "*We look at how the camera is moving. The camera can be translating, the camera can be rotating, and an object can be moving. All those three ingredients are part of the motion in the world which creates the optical flow. We first estimate the camera rotation and we subtract this off the optical flow, such that we only have a translational flow field, which has the camera translation and the object motion. If you just look on the direction of the optical flow, not on the magnitude, then you can see which objects are moving into a different direction. We developed a system which is considering the anti-optical flow and the derived flow likelihood, which computes how likely is a motion model given the observed optical flow.*"

In terms of next steps, Pia tells us that she will focus more on solving all those problems at the same time – **depth estimation, motion segmentation, object recognition, as well as estimating optical flow**. She predicts that if you solve all those problems together, they can support each other.

If you want to find out more about Pia's work, come along to her poster [H14] today, 10:10-12:30 in Halls C-E.

You have to throw data away instead of collecting it, because it is redundant. The basic idea is to reduce redundant data.

Federico Pernici is an assistant professor and researcher at the University of Florence. He talks to us before his poster presentation today.

Federico describes his work as a lifelong learner in which we are learning the **multi-object appearance of faces**. It's multi-target tracking, where we are tracking multiple faces, but able to learn their appearance in a lifelong way. The idea is once they exit the field of view and re-enter a few minutes later, we can provide the same ID.

He says the challenging part is that video does not provide IID data. Typically, in supervised learning, you have IID data. They are already labelled and are generated from an unknown distribution but in an IID. Humans have already selected the best patterns. In video, you are always viewing the same person with the same appearance, so after a few seconds, you don't need any more data on that person. **You have to throw data away instead of collecting it, because it is redundant**. The basic idea is to reduce redundant data.

It uses a memory-based algorithm, following the recent idea of memory-augmented networks in a very smart way. Typically, **backpropagation is very slow**. If you want to backpropagate your network at runtime, you have to wait too much time. The idea is to use a memory which is very fast to be loaded and unloaded at runtime. It's a challenge, because the system is running in real time.

The real-world application of this is **open world face recognition**. Federico explains: "*Typically, with face recognition you have the bad guys already labelled. In this approach, anyone can in principle be there. If I don't know you, you go into a gallery. It's a dynamic gallery, so the gallery of the bad guys is not fixed and is built at runtime. It's an open border.*"

In terms of next steps, Federico says that is about learning the representation in real time. In this approach, the representation – which is a descriptor taken from the **VGG (Visual Geometry Group) Face**, a huge deep neural network pre-trained at face recognition – is fixed. The idea is that once you have collected so many faces, you can fine-tune on top of these new faces.

You are all invited to learn more by coming along to Federico's poster [Q9] today, at 12:30-2:50 in Halls C-E.

In video, you are always viewing the same person with the same appearance, so after a few seconds, you don't need any more data on that person.

"Actually, the trickiest part was to extend this method for LSTM architecture. We solved it by normalising in time these backpropagated probabilities."

We speak to Andrea Zunino and Sarah Adel Bargal ahead of their poster presentation today. Andrea is a postdoc researcher at the Italian Institute of Technology in Genoa in the Pattern Analysis and Computer Vision department. His supervisor is Vittorio Murino. Sarah Adel Bargal is a PhD student at the University of Boston in the Image and Video Computing Group. Her supervisor is Stan Sclaroff....

Andrea and Sarah's work presents a top-down saliency framework for vision understanding. Portraying a **CNN-LSTM architecture** trained for action recognition in videos or video captioning, their work is able to highlight in the video frames which are the spatial temporal evidence that the model has used to classify the action or to generate the specific word in the caption.

The aim is to give an explanation about why these **deep learning techniques** are the best methods to use for video understanding and to help understand why they are working. CNN approaches are already working well for image understanding. What they have done is try to extend these explanations for CNN but also using LSTM methods, considering video as input.

Andrea says they have considered the excitation backprop, a CNN-based top down saliency framework devised for image understanding, and they have extended the approach for video understanding considering recurrent models (**RNNs**).

Their challenge was to extend this method for video understanding. In video understanding, the difficult part is that you have to also consider the temporal information, because there is not only the spatial information of the input. He says the most challenging part was to extend this method considering the LSTM part. The actual functioning of the LSTM is difficult because there are a lot of gates and other things inside the architecture. Since their approach is probabilistic that tried to backpropagate the probability inside the net, trying to see which are the excitatory connections that the model has used to classify the label.

Andrea says: "*Actually, the trickiest part was to extend this method for LSTM architecture. We solved it by normalising in time these backpropagated probabilities. The probabilities that you are backpropagating you have to preserve in time, because LSTM also considers the time. The most challenging part was to normalise this probability for all the time-steps to consider in the clip. Basically, we have added this temporal normalisation in the backpropagation.*"

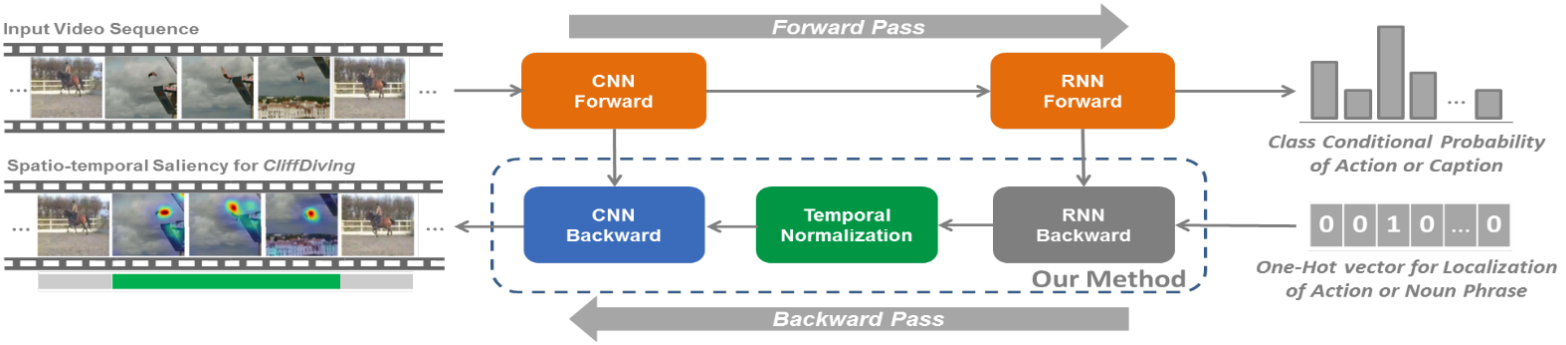
"…grounding can take other forms including: identifying evidence within the model, or identifying evidence within the training corpus."

Sarah adds: "*We know already the next steps for this work. In this work we focus on one form of grounding: identifying evidence of a prediction within the image or video input. However, grounding can take other forms including: identifying evidence within the model, or identifying evidence within the training corpus.*"

Andrea tells us that he has enjoyed working with Sarah on this project and feels that their skills are complementary. He is enthusiastic about their collaboration and that of the two institutions and hopes it will continue.

To find out more about their work, please visit Andrea and Sarah's poster [D22] today at 12:30-2:50 Halls C-E.

"What they have done is try to extend these explanations for CNN but also using LSTM methods, considering video as input."



"The reason we transfer the knowledge in the attribute space is because normally for the human re-identification, the ID labels from different domains, different data sets, are independent – they don't have overlaps – but in our study, we want to transfer the knowledge."

Jingya Wang is a final year PhD student at Queen Mary University of London. She speaks to us ahead of her poster presentation today.

Jingya tells us that her work is about **jointly learning the attribute and identity of person re-identification under an unsupervised setting**. Matching the person from non-overlapping camera views in different locations. Most existing person re-identification works under a supervised setting that needs a large amount of human annotation. Her work focuses on the transferable unsupervised setting, transferring knowledge from the existing dataset to the unseen new dataset.

The work focuses on the jointly-modelled global identity and the local attribute information, because there is a heterogeneous problem for multi-task learning, so she proposes a progressive knowledge fusion mechanism by encoder-decoder networks for intermediate space that progresses transfer for the domain adaptation.

Jingya explains that previous re-identification work has focused on the unsupervised setting that needed a larger amount of pairwise data. In this work, they want to jointly learn the attribute and identity space to get the better feature extraction for the person re-identification.

"This is an open-set recognition problem."

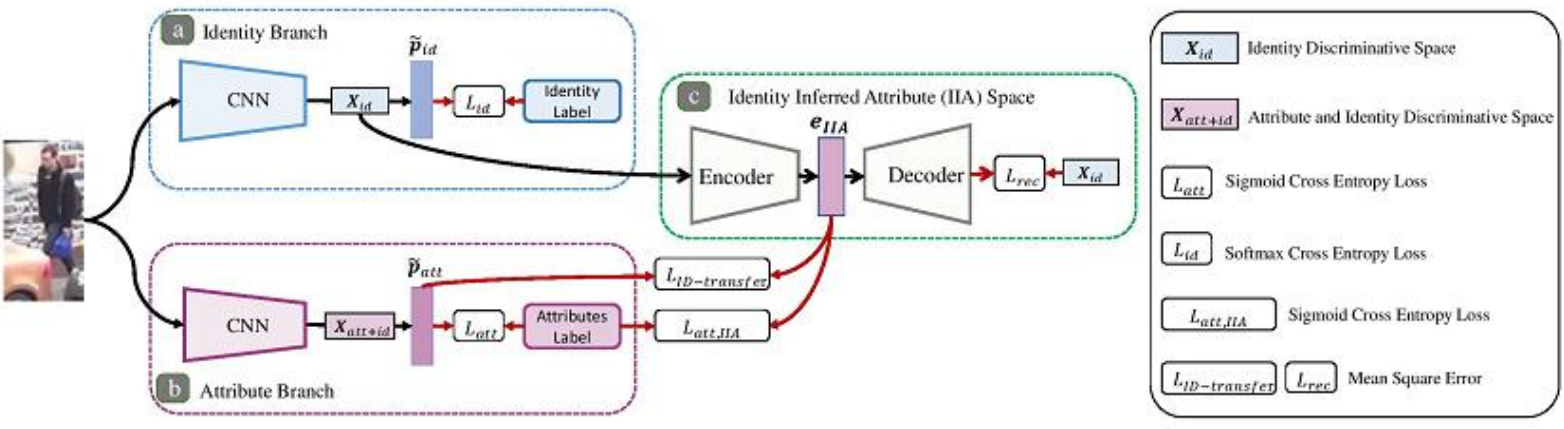
In terms of challenges, she says that because this work uses surveillance video, it's not like most computer vision works that use well annotated or well-structured annotation. **This is unsupervised**. Also, because it is surveillance video, it has poor image quality, the background is normally blurred, and sometimes there are different view conditions because of different camera locations. She adds that **person re-identification is more challenging than facial recognition**.

Jingya explains more about the computer vison methods used: "*The first one that we introduced was progressive knowledge fusion mechanism by encoder-decoder intermediate space. Second, we proposed a normal domain adaptation method, but it should be the consistency scheme, and because of this we can transfer the knowledge in the attribute space. The reason we transfer the knowledge in the attribute space is because normally for the human re-identification, the ID labels from different domains, different data sets, they are independent – they don't have overlaps – but in our study, we want to transfer the knowledge. This is an open-set recognition problem. In our work, we want to introduce attribute space, because it's more uniform space for transferring the knowledge, because they share the most common description.*"

The next step is to deploy a better domain adaptation model.

The next step is to deploy **a better domain adaptation model**. Being this an open-set domain adaptation problem, it is very challenging in the re-identification setting, and Jingya would like to explore more on that aspect.

To discover more about Jingya's work, come along to her poster [P16] today (Tuesday) at 12:30-2:50 in Halls C-E.



"…real-world applications could be in video surveillance, or in human-robot interaction, to understand what is going on just in front of the robot e.g. what is the person doing in real time?"

Fabien Baradel is a PhD student at INSA Lyon in France. He speaks to us about his poster presentation today. The co-author of the paper are Christian Wolf, an associate professor at INSA-Lyon and LIRS/CITI, Julien Mille, associate professor at INSA Centre Val de Loire and LI Tours, and Graham Taylor, associate professor at University of Guelph. The research project is funded by the ANR/NSERC DeepVision.

Fabien says his work is about **human activity recognition**, having a model which is able to infer what is going on in a video using only the RGB data. Its real-world application could be in **video surveillance**, or **in human-robot interaction**, to understand what is going on just in front of the robot e.g. what is the person doing in real time?

In terms of human activity understanding, he says the real problem right now is that they are dealing with **fine-grained understanding**. There are some really local parts which are important and discriminative for the task, and right now, all the video understanding problems are more about understanding what's going on in the context. This work is more about understanding a local part of the video.

He adds that most of the related work on this subject is more about using pose data information, so skeleton data which is given by the software, but in their case, they wanted to have a model which is able to run using only RGB data for understanding some fine-grained action done by humans.

"We wanted our method to be fully differentiable…"

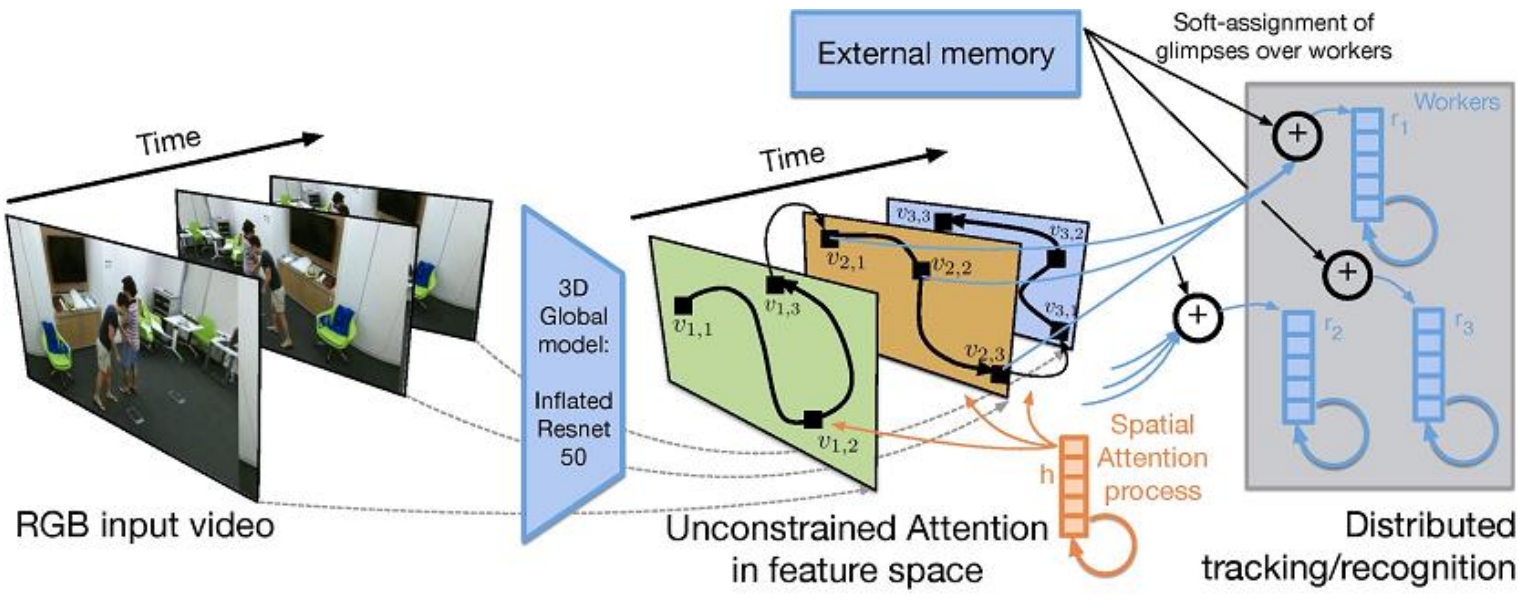
Fabien explains further: "*We wanted our method to be fully differentiable, to use backpropagation in order to learn our ways, so we have used a spatial transformer to do some cropping in the video in a differentiable way. This extracts a patch in the video, because we believe that extracting local information is important for the whole task.*"

He tells us that they have been focusing on trained video and that is already a difficult enough task, but in terms of next steps, the biggest challenge – and one which would have a lot of applications in the real world – would be to do an extension of this work on untrained video, so understanding when there is no action happening in the video and when the action starts and ends.

If you want more information about Fabien's work you can visit the project page and there is also a github repository where they will be releasing the code this week and you will be able to train and evaluate their work.

You can also learn more and ask Fabien any questions about this work by coming along to his poster [H2] today at 10:10-12:30 in Halls C-E.

"…in terms of next steps, the biggest challenge – and one which would have a lot of applications in the real world – would be to do an extension of this work on untrained video."



Nasrin Mostafazadeh

Nasrin Mostafazadeh is a senior AI research scientist at Elemental Cognition where she works on the next generation of AI systems that not only comprehend stories, but also explain themselves.

I am a senior research scientist at Elemental Cognition, which is this fundamental AI research startup that is focused on pushing the boundaries of AI forward through to AI systems that can explain themselves, collaborate with humans via dialogue, and machine reading to have a better shared understanding of the world.

You are here at CVPR to give a talk and just for one day. What is your relationship with computer vision?

I would say that I have a complicated relationship with computer vision.

We love complicated relationships!

[*laughs*] I'm joking.

[*laughs*] I'm not! [*both laugh*]

I got exposed to working on vision and language when I was at Microsoft Research for about a year. There they had an ongoing research project on a line of research which was mine which was narrative structure understanding and story understanding, which they wanted that to be multi-modal. Hence, through that, I got involved with vision. Before that, I had never done any vision and language work. That was the start. After that, I built up on top of that line of work. I kept working on different vision and language projects, which involved deeper language understanding and common sense reasoning, which is the field of AI that is really close to my heart. I'm just super passionate about doing something on common sense reasoning. All the vision and language work that I've done has been trying to push that forward. That's the complicated relationship. I'm not part of the vision community, I would say. I'm a part of the vision and language community. I was very happy to come to CVPR after all and become a part of the community. So that's the complicated relationship. [*laughs*]

You mentioned in your talk that ten years ago there were concepts that were difficult to explain to the community, and today it's much easier. Can you tell our readers what that was and why you think that changed?

So actually I made a point that the change happened and the kind of revolutionary effect that deep learning had on the community happened in other fields. I would characterize different AI challenges that we have in two categories. One being pattern recognition. The other being perception and reasoning. Deep natural language kind of falls beyond pattern recognition and into a task that requires reasoning and common sense. Anything like that, which deep natural language understanding is one of them, requires

lots and lots of knowledge and has not been really revolutionized in the past decade, I would say. Whereas a lot of pattern recognition such as image processing and speech recognition have been revolutionized, which is the example that I was making in my talk.

What made people change their approach? The deep learning revolution or something else?

What has happened in the vision community is the abundance of data which has enabled existing algorithms which have been around for decades and decades such as deep learning to work actually. On top of that, we have much more computing power such as a lot of GPU's which many big companies have access to these days. They have given rise to the so-called deep learning revolution, which has definitely made a huge change in pattern recognition problems.

Regardless of your excellent accent, I understand that you are not American-born. Where do you come from?

Oh, wow! My accent is excellent then? [*laughs*]

I think so!

Oh, thank you! I'm Iranian. I came to the US five and a half years ago. I am Iranian. I was born and raised in Iran. I went to school in Iran.

When did you discover that you would like to study language?

Ah well, the story goes all the way back to high school. Actually, I started working on natural language processing and natural language understanding in high school. The story that I told about how I got into natural language understanding from robotics was in high school. I started working on RoboCup competitions when I was like 15 or so. Then through that classical natural language example that I just told you, I really was excited to work on a problem that people called AI-complete. I can tell you the story of how I got exposed to natural language.

Please do!

It's just a bit long.

[*laughs*] We like long stories!

Here's the story. We were me and a couple of my amazing friends, we had this robotics team, as I said. Because of that team, we wanted to compete in international RoboCup competitions. It meant a lot of hard work. For about more than a year, maybe a year and a half, we just pushed on that agenda. We had a team, and we wanted to win. Through that, we didn't go to any of our classes in school. We had this amazing school that let us do this. The end of that part is that it was successful. We became actually the second in the world,

which was a dream at the time. We were kids coming second in the world.

Could you believe it?

We couldn't believe it! As I said, we skipped our classes. The year after, when I came back, we had this physical education teacher. She gave me a hard time. She said that I won't let you pass this course because you haven't shown up for about a year. She handed me this stack of papers for translation. They were English documents that she just wanted to be translated into Farsi. I was the manual labor that she had. She said: "*Unless you translate these documents, I won't pass you!*" Anyways, I take these documents home. Days pass, and I think "*This is too much!*" Then I was like, I've spent the past few years of my life just sitting in front of a computer and programming. What if there was a program that could do this for me? Then I looked up how to automatically translate English to Farsi, something like that. Then I remember the Wikipedia page was maybe the third hit with machine translation. That's how I started knowing that natural language processing was a thing. I didn't know that the field existed. Through that, I got exposed to natural language processing. I got interested in it.

Then it turns out that there is this task that is incredibly challenging for a system, but is super easy for a four- or five-year-old kid. I just chose to work on that moving forward. That's my story of how I got into it.

How did it develop over time?

It's kind of funny. I was one of these people who stuck with what I was doing and didn't let go. There are people who have done a minor in psychology or a major in philosophy and then it turned into computer science. I've been doing this forever. I've been doing this for many years.

Did you ever have any second thoughts?

I have been lucky, I would say, that I knew what I wanted to do. At the same time, I know there's a downside, right? I've never dived deep into history or geography or other topics that I know I don't like, but maybe I should have pursued. [*laughs*]

But you don't like them?

Maybe, but I've passed generic breadth courses.

So what can you tell to all of the PhD students who are having second thoughts about what they are learning?

There are two things in life, right? I would say, you either pursue something that you are great at or you pursue something that you are super passionate about. You're extremely lucky if those two collide.

And if they're useful, that's even better!

That's another thing. You think, you know, at this point you want to maximize your chances of getting a job in computer science. [*both laugh*]

The truth is, there should also be a demand,

right? You're the luckiest if there is also a demand for what you are doing. I would say that it's about balancing. Maybe it's an art of how to find something that you enjoy and, at the same time, you are good at. Sometimes you have to pay the price of doing something that you love so much that you can't let go. I think if you have second thoughts, you should really try to talk to many people in that particular doubtful area that you are at. I found that the most useful thing whenever you are going through something, you are not alone. So many people have experienced that. Just talk to people, and collect your data points. Go about making a meaningful and significant decision by having prior knowledge. It's okay to doubt. It's absolutely fine to not to know what you want to do. Everyone around me has been like that. It's just that I have been super lucky that I knew what I wanted to do from early on.

Did you see other students working with you that had second thoughts?

Everyone… I've had so many close friends that have dropped out of their PhD. I've had friends that have dropped out of college. I've had friends who, right now, they got a PhD. They got a job, but say that they don't like computer science. This happens all the time. I do try to do my best to help people make the right decisions for themselves, but this is very personal, right? Even the idea of should I get a PhD or not in the AI world is very personal. It really depends on what you want to do, if you're super passionate about research.

It sounds like you are extremely passionate about this subject.

Yes!

What is your second biggest passion?

In life? If I wasn't doing AI research, in an alternate universe, or maybe when I somehow get to a position where I could do this, I would have been Anthony Bourdain. I'm so sorry that he's gone now.

Ah, cooking…

I always said that I would have been a chef, traveled, and worked… and explore history and culture through food and understood humanity through food.

Tell me something about Iranian people that we don't know.

In the US, people don't know much. I think through the lens of media, and this is changing a lot with a younger population of Americans, it is a totally different perspective. I think on average, not in major cities of the US, but people have this image of Iran.

I'll tell you what I know: warmth and hospitality.

That's a good characterization. I think you know this too: women in Iran, in terms of university graduates, actually there are more women than men. They have a majority. There are higher percentage of women.

We see it also in our community.

Yes, in vision, I'm sure. There are many Iranian women. That's something that

people have asked me: were you able to study the same way that men did in Iran?

Well, our magazine has interviewed many impressive women from Iran.

I think you already know too much!

Sorry! [*both laugh*] What is the thing that you regret that you have in Iran that you don't have here?

Family!

They are there?

Of course. Iranians, almost all of them, or a lot of them, get single entry visas. Then they become students in the US. Because there is no US embassy in Iran. It's a huge risk to leave the US after you've come here to obtain a new visa. People end up staying here for long, long times without seeing their families. Also, getting tourist visas for families is very hard, if not impossible. I was lucky. I got a multiple entry for two years. I could go back to Iran twice, but I was a huge exception. Now, I've been here for four and a half years without being able to present my work at different conferences when they were outside of the US or to see my family. There are many other Iranians who have had it much worse than I have had. They have been here for eight years, nine years without seeing their families. It's just cruel, right?

It's super hard. People keep up with it because they have to, but it's really, if you think about it, it's just terrible. It's really hard.

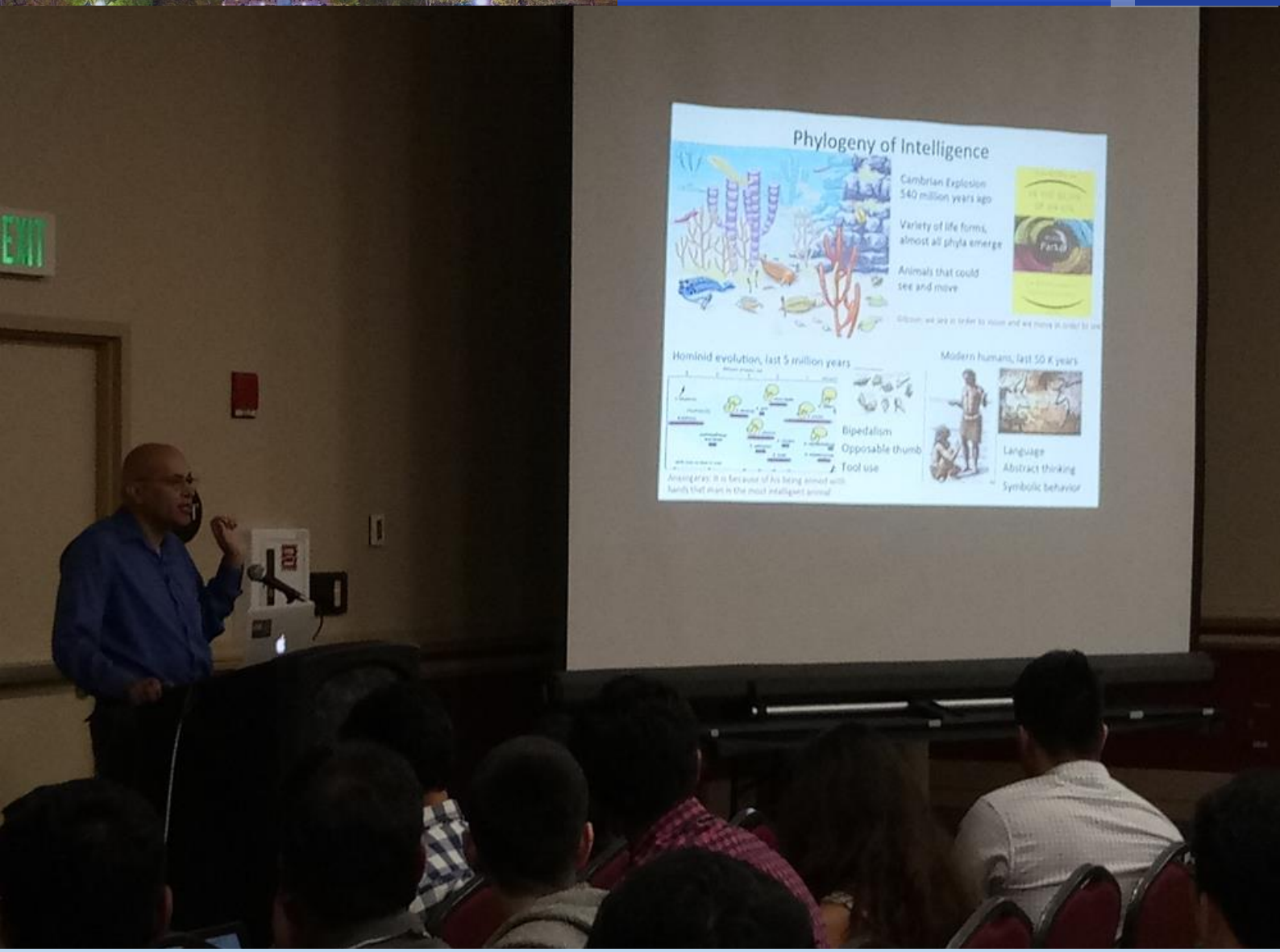
"It's just a blip..."

What about ups and downs?

I'll tell you this story. Once I read about a financial recession, decades ago. Many people really stressed about their job prospects. They thought that everyone would go down, and they would lose their houses. Now, you zoom out the graphs of the economic growth throughout the years, and look at that point in time: it looks like a little blip. A tiny blip, which at the time seemed like a disaster. I would say that in life, you have so many such blips. The only thing you can do is to make sure that when you are living it, be mindful of the fact that when you zoom out, it would have been nothing. Try not to stress. It's hard when you are in the middle of something bad that is happening to you. It's so easy to lose sight of the big picture, things that you have, and how much things really matter. Just saying "*shake it off*", doesn't work. I think sometimes reminding yourself of the fact that however desperate, helpless, and sad you feel, it's just a blip in your life. That can maybe help you to get through it more easily.

How long will it take until an AI can replace me in doing this interview, with laughs and all?

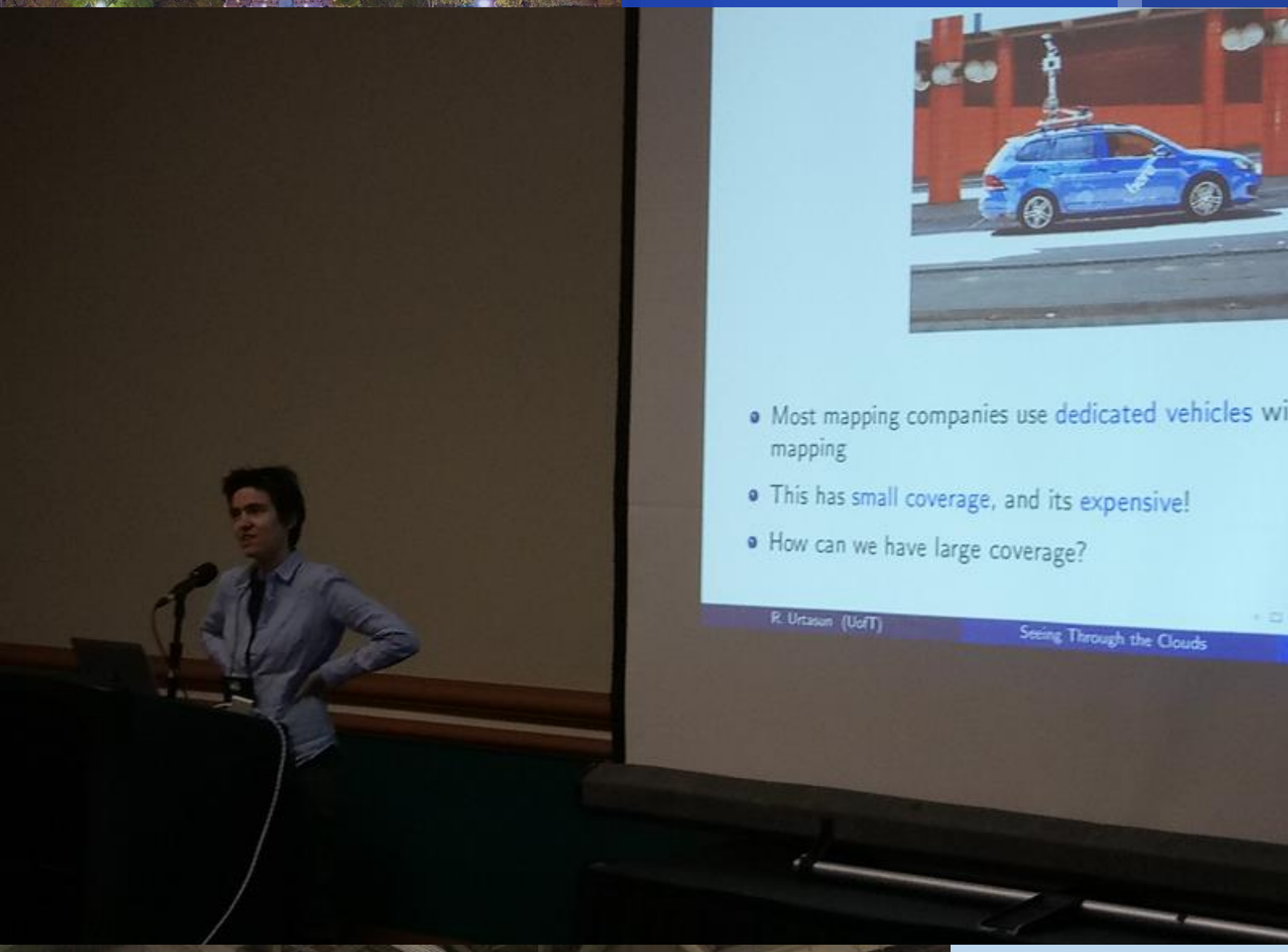
The full interview with all the breaks and pauses, in the same way that you did it? I will say 30 years from now!

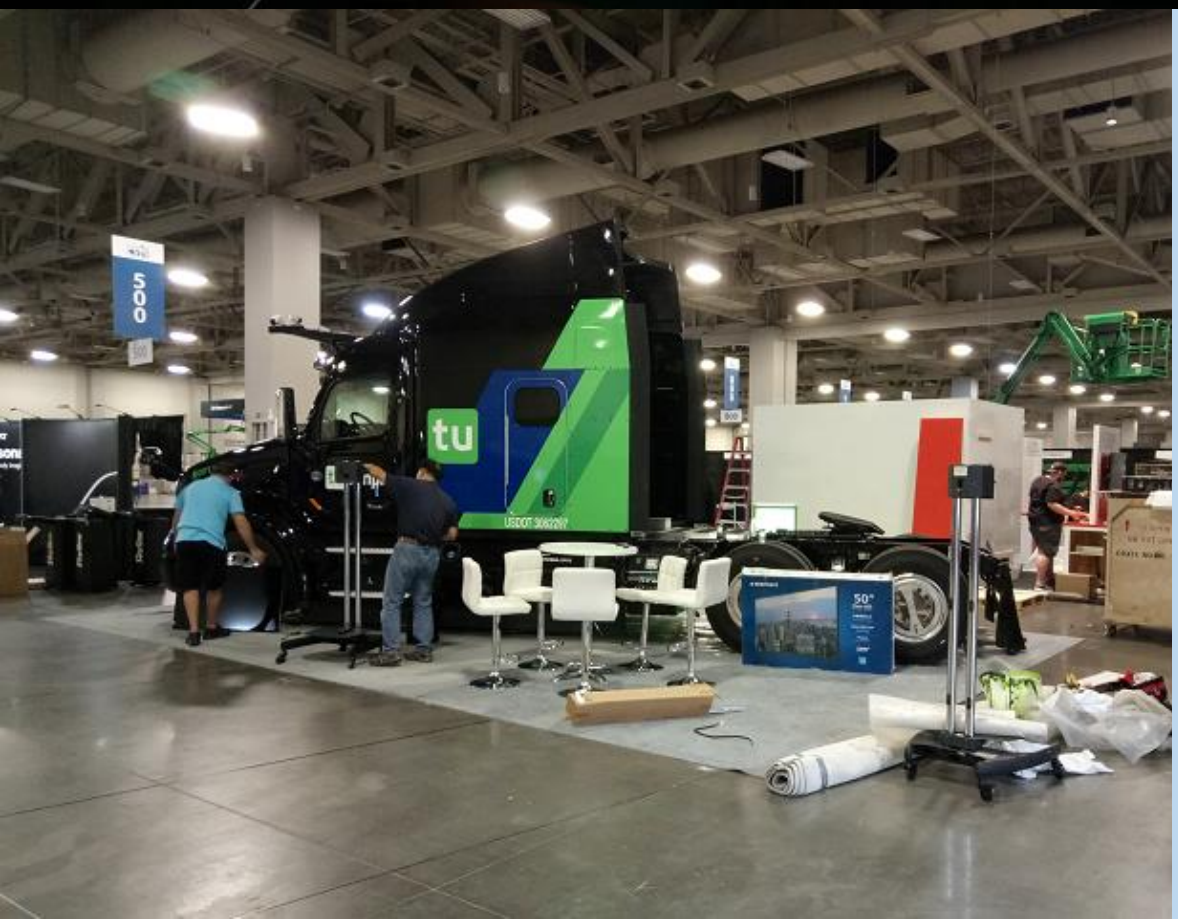


Jitendra Malik during his talk at the VQA workshop, yesterday.

Top image:
Mehdi Moradi (IBM) presenting at the Medical Computer Vision and Health Informatics workshop, yesterday. Left, co-organizer Tal Arbel.

Bottom image:
Tal Arbel introducing Alejandro Frangi. Tal was Program Chair at MICCAI 2017 and Alex will be General Chair at MICCAI 2018, later this year.





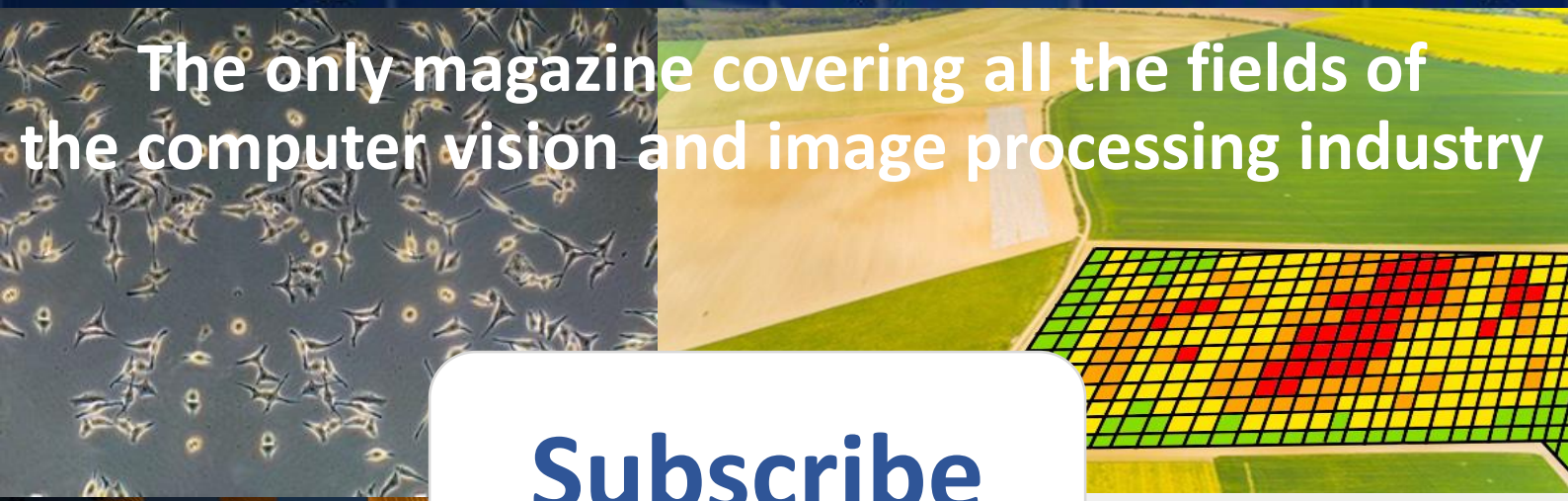
Top image: Raquel Urtasun (Uber ATG) presenting at the DeepGlobe workshop, yesterday. Read more at page WW

Bottom image: The Expo is being built. This is very probably an autonomous truck developed by TuSimple

Improve your vision with

Computer Vision News

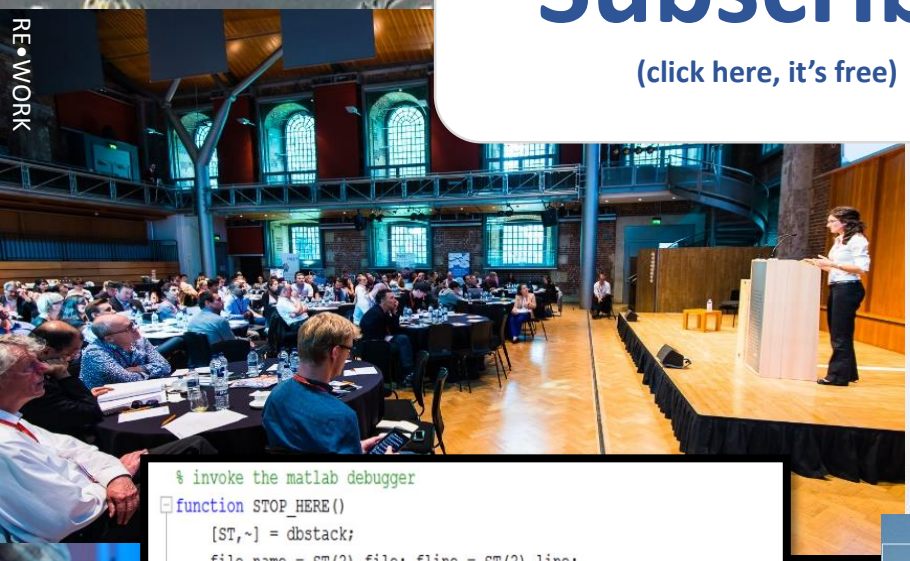
The Magazine Of The Algorithm Community

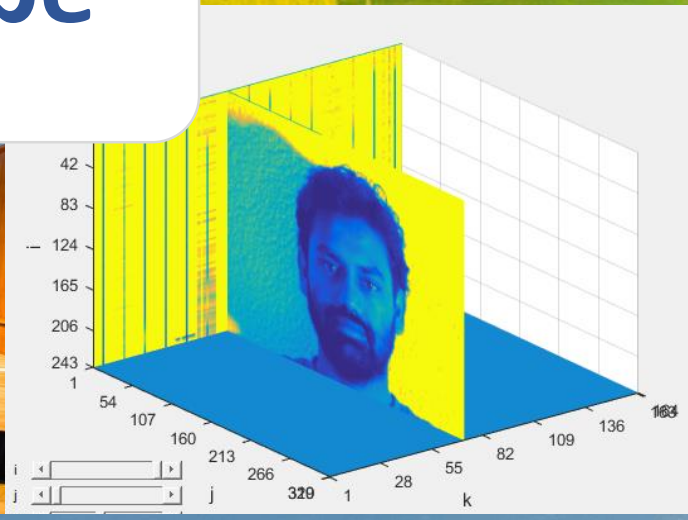


The only magazine covering all the fields of the computer vision and image processing industry

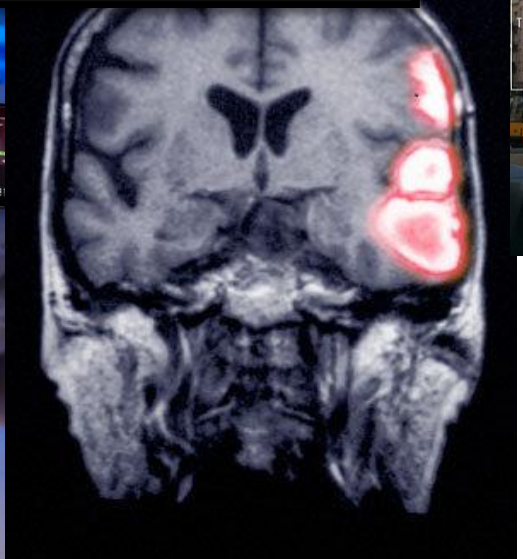
Subscribe

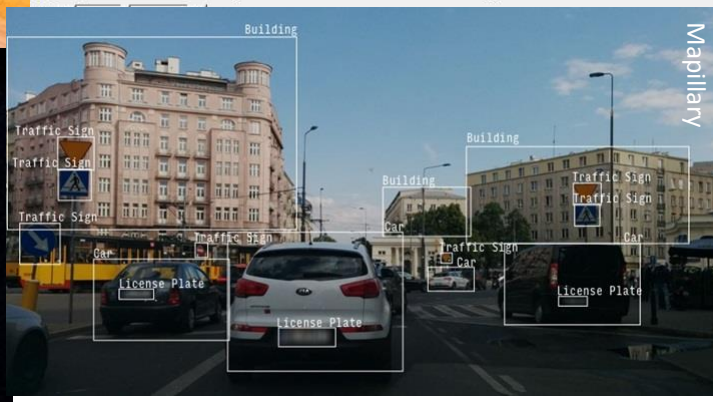
(click here, it's free)






```
% 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