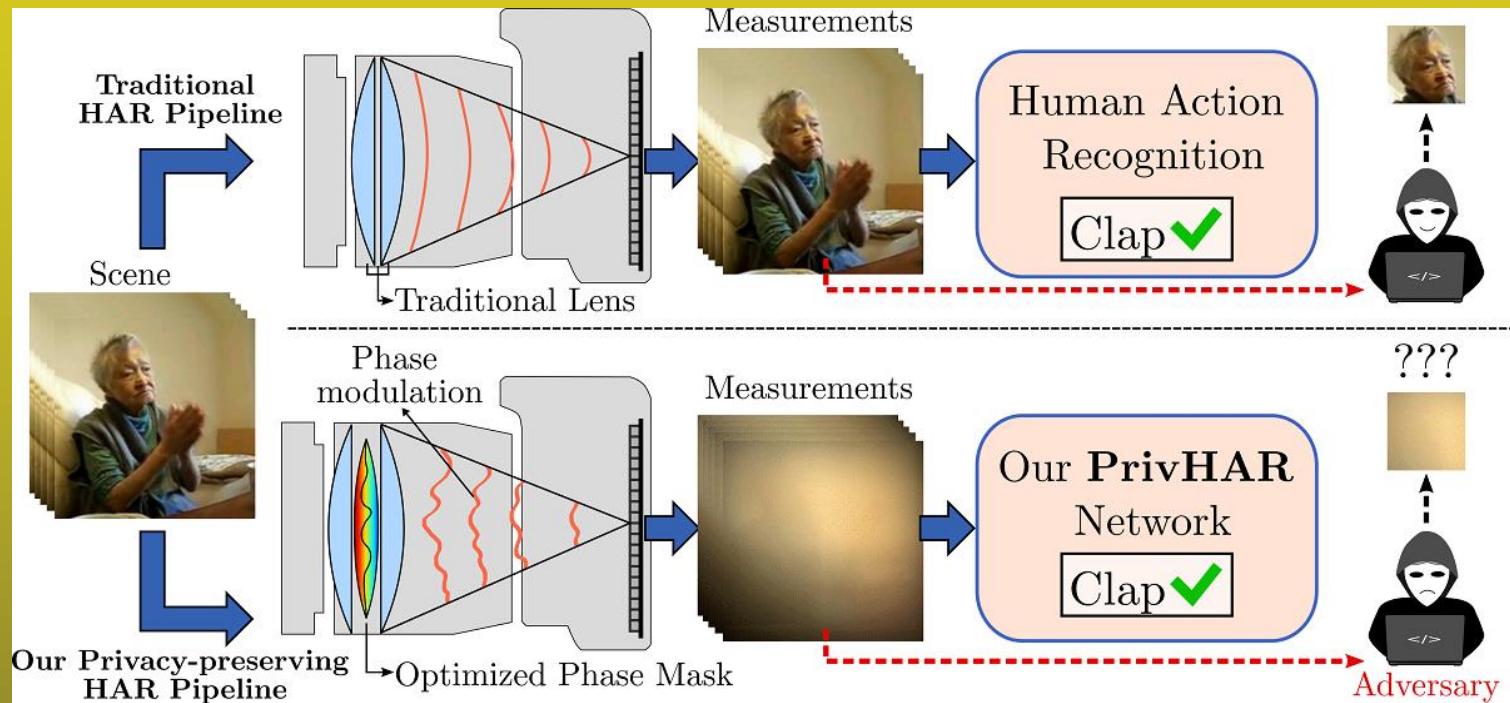




EUROPEAN CONFERENCE
ON COMPUTER VISION
TEL AVIV 2022

October 23-27, 2022



Brilliant Oral and Poster Presentations

Editorial with Program co-Chairs:
Tal Hassner and Shai Avidan

Exclusive Interview with:
Gerard Pons-Moll

Women in Computer Vision:
Angela Dai

Today's Picks by:
Silvia Bucci

My First ECCV:
Kate Yurkova

In Cooperation with

Computer Vision News

The Magazine of the Algorithm Community

A Publication By



Global Leader in Computer
Vision and Deep Learning

For today, Wednesday 26

Hello there! I'm Silvia Bucci, a PhD student at Polytechnic of Turin, Italy!

"In short, the purpose of my whole research can be summed up in one question:

Is it possible to build a general model that works consistently well with any data met at test time?

I'm trying to answer this challenging question by directing my research toward the development of transfer learning algorithms.



In particular, I focus on Domain Adaptation/Generalization and Anomaly/Unknown Detection techniques generally based on Self-Supervised Learning.



"Besides computer vision, one of my biggest passions is Astronomy.

I'm deeply curious and fascinated about the physics that governs astronomical phenomena such as space-time theory.

In the coming future, I would definitely love to apply my knowledge also in this exciting research field!"

You are super welcome to come take a look at our new work #5215 poster 3.B.39 - tomorrow (Thursday)!"

Silvia's picks of the day (Wednesday - only posters):

- 2.A.8 Self-Supervised Sparse Representation for Video Anomaly Detection
- 2.A.13 A Contrastive Objective for Learning Disentangled Representations
- 2.A.75 A Sketch Is Worth a Thousand Words: Image Retrieval with Text and Sketch
- 2.B.11 BayesCap: Bayesian Identity Cap for Calibrated Uncertainty in Frozen Neural Networks
- 2.B.98 Improving Generalization in Federated Learning by Seeking Flat Minima

Good morning ECCV 2022!

ECCV and **RSIP Vision** have once again partnered to bring this magazine to the community, showcasing some of the brilliant work being presented here in **Tel Aviv**. We saw some fantastic workshops and tutorials earlier in the week and a day of fascinating presentations and posters yesterday, with much more to come!

We are thrilled to be hosting a top-tier conference of this magnitude and importance in Israel for the first time. We made a non-trivial decision only to allow **in-person talks** this year, and it has been exhilarating to see everybody back together again. We thank **Ortra** for doing a great job and help people attending by solving 230 visa-related issues.

Out of 5,800 valid submissions, more than 1,600 papers were accepted, with 160 as orals. It was a monumental logistical effort to manage an army of **4,700 reviewers and almost 300 area chairs**, but everyone did a phenomenal job navigating all the inevitable challenges.

For many young students, **ECCV 2022** is their first major in-person conference. We always advise students to attend independently rather than with the friends they see all year round. Figure out the most suitable papers to your taste, and don't be afraid to engage with people. That's the best part of the conference. New lifelong friends may be here waiting to talk to you!

My (Tal) first conference was ECCV 2002 in Copenhagen, where I met a fellow Israeli student. Fast forward 20 years, and she is one of my closest personal friends. That person is **Lihi Zelnik-Manor**, one of our General Chairs. We came full circle as organizers of the same event, this time happening blocks away from the home I left to fly to Denmark all those years ago. Fast forward to 2042, and this could be your story!

Finally, we want to give a shout-out to our fellow Program Chairs, **Giovanni Maria Farinella** and **Gabriel Brostow**. It's been a group effort, but we all agree that without Gabriel in particular, we would not have been able to celebrate ECCV this year! We hope you all enjoy the conference as much as we've enjoyed putting it together for you.

We wish you a very successful ECCV! Please share the magazine with your friends and colleagues [and subscribe for free!](#)

Shai Avidan

Professor, **Tel-Aviv University**

Program co-Chair, **ECCV 2022**

Tal Hassner

Applied Research Lead, **Meta AI**

Program co-Chair, **ECCV 2022**

Ralph Anzarouth

Marketing Manager, **RSIP Vision**

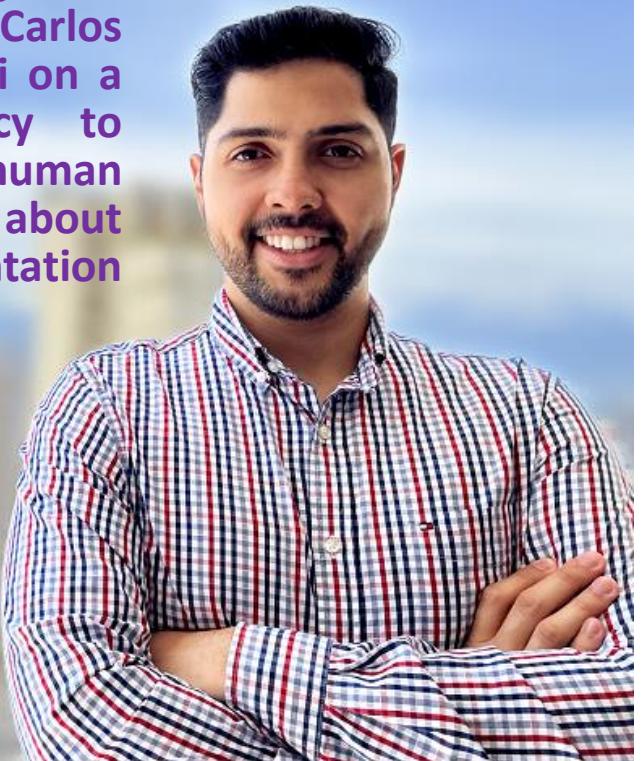
Editor, **Computer Vision News**



PrivHAR: Recognizing Human Actions from Privacy-Preserving Lens

Carlos Hinojosa just completed his PhD at the Industrial University of Santander in Colombia under the supervision of Henry Arguello.

In his final year, he had an internship at the Stanford Vision and Learning Lab, where he worked with Juan Carlos Niebles, Fei-Fei Li, and Ehsan Adeli on a project about preserving privacy to perform computer vision tasks in human action recognition. He speaks to us about that work ahead of his oral presentation this afternoon.

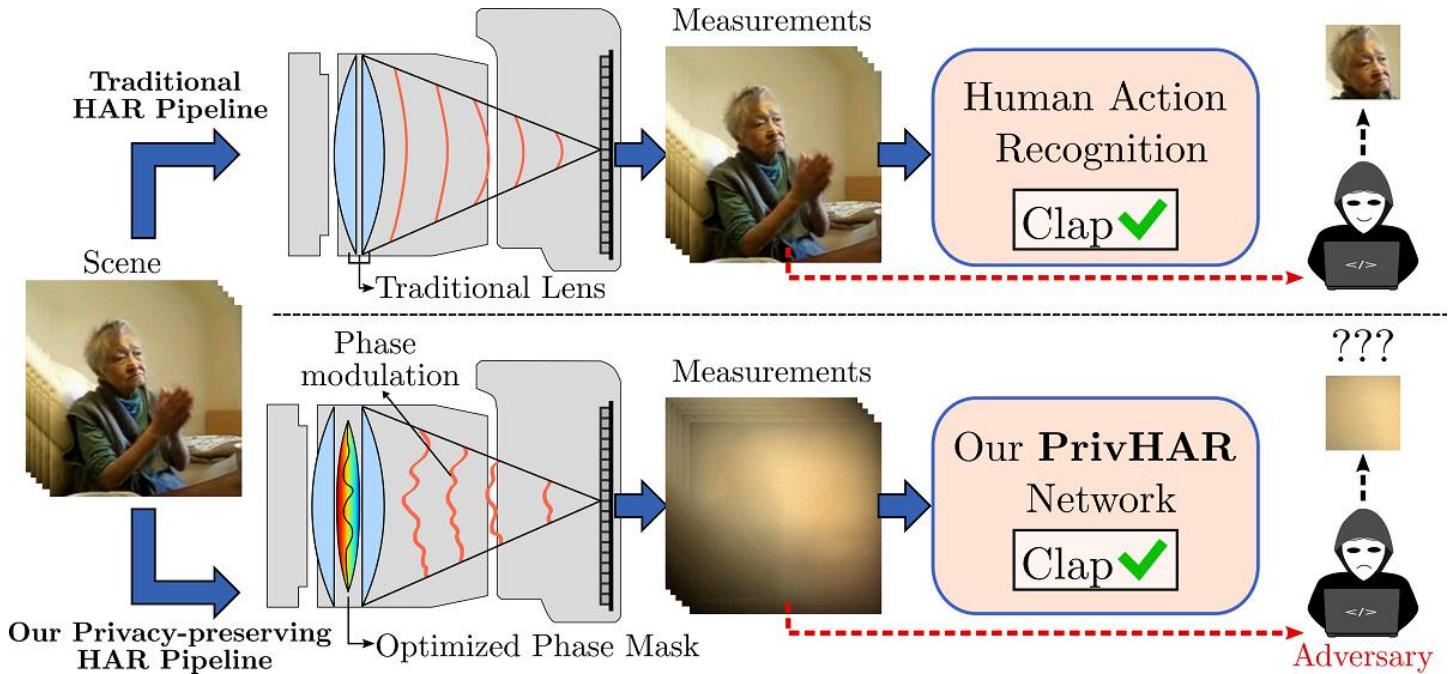


Digital cameras are fundamental to a range of intelligent systems that recognize relevant events and assist us in our activities, with multiple applications in healthcare and beyond. However, their ever-improving ability to imitate the human vision system and produce the highest-quality images has raised **concerns about privacy and security**.

Previous works have focused on software-level processing of videos after they have been filmed to resolve these concerns, but this can leave original videos unprotected. Carlos wants to address this problem within the camera hardware itself. He proposes a **new parameterized lens design that adds distortions optimized to preserve privacy** while retaining features that allow computers to perform human action recognition tasks.

"The main idea is to design the lens jointly with a human action recognition network and an adversarial network that performs attacks

to this lens," he explains. "One of the biggest challenges was optimizing the network because our approach started with models unconcerned about privacy. We began pre-training models on computer vision tasks, and as we fine-tuned, there was a trade-off between privacy and task performance."



The team tried two human action recognition networks: **C3D** and **Rubiksnet**. With C3D, it was difficult to achieve this trade-off, but Rubiksnet showed better results with modifications, **including temporal similarity matrices** to preserve the temporal information that could be lost when the input video is distorted.

There are several potential applications for this work. In hospitals, where cameras perform vital computer vision tasks, this model could help preserve patients' privacy, with the added benefit of enabling the collection of anonymized patient data that could be used for further research. It could also be used at home when a family may wish to monitor an elderly relative's activity to know if they have fallen, for example, without intruding on their privacy.

"We had a previous paper that designed a camera to preserve privacy for human pose estimation," Carlos adds.

"That could also be implemented in surgical rooms to monitor the movement of patients and doctors."

This work was conceived when the team met virtually during the height of the pandemic to discuss designing a privacy-preserving system. It's been a true collaboration, with Henry Arguello specializing in optics and

Juan Carlos Niebles specializing in computer vision and action recognition.

"My supervisor and I gave the idea of encoding the scene, and the idea of including all the temporal information and all the action recognition was given by Prof. Niebles," he reveals. *"I put all that together, and it works very well."*

The novel approach to **adding aberrations to the lens while optimizing the network** and the fact that the team built a proof-of-concept optical system in the lab is what Carlos suspects earned the work a coveted oral slot at this year's conference.

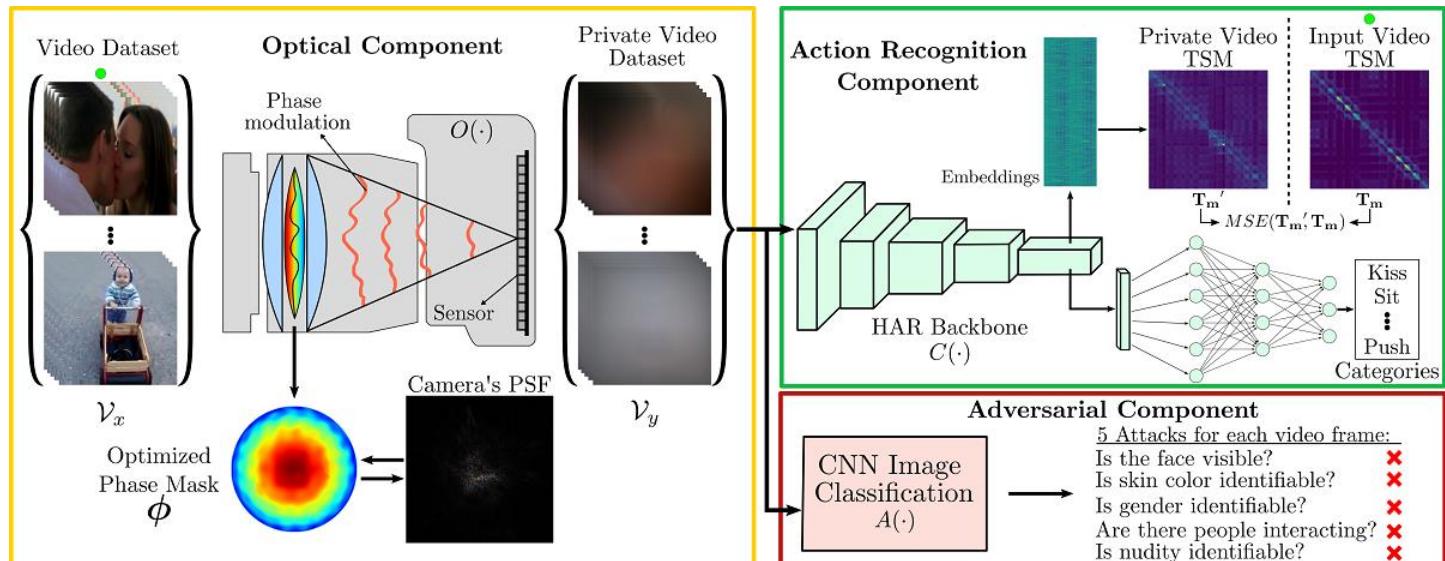
What led him to this magazine is another story. Last year, at ICCV 2021, a young gentleman from Colombia asked the organizers: How do I get featured in the ICCV Daily?

We told him it was difficult. You'll need at least a paper accepted at the conference. Even better if it's an oral. The rest, as they say, is history.

Does that young man have any advice for other scholars hoping to follow in his footsteps with an oral presentation at a big conference?

"Validation of your proposed method is very important," Carlos asserts. *"I'd recommend you perform several validations of your idea. Include all of these in the paper and use the space for supplementary material. I can say this because I had a paper rejected for not performing enough validation! Also, remember your idea should be very new, interesting, and different to be accepted as an oral."*

Carlos works in the **High Dimensional Signal Processing (HDSP) Research Group** at the **Industrial University of Santander**.



"We are a small group, and I think I'm the only guy working on computer vision, but without everyone's help, this work wouldn't have been possible," he tells us. "I'm not an expert in optics, for example, but I've been able to implement this with the group's help, so I would like to thank them."

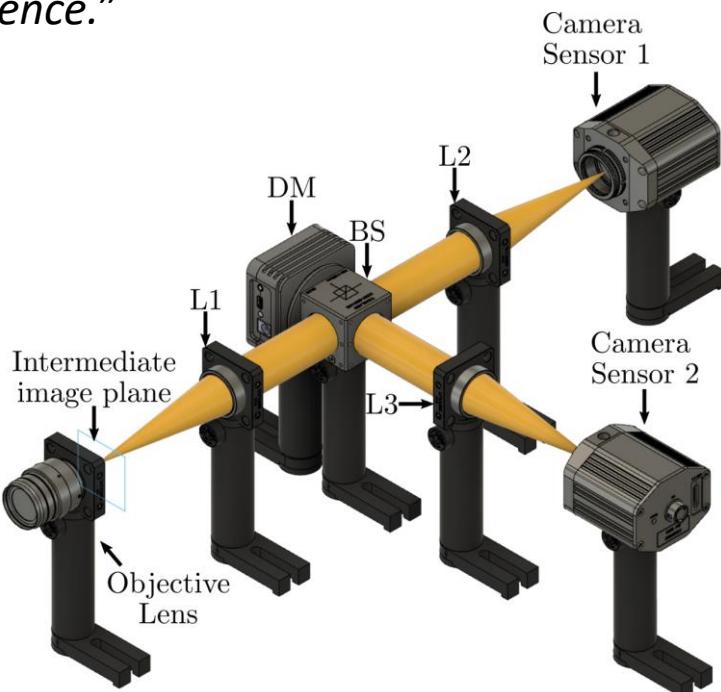
Looking to the future, he will be starting a new adventure as a postdoc at **KAUST** in Saudi Arabia with **Bernard Ghanem** next year.

"Prof. Ghanem is amazing, and I'm very excited to start this position. I'm currently doing all the paperwork for that."

Fei-Fei Li, Juan Carlos Niebles, and Ehsan Adeli were also co-organizers of the Compositional and Multimodal Perception workshop, which took place on Sunday, gathering researchers working on activity/scene recognition, compositionality, and multimodal perception.

What is it like to work with leading scientist Fei-Fei Li?

"I had only a few meetings with her, but they were very efficient because she gives so many good ideas, and that helps you to take the work forward and finish it," he tells us. "During my internship at Stanford, I worked mainly with my advisor Prof. Niebles, who helped me a lot, as did Ehsan Adeli. I learned a lot from that whole group. It was a very good experience."



Carlos, Ehsan and Henry have two US patents in progress (final stage) around these ideas of privacy-preserving for computer vision. To learn more about Carlos's work, you are invited to visit Oral 2.B.1: Scene, Action, and Video Understanding today from 13:30-15:30 and Poster 2.B from 15:30-17:30.

From the Workshops

Computer Vision for Metaverse



Devi Parikh of Georgia Tech and Meta AI gave an exciting presentation at the Computer Vision for Metaverse workshop on Multimodal Generative AI.

She showed a delightful video that used diffusion models to visualize lyrics of the song “My favorite things” that was created by Oron Ashual on her team - watch the video below.

Devi showed state-of-the-art image, audio and video generations - including from the recent [Make-A-Video](#) system from Meta AI.



***“Nothing is
as easy
as it looks!”***



**Read the full interview
with Gerard Pons-Moll
in the next pages!**



Gerard Pons-Moll is a Professor at the University of Tübingen and a Senior Researcher at the Max Planck Institute for Informatics.

Gerard, can you tell us about your work?

We work on roughly three areas. One is on modeling the geometry and appearance of people. More recently, we've been working on capturing behavior and how we interact with the 3D world around us. There are two phases to this. The capture part is where you have observations and want to understand what's going on, like in images, point clouds, or whatever data you have. The other side is wanting a compact model you can control to generate data. These two problems are interrelated because you can build better synthesis models with a good capture model. You can compare your beliefs against the observations and iterate this loop with a good synthesis model. We also work on general 3D representation learning, 3D scenes, and 3D rendering. Everything to do with 3D excites me!

What are the applications of all this work in the real world?

There are many. Imagine you deploy a robot, and the robot needs to understand what humans are doing in the scene. At the very coarsest level, it needs to know how to avoid them. All humans have these predictive models of the world, and we need to build these into intelligent systems. That's the most basic need, but there are more concrete ones. For example, any 3D content generation application, like 3D entertainment or video games, needs a way to synthesize humans and

scenes and the 3D world in an easy and controllable way. There are compact learned mathematical models underneath all these to do this.

3D has been an important part of the best years of your career. Can you tell us what you find so fascinating about it?

Firstly, I'm not sure these are the best years of my career yet – hopefully, it will get better and better! [he laughs] We live in a 3D world, which we're trying to fit into computers. Attempting to understand the 3D world around us without modeling the third dimension doesn't seem like the right approach. If you model things in 3D, many of the invariances desired in algorithms are already built in, like rotational invariance – if you rotate an object, it doesn't change its shape or appearance. Humans and animals learn to perceive and interact in the 3D world, so replicating that in machines just makes sense.

Do you see the same passion in new teachers and students, or is it unique to you?

We've seen a growth in interest in algorithms that incorporate 3D processing. We were more of a minority a few years back, but now there are as many papers doing 3D things as there are doing pure 2D things.

What is the most precious learning that you have learned from your teachers?

I learned different things from different people. One is to really focus on important problems and to think more about those problems than you do your next paper. I always try to go to the core of a problem. What's the thing that will make the biggest difference? What's the easiest way we can make that difference? What's the simplest idea? If we publish a paper and people read it and think, oh, that's obvious, but it produces excellent results, then we are successful because these are the key things we need to discover.

Also, I've learned that helping your students think about their careers is very beneficial for them, but



indirectly, it's beneficial for you too. That's something I implement as much as I can. I try to make everyone in my team successful.

Your group has an impressive number of papers at ECCV this year. Can you tell us about one of which you are particularly proud?

I don't like to choose because I think they're all excellent, but if I have to, can I pick two?

Of course!



One work I think is really cool is where we fit models to observations – not necessarily of humans but meshes. It's called Learned Vertex Descent. Instead of doing it the classical way, where you have an objective function, take gradients in parameter space, and then find local minima, we go much more direct. For every vertex in the model, we learn to displace it towards the ground truth iteratively. It's a super simple idea that leads to good results.

Another paper I find cool is about posed neural distance fields: Pose-NDF. Neural fields have been used for representing 2D surfaces embedded in 3D. In this paper, we thought that was limiting. You could generalize and consider those neural fields to represent more general manifolds in higher dimensions. We focus on modeling the manifold of poses. The manifold is like a hypersurface, where every real pose lives in this hypersurface. As soon as you leave this hypersurface, it's the wrong pose. We use this idea to learn the manifold using neural fields so that, for example, we can take a completely wrong pose and then project it back to the manifold with this machinery. [*Just after this interview, this paper won the Best Paper Honorable Mention at ECCV 2022. Kudos!*]

What can you tell us that is special about the work of your lab?

I want the lab to have a collaborative spirit so that everyone has their own project, but we have common long-term goals. That gives us some leverage because then people can help each other and reuse code for different projects. This collaborative spirit is a strength that I cherish very much.

How did an authentic Catalan like you become a prominent computer science scientist in Germany? What is the meaning of this?

[We laugh] That's a good question! I guess I spent some years in Germany. I set roots here. Now, my son is here. Yeah, it just happened! Germany provides many resources for researchers, which was attractive.

That is something we can say to every country: if you don't want the brain drain, invest in your science!

Yes, definitely. If Barcelona had the same resources, I probably wouldn't have left because I very much like being there.

Finally, our community has been separated for a couple of years – we reunited first at CVPR, and now here we are at ECCV. Many students will be attending their first in-person conference. Do you have a message for them?

I know the community has changed recently. There are a lot of papers, many more papers than we have time to read. That puts much stress on students. I want them to enjoy the journey of doing a PhD. If they work hard and believe in their ideas, eventually, it will pay off. But we all fail. It only works out a few times out of many. All those papers that we're seeing are exceptions. It's not like you just come up with the idea that works. That doesn't happen every day. They should know that we all have struggles at work and beyond work. Right now, I have to balance taking care of my son between trying to prepare presentations and this interview. Students will have their own struggles. Nothing is as easy as it looks!



Angela Dai is W2 professor at the Technical University of Munich, where she leads the 3D AI lab.

More than 100 inspiring interviews with successful Women in Computer Vision in our archive

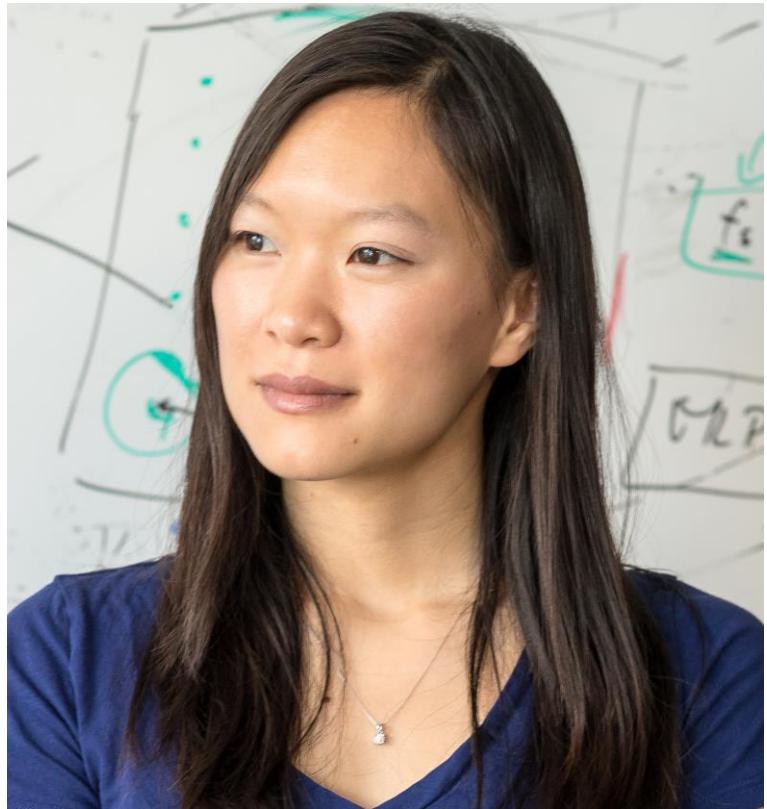
Angela, some of our readers may remember you from [our previous interview in 2017 in Hawaii](#). Can you share what happened in the last five years?

The last time I had the pleasure to speak with you, this was about the

work we had just done on ScanNet. That was really to build up a database that was available to the community that allowed people to get access to many examples of geometric reconstructions of indoor environments and also their semantics. So, of course, the stuff that I've done since then is trying to move towards how we can actually perceive real-world environments, typically indoor environments, from commodity kinds of data. So from an image, from an RGBD sensor, how can we

***“Just keep trying,
even if you’re terrified inside!”***





get out the complete geometry of that environment? How can we understand the individual objects that are observed there, even though they're not seen perfectly? The data is limited. It's imperfect. How can we sort of imbue machines with the same kind of 3D perception that we, as people, do?

Were all these developments expected then, or some of these fields or subfields came as a surprise for you?

There were definitely some developments that came as a surprise. In hindsight, they made sense, and when they came out, they were a surprise. But that's a sign of good research! Something that a lot of people were doing in this area is that they came and developed, for instance, the first

thing probably was these neural coordinate field implicit representations for representing 3D shape geometry. They were very, very effective. They made a lot of sense, and they go back and have ties to traditional geometric representations and implicit representations. That was quite powerful. It's not a perfect representation. It's still a bit of a challenge to see what's the proper way to represent a large scene and not just one object. But, this was cool. Probably everybody knows about NeRF and all of the amazing stuff that you can do with NeRF. And that's, of course, a huge development, particularly from the sort of photorealistic generation side.

I think more things have changed since the last time.

I was doing my PhD in Stanford and completed it at the end of 2018. I then moved to the Technical University of Munich, where they had these nice opportunities that you could apply for what they called junior research group positions. You can basically apply for funding for yourself and two students, which is actually quite nice. This presents a lot of opportunities to start building up this kind of research group of your own early on, prior to even becoming a professor. So that's what I did in Munich, and I

am now also a W2 professor, meaning I do have that official professor role now.

Can you share with us how that happened? And secondly, did you have any hesitation in accepting the offer?

How did it happen? I applied for the funding, and they were nice enough to provide me with the funding. [laughs] Maybe you're asking why I was looking for these kinds of positions? I initially thought I wanted to go into industry, so I did an internship at Google, and I still think industry research is super cool. You see all of this crazy stuff like Imagen and these amazing text-to-video things that come out of industry. But I found it was quite a different environment from the academic environment, and I had gotten so used to being in an academic environment. I wanted to try it out for a bit longer and figure out what the possibilities are there. Of course, the move from California to Munich was large, and I was probably over-eager [laughs].

You were born in California?

Yes.

What was the experience like to leave your native country and go to Europe?

If you want to be in academia, then you are really looking for a top institution with a lot of good

students. Universities only exist in certain locations, right? So, you need to be a little bit open about locations and where various opportunities might be. There are a lot of opportunities that I think are quite nice in Germany.

You certainly have a very strong American identity, being born there and being raised there. Do you feel also any Asian identity and have you started to feel any European identity?

It's an interesting question. I definitely felt very Chinese when I was a kid because my parents are basically first-generation immigrants to the U.S. I grew up with them speaking Chinese to me, and we went to visit China every two years. But as I got older, around my teenage years, I started to feel that this cultural difference between California and China was quite large. Many things were very confusing to me. I mean, just small, benign things like how do



you receive a gift? This is very different between the two cultures, so I was always very confused at birthday parties. [laughs]

From the outside eyes, it seems that you solve this contradiction very well and easily. Can you share some of your insights?

I think for me, it came a bit more naturally because when I was a kid, all of my friends were at least living in America, even if they didn't fully have American or Californian identities themselves. That helps tie you to a place, right? If all of your friends are there, the environment, routine, and habits. Even going to certain places for

food, the parks, movies, and other stuff for fun.

Did you ever feel treated like a non-American?

Well, here and there may be small things, but I would say I've been lucky enough that there was never anything significant, so that's definitely very nice.

The second part of my question was about Europe.

Europe? So that's actually funny because I think Europe and the U.S are much more similar in culture, but it meant that I noticed a lot of these small things that are different, much more than I would





have otherwise.

How do we receive gifts in Europe?

Much more similar to America. [laughs] That's not so much of a thing, but now that I've been there for a few years, I definitely feel more comfortable feeling like I have a home in Munich now. I guess getting used to living there, the general sentiment and attitude of people, and sort of the university environment, I definitely feel more Californian. There's definitely a small part of me that feels like, hey, this is my home in Munich!

Do you see yourself continuing your life in Europe in the midterm?

It's a definite possibility. I just want to do some cool research with some cool people, so I'll go where that is.

If I ask you where you see yourself ten years from now, do you feel more like a researcher or a teacher?

I have a very hard time predicting what will happen ten years from now. But probably, I'll still feel more like a researcher. I think the teaching part is fun in a different way, but the research part is what I think is really exciting. There's always new stuff coming, as you see, and there are always new surprises.

Do you like surprises?

I mean, sometimes you think, why didn't I think of that? [laughs] But it's always a good thing when a surprise comes, because it makes you think about new things, which is the cool part about research.

What is good about teaching?

It's nice to see other people's growth due to the university environment. That's scary at the very beginning because you feel like you have so much responsibility for what's going to happen. But that's kind of cool. It's something that you wouldn't really have too much of a chance to get in any other kind of environment. I think it's really just seeing people develop in this kind of close-contact fashion. Especially for PhD students when you're working with them all the time, close together, to see them develop over time. That's the best part.

I think it's a question that I asked [Sanja Fidler](#) five years ago, also in Hawaii, actually. I asked her, do



you feel more satisfied when you succeed or when your student succeeds?

I think they're kind of one and the same. I would feel far more satisfied if my students are succeeding now than I did in the past, even, let's say.

Angela, you are very upbeat and happy with what you are doing, and things are going your way. Can you tell our readers one of the tips about how to do it?

Actually, I'm a very afraid kind of person. Very shy, very introverted. So I think the part that was important for me was just to keep trying, right? Even though internally I'm terrified, I just keep trying to go for things. So just keep trying, even if you're terrified inside. [laughs]

Nobody notices that, so that's okay, you only know.

Yeah, maybe. [laughs] I think it is important to always be active and try out new things. I'm always a

little bit terrified inside when I do that, but I know that it's important and also that's the best part of it in the end.

You have already accomplished many things in your young career. Is there one thing that you have not accomplished yet, and what would be your next target?

The big thing that I think would be cool is to take a short cell phone video of an environment and then import it into a video game level because I like to play video games. [laughs] I think the video game sort of nature of things, the interactivity is really cool, and that's something that we don't really have remotely solved yet.

So you're that kind of millennial that likes video games?

Actually, I started playing video games in grad school, so I'm a bit late. [laughs]

Angela, tell me one thing about you that we don't know.

Besides the fact that I like to play video games, I'm trying to jog everywhere I go to a conference. So I went jogging along the beach in Tel Aviv this morning which was really nice. It's a bit warm for jogging, but it's super pretty.

I used to roller skate there. But then I broke too many bones to continue. I hope that your jogging is less brutal!

No broken bones yet! [laughs]

Let's talk about future projects... maybe one word about things that you have not done yet.

I think a lot of the stuff that we're looking at involves how we learn more efficiently. So, in 3D, we don't really have anywhere near as much data as you have with images or with text, which means that you do have to think about a more efficient learning paradigm and how we can learn with fewer labels or no labels.

Or create the data?

Or create the data! That is always an option to jumpstart a new area. It's obviously not a tractable solution to create a new data set every time you have a new problem.

Your message for the community?

I think this community is pretty awesome to be in right now. So many things are changing, and that's really exciting! Lots of new ideas and new tools.

Do you want to give one example of something that's impressing you at this moment?

The work that people have done leveraging diffusion models in order to provide supervision for 3D is pretty cool. So, for instance, DreamFusion, I think it's called, actually showed that you can get some notion of 3D structures out of image-based models. I don't think it will solve everything, but I think that it was pretty cool.

[More than 100 inspiring interviews with successful Women in Computer Vision!](#)





Kate Yurkova is a Deep Learning Research Engineer at Deci, where she works on Deci's proprietary Neural Architecture Search engine.

StretchBEV: Stretching Future Instance Prediction Spatially and Temporally



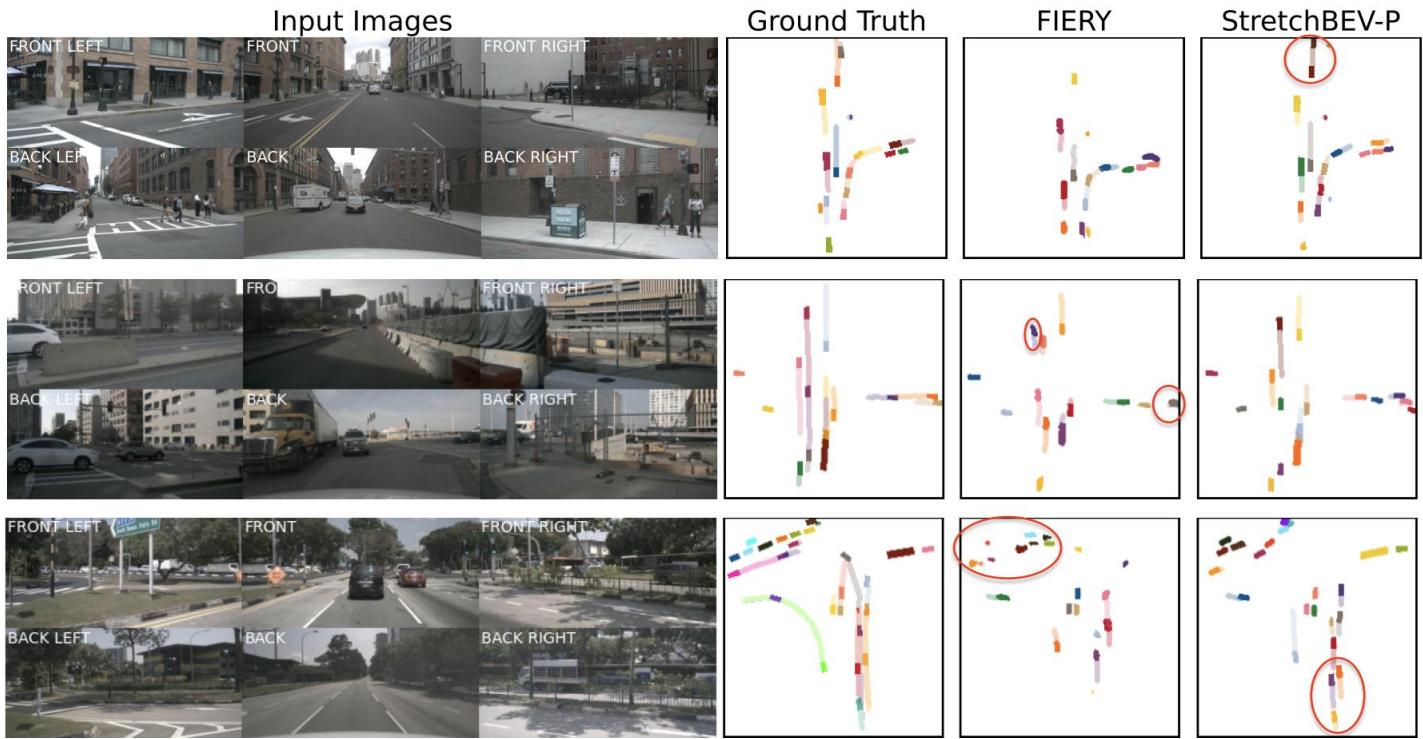
Adil Kaan Akan (*right in the photo*) has just graduated from his Master's at Koç University in Istanbul and is currently looking for PhD and internship opportunities.

He works on stochastic future prediction and video understanding under the supervision of Fatma Güney (*left*), an Assistant Professor at Koç. They speak to us ahead of their poster session this afternoon.

Future prediction is a critical planning requirement for self-driving, where it must consider the location and motion of all objects around a vehicle. Previous work has introduced the concept of **fusing information from multiple cameras into a compact bird's-eye view (BEV) representation** to perform this prediction. However, with numerous possible futures needing to be covered, the quality of predictions can degrade over time.

This work proposes a **future segmentation method in a compact BEV representation**, using time-dependent stochastic latent variables to model future uncertainty and the accuracy of predictions.

"Think about a case where you're coming into a junction, and a car in front of you can turn right, or it can go straight ahead," Kaan poses. *"You need to know both these situations for planning. If you know the possibilities, you can plan ahead and act accordingly. There are multiple possible futures, and the prediction models must cover most of them for*



self-driving. Using stochastic latent variables, we can model different modalities correctly, and our future predictions can cover several modalities rather than only one.”

The network must understand the **temporal dynamics in the video** to predict where each object is and where it will be in the next few seconds.

“Much of the work so far has been in the pixel space, but we’re using something more compact because self-driving vehicles have multiple cameras,” Fatma explains. *“We’re summarising the information from six cameras into this compact bird’s-eye view representation and trying to predict the future instances in that.”*

Alex Kendall gave a fascinating talk at the Workshop on **Uncertainty Quantification for Computer Vision** earlier this week on the Foundation Model he’s building for autonomous driving and how to make an end-to-end learning approach safe. Fatma also points to [**Yann LeCun**](#)’s recent position piece on world models as a great way to approach self-driving.

“Once you have the world model, you can generate arbitrary length futures,” she tells us. *“You can ask, given this state and this action, what’s going to happen next? With that kind of model, which can accurately predict the future, you can learn how to act based on its predictions. You don’t need a simulator. You don’t need something like **CARLA** to give you the next thing because your model can do that for you. It opens up many possibilities in terms of action.”*

Kaan tells us they are already working on an extension to the project focusing on the interaction between objects.

“By modeling the interaction between objects, we think it’ll improve the performance of each trajectory or each future prediction for the objects,” he reveals. “We’ll use intuition from trajectory prediction literature, like GNNs or attention type of connections, to model this.”

This work is Kaan’s second poster, but the first in person, having presented virtually at ICCV last year. How does it feel to have a poster accepted?

“Yeah, it’s really cool!” he smiles. “Especially having your poster accepted at a top-tier conference like ECCV – it’s great!”

Perhaps an oral at CVPR next?

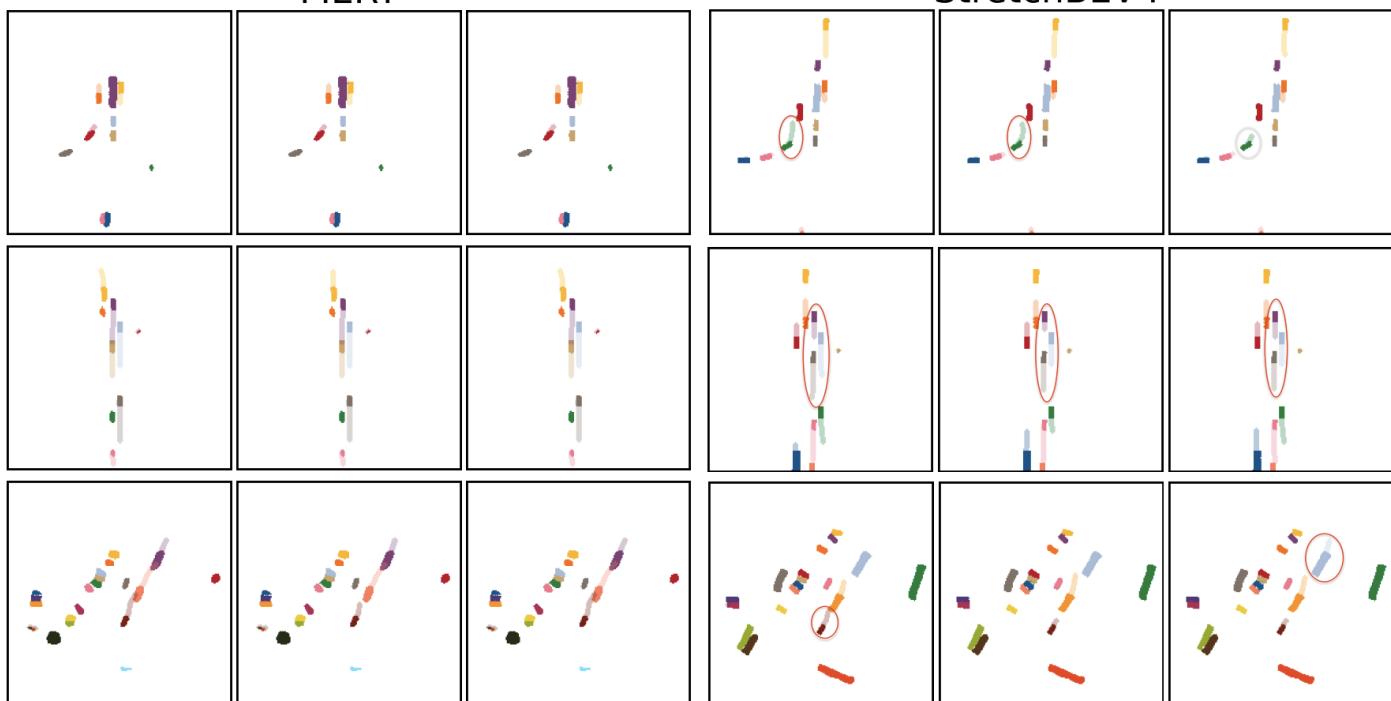
“We hope so!” he laughs.

Kaan works mainly on future prediction, but the representation changes depending on the project. He also works on *trajectory prediction* and has worked on both video prediction and bird’s-eye view prediction as well.

Previously, Fatma was a PhD student in Germany working with **Andreas Geiger**, then a postdoc at the **University of Oxford**, working with **Andrea Vedaldi** and **Andrew Zisserman**. As an Assistant Professor in Istanbul, her team focuses mainly on self-driving, and most of their

FIERY

StretchBEV-P



recent papers have been on future prediction in its many representations – pixels, bird's-eye view, and coordinates. She hopes they will have a submission for CVPR 2023 on **trajectory forecasting in the coordinate space**.

"We're also working on trying to segment and detect corner cases where an unknown object is on the road," she tells us. "On top of this, we're trying to connect it to robotics to use these world models in action. How can we learn to act based on the predictions of the world model? For that, you need action condition prediction. Given the current state and the action to perform, you predict a different future. For example, if you go straight, the future looks different from if you turn left. We're trying to predict these conditional futures based on the input action."

When real data is available, the team uses it, but when they cannot find real data and need actions in a particular form, they use a simulator like CARLA. For future prediction, they use **nuScenes**, and for trajectory prediction, **Argoverse** and the **INTERACTION** data set.





With brain drain being a hot topic in Turkey, does Fatma intend to stay put for now?

"Yes, everything looks great for now – I think I will stay for a few more years at least!" she laughs. *"I have friends, family, and amazing students here. There are lots of reasons to stay. Also, Istanbul is amazing. You should visit!"*

We ask Kaan and Fatma for one final pitch – why should our readers come to see their poster?

*"It's one of the **most accurate world models for self-driving** – actually, it's state of the art,"* Fatma asserts. *"It builds on great work from BEV and FIERY and improves the state of the art in all metrics, particularly in long horizons and further regions in the scene, which is still relevant for the car. We have an amazing poster!"*

Kaan adds:

"Our model predicts not only accurate but also diverse results, which means it can imagine different futures depending on the input."

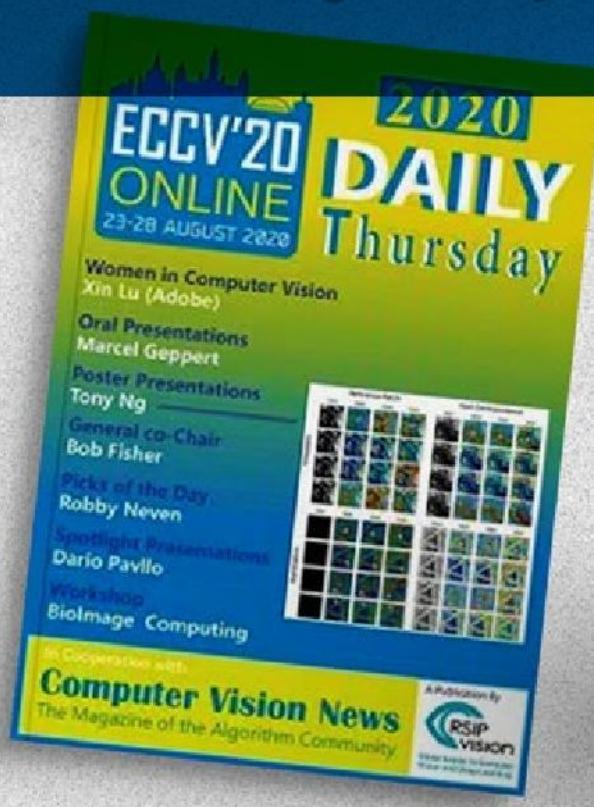
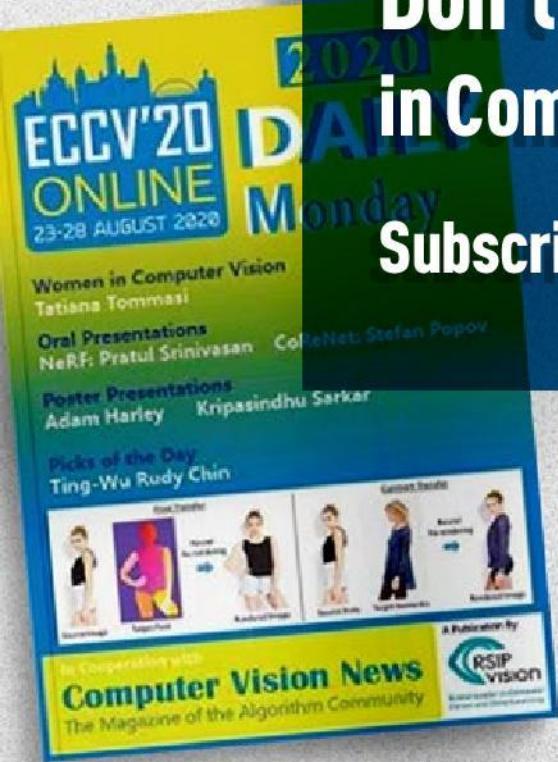
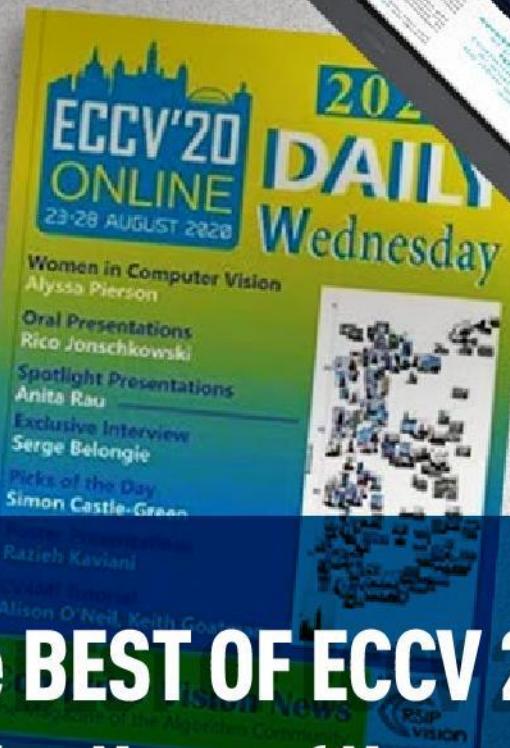
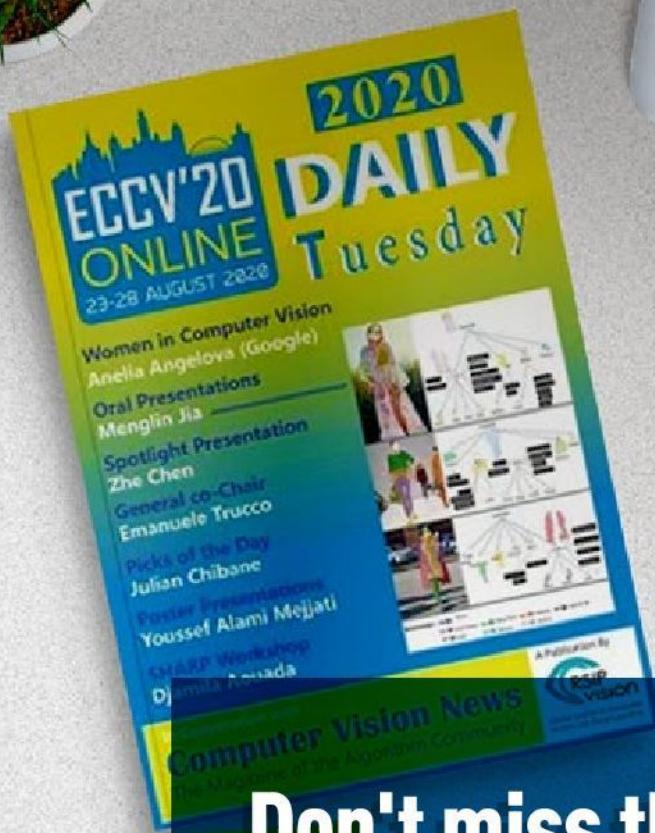
Before we finish, Fatma has a plea for the community:

"We need to increase diversity! I keep saying it, but more women, more minorities, and more internships for students, not only in Europe and the US but also in Turkey. I care about that deeply. We need diverse backgrounds."

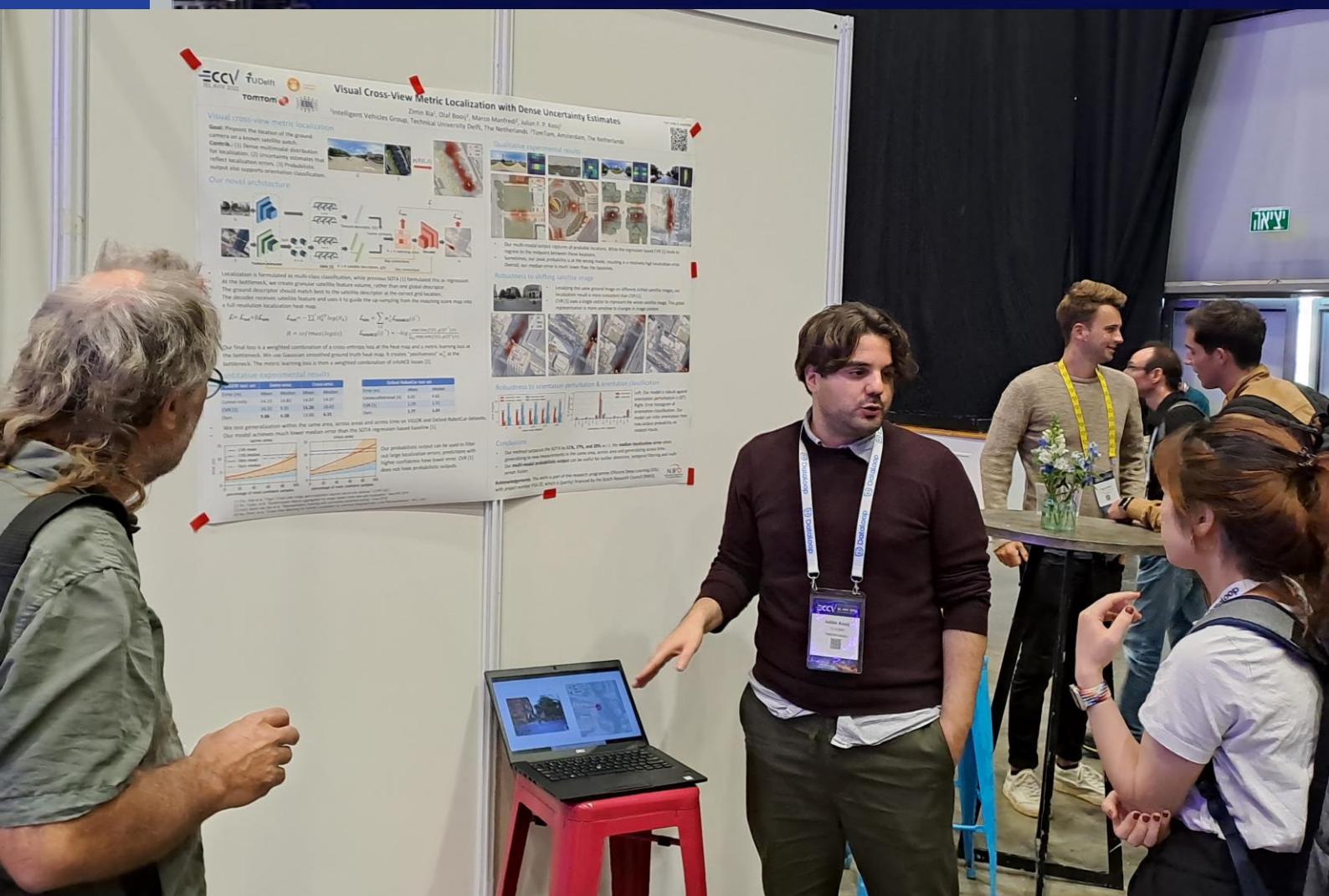
To learn more about Kaan and Fatma's work, you are invited to visit Poster 2.B today from 15:30-17:30.

**Don't miss the BEST OF ECCV 2022
in Computer Vision News of November.**

Subscribe for free and get it in your mailbox!

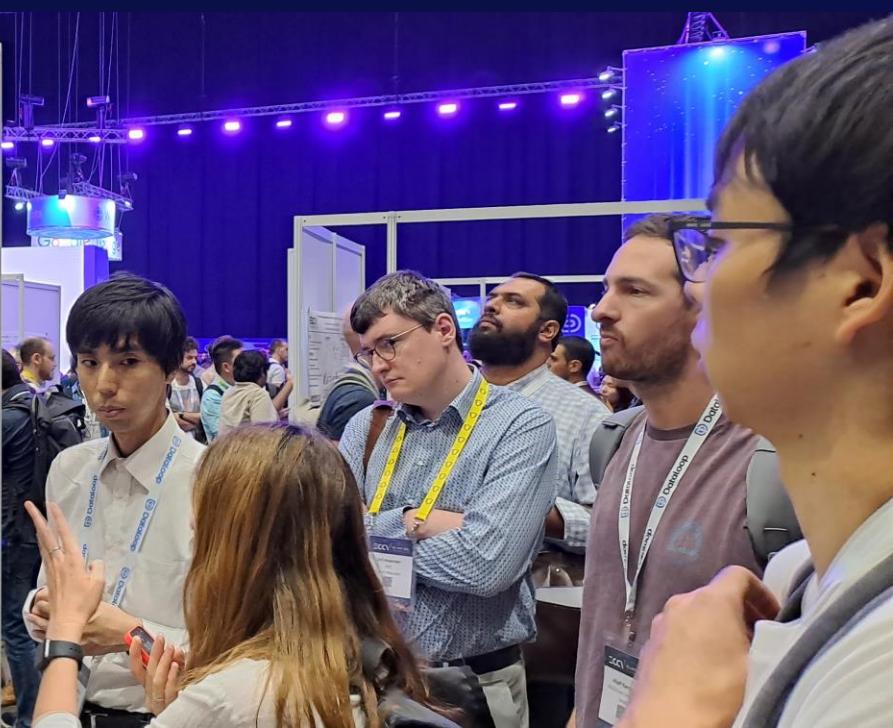


Poster Presentations



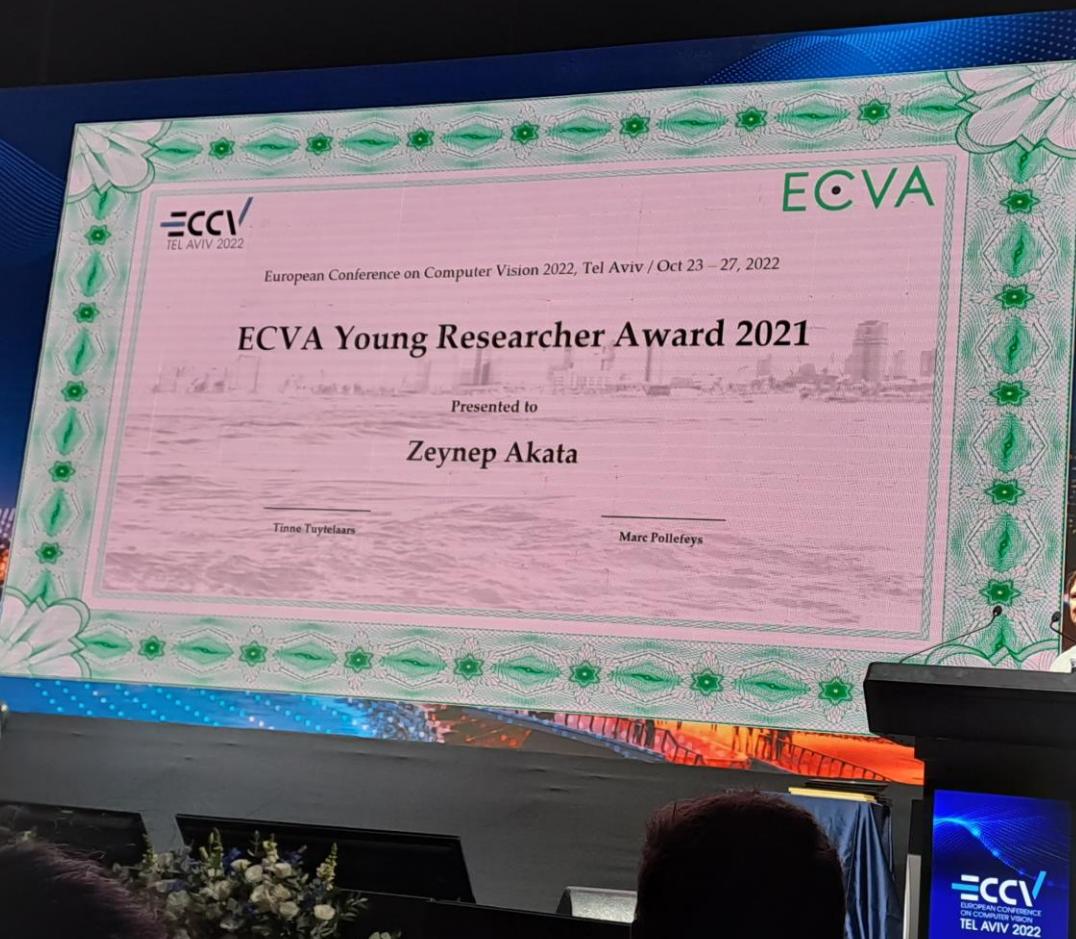
Julian Kooij (top) presents the work by his PhD student Zimin Xia on “Visual cross-view Metric Localization with Dense Uncertainty Estimates”.

Kengo Nakata (bottom) presents his paper about “Revisiting a kNN-based Image Classification System with High-capacity Storage”.





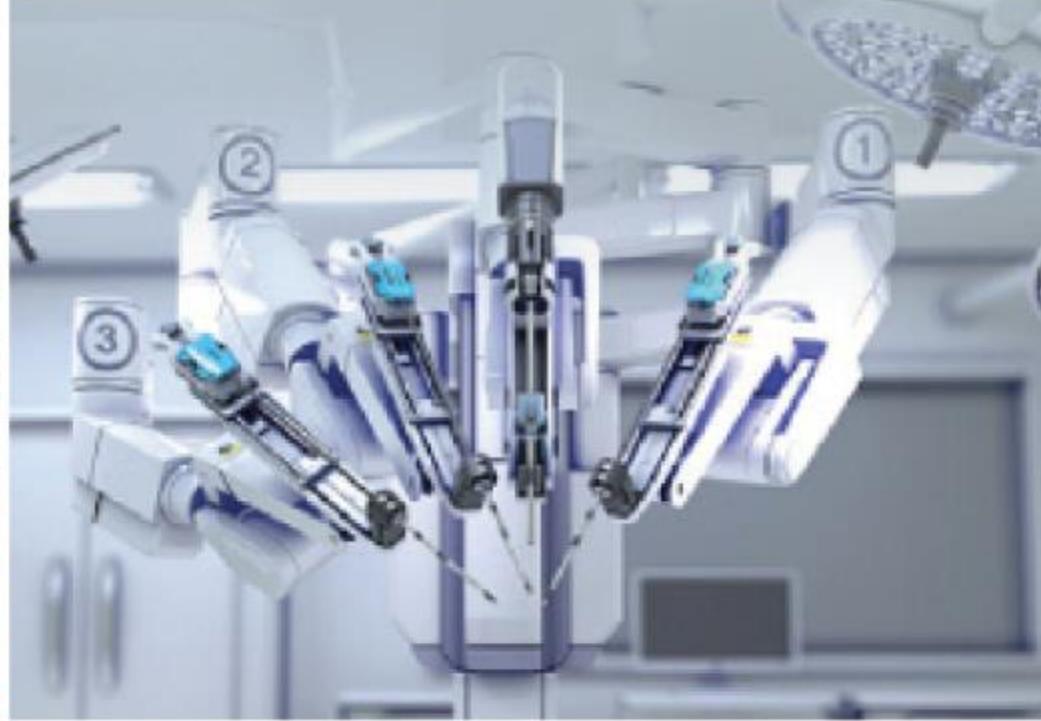
Congrats!



HAPPINESS IS... HOLDING TICKETS THAT SAY VANCOUVER!



CVPR 2023 will be held in Vancouver.
Come over. In person. It will be fun!



IMPROVE YOUR VISION WITH Computer Vision News

SUBSCRIBE

to the magazine of the
algorithm community

and get also the
new supplement
Medical Imaging News!

